SAFETY AND RELIABILITY – SAFE SOCIETIES IN A CHANGING WORLD

PROCEEDINGS OF THE 28TH INTERNATIONAL EUROPEAN SAFETY AND RELIABILITY CONFERENCE (ESREL 2018), TRONDHEIM, NORWAY, 17–21 JUNE 2018

Safety and Reliability – Safe Societies in a Changing World

Editors

Stein Haugen & Anne Barros
NTNU, Faculty of Engineering, Trondheim, Norway

Coen van Gulijk
University of Huddersfield, Faculty of Computing and Engineering, Huddersfield, UK

Trond Kongsvik
NTNU, Faculty of Economics and Management, Trondheim, Norway

Jan Erik Vinnem
NTNU, Faculty of Engineering, Trondheim, Norway

CRC Press
Taylor & Francis Group
Boca Raton London New York Leiden

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A BALKEMA BOOK

CRC Press/Balkema is an imprint of the Taylor & Francis Group, an informa business

Typeset by V Publishing Solutions Pvt Ltd., Chennai, India
Except:
'Failure Mode Effects & Criticality Analysis (FMECA) using Bayesian Dirichlet-multinomial conjugate pair' by W. Baun

'Lessons learned from an unexpected uranium accumulation event' by D.G. Harrison & A. Smith

Published by: CRC Press/Balkema
Schipholweg 107C, 2316 XC Leiden, The Netherlands
e-mail: Pub.NL@taylorandfrancis.com
www.crcpress.com – www.taylorandfrancis.com

ISBN: 978-0-8153-8682-7 (Hardback)
ISBN: 978-1-351-17466-4 (eBook)

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Table of contents

Economic analysis in risk management

Organizational factors and safety culture

Mathematical methods in reliability and safety

Resilience engineering

Risk assessment

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Preface

The 28th instalment of the European Safety and Reliability Conference contributes to a long-standing tradition of sharing and learning in safety and reliability in Europe and beyond. Academics and professionals from all over the world meet in Trondheim to share the state-of-the-art in safety and reliability and discuss collaborations and future work.

The annual European Safety and Reliability Conference (ESREL) is an international conference under the auspices of the European Safety and Reliability Association (ESRA). It is one of the largest, and most important safety and reliability events in Europe and has become a recognized conference all over the world.

ESREL aims to be an inclusive event where safety and reliability students can meet renowned professionals and partnerships are forged between participants from all parts of the globe. This inclusiveness has been a characteristic for ESREL over many years and it contributes to the continued success of ESREL and ESRA.

NTNU is a university with an international focus, with headquarters in Trondheim and additional campuses in Ålesund and Gjøvik. NTNU has a main profile in science and technology, a variety of programmes of professional study, and great academic breadth that also includes the humanities, social sciences, economics and medicine. The Department of Marine Technology (IMT) is a world leader in education, research, and innovation for engineering systems in the marine environment. Areas of research include safety and reliability of marine systems, dynamic modeling, marine engineering, transport and production for oceans and the Polar Regions.

The programme of ESREL 2018 offers a variety of arenas for discussion. In invited and plenary lectures world leading scientists explain what their contributions are and share their view for the future of safe societies in a changing world. Internationally recognized university teachers, researchers and practitioners support sharing and discussion in topic sessions and special sessions focus on topics that are of interest to special interest groups. Industry challenge sessions offer an arena to industry professionals to discuss direct industry interests in the field of safety and reliability. And finally, talented young researchers have an opportunity to share their work.

This volume contains more than 400 abstracts of the scientific and industry contributions from the ESREL conference; it is ordered according to the methodologies that form the backbone of the ESREL conference. The full papers are published as Open Access papers on the following website: @@@@@@

Stein Haugen
Anne Barros
Coen van Gulijk
Trond Kongsvik
Jan Erik Vinnem

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Acknowledgment

We would like to thank many people for their support and contributions to ESREL 2018.

We gratefully acknowledge the members of the ESREL 2018 Technical Programme Committee for their support of the scientific programme. We gratefully acknowledge the European Safety and Reliability Association Technical Committee Chairs and Co-Chairs, for volunteering their time and expertise to provide feedback as part of the contributed paper review process and for chairing the sessions at the conference. And last but not least we thank all authors and reviewers who have willingly given some of their time to ensure a high quality of papers for this conference. Every paper was reviewed by anonymous reviewers.

We would like to thank colleagues who organised special sessions of contributed papers and colleagues who organised workshops. We also thank the ESREL 2018 Plenary Speakers for offering their unique perspectives on safety and reliability at this conference.

The support of the ESREL 2018 sponsors and exhibitors is gratefully acknowledged.

Finally, we would like to thank the respective organisations for supporting the conference: It has been made possible by the close collaboration of ROSS Gemini Center, NTNU and the European Safety and Reliability Association.

Stein Haugen
Anne Barros
Coen van Gulijk
Trond Kongsvik
Jan Erik Vinnem

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Organisation

CONFERENCE GENERAL CHAIR

Stein Haugen, *Chair, Norway*
Coen van Gulijk, *Co-Chair*, *UK*

ORGANISING COMMITTEE

Jan Erik Vinnem, *Co-Chair, Norway*
Trond Kongsvik, *Co-Chair, Norway*
Anne Barros, *Co-Chair, Norway*

ESRA OFFICERS

Terje Aven, *Chair, Norway*
Radim Briš, *Vice Chair*, *Czech Republic*
Roger Flage, *General Secretary, Norway*
Piero Baraldi, *Treasurer, Italy*
Antoine Grall, *Conference Committee, France*

TECHNICAL PROGRAMME COMMITTEE

Stein Haugen, *Chair, Norway*
Ragnar Aarø, *Norway*
Eirik Abrahamsen, *Norway*
Eirik Albrechtsen, *Norway*
Vladimir Algin, *Belarus*
Reza Ahmadi, *Iran*
Ben Ale, *The Netherlands*
Petter Almklov, *Norway*
John Andrews, *UK*
Stian Antonsen, *Norway*
Terje Aven, *Norway*
Bjørn Egil Asbjørnslett, *Norway*
Piero Baralid, *Italy*
Anne Barros, *Norway*
Pierre Beauseroy, *France*
Christophe Berenguer, *France*
Lars Bodsberg, *Norway*
Nicolae Brinzei, *France*
Emanuele Borgonovo, *Italy*
Mario Brito, *UK*
Rolf Bye, *Norway*
Vu Hai Canh, *France*
Bruno Castanier, *France*
Marko Čepin, *Slovenia*
Terje Dammen, *Norway*
Yingjun Deng, *China*
Pierre Dersin, *France*
Martha De Souza, *Brazil*
Fateme Dinmohammadi, *UK*
Estelle Deloux, *France*
Henrik Hassel, *Sweden*
Jan Hayes, *Australia*
Marcelo Hazin Alencar, *Brazil*
Ivonne Herera, *Norway*
Peter Hughes, *UK*
Bob Huisman, *The Netherlands*
Tuan Huynh, *France*
Stig Ole Johnsen, *Norway*
Urban Kjellen, *Norway*
Bas Kolen, *The Netherlands*
Trond Kongsvik, *Norway*
Martin Kresja, *Czech Republic*
Nicolas Lefebvre, *Norway*
Gregory Levitin, *Israel*
Eric Levrat, *France*
Bjørn Ivar Kruke, *Norway*
Yves Langeron, *France*
Chaira Leva, *Ireland*
Yiliu Liu, *Norway*
Mary Ann Lundteigen, *Norway*
Benoit Iung, *France*
Jana Markova, *Czech Republic*
Effie Marcoulaki, *Greece*
Sebastian Martorell, *Spain*
Sophie Mercier, *France*
Ralf Mock, *Swizerland*
Jakub Montewka, *Poland*
Tor-Olav Nævestad, *Norway*
Thomas Nilsen, *Norway*
Knut Øien, *Norway*
Martin Rasmussen, *Norway*
Antoine Rauzy, *Norway*
Genserik Reniers, *Belgium*
Marilia Abilio Ramos, *Norway*
Eric Rigaud, *France*
Harald Rødseth, *Norway*
Willy Røed, *Norway*
Børge Rokseth, *Norway*
Tore Sagvolden, *Norway*
Giovanni Sansavini, *Swizerland*
Per Morten Schiefloe, *Norway*
Mahmood Shafiee, *UK*
Kari Skarholt, *Norway*
Snorre Sklet, *Norway*
Ann Britt Skjerve, *Norway*
Raphaël Steenbergen, *NL*
Fuqiang Sun, *China*
Trygve Steiro, *Norway*
Trine Thorvaldsen, *Norway*
Kristine Vedal Størkersen, *Norway*
Ingrid Utne, *Norway*
David Valis, *Czech Republic*
Pieter van Gelder, *The Netherlands*
Coen van Gulijk, *UK*
Do Van Phuc, *France*
Jørn Vatn, *Norway*
Jan Erik Vinnem, *Norway*
Zdenek Vintr, *Czech Republic*

Francesco Di Maio, *Italy*
Enrique Lopez Droguett, *Chile*
Rawia El Rashidy, *UK*
Serkan Eryilmaz, *Turkey*
Miguel Figueres, *UK*
Olga Fink, *Swizerland*
Roger Flage, *Norway*
Ingar Fossan, *Norway*
Knut Robert Fossum, *Norway*
Mitra Fouladirad, *France*
Gudveig Gjøsund, *Norway*
Antoine Grall, *France*
Bjørn Axel Gran, *Norway*
Ove Njå, *Norway*
Peter Okoh, *Norway*
Eivind Okstad, *Norway*
Girish Kumarm, *India*
Nicola Paltrinieri, *Norway*
Christian Paroissin, *France*
Edoardo Patelli, *UK*
Nicola Pedroni, *Italy*
Francois Peres, *France*
Kenneth Pettersen, *Norway*
Luca Podofillini, *Swizerland*
Thomas Porathe, *Norway*
Darren Prescott, *UK*
Aud Wahl, *Norway*
Lesley Walls, *UK*
Jin Wang, *UK*
Siri Wiig, *Norway*
Rune Winther, *Norway*
Xue Yang, *Norway*
Xiaojian Yi, *China*
Elena Zaitseva, *Slovenia*
Jifen Zhang, *China*
Enrico Zio, *Italy*
Wenjin Zhu, *China*

EDITORIAL TEAM

Anna Dong, *Norway*
Børge Rokseth, *Norway*
Siri M. Holen, *Norway*

ESRA TECHNICAL COMMITTEES AND CHAIRS

Methodologies Chairs	
Accident and Incident Modeling	Stig Johnsen, Nicola Paltrinieri
Economic Analysis in Risk Management	Eirik B. Abrahamsen
Foundational Issues in Risk Assessment and Management	Terje Aven, Enrico Zio
Human Factors and Human Reliability	Luca Podofillini, Maria Chiara Leva
Maintenance Modeling and Applications	Christophe Bérenguer, Mitra Fouladirad
Mathematical Methods in Reliability and Safety	John Andrews, Nicolae Brinzei
Prognostics and System Health Management	Piero Baraldi, Enrico Zio
Resilience Engineering	Ivonne Herrera, Eric Rigaud
Risk Assessment	Marko Čepin, Henrik Hassel
Risk Management	Lesley Walls, David Valis, Marcelo Hazin Alencar
Simulation for Safety and Reliability	Nicola Pedroni, Edoardo Patelli
Structural Reliability	Jana Markova, Martin Krejsa
System Reliability	Gregory Levitin, Serkan Eryilmaz
Uncertainty Analysis	Emanuele Borgonovo, Roger Flage

APPLICATION AREAS AND TECHNOLOGICAL SECTORS CHAIRS

Aeronautics and Aerospace	Darren Prescott
Chemical and Process Industry	Valerio Cozzani, Gabriele Landucci, Nima Khakzad
Civil Engineering	Raphael Steenbergen
Critical Infrastructures	Giovanni Sansavini, Enrico Zio
Energy	Michalis Christou
Information Technology and Telecommunications	Elena Zaitseva, Ralf Mock
Land Transportation	Olga Fink, Bob Huisman
Manufacturing	Benoit Iung, François Peres
Maritime and Offshore Technology	Jin Wang, Ingrid B. Utne, Mario Brito
Natural Hazards	Pieter van Gelder, Bas Kolen
Nuclear Industry	Sebastian Martorell, Francesco Di Maio
Occupational Safety	Ben Ale, Reniers Genserik
Security	Sissel H. Jore, Zdenek Vintr

Accident and incident modeling

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Comparing HFACS and AcciMaps in a health informatics case study—the analysis of a medication dosing error

O.O. Igene & C.W. Johnson
Department of Computing Science, University of Glasgow, Scotland

ABSTRACT: The utilization of Information Technology/software systems is considered a proactive measure for reducing medication errors, providing clinical efficiency and improving patient safety. However this has added a layer of risk that can potentially harm patients and compromise safety. A comparative study using specific accident models; the Human Factors and Classification Systems (HFACS) and Accident Mapping (AcciMaps) was utilized on a health IT related case study analysis of medication error relating to the Computer Provider Entry System (CPOE). The results (outcomes) of the analyses were compared using the usage characteristics criteria developed by Underwood and Waterson (2014). The usage criteria framework focuses on ease of learning (usability), data requirements, validity and reliability of analysis. The second objective of our study discusses the limitations of both models and proposes a way forward on enhancing the usability, validity of results and more importantly the reliability of the AcciMap approach for accident analysis in healthcare.

1 INTRODUCTION

The development and utilization of Information Technology (IT)/software systems within clinical settings in hospitals has helped to reduce medical errors and improve efficiency in health care delivery. However, its implementation and utilization has also introduced an additional layer of risks and unforeseen errors that can compromise patient safety (Koppel et al. 2005, IOM 2012, Magrabi et al. 2016). This is especially apparent within complex safety critical sociotechnical systems like healthcare (IOM 2012, Schneider et al. 2014). A sociotechnical system while regarded as an imprecise term (Klein 2014) consists of complex interactions between different entities (people, technology, process, organization and external environment) (Sittig & Singh 2010). Accident/errors can occur as a result of interactions involving people (clinicians/physicians) using software/IT systems. They typically stem from latent conditions (systemic factors) that stretch well beyond the "frontline" of healthcare service provision.

The development of accident causation methods has provided a means of investigating and analyzing failures. They focus on the subsequent development of counter measures to improve patient and system safety (Johnson 2004). These accident analysis techniques include but are not limited to linear/event, taxonomy and systemic based methods based on recognized accident causation theories like Swiss Cheese Model (SCM) (Reason 1995, Qureshi 2008). Systemic methods are considered to be more suitable for analyzing interactions in sociotechnical systems than linear and event based approaches (Qureshi 2008, Leveson 2011). One of such methods includes the AcciMap method (Svedung and Rasmussen 2000). Systemic analysis techniques are typically applied retrospectively in analyzing incidents/case studies and different outcomes are produced based on their component methods.

This paper presents a comparative study using a systematic taxonomy framework (HFACS) and AcciMaps on a published clinical case study report (Horsky et al 2005). The main purpose of the study is to identify the putative causes of errors resulting from interactions between the users and IT system. We are also motivated to identify contributing factors associated particularly at the systemic level to understand not only what happened but crucially why. Different components within the sociotechnical system cannot be analyzed in isolation from one another especially if system safety is to be achieved. The application of HFACS in healthcare incident analysis has been established but AcciMaps has not been utilized as an accident analysis tool for clinical investigation and analysis.

This paper is divided into sections highlighting the objectives of the comparative analysis,

the methodology applied as well as the description of an example clinical case study utilized. Each of the methods selected is also briefly described, applied to the case study and their respective results (outcomes) will be compared and discussed based on the usage characteristics (Horsky et al. 2005). Limitations of each method are highlighted leading to the discussion of the proposed method to address some of these limitations.

2 OBJECTIVES OF RESEARCH

The two methods were selected based on their differences in their methodological approach and systematic way of analyzing accidents. The objectives of the study are as follows:

I. Analysis of the medication dosing error case study involving the Computer Provider Order Entry system using the HFACS and AcciMap models.
II. A comparison of the resulting outputs of each accident method based on the usage characteristics criteria (Underwood and Waterson, 2014).
III. Identifying common causes and contributing factors from analyses relating to the adverse outcome.

3 RESEARCH METHODOLOGY

This study adopts a qualitative (case study) approach for identifying, understanding and analyzing a complex sociotechnical technical setting involving healthcare informatics. We aim to expose the contributing factors that are associated with an adverse event. A medical case incident was used in a comparative study of accident causation methods; the Human Factors and Classification Systems (HFACS) (Shappell & Wiegmann 2000) and the Accident Mapping (AcciMaps) method, a component of the broader Risk Management Framework (RMF) (Rasmussen and Svedung 2000).

This case incident details a sociotechnical scenario about risks relating to the house providers and the Computerized Provider Order Entry System (CPOE) (Horsky et al. 2009, IOM 2012). The selected accident methods were used to analyze the case incident and the outputs were further iterated, reviewed and validated with a HFACS and an AcciMap expert. The resulting outputs were then compared using the usage criteria framework consisting of its graphical representation, usability, validity and reliability (Underwood & Waterson 2014). Strengths and limitations of the methods applied are elaborated in section 8.

4 DESCRIPTION AND ANALYSIS OF CASE STUDY

The case study involves two providers (A and B) who were involved in the administration of Potassium Chloride (KCI) to a patient who was initially hypokalemic. The timeline of the events leading to the patient becoming hyperkalemic (receiving a high dosage of KCl) occurring over a period of three days are detailed in the work of Horsky et al. (2005). The patient was first examined by the first physician (Provider A) and administered an intravenous (IV) bolus injection thereby repleting the potassium. Provider A then realized that the patient already had an IV and so decided to administer the KCI as part of the treatment. The patient started receiving a higher KCI dosage than what was originally intended due to several events that took place. The initial dosage order was detected to be higher than what was allowed by the hospital's policy and was discontinued. A new dosage order had to be written. However this new dosage order was not entered correctly into the CPOE system and the maximum volume was not indicated for the fluid that was to be administered to the patient (Horsky et al. 2005).

A changeover between the first physician and the incoming physician (Provider B) took place the next day where the latter was notified about the patient's KCI levels from the system. However, the second provider did not know that the laboratory results were not current and it was before the last potassium repletion occurred (Horsky et al. 2005). This led to provider B to consider the KCI levels of the patient to be low and decided to order an additional IV injection despite the KCI still running. This eventually led to the patient to become severely hyperkalemic and the problem was immediately rectified and the patient was discharged (Horsky et al. 2005). Their study provided a comprehensive analysis of the contributing factors as a result of both human errors and system design issues that led to the patient experiencing an adverse event.

5 DESCRIPTION OF THE ACCIDENT ANALYTICAL METHODS

The methods selected for the case study are based on recognized accident causation theories and are briefly described below:

5.1 *Human Factors and Classification System (HFACS)*

The HFACS method is a systematic approach that is based on the Swiss Cheese Method (SCM) (Reason 1995). This approach compares different levels of 'causal categories' relating from active to

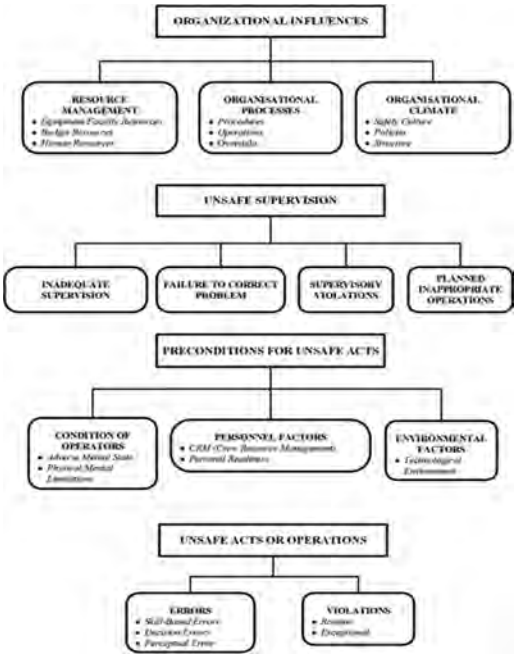


Figure 1. HFACS taxonomy (Shappell and Wiegmann, 2002).

latent conditions as shown in Figure 1 (Shappell & Wiegmann 2000). The categories include Unsafe Acts, Preconditions for Unsafe Acts, Unsafe Supervision, and Organizational Influences (Shappell & Wiegmann 2000). Each of them consists of specific subcategories that place emphasis on "who" and "what" rather than on "why" accidents or a significant adverse event occurred (Diller et al. 2013). This method can be utilized to analyze a comprehensive case study or a set of incidents to determine trends so as to develop countermeasures.

5.2 *Accident Mapping (AcciMaps) method*

The AcciMap method was developed as a component of the Risk Management Framework (RMF) and can be utilized as a standalone technique aside and as part of the broader RMF methodology consisting of the use of ActorMaps and Conflict Maps (Rasmussen & Svedung 2000). The AcciMap method allows analysts to graphically map causal relationships leading to an adverse event across each of the levels in a sociotechnical system (Branford 2011). The AcciMap levels as shown in Figure 2 depict interactions in a sociotechnical system and consist of Environment/Surroundings, Physical Processes/Actor activities, Organizational (Technical/Operational Management, Company

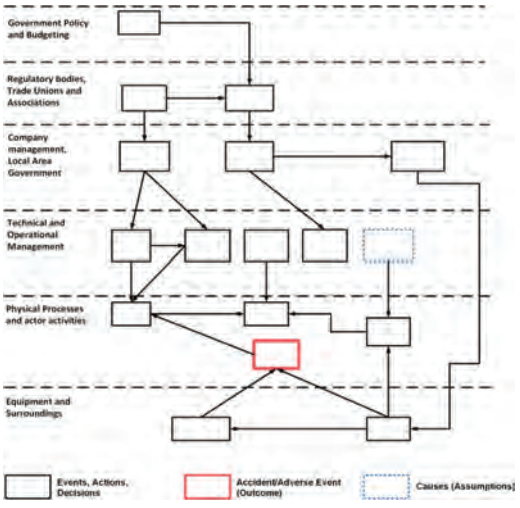


Figure 2. The AcciMap method adapted from Rasmussen and Svedung (2000).

management), and External (Regulatory bodies/ Trade unions, and Government Policy and Budgeting) (Rasmussen & Svedung 2000).

6 RESULTS

The resulting outputs from both methods are discussed below:

6.1 *HFACS output*

The results of applying HFACS were given to an experienced human factors specialist for verification and validation. The factors were identified and classified according to each level's defined categories and sub-categories of the HFACS taxonomy as shown in Figure 3 and described as follows:

6.1.1 *Unsafe acts*

Operators (clinicians) not using the CPOE system correctly was categorized under '*skill based errors*'. Lack of communication between the clinical providers (A and B) that led to the second provider administering additional KCI dosage to the patient is considered a '*decision error*'. While there wasn't any explicit evidence relating to '*perceptual errors*', violations regarding clinical handover procedures between Providers A and B was noted to have increased the risk of the patient becoming hyperkalemic.

6.1.2 *Preconditions for unsafe acts*

Several factors at this level include those relating to the design and usability of CPOE system

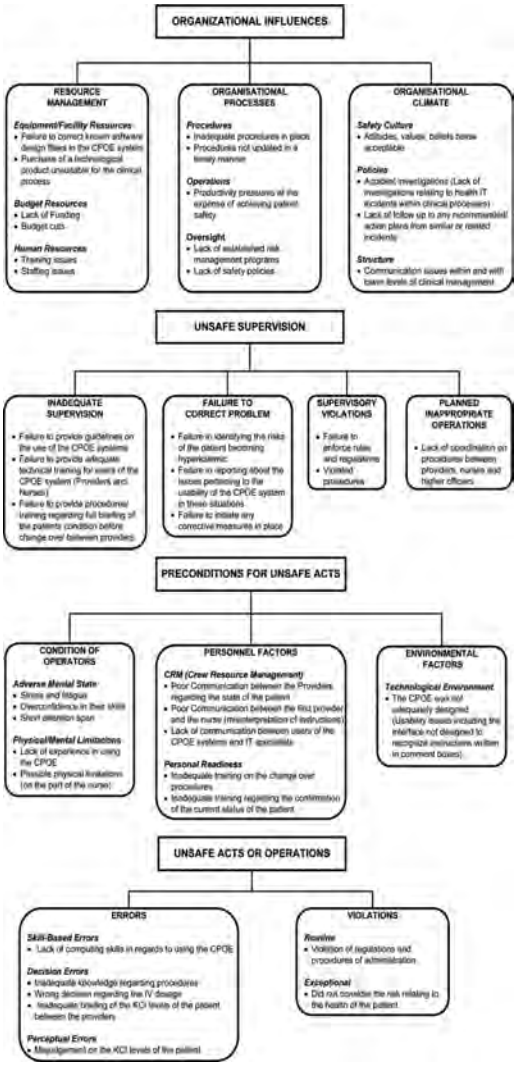


Figure 3. Application of the HFACS method on the medication dosing error incident.

(*Environmental factors*) and lack of user experience with the current version of the CPOE system (*Physical and Mental limitations*). Also at the '*Personal Readiness*' subcategory level, factors identified are attributed to issues regarding lack of communication between the providers as well clinical handover procedures not adequately carried out. This could be due to lack of effectiveness in training regarding the procedure.

6.1.3 *Unsafe supervision*

At both the '*supervisory violations*' and '*failure to correct problem*' subcategories, contributing factors arise from issues relating to inadequate training on the use of the current CPOE system. Furthermore, the system may not have been rigorously tested at the initial stage for its usability in a simulated environment (no explicit evidence in the case). In addition, there is a possibility of inadequate communication and coordination between the actors (providers A and B) and those at the management level including those responsible for IT systems which is classified under the '*Planned Inappropriate Operations*' subcategory.

6.1.4 *Organizational influence*

It was noted that within these subcategories, '*resource management*', '*organizational processes*' and '*organizational climate*', the case study did not indicate explicit evidence as to the contributing factors that enabled the accident. In examining the other subcategories within this category, it can be indicated that the organization may have underestimated the risks and severity relating to IT systems. In addition, there may not be any existing policies relating to testing the usability of the CPOE system within a simulated environment before its deployment. Other issues can include budget restrictions and low safety culture will be considered potential systemic factors.

6.2 *AcciMap output*

The original AcciMap method (Rasmussen and Svedung, 2000) and the standardized AcciMap format (Branford 2011) was applied on the case incident (See Figures 4 and 5). The latter was utilized by an experienced clinician (e-health pharmacy) and this method provided a simpler way of analyzing the case study particularly at the organizational level. The AcciMap outputs were externally validated by a human factors expert.

However for the purposes of study, the result of the AcciMap model is described using the original AcciMap format as follows:

6.2.1 *Equipment and surrounding*

In this case study, the equipment used here will be the CPOE system that allows providers to place orders regarding the administration of KCl to the patient. However human errors occurred as a result of issues relating to the design and usability of the system (relating to the interface of the system where instructions were difficult to read) leading to the eventual adverse outcome.

6.2.2 *Physical processes and actor activities*

This level refers to the actors involved in the adverse event (house providers, attending nurse and the patient). Here, provider A administered a high dosage of KCl which was entered incorrectly into the

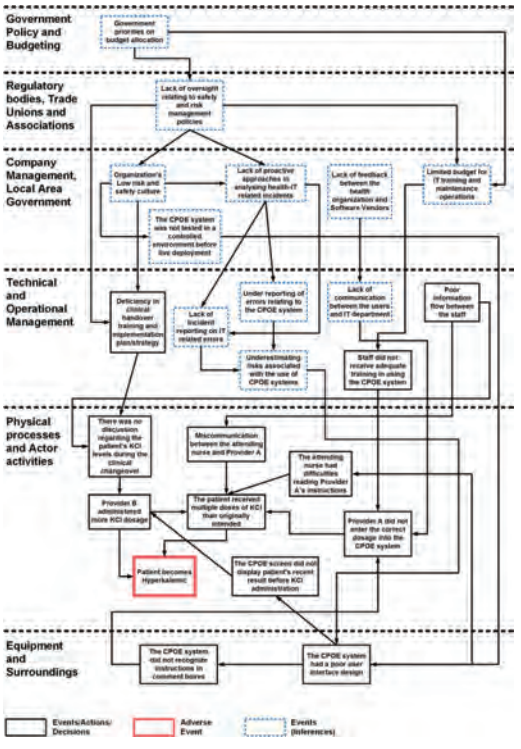


Figure 4. AcciMap analysis of the medication dosing error case study using the original format.

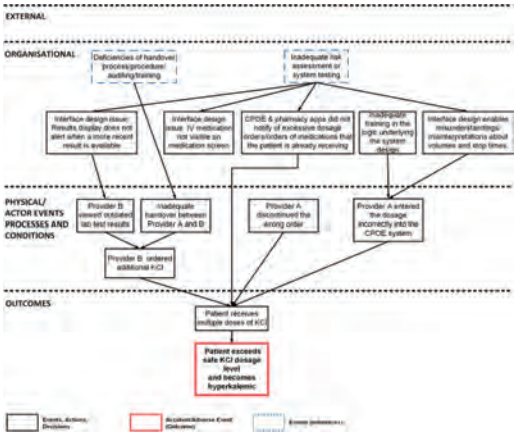


Figure 5. AcciMap analysis of the medication dosing error case study using Branford's standardized AcciMap format.

system. The second provider further implemented an additional dosage of KCl after failing to ascertain the current KCl levels of the patient by reading the outdated results from the CPOE system during the changeover process. Further problems occurred as a result of miscommunication between providers A and B on the patient's KCI levels during the clinical handover process and as a result, the patient became hyperkalemic.

6.2.3 *Technical and operational management*

At this level, issues relating to CPOE include specifically its interface design flaws which increased the probability of an error to occur. This could also be attributed to lack of rigorous software simulation testing of the system before it was deployed for clinical use. The IT vendor responsible for the development of the software may cite lack of training on the use of the software rather than the problem of the system design. Although this point can be inferred as a contributing factor, there was no explicit evidence from the reports that this was the case. Another inference that can be made is that similar incidents may not have been reported as an "IT-related" event and risks that can arise from system design of the CPOE was either underestimated or may not have been taken into consideration by the health organization.

6.2.4 *Company management and local government*

The case report did not contain explicit evidence regarding specific failures at this level. However, from the analysis of the preceding levels and some inferences, errors can occur due to lack of oversight, underestimation of the severity of risks associated with IT systems. Additionally, there may be issues relating to inadequate training and enforcement of existing polices relating to clinical handover process between providers can also contribute to the adverse event.

6.2.5 *Regulatory bodies, trade unions and associations*

At this level, contributing factors could include failure of implementing existing policies passed by the government and other relevant medical bodies carried out to ensure patient safety. Another inference could be that incidents relating to IT systems are not being adequately reported and lack of proactive measures may not be in place to handle such kind of incidents. However, it should be noted again that there was no explicit evidence in the case study to support these inferences.

6.2.6 *Government policy*

Issues like budget limitations, priorities on budget allocations and policies regarding the reporting of incidents relating to software/IT systems implemented in health organizations would be considered as systemic factors. However, a much more detailed report will need to include these systemic

factors that could contribute to the accident or any similar incidents. Based on the AcciMap outcome, there was no explicit cause indicated due to lack of evidence relating to systemic factors at this external level (government). This can be attributed to the fact that incident reports do not typically contain information especially on contributing factors in enabling the government to realize their role in creating latent conditions that can further enable similar incidents to occur again.

7 USAGE CHARACTERISTICS CRITERIA

The HFACS and AcciMap methods utilized for the case study is briefly described and the resulting outputs are discussed below:

7.1 *Data requirements*

The use of both methods requires data which can be collected in various formats including case incidents or a detailed case report of significance. Other data sources include documentations, interviews with actors (frontline workers, management) involved in the incidents and observing the events in a sociotechnical scenario. Both methods require a great deal of details from the actors directly involved in the incident and at the higher (organizational) levels to be to identify valid causes and contributing factors to accidents.

7.2 *Graphical representation of the accident*

HFACS and AcciMaps differ in this category. The HFACS only used a set of defined categories that classify causes to the accident and so lacks a way of graphically representing contributing factors and their relationships. However, the AcciMap method provides this advantage by enabling analysts to use different graphical symbols to represent different meanings; adverse event (accident), causes (evidence), causes (inferred) in relation to the adverse outcome (Branford 2011, Salmon et al. 2012). AcciMaps also allow the causes identified to be represented using causal relationships (represented by arrows) within and between each level of the sociotechnical system.

7.3 *Usability of method*

The use of both HFACS and AcciMaps is relative to the level of skill and experience of the analysts. While both methods require understanding of the accident causation theories they are built on, the HFACS method provides a framework of failure categories built for classifying failures according to the Swiss Cheese Model from either singular or multiple incidents. The AcciMap method provides users the basic understanding of the causes that interconnect in a vertical manner to enable understanding of why the accident took place. However, the use of these methods can be relatively time-consuming depending on the complexity and comprehensiveness of the case study, and due to its subjective nature of analysis from multiple users.

7.4 *Validity of method*

Each of the outputs from HFACS and AcciMaps were reviewed by an experienced and expert user of both methods for both content and face validity (hence external validation being required). As both methods have differing methodological approach (each based on a recognized accident causation theory), its results will only reflect the methods perspective as to the analysis of the accident and possible safety recommendations needed to be in place.

7.5 *Reliability of analysis*

In previous studies, the HFACS method was noted to provide a considerable measure of reliability due to the defined failure categories (Salmon et al. 2012, Ergai 2013). However as was demonstrated in the case study, the AcciMap method is limited in terms of considering failure categories that can exist at different levels including the external level as it relates to this case study. This is the reason why reliability of the AcciMap method is considered to be low, due to the subjective nature of its analysis (Salmon et al. 2012, Underwood & Waterson 2014) especially because multiple analysts can produce different causal outcomes from the same case incident. There is also the issue of bias including hindsight bias as to what contributed to the adverse outcome.

8 DISCUSSION

The implementation of both HFACS and AcciMaps shows both strengths and limitations they offer regarding accident analysis. Drawbacks of the AcciMap method include its subjectivity of analysis and the lack of failure categories in each of the levels (Salmon et al. 2012). The issue of subjectivity is due in part to lack of existing standard guidelines in addition to the need for external validation of results generated. This can ultimately affect the type of safety recommendations that are needed, and how effective they are in preventing an occurrence of the accident. Although Branford made an attempt to solve that issue through the development of the standardized AcciMap format,

it still lacks a formal way of classifying failures into specific categories. This limitation makes the AcciMap method not suitable for analyzing multiple incidents (Salmon et al. 2012, Goode et al. 2017). There is also the issue of bias including hindsight and outcome biases, which can also affect the results of the accident analysis, especially in determining why the adverse event occurred (Johnson 2004).

In comparing outputs from the analyses, similar causes from the analyses include communication issues between the actors (Providers A and B), software design issues with the CPOE system, issues relating to clinical handover process between the providers and inadequate/ineffective training relating the use of the current CPOE system. The difference however lies in identifying contributing factors at the higher levels especially as to how these factors at external levels (Regulatory bodies and the Government) can systemically contribute to the occurrence of the accident. For example, based on the AcciMap result, inferences was made as to why there was ineffective feedback between the health organization (i.e. IT department) and the software vendor responsible for the design of the CPOE system. Another inference could be made regarding inadequate policies relating to testing the software product (albeit in a simulated environment) before it was deployed live for clinical purposes. While both methods assist in identifying causes and contributing factors in a systematic way, the AcciMap method is not restrictive in terms of identifying and placing causes at each level while in the case of HFACS, the causes need to be classified according to the defined framework of failure categories (based on Reason's Swiss Cheese Model of accident causation). The HFACS taxonomy is also considered a generic-based method (initially developed for investigating accidents in the aviation system), but has also been adopted in the healthcare system (Diller et al. 2013). However, one of the limitations attributed to HFACS method is in the restrictive nature of the categories within the HFACS taxonomy. There is also the need to expand and adapt the taxonomy within specific healthcare scenarios including the addition of higher level related factors (External) (Salmon et al. 2012) as well as any other factors not included in the original format. The advantages of HFACS and/or any similar health-based error taxonomies and AcciMaps can potentially be combined together to improve the latter's reliability. This step was explored in investigating multiple led outdoor incidents through the development of an incident reporting and learning system based on the UPLOADS (Understanding and Preventing Led Outdoor Activities Data Systems) taxonomy (Salmon et al. 2015). Each AcciMap level had failure categories based on contributing factors identified from multiple incident reports as it related to the actors/decision makers in each of the level. Causal relationships, which are a very important feature of the AcciMap method, were also depicted between contributing factors (relating to the actors identified in the system) within and between each AcciMap level (Salmon et al. 2015, Goode et al. 2017).

This similar approach can potentially be applied in investigating significant incidents and near misses as a result of software issues within the healthcare context. This could not only potentially improve the validity of results but can also enhance its usability for accident analysis and adoption by safety practitioners in healthcare systems. This approach can also be utilized to investigate IT failures or near misses so that lessons learned can be used as a way of improving safety culture towards the implementation and utilization of health IT systems by its operators.

9 LIMITATION OF STUDY

This study has limitations in its identification of causes and in the validation of outcomes/recommendations from the analysis. It was particularly challenging identifying systemic factors at the higher levels (in the case of the AcciMap analysis) due to lack of evidence in the report; especially dealing with governance. Another issue is the knowledge and application of the methods. especially for first time users in identifying causes that are valid and that actually contributed to the occurrence of the adverse event. The last point is very important for the development of effective safety countermeasures and to which level this countermeasure is applied to. Future studies will require the use of a more structured approach by involving a multidisciplinary team ideally from all sociotechnical levels. This will involve safety practitioners, IT specialists, users of software systems, vendors and expert opinions in the analysis of health IT related accidents.

10 CONCLUSION

Both methods allow for a systematic identification and analysis of accidents in complex sociotechnical systems like healthcare. The purpose of the study was not to indicate that one method is better than the other, but to highlight the advantages and limitations of each method in the analysis of the case study. Isolating software related problems and taking steps to improve the functionality and reliability of the system does not necessarily improve

patient safety. Accidents where software systems played a role must be analyzed from a sociotechnical perspective. This is why human factors engineering plays a very important role as it involves analyzing complex interactions between clinicians and software systems utilized, and determining latent conditions based on decisions at the systemic level.

11 FUTURE WORK

Beyond the objectives of this present study, a current study is underway focusing on the development of a health-specific AcciMap taxonomy model. This model will then be used to analyze cases/incidents relating to software/IT related accidents in healthcare settings. The development of the model will comprise of examining existing health based error taxonomies similar to the HFACS and contributing factors framework. This will then be synthesized within the levels of the standardized AcciMap method. The purpose of this approach will be to determine if both the reliability and validity of the AcciMap method will be improved through the use of defined failure categories in a taxonomic structure.

ACKNOWLEDGEMENT

The HFACS result was reviewed by Dr. Suzanne Shale, trained in the use of the HFACS method by Scott Shappell, who offered comments and approach to analysis. Further acknowledgement goes to Dr. Kate Branford, a human factors and AcciMaps expert as well Mr. Iain Bishop, an e-Health Pharmacy Adviser, Information Technology of the National Services Scotland (NSS).

REFERENCES

Branford, K. 2011. Seeing the big picture of mishaps: applying the AcciMap approach to analyze system accidents. *Aviation Psychology and Applied Human Factors*, vol. 1(1), pp. 31–37.

Diller, T., Helmrich, G., Dunning, S., Cox, S., Buchanan, A., and Shappell, S. 2013. The Human Factors Analysis Classification System (HFACS) applied to healthcare. *American Journal of Medical Quality*, vol. 29, issue 3, pp. 181–190.

Ergai, A. 2013. Assessment of the Human Factors Analysis and Classification Systems (HFACS): Intra-rater and Inter-rater reliability. *All Dissertations*, paper 1231.

Goode, N., Salmon, P.M., Taylor, N.Z., Lenne, M.G., and Finch, C.F. Developing a contributing factor classification scheme for Rasmussen's AcciMap: reliability and validity evaluation. *Applied Ergonomics*, vol. 64, pp. 14–26.

Horsky J., Kuperman, G.J., and Patel, V.I. (2005). Comprehensive analysis of medication dosing error related to CPOE. *Journal of the American Medical Informatics Association*, vol. 12, pp. 378–381.

IOM. 2012. Health IT and patient safety: building safer systems for better care.

Johnson, C.W. An introduction to root cause analysis in healthcare. http://www.dcs.gla.ac.uk/~johnson/papers/Pascale_book/incident_analysis.PDF.

Klein, L. 2014. What do we mean by sociotechnical? On values, boundaries and the problems of language. Journal of Applied Ergonomics, vol. 45, pp. 137–142.

Koppel, R. 2005. Role of computerized physician order entry systems in facilitating medication errors. *Journal of the American Medical Association,* vol. 293, no. 10, pp. 1197–1203.

Leveson, N.G. 2011. Applying systems thinking to analyze and learn from events. *Journal of Safety Science*, vol. 49, pp. 55–64.

Magrabi, F., Ong, M., and Coiera, E. 2016. Health IT for patient safety and improving the safety of health IT. *Evidence-based Health Informatics*, pp. 25–36.

Qureshi, Z.H. 2008. A review of accident modelling approaches for complex critical sociotechnical systems. *Department of Science and Technology Organization, Department of Defense, Australian Government*, pp. 1–72.

Rasmussen, J. and Svedung, I. 2000. Proactive risk management in a dynamic society. https://www.msb.se/RibData/Filer/pdf/16252.pdf.

Reason, J. 1995. Understanding adverse events human factors: human factors. *Qual Health Care*, vol. 4, pp. 88–89.

Salmon, P.M., Corneliessen, M., and Trotter, M.J. 2012. Systems-based accident analysis methods: a comparison of AcciMap, HFACS and STAMP. *Journal of Safety Science*, vol. 15, issue 4, pp. 1158–1170.

Salmon, P.M., Goode, N., Taylor, N., Lenne, M.G., Dallat, C.E., and Finch, C.F. 2015. Rasmussen's legacy in the great outdoors: a new incident reporting and learning system for led outdoor activities. *Journal of Applied Ergonomics*, pp. 1–12.

Schneider, E.C. 2014. Promoting patient safety through effective health information technology risk management, RAND research report, https://www.healthit.gov/sites/default/files/rr654final_report_5-27-14.pdf.

Shappell, S. and Wiegmann, D. 2000. The Human Factors Analysis and Classification System—HFACS. https://www.nifc.gov/fireInfo/fireInfo_documents/humanfactors_classAnly.pdf.

Sittig, D.F. and Singh, H. 2010. A new sociotechnical model for studying Health Information Technology in complex adaptive healthcare systems. *Qual Saf, Healthcare*, 19, suppl. 2, pp. i68–i74.

Underwood, P. and Waterson, P. 2014. Systems thinking, the Swiss Cheese and accident analysis: a comparative systemic analysis of the grayrigg train derailment using the ATSB, AcciMap and STAMP models. *Journal of Accident Analysis and Prevention*, vol. 68, pp. 75–94.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Possibilities of using simulation software to estimate losses of industrial facilities and installations—critical analysis

J. Ryczyński, P. Mastalerz, K. Ksiądzyna & T. Smal
Tadeusz Kosciuszko Military University of Land Forces, Wroclaw, Poland

ABSTRACT: In the paper, simulation of severe failure of fuel terminal is presented. Results of the simulation are compared to the real situation, that took place on 11th December 2005 in Hertfordshire Oil Storage near the M1 motorway in England. That comparison shows how beneficial an accurate simulation of the accident is. Next, the risk analysis is presented. The Preliminary Hazard Analysis (PHA) and the fault tree are methods that show possible cause of the accident. Using simulation, the course of action before and during failure could be traversed. That kind of research shows how important is to do the simulation before the accident occur and assess the possible consequences of the adverse situation. That simple method is significant to increase awareness of the fuel terminal workers and anyone who is related to the technical process safety.

1 INTRODUCTION

It is commonly known, that already high demand for fuel products is growing year by year. What is more, this kind of products has to be stored properly and transported in special conditions. Because of that, the authors decided to look into this topic. Analysis of previous petrochemical failures shows, that most of accidents with fuel products occur during loading and unloading of the materials. Those operations are crucial and cannot be omitted during the whole transportation process. Thereupon, both operations, loading and unloading, should be performed in a way that ensures no unwanted situation occurs. To fulfill the safety rules and eliminate the possible risk, series of complex theoretical analysis are made before the transportation of the fuel products is possible. Nowadays, computer simulations are used to estimate possible consequences of failure. Simple computer programs are able to show many indicators, like the dangerous zones range, which is helpful especially for personnel responsible for safety of the technological process.

According to data, the requirement for fuel products rises year by year, hence the need to store and transport it. Without the suitable stock of liquid fuel storage the whole transportation would not be able to function. Security in each of these processes is a crucial element. Unluckily, it is often not possible to forecast events that menace safety in general, including the safety of fuel terminals. Figure 1. introduce the dependence of oil manufacturing and consumption of crude oil products

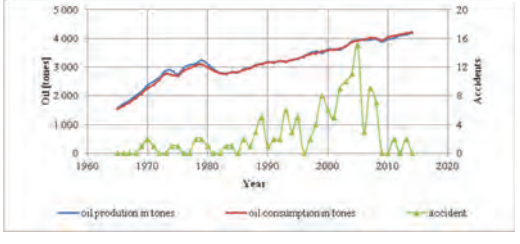


Figure 1. Oil production with consumption and amount of accidents between 1960 and 2016 (BP, Statistical Review of World Energy June 2016, edition 65-th).

on calamity rates between 1965 and 2016. Because of that, multi-stage risk analysis presents such a vital role in the petrochemical industry.

For risk analysis to reveal reality precisely, different databases including industry cases are utilized. In Poland there are two programs: the System Pomocy W Transporcie Materiałów Niebezpiecznych—transport of dangerous goods aid system (SPOT) and the Safe Chemistry, as an covenant between the Polish Chamber of Chemical Industry and the leading plants in the chemical industry in the Republic of Poland. One of the crucial elements of this collaboration is, among others, to form a common disaster database as a ground for sharing requests for unfavourable events.

The European equivalents are the Major Accident Reporting System (eMARS), Analyse, Recherche et Informations sur les Accidents (Aria) and Failure and Accidents Technical information

System (FACTS). This sort of database should be kept topical with new industry incidents, establishing a kind of precedent. The high consumption of fuel products and the evident safety magnitude have led the authors to become interested in the matter of fuel terminals.

In this paper, a simulated malfunction of the fuel terminal was shown via the usage of computer modeling. The simulation outcome was compiled with the positive situation that occurred on the Hertfordshire Oil Storage near the M1 at Hemel Hempstead in Hertfordshire in Europe (Mannan et al., 2009). This approach permit a similitude of both situations to project the accuracy of modeling. Based on the scores, we generated an assumptive simulation of the detonation of the whole storage base in order to settle the possible implications in the event of non-trip events in this type of facility.

2 SAFETY IN STORAGE LIQUID FUELS

Storage liquid fuels safety depends primarily on the proper storage infrastructure, legitimate handling of loading and unloading actions and compliance with minimum tank distance provisions (Ryng, 1989). Additionally, it is significant to notice that, in spite of expanded automation, qualified and experienced personnel are the major determinants of workplace safety.

High haze volatility, continuous fire and explosion hazards necessitate the precaution against exorbitant loss of stored substance. These losses are mainly caused by the so-called 'breathing tanks'. Attempts to limit the losses include (Ryng, 1989):

- the usage of "tanks operating under pressure within the limits held by valves";
- maintaining the maximum filling condition of the tank;
- protective coatings that reflect sunlight;
- the usage of isolating tanks;
- placing underground tanks.

Flammable liquid tanks are fitted with safety attachments, which can comprise of:

- respiratory valves protected with valves with a fire stop;
- fire fuses;
- throats for feeding foam from the fire extinguishing system;
- Pouring and Emptying Devices (PED) installations;
- control and measuring apparatuses;
- double bottoms;
- fire alarm systems (mechanical and automatic);
- sprinkler systems;
- Faraday's grids;
- tank control systems—Distributed Control System (DCS);
- emergency pools.

3 THE FUEL TERMINAL FAILURE IN BUNCEFIELD—COMPARATIVE SIMULATIONS

In Hertfordshire, at the daybreak of the 11th of December of 2005, a fuel terminal failure arose. The consequence of this event was the loss of about 1/3 of the entire stockpile of about 10 million liters of liquid fuels. The effects of the explosion of the gathered medium were considerable damage of up to 10 km from the site (several houses were grievously damaged, while hundreds had incurred smaller, non-injurious damages). Accordingly, there were no fatalities, but over 3,000 direct casualties are estimated (no fatalities, with more than 2,000 homes and 92 businesses evacuated in the neighborhood). As an outcome of the fire, massive amounts of smoke were emitted, spreading over southern England. The fire has been burning for 3 days and to quench it, gargantuan amounts of water and foam were spent. 180 firefighters took part in the rescue operations by at the peak of the event. The course of events (Buncefield major incident investigation. Initial Report, Hemel Hempstead, on 11 Dec. 2005, p. 7):

1. At around 3:00, the fuel gauge for tank 912 (Figure 2) indicated a steady level of filling, even though the tank was filled with 550 m^3/h; Around 5:20, the tank began to overflow as a result of overfilling.
2. At 5:38, approximately 300 t of gasoline flowed through the roof of the tank to the emergency pool, resulting in an explosive mixture (approx. 5:38 of the emergency pool).
3. About 5:50 mass flow of fuel to the tank was 890 m^3/h.
4. At 6:01 am the first explosion occurred, resulting in the fire of 20 storage tanks.

Simulations have been performed, based on the post-accident studies data, to determine three phases of toxic cloud formation. In (Krawczyszyn et al., 2017), authors used Buncefield case to show accompanying phenomenon and consequences of the accident, such as formation of release zones, the floodplain area prior to the explosion and explosion itself. In this paper, Buncefield catastrophe is used to introduce consequences for humans. It should be emphasized, that each result shown in this work is assumed, because there were no fatalities and injured in Buncefield accident. What is more, the topology of the area was not considered in current work. Possible injuries in case of fire in

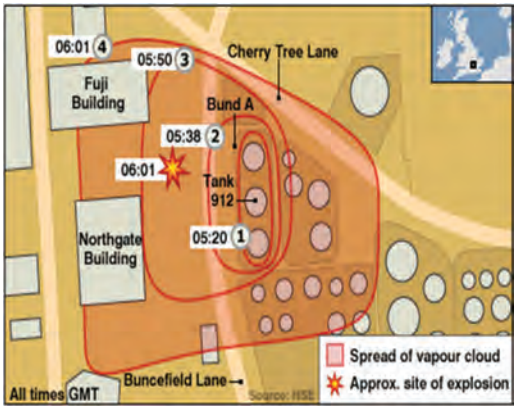


Figure 2. The spread of the toxic cloud according to accident reports (Mannan et al., 2009).

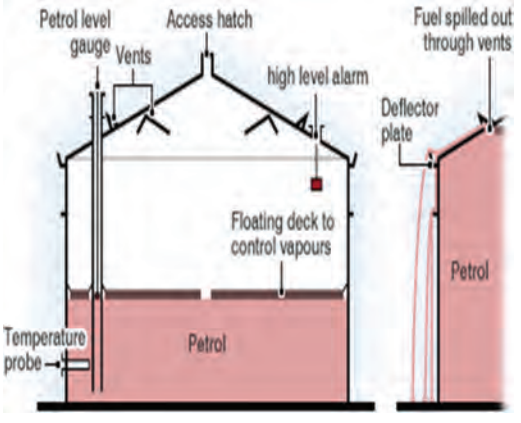


Figure 3. The spread of the toxic cloud according to accident reports (Mannan et al., 2009).

different stages of fuel spread are presented. Next part of the paper illustrates scathes that could be received during Buncefield catastrophe.

4 FIRST SIMULATION

Figures 4, 5 and 6 show dependence of burns degree in percent in function of exposure duration for 2 kW/m^2, 5 kW/m^2 and 10 kW/m^2 radiation heat. The mathematical model is based on value of probit function and percentage measure relation (Committee for the Prevention…, 1989):

$$Probit = -38.48 + 2.56 * \ln\left(t * q^{4/3}\right) \quad (1)$$

where: t–time of exposure (s), q–heat-flux absorbed (Wm^2).

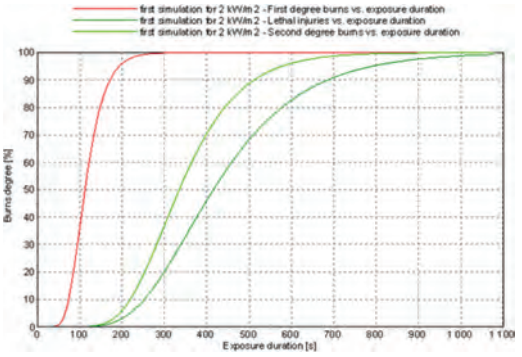


Figure 4. Burns degree in percent to exposure duration for 2 kW/m^2.

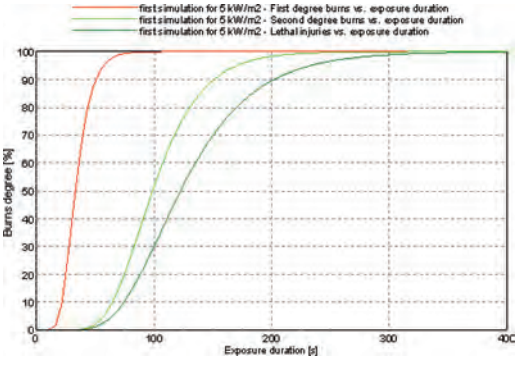


Figure 5. Burns degree in percent to exposure duration for 5 kW/m^2.

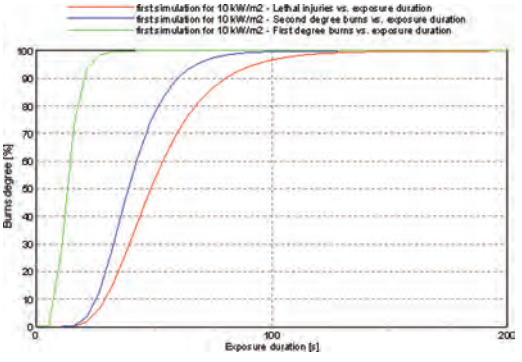


Figure 6. Burns degree in percent to exposure duration for 10 kW/m^2.

In all cases increase in exposure time generates higher burns degree. According to the figures, people exposed to the 10 kW/m^2 are more likely to get severe damages in shorter time (lethal injuries after 16 s) than people exposed to lower radiation heat (for 2 kW/m^2 lethal injuries after 130 s).

The graphs 4–6 depict that increase in first degree burn grown quicker than other injuries. The saltatory change of body damage appears in a different values for each simulation, the first degree burn curve saturate soonest and the lethal injuries curve has gradual transition and gain the plateau much later.

For 2 kW/m^2 the first degree burns plateau is gained after 328 s, the second degree burns after 900 s and the lethal injuries after 1080 s.

For 5 kW/m^2 the first degree burns plateau is gained after 98 s, the second degree burns after 280 s and the lethal injuries after 370 s.

For 10 kW/m^2 the first degree burns plateau is gained after 50 s, the second degree burns after 100 s and the lethal injuries after 125 s.

5 SECOND SIMULATIONS

The second simulation shows explosion of the entire fuel terminal and consequences for people. The average weight for man is 80 kg, and for woman – 55 kg. The calculations are based on Netherlands organization for applied scientific research (TNO) methodology (Committee for the Prevention…, 1989). The probit function for lung damage:

$$Probit = 5.0 - 5.74 * \ln S \tag{2}$$

$$S = \frac{4.2}{\bar{p}} + \frac{1.3}{\bar{i}} \tag{3}$$

$$\bar{P} = \frac{P}{P_0} \tag{4}$$

$$\bar{i} = \frac{i}{P_0^{1/2} * m^{1/3}} \tag{5}$$

where: P_0–atmospheric pressure, P–pressure exerted on the body, m–mass of the body.

The probit function for hearing damage:

$$Probit = -12.6 + 1.524 * \ln P_s \tag{6}$$

where: P_s–peak overpressure in the incident.

The probit function for the impact with the whole body:

$$Probit = 5 - 2.44 * \ln S \tag{7}$$

$$S = \frac{7.38 * 10^3}{P_s} + \frac{1.3 * 10^9}{P_s * i_s} \tag{8}$$

where: i_s–impulse of the incident pressure.

Figure 7 represents dependence of percentage of body damage and pressure in bars for average man, whereas Figure 8 depicts dependence of percentage of body damage for average woman. What is more, Table 1. Introduces comparison of body damage for average man and woman. Based

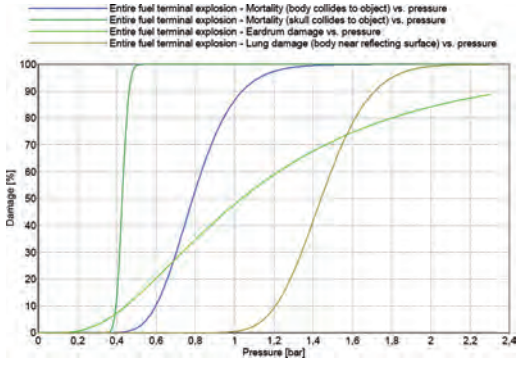


Figure 7. Percentage of body damage and pressure for average man.

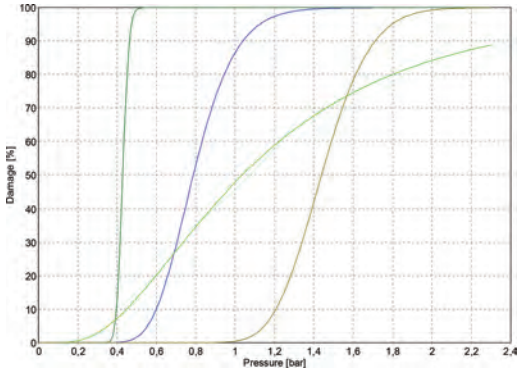


Figure 8. Percentage of body damage and pressure for average woman.

Table 1. Comparison of body damage caused by pressure for average man and woman.

Body damage	Man	Woman
Lung damage when body is perpendicular to blast direction (%)	18,88	19,2
Lung damage when body near reflecting surface (%)	99,97	100
Whole body displacement (when body collides to object) (%)	100	100
Whole body displacement (when skull collides to object) (%)	100	100
Eardrum damage (%)	88,78	88,78

on information in Table 1. We could assume, that potential explosion would have a similar effect on damage to the bodies of a woman and a man, there would be a slight discrepancy in the lung damage.

6 CONCLUSIONS

The purpose of the paper was to introduce the potential results of a fuel terminal failure. Choosing the Buncefield fire incident allowed to establish the correct mathematical model for a simulation of this type. The overriding goal was to identify potential consequences for people who were exposed to radiation heat generated in the event of an explosion of the entire storage base. Detailed analysis of the data and implementation of the risk analysis allowed the authors to view the issue of storage of dangerous substances in a tangible way.

It should be mentioned, that based on available data, authors were not able to define all parameters needed for the simulation, nor to fully anticipate the failure mechanism. Even though, it is valid to make such analysis and computer simulations to help us better understand the nature of the production process.

Analyzed simulations show, that humans near the centre of explosion would be suffer serious injuries or even death. What is more, there is only slight difference between percentage of body damage sustained by average man and woman exposed to pressure in the centre of the explosion. Probable cause is hidden in program limitation, because weight was the only factor that could be set to distinguish between man and woman. Such parameters as height or physique are not included and could potentially influence the results of simulation.

REFERENCES

BP, Statistical Review of World Energy June 2016, edition 65-th, [online cit.: 2017-05-25], Available from: https://www.bp.com/content/dam/bp/pdf/energy-economics/statistical-review-2016/bp-statistical-review-of-world-energy-2016-full-report.pdf.

Buncefield Major Incident Investigation Board, The final report of the Major Incident Investigation Board Volume 2, ISBN 978-0-7176-6318-7.

Control of Major Accident Hazards, Buncefield: Why did it happen?, 2011, [online cit.: 2017-05-04], Available from: http://www.hse.gov.uk/.

Krawczyszyn P., Ksiądzyna K., Ryczyński J., 2017. Simulating an industrial accident based on the example of gasoline terminal, *Proceedings of 21th International Scientific Conference*. Transport Means, p. 708.

Major Incident Investigation Board 2006. Buncefield Incident 11 December 2005, p. 5–12, ISBN 978-0-7176-6270-8.

Mannan M. S., O'Connnor M. K., A Technical Analysis of the Buncefield explosion and fire, 2009, Symposium series no. 155 IChemE, Hazards XXI, [online cit.: 29-04-2017], Available from: http://www.icheme-oncampus.org.

Ryng M., Technical Safety in the Chemistry Industry, Warsaw 1989, p. 282, ISBN 83-204-0664-1.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Analysis on factors of subway incidents for signal system maintenance improving based on a hybrid model

S. Zhang, T. Tang & R. Niu
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

F. Yan
National Engineering Research Center of Rail Transportation Operation and Control System, Beijing Jiaotong University, Beijing, China

L. Yue & L. Wan
Communication and Signaling Branch affiliated with Beijing Mass Transit Railway Operation Corp. Ltd., Beijing, China

ABSTRACT: Signal system is widely used to improve the safety and efficiency of subway system, while signal failure of it may lead to a huge breakdown of subway capacity. Most signaling operation and maintenance team still use the outdated checklists while signal operation and maintenance is proved to be a more and more difficult task because the internal and external interactions. This paper proposes hybrid accident/incident causal model of bowtie and cybernetic control loop to guide the design of incident data model. After that, the database is demonstrated on Beijing subway line No. 5. Based on the analysis, statistical characteristics of incidents are concluded. The trend of these factors could help companies improve the performance of subway system and the process of this analysis give them a new manner to record and learn from incident data.

1 INTRODUCTION

Many big cities rely on subway to solve the traffic problem, including Beijing. There are now 21 million people living in this city and 6 of the 18 subway lines have a passenger volume for over one million every working day (Beijing Mass Transit Railway Corp. Ltd., 2017). Signal system is widely used for the safety and efficiency in subway. While the advanced block mode and automatic controls provided by signal system can achieve shorter train interval that makes it possible for more trains operating at the same time, the failure of this system may lead to a breakdown of subway capacity.

An event with a single over-5-minute-delay train is classified as an operation accident by Beijing subway operation cooperation. Statistic shows that there were 36 operation accidents caused by signal failure of Beijing subway in the year 2016, which caused great social repercussions. For example, a failure of track circuit of Beijing subway Line No. 5 happened to January 26, 2016. Before the trouble removal 37 minutes later, the subway had to operate by telephone block, which caused 53 trains' over-5-minute-delay. So it is of vital importance to reduce the signal failure and control the loss during recovery period.

Learning from incident data is a way to help subway companies improve system performance and prevent an accident. While operation accidents seldom happened and learning from old accidents cannot catch up the operation state of the system, it is a good choice to reduce accident by analyzing minor incidents and minimizing the number of it according to Heinrich's Law (Jehring, 1959). There are papers that analyze subway incident with data. Zhang et al. (2016) builds a database to collect and analyze time, location, severity, causes and staff data of near misses in Shanghai subway. Ding et al. (2017) uses fault log database for safety management of subway. The data model of these databases is built based on existing incident record provided by subway companies and no safety science theory was mentioned to explain why they use this data and what other data could also be collected. Moreover, these databases are designed for the whole subway operation, concerning about aspects like fire, passenger fall downs, explosions, train collisions, vehicle derailments. And there is no database that illustrates the prevention and recovery of signal failure.

Signal operation and maintenance are proved to be a more and more difficult task not only because electronic equipment is sensitive to sever environment but because the interactions among signaling systems, surrounding equipment and human operators become much more complex. However, most signaling maintenance teams still use the outdated checklists, and only inspects and records the parameters of signaling itself. The lack of records of the working environment, related interfaces and operations makes it difficult to do incident cause-tracing and maintenance dynamic adjustment.

Accident/incident models provide insight to analysis the causes of an accident/incident. This paper proposes a signal system operation and maintenance database based on hybrid accident/incident causal model of bow-tie and control loop, in which the former provides an overview of what process should be under controlled to keep the performance of the system and mitigate the loss and the later explore how the control processes failed. Cooperated with Communication and Signaling Branch affiliated with Beijing Mass Transit Railway Operation Corp. Ltd., this research is based on real subway operation data. By the approach of building a signal-related incident database, this analysis is aimed at helping reduce the failure rate and improved recovery efficiency of signal system in the daily operation and maintenance of Beijing subway.

The structure of this paper is as follows. The literature review is in Section 2. Section 3 describes the proposed model and the demonstration of this data model in Beijing subway is in Section 4. Finally, Section 5 is the discussion and Section 5 is the conclusion of this paper.

2 LITERATURE REVIEW

2.1 *Bow-tie model*

Bowtie models demonstrate accidents and incidents as cause consequence diagrams. Nielsen (1971) invents the earliest bowtie by combining a fault tree and event tree into one diagram, in which the fault tree explains the cause and the event tree explains how the consequence happens. A top event (also called critical event) is put in the center of this kind of diagram while elements in the left end are causes and elements in the right end are called the consequences. Haddon (1973) put the concept of barrier to keep hazards from impacting a target, after which Reason (1990) use barriers as layers in his Swiss Cheese Model to explain the weakness in a system. Barrier then becomes an important part in bowtie models to explain the linear causal relationship between top event and causes or consequences. Indeed, fault tree and event are often abandoned in many bowtie models to be more suitable for qualitative analysis (Visser 1998, Zuijderduijn 1999, Ruijter et al. 2016). Fig. 1 displays a basic structure of Bowtie model. The barriers at the left of the critical event are control barriers and these at the right side are recover barriers (Visser 1998).

Control barriers prevent the cause from occurring as well as prevent the cause from leading to a top event. Recover barriers prevent the top event from leading to an undesired loss and mitigate the loss.

A bowtie model combines a cause consequence diagram with barriers, which means linking barriers of all operation process in the system. As is widely used for decision making purpose (Ruijter et al. 2016, Visser 1998, Rathnayaka et al. 2014), bowtie models can model the operation of an integrated system and provide an overview of what varieties of barriers should be placed.

2.2 *Control loop*

Control loop is a form of information flow diagram in the system, which aimed to ensure the outcome of a process can follow a specified target or setpoint. The basic structure of a control loop is shown in Fig. 2.

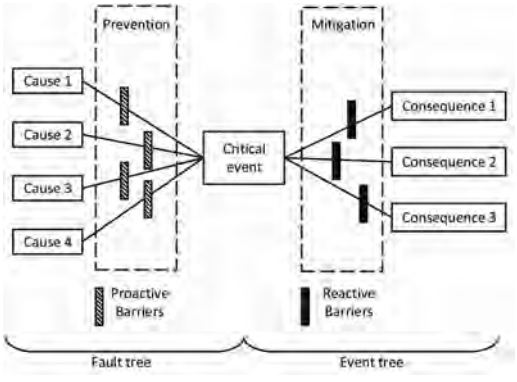


Figure 1. A basic structure of bowtie model.

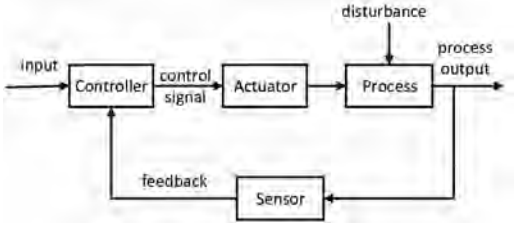


Figure 2. A basic structure of control loop.

Kuhlmann (1986) states that safety can be understood as a cybernetic problem because of the interactions among machine, human and environment. Behavioral cybernetic theory. Smith, K.U. (1972) and Smith, T.J. (1987) maintains that human behavior is guided as a closed-loop, feedback controlled process. Cybernetics control loops are used to analyze the performance of complex, socio-technical systems in which individuals and organizations operating together with advanced technology systems (Smith, T.J. et al., 1995). Sonnemans (2006) uses the control loops in accident analysis of the chemical industry to exam the control mechanism inside organizations. Dijkstra (2007) uses a cybernetics management theory in aviation to model the organizational structure of communication which should be able to adapt the changing environment. Kontogiannis (2012) uses control loops based on VSM to look into general patterns of breakdown related to structural vulnerabilities and gradual degradation of performance in an accident from a Helicopter Emergency Medical Service.

The control loops are also used in railway. Appicharla (2011) analyzes the system interfaces of railway with cybernetics loops because cybernetics studies organization, communication and control in complex systems. Kohda (2007) uses control loops in the accident analysis of railway protective systems. As can be seen, control loops can model a control process, providing details of subjects in the process and details of interactions among human, equipment and environment in a complex system, such as railway signal system.

3 METHOD PROPOSED

Beijing subway Line No. 5 is an important subway line in Beijing subway system, which contains 23 stations and 27.6 km mileages, providing over one million passenger trips every working day. While Line No. 5 is the 4th busiest line in Beijing, it takes the top in signal system failure records and in operation delay in incident records in the past 4 years. From the 2016, RCS lab of Beijing Jiaotong University has been working with Beijing Subway to improve the performance of the signal system and subway Line No. 5 is chosen to be the research subject in this paper.

3.1 *A hybrid data modeling method proposed*

Bowtie and control loop are combined into a hybrid model in this paper.

Signal system can guarantee the shortest train interval under the normal operation mode, while signal degradation resulting from signal failure could cause a traffic jam on track. The signal-related incidents have different causes and different degree of loss but all of them have an event that the normal operation mode of signal system failed. The responsibility of operators and maintainers is to keep signal system working under the normal mode and to recover the operation if a signal failure happens. A bowtie model divides an incidents into causation and effect and signal failure is put in the middle of it. Then the operation and maintenance tasks can be recognized as barriers besides the top event. Bowtie models can be valuable at high level to guarantee all various kinds of data is taken into consideration for the target incident data model.

Signal system is a complex social-technical system and the operation and maintenance can be influenced by the complex interactions among human, equipment and environment. While the bowtie model is built with high abstraction level and less specific information, the management of barriers takes the view from the whole picture to more detail of the system. Thus control loops can be used as a model for the management of barriers could tell what detail data should be collected or recorded for the database.

The hybrid model can be established by the following two steps:

Step 1 Establish a bowtie model based on practical subway operation and maintenance situation and gain the barriers.
Step 2 Establish a control loop specific for every barrier gained in the bowtie model.

The steps are carried out as follows.

3.2 *Establish the bow-tie model*

A qualitative bowtie model is used in this hybrid model, in which there is no fault tree or event tree. Various types of signal failure of normal mode are placed at the middle of the model as top events. Component failures of the system are placed at the left end of the bowtie as the cause of the top event while the loss is placed at the right end of this model. Then there should be a way to determine the barriers of the bowtie model.

RAMS concept is widely used in mainline railway and subway domain, which is defined in terms of reliability, availability, maintainability and safety and their interaction (CENELEC 1999). It provides a guide to assess and improve the quality of service. The RAMS is influenced in three ways, which are system conditions, operating conditions and maintenance conditions. Thus the three conditions can be used in the bowtie model to determine the categories of barriers of subway signal system.

These three conditions can be directly put on the left side of the top event to control the causes and can also be used in the right side of the top event for recovering within more specific scenarios. Thus, the six barriers include system conditions, maintenance conditions and operating conditions as well as degradation conditions, emergency maintenance conditions and operational response conditions.

The system conditions barriers focus on the maintainability and technical characteristics of the system, which are the inherent attribute of the system, as well as the internal and external disturbance of the system.

Maintenance conditions barriers include maintenance procedures such as the conditions of preventive maintenance and corrective maintenance, logistics, and also human factors in the maintenance.

Operating conditions barriers are a question of the operation parameter of the system before the signal failure. Factors like operation time, train interval and passenger flow volume can affect the system and human factors like driving manners and dispatchers' intervention can also contribute to signal failure.

Degradation conditions describe the system conditions without the normal operation mode. The trains have different driving modes and the ground system has different block modes. When the signal failure happens, it is important that a backup mode is ready for operation.

Emergency maintenance conditions are corresponding to maintenance conditions. After a failure happens, it is important to report, diagnose, and repair the system as quickly as possible. Enough well-trained maintainer in the nearby and sufficient spare parts help a lot.

Operational response conditions are corresponding to operation conditions while emphasize the recovery operation procedure. The operational response under the command of dispatchers should decide and carry out methods like degradation, making trains offline, canceling trains and detain trains at stations.

The cause consequence diagram and the 6 categories of barriers are combined and the bowtie part of the hybrid model is shown in Fig. 3.

3.3 *Establish the control loops diagram*

There is no unified control loops diagram for the six categories of barriers. The control loops diagram can be different for different systems and the complexity of these loops depends on how far the analysis is intended to go. The six categories of barriers are unfolded by control loops to reveal the detail factors.

System conditions and degradation conditions mainly talked about how the complex technical system works in normal mode and degradation mode. A control loops diagram that shows the entire conditions of signal system of Beijing subway Line No. 5 is shown in Fig. 4, including the control loops of normal system conditions and degradation conditions.

The control loops diagram of the rest four barriers are more like mission based control. As the

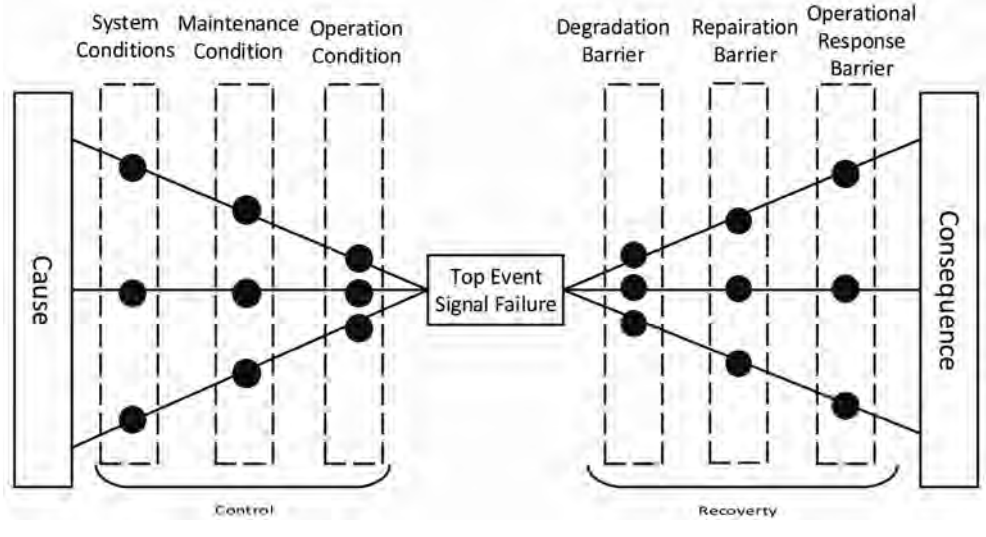


Figure 3. The bowtie part of the hybrid model.

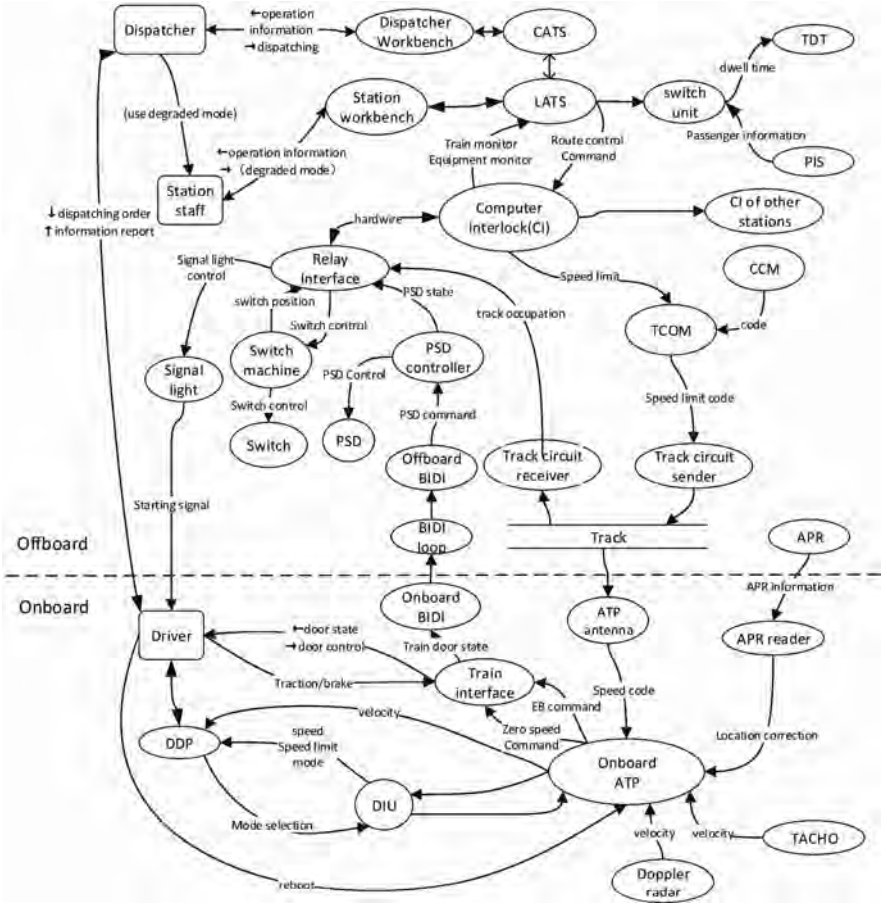


Figure 4. Control loops of system conditions.

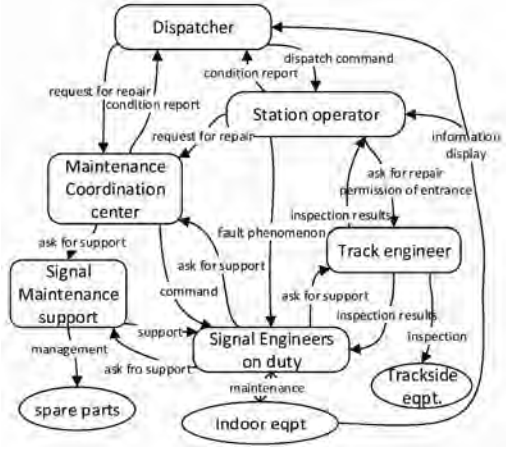


Figure 5. Control loops of EMER MAINT conditions.

control loops diagram of emergency maintenance conditions shown in Fig. 5, person, group and technical system work together to make the system operating.

Each control loop has three basic elements, which are human/equipment, interaction arrow and disturbance. Data can be obtained directly from these three elements.

4 DEMONSTRATION

4.1 *Data model*

Base on the hybrid model proposed in Section 3, a signal-related incident data model for Beijing subway Line No. 5 is built, which is shown in Fig. 6. There are 16 tables in this data model, in which the type of 4 tables are defined by the cause consequence diagram and the type of 6 tables are

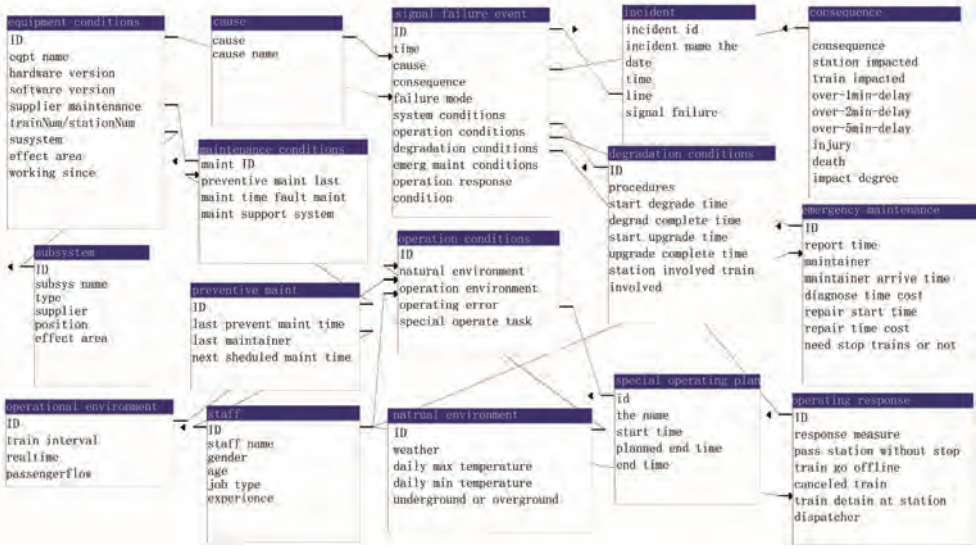


Figure 6. The data model for signal-related subway incidents.

defined by the 6 barriers categories, while the rest tables and the join between tables as well as fields in all these tables are explored in control loops. The data model can direct what kinds of data are needed for the incident data analysis.

Based on the data model, the signal-related subway incident database is built with Microsoft Access 2013.

4.2 *Data collection*

The data analysis in this paper is from the combination of varieties of existing records by Beijing subway, which includes operation accident reports, signal-related incidents records, signal system maintenance records and equipment maintenance manual. Base on the files above, a FMEA analysis is carried out and 65 usual subsystem level failure modes that affect the normal operation mode are found. The outcome is used to rebuilt the incident records, classifying the incident records firstly by the structure of system, subsystems and equipment, secondly by the subsystem failure mode, equipment failure mode and thirdly by reasons including hardware failure, software failure, human error and environment. The subsystem failure modes are defined to be the top events and the causes and consequences are also recognized.

Then files of other department are collected and attached. Dispatching control records are gathered from dispatching center so the dispatching information of every incidents are stored in the database. Then line arrangement figure is used to locate the underground and ground part of the line. Subway dispatching timetable is used to tell the train interval during the signal failure. What's more, weather data is downloaded from the website of China Meteorological Administration.

A total of 4352 incident records of Beijing subway signal related incidents from 2013 are input into the database, 69.5 percent of which have not a single 1 min delay while 0.67 percent have an effect of over-5 min-delays.

4.3 *Findings*

4.3.1 *Analysis based on RAMS*

The frequency of subsystem failure is shown in Fig. 7, in which 85 percent of subsystem failures happened in onboard system. Some factors are found with the analysis of the data.

The data analysis shows that the reliability, availability and maintainability of onboard system of Beijing subway Line No. 5 is low, for which some technical characteristics of the system should be responsible. In terms the reliability of the onboard system, the frequency of inner communication error and of APR reader failure are very high. And the structure of onboard ATP is 2 out of 2, which means no redundancy and a single failure will lead to an emergency break. In terms of availability, there is no efficient backup mode like there are in other lines. If the normal mode fails, the trains have to be operated under the instruction of signal light, which cannot deal with the train interval which is so short. In terms of maintainability, firstly there

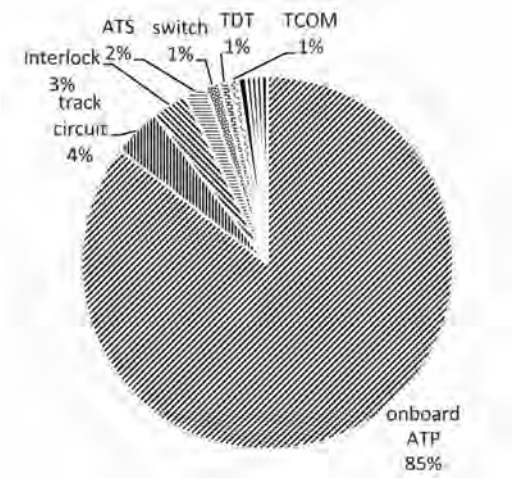


Figure 7. Frequency of incidents by subsystems.

is no assist information about what is wrong after an onboard system failure except an alarm light. So the drivers and dispatchers cannot take actions according to the specific failure and can only use the same procedure. Secondly, the signal supplier is now broken, and the reasons of many failures are too hard to figure out.

4.3.2 *Analysis on environment disturbance*

Track circuit is used in Beijing subway line No. 5 to send MA (movement authority). And MA is coded in low frequency and carrier frequency. In 2016 and the first half year of 2017, there are 64 emergency breaks that the fault code is 'LF error' or 'CF error'. So analysis is carried out to find if there are some environmental disturbances that affect this failure.

So when the LF/CF errors are compared with the location, as is shown in Fig. 8, it could be found that the frequency of these errors increased in track near Chongwenmen and Dongdan station. So the signal maintainers inspect these sections and found that the LF/CF might be disturbed by the traction return current circuit, which was maintained in early 2016.

4.3.3 *Analysis on human factors of maintainers*

Despite used for logic control for a long time, now relays are widely used in computer-based railway signal system in order to lead the system to the safety side after component failures. After the increment in the number of failures as a result of the shortening of train interval in 2014, Beijing subway increased the frequency of scheduled maintenance on the whole signal system in the late of that year.

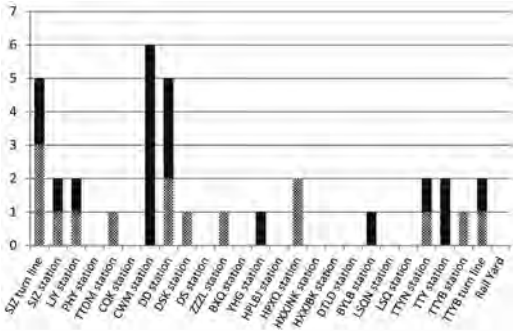


Figure 8. Track circuit code error.

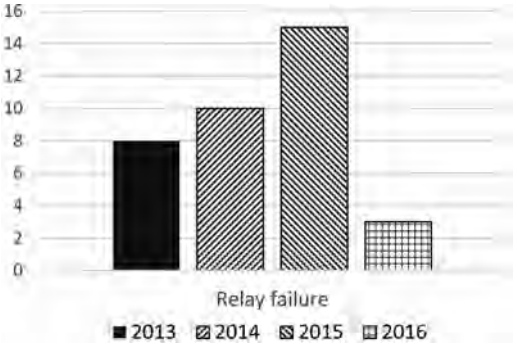


Figure 9. The trend of relay failure over years.

However, as can be seen in Fig. 9, the data shows that relay failures increased in the year 2015 and many of them were found to be or could be related to human errors. It seems that it is not a good choice to keep a high maintenance rate for every component. Relays, for example, have a large amount and every disassembling and assembling means potential of human errors. Then the maintenance frequency of relays was befittingly reduced and more importance are put on the quality of it and the failure rate reduced in 2016.

4.3.4 *Analysis on human factors of operators*

Since the reliability and maintainability of onboard signal system of Line No. 5 is not good, onboard systems often suffer from random failures and the drivers cannot read what is wrong. So there are two choices for the dispatchers to deal with the failure train before its recovery, one is to degrade the bock mode to station block, which is used as a standardized treatment but need time to the enable of signal lights, and the other is to keep the normal mode and guide the train by dispatcher themselves, which relay on dispatchers' efficiency.

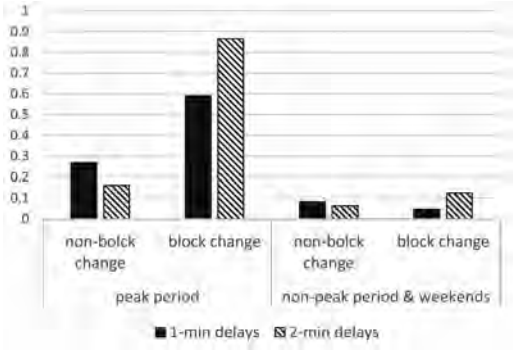


Figure 10. Average delays per onboard failure.

Both the choices are used and it often depends on the dispatchers' experience and preference.

Fig. 10 shows the average 1-min delays and 2-min delays in under different conditions. The data shows that it is related to the train interval. The signal failures with block changes have higher average delays in peak period, but the difference is not so significant during non-peak period and weekends. So it is recommended to the subway company that it is a better choice not to use block change for a single onboard failure in peak period but related regulation must be assessed to guarantee the safety.

5 DISCUSSION AND CONCLUSION

The hybrid model proposed in this paper provides a manner for maintainer and operator to gather and use data to analyze incidents and prevent operation accident that related to signal failures. The bowtie provides an overview and in a way can ensure the comprehensiveness of the categories of data. And the control loops can directly provide varieties of data in detail in a systematic way. And the two part are well bonded because control loops can describe the management of barriers in bowtie. The hybrid model is also costume to subway domains by using the concept of RAMS, which has been used as a guide in this industry for a long time. So there is no difficult for subway engineers to learn and their experience can be inherited in the structure of the database.

In this paper, the data model built from the hybrid model replaces the outdated checklist of Beijing subway and treats the signal system in the way of complex, social-technical system. The incident data analysis can help signal maintainer find the environmental disturbances of signal failure. And it can also help signal maintainer find the human factors that affect the system maintenance so as to adjust their manner of working. Operators can carry out optimistic changes to the operation manners in recovery period. The data analysis is proved to be useful for the maintenance and operation of subway signal system. What's more, the findings and changes can also be checked in future data.

The data model and data collection in this paper play four roles in helping record and collect data. Firstly, it can help subway companies to make a formal manner for signal-related incident data recording since recorders have different habit for recording incidents. Secondly, maintainers and operators record comprehensive have more comprehensive data to use that their analysis can be carried out immediately when they want to prove some measure can bring improvement the performance of signal system, but not starting from recording new data. That can improve the efficiency for them to acquaint the system. Thirdly, existing data collected by different department in a same subway company can be put together to analyze the interaction problems among them. And the environmental data recorded by other departments can be used to analyze and eliminate external disturbances. Furthermore, the model provides a way to analysis what kinds of data are valuable though they are hard to record so sensors could be bought or developed to record more system conditions data as well as advanced information system could be used to record human behaviors.

This paper proposed a hybrid model to recognize what kind of data should be recorded for maintainers and operators to improve the performance of subway signal system. Designing and establishing of this hybrid model are introduced in detailed. After that, the database is demonstrated for Beijing subway line No. 5. Based on the analysis, statistical characteristics of incidents are concluded. The trend of these factors could help companies improve the performance of subway system and the process of this analysis give them a new manner to record and learn from incident data.

In future research, an incident data analysis method could be proposed base on this hybrid model. The method may guide people or even computers to conduct data analysis.

ACKNOWLEDGEMENTS

This work was supported by the Fundamental Research Funds for the Central Universities under 2015 JBZ006, the Project U1434209 supported by National Natural Science Foundation of China.

REFERENCES

Appicharla, S.K. 2011. System for Investigation of Railway Interfaces. In proceedings of *IET International Conference on System Safety* (Vol. 32, pp. 7–16). IET.

Beijing Mass Transit Railway Corp. Ltd. 2017. Beijing subway passenger flow statistics. Available at: http://www.bjsubway.com/support/cxyd/klxx/.

CENELEC, 1999. Railway applications—The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS).

Dijkstra, A. 2007. *Resilience engineering and safety management systems in aviation*. KLM Royal Dutch Airlines/TU Delft.

Ding, X., Yang, X., Hu, H., & Liu, Z. 2017. The safety management of urban rail transit based on operation fault log. *Safety Science*, 94, 10–16.

Haddon, W., 1973. Energy damage and the 10 countermeasure strategies. *Injury Prevention*, 1(1), 40–44.

Jehring, J. 1959. Industrial accident prevention: a scientific approach, by h. w. heinrich. *Industrial & Labor Relations Review*, 4(4), 609–609.

Kohda, T. 2007. Accident Analysis of Protective Systems Based on System Control Concepts. *Reliability and Maintainability Symposium*, 2007. Rams '07 (pp. 414–419). IEEE.

Kontogiannis, T., & Malakis, S. 2012. A systemic analysis of patterns of organizational breakdowns in accidents: A case from Helicopter Emergency Medical Service (HEMS) operations. *Reliability Engineering & System Safety*, 99, 193–208.

Nielsen, D.S., 1971. *The Cause/Consequence Diagram Method as a Basis for Quantitative Accident Analysis*. Tech. Rep., Danish Atomic Energy Commission, Roskilde, Denmark.

Rathnayaka, S., Khan, F., & Amyotte, P. 2014. Risk-based process plant design considering inherent safety. *Safety Science*, 70, 438–464.

Reason, J.T., 1990. Human Error. Cambridge University Press.

Ruijter, A.D., & Guldenmund, F. 2016. The bowtie method: a review. *Safety Science*, 88, 211–218.

Smith, T.J., Henning, R.A., & Smith, K.U. 1995. Performance of hybrid automated systems—A social cybernetic analysis. *International Journal of Human Factors in Manufacturing*, 5(1), 29–51.

Smith, T.J. 1999. Synergism of ergonomics, safety, and quality—a behavioral cybernetic analysis. *International Journal of Occupational Safety & Ergonomics Jose*, 5(2), 247–278.

Smith, K.U. 1972. Cybernetic psychology. In *R.N. Singer (Ed.), The psychomotor domain* (pp. 285–348). New York: Lea and Febiger.

Sonnemans, P.J., & Körvers, P.M. 2006. Accidents in the chemical industry: are they foreseeable? *Journal of Loss Prevention in the Process Industries*, 19(1), 1–12.

Visser, J., 1998. Developments in HSE management in oil and gas exploration and production. In: *Hale, A., Baram, M. (Eds.), Safety Management: The Challenge of Change*. Pergamon, Oxford, UK, pp. 43–65 (Chapter 3).

Zhang, X., Deng, Y., Li, Q., Skitmore, M., & Zhou, Z. (2016). An incident database for improving metro safety: the case of shanghai. *Safety Science*, 84, 88–96.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

An investigation identifying trends between the enforcement of offshore safety case regulations and the occurrence of vessel to platform collision incidents

S. Loughney, J. Wang & B. Matellini
Liverpool John Moores University, Liverpool, UK

K. Pemberton
Health and Safety Executive, Bootle, Liverpool, UK

ABSTRACT: The regulatory beginnings of the modern offshore Safety Case (SC) are demonstrated by the release of the Health and Safety at Work Act 1974, where it set up two new organisations to oversee its implementation: The Health and Safety Commission (HSC) (which was dissolved in 2008) and the Health and Safety Executive (HSE). Following the public inquiry into the Piper Alpha disaster, the responsibilities for offshore safety regulations were transferred from the Department of Energy to the HSC through the HSE as the singular regulatory body for safety in the offshore industry (Wang, 2002) (Department of Energy, 1990). In response to this the HSE launched a review of all safety legislation and subsequently implemented changes. The propositions sought to replace the legislations that were seen as prescriptive to a more "goal setting" approach. Since these events the SC regulations of 1992 have seen several amendments in both 2005 and 2015, as well as the release of further regulations, such as: The Offshore Installations Prevention of Fire and Explosion, and Emergency Response (PFEER) in 1995 and the Safety Zones around Oil & Gas Installations in Waters around the UK in 2008. However, while Safety Cases are subject to review and updating as often as is required to keep them up to date, the process of change to the Safety Case may be slow and gives a monolithic appearance to the document. Subsequently, there have been several papers suggesting that a dynamic risk assessment method should be utilised to assist with SC regulation enforcement. This has led to the investigation presented in this report where 508 vessel to platform collision incidents between 1971 and 2015 have been analysed and compared with the release of key offshore SC regulations. The incidents have been sourced from the HSE, World Offshore Accident Databank (WOAD) and the MAIB. This analysis has identified a key trend between the reporting of offshore collision incidents and the release and enforcement of offshore SC regulations.

1 INTRODUCTION

Following the public inquiry into the Piper Alpha disaster, the responsibilities for offshore safety regulations were transferred from the Department of Energy to the Health and Safety Commission (HSC) through the Health and Safety Executive (HSE) as the singular regulatory body for safety in the offshore industry (Wang, 2002) (Department of Energy, 1990). In response to this the HSE launched a review of all safety legislation and subsequently implemented changes. The propositions sought to replace the legislations that were seen as prescriptive to a more "goal setting" approach. Several regulations were produced, with the mainstay being the Health and Safety at Work Act (HSE, 1992). Under this a draft of the offshore installations safety case regulations was produced. The regulations required operational safety cases to be prepared for all offshore installations. A SC is to be submitted by the "operator" for fixed installations and by the "owner" for mobile installations. Within this all new production installations require a design safety case and for mobile installations, the duty holder is the owner (Wang, 2002).

Offshore operators must submit operational safety cases (SC) for all existing and new offshore installations to the Health and Safety Executive's (HSE) Energy Division (formerly the Offshore Safety Division) for acceptance, but not approval, yet it is an offence to operate without an approved SC (HSE, 2006a). The SC must show that it identifies the hazards with potential to produce a serious accident and that these hazards are below a tolerability limit and have been reduced to the ALARP Level (As Low As Reasonably Practicable) (Wang, 2002).

Safety and risk assessment for offshore installations is vigorous and requires demonstration

from duty holders that all hazards with potential to cause major accident are identified, all major risks have been evaluated, and measures have been or will be taken to control the major accident risks to ensure compliance with the statutory provisions (HSE, 2006b). Furthermore, it used to be a requirement stipulated in the SC Regulations that suitable and sufficient quantitative risk assessment was undertaken. However, the 2015 regulations have actually removed the requirement for this specific form of risk assessment though in practice most duty holders still use QRA extensively, it should however, only be used as an evaluation tool (HSE, 2015).

This is vitally important as accidents in the offshore industry lead to devastating consequences, such as the explosion on board the Deepwater Horizon rig in the Gulf of Mexico which was caused by the failure of a subsea blowout preventer (BOP), with some failures thought to have occurred before the blowout (Jones, 2010). This solidifies the use of quantitative risk and reliability analysis, current ideas behind these models can perform predictive analysis and diagnostic analysis (Cai, *et al*., 2013).

After many years of employing the safety case approach in the UK offshore industry, the regulations were expanded in 1996 to include verification of Safety Critical Elements (SCE). Also, the offshore installations and wells regulations were introduced to deal with various stages of the life cycle of the installation. SCEs are parts of an installation and its plant, including computer programs or any part whose failure could cause or contribute substantially to or whose purpose of which is to prevent or limit the effect of a major accident (Wang, 2002) (HSE, 1996).

Recently, however, it is felt that an expansion on Safety Cases is necessary, especially in the offshore and marine industry, as they are static documents that are produced at the inception of offshore installations, despite containing a structured argument demonstrating that the evidence contained therein is sufficient to show that the system is safe (Auld, 2013) (Eleye-Datubo, A., *et al*., 2006).

While the authors were conducting research regarding dynamic risk assessment techniques and methods for offshore installations and SCEs, more specifically data gathering regarding offshore ship to platform collisions, a periodic pattern emerged between the release of SC regulations and the number of collision incidents on the UKCS.

2 BACKGROUND

It is known that Safety Cases are subject to review and updating as often as is required to keep them up to date, the process of change to the Safety Case may be slow and gives a monolithic appearance to the document. A SC is a relatively high level document which identifies the operator/owner, describes the installation and its layout, its environmental limits, the types of operation to be undertaken and how the health and safety aspects will be managed, and the maximum number of persons expected to be on the installation at any one time. Similarly, there also needs to be a description for the arrangements for detecting toxic or flammable gas and the detection, prevention or mitigation of fires and the arrangements for protecting people from the hazards of explosion, fire, heat, smoke *etc*. usually in the form of a temporary refuge, egress routes means of evacuation and means of monitoring and control for an incident. There should also be a demonstration by suitable risk assessment that the risks have been mitigated to a level that is ALARP and a description of the verification scheme.

An attempt to identify trends with safety case regulation and incidents regarding offshore SCEs, such as gas drive turbines and generators, was conducted. However, due to possible under reporting or the availability of data, it is difficult to demonstrate the trend of some SCEs, with the updating of offshore regulations. On the other hand, it is possible to demonstrate the effect of a number of possible influences, (slow updating or enforcement of regulations, and under reporting or improper recording of incidents), and the effects they may have on incidents on-board offshore platforms. A key area that can be assessed is the issue of ship to platform collisions.

Before any data is presented, it is important to understand the timeline of key offshore regulations and incidents that have shaped the modern-day safety case regulations. The following list identifies the key timeline of incidents that have built the current safety case regulations.

- *1974 Health & Safety at Work Act (HSWA)*

The HSWA adopted a holistic approach to the health, safety and welfare of workers. The Act focuses on the concept that any situations that may give rise to harm need to be recognised and suitable measures put in place to eliminate or reduce the potential for harm. It set up two new organisations to oversee its implementation: The Health and Safety Commission (HSC) and the Health and Safety Executive (HSE). The HSE is the executive organisation that enforces the provisions of the HSWA. However, in April 2008 the HSC was dissolved and merged with the HSE. The HSC used to protect health and safety at work in the UK by conducting research, training and providing advice and information. The Commission also used to propose new regulations and approved codes of practice under the authority of the Act. This is

all now conducted by the HSE (Inge, 2007) (The Stationary Office, 1974).

- *1988 Piper alpha disaster*

Piper Alpha was an oil production platform in the North Sea off the coast of Aberdeen, Scotland. The platform began production in 1976, initially as an oil platform but was later converted to accommodate gas production. Oil & Gas fires and explosions destroyed Piper Alpha on 6 July 1988, killing 167 people, including two crewmen of a rescue vessel and 61 workers aboard survived. Thirty bodies were never recovered. The total insured loss was about £1.7 billion ($3.4 billion), making it one of the costliest manmade catastrophes ever. At the time of the disaster, the platform accounted for approximately ten per cent of North Sea oil and gas production. The Cullen Inquiry was set up in November 1988 to establish the cause of the disaster, chaired by Judge William Cullen. After 180 days of proceedings, the "Public Inquiry into the Piper Alpha Disaster" or "Cullen Report" was released in November 1990. It concluded that the initial condensate leak was the result of maintenance work being carried out simultaneously on a pump and related safety valve. The report was critical of Piper Alpha's operator, which was found guilty of having inadequate maintenance and safety procedures (Inge, 2007) (Oil & Gas UK, 2008).

- *1989 Offshore installations (safety representatives & safety committee) regulations*

The document provides information on interpretation and enforcement of the Offshore Installations (Safety Representatives and Safety Committees) Regulations 1989. These regulations were made under the Mineral Workings (Offshore Installations) Act 1971. They allow the workforce on an offshore installation to elect safety representatives from among themselves, and confers on those functions and powers in relation to the health and safety of the workforce. They also provide for time off with pay for safety representatives so they can perform these functions and undergo relevant training (The Stationery Office, 1989).

- *1990 The Cullen report*

The Cullen Inquiry was set up in November 1988 to establish the cause of the disaster, chaired by Judge William Cullen. After 180 days of proceedings, the "Public Inquiry into the Piper Alpha Disaster" or "Cullen Report" was released in November 1990. It concluded that the initial condensate leak was the result of maintenance work being carried out simultaneously on a pump and related safety valve. The report critical of Piper Alpha's operator, which was found guilty of having inadequate maintenance and safety procedures. 106 recommendations were made calling for, amongst many matters, the requirement of a SCs, the transference of the discharge of offshore regulation from the Department of Energy to a discrete division of the HSE. The responsibility of implementing the recommendations was spread across the regulators and the industry with, the HSE overseeing 57, the operators were responsible for 40, the offshore industry were given 8 to progress and the final one was for the Standby Ship Owners Association. The industry acted urgently to carry out the 48 recommendations that operators were directly responsible for. The HSE developed and implemented Lord Cullen's key recommendation: the introduction of safety regulations requiring the operator/owner of every fixed and mobile installation operating in UK waters to submit to the HSE, for their acceptance, a SC (Inge, 2007).

- *1992 Safety case regulations*

The Offshore Installations (Safety Case) Regulations came into force in 1992. By November 1993 a safety case for every installation had been submitted to the HSE and by November 1995 all had had their safety case accepted by the HSE. The Safety Case Regulations require the owner/operator/duty holder of every fixed and mobile installation operating in UK waters to submit to the HSE, for their acceptance, a safety case. The safety case must give full details of the arrangements for managing health and safety and controlling major accident hazards on the installation. It must demonstrate, for example, that the company has safety management systems in place, has identified risks and reduced them to as low as reasonably practicable, has introduced management controls, provided a temporary safe refuge on the installation and has made provisions for safe evacuation and rescue (Inge, 2007) (HSE, 2005).

- *1995 Offshore installations Prevention of Fire and Explosion, and Emergency Response (PFEER)*

PFEER deals primarily with fire and explosion events but it also deals with any event which may require emergency response and includes systems that may rely on radar on a standby vessel or responsible staff on the installation monitoring incoming vessels. The Regulations, ACOP and guidance deal with: (a) preventing fires and explosions, and protecting people from the effects of any which do occur; (b) securing effective response to emergencies affecting people on the installation or engaged in activities in connection with it, and which have the potential to require evacuation, escape and rescue (Amended in 2005 and 2015) (HSE, 2015a).

- *1996 Offshore installation (design & construction) regulations*

DCR requires the installation to possess integrity at all times, as is reasonably practicable. It requires

the design of the installation to withstand such forces that are reasonably foreseeable and in the event of foreseeable damage it will retain sufficient integrity to enable action to be taken to safeguard the health and safety of persons on or near it. The duty holder also has to record the appropriate limits within which it is to be operated. Further duties can be found in the Offshore Installations and Wells (Design and Construction, *etc*.) Regulations 1996.

- *2005 Offshore installations (safety case) regulations (April 2006)*

The Offshore Installations (Safety Case) Regulations 2005 came into force on 6 April 2006. They replace the previous 1992 Regulations. The primary aim of the Regulations is to reduce the risks from major accident hazards to the health and safety of the workforce employed on offshore installations or in connected activities. The Regulations implement the central recommendation of Lord Cullen's report on the public inquiry into the Piper Alpha disaster. This was that the operator or owner of every offshore installation should be required to prepare a safety case and submit it to HSE for acceptance (HSE, 2005). These SC regulations have been replaced by the 2015 regulations (HSE, 2015).

- *2008 Safety zones around oil & gas installations in waters around the UK (HSE)*

While this document is not a regulation, it explains the purpose and significance of safety zones around offshore oil and gas installations and their effect on marine activities, particularly relating to fishing vessels. A safety zone is an area extending 500 m from any part of offshore oil and gas installations and is established automatically around all installations which project above the sea at any state of the tide. Subsea installations may also have safety zones, created by statutory instrument, to protect them. These safety zones are a 500m radius from a central point. Vessels of all nations are required to respect them. It is an offence (under section 23 of the Petroleum Act 1987) to enter a safety zone except under the special circumstances. (HSE, 2008b)

- *2015 Offshore installations (offshore safety directive) (safety cases etc.) regulations (July 2015)*

The Offshore Installations (Offshore Safety Directive) (Safety Case *etc*.) Regulations 2015 came into force on 19 July 2015. They apply to oil and gas operations in external waters, that is, the territorial sea adjacent to Great Britain and any designated area within the United Kingdom Continental Shelf (UKCS). They replace the Offshore Installations (Safety Case) Regulations 2005 in these waters, subject to certain transitional arrangements (HSE, 2015b).

These key events from 1974 to the present day have shaped the modern SC into what it is today. After identifying these key regulations and incidents it was possible to plot these against the number of ship to platform collision incidents. This is where the data gathering and statistical analysis is conducted.

3 STATISTICAL ANALYSIS & DISCUSSION

The current database of ship to platform collisions provided by the HSE is out dated as it was last published in 2001. Similarly, the OGP (Oil & Gas Producers) produced a document in 2010 of worldwide collision statistics and Oil & Gas UK produced a document of accident statistics on the UKCS for offshore units between 1990 and 2007 (HSE, 2003) (OGP, 2010) (Oil & Gas UK, 2009). However, the OGP document provides only the frequency of collisions of incidents over key offshore and shipping areas around the world and the Oil & Gas UK document lists all accidents statistics with overall frequencies per year. These three documents are sufficient enough to demonstrate the trend between offshore collision incidents and offshore regulations. Therefore, a statistical analysis is conducted for ship to platform collisions from 1971–2015 across the UKCS.

For this study and incident has been defined as a reported impact or contact between a vessel and a fixed or mobile installation in terms of the Reporting of Injuries, Diseases and Dangerous Occurrences Regulations (RIDDOR 2013) database, which utilizes reported incident information from the OIR/9b and F2508A forms.

The 2001 ship/platform collision incident database has been further cross-checked and extended. The complete list of incidents demonstrates a total of 508 incidents of vessels impacting or contacting both fixed and floating offshore structures. These incidents have been reported from 1st January 1971 to 31st December 2014.

The data has been recorded from a number of sources. In many cases the data is supplemented or confirmed from additional sources. Data across the whole study has been compiled from the following sources:

- RIDDOR 2013, utilizing search criteria "Collisions, or potential collisions", between "vessels and offshore installations". Information source is labelled as HSE in the database.
- World Offshore Accident Databank (WOAD) using the search criteria ("Collision", "Offshore Units" and "Europe North Sea".
- Marine Accident Investigation Branch (MAIB) using the search criteria "Offshore installations", "collision" and "contact".

- Global Integrated Shipping Information System (GISIS) using search criteria “collisions” and “North Sea”.
- World Energy Related Casualties (WREC) using search criteria “offshore installations”, “collisions” and “North Sea”.

In order to demonstrate a coherent analysis, an incident frequency and a cumulative incident frequency has been calculated based upon the approximate level of operating experience per year. Table 1 demonstrates the operating experience, the number of incidents that have been reported and

Table 1. List of all reported collision incidents on the UKCS, as well as operating experience and incident frequencies.

Year	Total experience (N)	Cumulative experience (N1)	All incidents (r)	Cumulative incidents (r1)	Frequency (λ)	Cumulative frequency (λ1)
1971		–	1	–	–	–
1972	–	–	0	–	–	–
1973	–	–	1	–	–	–
1974	–	–	2	–	–	–
1975	89	89	12	12	0.135	0.135
1976	93	182	12	24	0.129	0.132
1977	102	284	20	44	0.196	0.155
1978	103	387	19	63	0.184	0.163
1979	107	494	25	88	0.234	0.178
1980	113	607	17	105	0.150	0.173
1981	121	728	29	134	0.240	0.184
1982	132	860	23	157	0.174	0.183
1983	142	1002	20	177	0.141	0.177
1984	165	1167	12	189	0.073	0.162
1985	177	1344	18	207	0.102	0.154
1986	169	1513	13	220	0.077	0.145
1987	174	1687	6	226	0.034	0.134
1988	195	1882	7	233	0.036	0.124
1989	210	2092	18	251	0.086	0.120
1990	262	2354	24	275	0.092	0.117
1991	281	2635	19	294	0.068	0.112
1992	272	2907	25	319	0.092	0.110
1993	270	3177	14	333	0.052	0.105
1994	276	3453	12	345	0.043	0.100
1995	289	3742	3	348	0.010	0.093
1996	262	4004	8	356	0.031	0.089
1997	271	4275	16	372	0.059	0.087
1998	278	4553	16	388	0.058	0.085
1999	291	4844	14	402	0.048	0.083
2000	300	5144	13	415	0.043	0.081
2001	307	5451	10	425	0.033	0.078
2002	308	5759	7	432	0.023	0.075
2003	311	6070	5	437	0.016	0.072
2004	313	6383	4	441	0.013	0.069
2005	314	6697	7	448	0.022	0.067
2006	315	7012	6	454	0.019	0.065
2007	331	7343	11	465	0.033	0.063
2008	337	7680	8	473	0.024	0.062
2009	338	8018	4	477	0.012	0.059
2010	332	8350	5	482	0.015	0.058
2011	332	8682	6	488	0.018	0.056
2012	335	9017	3	491	0.009	0.054
2013	337	9354	7	498	0.021	0.053
2014	340	9694	3	501	0.009	0.052
2015	331	10025	3	504	0.009	0.050

the frequency of these incidents. It can be seen that 4 incidents were recorded from 1971 to 1974 however, they are not part of the overall frequency analysis as there is limited data regarding operating experience. Similarly, these numbers are only confirmed from WOAD as the HSE did not begin recording incidents until 1975.

Figure 1 demonstrates the number of ship to platform collision incidents from 1971–2015 as well as the key regulations and incidents as outlined previously (GL, 2017) (HSE, 2016) (MAIB, 2016).

It can be seen from Figure 1 that the number of ship to platform collision incidents from 1971 to 2015 is very turbulent, as more clearly demonstrated by the average trend-line. At a first glance, this trend seems to be rather erratic, following no logical pattern. However, when the milestones in the safety case regulation timeline are taken into consideration, patterns begin to emerge in the number of incidents each year in the UKCS.

Initially, from 1971 to 1974 the number of incidents is very low at one or two per year. A possible reason for this is that the data entries for 1971 to 1974 are from WOAD only, as the HSE began their ship to platform collision recordings from 1975. However, from 1975 onwards the number of incidents per year greatly increases until 1981 from 12 to 29 respectively. There are a number of possibilities that can cause this rapid increase. Firstly, the HSWA is enforced from 1974 hence, the recognition of dangerous incidents that can cause harm to personnel is increased. Secondly, as more and more dangerous incidents are being recognised, the need to report said incidents also increases. Therefore, it is safe to say that an increased awareness of dangerous situations coupled with the need to report these incidents gives rise to a dramatic increase in the number of collision incidents. Thirdly, according to the HSE, the approximate number of installations operating in the UKCS increases from 89 in 1975 to 121 in 1981. The increase in the number of operating platforms would statistically increase the number of collisions at that time.

From 1981, however, the number of incidents per year begins to decrease until 1987, from 29 to 6. This decrease is much greater that the increase in incidents from 1975 to 1981. It is possible that the enforcement of the HSWA had a large effect on the safety procedures on offshore platforms in the UKCS. This hypothesis would also be consistent with the approximate number of platforms operating in the UKCS which increases from 121 in 1981 to 174 in 1987. This contradicts the previous statement that the number of incidents would increase with the number of platforms in operation. However, in the 6-year period between 1981 and 1987 this is not the case. This further backs up the idea that the regulations from 1974 have been increasingly enforced and have reduced the number of incidents. However, it is also possible to state that

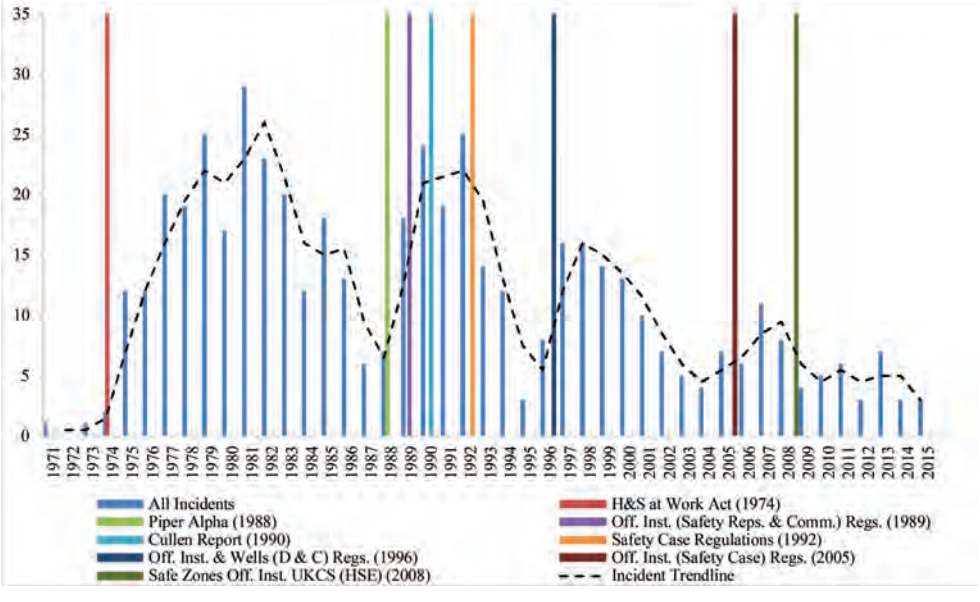


Figure 1. Graph demonstrating the number of ship to platform collision incidents per year, as well as the key regulations and events that formed the modern safety case.

the level of reporting of the collision incidents has decreased. This is a much more difficult claim to validate as there is not any possible way to determine whether an incident has happened and has not been reported. This is part of the reasoning behind an increase in research regarding offshore dynamic asset integrity modelling and autonomous systems, as the most common reason a detector or sensor will not detect and log any information is if it is faulty. On the other hand a human has the ability to choose not to carry out an action. Hence it is difficult to accurately determine the level of underreporting that would have taken place between 1981 and 1987.

Continually, the time period between 1988 and 1994, in terms of collision incidents, is very interesting. The year 1988 is well known in the offshore industry and indeed the world as the year of the Piper Alpha disaster in which 167 crew members lost their lives in the July of that year. When one examines the collision incidents that were reported in 1988, approximately 70% were reported after the loss of Piper Alpha on 6th July. This may suggest that a large-scale disaster, such as Piper Alpha, triggered an increase in the level of incident reporting. However, the number of collision incidents in 1988 alone are not enough to state this with any conviction. What is interesting however, is that the number of collision incidents increase in 1989 to 18, from 7 in 1988. This is a drastic increase in terms of the number of reported incidents in the UKCS, after a large-scale offshore disaster.

Furthermore, in 1989 the Offshore Installations (Safety Representatives & Safety Committee) Regulations were published. This stated that the workforce could elect safety representatives from amongst themselves. This may have increased the level of reporting of collision incidents in 1989. However, it appears to be too much of a drastic increase from the previous year to conclusively state that the new regulations in 1989 resulted in a considerable number of reported incidents. It seems much more likely that a combination of the Piper Alpha disaster and the release of the Offshore Installations (Safety Representatives & Safety Committee) Regulations contributed to the vast increase in reported collision incidents.

Continually, in 1990, the Cullen Report was published which was public enquiry into the Piper Alpha disaster. The report was heavily critical of the platform operators. Lord Cullen made a total of 106 recommendations within his report, all of which were accepted by industry. The responsibility of implementing them was spread across the regulators and the industry with, the HSE overseeing 57, the operators were responsible for 40, the offshore industry were given 8 to progress and the final one was for the Standby Ship Owners Association. The industry acted urgently to carry out the 48 recommendations that operators were directly responsible for. By 1993 all had been acted upon and substantially implemented. Furthermore, the HSE developed and implemented Lord Cullen's key recommendation: the introduction of safety regulations requiring the operator/owner of every fixed and mobile installation operating in UK waters to submit to the HSE, for their acceptance, a safety case. Hence, in 1992 the Offshore Installations (Safety Case) Regulations came into force. By November 1993 a safety case for every installation had been submitted to the HSE and by November 1995 all had had their safety case accepted by the HSE.

If the number of collision incidents is examined from the Cullen Report in 1990 to all installation Safety Cases being accepted in 1995, it can be seen that the number of incidents per year decreases rapidly from 24 to 3 respectively. This is again a massive fluctuation in the number of incidents following a series of key regulations being enforced. It shows that the release of new regulations prompts the level of incidents to decrease as the regulations are enforced. However, as 1995 is a number of years after the Cullen Report and the introduction of Safety Cases it is possible that an element of complacency in terms of reporting may occur. This can be seen from the number of incidents between 1995 and 2004. The number of collision incidents increases from 3 in 1995, to a peak in 1998 of 16, then to a new low of 4 in 2004. What is key is this fluctuation, actually does not exactly follow the common trend of more incidents given more operating installations. From 1995 to 1996 the approximate number of installations decreases from 289 to 262, then steadily increases to 313 in 2004. This again does not follow a pattern as the number incidents decreases after 1998 as the approximate number of installations increases. This could be attributed to a substantial level of regulation enforcement leading to a decrease in incidents, or there may be an anomaly in the approximate number of installations and incidents, given that the approximate number of installations increases steadily from 1975 to 1995, then suddenly decreases. Similarly, it is at this time that the HSE began utilising the RIDDOR regulations (from 1st April 1996).

What appears to be more likely is at the low point of 3 collisions in 1995, a new set of regulations are introduced and enforced, the Offshore Installations Prevention of Fire and Explosion, and Emergency Response (PFEER) along with the Offshore Installation (Design & Construction) Regulations in 1996. At that point, the number of incidents increases and peaks in 1998. It would be foolish to say that the introduction of new offshore

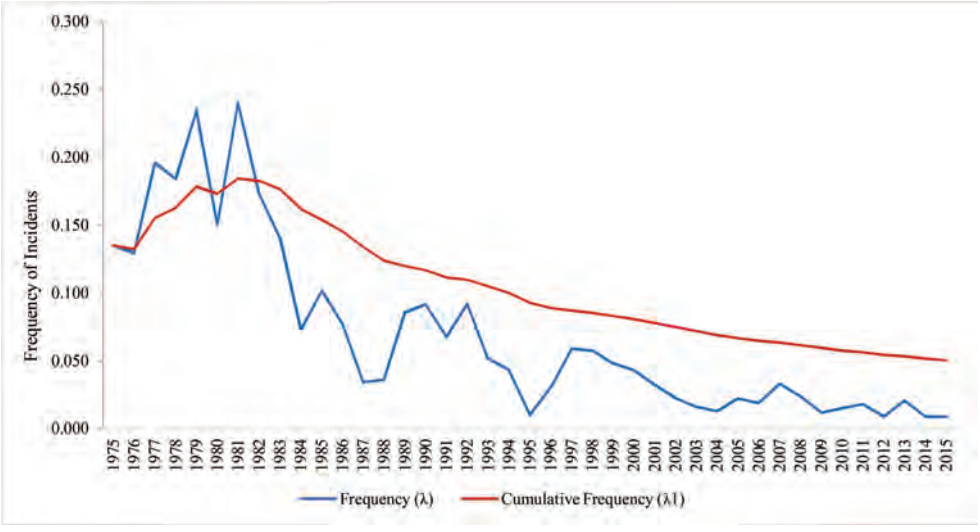


Figure 2. Incidents frequency and cumulative frequency for ship to platform collisions per year from 1975 to 2015.

regulations causes offshore incidents. What is far more likely is that the increase of new regulations prompts a more proactive response in the accuracy of incident reporting.

In addition, this trend can be seen yet again from 2004 to 2012, where the number of collision incidents per year increases from 4 in 2004 to 11 in 2007 then decreases to 3 in 2012. This could be attributed to the Offshore Installations (Safety Case) Regulations 2005 being enforced in 2006. As with the regulations in 1995 and 1996 the number of incident increases and begins to decrease. However, the number of collision incidents becomes much steadier and doesn't fluctuate as much as previous years, as the increase from 4 to 11 in 2013 is not a huge increase, and could be said to be anomalous when looking at the number of incidents in that period. However, it is an increase none the less.

Furthermore, in 2008 the Safety Zones around Oil & Gas Installations in Waters around the UK information sheet is introduced by the HSE. This specifically targets the area of offshore collisions, near misses and general safety when operating in an installations 500m zone. Therefore, it makes sense to state that this introduction has maintained a steady level of incidents between 2008, with 8, and 2015, with 3.

Furthermore, by applying the use of a calculated incident frequency and cumulative frequency, based upon the operating experience on the UKCS from 1975 to 2015, it can be seen that the number of incidents has drastically decreased over the 40 year period. This is demonstrated by Figure 2 which highlights the trend of incidents frequencies per year as well as the cumulative incidents frequency per year. What can be seen in Figure 1 and much more clearly in Figure 2 is that the average frequency of incidents has generally decreased since 1981. This clearly demonstrates that over the 40 year period from the introduction of the HSWA to the current amended Safety Case regulations, the enforced regulations have had a huge impact on installation safety in terms of collision incidents. As the general number of incidents has decreased, the approximate number of operating installations has increased.

4 CONCLUSION

From the information presented in Figures 1 and 2 as well as Table 1, it can be seen that the offshore industry can be said to be reactive in its approach to reporting incidents, especially in the area of ship to platform collisions. What is also apparent is that the fluctuation has become gradually smaller in more recent times. This shows that the effect of introducing and amending regulations over time has a positive effect on the overall trend of collision incidents. While this study identifies trends in ship to platform collisions, it would still be valid to state that the offshore industry would profit greatly from having a dynamic risk monitoring tool to aid with the continual enforcement of regulations across all areas of an offshore platform. In the near future, a widely accepted and integrated dynamic asset integrity monitoring tool could be a distinct possibility.

ACKNOWLEDGEMENTS

This research is partially supported by the HSE through a project: "Effective collision risk management for offshore installations and ship/platform collision database."

REFERENCES

Auld, H., 2013. A Safety Case development framework, Bristol: Atomic Weapons Establishment & Defence Science and Technology Laboratory.

Cai, B. *et al*., 2013. Application of Bayesian Networks in quantitative assessment of subsea blowout preventer operations. *Journal of Risk Analysis*, Volume 33, pp. 1293–1311.

Department of Energy, 1990. The public inquiry into the Piper Alpha disaster, London: Department of Energy.

Eleye-Datubo, A., Wall, A., Saadjedi, A. & Wang, J., 2006. Enabling a powerful Marine and Offshore Decision Support Solution Through Bayesian Network Technique. *Risk Analysis*, 26, pp. 695–721.

GL, D., 2017. World Offshore Accident Databank, WOAD, Det Norske Veritas, Germanischer Lloyd.

HSE, 1992. The offshore installations (safety case) regulations, London: Health and Safety Executive.

HSE, 1996. The offshore installations and wells (design and construction, *etc*.) regulations, London: Health and Safety Executive.

HSE, 2003. Ship/platform collision incident database (2001), Health and Safety Executive.

HSE, 2005. The offshore installations (Safety Case) regulations (2005), Health and Safety Executive.

HSE, 2006a. A guide to the offshore installations (safety case) regulations 2005, Health and Safety Executive.

HSE, 2006b. Guidance for risk assessment for offshore installations, Health and Safety Executive.

HSE, 2008a. A guide to the well aspects of the offshore installations and wells (design and construction, *etc*.) regulations 1996, Health and Safety Executive.

HSE, 2008b. Safety zones around oil and gas installations in waters around the UK, Health and Safety Executive.

HSE, 2015. The offshore installations (offshore safety directive) (safety cases *etc*.) regulations 2015, Health and Safety Executive.

HSE, 2016. RIDDOR database. Liverpool: Health and Safety Executive.

Inge, J.R., 2007. The safety case, its development and use in the United Kingdom.

Jones, C., 2010. The 2010 gulf coast oil spill. 1st ed.: BookBoon.com.

MAIB, 2016. Marine Accident Investigations Branch Reports. [Online]. Available at: https://www.gov.uk/maib-reports.

OGP, 2010. Risk assessment data directory—ship/installation collisions. International Association of Oil & Gas Producers.

Oil & Gas UK, 2008. Piper Alpha: lessons learnt, http://oilandgasuk.co.uk/.

Oil & Gas UK, 2009. Accident statistics for offshore units on the UKCS 1990–2007, London: Oil & Gas UK.

The Stationary Office, 1974. Health and safety at work *etc*. act 1974. [Online]. Available at: http://www.legislation.gov.uk/ukpga/1974/37/contents.

The Stationery Office, 1989. The offshore installations (safety representatives and safety committee) regulations 1989, www.legislation.gov.uk.

Wang, J., 2002. Offshore safety case approach and formal safety assessment of ships. *Journal of Safety Research*, Volume 33, pp. 81–115.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Understanding and effectively managing conservatisms in safety analysis

Steven Krahn
Vanderbilt University, USA

Mohammad Modarres
University of Maryland, USA

James O'Brien
U.S. Department of Energy, USA

ABSTRACT: Safety analysis for nuclear facilities are performed in a conservative manner so that there is additional assurance that the nuclear facilities are safe to operate given the potential events that may impact or challenge the facilities. The degree of conservatism included in the analysis should be well know and be increased if there is a large degree of uncertainty associated with the analysis. If uncertainty is decreased, through better data or more sophisticated and/or rigorous analysis, a decrease in conservatism can be made without impacting the margin of safety of the design.

1 INTRODUCTION

1.1 *Objective*

Providing conservatism in the safety analysis of a facility results in additional assurance that the facility is safe to operate. This paper explores how the degree of conservatism and the degree of uncertainty of the safety analysis are linked in determining the margin of safety of the facility design. The objective of this research is to improve the understanding of the theoretical and technical basis for ensuring the safety of a facility design.

1.2 *Approach*

This research included the following steps.

- Reviewing definitions for conservatism
- Reviewing national and international practices regarding use of conservative inputs into safety analyses
- Evaluating conservative and best estimate analyses and a potential method for reducing conservatism while maintaining safety margins

A follow-on paper is planned that will evaluate some potential applications of the method to reduce conservatism while maintaining safety margins.

2 CONSERVATISM DEFINITION

2.1 *Definition*

A discussion of conservatism, as it relates to safety analysis, is found in the U.S. Nuclear Regulatory Commission's (NRC) NUREG-2122, *Glossary of Risk-Related Terms in Support of Risk-Informed Decision-making* (NRC 2013). NUREG-2122 takes a holistic approach in describing conservatism; it defines the term conservative in combination with "analysis," and an additional qualifying term "demonstrably"–thus implying that an analysis must be evaluated as a whole. A "demonstrably conservative analysis" is: "An analysis that uses assumptions such that the assessed outcome is meant to be less favorable than the expected outcome."

The more detailed discussion in NUREG-2122 goes on to explain that a "demonstrably conservative analysis provides a result that *may not be the worst result* of a set of outcomes, but produces a quantified estimate of a risk metric that is *significantly greater* than a risk metric estimate produced using the most realistic information available" [emphasis added]. Thus, a conservative analysis can be described as being located between a bounding analysis and a best estimate analysis—but skewed towards the bounding case and significantly away from the results of the best estimate analysis. The relationship among the terms "conservative", "bounding", and "best estimate" is discussed further below.

A guide by the United Kingdom's (UK) Office for Nuclear Regulation (ONR), Safety Assessment Principles for Nuclear Facilities (UK ONR 2014) defines conservatism as:

> *In analysis, an approach where the use of models, data and assumptions would be expected to lead to a*

result that bounds the best estimate (where known) on the safe side. The degree of conservatism should be proportionate to both the level of uncertainty and the overall significance of the estimate to the safety case.

This definition is informative, as it links the level of uncertainty to the "significance of the estimate"; i.e., the importance of the parameter being addressed; however, what is the ONR meant by "bounds the best estimate" is not clear.

3 NATIONAL AND INTERNATIONAL PERSPECTIVES ON CONSERVATIVE SAFETY ANALYSES

3.1 *U.S. Nuclear regulatory commission*

The NRC's Safety Goal Policy Statement (NRC 1986) states that, "to provide adequate protection of the public health and safety, current NRC regulations require *conservatism* in design, construction, testing, operation and maintenance of nuclear power plants" [emphasis added]. Relative to conservatism in accident analysis, NRC Regulatory Guide 1.203, *Transient and Accident Analysis Methods* (NRC 2005) states that:

... results of an analysis can be conservative due to a combination of code input and modeling assumptions…. However, conservatism in just one aspect of [a model] … cannot be used to justify conservatism in the [model] as a whole, because other aspects of the model may be non-conservative and cause the overall model to be non-conservative. The degree of conservatism in the overall model must be quantified and documented. Showing the degree of conservatism in [a model] … may be accomplished by a relatively simple uncertainty analysis….The key to simplifying the uncertainty analysis is identifying the small number of parameters and physical phenomena that are important in determining the behavior of the accident.

The issue of conservatism is also tied somewhat to the issue of probabilistic versus deterministic approaches; in probabilistic approaches, conservatism and uncertainty can be specifically evaluated, whereas in a deterministic approach there is often less information to support an understanding of the degree of conservatism. This was discussed in the NRC's NUREG/CR-7168, *Regulatory Approaches for Accessing Facility Risks* (NRC 2015). A companion issue to that of probabilistic versus deterministic approaches is whether analyses should be based on data and computational methodologies that represent the best estimate of what might really occur, with uncertainty analysis to explore the effects of incorrect data or models, or should be based on demonstrably conservative data and models. Most regulations and license applications have used a conservative, deterministic approach. The NRC has identified problems with using this approach as discussed in Appendix C of NUREG-1909, *Background, Status, and Issues Related to the Regulation of Advanced Spent Nuclear Fuel Recycle Facilities* (NRC 2008). Two of the most important problems identified were: (1) that using very conservative assumptions can mask risk-significant items, and (2) that most conservative analyses are not accompanied by an uncertainty analysis.

The NRC has also addressed the issue of conservatism in thermal hydraulic code analysis and provide guidance on how best-estimate calculations can be utilized in place of conservative models. Specifically, in Regulatory Guide 1.157, *Best-Estimate Calculations of Emergency Core Cooling System Performance* (NRC 1989), the NRC states that:

the NRC staff amended the requirements of § 50.46 and Appendix K, "ECCS Evaluation Models" (53 FR 35996), so that these regulations reflect the improved understanding of ECCS performance during reactor transients that was obtained through the extensive research performed since the promulgation of the original requirements in January 1974. Paragraph 50.46(a)(1) now permits licensees or applicants to use either Appendix K features or a realistic evaluation model. These realistic evaluation models must include sufficient supporting justification to demonstrate that the analytic techniques employed realistically describe the behavior of the reactor system during a postulated loss-of-coolant accident. Paragraph 50.46(a)(1) also requires that the uncertainty in the realistic evaluation model be quantified and considered when comparing the results of the calculations with the applicable limits in paragraph 50.46(b) so that there is a high probability that the criteria will not be exceeded.

For the purpose of the above regulatory guide, the terms "best-estimate" and "realistic" have the same meaning. Both terms are used to indicate that the techniques attempt to predict realistic reactor system thermal-hydraulic response; though best-estimate is not used in a statistical sense in this guide.

The use of conservative values has been investigated in International Atomic Energy Agency (IAEA) Safety Report 52, *Best Estimate Safety Analysis for Nuclear Power Plants: Uncertainty Evaluation*, (IAEA 2008) and IAEA Publication 1428, *Deterministic Safety Analysis for Nuclear Power Plants* (2014). These documents discuss the practices, benefits and downsides associated with use of conservative and best estimate analyses. The IAEA notes that for accident scenarios with large estimated

margins to acceptance criteria, "it is appropriate for simplicity, and therefore, economy, to use conservative analysis"; however, "for scenarios in which the margin is smaller, a best estimate is necessary.…"

The IAEA goes on to describe the need for best estimate analysis in this instance is to "to quantify the conservatism"; that is, "to show the margins to the acceptance criteria that apply in reality." However, the IAEA discussion of best estimate analysis in IAEA Safety Report 52 is always combined with a focus on evaluation of uncertainties; therefore, the IAEA discussion is of "best estimate analysis together with evaluation of uncertainties." The use of this type of methodology is qualified with the caution that "realistic input data are used only if the uncertainties or their probabilistic distributions are known." For data that do not meet this test "conservative values" should be used.

An alternate approach to conservative analysis, referred to as "best estimate plus uncertainty," is discussed in IAEA Safety Report 52. The benefits of the best estimate plus uncertainty approach are described in IAEA Safety Report 52: (1) it provides more realistic information about the physical behavior of the facility and thus assists in identifying the most relevant safety parameters, (2) the use of conservative assumptions can lead to predicting an incorrect event progression or exclude relevant physical phenomena, and (3) it provides information about safety margins which is not always obvious in conservative deterministic analyses. However, moving toward best estimate analysis with uncertainty involves several challenges. For example, it is difficult to develop a relevant, validated best estimate computational methodology for the analysis. In addition, sufficient data on critical parameters must be available so that a probabilistic distribution function can be developed that is statistically valid. IAEA Publication 1332, *Safety Margins of Operating Reactors—Analysis of Uncertainties and Implications for Decision Making,* (2003) discusses how, in safety analyses, it is customary to demonstrate that adequate margins exist between the true (but unknown) values of important, safety-related parameters of interest and the corresponding regulatory limits (requirements or physical limits) that, if exceeded, would result in adverse consequences (e.g., release of radioactivity).

Safety margins traditionally are established by relying on conservative models, conservative assumptions, and conservative interpretation of the analysis results. This approach has served the nuclear industry well; however, it can lead to employing additional safety measures and barriers that may not be strictly required, resulting in costlier and "overbuilt" designs. Also, the conservative approaches are not routinely able to identify the amount or degree of conservatism, nor do they describe the degree of confidence in the resulting conclusions and safety features.

4 TECHNICAL EVALUATION OF CONSERVATIVE AND BEST ESTIMATE ANALYSIS

4.1 *Introduction to conservative analysis and safety margins*

To illustrate the conservative approach to safety analysis, consider Figure 1 (shown at end of this paper). In the conservative approach, the results are expressed in terms of a set of deterministically calculated values for the *safety parameters of interest* (e.g., radioactive dose) that are expected to be more pessimistic than the true values of these parameters. The difference between a conservative estimate of the safety parameter of interest and the regulatory/requirement limit is called the *safety margin*. Conservatism is intended to make the calculated deterministic value more limiting than the true (but unknown) value of a safety parmeter of interst, to assure that the estimated safety margin is smaller than the *true safety margin*. The difference between the true safety margin and the estimated safety margin has been described by the IAEA (IAEA 2014) as the *overbuilt safety margin*.

As illustrated in Figure 1 (figures are shown at the end of this paper), when deterministically analyzing the safety parameter of interest the conservative safety analysis approach finds a single value of that parameter to compare to the regulatory limit/requirement (referred to as the "acceptance criterion" by the IAEA) and to determine if it is below that requirement/limit with a minimum, but undefined, amount of safety margin. In the conservative deterministic approach the degree of conservatism, and the true value of the parameter in relation to the conservative estimates, remain unknown. Generally, analysts believe that the conservative estimate provides an estimate of the safety margin smaller than the true margin, which by definition leaves some "overbuilt" margin.

As a simple example, consider the following:

Say the *Regulatory Limit* for the dose to the maximally exposed offsite individual (MOI) is 25 rem Committed Effective Dose Equivalent (CEDE).

If a *Conservative Estimate* of the dose in a given accident scenario to the MOI is calculated to be 5 rem. The *True Value* of the dose to the MOI might be 0.5 rem (if the accident scenario, accident parameters, and phenomena were exactly known).

In this case then the:

- Design Safety Margin is 20 rem
- True Safety Margin is 24.5 rem
- Overbuilt Margin is 4.5 rem

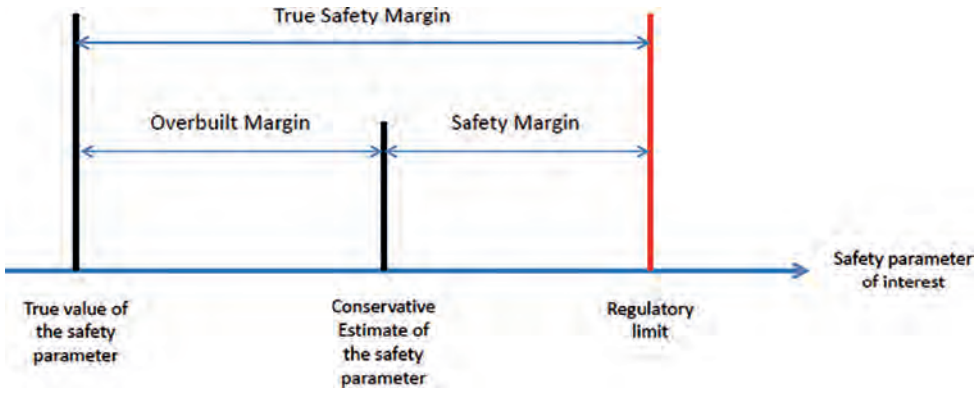


Figure 1. The concept of safety margins in conservative safety analysis.

Another way to look at the margins is the factor below the regulatory limit. In the above example, the Safety Margin is anticipated to be a factor of 5, the True Safety Margin is a factor of 50 and the Overbuilt Margin represents 90% of the True Safety Margin.

The means of assuring conservatism in deterministic safety analysis is the use of conservative models, codes, assumptions and data with the anticipation that these collectively yield pessimistic estimates of the safety parameters of interest, relative to the regulatory limit/requirements. An alternative approach to conservatism in safety analysis is the best estimate approach (as described in the NRC Regulatory Guide 1.157) to the assessment of the safety parameters of interest that includes formal quantification of associated uncertainties (or "best estimate plus uncertainty" to use the IAEA terminology).

In the best estimate plus uncertainty approach, best estimate models and computational methods, field and experimental data, and realistic assumptions are used to estimate the safety parameter of interest. Clearly, availability of such tools and data are critical to make this approach feasible. However, depending on the amount of data and information available, sometimes the best estimate approach can only provide a rough estimate of the uncertainties, which may need to be supplemented with some conservative assumptions. Conversely, when data and information are abundant, the best estimate results are frequently associated with less uncertainty. In the best estimate plus uncertainty approach, the amount of uncertainty may be expressed explicitly by obtaining the probability distribution of the safety parameter of interest and quantifying the safety margin. This concept will be discussed in more detail in the following section.

4.2 *Best estimate plus uncertainty approach to safety analysis*

In the best estimate plus uncertainty approach, the safety parameter of interest is treated as a random variable and the probability distribution within which the *true* value of the safety parameter of interest resides is estimated. Consider Figure 2, which shows a hypothetical distribution of a parameter. The true, but unknown, value of the safety parameter of interest resides somewhere within the span of this distribution. The regulatory limit/requirement of interest is shown within the range of this distribution.

Also, shown in Figure 2 is the *desired safety margin,* set to account for "unknown-unknowns." As such the safety parameter of interest should be below the regulatory limit/requirement *plus* this prescribed, desired margin.

In the best estimate plus uncertainty approach, the distribution of the safety parameter of interest, δ, is expressed by the probability density function (distribution function), $f(\delta)$, which is obtained by using realistic data, best-estimate models and codes. The best-estimate approach does not use conservative assumptions. Once developed, the distribution, $f(\delta)$, may be used to find *the probability* that the true (but unknown) value of the safety parameter of interest (expressed by the random variable, δ) is below the regulatory limit/requirement, D. In this approach it is this probability, and confidence associated with it, that forms the basis for safety decision-making. As such, in the best estimate plus uncertainty approach, the probability that the *true margin*, $(D - \delta)$, exceeds the *desired safety margin*, Δ, is expressed by:

$$Pr\left[(D-\delta) > \Delta\right] = Pr\left[\delta < (D+\Delta)\right] = \int_0^{D+\Delta} f(\delta)\, d\delta$$

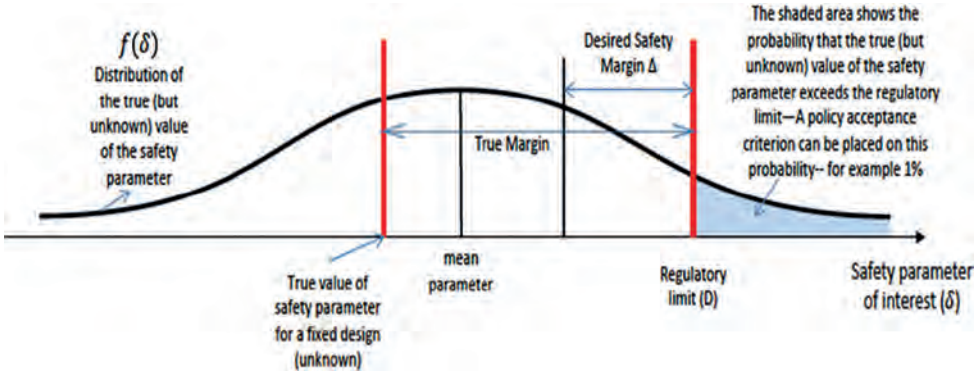


Figure 2. Conceptual depiction of the probability density function of a safety parameter.

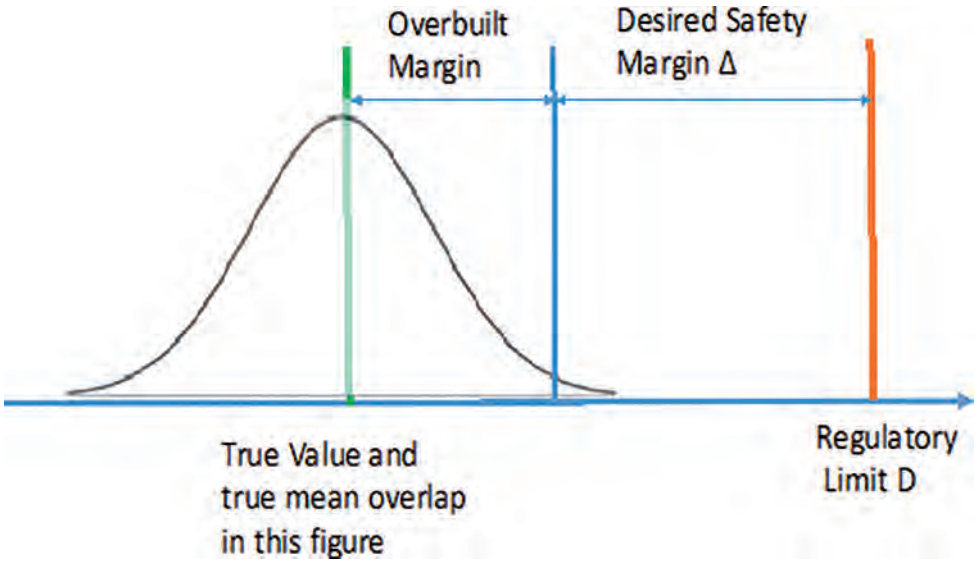


Figure 3. Potential effect of more information: True value of δ distributes narrowly.

Also, the probability that the requirement is not met would be:

$$\Pr(\delta > D) = \int_D^{\infty} f(\delta)\, d\delta$$

The benefit of the best estimate plus uncertainty method is in its characterization of the safety margin in light of the information, data and other evidence that is available. As the amount of such data increases, it is natural to expect that $f(\delta)$ becomes narrower (with smaller spread) and represents a small span within which the true value of the safety parameter of interest δ is most likely located. This concept is illustrated in Figure 3, assuming that more information about δ was available and resulting in much narrower $f(\delta)$, as compared to the probability distribution in Figure 2.

The best estimate plus uncertainty approach, as depicted in Figures 2 and 3, illustrates the cases where the true margin could become very large (for example, due to conservative initial design) by showing that $Pr\left[(D-\delta) > \Delta\right]$ could be close to unity, and highlights the presence of a large overbuilt margin. Such cases can provide a rationale to revisit the need for such large margins. This insight can only be achieved by a best estimate plus uncertainty approach, where the uncertainties associated with the margins are formally and quantitatively assessed–and confidence levels established–allowing

the overbuilt margins to be explicitly defined, explored and evaluated as to their necessity.

4.3 *Best-estimate models/codes versus conservative models*

In determining the probability distribution, $f(\delta)$, of the safety parameter of interest, one needs to have access to validated best estimate models and computational methodologies. In the conservative approach, estimates of the safety parameters of interest are obtained from conservative or bounding assumptions, and/or conservative models. To compare these two approaches, consider Figure 4. This figure shows a case where data about a safety parameter of interest are available. The conservative model (top-left branch), exemplifies an empirical (linear) model that *bounds* the data—meaning *all* the data fall below the model. As such, when the independent variable, x, takes the value x_1, the model estimates a dependent value (e.g., for a safety parameter of interest) of y_1 which is higher (more pessimistic) than all the data (evidence) that exist (note: if desired, it is possible to account for unobserved data, and draw the line above the cluster of the data with an additional margin, to cover for the unknown-unknowns).

Conversely, the best estimate approach would fit the empirical line (model) into the data using a regression analysis including the quantiles that describe the uncertainty about this model (top-right model and the bottom-right model are best estimate fits to the data). The model's upper and lower quantiles represent the model uncertainty using the residuals (expressed by the model error, ϵ). Unlike the bounding model, in the best estimate model, for a given value of a dependent variable x_1, produces a probability distribution for y_1. Similar to the conservative analysis, it may become necessary to make the best estimate model more conservative by introducing a bias to account for the unobserved data. The bottom left branch shows the same regression model of the lower right, but with an added conservative bias.

Impact of Conservatism When Multiple Parameters are Inputs to an Analysis.

When several parameters are inputs into an analysis used to determine a resulting "final" parameter or "figure of merit," which is then used to compare against regulatory limits, the amount of conservatism in the final parameter will be larger than the conservatism in each of the input parameters. This can result in large conservatisms in the resulting "final" parameter. In part, this reflects that fact that the uncertainty in the final parameter *does* increase as consequence of increases in the number of input parameters—each with its own level of uncertainty. When probability distribution

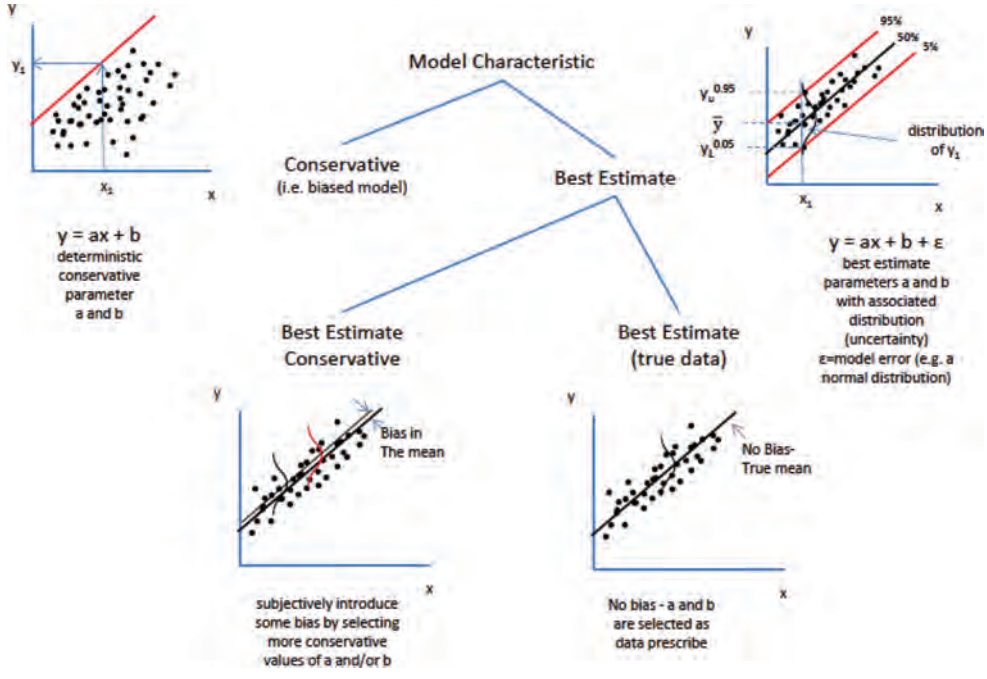


Figure 4. Comparison between best estimate and conservative models.

functions, along with associated confidence levels are not known (i.e., a "data-deficient" environment), industry standard approaches work to ensure—through conservative parametric values and/or modeling—that an appropriate level of conservatism is present in the final parameter. When this result is compared against the regulatory limit, several possibilities exist: (1) if the margin is small, more detailed analysis may be called for to more fully characterize the safety case, as discussed by the IAEA and NRC above; (2) if the margin is large and measures incorporated into system design and/or procedures are not onerous, further action may not be called for; and (3) if the margin is large and measures incorporated into system design and/or procedures are costly or result in overly complex operations, detailed analysis may be called for to assess the relative contribution of such measures to safety.

5 CONCLUSIONS

This paper first provided definitions of the term "conservative"—especially as it is applied to describe safety analyses. It then explored how the degree of conservatism and the degree of uncertainty from data supporting the safety analysis are linked in determining the margin of safety obtained in a facility design. It showed that, in theory, safety margins can be maintained with reductions in conservatism if corresponding reductions in uncertainty are made (through better data or improved analysis). This insight will help support decision-making on whether experimental or analytical resources would be best applied to making more detailed and sophisticated best-estimate analysis or by utilizing less sophisticated analysis with larger conservatisms which could lead to more resources spent on the facility design. A follow-on paper is planned to further investigate this with some example applications.

REFERENCES

IAEA (2003). Safety Margins of Operating Reactors—Analysis of Uncertainties and Implications for Decision Making. *IAEA Publication 1332.*

IAEA (2008). Best Estimate Safety Analysis for Nuclear Power Plants: Uncertainty Evaluation. *Safety Report Series Number 52.*

IAEA (2014). Deterministic Safety Analysis for Nuclear Power Plants. *IAEA Publication 1428.*

NRC (1986). Safety Goals for the Operation of Nuclear Power Plants. *51 FR 30028.*

NRC (1989). Best-Estimate Calculations of Emergency Core Cooling System Performance. *NUREG 1.157.*

NRC (2005). Transient and Accident Analysis Methods. *NUREG 1.203.*

NRC (2008). Background, Status, and Issues Related to the Regulation of Advanced Spent Nuclear Fuel Recycle Facilities. *NUREG-1909.*

NRC (2013). Glossary of Risk-Related Terms in Support of Risk-Informed Decision-making. *NUREG 2122.*

NRC (2015). Regulatory Approaches for Addressing Reprocessing Facility Risks: An Assessment. *NUREG-CR-7168.*

UK ONR (2014). Safety Assessment Principles for Nuclear Facilities.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Awareness and preparation of the population for emergencies

M. Vašková
The National Cyber and Information Security Agency, Brno, Czech Republic

M. Náplavová
The College of Regional Development, Prague, Czech Republic

J. Barta
University of Defence in Brno, Brno, Czech Republic

ABSTRACT: The paper deals with the level of citizen awareness in selected dangerous areas of Vysočina region and with their preparation for an emergency. In the theoretical part of the paper there was done an analysis of approaches to dealing with informing and training of inhabitants to emergencies in general as well as focusing on two selected subjects of Vysočina region. There were also discussed possibilities of training inhabitants and logistics information flows. In the practical part were created proposals for a preparation of the citizens for a potential emergency. First emergency we chose was a leakage of dangerous substances from ice skating park. There were also made a simulation of leaking of amoniak. The second emergency we focused on was a rupture of local dam. For both emergencies were made simulation to see its consequences.

1 INTRODUCTION

The article deals with the awareness of the population and their preparedness for a case of extraordinary event in 2 zones. Specifically, it focuses on emergency zones of winter stadium, where the dangerous substance, ammonia (NH_3), can escape. The second emergency zone is the flood territory of the dam Vír in the Vysočina.

According to §15 of Act No. 239/2000 Coll. "The municipal office familiarizes the population with the character of the possible threats in the region. Then the office also familiarizes them with the prepared rescue and liquidation works and the protection of the population. The office also organizes their training" (Act No. 239/2000). From the law it is clearly showen who has the task to inform, or to train the population. However there is not mentioned anywhere, how often the training should take place. This gap in the law is one of the aspects that cause the ignorance of the population. However, it is not possible to lay the blame only on the gap in the law, because the iniciation of the population is also very important nowadays. In today's world which is full of all kinds of inventions, and due to trends of travelling and learning about foreign countries, people often neglect the importance of knowledge area of their own residence. Sometimes it also happens that even when a citizen tries to get some news, it is impossible.

The paper is focused on the awareness of the population. So there was at the beginning of our work elaborated a questionnaire for the residents of vulnerable areas and for the authorities, under which the affected area belongs to. The question was if they carried out the training on possible scenarios of threats and whether the citizens know how to act in the case of an emergency or if they know where to go to get that information. The questionnaire was developed primarily to obtain an objective view on the situation and for the analysis of the current state in selected territories. The main goal of our work is to find an acceptable solution of the issue or at least to deepen the awareness and knowledge of the population in selected territories.

2 ANALYSIS OF APPROACHES TO ADDRESS THE AWARENESS AND TRAINING OF THE POPULATION

2.1 *The dam Vír*

The dam Vír is located in the Vysočina, location of the Vysočina region in the Czech Republic is shown in Figure 1, on the flow of the river Svratka. It serves as a source of drinking water for the area called Žďárské vrchy and the surrounding areas and a part of Brno. The area of flood planning for this reservoir is located along the river Svratka.

Figure 1. Location of the Vysočina region in the Czech republic.

It is reported that the possible flood wave would most likely reached up to the Brno dam, where it could cause considerable complications.

The total volume of the reservoir of the dam is 56,193 million m^3 and the size of the flooded area is 223,6 ha. The concrete gravity dam is in its crown 390 m long and 9 m wide. Harmless water drain under the dam is 55 $m^3.s^{-1}$. On the left side of the dam is located a water power plant.

According to the model of the passage of special flood caused by the breach of the dam was determined by the extent of the flood area. This model implies that the extent of the flood area lies in the stretch from the dam to the city Brno. Vulnerable place, which this work considers is the village Vír, which is located in the immediate vicinity of the dam Vír. Other endangered sites are in respectively the village Koroužné, Švařec, Štěpánov nad Svratka, Ujčov and Lower Čepí. Of course endangered sites also include other municipalities, but this work deals only with the territory of the Vysočina region.

Warning according to the operating plan for special flood is done by the owner of the dam Vír by using his own sirens and notifies the operational information centre of the fire rescue service (IRS) of the region about the dangers of specific floods. The IRS of the region then immediately notifies threatened population and also provides information about the development of the emergency (The operational plan for a special floods, 2017).

2.2 *Winter stadium*

The capacity of the winter stadium in Žďár nad Sázavou is 3500 visitors (Sportis, 2011). To the cooling of roller surface is used ammonia (NH_3). Ammonia is under normal conditions a colourless gas with a typical pungent odor; it is alkaline, irritant and caustic. Ammonia is very toxic for aquatic organisms (especially fish). Its very good solubility in water also plays an important role too. Also plants can be negatively affected if they are exposed to its higher concentrations both in air and in water. It participates in the acidification of soils (The integrated pollution register, 2017).

Due to the properties of ammonia there could occur an ecological disaster, because around the selected stadium is a river Sázava.

The ammonia has during short-term exposure of the person, irritating effect. It can burn the skin and eyes. It causes cough and difficulty breathing. In a concentration higher than 3.5 $g.m^{-3}$ is even short-term exposure the lethal. In the current environment is the concentration of ammonia so low that it does not entail almost any risk. Its advantage is an intense pungent odor, which highlights its presence in the air before it could rise to a dangerous level (The integrated pollution register, 2017).

In the cooling device of the winter stadium is the total charge of ammonia of 6000 kg (amount of ammonia is under the current emergency plan, in fact it is 1400 kg). This amount is divided into three parts—the condenser, the expansion tank and the ice-skating area (Trávník, 2016).

The zone of emergency planning for the winter stadium is about 130 m. In that zone there is a sports hall, 2 restaurants and a few family houses, which is located on the edge of the collision zone.

Due to the security and availability of hazardous substances and a large number of people at hockey games and other, may be the winter stadiums misused to commit a terrorist attack intentional discharge of a dangerous substance or destruction of the device (Zeman & Břeň & Urban, 2017).

Warning according to the emergency plan of the winter stadium in the release of ammonia is performed by the doorman of the winter stadium after the notification of the engineer. He begins to organize the measures in the premises of the winter stadium (according to existing emergency response plan). Alarms are divided into 3 groups according to the amount of leaked substance:

1. The first level of threat—a leakage of ammonia to 1 000 kg, if the spill threats only object of engineering.
2. The second level of threat—a leakage of ammonia to 2 000 kg, if the spill threats the entire object.
3. The third level of threat—a leakage of ammonia up to 3 000 kg, if the spill threats area of winter stadium and outside emergency zone in the direction of the wind (Trávník, 2016).

A liquidation of the accident in the first degree is undertaken by the staff of the ice rink with the use of a IRS of the winter stadium. Liquidation of the accident the second and third tier is governed by the emergency commission headed by the head of the winter stadium, who will call the appropriate personnel with cooperation with IRS (Trávník, 2016).

Due to the maximum range of the zone of emergency planning (130 m) it is necessary to warn about the emergency all persons who are in the zone of emergency planning by the signal general warning (Trávník, 2016).

Security protection of persons is carried out through escape routes (two kinds):

1. from the space of the winter stadium. There are two entrances in the main building and two entrances in the eastern and western bleachers. Escape routes from the machinery spaces are in addition to the main entrance to the lobby even to the back of the winter stadium (to cooling towers) and in front of the garage into the hall.
2. from the space of a vapour cloud of ammonia against the direction of the surface wind (Trávník, 2016).

Escape routes will be to the population advertised by radio device. All escape routes are properly marked and are kept passable. Movement of persons on the escape routes will be monitored by the riot service, which will guard the access to the infested area (Trávník, 2016).

To determine the state of awareness of the population in the emergency zone of the winter stadium was used questionnaire method. The residents of those zones and visitors as the winter stadium, sports hall, located next to the winter stadium, was submitted to a questionnaire of 9 questions.

3 THE RESULTS OF THE SURVEY

In the individual, above-mentioned zones of danger, it was examined, how are the residents prepared for emergency and whether they have enough information. Residents were submitted to a questionnaire of 9 questions.

3.1 *The results of the survey in the emergency zone of the dam Vír*

The number of persons surveyed in individual municipalities in the flood planning zone of the dam Vír is shown in the Table 1.

Table 1. Number of persons surveyed in the flood planning zone.

Municipality	Number of answered questionnaires
Vír	32
Koroužné	19
Štěpánov nad Svratkou	17
Ujčov	12
Total	80

On the first question, "Do you know that the place of your residence is located in flood territory?", replied to 100% of the respondents that they know about it.

On the second question, "Do you know where to get more information about the threat?" responded the majority of respondents positively. The answer "Yes, I know." checked 70% of the respondents. The remaining 30% didn't know where to get the necessary information.

The third question examined whether there is any training about what to do in case of emergencies. Approximately 57% of the respondents replied that the training is in progress, 37% did not know and the remaining 6% claimed that training is conducted. The answers of the respondents were compareted with the replies of the representatives of the municipalities in which the survey was conducted. It should be noted that municipal representatives about any training didn't know anything.

The respondents, who at the third question answered in the affirmative, they were asked about the evaluation of the appointed training. Respondents argued that the training was satisfied and beneficial. Unfortunately however, could not recall how often or when was the last time such training was held.

An important part of the questionnaire was to ascertain the opinion of citizens on their readiness for an emergency, whether they would welcome the introduction of the training and what form would this training should have. Only one fifth of the respondents had shown interest in possible training courses to attend. A total of 55% of respondents would prefer a combination of lectures and manuals, 22% only lecturing and the remaining 23% would meet the processed documents (manuals).

The fifth question examined, and vetted knowledge of the population, whether they know the manner in which they will be informed that an exceptional event has occurred. Interviewees had a choice of three options. The first option was that they hears from neighbors, the second option was that a warning signal will sound and the third, that they will start to ride the car of the IRS, especially car fire brigade and the Police of the Czech Republic. All of those interviewed, up to 8%, chose the answer that a warning signal will sound. Of those 8%, chose the answer: I learn it from the neighbors, with the argument that it learns more and earlier.

Question number 6 looked at whether the residents of the affected territory think that they are prepared on the emergency. To this question, respondents split almost in half. The first half is according to the response to emergency adequately prepared and the second on the contrary is not.

To the seventh question, answered all the questioned correctly. Had a task to choose the correct series of numbers to the IRS.

First aid can according to the eighth questions provide 92% of the respondents.

The last question examined whether the population knows what evacuation bag includes. It should be noted that in the questionnaire was the choice of just yes or no answer. However, all those who answered yes, they were verbally tested if they really know, what an evacuation bag should contain. The others were at least advised. About 52% of the respondents answered that knows what features to include evacuation bag.

3.2 *The results of the survey in the emergency zone of the winter stadium*

The number of persons surveyed in the emergency zone of the winter stadium is shown in the Table 2.

The first question was focused on whether the interviewees know about the potential dangers that winter the stadium represents. Only 30% of respondents know that it is winter stadium a source of danger.

The second question asked, whether the citizens know where to obtain more information about this risk. The answers were also alarming, only 17% knew where to get the information.

Due to the low number of informed respondents, it was almost unnecessary to ask the third question, whether they were ever trained on what to do in case of leakage of ammonia. Nevertheless, it was found that some training completed 5% of the respondents; it was for them in some way beneficial and indicated that such training is conducted about once every 2 years.

The fourth question asked respondents whether there would be interested in any training. Only 38% of the interviewed would be interested in training. They would prefer a combination of lectures and manuals or separate lectures.

To the fifth question "Do you know how you will be notified of the fact that there was an emergency?" a majority of those surveyed answered correctly. Only a small part of the chosed a different answer.

Table 2. Number of persons surveyed in the emergency zone of the winter stadium.

Place	Number of answered questionnaires
Winter stadium	56
Sports hall	44
Residential houses	3
Total	103

The sixth question dealt with the feeling of preparedness of the respondents to such emergency. A total of 83% of respondents do not feel sufficiently prepared for the emergency.

Any of the respondents do not own any protective agent for the case of leakage of ammonia.

On the contrary, everyone, as we found out in the eighth question, knows the emergency numbers.

First aid in case of contact with ammonia can provide, according to the answers to the ninth question, 26% of respondents.

From the responses it can be concluded that the awareness among the population regarding the aforementioned winter stadium is alarming and it is necessary to take care about this issue more.

From the questionnaire sent to the municipal authority, department of emergency management, in Žďár nad Sázavou, which was focused on acquiring information about ongoing or planned trainings. It was found that no training do not take place, even in the near future do not plan. The population was according to the responses informed about the issue of risk arising from the winter stadium a few years ago through the local press.

4 EVALUATION OF KNOWLEDGE OF THE POPULATION

Although education in this field at primary, secondary and higher professional schools required by law, in many cases, does not take place. Or it takes place only to a very low level.

From the results of the questionnaire survey under the dam Vír shows that the majority of the population, living in the village Vír, Koroužné, Švařec, Štěpánov nad Svratkou and Ujčov is well informed of the potential danger. However, there is the problem that there is no training due to this threat. The affirmative should be rated and that the population knows the emergency numbers and can provide first aid. But more than half of the respondents could not wrap evacuation luggage according to the investigation.

The positives

- Training is conducted in the framework of the voluntary fire brigade.
- Awareness is very good.
- People know where to get the information.
- Knowledge of emergency numbers.
- The majority of can provide first aid.

The negatives

- Official training in the scope of the village perimeters.
- Ignorance of the content of the evacuation luggage.

The results of the survey the emergency zone of winter stadium, follow that a substantial part of the population, whether living in the zone of emergency planning or visitors and sports hall, unaware of the potential danger.

The positives

- The population knows the emergency numbers.
- People know how they will be notified of the emergency event—warning signal.
- Some respondents were trained what to do in case of leakage of ammonia.
- Training sessions are carried out at least in the context of some of the schools in Žďár nad Sázavou.
- The municipality to warn population about the issue through the local press, several years ago.

The negatives

- The municipality does not perform training.
- Disinterest of a large number of citizens about the training.
- Training is not in the foreseeable future planned.
- People do not know where to get the necessary information.
- Doesn't know how to provide first aid in case of contact with ammonia.
- Insufficient readiness of the population to this extraordinary event. Obec neprovádí školení.

In determining the current state of the issue has been identified in the area of implementation of the statutory training of the population of the municipalities. This lack manifested itself especially on the ignorance of people. Surveyed population showed large gaps in knowledge of population protection, information, first aid and the total unpreparedness on the emergency.

In contrast, the population living in záplavovém territory under the dam Vír is sufficiently informed and ready for the emergence of floods. People know where to get information about the dangers, knows how to provide first aid. In this area it is necessary to arouse awareness about the content of the evacuation luggage.

After evaluation of the results, we have decided to undertake the modelling of leakage of hazardous substances from the winter stadium and, subsequently, to prepare and carry out an exercise on this to emergency.

5 APPLIED METHODS

An extraordinary event was chosen such that results from threats to the environment. To leakage of hazardous substances from refrigeration equipment occurs quite often and is still a current issue. Among other methods used for the purposes of the work are undoubtedly a literature review and questionnaire method, from which we emerged some threats.

Further there was used the method of simulation, in which were used the average values of the long-term monitoring of weather conditions in the site of the emergency and the expert estimation of the specified rate of leakage of hazardous substances. When using the method of constructive simulation is simulated entity controlled by the simulated operator. Constructive simulation is kind of simulation, when the model contains everything needed to during the simulation, replaced the original, including humans. Control of this type of simulation is implemented using the user interface. The display of the synthetic environment is similar to a topographic map. Constructive simulation is used in a variety of distinctive levels for different types of operations to deal with emergencies (Kanj & Flaus, 2015).

6 MODEL OF LEAKAGE OF HAZARDOUS SUBSTANCES FROM THE STADIUM

On the basis of the requirements for the format of the output was carried out the evaluation of the available simulation programs for the modelling of leakage of hazardous substances. There were taken into account the information about potential emergencies and the input data, which were known about the emergency. Other relevant data were long-term hydrometeorological data from the vicinity of the emergency. Due to the lack of input information was selected by the simulation program TerEX, which allows working with a minimum of input information. For more details of the evaluated programs were published in the article The Use of Simulation Programs of Leakage of Harmful Substances for Crisis management (Barta, 2015).

The main factor when entering the input data was the amount of the leaked dangerous substance. Due to the fact that the system of the winter stadium is designed so that it is divided into three independent parts with the possibility of swapping the cooling medium into any of them, it is not expected that in the event of a crash missed more than 60% of the amount that falls on 1/3 of the technology. This corresponds to approximately 280 kg of ammonia. Basic input data:

- Model: PUFF—Single release of boiled liquid with rapid cloud evaporation
- Substance: Ammonia
- Temperature: 7°C
- Total amount of escaped liquid: 280 kg
- Ground layer wind speed: 7 m/s
- Sky Overcast: 0%

- Time of incident event occurence and continuance: Night, morning or evening
- Type of atmospheric stability: D—isothermic
- Surface type in direction of substance spreading: Inhabited area

On the basis of the determined average temperature, the average values of the direction and force of the wind, was in the program TerEx performed a simulation whose result is shown in Figure 2.

For the determination of the extent of the emergency was output from the modeling program exported into the map base. As seen in the sector of Blue part in Figure 3, when the prevailing west wind was threatened residential area with several family homes. It was the basis for the processing of documents for the exercise called the leak of ammonia from the winter stadium in Žďár nad Sázavou.

For the realization of the exercise it was necessary to choose a suitable and available simulation program. In the framework of the project research we addressed this issue and then we have established the basic criteria, which has a simulation program for the implementation of practical exercises meet (Barta et al. 2016), (Urban et al. 2017), (Marana et al. 2016). Among the basic criteria belonged to the user friendliness and the variability of the simulator:

- Scene Editor;
- Implementation of External Data;

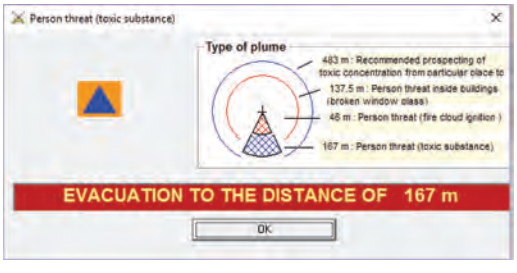


Figure 2. The output data from the program TerEx.

Figure 3. The plot of the leakage of hazardous substances into the map.

- Simulation Level (Teams or Individuals)
- Communication Possibilities;
- Continuity in Relation to the Surrounding Environment.

On the basis of comparison was evaluated as the most appropriate simulation program SIMEX, which is built on the platform of the simulator WASP (Barta & Vašková & Urbánek, 2016).

For the creation of the exercise was the best based on emergency, which become in the past. At the winter stadium in Žďár nad Sázavou there was a leakage of ammonia on 8. June 2011. On this basis, it was taken over and modified the scenario of the exercise.

6.1 *The scenario of the exercise*

8 June 2011 at 18:18 pm adopted Regional operational and information centre of the Vysočina IRS report of a leakage of ammonia from the winter stadium in the street Libušínská in Žďár nad Sázavou. To the location of the event rolls out unit of professional firefighters from the station Žďár nad Sázavou and volunteer firefighters. By carried out survey of the site firefighters found that the ammonia is leaking from a cracked pipe behind the stadium. The place was immediately marked as a danger zone. There were found two persons on the spot intervention with breathing difficulties and called an ambulance. Firefighters ordered to evacuate the population from the nearest houses downwind. They also very quickly managed to prevent further leakage of ammonia. Effluent water with a weak concentration of ammonia fell into the river, a direct threat to the environment and to the death of fish did not occur.

Before the termination of the intervention, there was carried out a final measurement with a negative result and units returned back to station.

On the basis of the scenario exercise of the ammonia leakage, it was necessary to define all entities that create, complement and comprehensively participated in the exercise. On the basis of the analysis carried out exercises with the leakage of a hazardous chemical (ammonia) to summarize each of the entities, which have been successively fed into the simulator (Oulehlová et al. 2016). The decisive step was creating lists, which contain basic clusters of entities, in particular for the area:

- Staff—as a threat to people in the winter stadium and in its vicinity, the crew of the responding units, people who are watching, etc.;
- Technical means—for example, vehicles of emergency units, auxiliary vehicle, vehicle simulating normal transport in the Žďáru nad Sázavou, etc.;
- Environment—map data of the place of leakage of hazardous substances and the terrain database with the required layers for a simulator SIMEX.

Table 3. Plan the connection of entities dealing the leakage of ammonia from the winter stadium.

Unit	Telefonne number	Work station
Regional operational and information centre	112	PS15
Fire brigade–Žďár nad Sázavou		PS02
The commander of the intervention		PS01
Units of the volunteer firefighters—Žďár nad Sázavou 2		PS03
Police of the Czech Republic	158	PS07
Emergency medical service	155	PS05
The mayor of the municipality of Žďár nad Sázavou		PS10
Management of the exercise		PS řídící
Members of the tip-off		PS13
Information line	1188	PS12
Emergency accommodation of persons		PS14
Evacuation center		PS09
Position available		
Position available		

In Table 3 are in the plan of merger provides basic entities for the resolution of an incident.

On the basis of defining the entities, there have been determined their role and activities envisaged in the framework of the scenarios dealing with emergencies (Okstad, 2016). The exercise is currently being prepared and will serve for the practical training of workers, emergency crews, and the general public to obtain information about a possible danger, its extent and consequences. In but not least, the citizens hear a lot of information how to maintain when the occurrence of an emergency with the leakage of dangerous substances.

7 CONCLUSIONS

Total awareness of the population about the risks and threats that are in their surroundings is very important. Also basic reaction and behavior of the population upon the occurrence of extraordinary events are dependent on the awareness of the population.

In the preparatory phase of the exercise had very proven freely available materials on the website of the municipality of Žďár nad Sázavou and Vysočina region. There were obtained very useful information for the preparation of exercises for informoování of the population.

This revealed that for obtaining sufficient information about the impending danger from the winter stadium (release of ammonia), inhabitants of the town of Žďár nad Sázavou have sufficient options, but these options are not used.

In the framework of the forthcoming exercises instructor received the basic information about the issue, increased knowledge about crisis management and theoretically prepared for work in the selected simulator. It was a very good basis for a workout that is focused on practice management functions, implementation of established procedures and clear communication between individual practitioners of entities.

Residents of the city may attend planned exercise, and to realize possible dangers and to obtain the necessary knowledge not only for the case of the ammonia leakage, but it will acquire a basic knowledge how to behave in emergencies, with the leakage of dangerous substances.

REFERENCES

Act. 2000. *No. 239/2000, on Integrated Rescue System, as amended*. In: Collection of Laws, no. 73, pp. 3461–3474. ISSN 1211-1244.

Barta, Jiří & Vašková, Michaela & Urbánek, Jiří. 2016. Evaluation of Simulation Programs Applicable to the Support of Decision-Making Processes in Crisis Management of Critical Infrastructure. *International Journal of Education and Learning Systems*, vol. 2016, no. 1, p. 74–80. ISSN 2367-8933.

Barta, Jiří. 2015. The Use of Simulation Programs of Leakage of Harmful Substances for Crisis management. *In: International Science Index.* Řím, Italská republika: International Scholarly and Scientific Research & Innovation, p. 856–861. ISSN 1307-6892.

Kanj, Hassan & Flaus Jean Marie. 2015. A simulation ap-proach for risk modeling and analysis based on multi-agents. Safety and Reliability of Complex Engineered Sys-tems - Pro-ceedings of the 25th European Safety and Reli-ability Conference, ESREL 2015. p. 3919–3926. ISBN 978-1-138-02879-1.

Marana, Patricia; Labaka, L & Sarriegi, J. 2016. Barriers that hamper the efficiency of Public-Private Partnerships (PPPs) in critical infrastructure protection. *Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016.* Glasgow: CRC Press. p. 532–539. ISBN 9781138029972.

Okstad, Eivind Halvard. 2016. Scenario approaches as a means of handling emerging risks in society. In *Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016.* Glasgow: CRC Press. p. 502–509. ISBN 9781138029972.

Oulehlová, Alena; Malachová, Hana; Kincl, Pavel & Navrátil, Josef. 2016. Simulated Exercise—"Gale" Crisis Scenario. In: *Vision 2020: Innovation Management, Development Sustainability and Competitive Economix Growth*. VOLS. I–VII. Seville: International Business Information Management Association (IBIMA), p. 3867–3876. ISBN 978-0-9860419-8-3.

Sportis. 2011. SPORTIS Žďár nad Sázavou [online]. [cit. 2017-04-24]. Available from: http://www.sportispo.cz/page.aspx?IDSekce=1&IDPage=2.

The fire and rescue service of the Vysočina region. 2017 Operational plan for a special floods the dam Vír: Plan of protection of the territory of the administrative district of the Vysočina region under the selected waterworks in front of a special flood. Žďár nad Sázavou: the Fire and rescue service of the Vysočina region, 15 p.

The integrated register of pollution [online]. [cit. 2017-02-26]. Ammonia. Available from: http://www.irz.cz/repository/latky/amoniak.pdf.

Trávník, Bohumil. 2016. Emergency plan: Winter stadium. Žďár nad Sázavou: Gymnasium unity ZDAS, 16 p.

Urban, Rudolf & Oulehlová, Alena & Malachová, Hana. 2017. Computer Simulation—Efficient Tool of Crisis Management. *In: International Conference Knowlwdge-Based Organisation.* Sibiu: "Nicolae Balcescu" Land Forces Academy, p. 135–141. ISSN 2451-3113.

Zeman, Tomáš; Břeň, Jan & Urban, Rudolf. 2017. Role of Internet in Lone Wolf Terrorism. Journal of Security and Sustainability Issues 7. https://doi.org/10.9770/jssi.2017.7.2(1).

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Safety climate and work conditions related to acute spills and hydrocarbon leaks in the offshore oil and gas industry—a repeated cross-sectional study

A. Aalberg, S.A. Kvalheim & I.B. Nilsen
Safetec Nordic, Trondheim, Norway

R.J. Bye
Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: In the present study, the relationship between survey data regarding work conditions and safety performance is investigated. We transform a questionnaire into several factors which involve work environment, safety climate and organizational aspects, and run a series of analyses on repeated cross-sectional data in order to predict occurrences of Hydrocarbon (HC) leaks and acute spills. We apply survey data from 49 500 respondents across eight distributions from 2001 to 2015. Methodically, we conduct Analyses of Variance (ANOVA), Principal Component Analyses (PCA) with Cronbach's alpha tests, and lastly multiple logistic regressions. Our results give some support to the hypotheses—that factors of safety climate and psychosocial aspects may be predictors of safety performance. The results and inherent qualities and weaknesses of the present study are discussed, and recommendations for further research are presented. This paper is a contribution to the development of proactive lead indicators appropriate for safety management.

1 INTRODUCTION

Preventing hydrocarbon (HC) leaks and acute spills to the environment are incidents that are of great importance to the industry and society. HC leaks can in the worst-case lead to major accidents, especially due to the inherent ignition and explosion risk. Acute spills have person injury potential, however mainly they pose a concern for the environment. Acute spills from the oil and gas industry on the Norwegian Continental Shelf is subject to high focus from both regulatory authorities and the civil society. The Petroleum Safety Authority annually issues a report on the risk level for acute spills and states that there is a need for more knowledge on the conditions that lead to acute spills (PSA, 2017). In the present study, we utilize precursor survey data and technical data related to offshore installations to investigate the conditions that are associated with HC leaks and acute spills. The survey data is obtained by the standardized questionnaire called the 'Norwegian Offshore Risk and Safety Climate Inventory' (NORSCI), which is constructed to measure health and safety climate and the risk for occupational health and accidents (Tharaldsen et al. 2008). The study focuses on individual and organizational aspects measured by a survey instrument, in order to contribute to the development of valid safety indicators that may be used as "early warnings".

1.1 *Safety indicators*

Safety indicators is used in this paper to denote the independent variables used in the analysis. An indicator may be defined as a measurable variable that can be used to describe the state of a phenomenon, when the actual state of the phenomenon might be unknown (Haugen et al., 2012). Safety indicators may be seen as observable measures that should give information concerning the safety performance (Kongsvik, 2012). An indicator does not necessarily imply that there is a causality between the content of the indicators and the safety performance. A measure may be used as an indicator as long as there is a correlation with the phenomena one want to gain knowledge about. One may differentiate between lead indicators and lag indicators. Lead indicators are considered proactive performance indicators (Dyreborg, 2009). Lag indicators are conceptualized as outcome measurements, usually represented as survey items measuring the respondent evaluation of the work practice or safety level in their organization (see e.g. Kongsvik, 2012) or different types of accident statistics (see e.g. Høivik, 2007).

The question whether survey data are lag indicators or lead indicators of safety performance should be addressed as a part of the explorations of the relationship between survey data and safety performance. It seems reasonable that both directions may apply; one could believe that earlier safety records may influence the responses of a respondent in a survey, but this may also be influenced by the conditions leading to the safety records.

Kilskar et al. (2016) have conducted a review of 174 publication addressing safety indicators, in order to gain information regarding the relation between indicators and accident risk. They conclude that there is a general lack of documented valid indicators, and that there is a need for more empirical research.

The present study's analysis is a contribution towards the ambitions within the community of safety research to close this knowledge gap. An innovative aspect of this research, in relation to previous research, is the attempt to explore and develop indicators for acute spills from the oil and gas industry.

1.2 *Organizational conditions and safety performance*

Vinnem (2012) have conducted an analysis of HC leaks in the Norwegian offshore industry based on reported incident data. In his analysis of operational circumstances for the leaks, he found that 55% of the leaks where related to human interventions. Among these, 82% of the leaks could be attributed to latent errors, e.g. errors due to maintenance and modifications. These findings support the quest to explore the possible relations between measurement of human and organizational factors and HC leaks.

Several of survey instruments have been developed in order to measure conditions and phenomena that are assumed to influence safety performance. Various concepts have been used to denote these phenomena. Concepts that are relatively common in use are *safety culture*, *safety climate, and organizational and psychosocial factors.* In addition, there are different concepts and survey instruments that are designed to measure phenomena that are not directly safety-related, but where researchers have used these instruments to test possible relations with safety performance (see e.g. Høivik 2007, Bergh et al. 2014, Olsen et al. 2016).

Measuring safety culture and safety climate has been extensively debated regarding both the definition of concept and the concept validity—what are we actually measuring? In spite of the lack of consensus regarding what is actually measured, it has become a relatively common practice in some industries to conduct these measurements.

Safety climate has been conceptualized as a representation or a subset of safety culture (for example Cooper & Phillips, 2004; Zohar, 2003). Others argue that it is a reflection of the safety culture, i.e. a kind of representation of the somewhat vague and intangible safety culture (see e.g. Guldenmund, 2000; Mearns & Flin, 1999).

Safety climate has been defined as the set of perceptions that employees share regarding safety in their environment (Zohar, 1980). A safety climate questionnaire consists of a broad range of items where the respondents are asked to give responses that are assumed to reflect their perceptions of safety related topics. Common features of a safety climate construct include management/supervision, safety competence, safety systems, work pressure and risk (Flin et al., 2000).

A safety climate questionnaire often consists of items aimed to reflect the respondents' perceptions of how safety is valued in their organization (Griffin & Neal, 2000); hence these perceptions should ideally form the frame of reference for employees about the behavior that is expected, supported, and rewarded (Zohar, 2010). However, some safety climate survey instrument consist of items that measure not only perceptions of how safety is valued, but also how they describe their own work practice, and perceive and evaluate organizational conditions such as e.g. procedures, leadership, communication, competence etc.

Whereas safety climate survey tend to address how employees make sense of their work environment, their values and attitudes, psychosocial surveys instrument seems to be more oriented towards conditions that influence their cognitive and physical capacity. Simplified, one may claim that psychosocial surveys focus more on conditions that theoretically are assumed to influence human error (reflected in the use of concept denoted as e.g. "stress", burnout", "mental exhaustion", see Bergh 2014), whereas safety climate address conditions that influence violations and lack of compliance.

1.3 *The relationship between safety climate and accident statistics*

There have been several previous attempt of exploring the relations between survey data and safety performance by the use of data from the Norwegian oil and gas industry. Several of these studies are based on survey data obtained by the NORSCI instrument.

Tharaldsen et al. (2008) have investigated the association between five safety climate dimensions and accident rates. The researchers found statistically significant, but rather weak correlations between safety climate and accident rates. Similarly, Hestad and Lilleheier (2009) found correlations

between safety climate and HC leaks. Kongsvik et al. (2011) explored the leading and lagging qualities of safety climate. In line with Hestad and Lilleheier (2009), they found that safety climate could be used as both a leading and lagging indicator for HC leaks; more negative safety climate scores were associated with an increased number of HC leaks the following 12 months. HC leaks one year before measuring safety climate also correlated negatively and significantly with the safety climate indicator; more HC leaks were associated with worse safety climate scores. The correlations were medium sized. Vinnem et al. (2010) found that a safety climate measure together with barrier failure data explained 37% of the variance in HC leaks on the installation level. They also found that the safety climate measure was the strongest predictor of leaks.

Gilberg et al. (2015) found a significant relationship between safety climate, and safety performance conceptualized as HC leaks and dropped objects. In their study, which consisted of data from 2001–2013, they also found that the leading effect was stronger than the lagging.

In additions to these analysis that have combined NORSCI survey data and safety records, there have also been conducted several studies where items in the NORSCI survey data has been used as both independent and dependent variables (Kvalheim & Dahl, 2016). They found dimensions in the survey data to be a strong predictor of accident precursors such as self- reported safety compliance in the oil and gas industry; explaining roughly 27% of the variance in safety compliance over a period of 7 years.

There has also been conducted explorative studies on Norway regarding the relationship between survey data and safety performance, by the use of other survey instrument than NORSCI and with other samples of respondents. Høivik et al. (2007) have analyzed possible relationship between items in a general work environment survey and health and safety records (occupational accidents) in one oil and gas company. They found that management style and trust in the manager are important factors for predicting personal injuries. Bergh et al. (2014) have analyzed the relationship between a psychosocial risk indicator and HC leaks frequencies, with a sample from one oil and gas company. Both survey data and some technical data (variables/indicators: age, weight, number of leakage) were used as independent variables (lead indicators). They found that the survey data significantly counted for variations in HC leaks. They found no significant relation between the technical indicators and leaks.

A general work environment survey within one single company was also used in a study by Olsen et al. (2016), in order to predict HC leaks. They found that several identified dimension in the survey data, by the use of factor analysis, where significantly related to HC leaks.

Internationally, there has been conducted some meta studies regarding relationship between survey data and safety performance. Meta studies by Clarke (2009) and Christian et al. (2009) on the relationship between safety climate and accident statistics/injuries demonstrate medium size correlations between (–0.22 and –0.39 respectively). Payne, Bergman, Beus, Rodríguez, and Henning (2009) found a negative correlation between safety climate and releases/contamination and property damage one year after measuring safety climate in the process industry. A meta-analysis considering the job-demands-resources model (Nahrgang et al. 2010) found that job resources like a supportive environment were related to safety outcomes (accidents, injuries, adverse events and unsafe behavior) in several industries.

Examples of one individual studies are Swaen et al. (2004). They conducted a cohort study and found that high psychologic job demands were a risk factor for being injured in occupational accidents in a wide range of companies and organizations. Swaen et al. (2004) investigated sleep among 47 860 individuals, and found a relationship with self-reported sleep and a risk of fatal occupational accidents.

Many of these previous studies, with notable exceptions, have conducted analyses on small data sets, often because either the number of installations has been low, or that the number of incidents has been low.

1.4 *Hypotheses*

The present study adds to the current research base by including new and updated data on oil and gas installations, and a new response variable with acute spills. Acute spills have a lot of incidents and thus serves as a response variable with better inherent statistical power. In addition, we broaden the scope of survey items, not only safety climate questions, but also organizational factors and psychosocial work environment factors combined.

Based on the discourse regarding relationship between survey data and safety records, we hypothesize the following:

H1: *Negative scores on factors concerning organizational, work environment and safety climate will be significantly related to higher probability of HC leaks the year after measurement*

H2: *Negative scores on factors concerning organizational, work environment and safety climate*

will be significantly related to high probability of acute spills the year after measurement

H3: *The factors will be significantly related to HC leaks and acute spills when controlling for operator, installation type and area of operation*

2 METHOD

2.1 *Data and variables*

The PSA gather data from the companies operating on the NCS each year: this is data regarding over 20 different defined scenarios of hazard and accident (DSHA) as well as maintenance data and barrier test data. As a part of this, there is a biannually survey, using 'Norwegian Offshore Risk and Safety Climate Inventory' (NORSCI) as instrument. This study make use of the survey data, accident records regarding HC leaks and acute spills, data regarding the types of installations and the area of the operations.

We have conducted extensive data cleaning and quality assurance. This was necessary due to different ways of reporting across data sources and operators.

As a part of the preparation, we have used the concept of *installation years*. Each observation in the analysis consists of one installation year, that is, for installation X in year Y we have survey data, HC leaks and acute spills. Further we have data for the same installation every two years later. This means that each installation included in the study have a maximum of 8 observations.

The installations are coded into types; fixed, floating and mobile units (rigs). Fixed is used as a reference category in the logistic regressions.

2.2 *Survey instrument (lead indicators)*

In total, the data consists of eight distributions of a safety and work condition survey through the RNNP study, from 2001 to 2015. The data is gathered by the PSA. The survey covers safety climate factors as well as general health, psychosocial work environment factors and background questions.

For our analyses, the data from catering and cleaning crew was excluded from the present study due to their expected small impact on the causality of HC leaks and acute spills. The total number of respondents in the survey data is 69 111. After excluding responses without reported installation name, N was 57 550, and after excluding catering and cleaning crew the final N totaled 49 500.

The questions generally are answered on a scale from 1 to 5, where 1 equals a "positive" answer. However, we have recoded questions so that a high score equals a "positive" answer.

2.3 *Safety performance (lag indicators)*

We include two variables measuring safety performance, from 2001–2016; HC leaks and acute spills.

The HC variable is a dichotomous variable with number of HC leaks over 0.1 kg per second throughout the year. For the analyses using HC leaks, we excluded mobile units, such as drilling rigs, which do not have process areas for hydrocarbon in the same extent as fixed installations like oil platforms and FPSO ships. We also excluded normally unmanned installations and data related to fields rather than installations. We coded the variable so that 0 = no HC leaks, and 1 = one or more acute spills.

Similarly, we dichotomized a variable of the number of acute spills throughout the year. The acute spills were divided in three types; *chemical, raw oil,* and *other oils*. We made a binomial variable where 0 = no acute spills and 1 = one or more acute spills. As in HC leaks, we also excluded normally unmanned installations and data related to fields rather than installations.

2.4 *Research design*

The present study is a repeated cross-sectional study with several data sources. This means that temporal variations may be investigated, and that common method bias is reduced. A design using accidents and incidents as a dependent variable also allow discussion of the measures' predictive validity.

Our analysis may be divided into four steps; exploratory bivariate tests with ANOVA, dimension reduction techniques with principal component analysis, tests of reliability with Cronbach's alpha, and lastly a multiple logistic regression analysis.

The explorative *Analysis of Variance* (ANOVA) was related to H1 and H2. A central question to ask when conducting these repeated cross-sectional studies was: Does negative response to organizational factors lead to accidents, or do accidents lead to certain perceptions of the work environment and organizational factors? Which one is the strongest? It may be argued for both directions of causality, and therefore we chose to test both directions of causality in the first steps of the study.

In the ANOVA test, we ran a series of analyses where we compared the means of the survey items by the two levels of the outcome variables (0 = no incidents, and 1 = one or more incidents).

The ANOVA test is similar to a Student's t-test as $F = t^2$, with identical p-value, for analyses with two groups.

In the following analytic steps, we chose to go further with the items significantly related to HC leaks and/or acute spills, but only if they were either only significantly related to the future (leading) or better as a predictor of the future than as a product of the history (lagging).

The *Principal Component Analysis* (PCA), commonly denoted as a *factor analysis,* was conducted in order to further investigate the findings in the ANOVA analysis. This is due to that a lot of the items significantly related to our outcomes were believed to represent common latent phenomena.

We conducted four PCA's, two on each target variable. The first iteration was an exploratory factor analysis using a criterion of eigenvalue > 1 and ibspection of the scree plot to decide the number of factors. Thereafter, we conducted a confirmatory PCA with the number of factors we chose to extract based on eigenvalues and scree plot.

A *Cronbach's alpha* (a) procedure was ran on all factors in order to ensure the internal consistency of the factors. The Cronbach's alpha procedure essentially is a way of calculating all inter-correlations between items of a scale or factor (Field, 2009). A common rule of thumb is that factors should be a > 0.70 (Nunnally, 1978).

Multiple logistic regression was performed to test H3, and further H1 and H2. The factors identified in the PCA step may be confounded by other factors like differences between installation types. To control for these factors, and especially the survey factors themselves, a multiple regression method was viable.

We conducted one multiple logistic regression model for each target variable. In the first step, we included the major companies as dummy variables. The second step consisted of installation types, and the third was sea locations (the Norwegian sea or the Barents Sea).

In the final step, we included the survey factors to see a) if they could explain differences in the target variables even when controlling for these background variables, and b) to identify which survey factors that were strongest when controlling for each other.

3 RESULTS

From the ANOVA analysis—we found that a total 44 out of 144 items were significant for acute spills, and 57 for the HC leaks. As mentioned in the methods section, a lot of these were stronger as a lagging indicator than a leading indicator, that is, the survey results was more correlated with the previous year's accident statistics than the next year's. When we excluded these items, as well as some questions loading on several or none of the factors in the PCA, we had 21 items for HC leaks and 14 for acute spills.

The iterations of PCA showed four factors for acute spill items and five factors of HC leak items. The first factor of both HC leaks and acute spills was by far the most explanatory. This factor, denoted as *Framework condtitions,* consisted of items regarding e.g. competence, training, safety procedures and manning.

One of the factors of HC leak items had a Cronbach's alpha level which was not satisfactory, thus excluded from further analysis. The results are presented below in Tables 1 and 2. In sum, the ANOVA and PCA findings give support to hypothesis 1 and 2.

Descriptive statistics and ANOVA results for the factors and target variables are presented in Table 3 below.

The results from the logistic regression presented in Table 4 show that the model explained 12% of the variation in HC leaks the next year, and 21% of the Acute spills. Operator 2 has significantly lower probability of both incident types when compared to operator 1. Operator 3 and 5 also have this relationship, but only for Acute spills. For HC leak, the factor Framework conditions (see Table 5) is significantly and negatively related to HC leaks, meaning that for higher (more positive) scores on this factor, the probability of one or more HC leaks the next year is lower. The same relationship

Table 1. Factor loadings for survey factors HC leaks.

HC leak Factor name	Factor loadings	Cronbach's alpha	Items
Framework C	0.46–0,76	0,87	10
Leadership	0,58–0,83	0,88	8
Work organisation	0,41–0,74	0,84	7
Coordination	0,43–0,82	0,93	3
Organsational risk awareness	0,44–0,69	0,45	3

Table 2. Factor loadings for survey factors acute spills.

Acute spills Factor name	Factor loadings	Cronbach's alpha	Items
Phys/quant WE	0.63–0,82	0,81	5
Noise and sleep	0,82–0,89	0,70	2
Psychosocial WE	0,49–0,82	0,72	4
Competence	0,74–0,82	0,70	3

Table 3. Means of the factors related to HC leaks (N = 211) and Acute spills (N = 520) next year.

	HC leak		Acute spill	
	0	1 or more	0	1 or more
Phys/quant WE	–	–	3,76	3,63**
Noise and sleep	–	–	3,83	3,74**
Psychosocial WE	–	–	3,88	3,94**
Competence	–	–	3,42	3,44
Framework cond.	4,19	4,07**	–	–
Leadership	4,03	3,89**	–	–
Work organisation	3,64	3,47**	–	–
Coordination	4,12	4,08**	–	–

**Significantly lower on a 0,01 level.

Table 4. Results from logistic regression predicting HC leaks (N = 211), and acute spills (N = 520).

	HC leak		Acute spill	
	OR	St. error	OR	St. error
Op 2	0,11*	1.06	0,18**	0.38
Op 3	0,74	0.61	0,21**	0.46
Op 4	0,00	882.74	0,90	0.54
Op 5	0,41	1.09	0,35*	0.42
Floating inst.	1,28	0.38	1,82	0.33
Mobile unit	–	–	0,99	0.33
Norwegian Sea	0,91	0.48	3,52**	0.43
Phys/quant WE	–	–	0,10**	0.52
Noise and sleep	–	–	2,86	0.44
Psychosocial WE	–	–	0,45	0.64
Competence	–	–	0,35*	0.37
Framework cond.	0,01**	1.57	–	–
Leadership	6,66	1.33	–	–
Work organisation	0,22	1.05	–	–
Coordination	0,37	0.86	–	–
Nagelkerke *r2*	12%		21%	

**Significant on a 0,01 level,
*Significant on a 0,05 level.

concerns the Physical and Quantitative Work Environment and the Competence factor with regards to Acute spills (see Table 5). The factors contains questions regarding exposure to chemicals, psychosocial demands and training. Installations operating in the Norwegian Sea was significantly positively related to acute spills the next year, when compared to the North Sea.

In sum, these findings give partial support to hypothesis 3.

The items of the two strongest significantly related questionnaire factors are presented in Table 5.

Table 5. Items in the most important survey factors.

Physical and quantitative work environment (acute spills)
Are you exposed to skin contact with for example oil, drilling fluids, cleaning fluids or other chemicals?
Can you smell chemicals or clearly see dust or smoke in the air?
Do you have difficulties seeing what you should see because of lack of, weak or blinding lighting?
Is it necessary to work in a high pace? Er det nødvendig å arbeide i et høyt tempo?
Do you consider the shift arrangement as demanding?
Framework conditions (HC leaks)
I have the necessary competence to conduct my work in a safe way
The HSE procedures covers well my tasks at work
I have had sufficient training in safety
I find it easy to find things in governing documents (requirements and procedures)
Does your work demand so much attention that you perceive it as demanding?
The manning level is sufficient so that HSE is taken care of in a good way
Risky work operations are always thoroughly assessed before they start
Information about unwanted events are effectively used to prevent repetitions
Safety as the highest priority when I do my job
Competence (acute spills)
I have had sufficient training in safety
Do you get the necessary training in the use of new ICT systems?
I find it easy to find things in governing documents (requirements and procedures)

4 DISCUSSION

The aim of the present study was to investigate the predictive effect towards HC leaks and Acute spills by using a survey addressing work conditions and safety climate, offshore installation and company information.

We found that our model explained 12% of the variation in HC leaks, and 21% of the Acute spills the next year.

Our results show that these factors could be used as indicators for future safety performance, although the total explained variance is modest. The overall findings are in line with several of previous studies (e.g. Gilberg et al., 2015; Kongsvik et al., 2011, Beus et al., 2010). We also found that the set of perceptions that employees share regarding safety in their environment (Zohar, 1980; Zohar, 2003; Zohar, 2010) are somewhat related to safety performance.

In addition, we see that not only safety climate factors are related, but the fact that the Physical and quantitative work environment and Noise and sleep factors are related to safety performance (although bivariately) corresponds well with particularly the research of Nahrgrang et al., (2010) and Åkerstedt et al. (2002) on work conditions.

In the following, the hypotheses are specifically discussed.

4.1 *Hypothesis 1 – HC leaks model*

All the factors were significantly related to HC leaks, thus giving support to hypothesis 1. In the multiple regression, the Framework factor was the strongest predictor of HC leaks. This should be highly relevant from the perspective of the supervisory authority, because these are aspects that can easily be inspected by regulators. Based on our findings, targeting documentation of training, manning level analyses and the availability and quality of HSE procedures on the installations can aid in identifying installations at risk of HC leaks. The content of the Framework factor is consistent with the content of several of the safety climate dimensions such as safety system, safety competence and work pressure presented in Kvalheim & Dahl (2016) as important predictors of safety compliance.

Considering the complexity of a HC-leak, the fact that we measure the safety climate the year before, that no technical condition data is in the equation, and the inherent validity and reliability issues present with surveys as a method, we conclude that these are indeed interesting results. However, the incremental explained variance is somewhat lower than earlier studies, which may be due to new control variables (for example, installation type and area of operation), or that the inclusion of more, new and better data reduces unwanted biases in the observations.

4.2 *Hypothesis 2 – Acute spills model*

All factors but one was significantly related to acute spills the next year as shown by the ANOVA analysis. Competence was the factor that did not significantly explain any variance in the ANOVA, but, however, it was significant as a predictor in the regression analysis. In sum, this give support to hypothesis 2. The strongest predictor was the Physical and quantitative work environment factor, and the Competence factor was also related to acute spills the next year.

Interestingly, the questions and factors related to acute spills were quite different from the questions and factors related to HC leaks. Whereas the acute spills aspects were related to work environment factors, and for example the subjective experience of having a demanding work hour, the HC leaks aspects were related to framework conditions such as the competence, safety system and HSE procedures. This difference reflects the assumption that HC leaks are related to more complex work processes, which involves e.g. more coordination of different tasks and people than acute spills which may be more related to more limited task conducted within a more limited time period.

Some of the questions may be related to the baseline risk on some installations. This particularly concerns the two questions addressing whether the respondents can see or smell chemicals, or are exposed to oil. This means that a reason for the strong effect could be due to that, whatever reason, some installations have a lot of acute spills, and this influences the results on these two questions, and therefore the correlations found here are to some degree spurious. However, since we included only the questions that were more relevant as a leading indicator, some predictive explanatory effect solely by the questions is present.

4.3 *Hypothesis 3 – Multiple regression*

When looking back to hypothesis 3, we assumed that the factors related to safety performance would be significant even when controlling for variables and conditions such as operator, installation type and area of operations.

This was indeed the case for Framework conditions (in regards to HC leaks), Physical and quantitative work environment and Competence (in regards to acute spills). The other factors were not significant. This gives some support to our hypothesis.

It may be that although we find different factors using a PCA procedure, that several of the survey factors to some extent in reality measure one underlying phenomenon, a common work condition/work environment response, and that the specific aspects do not have enough explanatory effect. Indeed, the first factor of the PCA's have a larger variance explained than the others. The common method bias of a broad survey is also somewhat present, which may explain that only one and two, respectively, factors are significantly related to safety performance.

5 CONCLUSIONS AND FURTHER RESEARCH

In line with theories and findings that show the complexity of accidents and incidents, we included a broad spectrum of survey factors, including safety climate, work environment and

psychosocial factors. We found that survey factors can indeed be related to future safety performance, and to some extent even when controlling for operator, installation type and are of operation. There are inherently different causal chains which lead to HC leaks than acute spills. This was also reflected in the findings—the factors from the PCA were quite different between HC leaks and acute spills.

Further research should consider treating the variables as continuous to cover more of the data variation. Additionally, comparing results across other types of safety performance would be interesting, for example dropped objects, HC leaks below 0.1 kg/sec, or person injuries. Where data are available, such comparisons as in this study would be interesting to conduct in other industries as well. Further, some installations may vary in terms of size, operation type and activity. Installation types could be nuanced according to these aspects in further research. Efforts to find a normalization variable for the target variables should be done (e.g. HC leaks per work hour, gas produced etc.).

Lastly, considering alternative approaches to the aggregations of a single installation score could be done. Aggregating from individual to installation level is necessary a data loss, which should be investigated using different techniques than the arithmetic mean.

To sum up, this study shows that there are modest, but interesting, relationships between safety climate and work environment inquires and future safety performance. These conditions should be recognized among authorities and companies operating in the oil and gas industry, and the factors may be used as risk indicators in safety management efforts.

REFERENCES

Åkerstedt, T., Fredlund, P., Gillberg, M., & Jansson, B. (2002). A prospective study of fata occupational accidents – relationship to sleeping difficulties and occupational factors. *J. Sleep Res., 11, 69–71.*

Bergh, L.I.V., Ringstad, A.J., Leka, S., & Zwetsloot, G.I. (2014). Psychosocial risks and hydrocarbon leaks: an exploration of their relationship in the Norwegian oil and gas industry. *Journal of Cleaner Production*, *84*, 824–830.

Beus, J.M., Payne, S.C., Bergman, M.E., & Arthur Jr., W. (2010). Safety climate and injuries: An examination of theoretical and empirical relationships. *Journal of Applied Psychology*, *95*(4).

Christian, M.S., Bradley, J.C., Wallace, J.C., & Burke, M.J. (2009). Workplace safety: a meta-analysis of the roles of person and situation factors. *The Journal of Applied Psychology*, *94*(5), 1103–27. doi:10.1037/a0016172.

Clarke, S. (2009). The relationship between safety climate and safety performance: A meta-analytic review. *Journal of Applied Psychology*, *94*(5), 1103–1127.

Cooper, M.D., & Phillips, R.A. (2004). Exploratory analysis of the safety climate and safety behavior relationship. *Journal of Safety Research*, *35*(5), 497–512. doi:10.1016/j.jsr.2004.08.004.

Dyreborg, J. (2009). The causal relation between lead and lag indicators. *Safety Science*, *47*(4), 474–475.

Field, A. (2009). Discovering statistics using SPSS. *Discovering statistics using SPSS 2nd ed* (2nd ed.). London: Sage Publications.

Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Measuring safety climate: identifying the common features. Safety Science, 34, 177–192.

Gilberg et al. (2015). Why measure safety climate? A longitudinal study on the relationship between safety climate measurements and safety performance. In Safety and Realibility of Complex Engineered Systems – Podofillini et al., (Eds). Taylor and Francis Group, London. ISBN 978-1-138-02879-1.

Griffin, M.A., & Neal, A. (2000). Perceptions of safety at work: A framework for linking safety climate to safety performance, knowledge, and motivation. Safety Science, 34(1–3), 99–109.

Guldenmund, F. (2000). The nature of safety culture: a review of theory and research. Safety Science, 34(1–3), 215–257. doi:10.1016/S0925-7535(00)00014-X.

Haugen, S; Seljelid, J; Nyheim, O M; Sklet, S; Jahnsen, E (2012). A generic method for identifying major accident risk indicators, Presented at ESREL 2012, June, Helsinki, Finland.

Hestad, J.A., & Lilleheier, T. (2009). Regresjonsanalyse av hydrokarbonlekkasjer mot andre indikatorer i RNNP – Norsk sokkel.

Høivik, D., Baste, V., Brandsdal, E., & Moen, B.E. (2007). Associations between self-reported working conditions and registered health and safety results. *Journal of Occupational and environmental Medicine*, *49*(2), 139–147.

Kilskar, S.S., Øien, K., Tinmannsvik, R.K., Heggland, J.E., Hinderaker, R.H., & Wiig, S. (2016). Major Accident Indicators in High Risk Industries—A Literature Review. In *SPE International Conference and Exhibition on Health, Safety, Security, Environment, and Social Responsibility*. Society of Petroleum Engineers.

Kongsvik, T., Fenstad, J., & Wendelborg, C. (2012). Between a rock and a hard place: Accident and near-miss reporting on offshore service vessels. *Safety science*, *50*(9), 1839–1846.

Kongsvik, T., Kjøs Johnsen, S.Å., & Sklet, S. (2011). Safety climate and hydrocarbon leaks: An empirical contribution to the leading-lagging indicator discussion. Journal of Loss Prevention in the Process Industries, 24(4), 405–411. doi:10.1016/j.jlp.2011.02.004.

Kvalheim, S.A. & Dahl, Ø., 2016. Safety compliance and safety climate : A repeated cross-sectional study in the oil and gas industry. Journal of Safety Research, 59, pp. 33–41.

Mearns, K.J., & Flin, R. (1999). Assessing the state of organizational safety—culture or climate? Current Psychology, 18(1), 5–17. doi:10.1007/s12144-999-1013-3.

Nahrgang, J.D., Morgeson, F.P., & Hofmann, D.A. (2011). Safety at work: a meta-analytic investigation of the link between job demands, job resources, burnout, engagement, and safety outcomes. The Journal of Applied Psychology, 96(1), 71–94. doi:10.1037/a0021484.

Nunnally, J.C. (1978). Psychometric Theory. New York: McGraw-Hill.

Olsen, E., Næss, S., & Høyland, S. (2015). Exploring relationships between organizational factors and hydrocarbon leaks on offshore platform. *Safety science*, *80*, 301–309.

Payne, S.C., Bergman, M.E., Beus, J.M., Rodríguez, J.M., & Henning, J.B. (2009). Safety climate: Leading or lagging indicator of safety outcomes? Journal of Loss Prevention in the Process Industries, 22(6), 735–739. doi:10.1016/j.jlp.2009.07.017.

Petroleum Safety Authority (PSA). (2017). Akutte utslipp. Utviklingstrekk 2016. Norsk Sokkel. Risikonivå i norsk petroleumsvirksomhet. Stavanger.

Swaen, G.M.H., van Amelsvoort, L.P.G.M., Bültmann, U., Slangen, J.J.M., & Kant, I.J. (2004). *Journal of Occupational and Environmental Medicine, 46*(6), 521–527.

Tharaldsen, J.E., Olsen, E., & Rundmo, T. (2008). A longitudinal study of safety climate on the Norwegian continental shelf. Safety Science, 46(3), 427–439. doi:10.1016/j.ssci.2007.05.006.

Vinnem, J.E. (2012). On the analysis of hydrocarbon leaks in the Norwegian offshore industry. *Journal of Loss Prevention in the Process Industries*, *25*(4), 709–717.

Vinnem, J.E., Hestad, J.A., Kvaløy, J.T., & Skogdalen, J.E. (2010). Analysis of root causes of major hazard precursors (hydrocarbon leaks) in the Norwegian offshore petroleum industry. Reliability Engineering & System Safety, 95(11), 1142–1153. doi:10.1016/j.ress.2010.06.020.

Zohar, D. (1980). Safety climate in industrial organizations: Theoretical and applied implications. Journal of Applied Psychology, 65(1), 96–102.

Zohar, D. (2003). Safety climate: Conceptual and measurement issues. In J.C. Quick & L.E. Tetrick (Eds.), Handbook of occupational health psychology (pp. 123–142). Washington, DC: American Psychological Association.

Zohar, D. (2010). Thirty years of safety climate research: reflections and future directions. Accident Analysis & Prevention, 42, 1517–1522.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Reliability of power system considering replacement of conventional power plants with renewables

Marko Čepin
Faculty of Electrical Engineering, University of Ljubljana, Ljubljana, Slovenia

ABSTRACT: Power system is one of the most complex systems ever designed. Determining its reliability is therefore a demanding process, which is mostly dealt with in a way that partial problems are stated and solved. A reliability method of loss of load expectation is selected and improved and the model of conventional power system is considered as a standpoint. Reliability of the initial conventional power system is evaluated. Then, the conventional power system is changed, where one conventional power plant is abandoned and new wind power plants stand to its place. The reliability of the initial and the changed power system is calculated and compared. Discussion shows the problems of decreased reliability of the power system with increasing the percentage of wind power plants within the power system replacing conventional plants, if not enough reserve is added to the power system at the same time. The other related problems are discussed in addition, to show that the problem is very complex and one cannot only consider a limited number of influencing parameters when deciding on power system configurations.

1 INTRODUCTION

Power system is one of the most complex systems ever designed (Cepin, 2011). Determining its reliability is therefore a demanding process, which is mostly dealt with in a way that partial problems are stated regarding its reliability and then the solutions are presented. Many methods exist, which each from its viewpoint give the representative answer to the stated problem. Some of them deal with system as a static system and the term adequacy is considered as the means of power system reliability. Some of them deal with the system as a dynamic system and the term security is considered as the means of power system reliability.

Loss of load expectation is only one of the methods identified (from the static point of view to power system reliability) which has reached a number of interesting applications (Calabrese, 1947, Billinton, Allan, 1996, Elmakias, 2008, Bricman Rejc, Čepin, 2014).

The objective of the paper is to extend the static method of power system reliability assessment named loss of load expectation into an improved version of the method and present realistic examples of the method, which can represent a standpoint for discussions for selections of new power plants in the system and the amount of reserve power needed in order to keep the specified level of power system reliability.

Loss of load expectation assesses reliability of conventional power systems with constant power and constant plant availability.

Renewable power plants are different from conventional in many aspects and the variable power is one if important ones.

The objective is to improve the method is sense to allow assessment of loss of load expectation performed in several steps considering the conventional plants and considering the renewable sources, which depends largely on weather parameters (such as river flows, wind speed, sun irradiance), which drive the power of the plant. In addition to consider the power variability also the power plant availability shall be a variable and not a constant.

2 METHODS

2.1 *Loss of load expectation*

A loss of load expectation is a method, which is based on probabilistic approach for determination of required reserves in the power system or to assess its static feature how probable is state, where the loads of the power system are too high to be powered from the available power (Calabrese, 1947, Billinton, Allan, 1996). The method analyses the probabilities of simultaneous outages of power plants, which based on a model of daily load diagram, determine the number of hours per period considered, e.g. per one year, of expected power production capacity shortages.

Loss of load occurs whenever the power system load exceeds the available generating capacity of power plants. Loss of load expectation expresses value representing the number of hours or days

in a certain time period considered, when power consumption cannot be covered considering the probability of losses of generating units. This time period is usually one year.

The power system generation planners can evaluate generation system reliability and determine how much capacity is required to obtain a specified level of loss of load expectation. As demand grows over time, additional generating units are included in a way that the loss of load expectation does not exceed the required criterion.

Actually, the initial objective of the method is more to determine the required reserve power in the system than to evaluate its reliability and in this sense the application of the method is realized.

2.2 *Mathematical model*

Figure 1 shows yearly load diagram showing the terms explaining the Loss Of Load Expectation (LOLE). Mathematical model includes the following parameters:

- number of power plants that contribute to the system,
- number of steps,
- number of hours in daily load diagram (or yearly load diagram),
- generating capacity of power plant i in step j,
- availability of power plant i in step j,
- yearly or daily load diagram of the power system,
- probability of state k in step j; it includes some available and some unavailable power plant—all of them considered with their availability or unavailability, accordingly,
- time duration of loss of capacity, i.e. the time period, when the power in the yearly (or daily) diagram is larger than the sum of generating power plants in the system—and it is related to specific state; it denotes the duration of loss of capacity on Figure 1,
- installed capacity of the power system is obtained through the sum of generating capacities of all power plants contributing to the system in step j,
- unavailability of power plant i is calculated as complement of its availability,
- number of system states is calculated as the function of the number of power plants that contribute to the system (st_stanj = 2^{no_el}); each state denotes that a particular power plant is available or unavailable; thus the number of states increases with the power of number of power plants that contribute to the system; alternative recursive algorithm can be a solution for larger systems, where the number of states would reach the number, which is not possible to be evaluated with the current computers,
- time duration of not providing enough load in daily load diagram (or yearly load diagram) for system state k and step j,
- available power capacity of all the power plants, which are available.

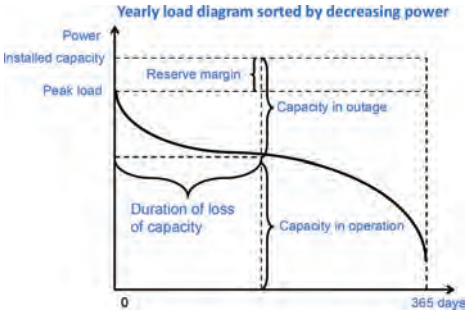


Figure 1. Yearly load diagram showing the terms explaining the loss of load expectation.

General equation for calculation of loss of load expectation is the following:

$$LOLE = \sum_{k=1}^{st\_stanj} p(k) \cdot tr\_izp\_procent(k) \qquad (1)$$

where: LOLE = loss of load expectation, p(k) = probability of state k, k = index of states, tr_izp_procent = time duration of not providing enough load in daily load diagram (or yearly load diagram), or $t(P_1 < P_2)$, where P_1 is the power capacity of power plants in specific state and P_2 is the power in the daily load diagram or yearly load diagram) or expressed on Figure 1: duration of loss of capacity.

The values of p(k) are obtained from capacity table, which needs to be prepared for evaluation.

Table 1. Example of capacity table for three power plants with powers: 40 MW, 30 MW, 10 MW and their respective unavailabilities: 0.1, 0.05, 0.04 also addressed as Forced Outage Rates (FOR).

Unit A	Unit B	Unit C	Capacity lost (MW)	Capacity in service (MW)	Probability of each capacity state
1	1	1	0	80	$0.90 \cdot 0.95 \cdot 0.96 = 0.8208$
1	1	0	10	70	$0.90 \cdot 0.95 \cdot 0.04 = 0.0342$
1	0	1	30	50	$0.90 \cdot 0.05 \cdot 0.96 = 0.0432$
0	1	1	40	40	$0.10 \cdot 0.95 \cdot 0.96 = 0.0912$
1	0	0	40	40	$0.90 \cdot 0.05 \cdot 0.04 = 0.0018$
0	1	0	50	30	$0.10 \cdot 0.95 \cdot 0.04 = 0.0038$
0	0	1	70	10	$0.10 \cdot 0.05 \cdot 0.96 = 0.0048$
0	0	0	80	0	$0.01 \cdot 0.05 \cdot 0.04 = 0.0002$

1 represents plant operable, 0 represents plant inoperable.

2.3 *Upgrade of the method*

The upgrade of the method goes in considering specifics of available power from power plants, which cannot generate the full power all the time, so their nominal power is not used in every step j, but their real power capacities are used.

Namely, the hydro power plants with not a lot of accumulation can produce the power related to the incoming flow, the wind power plants can produce the power related to the weather parameters, such as wind speed, and the solar power plants can produce the power related to the weather parameters and solar power density. Furthermore, the daily load diagram changes every day and for the more detailed model, the listed details are included.

Loss of load expectation changes through the days depending on power of power plants, depending on specific daily load diagram of the day. Its average can be calculated.

$$LOLE_{h/d} = \frac{1}{st\_cas\_k} \sum_{j=1}^{st\_cas\_k} LOLE_{h/d}(j) \qquad (2)$$

where: $LOLE_{h/d}$ = loss of load expectation (hours per day), $LOLE_{h/d}$ (j) = loss of load expectation in step j (hours per day), and st_cas_k = number of steps, j = index of steps.

At the same time LOLE expressed in hours per year can be calculated:

$$LOLE_{h/y} = \sum_{d=1}^{365} LOLE_{h/d}(d) \qquad (3)$$

where: $LOLE_{h/y}$ = loss of load expectation (hours per year), $LOLE_{h/y}$ (d) = loss of load expectation in day d (hours per day), and d = index of days.

Due to variability of power of renewable sources another dimension of power from renewable sources can be introduced to every time step of calculation, thus giving minimum and maximum of LOLE in addition to its mean value. Actually, a distribution of values instead of single value can be assessed.

2.4 *Software support*

Software for supporting calculations has been developed, which allows quick evaluation of many scenarios and which enables sensitivity studies and analyses of variations of power supply in future years with increased consumption, which calls for new production capabilities.

3 ANALYSIS AND RESULTS

3.1 *Models – base case*

Real conventional power system consists of 11 power plants: 1 nuclear, 6 thermal power plants and 4 hydro power plants and, in addition, a hydro pump storage power plant. The reliability model of selected system is developed for calculation of loss of load expectation considering the yearly load diagram as it appeared in the year 2016. This model represents the base case.

The data for the base case consists of name of the plant, nominal power (electric power delivered to the power system) and forced outage rate (or plant unavailability caused by unintentional causes). Exception is hydro pump storage power plant, which is modeled as changing directly the load diagram and its forced outage rate is not needed. When the hydro pump storage power plant operates in pumping state, its power is added to the load diagram, when it operates in generating power state, its power is reduced from the load diagram. The losses are neglected at this point, but should be considered in future.

Table 2 presents the power system data, where the hydro power plants are arranged into groups: identification, nominal power and Forced Outage Rate (FOR). The power of hydro power plants is not constant but varies significantly depending on water incoming flow. The average power of all hydro power plants through the year is around 40% of nominal power. The data is obtained from the real power system of the region.

The variants of the power system are as follows:

– Variant 1: nuclear power plant is replaced by three wind power fields, each consisting of several wind turbines. Each of wind fields has nominal power of 1154.9 MW (all three: 3470.7 MW). Their common nominal power suits the ratio of load factors for nuclear versus wind, which means that wind power plants need approximately 5 times larger nominal power than replaced nuclear power plant.

Table 2. Power system data – base case.

Power plant identification and net electrical power	FOR
Nuclear Power Plant, 696 MW	0.99
Thermal Power Plant (coal), TEŠ6, 544 MW	0.92
Thermal Power Plant (coal), TEŠ5, 345 MW	0.91
Thermal Power Plant (coal), TEŠ4, 275 MW	0.91
Thermal Power Plant (gas), TEŠG, 84 MW	0.94
Thermal Power Plant (coal), TETOL, 124 MW	0.94
Thermal Power Plant (gas), TEB, 297 MW	0.96
Hydro Power Plant, DEM, 590 MW	0.99
Hydro Power Plant, SENG, 321 MW	0.99
Hydro Power Plant, SEL, 118 MW	0.99
Hydro Power Plant, HESS, 156 MW	0.99
Hydro Pump Storage Power Plant, Avče, 185 MW, (180 MW for pumping mode)	n/a

– Variant 2: nuclear power plant is replaced by two wind power fields, each with power of 1735.35 MW (both wind power fields have together 3470.7 MW), each consisting of several wind turbines and a large power plant as a reserve power—gas power plant (e.g. 348 MW).

3.2 *Models – variant 1*

Table 3 presents the data for the variant 1 of the power system.

3.3 *Models – variant 2*

Table 4 presents the data for the variant 2 of the power system.

Table 3. Power system data for variant 1.

Power plant identification and net electrical power	FOR
Thermal Power Plant (coal), TEŠ6, 544 MW	0.92
Thermal Power Plant (coal), TEŠ5, 345 MW	0.91
Thermal Power Plant (coal), TEŠ4, 275 MW	0.91
Thermal Power Plant (gas), TEŠG, 84 MW	0.94
Thermal Power Plant (coal), TETOL, 124 MW	0.94
Thermal Power Plant (gas), TEB, 297 MW	0.96
Hydro Power Plant, DEM, 590 MW	0.99
Hydro Power Plant, SENG, 321 MW	0.99
Hydro Power Plant, SEL, 118 MW	0.99
Hydro Power Plant, HESS, 156 MW	0.99
Wind Power Plant 1, 1156.9 MW	0.99
Wind Power Plant 2, 1156.9 MW	0.99
Wind Power Plant 3, 1156.9 MW	0.99
Hydro Pump Storage Power Plant, Avče, 185 MW, (180 MW for pumping mode)	n/a

Table 4. Power system data for variations.

Power plant identification and net electrical power	FOR
Thermal Power Plant (coal), TEŠ6, 544 MW	0.92
Thermal Power Plant (coal), TEŠ5, 345 MW	0.91
Thermal Power Plant (coal), TEŠ4, 275 MW	0.91
Thermal Power Plant (gas), TEŠG, 84 MW	0.94
Thermal Power Plant (coal), TETOL, 124 MW	0.94
Thermal Power Plant (gas), TEB, 297 MW	0.96
Hydro Power Plant, DEM, 590 MW	0.99
Hydro Power Plant, SENG, 321 MW	0.99
Hydro Power Plant, SEL, 118 MW	0.99
Hydro Power Plant, HESS, 156 MW	0.99
Wind Power Plant 1, 1735.35 MW	0.99
Wind Power Plant 2, 1735.35 MW	0.99
Reserve Power Plant (Gas), 348 MW	0.99
Hydro Pump Storage Power Plant, Avče, 185 MW, (180 MW for pumping mode)	n/a

3.4 *Analysis and results – base case*

Loss of load expectation is calculated as 9.92 hours per year (0.027 hours per day) as the average of a number of calculations, where maximum loss of load expectation is 73.4 hours per year and minimum LOLE is 0.12 hours per year. The differences in performed calculations are due to the differences of hydro power plants, which are not assumed nominal all the time, but their power depends on incoming river flow through the year and changes respectively in the considered calculations.

If additional base load power plant of 348 MW is introduced, loss of load expectation is approximately ten times lower: 0.96 hours per year (0.0026 hours per day) as the average of a number of calculations, where maximum loss of load expectation is 8 hours per year and minimum is 0.008 hours per year.

If only the nuclear power plant is removed from the model and the rest stays as it is, the loss of load expectation increases to average of 419 hours per year (1.1 hours per day) with maximum of 2074.6 hours per year and minimum of 9.1 hours per year.

3.5 *Analysis and results – variant 1*

Nuclear power plant is replaced with three wind power plants in the model and model has been evaluated for 3 examples of wind power versus weather parameters. 3 examples of different functions of wind power as the function of time are considered, where the wind speed and other weather parameters direct the wind power.

Each of examples gives different wind power in configuration points, which can be seen also as time points and consequently the power of available power plants differs among configurations and thus loss of load expectation is different.

Figure 2 shows wind power for example 1. Figure 3 shows loss of load expectation for example 1.

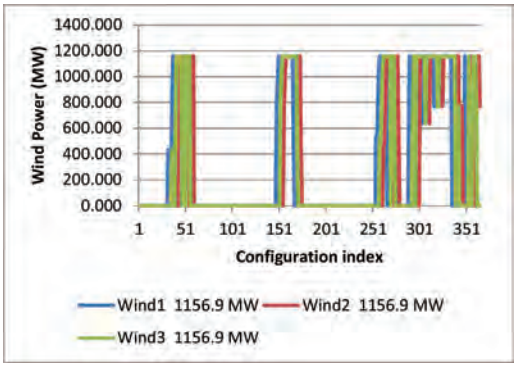


Figure 2. Wind power—variant 1 – example 1.

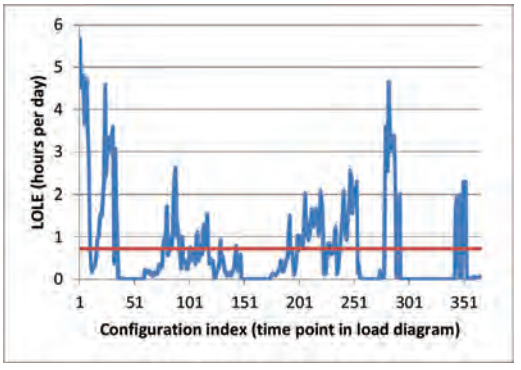


Figure 3. LOLE—variant 1 – example 1.

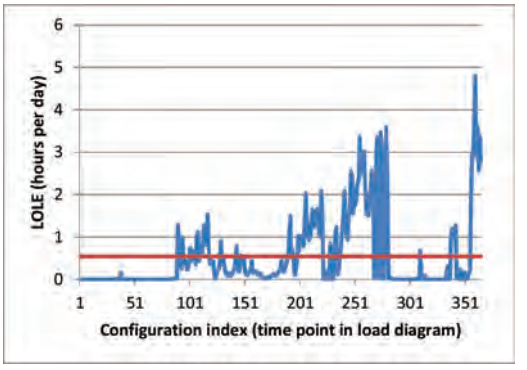


Figure 5. LOLE—variant 1 – example 2.

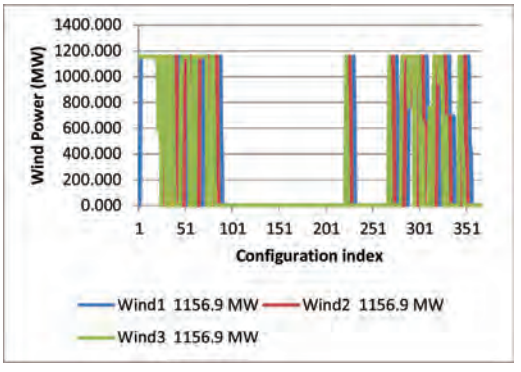


Figure 4. Wind power—variant 1 – example 2.

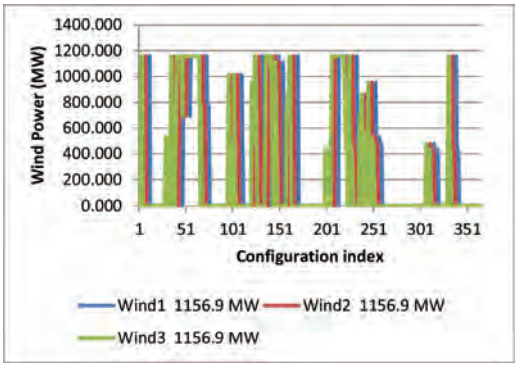


Figure 6. Wind power—variant 1 – example 3.

Loss of load expectation increased significantly with replacement wind for nuclear, which can indicate a significant decrease of power system reliability: for example 1 it is 262 hours per year (26 times increase) as the average of a number of calculations, where maximum loss of load expectation is 2075 hours per year and minimum is so low that it can be rounded to 0.

Figure 4 shows wind power for example 2. Figure 5 shows loss of load expectation for example 2.

It is 199 hours per year as the average of a number of calculations, where maximum loss of load expectation is 1755 hours per year and minimum is so low that it can be rounded to 0.

Figure 6 shows wind power for example 3. Figure 7 shows loss of load expectation for example 3.

It is 268 hours per year as the average of a number of calculations, where maximum loss of load expectation is 1755 hours per year and minimum is so low that it can be rounded to 0.

If the results of variant 1 are summarized, one can observe that the nuclear cannot be replaced by wind without significant other changes in power

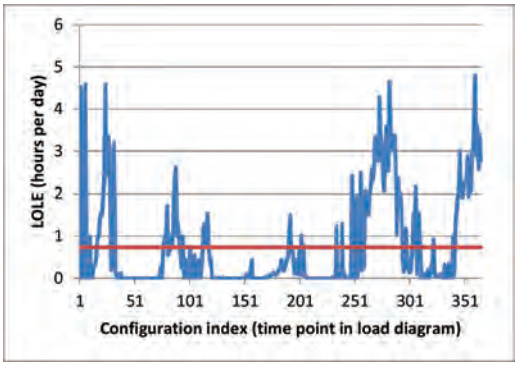


Figure 7. LOLE—variant 1 – example 3.

system such as more reserve power. In addition other factors, which are not discussed in this paper are certainly needed to be considered for such an exchange. At least some of the most important factors are listed here: frequency control of the power system, voltage control of the power system, improved rules for transmission system operation,

improved rules for distribution system, and improved rules for system operation in all aspects, which depend on intermittent power changes and the needed procedures.

3.6 *Analysis and results – variant 2*

Nuclear power plant is replaced with two wind power plants in the model and reserve power of its half power and model has been evaluated for 3 examples of wind power versus weather parameters.

Each of examples gives different wind power in configuration points, which can be seen also as time points and consequently the power of available power plants differs among configurations and thus loss of load expectation is different.

Figure 8 shows wind power for example 1. Figure 9 shows loss of load expectation for example 1.

Loss of load expectation increased notably with replacement of nuclear with wind and reserve power, which can indicate a significant decrease of power system reliability: for example 1 loss of load expectation is 47.8 hours per year (more than four times increase from base case) as the average of a number of calculations, where maximum loss of load expectation is 470 hours per year and minimum is so low that it can be rounded to 0.

Figure 10 shows wind power for example 2. Figure 11 shows loss of load expectation for example 2.

It is 46.7 hours per year as the average of a number of calculations, where maximum loss of load expectation is 376 hours per year and minimum is so low that it can be rounded to 0.

Figure 12 shows wind power for example 3. Figure 13 shows loss of load expectation for example 3.

It is 46.5 hours per year as the average of a number of calculations, where maximum loss of load expectation is 470 hours per year and minimum is so low that it can be rounded to 0.

If the reserve power is increased to 620 MW instead to 348 MW, the loss of load expectation is at approximate similar level as at the base case (with 696 MW of nuclear power).

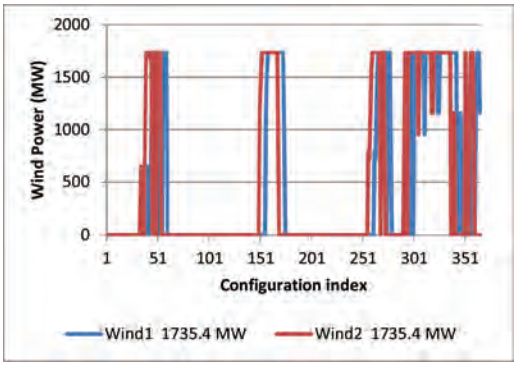


Figure 8. Wind power—variant 2 – example 1.

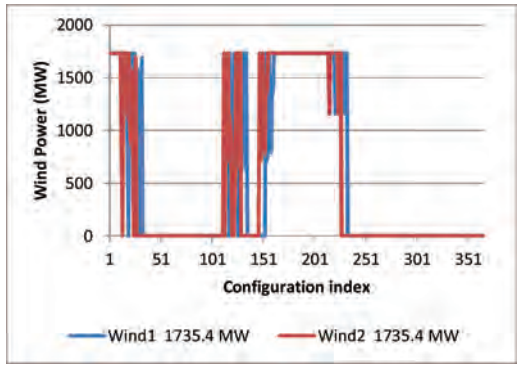


Figure 10. Wind power—variant 2 – example 2.

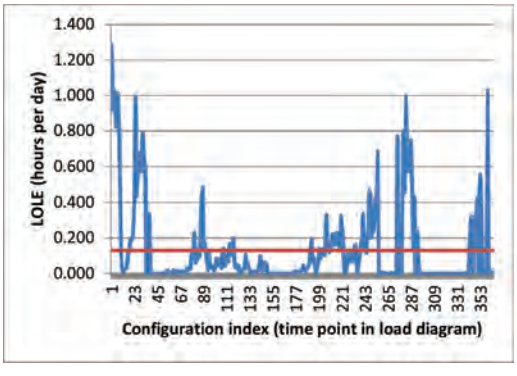


Figure 9. LOLE—variant 2 – example 1.

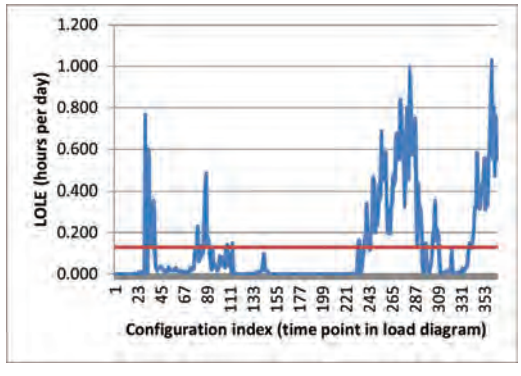


Figure 11. LOLE—variant 2 – example 2.

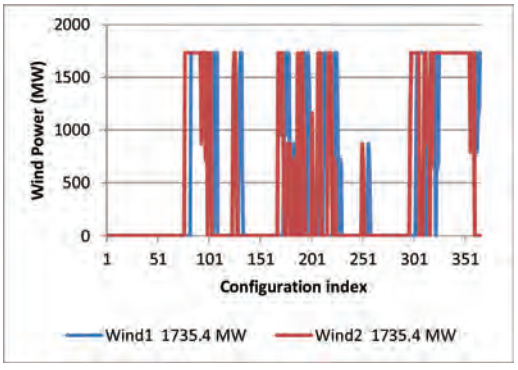


Figure 12. Wind power—variant 2 – example 3.

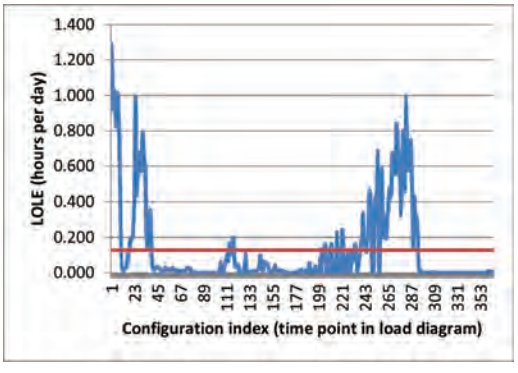


Figure 13. LOLE—variant 2 – example 3.

If the results of variant 2 are summarized, one can observe that the nuclear cannot be replaced by wind and half of its power with reserve power without significant reduction of power system reliability. In addition, other changes in the power system need to be performed (see the comments on that in the previous section).

4 CONCLUSIONS

The objective of the paper was to present realistic examples of the improved method for calculation of loss of load expectation to foster discussions about the amount of reserve power needed in the power system with increasing portions of renewable sources replacing conventional power plants.

The method was presented and improved. The computer code for the application on real examples was developed. The realistic cases have been analysed and the results were obtained.

The results show significant decrease of power reliability or in other words: a significant increase of loss of load expectation, in the case that nuclear power plant is replaced with the wind power plants of approximately five times larger joined nominal power.

The results show less significant increase of loss of load expectation, if replaced wind power goes along with additional reserve power, e.g. gas power plant of half of nominal power of nuclear power plant. But still several times of increase of loss of load expectation is present.

If the nuclear power plant is replaced with five times larger nominal wind power and with additional reserve (around of 90% of replaced nuclear power), the loss of load expectation shows that the power system reliability from the static point of view is not decreased. More sensitivity studies are needed for justifying this statement in more details.

If the number of wind power plants is larger, the variability of wind power may be smaller than shown at these examples (variant 1 and variant 2) and less reserve power may be needed (than indicated 90%). If the power system consists of several hydro power plants with large accumulations, which by themselves introduce the reserve accumulation in the power system, the problem can be solved with less additional power.

The probabilistic methods to calculate contribution of renewable sources in the power system may be used in this sense.

Future work can be oriented to consideration of probability distributions of renewable power to assess probability distributions of loss of load expectation.

ACKNOWLEDGMENT

The Slovenian Research Agency supported partially this research within the research program P2-0356.

REFERENCES

Anders G.J (1989) *Probability concepts in electric power systems*, John Wiley and Sons.

Billinton R., Allan R. (1996) *Reliability evaluation of power systems. Plenum Press*, New York.

Bricman Rejc Ž., Čepin M. (2013a). An improved method for power system generation reliability assessment (in Slovenian), *Elektrotehniški vestnik,* Vol. 80, no. 1/2, p. 57–63.

Bricman Rejc Ž., Čepin M. (2013b). Power system reliability assessment: implementation of common-cause failures. *Proceedings of Powertech,* p. 1–6.

Bricman Rejc Ž., Čepin M. (2014) Estimating the additional operating reserve in power systems with installed renewable energy sources. *International Journal of Electrical Power & Energy Systems*, Nov. 2014, vol. 62, str. 654–664.

Calabrese G. (1947) Generating reserve capacity determined by the probability method, *AIEE Trans* 66:1439–50.
Čepin M, (2011) *Assessment of power system reliability*, Springer.
Elmakias D. (2008) *New computational methods in power system reliability*, Springer Verlag Berlin Heidelberg.
IEEE Std 1366 (2003) Guide for electric power distribution reliability indices, IEEE.
Kirn B., Čepin, M., Topič M. (2017) Effective load carrying capability of solar photovoltaic power plants—case study for Slovenia, Safety & reliability: theory and applications: *Proceedings of the 27th European Safety and Reliability Conference, ESREL* 2017, Portorož, Slovenia, 18–22 June 2017, Taylor & Francis, p. 3231–3239.
Xiaoming F. (1990), *The probabilistic production simulation of electric power systems using equivalent load duration curve methods*, PhD Thesis, Ohio University.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Losing containment at high temperature and pressure—an experimental study with water-steam circuit

F. Heymes, P. Lauret & C. Lopez
LGEI, IMT Mines Ales, University of Montpellier, Ales, France

P. Hoorelbeke
Total SA

ABSTRACT: The design of pressure vessels aiming to contain water under high pressure and temperature conditions considers steady state behavior of the fluid. But if pressure decreases suddenly thanks to a leak or a rupture, water can get into a supercritical state. Supercritical state is a high energy and metastable state that could lead to a catastrophic accident such as explosion. The phase change of the superheated fluid is expected to entail a violent repressurization of the vessel, and may blow up the vessel. Little data can be found in literature about that repressurization process. The aim of this work was there to measure the repressurization dynamics of superheated water following a loss of containment. Experiments were performed at high temperature and pressure (up to 315°C, 100 bar) in order to understand the phenomenon. The pressure drop in the tank was fast and well below the saturation pressure at the given temperature. Then, a repressurization peak occurred, depending on the pressure drop history. The set of results is discussed and compared to literature data.

1 INTRODUCTION

Steam is widely used in industry to carry heat. After condensation, hot water flows usually back to a furnace to be reheated and vaporized. Steam is produced and used at different temperatures and pressure. Typically, steam below 3.5 barg is termed as low pressure steam. Steam above 3.5 barg but below 17.5 barg is termed as medium pressure steam and steam above 17.5 barg is termed as high pressure steam. Some users define their steam above 40 barg as ultra-high pressure steam.

A water-steam circuit will comprise sections of piping where water flows at high pressure and high temperature. In case of a leak the hot water will undergo a rapid transformation into steam due to violent boiling or flashing. This phenomenon is called a steam explosion. The higher the degree of superheat (i.e. the difference between the temperature of the water and the atmospheric boiling temperature of the water) the more violent will be the explosion. The water vaporizes from liquid to vapor with extreme speed, increasing dramatically in volume. A steam explosion sprays steam and boiling-hot water and the hot medium that heated it in all directions, creating a danger of burning. Some steam explosions appear to be special kinds of Boiling Liquid Expanding Vapor Explosion (BLEVE), and rely on the release of stored superheat.

This presentation presents (i) thermodynamics considerations about the superheated state of water, (ii) the experimental setup, (iii) results about the depressurization dynamics in the vessel and the conclusions that can be driven for the collected data.

2 THERMODYNAMIC CONSIDERATIONS

2.1 *Phase diagram (P-T)*

Water thermodynamic properties were widely investigated previously. Phase change lines, triple point and critical point are given on Figure 1. Experimental data about triple point and critical point are given in Table 1.

A specific state of water is called the superheated state. This state (sometimes referred to as boiling retardation or boiling delay) is usually described as the phenomenon in which a liquid is heated to a temperature higher than its boiling point, without boiling. Superheating is achieved by heating a homogeneous substance in a clean container, free of nucleation sites, while taking care not to disturb the liquid.

More generally, a liquid is said to be superheated when its temperature exceeds its saturation temperature of its pressure or when its pressure decreases below its saturation pressure of its temperature

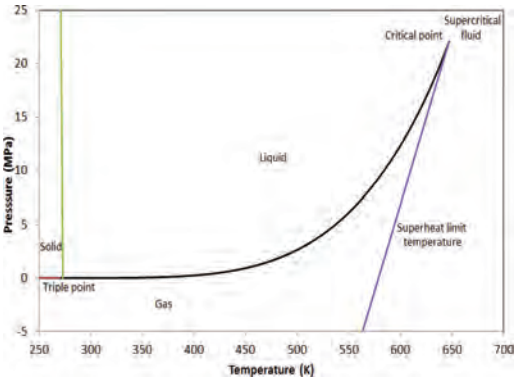


Figure 1. Phase change diagram of water in (T,P) coordinates.

Table 1. Specific thermodynamic points of water (Abbasi & Abbasi 2007).

	Temperature		Pressure	
	K	°C	Pa	atm
Triple point	273.15	0.01	661.73	0.00653
Critical point	647.37	374	22 120	218

while the liquid is still not boiling: $T_L > T_{sat}(P_L)$ or $P_L < P(T_L)$. Superheating may happen at any pressure below the critical point. Superheat domain is usually delimited by a line defined between the critical point and the superheat limit temperature at atmospheric pressure.

2.2 *The superheat state*

The superheat state may be described on the usual P-V diagram as following (Figure 2). C is the critical point. [BCB'] is the saturation curve or the binodal. The isotherm at temperature T = 580 K (306.85°C) is [ABEFB'D]. B and B′ are equilibrium states on the binodal. P_{sat} is the equilibrium pressure at T.

When the liquid state is between A and B, it is called the subcooled liquid. The liquid at point B is called the saturated liquid. When the liquid state is between B and E, it is called the superheated liquid because its temperature has been higher than the saturation temperature of its pressure or its pressure has been lower than the saturation pressure of its temperature.

When the liquid becomes superheated, it also becomes metastable which means its stability can be easily broken by external perturbations. If so, it can no longer maintain its liquid state and phase transition must occur. When the metastability of the liquid becomes larger (the liquid is approaching point E), the minimum perturbation required to break the stability of the liquid becomes smaller and finally at point E, the thermodynamic stability limit has been reached, which means phase transition will spontaneously occur without any external perturbations or without any suitable nucleation site. The stability of the liquid can be broken by the density fluctuations of the liquid itself.

2.2.1 *Kinetic Superheat Temperature (KSL)*

In experiments on superheating measurements of a liquid, bubble nucleation (generation of small bubbles) will start when point K on the isotherm in Figure 2 is reached. Point K is called the Kinetic Superheat Limit (KSL). The KSL can be measured experimentally provided early bubble generation by impurities of wall effects is prevented (no heterogeneous nucleation). Avedisian (1985) did a comprehensive work about the KSL measurement and reported values of KSL from the literature. A large discrepancy was observed due to different operating data.

2.2.2 *Thermodynamic Superheat Limit (TSL)*

Point E is called the Thermodynamic Superheat Limit (TSL) of the liquid and [CE] is called the superheated liquid spinodal. The phase separation occurring at the TSL is called spinodal decomposition. On Figure 2, [CE], [CF] and [BEFB'] curves are not experimental but only illustrative, since experimentally the spinodal decomposition has only been observed by light scattering techniques at a temperature very close to the critical point in binary mixture systems and the process is too fast to allow transient measurement of thermodynamic properties.

The thermodynamic superheat limit (TSL) curve is defined by the spinodal curve, given by the equation:

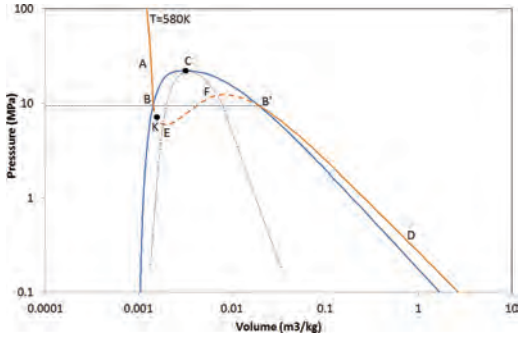


Figure 2. Coexistence line (BCB'), isotherm at equilibrium (ABB'D), liquid spinodal (CE) and vapor spinodal (CF) in case of water.

$$\left(\frac{\partial P}{\partial V}\right)_T = 0 \tag{1}$$

This equation is a condition of thermodynamic stability. It separates two regions on the PT diagram (Figure 1) separated by the superheat limit curve, defined by the value of the pressure gradient:

If $\left(\frac{\partial P}{\partial V}\right)_T < 0$ the system is stable

If $\left(\frac{\partial P}{\partial V}\right)_T > 0$ the system is unstable

TSL can be predicted by use of equations of state (EOS). There are several equations of state, depending on the considered pressure and temperature range and the molecular structure of the compound. (Reid 1979) used the Redlich Kwong EOS to predict the thermodynamic superheat limit temperature. (Kim-E 1981) proposed to use the Peng Robinson equation (PR EOS):

$$P = \frac{RT}{V-b} - \frac{a}{V(V+b)+b(V+b)} \tag{2}$$

By deriving this equation, the following relation has to be solved:

$$V^4 + \left(4b - \frac{2a}{RT}\right)V^3 + 2b\left(b + \frac{a}{RT}\right)V^2 + 2b^2\left(\frac{a}{RT} - 2a\right)V + b^4 - \frac{2ab^3}{RT} = 0 \tag{3}$$

where V is the molar volume, a and b are constants depending on which specie is being analyzed. The constants can be calculated from the critical point data of the gas. For a temperature below the critical one, the previous equation will yield four real roots. One will be less than the saturated liquid volume and another will be greater than the saturated vapor volume; these are disregarded. The two remaining roots are between the liquid and vapor saturation curves; the smallest root corresponds to the liquid spinodal curve and the largest to the vapor spinodal curve. The TSL can therefore be calculated at atmospheric pressure by solving this equation.

Other equations of state can be used. (Abbasi & Abbasi 2007) used 9 models to calculated TSL in case of water (Table 2): Van der Waals (VDW), Soave Redlich Kwong (SRK), Peng Robinson (PR), Twu-Redlich-Kwong (TRK) or Peng-Robinson-Mathias-Copeman (PRMC), Berthelot (B). According to the authors calculations, the TSL varied in the range [546.6–604.5] K, that is a range of ΔT = 58 K depending on the considered EOS.

2.2.3 *Superheat limit curve*

Where does the superheat limit curve drawn on Figure 2 locate? Reid (1979) claimed that the thermodynamic superheat limit TSL predicted by the Redlich-Kwong equation of state is reasonably in agreement with experimental data, but he doubted those results because "...no satisfactory correlation now exists to relate p, v and T in the superheated liquid region...". An equation of state is obtained by correlating experimental data outsides the saturation dome. Using an equation of state for metastable states inside the saturation dome (Figure 2) is equivalent to extrapolation of those experimental data. For slightly superheated liquid, the extrapolation is still reliable (Mengmeng 2013), but for highly superheated liquid states, Reid's worry is inevitable. The accuracy of an equation of state to predict the TSL cannot be demonstrated by such comparisons unless the measured KSL has been proven to be very close to the real TSL, which has not been done in any experiment yet. If the validity of an equation of state in predicting the thermodynamic superheat limit has not been proven, the close agreement of its predictions with the measured superheat limit cannot be interpreted as matching.

In case of water, Abbasi compared TSL calculations to experimental KSL at the atmospheric pressure and calculated absolute deviations. For water, the VDW-EOS gives the minimum average absolute deviation of 1.20% from the experimental value; The Twu-Redlich-Kwong equation of state (TRK-EOS), gives 5.90%; The PR-EOS and its modified version, the Peng-Robinson-Mathias-Copeman equation of state (PRMC-EOS) give the deviations of 7.90% and 8.02% respectively; The Berthelot equation of state (Berthelot-EOS) gives 7.53%; The RK-EOS gives 4.76% and the Soave-Redlich-Kwong equation of state (SRK-EOS) gives 9.27%. For more information refer to the study of Abbasi & Abbasi (2007) or Salla et al. (2006).

According to these considerations, it appears that the superheat limit temperature of water is still not well defined. The KSL seems more reliable since it can be measured experimentally, keeping in mind that the value depends strongly on the surface state of the apparatus.

Two values of KSL at atmospheric pressure were found in the literature: Avedisian et al proposed a KSL of 575.1 K (301.95°C) whereas Abbasi et al proposed a KSL for water at atmospheric pressure of 553.2 K (280.05°C).

2.3 *Phase change of superheated liquid*

The key point of this work is the violent phase change of the superheated liquid. Therefore a focus has to be done on the phase change

dynamics. Two kinds of phase change have to be considered: homogeneous and heterogeneous nucleation. When a liquid becomes superheated, vapor embryos can be formed because the excess energy of the superheating can be used to cover energy needed for phase change and to maintain surface tension. This process called bubble nucleation, has two forms, depending on the locations where vapor embryos form:

- Homogeneous nucleation, as the name indicates, occurs in the middle of the fluid where no phase boundaries are present;
- Heterogeneous nucleation occurs on phase boundaries such as rough walls or suspending solid impurities.

To form a vapor embryo with the same volume, heterogeneous nucleation requires less energy than homogeneous nucleation because the presence of the phase boundaries allows a lower interface area of the vapor embryo. Generally speaking, heterogeneous nucleation occurs at a lower degree of superheat than homogeneous nucleation.

Distinction has to be made between nucleation and bubble growth in the metastable state but still far away from the KSL and nucleation and vaporization close to the KSL. In conditions away from the KSL, homogeneous nucleation is much less probable than heterogeneous nucleation and the question how fast the nucleation is, also involves questions on availability and properties of surfaces for nucleation. Avedisian et al (1985) suggested that the nucleation rate J (nuclei/cm^3.s) depends on the superheat level ΔT (Figure 3). J_0 defines the minimum nucleation rate below which homogeneous nucleation is unlikely. The superheat ΔT_1 and ΔT_2 correspond to rates J_1 and J_2 respectively. The superheat at J_{max} is shown corresponding to the thermodynamic limit ΔT_t.

When a liquid is superheated but is still far away from the superheat limit, vaporization will start first on locations most favorable for the formation of initial small bubbles which is on solid surfaces or on dust or other solid particles in the fluid. Once bubbles are formed they grow according growth laws which have been well studied in the literature.

Approaching to the superheat limit state, the process generating smaller bubbles (nuclei) is different than away from the superheat limit state because the small 'vapor' nuclei originate homogeneously in the fluid and at a much faster rate. The smallest stable bubbles close to the KSL are much smaller than away from the KSL. Classical homogeneous nucleation theory has been developed to describe this case. It turns out that its predictions of nucleation rate are extremely dependent on details.

The growth of bubbles at the superheat limit also proceeds in a different manner than away from the superheat limit. The rapid growth of bubbles

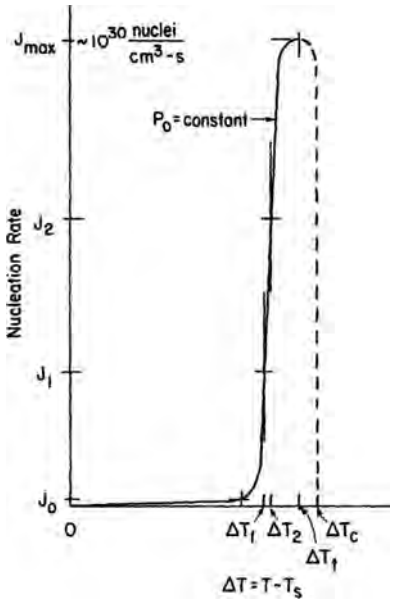


Figure 3. Schematic variation of nucleation rate with temperature at a given ambient pressure (Avedisian 1985).

may create a rise in pressure which counterbalances the initial pressure drop, and keeps the liquid away from the superheat limit. This phenomenon must be taken into account in the evaluation of the effects of rapid depressurization or rapid heating.

Moreover, during a depressurization of a superheated liquid, the decrease in pressure is not felt instantaneously all over the liquid but spreads in the form of a wave. The local temperature at different locations in the liquid may be different, depending on the features of the accident causing the vessel and the amount of heat used for vaporization.

The starting points at different locations in the liquid may be at different locations on the saturation curve. The decrease in pressure due to opening of the vessel will be felt at different locations at different moments in time and once vaporization starts the decreases in pressure can be stopped by the expansion of the mixture due to vapor generation. The rate of depressurization is also controlled by the time needed for vessel rupture.

This has been taken into account in a refinement of the superheat limit theory proposed by (Mcdevitt 1990). As summarized in the review proposed by Leslie & Birk (1991), the homogeneous boiling only occurs at the rupture location where the liquid sucked out of the breach first reaches atmospheric pressure. Their study focused on the liquid behavior inside the vessel (liquid hammer, pressure recovery etc.), before a possible total disintegration of the vessel. A vaporization front can have very complex shape and wildly fluctuating properties.

The break-up of existing bubbles in smaller bubbles in a wildly fluctuating vaporization front creates extra area and also new nuclei for heterogeneous vaporization. This can enhance the vaporization rate and the front propagation rate enormously. The question arises whether the vaporization in a propagating front can generate a sufficiently strong volume source for significant blast propagation.

These considerations indicate that apart from experiments and thermodynamic models to determine the superheat limit curve for various substances, also experiments and fluid dynamic models are needed to determine the propagation speed of vaporization fronts.

3 THE BOILING LIQUID EXPANDING VAPOUR EXPLOSION (BLEVE)

A superheated liquid is in a high energy state and a metastable equilibrium. Therefore it can release a large amount of energy in explosive behavior. A superheated liquid explosion requires that a large part of the liquid vaporizes in very short time. Different behaviors were described in literature. Two main categories can be made, according to the way the liquid becomes superheated:

- By sudden pressure loss, such as observed in boiling liquid expanding vapor explosion (BLEVE)
- By sudden temperature increase, such as observed in rapid phase transition (RPT)

BLEVE is the most common phenomena which resulted in many studies. The standard theory of BLEVE was originally proposed by (Reid 1979). The essential idea is illustrated Figure 4. Under normal conditions the content of the vessel containing a liquid and its vapor is in thermodynamic equilibrium and the pressure and temperature combination lies at the saturation curve (points A or C). In the case of vessel rupture the pressure suddenly decreases resulting in superheated liquid. There is a limit to the degree in which a liquid can get superheated. At constant pressure, the superheat limit temperature is the highest temperature that a liquid can sustain without undergoing phase transition and at constant temperature; the superheat limit pressure is the lowest pressure for a liquid to maintain its liquid state. According to Reid's theory, when the pressure of the liquid decreases from point C to D, the liquid reaches the superheat limit curve and a BLEVE will occur while in the process of A to B, the liquid does not reach the superheat limit curve, no BLEVE will occur. Based on previous part, the superheat limit referred in Reid's theory should be the kinetic superheat limit (KSL).

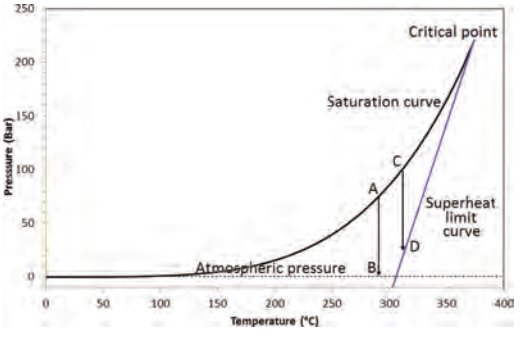


Figure 4. Schematic explanation of Reid's superheat limit theory for BLEVE in depressurization processes.

As explained previously, the situation in a real incident will be much more complicated than the simple trajectories from A to B or from C to D on Figure 4. In Reid's superheat limit theory (Reid 1976), the severity of the hazard of a BLEVE is attributed to the fact that the KSL has been reached, not the TSL. Throughout a study on BLEVE one should bear in mind the difference between the two superheat limits definitions and the difference between an experimentally observed phenomenon (KSL) and a theoretical thermodynamic property (TSL).

Direct correspondence between a BLEVE and the spinodal decomposition has never been proven by experimental data. Recently Birk (Birk et al. 2007) after an analysis of a number of medium scale BLEVE tests have come to the conclusion that in the case of rupture of high pressure vessels only partially filled with liquid propane (fill level in the range 13 to 61%) the shock waves observed in the far field seem rather produced by expansion of the vapor and not by the vaporization of the liquid, which is said to be a too slow process for generating a strong blast. In (Birk et al. 2007) it is mentioned however that the rapid vaporization process can produce significant dynamic pressure effects in a near field. These effects are of particular importance in case of a BLEVE in a confined space such as a tunnel. Their demonstration does not involve superheat limit theory but uses a thermodynamic estimate of the available energy. Such estimates can assume isentropic expansion (Prugh 1991) or, more realistic, adiabatic irreversible expansion (Planas-Cuchi et al. 2004). The second estimate is about half of the first (Abbasi & Abbasi 2007).

4 MATERIALS AND METHODS

The apparatus was designed to heat water at high temperature and pressure and to release pressure thanks to a ½″ fast ball valve. The specifications

of the vessel are given in Table 2. A sketch and a picture of the setup are given on Figure 5.

Eight K type thermocouples were set on a vertical axis in the vessel. A stirrer was put at the symmetry axis of the vessel, so the thermocouple axis was shifted 5 cm on the side of the stirrer axis. Location data are reported in Table 3.

The power of the heater was 8000 W, with a maximum wall temperature of 450°C, so one hour was necessary to reach the target temperature (300°C). The apparatus was be completely insulated to minimize heat losses.

Two water cooled dynamic pressure sensors (Kistler 601C) were put on the vessel to measure the transient pressure in the vessel. Data acquisition rate was 200 kHz. Two static pressure sensors were put at a distance from the vessel (250 bar) and remained cold during the tests. A half inch fast discharge pneumatic valve was put at the top of the vessel and controlled remotely.

Table 2. Experimental vessel specifications.

Pressure vessel standard	CODAP 2005
Volume	5 L
Steel	X2CrNiMo17-12
Maximum allowable working pressure	142 bar
Working temperature	0°C to 300°C
Design pressure	190 bar
Design temperature	300°C
Test pressure	305 bar
Safety valve pressure	183 bar
Internal diameter	125 mm
Internal height	395 mm
Wall thickness	11.5 mm
Vessel volume	5000 mL
Total volume (incl. pipes)	5090 mL

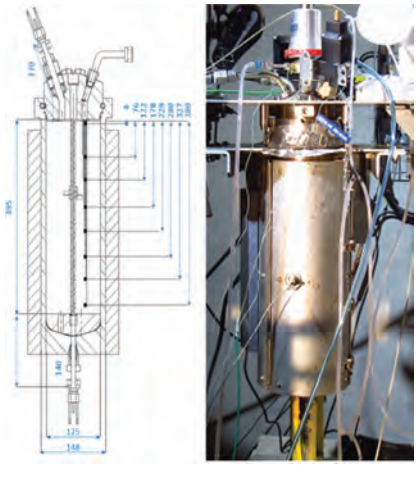


Figure 5. Detailed sketch of the experimental vessel.

Table 3. Location of the thermocouples in the vessel.

Thermocouple number	Distance from bottom mm	Distance from top mm
8	391	4
7	319	76
6	273	122
6	217	178
4	166	229
3	115	280
2	68	327
1	15	380

5 RESULTS AND DISCUSSION

A series of 8 experiments was performed. Operating conditions of the tests are given in Table 4.

5.1 *Test description*

The experimental vessel was filled with an accurate quantity of high purity water (milliQ grade). The gas space remained filled with air at atmospheric pressure. The heater was switched on. In order to purge the air, the relief valve remained open until vapor was observed at the exit of the valve, typically during 2–3 minutes. Then the valve was closed. Very little water was lost during the air purge. During this purge dissolved gases were also removed.

Figure 6 shows the water temperature evolution on a vertical axis during the test. T1 was located at the lowest level whereas T8 was located at the uppermost level in the tank. Temperature stratification occurred during the ten first minutes. Then, boiling at the wall created bubbles and turbulence that are clearly observable on the temperature curves. When the boiling was sufficient to provoke a complete mixing of the liquid, all temperature records converged to a single temperature curve. The internal pressure increased according to the vapor-liquid equilibrium law (Figure 7).

A perfect fit is observed with the literature data on the (T,P) diagram of water (Figure 8). When the temperature target was reached, the fast discharge valve was opened. A powerful and noisy steam jet was created. A noise recording was performed at the operator bunker and indicated a level of 102 dB. The temperature and pressure dropped very rapidly.

These data were recorded at an acquisition rate of 20 Hz. In order to record perfectly the fast pressure dynamics in the tank, two dynamic pressure gauges and a high speed recording equipment enabled to get data at a 200 kHz recording rate.

After a strong depressurization, a small repressurization occurred and was measured with both

Table 4. Summary of experiments.

Experiment	Water quantity (g)	Water temperature at relief time	Pressure at relief time P_R	Volume fraction at T_R
	g	°C	bar	%
1	3995	202	17	91%
2	3500	236	32	89%
3	3425	267	53	90%
4	3730	281	66	80%
5	3700	290	76	90%
6	3899	300	88	96%
7	3913	281	67	95%
8	3854	281	68	90%

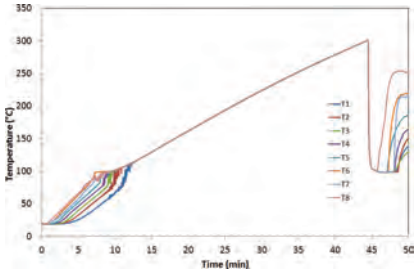


Figure 6. Temperature of the fluids (Test #6).

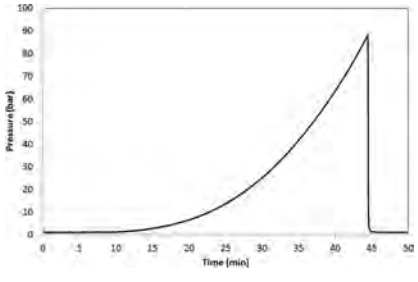


Figure 7. Pressure in the tank (Test #6).

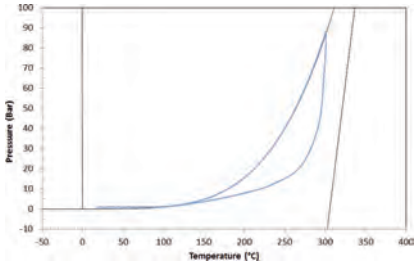


Figure 8. Thermodynamic transformation on phase diagram (Test #6).

pressure transducers. This repressurization was very short in time; quickly the depressurization overcame the repressurization phenomena and the

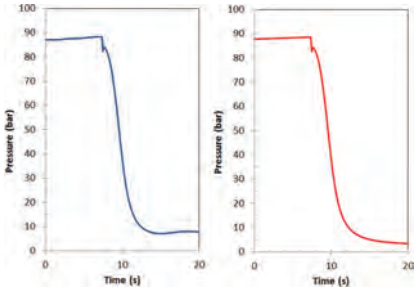


Figure 9. Dynamic pressure in the tank (200 kHz, test #6).

pressure restarted to decrease. In the considered case, a depressurization of 4.7 bar occured during 110 ms and was followed during 140 ms by a 1.8 bar pressure increase. Then, the pressure restarted to decrease to reach the atmospheric pressure. The data of both dynamic pressure transducers are reported on Figure 9. The data on the left corresponds to pressure sensor located at the bottom of the vessel, the data on the right corresponds to pressure sensor located at the top of the vessel. Both data match perfectly.

5.2 *Influence of superheating temperature*

According to the theory, nucleation and repressurization rate are expected to depend on the superheat level. Therefore five experiments were performed at different temperatures: 236°C, 267°C, 281°C, 290°C and 300°C.

The transient pressure data are given on Figure 10. In these tests, a repressurization peak was observed. The peak is more visible at higher temperatures (281°C, 290°C and 300°C), but a slight peak is also noticeable at 267°C and some oscillation can me observed at 236°C.

5.3 *Discussion*

Experimental data about kinetic superheat limit are reminded in Table 5.

Tests performed at 281, 290 and 300°C tests (Figure 10) were above the KSL and indicated a repressurization peak which seems to corroborate the proposal of (Avedisian 1985). Indeed, the pressure drop due to venting was countered by the nucleation rate in the vessel. It is not possible to check if nucleation was homogenous of heterogeneous.

The data was computed in order to get the numerical value of the pressure drop and the consequent repressurization pressure peaks (Table 6). According to the theory of (Avedisian 1985), the difference between superheated temperature of water before release and the standard boiling temperature ΔT was calculated and reported on Figure 11. This figure has to be compared with Figure 3, with a dif-

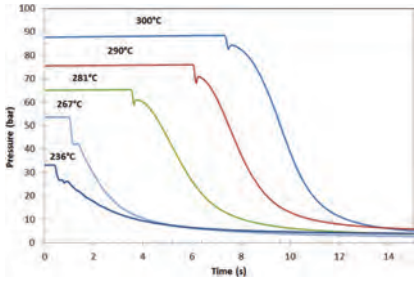


Figure 10. Temperature influence on transient pressure after depressurization.

Table 5. Experimental values of water KSL.

	KSL
	°C
Avedisian et al.	301.95
Abbasi et al.	280.05

Table 6. Pressure drop and rise after depressurization.

	Temperature	Target pressure	Pressure drop	Pressure peak
Test	°C	bar	bar	bar
3	26	53	11.68	0.42
4	28	65	6.28	1.97
5	290	76	7.72	2.62
6	300	88	5.89	1.64

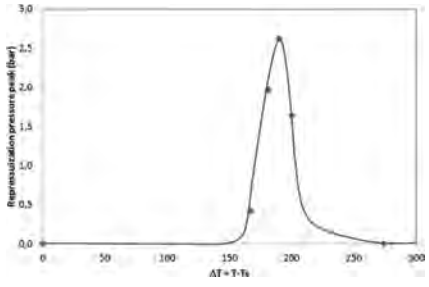


Figure 11. Repressurization peak as a function of superheated level.

ference that the nucleation rate was unknown and replaced by the repressurization peak on the y-axis. Both values are linked by non-linear relations that were not investigated in this work.

A similar trend was observed, but further work will be required to investigated more accurately the nucleation rate for comparison with the work of (Avedisian 1985).

In the frame of pressure vessels safety, results indicate that the pressure peaks remained small (< 3 bar) in comparison with the pressure in the tank at relief time. Therefore, if an industrial vessel or pipe containing superheated water is depressurized by a little hole due to corrosion of mechanical failure, no significant pressure peak able to break the vessel should be feared for the conditions studied in this article.

ACKNOWLEDGMENTS

The authors are grateful to TOTAL SA for proposing and supporting this study.

REFERENCES

Abbasi, T. & Abbasi, S.A., 2007. Accidental risk of superheated liquids and a framework for predicting the superheat limit. *Journal of Loss Prevention in the Process Industries*, 20(2), pp. 165–181. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0950423005001853 [Accessed November 6, 2013].

Avedisian, C.T., 1985. The homogeneous nucleation limits of liquids. *Journal of Physical and Chemical Reference Data*, 14(4), pp. 695–729.

Birk, A.M., Davison, C. & Cunningham, M., 2007. Blast overpressures from medium scale BLEVE tests. *Journal of Loss Prevention in the Process Industries*, 20, pp. 194–206. Available at: http://linkinghub.elsevier.com/retrieve/pii/S095042300700006X.

Kim-E, M.E., 1981. *The possible consequences of rapidly depressurizing a fluid*.

Leslie, I.R.M. & Birk, A.M., 1991. State of the art review of pressure liquefied gas container failure modes and associated projectile hazards. *Journal of Hazardous Materials*, 28, pp. 329–365.

Mcdevitt, C.A., 1990. Initiation step of boiling liquid expanding vapour explosions. *Journal of Hazardous Materials*, 25, pp. 169–180.

Mengmeng, X., 2013. *Thermodynamic and Gasdynamic Aspects of a Boiling Liquid Expanding Vapour Explosion.* Delft University.

Planas-Cuchi, E., Salla, J.M. & Casal, J., 2004. Calculating overpressure from BLEVE explosions. *Journal of Loss Prevention in the Process Industries*, 17(6), pp. 431–436. Available at: http://linkinghub.elsevier.com/retrieve/pii/S095042300400066X [Accessed November 6, 2013].

Prugh, R.W., 1991. Quantitative evaluation of BLEVE hazards. *Journal of Fire Protection Engineering*, 3(1), pp. 9–24.

Reid, R.C., 1979. Possible mechanism for pressurized-liquid tank explosions or BLEVE's. *Science*, 203(4386), pp. 1263–1265.

Reid, R.C., 1976. Superheated liquids. *American Scientist*, 64:2, pp. 146–156.

Salla, J.M., Demichela, M. & Casal, J., 2006. BLEVE: A new approach to the superheat limit temperature. *Journal of Loss Prevention in the Process Industries*, 19(6), pp. 690–700. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0950423006000180 [Accessed November 3, 2014].

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

An investigation and statistical analysis into the incidents and failures associated with dynamic positioning systems

O. Olubitan, S. Loughney & J. Wang
Liverpool John Moores University, Liverpool, UK

R. Bell
Engineering Safety Consultants Ltd., UK

ABSTRACT: This investigation highlights the safety of DP (Dynamic Positioning) systems by briefly discussing the existing risk assessment methods and risk control measures provided by the International Maritime Organization (IMO). Following a study regarding DP systems and the loss of position incidents reported to the International Marine Contractors Association (IMCA) and contained in the World Offshore Accident Databank (WOAD), the relevant hazards are identified and collated. The primary causes of loss of position incidents were found to be: the positional reference system failures and thruster failures, with both contributing to 20.6% each of the total incidents from 2000 to 2016. Similarly the time period of the analysis is the 17 years between 2000 and 2016. Given the two stated primary causes, the positional reference system failures and thruster failures, the thruster failures are analysed further. This is due to thruster failures occurring at an increased rate within DP incidents from 2011 to 2016. Hence, this trend, between 2011 and 2016, is analysed further. Three undesired events for loss of position due to thruster failures were identified, these are as follows: "drive off", "drift off" and "time loss". Further investigation of incidents in more recent years, in this case, 2012 to 2016, identified that the DP control system accounted for 33.7% of the initiating incidents that led to thruster failures. Furthermore, the undesired event "time loss" accounted for 71.2% of total incidents caused by thruster failures.

1 INTRODUCTION

Many marine vessels are now equipped with Dynamic Positioning (DP) systems. These vessels have different functions and come in different sizes. Some vessels are used for Simultaneous Operations (SIMOPS) and Combined Operations (COMOPS), bringing them in close proximity with other vessels and/or offshore installations. Some are used to conduct diving operations, with a huge number of vessels also used for drilling operations. The functions of these DP vessels are very numerous, challenging and distinct. To cater for these functions, their level of safety was required to be increased and classified. Through time, with support from different organizational bodies and from lessons learnt from past DP failures, the safety of DP systems and vessels has been increased. These DP systems now utilize a redundancy scheme to ensure that a single failure would not lead to an accident. In this research the term safety is derived from the definition of functional safety outlined by Bell (2010). This states that "*functional safety is a part of the overall safety that depends on a system or equipment operating correctly in response to its inputs. In essence, this means the achievement of safety through application of control systems. This requires identifying what has to be done and how well it should be done*" (Bell, 2010).

DP has evolved over the years, from use as a tool for mobile offshore drilling units, for maintaining position over offshore wells, to being employed for a wide range of position keeping operations. We see DP systems being fitted on an increasingly large number of new and diverse vessels, from offshore units to shuttle tankers to passenger vessels (IMO 1994).

The increase in the number of diverse applications results in an increase to the estimated risks involved with dynamic positioning vessels, further requiring an increased level of safety. To this end there has been a growth in the development of the dynamic positioning systems. There are now several classes from DP0 to DP3 (DNV 2012), each class with a different level of safety.

The rationale behind this topic is to identify, investigate and analyse the loss of position incidents due to DP system failure. These incidents are analysed

according to the seasons in which they occurred, types of vessels, main causes and initiating causes.

2 BACKGROUND

2.1 *Dynamic positioning*

DP can be defined as "a means of holding a vessel in a relatively fixed position with respect to the ocean floor, without using anchors, accomplished by two or more propulsive devices, controlled by inputs from sonic instruments on the sea bottom and on the vessel, by gyrocompass, by satellite navigation or by other means" (Holvik 1998).

The launch of the Global Positioning System Satellite network brought new ideas and new technology to be integrated into the DP vessels for more efficient performance. In 1981, the Nautical Institute began working on certification process for DP operators. This was done to reduce accidents and failures caused by human error. In 1983, the Department of Energy and the Norwegian Petroleum Directorate produced guidelines for diving from DP vessels. Howard Shatto in 1983 further improved on the DP systems, allowing greater water depths of 7,500 m and even rougher seas to be achievable. From his knowledge of satellite positioning and through his participation in the first use of Failure Mode and Effects Analysis (FMEA) for DP systems in 1983, the Mean Time Between Failures (MTBF) for DP systems was improved six-fold (DPC-MTS 1996).

By 1985, the number of DP capable vessels had increased to over 150. At this time, the vessel types equipped with DP systems had also increased. The following are some of the types of DP vessels that were available by 1985 based on their functions: Drillships, Mobile Offshore Drilling Units (MODU), Diving Semi-Submersibles, Diving and Emergency Response Vessels, Remote Operated Vessels, Diving Support Vessels, Shuttle Tankers and Accommodation Vessels (Flotels).

The following years saw an increase in the use of DP systems various functions and vessel types. In 1990, the first DP Floating Production Storage and Offloading (FPSO) vessel, Seillean, was launched by British Petroleum. The Dynamic Positioning Vessel Owners Association (DPVOA) was formed in the same year. In 1994, IMO provided guidelines for DP systems, MSC/Circ.645, the same year, the American Bureau of Shipping (ABS), introduced their first DP rules. The following year, 1995, the International Marine Contractors Association (IMCA) was formed through the merger of the Association of Offshore Diving Contractors and the DPVOA. In a bid to encourage exchange of information, foster improvement of DP reliability, develop guidelines, train and educate, and address any other issues pertinent to DP that encourage an incident free operation of DP systems. The DP committee was founded in 1996 as a Professional Committee of the Marine Technology Society (Sean 2009).

DP systems have been improved considerably since the installation of the first DP drill ship, Eureka. Vessels used for a variety functions are now equipped with DP systems. Some functions have been stated previously.

2.2 *Classification of the DP systems*

The different classes in the IMO and the ABS regulations are examined. The only difference between the two is that there is not a Class 0 for IMO regulations. ABS uses DPO to refer to vessels without DP systems. IMO does not put this as a class of DP vessels. IMO classes start from Class 1.

1. Class 0: DPS-0
   "For vessels, which are fitted with centralized manual position control and automatic heading control system to maintain the position and heading under the specified maximum environmental conditions" (ABS, 2013). Class 0 does not exist on the IMO classification.
2. Class 1: DPS-1
   "For vessels, which are fitted with a dynamic positioning system, which is capable of automatically maintaining the position and heading of the vessel under specified maximum environmental conditions having a manual position control system" (ABS, 2013).
3. Class 2: DPS-2
   "For vessels, which are fitted with a dynamic positioning system, which is capable of automatically maintaining the position and heading of the vessel within a specified operating envelope under specified maximum environmental conditions during and following any single fault, excluding a loss of compartment or compartments" (ABS, 2013).
4. Class 3: DPS-3
   "For vessels, which are fitted with a dynamic positioning system, which is capable of automatically maintaining the position and heading of the vessel within a specified operating envelope under specified maximum environmental conditions during and following any single fault, including complete loss of a compartment due to fire or flood" (ABS, 2013).

DPS-1, DPS-2 and DPS-3 classification notations stated by ABS (2013) are structured to conform to IMO (2017). Therefore, the classifications DPS-1, DPS-2 and DPS-3 relate to IMO's equipment classifications 1, 2 and 3 (ABS, 2013).

3 STATISTIC ANALYSIS

For the purpose of hazard identification, incident data has been gathered from two sources: the World Offshore Accident Database (WOAD) and the IMCA. The data from IMCA has been utilised for the majority of the analysis as it is consistent and covers a substantial time period.

3.1 *Terms used by IMCA*

The incidents and events listed in the various reports have been categorized by IMCA into three areas, where PL stands for Position Loss. These categories are:

1. *DP Incident (PL 1): This is the loss of automatic DP control, loss of position or any other incident which has resulted in or should have resulted in a RED Alert status* (IMCA 2017). In other words, these are incidents of a serious nature.
2. *DP Undesired Event (PL 2): Loss of position, loss of stability, or another event which is unexpected or uncontrolled and has resulted in or should have resulted in a Yellow Alert status* (IMCA 2017). As such, they are incidents of a less serious nature.
3. *DP Downtime: Position keeping problem or loss of redundancy which would not warrant either a 'Red' or 'Yellow' alert, but where loss of confidence in the DP has resulted in a stand-down from operational status for investigation, rectification, trials, etc*. (IMCA 2017). From an operational point of view, a loss of time or downtime can be seen as an undesired event which should be avoided at all cost, so as to save money.

For the purpose of this research, further analysis focuses on the prior two incidents; DP Incidents (PL 1) and Undesired Incidents (PL 2).

There are two common types of position loss, these are as follows:

1. Drive Off: This is characterised by the thrusters going to high unwanted thrust usually because the DP control system believes the position is wrong (Jenman 1998).
2. Drift Off: This is caused by the lack of sufficient power or thrust. For example, a total blackout on a ship on high seas, combined with strong currents. This would lead to a drift off (Jenman 1998).

There was a debate as to how one would differentiate a "drift off" and a "drive off" of vessels that were quickly recovered and returned to their original position as opposed to vessels which travelled far from their original position and could not be easily recovered (Jenman 1998). These arguments led to the addition of the third type of position loss, Large Excursion.

3. Large Excursion: This is an excursion that takes the DP vessel beyond its normal excursion characterised by its footprint. The footprint is the outline of the vessels movement in a particular sea state (Jenman 1998).

3.2 *IMCA reporting*

Vessels from all over the world, report DP related incidents to IMCA, who ensure that the vessel reporting is kept anonymous, thereby ensuring safety and keeping company integrity. Also, there are a range of export regulations and restrictions which affect business and trade with many countries of the world. These include restrictions on dual-goods and technology as well as the various sanctions regimes (UN, US and EU) targeting individual states including, regulation (EU) No. 833/2014 which is directed at Russia.

The data used in this research is available on the IMCA website, to IMCA members only. However, there are exceptions so that the information can be released to other interested parties who are not IMCA members. IMCA requires both its members and non-members to confirm that they will not use any IMCA document in breach of the Restrictions.

The data provided spanned over 17 years (2000–2016). From the reporting styles over the years, some improvements in reporting are visible, thereby creating differences in the data provided.

From Table 1, a total of 1,163 incidents were analysed and documented by IMCA between the 17-year periods. From Table 1, it is possible that the incident reporting by vessel owners is not consistent. It declined from 2000 to 2004 then picked back up, to peak in 2008 and fall again in 2010. Then there was a steady increase throughout the remaining years.

More emphases should be placed on incident reporting, including potential near misses.

Table 1. Incident data analysed from 2000 to 2016.

Incidents reported			
2000	110	2008	102
2001	98	2009	75
2002	64	2010	56
2003	51	2011	54
2004	34	2012	64
2005	36	2013	64
2006	59	2014	71
2007	67	2015	80
2008	102	2016	78

From the 1,163 incidents analysed, 633 of those incidents indicated the month in which they occurred. This was from the year 2007 to 2015. Based on these IMCA reports, Table 2 was developed.

It can be seen from Table 2 that in the spring and summer of the 9 years, recorded the highest amount of incidents reported. Taking a closer look at the individual years and months, it is evident that even though spring has the highest number of incidents reported over the 9 years, there were more years when summer had more incidents reported than spring. From this, one can state that the majority of the incidents reported occur during summer and spring time.

Looking at the seasons of the year and looking at the months where natural disasters are prominent, this falls within the same seasons as the one stated above. According to AccuWeather Inc., an American media company that provides commercial weather forecasting services worldwide, hurricanes occurs at different times of the year for different regions in the world. For the Atlantic Ocean, hurricane season runs from June to November. In the Eastern Pacific Ocean, it occurs in the months of May through to November. The hurricane season for the Western Pacific Ocean runs from July to November, while in the South Pacific Ocean, it runs from October through to May, reaching a peak in late February or early March. The Indian Pacific Ocean's hurricane season runs from April to December in the Northern Indian Ocean and October to May in the Southern region (Mummey 2010).

Research conducted by the University of Manchester, showed that tornado season in the UK occurs within the months of May to October (Mulder & Schultz 2015). Mother Nature Network on tornado season in America stated that it occurs between March and Early June in the Southern regions. They also state that in the Gulf Coast, tornadoes occur during the spring, while peaking in June and July in the Northern regions. Some states are also mentioned to experience a later tornado season from October to December (Sarah 2011).

From the analyses above, it is evident that most of the adverse weather conditions occur during the spring to autumn seasons. It is safe to attribute this to the reason why there is a higher frequency of incidents reported within the spring to autumn seasons.

Furthermore, the IMO does not state at which time of the day these incidents have occurred. Similarly, it should be noted that these incidents are provided from all around the globe, not one specific region.

For legal, trade and restriction reasons, the number of incidents occurring in different areas of the world has been withheld.

3.3 *Incidents according to vessel type*

Several DP vessels with different functions report incidents to IMCA. Inconsistency in reporting and change of reporting style can be seen in this area. The years stated in this report are the years that were clearly documented by IMCA. The following types of vessels that reported incidents over the aforementioned years are as follows:

1. Remote Operated Vehicles (ROVs)
2. DP Diving Support Vessels (DSV)
3. Drilling Vessels
4. Pipe/Cable Lay Vessels
5. Offshore Loading Vessels
6. Standby Vessels
7. Well Operations Vessels
8. Seismic Vessels
9. Multi Service Vessels (MSV)
10. Shuttle Tanker

Table 2. Incident data analysed from 2000–2016.

	2007	2008	2009	2010	2011	2012	2013	2014	2015	Total	Season	
Dec	10	10	7	3	11	1	2	3	6	53	136	Winter
Jan	5	6	4	1	5	4	6	5	4	40		
Feb	6	2	5	10	2	4	2	5	7	43		
Mar	5	7	7	4	5	10	8	7	10	63	179	Spring
Apr	6	7	12	4	2	6	3	8	10	58		
May	9	14	3	3	5	2	8	8	6	58		
Jun	4	11	8	6	5	8	11	5	7	65	172	Summer
Jul	5	7	7	5	4	7	6	5	7	53		
Aug	6	11	9	5	3	3	5	8	4	54		
Sep	6	9	3	6	4	4	6	5	5	48	145	Autumn
Oct	4	7	2	4	2	10	3	7	5	44		
Nov	1	11	7	5	6	5	4	5	9	53		
Total	67	102	74	56	54	64	64	71	80	632		

11. Flotels
12. Construction Vessels
13. FPSO
14. Rock Dumping Vessels
15. Crane Vessels
16. Supply Vessels

Since there are different types of vessels of different functions and not all these vessels reported incidents every year, Table 3 shows a list of the vessels that reported incidents consistently through the stated time period.

Table 3 shows that DSVs have reported the most incidents in the 10 years stated above. From 2008 to 2014, the reporting style changed and the various vessels that had experienced incidents were not reported.

Table 3. Types of vessels reported 2000–2016.

	DP DSVs	Drilling	Pipe/Cable Lay
2000	21	22	4
2001	35	5	15
2002	11	5	12
2003	15	10	4
2004	4	13	2
2005	9	15	2
2006	13	9	10
2007	16	15	11
2015	10	20	9
2016	18	2	19
Total	152	116	88

3.4 *Incident main causes*

In the IMCA reports, they go further by indicating the primary cause of each incident, making the statistical analysis very easy for the user. Table 4 shows the incident causes over the past few years

As shown in Table 4, the amount of incidents caused by the various DP elements are visible. On further study of Table 4, the elements that cause the most incidents per year can be identified. These elements have been highlighted in bold in Table 4. Judging from this, it is evident that the major main causes of the DP incidents that have been reported are caused by either Reference element or the Thruster Element. It should be noted that the reference element mentioned in Table 4 includes the sensors while the thruster element mentioned includes the propulsion system.

It has been found that in the years prior to 2010, reference system failures were the main cause of incidents. However, from 2011 to 2016, it can been seen that thruster failures have consistently been identified as the main cause of DP incidents. In 2000, there was a drastic drop in incidents related to reference systems. The number of reference related incidents then peaks again in 2008 falls drastically again up to 2016 and the reasons for this has not been identified here. It may be assumed that multiple DP regulations were adopted and enforced across the 17-year period which led to the decrease in reference related incidents.

Looking at the whole 17-year period, Table 4 sums up the total incidents caused by the different

Table 4. Incident main causes from 2000–2016.

Main causes	Computer	Environment	Power gen.	Operator	Reference	Thruster	Electrical	Other	Total
2000	18	5	12	18	**34**	17	3	3	110
2001	**23**	18	8	8	14	14	10	3	98
2002	11	3	13	8	**14**	12	3	0	64
2003	4	7	8	**14**	6	11	0	1	51
2004	1	2	8	6	**12**	4	1	0	34
2005	**8**	3	4	5	5	7	4	0	36
2006	7	4	12	**13**	11	4	6	2	59
2007	**18**	4	11	7	13	8	5	1	67
2008	22	3	9	5	**27**	21	10	5	102
2009	8	2	13	10	**18**	12	10	2	75
2010	6	4	5	3	**21**	2	12	3	56
2011	**14**	5	7	3	9	13	3	0	54
2012	8	2	6	11	11	**20**	4	2	64
2013	6	3	13	7	15	**20**	0	0	64
2014	13	2	9	7	12	**26**	0	2	71
2015	13	11	10	10	11	**24**	1	0	80
2016	15	0	15	16	7	**24**	0	1	78
Incidents	195	78	163	151	**240**	**239**	72	25	1163
Percentage	16.8%	6.7%	14.0%	13.0%	**20.6%**	**20.6%**	6.2%	2.1%	100%

systems and elements, showing their various percentages as they relate to each other.

3.5 *Thruster statistics analysis*

From the analysis of Table 4, it is evident that the thruster system in recent times (2012–2016) has been the major cause of incidents. Looking further into this, it is possible to define initiating causes of undesired events relating to thruster failures. For this, a more specific analysis of incidents caused by thruster failures, that occurred within the years of 2012–2016, was conducted. It can be seen that for the period in question, a total of 357 incidents occurred, of which, 114 incidents were thruster-caused (31.9%), which is extremely high compared to other elements. Of the 114 incidents, 105 were categorized into the different undesired events; "drift off", "drive off" and "time loss" events. Table 5 shows the undesired incidents that occurred and their initiating events.

From Table 5, it is clear that the majority of thruster related incidents begin with a fault in the DP control system, followed by the electrical system and the mechanical system. It is worth noting here that some minor failures have been grouped together to make up the failures mentioned above. For control system, there are failures such as, feedback error, loss of control, wrong DP operator input, *etc.* These are the failures that are rectified when the DP control system is restarted. Software errors can also fall under DP control system errors. Under electrical errors, there are occurrences such as: loose wiring, fibre optic fault, low/high voltage supply, loose fuse, field circuit failure and DC motor failure. In this case, when the loose wire has been fixed or the electrical problem fixed, operation continues.

In the case of mechanical initiating failures, there are faults such as, low oil pressure, hydraulic pump failure, faulty valve, engine failure, oil pipe leakage, cooling motor failure, brake failure, *etc*. In any of these mechanical event scenarios, a redundant system will have to be used or the system will be taken off DP control and returned to port for fixing. For the category of human error, there are a number human failures, such as: wrong procedures, lack of maintenance, inexperience and late response. Some human errors fall under the DP control system, because there has to be an interface between the control system and the human. Here, depending on the fault, the vessel is regained as soon as possible to avoid a worse undesired event. For the reference system, there are errors such as; wind sensor errors, tachometer errors, DGPS errors, *etc*. These errors coupled with some hidden errors can cause the thruster to fail at an odd angle (pitch). Finally, there are failures relating to the power generators. These failures are explanatory on their own and can cause more than the thrusters to fail. They also include the failure of the Uninterrupted Power Supply system (UPS). When these fail, they can cause an immediate failure of the thruster, which can be fixed by a redundant power generation system.

Table 5. Undesired incidents identified from thruster failures and their initiating causes from 2012 to 2016.

Undesired events	Drive off	Drift off	Time loss	Total
Reference	0	0	2	2
Electrical	5	3	21	29
Mechanical	2	2	21	25
Generator	0	0	1	1
Control system	7	6	22	35
Human error	3	2	7	12
Total	17	13	74	104

4 CONCLUSION

From the statistical analysis presented in this study, it was possible to determine the number of incidents that occurred in the different seasons of the year, over a 9-year period. From the results obtained, it was observed that there was an increase in the incidents in the spring and summer seasons. From research it was discovered that this could be as a result of the harsh natural weather conditions that occur during those seasons of the year, such as, tornadoes, hurricanes, *etc*. The tides that occur during the spring, resulting in high tide were also mentioned as a factor for increased incidents during these seasons.

The vessel types with the most incidents were also analysed. It was found that of all the vessels recorded, the three types; DSV, Pipelay and Drilling Vessels had the most incidents recorded. The DSV had the most recorded incidents, with 152 consistently recorded (see Table 3). The reason for this was not fully identified, however it can be argued that pipelay vessels are not used as frequently the DSVs or Drilling vessels. Hence, this has resulted in large difference in the number of reported incidents relating to these vessel types.

Following this, the main causes were analysed and separated into several categories. It was found that the two categories with the highest number of incidents in the 17 year period (2000–2016) were the reference and thruster systems. On further investigation of the data, it was found that within

the years 2000 to 2002 there was a drastic drop in reference caused incidents. Although not found, it is suspected that a regulation was brought into force during that time that either checked incident occurrence or made the operators not to be able to report incidents.

Failure of thruster systems were further analysed because in previous years (2012–2016), they were the major cause of DP incidents. The analysis led to the discovery of the initiating events of the thruster failures and their final consequences (Drift-Off, Drive-Off and Time Loss). This helped to identify the hazards and establish the basic events that would cause the release of such hazards and provided a base to conduct a risk assessment.

Although there were changes in the style of reporting, which led to the inconsistency of some information, the data sourced from IMCA was sufficient in identifying the hazards relating to DP operations.

Finally, from the hazard analysis, reference failures and thruster failures were identified as the major causes of DP incidents, with control failure being the most frequent initiating event leading to thruster failures.

ACKNOWLEDGEMENTS

This work is partially supported by an EU Marie Curie RISE project RESET (reference no. 730888). Similarly, funding has been supplied by the Royal Academy of Engineering under reference number VP1415/1/22.

DISCLAIMER

This paper is the opinion of the authors and does not necessarily represent the belief and policy of their employers.

REFERENCES

ABS (American Bureau of Shipping), 2013. Guide for Dynamic Positioning Systems, November 2013 (updated July 2014), ABS, Houston, Texas.

Bell, R., 2010. Assessment, certification and other assurance measures. *Engineering Safety Consultants Limited,* IEC61508.

DNV, 2012. Dynamic Positioning Systems. *Rules for Classification of Ships*. Available at: www.dnv.com.

DPC-MTS, 1996. Dynamic Positioning Timeline. Available at: http://dynamic-positioning.com/timeline/ [Accessed June 5, 2017].

Holvik, J., 1998. Basics of Dynamic Positioning. *MTS* Dynamic *Positioning Conference*. Available at: http://dynamic-positioning.com/proceedings/dp1998/BHolvik.PDF.

IMCA, 2017. International Marine Contractors Association—Statistics Archives. Available at: https://www.imca-int.com/briefingseries/statistics/ [Accessed July 4, 2017].

IMO, 2017. International maritime organization—Guidelines for Vessels and Units with Dynamic Positioning (DP) Systems. *MSC.1/Circ.1580*, (June).

Jenman, C., 1998. Quantification of the Frequency of an Unsuccessful Disconnection because of a DP Problem. *MTS Dynamic Positioning Conference*. Available at: http://dynamic-positioning.com/proceedings/dp1998/RJenman.PDF.

Mulder, K.J. & Schultz, D.M., 2015. Climatology, Storm Morphologies, and Environments of Tornadoes in the British Isles: 1980–2012. *Monthly Weather Review*, 143(6), pp. 2224–2240. Available at: http://journals.ametsoc.org/doi/10.1175/MWR-D-14-00299.1 [Accessed August 22, 2017].

Mummey, 2010. When and Where Do Hurricanes Occur? *AccuWeather*. Available at: https://www.accuweather.com/en/weather-blogs/hurricanefacts/when-and-where-do-hurricanes-o/31028 [Accessed August 22, 2017].

Sarah, B., 2011. When is tornado season? | MNN – Mother Nature Network. *Mother Nature Network*. Available at: https://www.mnn.com/family/protection-safety/stories/when-is-tornado-season [Accessed August 22, 2017].

Sean, 2009. A Brief History of Dynamic Positioning—Captain. Available at: http://gcaptain.com/history/ [Accessed May 25, 2017].

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A heterogeneous ensemble approach for the prediction of the remaining useful life of packaging industry machinery

F. Cannarile, P. Baraldi & M. Compare
Energy Department, Politecnico di Milano, Milano, Italy
Aramis Srl, Milano, Italy

D. Borghi & L. Capelli
Tetra Pak Packaging Solutions S.p.A., Modena, Italy

E. Zio
Energy Department, Politecnico di Milano, Milano, Italy
Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy-Electricité de France, Ecole Centrale Paris and Supelec, France

ABSTRACT: We present a method based on heterogeneous ensemble learning for the prediction of the Remaining Useful Life (RUL) of cutting tools (knives) used in the packaging industry. Ensemble diversity is achieved by training multiple prognostic models using different learning algorithms. The combination of the outcomes of the models in the ensemble is based on a weighted averaging strategy, which assigns weights proportional to the individual model performances on patterns of a validation set. The proposed heterogeneous ensemble has been applied to real condition monitoring knife data. It has provided more accurate RUL predictions compared to those of each individual base model.

1 INTRODUCTION

As the digital, physical and human worlds continue to integrate, the 4th industrial revolution, the internet of things and big data, the industrial internet, are changing the way we design, manufacture, deliver products and services. In this fast-pace changing environment, the attributes related to the reliability of components and systems continue to play a fundamental role for industry. On the other hand, the advancements in knowledge, methods and techniques, the increase in information sharing and data availability, offer new opportunities of analysis and assessment for reliability engineering. Based on this increased knowledge, information and data available, we can improve our reliability prediction capability. Particularly, the increased availability of data coming from monitoring the relevant components and systems parameters and the grown ability of treating these data by intelligent algorithms capable of mining out information relevant to the assessment and prediction of their state, has open wide the doors for Prognostics and Health Management (PHM) and predictive maintenance in many industrial sectors, for improved operation and maintenance (Zio, 2016). Approaches for RUL estimation can be generally categorized into model-based and data-driven (Baraldi et al., 2015a). Model-based approaches use physics-based models to describe the degradation behavior of the equipment (Baraldi et al., 2015a). On the other side, data-driven methods are of interest when an explicit model of the degradation process is not available, as they rely on the availability of field data collected during the operation of one or more similar components. Among data-driven methods one can distinguish between (i) degradation-based approaches, modeling the future equipment degradation evolution and (ii) direct RUL prediction approaches, directly predicting the RUL.

Degradation-based approaches are based on statistical models that learn the equipment degradation time evolution from time series of the observed degradation (Baraldi et al., 2017). The predicted degradation state is, then, compared with a failure criterion, such as the value of degradation beyond which the equipment fails performing its function (failure threshold). Examples of modeling techniques used in degradation-based approaches are Auto-Regressive models (Gorjian et al., 2009), Relevance Vector Machines (Di Maio et al., 2012) and Semi-Markov Models (Cannarile et al., 2017a) (Cannarile et al., 2018).

Direct RUL predictions approaches, instead, typically resort to machine learning techniques

that directly map the relation between the observable parameters and the equipment RUL, without the need of predicting the equipment degradation state evolution towards a failure threshold (Schwabacher et al., 2007). Techniques used in direct RUL prediction approaches are, for example, Artifical Neural Networks (Wang & Vachtsenavos, 2001), Extreme Learning Machines (ELM) (Yang et al., 2017), Gaussian Processes (GP) (Baraldi et al., 2015b), etc.

When few run-to-failure degradation trajectories are available, direct RUL approaches may overfit, i.e., these algorithms customize themselves too much to learn the relationship between the observable parameters and the corresponding RUL in the training set. Therefore, these methods tend to lose their generalization power, which leads to poor performance on new data. To overcome this, ensemble approaches, based on the aggregation of multiple model outcomes, have been introduced (Baraldi et al., 2013a). The basic idea is that the diverse models in the ensemble complement each other by leveraging their strengths and overcoming their drawbacks.

Thus, the combination of the outcomes of the individual models in the ensemble improves the accuracy of the predictions compared to the performance of a single model (Brown et al., 2005) (Baraldi et al., 2013a). Different methods, such as ANN (Baraldi et al., 2013b), Support Vector Machine (SVM) (Liu et al., 2006) and kernel learning (Liu et al., 2015), have been used with success to build the individual models. For example, an ensemble of feedforward Artificial Neural Networks (ANN) has been embedded into a Particle Filter (PF) for the prediction of crack length evolution (Baraldi et al., 2013b) and an ensemble of data-driven regression models has been exploited for the RUL prediction of lithium-ion batteries (Xing et al., 2013). In (Rigamonti et al., 2017) a local ensemble of Echo State Networks (ESN) has been proposed to improve the RUL prediction accuracy of turbofan engines.

The objective of this work is to predict the RUL of knives installed on Tetra Pak® A3/Flex filling machines used to cut package material. The prognostic task is complicated by the fact that few run-to-failure degradation trajectories are available, and a failure threshold is not available. To cope with these issues, this work proposes an ensemble formed by multiple data-driven direct RUL prediction models, capable of aggregating the RUL predictions for good performance throughout the entire degradation trajectory of a knife. Ensemble diversity is achieved by *heterogeneous* ensemble generation, i.e., by training the models using different prognostics algorithms. Aggregation is obtained by averaging the output of the individual base models with weights proportional to the inverse of their Empirical Generalization Error (EGE) on retrieved patterns in a validation set. The application of the proposed heterogeneous ensemble method to real condition monitoring knife data has shown to provide more accurate RUL prediction compared to that of each individual base learner in the ensemble.

The paper is organized as follows: in Section 2, the objectives of this work and the assumptions are discussed; in Section 3, ensemble learning main concepts for data-driven direct RUL prediction are illustrated; in Section 4, performance metrics to compare different prognostic models are discussed. The application of the methodology to Tetra Pak® A3/Flex filling data is described in Section 5, whereas Section 6 draws the work conclusions.

2 ASSUMPTIONS AND OBJECTIVES

We assume to have available run-to-failure degradation trajectories of N pieces of equipment similar to the one currently monitored (test equipment). Let $\boldsymbol{x}_i(\tau_i) \in \mathbb{R}^m, i = 1,..,N; \tau = 1,...,n_i$ be the vector of m features extracted from signal measurements performed at time τ_i on the ith i^{th} equipment, with ni indicating the total number of data acquisitions performed on the ith equipment before its failure. The ground truth RUL of the ith piece pf equipment at time τ_i will be referred to as $y_i(\tau_i), i = 1,\ldots,N; \tau_i = 1,\ldots,n_i$. We consider a case in which the failure thresholds for the extracted features are not known. In this setting, fault prognostics is framed as a regression problem: given the historical dataset U formed by N realizations (degradation trajectories) $\{\boldsymbol{x}_i(\tau_i), y_i(\tau_i), \tau_i = 1,\ldots,n_i\}, i = 1,\ldots,N$, of a stochastic process $(\boldsymbol{X}(\tau), Y(\tau)) \in \mathbb{R}^m \, x(0,+\infty)$, our task is to find a function $\tilde{f}: \mathbb{R}^m \to (0,+\infty)$ such that it associates to a test pattern $x_{test}(\tau_{test}) \in \mathbb{R}^m$, the corresponding output $y_{test}(\tau_{test})$. In what follows, we refer to $\tilde{f}$ as base model or base learner (Zhou, 2012).

3 ENSEMBLE LEARNING FOR FAULT PROGNOSTICS

In contrast to ordinary learning approaches which try to construct one base learner from training data, ensemble methods try to construct a set of learners $\widetilde{f_1},\ldots,\widetilde{f_H}$ and combine them to obtain an ensemble learner $\widetilde{f_{ens}}$. In this work, we consider combination of base learners based on weighted averaging (Zhou, 2012), i.e., the combined output $\widetilde{f_{ens}}$ is obtained by averaging the output of the individual learners with different weights α_h,

which implies that the different learners have different importance

$$\widetilde{f_{ens}}\left(\boldsymbol{x}(\tau)\right) = \sum_{h=1}^{H} \alpha_h \widetilde{f}_h\left(\boldsymbol{x}(\tau)\right) \quad (1)$$

where

$$\sum_{h=1}^{H} \alpha_h = 1; \quad \alpha_h \geq 0; \; h = 1,\ldots,H \quad (2)$$

3.1 *Error ambiguity decomposition*

In this Subsection, we motivate the use of ensemble learning to enhance RUL predictions of a test equipment. Referring to the ensemble generalization error as $GE\left(\widetilde{f_{ens}}\right)$, one can show that the following error-ambiguity decomposition holds (for more details, see the Appendix):

$$GE\left(\widetilde{f_{ens}}\right) = \overline{GE}(h) - \overline{ambi}(h) \quad (3)$$

where $\overline{GE}(h) = \sum_{h=1}^{H} \alpha_h GE\left(\widetilde{f}_h\right)$ is the weighted average of the hth individual base learner generalization error $GE\left(\widetilde{f}_h\right)$; and $\overline{ambi}(h) = \sum_{h=1}^{H} \alpha_h ambi\left(\widetilde{f}_h\right)$ is the weighted average of the hth individual base learner ambiguity $ambi\left(\widetilde{f}_h\right)$ defined in Appendix. The quantity $ambi\left(\widetilde{f}_h\right)$ quantifies how much the hth base learner predictions, $\widetilde{f}_h$, differ from the ensemble predictions. On the right-hand of Eq. (3), the first term $\overline{GE}(h)$ represents the individual learner average error, which depends on the generalization ability of individual base learners whereas the second term $\overline{ambi}(h)$ represents the ambiguity, which depends on the ensemble diversity. Since the second term is always positive, and it is subtracted from the first term, it is clear that the error of the ensemble will never be larger than the average error of the individual base learners. Further, Eq. (11) shows that the more accurate and the more diverse the individual learners, the better the ensemble.

3.2 *Ensemble generation*

According to the error-ambiguity decomposition discussed in Subsection 3.1, ensemble diversity, i.e., the difference among the individual base learners is a fundamental issue in ensemble learning. Therefore, since complementarity is more important than pure accuracy (Zhou, 2012), an ensemble formed by only very accurate learners can provide worse performances than one formed by also some relatively weak learners. Two approaches are typically used to generate diverse base learners:

- *Homogeneous ensemble generation*: different base learners are generated using the same prognostic algorithm and diversity is achieved by manipulating data in different ways: subsampling from the training set (e.g., bagging ((Zhou, 2012))) or using different subsets of features.
- *Heterogeneous ensemble generation*: different base models are generated using different prognostic algorithms.

In this work, we have resorted to heterogeneous ensemble generation since it has been shown able to provide better performance than homogenous ensemble methods in cases of few low-dimensional data (Rathore & Kumal, 2017).

3.3 *Setting the ensemble base model weights α_h*

The data extracted from the available N run-to-failure degradation trajectories of similar components are divided into training, validation and test subsets, formed by P_{train}, P_{valid} and P_{test} instances, respectively. The training subset is used to build the H individual base models, the validation subset to assign them weights to be used for the aggregation of the individual model outcomes (Eq. (1)) and the test subset to verify the final ensemble performance. The weight α_h associate to the hth base learner is calculated based on its performance in predicting the RUL of the validation set patterns. Performance is measured resorting to the Empirical Generalization Error (EGE), which for the hth base learner is defined as the mean squared error on validation set patterns:

$$\widehat{GE}\left(\widetilde{f}_h\right) = \frac{1}{P_{valid}} \sum_{p=1}^{P_{valid}} \frac{1}{n_p} \sum_{\tau_p=1}^{n_p} \left(y_p(\tau_p) - \widetilde{f}_h\left(\boldsymbol{x}_p(\tau_p)\right)\right)^2 \quad (4)$$

In this work, we have considered weights proportional to the inverse of the EGE, i.e.,

$$\alpha_h = \frac{\dfrac{1}{\widehat{GE}\left(\widetilde{f}_h\right)}}{\displaystyle\sum_{l=1}^{H} \frac{1}{\widehat{GE}\left(\widetilde{f}_l\right)}} \quad h = 1,\ldots,H \quad (5)$$

4 PROGNOSTIC PERFORMANCE METRICS

In addition to EGE, we have considered other performance metrics, which are typically considered (Rigamonti et al., 2017) for quantitatively assessing and comparing the point prediction performance of different prognostic algorithms (Saxena et al., 2009). A brief description of the implemented metrics is

given hereafter considering a generic test trajectory $\left(\boldsymbol{x}(\tau), y(\tau)\right), \tau = 1, \quad, n$ and a general base learner $\tilde{f}$.

- *Relative Accuracy (RA)*:

$$R\left(\tilde{f}\right) = \sum_{\tau}^{n} \exp\left(-\frac{\left|\tilde{f}(x(\tau) - y(\tau)\right|}{y(\tau)}\right) \tag{6}$$

Notice that $R\left(\tilde{f}\right)$ is in the range $[0,1]$ and the larger the relative accuracy the more accurate is the model.

- *Precision*:

$$P = \sqrt{\frac{\sum_{\tau=1}^{n}\left(e(\tau) - \bar{e}\right)^2}{n}} \tag{7}$$

$$e(\tau) = \tilde{f}\left(x(\tau)\right) - y(\tau) \tag{8}$$

$$\bar{e} = \frac{1}{n}\sum_{\tau=1}^{n} e(\tau) \tag{9}$$

This measure quantifies the dispersion (stability) of the prediction error around its mean. Closer to zero is the precision, more stable is the model.

5 CASE STUDY

This Section presents the results of the application of the proposed method to Tetra Pak® A3/Flex filling knife condition monitoring data.

We have available run-to failure-degradation trajectories from N = 10 different knives. For each knife, we have available m = 2 health indicators which have been extracted using the procedure presented in (Cannarile et al., 2017b).

In this work, a heterogeneous ensemble generation has been developed considering H = 4 prognostic algorithms:

- Gaussian Process Regression with Squared Exponential (GPRSE) covariance function;
- GRP with Matern 3/2 (GRPM) covariance function;
- Support Vector Regression with Gaussian Kernel (SVRGK);
- SVR with Quadratic Polynomial Kernel (SVRQPK).

These algorithms have been selected, since they have proved to be effective also when few training data with no clear patterns of regularity are available for training (Domingos, 2012). To properly compare the performance of the ensemble model with that of each base model, we have resorted to a twice nested Leave-One-Out-Cross-Validation (LOOCV) approach. The outer loop is to assess the performance of the ensemble and the single base learners, whereas the inner loop allows setting the weights $\alpha_h, h = 1,..,4$. In practice, the weights associates to the base learners are computed on each outer-validation set (using the inner LOOCV loop) and the final performance is measured on the corresponding outer-testing set (see Figure 1).

Table 1 compares the performances of the developed ensemble model with that of the GRPM model, which has resulted to be the best performing individual model.

Notice that the ensemble model performs better than GRPM in all the considered metrics. In particular, the average EGE is 11.14% lower (more satisfactory) than that of GRPM, the relative accuracy of the ensemble is 3.34% larger (more satisfactory) than that of GRPM, whereas, the two methods are comparable from the point of view of the precision. Finally, Figs. 2 and 3 show the RUL predicted by the ensemble and GRPM for two representative test trajectories.

The most satisfactory ensemble predictions tend to be at the begininning of the life of the test knife. This is reflected by the great improvement of the EGE metric, which is more sensible to errors at the beginning of the run to failure trajectory than the relative accuracy.

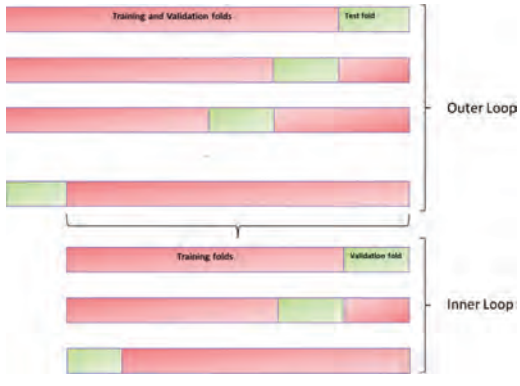


Figure 1. Twice nested LOOCV.

Table 1. Comparison between the ensemble and the GRPM performances.

	Ensemble	GRPM
Empirical Generalization Error (EGE) (best value 0)	3.2991	3.7127
Relative Accuracy (RA) (best value 1)	0.8149	0.7804
Precision (best value 0)	0.0569	0.0633

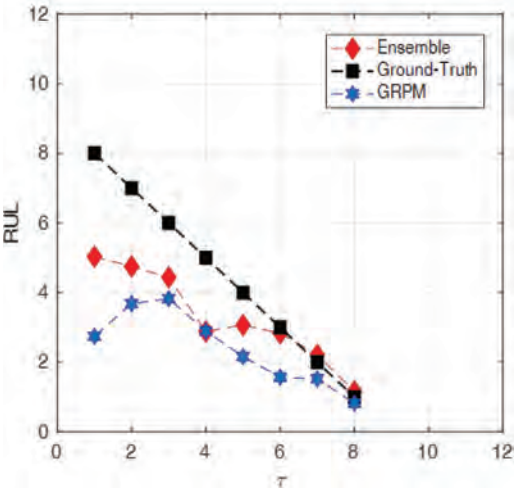


Figure 2. Predicted RUL by the ensemble (diamonds) and GRPM (exagon) for a test trajectory.

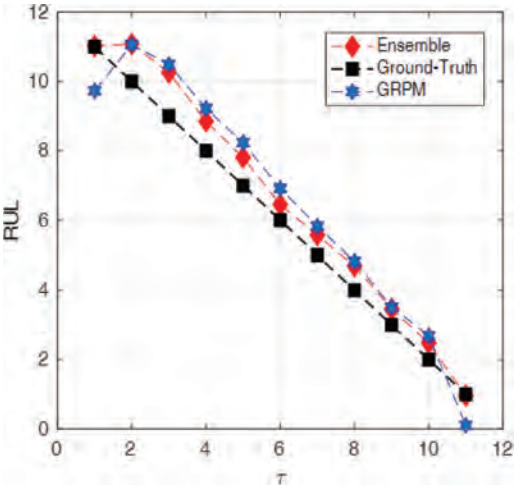


Figure 3. Predicted RUL by the ensemble (diamonds) and GRPM (exagon) for a test trajectory.

6 CONCLUSIONS

In this work, we have developed a heterogeneous ensemble model for enhancing the accuracy of the RUL prediction of knives used in the packaging industry. Thanks to the diversity of the base learner algorithms, the proposed approach has been shown capable of reducing the generalization error and providing more accurate RUL predictions compared to that of each individual base learner in the ensemble.

REFERENCES

Baraldi P., Cadini F., Mangili F., Zio E., 2013a. Model-based and data-driven prognostics under different available information. *Probabilistic Engineering Mechanics*, 32, pp. 66–79.

Baraldi, P., Compare, M., Sauco, S., & Zio, E., 2013b. Ensemble neural network-based particle filtering for prognostics. *Mechanical Systems and Signal Processing*, 41(1), 288–300.

Baraldi, P., Mangili, F., Zio, E., 2015a. A belief function theory based approach to combining different representation of uncertainty in prognostics. *Information Sciences*, 303, pp. 134–149.

Baraldi, P., Mangili, F., Zio, E., 2015B. A prognostics approach to nuclear component degradation modeling based on Gaussian Process Regression. *Progress in Nuclear Energy*, 78, pp. 141–154.

Baraldi, P., Di Maio, F., Al-Dahidi, S., Zio, E., Mangili, F., 2017. Prediction of industrial equipment Remaining Useful Life by fuzzy similarity and belief function theory. *Expert Systems with Applications*, 83, pp. 226–241.

Brown, G., Wyatt, J., Harris, R., & Yao, X., 2005. Diversity creation methods: a survey and categorisation. *Information Fusion*, 6(1), 5–20.

Cannarile, F., Compare, M., Rossi, E., Zio, E., 2017a. A fuzzy expectation maximization based method for estimating the parameters of a multi-state degradation model from imprecise maintenance outcomes. Annals of Nuclear Energy, 110, pp. 739–752.

Cannarile, F., Baraldi., P., Compare, M., Borghi, D., Capelli, L., Cocconcelli, M., Lahrac, A., Zio, E., 2017. An unsupervised clustering method for assessing the degradation state of cutting tools in the packaging industry. *Safety and Reliability: Theory and Application—Proceedings of the European Safety and Reliability Conference*, ESREL 2017.

Cannarile, F., Compare, M., Baraldi, P., Di Maio, F., Zio, E., 2018. Homogeneous continuous-time, finite-state, hidden semi-Markov modelling for enhancing Empirical Classification System diagnostics of industrial components. *Probabilistic Engineering Mechanics,* under review.

Di Maio, F., Tsui, K.L., Zio, E., 2012. Combining Relevance Vector Machines and exponential regression for bearing residual life estimation. *Mechanical Systems and Signal Processing*, 31, pp. 405–427.

Domingos, P., 2012. A few useful things to know about machine learning, *Communications of the ACM*, 55 (10), pp. 78–87.

Gorjian, N., Ma, L., Mittinty, M., Yarlagadda, P., Sun, Y., 2009. A review on degradation models in reliability analysis. Engineering Asset Lifecycle Management – *Proceedings of the 4th World Congress on Engineering Asset Management*, WCEAM 2009, pp. 369–384.

Liu, Y., An, A., Huang, X., 2006. Boosting Prediction Accuracy on Imbalanced Datasets with SVM Ensembles. In PAKDD 6, pp. 107–118.

Liu, Y., Zhang, Z., & Chen, J., 2015. Ensemble local kernel learning for online prediction of distributed product outputs in chemical processes. *Chemical Engineering Science*, 137, 140–151.

Rathore, S.S., Kumar, S., 2017. Linear and non-linear heterogeneous ensemble methods to predict the

number of faults in software systems. *Knowledge-Based Systems*, 119, pp. 232–256.
Rigamonti, M., Baraldi, P., Zio, E., Roychoudhury, I., Goebel, K., Poll, S., 2017. Ensemble of optimized echo state networks for remaining useful life prediction. *Neurocomputing*, Article in Press.
Saxena, A., Celaya, J., Saha, B., Saha, S., Goebel, K., 2009. Evaluating algorithm performance metrics tailored for prognostics. *In Aerospace conference, 2009 IEEE*, pp. 1–13.
Wang, P., Vachtsevanos, G., 2001. Fault prognostics using dynamic wavelet neural networks. Artificial Intelligence for Engineering Design, *Analysis and Manufacturing: AIEDAM*, 15 (4), pp. 349–365.
Xing, Y., Ma, E.W., Tsui, K.L., Pecht, M., 2013. An ensemble model for predicting the remaining useful performance of lithium-ion batteries. *Microelectronics Reliability*, 53(6), 811–820.
Yang, Z., Baraldi, P., Zio, E., 2017. A comparison between extreme learning machine and artificial neural network for remaining useful life prediction. *Proceedings of 2016 Prognostics and System Health Management Conference, PHM*-Chengdu 2016.
Zhou, Z., 2012. Ensemble Methods: Foundations and Algorithms, Chapman Hall/CRC.
Zio, E., 2016. Some challenges and opportunities in reliability engineering. *IEEE Transactions on Reliability*, 65 (4), pp. 1769–1782.

APPENDIX

Given an instance $\boldsymbol{x} = \boldsymbol{x}(\tau)$, the ambiguity of the individual base learner $\widetilde{f}_h$ is defined as

$$ambi\left(\widetilde{f}_h|\boldsymbol{x}\right) = \left(\widetilde{f}_h(\boldsymbol{x}) - \widetilde{f_{ens}}(\boldsymbol{x})\right)^2 \quad h = 1,\ldots,H \quad (10)$$

and the ambiguity of the ensemble is

$$\overline{ambi}\left(\widetilde{f_{ens}}|\boldsymbol{x}\right) = \sum_{h=1}^{H} \alpha_h ambi\left(\widetilde{f}_h|\boldsymbol{x}\right) = \sum_{h=1}^{H} \alpha_h \left(\widetilde{f}_h(\boldsymbol{x}) - \widetilde{f_{ens}}(\boldsymbol{x})\right)^2 \quad (11)$$

The ambiguity term measures the disagreement among the individual base learners on instance $\boldsymbol{x}$. If we use the Squared Error (SE) to measure the performance, then, the error of the individual base learner $\widetilde{f}_h$ and the ensemble $\widetilde{f_{ens}}$ are, respectively,

$$SE\left(\widetilde{f}_h|\boldsymbol{x}\right) = \left(\widetilde{f}_h(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2 \quad h = 1,\ldots,H \quad (12)$$

$$SE\left(\widetilde{f_{ens}}|\boldsymbol{x}\right) = \left(\widetilde{f_{ens}}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2 \quad (13)$$

Then, one can show that (Zhou, 2012)

$$\overline{ambi}\left(\widetilde{f_{ens}}|\boldsymbol{x}\right) = \overline{SE}\left(\tilde{h}|\boldsymbol{x}\right) - SE\left(\widetilde{f_{ens}}|\boldsymbol{x}\right) \quad (14)$$

where $\overline{SE}\left(\tilde{h}|\boldsymbol{x}\right) = \sum_{h=1}^{H} \alpha_h SE\left(\widetilde{f}_h|\boldsymbol{x}\right)$ is the weighted average of the individual base learner errors. Since Eq. (14), holds for every instance $\boldsymbol{x}$, after averaging over the input distribution $p(\boldsymbol{x})$ from which the instances are sampled, it still holds that.

$$\sum_{h=1}^{H} \alpha_h \int ambi\left(\widetilde{f}_h|\boldsymbol{x}\right) p(\boldsymbol{x})\, d\boldsymbol{x} = \sum_{h=1}^{H} \alpha_h \int SE\left(\widetilde{f}_h|\boldsymbol{x}\right) p(\boldsymbol{x})\, d\boldsymbol{x} - \int SE\left(\widetilde{f_{ens}}|\boldsymbol{x}\right) p(\boldsymbol{x})\, d\boldsymbol{x} \quad (15)$$

The generalization error and the ambiguity of the individual base learner $\widetilde{f}_h$, can be written as, respectively,

$$GE\left(\widetilde{f}_h\right) = \int SE\left(\widetilde{f}_h|\boldsymbol{x}\right) p(\boldsymbol{x})\, d\boldsymbol{x} \quad h = 1,\ldots,H \quad (16)$$

$$ambi\left(\widetilde{f}_h\right) = \int ambi\left(\widetilde{f}_h|\boldsymbol{x}\right) p(\boldsymbol{x})\, d\boldsymbol{x} \quad h = 1,\ldots,H \quad (17)$$

Similarly, the generalization error of the ensemble reads

$$GE\left(\widetilde{f_{ens}}\right) = \int SE\left(\widetilde{f_{ens}}\,|\,\boldsymbol{x}\right) p(\boldsymbol{x})\, d\boldsymbol{x} \quad h = 1,\ldots,H \quad (18)$$

Based on the notation just introduced and Eq. (14), we obtain the error-ambiguity decomposition (Zhou, 2012):

$$GE\left(\widetilde{f_{ens}}\right) = \overline{GE}(h) - \overline{ambi}(h) \quad (19)$$

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Strength of knowledge assessment for risk informed decision making

Tasneem Bani-Mustafa & Zhiguo Zeng
Chair on System Science and the Energetic Challenge, EDF Foundation, Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay, France

Enrico Zio
Chair on System Science and the Energetic Challenge, EDF Foundation, Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay, France
Energy Department, Politecnico di Milano, Italy

Dominique Vasseur
EDF R&D, PERICLES (Performance et Prévention des Risques Industriels du Parc par la Simulation et les Etudes), EDF Lab Paris Saclay, France

ABSTRACT: Risk Informed Decision Making (RIDM) is based on risk metrics obtained from a Probabilistic Risk Assessment (PRA). For plants exposed to multiple hazards, Multi-Hazards Risk Aggregation (MHRA) is necessary to inform decisions. In practice, this is often done by a simple arithmetic summation over the different risk contributors, without taking into account that the state of knowledge of the risk models of the different hazards can be quite different. In this paper, we provide a hierarchical framework to assess the strength of knowledge that PRA models are based upon. The framework is organized in three attributes characterizing the knowledge which a PRA model is based upon (assumptions, data, phenomenological understanding). These attributes are further broken down into sub-attributes and, finally, "leaf" attributes that can be evaluated. The PRA models of two hazards groups for Nuclear Power Plants (NPPs) are considered and the strength of knowledge behind each model is assessed using the developed framework.

1 INTRODUCTION

In risk assessment, quantities are calculated to describe the magnitude and likelihood of the consequences from accidents that may develop from known hazards [1]. The confidence on the calculated risk indexes depends on the knowledge available to support the risk assessment [3–5]. For example, in the risk assessment of Nuclear Power Plants (NPPs), there is more experience and knowledge on internal events than other hazard groups like external flooding [1]. Evaluating the strength of knowledge of a risk assessment, is, then, important to evaluate how much confidence we can put on the risk outcomes, that are, then, used to inform decision making [2].

Research efforts have been conducted, recently, for linking knowledge, knowledge evaluation and knowledge management to Risk-Informed Decision-Making (RIDM) [4–7]. For example, in the nuclear industry, knowledge management has been identified as a key factor in sustaining nuclear power programs and maintaining their safety and security [3]. However, most of the existing works are qualitative in nature. A semi-quantitative method for evaluating the strength of knowledge has been proposed by Flage and Aven [4], where the strength of knowledge is evaluated in terms of four attributes: (i) phenomenological understanding and availability of trustable predicting models; (ii) reasonability and realism of assumptions; (iii) availability of reliable and relevant data, and information; (iv) agreement/disagreement among peers. The four attributes were assessed in three levels (minor, moderate and significant) and aggregated for strength of knowledge assessment [4]. Although the knowledge attributes proposed are plausible and reasonably complete, their definitions remain ambiguous. In addition, the evaluation of these attributes is somewhat intangible in practice, since it is done by simple scoring based on a plain description of the attributes. To overcome this problem, we expand the work in [5] and introduce a hierarchical tree-based framework for evaluating the state of knowledge.

The rest of the paper is organized as follows. In Sect. 2, we present the developed framework for strength of knowledge assessment. Section 3

applies it on a case study of two hazard groups considered in NPPs risk assessment. Finally, in Sect 4, the paper is concluded with a discussion on potential future developments.

2 ASSESSMENT FRAMEWORK

We consider the strength of knowledge assessment of event tree models which are widely applied in PRA of NPPs. The events probabilities in the event tree model might are typically calculated by fault tree models. The risk index associated to a given consequence (e.g. the probability of core damage) is calculated by summing the values of the risk index from several risk models:

$$R = \sum_{i=1}^{n_O} \sum_{j=1}^{n_{S,i}} R_{i,j} \quad (1)$$

where n_O is the number of operation states (O), $n_{S,i}$ is the number of accident sequences (scenarios, S) that in operation state i can lead to the given consequence. Each $R_{i,j}$ in (1) quantifies the specific risk index under scenario j (e.g., medium flood level) in operation state i (e.g., emergency shutdown).

The risk models used to calculate the risk index $R_{i,j}$ values are characterized by Initial Events (IEs), Basic Events (BEs) and the combinations of the latter into Minimal Cut Sets (MCSs). In practice, it can often be assumed that the MCSs are mutually exclusive; then, $R_{i,j}$ can be calculated by [5]:

$$R_{i,j} = \sum_{k=1}^{n_{MCS,i,j}} \prod_{q=1}^{n_{BE,k}} P_{BE,q} \quad (2)$$

where $n_{MCS,i,j}$ is the number of minimal cut sets in the risk model for operation state i and scenario j, $n_{BE,k}$ is the number of basic events in the kth minimal cut set, and $P_{BE,q}$ is the probability of having the qth basic event. The five elements, S, O, IE, BE and MCS, fully define a PRA model, as shown in Figure 1. In this paper, we refer to these five elements as "atomic elements".

To assess the strength of knowledge of a PRA model, all the five atomic elements need to be considered. In practice, however, PRA models are very complex and contain many scenarios and operation states combined in large and complex fault trees and event trees, that consist of thousands of BEs and MCSs [6]. For such a complex risk assessment model, it is not practical to consider all atomic elements for evaluating the strength of knowledge. To address this problem, in this work, we first develop a reduced-order model for (1), in order to limit the number of atomic elements that need to be analyzed.

A flowchart of the developed knowledge assessment method is given in Figure 2. The first step

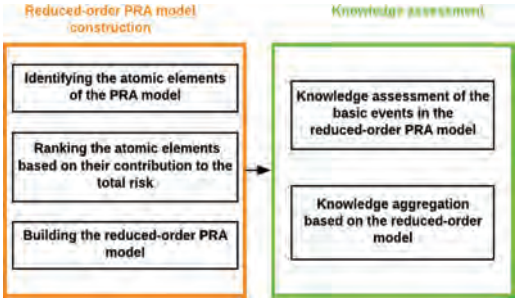


Figure 1. Steps of PRA model knowledge assessment.

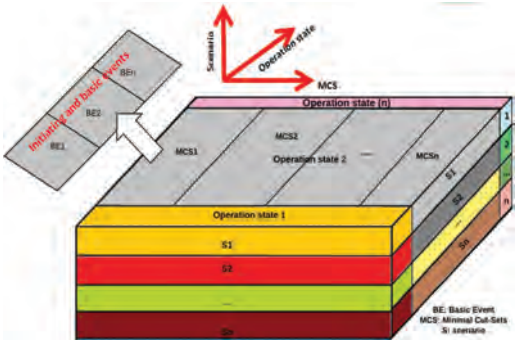


Figure 2. Atomic elements of a PRA model.

involves developing a reduced-order model for the original risk assessment model. A detailed discussion on how to construct the reduced-order model is given in Sect. 2.1. Then, the strength of knowledge supporting each atomic element in the reduced-order model is assessed by an Analytical Hierarchy Process (AHP), as illustrated in Sect. 2.2. Finally, the strength of knowledge of each element is aggregated to evaluate the strength of knowledge of the entire PRA model. A detailed discussion is given in Sect. 2.3.

2.1 *Reduced-order PRA model construction*

It is often observed in PRA models that most of the contribution to the total risk is due to a small number of elements of the problem (known as "*Pareto principle*") [7]. We can, then, reduce the PRA model into a reduced-ordered model, which consists of only the most important "*elements*".

The procedure for constructing the reduced-order model comprises of three steps. Firstly, the number of operation states n_O is reduced to $n_{O,Red}$, as follows:

- Calculate the risk R_{O_i} for each operation state:

$$R_{O_i} = \sum_{j=1}^{n_{S,i}} R_{i,j}, \quad 1 \le i \le n_O \tag{3}$$

where $R_{i,j}$ is calculated by (2).

- Rank R_{O_i} in descending order.
- Find the minimal $n_{O,Red}$, so that

$$\frac{\sum_{i=1}^{n_{O,Red}} R_{O_i}}{R} \ge \alpha \tag{4}$$

where α is the fraction of total risk that can be reproduced by the operation states in the reduced-order model (in the case study in Sect. 3.2.1, we assume that $\alpha = 0.8$).

- Keep only operation states for $i = 1,\cdots,n_{O,Red}$; operation states with $i > n_{O,Red}$ are eliminated.

The second step is to define the reduced number of scenarios $n_{S,Red,i}$ for each operating state i in the reduced-order model, where $i = 1,\cdots,n_{O,Red}$:

- For $i = 1,\cdots,n_{O,Red}$, calculate the risk $R_{i,j}$, $1 \le j \le n_{S,i}$ by (2).
- Rank $R_{i,j}$ in descending order.
- Find the minimal $n_{S,Red,i}$ so that,

$$\frac{\sum_{j=1}^{n_{S,Red,i}} R_{i,j}}{R_{O,i}} \ge \beta \tag{5}$$

where R_{O_j} is calculated by (3) and β is the fraction of total risk that can be reproduced by the scenarios in the reduced-order model (in the case study in Sect. 3.2.1, we assume that $\beta = 0.8$).

- Keep only scenarios for $j = 1,\cdots,n_{S,Red,i}$; scenarios with $j > n_{S,Red,i}$ are eliminated.

Finally, the number of minimal cut sets $n_{MCS,i,j}$ is tailored to $n_{MCS,Red,i,j}$, $i = 1,...,n_{O,Red}, j = 1,...,n_{S,Red,i}$:

- Calculate $R_{i,j,k}$ by:

$$R_{i,j,k} = \prod_{q \in MCS_{i,j,k}} P_{BE,q}, \begin{matrix} 1 \le i \le n_{O,Red} \\ 1 \le j \le n_{S,Red,i} \\ 1 \le k \le n_{MCS,i,j} \end{matrix} \tag{6}$$

- Rank $R_{i,j,k}$ in descending order.
- Find the minimal $n_{MCS,Red,i,j}$ so that,

$$\frac{\sum_{k=1}^{n_{MCS,Red,i,j}} R_{i,j,k}}{R_{i,j}} \ge \gamma \tag{7}$$

where $R_{i,j,k}$ is calculated by (6) and γ is the fraction of total risk that can be reproduced by the minimal cut sets in the reduced-order model (in the case study in Sect. 3.2.1, we assume that $\gamma = 0.8$).

- Keep only minimal cut sets for $k = 1,\cdots,n_{MCS,Red,i,j}$; minimal cut sets with $k > n_{MCS,Red,i,j}$ are eliminated.

Assuming that the MCSs are mutually exclusive, the total risk of the reduced-order PRA model can be calculated by:

$$R_{Red} = \sum_{i=1}^{n_{O,Red}} \sum_{j=1}^{n_{S,Red,i}} \sum_{k=1}^{n_{MCS,Red,i,j}} \prod_{q \in MCS_{i,j,k}} P_{BE,q} \tag{8}$$

Note that from (4), (5) and (7), the reduced order risk R_{Rec} can reconstruct $\alpha \cdot \beta \cdot \gamma$ of the total risk R. Only the events that are contained in the reduced-order model (8) are used for assessing the strength of knowledge of the PSA.

2.2 *Knowledge assessment for the risk elements*

Once the reduced-order model is constructed, the strength of knowledge of each atomic element in such model is evaluated. In Section 2.2.1, we present a tree-based hierarchical framework for knowledge assessment. Then, in Section 2.2.2, we show how to proceed with the evaluation using the Analytical Hierarchy Process (AHP) method.

2.2.1 *Knowledge assessment framework*

A tree-based hierarchical framework is here developed for knowledge assessment, as shown in Figure 3. The strength of knowledge, represented by K (Level 1), represents the solidity of knowledge that supports a PRA model. A higher value of strength of knowledge indicates that the PRA model is supported by trustable evidence and reliable knowledge, and, therefore, its results can be taken with confidence.

As in Flage and Aven [4], we evaluate the strength of knowledge in terms of three attributes: assumptions (K_1), data (K_2) and phenomenological understanding (K_3) The attribute K_1 represents the adequacy, solidity and plausibility of the assumptions upon which the model is based; K_2 represents the amount and quality of the available data that are used to estimate the parameters of the model; K_3 represents the knowledge behind the phenomenon described in the model.

For their evaluation, the three attributes are further decomposed into sub-attributes. In particular, assumptions (K_1) is evaluated in terms of quality of assumptions (K_{11}), value ladenness (K_{12}) and impact (K_{13}); data (K_2) is evaluated in terms of the amount of data (K_{21}) and the reliability and consistency of data (K_{22}); phenomenological understanding (K_3) is evaluated in terms of years of experience of the experts involved in the model development (K_{31}), number of experts involved (K_{32}), academic evidence (K_{33}) and industrial evidence (K_{34}). Value ladenness and reliability and consistency of data are further decomposed into "leaf" sub-attributes in level 4 for their evaluation, as shown in Figure 3.

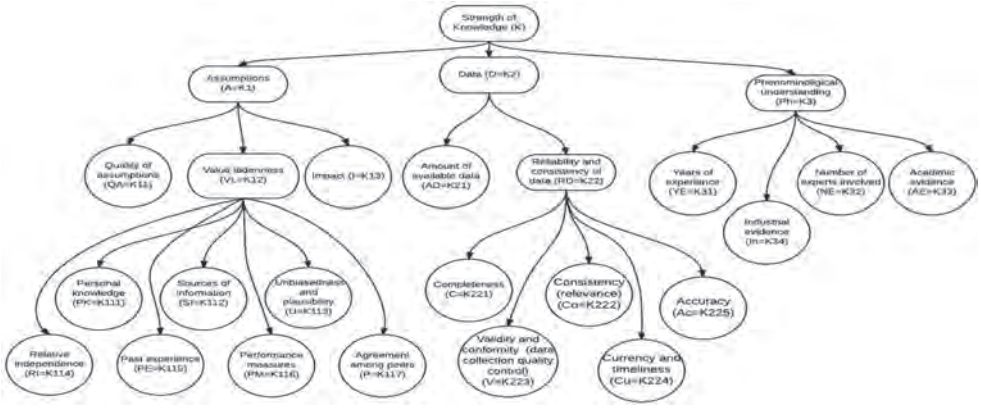


Figure 3. A hierarchical tree-based framework for knowledge assessment.

Table 1. References that justify the model in Figure 3.

Attributes	References
Strength of knowledge is evaluated by K_1, K_2 and K_3.	[4]
Realism and plausibility of assumptions (K_1) is evaluated by the quality of assumptions (K_{11}), the value ladenness and subjectivity of the experts (K_{12}) and sensitivity analysis on the assumptions (K_{13}).	[9]; [10]; [11]
Value ladenness (K_2) is defined by K_{111}–K_{117}.	[10]; [9]; [12]; [13]; [14]
Data (K_2) is evaluated in terms of amount of available data (K_{21}) and reliability of data (K_{22}).	[4]
Reliability of data is defined by: (i) completeness; (ii) consistency; (iii) accuracy; (iv) validity; (v) timeliness.	[15]; [16]; [17]

The tree structure in Figure 3 is constructed based on a thorough literature review related to trustworthiness and validity assessment of PRA/QRA. References related to the construction of the tree model are given in Table 1. It should be noted that for phenomenological understanding, few references directly consider its assessment. A comprehensive understanding of phenomena requires its explanation [8], which depends on the capability of the experts involved in the risk modeling and analyses. Then, four sub-attributes are proposed for the assessment of phenomenological understanding: (i) industrial evidence; (ii) academic evidence; (iii) number of experts involved; (iv) number of years of experience in the domain).

2.2.2 *Evaluation using AHP*

Given the hierarchical tree in Figure 3, the assessment of the strength of knowledge is carried out within a Multi-Criteria Decision Analysis (MCDA) framework. AHP is adopted [18], as it is fit for both quantitative and qualitative evaluation of attributes and factors [19] and for group decision making [20].

A first step in applying AHP is to evaluate the "leaf" attributes (the non-decomposable attributes in Figure 3). A score between 1 and 5 is used to represent the strength of knowledge with respect to each "leaf" attribute, where 1 represents the lowest knowledge level and 5 represents the highest knowledge level. The score is evaluated based on some predefined evaluation criteria. Due to page limits, we only present the evaluation criteria for K_{11} as an example (See Table 2).

Then, the inter-level priorities (weights) are determined for each attribute, sub-attribute and "leaf" attribute, denoted by $W(K_i), W(K_{ij})$ and $W(K_{ijk})$, respectively. Based on [14] and [20], a scale of 1–9 is used for evaluating the importance of each of these attributes relative to each other, with reference to their contribution to the parent attribute: a value of 1 is assigned when two attributes of the same level of the hierarchy are equally important and 9 is assigned when one attribute is significantly more important than the other.

The strength of knowledge of the ith atomic element, denoted by K_i, is, then, calculated as a weighted average of all the scores of the "leaf" attributes. The value of K_i is between 1 and 5 and a high value indicates that we have stronger knowledge on that atomic element.

Table 2. Quality of assumptions scoring guidelines.

Attribute \ Score	1	3	5
Quality of assumptions	The assumptions are based aon weak knowledge and not realistic (conservative assumptions or over-optimistic)	The assumptions are acceptable based on moderate knowledge, simple model and extrapolated data	The assumptions are based on strong knowledge and established theory, verified by peer review and very plausible

2.3 *Knowledge aggregation*

From (8), the risk index of the reduced-order PRA model is the sum of $n_l = \sum_{i=1}^{n_{O,Red}} n_{S,Red,i}$ risk index values $R_{Red,l}$ from the corresponding elementary risk model, where each elementary risk model is further composed of MCS and BEs, as shown in (2):

$$R_{Red,l} = \sum_{k=1}^{n_{MCS,Red,l}} \prod_{q \in MCS_{l,k}} P_{BE,q} \tag{9}$$

In (9), $R_{Red,l}$ is the risk index of the l-th reduced elementary risk model, where $l = 1,2,\cdots,n_l$ and $n_{MCS,Red,l}$ is the number of MCS in the l-th reduced elementary risk model.

Let $K_{BE,l,q}$ denote the strength of knowledge of the q-th BE in the reduced elementary risk, where $K_{BE,l,q} \in [1,5]$ and a large value of $K_{BE,l,q}$ indicates strong knowledge of BE. The $K_{BE,l,q}$s are assessed using the procedures described in Sect. 2.2.

The next step is to aggregate the $K_{BE,l,q}$s to assess the strength of knowledge of the whole risk assessment model. The aggregation should consider the difference in each atomic element's contribution to the total risk.

Different importance measures can be used to evaluate the contribution of the atomic elements with respect to the total risk. Since the elementary reduced-ordered risk model is constructed by the BEs through MCSs, the weights of the BEs are calculated based on Fussell-Vesely importance measures:

$$W_{BE,l,q} = \frac{I_{BE,l,q}}{\sum_{q=1}^{n_{BE,l}} I_{BE,l,q}} \tag{10}$$

where $I_{BE,ij}$ is the Fussell-Vesely importance measure of the corresponding BE in the elementary risk model l.

The strength of knowledge for the l-th elementary reduced order risk model, denoted by K_l, is calculated by:

$$K_l = \sum_{q=1}^{n_{BE,l}} W_{BE,l,q} \cdot K_{BE,l,q} \tag{11}$$

The importance of the elementary reduced-order model is evaluated by its contribution to the total risk:

$$W_l = \frac{R_{Red,l}}{\sum_{l=1}^{n_l} R_{Red,l}} \tag{12}$$

where $R_{Red,l}$ is the risk index value of the l-th elementary reduced order model and is calculated by (9).

To calculate the total strength of knowledge K_{Red} of the reduced-order risk model, the knowledge indexes K_ls of the reduced-order elementary risk models are further aggregated by considering their contributions:

$$K_{Red} = \sum_{l=1}^{n_l} W_l K_l \tag{13}$$

The index K_{Red} is, then, used to represent the strength of knowledge of the entire PRA. Its value is between 1 and 5, and a high value indicates that we have strong knowledge in support of the PRA model and its risk outcomes.

3 CASE STUDY

3.1 *Problem description*

In this section, we apply the developed method to assess the strength of knowledge of NPPs PRAs. Two hazard groups, i.e., external hazards and internal events are considered in this case study.

External hazards refer to the undesired events originating from sources outside the NPPs such as: external flooding, external fires, seismic hazards, etc., [21]. In particular, external flooding is a naturally induced hazard that might be caused due to different reasons such as: tides, tsunamis, dam failures, snow melts, storm surges and etc., (see [22] for more examples). The choice of these initiating events to be a part of the external flooding risk assessment models is site-specific and some guidance should be provided for this purpose [23]. In general, for external flooding, the state of PRA practice is considered less mature than for internal events [24]. For

example, the flood frequencies are obtained using statistical models and by extrapolating design basis flood levels to the fitted historical data (usually limited), which results in a very high uncertainty [24]. Moreover, for extreme floods, the probability of occurrence is very low but, on the other hand, the potential consequences can be catastrophic [22]. The low probability and the consequent lack of data experience introduces large uncertainties in the risk analysis of this type of events [22].

Internal events refer to the undesired events that originate within the NPPs itself, which cause initiating events that might lead to loss of important systems and might eventually result in core meltdown [1]. The internal events are mainly [25]: (i) different types of components, systems and structures failures, missiles and fires; (ii) safety systems operation and maintenance errors. These types of internal events can cause other initiating events such as turbine trip or Loss of Coolant Accidents (LOCAs). The risk assessment of internal events has been significantly developed and considered to have lower uncertainty compared to other hazard groups [1].

3.2 *Evaluation of hazard group strength of knowledge*

In this case study, we consider the risk analysis models of two hazard groups developed by Electricité de France (EDF) using Risk Spectrum Professional software [26]; [27]. The knowledge assessment framework developed in Section 2 is applied to evaluate the strength of knowledge of the risk models for both internal and external events. Technical reports were provided to the experts to support the knowledge assessment with the needed data and information. For simplification, we only present the case of the external events (specifically flooding). For internal events, we only show the results of the application.

3.2.1 *Reduced-order model construction*

Based on Eq. (4) with $\alpha = 0.8$, we found that only one out of six operating states (NS/SG-normal shutdown with cooling using steam generator-NS/SG) is needed for the reduced-order model, which contributes to 86% of the total risk index. Similarly, based on Eq. (5) with $\beta = 0.8$, only one out of ten scenarios (water levels) is needed for the reduced-order model, whose risk contribution is 98.7% Based on Eq. (7) with $\gamma = 0.8$, the number of MCSs needed for the reduced-order model is 5 out of 3102, and the risk contribution is 80.1%. Therefore, a reduced-order model is constructed based on the atomic elements in Table 3, as shown in Figure 4.

3.2.2 *Strength of knowledge assessment*

After constructing the reduced-order model, the knowledge assessment framework in Section 2.2 has been applied on each of the atomic elements and, then, AHP is used to compare the overall model strength of knowledge. The strength of knowledge for external flooding turns out to be $K_{EF} = 2.78$. The results of the knowledge assessment for both hazard groups are graphically illustrated in Figure 5. It can be seen from the Figure that the strength of knowledge on the internal events is higher than that on external flooding. In fact, these results confirm our expectations. In addition, most of the risk assigned to the external flooding is due to two basic events (failure to close the isolating valve in the auxiliary feedwater system and failure of the containment spray system), whose strength of knowledge is very weak.

Table 3. Reduced-order model constituents.

Operating state	Scenarios	Number of MCS	Number of basic events BE	Total risk contribution
NS/SG	Water level A	5	10	$0.86 \times 0.987 \times 0.801 = 0.6799$

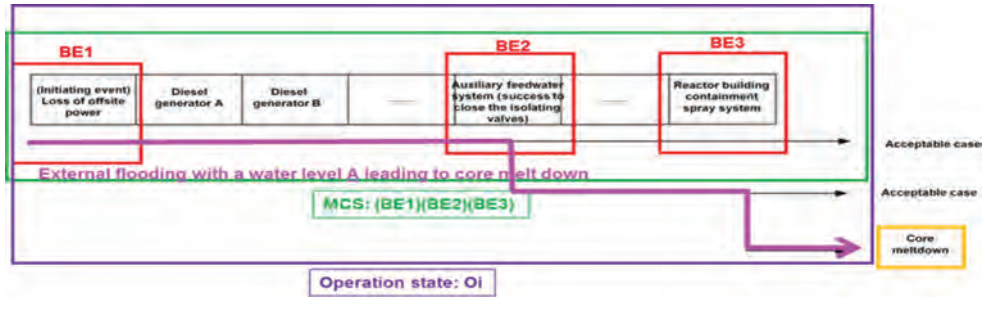


Figure 4. An illustration of the reduced-order model.

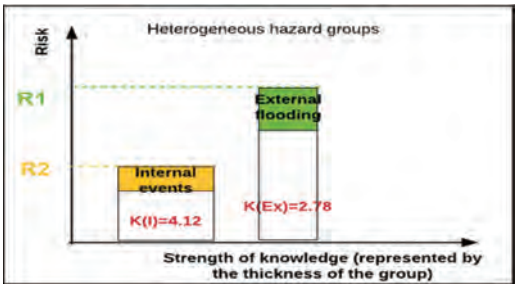


Figure 5. Representation of hazard groups levels of risk and strength of knowledge.

4 CONCLUSION

An analytical hierarchy process-based framework has been proposed for assessing the strength of knowledge of PRA models. The framework is based on three main attributes (assumptions, data, and phenomenological understanding), which are further decomposed into sub-attributes and "leaf" attributes. A reduced-order PRA model is constructed, that reduces the number of atomic elements to be analyzed. The framework has been applied on two hazard groups in NPPs.

In the future, model uncertainty in the PRA model will be considered for a more comprehensive knowledge assessment. In addition, in the current framework, the weights of the attributes in the AHP were subjectively evaluated. Future investigations will be devoted on how to more objectively evaluate the weights.

REFERENCES

[1] EPRI, "An Approach to Risk Aggregation for RisK-Informed Decision-Making," Palo Alto, California, 2015.

[2] T. Aven, "Practical implications of the new risk perspectives," *Reliab. Eng. Syst. Saf.*, vol. 115, pp. 136–145, 2013.

[3] IAEA, "Technical Meeting on Managing Nuclear Safety Knowledge—Approaches and National Experiences," 2017.

[4] Flage and T. Aven, "Expressing and communicating uncertainty in relation to quantitative risk analysis," *Reliab. Risk Anal. Theory Appl.*, vol. 2, no. 13, pp. 9–18, 2009.

[5] E. Zio, "An introduction to the basics of reliability and risk analysis. Series in Quality," *Reliab. Eng. Stat.*, vol. 13, 2007.

[6] RELCON AB, "Theory Manual," 2005.

[7] R. Koch, *The 80/20 principle: the secret to achieving more with less*. Crown Business, 2011.

[8] C. Kelp, "Understanding phenomena," *Synthese*, vol. 192, no. 12, pp. 3799–3816, 2015.

[9] I. Boone *et al.*, "NUSAP: a method to evaluate the quality of assumptions in quantitative microbial risk assessment," *J. Risk Res.*, vol. 13, no. 3, pp. 337–352, 2010.

[10] P. Kloprogge, J.P. Van der Sluijs, and A.C. Petersen, "A method for the analysis of assumptions in model-based environmental assessments," *Environ. Model. Softw.*, vol. 26, no. 3, pp. 289–301, 2011.

[11] P.A. Stirling, "On Science and Precaution in the Management of Technological Risk: Volume II-case studies," 1999.

[12] E. Laes, G. Meskens, and J.P. van der Sluijs, "On the contribution of external cost calculations to energy system governance: The case of a potential large-scale nuclear accident," *Energy Policy*, vol. 39, no. 9, pp. 5664–5673, 2011.

[13] J.P. Van Der Sluijs, M. Craye, S. Funtowicz, P. Kloprogge, J. Ravetz, and J. Risbey, "Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System," *Risk Anal.*, vol. 25, no. 2, pp. 481–492, 2005.

[14] E. Zio, "On the use of the analytic hierarchy process in the aggregation of expert judgments," *Reliab. Eng. Syst. Saf.*, vol. 53, no. 2, pp. 127–138, 1996.

[15] M. Bergdahl, M. Ehling, E. Elvers, and E. Földesi, *Handbook on Data Quality Assessment Methods and Tools*. 2007.

[16] DAMA, "The six primary dimensions for data quality assessment: defining data quality dimensions," 2013.

[17] IAEA, *Data Collection and Record Keeping for the Management of Nuclear Power Plant Ageing*, no. 50. 1991.

[18] T.L. Saaty and L.G. Vargas, *Models, methods, concepts & applications of the analytic hierarchy process*, vol. 175. Springer Science & Business Media, 2012.

[19] M. Alexander, "Decision-Making using the Analytic Hierarchy Process (AHP) and SAS/IML," *United States Soc. Secur. Adm. Balt.*, pp. 1–12, 2012.

[20] T.L. Saaty, "Decision making with the analytic hierarchy process," *Int. J. Serv. Sci.*, vol. 1, no. 1, p. 83, 2008.

[21] IAEA, "Development and Application of Level 1 Probabilistic Safety Assessment for Nuclear Power Plants," 2010.

[22] IAEA, "External Events Excluding Earthquakes in the Design of Nuclear Power Plants," 2003.

[23] IAEA, "Meteorological and Hydrological Hazards in Site Evaluation for Nuclear Installations," 2009.

[24] EPRI, "Practical Guidance on the Use of Probabilistic Risk Assessment in Risk-Informed Applications with a Focus on the treatment of Uncertainty," Palo Alto, California, 2012.

[25] IAEA Safety Standards Series, "Deterministic Safety Analysis for Nuclear Power Plants," 2009.

[26] EDF, "Technical report: Qualitative Analysis for LOOP due to internal cause."

[27] EDF, "Report technique: Approche Graduée - Analyse Probabiliste du Risque pour l â€™ Inondation Externe : Quantification des Scénarios Accidentels," vol. 33, no. 1, 2015.

Economic analysis in risk management

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Formalization of RAM contracts for advanced consistency and completeness checking

A. Joanni & D. Ratiu
Siemens AG, Corporate Technology

ABSTRACT: Non-compliance with contractual terms concerning Reliability, Availability and Maintainability (RAM) can have significant financial impact on the profitability of technical projects due to e.g. penalties or warranty costs. Often, these terms are stipulated in complex (multi-party) contractual agreements. In practice, these contracts are captured as natural language prose and this can lead to ambiguities, inconsistencies or incompleteness. In this paper, we present the benefits obtained when RAM contracts have a precise formal description expressed in a semantically rich language. Examples of benefits are advanced validation and consistency checks such as overlap of time periods, gaps in warranties between parties, transfer of risks to other parties, even before a numerical evaluation is performed. We present a brief theoretical background of the approach, an illustration by means of exemplary contractual warranty terms, and give an outlook on further possibilities and advantages.

1 INTRODUCTION

Non-compliance with contractual terms concerning reliability, availability and maintainability (RAM) characteristics can significantly impact the profitability of technical projects. Complex projects involve contractual agreements between different par-ties working in different stages of the supply chain. These contracts must be perfectly synchronized and orchestrated to avoid unnecessary non-conformance costs for the contractor, for instance due to failures of the integrated components.

Common practice today is that contracts are written in plain natural language and thereby are often ambiguous and incomplete. Furthermore, natural language contracts are not amenable to automatic analyses.

A typical example of complex RAM contracts is related to warranties. Servicing of warranty cases involves additional (non-conformance) costs to the contractor and these costs depend on product reliability and contractual warranty terms between the contractor and the customer. Furthermore, these costs can be (fully or partially) covered by warranty contracts between the contractor and the individual suppliers. As extensively discussed in literature, see e.g. Murthy & Jack (2003), there is a wide variety of warranties policies determined by different warranties periods and associated costs. Examples of different periods are simple periods, extended warranties, life-time warranties, physical life, technological life; examples of different forms of penalties are product renewal, repairing and looping of warranty. Taking informed decisions about liability due to warranties requires transparent assessment of contractual requirements between all stakeholders (customer, contractor, suppliers). Other examples of complex RAM contracts are related to performance guarantees of technical systems.

Furthermore, automatic analysis of "what if" questions like "what additional costs would occur if the contractor offered extended warranty to customers" or "what is the longest warranty a contractor can offer such that the risk budget is under a certain threshold with a given confidence" require precise modeling and analysis of contracts.

In this paper we present our work in progress on using domain specific languages to model contracts in a precise and unambiguous manner. Once semantically rich contract models are available, they represent a basis for further analyses for consistency and completeness. As an example we model warranty clauses of multi-party contracts.

2 MODELING OF RAM CONTRACTS

2.1 *Previous work*

Formalization of (business) contracts using formal models and languages with precise semantics has been subject of research for about two decades; see e.g. Peyton Jones et al. (2000), Hvitved (2011) or Farmer & Hu (2016). In particular, these approaches have been adopted early on by the financial industry for pricing of contracts that

were traded e.g. in the financial derivative markets. It was realized that formalization of contracts not only gives a precise way to describe complex contracts, which may immediately reveal ambiguities of contracts expressed in a natural language. Moreover, the formal semantics provides the basis for analyzing and integrating formal methods into contracts, as well as for automated contract management. Last but not least, it provides a means to valuate contracts over time. All of this is, naturally, compositional and comes with a high degree of reusability, modularity and abstraction.

Modeling and valuation of contractual RAM requirements has been demonstrated in a previous paper by Joanni & Ratiu (2018), based on the pioneering work by Peyton Jones et al. (2000). Namely, a compact library of so-called contract primitives is used to perform compositional analysis of contractual RAM requirements expressed in the library, and based on this it is shown how to perform valuation of a broad range of contractual requirements. The contract primitives are supplemented by the concept of so-called observables, which determine how the meaning (and value) of a contract evolves over time. Observables, for instance, define when certain conditions become true that entail a certain payment at that time, or what the precise amount of a payment is. In the context of contractual RAM requirements: whether a given component has failed or not may constitute an observable; another example of an observable is the average availability of a module of a process plant over a given time period. In essence, observables are concepts that glue together the formal contract model and the model of the technical system.

2.2 *Modeling of technical and contractual RAM aspects*

This section briefly exemplifies and reiterates how contractual RAM requirements are expressed in terms of contract primitives and observables as demonstrated by Joanni & Ratiu (2018), in order to set the stage for advanced consistency and completeness checking of RAM contracts which is the focus of this paper.

For the purpose of contractual RAM requirements, the modeling of technical RAM aspects is one source of observables that determine how the meaning (and value) of a contract evolves over time. For instance, suppose that the warranty risks for component failures arising from a contract for a major technical project are to be analyzed, and the event of a control system failure of a given subsystem, say system 2, constitutes an observable whose value becomes true as soon as a control system failure occurred. Let this observable be denoted by *system2:controlFailure* in the following.

Some of the contract primitives are introduced next, in order to show how contractual RAM requirements can be composed in terms of these generic contract primitives and linked to observables. First, the contract *(one EUR)* represents a contract where, once acquired, the holder immediately receives one unit of the currency Euro. The contract *(give c)* is a contract where all rights and obligations arising from contract *c* are reversed. If the contract *(or c1 c2)* is acquired, then the holder must immediately choose whether to acquire *c1* or *c2*, where the choice will be such that the benefit is greatest. Finally, once the contract *(and c1 c2)* is acquired, both of the contracts *c1* and *c2* are immediately acquired. As an example, if the contract.

(give (and (one EUR) (one EUR)))

is acquired, then holder is obligated to immediately pay two Euros. Consequently, in order to be willing to acquire the contract above, one would expect to receive two Euros in return as a fair price. Generally, we consider only a fair price when we talk about contract valuation, and the usual profit margin would have to be added afterwards.

Up to now, the contract primitives did not refer to any observables, and all rights and obligations took effect immediately. These limitations will be overcome with the following contract primitives. If the contract *(scale o c)* is acquired, then the contract *c* is immediately acquired where all payments arising from the contract *c* are multiplied by the value of observable *o* at the moment of acquisition. The contract *(when o c)* implies that, once acquired, the holder must acquire the contract *c* as soon as the observable *o* becomes true.

It should be noted that, while both observables introduced in the previous paragraph are in fact stochastic processes, the first one can be seen as a numerical value, whereas the second one is a stochastic process that assumed the values true or false. Other examples of the latter kind are observables like *(at date d)*, which becomes true at a certain date *d*, or *(between d1 and d2)*, which assumes the value true between the dates *d1* and *d2* (and false otherwise).

Using the contract primitives introduced so far, we will now give examples how these can be linked with observables, and how contractual RAM requirements of practical relevance can be expressed by exploiting the compositional aspects. For instance, the contract.

(when (at date 2017-01-31) (scale 100 (one EUR)))

pays the holder of the contract the amount of 100 Euros on January 31, 2017. In order to determine a fair price for the contract at the time of acquisi-

tion, one would have to calculate the discounted present value at the time of acquisition depending on interest rates (and possibly currency exchange rates). Certainly, the value of the contract is zero if acquired after January 31, 2017. Similarly, we can define a contract.

(when system2:controlFailure (give (scale 100 (one EUR))))

which implies that the holder of the contract is obligated to pay 100 Euros on occurrence of the event *system2:controlFailure* (see beginning of the section). This already constitutes a simple contractual RAM requirement which could potentially be found in a real contract. The value of this contract at the time of acquisition, i.e., the fair amount of money the contractor would receive once it has entered into the contract, depends here on the number of occurrences within the contract duration, and on the time instants the event occurs (in case of discounting).

Next, consider a typical warranty clause from a technical contract that requires that a certain amount be paid to the customer on occurrence of failures of respective components (or, equivalently, the components must be replaced at the contractor's expense). This only applies within a given warranty period, and the total amount is limited by a cap of 20,000 Euros. Expressed in terms of contract primitives, this could take the following form.

(or (scale 20000 (give (one EUR))) (and (when (system2:controlFailure && between startOfProject and November 03, 2021) (give (scale 800 (one EUR)))) (when (system3:valveFailure && between startOfProject and November 03, 2021) (give (scale 130 (one EUR)))) …))

where a Boolean expression is used in the condition, and the date *startOfProject* is a variable that was defined elsewhere in the model.

It is noted here that the possibility to compose more complex contractual RAM requirements from contract primitives is a key point and adds great value to our approach. Without it, it would be hardly practicable to capture and analyze the various types of contractual requirements that may be encountered in realistic contracts.

It is also important to note that, since our approach targets persons in charge of technical, contractual and commercial aspects who are domain specialists and do not necessarily have any programming background, the expressions in terms of contract primitives can be just as well automatically generated from higher-level expressions of contractual requirements, which are closer to natural language formulations. The reader is referred to Joanni & Ratiu (2018) for details.

3 ADVANCED CONSISTENCY AND COMPLETENESS CHECKING

3.1 *Formalization of the contract model*

What has been presented in Section 2 is essentially a model-based form of a contract. This can be expressed in form of a domain specific language with constructs which directly represent concepts from the domain of contracts. Individual contracts can thereby expressed as models written using the domain specific language. The model can be represented as a tree (see Figure 1 above for the example from the previous section). The value of the contract at a certain point in time is determined by a precisely defined, so-called *valuation semantics*; see Peyton Jones et al. (2000). Moreover, prior to numerical evaluation, it may be useful to apply certain transformation rules in order to obtain the contract in an optimized form (ibid.).

Typically, a project involves contractual obligations that are not restricted to one party, but are stipulated in several contracts with multiple parties. Several contracts can be merged in a single one where the individual contracts are subsumed by an *and* contract. Care must be taken, though, to correctly express the rights and obligations arising from the individual contracts from the point of view of the party of interest (e.g. the main contractor). An observable such as a failure of a specific component then usually affects more than one contract, if the event gives rise to contractual rights or obligations for more than one of the parties involved.

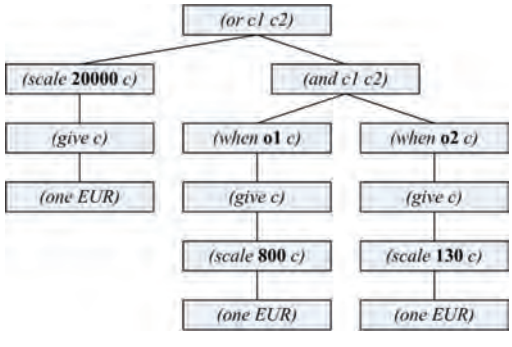


Figure 1. Tree representation of a contract. Observables are shown in bold face (the observables related to failure events are replaced by **o1**, **o2**).

3.2 *Enriching the contract model with additional semantics*

Generic contracts can be enriched with additional semantics, such as whether a given branch in the tree represents the contract with a specific party, for instance the customer and the various suppliers. One possible way to implement this is by annotating certain branches of the tree with information on the respective contractual party. Moreover, certain parts of the tree can be annotated with information on the specific type of payments arising from this part of the contract, for instance contractual penalties or warranty payments. In this way, information on the type of payment and the contractual parties involved can be retained and recorded during the numerical evaluation, e.g. through stochastic simulation.

Once contracts have a precise formal description expressed in a semantically rich language, it is possible to implement advanced validation and consistency checks such as overlap of time periods, gaps in warranties between parties, transfer of certain risks to other parties etc., even before a numerical evaluation is performed. Some more concrete examples are given in the following:

- In the context of warranty costs, a formal check can reveal if warranty payments owed to the customer in case of component failures is always (fully or partially) balanced by warranty payments received from the respective suppliers (back-to-back coverage). This is not always obvious for complex systems with many components, where warranty periods may be even be subject to looping conditional on certain events.
- In case of random events which entail payments and/or liabilities from different parties, a formal analysis can help determine bounds on the overall financial impact. This is extremely useful particularly if the frequency of occurrence is based on uncertain estimates, but determining bounds may not be an easy task to do manually if certain payments or liabilities are subject to further conditions.

This list is certainly not exhaustive and depends on the specific technical system and contractual aspects under consideration.

4 TOOLING AND IMPLEMENTATION

In our implementation of the approach, the financial and technical aspects of RAM contracts are expressed with the help of domain specific languages (DSLs). A DSL provides a formal syntax (meta-model) for representing the domain concepts and relations between them. Once appropriate constructs are available, simple consistency checks which define the set of well-formed contracts can be implemented easily. In addition, advanced consistency and completeness checks can be implemented as described in this paper.

The technical basis for our implementation is language engineering technologies as provided by the Jetbrains MPS language workbench (www.jetbrains.com/mps). MPS allows creation and composition of different modular DSLs, which enables domain experts to directly and explicitly express concerns from their business domain. We created DSLs to describe both the technical and commercial aspects of contractual RAM requirements, so they are available as models. Compared with previous work on (business) contract formalization, which was e.g. implemented in Haskell and requires know-how about functional programming, our approach targets persons in charge of technical, contractual and commercial aspects who are domain specialists, and it does not require any programming background.

Moreover, our implementation gives confidence in evaluation by validation of input and advanced consistency checks; the models are "correct by construction" and allow time saving due to efficient specification of the technical, contractual and commercial aspects of contractual RAM requirements.

5 EXAMPLE

In order to demonstrate the modeling and advanced checking of contractual RAM requirements, a simple example in the context of product warranty is given in the following. For a comprehensive review of warranty policies see e.g. Murthy & Jack (2003) or Rahman & Chattopadhyay (2006). In this paper, we restrict ourselves to the case of so-called *Free Replacement Warranty* without renewal, where a certain amount is to be paid to the customer on occurrence of failures of certain components within a fixed time interval. This amount may comprise the costs for replacement of the failed components as well as additional servicing costs. The corresponding model of the customer contract is shown in the top pane of Figure 2, which repeats the exemplary warranty clause from Section 2.2. The model is annotated by a label "CustomerWarranty" in order to indicate that, within this branch of the tree, any observables related to component failures incur a liability towards the customer. The specific amount to be paid, the warranty periods and the fact that a cap was agreed are not considered here, but will certainly influence the value of the contract. The intention here is to check whether warranty costs due to component

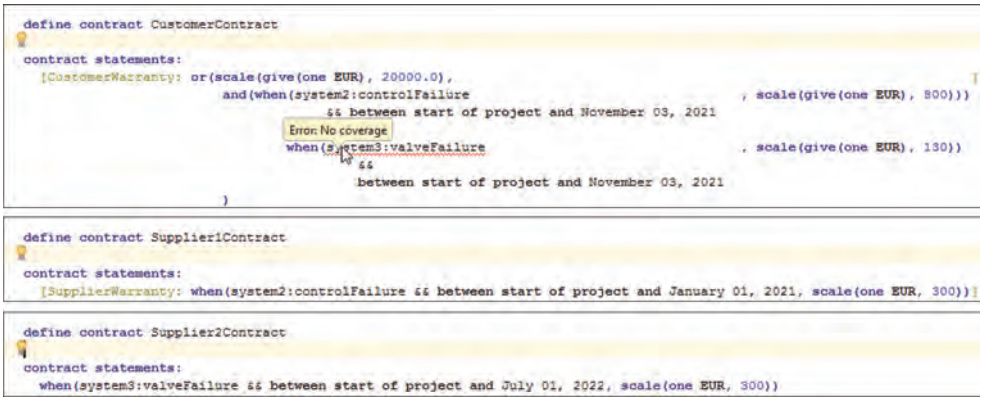


Figure 2. Exemplary customer contract (top pane) where the warranty clause indicates an incomplete coverage by an error message. This is because only the contract with supplier 1 (middle pane) is annotated as a supplier warranty, but the contract with supplier 2 (bottom pane) is not.

failures can, in principle, be balanced by warranty payments received from the respective suppliers, based on the contracts entered.

Similarly, the middle pane of Figure 2 shows a model for a supplier contract that states that failure of the component *system2:controlFailure* incurs a payment received from the supplier of this component. Here, the corresponding model is annotated by a label "SupplierWarranty" in order to indicate that, within this branch of the tree, any observables related to component failures imply that a claim can be made towards the supplier. The bottom pane of Figure 2 shows a model for another supplier contract for payments received from the supplier on occurrence of failures of component *system3:valveFailure*. Here, however, the model was not annotated by a label "SupplierWarranty", which will be revealed by the checking rule as shown below.

An advanced checking rule was implemented in order to identify component failures that only incur a liability towards the customer but no corresponding claims towards the supplier, based on all contracts entered. This was done by traversing the tree of the contract model and searching for observables related to component failures that are annotated with the label "CustomerWarranty", but not "SupplierWarranty". Consequently, an error message is shown in the model of the customer contract that points out the respective components which failures are not covered by a warranty clause in any of the supplier contracts. In this example, the obvious remedy would be to annotate the contract with the second supplier with the label "SupplierWarranty". However, it may well be the case for complex projects that it has been overlooked to agree on a warranty clause for one or more components. In this case, the error message suggests that it would be advisable to renegotiate appropriate warranty agreements with the suppliers of the respective components.

6 DISCUSSION AND OUTLOOK

Based on our approach to express contractual RAM requirements by means of a precise formal description with a semantically rich language, we have explained the principles of performing advanced validation and consistency checks of certain contractual characteristics of interest (in addition to providing a way to efficiently evaluate the contractual requirements as demonstrated in Joanni & Ratiu (2018).) The example presented in this paper demonstrates the implementation for a rather simple example of warranty costs. However, the contract model can, in principle, also be transformed into other mathematical models that allow for more sophisticated checking algorithms that cover various aspects such as completeness of conditions. This is the subject of our ongoing research.

In our opinion, formalization of contractual RAM requirements is not only of theoretical interest, but is of practical applicability and provides significant benefits such as efficient assessment and potential reduction of non-conformance costs, in addition to increased transparency particularly when dealing with complex contractual agreements for projects where multiple parties are involved.

REFERENCES

Farmer, W. M. & Hu, Q. 2016. A Formal Language for Writing Contracts. In Proc. *IEEE 17th International*

Conference on Information Reuse and Integration (IRI).

Hvitved, T. 2011. *Contract Formalisation and Modular Implementation of Domain-Specific Languages*. PhD thesis, Department of Computer Science, University of Copenhagen.

Joanni, A. & Ratiu, D. 2018. Modeling and Valuation of Contractual RAM Requirements Using Domain-Specific Languages. In *Proc. 64th Annual Reliability and Maintainability Symposium, Reno, NV, January 22-25, 2018.* IEEE.

Murthy, D. N. P & Jack, N. 2003. Warranty and Maintenance. In Hoang Pham (ed), *Handbook of Reliability Engineering*, Springer, London, 2003.

Peyton Jones, S., Eber, J.-M., Seward, J. 2000. Composing Contracts: An Adventure in Financial Engineering. In *Proc. ACM SIGPLAN 5th International Conference on Functional Programming*.

Rahman, A. & Chattopadhyay, G. 2006. Review of long-term warranty policies. *Asia-Pacific Journal of Operational Research* 23(4): 453–472.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Cost-benefit analysis for non-structural flood risk mitigation measures: Insights and lessons learnt from a real case study

G. Pesaro, M.T. Mendoza, G. Minucci & S. Menoni
Politecnico di Milano, Milan, Italy

ABSTRACT: Cost-Benefit Analysis (CBA) in flood risk management is becoming increasingly popular as a tool that makes the relationship between investments in mitigation measures and the related effects more comprehensible. Even if the application of CBA in flood risk management is nothing new, there is still limited evidence on its use conditions and possible limits when applied in real. The essay first suggests a new classification of Flood Mitigation Measures (FMM), towards a new taxonomy that goes beyond the distinction between structural and non-structural. This distinguishing between risk and damage reduction measures, public and private decision-making processes, mandatory and voluntary actions and allocating more importance to the value of the avoided damages related to the commons (i.e. cultural heritage) and other non-renewable, non-reproducible or non-restorable territorial resources. Second, the contribution of CBA as an ex-ante decision-making tool in flood risk reduction is discussed. Moreover, the essay offers a methodological insight and operational elements as results of a case study developed in the European research project IDEA, where a CBA for a non-structural mitigation measure was performed using the data and information available after the 2012 and 2013 flood events in the Umbria Region (Italy). Here, dams built to produce hydroelectric power have been used for laminating the floods, as a non-structural risk mitigation measure. The experimental application of the CBA to this measure, including the combination of real damage data collected after the floods and damage modeling for the alternative scenario, provided methodological and operational evidence of its capability to reduce/avoid a part of the damage. Finally, the essay presents the lessons learnt, the open problems and future developments required for an effective CBA, in reference to the technical and scientific perspective and to the difficulties in the understanding and interpretation of the whole of the cause-effect chains and externalities.

1 INTRODUCTION

In recent years, Cost Benefit Analysis (CBA) has been increasingly proposed as a key tool to support decision-making processes for flood risk prevention and mitigation, including in the European Floods Directive 2007/60/EC. It serves the purpose of comparing the costs of mitigation measures with the potential benefits interpreted as reduction of potential Damage and Losses (D&L). However, there are a number of limitations implied by most common C&B analyses regarding a variety of issues (Ale et al. 2015). There are difficulties in assessing the entirety of costs, for example indirect ones, questionable values are attached to both costs and benefits without adequate empirical supporting evidence; ethical and equity concerns arise, for example, when intangibles are assessed in monetary terms. In order to address some of those limitations, we have proposed some innovative procedures while carrying out C&B analyses according to a more rigorous economic thinking. At first, a better classification of Flood Mitigation Measures (FMM) with respect to most C&B applications, is proposed, as the variety of the possible solutions and the related results when implemented is very high. This because both costs and benefits can be better assessed if mitigation measures are better identified and distinguishable one from the other.

In fact, a better classification of FMM has a twofold purpose. On the one hand, if exhaustive enough, it offers planners a wider variety of options which to choose from (Yevjievich 1994). On the other hand, diverse FMM typologies have different costs and therefore require different degree of investment. A better classification of FMM may also help in highlighting for each combination of measures the different mix of direct and indirect costs that are associated.

Bouwer et al. (2014) underline the fact that there is often a solid base of information on the direct cost in hard mitigation measures, whereas there are few cost assessment methodologies for non-structural measures. Moreover, Mechler's review (2016) on available studies on CBA for DRR measures shows that mostly structural measures are considered.

In this context, we first propose a new classification framing the different types of FMM. Then a focus is made on the analysis of a non-structural measure, namely the use of dams for electric hydropower production to retain potential overflow due to heavy rains. This case-study introduces a second level of novelty in the arena of most widely applied C&B methodologies, in that it uses real damage and costs experienced after events that have actually occurred as a starting point for establishing monetary values of benefits intended as avoided damage. This in line with the methodological approach taken by the European funded project, IDEA (Improving Damage assessments to Enhance cost-benefit Analysis).

2 FLOOD MITIGATION MEASURES (FMM)

2.1 *Scientific background: FMM classification*

Yevjievich et al. (1994) classify FMM on the basis of a geographic approach according to (i) adjustment to natural hazards (ii) typology of flood damage prevention; (iii) typology of flood damage reduction; (iv) typology of flood policymaking and (v) basic categories of individual measures.

The Australian Bureau of Transport and Regional Economics in 2002 proposed to classify FMM from an economic perspective according to (i) flood modification, (ii) property modification and (iii) response modification. On the basis of the latter, Hawley et al. (2012) categorize FMM by (i) structural and non-structural flood control, (ii) exposure and property modification and (iii) behavioral response modification. Considering the time scale before the damaging event, Mechler (2005) differentiates between FMM that reduce risk (mitigation/prevention and preparedness) or transfer and spread it on a larger basis (risk financing).

The FLOOD-ERA Joint Report (Schanze et al. 2008) enlarges the economic perspective by considering both ex-ante and ex-post evaluation of structural and non-structural flood mitigation measures.

A multi-sectoral perspective on FMM ex-post evaluation is provided in the FLOOD*site* project (Klijn et al. 2009), which takes into account not only the economic efficiency but also the effectiveness, robustness, and flexibility of selected measures. Accordingly, measures (i.e. physically tangible interventions) are differentiated from instruments (i.e. interventions that cause effects indirectly).

2.2 *Towards a new taxonomy*

On the base of the literature review, a preliminary list of a large selection of possible flood mitigation measures and a new FMM classification have been developed (Table 1). These measures have been classified according to the following criteria: (i) typology, i.e. structural or non-structural measures, (ii) purpose, i.e. "risk" or "damage" mitigation, (iii) who takes the decision, (iv) typology of action. The definition for structural and non-structural measures adopted in this work is the one by the UNISDR (see Terminology on DRR, 2017).

Regarding non-structural measures, we propose as sub-categories: riverine environment-based (e.g. river management), built environment based (e.g. building regulations), social involvement-based (e.g. education programs) and economic-based (e.g. risk transfer through insurance).

Concerning risk mitigation, investments are directed to the reduction of flood exposure, no matter how vulnerable the exposed subjects and objects are. As for D&L, reduction may mean the capability: (i) to reduce exposure, (ii) to recognize and intervene on the vulnerability of a great variety of individual situations, and (iii) identify the potential loss of territorial resources/values that are either unreplaceable or uneasily restored/rebuilt.

Moreover, flood mitigation measures involve, in different ways, public and private actions, decision-making and investments. It is therefore important to address attention not only to risk or damage mitigation measures or to structural and non-structural interventions, but also to the main decision-makers involved in their implementation. These were categorized as "public", "public-private", "economic subject" (private firms) and "private individuals". Regarding the typology of action, three sub-categories were defined: "regulated" (when regulation exists with defined procedural aspects), "mandatory by law" (binding action by law) and "voluntary" action. Of course, particularly in this last category, there might be substantially different results for different territorial areas and countries, according to the different regulation and institutional systems.

Table 1. The proposed new FMM classification.

	Mitigation measures	Mitigation type		Main decision-maker				Action typology		
Typology	Description	Risk	Damage	Public	Public-Private	Private firms economic subject	Private-Individual	Regulated	Mandatory by law	Voluntary
Structural	Levees	X		X				X		
	Dams	X		X	X	X		X		X
	Diversions and channel improvements	X		X				X		
	Flood walls	X		X				X		
	Detention Basins	X		X				X		
	Urban Drainage and pumping	X		X				X		
	Building Drainage and pumping or flood gates	X				X	X	X		
NonStructural	Riverine environment-based									
	Restoration of Flood Plain	X	X	X					X	
	Environmental protection designations		X	X					X	
	River management (i.e.: dredging of sediments)		X	X				X	X	
	Retention areas designation	X	X	X	X			X		
	Dam Management (i.e.: use of existig dams for flood lamination)	X	X		X	X		X		X
	Built environment-based									
	Zoning and land use planning (plans and relocation)		X	X				X	X	
	Purchase or acquisition	X		X	X			X		X
	Building regulations and enforcement		X	X					X	
	Building flood proofing (i.e: house raising, green roof, etc.)	X	X				X	X		
	Infrastucture flood-proofing		X	X		X	X	X		
	Social involvement-based									
	Information on risk management plans		X	X					X	
	Education programmes		X	X						X
	Preparedness (i.e.: emergency planning/state and individual)		X	X		X	X		X	X
	Forecasts and warining systems		X	X					X	X
	Evacuation of “assets”		X	X	X	X	X	X		X

(*Continued*)

Table 1. (*Continued*).

	Mitigation measures	Mitigation type		Main decision-maker				Action typology		
Typology	Description	Risk	Damage	Public	Public-Private	Private firms economic subject	Private-Individual	Regulated	Mandatory by law	Voluntary
	Economic-based									
	Economics (Financial) incentives/disincentives for risk management (i.e.: subsidies for relocation and adaptation, insurance premium according to flood zone)		X	X					X	
	Risk transfer (by means if (re-) insurance public assets		X	X					X	X
	Risk transfer (by means if (re-) insurance or relief funds) for private assets		X			X	X		X	X
	National and local reserve funds		X	X						X
	Loss compensation by public funds		X	X					X	X
	Administrative changes (i.e.: merging and segregatin emergency management departments, or changing the work culture/processes within/between them, to increase efficency in reducing risk)		X	X		X				X
	Health regulations		X	X					X	
	Tax adjustments		X	X					X	
	Business continuity planning (for infrastructure)		X			X				X

3 CBA FOR FMM

3.1 *Economic thinking and CBA in flood risk management*

The assessment methods and tools to quantify some of the damage categories are central elements when referring to the economic approach to disaster risk reduction. The debate developed during the last two decades converge to that damages suffered by populations and the built environment are nowadays better understood than in the past. In the contrary, less evidence is available on damage suffered by economic sectors and by the entirety of territorial resources, encompassing cultural heritage, the natural environment, and shaping communities' identity and social models (some of these elements are discussed in Mechler 2016). This is mainly due to the incidence of indirect and systemic damage (Cochrane 2004a,b), that are instead often underestimated (Shreve & Kelman 2014) but also to the difficulties in assessing the monetary value of intangibles, such as public resources, cultural and historical heritage (Pesaro 2005).

Looking at the whole of the resources a territorial area can rely on to sustain its production and consumption models, qualitative and quantitative growth, the economic perspective suggests to account not only direct quantitative damage, using money as the measure unit, but also indirect and systemic D&L even if not easily quantifiable and measurable. Monetization models have been developed in the field of environmental economy, first, and in risk and damage matters afterward, for instance, for cultural heritage and natural environment, but still, monetary evaluation remains very controversial (see, among others, Meyers et al. 2013).

Cost-Benefit Analysis (CBA) for the evaluation of flood mitigation measures is an important and widely used tool for decision makers to rely on, even though sometimes the incidence of non-monetizable values limits its potential. In a decision-making framework, the main question is how to choose among different possible action/intervention alternatives using an economic-based toolbox, whose strength also lies in the use of quantitative measure units, able to offer clear-to-read results to support and address a selection process. Moreover, it enables to look at the results envisaged for different mitigation measures not only concerning the technical and operational performance but also in terms of investments effectiveness.

CBA can be assessed following an economic approach, which is preferred and adopted in the preset article, and a purely financial one. According to the economic approach, costs and benefits associated with a policy or a project ensure a more exhaustive assessment of damage and effects, including those to third parties. Whereas the financial approach to CBA is mainly based (if not exclusively) on a cash flow analysis. Consequently, economic CBA aims at highlighting the whole range of values and externalities implied by a certain mitigation measure.

Several studies are available introducing cost-benefit analysis as a tool to assess the economic feasibility of flood management strategies (Botzen et al. 2017). Brouwer & van Ek (2004), for instance, justify changes in land use and floodplain restoration in the Netherlands based on both cost-benefit and multi-criteria analyses when ecological, social and economic benefits in the long term are considered. Jonkman et al. (2004) investigate the application of CBA in decision-making on flood protection measures in the Netherlands as well. Joseph et al. (2014) apply conceptually the cost-benefit analytical framework to flood risk adaptation measures taken at the property level in the UK and show the involved costs and benefits. Moreover, the EU Floods Directive (2007/60/EC) requires CBA for supporting public decision making to tackle flood risk in all member states of the European Union.

The use of CBA to enhance the implementation of disaster mitigation measures relies on the concept of intervention profitability, which, in its turn, refers to the capability of investments in mitigation measures to obtain the expected outcomes in terms of risk prevention and/or damage mitigation. It is a method to compare the whole range of direct costs associated with each typology of mitigation measure and their relative outcomes, measured as the total value of the avoided damage and losses plus the benefits coming from an increase in safety and territorial quality.

In this conceptual framework, CBA is seen as an ex-ante decision-making tool, which calls for the capability to develop the assessment process in an ex-ante perspective. In this case, mitigation measures in flood risk allow the system to act on both risk and damage mitigation and reduction.

The results of a CBA might be effective also in the light of a common problem decision-makers encounter very often, that is the perception of direct and quantifiable costs related to mitigation measures implementation. Quantified costs, even when related to damage prevention, are much more evident and easy to perceive than the possibility of a reduction of potential D&L in an uncertain and or known future. The clearer today's mitigation costs but not directly comparable to future damage decrease, the more difficult is to obtain consensus on expenditure (both public and private) in time of peace.

In this respect, CBA should be regarded as an effective tool as it is able at least to elicit relevant knowledge and information to support policy-makers in choosing among different alternative

measures and interventions to reduce damage and risk, even when some elements cannot be directly assessed in monetary terms. In this context, the question is how to recognize the "list" of costs and benefits suitable for describing the risk in a given area and how to evaluate and monetize them. The high variety of local conditions in terms of exposure and vulnerability, which influences the impacts of events on territorial subjects and objects (even when considering just floods) makes this task challenging.

3.2 *The costs side*

An effective synthesis focused on the comparison among the costs of different available mitigation measures assessed by way of different methodologies is offered in Hawley et al. (2012). Likewise, Bouwer et al. (2014) use a cost classification that distinguishes between direct and indirect costs, tangible and intangible.

Besides the direct costs of FMM, such as construction costs of physical infrastructure or evacuation costs, negative externalities or "co-costs" (Rose 2016) can be generated, e.g. interferences with the natural environment and landscape amenities. The externalities that are usually included in economic CBA are the costs or benefits resulting from a project that influences third parties without monetary compensation (European Commission 2014). Negative externalities, or external costs, are considered in social analyses because they impact on the social welfare, causing market failure when not included (Pesaro et al. 2016).

3.3 *The benefits side*

In CBA, benefits are evaluated as the "positive impacts obtained because of the reduced or avoided D&L" due to the different mitigation measures to be implemented. Flood damage assessments are then crucial to measure avoided damages, which may include primarily health problems and loss of lives, physical damage to buildings, infrastructures and lifelines, environmental assets, cultural heritage, and business interruption. Such assessment, along with the costs side mentioned in the above paragraph, can be developed both ex-ante and ex-post the flood occurrence.

Concerning avoided damage, economic subjects and businesses are still less considered than households or public infrastructures, even though economic activities are core elements for the functioning and development of a territory. The problem is the great variety characterizing the economic sector, production activities and assets which make them differently prone to damage and losses.

As for positive externalities or "co-benefits" of disaster mitigation investments, such as, among others, environmental conservation, improved social cohesion and increased agricultural productivity, they may materialize even in the absence of a disaster (Tanner et al. 2015). Discussions developed during the IDEA Project revealed that civil protection representatives, for instance, carefully consider as positive externalities the positive reputational effect resulting from the enhancement of prevention and mitigation measures.

It is finally important to highlight also the "system benefits" arising from stakeholders' perception of safety, which have to be better integrated into the "benefit scenarios". If safety were perceived as a strength/value of a place to live or to locate economic activities, a "safer" place might become more attractive. This, of course, would produce an increase in the demand for safer space and properties, which, in its turn, not only would generate a rise in the real estate values but also a more dynamic environment for households and businesses in such safer zones. However, over time, as a drawback, exposure to residual flood risk would increase as well. Therefore, also other system benefits should be acknowledged, like for example improved knowledge and information dissemination, leading to changes in households and economic subjects' behavior in case of flood. An increased awareness of flood risk would make communities better able to respond rapidly to any event, reducing damages, losses and psychologic distress.

4 THE DAM COST-BENEFIT ANALYSIS FOR FLOOD EVENTS IN THE UMBRIA REGION

4.1 *Case study background*

After a national directive on national and regional warning system for hydro-geologic risk for civil protection purposes was issued in Italy in 2004, the National Department of Civil Protection established a technical panel on this topic focusing on the Tiber river basin. As a first result of this activity, an "informal agreement" on the use of dams to retain water in case of heavy rains was signed among different authorities (Tiber river Authority, Dam Operators, etc.) in 2005. On the base of this, dams were emptied according to early warning and could reduce the river flow during flood events in Umbria in 2005, 2008, 2010, 2012 and 2013. Furthermore, the Umbria and Lazio Regions adopted a framework agreement and a flood management plan in June 2016.

4.2 *The "dam exercise": data collection, maps design, data processing and CBA assessment*

During both the 2012 and 2013 events, three dams (Montedoglio, Valfabbrica/Casanuova, and

Corbara) were used to retain water and consequently reduce the river flow. The approach proposed in this paper for the development of CBA is the economic one, as in Brouwer & van Ek (2004), while avoided damage has been derived from the real damage and costs as evaluated by the Civil Protection of the Umbria Region, that is by a public actor.

In order to understand the net benefits thanks to the application of the dams as non-structural measures 5 main steps were followed: (i) identification of the "avoided event" and its associated damages; (ii) identification of damages of the "occurred event"; (iii) calculation of "avoided damages", (iv) identification of costs; (v) identification of the net benefits.

As "avoided event", we mean the flooding that had not occurred thanks to the water retained by dam reservoirs. The "avoided event" was assessed by analysing the rainfall data and the water depth measured upstream of the dams from the reports done by the Civil Protection of Umbria Region, for the 2012 and 2013 flood events.

The damage that the "avoided event" could have provoked were estimated with the Flood-IMPAT procedure (see Molinari et al. 2016), which performs a damage assessment at the meso-scale level by depth-damage curves in a GIS environment. Flood-IMPAT allowed estimating direct damages to the residential, industrial/commercial and agricultural sectors, and consequently, only these typologies were considered for the analysis (Table 2a).

The damages due to the "occurred event" (Table 2b) were assessed by using the real flood damage values collected by the Civil Protection through the damage declarations filled in by households and economic subjects (see Menoni et al. 2016). The avoided damages due to the operation of the dams during the flood event result from the difference between the estimated damages of the "avoided event" and the damages of the "occurred event" (Table 2c).

The assessment of the costs associated with the use of the dams for flood lamination was based on the hypothesis that the dams of Montedoglio and Casanuova were used exclusively for energy generation. These costs (Table 3a) were estimated based on the loss of revenue due to the unsold energy as a consequence of the lack of whirl of the dams during the whole operation (considering also the days before the flood when the dams were emptied).

Finally, the net benefits associated with the flood lamination (Table 3b) for each event were calculated with the difference between avoided damages (benefits) and the loss of revenue (costs). As a further step, it has been computed the absolute net benefits in order to understand the relevance of the net benefits in relation to the damages of the avoided event. The absolute net benefit for the 2012 and 2013 flood events is respectively 32% and 73%,

Table 2a. Estimate of the total damage of the "avoided" events in 2012 and 2013.

	Damages without dam management (Flood-IMPAT) [M€]			
Event	All sectors	Residential	Agriculture	Industry/ Commercial
2012	55.5	5.5	19	30.5
2013	12.5	1.5	6.5	5

Table 2b. Occurred total damage per sector for the 2012 and 2013 events.

	Occurred Damages [M€]			
Event	All sectors	Residential	Agriculture	Industry/ Commercial
2012	36	2.5	12	21
2013	2	1.5	0	0.5

Table 2c. Total avoided damages due to the use of the dams in the 2012 and 2013 events.

	Avoided damages with the use of dams [M€]			
Event	All sectors	Residential	Agriculture	Industry/ Commercial
2012	19.5	3	7	9.5
2013	10.5	0	6.5	4.5

Table 3a. Costs beard by the energy operator (operators' calculations).

Event	Dam	Laminated volume [m^3]	Loss of revenue [€]
2012	Corbara	70 M	2
	Montedoglio	25 M	0.7
	Casanuova	20 M	0.5
2013	Montedoglio	25 M	0.7
	Casanuova	21 M	0.6

Table 3b. Net benefit from the CBA assessment.

Event	Benefits [M€]	Costs [M€]	Net benefits [M€]
2012	19.5	1.2	18.3
2013	10.5	1.3	9.2

which means that the use of the dams in both cases clearly allowed reducing damages, up to such a high performance as in the second case. Such a significant result is obtained because a risk mitigation and not just a damage mitigation measure has been applied, reducing the severity hazardous event at the origin. A damage reduction degree even higher if the indirect and systemic effects of the avoided direct losses had been evaluated.

4.3 *Lessons learnt*

The economic advantages obtainable by means of dam management as a non-structural risk mitigation measure proved to produce high savings in terms of avoided damage. The case study allowed to identify a variety of factors related to the use of the CBA methodology and to the uncertainties linked to the processing of technical and scientific elements.

Adopting the economic approach perspective, problems arise from the difficulties in assessing the cause-effect chains and externalities. This is mainly due to the presence of indirect and systemic damage, suffered by the territorial elements and subjects, from the one side, and by the private owners of the dam on the other side. This means a potential lack of damage information from both the benefits side, in terms of avoided damage, and in terms of costs, due to the losses produced because of the implementation of the measure. In this particular case, the production of electrical hydropower is concerned, which means that the accounting for costs depends on the energy costs, the tariffs profiles, the customers typology, the variable demand for power during different periods of the day and the related energy production functions. On the other hand, negative externalities could occur because of the interruption of energy production, depending on the overall power production sources, and the amount of the power demand covered by the dam production.

From a technical and scientific perspective, flood damage estimation by means of real damage assessment presents a set of shortcomings similar to those implied by fully modelled damage, depending on the availability of context-based functions and local specific characteristics that hamper the possibility to transfer results in space and time (see Merz et al. 2010).

In addition, the case study shows that there is a huge space for non-structural measures negotiated among territorial actors, both ex-ante and ex-post events, whose costs are difficult to be assessed but whose benefits might be important. Although not directly considered in the case study, the use of the dam was possible through the negotiation process and collaboration between private and public subjects. Therefore, negotiation costs should be better taken into consideration, together with the different negotiation models and related power degree. In the dam example, for instance, the private subjects, that is the private dam managers, have been obliged to accept the demand for intervention under the specific regulation mentioned above, the Umbria Region's *Piani di laminazione* (flood retaining plans). In principle, private subjects should be compensated for the costs undertaken for public interest purposes, however their right has to be balanced against the fact that they are beneficiaries of a license arrangements to use water for energy production (Pesaro 2007).

5 OPEN PROBLEMS AND FUTURE DEVELOPMENTS

The proposed FMM classification shows that, even if traditionally less utilized (WMO 2009), non-structural measures are more numerous and variate than structural ones. A discussion about Flood Mitigation Measures (FMM) is central to better understand the potentials of CBA and improve it. Furthermore, structural measures are usually less cost-effective (Kelman 2013). Non-structural measures should then be considered avoiding professional biases or limitation in the appraisal process (Mechler 2016). Furthermore, the use of CBA as proposed here has been based on the comparison between "near to real" costs and benefits, derived from damage evaluated by the Regional Civil Protection and not on theoretical monetization methods. The aim of the exercise was to prove the evidence of the potential effectiveness of non-structural measures to reduce damages, whose dimensions have been calculated based on the real damage data collected on the ground after the described event. Moreover, disaster risk reduction being the main goal, attention was driven toward costs in terms of interventions able to reduce the territorial impacts of floods and on non-structural measures as investments able to reduce implementation costs and related territorial and landscape impacts.

Following the proposed example of CBA applied to the use of dam reservoir to store part of the peak volume, further efforts are required to estimate the economic efficiency of this type of measures, as well as to identify and assess their related systems of costs.

Concerning future developments from the "benefits side", avoided damages should also consider the high potential weight of the indirect and systemic damage to economic activities. In fact, more attention should be devoted as well to show how non-structural measures entail larger positive systemic effects on economy, by taking into consideration local and over-local links and interactions.

Investments in FMM might produce a system of other positive externalities, which should be better taken into consideration in decision-making processes about the selection of the FMM. Such additional side effects might therefore maximize the effectiveness of the investments from a system's perspective, producing win-win effects. This, for instance, when, even in absence of flood events, territorial and natural environment management and conservation are considered and, where, therefore, consensus can be achieved more easily.

From the "costs side", the presence of often neglected negative externalities from flood mitigation interventions, mainly the structural ones, should be included in the analysis (monetary or not) because they may make other measures more desirable at the system level. A reason why, finally, there is the need for a more interdisciplinary perspective to better integrate the economic thinking when assessing the impacts of mitigation.

ACKNOWLEDGMENTS

This research was funded by the European Commission IDEA project G.A.N. ECHO/SUB/2014/694469 H2020 – Prevention and Preparedness in Civil Protection.

REFERENCES

Ale, B.J.M., Hartford, D.N.D. Slater, D. 2015. ALARP and CBA all in the same game, *Safety Science* 76: 90–100.

Botzen, W.W.J., Monteiro, E., Estrada F., Pesaro G., Menoni S. 2017. Economic Assessment of Mitigating Damage of Flood Events: Cost–Benefit Analysis of Flood-Proofing Commercial Buildings in Umbria, Italy. In *The Geneva Papers on Risk and Insurance – Issues and Practice* 42 (4): 585–608. doi:10.1057/s41288-017-0065-0.

Bouwer, L.M., Papyrakis, E., Poussin, J., Pfurtscheller, C., and Thieken, A.H. 2014. The Costing Of Measures for Natural Hazard Mitigation in Europe. *Natural Hazards Review* 15 (4): 04014010. doi:10.1061/(asce)nh.1527–6996.0000133.

Brouwer, R., van Ek, R. 2004. Integrated ecological, economic and social impact assessment of alternative flood control policies in the Netherlands, *Ecological Economics*, 50(1–2): 1–21.

BTRE. 2002. *Benefits of Flood Mitigation in Australia*, 177 pp, Bureau of Transport and Regional Economics, Commonwalth of Australia: Canberra.

Cochrane H.C. 2004a. Economic loss: myth and measurement, *Disaster Prevention and Management*, 13, 4:290–296.

Cochrane, H.C. 2004b. Indirect Losses from Natural Disasters: Measurement and Myth. In Okuyama Y. & Chang S.E. (eds) *Modeling the Spatial and Economic Effects of Disasters*, New York, Springer.

European Commission. 2014. *Guide to Cost-Benefit Analysis of Investment Projects*. Luxembourg: Publications Office of the European Union.

Hawley K., Moench M., Sabbag L. 2012. *Understanding the economics of flood risk reduction: a preliminary analysis*, Institute for Social and Environmental Transition-International, Boulder.

Jonkman, S.N., Brinkhuis-Jak, M. & Kok, Matthijs. 2004. Cost benefit analysis and flood damage mitigation in the Netherlands, *Heron*, 49 (1). 49.

Joseph, R., Proverbs, D., Lamond, J. & Wassell, P. 2014. Application of the concept of cost benefit analysis (CBA) to property level flood risk adaptation measures, *Structural Survey*, 32(2): 102–122.

Kelman, I. 2013. Disaster Mitigation is Cost Effective. World Bank, Washington, DC. World Bank. https://openknowledge.worldbank.org/handle/10986/16341 License: CC BY 3.0 IGO.

Klijn, F., Olfert, A., Schanze, J. 2009. Methodology for ex-post evaluation of measures and instruments in flood risk management (postEval) – Executive Summary, Leibniz Institute for Ecological and Regional Development (IOER), FLOOD*site* Report T12–07–01, Dresden.

Mechler, R. 2005. *Cost-benefit Analysis of Natural Disaster Risk Management in Developing Countries: Manual*, GTZ.

Mechler, R. 2016. Reviewing estimates of the economic efficiency of disaster risk management: opportunities and limitations of using risk-based cost–benefit analysis, *Nat Hazards*, 81(2016): 2121–2147.

Menoni S., Molinari, D., Ballio, F., Minucci, G., Mejri, O., Atun, F., Berni, N., Pandolfo, C. 2016. Reporting flood damages: A model for consistent, complete and multi-purpose scenarios, *Natural Hazards and Earth Systems Sciences*, 16: 2783–2797.

Merz, B., Kreibich, H., Schwarze, R. & Thieken, A. 2010. Review Article "Assessment of Economic Flood Damage", *Natural Hazards And Earth System Science* 10 (8): 1697–1724. doi:10.5194/nhess-10-1697-2010.

Meyers, V., Becker, N., Markantonis, V., Schwarze, R., van den Bergh, J.C.J.M., Bouwer, L.M., Bubeck, P., Ciavola, P., Genovese, E., Green, C., Hallegatte, S., Kreibich, H., Lequeux, Q., Logar, I., Papyrakis, E., Pfurtscheller, C., Poussin, J., Przyluski, V., Thieken, A.H., Viavattene, C. 2013. Review article: Assessing the costs of natural hazards – state of the art and knowledge gaps, *Natural Hazards and Earth System Sciences*, 13: 1351–1373.

Molinari D., Menoni S., Aronica G.T., Ballio F., Berni N., C. Pandolfo C., Stelluti M., Minucci G. 2014. Ex post damage assessment: an Italian experience, *Natural Hazards and Earth System Sciences*, 14: 901–916.

Molinari, D., Minucci, G., Mendoza, M.T., & Simonelli, T. 2016. Implementing the European "Floods Directive": The Case Of The Po River Basin, *Water Resources Management* 30 (5): 1739–1756. doi:10.1007/s11269-016-1248-3.

Pesaro G. 2005. La conservazione dei centri storici in zona sismica: un approccio economico. In S. Lagomarsino S., Ugolini P. (eds), *Rischio sismico, territorio e centri storici*, Milano, FrancoAngeli.

Pesaro G. 2007. Prevention and mitigation of the territorial impacts of natural hazards: the contribution of economic and public-private cooperation instruments.

In Aven T., Vinnem J.E. (eds.) *Risk, Reliability and Societal Safety – Vol.1 Specialisation Topics*, London, Taylor&Francis.

Pesaro G., Mendoza, M.T., Menoni, M., Minucci, G., Bezzam, V., Russo, F., Botzen, W., Monteiro, E., Estrada, F., Hudson, P. 2016. *Cost benefit analysis of mitigation measures to pilot firms/infrastructures in Italy*, IDEA Project, Deliverable D.4, http://www.ideaproject.polimi.it.

Rose A., Huyck, C.K. 2016. Improving Catastrophe Modelling for Business Interruption Insurance Needs, *Risk Analysis*, DOI: 10.1111/risa.12550.

Schanze J., Hutter, G., Penning-Rowsell, E., Nachtnebel, H-P, Meyer, V., Werritty, A., Harries, T., Holzmann, H., Jessel, B., Koeniger, P., Kuhlicke, C., Neuhold, C., Olfert, A., Parker, D., Schildt, A. 2008. Systematisation, evaluation and context conditions of structural and non-structural measures for flood risk reduction. FLOOD-ERA Joint Report, published by ERA-NET CRUE, http://www.crue-eranet.net.

Shreve C.M. & Kelman, I. 2014. Does mitigation save? Reviewing cost-benefit analyses of disaster risk reduction, *International Journal of Disaster Risk Reduction*, 10(2014): 213–235.

Tanner, T.M., Surminski, S., Wilkinson, E., Reid, R., Rentschler, J.E., & Rajput, S. 2015. The Triple Dividend of Resilience: Realising development goals through the multiple benefits of disaster risk management, Global Facility for Disaster Reduction and Recovery (GFDRR) at the World Bank and Overseas Development Institute (ODI), London. www.odi.org/tripledividend.

UNISDR (United Nations International Strategy for Disaster Reduction) 2017. Terminology on DRR, https://www.unisdr.org/we/inform/terminology.

WMO 2009. Flood Management in a Changing Climate, APFM Technical Document No. 14, *Flood Management Tools Series*.

Yevjevich, V. 1994. In Rossi, G. and Harmancıoğlu, N. (eds.) *Coping With Floods*: 573–576. Dordrecht: Kluwer Academic Publishers.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Behavioural modelling of attackers' choices

S. Panda, I. Oliver & S. Holtmanns
Cybersecurity Group, Nokia Bell Labs, Finland

ABSTRACT: This paper examines a cyber environment involving attackers and telecommunications operators from attackers' perspective. We incorporate a behavioural approach to understanding attackers' behaviour during the attack process. Traditionally, security games have been analysed assuming the attackers to be of strictly bounded rationality or strategy-less. Furthermore, studies consider attackers do aim to maximise their expected gain which contradicts the assumption of bounded rationality of attackers. We have analysed security interactions considering attackers as rational entities with attack strategies. To understand the thought process and behavioural decision-making choices of attackers, we utilise a decision analysis model capturing the attack process. Based on our analysis, we propose a framework providing a way to enhance attack strategies against cooperating and non-cooperating (competing) operators. This study is intended to capture essential characteristics of an attacker to comprehensively understand and predict their expected behaviour assisting cybersecurity.

1 INTRODUCTION

The major concerns in cybersecurity are to measure the security risks and to determine the effectiveness of one's security investments against perceived threats. As in cybersecurity, security is defined by not only on an individual's security-related investments but also by others' security investments (Anderson and Moore 2006, Laszka et al. 2015). This security interdependence adds additional complexities in quantifying the security risks and crafting appropriate measures against it.

Game theory, being a mathematical modelling tool, has been widely used to study varied aspects of security (Roy et al. 2010, Merrick et al. 2016, Altman et al. 2006, Liang and Xiao 2013) and privacy (Manshaei et al. 2013). Most of the work focused on studying defenders' behaviour and have proposed strategic recommendations which include stochastic approaches, frameworks, cognitive and behavioural models strengthening defenders' chances of successfully defending against attempted attacks. Studies have often assumed strategy-less behaviour of adversaries with a prescribed set of actions consistent with the threat models. However, alongside defenders, attackers are also intelligent entities and this assumption is not ideal in real-world situations which consists human adversaries (Camerer 2011).

In cybersecurity, attackers' behaviour has been less explored due to lack of reliable data on their intentions and interactions limiting our understanding of their characteristics and behaviours. (Veksler & Buchler 2016) and (Anderson 2009) have indicated that cognitive approaches can aid in predicting attackers' behaviour addressing real-world security problems.

In addition, over the past years, adversaries have become more financial oriented (Gordon 1994, Gordon 2000, Franklin et al. 2007) making them highly unpredictable. Some intentions behind these malicious activities are instigated by curiosity, or for peer recognition, and are often undecided in terms of ethical legitimacy (Gordon 1994, Gordon 2000). The possibilities of using illegal methods provoke new classes and strategies of attacks creating a need in studying and analysing attackers' behaviour to understand their intentions and decision making criteria.

We performed a game-theoretic investigation on attackers' strategies in the context of cybersecurity. The examined scenarios illustrate security games between attackers (cybercriminals, hackers) and telecommunications operators (defenders). An attacker is an external entity with malicious objectives attempting to break through the security of the targeted entity/system with an intention to hamper the existing state of the target.

A behavioural approach is utilised to anticipate decision-making behaviour of attackers. We intend to determine attack strategies optimising attackers' effort in performing an attack and improving their perceived utility. A viewpoint this paper aims to highlight is when attack strategies are taken into consideration, what can the choice of not attacking signify?

This study is a step towards understanding the mentality of attackers and their decision-making

behaviour from a cybersecurity perspective. Lack of decisive information on adversaries along with the available security information being highly asymmetric—favouring the attackers; results from this study can be used by defenders in assessing their conditions and perceiving the most expected attack strategies.

The rest of the paper is organised as follows. Section 2 covers the relevant literature and highlights the relationship of our work with existing research. Section 3 discusses the behavioural aspects of attackers and presents an attack framework disintegrating the efforts required in an attack process. An analysis of how attack strategies can be optimised using the attack framework is explained in Section 4. Section 5 discusses our findings and concludes this paper.

2 RELATED STUDIES

Even though this paper is confined towards understanding attackers, the complementarity of attackers' behaviour on operators' state is such that modelling them without an underlying operators' state is complex and unrealistic. An operator's state is defined by his security-related investments and relationships with other operators—cooperation (Laszka et al. 2015, Kunreuther and Heal 2003, Varian 2004, Hota and Sundaram 2015) and competition (Jiang et al. 2008, Sun et al. 2008, Khouzani et al. 2014, Panaousis et al. 2014), which induces additional security dependencies.

Attackers, alike operators (defenders), have strategic incentives and work towards maximising their expected utility (Laszka et al. 2015, Hausken 2006). The expected utility is a critical influencer in any decision-making process. For example, an operator invests in a particular security technology only after acknowledging that the investment will attain the expected returns (Hausken 2006). Similarly, the expected utility moderates attackers' strategic choices, especially the motivation (Hausken 2006, Herley 2010), behind attacks. The strategic choices of attackers are also influenced by available resources (Hausken 2006), the context of the interaction and the targeted operator's state which shapes the expected utility.

From an economic perspective, Herley (Herley 2010) pointed out that an attack strategy should be defined by the economics of attacks. He proposed attack strategies distinguishing attacks into scalable attacks and targeted attacks. In scalable attacks, the effort is independent of the number of users attacked. While in targeted attacks, the effort depends on per-user attacked suggesting targeted attacks must be on users with higher than average expected value. A profitable attack strategy involves accurately distinguishing viable from nonviable targets and deciding which viable target to attack based on the expected value (Herley 2012).

Grossklags et al. (Grossklags et al. 2008) analysed the Nash equilibria and social optima for different classes of attacks and defences in weakest-link, best-sort, and sum-of-effort security games. They introduced a weakest-target game "where the attacker will always be able to compromise the entity (or entities) with the lowest protection level but will leave other entities unharmed." Florencio et al. (Florêncio and Herley 2013) refined this criterion by incorporating the concept of free-riding (discussed by (Varian 2004)) to the lowest protected entity (or entities) stating that even though there exist economically profitable targets, many attacks are extremely difficult to turn into profitable ones grounding it to the economics of attacks.

From an extensive literature review, Hausken and Levitin (Hausken and Levitin 2012) categorised attack tactics on plausible types of attacks such as attacking a single element, attacks against multiple elements, consecutive attacks, random attacks, attacks involving a combination of intentional and unintentional impacts, attacks with incomplete information, and attacks with variable resources. However, a critical difficulty in modelling opponents in general, specifically in the security domain, is due to lack of decisive information regarding potential adversaries and attackers-defenders interactions being highly complex and extensive (Pita et al. 2012).

To understand the behavioural aspects of participants in cybersecurity, (Kusumastuti et al. 2015) and (Ryutov et al. 2015) have studied not only technical aspects but also psychosocial aspects through a three-player cybersecurity game. Kusumastuti et al. (Kusumastuti et al. 2015) used mini-max solution to identify game parameters and their influence on a player's behaviour. Ryutov et al. (Ryutov et al. 2015) aimed at understanding and modelling roles, motivations and conflicting objectives of players. In addition, (Anderson 2009), (Tambe et al. 2014) and (Veksler and Buchler 2016) have demonstrated improvement in the predictability of attackers' behaviour by using behavioural/cognitive modelling in a repeated security game environment.

To address adversary's bounded rationality, researchers have been pursuing alternative approaches. One approach includes robust optimisation techniques *avoiding adversary modelling* (Yang et al. 2011, Pita et al. 2012, Pita et al. 2010), while the other approach incorporates human decision-making models for computing *defend strategies* (Nguyen et al. 2013). Our work utilises the later approach and differs from the existing research by focusing on modelling adversaries

rather than defenders. Firstly, instead of strictly bounded rationality of attackers, we consider attackers with strategic incentives working towards maximising their expected utility. More precisely, we pose a decision model with an intention to understand the mentality of attackers and their decision-making behaviour. We also introduce a generalised attack framework distinguishing the effort required during an attack process. The attack framework is used to evaluate and refine attack strategies. In addition, this framework also facilitates a way of addressing the abstract states of the decision model, which we believe applies to a whole class of security scenarios.

3 BEHAVIOURAL ANALYSIS AND FRAMEWORK CHARACTERISATION

We define the attackers-operators interaction as a game-theoretic model that captures essential characteristics of strategic decision making. The essence of game theory is to study factors influencing behaviour by reasoning what players think other players will do. However, in reality, having complete and perfect information regarding your opponents is never feasible. This applies particularly in the context of security, where the threat is almost always unknown and effectiveness of security investments are very hard to quantify (Laszka et al. 2015). So, every interaction is considered to involve certain degrees of uncertainty in committing to decisions. For example, attackers might have knowledge regarding an operator's investment in security but have no decisive information related to the extent an operator has invested or on what kinds of secure technologies has the operator invested in.

Attackers, alike defenders, being a deficit in resources (Florêncio and Herley 2013) have to act strategically maximising their expected gain and optimising their investment of resources. Aiming this, we analyse the attackers assuming their end goal is to successfully attain the expected results while minimising their effort in the attack process. First, this assumption alleviates the strictly bounded rationality of attackers and facilitate them with diverse attack strategies in contrast to the only choices of attacking or not attacking in traditional game-theoretic modelling approaches. In addition, it supports analysing interaction environments under conditions when attackers do not react—ignore or watch the target; diverging from the traditional approach where attackers follow a prescribed path of invariably attacking.

Introduction of strategic attackers expands the possibilities where an action can bear latent objectives and motives raising concerns regarding the admissibility of proposed defence strategies. To take a concrete example, consider a case of distributed denial-of-service (DDoS) attack, where an attacker attempts to prevent an operator from delivering information or services. With a strategy-less attacker, the resultant action for the operator would be to invest resources in countering the attack with full capacity to minimise the damage. The attacker being strategy-less an attack would precisely be an attempt to harm the existing state of the operator. However, for a strategic attacker, the DDoS attack might merely be a probing attack to assess the strength of the operator or it can be a diversion ahead a powerful targeted attack.

Figure 1 presents a glimpse of an extended action set for strategic attackers. However, further characterisation of attacks based on the severity of attacks and dependencies between attacks are beyond the scope of this paper. Additionally, we consider investment in security as discrete (Grossklags et al. 2008, Kunreuther and Heal 2003, Lelarge and Bolot 2008), providing insulation towards all forms and degrees of attack, with no further distinctions in their capabilities to defend specific attacks.

Extracting the intent and decisions made to achieve the intended results from an instance of a highly restricted interaction scenario achieved through classical game theory is challenging. Indeed, it is the very problem in decisively predicting the behaviour of human players administring strictly bounded rationality (Camerer et al. 2004), especially while addressing human adversaries (Camerer 2011) where there is no evidence on generated forms of motivation and intention behind attacks. One approach to tackle this problem is by understanding the context of interaction, as

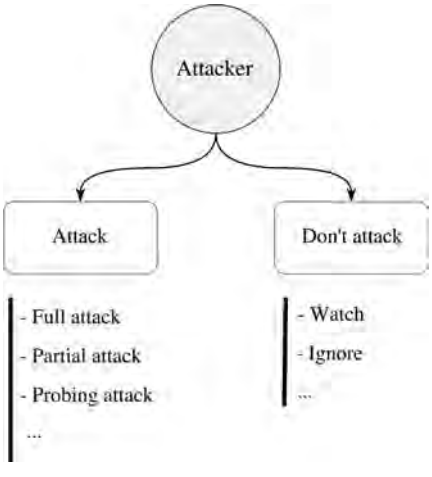


Figure 1. Attacker's decision space.

a decision must be made within a specific context and can be best represented through a hierarchy of decision states (Lewis 2013). Figure 2 is a hierarchical decision analysis tree capturing the mentality of attackers. The lowest level of the hierarchy represents concrete actions or choices of an attacker. As we ascend the hierarchy, states become increasing abstract and can be further fragmented into transitional stages precisely representing and supplementing an interaction scenario.

The hierarchical decision analysis tree is a highly simplified decision model representing cognitive workflow of attackers, initiating from the thought of an attack and terminating on a definitive decision on attacking or not attack, replicating an attack process. The thought of an attack is supported by deciding on whether to search for vulnerabilities or the type of attack to perform within the attacker's capabilities. Based on the context of interaction there could be numerous other decision paths to choose from for attackers. These intermediate choice of paths are latently, or innately, or precisely influenced by factors backing the intended goal. The subsequent steps down the hierarchy include designing attack strategies and then deciding whether to commit to an attack. The low-level decisions which demonstrate certain behaviour are being modelled using game theory. Game theory being a mathematical modelling tool supplements in determining and quantifying elements influencing decisions and assist in predicting behaviour (Burke 1999). The low-level behaviour can be used to infer higher-order objectives that are likely driving such behaviour augmenting our understanding of the intention behind attacks. An improved understanding of intention and motivation will support rigid estimations of attackers' behaviour. However, understanding the abstract states—top levels of the hierarchy demands a multi-disciplinary approach with effective application of concepts from behavioural psychology and cognitive science.

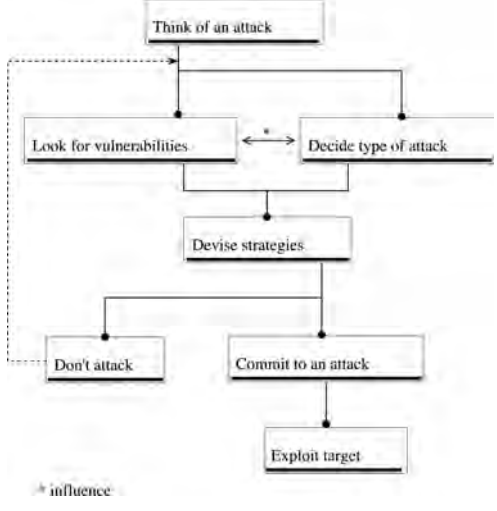


Figure 2. Attacker's decision analysis.

Figure 3. Attack framework.

We have characterised the attack process into different efforts required in performing an attack, acknowledging attackers to be rational entities. Figure 3, presents the attack framework demonstrating the efforts required in the attack process. The overall effort required can be broadly divided into *searching effort* and *breaking-in effort*. The searching effort includes efforts required in searching victims (targets), gathering information and searching vulnerabilities to exploit. Breaking-in effort represents the efforts required to compromise a system after choosing a target and a vulnerability to exploit. Based on the total effort required to compromise the target, an expected value can be derived. The expected value is one of the crucial factors moderating an attacker's decision (Herley 2010, Laszka et al. 2015).

The conversion of attackers' decision model into efforts required in the attack process disintegrates the top abstract levels of the decision model into modellable units. These modellable units can be used to quantify the expected utilities revealing the incentives behind attacks offering a better understanding of attackers' decision-making behaviour. Additionally, the attack framework assists in evaluating and enhancing attack strategies by effectively regulating the efforts strengthening the efficacy of attacks and ensuring better profits.

4 OPTIMISING ATTACK STRATEGIES

Lack of complete and perfect information against target induces uncertainty in attackers' decisions. The attack process eventually converges to a point of choice where an attacker has to decide on whether to attack or not to attack. The fate of an attempted attack depends on the target's capabilities to defend against attacks which further

depends on the extent of security investments. Figure 4, represents the expected payoffs of an attack against a targeted operator. We consider the investment in security as discrete—successfully preventing all forms and degrees of attacks. In the Figure 4, secure represents a system capable to successfully defend an attack and insecure represents the alternate.

In reality, defenders outnumber attackers. A critical problem attackers face is identifying targets such that a committed attack would yield something. For example, let's say the telecommunications domain has a total of N operators with N_c cooperating operators sharing security dependencies and N_n non-cooperating operators competing against security. It is an extremely expensive task for attackers to choose a viable target from such an intertwined mesh of operators. Here, the number of systems under each operator is ignored as considering it magnifies the complexity by many folds. Possible tactics attackers might adopt addressing this situation are

1. To randomly choose an operator and try breaching through the operator's defences. This approach would involve a heavy searching effort and a heavy breaking-in effort. This approach adds additional uncertainty as the attacker is unsure regarding his capabilities in successfully compromising the target.
2. To search for a certain type of vulnerability and then trying breaching it. This approach would involve heavy searching effort but a small breaking-in effort. Even though this approach involves high searching effort, chances of successfully compromising the chosen operator are very high.

The expected Utility (U) represents the probable payoff an attacker will receive on attacking the chosen operator. Based on the attack framework in Figure 3, the expected Utility for an attack can be determined as

$$U = cost\,(Information\_searching + Target\_searching + Vulnerability\_searching + Breaking\_in) - expected\ Value$$

where from (Herley 2010),

$$cost\,(Information\_searching) < cost\,(Target\_searching)$$

and any other forms of relationships cannot be defined from the existing literature.

Gathering and sharing of security-related information is a key factor heightening cybersecurity in both cooperating (Hausken 2017) and non-cooperating (Khouzani et al. 2014) environments. However, it is a known fact that the proposed information by defenders supports attackers in strategic decision-making. The following analysis illustrates how commonly available knowledge on operators can be used to reduce the cost of an attack. The use of available information reduces the information-searching effort to a static cost, represented as C_i, rather than a variable cost. In addition, attackers must bear the vulnerability-searching costs, represented as C_v, as a common cost irrelevant to any choice of target. t represents the choice of a target from the set of operators.

In a cooperating environment, the state of an operator is not only influenced by his decision but also by other cooperating operators' decisions. An attacker knowing that a set of operators (N_c) are cooperating refines the target-searching scope from N operators to N_c operators, where, $N_c < N$, reducing the effort to an extent. The expected Utility (U_c) for attacking cooperating operators can be defined as

$$U_c = C_i + cost\,(Breaking\_in) + C_v \sum cost\,(Target\_searching)\, N_{c(t,-t)} - expected\ Value$$

Whereas, in a non-cooperating environment, an operator's security investment might encourage competing operators to invest in better security measures. On the other hand, it might also increase the likelihood of attacks on competing operators as the attacker will prefer a victim will lower resistance. Knowing operators are competing reduces the victim-searching effort considerably, as it is economically beneficial to attack the losing operator. Reduced victim-searching effort can facilitate in allocating additional resources for vulnerability-searching and for breaking into the operator's defences. The expected Utility (U_n) for attacking competing operators can be defined as

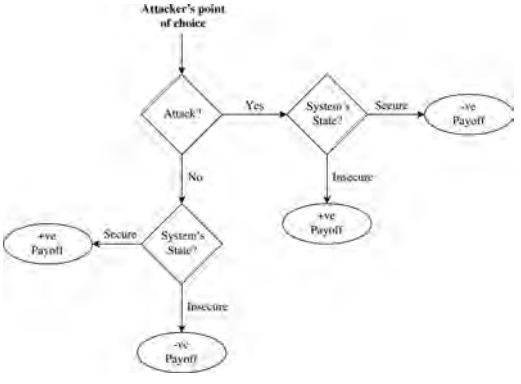


Figure 4. Attacker's expected payoff.

$$U_n = C_i + cost\,(Breaking\_in) + C_v \sum cost\,(Target\_searching)\, N_{n(t,-t)} - expected\ Value$$

Desired Gain represents the amount of gain the attacking wants from an attack. From an economic perspective, an attacker would prefer the attack that maximises his desired Gain. That is, from the available range of attacks which would successfully compromise the found vulnerability, he chooses the attack which max(U – expected Gain). This indicates co-existence of several classes of attacks on a point of attack. The expected payoff and the desired gain from an attack would moderate the decisions of the attacker. As

$$\text{decision} = \begin{cases} Attack, & \text{if } U \geq \text{expected Gain} \\ Do\ not\ attack, & \text{if } U < \text{expected Gain} \end{cases}$$

5 CONCLUSION AND FUTURE WORK

We investigated cybersecurity environment from attackers' perspective. Our results show that taking into consideration and admitting that attackers have strategies, incentives etc, implies that defenders (telecoms operators in our studies) need to change how they perceive, defend and react to attackers. The implications given the rise of targeted/coordinated attacks versus uncoordinated attacks (eg: DDoS) mean that operators must significantly reassess their investment in security technologies towards the former, despite the latter having better 'security theater'[1].

Traditionally, security games have been analysed **assuming the attackers to be of bounded rationality with limited set of prescribed choices**. Furthermore, studies consider attackers do aim to maximise their expected gain and this consideration contradicts the assumption of bounded rationality of the attacker. We study the attackers considering they **share similar characteristics as defenders** with attack strategies maximising their expected gain.

In particular, we model security interactions with an extended set of actions available to the attackers. This expands the possibilities where an action can bear latent objectives and motives raising concerns regarding the admissibility of proposed defence strategies. We present a hierarchical decision tree capturing the mental model of attackers during the attack process. The decision model is supported by a generalised framework representing the attack process in terms of efforts required by the attackers addressing the abstract levels of the decision model. Using this framework attack strategies against cooperating and competing operators are derived optimising attackers' effort resulting in a better gain. Furthermore, it facilitates a way of understanding the strategic decision-making abilities of attackers.

Not all attacks are intended towards achieving economic targets. A novice attacker might not aim to maximise his economic payoff rather aim in gaining experience, or reputation and the interaction might end on an attempted attack. However, it might be a completely different picture for an experienced attacker. When such personality traits of the attackers are considered, specifically the strategic option of not attacking, it unsettles the traditional security modelling approach, particularly the Stackelberg approach (Kar et al. 2017), where the game proceeds with the assumption that the attacker acts (invariably attacks). This raises a number of research questions challenging the traditional approach used in modelling cybersecurity. For example

- Is every interaction between an attacker and defender a repetitive process or is it a single-point interaction which ends on an attempted attack?
- Is using Stackelberg Security Games to model security interactions an appropriate choice?

Considering the economics of scalable and targeted attacks discussed by (Herley 2010), would it be an effective strategy to launch a small scalable attack to determine the strength of the target and then launch a specific attack incapacitating the target?

In (Kusumastuti et al. 2015), attackers are facilitated with an option to not attack and invest in enhancing their capabilities enabling in launching stronger attacks in the future. This consideration would reduce the breaking-in effort and vulnerability searching effort. From a psychological perspective, Rogers (Rogers 2000) classified hackers depending on their expertise (from novice to experienced), areas of interests (software, hardware, etc.) and behavioural patterns.

Modelling players to be able to predict expected behaviour in a more realistic way requires a profound understanding of their incentives, motives and the context of the interaction. Behavioural modelling of the attacker will not only assist in understanding the expected intentions and behaviour of attackers but will also assist in devising comprehensive defences against such characteristics of attacks.

However, each instance of an interaction is unique, with a unique set of parameters characterising and moderating it. Modelling these unique

1. Bruce Schneier—Beyond Security Theater: https://www.schneier.com/essays/archives/2009/11/beyond_security_thea.html.

interactions under common grounds is highly ineffective. They demand to be modelled based on the context of the interaction and using only game-theoretic concepts restricts the context and the interaction environment to a larger extent through biases, heuristics, and convenience.

This preliminary exploration will guide our future studies in aptly modelling behavioural aspects of attackers and in refining the attack strategies and characteristics of attackers by incorporating proposed concepts from the existing research work. This would further aid in comprehensively modelling the behavioural aspects of the participants in the context of information-cyber security.

ACKNOWLEDGEMENTS

The work was made in conjunction with the EU SCOTT and SECREDAS Projects and has received funding from the Electronic Component Systems for European Leadership Joint Undertaking under grant agreement No 737422.

REFERENCES

Altman, E., T. Boulogne, R. El-Azouzi, T. Jiménez, & L. Wynter (2006). A survey on networking games in telecommunications. *Computers & Operations Research 33*(2), 286–311.

Anderson, J.R. (2009). *How can the human mind occur in the physical universe?* Oxford University Press.

Anderson, R. & T. Moore (2006). The economics of information security. *Science 314*(5799), 610–613.

Burke, D.A. (1999). Towards a game theory model of information warfare. Technical report, AIR FORCE INST OF TECH WRIGHT-PATTERSONAFB OH.

Camerer, C.F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C.F., T.-H. Ho, & J.-K. Chong (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics 119*(3), 861–898.

Florêncio, D. & C. Herley (2013). Where do all the attacks go? In *Economics of information security and privacy III*, pp. 13–33. Springer.

Franklin, J., A. Perrig, V. Paxson, & S. Savage (2007). An inquiry into the nature and causes of the wealth of internet miscreants. In *ACM conference on Computer and communications security*, pp. 375–388.

Gordon, S. (1994). The generic virus writer. In *Proc. Intl. Virus Bulletin Conf*, pp. 121–138.

Gordon, S. (2000). Virus writers: The end of the innocence? In *10th Annual Virus Bulletin Conference (VB2000), Orlando, FL*.

Grossklags, J., N. Christin, & J. Chuang (2008). Secure or insure?: a game-theoretic analysis of information security games. In *Proceedings of the 17th international conference on World Wide Web*, pp. 209–218. ACM.

Hausken, K. (2006). Income, interdependence, and substitution effects affecting incentives for security investment. *Journal of Accounting and Public Policy 25*(6), 629–665.

Hausken, K. (2017). Security investment, hacking, and information sharing between firms and between hackers. *Games 8*(2), 23.

Hausken, K. & G. Levitin (2012). Review of systems defense and attack models. *International Journal of Performability Engineering 8*(4), 355–366.

Herley, C. (2010). The plight of the targeted attacker in a world of scale. In *WEIS*.

Herley, C. (2012). Why do nigerian scammers say they are from nigeria? In *WEIS*.

Hota, A.R. & S. Sundaram (2015). Interdependent security games under behavioral probability weighting. In *International Conference on Decision and Game Theory for Security*, pp. 150–169. Springer.

Jiang, L., V. Anantharam, & J. Walrand (2008). Efficiency of selfish investments in network security. In *Proceedings of the 3rd international workshop on Economics of networked systems*, pp. 31–36. ACM.

Kar, D., T.H. Nguyen, F. Fang, M. Brown, A. Sinha, M. Tambe, & A.X. Jiang (2017). Trends and applications in stackelberg security games. *Handbook of Dynamic Game Theory*, 1–47.

Khouzani, M., V. Pham, & C. Cid (2014). Strategic discovery and sharing of vulnerabilities in competitive environments. In *International Conference on Decision and Game Theory for Security*, pp. 59–78. Springer.

Kunreuther, H. & G. Heal (2003). Interdependent security. *Journal of risk and uncertainty 26*(2–3), 231–249.

Kusumastuti, S., J. Cui, A. Tambe, & R.S. John (2015). A behavioral game modeling cyber attackers, defenders, and users. Research paper presented at the AAAI Spring Symposium, Stanford University, Palo Alto.

Laszka, A., M. Felegyhazi, & L. Buttyan (2015). A survey of interdependent information security games. *ACM Computing Surveys (CSUR) 47*(2), 23.

Lelarge, M. & J. Bolot (2008). A local mean field analysis of security investments in networks. In *Proceedings of the 3rd international workshop on Economics of networked systems*, pp. 25–30. ACM.

Lewis, M.J. (2013). Hierarchical decision making. In *STIDS*, pp. 162–165.

Liang, X. & Y. Xiao (2013). Game theory for network security. *IEEE Communications Surveys & Tutorials 15*(1), 472–486.

Manshaei, M.H., Q. Zhu, T. Alpcan, T. Bacs¸ar, & J.-P. Hubaux (2013). Game theory meets network security and privacy. *ACM Computing Surveys (CSUR) 45*(3), 25.

Merrick, K., M. Hardhienata, K. Shafi, & J. Hu (2016). A survey of game theoretic approaches to modelling decision-making in information warfare scenarios. *Future Internet 8*(3), 34.

Nguyen, T.H., R. Yang, A. Azaria, S. Kraus, & M. Tambe (2013). Analyzing the effectiveness of adversary modeling in security games. In *AAAI*.

Panaousis, E., A. Fielder, P. Malacaria, C. Hankin, & F. Smeraldi (2014). Cybersecurity games and investments: a decision support approach. In *International Conference on Decision and Game Theory for Security*, pp. 266–286. Springer.

Pita, J., M. Jain, M. Tambe, F. Ordóñez, & S. Kraus (2010). Robust solutions to stackelberg games: Addressing bounded rationality and limited observations in human cognition. *Artificial Intelligence 174*(15), 1142–1171.

Pita, J., R. John, R. Maheswaran, M. Tambe, & S. Kraus (2012). A robust approach to addressing human adversaries in security games. In *Proceedings of the 20th European Conference on Artificial Intelligence*, pp. 660–665. IOS Press.

Rogers, M. (2000). A new hacker taxonomy. *University of Manitoba*.

Roy, S., C. Ellis, S. Shiva, D. Dasgupta, V. Shandilya, & Q. Wu (2010). A survey of game theory as applied to network security. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pp. 1–10. IEEE.

Ryutov, T., M. Orosz, J. Blythe, & D. von Winterfeldt (2015). A game theoretic framework for modeling adversarial cyber security game among attackers, defenders, and users. In *International Workshop on Security and Trust Management*, pp. 274–282. Springer.

Sun, W., X. Kong, D. He, & X. You (2008). Information security investment game with penalty parameter. In *Innovative Computing Information and Control, 2008. ICICIC'08. 3rd International Conference on*, pp. 559–559. IEEE.

Tambe, M., A.X. Jiang, B. An, & M. Jain (2014). Computational game theory for security: Progress and challenges. In *AAAI spring symposium on applied computational game theory*.

Varian, H. (2004). System reliability and free riding. In *Economics of information security*, pp. 1–15. Springer.

Veksler, V.D. & N. Buchler (2016). Know your enemy: Applying cognitive modeling in security domain. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, pp. 2405–2410.

Yang, R., C. Kiekintveld, F. Ordonez, M. Tambe, & R. John (2011). Improving resource allocation strategy against human adversaries in security games. In *IJCAI Proceedings—International Joint Conference on Artificial Intelligence*, Volume 22, pp. 458.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Asset replacement decisions: A Markowitz efficient frontier approach to evaluate the trade-off between total costs and system availability

A.M. Teodoro-Filho, G.A. da Costa-Lima & L.A.N. Costa
Aremas, Campinas, Brazil

F.C. Marinho & A. Prestes
Brookfield Renewable, Rio de Janeiro, Brazil

ABSTRACT: In this work, an approach to estimate the optimal strategy to replace groups of assets is discussed. The model includes parameters such as cost of economic depreciation, cost of decommissioning and probabilistic distributions to represent random variables (for instance, time between failures) and uncertainty in cost of failures. This simulation model helps to understand the trade-off relationship between cost and availability according to the possible replacement strategies available. We present and discuss how Markowitz effective frontier can be created based on the simulated values for different replacement strategies. This work can fill a gap in the literature concerning the problem of asset group replacement, which are not well explored, but is important for decision-makers dealing with real world problems. The approach presented also helps the asset replacement strategy, which is part of the operational strategy, to be more flexible to support the high level business strategy.

1 INTRODUCTION

One of the main tasks of a plant and maintenance engineer is related to the replacement of existing assets, which are especially important in industries such as mining, petroleum, power generation, etc. The problem faced by companies, though, is not only related to finding the economic life of each asset individually, but it is also related to assessing the impact of different asset replacement strategy in the system performance by techniques such as RAM (Reliability Availability and Maintainability) modeling. In this paper, we employ the Reliability Block Diagram (RBD) and Monte Carlo simulation to evaluate the impacts of asset replacement strategies in system performance variables such as total cost—which include OPEX (Operational Expenditure) and CAPEX (Capital Expenditures) – and downtime. These are conflicting objectives. In this context, the main objective is to understand the trade-off between cost and availability according to the possible replacement strategies available. A numerical example of a generation unit system of a power generation company in Brazil is used to define the Markowitz effective frontier.

In order to fulfill the goals of the paper, the modeling technique is Reliability Block Diagram (RBD) employed together with the tool Monte Carlo simulation as discussed in section 2. Section 3 contains the results (simulated and calculated). Section 4 presents finals comments.

2 ASSET REPLACEMENT

Authors like Eilton et al. (1966) Bazargan & Hartman (2012) and Al-Chalabi et al. (2015) use analytical models to find the asset replacement age in order to minimize the Total Cost of Ownership (TCO). Also, others like Shi & Min (2014) and Adkins & Paxson (2017) use Brownian geometric motion to simulate the costs of an asset to be applied to asset replacement management models.

The weakness of this approach to minimize the TOC is that its focus is on asset level and does not take into account the importance of performance index like availability and reliability at system level, which is the real problem. In this matter, Leung & Tanchoco (1990) mention that if multiple assets are employed in an integrated system, then the replacement decision for each equipment of the system must be analyzed simultaneously.

In spite of the idea of this integrated analysis has been a not so recent concernment, there is still a lack models studying real problems in the asset management literature.

As discussed, the isolated analysis sometimes is not enough to give the decision maker the correct insight about the reality, because it does not consider the complex configuration of a production system found in real world. In this isolated analysis, it does not matter if the equipment is in series, active parallel or in standby in the system by the reliability point of view. This is the reason

why RBD was employed in the current paper. Simulation is employed to allow more flexibility in this problem with conflicting objectives. The final model may contribute to feel part of the gap of models to solve real replacement problems in asset management.

3 MARKOWITZ EFFICIENT FRONTIER

There are N strategies in asset replacement with different and conflicting relation between availability and cost. Then, there is efficient frontier for management decision making.

In financial world, the objective of a portfolio of stocks is (a) maximize the average return and (b) minimize risk. There is a conflict between these two variables—the higher the return, the higher the risk. As explained by Powel & Baker (2009), both objectives cannot be met simultaneously, but optimization techniques can be used to exploit the trade-offs.

For financial analysis, Markowitz (1952; 1959) developed an approach to examine the trade-off between risk and return of a portfolio where the random variable is rate of return over time. The optimization consists in the selection of stocks in order to give the highest expected return (measure by average) for a given risk (measured by the variance of return).

In the present work of replacement strategy, the attention is not in the trade-off between risk and return, but in the one between cost and availability for a system composed of a number of assets that suffer degradation over time with use.

The managerial flexibility is the replacement decision of each individual equipment. Each managerial flexibility give different cost and availability at system level. Then, by simulation a number of managerial alternatives it is possible to construct a Markowitz efficient frontier with solutions that are not overcome by others. From this information, managers will be able to choose de replacement strategy according to the companies risk profile.

4 MODELING

This paper considers the Markowitz efficient frontier of a Power Generation Unit. Its Reliability Block Diagram is in Figure 1.

In Figure 1, equipment is in series, that is, each one asset can turn the system unavailable. This type of system demands less initial capital investment, but has more risk of failure.

In discussing modeling by reliability block diagrams, Zhang et al. (2012) mentions that there are two approaches: (1) Top-down when the stochastic behavior modeling of the failure time of an system does not consider its components individually and (2) Down-top, when the simulation of the moment of occurrence of failure of an system considers the premise that the statistical modeling of the time until the failure of the individual components of this equipment is known.

In an RBD analysis, the choice of the Top-down or Down-top approach is influenced by the availability of historical data. Thus, for some equipment the amount of data available allows for a more detailed representation of the time until the individual components fail, but for other cases this is not possible.

Model presented in Figure 1 follows the approach Top-down. Noteworthy that the RAM analysis based on the flexibility of Monte Carlo simulation gives flexibility enough to become the model as complex as the judgment of the analyst requires. Table 1 summarizes the code, description and current age of each asset.

According to Table 1, Auxiliary Service Transformer, for example, is indicated by the asset code AST and its current age is 10 years of operation (87.600 hours).

The current age of each asset affects directly its failure rate and consequentially its availability and maintenance cost. Table 2 shows the probability distribution of time-to-failure of assets of Power Generation System.

As shown in Table 2, for example, the lifetime of Auxiliary Service Transformer is modeled using Weibull distribution with scale parameter (eta) equal to 56.940 hours and shape parameter (beta) equal to 1,8. It is interesting to note that all assets

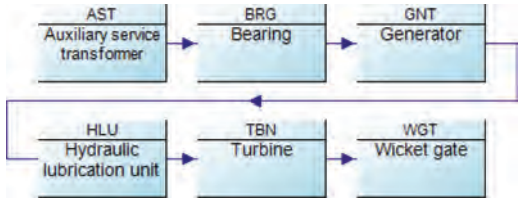


Figure 1. Reliability block generation unit system.

Table 1. Assets characterization for RDB modeling.

Asset		
Code	Description	Current age (hours)
AST	Auxiliary service transformer	87,600
BRG	Bearing	87,600
GNT	Generator	131,400
HLU	Hydraulic lubrication unit	87,600
TBN	Turbine	131,400
WGT	Wicket gate	43,800

Table 2. Probabilistic modeling of time-to-failure of each asset.

Asset code	Weibull distribution parameters	
	Eta (hours)	Beta
AST	56,940	1,8
BRG	30,660	4,2
GNT	61,320	3,8
HLU	78,840	3,2
TBN	70,080	3,6
WGT	43,800	1,4

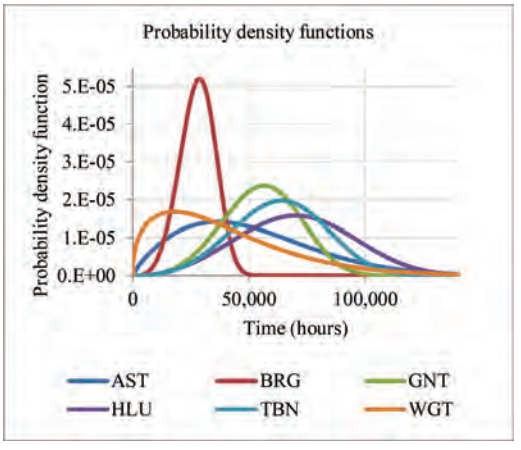


Figure 2. Probability density functions.

have an increasing failure rate, giving rise to expect find degradation as the main failure rate.

The probability density functions based on the parametes of Table 2 can be observed in Figure 2.

Each curve in Figure 2 represent the probability density function of the first failure of each asset.

After the failure of each asset, management applies corrective maintenance. An important assumption is that after maintenance, the probability density function of second failure due is not the same of the first one. Table 3 contains the characteristic of the corrective maintenance task of each asset.

Considering again the Auxiliary Service Transformer (ASF), the task duration of the corrective maintenance is 120 hours, the cost of this task is $ 66,000 and the age reduction factor is equal to 5%. To understand the concept of age reduction, consider that after a maintenance task, the asset can return to a condition as good as new, as good as old or intermediate. To quantify this intermediate condition the following model can be used (Malik, 1979):

$$I_d = I_a * (1 - ARF) \quad (1)$$

where I_d is the age of the equipment after maintenance, I_a is the age before maintenance and ARF is the age reduction factor. The as good as new condition is represented by ARF = 100% and the as good as old condition is represented by ARF = 0%. The condition after the maintenance depends on the complexity of the asset, among others things.

The modeling of the asset failure rate per hour appears in Figure 3 according to just one simulation.

In the simulation presented in Figure 3, one failure occurred after 67,991.3 operations hours, then a corrective maintenance was performed and the failure rate was affected by the age reduction factor. Other failures occurred after 82,065.8 and 93,971.2 hours and again it is possible to see the impact on failure rate. The asset replacement was performed in 96,360 hours, when the failure rate returned to as-good-as-new condition. The last simulation of failure occurrence was around 158,454.5 hours.

The problem of finding the right ARF is complex and requires additional data for modeling. In this context, the present work does not have make inferences about the unknown parameters of an

Table 3. Maintenance corrective characteristics.

Asset code	Corrective maintenance		
	Task duration (hours)	Cost ($)	Age reduction factor
AST	120	66,000	5%
BRG	72	21,000	85%
GNT	144	72,000	5%
HLU	96	60,000	5%
TBN	168	114,000	5%
WGT	120	90,000	5%

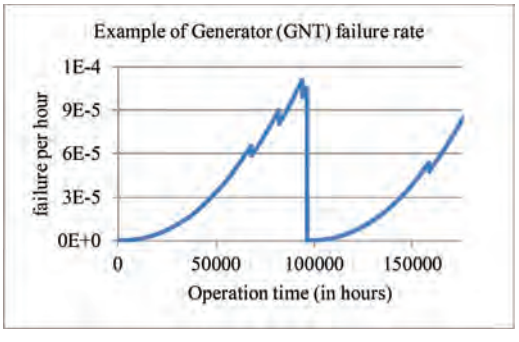


Figure 3. Profile of failure rate per hour using simulation.

Table 4. Task and cost used in asset replacement.

Asset code	Asset replacement	
	Task duration (hours)	Cost ($)
AST	24	300,000
BRG	48	90,000
GNT	240	750,000
HLU	120	270,000
TBN	480	630,000
WGT	240	255,000

imperfect repair model as in Melchor-Hernández et al. (2014) and Toledo et al. (2015).

The activity of asset replacement implies in a replacement duration (downtime) and Capital Expenditure (CAPEX). In Table 4 there are some data of tasks duration and cost of equipment of Figure 1.

In Table 4, the replacement of the Auxiliary Service Transformer has task duration equal to 24 hours and investment cost equal to $ 300.000. Data for other assets are also in Table 4.

5 RESULTS

Based on parameters discussed in section 2, excluding current age and asset replacement characteristic, Figure 4 contains the maintenance cost profile simulated of each asset for a period of 20 years.

Each profile is the result of the simulated mean maintenance cost after 10,000 simulations using the software Isograph Availability Workbench (AvSim®) considering that each equipment is new in year zero.

From Figure 4, it is easy to notice that different assets have different corrective maintenance cost profile over time. For example, considering only the year 20, the Turbine (TBN) has the highest simulated mean maintenance cost. Then, it is expected that the optimal replacement strategy is not equal for all assets.

The model of economic life of an asset based on cost minimization depends strongly on maintenance cost profile and other economic variables. In this case, consider that the opportunity cost of capital equal to 12% per year, fiscal depreciation equal to 10% per year, income tax rate equal to 25% a year and economic depreciation of asset value of 10% per year.

Figure 5 shows the relationship between Total Cost Ownership and time of replacement from Costa Lima & Teodoro- Filho2013) for different assets.

Figure 5 highlights that each asset has its own economic life. Turbine (TBN) and Generator

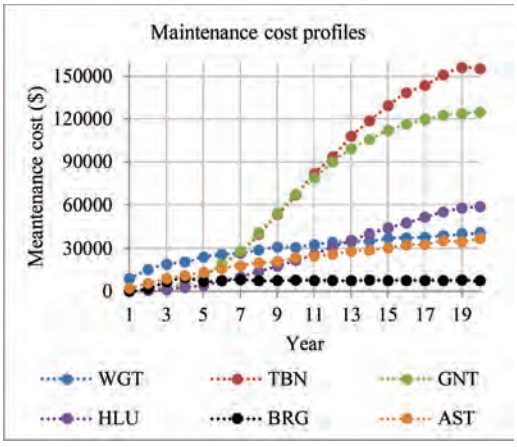


Figure 4. Simulated mean cost of maintenance of different asset.

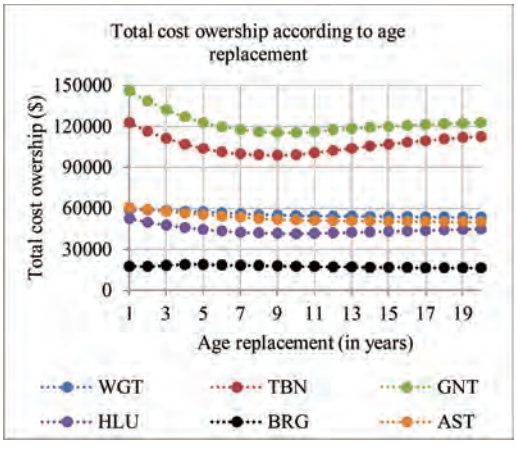


Figure 5. Total cost ownership according to age replacement.

(GNT), for example, have economic life of 9 and 10 years, respectively. By other hand, for the Auxiliary Service Transformer (AST), its economic life is at least for a period of 20 years.

In reality, the problem faced by the companies is related not only about finding the economic life of each asset, but about assessing the impact of different sets of asset replacement (strategies) in system performance. In order to illustrate the question, consider the model discussed in section 2 simulated for 3 years. In case of no replacement at the beginning of the period, the simulated present value of mean total cost is equal to $ 1,389,415 (considering opportunity cost of capital equal to 12% per year) and mean system downtime equal to 2,749 hours as shown in Table 5.

Table 5. Results of simulation present values of cost and downtime—Part 1.

Replacement decision	Mean total OPEX ($)*	Mean total cost ($)*	Mean system downtime
Replace all	61,874	1,692,042	611.00
No replacement	1,389,415	1,389,415	2,749
AST	1,330,456	1,525,853	2,658
BRG	1,354,562	1,413,181	2,688
GNT	834,091	1,429,673	1,775
HLU	1,304,659	1,480,516	2,723
TBN	823,115	1,323,404	2,319
WGT	1,352,345	1,456,770	2,935

*Present value.

In Table 5, replacing all assets at the beginning of the period results in a mean total cost equal to $ 1,692,042 and mean system downtime equal to 2,749 hours. Replacing only the Auxiliary Service Transformer (AST) at the beginning of the period results in a mean total cost equal to $ 1,525,835 and mean system downtime equal to 2,658 hours.

In short, Table 5 presents the results of the following replacement strategies: (a) replacing all assets at the beginning of operation, (b) no replacement at beginning and (c) replacement of only a specific asset and the others not. The mean total cost in present value is calculated based on the mean OPEX (represented by the corrective maintenance cost) plus the investment cost in new assets less the resale value of the assets replaced. Both result is generated based on the simulation. In this model the resale value is calculated considering an economic depreciation of 10% in relation to the last year.

In contrast, in Table 6 contains all 15 combinations of two assets replacement.

Considering the replacements of two assets, to replace GNT and TBN results in the lowest mean system downtime. To replace BRG and TBN results in the lowest mean total cost in present value.

In Table 7 the results of all possible 3 assets replacement are presented.

To replace GNT, HLU and TBN results in mean system downtime equal to 900 hours, which is even lower the replacing only GNT and TBN. None of 3 asset replacement strategy results in lower mean total cost than replacing only BRG and TBN or replacing only TBN.

All 4 assets replacement simulated performance is presented in Table 8.

Replacing AST, GNT, HLU and TBN results in mean total downtime is 780 hours, but has a mean total cost in present value equal to $ 1,595,473. Again, no option is the best in both criteria simultaneously.

Table 6. Results of simulation of present values of cost and downtime—Part 2.

Replacement decision	Mean total OPEX ($)*	Mean total cost ($)*	Mean system downtime
AST; HLU	1,245,466	1,616,719	2,609
AST; WGT	1,297,708	1,597,529	2,832
BRG; AST	1,297,233	1,551,248	2,577
BRG; GNT	805,897	1,460,098	1,671
BRG; HLU	1,273,306	1,507,782	2,620
BRG; TBN	793,179	1,352,087	2,214
BRG; WGT	1,316,862	1,479,906	2,823
GNT; AST	774,217	1,565,195	1,655
GNT; TBN	268,276	1,364,146	1,058
GNT; WGT	797,726	1,497,733	1,722
HLU; GNT	749,450	1,520,888	1,622
HLU; TBN	739,361	1,390,800	2,168
TBN; AST	768,240	1,463,925	2,210
TBN; WGT	791,379	1,396,092	2,273
WGT; HLU	1,274,912	1,555,194	2,801

*Present value.

Table 7. Simulated performance—Part 3.

Replacement decision	Mean total OPEX ($)*	Mean total cost ($)*	Mean system downtime
AST; BRG; GNT	748,188	1,597,785	1,553
AST; BRG; HLU	1,216,724	1,646,597	2,510
AST; BRG; TBN	735,067	1,489,371	2,098
AST; BRG; WGT	1,264,536	1,622,976	2,723
AST; GNT; HLU	692,638	1,659,473	1,505
AST; GNT; TBN	210,166	1,501,433	935
AST; GNT; WGT	748,188	1,643,591	1,553
AST; HLU; TBN	683,598	1,555,140	2,055
AST; HLU; WGT	1,217,433	1,693,111	2,692
AST; TBN; WGT	730,615	1,530,725	2,152
BRG;GNT; HLU	717,374	1,547,431	1,509
BRG; GNT; TBN	234,921	1,389,410	940
BRG; GNT; WGT	765,303	1,523,929	1,609
BRG; HLU; TBN	707,366	1,442,130	2,059
BRG; HLU; WGT	1,237,202	1,576,103	2,687
BRG; TBN; WGT	759,511	1,422,843	2,166
GNT; HLU; TBN	184,808	1,456,535	900
GNT; HLU; WGT	716,946	1,592,809	1,574
GNT; TBN; WGT	233,956	1,434,252	1,005
HLU; TBN; WGT	709,079	1,489,649	2,126

*Present value.

In Table 9 the results of all possible 5 assets replacement is presented.

With the results of Table 9, all possible asset replacement strategy to the beginning of the period was mapped. It's possible to identify that some strategies benefits one criteria and others

Table 8. Simulated performance—Part 4.

Replacement decision	Mean total cost in present value ($)	Mean system downtime
AST; BRG; GNT; HLU	1,685,685	1,632
AST; BRG; GNT; TBN	1,526,789	818
AST; BRG; GNT; WGT	1,667,167	1,501
AST; BRG; HLU; TBN	1,579,957	1,944
AST; BRG; HLU; WGT	1,718,868	2,583
AST; BRG; TBN; WGT	1,558,509	2,044
AST; GNT; HLU; TBN	1,595,473	780
AST; GNT; HLU; WGT	1,729,547	1,454
AST; GNT; TBN; WGT	1,572,443	884
AST, HLU; TBN; WGT	1,625,435	2,005
BRG; GNT; HLU; TBN	1,483,405	785
BRG; GNT; HLU; WGT	1,621,005	1,464
BRG; GNT; TBN; WGT	1,461,611	891
BRG; HLU; TBN; WGT	1,516,077	2,016
GNT; HLU; TBN; WGT	1,528,600	851

*Present value.

Table 9. Simulated performance—Part 5.

Replacement decision	Mean total cost ($)*	Mean system downtime
AST; BRG; GNT; HLU; TBN	1,620,627	662
AST; BRG; GNT; HLU; WGT	1,758,903	1,347
AST; BRG; GNT; TBN; WGT	1,598,189	767
AST; BRG; HLU; TBN; WGT	1,652,005	1,896
AST; GNT; HLU; TBN; WGT	1,664,867	726
BRG; GNT; HLU; TBN; WGT	1,553,340	732

*Present value.

benefits the other criteria, but some strategies can be discarded for being surpassed by at least one other strategy in both criteria.

All 64 assets replacement combination was simulated, but the plot of Figure 5 highlights only 26 solutions. These solutions were chosen to the plot for being between the 14 best solutions to minimize cost or unavailability. In Figure 6 the plot indicates the mean total down time and mean total cost in present value to each strategy.

It is clear that some strategies are overcome by at least one other strategy and others are not. The strategy of replacing BRG and GNT (green circle), for example, is surpassed by replacing BRG, GNT and TBN not only in terms of cost, but in terms of availability too.

The unsurpassed strategies form the efficient frontier as in Table 10.

The strategy of replacing only the Turbine in the beginning of the period results in the minimum

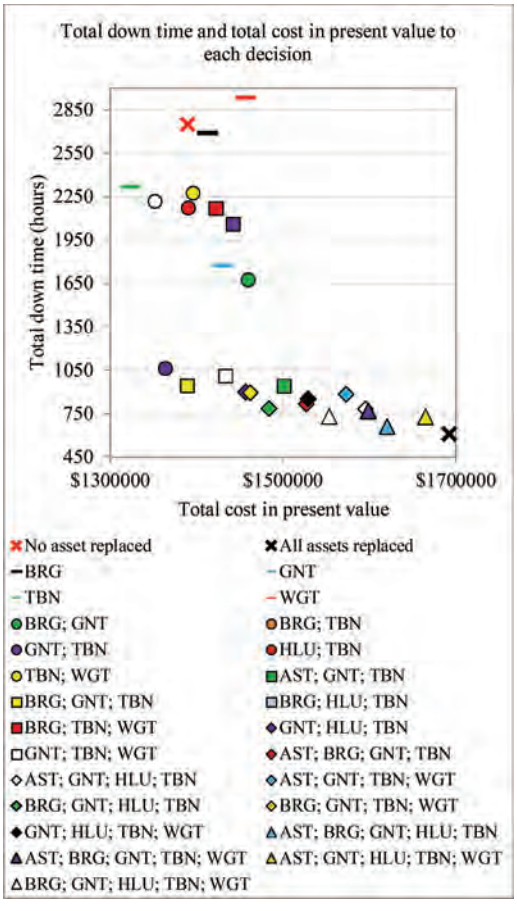


Figure 6. Mean total down time and mean total cost in present value to each decision.

Table 10. Markowitz effective frontier.

Replacement decision	Mean total cost ($)*	Mean system downtime
All assets replaced	1,692,042	611
TBN	1,323,404	2,319
GNT; TBN	1,364,146	1,058
BRG; GNT; TBN	1,389,410	940
GNT; HLU; TBN	1,456,535	900
BRG; GNT; HLU; TBN	1,483,405	785
BRG; GNT; TBN; WGT	1,461,611	891
AST; BRG; GNT; HLU; TBN	1,620,627	662
BRG; GNT; HLU; TBN; WGT	1,553,340	732

*Present value.

total cost in present value, but results in an high unavailability. The choice for one strategy must rely on budget to maintain the system and availability

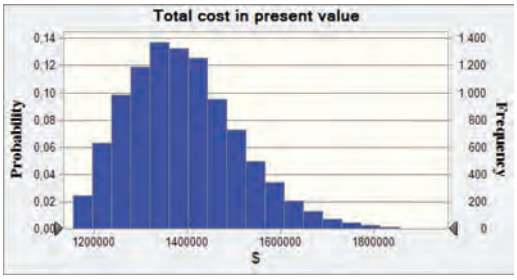


Figure 7. BRG, GNT, TBN replaced.

targets. So, if the companies do not tolerate total down time in hours to 3 years surpassing 650 hours, the strategy of replacing all assets can be a good option.

As the value discussed so far considers only the averages, we can also consider the uncertainty in costs. So the mean failure frequency is used to simulate the number of failure based on a Poisson distribution and considering each corrective maintenance cost as a uncertain variable which can assume a value between 80% and 120% of the value presented in Table 3.

In Figure 7 the histogram of the total cost in present value simulated considering replacement of Bearing, Generator and Turbine is presented.

Mean total cost in present value is equal to $ 1,389,261 similarly to the value presented in Table 10. The advantage of this simulation is to show the dispersion of the total cost represented by a standard deviation equal to $ 119,971, 10th percentile equal to $1,246,437, 90th percentile equal to $ 1,549,009, minimum value equal to $ 1,154,489 and maximum value equal to $ 1,937,684. So, the manager even making a decision based on the mean values can investigate the dispersion of each variable to have a better understanding about the risks.

6 CONCLUSIONS

Different strategies have different impacts in total cost and in total down time. Considering the existence of trade-offs between the two variables, an effective frontier can be created based on the simulated values for different replacement strategy, similarly to Markowitz (1952).

A big system contains a lot of asset, which results in a lot of combinations of replacement (strategies). This problem easily falls in the curse of the dimensionality, so, to big systems, the modelers must consider optimization algorithms in order to find a good solution or the efficiency frontier.

REFERENCES

Adkins, R.; Paxson, D. 2017. Replacement decisions with multiple stochastic values and depreciation. European Journal of Operational Research, v. 257, n. 1, p. 174–184.

Al-Chalabi, H. 2015. Economic lifetime of a drilling machine: a case study on mining industry. Int. J. Strategic Engineering Asset Management, Vol. 2, No. 2.

Bazargan, M.; Hartman, J. 2012. Aircraft replacement strategy: model and analysis. Journal of Air Transport Management, 25, 26–29.

Costa-Lima, G. A.; Teodoro-Filho, A. M. 2013. Economic and risk replacement model for physical assets in a natural gas distribution's system. In: 22nd European Safety and Reliability Conference, 2013, Amsterdam. 22nd European Safety and Reliability Conference.

Eilon, S.; King, J.R.; Hutchinson, D.E. 1966. A study in equipment replacement. Journal of the Operational Research Society, 17(1), 59–71.

Leung, L.C.; Tanchoco, J. M.A. 1990. Multiple machine replacement analysis. Engineering Costs and Production Economics, v. 20, n. 3, p. 265–275.

Malik, M.A.K. 1979. Reliable preventive maintenance policy. AIIE Transactions, 11(3), 221–228.

Markowitz, H. 1952. Portfolio selection. The journal of finance, v. 7, n. 1, p. 77–91.

Markowitz, H. 1959. Portfolio selection: Efficient diversification of investments. Cowles Foundation monograph. n. 16.

Melchor-Hernández, C.L.; Rivas-Dávalos, F.; Maximov, S.; Coria, V.H.; Guardado, J.L. 2015. A model for optimizing maintenance policy for power equipment. Electrical power and energy system, No. 68, PP. 304–212.

Powell, S.G.; Baker, K.R. 2009. Management science: The art of modeling with spreadsheets. Wiley.

Shi, W.; Min, K. J. 2014. Product Remanufacturing and Replacement Decisions Under Operations and Maintenance Cost Uncertainties. The Engineering Economist, 59:154–174.

Toledo, M.L.G.; Freitas, M.A.; Colosimo, E.A. 2015. ARA and ARI imperfect repair models: Estimation, goodness-of-fit and reliability prediction. Reliability engineering and system safety. No. 140, PP 107–115.

Zhang, C.; Guo, L.; Xiao, B.; Kang, R. 2012. Complex system fault sampling under conditon imperfect repair. Prognostics & system health management conference, Beijing, China.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Considerations related to insurance of cruise traffic in the arctic waters

K. Trantzas, O.T. Gudmestad & E.B. Abrahamsen
University of Stavanger, Norway

ABSTRACT: The wish of humans to explore new areas, environmental changes and growing worldwide demand have led to the Arctic region's increasing popularity in recent years. The cruise industry is continuously evolving in this area, creating an important need for more research into the risk of operating in the Arctic Ocean. However, most insurance firms do not yet have standard procedures for evaluating risk and policies to build insurance premiums for the Arctic cruise ship industry. The paper refers to our participation in a survival exercise in Arctic waters during May 2017, where the objectives were to assess the capability of rescue means in cold regions. Using our experience of the exercise as a basis, in this paper we are presenting the limitations related to cruise voyages in the Arctic. We suggest an insurance process that should be followed and finally discuss on some key factors that drive an insurance premium's cost for the cruise ship industry in the Arctic.

1 INTRODUCTION

In recent years, the Arctic has gained more and more popularity due to the extraordinary environmental and developmental changes that have taken place in this region. Meanwhile, climate change has led to extensive thinning of sea ice, making marine access in the Arctic Ocean much easier. It is obvious that this ice reduction extends to all seasons of the year, giving the maritime industry the opportunity for extended seasons of navigation and access to new areas that were previously difficult to reach. (Guy, 2006, Lasserre, 2011, Østreng et al., 2012, Sarrabezoles et al., 2014, Lasserre and Pelletier, 2011). At the same time, global marine tourism is rising and a place of extraordinary beauty like the Arctic could not stay unaffected by this trend. The potential impacts of these new marine uses—social, environmental and economic—are unknown but will be significant. Thus, there is great interest from both cruise ship owners and insurance companies regarding these trips. Safety is the main challenge that should be addressed by the owners of cruise ships operating in a remote and isolated area, with harsh weather conditions and poor infrastructure and communications. When a company needs to manage the negative consequences of an accident, it can: a) take all the consequences if/when an accidental event occurs, b) reduce the probability of an accident and/or its consequences by safety measures or c) transfer the consequences of the occurrence to parties better able to carry them (i.e. buying insurance) (Abrahamsen and Asche, 2011). As during any other operation, when planning a cruise, especially in an in-hospitable environment like the Arctic, both the ship owners and the passengers must be insured. Thus, marine and travel insurance companies are keen to increase their involvement in Arctic cruising, and this paper aims to give some considerations, in relation to insurance policies, that should be followed in the Arctic region. The limitations that the insurance companies should place on the ship owners, in order to offer insurance for their vessels, are discussed. In addition, this paper presents a standardized procedure that the insurance companies should follow before putting a value on the insurance contract. Finally, we debate which are the cost drivers and how could they affect the price of an insurance premium.

2 CRUISE INSURANCE

Cruise insurance is not considered marine insurance (Burke, 2000) but is better described as the sum of two categories: marine insurance and travel insurance.

With the term 'marine insurance', we mean the contract offered by an insurance company to the ship owner. With this contract, the insurer undertakes to indemnify the assured, in manner and to the extent thereby agreed, against marine losses, that is to say, the losses incident to marine adventure (Marine Insurance Act, 1906). There are specific marine insurance types and policies, which the cruise ship-owners are obliged to have, according to the international law and the national regulations of each country, depending on the specifications and the details of the upcoming voyage of their vessel.

Figure 1. Cruise insurance.

On the other hand, travel insurance is considered as the insurance product designed to cover the costs and losses of the passengers and to reduce the risk associated with unexpected events that someone might incur while traveling. Here again, there are different types that cover the specific needs of travelers.

In this subchapter, we discuss the policies of Arctic cruise insurance. The limitations that the insurance companies should place on the ship owners in order to offer insurance for their vessels, are presented. Finally, we debate which are the cost drivers and how could they affect the price of an insurance premium.

3 LIMITATIONS

The insurance companies decide whether they should offer insurance to a shipowner, according to their policies. Whether they offer an insurance premium to a cruise vessel owner depends on the risk appetite, meaning that it depends on the 'amount' and type of risk that a company is willing to take in order to meet their strategic objectives. An insurance company can be defined as: a) risk-averse when the company dislikes risk and will stay away from adding high risk investments to their portfolio, b) risk-seeking when the company prefers to take some high-risk investments and c) risk-neutral when the company seeks high-risk investments but at an average level.

Insurance is considered an alternative to investing in safety measures offered in order to transfer risk to a third party (insurance company) (Abrahamsen and Asche, 2011). However, the Arctic is a very hostile environment, about which we lack sufficient knowledge. Although there are no specific limitations stated by the regulatory framework in the Arctic, in order for a ship-owner to obtain insurance, we strongly believe that there should be some minimum requirements that should be covered by the cruise ship owners prior to an Arctic expedition.

Quite often ship-owners try to save from the costs of investing in safety measures by buying insurance. To make Arctic expeditions safer, there should be some limitations regarding their ability to obtain insurance. The main requirement that they must meet is to have an adequate polar- or ice-class-certified vessel for traveling in the Arctic. A vessel should not be allowed to operate in the Arctic area unless it is certified as eligible for the area. In August 2006, the International Association of Classification Societies (IACS) released a document, titled the "Unified Requirements for Polar Ships", which standardized global ice classification specifications for vessels (Table 1).

Another limitation should be the shipping firm's competence in Arctic shipping. There should be a strict investigation of the firm's past behavior in relation to safety policies. For example, if a shipping firm has neglected safety issues in previous trips and put passengers' lives at risk, then it should not qualify for an insurance contract.

Furthermore, before a ship-owner is given the right to insure his vessel for an Arctic voyage, there should be a thorough inspection of the vessel and the extent to which the ship owner has covered the requirements for survival equipment. An insurance company must deny insurance to a ship that does not have sufficient lifeboats and life rafts for all the passengers, modified to suit the Arctic needs (i.e. winterized). Another example could be the lack of adequate insulated survival suits, as well as Personal Survival Kits (PSKs) and General Survival Kits (GSKs). An insurance company could protect itself against wrongdoing or neglect on safety issues by the shipping firm, by stating specific terms in the insurance premium that place the blame on the ship-owner in the case of an accidental event which occurred due to the shipowner's negligence.

But then an important question arise: Should those limitations regarding the shipping firm

Table 1. Polar Class descriptions (International Association of Classification Societies, 2016).

Polar class	Ice descriptions (based on WMO Sea Ice Nomenclature)
PC 1	Year-round operation in all polar waters
PC 2	Year-round operation in moderate multi-year ice conditions
PC 3	Year-round operation in second-year ice which may include multiyear ice inclusions
PC 4	Year-round operation in thick first-year ice which may include old ice inclusions
PC 5	Year-round operation in medium first-year ice which may include old ice inclusions
PC 6	Summer/autumn operation in medium first-year ice which may include old ice inclusions
PC 7	Summer/autumn operation in thin first-year ice which may include old ice inclusions

depend only on the insurance company's interpretation? As already stated, certain insurance companies are considered to be risk seeking, meaning that they are willing to take high risks in order to gain profits. Thus, it is of great importance that these limitations be included in a legally binding agreement. There must be a regulatory framework between the Arctic countries and the ship owners interested in operating in Arctic waters that states specific limitations for the vessels, in terms not only of operating in the Arctic but also of obtaining insurance coverage. This way, the ship owner can be made responsible in the case of an accidental event that involves neglect or wrongdoing on the shipping firm's side.

4 INSURANCE PREMIUMS IN THE ARCTIC

The lack of data and standardized methods regarding the assessment and the modeling of the risks, as well as the poor background knowledge, create a big challenge for insurance companies in the Arctic. This leads the underwriters to work on a case-by-case basis that vastly increases the cost of the insurance premium. Thus, the sustainability of Arctic expeditions is strictly dependent on the cost of marine insurance, and the industry is calling for more standardized procedures. Here, a standardized procedure for the insurance companies is suggested (Figure 2).

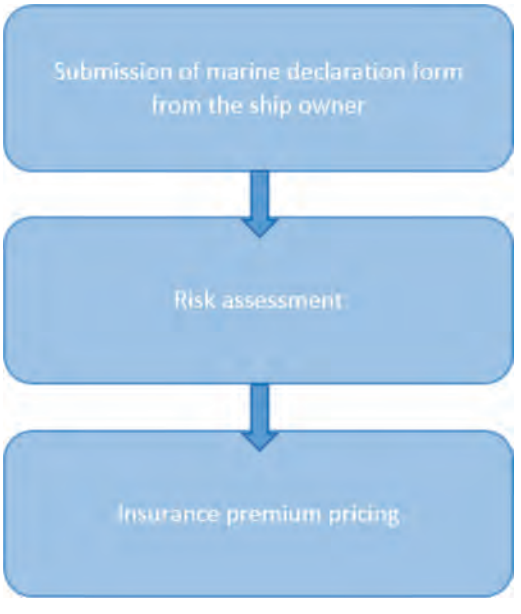


Figure 2. Phases of suggested insurance premium procedure.

Initially, and after receiving all the relevant information regarding the vessel and the trip to be insured, from the ship owner through the marine declaration form, an insurance company should check whether the shipping firm covers the limitations regarding obtaining insurance (polar/ice class, life-saving equipment, etc.). If those limitations are covered, then a risk assessment should follow, with focus on the consequences and the associated probabilities. The risk assessment starts with a qualitative risk analysis, in which the hazard identification and their categorization take place. After identifying the possible hazards and their related consequences, a quantitative risk analysis should follow, in which, according to the data for the region of the trip concerned and the background knowledge of previous incidents, a probability number will be assigned to each accidental event. Then, by using the specific information obtained from the shipping firm regarding the vessel and the trip (e.g. ship design information, winterization of the ship, quality and quantity of survival equipment, etc.), adjustments should be made to the probabilities linked with the specifications of the shipping firm. Special attention must be given to the strength of knowledge and the uncertainties, as these may lead to a higher insurance premium. In general, the insurance premium is to a large extent equal to the expected costs. Then, the insurance premium is considered fair. If an insurance company does not offer fair insurance another company can attract their customers. Thus, high uncertainties or poor background knowledge can result in higher insurance premiums. For example, one important finding of our participation in the survival exercise in the Arctic was the influence of an integrated heating system in the lifeboats on the survivability rate of the passengers (Gudmestad et al., 2017). A vessel that does carry lifeboats with an integrated heating system will be assigned a higher probability for the risk of people getting cold due to low temperatures than a vessel equipped with lifeboats with a heating system. Thus, the insurance premium will be lower for the second vessel, as the passengers have a higher probability of surviving and the insurance company is taking a lower risk.

After finishing the risk assessment, the pricing of the insurance premium follows. In this phase we identify two main steps. During the first step, the expected costs are calculated by using information from the risk assessment. The expected cost for each accidental event, $E[C|A_i]$, is multiplied by the associated probability $P(A_i)$. This is done for all the accidental events, and then summarized to give an initial price at the insurance premium. During the second step, the insurance premiums calculated on the previous step based on the expected costs, are

adjusted in order to take the strength of knowledge and uncertainties into consideration. Finally, the administration costs and a percentage of the profit that the company wants to make should be added to the price of the premium decided by the insurance company on the first step. As already mentioned, an insurance premium is considered fair when it is to a large extent equal to the expected costs. As a result, the added profit that an insurance company wishes to obtain should not be too high. It is also recommended that this later adjustments of the second step should be made through a broad managerial review. It is also of great importance that the insurance company executes a self-assessment of the vessel, sending its own surveyors to check the design and the equipment of the cruise ship.

A cruise ship operator is concerned with lowering the costs to increase the profits and will, most likely, only want to fulfill the minimum requirements stated by the legal framework. However, if the requirements are very strict and conservative, it might make the vessel owners stop their Arctic cruise business. Having first-class survival equipment that could fit every passenger on a cruise ship of more than 2000–3000 passengers could be quite expensive. Thus, there must be a conscious effort by the legislators to secure top levels of safety in Arctic cruise traffic, while, at the same time, not acting conservatively and overestimating the presence of extra safety measures.

5 ARTIC REGION COST DRIVERS OF MARINE INSURANCE

After presenting the limitations that should apply to Arctic voyages, and our suggestion for a standardized procedure that should be implemented on determining an insurance premium in the Arctic, we will now discuss the influence of specific factors on the insurance premium.

The most important criterion that influences the cost of the insurance premium is the ice class of the vessel. As previously mentioned, insurance companies should not offer insurance contracts to firms whose ships are not designed for navigation in potentially iced waters. However, there are different polar and ice class categories and, thus, the higher the polar and/or ice class of a vessel is, the lower the price of the insurance premium offered by the insurance company. This is reasonable, as the higher the ice class of a vessel, the higher the capacity of the vessel to withstand harsh weather conditions and avoid any accidental events.

Another factor that could influence the cost of an insurance premium in the Arctic is the winterization of the vessels. Winterization is a process, which enables vessels to operate in extreme sub-zero temperatures without suffering loss of equipment operability, vessel stability and power, and personnel habitability, and which permits crew operations to be performed safely (Ghosh and Rubly, 2015). Class regulations require equipment and systems to be winterized. Some of those winterizations procedures are presented by Hasholt (2011). Sheltered pathways, under-deck heating, heated mooring equipment, low temperature emergency generators, ice navigation radar, ice searchlights, etc. are all elements that could drive the cost of an insurance premium down. The more of the aforementioned elements a cruise ship has, the more 'winterized' it is and the less is the price of the premium. Each element has specific importance for operating in the Arctic and together they constitute the winterization of the vessel. For example, if a vessel is considered 40% winterized then there could be an agreement with the insurance company to lower the insurance premium price by 5%–10%.

Communication systems and stability play an important role in the Arctic region. The remoteness of the area highlights the need for survival equipment able to help the passengers to survive for the period of five days stated in the regulatory framework for ships operating in the Arctic, The Polar Code (IMO, 2016). The presence of enhanced communication systems and adequate survival equipment could dramatically lower the overall price of an insurance premium. During harsh weather, improved communication systems could prevent loss of communication, which could be vital for the cruise ship. Being equipped with sufficient lifeboats, life rafts, survival suits, PSKs and GSKs is one of the previously stated limitations. However, during our survival exercise in Arctic waters, it was noted that further improving the survival suits, by adding integrated woolen underwear, and providing sizes for all the passengers could lead to an insurance company's decision to lower their offer, as the severity of the consequences in case of an unexpected event are reduced (Gudmestad et al., 2017, Solberg et al., 2016).

The time of the year and the route that the cruise ship is planning to take can also influence the insurance premium. For instance, a trip planned during August would have a lower insurance premium than one trip planned during the beginning of the Arctic cruise season in May. This is because the probable ice concentration and ice movement during August would be lower than the corresponding ice concentration and ice movement during May (Lasserre, 2014).

One of the most important criteria for an area like the Arctic, where we lack sufficient knowledge, is the experience of the captain and crew members'. According to Sarrabezoles et al. (2014), after inter-

viewing several companies that offer marine insurance for the Arctic, they found most firms said that trust in the shipping company is important and that they therefore might be reluctant to insure firms that do not have experience in Arctic shipping. Some of them stated that they would examine every submission but evaluate the preparedness of the shipping firm, the crew's experience, charts' accuracy and contingency planning in case of problems. However, for the majority of firms, it is clear there is a strict inspection of the shipping firm's past behavior and safety-related policy. This means that the insurance company expects to see proof that the shipping firm is able to perform well in Arctic waters.

Finally, each traveler should have private travel insurance. However, most private travel insurance companies require travelers to have separate search and rescue cover. This can raise the price for the traveler by two or sometimes three times, compared to the initial travel insurance cost. Furthermore, it is not unusual for those third-party companies that offer search and rescue cover, to put extra limitations on travelers, such as age or the area in which they provide cover. Thus, it is of significant importance that the ship-owners should acquire such search and rescue cover for all their passengers. This would decrease, the insurance premium provided by the insurance company, as there would be higher chances of survival in the case of an accidental event and, thus, lower severity of consequences. The private travel insurance premium of the travelers would also decrease. However, the ship owners will have to undertake extra cost for the search and rescue cover.

6 CONCLUSIONS

Although Arctic cruises are gaining in popularity, the Arctic region remains a hostile environment, where we lack sufficient background knowledge and data to determine insurance premiums. There are many hazards related to cruise traffic in the Arctic, and even more challenges arise in the case of an accidental event during the voyage.

Through a legislative framework, specific limitations must be implemented that, unless met by the ship owners, forbid the signing of insurance cover. Ice and/or polar class certification, thorough background check, inspection of the vessel and adequacy of survival equipment must all be included as mandatory limitations before insuring a cruise vessel for an Arctic voyage.

A standardized procedure should be followed by underwriters to determine insurance premiums in the Arctic. Thus, the creation of a platform, where Arctic data would be gathered and easily accessed is of high importance.

Different factors can influence the cost of an insurance premium between a shipping firm and an insurance company, the most important being the level of winterization of the cruise vessel, the level of training of the shipmaster and the crew members and the coverage of the search and rescue procedure for the passengers in case of emergency.

REFERENCES

Abrahamsen, E.B. & Asche, F. 2011. On how access to an insurance market affects investments in safety measures, based on the expected utility theory. *Reliability Engineering & System Safety,* 96, 361–364.

Burke, D.D. 2000. Cruise lines and consumers troubled waters. *American Business Law Journal,* 37, 689.

Ghosh, S. & Rubly, C. 2015. The emergence of Arctic shipping: issues, threats, costs, and risk-mitigating strategies of the Polar Code. *Australian Journal of Maritime & Ocean Affairs,* 7, 171–182.

Gudmestad, O.T., Solberg, K.E. & Skjærseth, E. 2017. SARex2: Surviving a maritime incident in cold climate conditions. *Rapporter fra Universitetet i Stavanger; 69.* University of Stavanger.

Guy, E. 2006. Evaluating the viability of commercial shipping in the Northwest Passage.

Hasholt, S. 2011. Rules for Ice and Cold Operations: 'Winterization' of Vessels (Corporate presentation). London: Lloyd's Register.

IMO 2016. International Code for Ships Operating in Polar Waters (Polar Code). *In:* IMO (ed.).

International Association of Classification Societies 2016. *Requirements concerning POLAR CLASS*.

Lasserre, F. 2011. Arctic shipping routes: from the Panama myth to reality. *International Journal,* 66, 793–808.

Lasserre, F. 2014. Case studies of shipping along Arctic routes. Analysis and profitability perspectives for the container sector. *Transportation Research Part A: Policy and Practice,* 66, 144–161.

Lasserre, F. & Pelletier, S. 2011. Polar super seaways? Maritime transport in the Arctic: an analysis of shipowners' intentions. *Journal of Transport Geography,* 19, 1465–1473.

Marine Insurance Act 1906. English Marine Insurance Act 1906 – An Act to codify the Law relating to Marine Insurance. England: UK Parliament.

Sarrabezoles, A., Lasserre, F. & Hagouagn'rin, Z. 2014. Arctic shipping insurance: towards a harmonisation of practices and costs *Polar Record*.

Solberg, K.E., Gudmestad, O.T. & Kvamme, B.O. 2016. Search and rescue exercise conducted off North Spitzbergen: Exercise report. *Rapporter fra Universitetet i Stavanger; 58.* University of Stavanger.

Østreng, W., Eger, K.M., Fløistad, B., Jørgensen-Dahl, A., Lothe, L., Mejlænder-Larsen, M. & Wergeland, T. 2012. *Shipping in Arctic waters: a comparison of the northeast, northwest and trans-polar passages.* New York; Springer.

Organizational factors and safety culture

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

The impact of personal liability concerns on incident reporting in engineered systems

Jan Hayes, Janice Wong & Christina Scott-Young
School of Property, Construction and Project Management, RMIT University, Melbourne, Australia

Sarah Maslen
School of Government and Policy, University of Canberra, Canberra, Australia

ABSTRACT: Previous research on aviation and health sectors has found that individual blame for small failures discourages incident reporting and so adversely impacts disaster prevention. This finding has widely influenced practice in organizations relying on engineers. Based on a survey of Australian engineers (n = 275) this paper examines how personal legal liability considerations impact on hazard reporting and other forms of knowledge sharing. We found that 48% of engineers are more likely to report hazards despite changes in societal expectations and the tendency to blame. Only 5% indicated that they were less likely to report hazards as a result of their liability concerns. We suggest that these findings are due to the nature of engineering work, where decision-making is distributed across time, place and people. In this environment, blame and responsibility are less attributable to individual actors. Equally, reporting a hazard may act to transfer responsibility and so limit one's personal liability.

1 INTRODUCTION

Excellence in organizational safety performance requires that all personnel are willing to learn from small faults and failures (Hopkins, 2009) and also from past accidents (Snook, 2000) in order to prevent recurrence. This leads to a range of strategies for collecting and sharing information ranging from formal incident reporting to informal sharing of stories within professional groups (Hayes and Maslen, 2015, Maslen and Hayes, 2016).

When it comes to formal incident reporting systems, a key challenge is encouraging people to report what they know so that systemic problems can be identified and addressed. A culture that encourages reporting is thought to be a just culture—one in which people can report genuine mistakes without fear of retribution (Dekker, 2007, Reason, 1997). Much of the research that led to these conclusions was originally grounded in the health and aviation sectors although the findings have been assumed to generalize into any complex socio-technical system including those where work is largely carried out by engineers. Our study challenges this paradigm within the engineering context.

We have studied the impact of personal legal liability concerns on high stakes decision making by engineers. The results of our survey show that personal legal liability supports good decision-making when legal responsibility aligns with issues over which engineers have control (Hayes et al., 2017). Where these are not in alignment, there is evidence that 'defensive engineering' practices can develop analogous to 'defensive medicine'—a set of practices that have grown up in the health sector in response to increased liability concerns (Catino, 2009).

In this article we focus specifically on those aspects of the survey related to the impact of personal liability concerns on reporting and knowledge sharing. Knowledge sharing practices include incident reporting, the sharing of stories about past incidents and the release of findings from high quality accident investigations that support learning. First, we engage with the literature on safety culture and responsibility in engineering. We then describe our survey method and the survey results as they apply to learning from accidents and incidents. The paper concludes with a discussion of the implications of our findings and key conclusions.

We argue that the nature of engineering work means that knowledge sharing behaviors (reporting, storytelling) are less impacted by personal liability concerns in engineering-based industries than in healthcare and aviation. Rather than inhibiting reporting, concerns about blame may encourage reporting. There is some evidence of impacts of personal liability concerns on storytelling among our respondents' colleagues, though this is not as pronounced as in healthcare and aviation. The implications of this result warrants further inquiry.

2 JUST CULTURE AND ENGINEERING RESPONSIBILITY

Much organizational learning uses a strategy of trial and error, but learning in order to prevent disasters requires more sophistication. Reason's (1997) famous Swiss cheese model of accident causation encourages us to seek out those holes in the cheese not only so they can be fixed, but because they provide important information about organizational weaknesses. Similarly, high reliability theorists tell us that high performing organizations are preoccupied with finding small errors and faults, again for the insights they provide on underlying organizational issues that have the potential to result in more serious failures (Weick and Sutcliffe, 2001).

Such learning strategies rely critically on professionals throughout organizations being prepared to report small errors and faults but this is an inherently difficult undertaking when, at least in some circumstances, people are effectively being asked to report themselves. As Reason (1997, pg. 196) says, 'human reactions to making mistakes take various forms, but frank confession does not usually come high on the list'.

In response to this problem, a significant research literature and safety practice has developed around what is known as 'just culture' which attempts to balance accountability for current problems with learning in order to prevent more problems in the future (Dekker, 2007). Enacting a just culture within an organization is seen as the ultimate goal in fostering open reporting of small faults, failures and errors and so maximizing opportunities for learning (Aveling et al., 2016, Gerede, 2015, Khatri et al., 2009). These ideas have been taken up with gusto, particularly in the health sector. A Scopus search for 'just culture' returns 175 items of which 71% are health sector publications and 12% are air traffic management-related. Only six articles are based in engineering-related industries and all of these focus on application of the just culture concept.

Despite this lack of specific evidence, the proposition that a just culture increases reporting is assumed to apply in engineering-based industries, too. In particular, it has become common for safety culture survey tools to include questions based on just culture principles (see for example Kines et al., 2011, Shirali et al., 2013).

In addition to formal reporting systems, informal practices including sharing stories are also important for learning and accident prevention. Engineering professionals share stories among their professional group with the explicit purpose of fostering expertise (Hayes and Maslen, 2015, Vastveit et al., 2015). This also provides a mechanism for learning from past failures, be they small faults or major disasters. Through stories, workers come to understand the potential consequences of their decisions (Maslen, 2014, Maslen and Hayes, 2016, Størseth and Tinmannsvik, 2012). They support the development of a 'safety imagination'—an ability to link one's actions to the potential consequences (Pidgeon and O'Leary, 2000). When then faced with potentially minor operating anomalies, engineers are able to draw connections to past major events and so dig deeper into the state of the system (Macrae, 2009).

Professionals taking responsibility for their own expertise and learning to ensure the best decision making is an example of what the engineering ethics literature calls forward-looking responsibility: 'a virtue or a moral obligation to see to it that a certain state-of-affairs applies' (Doorn and van de Poel, 2012, pg. 10), in this case accident-free operations. Some authors have called for an increased focus on this aspect of engineering responsibility pointing out that the discourse on engineering ethics and responsibility tends to be dominated by a framework of responsibility as blameworthiness (Coeckelbergh, 2012, Doorn, 2012, Kermisch, 2012). Questions of engineering responsibility therefore link directly to the issue of learning from incidents and accidents as past errors can be interpreted either as deserving of punishment or as key lessons to prevent recurrence (Fahlquist, 2006).

We find these issues occupying the attention of survey respondents, but in ways that may not be expected based on the body of published research.

3 METHOD

To assess the prevalence and impacts of liability concerns among engineers in Australia, we developed and administered a survey with a mixture of closed- and open-ended questions. It was designed for self-completion via an internet browser on a computer, tablet or smartphone and was programmed and administered using Qualtrics software.

The questionnaire addresses six main issues:

- Risks and benefits of considering personal liability.
- Impact of serious incidents on personal liability concerns.
- Impact of others' individual experiences on personal liability concerns.
- Impact of personal liability concerns on reporting and knowledge sharing.
- Impact of personal liability concerns on professional board registration.
- Impact of personal liability concerns on career.

Survey participants were recruited from two sources. Email invitations to participate in the survey were sent to Australian Pipeline and Gas Association Research and Standards Committee members

and to Engineers Australia's Brisbane area members. A total of 275 Australian engineers participated in the survey during the period July to September 2016. Due to the way in which Australian engineers were recruited for this survey (i.e. a snowballing approach was used whereby one group of targeted respondents was asked to circulate the survey throughout their organisation in addition to participating themselves), it is not possible to accurately identify the total number of Australian engineers invited to participate. Therefore, it is not possible to accurately calculate the survey response rate nor to comment quantitatively on the extent to which the results might be indicative of engineers in Australia overall.

This work has been approved by the relevant university human ethics committees. The survey data were processed using the IBM SPSS-23 statistical software package and both IBM SPSS-23 and Microsoft Excel were used to produce percentages. Graphical representations of the data were produced in Microsoft Excel.

The percentage results presented against each survey question have been calculated using the total number of valid responses received for that question. The number of valid responses (as shown in each figure) varies due to individuals (either intentionally or unintentionally) not providing a response to a question and/or individuals not being eligible to complete a question given their previous responses in the questionnaire.

4 IMPACT OF PERSONAL LIABILITY CONCERNS ON REPORTING AND KNOWLEDGE SHARING

The extent to which respondents are concerned about personal legal liability provides important context for these results. Thirty-two per cent of respondents indicated that they think about personal liability to a great extent when making decisions at work. Further, 89% of respondents indicated that they consider personal liability at least to some extent. These findings show that personal liability is an important issue for respondents.

To investigate the impact of liability concerns on formal hazard reporting and informal practices for sharing knowledge about things that have gone wrong, we asked a series of questions regarding hazard reporting practices and other practices about learning from past failures.

Over four fifths of respondents (83%) said their organization provided a formal system for employees to report hazards. Among these respondents, close to all (93%) of those who had encountered a hazard in their organization said they reported it into the organization's formal system. Only 12 of the respondents (out of a possible 107) who had encountered a hazard in their organization had not formally reported it.

Members of this small group were then prompted to explain why, giving the following reasons for not making a formal report:

- The incident was too small to report. This was the most frequently noted reason ('I felt the hazard was minor'; 'it was isolated…it was personal factors, not systemic'; 'trivial problems—can't cotton ball everything and everyone').
- Time to report the incident ('no time'; 'not worth the time to report').
- The incident was already known. One respondent observed '[I] didn't want to create a fuss over something that was already known'.
- The incident was immediately fixed ('the people involved investigated and steps [were] taken to redress the problem immediately'; 'I was able to do something about it rather than just report it for someone else to deal with').

In the context of the survey, it is noteworthy that no-one indicated that they had failed to report a hazard as a result of concerns about liability and blame.

Building on this, Figure 1 shows responses regarding whether personal liability concerns impact on whether a respondent is likely to report hazards into their organization's formal system. Forty-seven per cent of respondents indicated that personal liability concerns would have no impact. Only 5% of respondents indicated that they were less likely to report hazards as a result of liability concerns. In contrast, 48% of respondents indicated that they were more likely to report hazards including 21% who said they were much more likely to report hazards. In other words, personal liability concerns promote rather than hinder likelihood to report. This is a critical finding because it is contrary to what we would anticipate based on previous research that emphasizes the importance of a just culture in fostering hazard reporting.

Another way in which learning about problems occurs is via informal sharing of stories. Figure 2 shows responses regarding whether personal liability concerns impact on the likelihood that a respondent will share stories of things that have gone wrong with colleagues so that lessons can be learned. Forty-two per cent of respondents indicated that personal liability concerns would have no impact. Sixteen per cent of respondents indicated that they were less likely to share stories as a result of liability concerns. While this is more than the 5% who indicated they were less likely to report hazards as a result of liability concerns it is still a low figure. Again, contrary to what might be expected from the broader literature on organizational learning and

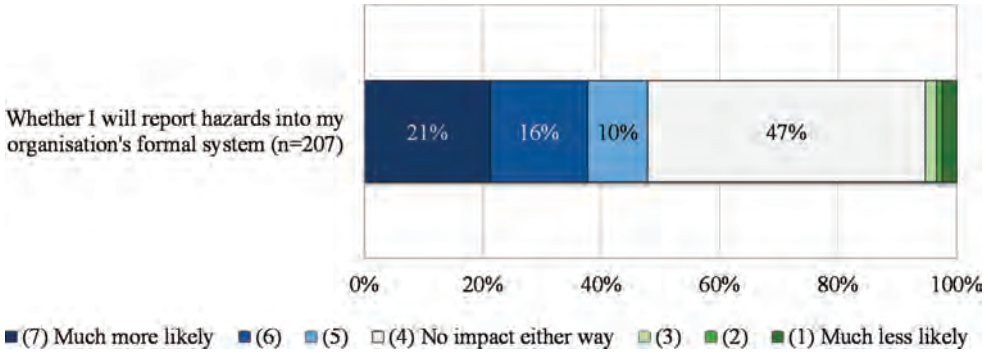


Figure 1. Does the risk of your personal liability make it more or less likely that you will report hazards into your organisation's formal system?

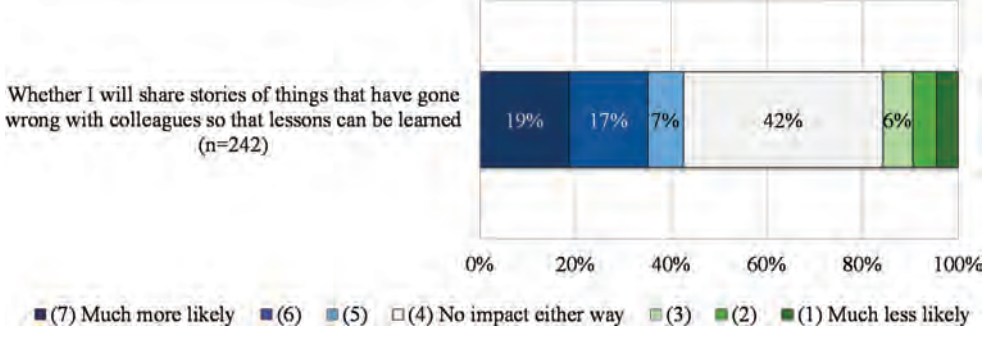


Figure 2. Does the risk of your personal liability make it more or less likely that you will share stories of things that have gone wrong with colleagues so that lessons can be learned?

blame, 43% of respondents indicated that having personal liability concerns made them more likely to share stories, including 19% who said the concerns made them much more likely to share stories.

While personal liability concerns appear to have only a minor impact on willingness to share stories, it is important to note that these figures were higher when engineers reflected on sharing of stories by others. Figure 3 indicates that almost one third of respondents (31%) were aware of other engineers being unwilling to share stories of things going wrong due to concerns about personal liability.

Those aware of such cases were asked to provide more details. Thirty seven respondents chose to do so.

Key reasons cited that influenced people's perceived willingness to share stories of things going wrong are:

- Contractual relationships and concerns about blame and financial liability;
- Failures being indicative of incompetence and so those involved being unwilling to admit what has occurred;

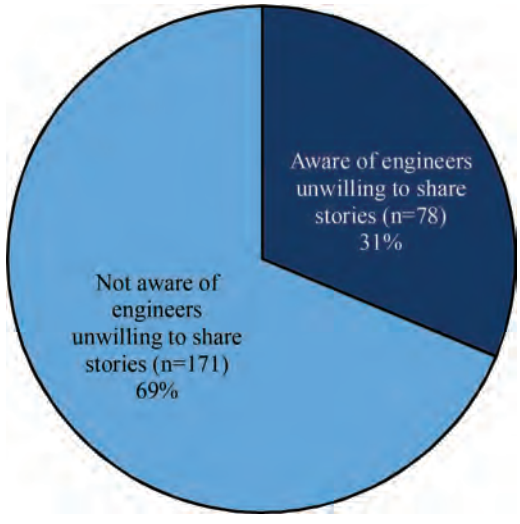


Figure 3. Are you aware of any engineers (other than yourself) who were unwilling to share stories of things going wrong due to their concerns about their personal liability?

- Fear of blame linked to job security; and
- Fear of blame linked to legal processes.

We would expect unwillingness to share stories on these grounds based on the just culture research. The reasons that engineers feel that their peers may be less likely to share stories than themselves warrants further inquiry.

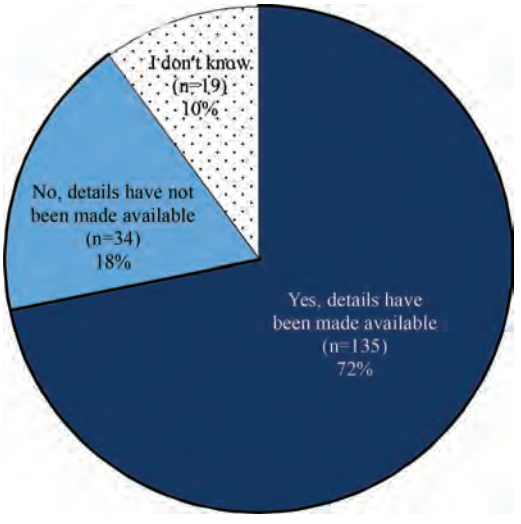


Figure 4. [Regarding the serious accidents that occurred in your sector, either in Australia or overseas] Have details of those serious accidents been made available in a way that allows lessons to be learned?

Other sources that may be drawn upon to learn about disaster prevention include investigations into the causes of serious accidents. As shown in Figure 4, 72% of respondents reported that details of serious accidents in their particular sector have been made available in a way that allows lessons to be learned. While this is positive, it also shows room for improvement. Twenty-eight per cent of respondents indicated either that information had not been made available or they were not aware whether it had been made available. In either case, respondents were not able to learn from serious accidents. Opportunities to learn lessons from major failures are mercifully rare. However, this makes it all the more important that lessons are maximized when accidents occur to prevent future occurrences. In this context, it is problematic that almost a third of respondents had observed information not being made available post-disaster.

Looking into this issue further, respondents were asked if they knew why information had not been made available for learning. As shown in Figure 5, in more than two thirds of cases (68%), respondents reported that lessons had not been generated and/or not released. In only 12% of cases, was the issue one of timing in that lessons would be released in the future. In 21% of cases respondents were not aware why lessons had not been released.

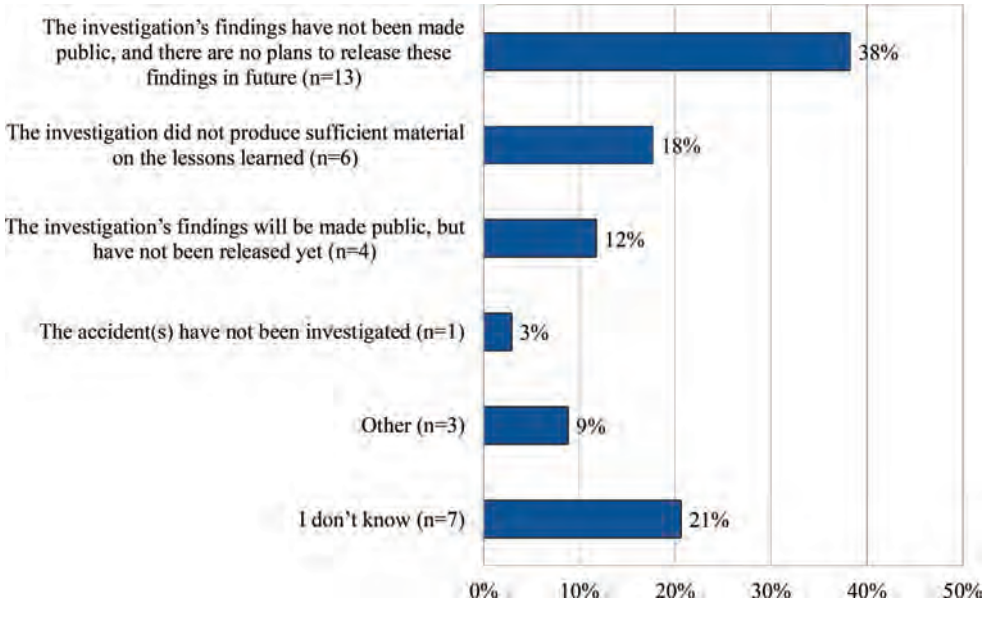


Figure 5. Do you know why the details [of the serious accidents] are not available?

5 DISCUSSION

Based on the small survey population it appears that most engineering organizations have in place a system for reporting hazards and that engineers use such systems with very few exceptions. Despite the expectations of the just culture literature that concerns about blame would inhibit reporting, this does not seem to be the case for these small and relatively common workplace problems. This surprising result is also borne out by our previous work in incident reporting in a nuclear power station which exhibited over 2000 reports per annum with the primary advantage to those making reports being seen as quick action to remediate issues raised (Hayes, 2009).

When asked specifically about legal liability and hazard reporting, many respondents saw liability as a reason to increase, rather than decrease, reporting. Respondents reported similar behavior when it came to more informal professional learning practices regarding sharing stories with colleagues. This raises the question as to why liability is of less concern in the engineering sector than in health. Doorn and van de Poel (2012) have highlighted the extent to which responsibility is a key issue in engineering ethics (in contrast to medical ethics, for example). They maintain that this is in response to the collective nature of engineering activity and also the uncertainty and indirect causality of major engineering failures. As Coeckelbergh describes, '… between the actions of an engineer and the eventual consequences of her actions lies a world of relationships, people, things, time and space (2012, pg. 37).

It is possible that this goes some way to explain our surprising results on this issue. In an environment where work is both distributed and collective, responsibility for small problems is also distributed. Perhaps reporting small problems could be seen as 'insurance' against blame beyond one's sphere of direct control in the event that something goes wrong. Reporting possible warning signs may provide something to point to as a defense against blame when collective endeavors fail to give the required outcomes. This explanation is somewhat speculative, but it is consistent with evidence to date.

Overall, the positive link between liability and reporting appears to be good news. Concerns about legal liability are perhaps raising awareness of the need to learn and so increasing forward-looking responsibility. At its best, forward-looking responsibility is conceptualized by ethics scholars as going beyond simply doing one's duty. Duty implies rule-following, whereas responsibility is more concerned with getting the best outcome. As Goodin (1986, pg. 50) explains, '[d]uties dictate actions. Responsibilities dictate results.' Other safety researchers have warned of the dangers of decision making grounded only in compliance where obeying rules becomes an end in itself, rather than a means to safe operations (Bieder and Bourrier, 2013). It seems at least possible that a link between learning and responsibility mediated by an increased focus on liability may support the development of forward-looking or virtue responsibility, rather than simply a desire to avoid blame.

The potentially positive impact of liability concerns on learning is tempered by the results obtained when questioning moves from self-reported behavior to observations about the behavior of professional colleagues. Almost one third of respondents reported seeing colleagues who are unwilling to share stories of things going wrong for reasons linked to liability. Possibly in these cases the potential storyteller has a more personal link to the problem at hand and so feels more acutely a sense of backward-looking responsibility as blameworthiness outweighing a sense of forward-looking responsibility and virtue in efforts to prevent accidents.

This could also be a reflection of the reflexivity of respondents and their discomfort confessing their own perceived shortcomings. They are uncomfortable talking about their own shortcomings, but are much more talkative about the shortcomings of their colleagues. In another study, in which we examined decision making in the oil and gas sector, we found that respondents claimed that their decision making was not influenced by financial incentives such as bonus payments awarded by their employer (Maslen and Hopkins, 2014). Incentives would have a negative impact on the quality of decisions from a safety perspective, in the same way that failing to share stories due to personal liability concerns has the potential to negatively impact disaster prevention. In contrast, many felt that their colleagues' decisions could be influenced by incentives. In an interview context, as opposed to a survey, the interviewer has an opportunity to explore the reasons for this inconsistency. Is it perhaps that liability concerns (like incentives) do influence behavior in a way that is not readily acknowledged? This is a critical question that warrants attention using qualitative methods such as interviewing and observation.

The final issue of interest is learning from significant disasters that have been the subject of formal inquiries. The survey results indicate that there is room for improvement in this area. Respondents reported a significant proportion of cases in which lessons are not made available for learning. Dekker (2015) describes four psychological purposes of accident investigations: epistemological (establishing what happened), preventative

(identifying pathways to avoidance), moral (tracing the transgressions that were committed and reinforcing moral and regulatory boundaries) and existential (finding an explanation for the suffering that occurred). These different orientations to an investigation result in different findings as shown by Gephart (1993) in his detailed analysis of a Canadian pipeline failure. As our survey respondents indicate, not all orientations result in effective lessons to prevent recurrence.

6 CONCLUSIONS

Much published research focuses on the adverse impact of personal liability and blame on learning from accidents and incidents. The conventional wisdom is certainly that blame inhibits learning. Our study in the engineering profession reveals a more complex relationship. It seems that, at least for our respondents, a focus on personal legal liability may act to increase practices related to learning as a result of an increased sense of virtue responsibility.

This finding is tempered however by reports that a significant minority of engineers are unwilling to share stories as a result of liability concerns. Legal constraints may also be responsible for a reluctance to share lessons for major accident investigations in some cases.

It appears that liability is a double-edged sword with both positive and negative effects on learning which are surely worthy of further investigation, particularly in this engineering context where such issues are little studied by comparison with the health sector.

It must be emphasized that this is effectively a pilot study with a relatively small cohort of respondents. As such, we make no specific claims that these findings are representative of the engineering profession in Australia or globally. Nevertheless, the findings are surprising and deserve further investigation in the context of organizational learning for accident prevention.

ACKNOWLEDGEMENTS

This work was funded by the Energy Pipelines CRC, supported through the Australian Government's Cooperative Research Centres Program. The cash and in-kind support from the APGA RSC is gratefully acknowledged.

We acknowledge the support of Engineers Australia Queensland Division in distributing the survey and the interest they have expressed in the results.

We also acknowledge the survey participants who contributed to this study. Many of the respondents expressed a desire to contribute to the future of their profession and they saw our survey as a vehicle to do so. They deserve our sincere thanks.

REFERENCES

Aveling, E.-L., Parker, M. & Dixon-Woods, M. 2016. What is the role of individual accountability in patient safety? A multi-site ethnographic study. *Sociology of Health and Illness,* 38, 216–232.

Bieder, C. & Bourrier, M. (eds.) 2013. *Trapping safety into rules: How desirable or avoidable is proceduralization?*, Farnham: Ashgate.

Catino, M. 2009. Blame culture and defensive medicine. *Cognition, Technology and Work,* 11, 245–253.

Coeckelbergh, M. 2012. Moral responsibility, technology, and experiences of the tragic: From Kierkegaard to offshore engineering. *Science and Engineering Ethics,* 18, 35–48.

Dekker, S. 2007. *Just culture: Balancing safety and accountability*, Aldershot: Ashgate.

Dekker, S. 2015. The psychology of accident investigation: epistemological, preventive, moral and existential meaning-making. *Theoretical Issues in Ergonomics Science,* 16, 202–213.

Doorn, N. 2012. Responsibility ascriptions in technology development and engineering: Three perspectives. *Science and Engineering Ethics,* 18, 69–90.

Doorn, N. & Van De Poel, I. 2012. Editors' overview: Moral responsibility in technology and engineering. *Science and Engineering Ethics,* 18, 1–11.

Fahlquist, J.N. 2006. Responsibility ascriptions and Vision Zero. *Accident Analysis and Prevention* 38, 1113–1118.

Gephart, R.P. 1993. The textual approach: Risk and blame in disaster sensemaking. *The Academy of Management Journal,* 36, 1465–1514.

Gerede, E. 2015. A study of challenges to the success of the safety management system in aircraft maintenance organizations in Turkey. *Safety Science,* 73, 106–116.

Goodin, R.E. 1986. Responsibilities. *The Philosophical Quarterly,* 36, 50–56.

Hayes, J. 2009. Incident reporting: A nuclear industry case study. In Hopkins, A. (ed.) *Learning from High Reliability Organisations.* Sydney: CCH.

Hayes, J. & Maslen, S. 2015. Knowing stories that matter: Learning for effective safety decision-making. *Journal of Risk Research,* 18, 714–726.

Hayes, J., Maslen, S., Scott-Young, C. & Wong, J. 2017. The rise of defensive engineering: How personal liability considerations impact decision-making. *Journal of Risk Research,* DOI: 10.1080/13669877.2017.1391319.

Hopkins, A. 2009. Identifying and responding to warnings. In Hopkins, A. (ed.) *Learning from High Reliability Organisations.* Sydney: CCH.

Kermisch, C. 2012. Risk and responsibility: A complex and evolving relationship. *Science and Engineering Ethics,* 18, 91–102.

Khatri, N., Brown, G.D. & Hicks, L.L. 2009. From a blame culture to a just culture in health care. *Health Care Management Review,* 34, 312–322.

Kines, P., Lappalainen, J., Mikkelsen, K.L., Olsen, E., Pousette, A., Tharaldsen, J. & Törner, M. 2011. Nordic safety climate questionnaire (NOSACQ-50): A new tool for diagnosing occupational safety climate. *International Journal of Industrial Ergonomics,* 41, 634–646.

Macrae, C. 2009. From risk to resilience: Assessing flight safety incidents in airlines. In Hopkins, A. (ed.) *Learning from High Reliability Organisations.* Sydney: CCH.

Maslen, S. 2014. Learning to prevent disaster: An investigation into methods for building safety knowledge among new engineers to the Australian gas pipeline industry. *Safety Science,* 64, 82–89.

Maslen, S. & Hayes, J. 2016. Preventing black swans: Incident reporting systems as collective knowledge management. *Journal of Risk Research,* 19, 1246–1260.

Maslen, S. & Hopkins, A. 2014. Do incentives work? A qualitative study of managers' motivations in hazardous industries. *Safety Science,* 70, 419–428.

Pidgeon, N. & O'Leary, M. 2000. Man-made disasters: Why technology and organizations (sometimes) fail. *Safety Science,* 34, 15–30.

Reason, J. 1997. *Managing the risks of organizational accidents*, Aldershot: Ashgate.

Shirali, G.A., Mohammadfam, I. & Ebrahimipour, V. 2013. A new method For quantitative assessment of resilience engineering by PCA And NT approach: A case study in a process industry. *Reliability Engineering & System Safety,* 119, 88–94.

Snook, S.A. 2000. *Friendly fire: The accidental shootdown of US Black Hawks over Northern Iraq*, Princeton: Princeton University Press.

Størseth, F. & Tinmannsvik, R.K. 2012. The critical reaction: Learning from accidents. *Safety Science,* 50, 1977–1982.

Vastveit, K.R., Boin, A. & Njå, O. 2015. Learning from incidents: Practices at a Scandinavian refinery. *Safety Science,* 79, 80–87.

Weick, K.E. & Sutcliffe, K.M. 2001. *Managing the unexpected: Assuring high performance in an age of complexity*, San Francisco: Jossey-Bass.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

On the level of safety knowledge in the general public

G. Baldissone, M. Demichela, L. Comberti & E. Pilone
Department of Applied Science and Technology, Politecnico di Torino, Torino, Italy

J. Geng
College of the Quality and Safety Engineering, China Jiliang University, Hangzhou, Zhejiang, China

L. Maida
Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Torino, China

ABSTRACT: Following the implementation of international and national regulations and standards pertaining the control of risks in different industrial and civil domains, a general increase of the safety knowledge was noticed and recognized in literature (e.g. Lee & Harrison, 2000). Indeed, different campaigns to rise the awareness of the general public were developed and implemented. However, even if numerous studies and researches measured the safety culture in occupational domains (Choudhry et al., 2007), a measurement of the safety culture in the civil society is far to be defined. The present paper shows an attempt of measuring the public safety knowledge: a simple methodology based on gaming (safety quiz related to darts playing) was developed to collect data on the safety knowledge in the population, both for children and for adults. Different sets of questions were established for children (aged from 4 to 15) and adults. The quiz was proposed during the European Researchers' Night in 2015, 2016 in Turin (Italy), collecting about 250 replies. The present paper presents the analysis of the data collected, together with some observations both on the diffusion of the safety culture in the general public and on the possible improvement of the data collection approach.

1 INTRODUCTION

Today, the occupational incidents are an economical and social problem (Hamalainen et al., 2006). I.e., in 2004 the cost of the occupational accidents in the European Union was around 35,000,000,000 €, with 3.2% of workers involved (Battaglia et al., 2014). For this reason, since 1989, UE promoted policies to reduce the occupational accidents, as the Council Directive 89/391/EEC (1989).

Between 1994 and the 2012, ESAW data (European Statistics on Accidents at Work) showed a decrease trend in the occupational accidents, however in 2013 (Eurostat, 2013) registered 3647 fatal accidents and 3,100,000 non-fatal accidents, which cannot be considered negligible values. Lehtola et al. (2008) explained the decreasing trend as the product of multiple factors, such as the regulation change, the demographical change (Morse et al., 2009) or the deindustrialization process (Loomis et al., 2004).

The data collected in Italy confirm the European trend (Figure 1). In Italy, the decrease trend started around 50 years ago, and accidents occurring demonstrated to be strictly correlated with the socio-economical evolution in Italy (Comberti et al. 2017).

The numbers of domestic and non-work related accidents are more difficult to be defined; however,

Figure 1. Non-fatal occupation accidents in Italy 1951–2005.

INAIL – Italian institution for insurance against accidents at work evaluate the Italian domestic accidents in around 3,000,000 (Ferrante et al. 2012), with a stable trend in the decade 1998–2008 (Figure 2). This value highlights the socio-economic importance of domestic accidents, most of all considering that they can involve also children.

Different accidents, both occupational and domestic, are caused by the lack of knowledge on the basic safety rules. For the occupational safety, the Italian Legislative Decree 81/08 required the

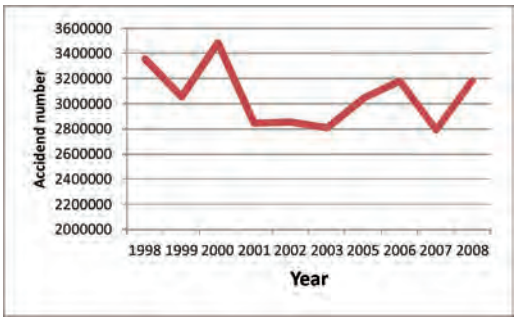


Figure 2. Domestic accident in Italy.

employers to develop mandatory safety training sections for the workers. At the end of each safety training, the workers' acquired knowledge is tested. However, since the test is not repeated after a while, it is not possible to verify if the workers maintain, apply and consolidate the lessons learned.

Literature provides several studies on the safety knowledge (Choudhry et al., 2007) in specific working sectors such as: Australian manufacturing and mining sector (Griffin & Neal, 2000); Iranian meat processing (Ansari-Lari et al., 2010); medical sector (Kolade 2002); nuclear power plant (Lee & Harrison, 2000), etc. However, scarce studies are available on the safety knowledge among the population: i.e., Kennedy et al. (2005) analyzed safety knowledge of the customers about food management.

In order to increase the safety knowledge and safety culture of the population, it is fundamental to start teaching the basis of the safety rules to children. For this reason, some tools were promoted and developed by EU and the member states, such as "Napo for teachers" (napofilm.net, 2017), or in Italy, the projects "Scuola e sicurezza – School and safety" (Dettoni, 2011), applied in Piedmont region, and "A scuola di sicurezza! – At safety school!" (Bove et al., 2002), applied in Lombardy region. Testing the safety knowledge acquired by the children can provide interesting information on the proper functioning of these programs.

This paper presents the data on the safety knowledge of a sample of Turin population: both adults and children, were tested through the participation to a simple game, in order to verify the continuity and solidity of the safety information acquired at school, at work, or through the media.

The data were collected during the "European Researchers' Night" in Turin, in 2015 and 2016; 100 stands and more than 1000 researchers were present (CLOSER, 2016). This choice allowed to attract and test a large number of people, representing a wide range of age, gender and work conditions. In the end, 115 adults (between 10 to 73 years old) and 132 children (between 5 to 15 years old) participated to the safety game, providing meaningful indications on the safety knowledge levels of the civil society.

2 MATERIAL AND METHODS

The data collection was carried out through a game, developed and then applied by the researchers of the SAFER team of Politecnico di Torino (Centro Studi su Sicurezza Affidabilità e Rischi), led by prof. Micaela Demichela. The game was part of an entire stand dedicated to safety, aimed at disseminating safety rules and showing the result of research in the safety field. For attract more participant at the data collection was used a gaming approach. During the game at the participant are proposed 3 questions on the safety. The questions proposed at each participant are chosen in the random way. And the answers are used for test the safety knowledge in the population.

2.1 *Gaming*

The data collection was structured like a game with the aim of attracting more participants, stimulating their competitiveness and their willingness to get involved. The game consisted in a simple darts-playing; depending on the score made on the dartboard, the participant is subjected to questions related to safety, with an increasing level of difficulty.

The 120 questions, with 5 different categories of difficulty, were prepared by the research team, but only 3 questions were asked to each participant, chosen in a random way. The participants could play alone or proceed with a sort of "race", verifying who could provide more right answers.

The questions tested the knowledge of the participants; the collected answers constituted a useful sample of data on the diffusion and level of safety awareness of the population. In particular, the number of correct and incorrect answers was analyzed. The safety game and the related data collection also gave the possibility to the researchers to carry out a disseminating activity about safety, in particular in case of incorrect answers.

2.2 *Questions*

The questions allow a multiple-choice: 3 possible choices are provided, only one is correct. 2 groups of questions were prepared on the basis of the age of the participants:

- Junior (less than 15 years old);
- Senior (over 10 years old).

The 2 categories present a common age range, between 10 to 15 years old. This is due to the fact that the questions in the category "Junior" were designed for younger children, centered in the primary school age, while the questions of the "Senior" category were more work-oriented (even

if not only regarding safety in the workplace). Therefore, the participants between 10 to 15 years old could find the "Junior" questions too easy, and the "Senior" questions too difficult. In the end, it was decided to let the children choose, before starting the game, if answer to "Senior" or "Junior" questions. Many children decided to answer "Senior" questions, demonstrating to have in any case an adequate knowledge to correctly deal with them.

For each participant, were collected:

- General data for statistical purposes (gender, age);
- ID code of the question proposed;
- Answer obtained.

The data were collected in a dedicated form in anonymous way.

2.2.1 *Senior questions*

Around 70 questions were developed for the "Senior" categories, covering different safety fields, such as:

- Safety rules in the workspace;
- Personal Protective Equipment and its use;
- Properties of chemical substances and their classification;
- Safety signals and symbols;
- Other safety fields.

An example of "Senior" question is:
"When you work on computer, how far from you is the monitor?

A. 35–50 cm;
B. 25–40 cm;
C. 50–70 cm."

2.2.2 *Junior questions*

Around 40 questions were developed for "Junior" category, to verify the awareness of the children in relation to:

- Safety rules in the school;
- Road safety;
- Safety during free time (sport, game, …);
- Rudiments of emergency management (emergency number, behavior in case of earthquake, …);
- Other safety fields.

The questions present simple scenarios and the children have to choose a possible solution. An example is: "You are at a friend's house and you are playing together. Since you are thirsty, you go in the kitchen and you find a bottle without labels. What are you going to do?

A. Drink;
B. Open the bottle, sniff and then drink;
C. Drink from the tap."

3 RESULTS

247 people from 4 years old to 73 years old were involved in the data collection between 2015 and 2016; 115 people were tested in the "Senior" category; while 132 children were tested in the "Junior" one.

The following tables and figures represent gender and age distributions for the "Senior" category (Table 1 and Figure 3) and "Junior" category (Table 2 and Table 3); for adults, it resulted that the majority of participants was younger than 25 years old and 55% was male. For the children, the gender of the participants was equally distributed, and the majority was between 7 and 12 years old.

3.1.1 *"Senior" category results*

The 61.3% of the people involved in the "Senior" category provided corrected answers. The global

Table 1. Gender distribution in "Senior" category.

	Participants		
	Total	Male	Female
2015	40	22	18
2016	75	41	34
Total	115	63	52

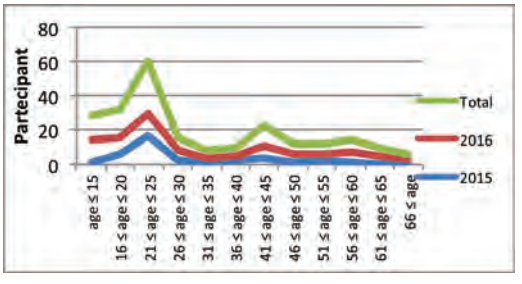


Figure 3. Age distribution in the "Senior" category.

Table 2. Gender distribution in "Junior" category.

	Participants		
	Total	Male	Female
2015	74	36	38
2016	58	30	28
Total	132	66	66

Table 3. Age distribution in "Junior" category.

	Participants			
	age ≤ 6	7 ≤ age ≤ 9	10 ≤ age ≤ 12	13 ≤ age ≤ 15
2015	8	24	40	2
2016	5	20	26	7
Total	13	44	66	9

value of correct answers was almost constant in 2015 (61.9%) and 2016 (61.2%) and apparently there is no difference between male and female, as shown by the last bars of Figure 4. However, in 2015 women gave more correct answers than men, while in 2016 the opposite trend occurred. Figure 5 shows the percentage of correct answers on the basis of the age of the participants: the fraction of correct answers is higher (around or over 60%) for participant younger than 45 years old. A negative peak is registered after 45 years old, and then the correct answers increase again. For some age groups, the uncertainty boundary is higher: give the small number of participants, wrong answers assume a high impact on the results. On the contrary, when the number of participants is high for both the years analyzed, the uncertainty is lower.

3.1.2 *"Junior" category results*

The 70.3% of the children provided correct answers, but the percentage of correct answer was higher in 2015 (72.9%) than in 2016 (67.1%).

About the gender difference, in the "Junior" category, girls gave more correct answers than boys, a trend particularly evident in 2015 (Figure 6).

Concerning the correlation with the age distribution, Figure 7 shows that the correct answer trend grows with the age of the participants.

Figure 4. Correct answer in the "Senior" category divided for gender.

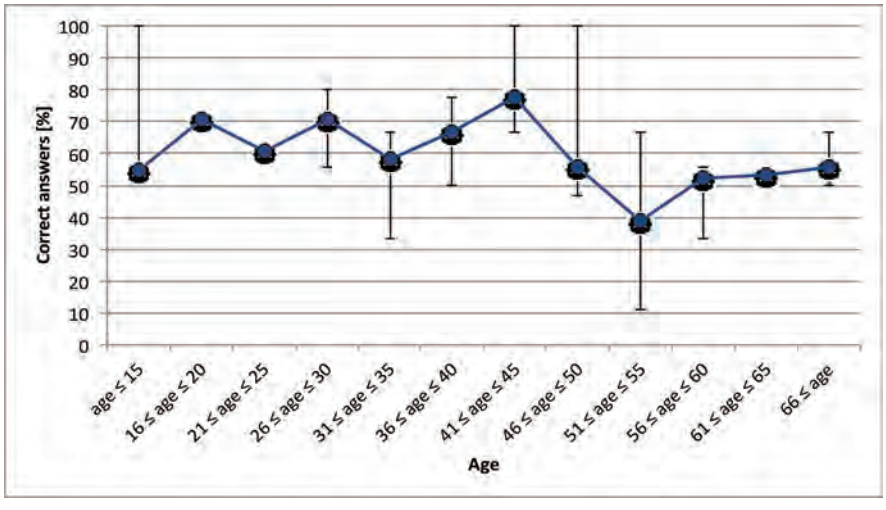


Figure 5. Trend of correct answer by age of participants, in the "Senior" category.

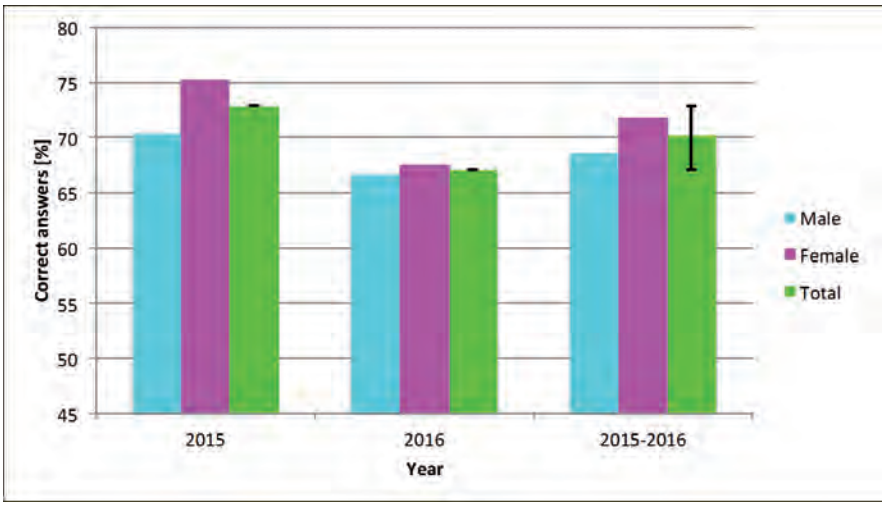


Figure 6. Correct answer in the "Junior" category divided for gender.

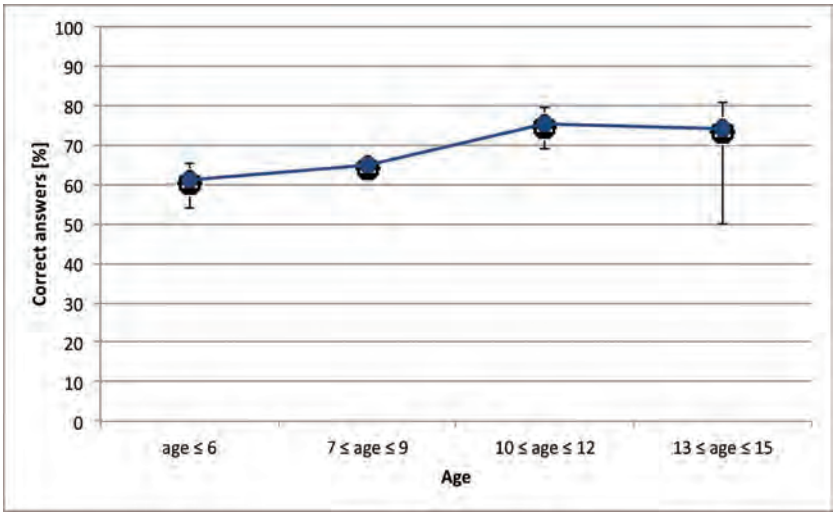


Figure 7. Trend of correct answer by age of participants, in the "Junior" category.

4 DISCUSSION

On the basis of the data collected and presented in the previously paragraphs, some consideration can be done.

For the adults, the mean values obtained from the test underline that the population know the basis of safety knowledge; but the level needs to be increased to reach a sufficient quality and diffusion of knowledge.

The data show no gender gap in the safety knowledge, while the age trend gives some encouraging signal for the future. In fact, the younger participants in the "senior" category gave more correct answers than the older ones. This result probably evidences the importance of the several activities of sensitization about safety conducted in the last 20 years during the school period (primary, secondary and high school). On the contrary, the worse results for the older participants mark the resistance in learning and adapting to safety rules, a behvaoiur that is still quite common in the Italian culture.

In the "Junior" category, the mean value of safety knowledge is higher than for "Senior", reaching values in line with those of the younger "Senior" participants. This good result probably underlines the effect of the safety sensitization campaign done during school age; it represents an

encouraging signal of a possible radical change in the culture of workers and citizens, but this trend has obviously to be monitored in the future. At the moment, the answers given by the younger "Senior" participants seem to confirm this increased sensitivity of the younger generations to the themes of the safety culture.

Finally, the last interesting date emerged from the analysis is related to gender: "Junior" females showed a more in-depth safety knowledge than males. Further investigations should be executed to strengthen the definition of this trend, but in case of confirmation this could be the basis for a multidisciplinary investigation on possible cultural differences in learning mechanisms.

5 CONCLUSION

In this paper, a method for data collection on safety knowledge of the population is presented. A gaming approach was settled to attract and involve more participants: series of multiple—choice questions were developed and exposed to the users as a quiz. The game was applied during the "European Researchers' Night" in Turin in 2015 and 2016, with different questions for adults and children.

The data collected showed that the adults have a discrete level of safety knowledge, and that younger participants have an in-depth safety knowledge than the older ones. Children demonstrated to have a good safety knowledge level.

The data collected prove that the safety-related projects applied in schools are producing good results on the population, but it is fundamental to continue with the activity of training both for young and older people. In particular for the latter, it should be essential to recover the gap in knowledge for people between 50 and 60 years old; in fact, these persons could be in influent positions inside their companies and jobs, therefore their awareness and knowledge in relation to safety themes has a wider influence than the foreseen.

ACKNOWLEDGEMENT

Special thanks to the organization of the "European Researchers' Night" in Turin for the fundamental and excellent job that allowed the data collection; and most of all, to all the participants for their collaboration.

REFERENCES

Ansari-Lari M., Soodbakhsh S., Lakzadeh L., 2010. Knowledge, attitudes and practices of workers on food hygienic practices in meat processing plants in Fars, Iran. *Food Control* 21(3). 260–263.

Battaglia M., Frey M., Passetti E. 2014. Accidents at work and costs analysis: a field study in a large Italian company. *Industrial Health* 52. 354–366.

Bove E., Rocca G., Peracchi M., 2002. A scuola di sicurezza! Milano, Regione Lombardia.

Choudhry R., Fang D., Sherif M., 2007. The nature of safety culture: A survey of the state-of-the-art. *Safety Science.*

CLOSER 2016, http://nottedeiricercatori.piemontevalledaosta.it/2016/ access 29/11/2017.

Comberti L., Baldissone G. Demichela M., Patrucco M, Maida L., 2017. Investigation on the impact of National regulations on the occupational safety: 45(10). 993–1012.

Council Directive 89/391/EEC of 12 June 1989 on the introduction of measures to encourage improvements in the safety and health of workers at work. *In ESEREL 2017 Portoroz, Slovenia, 18–22 June, 2017.*

Decreto Legislativo 9 aprile 2008, n. 81 *"Attuazione dell'articolo 1 della legge 3 agosto 2007, n. 123, in materia di tutela della salute e della sicurezza nei luoghi di lavoro".*

Dettoni L. 2011. Scuola e sicurezza: dall'esperienza di un lavoro in rete raccomandazioni pratiche e support della progettazione. Pienerolo: Centro Regionale di Documentazione per la Promozione della Salute.

Eurostat 2013. European social statistics. European Commision. Luxembourg.

Griffin M.A., Neal A., 2000. Perceptions of safety at work: a framework for linking safety climate to safety performance, knowledge, and motivation. *Journal of Occupational Health Psychology* 5(3). 347–358.

Ferrante P., Massari S., Buresti G., Iavicoli S., 2012. *Infortuni Domestici: epidemiologia del fenomeno ed approfondimenti sulla popolazione infortunata*. Milano INAIL.

Hamalainen P., Takala J., Saarela K.L. 2006. Global estimates of occupational accidents. *Safety Science* 44(2). 137–156.

Kennedy J., Jackson V., Blair I.S, McDowell D.A., Cowan C., Bolton D., 2005. Food Safety Knowledge of Consumers and the Microbiological and Temperature Status of Their Refrigerators. *Journal of Food Protection* 68(7). 1421–1430.

Kolade E.S. 2002. Injection safety: knowledge and practice among health workers. *West African journal of medicine* 21(1). 70–73.

Lehtola M.M., van der Molen H.F., Lappalainen J., Hoonakker P.L., Hsiao H., Haslam R.A. Hale A.R., Verbeek J.H. 2008, The Effectiveness of Interventions for Preventing Injuries in the Construction Industry: A Systematic Review. *American Journal of Preventive Medicine* 35(1). 77–85.

Lee T., Harrison K., 2000. Assessing safety culture in nuclear power stations. *Safety Science* 34(1–3). 61–97.

Loomis D., Richardson D.B., Bena J.F., Bailer A.J. 2004. Deindustrialisation and the long term decline in fatal occupational injuries. *Occupational & Environmental Medicine* 61(7). 616–621.

Morse T.F., Deloreto A., Louis T.S., Meyer J.D. 2009. Are employment shifts into non-manufacturing industries partially responsible for the decline in occupational injury rates? *American journal of industrial medicine* 52(10). 735–741.

Nepofilm.net 2017, https://www.napofilm.net/en/using-napo/napo-for-teachers, access 24/11/2017.

Notte dei Ricercatori 2015, http://piemonte.nottedeiricercatori.it/index.php# access 29/11/2017.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Safety climate and compliance in the Norwegian aquaculture industry—employees' perceptions at different company levels

T.Ø. Kongsvik
Department of Industrial Economics and Technology Management, NTNU, Norway

T. Thorvaldsen & I.M. Holmen
SINTEF Ocean, Norway

K.V. Størkersen
NTNU Social Research, Norway

ABSTRACT: The aquaculture industry is economically important in Norway, and the production is expected to increase in the future. Employees at the fish farms face a high risk of accidents compared to employees in other industries and the focus on safety from both industry and researchers has increased during the last decade. Adding to the knowledge on safety in aquaculture, the objective of this paper is to study employees' perception of safety climate, and whether aspects related to safety climate may predict employees' compliance. Findings from two surveys aimed at managers and employees in different companies are analysed. The first is a telephone survey targeting employees and managers at the fish farms. The second is a web-based survey involving the onshore management level. The results show that employees at all levels have a positive perception of the safety climate, but they also illustrate challenges related to work pressure, maintenance and employee participation. Furthermore, the analysis shows that work pressure affects compliance negatively while participation and competence have positive associations with compliance. These results give input to some practical measures for safety management in the industry.

1 INTRODUCTION

Norway is the second largest exporter of fish worldwide, and the largest producer of finfish (FAO, 2016). Since the 1970's aquaculture, and fish farming in particular, has become a significant contributor to the national value creation. In 2016, the total production in the aquaculture industry was 1,3 mill metric tons, equal to a value of 65 billion NOK (Norwegian Directorate of Fisheries, 2017). 93% of this was Atlantic salmon.

There are both large and small companies in the industry, and the structure of management levels depends on the size. Smaller companies may have personnel that serve different roles, combining the responsibility for safety with other areas such as quality management. Larger companies may have dedicated management working solely with health and safety. The lowest management level is the operational managers, who are responsible for biological production and personnel at one or two fish farms, typically manned by three to six employees (fish farmers).

There are typically six to 12 circular plastic collar net cages in one fish farm (Jensen et al., 2010, Holen et al., 2017a). The fish farm also has a feeding barge for equipment and feed storage, the feeding system, as well as manager offices, meeting rooms and accommodation for shift workers.

Fish farmers are decreed to perform daily inspections to assess fish welfare and document that the net cages are in order. Fish farmers often use boats and cranes in their work, but increasingly rely on specialized service vessel crews to perform tasks such as mooring operations and delousing. The safety for fish, material assets and personnel is regulated by an extensive set of statutory requirements (Holmen et al., 2017b), which are audited by five different regulatory authorities (Holmen et al., 2017c).

The aquaculture production has the potential to increase in the future (Olafsen et al., 2012). Technology innovations aim to enable production at areas more exposed with respect to climate, wind and currents. This raises new challenges when it comes to fish welfare and operational safety (Bjelland et al., 2015).

The attention to occupational and operational safety in the aquaculture industry has increased during the last decade. Recent studies indicate that safety and risk management systems need to

be improved to maintain a sustainable food production and ensure a safe work environment at harsher locations (Størkersen, 2012, Holmen et al., 2017b, Holmen et al., 2017c).

Being a fish farmer is the 2nd most risk exposed occupation in Norway after being a fisherman, according to the rate of fatal accidents (McGuinness et al., 2013, Holen et al., 2017b). Fall, blow by object, entanglement/crush and cuts are the most common modes of injuries in the Norwegian aquaculture industry (Holen et al., 2017a). In a recent study among aquaculture workers, three out of four respondents reported to have knowledge of near accidents (Thorvaldsen et al., 2017). Organisational aspects and safety indicators have been the topic in several studies (Fenstad et al., 2009, Størkersen, 2012, Thorvaldsen et al., 2015, Holmen et al., 2017b). Looking back, Allred et al. (2005) found that many companies had implemented some systematic HSE work, although production often was given priority.

In this article we will explore the following problems: 1. How do fish farmers, operational managers and onshore management/staff perceive the safety climate? 2. To what extent can safety climate predict safety compliance?

1.1 *Safety climate*

Safety culture has been a defined area of research since the late 1970s, and there is still much research activity related to the issue. Up until 2015, 1789 research publications related to safety culture has been published (van Nunen et al., 2017). In the beginning it was applied as a construct related to causal analyses of major accidents, with Barry Turner's (1978) contribution as an important starting point. Later, the research interests spread over different topics, and involved different disciplines, including anthropology, psychology, sociology and the engineering sciences. Today it is considered a multi-dimensional and cross-disciplinary field of research (van Nunen et al., 2017)

The contributions from psychology is in particular related to *safety climate*, a construct which is in many instances used as synonymous to safety culture and defined in similar ways (Guldenmund, 2007). A much used definition of safety climate is provided by Zohar (2003), who sees it as shared perceptions about safety policies, procedures and practices in a work community. Questionnaire surveys are commonly used to measure safety climate in a work community, involving assessments of topics such as the safety system, work pressure, the safety competence and leadership/supervision (Flin et al., 2000).

Much research on safety climate has been related to identifying causal links between safety climate as an independent variable and different safety outcomes, summed up in review studies (Clarke, 2006, Christian et al., 2009). This includes exploration of safety climate as an indicator (Kongsvik et al., 2010). The review studies also point to many studies that find a positive relationship between safety climate and safety compliance (adherence to safety instructions, rules, and procedures).

A general finding in the research is that the safety climate in a work community is associated with the work practices in the same community; a positive safety climate is related to compliance and participation in safety-promoting activities, and also mindful safety practices (Dahl and Kongsvik, 2018).

2 METHOD

2.1 *The surveys*

The results are based on two different surveys, that were later combined. The first included fish farmers, operational managers and service vessel crew members as well as some employees in other positions. Representatives at the management level in a selection of 40 companies were informed about the aim of the survey, and invited to share employees' phone numbers. A professional polling company conducted the survey, and a total of 447 out of 735 employees participated. Here, answers from 258 fish farmers and 110 operational managers are used.

The second survey included onshore management and staff. Companies were contacted and asked to provide e-mail addresses to their employees who then got an e-mail linking to the digital survey. Some companies distributed the link to their employees themselves, so the response rate cannot be estimated. A total of 135 persons responded. Here, the net sample includes 92 onshore managers or staff.

The questionnaires were developed on basis of earlier surveys, but were tailor made to the aquaculture industry. In both surveys, the respondents were asked to state their agreement to different statements related to safety climate, on a 5-point Likert scale, ranging from totally disagree to totally agree.

2.2 *Analyses*

Answers from both surveys were extracted and combined in one data file. Some items had divergent wording, mirroring their position.

The items common for the two samples were thematically sorted into four categories: Work pressure, Participation, Competence and resources and Compliance. The first three categories are directly related to the safety climate construct, while Compliance is related to safety practices and whether employees live up to requirements given by the companies.

The statistical analysis aimed at comparing the perceptions of three groups, involving comparing means related to the responses. One-way ANOVA was applied for comparing the means. The limit for statistical significance (P-value) was set to 1%.

We performed a multiple regression analysis to explore if compliance could be predicted by the safety climate factors. The analysis was restricted to the fish farmers (n = 258) to ensure homogeneity in the work situation for those included. A *Compliance scale* was constructed by combining three items: 1. I use the required protective equipment 2. If I see dangerous situations at work, I report them. 3. Safety has first priority when I do my job. The Cronbach's alpha for the scale was .71. The *Work pressure scale* consisted of three items: 1. Sometimes I feel a pressure to continue working, although safety can be compromised 2. In practice, consideration to production is prioritized at the expense of safety 3. Inadequate maintenance has reduced the safety level. The Cronbach's alpha for the scale was .67. The *Participation scale* also included three items: 1.I participate in making new procedures 2. I get involved when new procedures are to be introduced 3. My manager appreciates that the employees take up safety issues. The Cronbach's alpha for the scale was .70. The *Competence scale* included two items: 1. I have the necessary competence to handle my work tasks safely. 2. I have received the necessary training for handling critical or dangerous situations. The Cronbach's alpha for the scale was .60.

The independent variables in the regression analysis were introduced by forced entry. Missing values were excluded list wise. The variance inflation factor (VIF) varied between 1.141 and 1.276, and tolerance values varied between 0.783 and 0.876. This gives no indications of multicollinearity. To check the assumption of independent errors, the Durbin–Watson test was performed. This test shows there was no concern regarding autocorrelation, with a test statistic of 2.063 (Field, 2005).

3 RESULTS

Here, the mean values reported by the respondents of the surveys are compared. Results are presented by the three following groups: onshore management (M), operational managers (OM) and fish farmers (F).

3.1 *Perceived safety climate*

The safety climate can be expressed by the survey results about work pressure; competence and resources; and compliance.

3.1.1 *Work pressure*

Table 1 reports the results from three items considering work pressure.

The first item is phrased differently for operational and onshore personnel. Yet, all groups somewhat disagree that production pressure makes operational personnel break safety rules and continue unsafe work. Onshore management agree more than operational personnel that employees will compromise on safety because of production pressure.

On the next item, this controversy is partly reversed. On average, fish farmers neither agree nor disagree that production is prioritized over safety, while both types of managers disagree more. Still, analysis show that 22.9% of the fish farmers and 13.6% of the operational managers agree or totally agree that consideration to production is prioritized at the expense of safety.

All three groups' mean values show they somewhat disagree that inadequate maintenance has reduced the safety level. 19.8% of the fish farmers and 18.2% of the operational managers agree or totally agree that safety has been reduced due to inadequate maintenance.

3.1.2 *Participation*

Table 2 includes three items about employee participation.

All groups agree that managers appreciate employees' safety engagement. Onshore management appreciate that employees take up safety issues more than considered by the employees.

Table 1. Perceptions of Work pressure – means on a scale from 1 (totally disagree) to 5 (totally agree).

Items	Groups	Mean	P-value
F/OM: Sometimes I feel a pressure to continue working, although safety can be compromised	Fish farmers	2.02	0.000
	Operational managers	2.16	
	Onshore management		
M: Owing to the company's production demands, the employees sometimes have to break the safety rules		2.63	
In practice, consideration to production is prioritized at the expense of safety	Fish farmers	2.56	0.001
	Operational management	2.09	
	Onshore management	2.19	
Inadequate maintenance has reduced the safety level	Fish farmers	2.40	0.609
	Operational managers	2.30	(NS)
	Onshore management	2.47	

Table 2. Perceptions of Participation – means on a scale from 1 (totally disagree) to 5 (totally agree).

		Mean	P-value
F/OM: My manager appreciates that the employees take up safety issues	Fish farmers	4.20	0.000
	Operational managers	4.32	
M: As manager, I appreciate that the employees take up safety issues	Onshore management	4.92	
F/OM: I participate in making new procedures	Fish farmers	2.77	0.000
	Operational managers	3.54	
M: The employees participate in making new procedures	On shore management	3.61	
F/OM: I get involved when new procedures are to be introduced	Fish farmers	3.35	0.006
	Operational managers	3.73	
M: The employees get involved when new procedures are to be introduced	On shore management	3.76	

The mean results about employees' participation in development of new procedures lies around neither/nor. Fish farmers lean towards disagreement, while both management groups agree somewhat that employees (including the operational managers themselves) participate in making procedures.

Management agree that employees get involved in introduction of procedures, Analysis shows that 24.4% of the fish farmers disagree or totally disagree that they are involved when new procedures are introduced, compared to 19.1% of the operational managers.

3.1.3 *Competence and resources*

The three items considering Competence and resources are presented in Table 3.

All groups agree that the operational personnel have the competence to work safely and that manning is sufficient for safe work. It is 12.4% of the fish farmers and 12.7% of the operational managers that disagree or totally disagree that manning was sufficient.

There are larger differences regarding learning from unwanted events. Fish farmers and operational managers agree that information regarding unwanted events is utilized to prevent recurrence, while onshore management state neither/nor.

3.1.4 *Compliance*

All respondent groups' in average *agree* regarding the three items about compliance, but there are some differences (Table 4).

Table 3. Perceptions of Competence and resources – means on a scale from 1 (totally disagree) to 5 (totally agree).

		Mean	P-value
F/OM: I have the necessary competence to handle my work tasks safely	Fish farmers	4.49	0.578 (NS)
	Operational managers	4.56	
M: Our employees have the necessary competence to handle their work tasks safely.	Onshore management	4.47	
The manning is sufficient to maintain the safety	Fish farmers	3.70	0.235 (NS)
	Operational managers	3.62	
	Onshore management/ staff	3.87	
F/OM: Information regarding unwanted events is utilized adequately to prevent recurrence	Fish farmers	4.11	0.000
	Operational managers	4.13	
M: We utilize the information from reported unwanted events sufficiently in the preventive work	Onshore management	2.91	

The analysis displays no significant differences regarding perceptions of reporting of dangerous situations. However, 31.4% of the fish farmers answered that they agreed or totally agreed with the statement: *I think it is uncomfortable to point out lack of compliance to safety rules and procedures*. 16.4% of the operational managers said the same. Onshore management were not asked about this aspect, as it relates to the operational context and the everyday interaction between workers at the fish farms.

Regarding use of protective equipment operational personnel totally agree that they use it, while onshore staff only agree.

Operational personnel agree (4.30), while onshore management almost totally agree (4.71) that safety has the first priority.

3.2 *Safety climate's relation to compliance*

A linear regression analysis restricted to the fish farmers in the sample was performed. Work pressure, Participation and Competence were used to predict Compliance.

Work pressure, Participation and Competence explained 30.4% of the variance in Compliance ($p < 0.000$). Each of the predictor variables had a significant contribution to the model.

Table 4. Perceptions of Compliance – means on a scale from 1 (totally disagree) to 5 (totally agree).

		Mean	P-value
F/OM: If I see dangerous situations at work, I report them.	Fish farmers	4.38	0.209 (NS)
	Operational managers	4.54	
M: The employees use the reporting system adequately when it comes to personal injuries and other serious events.	Onshore management/ staff	4.32	
F/OM: I use the required protective equipment	Fish farmers	4.55	0.000
	Operational managers	4.72	
M: Our employees always use the required protective equipment	Onshore management	4.20	
Safety has first priority when I do my job	Fish farmers	4.30	0.000
	Operational managers	4.30	
	Onshore management	4.71	

Table 5. Linear regression analysis predicting 'Compliance': Beta-values (B), standard errors (SE B), standardised betas (β) and explained variance (Adjusted R^2) (N = 253).

	B	SE B	B	Adjusted R^2
Constant	2,086	0.301		0.304
Work pressure	–0.150	0.044	–0.197	
Participation	0.112	0.043	0.156	
Competence	0.366	0.054	0.381	

Work pressure had a negative association with Compliance, while Participation and Competence had a positive association. High perceived work pressure was thus associated with lower Compliance, while higher degrees of participation and Competence, was associated with higher Compliance.

4 DISCUSSION

Considering the results, variations in perceptions of safety climate related to the three different groups and company levels are discussed.

4.1 *Regarding work pressure*

All groups recognize some pressure to continue unsafe operations or violate rules and prioritize production over safety (two first items, Table 1). Almost ¼ of the fish farmers agree that production sometimes trumps safety. Unsurprisingly, managers disagree more as they often have a more positive (Fenstad et al., 2009) or less realistic view of operations (Hale and Borys, 2013b, Hollnagel, 2011). When considering if *production pressure will make employees compromise on safety,* managers agree more than operational personnel do. One explanation may be that management have HSE as their area of expertise, and might have learned, in practice and theoretically, that operational personnel will feel pressured to work efficient and not thorough (Hollnagel, 2009) or prioritize production over protection (Reason, 1997) with a drift towards unsafe performance (Dekker, 2011, Vaughan, 1997, Rasmussen, 1997).

This resonates with earlier findings in Norwegian aquaculture. In Allred et al. (2005), 21% agreed that they were pressured to work in a way that could threaten safety. In addition, 27% agreed that the operational manager did not have the time to sufficiently manage employees' HSE – meaning that they focused on production to make the ends meet (ibid). Priority of the biological product, the fish, is also qualitatively described (Fenstad et al., 2009, Størkersen, 2012).

A more recent study found that work pressure may be caused by poor planning of operations, insufficient staffing, time pressure and working long hours (Thorvaldsen et al. 2015). And even though their personal safety may be threatened, employees are very conscious about following up on their work responsibilities.

An increased efficiency pressure coincides with the statements about maintenance (Table 1). 1/5 of the fish farmers and operational managers think that inadequate maintenance has reduced the safety level. Maintenance on existing equipment might be postponed to times with less activity and potential earnings. It is shown that coastal vessels running for the prosperous aquaculture industry have tighter schedules and more stress than vessels operating in slower markets (Størkersen, 2017).

Economic priorities are relevant here, as it may affect maintenance of existing equipment at a given fish farm or vessel, but also limit investments in new technology (Thorvaldsen et al. 2015). A previous study also found that fish farmers experienced increased profit as more important than workers' safety (Fenstad et al. 2009).

4.2 *Regarding participation*

Personnel on all levels agree that managers appreciate employees' safety engagement. However, mangers seem to appreciate employees taking up safety issues more than fish farmers think. The interaction between company levels as well as differences between formal and informal communica-

tion may be reflected here. Fish farmers may take up safety issues with the operational managers, who then decides whether and how to address a specific issue. In larger companies, formal reporting systems are also used. As operational managers give a higher score than fish farmers, fish farmers may think about day-to-day interaction when they disagree with the perception of the operational managers, or they may be thinking about the (lack of) response they get when they address safety issues through formal reporting systems.

Allred et al. (2005) found that 81% agreed that employees could influence the HSE conditions to a large degree. Still, a third of the respondents said they did not have the chance to participate in HSE strategies.

When it comes to new procedures, and employees' participation in developing them, the answers gave a neither/nor result. Fish farmers disagree more than the managers. Here, a likely explanation will be that the onshore managers involve the operational mangers and some of the fish farmers when procedures are made. Even though not all fish farmers are involved in such processes the onshore managers do involve some of the employees.

The situation seems to be somewhat similar when it comes to new procedures. Almost 25% of the fish farmers and 19% of the operational managers disagree. As procedures are often sent to employees via computer-based systems, this answer may indicate that this is not seen as involvement, but rather as information.

4.3 *Regarding competence and resources*

Over all, the participants agree that operational personnel at the fish farms have the competence to work safely. Several companies also provide external safety courses for employees, and companies use internal procedures to document how operations should be performed. Experience is also highly valued at the fish farms (Holmen et al., 2017a, Thorvaldsen et al., 2015).

In 2005, 68% stated that employees got adequate safety training, but 41–44% wanted more training in sickness and injury preventing work and safety routines (Allred et al., 2005).

When it comes to staffing, about 12% of both fish farmers and operational mangers disagree that manning is sufficient. Safety issues when workload is increased without additional personnel has been discussed in previous studies in aquaculture (Thorvaldsen et al. 2015) and maritime industries (Hetherington et al., 2006, Österman and Hult, 2016).

Fish farmers and operational managers agree that they use information from previous events for prevention, but onshore staff answers neither/nor. It may be that onshore management see more potential for learning and prevention both on their part and from the operational personnel. Onshore management must follow up on non-compliance reports from many fish farms, and may lack time and resources do to this optimally. The operational personnel on the other hand, may answer more positively based on activities and measures on their specific fish farm, and not what comes from the onshore management. With more formalized systems for reporting, one might think that this was an area that had improved a lot during the last decade. Looking back, however, 72% of the participants in 2005 (Allred et al.) answered that information about accidents and unwanted events was actively used by the companies.

4.4 *Regarding compliance*

All groups agree they report dangerous situations (the first item, Table 5), but a third of the fish farmers also are uncomfortable pointing out non-compliance. Not every company meets deviance from rules with an understanding that most personnel follow rules, but that some rules can be difficult to follow because they are contradictory to other rules, the context, or resources (March, 1994). It may even be necessary to break a rule to get the job done (Reason, 1990). Thus, the safety literature has emphasized that safety is not attained by blind rule-following (Hollnagel et al., 2006, Hale and Borys, 2013b). Still, compliance might be the safest option if the rules can be followed. A literature review of quantitative studies indicates "a positive linear relationship between safety compliance and safety. That is, the more compliance the better for the state of safety" (Dahl, 2014: 31). In the study of Allred et al. (2005), the findings were at least as positive as in the current study: 65% stated that employees always reported safety issues and dangerous conditions, and ¾ meant operational manager encouraged employees to report such conditions.

Both the current and previous studies (Allred et al., 2005) find that protective equipment mostly is used.

All groups prioritize safety in most situations, the management respondents are most certain. This is related to the statement that the production is prioritized over safety in some operations, although management disagree that production is prioritized over safety (Table 1). Management is commonly looser coupled to the negotiations in the operations. A relevant point here, is that our survey has targeted management respondents with health and safety as their responsibility. Furthermore, as Allred et al. (2005) also discuss, we do not know if it is only the most positive representatives working in the most HSE focused companies who have answered the surveys (Hollnagel et al., 2006, Hale and Borys, 2013).

4.5 *Safety climate and compliance*

The regression analysis revealed that safety compliant behaviour was predicted by safety climate

measures among fish farmers. In our context, compliant behaviour involves adherence to safety rules, and procedures, such as the use of the required protective equipment, reporting of dangerous situations if they are observed, and prioritizing safety when they do their job.

Among the three safety climate factors, Competence was the most important predictor, followed by Work pressure and Participation. Competence and Participation were positively related to Compliance, while Work pressure was negatively related to Compliance.

Several studies have shown similar results, adding up to a quite robust relationship between safety climate and safety behaviour (Clarke, 2006, Christian et al., 2009). In general, this research show that those who perceive safety as valued and prioritized in their work community, display a more positive safety behaviour, including compliant behaviour, than those who perceive safety as less valued (Dahl and Kongsvik, 2018).

This relationship has not been studied in the aquaculture industry previously. The results indicate that the relationship can be valid also in this context. This gives input to some practical measures for safety management in the industry. The competence scale included items on training, including training for emergencies. Providing such training might increase compliance through increased knowledge of the procedures. Further, avoiding work pressure that goes at the expense of safety might also increase compliance and reduce exhaustion for the individual employees. This may be a challenge, as there are some very labour-intensive periods related to delousing operations etc. Still, organizing these periods as to avoid long hours and heavy workloads might have a positive influence on compliance and safety climate. Lastly, involving employees in the construction of procedures, and having an 'open door' policy regarding safety issues can also have a positive influence on ownership, feeling of involvement and compliance.

Although safety compliance is one important aspect, it is also true that procedures and rules tend to be underspecified and cannot cover all eventualities in complex systems (Hollnagel, 2009). On the one hand, the work at fish farms include many routine tasks, where the applicability of clear procedures is evident. On the other hand, flexibility, situational awareness, practical experience, and problem-solving skills are also vital qualities in this context (Thorvaldsen et al. 2015). Consequently, rules and procedures should be dynamic, and involve sharp-end workers in formulating and evaluating them (Hale and Borys, 2013 a). So even if compliance in many instances is a basic foundation for many work operations in high-risk industries, performance variability is also a valuable asset that might increase the resilience in a sociotechnical system (Hollnagel, 2009, Haavik et al., 2017).

The results give grounds for further exploring the relationship between safety climate and safety compliance in the aquaculture industry. Future research could include onshore personnel, and suitable measures for safety climate and safety behaviour. This could broaden the view on how accidents can be prevented in the industry. Also, the cause and effect relationship between climate and safety outcomes can be explored. Other studies indicate that this relationship might be reciprocal (Kongsvik et al., 2011).

5 CONCLUSION

This article explores perceptions of safety climate at different company levels based on two surveys amongst employees in the aquaculture industry. Over all, perceptions of the safety climate are positive at all levels. This may reflect an increased focus on workers' health and safety during the last decade. Still, there are challenges related to work pressure, maintenance and employee participation. While aspects related to compliance such as reporting, wearing protective equipment and prioritizing safety get a high score, many fish farmers are uncomfortable with pointing out colleagues' lack of compliance.

The analyses further reveal that fish farmers' compliance to safety requirements is predicted by safety climate, and in particular by competence. Training, including emergency exercises, will be valuable for increased safety. The same goes for reduced work pressure. Work pressure relates negatively to compliance, and almost one quarter of the fish farmers agree that production is prioritized over safety.

Differences between company levels reflect different points of view and responsibilities within the companies. It is important that fish farmers are involved when their work procedures are created and introduced. Fish farmers and operational managers who work at the sharp end are physically closest to the occupational hazards, and rely on the onshore management to get the necessary means to mitigate the risks.

REFERENCES

Allred, K., Lie, T., Lindøe, P. & Østerhus, S. 2005. Systematisk HMS-arbeid i havbruksnæringen. RF – Rogalandsforskning.

Bjelland, H. V., Føre, M., Lader, P., Kristiansen, D., Holmen, I. M., Fredheim, A., Grøtli, E. I., Fathi, D. E., Oppedal, F., Utne, I. B. & Schjølberg, I. 2015. Exposed aquaculture in Norway: Technologies for robust operations in rough conditions. *OCEANS'15 MTS/IEEE Washington, Washington DC, 19-22 October, 2015.* IEEE conference proceedings.

Christian, M. S., Bradley, J. C., Wallace, J. C. & Burke, M. J. 2009. Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology,* 94, 1103–1127.

Clarke, S. 2006. The relationship between safety climate and safety performance: A meta-analytic review. *Journal of Occupational Health Psychology,* 11, 315–327.

Dahl, Ø. 2014. *Behind Safety Violations: Understanding the antecedents of safety-compliant behaviour in the oil and gas industry.* (Doctoral dissertation PhD), Norwegian University of Science and Technology, Trondheim, Norway. Retrieved from http://brage.bibsys.no/xmlui/handle/11250/268795.

Dahl, Ø. & Kongsvik, T. 2018. Safety climate and mindful safety practices in the oil and gas industry. *Journal of Safety Research,* 64, 29–36.

Dekker, S. 2011. *Drift into failure: from hunting broken components to understanding complex systems,* Farnham, Surrey, England, Ashgate.

FAO 2016. The State of World Fisheries and Aquaculture 2016. *In:* FAO (ed.). Rome.

Fenstad, J., Osmundsen, T. & Størkersen, K. V. 2009. *Fare på merde?: behov for endret sikkerhetsarbeid ved norske oppdrettsanlegg,* Trondheim, Norway, NTNU Samfunnsforskning.

Flin, R., Mearns, K., O'Connor, P. & Bryden, R. 2000. Measuring safety climate: identifying the common features. *Safety Science,* 34, 177–192.

Guldenmund, F. W. 2007. The use of questionnaires in safety culture research – an evaluation. *Safety Science,* 45, 723–743.

Haavik, T. K., Kongsvik, T., Bye, R. J., Dalseth Røyrvik, J. O. & Almklov, P. G. 2017. Johnny was here: From airmanship to airlineship. *Applied Ergonomics,* 59, 191–202.

Hale, A. & Borys, D. 2013a. Working to rule or working safely? Part 2: The management of safety rules and procedures. *Safety Science,* 55, 222–231.

Hale, A. R. & Borys, D. 2013b. Working to rule, or working safely. *In:* Bieder, C. & Bourrier, M. (eds.) *Trapping safety into rules. How desirable or avoidable is proceduralization.* . Farnham, United Kingdom: Ashgate.

Hetherington, C., Flin, R. & Mearns, K. 2006. Safety in shipping: The human element. *Journal of safety research,* 37, 401–411.

Holen, S. M., Utne, I. B., Holmen, I. M. & Aasjord, H. 2017a. Occupational safety in aquaculture – Part 1: Injuries in Norway. *Marine Policy,* In press.

Holen, S. M., Utne, I. B., Holmen, I. M. & Aasjord, H. 2017b. Occupational safety in aquaculture – Part 2: Fatalities in Norway 1982–2015. *Marine Policy,* In press.

Hollnagel, E. 2009. *The ETTO principle: Efficiency-Thoroughness trade-off. Why things that go right sometimes go wrong,* Farnham, Ashgate.

Hollnagel, E. 2011. *Resilience engineering in practice: A guidebook,* Farnham, United Kingdom, Ashgate.

Hollnagel, E., Woods, D. D. & Leveson, N. 2006. *Resilience engineering: Concepts and precepts,* Farnham, United Kingdom, Ashgate.

Holmen, I. M., Thorvaldsen, T. & Aarsæther, K. G. 2017a. OMAE2017-62023 Development of a Simulator Training Platform for Fish Farm Operations. *OMAE 2017.* Trondheim.

Holmen, I. M., Utne, I. B. & Haugen, S. 2017b. Organisational safety indicators in aquaculture – a preliminary study. *Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016 (Glasgow, Scotland, 25–29 September 2016).* CRC Press.

Holmen, I. M., Utne, I. B., Haugen, S. & Ratvik, I. 2017c. The status of risk assessments in Norwegian fish farming. *Safety & Reliability, Theory and Applications.* CRC Press.

Jensen, Ø., Dempster, T., Thorstad, E. B., Uglem, I. & Fredheim, A. 2010. Escapes of fishes from Norwegian sea-cage aquaculture: causes, consequences and prevention. *Aquaculture Environment Interactions,* 1, 71–83.

Kongsvik, T., Almklov, P. & Fenstad, J. 2010. Organisational safety indicators: Some conceptual considerations and a supplementary qualitative approach. *Safety Science,* 48, 1402–1411.

Kongsvik, T., Kjøs Johnsen, S. Å. & Sklet, S. 2011. Safety climate and hydrocarbon leaks: An empirical contribution to the leading-lagging indicator discussion. *Journal of Loss Prevention in the Process Industries,* 24, 405–411.

March, J. G. 1994. *Primer on decision making: How decisions happen,* New York, Simon and Schuster.

McGuinness, E., Aasjord, H. L., Utne, I. B. & Holmen, I. M. 2013. Fatalities in the Norwegian fishing fleet 1990–2011. *Safety Science,* 57, 335–351.

Norwegian Directorate of Fisheries. 2017. *Statistics for aquaculture 2016* [Online]. Available: https://www.fiskeridir.no/English/Aquaculture/Statistics [Accessed 2017-12-08].

Olafsen, T., Winther, U., Yngve, O. & Skjermo, J. 2012. Value created from productive oceans in 2050. Trondheim, Norway: SINTEF Fisheries and Aquaculture.

Rasmussen, J. 1997. Risk management in a dynamic society: A modelling problem. *Safety Science,* 27, 183–213.

Reason, J. 1990. *Human error*, Cambridge university press.

Reason, J. 1997. *Managing the risks of organizational accidents,* Aldershot, Ashgate.

Størkersen, K. V. 2012. Fish first. Sharp end decision-making at Norwegian fish farms. *Safety Science,* 50, 2028–2034.

Størkersen, K. V. 2017. Coastal cargo work: How can safety shout instead of whisper when money talks? *In:* Cepin, M. & Bris, R. (eds.) *Safety and Reliability. Theory and Applications.* Contributions presented at the 27th European Safety and Reliability Conference (ESREL 2017, Portorož, Slovenia, June 18–22, 2017): CRC Press.

Thorvaldsen, T., Holmen, I. M. & Kongsvik, T. 2017. HMS-undersøkelsen i havbruk 2016. Trondheim: SINTEF Ocean.

Thorvaldsen, T., Holmen, I. M. & Moe, H. K. 2015. The escape of fish from Norwegian fish farms: Causes, risks and the influence of organisational aspects. *Marine Policy,* 55, 33–38.

Turner, B. A. 1978. *Man-made disasters,* London, Wykeham.

Van Nunen, K., LI, J., Reniers, G. & Ponnet, K. 2017. Bibliometric analysis of safety culture research. *Safety Science,* In press.

Vaughan, D. 1997. *The Challenger launch decision: Risky technology, culture, and deviance at NASA,* Chicago, Illinois, University of Chicago Press.

Zohar, D. 2003. Safety climate: Conceptual and measurement issues. *In:* Tetrick, J. C. Q. L. E. (ed.) *Handbook of occupational health psychology.* Washington, DC, US: American Psychological Association.

Österman, C. & Hult, C. 2016. Administrative burdens and over-exertion in Swedish short sea shipping. *Maritime Policy & Management,* 43, 569–579.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Societal threat landscapes of petroleum industry activity in the high north

E. Okstad, T.O. Grøtan & A. Øren
SINTEF Technology and Society, Safety and Mobility Research, Trondheim, Norway

ABSTRACT: Today, industrial and societal systems interact and become more complex than ever, and hidden, dynamic and emerging threats and vulnerabilities evolve. By observing the petroleum activity in the north (west Barents Sea) along with other developments and societal trends in the region, it is possible to sketch out a threat landscape of the high north. A threat landscape is formed from interconnected 'pictures' of threats and actors, given some external conditions. These conditions are 'on the move', and requires that the risk pictures as well as the landscapes must be maintained and revised. Examples of such are effects of declining oil prices, or a changing climate. It is important to assess the validity of any picture contributing to the landscape. Threats may lead to 'spillover effects' and unexpected events if pictures become saturated. Scenarios applicable for stress-tests may be derived from such threat landscapes. The severity and urgency of a scenario is strengthened by events occurring simultaneously, or as combined events. Risk mitigation should thus be handled more in collaborating teams of interconnected actors instead of by single entities themselves. Actors involved in the 'oil in high north' threat landscape either take part in, support, or are affected by the petroleum activity in some way. Each party should consider their situation and role in view of the evolving threat landscape, and look for alternative handling of emerging risks. As part of a case work in the New Strains of Society project, key actors have been interviewed. In addition to the interviews, associated workshops have been held. This paper summarizes the foundations of a preliminary threat landscape based on these interviews and workshops, with recommendations for further elaboration by researchers as well as practitioners. Main challenges associated with the aforementioned threat landscape are outlined. The concepts of robust organizations and resilience are reflected upon as alternatives to the prevalent failure-oriented safety approaches. Knowledge from the current study serve as input to the final New Strains of Society framework.

1 INTRODUCTION

A threat could be explained as a possible danger that might exploit vulnerability of a system and cause possible harm. Threat landscapes then cover threats and vulnerabilities of a thematic area that involve a network of agents and stakeholders, interacting normally or more randomly (Grøtan & Antonsen, 2016). The landscape metaphor involves an aggregate of the more prevalent notion 'threat pictures'. A 'threat picture' is confined and focused on a narrowed set of topics, either thematic-centric (e.g., oil spill) or actor-centric. A picture comprises a frame or a border that represents a clear demarcation of the relevant vs. non-relevant issues belonging to the threat, including the relevant vulnerability and presumptions on operational conditions. Such presumptions can, however, be turned the other way around. They can be recognized both as demarcation lines for the validity of a picture in relation to hidden, dynamic and emergent effects, but also as a key to understanding when a picture will become saturated with unaccounted or unprecedented conditions. The overall idea is that as threat pictures mature, scenarios for stress testing can be built.

The finalized project New Strains of Society (New Strains) dealt with these matters. Empirical studies of societal threats and vulnerabilities were focused on gradual developments of threat landscapes for thematic areas like the petroleum activity in the high north. Information was gathered from documentation reviews, followed up with guided interviews and workshops. By utilizing a joint interview guideline, the representatives from involved actors were incited to elaborate on threat pictures and scenarios that normally fall outside their normal belief of system behavior. Okstad et al. (2017) described a basis for the design of empirical case studies of which the results should support to the New Strains framework.

1.1 *Objective*

The main objective of this paper is to summarize generic findings gathered from interviews and workshops with involved actors concerning threat

landscapes arising from the thematic area 'oil in high north', and to organize the finding into a rudimentary threat landscape that can be subject for elaboration and extension at a later stage, involving more actors.

2 INFORMATION GATHERED

2.1 *Description of case studies*

The case 'oil in high north' is about possible implications of the offshore petroleum business in the high north (Barents Sea). In addition to this case, there are two other cases covered in the New Strains project; the 'Pandemic' and 'ICT infrastructure'. The latter is much about cyber safety, security and resilience and is as such, integrated in the other two thematic areas. In fact, it deals with vulnerability and risk handling of critical ICT infrastructures, of most importance in maintaining any societal function.

The 'oil in high north' case defines threat landscapes much as societal impact of the petroleum activity and its presence in the region. Examples are environmental concerns, access to limited services in emergencies (e.g., SAR), like public hospitals and air-borne ambulance, as well as the dependability of shared infrastructures. The latter could be communication infrastructure, energy supply, transportation systems for goods and/or services to offshore installations as well as to local communities.

'Pandemic' is about dealing with a global epidemy that crosses national borders, affecting the public widely and thus, indirectly threatening important societal functions and infrastructures, herein questioning the deliberations of saturation points and interdependencies of the preliminary threat landscape presented in this paper.

2.2 *General aspects; method and interview guide*

In this paper, we try to delineate a preliminary threat landscape, based on tentative threat pictures, interviews and available public information. The key activity is to encircle and 'carve out' possible 'pictures' and their saturation characteristics. What external events could 'shake' or interrupt them? Possible spillover effects are here those effects that carry the potential to influence other (external or overlapping) pictures, and ultimately the whole 'landscape'. The strategy for doing this is based on the landscape representation of Grøtan and Antonsen (2016). Separate interviews were carried out and pragmatically used for this purpose, providing crucial grounds for the New Strains framework development aiming for stress-testing as the ultimate objective. Each interview was aligned to contribute in an optimal matter to the framework, and based on a joint interview guide (Grøtan and Antonsen 2016). Key players were interviewed at both a general level, and on specific topics related to their presumed or self-declared 'picture'. The initial aim was to identify the interviewees' home ground with respect to risk pictures and the risk management been implemented. What kind of events were covered by the risk analysis normally, and what have been the criteria for establishing emergency- and contingency plans?

By going through past events and how the organization responded, characteristics of how they were surprised are revealed, in addition to information about situation handling. By going through these experiences, one should touch upon the outer edge of what (situations) are possible to handle by the organization. This could be explained as the thickness of the threat pictures' frame. Next, the aim was to challenge actors to look for the bigger picture given the situations experienced, how could the situations possible escalate? What may have happened if the initial event occurred slightly different, or conditions were some otherwise for the scenario? At this stage, it was expected that actors used their imagination and took the opportunity to respond actively on both thinkable, and less thinkable scenario escalations. Important questions in this phase were:

- What has been done after the incident to be better prepared (learned by experience)?
- What kind of effects in sense of improved interaction between actors are seen?
- What kind of overlap with other threats are seen in the hypothetical crisis?
- How could such kinds of interactions and overlap be handled in practice?

Threat landscapes for 'oil in high north' have been formulated on basis of adapted interviews and meetings. The strategy was to pursue initiating events further and challenge operators and actors with respect to possible ways of escalations. By exerting pressure on a given threat landscape, one or more threat pictures were approaching a saturated state which triggered thinkable events. These events escalated from some defined 'overlaps' between the set of threat pictures and/or actors in the landscape.

One possible escalation relates to the production installation's dependence of electric power supply from shore in normal operation. A scenario, with overlapping threat pictures, e.g. challenging weather, or vulnerable power supply, could initiate from a winter storm caused by a polar low pressure. The storm knocks out the power supply to the offshore production facility, while at the same time hampers evacuation of personnel from installation by helicopter.

Then, we ask what could be done to avoid similar experiences, or to be better prepared. The risk management should be updated accordingly. Finally, taking as a premise that surprises anyhow might occur in the future—what could be done, or should happen to successfully deal with these surprises? Here, we identify some resilience capabilities. The interviews were structured according to the predefined sequence of issues starting with the interviewee's home ground, i.e. a description of the risk management- and emergency preparedness system in service.

2.3 *Actors involved in 'oil in high north'*

A list of potential interviewees was established based on a predefined actor landscape, shown in Figure 1.

These actors, or organizational units, were assumed to play active roles in the threat landscape. Hence, they will also be prime candidates for being actors in a future 'resilience landscape' (Grøtan & Antonsen 2016). Founded on the New Strains principles of building such landscapes (Grøtan & Bergström, 2016), Grøtan (2018) utilizes a New Strains discursive-support structure that enables such a polycentric resilience landscape to evolve over time. Based on this, a preliminary threat landscape may be elaborated and evolving further.

2.4 *Interviews and workshops accomplished*

During the project work, we managed to get in touch with an oil company, the council of county emergency preparedness in Finnmark, the regional health company and a local health company. We also planned to interview representatives from the police, the national electric power net agency, telecommunication companies and the local electric power company. This turned out to be difficult, inter alia because we easily run into asking for classified information that we cannot receive. There have been three workshops held in 2016–2017 on the topic cyber safety, security and resilience of critical infrastructure. Information from these workshops is elaborated on in the discussion part of the paper.

In addition to the obstacles of classified information, we also come into or derive information and knowledge that is obviously sensitive (but unclassified), and therefore is omitted.

However, it is emphasized here that the conception of 'sensitive' is based own judgements.

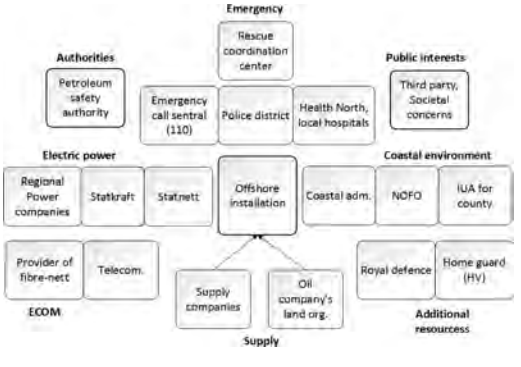


Figure 1. Actor landscape: 'Oil in High North'.

2.5 *Findings from narrative-structured interviews*

The 'narrative structure' is based on Grøtan and Antonsen (2016).

2.5.1 *General issues*

General characteristics of the high north involving the petroleum industry activity and maritime traffic in the area were i.a. as follows:

- A situation with limited 'resources' in the high north put constraints on crisis mitigation.
- Actors have experienced limited information provided by externals in early phases of situations, combined with bad communication.
- The region is known for rough weather conditions and large distances between parties and service providers in reaching the necessary resources.
- There are still conflicting interests between nations in the Barents region.
- Societal events, and/or global changes to society seem to have effect on business as usual.
- Cyber-security, hidden, dynamic and emerging (h/d/e) risks are evolving. These issues are incorporated in emergency exercises, initiated by the county administration, that involve actors and stakeholders starting to learn by collaboration.

2.5.2 *Actors 'state of the art' in risk management*

Oil companies give lower priority to security issues than pure (physical) safety. The petroleum business in Norway is not defined as a CI and thus, not subject to the Security Act like other critical infrastructures.

Misleading, cyber-related actions, or events with purposes others than what seems to be the case at first place are challenging tasks for the industry. It is generally difficult to improvise on such ICT-problems when lacking overview or knowledge to problem origin in the early phases of a scenario/event. Achieving a delayed overview of the situation, or late awareness of the main purpose of an attack is normal, not exceptional. Generally, there are limited roles, or dedicated activities in strengthening the security field in the petroleum business. Measures are prescribed for physical protection mainly, e.g.

like to secure persons from becoming violent offshore, implement 100% check of personal luggage before offshore flights, etc. However, the industry is aware of cyber threats like the 'denial of service'.

From the authority's point of view, cyber security is part of the overall emergency preparedness in society. Finnmark county has included these issues in their emergency-preparedness plans, and constitute an important part of their training program.

Geographical coverage of the health emergency-preparedness is defined in dialogue with national authorities. Focus is on robust systems for alerting, assessing and managing unexpected situations. The regional health company (Finnmark) is set for handling of major accidents with many injuries. It incorporates plans for the critical infrastructures. Emergency plans, that includes infrastructures, are aligned with the relevant actors, like oil companies, hospitals, police, coast guard, national defense, etc. The ambulance services are also crossing national borders to Russia and Finland.

2.5.3 *Experiences from h/d/e events in own business*

The following draws an overview of some threat pictures that are relevant for the 'oil in high north'. These pictures mainly are derived from the interviews.

Offshore installations may become vulnerable to electrical-power shut down from shore, given the new Norwegian policy that requires electrical power supply from shore. The tense political climate, with environmental activists and Russian collaboration makes operations challenging to a certain extent. The public society in the northern part of Norway, including the petroleum industry, have been surprised by several big events lately. Three examples were:

- The refugee situation in 2015 with large groups of people crossing Norwegian border at Storskog.
- The big avalanche at Svalbard in 2016 came as a surprise and challenged the health company's preparedness, including transportation of personnel.
- The ash cloud came from the Iceland volcano in 2010 led to major ambulance-flight restrictions.

Oil companies typically expect only minor, or none surprising events from own business. As mentioned, cyber security of critical infrastructures and lack of overview/delayed knowledge about emerging problems are general issues given implementation of new/innovated technology in society.

Finnmark as a large county with resources geographically spread is also a challenge. There is limited capacity of the local hospitals to handle big events with lots of injuries. Anyway, unforeseen events happen all the time in the north, often related to the challenging weather conditions and long distances. The time factor in an acute situation often become critical, and the required resources in an emergency may be far away, and take 2–3 hours to mobilize. An example of such is the fire-fighting service.

There are periods of the year with less access to back-up health personnel, e.g. during the Easter with people spending their holiday time out in the terrain. Personnel resources to solve ICT-problems may also be scarce locally, and it becomes a critical aspect if communication gets down during handling of a crisis.

Finally, the electric power situation depends on a power distribution grid crossing borders to both Finland and Russia, which makes it even more vulnerable. The electrical power net covering the northern parts of Norway is operated by 7 companies. Although, Norway, Russia and Finland have agreed on training the emergency preparedness against events (including power supply) that involve all three countries, the handling of crises near the borders to each country with utilization of the common resources is a critical issue.

2.5.4 *Pandemic as a threat*

The national preparedness plan for pandemic flu describes how to prevent and reduce the spread of infection, morbidity and death, and to provide good treatment and care for people affected by the flu pandemic. Municipalities in Norway are obligated through the Civil Protection Act, to actively work with emergency preparedness, such as e.g. to have a plan for how to handle possible pandemics. Some of our project group members were allowed to observe and challenge participants in the crisis response team during one specific manoeuvre, 'exercise Virus', in one municipality during the autumn 2017. The type of virus was in this case a seasonal influenza virus. In this exercise, all municipal offices were involved.

A pandemic is a complicated situation to handle for a municipality. It involves the administrative and operative management, and several different professionals such as general practitioners, hospitals, nursing homes, transportation, schools, day-care, groceries, to mention some. Moreover, a pandemic flu situation is highly different than an emergency event as e.g. a major explosion, that leaves an unexpected complex situation with a high degree of uncertainty. While many emergencies occur suddenly, a pandemic flu spreads and escalates in a slower manner. It is possible, to a certain degree, to be prepared and prevent the severity of the outbreak. However, it should be mentioned that the lethality of a flu pandemic could be abnormally high, in which could leave the situation even more complex and difficult to solve. Ebola pandemic, with the high degree of lethality, is another example of complex pandemic.

During the 'exercise Virus' it was observed that the organisation involved in handling the crisis is complex, the distance between the members in the

crisis management, the crisis response team, the administrative and operative management, as well as the many other actors involved, could lead to a non-optimal communication dialogue between professional institutions, external actors and decision makers. At some offices, the person of contact could have changed, without notifying in the emergency preparedness plan or in the digital crisis management tool, in which led to further communication problems. Misunderstandings could easily take place; degree of severity, underestimation of time consumption, and which resources are available. Moreover, since also neighbouring municipalities are affected in case of a pandemic, competition for the same resources as e.g. beds for a temporary emergency room/hospital, could occur. Other important issues are the complex logistics for handling sick persons, to continue with safe operation of critical societal functions and to have enough qualified personnel to run daily operations.

On the positive hand, since the pandemic situation often escalates in a slow matter, the management can reallocate human resources. E.g., personnel working at the municipal cultural school could be reallocated to work at vaccination posts or nursing homes, as there most likely would be shortage of health personnel in a pandemic situation. Still, because of the high numbers of persons involved in a pandemic, even a pandemic with low lethality could have the potential to paralyse—to a certain degree, a whole society.

From this, we hypothesize that a pandemic component in 'The High North', could affect the capability and function of persons directly involved at offshore installations as well as the support network on-shore from the Finnmark county. This could again threaten the safety for the work and the environment in 'The High North'. Although we presume that 'pandemic' aspects already have been considered by the 'Oil in High North' actors, we find reason to issue a warning that the 'crippling' effect of a pandemic is likely to jeopardize any assessment of saturation effects related to almost any risk picture. This is probably even more relevant for scarcely populated areas, where competent people probably have several roles, related to e.g., authorities, companies, NGOs and the civic society. The latter is especially important for reliance on community resilience, despite the assumption that the population is more robust compared to rural areas (see later discussion on resilience).

2.5.5 *Emergence, escalation into 'bigger pictures'*
The following gives some examples of possible escalation routes into 'bigger pictures'. Terror, where the health services can be targets themselves (physical/cyber), is thought of as a possible scenario. Sensitive patient information may get lost, although health companies are subjects to the Security Act.

A refugee situation could escalate with people disappearing after crossing the national border, with a pandemic coming up. At the same time, some might bring forward violent attitudes and crime.

With changing traffic patterns in the north, mainly enforced by increased tourisms, etc. the health companies are challenged on capacity and logistics in case of simultaneous events. There is limited capacity in local hospitals to handle big events with lots of injuries. These situations may as well effect on the health preparedness served to the petroleum business.

2.5.6 *Expectations and revised risk governance*
Of the most important learning from incidents by the oil companies is:

- Risk governance being updated based on learning from incidents.
- Outsourcing of ICT services requires improved focus on contracts.
- Improved (risk) knowledge is achieved among actors involved in the (oil) business.
- Hospitals, police, etc. are more aware of possible new scenarios offshore.

From a Finnmark county perspective, the new 'cyber scenarios' are to be included in the 'national risk picture', but the emergency preparedness should be made Finnmark-specific. The council of county emergency preparedness plays an important role in improving the society's risk awareness by maintaining an open dialog, and arranging training exercises.

For the health companies, review of emergency plans is expected to adapt at new events and developments of the society. New emergency plans should also cover 'hybrid-war' scenarios. Here, the rescue function of the armed forces need close cooperation with the civilian health service to handle crises. Interaction competence will thus be more important. The 22. July terror attack gave us new views in that respect (Johnsen & Øren, 2015).

2.5.7 *Recognizing the usefulness of resilience*
When it comes to specific resilience capabilities the county emergency preparedness is used as an example. The county expresses they adapt to situations, and the degree of improvisation often occurs there and then. It is however, more difficult to improvise on ICT-problems due to lack of overview/knowledge in early phases of a situation. Each sector/actor carry out resilience, or resilient behavior in different ways, either as an organization or individuals. Early sharing of information to society (e.g. the municipalities) is one of the county governor's main responsibilities and should support such a behavior.

Safety and emergency events offshore are handled by the Norwegian rescue coordination center (RCC) in first instance. The oil business itself may

contribute by its own transportation resources to assist the health companies in specific rescue operations, etc. The Seeking helicopter squadron has become an important resource for the health companies as well.

Experience from a gas-leak event on an installation in the north lately tells that health companies nowadays are notified earlier without occurrence of a serious event with injuries, or having a real evacuation situation. The threshold level to establish preparedness is maintained as before, but notification to the health company comes earlier than before. The main reason is that the crisis management at health companies is more able to judge, or assess the seriousness of situations during the early stages than operators at the alarm receiving center.

Another aspect of resilience is that inhabitants of Finnmark tend to try out their premises extensively in crisis situations. The northern communities are characterized by their robust populations. Due to the region of small, often isolated communities, the ability to handle unforeseen situations is spelled out when events occur. There is, however, a variety of such skills. Living in the north involves some risks and the inhabitants are more used to stand-in the situation when required. Next, there is also a reliance of our neighbor countries on Norwegian assistance in crises, e.g. in north Finland and Russia. An accident near the border to Russia can also be supported by resources from Russia. There is no experience from real events on such, but the countries exercise together regularly.

3 THREAT LANDSCAPES OVERVIEW

Generally, emergency preparedness in oil companies is mainly designed to handle single accident events at a time only, e.g. a hydro carbon-leak or fire. However, problems may escalate if e.g. a production upset and a hydro carbon leak occur at the same time and in combination with outfall of critical infrastructures. An example of the latter is loss of electric power from onshore, or technical failures on critical communication systems. The following summarizes input to threat landscapes derived from the interviews:

- Long distances in the north between critical (emergency preparedness) resources may induce logistic challenges when facing emergency situations, e.g. easy/in time access to onshore equipment in case of an unexpected oil spill.
- The petroleum industry is not subject to the 'security act', which makes the involved actors less accountable to cyber-attacks than other domains, e.g., 'denial of service' types of viruses.
- Possible 'Spillover effects' from externals during a crisis, e.g. environment activists, political interests may interrupt a proper handling of situations.
- Critical operations at the same time of an infection disease spread among key offshore personnel may reduce the problem-solving capacity dramatically.
- Terror attacks in combination with bad weather makes it difficult with respect to accessing accurate information, and handle multiple demands to limited resources for normalization.
- 'Oil in high north' is currently a Vest-Finnmark domain. A major accident offshore may however, require the whole capacity of the regional health company. That means the capacity of single health companies might be too scarce in those situations.
- Extraordinary events (e.g. the refugee situation at Storskog) combined with a major offshore accident could cause trouble for the emergency assistance from the health companies due to limited resources.
- Ambulance services by air is a critical function in emergency situations offshore, and may easily be interrupted by bad weather or natural disasters, like the ash cloud in 2010.
- The avalanche at Svalbard in 2016 came as a surprise and challenged the health preparedness of health companies, especially with respect to transportation of personnel. An offshore accident at the same time would have been difficult to handle locally.
- Health companies are always challenged on capacity and logistics by simulations events. The companies are decentralized, often small units with focus on the 'acute' functions during events.
- Experience from an exercise at a hospital with a CBR-incident (Chemical, Biological, Radiological), and at the same time a real failure in the communication system did occur. Then there was a need to put the exercise on 'hold' to solve it.
- Resources to solve ICT-problems become scarce with communication difficulties. Access to troubleshooting capacity also becomes a scarcity resource by constantly introducing new technology.

The above aggregate to three major treat landscapes that could be elaborated on further (see Figures 2–4):

- Production upsets
- Cyber security
- Emergency preparedness at sea

A further elaboration of these landscapes is conceivable, drawing on a discursive support scheme (Grøtan, 2018) that supports development of polycentric resilience landscape, including cyber vulnerability issues.

3.1 *External conditions*

In addition to threat landscapes, the following issues were registered through the interviews as

major external conditions, i.e. kinds of 'labyrinths or moving horizons' (Grøtan & Antonsen, 2016):

- Changing conditions for the offshore petroleum industry in general, mainly due to lower oil prices, a continuous need for cost cutting with structural and organizational consequences.
- Hidden vulnerabilities in the 'system' for electrical power supply from shore.
- Development strategies for the electrical power grid are not fully predictable for the industrial actors. Uncertainty relates to further development, of which, is partly geographically conditioned.
- Uncertainty relates to the resilience capabilities and degree of informal collaboration between actors involved in a potential (offshore) crises.
- The degree of information and communication constraints occurring between governmental levels and actors in crisis situations that evolve.

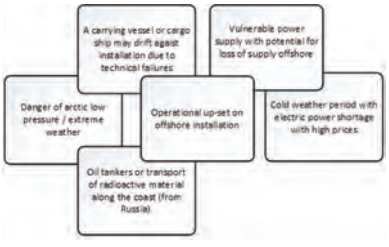


Figure 2. Threat landscape: Production upset.

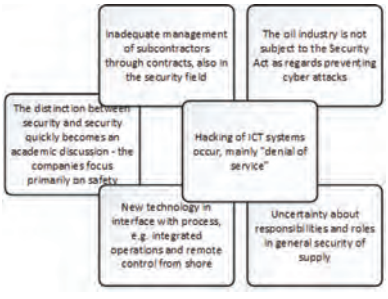


Figure 3. Threat landscape: Cyber security.

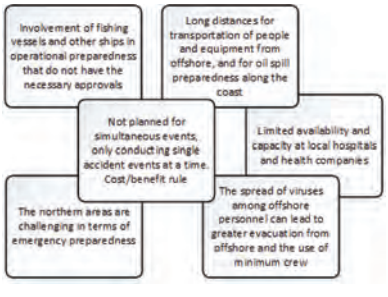


Figure 4. Threat landscape: Emergency preparedness at sea.

- The availability of personnel varies over time in the regional health companies depending on changing regional policies that may have effect on the quality of the health preparedness.
- The degree of shared information and communication between the actors in the petroleum business and health companies, has developed into new levels lately.
- The development of the expected 'total-defense' concept is uncertain. To what degree (and when) can actors rely on sufficient and steady interaction between actors according to such a concept?
- The technology change and degree of innovation in the hospital sector imply varying dependency on the Internet in rescue and health preparedness.
- Access to troubleshooting capacity (ICT related) may become scarce for the health companies in times of new implementations.
- Required manning in the northern health companies, and accessibility to personnel vary during the year due to individual work- and shift schemes.

4 DISCUSSION AND FURTHER WORK

Threat scenarios are elaborated on basis of threat landscapes with 'saturated' threat pictures and 'moving horizons' in Section 3 and 3.1. Scenarios may initiate from the offshore installation itself, from a passing ship in the area, enforced by bad weather conditions, or loss of critical infrastructures like electric power from shore. In combination with other events, scenarios may escalate into severe accidental events.

The important aim of scenario framing/delimitating in New Strains was to establish an appropriate foundation and discussion basis for the actors and stakeholders involved (Okstad, 2016). Any actor should take the opportunity to respond on knowable escalation factors of the identified scenarios, if they were involved or affected according to the given threat landscape.

The scenario should neither be too specific, nor to general. In case of the first it probably requires too detailed knowledge and extensive verification of its implications for the final acceptance as a scenario among actors. If scenarios become too general and vague, they will be counterproductive and of minor value in clarifying possible escalation factors and effects of such within a given threat landscape.

Another challenge was how far one should go to look for interconnections between contemporary threats and adverse circumstances as reinforcements to possible threat escalations. The right balance was found to be just making illustrations of possible spill-over effects, or the added effects of several overlapping events or vulnerabilities in the given threat landscape.

Threat scenarios that have been discussed related to 'oil in high north' are:

– Critical exploration drilling activities goes on in the North Barents Sea, and at the same time, a major ICT-system failure occurs. There is limited ICT infrastructure in the area. A well control situation escalates into a critical situation that require shutdown and immediate evacuation of personnel from the rig.
– A large passenger boat gets into fire at open sea, at the same time an ICT-problem strikes the health company. Combined with large discharges of oil/chemicals at a site of damage (CBR), situations quickly get out of control. Viewed from the health company's point of view, the size of the accident in sense of the number of injured people becomes the biggest stress factor.
– Simultaneous energy-power cut to an offshore installation and an oil spill, or operational problem at site. Challenges then occur related to fast and effective mitigation of consequences. Finnmark county is large, with the onshore recourses geographically spread. Then it takes time to move, e.g. by helicopter between west and east.
– Offshore event, international cyber-attack, critical failure in communication systems/infrastructure (random failure/out-fall, or a deliberated action, e.g. cut of cables).
– Ship collision with an installation, and at the same time an outfall of electric power from shore occurs. Finally, an escalating pandemic flu strikes the area. With a trend of increasing ship transportation with new production- and exploration drilling taking place in the north-east Barents Sea. In addition, system vulnerabilities exist connected to ICT and Internet of Things incorporated in every critical infrastructure and logistic function.

Our experience from the interviews is generally positive with respect to the interviewees willingness and engagement to share their opinions and thoughts on these aspects. However, we faced a challenge regarding confidentiality and security clearance at a degree that limited our contact with some actors.

Regarding the workshops, the participating people were motivated to contribute there and then. Discussions in interdisciplinary work groups was a success. People from authorities, research units and practitioners worked very well and the impression was that it was valuable spending of time. However, the discussions regarding threat pictures, and especially related to cyber security, quickly became technical. In some discussion groups, therefore, the discussion was raised a level and used more to elaborate around possible effects or consequences of the cyber threats. One drawback was the tendency of pulling out the consequences completely (to wide, or thinking 'worst case') instead of thinking possible event chains or sequences leading to such. Maybe we should have emphasized to concentrate on a limited set of consequences, i.e. linked to the given organization.

5 CONCLUSION

In line with the suggested approach to encircling and building threat landscapes, a preliminary threat landscape for 'Oil in High North' has been developed. Results from a study of pandemic handling in a municipality has also been included in the landscape. This threat picture sensitizes the 'oil in high north' scenario to be further scrutinized concerning operational presumptions and considerations of saturation effects.

For a further development to take place, crucial issues about not only classified information, but also sharing of obviously 'delicate' information on (e.g., cyber) vulnerabilities would have to be addressed. The authors judgement is that this should be possible, laying the foundations for stress-testing according to the New-Strains objective (Okstad et al. 2016).

ACKNOWLEDGEMENTS

The authors like to acknowledge contributors in interviews and workshops for providing valuable information to the discussion. The Norwegian Research Council through the research contract No. 238093/H20 sponsors the paper.

REFERENCES

Grøtan, T.O. 2018. Building Cyber Resilience through a discursive approach to 'Big Cyber' Threat Landscapes. Paper submitted to ESREL 2018.

Grøtan, T.O., Antonsen, S. 2016. Take it to the limits. An empirical strategy for exploring the new strains of society in terms of hidden, dynamic and emergent vulnerabilities. ESREL 2016, Glasgow, Scotland.

Grøtan, T.O., Bergstrøm, J. 2016. Calibrated resilience landscapes of composite protection: Theoretical grounding of an empirical approach. ESREL 2016, Glasgow, Scotland.

Johnsen S.O., Øren, A. 2015. Is complacency creating a reactive regulatory regime? ESREL 2015, Zürich, Switzerland.

Okstad, E. 2016. Scenario approaches as a means of handling emerging risks in society. ESREL 2016, Glasgow, Scotland.

Okstad, E., Dahl, Ø. 2017. Horizon scanning approaches for early sensing of cyber-physical threats to water utilities. 53rd ESReDA Seminar: Enhancing Safety: The Challenge of Foresight.

Okstad, E., Grøtan, T.O., Paltrinieri, N. 2017. An empirical case design and stress test addressing hidden, dynamic and emergent vulnerabilities of society. ESREL 2017, Portorz, Slovenia.

Øren, A., Grøtan, T.O., Dahl, Ø. Pandemic landscape of new strains of society. ESREL 2016, Glasgow, Scotland.

Weick, K.E. & Sutcliffe, K.M. 2007. Managing the unexpected, resilient performance in an age of uncertainty. Second edition, John Wiley & Sons, Inc.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Interorganizational complexity—main challenges and opportunities in the petroleum industry

V. Milch & K. Laumann
Research Group for Safety and Human Factors, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

ABSTRACT: With growing outsourcing in the petroleum industry, operations are becoming increasingly interorganizationally complex. This raises the question of how interorganizational complexity influences safety. The aim of the research project "interorganizational complexity and risk of major accidents" was to gain a better understanding of safety challenges associated with interorganizational complexity. Three studies have been completed. In this paper, we present the main findings from the research project and discuss how the resulting knowledge can best be applied in the petroleum industry. The findings call for more awareness concerning interorganizational safety challenges in the industry, and points to specific areas in which challenges may arise. In particular, coordinating work processes, ensuring sufficient levels of competence and transferring lessons learned and best practices across companies, are identified as such areas. However, the findings also imply practices that help manage interorganizational complexity. The presence of high-quality work relations appear to be important to achieve safety performance across collaborating companies. In this regard, middle managers appear to play a central role in terms of aligning employees from collaborating companies towards a shared focus on operational and safety related goals.

1 INTRODUCTION

In the Norwegian petroleum industry, the outsourcing of safety-critical activities and services to contractors in petroleum operations, have increased in recent years. As a result of increasingly complex drilling operations and recent technological developments, work processes in oil and gas production have become more and more specialized over the years. Contractors and subcontractor companies hold specialist competence within specific domains and are hired to provide various types of equipment, such as drilling equipment or valves, or specialist services in operation, such as cementing or drilling fluids. As a result, work processes in the petroleum industry are fragmented across a large number of companies with different areas of responsibility and varying levels of involvement. Outsourcing is undoubtedly an advantageous strategy in that it contributes to enhance performance by expanding the pool of available expertise, enabling organizations to become more competitive in an increasingly high-paced marked. However, an increase in the number of stakeholders also entails increased complexity and added possibilities of how human, technological and organizational components within the system can interact (Perrow, 1984, Rasmussen, 1997). This does not only have practical implications in terms of organization by placing greater demands on aspects such as communication and coordination of work processes, but may also have significant implications for safety. With multiple stakeholders that operate independently, yet interdependently, within their respective areas of expertise, it is challenging to maintain a complete understanding of what is going on. Accordingly, discovering and responding to safety threats becomes more challenging.

The Norwegian Petroleum industry is known to demonstrate enhanced focus on safety and invest significant effort towards strengthening collaboration and trust across actors in the industry. However, there seem to be less focus on understanding safety challenges that arise at the interfaces between collaborating companies. The developments we see in the industry in terms of increasingly demanding drilling conditions due to depleted oil fields and the onset of petroleum production in the Arctic, with more demanding weather and temperature conditions, inadequate infrastructure, and larger distances to onshore facilities (Petroleum Safety Authority Norway, 2012) calls for more complex and technologically advanced drilling operations in the future. Such conditions will require more in terms of cooperation between the companies involved, which means that interorganizational complexity can have consequences for how well

safety is managed. It is therefore important to gain more knowledge about the safety implications of interorganizational complexity.

The research project *Interorganizational complexity and risk of major accidents*, funded by the Research Council of Norway (Grant nr: 220798), was initiated to develop knowledge about the safety implications of interorganizational complexity in petroleum operations, focusing particularly on major accident risk. The main objective was to investigate how interorganizational complexity contributes to safety challenges, and explore and discuss how challenges can be reduced. The project was commenced in 2013, and completed in 2017. During this period of time, we have completed three studies exploring the influence of interorganizational complexity on safety from different angles. These include: a review of empirical literature addressing interorganizational safety challenges (Milch and Laumann, 2016), a qualitative study of safety challenges and practices that help manage interorganizational complexity on a Norwegian petroleum installation (Milch and Laumann, 2018; Milch and Laumann, 2017), and finally, a qualitative analysis of investigation reports, exploring the influence of interorganizational factors on incidents in the Norwegian petroleum industry (Milch and Laumann, submitted). In this paper, we present the main finding from the tree studies undertaken in this project and intend to discuss practical implications of the current project for the petroleum industry. In particular, we would like to discuss how the knowledge from this project can best be utilized and applied in the petroleum industry.

In this project, to explore the findings in a nuanced manner, we have employed different theoretical lenses that represent long-held views as well as more recent theoretical developments. These are: The High Reliability Organizations Perspective (Weick et al., 2008, Weick and Sutcliffe, 2007, Weick and Sutcliffe, 2015), Reason's perspective on organizational accidents (Reason, 1997) and the perspective of migration towards the boundary/drift into failure (Rasmussen, 1997, Dekker, 2012).

2 METHOD

Because of the sparsity of research within the field and the topic's complexity, we opted for a qualitative approach. Qualitative research is particularly suitable for investigating complex phenomenon, particularly in circumstances where prior knowledge is limited (Miles et al., 2014, Elliott et al., 1999, Silverman, 2006). To illuminate different angles, the studies in this research project draws on different sources of data. Study 1 is a literature review of empirical literature to identify interorganizational safety challenges (Milch and Laumann, 2016). The aim of this work was to get a sense of key challenges that can arise due to interorganizational complexity. A literature search was performed to identify empirical literature describing interorganizational issues that negatively influence safety. The search resulted in 22 articles, which shows that existing research on this topic is limited. In study 2, we explored how interorganizational complexity was managed on a Norwegian petroleum producing installation (Milch and Laumann, 2018; Milch and Laumann, 2017). The aim of the study was to gain a better understanding of the main safety challenges. In addition, we were interested in exploring practices that help manage interorganizational complexity. Semi-structured interviews with informants representing various affiliations and hierarchical positions were combined with observations on site and onshore, and analyses of relevant formal documents. In total, 14 interviews were conducted. In the third and final study, we investigated how interorganizational factors have contributed to accidents and incidents in the petroleum industry through examining investigation reports issued by the petroleum safety authority between 2006–2016 (Milch and Laumann, submitted). The aim of this study was to explore how interorganizational factors are related to incidents, and gain knowledge about of what factors contribute to unwanted occurrences. 22 reports were identified in which interorganizational issues could be linked to incidents and accidents.

In all three studies, data were analyzed using thematic analysis as described by Braun and Clarke (2006). Thematic analysis is a qualitative method used to identify overarching patterns or themes in the data. The analytical procedure consists of five recursive analytical phases. The first phase is simply familiarizing oneself with the content, reading through all of the material to get a sense of its main features. Following that, the material is coded, whereby the aspects of the material relevant to the research question are examined line-by-line, and smaller segments are given informative and short labels. The next phase involves actively searching for themes. Here, resulting codes are examined and compared and similar codes are clustered together to form candidate themes. The thematic structure at this point is not fixed, rather, the entire analytical process is one of constant comparison, in which the content of codes and themes are constantly compared against each other for deviations and adjusted accordingly. When a list of candidate themes has emerged, the next phase involves a more thorough review of the candidate themes to identify what themes represent main themes and what themes count as sub-themes. In the last phase, a final review is made before the thematic structure is finalized and a thematic map is developed.

3 RESEARCH FINDINGS

3.1 *Study 1*

The findings from this study show that safety challenges arising from interorganizational complexity falls into four categories: *economic pressures between companies, disorganization of work processes, dilution of competence* and *organizational differences* (Milch and Laumann, 2016).

The literature suggests that economic pressures between companies can be a source of safety issues. This is because stakeholders pursuing different and perhaps conflicting goals in addition to collective operational goals, contribute to goal conflicts for employees at the sharp-end and a fragmented overall safety focus. Moreover, due to the potential economic consequences of errors or accidents for individual companies, the focus on assigning blame amongst involved companies can often displace the focus on learning from what happened. Another aspect that becomes problematic when multiple companies are involved is the organization of work processes. Processes such as coordinating tasks and responsibilities and communication demands more effort the more interfaces that need to be managed. The literature indicates that interorganizational complexity often results in more disorganized and fragmented work processes. Findings also suggest that interorganizational complexity often contributes to reduced quality in available competence. Contractor employees, who spend shorter periods of time on the site, tend to be unfamiliar with the workplace, and can often be inexperienced and have little prior training. As such, they are not necessarily equipped with the knowledge to deal with unexpected safety-critical situations or may not be familiar with the hazards and safety constraints in the workplace. The final theme identified in the literature concerns organizational differences between collaborating companies as a source of safety issues. Collaborating companies can vary greatly in terms of organizational culture and work practices. Such variations can hinder efficient collaboration and a shared orientation towards operational goals. Moreover, organizational differences can also create conflicts and distrust between employees from different companies, which hinders open communication about safety issues.

The interorganizational issues that we found in this paper can in various ways impede the ability of the organizational system to identify and respond to safety treats, as they are associated with the development of latent conditions and fragmented operational focus. Therefore, challenges that arise from interorganizational complexity may increase the risk of major accidents.

3.2 *Study 2*

With regard to safety challenges, informants generally expressed few challenges with the interorganizational elements of their work. The perceived challenges largely reflect structural aspects relating to managing and organizing operational processes between involved companies. Coordination of work processes, information flow and varying levels of experience among contractor personnel were identified as the most important challenges (Milch and Laumann, 2018).

The study also identified several practices that help manage interorganizational complexity. Employees from collaborating companies do not necessarily form close collegial relationships, but still engage in close collaborations. As such, high-quality relations, underpinned by friendly interaction and mutual respect, was deemed crucial for maintaining well-functional collaboration across organizational boundaries. High-quality work relations appear to promote a shared focus on operational goals, and were also found to stimulate constructive and open communication about safety matters across organizational boundaries. Findings also suggest that long-term organizational relationships between collaborating companies, worker involvement and managers' interactive role in the sharp-end represent important organizational elements stimulating high-quality work relations in this context (Milch and Laumann, 2018; Milch and Laumann, 2017). In addition, the similarities found among companies in terms of safety philosophies and practices also appear to be important for aligning companies in their efforts to achieve collective operational goals. Moreover, similar safety practices and philosophies also appears to be conducive to reporting behavior, as employees from collaborating companies seem to have similar perceptions about safe operational conduct.

The findings show that high-quality work relations may be a particularly important element in the pursuit of achieving and sustaining safety across collaborating companies in an interorganizational context. Moreover, since many of the challenges identified in the literature was not identified in the current study, it could suggest that the presence of high-quality relations, combined with shared safety philosophies and practices across companies, may potentially counteract some of the safety challenges that have been found to arise in interorganizational collaborations.

3.3 *Study 3*

Findings show that interorganizational issues contribute to both occupational incidents and major near accidents. Four themes were identified that describe interorganizational factors contributing

to incidents: *Ambiguities in roles and responsibilities between personnel from different companies*, *Inadequate processes to ensure sufficient competence across interfaces*, *Inadequate quality control routines across organizational interfaces* and *communication breakdowns between companies* (Milch and Laumann, submitted).

Roles and responsibilities between personnel from collaborating companies were often insufficiently clarified, which resulted in confusion regarding what company or organizational unit was in charge of what. Such ambiguities had consequences in the form of disruptions in the follow-up of operational processes and had in some cases also resulted in omissions of safety-critical activities. Another interorganizational issue identified was inadequate processes to ensure sufficient competence across interfaces. In several of the reports, the lack of installation-specific training and experience in personnel, and particularly in contractor personnel, was identified as a contributing factor to sharp-end incidents. Moreover, failure to ensure sufficient competence in planning of safety-critical activities was also evident in serious near major accident, where relevant contractor personnel with key expertise had not been involved in planning.

The study also suggest inadequate quality control routines across organizational boundaries, contributed to incidents. This was identified as a problem in the handover of equipment delivered by third party companies, and also with assembled structures made up of components delivered from multiple companies.

Finally, communication breakdowns between companies were identified as challenge in the incident reports. Important information is lost or not communicated. A central issue in this regard is that companies fail to communicate experiences and lessons learned from previous incidents. This means that known problems and hazards are not sufficiently communicated between companies, resulting in the repeated occurrence of similar incidents.

The study demonstrates the relevance of including interorganizational factors in the investigation process. This is important not just from a preventative perspective by increasing the learning potential from investigation reports, but more attention towards interorganizational factors is equally important from a proactive perspective, in terms of making such issues more explicit and finding ways to cope.

4 DISCUSSION

The aim of the research project *Interorganisational complexity and risk of major accidents* has been to gain knowledge about safety implications of interorganizational complexity by investigating how interorganizational complexity contributes to safety challenges, and exploring and discussing how challenges can be reduced. The findings from this project offer several practical implications. In the following, we will address the project's main implications and discuss how the current findings can best be applied and utilized in the industry.

The industry demonstrates a strong safety focus and a lot of effort has been made to strengthen collaboration and trust across actors in the industry. However, there seems to have been limited focus on safety challenges that arise at the interfaces of companies. The findings from this study show that interorganizational complexity contributes to several safety challenges in the petroleum industry that can increase the risk of major accidents. Even though the occurrence of adverse events is rare in the industry, awareness concerning such issues is important. Knowing what to look for can increase the capacity of petroleum organizations to recognize and respond to weak signals before they develop into major accidents (Weick and Sutcliffe, 2007, Weick and Sutcliffe, 2015).

One area that appears to be challenging is successfully coordinating work processes between companies. The sheer number of interfaces that must to be coordinated on a petroleum producing installation, involving interfaces onshore/offshore, as well as interfaces between the operator company, the drilling contractor and various contractor and subcontractor companies require continuous effort, which makes it difficult to ensure sufficient information flow and maintain a complete overview of operational processes. The findings imply that misunderstandings and confusion regarding roles and areas of responsibility between companies can occur. This not only impedes the quality of monitoring and follow-up efforts, but can even lead to the omission of safety-critical activities. This is because companies are largely focused on their own areas of responsibility, and may be unaware of what the other companies are doing. As such, the negligence of important activities or tasks can go unnoticed. Such fragmentation can contribute to the build-up of latent conditions in the system, which in combination with triggering events can cause adverse events (Reason, 1997).

The coordination challenge that we find appears in some part related to lack of clarification about roles and responsibilities among companies before projects are commenced. Moreover, it was also found that formal documents describing the distribution of roles and responsibilities among companies were in some cases inaccurate or not up to date. This implies that operating companies in the Norwegian petroleum industry do not necessarily have

adequate routines to ensure that work processes are sufficiently clarified and understood across organizational boundaries. This might suggest that there is a need for better systems to achieve more successful coordination of organizational and operational processes between involved companies in the petroleum industry. In practice, this may require that more effort is made by companies to ensure that formal documents such as organizational maps and documents describing roles and responsibilities across companies are accurate and up to date, and consistent with how the work is actually organized among involved actors in practice. Following up on these aspects formally fall under the responsibility of the operator company, as the actor that manages the operational processes. However, it could be questioned whether it is reasonable to assume that operator companies are able to oversee and check that all these documents are up to date. According to Perrow (1984), maintaining centralized control is problematic with high degrees of complexity. In this regard, a more decentralized form of control may be appropriate for these processes, whereby each individual company is responsible to ensure that formal documents are up to date and roles and responsibilities are adequately described.

Another way to reduce ambiguities and confusion between companies is through initiatives to ameliorate current communication practices between companies in the industry. The industry may benefit from developing better communicational strategies to ensure that all parties have the same understanding of their roles and areas of responsibility. Research on teamwork has shown that the use of closed-loop communication, where the receiver communicates the message back to the communicator for confirmation, contributes to better team performance (Salas et al., 2005). Similarly, collaborative cross-checks, where employees with different perspectives actively discuss and explore each other's assumptions, appear to be a promising strategy in terms of making processes more observable among employees with different perspectives (Patterson et al., 2006).

The second challenge that we find is the failure of experience transfer related to lessons learned from incidents, known problems or sources of hazards and best practices among collaborating companies. The challenge with transferring experiences and learning from incidents is not a new one, and has been quite frequently debated in the industry. The Petroleum Safety Authority keeps asking, why aren't we learning? The current findings suggest that interorganizational complexity may be one part of the explanation. Learning must not only occur in one organization, but across multiple organizations that differ in their expertise and operational focus, and vary in degrees of involvement. Ensuring that lessons learned are absorbed equally in all involved companies can be challenging considering that companies focus on different operational domains, and the resources and time available to follow up on lessons learnt may vary. Another obvious issue is information overload. Experiences and lessons learned can easily be lost in the vast amount of information that is transferred between companies. With an ever growing information load, telling relevant information from noise can be difficult (Edmunds and Morris, 2000). There is no simple solution to this challenge. However, industry-wide cooperative initiatives such as "Safety forum" and "Working together for safety" have shown to be successful in terms of promoting openness and collaboration across companies in the industry about best practices and sharing experiences about safety issues (Haukelid, 2008, Wiig and Karlsen, 2006). Several informants in study 2 emphasized these forums as important arenas for experience transfer between companies. Accordingly, such collective arenas appear to be very fruitful for supporting experience transfer and learning across companies. The industry may benefit from developing these areas even further to facilitate learning across companies.

Ensuring sufficient levels of competence across companies also appear to be a challenge in petroleum operations. Findings from analyzing investigation reports suggest that the competence requirements for doing specific tasks or operating specific equipment are sometimes not met. Possible explanations may be that competence requirements are not sufficiently appreciated or understood, that competence requirements are not clear enough, but it could also be a matter of capacity. In study 2, some informants mentioned that lacking experience and training among contractor company personnel could be a problem in times of high activity, which would suggest that it is also related to availability. This was also identified as a challenge in a study that looked at safety issues related to capacity and competence in the Norwegian petroleum industry (Skarholt et al., 2014). In the investigation reports, there were several examples where new and inexperienced contractor personnel had been put to operate equipment that they were not trained to handle. This does not only increase the risk of erroneous handling of equipment, however, the presence of new and inexperienced personnel can constitute a risk in itself because such personnel will most likely be unaware of local conditions and hazards that can influence safety. Moreover, inexperienced personnel will have reduced capacity to improvise in unforeseen or unfamiliar situations.

Companies in the petroleum industry normally have a buddy system in which new personnel are appointed a "buddy" responsible for familiar-

izing them with the installation and with their work station. In the investigation reports, there were some indications to suggest that the buddy program on some installations had a more social focus, and did not sufficiently equip new personnel with relevant safety information. Moreover, there were indications that training programs in some cases were too general, and did not detail specific tasks or equipment. The level of specialization we see today, which will likely increase in the future, entails that expertise is key to safe performance (Weick and Sutcliffe, 2015, Weick and Sutcliffe, 2007). Ensuring safe operations require that employees have specific training and experience to perform their work tasks, so that they can recognize when something is wrong. In fact, the ability to track small failure has been identified as an important characteristic underlying the reliable and stable safety performance of high reliability organizations (Weick et al., 2008).

While the findings from this project points at areas in which interorganizational safety challenges occur, the current research also illuminate several factors that appear to be important to achieve and sustain safety across organizational boundaries. The findings suggest that inter-personal factors are central in this regard. The presence of high-quality work relations between employees from collaborating companies appear to contribute to strengthen a collective orientation towards operational goals as well as being conducive to open communication about safety related matters. By the majority of informants in study 2, high-quality relations were pointed out as an important ingredient for well-functioning collaboration with employees from collaborating companies. Because employees from collaborating companies do not necessarily form long-lasting collegial relationships, but still collaborate closely in safety-critical activities, the character of work-relationships can be crucial in terms of how well the collaboration will work. The emphasis on ensuring high-quality work relationships that we find in our research, appear to reflect the efforts in the industry the recent decades to improve safety culture and strengthen collaboration and trust between actors in the industry. This may suggests that initiatives directed towards promoting high-quality work relations across collaborating companies may be beneficial to enhance collaboration. However, cultural similarities may also be an important contributing factor in this regard as the majority of employees in the Norwegian petroleum industry share the same cultural background. With a high level of cultural diversity, it may be more challenging to form high-quality work relationships when employees have different cultural values and beliefs and speak different languages. We still do not know enough about the specific preconditions and factors that are central to forming high-quality work relations across organizational boundaries. More research is needed to further explore how high-quality relations are formed across collaborating companies.

With the current level of specialization of work among companies that we see in today's petroleum industry, middle managers seem to play an increasingly important role in terms of achieving a joint focus among involved companies towards mutual operational goals, but also for monitoring concurrent work processes. In particular, the interactive role managers have, spending a lot of time in the sharp-end, was regarded as an important element contributing to well-functioning collaborations. Middle-managers offshore saw it as the most important part of their work, and reported that being present in the field was important to build trust with personnel from various companies, and to get a sense of what was going on in the operation. Sharp-end workers emphasized that managers being present and approachable, eased bringing up safety-related matters. A potentially worrisome tendency in the petroleum industry is the increasing amount of paperwork required of middle managers. Informants with management responsibilities reported concern about the extent to which increasing amounts of paperwork ate away the time they could spend in the sharp end, and were worried about the safety effects this may have in the long run. Research has shown that managers play an important role in terms of shaping safety performance (Mearns et al., 2003, Flin, 2003). While increasing paperwork in many ways is the result of a stronger focus on documentation of safety-management in the industry, our findings suggest that this may have the opposite effect than intended, at the expense of activities that have important, but perhaps less obvious safety functions. Arguably, the safety functions embedded in middle-managers' interactive role in terms of managing interorganizational complexity appear to be important, but seem to be poorly understood.

5 CONCLUSION

The aim of this paper has been to present findings from the project "interorganizational complexity and risk of major accidents" and discuss how the implication from this project can be applied in the industry. Findings call for more awareness concerning interorganizational safety challenges in the industry, and points to specific areas in which challenges can arise. In particular, coordinating work processes, ensuring sufficient levels of competence and transferring lessons learned and best practices across companies are identified as areas

in which challenges can arise. However, the findings also suggest that interpersonal factors and the presence of high-quality work relations are important to achieve safety performance across collaborating companies. In this regard, middle managers appear to play a central role in aligning employees from collaborating companies towards a shared focus on operational and safety related goals.

REFERENCES

Braun, V. & Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3, 77–101.

Dekker, S. 2012. *Drift into failure: from hunting broken components to understanding complex systems,* Farnham, Ashgate Publishing Limited.

Edmunds, A. & Morris, A. 2000. The problem of information overload in business organisations: a review of the literature. *International Journal of Information Management,* 20, 17–28.

Elliott, R., Fischer, C. T. & Rennie, D. L. 1999. Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology,* 38, 215–229.

Flin, R. 2003. "Danger – men at work": Management influence on safety. *Human Factors and Ergonomics in Manufacturing* 13, 261–268.

Haukelid, K. 2008. Theories of (safety) culture revisited—An anthropological approach. *Safety Science,* 46, 413–426.

Mearns, K., Whitaker, S. M. & Flin, R. 2003. Safety climate, safety management practice and safety performance in offshore environments. *Safety Science,* 41, 641–680.

Milch, V. & Laumann, K. 2018. Sustaining safety across organizational boundaries: a qualitative study exploring how interorganizational complexity is managed on a petroleum-producing installation. *Cognition, Technology & Work,* accessed 15 February 2018, https://link.springer.com/article/10.1007/s10111-018-0460-8.

Milch, V., & Laumann, K. 2016. Which types of leadership behaviors can promote safety in an interorganizational context? In: Walls, L., Revie, M., & Bedford, T. eds. Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016, September 25–29, CRC Press: Glasgow, pp.1873–1879.

Milch, V. & Laumann, K. 2016. Interorganizational complexity and organizational accident risk: A literature review. *Safety Science,* 82, 9–17.

Miles, M. B., Huberman, M. A. & Saldaña, J. 2014. *Qualitative data analysis. A methods sourcebook,* Thoasand Oaks, Sage Pupbilcations.

Patterson, E. S., Woods, D. D., Cook, R. I. & Render, M. L. 2006. Collaborative cross-checking to enhance resilience. *Cognition, Technology & Work,* 9, 155–162.

Perrow, C. 1984. Normal accidents: Living with high risk systems. Princeton: Princeton University Press.

Petroleum Safety Authority Norway. 2012. The far north: Moving with caution.

Rasmussen, J. 1997. Risk management in a dynamic society: A modelling problem. *Safety Science,* 27, 183–213.

Reason, J. 1997. *Managing the risks of organizational accidents,* Hampshire, Ashgate Publishing Company.

Salas, E., Sims, D. E. & Burke, S. 2005. Is there a "Big Five" in Teamwork? *Small Group Research,* 36, 555–599.

Silverman, D. 2006. *Interpreting qualitative data,* London, Sage Publications Limited.

Skarholt, K., Lamvik, G. M., Evjemo, T. E. & Rosness, R. 2014. Kapasitet og kompetanse i riggnæringen. Er mangel på kompetanse en sikkerhetsrisiko? Trondheim: SINTEF.

Weick, K. E. & Sutcliffe, K. M. 2007. *Managing the unexpected: resilient performance in an age of uncertainty,* San Francisco, John Wiley & Sons.

Weick, K. E. & Sutcliffe, K. M. 2015. *Managing the unexpected: Sustained performance in a complex world,* San Francisco, John Wiley & Sons.

Weick, K. E., Sutcliffe, K. M. & Obstfeld, D. 2008. Organizing for high reliability: Processes of collective mindfulness. *Crisis management,* 3, 81–123.

Wiig, E. & Karlsen, K. 2006. Working Together for Safety: A successful story of cooperation. *SPE International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production.* Abu Dhabi: Society of Petroleum Engineers.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

False alarm? Effects of reducing unnecessary dispatches by fire and rescue services

G. Gjøsund
NTNU Social Research, Norway

P.G. Almklov
NTNU Social Research, Norway
SINTEF, Norway

C. Sesseng
RISE Fire Research AS, Norway

ABSTRACT: Fire and rescue services in Norway dispatch more often to false and unnecessary alarms than to real fires and accidents. In 2016, 60% of the emergency dispatches were conducted on the basis of false or unnecessary alarms. These unnecessary dispatches are costly in terms of time and resources spent, and can in some cases lead to a weakened preparedness towards real incidents. Also, the risk for traffic accidents increases when big vehicles rush through the streets on their way to where the alarm was triggered. Hence, there are good reasons to work to reduce the number of these kind of dispatches. On the other hand, one may also argue that there can be some positive effects of a certain number of mobilizations for the fire crews. Based on interviews with relevant actors connected to fire and rescue services, as well as on statistics collected through the BRIS reporting system, we will discuss possible consequences of reducing the number of false and unnecessary alarms and potential effects of implementing measures for decreasing unnecessary dispatches.

1 INTRODUCTION

Fire and rescue services in Norway more often dispatch to different types of false or unnecessary alarms than to real fires and accidents. In 2016, 59% of the emergency dispatches were conducted on the basis of false or unnecessary alarms (DSB 2017). The numbers are increasing, from 20 000 in 2013 to 50 000 in 2016. An interesting point is that the number of emergency dispatches conducted on the basis of false or unnecessary alarms varies a lot between different municipalities in Norway; some have 40% and some as high as 75% (numbers from 2010). There is currently little knowledge about why this is so.

Unnecessary dispatches are costly in terms of the time and resources spent, and can, in some cases, lead to a weakened preparedness towards real incidents. Also, the risk for traffic accidents increases when big rescue vehicles rush through the streets on their way to where the alarm was triggered. Hence, there are good reasons to work to reduce the number of these kind of dispatches. On the other hand, one may also argue that there can be some positive effects to be derived from a certain number of mobilisations for fire crews.

As a response to the increasing number of false alarms, several fire departments in Norway are seeking to develop measures to reduce these numbers. This includes different technical measures and operational measures, as well as fines for building owners. We will discuss the effects of these measures, both in terms of reducing the number of unnecessary and false alarms, but also in terms of how they may affect the overall safety.

This paper is based on interviews with firefighters and other relevant professionals, as well as on existing statistics and reports. We will discuss the effects of false alarms from an organizational perspective. We will also discuss DSB's (Directorate for Civil Protection and Emergency Planning) reporting system BRIS, which was introduced in 2015 and contains all assignments registered at the alarm centrals. With this data as a starting point, we discuss how projects to reduce the number of unnecessary dispatches may be designed in a directed manner, so as to retain the positive effects of these dispatches in terms of training and reduction of risk and avoid the negative effects of implemented measures in terms of the response to real events.

2 BACKGROUND

The following section will provide some facts about the background of the paper and the kind of information and data it is built upon. It will give the definitions of false alarms and unnecessary dispatches, describe the new reporting system BRIS and also describe the qualitative data collected in order to secure a deeper understanding of the data collected in the BRIS reporting system.

2.1 *Definitions and terms*

Several different terms and concepts are used to address the issue of unnecessary and false alarms. In English-language literature, it is more common to talk about 'false alarms' for all alarms resulting in unnecessary dispatches (Chagger and Smith 2014; Karter 2013). In Denmark, it is also common to use the term 'alarm', and not 'dispatches', but there they differ between 'false' and 'blind' alarms (blind alarms being the same as unnecessary alarms). In Norway, the use of the term 'false alarm' is considered imprecise, and it is recommended to distinguish between 'false alarms' (intended), and 'unnecessary alarms' (unintended). Since the term 'alarm' can result in some confusion, the preferred mode of referring to this issue in Norway is to talk about the need to reduce 'unnecessary dispatches'. Since fire brigades responds to all types of alarms, reduction of unnecessary dispatches must involve the reduction of both unnecessary and false alarms.

This paper, will mainly use the term 'unnecessary dispatches' when addressing this issue. We have also chosen to use the term 'dispatch' and not 'response'. The term 'response' covers a range of actions instigated because of an alarm, while we, in this paper, only want to concentrate on response dispatches—when fire and rescue services vehicles rush out due to an alarm. The common practice in Norway is for the alarm central to dispatch a basic fire response unit, a standard vehicle with four firefighters, who investigate an incident at site, irrespective of their expectations of whether the alarm is genuine or not.

Another term that will be used frequently is Automatic Fire Alarm systems, or AFA-systems. An AFA system is an installation comprising detector units connected to a central. This is in contrast to a smoke alarm, which is a standalone unit. The AFAs described in the current paper are also connected to an alarm central. In Norway, these alarm centrals are called '110-centrals', and their primary task is to address emergency messages for the fire and rescue services, alerting and calling out crews, establishing connections between relevant emergency actors and logging events. The AFA system immediately, or after some pre-set delay, transfers the alarm, which in turn leads to a dispatch from the local fire and rescue service. AFA-systems are mandatory in a number of buildings where there are many people, such as nursing homes, schools and public buildings. In addition, companies voluntarily choose to install AFA-systems that are directly connected to the alarm central.

2.2 *BRIS—Reporting system*

DSB, the Norwegian Directorate for Civil Protection, introduced the reporting system BRIS (acronym for the Norwegian words for Fire, Rescue, Reporting and Statistics) from the first of January 2016. BRIS is a national reporting system, gathering information about the Norwegian fire brigades' assignments. The overarching goal of this system is to give the fire brigade a good foundation for developing targeted measures in their preventive work, develop emergency work and increase data quality in order to give local and national decision makers a better knowledge base for learning and improvement. Another goal of the system is to ascertain a more suitable user interface that allows reporting to be conducted at the event site or in the fire truck on the way back to the station.

Over the last few years, DSB and the Norwegian fire brigade have become aware of the high and increasing number of false and unnecessary alarms which leads to unnecessary dispatches. With BRIS, they have attained better statistics and data on what causes these alarms. Still, there are issues to be sorted out in terms of data quality, as the BRIS data depends on the categorisation of dispatches, and, in this respect, there are some differences between fire departments and alarm centrals.

Later on, examples will be provided as to how fire brigades can use this data to develop measures to counter the increase of unnecessary dispatches.

2.3 *Data collection*

This paper employs statistics from BRIS which can tell us something about the situation of false and unnecessary alarms and unnecessary dispatches. In addition, from March till November 2017, we collected qualitative data in:

- one large fire brigade; 4 interviews (group and single) in emergency department, alarm centre (110-central) and preventive department,
- two small fire brigades; interviews with the two fire chiefs officers,
- one emergency exercise in a small fire brigade; observations,
- the work done by a project group in a large fire and rescue service, established for developing

measures for reducing the number of unnecessary dispatches; observations, discussions and document study.

This paper also builds upon the knowledge gained in earlier projects; on the organisation of Norwegian fire brigades, and on measure development for reducing the number of fatal fires amongst vulnerable groups (Fenstad et al, 2013; Storesund et al. 2015; Gjøsund et al. 2017; Gjøsund & Almklov 2016; Halvorsen et al. 2016).

3 UNNECESSARY DISPATCHES AND ITS CONSEQUENCES

3.1 *The situation in Norway*

Data from BRIS shows that 60% of the Norwegian fire brigade's dispatches are unnecessary. On average, fire brigades in Norway had 137 unnecessary dispatches each day. The number of emergency dispatches conducted on the basis of false or unnecessary alarms varies a lot between different municipalities in Norway. Some have a proportion of 40%, whereas others have a proportion as high as 75% (numbers from 2010). There is currently little knowledge about why this is so, but it could be due to variation in the number of buildings that are directly connected to the alarm central, or to emergency controllers having different formal procedures and different assessments of response to incoming alarms. Also, variations in the demographics, housing stock and municipal organisation may count for some of the variation.

Norway consists of more than 400 municipalities that range in population from 200 inhabitants to 650 000 (Oslo). These municipalities are highly diverse in terms of their demographic profile, organisational structure and available resources. Equally diverse are the fire and rescue services, in terms of ownership, management and organisation. Some municipalities own and run their own fire and rescue services, while others collaborate with neighbouring municipalities on all or parts of the services. Large fire brigades consist of mostly full-time employees, while the small fire brigades have mostly part time employees. There are 335 fire and rescue service brigades and 620 fire stations, but we see a trend for merging brigades in order to attain larger and more specialised services. In Norway, the fire and rescue service is the only emergency agency that is municipal. It has a higher density, clear demands to response times, and thereby exhibits shorter response times than the police and health service.

Most of the alarms resulting in unnecessary dispatches typically come from automatic fire alarm (AFA) systems going off due to different perturbations or errors, smoke detectors reacting to cooking, steam or dust, people triggering alarms by error, or, in some cases, intentional triggering. Not surprisingly there is a correlation between the size of the municipality in terms of number of inhabitants, and relative number of unnecessary alarms due to AFA-systems, and the main reason thereof is most likely that small places have fewer objects directly connected to the alarm central than larger places do.

3.2 *Comparison with Denmark*

In Denmark, they have a reporting system very much like BRIS. It is called ODIN (acronym of the Danish words Online Data Registration and Reporting System). The latest version, complete with a registration of information about AFA-systems, was released July 2015. This means that the first year of complete data is 2016, the same as in Norway. In Denmark, 44% of the dispatches were due to false or unnecessary alarms, and 42% out of all alarms came from AFA systems. Of these AFA alarms, only 9% were real, the rest were categorised as blind (or unnecessary). In 2016, there were dispatches to over 6000 addresses. 143 addresses had 2507 dispatches alone. This means that 2.3% of the addresses had 16.3% of all dispatches.

As in Norway, the blind alarms (or unnecessary dispatches) are unevenly distributed throughout the country. The variation is probably an expression of uneven distribution of institutions, and thereby buildings with AFA-systems, throughout the country (Beredskabsstyrelsen 2017). Not surprisingly, the amount of blind alarms from AFA-systems have also increased from 9 000 in 2007 to 15 500 in 2016. The interesting thing is that even if the absolute amount of blind alarms has increased, the relative amount of blind alarms has decreased, from 8 blind alarms in 2007 to 5.7 in 2016 out of 1000 detectors (Beredskabsstyrelsen 2017). This could mean that AFA-systems in general are more reliable than they were 10 years ago. We have not been able to find good data on how many AFA-systems are in use at any time in Norway, and how many of these are connected to an alarm central. (It may be possible to get approximate numbers by contacting all the alarm centrals in Norway, but his has not been part of the project's scope.) It is likely that there has been a relative decrease in the amount of unnecessary alarms from AFA-systems in Norway as well.

3.3 *Consequences of unnecessary dispatches*

The negative consequences of unnecessary dispatches are both well-known, as well as the background for trying to reduce the number of these dispatches. First of all, it is resource demanding, as an unnecessary dispatch takes a crew and a vehicle

away from other duties such as training, maintenance and makes them unavailable for other call-outs. In the case of part-time personnel, it also costs additional money, since the crew is paid extra for dispatches. There is also a direct risk connected with the traffic hazard posed by vehicles speeding with blue lights through traffic. A longer term negative effect is the possibility, and a concern noted by several of our informants, of a cry wolf effect, both in the sense of a delayed evacuation of the buildings concerned but also for the fire brigade. However, possible positive consequences of unnecessary dispatches have been very little discussed.

4 SOME FINDINGS IN NORWEGIAN FIRE BRIGADES

While it has been generally recognised that a large and increasing amount of the dispatches for the Norwegian Fire and Rescue services have been unnecessary, the introduction of the national BRIS database represented a stepwise change in the documentation of dispatches. In BRIS, all call outs from the alarm centrals are given a preliminary code by the operator, and are later given a final code by the responding unit. While there are limitations to this database as well, particularly when it comes to how to categorise the different dispatches (something we will discuss later on), it has laid the foundation for directed efforts to reduce the number of unnecessary dispatches, both at a national level and for individual fire services. The statistics clearly demonstrate that a majority of dispatches, particularly for urban fire brigades, are triggered by automatic detectors, and that in the vast majority of these cases there is no fire triggering them.

4.1 *Developing and implementing measures*

There have been increasing concerns in the fire community about unnecessary dispatches over the past years; however, it is after the implementation of BRIS that the reporting and the statistics have become more accurate, and it has been possible to attain facts about the actual causes of the unnecessary dispatches. Based on this data, the directorate of Civil Protection (DSB) has requested the local fire brigades to try to find ways to reduce these unnecessary dispatches, and several fire brigades have started this work. Since the complete BRIS-data is quite new, few measures have subsequently been implemented.

4.2 *Large fire and rescue services*

We know that a large fire and rescue service in a large city in Norway introduced an increase in the fines for unnecessary dispatches from 5500 to 8000 N.kr. They hoped this would motivate house owners to better maintain their alarm systems in order to reduce unnecessary alarms. After 8 months, there was no registered improvement, but they still hope for a long-term effect. There is however a fear that increasing fees may result in lower fire safety, which will be discussed later on.

Another big fire brigade in Norway appointed a project group to develop measures which would reduce the amount of unnecessary dispatches by 20% by the end of 2018. The average measure of unnecessary dispatches in Norway in 2016 was 59%, while the average for this fire brigade was as high as 64%. They had, in several real fires, experienced that persons living in the building did not evacuate or follow instructions because they had lost respect for fire alarms due to too many unnecessary alarms. Through BRIS-data, they found that they had dispatched almost 2400 times to unnecessary or false alarms throughout 2016, and that less than 10% of these dispatches were the result of false alarms. Further, they found that a majority of the unnecessary dispatches were to addresses with Automatic Fire Alarms (AFA). There were 532 unnecessary responses to only 63 addresses, between 5 and 28 dispatches to each address. They found that the most common causes for unnecessary dispatches were: wrong use or placement of detector (34%), technical or unknown error in the AFA-system (19%), intentional false alarm (11%), dust due to construction work (8%) and water steam (7.5%). The project group in this fire brigade therefor decided to concentrate on and target measures towards countering these worst cases, noting that, if they succeeded, a reduction in such cases alone would make them reach the target of 20% decrease in unnecessary dispatches. The project group developed targeted measures for reducing unnecessary dispatches. The three most important are: 1) The provision of information and guidance to house owners/residents on how to avoid unnecessary alarms, 2) Focus on unnecessary alarms under the fire and rescue services supervision of AFA-systems, and 3) Particular and direct follow-up on the worst cases.[1]

The tendency of there being a large amount of dispatches to a few addresses with AFA-systems installed is the same as that found in Denmark.

4.3 *Small fire and rescue services*

None of the small fire and rescue brigades we talked to had started to implement measures for reduc-

[1]These number and facts are taken from BRIS-database and from a presentation the mentioned project group held in Forum for Fire Safety in Trondheim, Norway, 14th of November 2017.

ing unnecessary dispatches. In contrast to the big fire and rescue services, they did not see the same need to reduce them. The main reason for this was that the total amount of alarms was not very high. When the total number of dispatches is low, smaller fire brigades do not get the same opportunity as large fire brigades to maintain their skills through dispatches. Since most of the firemen in the small brigades are part-time firefighters, unnecessary dispatches are an important part of their training, much more than for the full-time fire fighters who meet and train together each day. If the number of unnecessary dispatches were reduced, they were afraid that their skills would decrease. Part-time fire personnel can earn as little as 20 000 N.kr a year, even if it is a requirement that they have the same skills as full-time fire fighters. However, they earn extra pay for each dispatch. Even if it is an assumption that small fire brigades are reluctant to reduce the amount unnecessary dispatches because of the earnings, this was never mentioned as a reason by the small brigades.

The regulation of the dimensioning of the fire department scales the size of the brigade depending on the size of the population. Places/municipalities with less than 3000 inhabitants have small fire brigades consisting of part-time personnel without fixed schedules, while big cities only employ full time fire personnel. For the small fire brigades, it is very difficult to possess the same specialised knowledge and skill-set as the big fire and rescue services with full-time employees. But, one advantage that public services in small places have is their situated and local knowledge. Part-time fire personnel have knowledge about houses and people from their regular jobs, as well as through neighbours and friends living in the small place. They are very much aware of this advantage, and try to strengthen it. In the interviews, they said that the unnecessary dispatches are a valuable source for gaining an even deeper local knowledge of the special objects and people in their neighbourhood.

5 UNINTENDED IMPLICATIONS?

As we have seen, the overall goal for the Norwegian Directorate for Civil Protection (DSB) is to decrease the numbers of unnecessary dispatches. Also, for large fire and rescue services this is an explicit goal, and they have started to develop and implement measures in order to reduce these kinds of dispatches. There are, however, some questions that it is necessary to reflect upon when it comes to deciding which kind of dispatches are to be reduced, and also if some measures could have unintended negative consequences.

5.1 *The big increase and uncertainty in categorisation*

As mentioned, there has been a large increase in unnecessary dispatches in Norway, from 20 000 in 2013 to 50 000 in 2016. Differences in reporting practices are central in order to understand this increase and also the implications of reducing the unnecessary dispatches. It is reasonable to think of at least three reasons for the increase: 1) The increase correlates with the number of detectors connected to the alarm centrals, which explains the majority of the increase, 2) The introduction of the BRIS registration system has led to a more accurate and systematic registration of not only real alarms but also of the unnecessary alarms and dispatches, and 3) Those who register the dispatches in the BRIS system are uncertain of how to categorise them, and systematic biases can occur. E.g. there may be different assessments of whether an alarm is unnecessary when the alarm is the result of smoke related to making food. In interviews with firemen and employees at alarm centrals, we have been told that it can difficult to categorise borderline cases, and that systematic tendencies can occur in the reporting of such incidences. One concrete example concerns food making—if the dish is still edible, the dispatch is to be categorised as unnecessary, and if it is not possible to eat the food, the dispatch is to be categorised as necessary, even in cases where there is no fire, since it can be seen as a preventive dispatch. But who is to decide if the food is edible or not? Further, is the dispatch unnecessary if it prevented further development of the situation or if the dispatch reduced the probability for this kind of scenario happening again?

Since the complete BRIS-data is available only for 2016, it is not yet possible to determine how much of the increase in unnecessary dispatches is due to these different variables. In a few years, however, it will be possible to evaluate the numbers and variables available in this registration system.

5.2 *Possible positive consequences of unnecessary dispatches*

An important question to ask when considering measures to reduce the number of unnecessary dispatches is whether there are also some positive effects. In some cases, one may expect that a certain number of mobilisations can be useful for training fire crews. As long as the dispatches are not too repetitive, they may provide the organisation with important training and expose it to a variability of scenarios that strengthens its resilience beyond the limits of normal training procedures. This is especially important for smaller fire and rescue departments with part-time employees and few dispatches.

Since real fires are rare, everything they can learn about emergency preparedness and risk is considered useful, and the unnecessary dispatches are used in the mobilisation phase for training in basic skills and catching up with colleagues. In large fire and rescue services, on the other hand, where the employees are with their brigade each day, unnecessary dispatches are seen as disturbing. The positive effects of unnecessary dispatches are therefore more obvious in small, rather than in large brigades.

Also, many of the unnecessary alarms are triggered in buildings (such as care centres for the elderly) and areas where the fire risk is high, or where there are vulnerable groups, so the dispatch to these locations might provide useful knowledge for future scenarios. It might provide information regarding incumbent risk, as in Turner's (1976:381) "incubation period", strong or weak signals that might trigger pre-emptive measures, or other forms of learning for the fire department. As such, it might help the fire crews prepare for future genuine alarms and can also be a way to prevent false or unnecessary alarms in the future. It can also be considered as an exercise for people who have to evacuate buildings. Studies has shown a reduction in evacuation times through the repetition of evacuation drills (Hamilton et al. 2017).

Thus, though a dispatch may be unnecessary, seen as an isolated response, it might contribute to reducing risk in some ways, and in some cases. Measures towards reducing the numbers of unnecessary dispatches must be seen in the light of these possibilities for learning and risk reduction, as well as the cost and risk associated with the dispatches.

5.3 *Possible negative consequences of measures*

It is possible that measures implemented to reduce unnecessary dispatches may directly or indirectly harm the emergency preparedness or the response time to some types of fires. Fines may, for example, lower the threshold for disconnecting automatic sensors, or for not installing them in the first place. As mentioned above, if the measures are successful in buildings where there is a greater probability of a fire starting, like buildings housing vulnerable groups (Gjøsund et al. 2017), the measures for reducing unnecessary dispatches could indirectly result in the loss of valuable knowledge, possibilities to detect fire hazards and chances to suggest and implement fire preventive measures amongst vulnerable groups and exposed houses.

Because of the reporting system, BRIS, it has been possible to develop more appropriate and targeted measures for reducing unnecessary dispatches. Since the measures for large fire and rescue services are in an early phase, and because BRIS was implemented relatively recently, there are currently no clear results indicating the impact or the implications of the measures. What is certain is that the reporting system BRIS will be a useful tool when analysing the results and implications of the implemented measures.

6 CONCLUDING REMARKS

By developing measures, like those of the large fire brigades, which are directed towards the "worst cases" (i.e. those with several unnecessary alarms due to dysfunctional AFA-systems), the chance of succeeding is greater than if there were no such measures. It is also likely that this strategy for reducing unnecessary dispatches will have effects on the recurring incidents that will not give valuable knowledge, while the dispatches which can give valuable knowledge (for instance, about vulnerable groups or local and demographic knowledge) are more likely to persist.

One should not regard unnecessary dispatches as a homogenous category and implement measures blindly aimed at reducing the numbers. Rather, the detailed statistics of BRIS provide the opportunity for more targeted and more effective reduction measures, by which one can also avoid some possible pitfalls that might lead to an increase in overall risk. We have suggested two such categories: 1) Measures may lead to increased fire related risk if they lead to responses on the user side to reduce the number of alarms, for example by removing sensors altogether, by reducing their sensitivity or increasing the time before they are triggered, and 2) Measures may lead to a decrease in dispatches that are unnecessary in the sense of putting out an actual fire, but that may have other positive consequences in terms of training for the personnel, or as opportunities to obtain a better knowledge of vulnerable buildings.

Still, as our case study has shown, this leaves many unnecessary dispatches to be eliminated. The project we studied addressed frequently recurring dispatches to certain buildings in a targeted manner, dispatches that clearly did not fall under the categories where the side effects might be negative. In sum, then, we conclude that measures to reduce the number of unnecessary dispatches is important, but that they should be implemented based on a careful evaluation of the potential negative effects they may have. Evaluating the different categories of reported data in terms of the positive and negative effects that measures have on a total fire risk, regarding both the likelihood and consequences, is an important first step.

REFERENCES

Beredskabsstyrelsen [Danish emergency Management Agency], 2017. *Analyse: Automatiske brandalarmanlæg* [Analysis: Automatic fire alarm systems]. Report July 2017.

Chagger, R. & Smith, D. 2014. *The causes of false fire alarms in buildings. Briefing paper, report number BC 2982.* BRE Global Ltd.

DSB [Directorate for Civil Protection and Emergency] 2017: Oppdragsstatistikk fra BRIS [Assignment statistics from BRIS]. In *DSB-report 2017.* Available at: https://www.dsb.no/rapporter-og-evalueringer/oppdragsstatistikk-fra-bris-forste-halvar-2017/.

Fenstad, J. et al. 2013. *Framtidens brann-og redningsvesen*.

Gjøsund, G. et al. 2017. *Vulnerability and prevention of fatal fires*. In Walls, Lesley, Matthew Revie & Tim Bedford (eds.), Risk, reliability and safety: Innovating theory and practice. Proceedings of ESREL 2016. Taylor & Francis Group, CRC Press.

Gjøsund, G. & Almklov, P. 2016. Pilotprosjektet brannsamarbeid i bergensregionen. evalueringsrapport. [The pilot project pire cooperation in the Bergen region. An evaluation.] In *Rapport for direktorat for samfunnssikkerhet og beredskap (DSB). NTNU samfunnsforskning: Rapport 2016*.

Halvorsen, K. et al. 2017. Fire safety for vulnerable groups: The challenges of cross-sector collaboration in Norwegian municipalities. *Fire Safety Journal* 92: 1–8.

Hamilton, G.N. et al. 2017. Human behaviour during evacuation of primary schools: Investigations on pre-evacuation times, movement on stairways and movement on horizontal plane. *Fire Safety Journal* 91: 937–946.

Karter, M.J. 2013. *False alarm activity in the US 2012.* National Fire Protection Association.

Storesund, Karolina et al. 2015. Rett tiltak på rett sted. Forebyggende og målrettede tekniske og organisatoriske tiltak mot dødsbranner i risikogrupper [The right measures on the right place. Preventive technical and organizational measures against deadly fires for vulnerable groups]. In *Report for the Norwegian directorate for civil protection*.

Turner, B.A. 1976. The organizational and interorganizational development of disasters. *Administrative science quarterly*: 378–397.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Role multiplexity and home-grown resilience: A study of part-time firefighters in rural emergency management

Petter G. Almklov
SINTEF Technology and Society, Norway
NTNU Social Research, Norway

Marie Nilsen & Gudveig Gjøsund
NTNU Social Research, Norway

ABSTRACT: We discuss the role of part-time firefighters as a resource for local emergency management in Norway. Informal social relations, the trust between practitioners and the social capital of the organization, has been recognized as a resource for emergency management, particularly as it contributes to improvisation and coordination between actors belonging to different professional groups. Likewise, social capital, the trust among citizens, has been identified as a resource for societal resilience in crises. We discuss a combination of these forms, how the social embeddedness of the emergency practitioners in the community and the multiplexity of roles is important for community resilience. These professionals know each other through several different social roles, and have resources beyond the formal capacities their position should suggest. Thus, role multiplexity and social networks provides a functional redundancy and is a resource for resilience in the management of incidents and emergencies. These abilities are hard to make visible in a work plan and challenging to include in exercises. Moreover, these abilities are affected by recent developments towards professionalization of and centralization.

1 INTRODUCTION

1.1 *Background of our study*

Based on studies of part-time firefighters we discuss how the combination of professional roles and embeddedness in the community of this group is a resource for community resilience. This is done through discussing how local knowledge, social capital and role multiplexity influences practice and decisions in emergencies.

Our paper supplements the literature on community resilience and emergency management by studying the role of a "hybrid" group. Their role as members of the community and their professional roles are both important parts of their capabilities in the prevention and management of emergencies.

We understand community resilience as the adaptive capacity of a community when faced with emergencies. Key elements of this are improvisation and redundancy (both in terms of resources and competence). This will be elaborated in the theory section.

What we describe here as role multiplexity is the fact that these professionals have different professional and social roles, that their "hat", or in the case of the firefighters, helmet, does not represent their only relevant role for the way they solve their tasks in emergencies.[1] Moreover, role multiplexity is an important contributing factor for the firefighters' local knowledge, a competency that is well-recognized in the emergency management community.

Thirdly, the notion of social capital is used to describe the networks of trust relations as a resource for coordination in emergencies.[2]

These characteristics, we will argue, have proven to be important in the management of several

[1]The notion of role multiplexity is inspired by sociological theory on modernity and bureaucracy. Whereas one individual in a bureaucratic organization has one role only, and thus uniplex relations to the ones he interacts with, in small scale communities most people have several relations to each other. (Durkheim discussed in Brøgger (1993: 26ff), see also Almklov et al, 2017).

[2]The individual sense of the term social capital is inextricably related to the works of Bourdieu (1986), viewing social capital as a source of power individuals possess and use to further their interest. The collective view on social capital is particularly associated with Robert Putnam (1995), viewing it more as a property of a group, a community or a society. We are here referring to the latter, as a descriptor of how trust based networks are resources for collective action.

emergencies in rural districts in Norway. Interestingly, they tend to elude description in formal documents, and thus risk being undermined by administrative reforms in the domain of societal safety and emergency preparedness, such as developments towards centralization and professionalization of the fire departments. Understanding and documenting the specific competence and community role of this group is important input to such processes.

1.2 *Part-time firefighters in Norway, ongoing changes*

Norway consists of more than 400 municipalities that range in population from 200 inhabitants to 650.000 (the capital, Oslo). The municipalities are highly diverse in terms of demographic profile, geography, size, organizational structure, and available resources. Equally diverse are the fire and rescue services, in terms of ownership, management, and organization. Some municipalities own and run their own fire and rescue services, while others collaborate with neighbouring municipalities either by having joint fire brigades or just in providing parts of the services. The fire departments and placement of fire stations are dimensioned after specific criteria for response times, leading to a relative high density of fire stations and shorter response times in most areas compared to other emergency services.

Whereas large fire brigades in cities and towns rely mostly on full time personnel, smaller fire brigades are largely dependent on personnel in different forms of part-time employment. For a large fraction of the latter, their regular payment only covers mandatory training and a small reimbursement for being on call, plus additional pay for dispatches. This means that they typically have full employment in other trades. The composition of the fire and rescue services in Norway today are roughly 3500 full time firefighters, and around 8000 part-time firefighters. In principle, the competence demands for part-time firefighters are supposed to be equivalent with the basic requirement for full time personnel. However, in terms of technical skills and training, they tend to lag behind these demands, while full time personnel on the other hand train and rehearse their skills way beyond them. Recruitment, both for full time and part-time personnel, has often sought people with relevant technical skills from other trades, such as carpenters, electricians and people with military training. For part-time personnel, an additional requirement is often that they live close to the fire station, to be able to mobilize quickly. In rural districts, farmers often make up a significant portion of the crew.

1.3 *Societal safety and emergency preparedness*

In Norway, the municipality has a key responsibility in terms of risk management and emergency preparedness. In principle, the municipality "owns" the total risk picture within its borders, and should have updated all hazards risk and vulnerability analyses and emergency plans. The responsibility for this in small municipalities is usually given to an emergency management coordinator, typically an official in the technical department of the municipality that has this as a fraction of their position. In terms of operative resources, the Fire and Rescue Service (FRS) is a crucial first line of response, but the emergency plans also include other municipal personnel, as well as external actors (volunteers, industry, municipal technical department etc.).

1.4 *The fire & rescue service, the only remaining generalists in local emergency management?*

Societal safety as a term and policy area came about after the end of the cold war. Gradually, the resources available for emergency preparedness in the public sector in Norway in particular, and in Western Europe more generally, have been reduced since then. Moreover, as measures have been implemented to make the public sector more effective and goal oriented, through outsourcing and market based restructuring, generalist capabilities and functional redundancy have systematically been reduced. Other operative capacities in the public sector are trimmed (army, home guard, publicly owned technical services such as roads, energy, port authorities). Within this picture, the FRS is one of very few remaining generalists with a substantial redundancy. A result of this has, according to our informants in the FRS, informants from other sectors and public reports (DSB, 2013, Øren et al. 2016) that the scope of tasks for the FRS is expanding.

In addition, in the rural areas, police services and emergency health services are generally sparse, so the FRS tend to be first on site for accidents and incidents of all sorts. This means that they sometimes must fill in for medical personnel or the police while waiting for ambulances and police patrols.

The FRS still put out fires, but increasingly they respond to other accidents (in particular traffic accidents) and other emergencies. Due to the reduced operative redundancy in the public sector generally they are becoming increasingly important as first responders to other forms of emergencies, such as landslides, floods and storms and search and rescue operations. For many areas, climate change leads to increases in flash floods and associated landslides as well as an increasing risk of forest fires.

Another important development for the FRS is the implementation of the joint communication network for emergency services in Norway called Nødnett (lit. "emergency net"). This means that all part-time firefighters have a communication radio at home, serving both as a call out terminal and as a communication tool in operations. Their training in using these is an important resource for the coordination in emergencies both for the FRS itself, but also as they may act as liaisons with other municipal professionals (Tilset et al. 2015). In some FRS, the firefighters are supposed to have the radio nearby at all times. These radios further integrate the FRS with other emergency services, as they may communicate in shared working groups with the police, health services and other relevant actors. The FRS are owned by the municipalities, but often, and to an increasing degree, they are parts of inter-municipal collaborative arrangements. It is an explicit national strategy to increase the size and professionality of the FRS, and there are several ongoing changes in the sector. Understanding the unique role and competencies of the part-time fire fighters will be important to ensure that these changes are successful.

2 THEORY

2.1 *Resilience and robustness*

Resilience is employed here to describe how an organization, community or society absorbs shocks and 'bounces back' after a disturbance (Boin & van Eeten 2013). Resilience has become a central concept in the safety theory the last decades. One early contribution was Wildavsky's (1987) insistence that a "search for safety" should go beyond trying to mend known weaknesses, by including a creative exploration of ways to improve society's ability to sustain new challenges. Also, the descriptions and analyses of High Reliability Organizations (LaPorte & Consolini 1991, Weick and Sutcliffe 1995, see also Roe and Schulman 2008) stressed that designing robust systems was only one step of the way to achieve high reliability, stressing the need for redundancy, flexible organizing and organizational mindfulness to be able to cope with variability. The most prominent theory on resilience within safety research is found in the "resilience engineering" strand of research, where an intense focus on the management and learning from variability as a resource for safety has been a cornerstone (Hollnagel et al. 2006).

Outside the safety literature, the concept of resilience has also been important in studies on a societal and community level (e.g. Boin & van Eeten 2013), then often referring to the community or society's ability to bounce back (or even forwards) when confronted by major disasters. The literature on community resilience is broad and diverse within several research fields. (See e.g. Norris et al. 2008 for some background). In contrast, Resilience Engineering focuses heavily on the importance of resilience as a way of avoiding accidents.

Based on a study of the response to the 9/11 terror, Kendra & Wachtendorf (2003) identify some characteristics of resilience. These are redundancy, resourcefulness, effective communication, and the capacity to self-organize, undeterred by extremely challenging circumstances. They point out that resilience is essentially a set of attitudes concerning expediency of actions and the propensity to acquiring new capabilities.

There is a big difference between a well laid out plan and a plan that is well played out. The former points to the ability to foresee and predict while the latter refers to the ability to act when the situation calls for the use of a plan. Plans cannot guarantee the success of how emergencies are handled. They can only provide the backbone of an emergency response. A common example employed to explain resilience is to compare hard wood to bamboo. The former is strong and does not easily break when it encounters strong winds. This is, of course, up to a certain threshold. The robust hardwood tree will eventually break if the wind blowing is at hurricane-strength. In comparison, the bamboo sways with the wind. It has the ability to bounce back into place. This ability to bounce back is what defines resilience. It is able to adapt. A plan cannot be made for every single possible emergency situation in a community. It is the ability to adapt the plan according to the situation as it unfolds which will help a community to bounce back after a disturbance (Boin & van Eeten 2013).

The part-time firefighters add flexibility in the community response to local emergencies. There are at least three aspects that contribute to this flexibility: the firefighters' diverse backgrounds and experience (providing a functional redundancy), their local knowledge and proximity to the hazards and, that they possess a rudimentary organizational structure and means for communication and their social embeddedness in the community (easing swift coordination with volunteers and other external resources). Interestingly there is a good overlap between these characteristics and those identified by Kendra & Wachtendorf (2003). Though this is interesting, one should also draw comparisons between such different contexts with caution.

2.2 *Role multiplexity, social capital and community resilience*

Part-time firefighters have many ties in the community. They are members of their local fire

brigade. They are parents of children attending the local school, a colleague in the municipal organization or electricians, plumber, factory workers or farmers. They may also be members of the sports clubs, hunting groups, health professionals, janitors or a loyal customer of the local grocery store. They have social relations throughout the community and local knowledge of threats and resources. The repeated interaction and networks built in their local community, over time, develops and strengthens their social capital. Social networks, reciprocity and interpersonal trust are aspects that are critical to building social capital (Patterson et al. 2010). Social capital and networks among citizens are recognized to be critical to disaster survival and recovery (Aldrich & Meyer 2015) on a societal level. Importantly, here we also include the networks that go between the response organizations and the community, and that criss-cross organizational boundaries in the community (Almklov et al. 2017).

Local knowledge is an important element in disaster management. For instance, it can help build resilience to flooding in local communities by providing local information on actual flood patterns, frequency, and risk perceptions in the community (Ramsey et al. 2016).

While bureaucratic organizations are built around uniplex roles, where the person is his role and that is the only relevant feature (see Almklov et al. 2017, Brøgger 1993). In practice, however, we see, particularly in small communities that there are spill-over effects from other roles that the emergency professionals have in the community, and that these are often key both in terms of establishing trust relations that go beyond the formal relations and also that the multiplexity of roles provides the individual with a functional redundancy, in terms of knowledge and capacities. In the empirical section, we will give some brief examples of this.

3 METHODS AND DATA

This paper is based on an aggregate of data from several projects inspecting the roles of municipal emergency preparedness and the organization of the FRS in Norway: A study with scenario analyses for the future organization of the FR services in Norway (Fenstad et al. 2013), a study of different approaches for intersectoral collaboration between the fire departments and municipal services in prevention of deadly fires (Gjøsund et al. 2016, Halvorsen et al. 2017), a study of the implementation of Nødnett in Norwegian municipalities (Tilset et al. 2014), a process analysis for a project to improve regional collaboration between large and small FR services in Western Norway (Gjøsund & Almklov, 2017), and a study of municipal emergency preparedness (Øren et al. 2016). All these projects have been based on interviews with firefighters and personnel that they interact with on a daily basis. While the scopes of these projects have been diverse, they have all contributed pieces to the puzzle regarding the role and qualities of part-time firefighters.

We have, for the purpose of writing this paper, conducted five directed interviews with key informants, two leaders and one firefighter at part time fire department and two with fire fighters in an urban fire department that has some part time personnel affiliated. We also conducted an observation study (with some informal discussions along the way) of a training session with a part-time FRS. During a one day visit (observation and interviews) in a regional dispatch central, our discussions included the capacities, call out procedures and response times of the part time FRS under their control, and also their typical assignments. In addition, we studied reports from a selection of recent emergencies in Norway.

4 EMPIRICAL EXAMPLES

In this section we give some examples from our data, supplemented with reports from recent emergencies, to illustrate how the role of municipal firefighters in contributing to community resilience can be characterized by the concepts of role multiplexity and social capital.

4.1 *Two illustrative examples*

One fire chief leading a fire brigade in two rural municipalities explicitly valued of the varied competencies of his crew. The long distances meant that his part-time firefighters had to mobilize for all sorts of accidents, and he had employed personnel that who's day jobs were in the medical sector, i.e. nurses, adding to their ability to respond to both medical emergencies and to take better care of elderly people in trouble. But also, other occupations had qualities he valued. He told us about an accident on a farm where an old farmer had fallen and had to be rescued from a silo by the fire department and be evacuated by helicopter. When the helicopter and ambulance had left and the firefighters were demobilizing and consoling the old man's wife, they heard the cattle were in distress. They had not yet been milked. His crew, consisting of several farmers, would not leave the site before milking the cows, he said. Farmers just don't leave cows in distress! This example might seem little relevant, too trivial, for grand discussions of emergency management. However, it illustrates how their knowledge and professional values from other occupations spill over into their role as firefighters.

Another example from the same fire department illustrated the role of local knowledge and role multiplexity. An avalanche had hit a road, blocking the road and possibly covering some cars. While the police's operational leader, formally in charge of the rescue operations, was speeding to the site from the closest (yet distant) city, the fire department was first on site, starting a search and rescue on their own initiative. The leader of the first vehicle, realizing that they needed equipment to search under the snow, went to his other workplace, a skiing facility, to pick up search and rescue gear there. Thus, the firefighters had mobilized the necessary equipment before the other services even reached the site, again because of the knowledge and access to resources provided by the firefighters' day-jobs. It also illustrates the make-do attitude and improvisational skills they have to their job.

4.2 *The fires of 2013*

Some very prominent examples in the recent discussions of part-time firefighters and municipal emergency manager's improvisational skills in Norway are the fires in 2013. That winter had had a very rare weather situation, with a very dry winter with little snow in the normally humid coastal areas, leading to bushfires in winter, and a major fire in the wooden town of Lærdal. All these fires were testing to the local fire departments ability to mobilize, organize and execute an efficient response, and in the aftermath their effectiveness has been subject to debate (See Andresen 2017, PWC 2014).

Even before the fire in Lærdal started some of the firefighters had concerns to the fire risk due to the combination of strong eastern winds and dry weather. Thus, there was already an increased awareness before the fire started. This is underscored by the fact that the municipality on earlier occasions had implemented fire watches when this weather combination occurred, so this type of weather in was a risk recognized by the locals (Andresen 2017).

When the firefighters mobilized, the response had a very improvised nature, seamlessly integrating volunteers (such as farmers with manure spreaders) in the response. As many firefighters were municipal employees, they also had good knowledge of available resources, such as access to the waterworks (to get pumps started when the electricity failed) and equipment. When telecomservices and electricity was lost, coordination was done by improvised means. Evacuation was greatly helped by the firefighters' and volunteers' knowledge of where vulnerable people lived.

The response was improvised and organic. Representatives from the larger fire brigades that eventually assisted the firefighting, and the national fire authority (DSB), noted the lack of organization as a shortcoming, while the local community highlight the effectiveness in the improvised response. Both views have some have support in the eventual investigations. The initial response was clearly effective, and several problems were solved in creative ways, but as the fire grew and as more and more resources arrived and needed to be coordinated (without effective means of communication as the Nødnett broke down), the coordination based on local knowledge and social networks became less efficient.

The response to the Lærdal fire clearly shows how part-time firefighters may act as an integrated part of a closely-knit community, and how their role multiplexity and social embeddedness in the community proved invaluable resources for their response. However, it also shows that this mode of organizing has shortcomings in terms of tactical leadership and coordination of resources when the control span grows.

Similarly, an external evaluation of the response to the bushfire in Flatanger the same winter pointed to a deficient plan compensated by good local knowledge and well-oiled collaboration machinery in the region. According to the investigation report, the crew exhibited their willingness to go beyond what was expected of them despite the notably harsh conditions. Some of the firefighters even lost their homes while extinguishing fire to save other people's homes (PWC 2014:51).

4.3 *Responses to other emergencies*

An ever more common type of emergency in Norway the last decades are seasonal floods and flash floods, and water induced landslides. (DSB 2013, Fenstad et al. 2013) In particular flash floods and water induced landslides are commonly associated with climate change, as this leads to more intense precipitation.

In rural communities, the combination of part-time firefighters and volunteers (farmers and others with access to machinery) are the core first responders to such events. Again, the personal networks and role-multiplexity of firefighters and municipal employees, provide a combination of a rudimentary organization and access to resources beyond the standard gear possessed by the fire department. Moreover, floods and landslides are events that typically happen on locations that are known by the locals to be risky, so local knowledge is important both in prevention and response. During floods, the local fire department usually assists in pumping out water that has flooded buildings. Knocking from door-to-door, firefighters also often perform the task of informing the locals of possible flooding in the area, and hazards

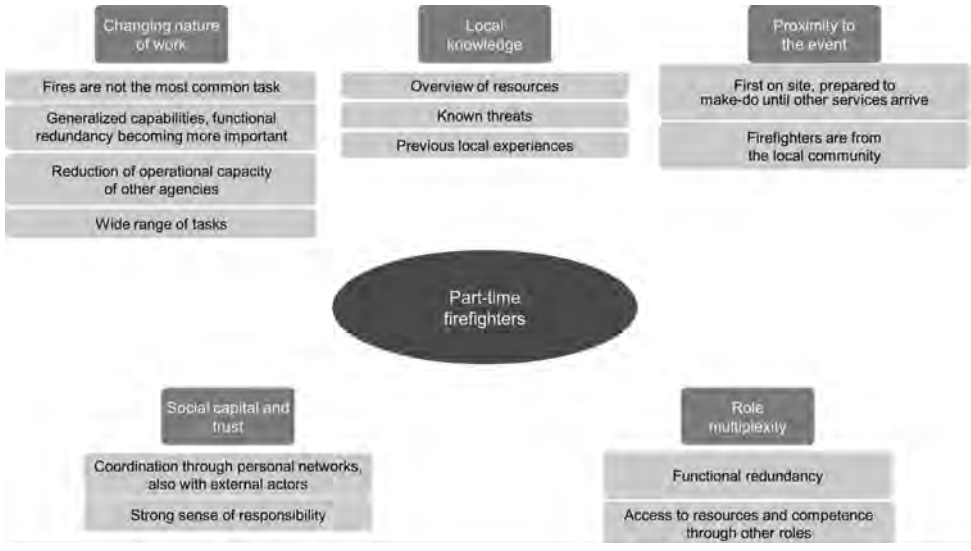


Figure 1. Summary of our description of the characteristics of the part time firefighters.

posed by landslides. The FRS also clears out trees that may pose a risk to the public or impede traffic. Other examples of notable cases of the rural fire department's new challenges is the 2013 triple murder in on a regional bus in Årdal. There, the fire department, together with ambulance personnel, managed to keep the perpetrator under control while the police response was severely delayed. In September 2011, a train at Rørosbanen was derailed. The first emergency personnel to arrive on the scene was the local part-time FRS. They started the evacuation of injured passengers and cared for them until the paramedics arrived. They also cleared out an evacuation path and organized an assembly point where the evacuated passengers were registered before they were allowed to leave the area.

The cases here illustrate some of the variation and complexity of the tasks facing these fire and rescue workers. City firefighters face some of the same complexity, but they have more support from police and health services and other professionals and experts. This difference is also a source of the respect city firefighters have for part time crews. The demands for generalist competencies are higher for the part time crews. The part time personnel are sometimes (incorrectly) referred to as volunteer firefighters, but their response is based on a rudimentary organizational structure and basic training, and it is also better integrated with more professional responders than most volunteers. In a discussion of volunteers' response to the large storms in the southern US (Katrina and Harvey), Wachtendorf & Kendra (2017) stress the importance of such coordination for the efficiency of volunteer responses. Though the part time FRS are not volunteers in the strict sense of the word, the part time FRS can be regarded as a hybrid form of response, connecting the volunteer community and official response to events.

5 DISCUSSION

5.1 *Are part-time firefighters only part firefighters, or do they bring some unique resources to the table?*

Based on our studies in different fire departments in Norway, it is clear that the part-time firefighters are less skilled and have less training for advanced firefighting than their full-time counterparts. The part-time fire departments also have shortcomings on formal communication procedures and on the management side, particularly when faced with larger incidents requiring coordination outside of their personal networks.

One should be very surprised if this was not the case that these firefighters lacked some skills, as they have highly limited time for courses and training. This is also noted in the national "Fire Study" (DSB, 2013). They also lack resources in terms of equipment, and several firefighters lack formal qualifications. Moreover, due to the low frequency of call outs, many of them struggle to gain practical experience with firefighting. There is no shortage of problems in these fire departments. Notorious underfunding has also led to some FRS

operating antiquated vehicles. Still, when we talk with full time firefighters and other professionals in the emergency community, they generally have great respect for the part-time crews for their general skills and for their ability to solve their tasks. One city fireman described how impressed he was by a nearby part-time FRS near an accident-ridden highway, how they responded quickly to horrible accidents, and how they on their own initiative had started taking first aid courses as a response to the slow response of the ambulances in that area. Their high motivation, and sense of responsibility was generally recognized by several of our informants.

The part-time firefighters are seen, by full-time firefighters, as generalists with improvisation skills based on their additional occupations. Also, the full-time fire departments have traditionally strived for this quality, by actively recruiting people with a variety of professional backgrounds, but this is even more pronounced in the part-time corps. Their local knowledge (of terrain, threats, resources, people and buildings) and embeddedness in the social fabric of the community is important for their ability to respond to emergencies, and their day job is sometimes an important resource.

As professional firefighters, they are inferior to the well drilled crews that practice every day, but they have other qualities that should not be underestimated, and that should be evaluated in the larger context of societal safety and community resilience, not only as actual fire fighting, which is only a minor part of their task portfolio.

5.2 *Role multiplexity and part-time resilience*

The fire-department in general, and the rural ones in particular, are organic parts of their communities. The qualities of the part-time firefighters that we have discussed here are important parts of what we have labeled community resilience. Their social relations in the community make them more effective than their fractions of positions may suggest, and they make up a critical part of the local communities' ability to withstand and respond to emergencies of a highly varied nature. We introduced two sociological explanations for this:

Role multiplexity: They have many hats, many forms of competence which give them a broad skill-set, competence and access to resources when faced with novel situations. In particular professional roles such as jobs in the municipality's technical services, farmers, carpenters or medical professions gives highly valued additional competencies.

Social capital: They have social networks, trust relations, that can be very useful for coordination in emergencies, and also for mobilizing equipment and resources. This also contributes to high motivation. Informants throughout the sector are clear that the economic incentives, the paycheck, is not the primary motivation for most of the firefighters. Rather, it is the sense of doing an important job for the community that is the main motivation for most.

6 CONCLUSION

One conclusion, not a very daring one, is that we (researchers and especially policy makers) need to know more about the specific role of these firefighters as Norway is about to restructure our fire departments. They might not be as competent as full-time fire fighters, but they are different and fill other roles and are organically involved in community preparedness and response. From a societal safety perspective, the part-time fire fighters are possibly the most cost-efficient operative emergency management resource in Norway.[3]

Beyond the discussion of firefighters in Norway, our paper emphasizes the importance of role multiplexity and social capital in the management of societal emergencies. Part time firefigthers are not volunteers in the traditional sense but not fully professional actors either. For effective emergency management, they represent an important hybrid resource as they both possess rudimentary means in terms of coordination, communication (most importantly by being equipped with and trained to use radio communication terminals) and leadership while simultaneously being well engrained in the social fabric of the communities they serve.

REFERENCES

Almklov P., Antonsen, S, Bye R. & Øren, A (2017) Organizational culture and societal safety: Collaborating across boundaries. Accepted for *Safety Science*. Special issue on societal safety.

Aldrich, D. P. and M. A. Meyer (2015). "Social capital and community resilience." *American Behavioral Scientist* 59(2): 254–269.

Andresen, S. A. (2017). In the heat of the moment: A local narrative of the responses to a fire in Lærdal, Norway. *International Journal of Disaster Risk Reduction*, *21*, 27–34.

Brøgger, J. (1993). *Kulturforståelse*. Oslo: NW Damm & Søn Forlag.

DSB (2013) Brannstudien [The Fire Study] Report from the Norwegian Directorate of Civil Protection (DSB).

[3]We have not investigated the economics of this, but to illustrate: A fire chief we interviewed stated that he had a yearly budget of roughly 4 mill NOK (around 400.000 Euros) and could mobilize 80 part-time firemen.

Fenstad, J., Almklov, P., Ishol, H., Storesund, K., & Albrechtsen, E. (2013). Framtidens brann-og redningsvesen.

Gjøsund, G. and Almklov, P. 2016. Pilotprosjektet Brannsamarbeid i bergensregionen. Evalueringsrapport. [The pilot project Fire Cooperation in the Bergen region. An evaluation.] Rapport for Direktorat for samfunnssikkerhet og beredskap (DSB). NTNU Samfunnsforskning: Rapport 2016.

Gjøsund, G., Almklov, P., Halvorsen K. and Storesund, K. 2017. Vulnerability and prevention of fatal fires, In: Walls, Lesley, Matthew Revie & Tim Bedford: *Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016*. Taylor & Francis Group, CRC Press.

Halvorsen, K., Almklov, P. and Gjøsund, G. 2017. Fire safety for vulnerable groups: The challenges of cross-sector collaboration in Norwegian municipalities *Fire Safety Journal*, 92 (2017) 1–8.

Kendra, J. M. and T. Wachtendorf (2003). "Elements of resilience after the world trade center disaster: reconstituting New York City's Emergency Operations Centre." *Disasters* 27(1): 37–53.

Norris, F. H., Stevens, S. P., Pfefferbaum, B., Wyche, K. F., & Pfefferbaum, R. L. (2008). Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. *American journal of community psychology*, 41(1–2), 127–150.

Patterson, O., Weil, F., & Patel, K. (2010). The role of community in disaster response: conceptual models. *Population Research and Policy Review*, 29(2), 127–141.

Putnam, R. D. (1995). Bowling alone: America's declining social capital. *Journal of Democracy*, 6(1), 65–78.

PWC (2014) Evaluering av brannene: Lærdal, Flatanger og Frøya [Evaluation of the fires: Lærdal, Flatanger and Frøya] Report by PriceWaterhouseCooper for the Department of Justice and Emergency Preparedness.

Roe, E., & Schulman, P. R. (2008). *High reliability management: Operating on the edge* (Vol. 19). Stanford University Press.

Tilset, H. D., Fagerholt, R. A., Almklov, P., Bisio, R., & Reegård, K. (2014). Nødnett i norske kommuner: Erfaringer fra de første fasene. [Emergency communication networks in Norwegian municipalities.] Report for KS.

Øren, A., Wasilkiewicz, K., Mohammad, A. B., Almklov, P. G., Albrechtsen, E., Antonsen, S., & Schiefloe, P. M. (2016). Kommunal beredskap-hva mener kommunene? [Municipal emergency preparedness. What do the municipalities mean?] SINTEF Report.

Wachtendorf, Tricia (2017) "Cajun Navy" rescuers in Hurricane Harvey show vital role of volunteer boats. Op ed in Salon magazine. http://www.salon.com/2017/09/03/cajun-navy-rescuers-in-hurricane-harvey-show-vital-role-of-volunteerboats_partner last accessed 19/9–17.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Reorganization and downsizing in the petroleum sector

L.I.V. Bergh, R. Høydal & J.E. Tharaldsen
Petroleum Safety Authority Norway, Norway

C. Aagestad & T. Sterud
STAMI—The National Institute of Occupational Health, Norway

ABSTRACT: In the petroleum sector low oil prices, high cost pressures and rapid technological development have led to higher demands for cost reductions, reorganisations and employee downsizing. The aim of this paper was to: 1) Assess trends and development related to reorganisations, downsizing, job insecurity, psychosocial factors and safety climate from 2013–2015; 2) Explore the correlation between reorganisations, downsizing and psychosocial factors, safety climate and occupational injury in 2015. The study is based on cross-sectional survey data collected every 2nd year, involving all offshore and land based personnel in the petroleum sector. Results show that employees have experienced an increase in reorganisations, downsizing and job insecurity in the period 2013–2015. Furthermore, the results show that employees who have experienced reorganisation and downsizing report higher risk of occupational injuries, poorer safety climate and psychosocial work environment, compared to employees who do not report such changes. The analyses also indicate that the higher risk of injuries reported by those who have been affected by downsizing and reorganisation may be associated with a poorer safety climate and psychosocial work environment.

1 INTRODUCTION

Over the years' significant changes have taken place in the petroleum industry. Technological changes and fluctuations in market dynamics have caused demands for cost reductions, organisational restructuring and downsizing (PSA 2017a, ILO 2013). Organisational changes at the workplace is a wide-ranging concept which often have implications for the way work is designed, organized and managed, also referred to as psychosocial work environment (Leka & Jain 2010). Over the years, a growing amount of research has explored the link between psychosocial and organisational factors and incidents in the petroleum industry (Bergh et al., 2014, Cricthon 2005, Mathisen & Bergh 2016, Olsen et al. 2015, Kongsvik et al. 2011).

Organisational change that influences the employees' job situation and involve modifications of the core system of an organization, including values, work practices, organizational structure and strategy, are likely to affect employees' well-being more than less-pervasive changes (Bamberger et al. 2012). Among the most dramatic organisational changes are restructuring, because these type of processes may involve aspects such as relocations, offshoring, closure, mergers/acquisition, outsourcing and internal restructuring including downsizing (Eurofound 2014, Mathisen et al. 2016). A review study performed by Quinlan and Bohle (2009) found that downsizing was associated with job insecurity and negative impact on health and safety. Furthermore, organisational change processes have been regarded as an important aspect in accident investigations, and might be highly relevant to the level of both process and worker safety (Baker, 2007; Grote, 2008). However, research on the relationship between organisational change processes, such as restructuring and downsizing, and risk of major accident is still limited (Grote, 2008; Koukoulaki 2009, Lofquist 2008).

Changes in the work environment are an everyday reality in the petroleum industry, partly due to demands for efficiency and cost-reductions. Development in knowledge, technology, and society brings about unavoidable changes in our working practices and experiences. Based on what we know from research, there is reason to assume that these changes can lead to increased risk level in the petroleum industry. It is important to understand the impact of the changes, which furthermore enable us to implement appropriate actions to respond and adapt appropriately to these changes, learning from experiences. The RNNP questionnaire survey is a data source that can help to highlight key issues related to these concerns.

In light of this, this study assessed the impact of organisational changes on psychosocial work

environment, safety, health and occupational injuries develop in the petroleum industry. To explore this topic the results from the RNNP survey (Trends in risk level in the petroleum activity) was utilized to:

1. Assess trends and development related to reorganizations, downsizing, job insecurity, psychosocial factors and safety climate from 2013–2015.
2. Explore the correlation between reorganizations, downsizing and psychosocial factors, safety climate and occupational injury in 2015.

2 METHOD

2.1 *Sample and measures*

This study utilized survey data from personnel on offshore facilities on the Norwegian Continental Shelf (NCS) and onshore installations, covering all employees offshore and onshore in the petroleum industry. The RNNP questionnaire-based survey is conducted as a cross-sectional study every other year, for the first time in 2001 and most recently in 2015. The analyses in our study are based on data from 2013–2015 with a response rate in 2015 of 27% for onshore installations and 29.7% for offshore facilities (PSA, 2016a). In spite of a rather low response rate, from year to year the sample is relatively stable over a range of variables such as gender, age, facility, the area of work, ratio between operators and entrepreneurs, permanent and temporary employees and proportion with managerial responsibilities. The RNNP data provide a good comparative basis for survey analyses from year to year. The large number of responders in the survey contributes to make the results more robust. It should be noted that the regression analysis used data from the last offshore surveys in 2015 (n = 8509).

The analyses included indicators related to reorganization, downsizing, job insecurity psychosocial factors and safety climate (RNNP, 2016). The following item measured reorganizations: "During the last year, have you experienced reorganisations that affect the way you plan and/or carry out your work on the facility?". The response scale for this item were: "Yes" or "No". Furthermore, the following item measured downsizing: "During the last year, has your workplace been subjected to workforce reductions or redundancies?". The answer categories were: "I have experienced reorganisations with significant consequences" to "I have not experienced reorganisations".

Based on the results from a factor analysis described in the RNNP-report from 2014, four factors that measure the psychosocial work environment were established. The psychosocial work environment dimensions measures employee's subjective experience of job demands, job control, supporting leadership and supporting colleges and was measured with five scales. *Job demands* was measured with the three items ($\alpha = 0.65$). *Job control* was measured with three items ($\alpha = 0.77$). *Supportive leadership* was measured with the three items ($\alpha = 0.76$). Finally, social support from colleague was measured with two items ($\alpha = 0.62$).

The answer categories were: "Very rarely or Never" to "Very often or always". The mean scale score was converted into three categories: Low (1.0–2,0), medium (2.1–3.0) and high (3.1–5.0). By dichotomizing job control (low = 1, medium and high = 0) and quantitative demands (high = 1, medium and low = 0), we constructed a job strain variable of the combination of low job control and high job demands (yes = 1, no = 0). All variables were coded so that high exposures indicate assumed negative exposure such as high job demands, low job control, low supportive leadership, and low support from colleagues.

Safety climate was measured by using an index consisting of 11 items divided into three dimensions. The items are described in greater detail in (Birkeland, et al, 2013; Tharaldsen, et al, 2008). These dimensions measure the employee's assessment of: 1) *Individual intentions and motivation* ($\alpha = 0{,}73$) 2) The *management's prioritization* of safety ($\alpha = 0{,}73$) and 3) *Safety routines* ($\alpha = 0{,}74$). The response scale for each item was: "Fully agree" "Partially agree" "Neither agree nor disagree" "Partially disagree" "Fully disagree". The statements are both positive and negative. The scales are therefore reversed on some of the items and higher value indicates a poorer safety climate. The mean scale score of the 11 items was converted into three categories: Good (1–1.53), medium (1.54–2.04) and poor (2.05–5). The cut-off point (2.05) for being included in the "poor category" is the 66.66 percentile. That is, respondents who are among the third with the highest score on the index in the 2015-material are included in the poor category. The cut-off point in 2015 is as such set as a reference. In addition, occupational injuries were measured with a single item "Have you been injured in a work accident while at the facility during the last year?". For further information about the items can be found online in the RNNP summery report (RNNP, 2016a).

2.2 *Analysis*

We used the chi-square tests to determine whether there were significant changes in the levels of reorganization, downsizing, job insecurity, psychosocial factors, safety climate and occupational injuries in offshore facilities in the period 2013–2015.

RNNP data from 2015 were used to compare psychosocial factors, safety climate and occupational injuries between employees reporting downsizing and reorganizations and employees not reporting such changes in the petroleum industry (offshore and onshore facilities). For the trend analyses data from the offshore population was utilized. Furthermore, for the multiple logistic regression analysis both the offshore and onshore data was assessed. Associations between reorganization or downsizing and work injuries were calculated as the odds ratio (OR) and 95% confidence interval (95% CI). Statistical analyses were conducted using SPSS Statistics for windows version 23.0 (IBM Corporation, Armonk, NY, USA).

3 RESULTS

3.1 *Reorganizations and downsizing*

Reorganization and downsizing vary over time, however results from this study show that there has been a considerable increase in the proportion of employees who have experienced reorganizations, downsizing and job insecurity in the period 2013–2015 (Table 1). Results show similar development for the operators, entrepreneurs and the mobile facilities.

3.2 *Psychosocial work environment, safety climate and occupational injuries*

The trend analyses for 2013–2015 show a significant change in employees who report increased levels of job demands (19% to 22%) low job control (10% to 12%) and the combination of high job demand and low job control (4,3% to 5,6%), in the industry as a whole. If we look further into offshore facilities (Table 1), we see that between 2013 and 2015, there is a significant change in the percentage reporting low job control among entrepreneurs (6.8% to 9.8%). However, among operators, no change is observed in the same period, and about 10% report low job control in 2015 (Table 1). Table 1 show a significant change among entrepreneurs with regards to high job demands (16% to 20%). While for employees working for operators, no change is observed in the same period. Furthermore, entrepreneurs also report a significant change related to the combination of high job demands and low job control (2.8% to 5.4%). At the same time, no change is observed among operators. There are no significant changes in the prevalence of low supportive leadership and low colleague support among operators, entrepreneurs and among workers on mobile facilities. However, it should be noted that operators report lower leadership support then entrepreneurs and workers at mobile facilities. There are however only smaller differences in the psychosocial working environment for entrepreneurs and operators on offshore facilities in 2015. As such, there are few changes in the psychosocial working environment on offshore facilities overall, but in the period 2013–2015, a deterioration of the psychosocial work environment is observed among entrepreneurs and workplaces with a large number of entrepreneurs. Table 1 show that the proportion of employees reporting poor safety climate (see description) decreased overall until 2013, while there was a significant increase in 2013–2015 among entrepreneurs (30% to 36%) and operators (from 34% to 38%) on offshore facilities. A similar change is however not observed for employees on mobile facilities (27% to 28%). In the period 2013–2015 no significant changes in occupational injuries were

Table 1. Change in reorganization, downsizing, psychosocial factors, safety climate and occupational injuries 2013–2015, offshore facilities (2013 n = 7952/2015 n = 6675).

	Operators		Entrepreneurs		Mobile facilities	
	2013	2015	2013	2015	2013	2015
Downsizing	17.0	67.4***	35.7	79.0***	5.0	77.0***
Reorganization	40.3	51.2***	31.5	47.3***	24.4	44.6***
Job insecurity	7.2	12.2***	9.0	33.2***	4.5	40.4***
High Job-demand	20.9	20.9	15.5	20.4***	19.2	24.7***
Low Job-control	10.0	10.9	6.8	9.8***	11.0	13.8*
Job strain	5.2	5.0	2.8	5.4***	4.0	6.7***
Low supportive leadership	21.3	22.6	14.1	14.8	14.3	15.8
Low colleague-support	3.2	3.5	2.9	2.8	3.0	3.6
Poor safety climate	34.4	38.1*	29.9	35.7***	27.1	28.4
Occupational injuries	3.1	3.4	5.3	4.5	4.5	3.3*

*P ≤ 0.005; **P ≤ 0.01; ***P ≤ 0.001.

reported among entrepreneurs and operators, but a significant reduction in injuries was found among employees on mobile facilities (Table 1).

3.3 *Reorganization, downsizing and occupational injuries*

In total, 4.1% (246/5977) of employees in the petroleum industry (offshore and onshore facilities) experienced downsizing and 2.8% (61/2159) not experiencing downsizing reported occupational injuries. In total, 4.4%, (175/3939) experienced reorganization and 3.1% (130/4163) not experiencing reorganizations reported occupational injuries.

Employees reporting reorganization or downsizing reported also a significant higher prevalence for all psychosocial exposures and poor safety climate with the exception for low colleague support compared to employees not reporting such changes (Table 2).

3.4 *Logistic regression analysis*

A logistic regression analysis were conducted to examine if downsizing and reorganizations are associated with an increased risk of occupational injuries, and to examine whether this association may be mediated by psychosocial factors. In the initial model (adjusted for age, gender and education) employees who had experienced downsizing had a significantly higher risk of occupational injuries compared with employees not affected by downsizing (OR = 1.54; 95% CI 1.14–2.06) (Table 3) Adjusting for psychosocial factors and safety climate reduced the OR by 44%. The most important factor was safety climate.

Table 2. Description of occupational injuries and explanatory variables for employees in the petroleum industry (n = 8102).

	Reorganization			Downsizing		
	Yes	No	p-value[a]	Yes	No	p-value[a]
Job insecurity	29.9	22.2	0.000	29.7	16.1	0.000
High job demand	28.8	15.6	0.000	23.2	18.7	0.000
Low job control	14.9	8.7	0.000	12.6	9.5	0.000
Job strain	8.0	3.4	0.000	6.2	4.1	0.000
Low supportive leadership	22.9	14	0.000	19.2	15.9	0.000
Low colleague support	4.6	2.8	0.000	3.7	3.6	0.408
Poor safety climate	42.2	28.2	0.000	37.5	28.3	0.000
Occupational injuries	4.4	3.1	0.001	4.1	2.8	0.005

[a]Variables were tested with chi-square test.

Table 3. Multiple logistic regression for occupational injuries and the effects of adjusting for psychosocial factors and safety climate.

	Occupational injuries	
Initial model[a]	Odds Ratio (95%CI)[a]	Change[c]
No downsizing (n = 2159 (2.8))[b]	1	
Downsizing (n = 5977 (4.1))[b]	1.54 (1.14–2.06)[d]	
Job insecurity	1.43 (1.06–1.93)	–0.20
Psychosocial factors		
High Job demand	1.46 (1.08–1.96)	–0.15
Low Job control	1.47 (1.09–1.98)	–0.13
Low Supportive leadership	1.46 (1,09–1.97)	–0.15
Low collegue support	1.48 (1,10–1.99)	–0.11
Poor safety climate	1.38 (1,03–1.87)	–0.30
All factors included	1.30 (0,96–1.76)	–0.44

[a]Adjusted for age, gender, education.
[b]Number of respondents (cases of injuries).
[c]Percentage change in OR after comparing the initial OR with further adjusted OR (i.e., the initial OR adjusted for work-related factors).
[d]p = 0.005.

Table 4. Multiple logistic regression for occupational injuries and the effects of adjusting for psychosocial factors and safety climate.

	Occupational injuries	
Initial model[a]	Odds Ratio (95%CI)[a]	Change[a]
No reorganization (n = 4163 (3.1))[b]	1	
Reorganization (n = 3939 (4.4))[b]	1.46 (1.15–1.84)[d]	
Job insecurity	1.39 (1.09–1.77)	–0.14
Psychosocial factors		
High Job demand	1.31 (1.03–1.66)	–0.33
Low Job control	1.38 (1.09–1.75)	–0.17
Low Supportive leadership	1.35 (1.07–1.72)	–0.23
Low collegue support	1.38 (1.09–1.75)	–0.17
Poor safety climate	1.25 (0.98–1.57)	–0.46
All factors included	1.13 (0.88–1.44)	–0.72

[a]Adjusted for age, gender, education.
[b]Number of respondents (cases of injuries).
[c]Percentage change in OR after comparing the initial OR with further adjusted OR (i.e., the initial OR adjusted for work-related factors).
[d]p = 0.002.

The same procedure was applied when analysing the association between reorganizations and the risk of occupational injury, and to determine if this association may be mediated by psychosocial working conditions and safety climate. In the initial model (adjusted for age, gender and education) employees who had experienced reorganizations had a significantly higher risk of occupational injuries compared to employees not affected by reorganizations (OR = 1.46; 95% CI 1.15–1.84) (Table 4) Adjusting for psychosocial factors and safety climate reduced the OR by 72%. The most important factor was safety climate.

4 DISCUSSION

Our study set out to explore whether exposure to downsizing and reorganization increase the risk of occupational injuries and to test whether the association between downsizing, reorganizations occupational injury may be mediated by psychosocial working conditions and safety climate. As expected, the results showed that employees in the petroleum industry report a considerable increase in reorganizations, downsizing and job insecurity in the period 2013–2015. In the same period, significant changes in the psychosocial work environment offshore are observed among entrepreneurs and among employees on mobile facilities but not among employees working for operating companies. In terms of safety climate items, a deterioration on production facilities is observed, but not among employees on mobile facilities, and during the same period there have been minor changes in the occurrence of self-reported injuries.

The analysis also indicate that a higher risk of injuries is reported by those who have experienced reorganizations and downsizing and that is again associated with a poorer safety climate and psychosocial work environment. The results show that employees who have experienced reorganization and downsizing report higher risk of injuries, poorer safety and psychosocial work environment compared to employees who do not report such changes. Organizational changes, such as restructurings involving issues such as relocations, offshoring, closure, mergers/acquisitions, outsourcing, and internal restructuring including downsizing (Eurofound 2014) have been found to be linked to of ill-health as well as incidents (de Jong et al., 2016, Serck-Hanssen 2002).

Overall, the changes in the psychosocial working environment appear to be somewhat more pronounced on mobile facilities and may be seen in the context of ongoing change processes. Looking at the differences between operators, entrepreneurs and mobile facilities it might be argued that employees from operating companies does not show a negative development with regards to psychosocial risk and safety climate because they have not been affected by change in the same time period. Operating companies are now moving into a period with more organisational changes being implemented, which might have implications for psychosocial and organisational factors such as job demands, job control, leadership support and social support. Research have found that psychosocial

risks factors, such as high workload, inadequate resources, increased time pressure, staff conflicts and lower levels of autonomy or loss of job control may be exacerbated as a result of organisational change (Leiter & Maslach 2009, Rafferty and Griffin 2006, Rabatin et al. 2015, Day Crown and Ivany 2016). As such, performing the same analysis for the upcoming RNNP survey in 2017 might shed some light on this particular issue.

4.1 *Strengths and limitations*

This study is cross-sectional and provide a snapshot of a particular group at a given time. In research, cross-sectional studies are used in order to determine prevalence. Cross sectional studies are the best way to determine prevalence and are useful at identifying associations that can then be more rigorously studied using a cohort study or randomized controlled study (Mann 2003). As such, this may limit the ability to draw firm conclusions about relationships observed between exposure (for example, downsizing and reorganization) and outcomes (for example, occupational injuries), and the extent to which exposure precedes outcome (for example, occupational injuries) in time. Given that the data was cross sectional, we cannot draw conclusions about causal relationships. Thus, there is a possibility that those who have been injured during the reorganization are somewhat more "negative" when they answer the survey or that the changes have more and more affected groups that initially reported poorer psychosocial work environment and safety. However, it is important to note that research over at least the last decade, including longitudinal studies, has shown that psychosocial hazards can have a negative impact on health and safety (Leka & Jain 2010, Mackey, Palverman Saul et al. 2012). Another important limitation is that those who are absent from work due to sickness, health complaints or injury during the survey are not included in the data. As such, studying potential consequences of downsizing or termination of employment is difficult, because those who has been terminated are not included in the data. In other words, those remaining after the downsizing processes are those included in the data.

A strength of this study is that the questionnaire survey contains questions that cover many background variables that can be explored in relationship to downsizing, reorganization and HSE outcomes. The sample is also large enough to permit analyses of different sub-groups of employees, i.e. employees in different areas of work and job categories, and in risk groups defined in the RNNP report from 2014. It is also important to note that research over at least the last decade, including longitudinal studies, has shown that psychosocial hazards can have a negative impact on health and safety at the workplace. In the way forward these analysis should be replicated in order to explore whether the changes the industry are currently undergoing will have a more pronounced long-term effect on psychosocial work environment and safety climate.

4.2 *Implications for follow-up of the industry*

Psychosocial and organisational factors in relation to health and safety have been the subject of research for the last twenty years. The Petroleum Safety Authority (PSA) have also over the years focused on areas and themes relevant to the promotion of psychosocial work environment. One example are the theme collaboration and employee involvement highlighted in PSA's main topic for 2017 "Reversing the Trend". Other examples are PSA's followed-up of company's organisational change processes over the last years and the focus on organisational and human factors in barrier management (PSA 2016b, PSA 2017b).

Despite the amount of evidence collected and described in the literature, one of the biggest challenges may seem to be for companies to translate knowledge into workplace interventions. For employers it is a legal obligation to assess and manage all types of risk to workers' health and safety (Arbeidstilsynet 2017). Organisational changes represents such a process where the risk should be assessed and managed. As such, for employers and business owners it is vital to see psychosocial and organisational factors as an important element of managing risk and ensuring a sound change process. Risk management of these issues through risk identification, assessment and prioritization including allocation of resources to minimize, monitor, and control negative impact from occurring support the organisation's objectives for implementing lasting changes (Langenhan et al. 2013). As with health and safety in general, the successful promotion of the psychosocial and organisational work environment requires integration into daily work processes, avoiding treating it as a separate project. At last, employers must ensure the interventions implemented should target the work environment as well as the individual, in order to create safer workplaces and to improve the capacity of workers to protect their safety and health and to maximize their overall effectiveness.

Since both risk of injury and exposure of psychosocial hazards are unevenly distributed among groups, the follow up should be able to reflect this. Informative and user-friendly risk profiles will support the promotion of risk-based initiatives in times of change. Nevertheless, in the way forward it is important to monitor the developments

from an inspection authority perspective especially when there is a significantly higher risk of injury, reports of poorer psychosocial work environment and safety climate environment among employees affected by changes.

ACKNOWLEDGEMENT

We gratefully acknowledge the contributions to the project from all the project members and numerous other discussion partners.

REFERENCES

Arbeidstilsynet 2017. *The Working Environment Act.* https://www.arbeidstilsynet.no/contentassets/e54635c3d2e5415785a4f23f5b852849/working-environment-act-october-web-2017.pdf.

Baker Report 2007. The Report of the BP US Refineries Independent Safety Review Panel.

Bamberger, S.G. et al. 2012. Impact of organisational change on mental health: a systematic review. *Occupational Environmental Medicine,* 69 (8), 592–598.

Bergh, L.I.V. et al. 2014c. Psychosocial risks and hydrocarbon leaks: an exploration of their relationship in the Norwegian oil and gas industry. *Journal of Cleaner Production*, 84, 824–830.

Birkeland, M.N. et al. 2013. A brief safety climate inventory for petro-maritime organizations. *Safety Science*, 58, 81–88.

Crichton, M. 2005. Attitudes to teamwork, leadership, and stress in oil industry drilling teams. *Safety Science*, 43, 679–696.

De Jong, N. et al. 2016. The impact of restructuring on employee well-being: a systematic review of longitudinal studies. *Work Stress*, 30 (1), pp. 91–114.

Day, A. et al. 2016. Organisational change and employee burnout: The moderating effects of support and job control. *Safety Sceince*, 100, 4–12.

Eurofund, 2014. Restructuring. Retrieved November 23, 2017.

Grote, G. 2008. "Diagnosis of safety culture: A replication and extension towards assessing "safe" organisational change processes." *Safety Science,* 46(3): 450–460.

Kongsvik, T. et al. 2011. Safety climate and hydrocarbon leaks: An empirical contribution to the leading and lagging indicator discussion. *Journal of Loss Prevention in the Process Industries*, 24(4), 405–411.

Koukoulaki, T. 2009. "New trends in work environment—New effects on safety." *Safety Science.*

Langenhan, M.K. et al. 2013. Psychosocial risks: Is risk management Strategic enough in business and policy making? *Safety and Health at Work*, 4, 87–94.

Leka, S. & Jain, A. 2010. *Health impact of Psychosocial hazards at work: An overview*. Geneva: World health organisation (WHO).

Mann, C.J. 2003. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. Emergency medicine Journal, 20(1), 54–60.

Mathisen, G.E. & Bergh, L.I.V. 2016. Action errors and rule violations at offshore oil rigs: The role of engagement, emotional exhaustion and health complaints. *Safety Science*, 85, 130–138. doi: 10.1016/j.ssci.2016.01.008.

Mearns, K. et al. 2001. Human and organisational factors in offshore safety. *Work and Stress*, 15(2), 144–160.

Olsen, E. et al. 2015. Exploring relationships between organisational factors and hydrocarbon leaks on offshore platform. *Safety Science*, 80, 301–309.

Petroleum Safety Authority Norway (PSA) 2016a. *RNNP—Risikonivå i Norsk Petroleumsvirksomhet. Hovedrapport—Utviklingstrekk 2016* – Norsk Sokkel (RNNP—Risk level in the Norwegian Petroleum Industry—Main report—development trends 2016.

Petroleum Safety Authority Norway (PSA) 2017a. *Trenden skal snus—Reversing the trend*.

Petroleum Safety Authority Norway (PSA) 2016b. *Sikkerhet, status og signaler*. http://www.ptil.no/sss2016/category1216.html.

Petroleum Safety Authority Norway (PSA) (2017b). *Prinsipper for barrierestyring I Petroleumsindustrien—Barrierenotatet*. http://www.ptil.no/getfile.php/1343444/PDF/BARRIERENotat%20%202017.pdf.

Rafferty, A.E. et al. 2006. Perceptions of organisational change: a stress and coping perspective. *Journal of Applied Psychology*, 91 (5), 1154.

Rabatin, J. et al. 2015. Predictors and outcomes of burnout in primary care physicians. *Journal of Primary Care Community Health*, 1–3.

Serck-Hanssen, C. et al. 2002. *Safe Change: Methodology on Change in Norwegian Oil Industry*. Det Norske Veritas.

Tharaldsen, J.E. et al. 2008. A longitudinal study of safety climate on Norwegian continental shelf. *Safety Science*, 46, 427–439.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Reviewing macro level factors as a foundation for understanding quality and patient safety improvement efforts across countries

T. Johannessen, E. Ree & S. Wiig
University of Stavanger, Norway

H. van de Bovenkamp & R. Bal
Erasmus University Rotterdam, The Netherlands

ABSTRACT: This paper takes a macro perspective on quality and safety in elderly care, focusing on the role of the macro level context in two European countries; Norway and the Netherlands. The aim is to conduct a comparison of the healthcare systems as a foundation to understand quality and patient safety improvement efforts in elderly care across the two countries. Our methodological approach is to conduct a review of macro level factors as suggested by previous studies such as governance and financing structure, role of different actors, and types of initiatives taken. Data was collected from open sources such as national statistics, governmental reports, and web pages of key actors at the macro level. The similarities and differences are discussed to highlight the main areas where improvement efforts need to take country-specific factors into account when learning across countries.

1 INTRODUCTION

1.1 *Background*

In many European countries, there are continuous policy demands for efforts to improve healthcare quality and patient safety. Why improvement interventions and efforts are successful in one context and not in another, and the role of macro level factors are still under-researched (Krein et al. 2010). Efforts to improve quality may work better in some situations than in others. Thus, quality improvement strategies requires information about context (McDonald, 2013). There are large variations in the organization, financing and delivery of the healthcare services in elderly care between countries and these variations can be expected to have an impact on the quality and safety work (Ennoo et al. 2015). European healthcare policy makers looking for sustainable ways to organize healthcare should take these differences in context into account (Eenoo et al. 2015).

This paper takes a macro perspective on quality and safety in elderly care, focusing on the role of the macro level context in two European countries; Norway and the Netherlands. The aim is to conduct a comparison of the healthcare systems as a foundation to understand quality and patient safety improvement efforts in elderly care across the two countries. Improvement efforts have to be adjusted to the specific contexts in which organizations are embedded in order to be effective. The following research question will guide this paper: What do differences in macro level factors mean for quality and safety improvement efforts in elderly care in Norway and the Netherlands?

2 METHODS

Our methodological approach was to conduct a review of macro level factors as suggested by previous studies (Wiig et al. 2014; Wendt, 2009; Grosse-Tebbe, 2005; Leijten et al. 2017) such as governance and financing structure, role of different actors, and types of initiatives taken. Data was collected from open sources such as national statistics, governmental reports, and web pages of key actors at the macro level (e.g. regulators, inspectorates, knowledge centers). We mapped and compared data on governmental expenditures on health, actors and types of community health care services, funding, governmental vision and regulation on community care, involvement of informal care, and national quality and safety indicators. Elderly care in this paper is limited to quality and safety improvement efforts in nursing home and home care.

3 HEALTH SYSTEM IN NORWAY

3.1 *Governance of elderly care*

Norway is a parliamentary democracy, divided into three different administrative levels, the state, the 19 counties and the 426 municipalities (Kartverket, 2017). The organizational structure

of the Norwegian health-care system is built on the principle of equal access to services for all inhabitants, regardless of their social or economic status, country of origin and geographical location. This overarching goal has been a long-standing feature of the Norwegian welfare system and has been embedded in the national health-care legislation and strategic documents (Ringard et al. 2013).

Since the beginning of the 2000s, the emphasis was both on achieving structural changes and to organize health care better. In addition, reforms have been carried out to strengthen patient and user involvement. In recent years, efforts have been directed towards better coordination of health services, while quality and patient safety has received increasing attention (Ringard et al. 2013).

The Norwegian health care system can be characterized as semi-decentralized. At the national level, parliament serves as the political decision-making body. The responsibility for specialist care lies with the state, administered by the four Regional Health Authorities (RHAs), which in turn own the hospital trusts. Municipalities are responsible for primary care including rehabilitation, physiotherapy, nursing homes, midwife, home care and after-hours emergency services. There is no direct command and control line from central authorities down to the municipalities and the municipalities have a great deal of freedom in organizing primary care services (Ringard et al. 2013).

The principles for determining who is responsible for health services varies in different parts of the system. Local politicians are therefore accountable to local inhabitants through local elections. Municipalities are also responsible to the government to follow policies and regulations. Further internal auditors control municipality's budgets and have access to other economic factors of importance.

The Coordination reform was introduced in 2012, and was part of an effort to improve care coordination and quality of the services. It implied that municipalities became responsible for more treatment tasks. Because of the reform, the municipalities now receive more patients with need for long-term treatment, but there are still challenges regarding capacity and competence of the staff at the municipal level (Meld. St. 11. 2014–2015; Gautun et al. 2013).

The main task of the central government is to assure the high quality of services across the municipalities through funding arrangements and legislation (e.g. the 2011 Municipal Health and Care Act). The Directorate of Health issues an annual circular letter to the municipalities. The circular letter of 2016 contained recommendations on preventive home visits in the municipalities. The letter described how the municipalities could use preventive home visits as part of their service provision to the elderly. The purpose of preventive home visits is to strengthen the individual's ability to stay healthy and active as long as possible (Rundskriv, 2016).

3.2 *Funding and legislation*

In 2017, health-care expenditure accounted for approximately 10.5% of Norway's GDP, placing it in 16th place in the Western Europe in terms of the share of GDP spent on health (OCED, 2017).

Government spending accounted for 85% of overall health spending, among the highest in the OECD mostly comprising financing from the central and local governments (73%) and from the National Insurance Scheme (12%) through fee-for-fee service payments and reimbursement of user fees (Ringard et al. 2013).

Primary care is funded from municipal taxes, block grants from the central government, and earmarked grants for specific purposes. Private sources mainly in the form of out of pocket payments account for approximately 15% of health expenditure (Ringard et al. 2013; OCED, 2017).

Private actors are involved in the provision of primary care services. The majority of general practitioner are self-employed but are in most cases fully embedded in the public system through contracts with the municipalities. Services provided by for-profit institutions include long-term nursing care (about 10% of nursing homes in 2010) (Ringard et al. 2013).

The Norwegian healthcare system is regulated through a large number of acts and secondary legislation. Legislation broadly reflects the decentralized nature of the healthcare system. Specialized healthcare, organized at the level of the RHAs, is regulated by the Specialist Care Act of 1999 and the Health Authorities and Health Trusts Act of 2001. Primary care which in Norway comprises long term elderly care is regulated by the Municipal Health and Care Act of 2011. A key act is the Patients' and Users' Rights Act (1999) with the aim of ensuring "equal access to health services of high quality". Strengthening the role of patients and next of kin has been a policy priority since the turn of the millennium, for example, through stronger patient choice and complaint procedures (Ringard et al. 2013).

The central government is currently focusing on "the patient's health and care service" putting the patient's need in the center of healthcare provision. (St. Meld. 34. 2015–2016). There are no dedicated rights for the elderly today. The responsibility for the elderly rights is divided between a variety of administration levels, supervision, and agencies. Elderly rights are mainly regulated by the Patients' and Users' Rights Act, which applies to all health services in Norway. Municipalities often coordinate the services of the elderly (Rundskriv, 2015).

There is no specific patient safety act in Norway. Demands for improving quality and safety in healthcare are incorporated in several acts and regulations. The Health and Care Act (2011), covering all health and care services provided by municipalities, requires service providers to work systematically to improve quality and patient and user safety (Health and care Act (2011) § 4–2).

A recent regulation on leadership and quality improvement in the health and care services (2017) requires all service providers to establish a management system to ensure internal control of the service provision. This includes overview of risk areas and how to plan activities, carry them out, evaluate, and correct deviances. Municipalities have the responsibility to supervise own activities in line with internal control regulations (Ringard et al. 2013), but there are also external supervisory bodies with mandate to supervise the service provision. The Norwegian Board of Health Supervision (NBHS) is the national regulatory body for health and care services. It is a public institution organized under the Ministry of Health and Care Services. At the regional level, 18 county governors oversee services within primary and specialized healthcare.

3.3 *Knowledge infrastructure*

National professional guidelines exist to help ensuring that health and care services have high quality, proper priorities, not unwanted variety in service provision, resolve coordination challenges and provide comprehensive patient care. The Directorate of Health develops, disseminates and maintains national professional guidelines. National professional guidelines provide recommendations for service provision, but are not legally binding. The health service providers are responsible for organizing the services to ensure that recommendations given in national professional guidelines can be followed (Helsedirektoratet, 2014).

There is a National Patient Safety Program (2014–2018) working on quality and safety improvement by strengthening the competence of health professionals and managers, and implementation of measures to reduce patient injuries. Intervention bundles are offered in areas such as medication list, prevention of fall and pressure ulcers, prevention and treatment of malnutrition, and early detection of deterioration (Pasientsikkerhetsprogrammet, 2017).

It is voluntary for municipalities to participate in the national patient safety program activities. Five municipalities currently participate in a pilot "patient and user-safe municipality", which aims at ensuring systematic and sustained work on patient and user safety at all levels in the municipal health and care services (Pasientsikkerhetsprogrammet, 2017). The National Patient Safety Program is co-operating with the developing centers for Nursing Home and Home Services (USHT). There is one USHT in each county and they establish networks with the municipalities. The development centers play a key role in supporting nursing homes and home services in quality improvement efforts, including organizing of learning networks (Pasientsikkerhetsprogrammet, 2017).

There is a National Quality Indicator System in Norway and the quality indicators for the municipal health and care services have increased the last years. Examples of quality indicators in nursing homes and home care are follow-up of nutrition; health services associated with infections; available medical time; medication review; and occurrence of readmission among elderly 30 days after discharge to the municipality (Helsedirektoratet, 2014).

4 HEALTH SYSTEM IN THE NETHERLANDS

4.1 *Governance of elderly care*

The Netherlands is a parliamentary democracy. The two governance levels that are most important for health care are the national and the local (municipal) level. The Dutch health care system can best be described as a hybrid governance system, with a mixture of top-down government regulation, market elements (e.g. free choice of provider and insurer), self-regulation by professionals and consultation amongst relevant actors (Van de Bovenkamp, De Mul et al. 2014).

Since the 1980s different themes figured on the national agenda. From the mid 1980s onwards, controlling health care costs, improving quality and safety and strengthening the position of patients have figured prominently on the agenda. Already in the 1980s a system of regulated competition was proposed which was officially implemented in 2006 after a process of incremental change (this applies to care covered under the Health Insurance Act). During the 1990s several acts were introduced aimed at regulating health care quality, such as the Medical Treatment Agreements Act (1995), the Quality of Care Act (1996) and the Individual Health Care Professions Act (1997). These acts came about as a result of consultation with the field and reflected an emphasis on self-regulation. The Quality of Care Act for instance dictated that health care organizations, including those providing elderly care, should have a 'quality management system' installed. However, what these systems should look like was largely left up to the field (Grol 2006). In the last decade, mostly in response to quality incidents, top down regulation has become more stringent which caused the health care inspectorate (responsible

for quality supervision) to play a more stringent role.

In addition to more stringent supervision, many national quality projects have been implemented also in elderly care. These national quality projects stimulate organizations to improve their quality and safety and provide funding to work on quality and safety projects. The short term focus of these programs (they only last for a couple of years) has caused the problem of fragmentation and projectification as organizations move from one project to the next which limits the sustainability of improvements (Slaghuis 2016).

Controlling costs and stimulating patient participation have been key issues of health care policy for some time now. These principles were also the goals of a large decentralization process that also affected elderly care. While part of elderly care is still governed at the national level (care that falls under the Health insurance act and Long-term care act) other parts have become the responsibility of the 388 municipalities from 2015 onwards (care that falls under the Social support act). These acts are further elaborated on in the next paragraph.

Dutch quality and safety policy of the last decade has strongly focused on making quality of care transparent through the use of indicators. This means that health care organizations have to put a lot of effort in registering and providing actors such as the inspectorate and insurers with performance information. Partly because of this, there is now much debate about regulatory pressure in health care. Professionals are calling for attention to the fact that they spend too much time on administration which limits their ability to provide care (Bovenkamp, Stoopendaal et al. 2017). In response to this, there is now much debate on working on quality in a different way that allows for local variation and focus on reflexivity (ibid.). In elderly care this has also been taken up in the latest national quality framework.

4.2 *Funding and legislation*

In 2016 13.8% of GDP, 96 billion euro, was spent on health care (including welfare) in the Netherlands.[1] This makes the Dutch healthcare system one of the most expensive worldwide. Hospital care comprises the largest part of the health care budget, 27 billion. Elderly care comes in second with almost 17 billion.[2]

Health care and elderly care in the Netherlands are regulated and financed through a variety of acts. Partly these acts determine how care is financed and who is responsible for the provision of care. In this section, we focus on the legislation especially relevant for elderly care. The Heath insurance Act (*Zorgverzekeringswet*), works through private insurance (and for a small part through out of pocket payments). The basic package is obligatory for all citizens (this package is set by the Dutch Health institute). Citizens can voluntarily choose to buy additional packages (e.g. for dental care, more physiotherapy). Under the Health insurance act a system of regulated competition is established. Health insurers compete for insured and can selectively procure care. Health care providers compete for patients. The care offered under this act includes acute and curative care (primary, secondary & tertiary, mental health).

The Long-term care act works through social security and is paid for through taxes and for some services small out of pocket payments. A 'care indication' is needed through an indication office to get access to this care (which includes 24 hour institutional or home care). The Act also provides for the possibility of personal budgets, empowering patients to organize their own care.

The Social support act, for which the municipalities are responsible also works through a combination of taxes and some out of pocket payments. For this type of care, which includes home care, special services and public mental health care, municipalities are responsible for care indications and financing. Municipalities can selectively contract health providers.

Elderly can be subject to all three laws, which means that a lot of coordination is needed between insurers, municipalities and providers.

As stated above there are also different acts to regulate health care quality, ensuring amongst others that organizations have a quality system in place and professionals comply to certain norms.

4.3 *The position of patients*

There is much focus in the Dutch system on strengthening the position of patients. Much effort, also in the national quality programs, is paid to let patients participate more in their care. This includes elderly care where the latest national program Dignity and Pride (*Waardigheid en Trots*) focuses on providing care according to the wishes of patients. The expectation is also that patients are active as much as possible themselves. For instance, the Social support act argues that people have to be as self-reliant as possible; only if people cannot take care of themselves, or cannot find help in their social network they are eligible to professional care.

[1]http://statline.cbs.nl/Statweb/publication/?VW=T&DM=SLNL&PA=83075NED&D1=a&D2=0,3,8,13,18,23,29,34,39,(l-2)-l& HD=160517-1206&HDR=G1&STB=T.
[2]http://statline.cbs.nl/Statweb/publication/?VW=T&DM=SLNL&PA=83038NED&D1=a&D2=a&D3=(l-2)-l&HD=170512-1505&HDR=T,G2&STB=G1.

There is also legislation ensuring patients' rights such as the right to informed consent, to choose providers and insurers and to file complaints. In addition, all health care organizations have to have a client council that has been given far reaching rights of advice (Van de Bovenkamp 2010, Schillemans, Bovenkamp et al. 2016). Furthermore, patients are increasingly involved in quality improvement projects in health care organizations. Next to client councils, methods used include patient surveys, mirror meetings, focus groups and experience based co-design (Bovenkamp, Grit et al. 2008, Schipaanboord, Delnoij et al. 2011, Vennik, Bovenkamp et al. 2016).

4.4 *Knowledge infrastructure*

The Dutch healthcare system has an elaborate knowledge infrastructure (Bekker, van Egmond et al. 2010). For elderly care, specifically, a knowledge institute exists at the national level (http://www.vilans.nl/) which is especially concerned with quality improvement. In the early 2000s, the Health Council of the Netherlands nevertheless argued that an intensification of research directed at elderly (care) was called for, which resulted in the establishment of a National Program Elderly Care (Grit, Dwarswaard et al. 2012, Wehrens, Oldenhof et al. 2017). The program, with a budget of almost 90 Million Euro, stimulated regional network building between elderly care organizations, municipalities and universities and the experimentation with and implementation of many quality programs. Participation of elderly was an explicit aim of the program. In addition, the national improvement project Dignity & Pride offers a platform for knowledge exchange between care organizations. Professional organizations also play a role in knowledge exchange. One of these organizations organizes a yearly benchmark for example based on quality indicators with the aim to help elderly care organizations to improve their care. However, this data is (as of yet) not public.

5 DISCUSSION

The similarities and differences between countries are discussed to highlight the main areas where quality and safety improvement efforts in elderly care need to take country-specific factors into account.

5.1 *Similarities and differences*

The healthcare systems in Norway and the Netherlands are listed as some of the best in Europe (OCED, 2017). Both countries have high health expenditures, but there are major differences in how healthcare services are organized and funded. The Dutch system is based on regulated competition and insurance, while the Norwegian system is based on taxes and public funding. For the Netherlands, this could have an impact on how working on quality and safety improvement is conducted, because it involves a high number of external actors in the market, and a more competitive system. In Norway there are less actors in the market and the public role is stronger. As a consequence, quality work in Dutch elderly care tends to be more fragmented, with many organizations pushing quality agendas and elderly care providers taking part in a diversified set of improvement initiatives (or not). While this on the one hand means that there is a very diversified set of quality and safety initiatives, which can be attuned to local contexts, there is also a lack of a building up of knowledge about quality and safety improvement across the sector. Several national initiatives that enable cross-organizational learning now try to fill this caveat, e.g. by organizing national meetings on specific quality subjects. In addition the national quality framework also states that organizations should develop learning networks as part of their quality policies.

In both countries, the patients have a strong position. The Netherlands had early on the agenda (around 1990) to strengthen patients' position, and Norway established a specific Patient and User Rights act in 1999.

There are no specified rights for the elderly, but the rights of this user group are governed by the strong patients' rights and legislations in both countries. In the Netherlands, there is a stronger focus on informal healthcare, compared to Norway. The expectation is also that patients themselves are active as much as possible. For instance, the Social support act argues that people have to be as self-reliant as possible; only if people cannot take care of themselves, or cannot find help in their social network they are eligible to professional care.

Both countries require that service providers develop quality management systems. The Quality of Care Act in the Netherlands, and the Health and Care Services Act in Norway require a systematic quality improvement work and systems in place. But, what these systems should look like is largely left to the field in both countries (Grol 2006). However, over the past decade, mostly as a result of quality assaults, regulation has become stricter in the Netherlands, which has led to a stricter supervisory role by the healthcare inspectorate. In terms of national quality indicators, knowledge structure, and national quality improvement projects the Dutch context has experienced a stronger pressure over years compared to Norway. The Norwegian indicator system has developed an increasing number of indicators the last years. Until a few years ago,

it mainly consisted of indicators covering the specialized healthcare services (Wiig & Lindahl, 2015). For the Netherlands, although there is much debate about indicators, a quality standard for elderly care has been developed and the sector engages in benchmarking in order to learn from best practices.

5.2 *Maturity*

By comparing the countries on macro level factors, the results indicate that the Netherlands has come further in their method development in user and next of kin involvement and in terms of initiating national improvement projects and indicators. At the same time, this can make healthcare providers more reserved to additional quality and safety improvement interventions and efforts since they might already have knowledge and tools in place, while this is still under-developed in the Norwegian context, according to our mapping (Meld.St. 26. 2014–2015). It appears as a challenge for the Dutch system, that nursing homes and home care services experience a strong pressure from external short-term projects, implying fragmentation and possible difficulties for implementation and sustainability. While policy makers in Norway emphasize the need for increased focus on quality and safety improvement and strengthen external knowledge support structures for nursing homes and home care, the Dutch system has national proposals on how to prevent these organizations from feeling such a strong external pressure that requires a lot of documentation and takes time from daily patient care.

6 CONCLUSION

In this study, we have mapped and compared macro level factors in Norway and the Netherlands as a foundation for understanding and learning from quality and safety improvement efforts across countries. Having an overview of funding mechanisms, patient rights, quality and safety requirements in the law, and the maturity of the knowledge infrastructure supporting the institutions in elderly care, are fundamental macro level contextual factors that should be mapped to understand the responses from institutions in quality and safety improvement efforts.

ACKNOWLEDGEMENT

The study Improving Quality and Safety in Primary Care-Implementing a Leadership Intervention in Nursing Homes and Homecare (SAFE-LEAD Primary Care) has received funding from the Research Council of Norway's programme HELSEVEL, under grant agreement 256681/H10, and the University of Stavanger.

REFERENCES

Bekker, M., van Egmond, S., Wehrens, R., Putters, K., & Bal, R. (2010). *Linking research and practice in Dutch healthcare: infrastructure, innovations and impacts*. Evidence & Policy, 6(2), 237–253.

Bovenkamp, H. M. v. d., Grit, K. J., & Bal, R. A. (2008*). Inventarisatie patiëntenparticipatie in onderzoek, kwaliteit en beleid.*

Bovenkamp, H. M. v. d., Stoopendaal, A. M. V., Oldenhof, L. E., & Bal, R. A. (2017). *Eigen Regie, Regeldruk en Regelruimte.*

Eenoo, L., Declercq, A., Onder, G. et al. (2015) *Substantial between-country differences in organizing community care for older people in Europe—a recview*. European Journal of Public Health. Vol 26, No. 2, 213–219.

Gautun, H., Syse A. (2013) *Samhandlingsreformen*. Novarapport, nr 8. Oslo: Norsk institutt for forskning, oppvekst, velferd og aldring.

Grit, K., Dwarswaard, J., & Bal, R. (2012). *Adviseren met beleid. Een onderzoek naar de doorwerking van adviezen van de Gezondheidsraad.*

Grol, R. (2006). Quality *Development in Health Care in the Netherlands.* Retrieved from Grosse-Tebble, S., Figueras, J. (2005) *WHO European Observatory on Health Systems and Policies.* Snapshots of Health Systems.

Health and care Act (2011) Kap 4 Reguirements for sound pratice, patient safety and quality § 4–2.

Helsedirektoratet. (2014) Available: https://helsenorge.no/Kvalitetsindikatorer/kvalitetsindikator-pleie-og-omsorg (Accessed at 12.12.17).

Kartverket (2017) Fylke og kommune oversikt. Available:https://www.kartverket.no/Kunnskap/Fakta-om-Norge/Fylker-og-kommuner/Tabell/ (Accessed at 29.11.17).

Krein, S. L., Damschroder, L, J., Kowalski, C, P. et al. (2010) The influence of organizational context on quality improvement and patient safety efforts in infection prevention: A multi-center qualitative study. Social Science & Medicine 71: 1692e1701.

Leijten, F., Struckmann, V., van Ginneken, E., et al. (2017) The SELFIE framework for integrated care for multi-morbidity: Development and description. Health policy: HEAP-3753, P 11.

Mcdonald, K, M. (2013) Considering Context in Quality Improvement Interventions and Implementation: Concepts, Frameworks, and Application. Academic Pediatrics; Vol 13: N 6S.

Meld. St. 11. (2014–2015) Kvalitet og pasientsikkerhet 2013. Oslo: Helse- og omsorgsdepartementet.

Meld. St. 26. (2014–2015) Fremtidens primær helstjeneste—nærhet og helhet. Oslo: helse- og omsorgsdepartementet.

Meld. St. 34. (2015–2016) Verdier i pasientens helsetjeneste—Melding om prioritering. Oslo: Helse- og omsorgs departemetet.

OECD (2017), Health spending (Indicator). Doi: 10.1787/8643de7e-en.

Pasientsikkerhetsprogrammet (2017). Available at:http://www.pasientsikkerhetsprogrammet.no/om-oss/i-kommunene. (accessed 17.11.2017.).

Ringard, Å., Sagan, A., Saunes, I. S et al (2013) *Norway Health System review*. Health System in Transition Vol. 15 No. 8.

Rundskriv, IS-8 (2015) *Pasient og brukerrettighetsloven med kommentar*. Oslo: Helse- og omsorgsdepartementet.

Rundskriv, I–2. (2016) *Om forebyggende hjemmebesøk i kommunene*. Oslo: Helse- og omsorgsdepartementet.

Schillemans, T., Bovenkamp, H. M. v. d., & Trappenburg, M. J. (2016). *From 'Major Decisions' to 'Everyday Quality'. Direct Accountability to Clients. In P. Mattei* (Ed.), Healthcare Governance and Accountability: Palgrave MacMillan.

Schipaanboord, A., Delnoij, D., & Bal, R. (2011). *Patient empowerment in the Netherlands*. In H. Lofgren, E. de Leeuw, & M. Leahy (Eds.), Democratising health. Consumer groups in the policy process (pp. 111–126). Cheltenham: Edward Elgar.

Slaghuis, S. (2016). *Riding the waves of quality improvement. Sustainability and spread in the Dutch quality improvement program for long-term care.* (PhD), Erasmus University, Rotterdam.

Van de Bovenkamp, H. (2010). *The limits of patient power: examining active citizenship in Dutch health care.* Instituut Beleid en Management Gezondheidszorg (iBMG).

Van de Bovenkamp, H., De Mul, M., Quartz, J., Weggelaar-Janssen, A. M., & Bal, R. (2014). *Institutional layering in governing healthcare quality.* Public Administration, 92, 208–223.

Vennik, F. D., Bovenkamp, H. M. v. d., Putters, K., & Grit, K. J. (2016). *Co-production in healthcare: rhetoric and practice*. International Review of Administrative Sciences, 82, 150–168.

Wehrens, R., Oldenhof, L., Verweij, L., Francke, A., & Bal, R. (2017). *Experimenteel sturen in netwerken: een evaluatie van proces en structuur van het Nationaal Programma Ouderenzorg.*

Wendt, C. (2009) *Mapping European healthcare systems: A comparative analysis of financing, service provision and access to healthcare.* J Eur Social Policy:19:432.

Wiig, S. Lindahl, AK. (2015). *Indikatorer for sikkerhetsarbeid—hvorfor og hvordan?* (Indicators for safety work—how and why?) In: Aase (eds) 2.nd edition. Pasientsikkerhet-teori og praksis. Universitetsforlaget.

Wiig, S., Aase, K., von Blessen, C., Burnett, S et al. (2014) *Talking about quality*. BMC Health Services Research. 14:478.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Multicultural workplaces: A state of the art study of the Norwegian construction industry

K. Wasilkiewicz
Norwegian University of Science and Technology, Trondheim, Norway
SINTEF Technology and Society, Trondheim, Norway

S.S. Kilskar, A. Øren & R.K. Tinmannsvik
SINTEF Technology and Society, Trondheim, Norway

I. Kilanowska
The Federation of Norwegian Construction Industries (BNL), Oslo, Norway

ABSTRACT: A multi-method case study, including interviews, a survey and participatory observations, was undertaken to describe opportunities and challenges related to multicultural workplaces in the Norwegian construction industry, and the consequences for occupational safety, working environment and work performance. The findings show that challenges related to language and cultural differences are the most common, and that employment relationships and duration of relations have proven to affect many of the challenges. Consequences due to challenges related to language and culture were found more prominent for working environment and work performance than for safety. The focus on opportunities is limited and the potential is not fully exploited. Measures implemented to improve the conditions at multicultural workplaces mainly focus on solving the challenges and mitigating the consequences in a short-term perspective. A more future-oriented focus is needed, including measures that may lead to long-term gains for the industry.

1 INTRODUCTION

With the free flow of labour in most of Europe, companies in Norway are experiencing potential for innovation and effectiveness, but also some challenges. In the Norwegian construction industry, it is estimated that around one-third of the work force are migrant workers (BNL, 2017). This paper discusses the opportunities and challenges found in a study on multicultural workplaces in the Norwegian construction industry which was conducted between August 2016 and August 2017 (Kilskar et al., 2017).

The study was carried out to bring more knowledge on how internationalisation influences the working environment and safety in the construction industry, including culture, management, attitudes, expectations, communication and behaviour amongst both workers and their employers.

The overall objective of the study was to enable organisations and businesses in the Norwegian construction industry to work systematically in improving performance and productivity through good working environment, collaboration and increased safety at multicultural workplaces. An important aspect has been to not only look for problems and challenges at multicultural workplaces, but also to find the potential benefits of having a multicultural working environment.

The study aimed at solving three research questions:

1. What are typical characteristics, strengths and challenges related to safety and working environment by using multicultural labour in the construction industry?
2. How do challenges related to multicultural work force influence safety, working environment and work performance?
3. What measures exist and are proposed by the construction industry to create good multicultural workplaces?

1.1 *Concepts and definitions*

Construction can be used as a collective term for more sectors. In Norway, the construction industry is divided mainly in the building sector (e.g. building houses, commercial buildings) and construction sector (e.g. construction of roads, railways).

This study was limited to the building sector; however, the term *construction* is used in this paper.

The study mainly focused on migrant workers from eastern Europe, but interviews were also conducted with workers from other nationalities to get a broad understanding of multicultural workplaces in the construction industry. In this study, *eastern European workers* are defined to include nationalities from countries in the old Eastern-Block (i.e. former Soviet countries; Bulgaria, Czech Rep., Hungary, Poland and the Balkans). However, the large focus of the study was on workers from countries which joined the European Union (EU) in 2004; Czech Republic, Estonia, Hungary Latvia, Lithuania, Poland, Slovakia and Slovenia.

2 BACKGROUND

2.1 *The Norwegian construction industry*

The construction industry is one of the fastest growing industries in Norway and it stands for approximately 12.5 percent of the Norwegian BNP. In 2016, the industry turnover reached 521.8 billion NOK, and the combined work force was at almost 235 thousand employees (SSB, 2017a). It is characterised by many small and medium sized companies, with union density lower than average in Norway (43 percent in the construction sector versus 54 percent for other industries in Norway) (Nergaard, 2016). The industry is project based with temporary organisations consisting of multiple actors, which are forming and dissolving with each project. The large companies account for a smaller share of the construction output.

Unemployment rates in Norway have been stable on a low level, which has resulted in a shortage of skilled workforce. Traditionally, the majority of Norwegian workers have been permanently hired, and possibilities for temporary employment are somewhat limited through the Norwegian Working Environment Act. However, in recent years, there has been an increase in use of temporary staffing agencies hiring migrant workers in the construction industry.

2.2 *Working environment and occupational safety in the construction industry*

The construction industry in Norway is one of the main-land industries in Norway with the highest number of fatal accidents. In 2016, 8 out of 45 work-related fatal accidents occurred within this industry (NLIA, 2017a). There has also been reported 5.8 accidents at work that resulted in prolonged absence from work per 1 000 employees in the construction industry, whereas the average for other industries was 3.7 (SSB, 2017b). The Norwegian Labour Inspection Authority (NLIA) focuses on controlling and guiding companies and workers in the construction industry, highlighting the importance of preventive measures.

2.3 *Migrant workforce*

With low unemployment rates and shortage of skilled workers, the enlargement of the EU in 2004 gave companies access to recruit migrant workers, while at the same time foreign service providers got access to the Norwegian construction marked. Between 2007 and 2015, work has been the most common reason for immigration to Norway, followed by family reunions. The largest groups of immigrants are from Poland, Somalia and Lithuania (SSB, 2017). Many of the male immigants that came to Norway from eastern Europe after 2004, came to work in the construction industry. The process was not free of tension, as the work migration led to many challenges concerning migrant workers' wages and working conditions, accommodation standards, undeclared work, examples of exploitation and so called "social dumping" (Dølvik et al., 2005).

One of the serious concerns is the question whether migrant workers are more prone to work related accidents, which previous research both abroad and in Norway suggests (NLIA, 2012; Salminen, 2011). To look further into the background of such numbers, a mixed approach was chosen with a focus on the qualitative studies.

3 METHODS

Triangulation was used to combine the advantages of qualitative and quantitative methods; collecting data through semi-structured interviews, participatory observations and a survey. Additionally, a literature study was performed by searching in Scopus and Google Scholar for scientific documentation prior the data collection. Further, searches on the internet were done for news-articles. Search words used were "multicultural workforce", "building/construction industry", "management" and "migrant workers". The focus was on Norwegian publications.

As workers from Poland constitute the largest group of migrant workers in Norway, this is also the nationality most represented in the interviews and observations.

3.1 *Interviews*

Interviews were undertaken with 35 persons in seven cases, each case being either a construction

project or a construction company. The majority of the informants were from Norway (17) and Poland (10), while the remaining (8), came from other countries; Germany, Denmark, Russia, Estonia, Afghanistan and Ethiopia. The interviewees were managers, work team leaders, safety deputies and workers (including hired labour and apprentices).

A semi-structured interview guide was developed in Norwegian, and questions were adjusted for managers, Norwegian workers and migrant workers. The interview guide was translated into Polish and English for interviews with workers, and to English for managers. As a quality assurance, the interview guide was discussed with persons from the construction industry, and a pilot study was performed.

The interview guide was divided into five sections. Section A included questions about the informants' background, views on safety and working environment in general, and general questions about migrant workers in Norway. Section B consisted of questions about characteristics, opportunities and challenges at multicultural workplaces. Section C dealt with implications of having multicultural work teams, e.g. communication and cooperation between workers and managers, management and follow-up at the worksite, and characteristics of good multicultural workplaces. In section D, the focus was on good practices and solutions for creation of good multicultural work places. Finally, section E included questions to summarise the interview.

Interviewees were identified by the Federation of Norwegian Construction Industries (BNL). Factors addressed in the selection process of the interviewees were persons from different companies, disciplines, countries, and according to geography in Norway. Additionally, size of the project/company was one criterion. The projects where the interviewees worked varied from being small projects to large housing complexes.

The interviews were performed in the period from November 2016 to May 2017.

3.2 *Observations*

To verify data from the interviews, participatory observations were conducted at two construction sites over a period of three weeks. BNL facilitated contact with observation objects. The field work was performed by two anthropology students supervised by one of the project members. Two different methodological approaches were used; 1) observation with a high degree of participation in the daily work, supplemented with a limited number of interviews, 2) interviews, as well as observation, but less participation in the daily work.

The observations were important to gain knowledge about interactions and cooperation between Norwegian and migrant workers. The focus in the observations was on eastern European workers.

3.3 *Survey*

A survey was performed among managers in companies that are members of BNL.

The questionnaire was based on the results from the interviews and participatory observations, thus making it possible to lift the findings up to a general/national level. The questionnaire was based on the same topics as the interview guide (see 3.1).

A pilot test of the questionnaire was performed as a quality assurance of the content. BNL performed the submission of the questionnaire to the respondents electronically.

In total, 5 774 managers were invited by e-mail to answer the survey. 350 e-mails were returned because of "unknown recipients". Moreover, several persons replied and told they were not managers, but in the administrative or economical department. This was estimated to be about 870 persons. Finally, 886 persons in 562 distinct construction companies answered the survey, resulting in an approximate response rate of 19.5 percent for individual respondents and 21 percent for companies.

3.4 *Limitations*

BNL is a business and employer policy organisation for companies in the construction industry, and an umbrella organisation for 15 industries that organise a wide range of companies (in total over 4000 companies with nearly 70,000 employees). The companies range from the smallest companies to the largest in the industry including manufacturing companies, plumbers, carpenters, landscape gardeners, masons, painters and entrepreneurs (BNL, 2018). Most of BNL's branch associations have defined requirements for a company to apply for membership. The organisation advocates that Norwegian society should be built by serious, honest companies and that customers should be able to trust their suppliers.

Most of the companies that participated in the study were members of BNL. These companies are thought to be performing in a serious way, as well as ensuring health, safety and environment (HSE) according to regulations. This could have influenced the results.

The survey was directed only towards managers, since BNL has contact information only to managers and administrative personnel among their members.

4 RESULTS AND DISCUSSION

The results of the study show the importance of several factors for ensuring the working environment, cooperation and safety at construction sites. The main findings of the study are presented in the following.

4.1 *Differences between migrant and Norwegian workers*

The study looked at perceived differences between migrant and Norwegian workers. In many aspects, including skills, competence and quality of performed work, the differences were found to be minor. However, it was seen that groups of migrant workers and Norwegian workers often have different views on each other. In example, several of the Polish workers that were interviewed perceived themselves as more efficient and solution oriented than Norwegian workers. The Norwegian workers on the other hand, had the same perception, but in their own favour. This can be a source of misunderstandings as their perception differs. However, the survey showed that there is, in many aspects, a minor difference in how managers perceive Norwegian and migrant workers, respectively.

Even though there were many similarities between workers, one aspect appeared to be very different. When it comes to reporting of unwanted events there was a major difference between eastern European workers and Norwegian workers. The eastern European workers report fewer unwanted events and dangerous conditions. One reason the workers gave was that they do not wish to report on others to the management. These findings also coincide with previous research (Wasilkiewicz et al., 2016). However, these major differences were not found for all aspects related to safety. For compliance with safety regulations the results from the survey showed that the managers observe differences, however in much smaller degree than for reporting unwanted events.

Findings from the survey further shows that perceptions on migrant workers to a large degree depend on managers' experiences with migrant workforce. Managers from companies with permanently employed migrant workers were less critical towards migrant workers, than managers from companies without permanently employed migrant workers. As an example, the leaders were asked the following question: "To what degree do you experience that migrant workers do not comply with existing safety requirements (e.g. do not use required protection equipment)?" Nearly 44 percent of the respondents from companies with no permanently employed migrant workers answered this question with "to a large degree" or "to a very large degree". The corresponding proportion among the respondents from companies that do have permanently employed migrant workers was only close to 23 percent. The interviews also supported this finding, as managers with a lot of experience with permanently hired migrant workers were the most positive when discussing this topic.

The perceptions of managers on migrant workers also appear to be somewhat influenced by the size of their respective companies, as managers in medium sized companies (which in the study was defined as companies with 22–51 employees) answered in more positive terms than the managers from smaller and larger companies did.

This shows that there are large individual differences between workers within a nationality, as well as individual differences between managers (e.g. their expectations and requirements) and their experiences with migrant workers, resulting in different opinions.

4.2 *Potential not fully exploited*

The interviewees were asked what Norwegian workers and managers and migrant workers can learn from each other. Additionally, it was looked at what the benefits are, related to having migrant workforce, as well as possible opportunities.

The opinions varied; some stated that there is nothing to learn, while others had examples of what they had learned. The opinions in large degree reflected personal experiences.

When Norwegian managers were asked about what they think is positive with migrant workers, their *flexibility* was accentuated, e.g. migrant workers' higher willingness to work overtime, stability in the workforce (e.g. not taking work days off due to sick children), and willingness to continue work to complete tasks before the day is over. According to the managers, these points were in large degree related to culture. However, they can also be related to other factors, such as the fact that many migrant workers have their families abroad, and therefore want to work as much as possible while they are in Norway. It is important to be aware that the described behaviour can also be related to power relations. Persons form different cultures can have different expectations and relations to mangers (House et al., 2004; Warner-Søderholm, 2012a; 2012b), and thus perceive mangers' requests, such as working over-time, as duties rather than options.

When it comes to learning potential in general, most Norwegian interviewees did not see that they could learn anything from the migrant workers, however several managers pointed out that Norwegian workers could take advantage of learning better *working moral* from the migrant workers and to be "less lazy". A few pointed out that migrant

workers could teach local workers other *working techniques* and vice versa.

Some of the migrant workers pointed out that they value the *characteristics of the Norwegian working life*, e.g. "calmness" and less stress, thinking things over before starting a work task, and that they could learn this from Norwegians both for work, as well as for their personal life.

The industry is not highly conscious of the benefits and learning opportunities, and utilisation of the potential that migrant workers might bring in. Earlier research shows that multicultural and diverse groups can be both more efficient and less efficient than groups consisting of one single culture (Adler & Gundersen, 2008, p. 140). Through good collaboration, multicultural teams can be more efficient and innovative than homogeneous teams. By sharing knowledge and having different perspectives and experiences, such teams can achieve better solutions for common problems.

From the free text field in the survey undertaken, an example was found of how the construction industry can benefit from the knowledge of migrant workers:

> *"Migrant workers and Norwegians have discussed product alternatives from abroad (…) with the consequence that products or alternative products are being imported from abroad now. (…) This resulted in economic benefits in the company because of a better calculation of the price, and often easier handling in production (a win-win situation for workers and the company)".* (Survey respondent).

Migrant workers in the construction industry represent a larger resource than what is being utilised. The full utilisation of their potential is often prevented by poor linguistic skills (in Norwegian or English) and insufficient recognition of their formal competence from abroad.

4.3 *Challenges affected by employment relationships and duration of relations*

Results from the interviews, the participant observations, as well as the survey indicate that language is the most common challenge related to multicultural workplaces. The Norwegian informants did, however, consider this more of a problem than many of the migrant workers did. Shortage in Norwegian skilled workforce is part of the issue, as illustrated by the following response when asked whether speaking Norwegian should be an absolute request:

> *"Yes, I believe so. It is, however, hard to get hold of enough people. Finding a man who knows both the language and how to do the job is not easy".* (Eastern European work team leader).

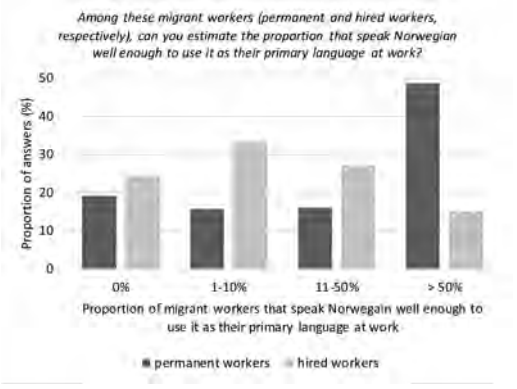


Figure 1. Distribution of manager's answers regarding linguistic skills among their permanent and hired migrant workers.

Figure 1 shows the distribution of managers' answers regarding linguistic skills among their migrant workers.

Close to 50 percent answered that more than half of their permanent migrant workers know Norwegian well enough to use it as their primary language. However, only 15 percent said the same about their hired migrant workers. In fact, close to 60 percent answered that less than one in ten hired migrant workers speak Norwegian well enough.

In addition to language related challenges, issues related to cultural differences were often mentioned. As an example, many migrant workers and eastern European workers particularly, tend to say "yes" and give the impression that they understand messages when they do not. The following quote from a worker from Poland is illustrative:

> *"I sometimes make mistakes when I don't understand. I don't know why I don't ask again. I don't know people working here, and I think I should understand".* (Eastern European worker).

Thus, in several cases where communication fails, language related issues cannot be seen irrespectively of those related to cultural differences. The respondents of the survey mostly agreed that managing migrant workers is more challenging than managing Norwegian workers. It is thus important both for managers and fellow workers to gain necessary understanding of such differences, and how, for example, differences in perceiving hierarchy and power distance can make Norwegian and eastern European workers act differently.

Other challenges included some migrant workers feeling that they were treated differently from their Norwegian co-workers, arguing that they were assigned to "harder" and more boring work.

In addition to specific challenges that may occur at multicultural workplaces, several of the Norwegian managers and workers raised a concern that an extensive use of migrant workforce may negatively affect the recruitment of young, Norwegian workers to the industry.

What is common for many of the identified challenges, is that they appear to be enhanced by employment relationships and conditions. The study shows that employment relationships are of greater importance than nationality when it comes to the challenges at multicultural workplaces. Hired workers from staffing agencies are less likely to learn the language, and companies are equally less likely to invest in language courses for these workers.

As mentioned, many managers find it harder to manage migrant workers, than Norwegian workers. Findings from the interviews, however, indicate that leadership challenges are greater when involving hired workforce rather than permanent workers. It is also the hired workers that most often claimed to be treated differently. As an example, some had experienced having to accept poor working conditions due to the fear of not being hired again.

The findings also show that the challenges seem to be affected by the duration of relations; that is, when workers are staying at a building site for just a short amount of time, which may cause problems:

> *"You know, when a subcontractor comes to the construction site, like those balcony fitters, they are staying for two, maybe three weeks before finishing their work. You don't have the time to get to know them. You don't know what they stand for and what their interests are for doing the work in a safe manner".* (Norwegian project manager).

4.4 *Greater consequences for work performance and working environment than for safety*

Some argued that the number of unreported unwanted events is larger among migrant workers, and quite a few imagined migrant workers being harmed at work more often than Norwegians. Despite this, most of the informants said that they had no basis for stating that migrant workers are overrepresented in their company's accident records.

Results from the survey clearly indicate that language related challenges and culture related misunderstandings cause building errors or disagreements at work more often than they cause accidents or near accidents. This implies that language and culture related challenges cause greater consequences for work performance and working environment than for safety. The distribution of the answers is shown in Figure 2. The proportion of respondents answering that they have experienced construction errors due to language related challenges and culture differences were 66 and 44 percent, respectively. Some would argue that these numbers, and the corresponding numbers regarding the working environment and safety are unacceptably high. Still, when asked, very few could tell of any concrete examples in which any of these were the case. It is also important to note that the percentages represent the number of respondents that have either experienced or observed any of these things within the *past three years*. Thus, they do not necessarily indicate that these are common problems over time. Also, when asked related questions in the interviews, almost no one could tell of any concrete events in which language or culture related challenges caused accidents or near accidents.

Segregation at the construction site leaves many workers with an experience of being divided into "us and them". The various nationalities typically keep to themselves during lunch and other breaks; especially in cases where the different professions constitute pure national working groups. This minimises the chances of extensive inclusion and involvement. The study also revealed several examples of Norwegians that felt alone among eastern European co-workers.

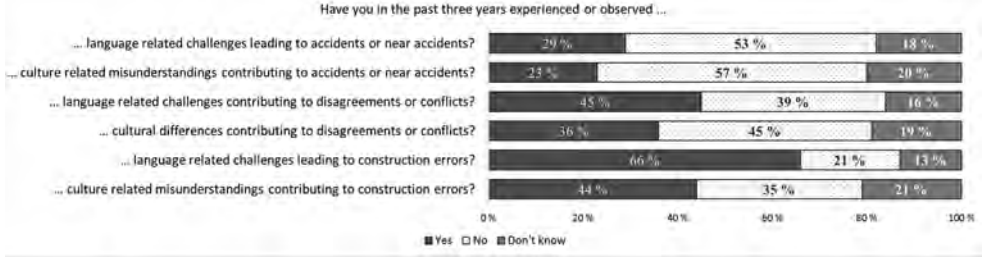


Figure 2. Results from survey. Managers' answers as to whether they have experienced or observed construction errors, disagreements/conflicts or accidents/near accidents due to culture or language in the past three years.

As described in 4.3, challenges may be affected by the duration of relations. As such, focusing on duration of relations is also important to maintain safety and working environment at multicultural workplaces. For the workers, getting to know each other may affect whether one is able to communicate well about the daily work. This may again influence everyday working environment and safety. For employers, on the other hand, the duration of relations affects how much one is willing to invest in the workers, for example through training and competence building. This, in turn, influences the prerequisites for the individual worker to work safely and in accordance with expectations from the employer and colleagues.

4.5 *Many measures focus on short-term solutions*

The study explored measures that exist to utilise the opportunities and cope with the challenges that come with using migrant workforce. It was found that most employers and construction sites do not put in direct efforts to promote the opportunities and fully utilise the resources that are represented by migrant workers. Measures in large degree aim to cope with specific challenges in the present. As language is mostly brought forward as a challenge, most measures are also related to cope with linguistic issues. These include among other, translations of safety materials, HSE training in other languages, language courses, language requirements at work sites, and arrangements of working groups by language.

In practice, some measures, such as language requirements can be difficult to comply with due to available workforce. Further, measures such as arrangement of working groups by language contribute to a bad spiral where migrant workers do not get the opportunity to learn Norwegian as they only work with people that speak the same language as themselves. And again, they do not get to work with Norwegians because they do not speak Norwegian. This is an example of a measure that is reactive in nature and contribute in a short-term perspective to reduce linguistic challenges at a specific construction site, but do not focus on coping with the challenges in the long run.

Many of the informants highlight communication as a key to success at multicultural workplaces. Language is a large part of communication, however not the only part. Cultural differences also influence communication at workplaces.

In the study, it has been seen that duration of relations is important for successful construction sites. Knowing each other and having a relation to each other is also important for understanding and communication between workers, and between workers and managers.

The study indicates both opportunities and challenges with multicultural workplaces. However, to reduce the challenges and to benefit from the potential that lies in a migrant workforce, further solutions need to be developed, e.g. tools and measures. The study shows that there are already many measures in place, especially to cope with challenges, but many of them aim to solve problems in a short-term perspective. There is also a need for measures and tools which aim to reduce the challenges in the future, rather than reducing consequences here and now. Further, there is a need to look for opportunities that come with migrant workers.

5 CONCLUSIONS AND FURTHER RESEARCH

The challenges with migrant workers are quite well documented and studied both internationally and recently in Norway. There has, however, been many unjustified assumptions when it comes to multicultural labour and consequences for safety, working environment and work performance. This study has documented both challenges and strengths by using migrant workers, and also a construction industry that has introduced different measures to create well-functioning multicultural workplaces.

The findings in this study cannot conclude that migrant workers are involved in more accidents than Norwegian workers. However, when it comes to reporting of unwanted events, significant differences were found between migrant and Norwegian workers.

Rather than resulting in many accidents or near accidents, challenges related to language and culture were found to lead to more consequences for work performance and working environment. Several informants said they imagined language and culture causing safety related consequences, but few could provide concrete examples of events in which this had been the case.

Further, it was found that employment relationships are of greater importance than nationality when it comes to challenges related to multicultural workplaces. Duration of relations is important to maintain safety and working environment at multicultural workplaces.

The study also shows that the potential for benefitting from knowledge and experiences among migrant workers is to a small degree realised, and the obvious focus in the industry is on the challenges.

Measures to cope with the challenges were found to mostly be of a short-term nature handling challenges right here and now, whereas few measures are of such a nature that they improve long-term conditions.

Migrant labour has become a natural part of the Norwegian construction industry, and the industry partners must take responsibility and work together to cope with the challenges and promote the opportunities. Measures to promote good multicultural workplaces include improving managers' understanding of cultural differences, adapting leadership style, as well as making conscious decisions regarding organisation of the work. A future-oriented focus is needed, including measures that may lead to long-term gains for the industry, and measures to lift the opportunities and exploit the potential that lies in the inequalities between different groups of workers at multicultural workplaces.

ACKNOWLEDGEMENTS

The project was initiated by BNL and funded by the Confederation of Norwegian Enterprise's (NHO's) Working Environment Fund.

Observations were led by Bjørn Nygaard from ideThandling in Denmark.

REFERENCES

Adler, N., & Gundersen, A., 2008. *International dimensions of organizational behavior (5th ed.)*. Mason, Ohio: South-Western.

BNL, 2018. *About BNL. Accessed: 08.02.2018. Available from: http://www.bnl.no/dette-er-bnl/english/*

BNL, 2017. HMS-kort statistikk 2017. [In Norwegian HSE-card statistics 2017] Downloaded: 26.08.2017. Available from: http://www.bnl.no/dette-er-bnl/aktuelt/hms-kort-statistikk-2017/.

Dølvik, J. E., Eldring, L., Ødegård, A.M., 2005. Lavlønnskonkurranse og sosial dumping: Utfordringer for det seriøse arbeidslivet [In Norwegian: Low wage competition and social dumping: Challenges for the serious working life]. Fafo-report 485. Fafo: Oslo.

House, R.J., Hanges, P.J., Javidan, M., Dorfman, P.W., Gupta, V., 2004. Culture, Leadership, and Organizations. The GLOBE Study of 62 Societies. SAGE Publications, Inc.

Kilskar, S.S, Wasilkiewicz, K., Nygaard, B., Øren, A., 2017. *Flerkulturelle arbeidsplasser i byggenæringen: Kartlegging av muligheter og utfordringer* [In Norwegian: Multicultural workplaces in the construction industry: Mapping of possibilities and challenges] SINTEF-report: 2017:00352 SINTEF: Trondheim.

Nergaard, K., 2016. Organisasjonsgrader, tariffavtaledekning og arbeidskonflikter 2014 [In Norwegian: Organisational degree, tariff agreements coverage and labour conflicts 2014]. Fafo-note: 2016:07. Fafo: Oslo.

NLIA, 2012. KOMPASS Tema nr. 2 2012. Arbeidsskader blant utenlandske arbeidstakere. [Work injuries among migrant workers] Trondheim: Norwegian Labour Inspection Authority.

NLIA, 2016. KOMPASS Tema nr. 8 2016 Ulykker i bygg og anlegg 2015 [Accidents in the construction industry 2015] Trondheim: Norwegian Labour Inspection Authority.

NLIA, 2017a. Statistikk—arbeidsskadedødsfall [In Norwegian: Statistics—work related fatalities] Downloaded: 06.12.2017. Available from: https://arbeidstilsynet.no/om-oss/statistikk/arbeidsskadedodsfall/.

NLIA, 2017b. KOMPASS Tema nr. 1 2017 Helseproblemer og ulykker i bygg og anlegg [Health problems and accidents in the construction industry] Trondheim: Norwegian Labour Inspection Authority.

Salminen, S., 2011. Are Immigrants at Increased Risk of Occupational Injury? A Literature Review. *The Ergonomics Open Journal,* I4: 125–130.

SSB, 2017a. Key figures for immigration and immigrants [Internet] Statistics Norway. Accessed: 06.11.2017. Available from: https://www.ssb.no/en/innvandring-og-innvandrere/nokkeltall.

SSB, 2017b. Accidents at work. Published: 03.10.2017. Downloaded: 06.12.2017. Available from: https://www.ssb.no/en/helse/statistikker/arbulykker.

Warner-Søderholm, G., 2012a. But we're not all Vikings! Intercultural Identity within a Nordic Context. Journal of Intercultural Communication. Issue 29: 19pp.

Warner-Søderholm, G., 2012b. Culture Matters: Norwegian Cultural Identity Within a Scandinavian Context. SAGE OPEN, October-December 2012: 1–12. DOI: 10.1177/2158244012471350.

Wasilkiewicz, K., Albrechtsen E., & Antonsen, S., 2016. Occupational safety in a globalized construction industry: a study on Polish workers in Norway*, *Policy and Practice in Health and Safety,* 14:2, 128–143, DOI: 10.1080/14773996.2016.1256553.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Reliability and safety engineering: The principles innovation and optimisation of German and Japanese product constructions

S. Bracke
Chair of Reliability Engineering and Risk Analytics, University of Wuppertal, Wuppertal, Germany

M. Inoue
Department of Mechanical Engineering Informatics, Meiji University, Kanagawa, Japan

ABSTRACT: The industrialisation of Europe and Asia leads to the development of innovative and reliable technical complex products. Industrial nations like Germany and Japan took separate ways to highly developed industrial nations. This leads to different goals and philosophies in terms of reliability and safety design regarding product and manufacturing process development. This paper shows essential aspects of the differences in German and Japanese reliability and safety engineering. Base of operations is the fundamental difference of the realisation of the principles "innovation" versus "optimisation": German product functionality contains often innovative solutions; Japanese product functionality shows often optimised solutions. These different designs are caused by different engineering philosophies and factors: The paper discusses the historical factors, cultural factors and geographical conditions, which lead to the different technical reliability and safety engineering solutions in Germany and Japan. Finally, the paper shows two examples of German and Japanese technical products and engineering solutions to illustrate the different strategies in German and Japanese product design.

1 INTRODUCTION

The industrialisation of Europe and Asia leads to the development of innovative and reliable technical complex products. Industrial nations like Germany and Japan took separate ways to highly developed industrial nations. Both countries have the ability to develop and manufacture technical complex products on a high reliability and safety standard. But a detailed look at Japanese and German products shows different characteristics regarding the principles "Innovation" and "Optimisation", especially in case of reliability and safety aspects. New developed German product generations include often innovations regarding functionality. Japanese product engineering considers very often the optimisation in relation to the previous product generation. An example for illustration is the comparison of the Japanese and German railway system—including trains, railway net and ticket machine logistics—with the focus on criterions like innovation, optimisation, efficiency and reliability. The railway system was invented in Europe (Great Britain, afterwards also Germany; cf. (Griffin 2010) and (Pollard 1981)). Subsequently the U.S. established the railway system in Japan after opening the country for trade. The Japanese engineers adapted and optimised the railway system: The concept (e.g.: Separation of high-speed and regional train system, as well as rail cargo), the logistics (e.g.: entering and leaving trains by passengers) and the ticket machine (e.g.: parallel paying, non-limited bank-note); cf. (Nussbaum and Roth 2005).

However, both approaches "Innovation" and "Optimisation" lead to different engineering solutions regarding technical complex products in terms of reliability and safety. Both reliability and safety design approaches have advantages and open the chance to learn from each other. This paper shows essential aspects of the differences in German and Japanese reliability and safety engineering within a research study. Base of operations is the fundamental difference of the approaches "innovation" versus "optimisation" within the product development process. Subsequently the paper shows different impacts—e.g. historical, culture and geographical—on innovative and optimised engineering.

2 GOAL OF RESEARCH STUDY

In this research study, the main influences which support product innovation and optimisation within an engineering process are analysed. The goal of the

research study is to work out the reasons, why German product functionality is mostly characterised by innovative solutions, whereas Japanese products often show optimised solutions.

In a first step, this paper illustrates the principals and influences with supporting or restraining impacts on creative engineering and product innovations (cf. section 3.1 and 3.2). Furthermore indicators for the measurement of innovation are shown (cf. section 3.3). Subsequently, the paper discusses main influences and aspects of Japanese and German historical factors, culture and education factors, as well as geographical conditions, which lead to the engineering preference regarding the approach "Innovation" or "Optimisation" (cf. chapter 4). Lastly, German and Japanese product examples are explained to illustrate the applied approaches with focus on innovative or optimised engineering.

3 BASE OF OPERATIONS: INFLUENCES ON CREATIVITY AND INNOVATION

The principals and main influences, which are supporting creativity and innovative thinking, are worked out by several authors, e.g. Linneweh, Rammert et al. and Dehio et al. Section 3.1 shows a summary of influences, which support creativity and innovative thinking within engineering processes. Section 3.2 shows a short overview regarding influences with restraining impacts on innovations. Finally, section 3.3 shows essential indicators, which can be used for the measurement of innovation capability.

3.1 *Influences with supporting impacts on creative thinking and innovative engineering*

The main influences supporting creative thinking and the development of product innovations are as follows:

1. Possibility of free thinking (Linneweh 1984),
2. Possibility for realisation, especially financial support, state subsidy; cf. (Rammert et al. 2016) and (BDI et al. 2017),
3. Consumer society based on customer expectations based on life style thinking (Rammert et al. 2016). Note: In the past, the customer expressed requirements; in the present the life style is an essential part of the decision making additionally,
4. Supply and demand, the base of a free market economy; cf. (Rammert et al. 2016),
5. Social innovations based on customer's expectations; the customers are familiar with dynamic changes, cf. (Rammert et al. 2016),
6. Motivation and incentive (e.g.: cash/money, reward, appreciation, cf. (Linneweh 1984)),
7. Knowledge and new technologies lead to new product functionalities; cf. (Rammert et al. 2016),
8. Digitisation (BDI et al. 2017) and
9. Wish for diversity (e.g.: Gender, Ethnos, Religion; cf. (BDI et al. 2017) and (Linneweh 1984).

3.2 *Influences with restraining impacts on innovations*

The main influences, which are restraining or stopping creative thinking and the developing of product innovations, are listed below:

1. Organisation with rules, structure, norms and regulations (Linneweh 1984),
2. Bureaucracy (Linneweh 1984),
3. Huge enterprises (combined with hierarchical structure); cf. (Dehio et al. 2006),
4. No possibility for creative thinking, combined with the hindrance of personal development (BDI et al. 2017),
5. No possibility for experiments and mistakes; cf. (BDI et al. 2017).

3.3 *Indicators to measure innovation capability*

The measurement of the innovation capability is based on fundamental indicators in relation to a country/state, which are shown in the following list (cf. (Dehio et al. 2006) and (BDI et al. 2017)):

1. Patents per inhabitant,
2. Scientific publications per inhabitant,
3. Investments regarding science in relation to the gross domestic product (GDP),
4. Turnover/sales based on new or improved products and
5. Balance of trade.

4 COMPARISON OF GERMAN AND JAPANESE ENGINEERING

Industrial nations like Germany and Japan took separate ways to highly developed industrial nations leading to different focuses regarding the approaches "Innovation" and "Optimisation" within engineering processes. This chapter works out the impacts leading to different philosophies and designs in terms of "Innovation" and "Optimisation" within product and manufacturing engineering in the respective countries. Section 4.1 discusses the historical impact in relation to the different ways of the industrialisation of Japan and Germany. The culture and society impact—especially the value

of education—is shown in section 4.2. The section 4.3 gives an overview regarding the main geographical and locational impacts.

4.1 *Historical impacts: Industrialisation in Germany and Japan*

The industrialisation of Europe and Asia led to the development of innovative and reliable technical complex products. Industrial nations like Germany and Japan took separate ways to the level of highly developed industrial nations. The industrialisation started in Europe within the second half of 18th century in England, and subsequently in Germany; cf. (Griffin 2010) and (Pollard 1981). Within the time of the German empire, the industrialisation was growing rapidly. Exemplary, the industrialisation process can be explained by the development of the railway system. The beginning of the railway history was in 1804 with the introduction of the first steam locomotive by Richard Trevithick. Shortly afterwards in 1825, the first public railway system was opened by Stockton and Darlington Railway in England (Tomlinson 1915), enabling the transport of people and rail cargo. Due to the fact that traveling times were reduced in Europe and North America, the industrial revolution itself was supported by the railway system. Furthermore it was pushed by the European wars, especially the German-French war from 1870 to 1871, where the transport of war goods was important. Today, the high-speed train ICE ("Inter City Express") is an essential part of the German traffic system: The four ICE generations were developed between 1990 and 2017 (Jehle et al. 2006).

However in Japan the industrialisation starts in the end of the 19th century caused by external pressure from the United States. The external pressure led to the end of the Edo era (1603 until 1867) starting the new Meiji area (1868 until 1912) with the rapidly growing industrialisation (Nussbaum and Roth 2005). Within the Edo era, Japan was encapsulated to the rest of the world. Entry and departure for foreigners were not allowed (exception: small exchange with china and Netherlands; colony island Dejima) and contact to other states were on a minimum level. In 1854, four war ships led by Matthew Perry were entering the harbour of Uraga. Based on a message from US president Millard Fillmore, the Tokugawa government was pressurised to start trade with the United States (otherwise U.S. go to war); cf. (Morinosuke 1976). The arrangement included the opening of the harbours Shimoda and Hakodate and the possibility for trading. Parallel, the Edo era ended and the reform of the empire led to the Meiji era, starting in 1868. The gentry of war ended, the constitution was modernised and the development of the country—especially from the engineering point of view—started. This point is exemplary illustrated with the Japanese railway system; cf. (Givoni 2006): Within the Meiji era, the railway system was developed—but limited to a track range of 1067 mm. Therefore, the speed level was limited to 100 km/h until 1950. Huge steps of the railway system development followed in the sixties and subsequently resulting in the development of the high-speed trains ("Shinkansen" – high speed) in the nineties parallel to Europe (1992: Series type 400, 1997: E3; 1999 E3–1000). Today the railway system in Japan is the most reliable and accurate system worldwide (Givoni 2006).

The other historical aspect impacting Japanese product development is World War II. Japanese companies trained their engineers for multiple tasks and skills (called "Tanoko" in Japanese) based on long-term employment and long-term trading with suppliers because of the labour shortage after World War II. As a result Japanese engineers are able to work in many different teams on a very efficient level. Furthermore, Japanese engineers are able to find optimum design solutions faster than their colleagues in other countries through a trial and error processes by "*Suriawase*" which is the Japanese way for negotiation among several teams of engineers including design teams, production divisions, and parts manufacturers (suppliers) from the initial design phase to the detail design phase concurrently (Inoue et al. 2011).

As a comparison between Germany and Japan: In Germany, the innovative engineering was pushed by the industrialisation under environmental conditions, which allows free thinking. Prussia and the subsequently following German empire's education system ("Humboldtian Model"; (Baumgart 1990)) allow free thinking and development of creative solutions (cf. section 4.2). However Japanese engineering was pushed to open the country, caused by external pressure from the U.S. Many technologies were imported during Meiji era from Europe and U.S. and adapted to Japanese environment (Nussbaum and Roth 2005). This was a part of the base of operations for Japanese focus on optimisation of imported technical products and processes.

4.2 *Culture impacts: Education system, society and work environment*

The culture and society of Germany and Japan is very different. Germany society is influenced by a hierarchical system parallel to an established social market economy. As mentioned in section 4.1, the Prussia educations system, founded by Wilhelm Humbold, focusses on a comprehensive education based on a combination of research and teaching. The

"Humboldtian Model" (Baumgart 1990) includes arts and sciences with research. The goal is comprehensive general learning and cultural knowledge. Nowadays, the model is still base of the German education system and is the fundament for creative, innovative thinking. Furthermore the social market economy is established since 1949 in Germany. Social market economy allows free entrepreneurial acting while social compensation is considered. The social market economy was pushed by Chancellor Ludwig Erhard and is an important factor to ensure free thinking in a secured environment.

The Japanese society is influenced by a strong hierarchical structure. Acting within the society is characterised by discipline. Furthermore, education has a high value within Japanese society (like in all Confucian orientated Asian countries).

Important character of the Japanese working environment is the principal of life long employment (until 20th century); cf. (Japan 2012). Furthermore, the loyalty regarding the employer is very important. Loyalty, discipline and healthy food lead to low numbers of employees absence due to illness (Japan: Average 1%; Germany: Average 3.2%). The strong connection between working environment and personal life style leads to a strong identification of Japanese employees with their work.

Both countries have an education system, which ensures a high educational qualification level. The differences are the social market economy, hierarchical system, employee identification with work and discipline. These aspects have a high influence on different engineering philosophies regarding product innovation and product optimisation. On the one hand strong hierarchical structure, discipline and loyalty support product optimisation. On the other hand, free thinking within a social market economy supports creative product innovations. The requirement for the ability to create product innovation or optimisation is the educational fundament, which is given in both countries.

4.3 *Geographical and location impacts*

A strong impact on innovative or optimised solutions are geographical impacts. The comparison between Japan and Germany shows obvious differences: Japan is located at the rupture zone of four tectonic plates of the earth crust:

- North American plate in the north,
- Eurasien plate in the west,
- Philippines plate in the south and
- Pacific plate in the east.

Each year, the plates are drifting within a range of some centimetres. The pacific plate moves below the Eurasien plate which leads to volcano eruption and many earthquakes. In average, 73 earthquakes with a Richter magnitude of 4 or higher occur in Japan per month (thereof nine per month with magnitude 5 or higher; 1.4 earthquakes per month with magnitude 6 or higher), cf. (Earthquake Research Committee et al. 2011).

However Germany is located in the centre of Europe, nowadays without any influences regarding earthquakes and volcano eruption.

This facts lead in Germany and Japan to very different strategies in case of safety and reliability engineering in many subjects, like facility engineering (e.g. buildings), mechanical engineering (e.g. railway systems) and electromechanical engineering (e.g. consumer goods). Japanese risk- and reliability analysis always considers safety aspects based on earthquake-vibration impacts in comparison to German products. This is reflected e.g. in constructional, material and electrical aspects as well as shutdown strategies for respective product concepts.

5 COMPARISON OF JAPANESE AND GERMAN CONSUMER PRODUCTS CONSTRUCTION

The comparison of Japanese and German consumer products—e.g. water boiler (section 5.1) and ticket machine (section 5.2) – shows the influences on historical, culture and geographical impacts on reliability and safety engineering solutions.

5.1 *Safety related construction: Water boiler*

Electric consumer goods like water boiler, micro waves and other electric devices, need electricity within the operation mode. The stressing of the connection cable between power source and electrical device—e.g. tensile force, folding force or torsion—can lead to a damage. As a consequence, a cable damage can lead to fire. The construction of Japanese electrical consumer products often includes an easy detachable connection between electric device and power source (cf. Fig. 1).

The reason regarding this construction is the geographical impact (cf. section 4.3): In case of an earthquake and the involved vibrations, the electrical devices within households are directly disconnected from the power source. This leads to a prevention of damage cases based on an electrical failure rout cause (e.g. damage of cable or technical device, which can cause fire). In comparison, the construction of electric consumer goods based on European respectively German standards mostly do not have such a safety related construction (cf. Fig. 2). The geographical location of Germany or Europe allows the negligence of possible earthquake impacts.

Figure 1. Japanese safety related construction: Easy detachable connection to avoid electric cable damage in case of force effect (e.g. caused by earthquake and involved vibrations).

Figure 2. Fixed—non-detachable—connection between electrical device and power source (European standard).

5.2 *Functionality related optimisation: Ticket machine in railway system*

Part of the industrialisation in Great Britain was the establishment of the railway system (cf. section 3.1), including the train, net, logistics and ticket machine. This transportation concept was an innovation regarding the transport of goods and people. It was built in Europe (especially Great Britain and Germany) and afterwards in North America. Subsequently the railway system was transferred to Japan at the end of 19th century. Each component was optimised by Japanese engineering, e.g. the functionality of the ticket machine.

A direct comparison of the basic function "paying process" shows the aspects of the European respectively German standard construction and the optimised Japanese construction.

European/German standard (cf. Fig. 3):

- Paying with coins, procedure is one-by-one,
- Paying with bank-note in limited form (e.g. 20 or 50 Euro), procedure one-by-one,
- Money return are coins in case of bank-note paying.

Japan optimised construction (cf. Fig. 4):

- Paying with coins can be parallel, because the slot has a hopper form, furthermore the automat has a sorting function,
- Paying with bank-note 10,000 Yen (approximately 100 Euro) is possible; procedures are "parallel" or "one-by-one",
- Money return in case of bank-note paying are bank-notes in optimal standard units.

Figure 3. Ticket machine: Standard solution, focus: single coin slot regarding paying by coins, procedure is one-by-one.

Figure 4. Ticket machine: Optimised solution based on Japanese Engineering. Focus: Parallel paying (e.g. coin slot hopper form).

6 SUMMARY

This paper shows essential aspects of the differences in German and Japanese reliability and safety engineering. The engineering approaches are based on the fundamental difference of the realised technical approaches "innovation" versus "optimisation": German product functionality is often characterised by innovative solutions, Japanese products show often optimised solutions regarding the previous product generation.

Base of operations are the influencing factors, which support or restrain creative thinking and, as a consequence, support product innovations or optimisations. Afterwards, the historical factors, culture factors and geographical conditions in Germany and Japan are analysed respectively, which leads to the different technical reliability and safety engineering solutions of German and Japanese engineering solutions.

Finally, the paper shows two examples of technical products and engineering solutions in Germany and Japan to illustrate the different strategies in German and Japanese product design.

However, both product design approaches—innovative on the one hand and optimised on the other have advantages and gives the chance to learn from each other.

REFERENCES

Baumgart, F. 1990. Between Reform and Reaction. Prussian school politics 1806–1859. Wissenschaftliche Buchgesellschaft, Darmstadt.

BDI & Technikwissenschaften 2017. Innovationsindikator. www.innovationsindikator.de release on 10.10.2017

Dehio, J., Engel, D., & Graskamp, R. 2006. Research and Innovation: Rank of Germany? Wirtschaftsdienst.

Earthquake Research Committee 2011. Supplementary Information to the Evaluation of Seismic Activity for January 2011. Earthquake Research Committee, 9. February 2011; release on 12. März 2011.

Griffin, E. 2010. A Short History of the British Industrial Revolution. London, Palgrave.

Inoue, M., Mogi, R., Nahm, Y.-E. & Ishikawa, H. 2011. Design Support for "Suriawase": Japanese Way for Negotiation among Several Teams. In Dan Frey, Shuichi Fukuda and Georg Rock (eds.), *Improving Complex Systems Today*, Springer-Verlag, pp. 385–392.

Japan Statistical Yearbook 2012. Statistics Bureau, Ministry of Internal Affairs and Communication.

Jehle, P., Naumann, R. and Schach, R. 2006. Transrapid and Wheel-Rail-High speed trains: A comprehensive comparison of systems. (in German). Springer.

Linneweh, K. 1984. Creative Thinking. Nadolski.

Morinosuke, K. (1976). History of japanese external relations. Volume 1: Country opening to Meiji Restauration. (in German) Steiner, Wiesbaden.

Moshe Givoni: Development and Impact of the Modern High-speed Train: A Review. In: Transport Reviews. 26, Nr. 5, 2006, ISSN 0144–1647, S. 593–611.

Nussbaum, L.-F. and Roth, K. 2005. Japan encyclopedia. Cambridge: Harvard University Press.

Pollard, S. 1981. Peaceful Conquest. The Industrialisation of Europe 1760–1970, Oxford.

Rammert, W., Windeler, A., Knoblauch, H., & Hutter, M. 2016. Innovation and Society today. Wiesbaden: Springer.

Tomlinson, W. W. 1915. The North Eastern Railway: Its rise and development. Andrew Reid and Company.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Safety and risk management in oil & gas industry: Development of safety x-factor model

Deshai Botheju
Independent Researcher, Sandefjord, Norway

ABSTRACT: It is noted that difficult market conditions faced by the oil industry during last several years have been manifested in a negative safety performance. No direct relationship exists to explain this trend. Even though many stakeholders instinctively believe that extreme cost-efficiency drives seen within the industry are somehow responsible for this outcome, any clear-cut mechanisms or pathways have not yet been proposed. This paper presents the preliminary development of a schematic model basis intended to explain the impacts of economic pressure on safety performance of a profit oriented organization when faced by market challenges. Further development of this model basis is expected to provide a clearer picture of this interrelation between safety performance and economic performance.

1 INTRODUCTION

Due to the downturn of oil industry ensued during last several years, many dramatic changes have been seen within the management structures of commercial entities engaged in the business. Some of these changes have unintentional consequences on the safety culture, barrier management, and the overall safety performance of organizations. These consequences can manifest as long lasting, and sometimes delayed, impacts and could lead to catastrophic major accidents as well as gradually deteriorating HSE performance. Therefore, timely recognition of these impacts and the pathways are crucial to avert short-term and long-term loses. Reasons (1998) pointed out that technologically complex high-tech industries are more vulnerable to organizational (major) accidents due to their intricate systems and subsystems that no single person could comprehend in isolation. Accordingly, weaken barriers or latent flaws accumulated during an economically hard time period could stay dormant for years or decades before they come into play a role in a major accident.

This paper explains possible mechanisms behind the recent negative trend in safety performance observed within the oil & gas industry proposedly instigated by the dramatic downturn of the crude oil economy seen during the last couple of years (Botheju & Abeysinghe, 2017). While recognizing the necessity of adopting certain organizational changes in order to face the new economic reality of the industry, the paper highlights the importance of understanding the drivers behind this discomforting trend that is threatening the prudent safety management procedures established over decades. It is recognized that there can be certain chain-linked relations between some innocent looking cost-cutting measures and the organization's safety culture dictating the overall behavior towards its safety performance and barrier management.

This article can also assist in developing prudent guidelines that could be used to implement a robust safety management system that performs even under challenging economic circumstances without compromising the safety and well-being of the organization.

The article is based on long-term experience of the authors, and careful observation of industrial dynamics related to safety and risk management. It is intended that this paper provides a much needed insight into the driving factors behind safety performance change currently being observed within the oil & gas industry, while establishing a schematic model basis to comprehend its safety dynamics under cost-efficiency pressure.

2 RECENT TRENDS IN SAFETY PERFORMANCE

The previous works of the authors (Botheju & Abeysinghe, 2017; Botheju & Abeysinghe, 2016) have argued that the downturn of oil industry has manifested itself as a sudden nosedive of the overall safety performance. Even more threateningly, some of the resultant impacts, especially regarding the process safety risks, could come into effect years later from now. In relation to Norwegian petroleum industry, Engen et al. (2017) have pointed out

that, while the level of safety and working environment conditions still remain relatively high, several safety challenges and serious conditions were starting to be manifested during the last few years.

The existence of an apparent correlation between economic pressure and the safety performance had previously been identified by other authors as well (Rasmussen, 1997; Coles, et al., 2000; Barden, and Lodestone, 2006; Young, 2015). However, many of the past case examples that were relating major accidents to cost-cutting measures, had straightforward links connecting key management decisions to poor safety barrier management (Chauhan, 2005; US Chemical Safety and Hazard Investigation Board, 2007). The current trend in the oil industry, that we are experiencing right now, is more intricate where such clear-cut pathways are still difficult to be observed. Among previous modelling attempts, Rasmussen's (1997) migration model is quite unique. He proposed the existence of a boundary of functionally acceptable performance; operating outside of that would lead to accidents. Rasmussen further theorized that there exists a gradient created by management's pressure directing the organization towards higher efficiency. This gradient, unless sufficiently counterbalanced by a gradient of safety culture, can gradually migrate the activities towards the functionally acceptable performance boundary.

The challenge, therefore, is to exactly recognize the driving mechanisms behind this recent trend. The most of the commercial entities continue to emphasize that they are still prioritizing safety as a paramount factor during their operations, and refuse to accept that any of their management actions could have led to a deteriorated safety performance.

In a way, what companies are claiming has a surface truth. Unlike in the past eras, the modern socio-ethical context and the associated legal and regulatory frameworks leave only a limited room for management bodies to initiate direct actions that could openly jeopardize safety. And above all, the most companies understand the gravity of such detrimental actions nowadays. Therefore, no sensible management would consciously support any action that clearly leads to poor safety performance.

Nevertheless, this paper argues that there are certain pathways linking the dramatic cost-efficiency actions and deteriorating safety performance. We denote these links as "Safety X-factors". The schematic model so named as "Safety X-factors Model", which is currently in its development stage, tries to explain these enigmatic connections and establish them within a model structure.

3 SAFETY X-FACTOR MODEL

As indicated in section 2, the overall safety performance within the oil and gas industry is being influenced in a negative trend aligned with the market downturn, with much visible evidence. Nevertheless, no management is accepting that they are actually driving this trend. This raises the question "what mechanisms are responsible for this trend then ?". The Safety X-factor model (abbreviated as SXF model) is presented as a basic attempt to answer this question. Note that this paper only presents its preliminary development. Figure 1 provides a schematic illustration of this model showing some of the key components and pathways proposed.

3.1 *What is SXF model*

SXF can be introduced as a basic schematic model in development aimed at explaining the safety performance outcome of an organization (or an entire

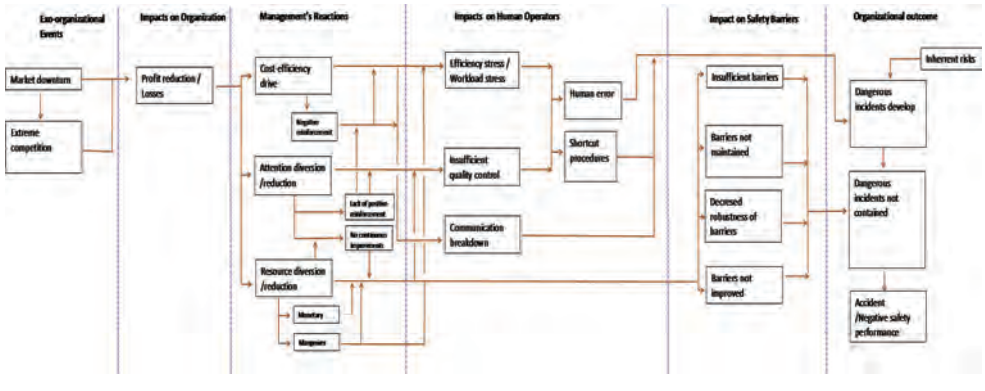


Figure 1. Schematic Illustration of the *Safety X-factor Model* (Preliminary).

industry in a broader sense) when faced by one or more exo-organizational events bearing certain economic impacts to the organization. In the current context, the relevant exo-organizational events are the market downturn triggered by low crude oil prices, and the extreme competition (the second is actually, to a larger degree, a resultant of the first in this case). The consequential prime organizational impact is the reduction of profit or even losses.

3.2 *The model structure*

The model is included with the behavioral dynamics of three (3) entities; namely, (i) Management, (ii) Employees/operators (Executors), and (iii) the technical safety barrier systems. In the current version of the model, the organizational and procedural safety barriers are not separately addressed, but considered to be embedded within the human operators behavior.

The model describes various influence pathways between the above 3 entities, that eventually lead to negative safety performance materialized as near misses, accidents, poor health, or degrading environmental impacts.

3.3 *Management reactions*

The SXF model describes the apparent management reactions as a threefold approach; i.e. (i) cost-efficiency drive, (ii) Attention diversion/reduction, (iii) Resource diversion/reduction; the each of these is briefly described below.

3.3.1 *Cost-efficiency drive*

This is the natural "panic action" by most managements when faced by market threats. Loss of profits forces top management to run for extreme cost reduction /efficiency enhancement measures. While the necessity of some such actions is justified, some extreme measures could significantly change an existing positive safety culture as explained by Botheju & Abeysinghe (2017). It is argued that such cost-efficiency measures could lead to negative reinforcements on behavioral safety under certain situations.

3.3.2 *Attention diversion/reduction*

When good economies exist backed by favorable market conditions, top managements usually have high attention to HSE aspects. This is the normal behavior of organizations possessing an adequately good safety performance history. However, a management is tested when it faces difficult market conditions and poor economic performance. Will they be capable of maintaining the same level of commitment to safety under pressing economic situations? Often, many managements yield under such situations and their attention is significantly drawn from their usual emphasis on safe performance to other more urging matters such as economic issues and profitability of operations. This in turn reduces the positive reinforcement previously received by operators for their good safety performance.

3.3.3 *Resource diversion/reduction*

Diversion of monetary and human resources to other purposes, than for the continuous improvements of safety systems as well as for proper maintenance of existing safety barriers, is a natural trend that can be observed under economically challenging periods.

When it comes to many technical safety barriers, they are costly to establish and their benefits are not immediately apparent, or may be perceived as "it can wait". Meanwhile, the actual cost components associated with such safety systems are very real and will have to be immediately dispatched from the existing economic resources. Under this scenario, many managements could be tripped into abandon or delay various continuous safety improvement actions and maintenance/upgrade actions. This forever conflict between production vs. protection (Reasons, 2000) can lean heavily towards production when the resources are more limited.

3.4 *Human operator impacts*

The above described management reactions generate multiple responses from human operators, as briefly described below.

3.4.1 *Efficiency stress and workload stress*

The extreme cost-efficiency drives combined with negative reinforcements, and the lack of positive reinforcements originating from attention diversion reaction leads to high level of worker stress which is further accelerated by reduced amount of human resources (Resource Diversion/Reduction).

3.4.2 *Insufficient quality control*

Both the attention diversion management reaction and the resource diversion/reduction reaction generate this impact on human operations. The lack of positive reinforcements further aids safety quality control barriers.

3.4.3 *Communication breakdown*

The extreme cost-efficiency drives coupled with associated negative reinforcements lead to increased rift between the management and the executors leading to the breakdown of efficient two-way communication. A coherent safety management becomes increasingly difficult under such communication breakdown situation.

3.4.4 *Human error*

A human error is an inadvertent event generated through the actions of human operators while trying to follow a preplanned course of actions. The likelihood of human error rapidly increases when operators are under stress. The probability of discovering such error is also diminishes in the face of insufficient quality control.

3.4.5 *Shortcut procedures*

A shortcut procedure is an intentional diversion from the planned (safe) course of action. Operators resort to such short-cut procedures either because such actions are indirectly promoted by the organizational culture or else as a way-out from the high workload and stress. The lack of safety quality control would further contribute to this situation. The short-cuts may work during most of the times but eventually can trigger chain reactions leading to dangerous incidents/accidents.

3.5 *Performance of technical safety barriers*

Stemming primarily from the resource diversion management reaction, technical safety barriers experience following impacts described below.

3.5.1 *Insufficient barriers*

If the amount of barriers are not sufficient to cover all the high-probability accident scenarios, then the risk of an incident developing into a full scale accident is high.

3.5.2 *Barriers not maintained*

All technical safety barriers require certain maintenance to keep them under optimum performance level. The lack of maintenance leads to their degradation over time and therefore their reliability decreases.

3.5.3 *Decreased robustness*

The robustness can be defined as the spare capacity of a safety barrier to handle accident scenarios beyond the normally expected magnitudes, frequencies, and operational conditions. A more robust safety system has a high tolerance for errors, so that it can still break the propagating incidents originating from significant human errors.

3.5.4 *Barriers not improved*

All safety systems need continuous improvements /adjustments over time. The facilities face different kinds of risks during their lifetimes. For example, an old facility may have a different risk picture compared to a similar but newer facility. Similarly, plant modifications lead to changed risk scenarios. Therefore, the safety barriers must continuously be adopted or upgraded according to the changing conditions.

3.6 *Organizational outcomes*

On top of the existing inherent risks of a facility, additional pathways leading to the development of dangerous incidents are generated as a result of the aforementioned human operator impacts. Meanwhile, due to the simultaneous weakening of the safety barriers, the possibility of containing/resisting dangerous incidents under propagation becomes increasingly difficult. This makes the likelihood of accidents higher. The term "accident" here also embodies other slow phase outcomes such as poor health and weak environmental performance.

4 CONCLUSIONS

This paper briefly presents the preliminary development of a schematic model aimed at recognizing mechanisms behind the apparent correlation between economic pressure vs. safety performance of a profit oriented organization. It is theorized that there exists several indirect pathways leading to a deteriorated safety performance initiating from an economically stressed management, even when the management may not intentionally compromise safety for economic gains.

So named Safety X-factor (SXF) Model is to be further expanded in order to fully explain the negative safety performance trends observed under market downturn situations. The eventual aim is to describe an organization's safety culture using more concrete and measurable terms.

REFERENCES

Barden, R.H., & Pacific, Lodestone. (2006). Cost savings at the expense of quality, safety, and the environment; A plastic molding example. 2nd International Conference on Power Electronics Systems and Applications, Hong Kong, China. doi: 10.1109/PESA.2006.343063.

Botheju, D., & Abeysinghe, K. (2016). Safety and environmental management under cost pressure: Threats, challenges, and solutions. In Proceedings of the SPE International Conference and Exhibition on Health, Safety, Security, Environment, and Social Responsibility, Stavanger. Society of Petroleum Engineers (SPE). doi: 10.2118/179467-ms.

Botheju, D., & Abeysinghe, K. (2017). New directions in safety & environmental management and policy: A brief update on petroleum industry. *Safety & Reliability; Theory and Applications* (Book), Taylor & Francis Group, London. ISBN 9781138629370.

Chauhan, T.R. (2005). The unfolding of Bhopal disaster. Journal of Loss Prevention in the Process Industries, Vol. 18, pp. 205–208.

Coles, E., Smith, D., & Tombs, S. (2000). Risk management and society (Book). Kluwer Academic Publishers.

Engen, O.A., Nistov, A., Håland, Ø.A., Joranger, Ø., Borthne, M., Bjerkeli, H., A., Sjåland, C., Furre, R.E., Kveim, M., Herland, T., Jonassen, Ø., Andersen, E., G., Lindheim, I., Skogesal, T., Sabel, P., Knudsen, S., Holhjem, A. (2017). Helse, arbeidsmiljo og sikkerhet i petroleumsvirksomheten (In Norwegian, a summary in English). Report available online https://www.regjeringen.no/contentassets/0a217a1b53a84a5b877bc526d67a5c5f/helse-arbeidsmiljo-og-sikkerhet-i-petroleumsvirksomheten.pdf.

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. Safety Science, Vol. 27, No. 2/3, pp. 183–213.

Reasons, J. (1998). Achieving a safe culture: Theory and practice. J. Work and Stress, Vol. 12, No. 3, pp. 293–306.

Reasons, J. (2000). Safety paradoxes and safety culture. J. Injury Control & Safety Promotion, Vol. 7, No. 1, pp. 3–14.

US Chemical Safety and Hazard Investigation Board. (2007). Investigation report; Refinery explosion and fire, BP Texas City, Texas.

Young, C. (2015). Process safety and low oil prices. Online article, available at http://www.jmcampbell.com/tip-of-the-month/2015/03/process-safety-and-low-oil-prices.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Contributors to successful safety level in the Norwegian railway sector

Dag Wilhelm Aarsland
The Norwegian Railway Directorate, Oslo, Norway

Jørn Vatn
Department of Mechanical and Industrial Engineering, NTNU, Trondheim, Norway

ABSTRACT: The Norwegian government is reforming the railway sector. The reform was formally implemented 1st of January 2017. This has transformed the structure of the sector to ensure clearer division of responsibility, better coordination of service and infrastructure improvements, better aligned incentives and a more customer focused industry. As with any major change in any sector a concern is raised regarding the potential negative safety implications. This paper investigates safety implications of the reform and overall safety responsibilities as the reform progresses. Focus is on factors that can contribute to a future successful safety level. A risk analysis of safety implications of the reform, lessons learned from a similar restructuring in the Norwegian aviation industry and general safety literature, are combined and compared to factors that contributed to a railway accident. A railway accident will be investigated to search for correlation with risk factors in a reformed sector. Special focus will be on the responsibilities of the Norwegian Railway Directorate.

1 INTRODUCTION

1.1 *Background*

To reduce costs within the railway sector and to help build a better railway for its customers, the Norwegian government is reforming the railway sector. The reform has transformed the structure of the sector into several government owned shareholding companies for a better coordination of service and infrastructure improvements. To put pressure on the Infrastructure Manager (IM) and Railway Undertakings (RU) to improve for a more customer focused industry, an incentive regime is implemented.

The Norwegian Railway Directorate was established to ensure coordination and to achieve political goals within the sector. To open for a competing passenger transport marked, the Directorate have requested for quotation for future operation from RUs within the international market.

The Directorate shall have the overall responsibility for coordination and safety measures of the safety level in the sector.

Bane NOR SF is the government owned infrastructure manager and shall carry out work according to budgets. Contracts between the Directorate and Bane NOR regulate the overall responsibilities for Bane NOR, and follow-up of Bane NOR is based on long term effect achievement. This gives Bane NOR a flexibility for future development and to manage their construction-, operation-, maintenance- and safety work within budget limits.

Statens Jernbanetilsyn (SJT) is the National Safety Auhority (NSA), and is responsible to supervise and follow-up the actors according to the railway safety regulations. SJT has certified Bane NOR and existing Norwegian RUs accordingly.

The Railway Undertakings (RUs) such as, passenger transport and cargo transport use the railway for the purpose which is stated in their certificate.

The responsibility for safety level delegated to the Directorate is not well defined, and what kind of responsibility the Directorate can or shall have concerning safety aspects is still under considerations.

To adapt to the reform, Bane NOR implemented a significant organizational change in 2016. To be more cost efficient, costs during the period 2016–2017 are reduced by NOK 750 Mill. For the next period of 4 years, an increase in cost-efficiency is planned. This may increase the organization financial pressure. Focus areas will be economy, staffing, new technology, increase of productivity at all levels and extended use of contractors.

The complexity within the sector has increased due to increased number of interfaces, contract arrangements, splitting of competence among companies and potential reluctance to share information.

The following 3 findings is a backdrop and motivation to proceed with the search for preventive safety factors:

1. To study safety implications before implementation of the reform, the Government ordered an overall risk analysis from SafeTec AS. SafeTec AS identified nineteen main hazards. Three of these safety hazards are identified as; (i) the interaction capability within the sector, (ii) importance and necessity to control the overall risk picture, and (iii) postponed maintenance due to financial pressure.
2. SJT reports 13 serious accidents in the first eight months of 2017, compared to 10 in 2016. Five more accidents are under investigation.
3. SINTEF (2005) investigated safety aspect during restructuring in the Norwegian aviation industry in the period 2000–2005. Lessons learned from this restructuring are used as a starting point for the investigations. Four of the main threats which affect safety were found to be:

   - Reduced organisational ability to discover and interpret development of hazards.
   - Loss of barriers due to loss of organisational capability to make decisions critical to safety.
   - Uncontrolled reduction of safety margins.
   - Problems connected to interaction

Special focus will be on the responsibilities of the Norwegian Railway Directorate which holds a sector responsibility for coordination and development of safety level. Companies that operates within a competing market may focus on the incentives of decision makers with a short-term finance and survival criteria instead of the focus of long term safety effects. This may lead to a systematic and silent migration toward an unacceptable safety level. Accidents or faults which affect a cost-effective operation will create conflicts and probably lead to an issue among the actors.

1.2 *Objective*

The objective of this paper is to focus on ways to structure the risk management, the importance of an overall risk-picture, interaction and cooperation among the different actors to achieve successful operation.

1.3 *Approach*

A risk analysis of safety implications of the reform, lessons learned from a similar restructuring in the Norwegian aviation industry and general safety literature, are combined and compared with factors that contributed to a railway accident. As a case study a railway accident is investigated to search for correlation with general findings in a reformed sector.

Special focus will be on the responsibilities of the Norwegian Railway Directorate.

2 A MOTIVATING EVENT FOR THIS STUDY

The event is a cargo train on a route at line Hovedbanen from Åndalsnes to Oslo, weighting 600 tons loaded with dangerous goods and driving at normal speed limit. Close to Bøn station the cargo train four rearmost wagons derailed caused by rail buckling. Both rolling stock and infrastructure were badly damaged, and the line was closed for 4 weeks.

We will return to the course of events considering obtained relevant literature.

The accident was investigated by the Accident Investigation Board Norway (Accident Investigation Board Norway, 2017). The intention from this investigation is to use and draw points for improvements of a future safety level in the railway sector.

3 THEORY AND BASIC TERMS

The objective of this section is to define the term interaction considering organisational and inter-organisational interface within safety of the railway sector. Conflicting goals in the interface in an organisation and between organisations and the potential of incremental drift towards an unsafe state of operation, need to be explained. The need for an overall coordination of the actors within the sector also need to be explained.

3.1 *Relevant safety perspective*

In theory and practice on safety management related to organisations and to prevent accidents, it is in the past decades developed several safety perspectives. However, accidents still happen. The perspective of Resilient Engineering (RE) is developed from earlier safety/accident perspectives as the 6th perspective (SINTEF, 2010).

> Hollnagel (2008) claims that '*… a resilient system….the intrinsic ability of an organization (system) to adjust its functioning prior to or following disturbances to continue working in face of the presence of a continuous stress or major mishaps*'

This ability to adapt can be done both before and in response to changes and disturbances. This system-change capability allows continuity of operation, even in the case of major accidents or continuous disturbances.

The Resilient Engineering perspective, suggests that an organization must follow, thus observing the

daily normal situation of change, thereby identifying changes that may be a chime for an undesired event. The uniqueness of the RE is not the focus on proactivity, but the ambition of being proactive in a system sense. In other words, a resilient organization can be summed up as resilient when knowing that unexpected problems are present and unpredictable. This involves a form of proactive control where the system (organization or sector) anticipates what is coming and minimizing or eliminating unwanted variation and taking advantage of productive variability (SINTEF 2010).

3.2 *Investigation method and root-cause analysis*

The situation is analysed by the STEP investigation method (Sequentially Timed Events Plotting, see Hendrick & Benner, 1987). STEP is a multi-sequential accident analysis method and is a good method that helps to illustrate events and hazards that causes an incidents/accident (Herrera & Woltjer, 2009).

A Fault Tree Analysis (FTA) is used to investigate the root cause of the basic event Risk Influencing Factors (RIFs). A fault tree is a logic diagram that displays the interrelationships between a critical event (accident) in a system and the causes for this event. The strength of FTA is the deductive reasoning to establish the basic events. However, the influences of contributing factors cannot be handled in the FTA. Therefor a hybrid approach motivated by see Gran et al. (2012) is used where the RIFs obtained from the STEP is linked to the basic events.

3.3 *Risk Influencing Factors (RIFs)*

A RIF is a set of relatively stable conditions influencing the risk. It is not an event, and it is not a state that influences over time. RIFs are thus conditions that may be influenced or improved by specific actions.

A risk model often comprises a formal logical representation of the system. Fault- and event tree analysis is often building blocks in such a representation. Barriers, safety functions and/or layers of protection are typically represented by basic events. Probabilities are assigned to the outcome of the basic events, i.e., success or failure. A wide range of factors and conditions will influence the outcome of the basic events, and these need to be considered. In the literature we find many approaches to link such "soft" factors to the formal logical representation. The literature also presents different names for such factors. In human reliability analysis (HRA) the term performance shaping factors (PSFs) and error producing conditions (EPCs) are used (see e.g., Rausand, 2011). The term contributing success factors (SCFs) has also been used in an attempt to establish resilience based early warning indicators (Øien et al., 2010). To avoid different set of factors for positive and negative issues, we will in the following use the term risk influencing factor (RIF) as a general tem, see e.g., Vinnem et al., (2012) for risk modelling of planning and execution of critical tasks. Interaction and coordination as factor for successful operation

The term interaction in terms of safety can be explained as the way the actors cooperate to coordinate their actions, to obtain a safety achievement, e.g. proactively cooperate to reduce a risk for an incident or accident.

> Rasmussen (1997) claims that '*individual decision makers cannot see the complete risk picture and judge the state of multiple defences conditionally depending on decisions taken by other people in other departments and organisations.*'
>
> Maidment (1998) claims; '*to keep safety levels intact, safety must have a very high-profile aspect in the contractual arrangements between the separate companies.*' To monitor the overall risk-picture in a sector it will be important to coordinate all assigned contracts on a higher level.

3.4 *Monitoring safety*

The safety in restructuring processes can be monitored most effectively by following up on all levels, from individual and technical equipment to sector level. Symptoms of security issues may appear on all of these levels. Security researchers have until now been interested in technical conditions, human factors and organizational levels and to some extent the level of interaction. Today, therefore, there is a lack of knowledge and methods at inter-organizational level/sector level, as well as knowledge and methods that embrace all levels (Rasmussen, 1997).

Problem connected to interaction might arise in connection with restructuring processes with a high level of conflicts. Conflicts will arise, and to avoid safety problems the conflict level need to be regulated to an overall acceptable level.

Management and work planning in any organization apply different control strategies, depending on time horizon, stability of systems, and predictability of disturbances. Management is based on monitoring of performance with reference to plans, budgets and schedules, and control is aimed at removing deviations (Rasmussen 1997). Risk management in a dynamic market in which all actors continuously strive to adapt to changes and the dynamic markets, require an explicit identification of the boundaries of safe operation together

with efforts to make these boundaries visible and to learn and cope with these boundaries.

To give a mental model as a basis we borrow the "layer of protection" concept from the process industry. A layer of protection can be seen as safety barriers, and they are often structured in the way they will be activated in an accident scenario. Compared to the presentation given by e.g., Rausand (2014) and CCPS (2007) we simplify and distinguish between:

a. Process design (by using inherently safe design principles).
b. Control, using basic control functions, alarms, and operator responses to keep the system in normal (steady) state.
c. Prevention, using safety-instrumented systems (SISs) and other safety barriers to act upon deviations from normal state and thereby prevent an undesired event from occurring.
d. Mitigation, recovery and emergency response in case of control is lost.

Figure 1 shows the layers of protection.

Having these categories of layers as a basis we now define: A normal operation corresponds to layer (b). Normal operation may involve adaption to varying conditions as emphasized by resilience engineering. But normal operation does not involve use of extraordinary safety functions representing layer (c). If the primary control is lost and there is a deviation from the normal state, layer (c) is intended to bring the system into a safe state. We define successful operation as a situation where layer (c) is invoked, and can bring the system back to the primary control state. In some situation layer (c) is invoked but without being able to bring the system back to the primary control state. This is often referred to as a state of system outside engineering control. If this is the case, we define successful recovery as the situation where the system could be brought under control and back to its normal state (after repairs, clean-up and so forth).

Figure 1. Protection layers (adapted and simplified from CCPS, 2007).

In the literature, we also find other mental models for accident avoidance than the protection layers understanding shown in Figure 2. One class of such models is the safety envelope model shown in Figure 2, the inner ellipsis limited by the defined operation boundary (DOB), corresponds broadly to what is controlled by protection layers (a) and (b) in the protection layer model in Figure 2.

We may also treat the elliptical band limited by the boundary of controllability (BC) very similar to what is controlled by protection layer (c) in Figure 2. However, there might be differences since the layer of protection concept in Figure 2 is usually based on very explicit identification and design of safety functions, whereas the safety envelope perspective in Figure does not have such an explicit understanding of safety functions. The outer elliptic band limited by the safe envelope boundary (SEB) represents extra safety margins with is a conceptual link to what is controlled by protection layer (d) in Figure 2 but at this level, similarities are not that evident.

A conceptual advantage of the safe envelope model over the layer of protection model is the visualization of the track of the activity over time. In the safe envelope, safe operation is defined as an operation where the operating boundaries are exceeded, but where the boundary of controllability is not exceeded (as illustrated in Figure 2). We will consider an operation to be safe if we are able to remain in the inner ellipsis of Figure 2. We will, however consider two special cases (sub sets) of normal operations as successful operation. The first situation is where there are extraordinary conditions making it very difficult to operate within defined operating boundaries. Examples could be a situation where a long-term development of a fault is hard to control for example due to a situation where the energy in a track is building up close to a rail buckling, or it could be related to extreme weather conditions. If we, despite these challenges, are able to operate within the defined boundaries

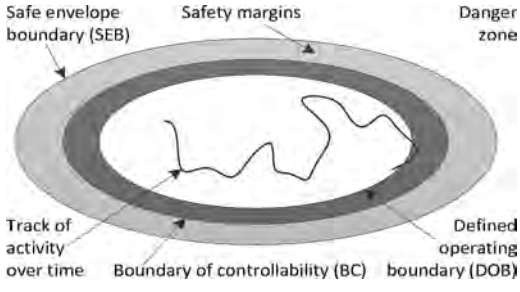


Figure 2. Safe envelope (Adapted from Hale et al., 2007).

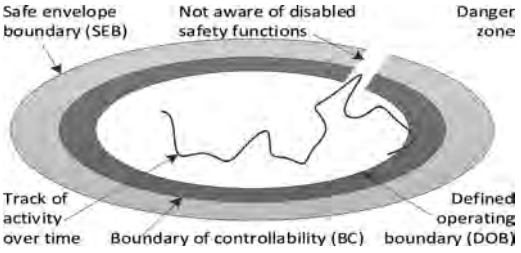


Figure 3. Safe envelope with broken boundaries due to unnoticed deficiencies or problems (Based on Hale et al. 2007).

we will define that the particular operation to be within safety margins. This will most likely represent a situation where operation runs smoothly and production performance will be high. A smooth process represents a situation where we use minimal effort to control the operation, and presumptively better prepared to handle any process upset that might occur. However, there is an argument in this latter situation that would question how safe such an operation is. The fact that there are no process upsets or situations to act upon, could be passivizing the operations crew in the sharp end limiting their situation awareness (see, e.g., Endsley, 1995, for a thorough discussion on situation awareness).

In the safety envelope model in Figure 2 the boundaries are illustrated as static in time. For many operations the boundaries or safety functions cannot be seen as static. We will in the following distinguish between a situation where (i) operating personnel is aware of this situation, and (ii) operating personnel is *not* aware of the status of a safety function or the limits for safe operation. A very easy example of situation (ii) is car driving where the temperature has decreased below zero degrees Celsius on a wet road without the attention of the driver.

Another situation is obvious very critical.

In a situation where the operation is following a trajectory towards an accident without the attention of operating personnel for a while, and then operating personnel become aware of the situation, and succeed in bringing the situation under control. This is visualised in Figure 3. If personnel become unaware of the situation, and unable to bring the situation under control, will cause the operation to move into the danger zone.

4 ANALYSIS OF THE EVENT

The Hovedbanen line is the oldest line in Norway and was opened in 1854. During the last 20 years it has been reported that the line has had several defects and geometry faults, and was planned for renewal in 2020. The national network track geometry condition is regularly measured with a special inspection vehicle. The inspection vehicle takes pictures every 10 meter of tracks during inspections, these are available to all employees in Bane NOR. To detect long term changes or establish fault trends to track and surroundings, Bane NOR need to compare pictures from different periods. The local staff and train drivers often observe changes, and their knowledge is an important input to achieve a good safety level.

The vegetation clearing along the railway increased exposure for direct sunlight to the rails, increased temperature and tension in the track. Missing reference points to be able to measure track position was unknown to the management.

The investigation of the situation is split into two STEP diagrams: One (long-term) shows the period from 2014 until May 31st in 2016, and the other shows the situation causing derailment. The STEP diagrams show that track faults were present already from April 2015 and further escalated until the accident occurred in Mai 2016. STEP diagrams will be published in a master thesis in September 2018 (Aarsland, 2018).

Track irregularities escalated due to construction work, and reported in August 2015, this was interpreted to be within acceptable limits.

From the *long-term* STEP-diagram the following safety problems were obtained during the analysis:

1. Economy: Maintenance back-log, renewal planned in 2020. Missing overall risk-situation picture. Late plan for setting out lacking exact track position marks.
2. Inadequate risk-assessment before construction work to be on an already faulty track.
3. Construction work affects track stability, and vegetation clearing along track increases sun exposure to rails.
4. Project management reports potential buckling problem to the Hovedbanen line management.
5. Several reports from drivers of irregular track conditions.
6. Inspection vehicle 4th measure, again reports geometric track fault.
7. Missing overall risk-situation control. Temperature rise and direct sun exposure to rails results in increased compression forces in track.
8. Track lateral forces on limit to buckling.
9. Overall cost-reduction scheme implemented in 2016.

From the *short-term* STEP diagram (situation causing derailment that day) the following safety problems were obtained during the analysis:

10. Loaded cargo train at regular line speed releases track forces.
11. Track buckles and results in derailment.

The fault tree in Figure 4 displays the interrelationships between the accident and causes of accident. The purpose of this FTA is to look at and

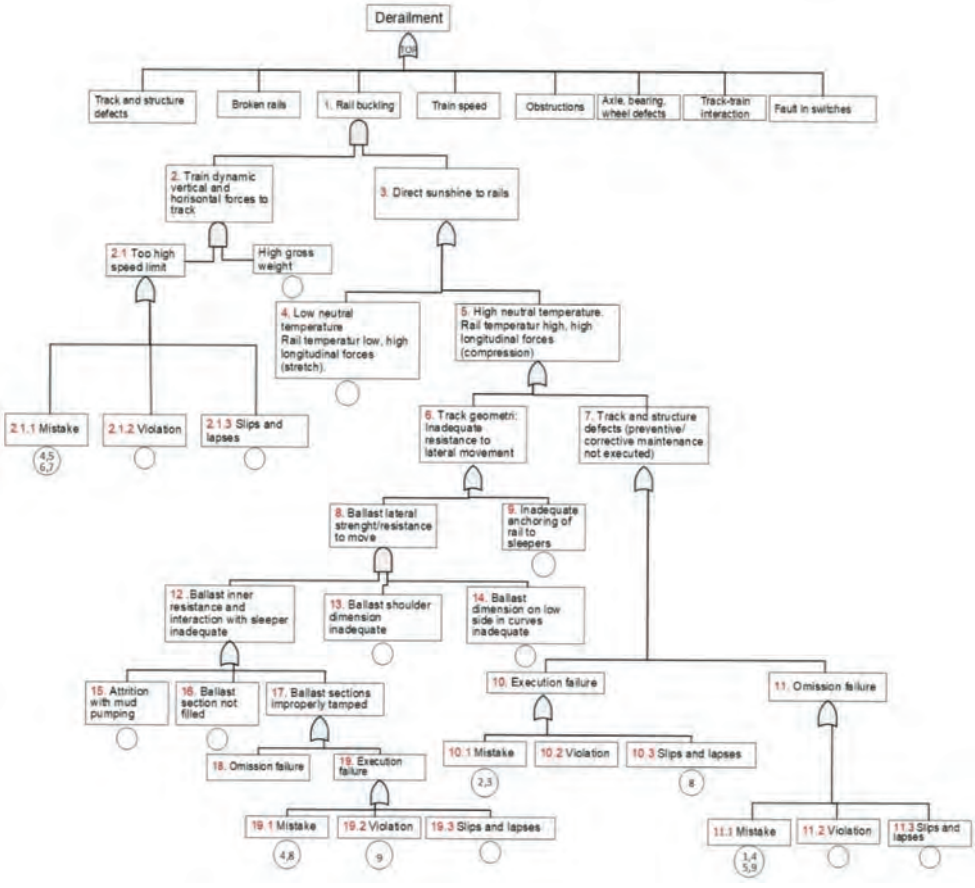


Figure 4. Fault tree with listed RIFs.

evaluate the prerequisites that led to the TOP event 'derailment'. The TOP event is deducted to causal event and further to basic events.

The RIF model developed by Gran et al. (2012) shows that it is advantageous to split basic events into mistake, violation and slips & lapses. We have considered this expedient and used the model in this study. RIFs found as contributors to the accident are shown in the diagram with the following corresponding numbers:

1. Missing overall risk-situation picture as a fundament for renewal investment.
2. Competence to carry out risk assessment.
3. Competence to understand risk of construction work on a degraded and faulty line.
4. Knowledge about neutral state of forces in the track (neutral rail temperature).
5. Missing overall risk-picture to prevent derailment
6. Missing management situation awareness and hence a wrong decision to keep planned traffic level and at regular train speed limit.
7. Management situation awareness, when track geometry is at critical level and in case of increased rail temperature, failure to establish barriers against buckling and derailment e.g. reduce train speed or close line.
8. Management risk awareness of construction work on an already degraded and faulty track.
9. Inadequate shared risk understanding at organisational management levels.

Based on the railway line quality, each of the network line managers reports annual renewal requirements. Requirements are considered and given budget priority in a national network renewal plan by Bane NOR senior management. The management policy is that safety measures always have first priority.

The track geodetic control marks along the railway line were not established, and made it difficult for the line management to know position of the track and hence critical rail temperature for buckling.

Such a condition requires special attention and measures. Risk assessment was carried out by the project management, but did not consider local aging condition, possible weakening lateral stability due to piling of a cable duct or the curvature of the track. Clearing of vegetation causing increased sun exposure was neither considered in the risk assessment.

It seems clear that a missing overall risk-situation picture, situation awareness, competence level, and interaction in the interface between the actors were contributing factors to the accident.

5 DISCUSSION

Comparing previously described SafeTec AS hazards and the SINTEF (2005) bullet points with the contributing factors for the derailment accident case, the ability to control overall risk situation is important to discover, interpret and control safety level.

A premise in a restructured sector, is that all actors shall have the responsibility for their own safety under the supervision of SJT. SJT shall ensure that Bane NOR and the RU's operation in the interface complies with railway safety regulations. However, not all actors in the sector are under the supervision of SJT.

Cooperation between and contribution from all actors are important input to achieve a dynamic risk-picture for managing risk in a sector. In the contract agreement between the Directorate and Bane NOR, Bane NOR shall establish and maintain a railway sector risk-picture, reflecting types of hazards and locations, and be based on risk assessment and incidents.

The risk of drifting away from safety margin, and 'Silent Drifting' due to financial pressure are indicated both in the SafeTec study and as uncontrolled reduction of safety margins in the SINTEF (2005) investigation. In the accident case, missing situation risk picture and ability to understand risk development, and decide priority to renewal of the line, resulted in long-term track fault development. Increased competition, and the future arrival of new railway undertakings and railway maintenance companies, will increase financial pressure. This can cause a risk of "Silent Deviation" in the operative sharp-end.

The term "Silent Deviation" is used by companies in the oil industry to describe the mismatch between procedures and actual work practices (Tinmannsvik, 2008). This is an attitude among the staff to reduce the conflict between getting the job done and compliance with procedures, and is an expression of how routine violations of written procedures tacitly become accepted practice. Reason (1997) uses the term "necessary violations", where operators adjusts the balance between procedures and knowledge based problem solving to get acceptable workload.

Problems connected to organizational interaction (SINTEF. 2005), SafeTec hazard 3., and the contributors to the accident: Lack of interaction between project, railway line management, Bane NOR management and involvement of maintenance personnel resulted in lack of shared safety related information.

In the restructured sector with an increase of organisational and inter-organisational interfaces, with increased financial pressure, might these challenge the safety margins. To understand how safety are affected at all interfaces, interaction and coordination at all levels is crucial. According to Rasmussen (1997) individual decision makers cannot see the complete risk picture and judge the state of multiple defences conditionally depending on decisions taken by other people in other organisations. Therefor monitoring of safety level depends on sharing of information, interaction and cooperation.

The Directorate has the responsibility for the development of a good sector safety level. SJT shall ensure that certified actors have a system for continuous improvement of safety, but do not have the mandate to set specific goals. The Directorate can however, set goals to be achieved in the contract agreements.

Exception from this is railway vehicle maintenance workshops, entities in charge of maintenance, vehicle owners or manufacturers. According to Maidment (1998) contracts must be coordinated on a high level in the sector. This is the responsibility of the Directorate.

6 SUMMARY AND CONCLUSIONS

A case study is used to shed light on overall safety responsibility, and possible risk factors in a reform with high political focus on the main success factors 'cost efficiency' and 'safety level' in the railway sector. An accident has been analysed with a traditional investigation method to reveal safety problems. The STEP method (Hendrick & Benner, 1987) has been used as a basis for the investigation. A fault tree analysis is carried out as a basis to find the Risk Influencing Factors contributing to the accident.

The case study shows that interaction in the interface between different organisational units is important to ensure safety. The study also shows that a situation risk-picture is required as an input to make safety related decisions. Risk management in a dynamic market in which all actors continu-

ously strive to adapt to changes, require an explicit identification of the boundaries of safe operation. Main threats learned from the investigation after restructure of the Norwegian aviation industry can be closed as follows:

The Directorate should contribute to strengthen overall cooperation within the sector for Bane NOR to be able to implement and monitor a process for interaction in the different interfaces, and manage boundaries of acceptable safety level. The following points are essential for organising and to allocate responsibilities.

- Ability to regulate safety at an overall level. A resilient railway sector requires a dynamic risk-picture with contribution of risk-assessments from all actors at all levels. Bane NOR should develop and to actively use the risk-picture as a sector risk management tool.
- The Directorate should contribute to strengthen level of safety among actors in the contract agreements within the sector.
- Bane NOR should have responsibility to establish an appropriate level of information sharing, coordinate and propose safety measures in the interfaces between all actors in the sector.
- Bane NOR should establish an expedient level of risk monitoring.

Actor independent safety aspects that are important to follow up:

- Success Contributing Factors (SCFs) should be used to establish resilience based early warning indicators.
- High level of competence to understand RIFs at all levels in the sector, and as part of risk assessment.

It is recommended to extend the investigation to search for visible explicit boundaries as a tool for monitoring overall safety within the railway sector.

REFERENCES

Aarsland, D.W. (2018). *Contributors to successful safety level in the Norwegian railway sector*. Master thesis available September 2018. Norwegian University of Science and Technology.

Board Norway; AIBN (2017). Rapport om avsporingen nord for Bøn stasjon på Hovedbanen 31. Mai 2016. *Rapport nr.: JB 2017/03.*

CCPS (2007). *Guidelines for Safe and Reliable Instrumented Protective Systems*. Wiley (Centre for Chemical Process Safety of AIChE), Hoboken, NJ.

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors.* 37(**1**), 32–64.

Gran, B.A., Bye, R., Nyheim, O.M., Okstad, E.H., Seljelid, J., Sklet, S., Vatn, J. and Vinnem, J.E. (2012). Evaluation of the Risk OMT model for maintenance work on major offshore process equipment. *Journal of Loss Prevention in the Process Industries,* 25(3), 582–593.

Hale, A.R., B.J.M. Ale, L.H.J. Goossens, T. Heijer, L.J Bellamy, M.L. Mud, A. Roelen, H. Baksteen, J. Post, I.A. Papazoglou, A. Bloemhoff, J.I.H. Oh. (2007). Modelling accidents for prioritizing prevention. *Reliability Engineering & System Safety*, **92**(12), 1701–1715.

Hendrick, K., and Benner, L. (1987). *Investigating accidents with STEP*. Marcel Dekker Inc. MY.

Herrera, I.A & Woltjer. R (2009). *Comparing a multi-linear (STEP) and systemic (FRAM) method for accident analysis.* Safety. Reliability and Risk Analysis: Theory, Methods and Applications—Martorell et al. (eds.).

Hollnagel E. (2008) "Preface" and "Safety management, looking back or looking forward" In: Hollnagel E, Nemeth CP, Dekker S, editors. Resilient Engineering Perspectives, Aldershot: Ashgate.

Maidment, D. (1998): *Privatisation and division into competing units as a challenge for safety management.* In Hale, AR & Baram, M. (Eds) (2000) *Safety Management: The Challenge of Organisational Change.* Pergamon, Oxford.

Øien, K., S. Massaiub, R.K. Tinmannsvik and F. Størseth (2010). Development of Early Warning Indicators based on Resilience Engineering. PSAM 10, June 7–11 2010, Seattle, USA

Rasmussen, J. (1997). *Risk Management in a Dynamic Society: A Modelling Problem.* Safety Science Vol 27, No.2 183–213.

Rausand, M. (2014). *Reliability of Safety-Critical Systems:*

Reason, J. (1990). *Human error*. Cambridge University Press. NY.

Safetec (2016). *Samferdselsdepartementet. Risikoanalyse av Jernbanereformen.* Hovedrapport ST-11574-1.

SINTEF (2005). *SINTEF RAPPORT: Flysikkerhet under omstillingsprosesser.* Report nr. STF50 A05102

SINTEF (2010). Organisational Accidents and Resilient Organisations: Six Perspectives. Revision 2.

Statens Havarikommisjon for Transport, Accident Investigation *Theory and Applications.* Wiley. Hoboken, NJ.

Tinmannsvik, R.K. (2008). 'Stille Avvik' – trussel eller mulighet? In Tinmannsvik R.K. (ed) *Robust arbeidspraksis:* 133–146. Trondheim: Tapir Akademiske Forlag.

Vinnem, J.E., Bye, R.J, Gran, B.A., Kongsvik, T, Nyheim, O.M., Okstad, E.H., Seljelid, J. and Vatn, J. (2012). Riskmodelling of maintenance work on major process equipment on offshore petroleum installations. *Journal of Loss Prevention in the Process Industries*. **25**(2). 274–292.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Tough men cry—learning from sharp end military aviation II

T.J. Steiro
Institute for Teacher Education, The Norwegian University of Science and Technology, Trondheim, Norway

C. Moldjord
The Norwegian Defence College, Department of Air Force Academy, Trondheim, Norway

ABSTRACT: The current study is scaffolding on the paper *"Tough Men Cry—Learning from Sharp End Military Aviation",* and presented at ESREL 2010. The previous paper reported on findings from a specialist course at the Royal Norwegian Air Force Academy (RNoAFA) termed *"Effectiveness in the Cockpit—Developing and Taking care of the Man in the Machine"* held in 2005. In 2010, five of the pilots were re-interviewed in order to understand more of the human side of military aviation and see more of the effects out in the squadrons. In the data from 2010, we see a transfer of knowledge taken back to the squadrons. We see from the interviews strong focus on performance and improvement. On the same time, there is hierarchy to reflect on regardless of hierarchical position. There is a need to protect ones position (professional hierarchy. However, the professional hierarchy can also be used positively to lead by example and pave the way for learning.

1 INTRODUCTION

The current study wanted to follow up on a previous study that was named "*Tough Men Cry—Learning from Sharp End Military Aviation*" that was and presented at ESREL 2010 (Steiro, Moldjord, Fredriksen & Firing, 2010). The previous paper reported on findings from a specialist course at RNoAFA named "*Effectiveness in the Cockpit—Developing and Taking care of the Man in the Machine*" in 2005 (Moldjord, 2007). All seven pilots participating in the course were interviewed and through a thematic analysis, the following topics were identified and discussed:

1. Challenging events
2. Emotions on the table
3. The role of the teacher/coach
4. Group based dialogue
5. Meaningful narratives
   (Firing & Moldjord, 2007; Steiro, Moldjord, Fredriksen & Firing, 2010).

In 2010, five of the pilots were re-interviewed in order to understand more of the human side of military aviation and see more of the effects out in the squadrons.

2 THEORETICAL FRAMEWORK

Historically, the educational practice in the military can be characterized by a behavioristic view of learning where teachers and instructors transmitted knowledge to the students. A wide set of reinforcements were used; both negative and positive consequences were systematically related to bad or good behavior respectively (Skinner, 1953). In the last 20 years process-oriented coaching seems to have turned the attention within educational practice in the direction of each student's thinking and feeling. Interviews indicate that students at the RNoAFA value practical exercises and cases and post-action reflection in groups as the most valuable learning for them (Steiro & Firing, 2009). Each student is addressed, stimulated and challenged to engage in their own process of assimilation and accommodation (Piaget, 1977). This method emphasizes the process of raising students' awareness and the development of each student's own thinking, emotions and reflection. In the same process a culture of coaching has arisen which supports this educational philosophy. Experienced officers serve as coaches and mentors in the process of constructing new knowledge. Their intention is to help the students in their own process of growth and realization of their potential as officers (Vygotsky, 1978). Today, The RNoFA has founded its educational practice in the concept of learning from experience as educational philosophy (Dewey, 1997, 1961; Skjevdal, Solheim & Henriksen, 1995). The RNoAFA has chosen to build its educational philosophy on three pillars:

1. Theory
2. Practical training
3. Reflection
   (Firing & Lien, 2007; Steiro & Firing, 2009)

The course "Effectiveness in the cockpit" was a single semester course offered in the autumn 2005 and winter 2006. The level is on both the individual and on a group level. The same level as we have kept on the analysis in this paper.

The course the cadets were provided aligned with the pedagogic model at the RNoAFA. I In accordance with how Bruner addressed the link between cognition, action and emotion (1986), these three pillars were also included in the curriculum as a second triangle. This is based on Bruner's thoughts (Bruner, 1986). Bruner (1986) addressed the link between cognition, action and emotion. One problem, however, with most military debriefing is that it often lacks the focus of emotions and individual inner experience (Folland, 2009). One way to get more aligned with Bruner is to broaden the debriefing by a more holistic reflection of thoughts, emotions, actions, team relations, and communication. In a Holistic Debrief (Moldjord 2015, Moldjord and Fredriksen 2017) we asked questions such as "how did you experience the event?", "what affected you most?", "how did we communicate?", and "how did we cooperate?" In addition to provide feedback to each other. In order to establish a sound safety culture, Reason (1997) pointed out the following elements as central for a safety culture in an organization:

- A reporting culture
- A just culture
- A learning culture
- A flexible culture

In order to obtain a sound safety culture, sensemaking can be seen as an important framework for safety. Sensemaking denotes processes of interpretation and meaning production through which individuals and groups engage their worlds on an ongoing basis (Cation & Patriotta; 2013; Patriotta & Brown, 2011; Weick, Sutcliffe, & Obstfeld, 2005). Sensemaking arises when people encounter situations that defy their current understandings and call for adjustment.

Several scholars have pointed out that learning from errors allows organizations to improve safety, reliability, and resilience (e.g., Reason, 1997; Ron, Lipshitz, & Popper, 2006; Weick, 1987; Weick & Sutcliffe, 2007; Catino & Patriotta, 2013). *"Errors, whether actual or anticipated, provide an empirical intersection between sensemaking and learning. In fact, learning—the detection, reporting, and correction of errors—is contingent upon the way in which individuals or groups interpret a problematic situation to themselves. In this respect, errors play an essential role in processes of construction of reality. This construction revolves around cognitive, emotional, and cultural sensemaking, which may or may not lead to processes of learning from errors"* (Catino & Patriotta, 2013:462). Zhao (2011) found that negative emotions can stimulate motivation to learning from errors when leaders encourage a positive and constructive view of errors and thereby alleviate individual tension and stress.

Bennis and Thomas (2007) claim that defining moments in a persons life is an opportunity to learn and to grow. Folland (2009) found that all the staff at the helicopter squadron had some challenging events that they brought up. Several had never shared the incidents or more precisely how the incidents had influenced on them. In our study, putting challenging events on the table made it more clearly for the other pilots to talk about each others experiences. Language does not transfer but also create and construct knowledge or realities (Bruner, 1986). Through the course, the cadets have transferred the experiences to meaningful narratives. This might have a positive health effect. Pennebaker (1997) argue that narratives organize overwhelming events to smaller units that hare easier to handle (Pennebaker, 1997). By constructing a narrative, the person goes through a process that helps the person to better understand both the experience and himself. The experiences have now got a structure and a meaning, which in turn may provide a feeling of solution. Something that has been viewed negatively might now be viewed more positively.

3 METHODS AND MATERIALS

We observed systematically throughout the course. The military psychologist took notes and recorded reflections, and the students also used writing reflection as a tool. These records of reflections were made available for a researcher at the RNoAFA who was not a part of the course. The same researcher conducted semi-structured interviews. The interviews with each of the seven pilots were carried out at the end of the half year course by one of the authors. The interviews were taped and later transcribed for analysis (Yin, 1994; Creswell, 1998; Thagaard, 1998). The research team consists of both teachers and researchers. One potential disadvantage is that the work can be viewed as action research (Greenwood & Levin, 2007), meaning that certain norms are imposed on the group and that the research then of course is not objective. These interviews were used as a background for planning and executing the research in 2010.

In the spring, just after the paper to ESREL 2010 was submitted, all pilots were invited to a new interview. Two of the pilots were not available due to practical reasons. Five of the pilots were available and agreed to be interviewed. The interviews were conducted at the pilots' respective workplaces

during 2010. The study of 2010 demonstrated that new learning had come up due to the educational framework created in the RNoAFA and in particular in the course "Effectiveness in cockpit". Military pilots are often performance orientated and aiming at high professional performance. Talking about incidents or accidents can put them in difficult position. All informants were informed of the purpose of the study and signed an informed written consent. All the informants also agreed that the interviews could be recorded. Most of the interviews lasted approximately 70 minutes, while one interview lasted 120 minutes. The recordings were later fully transcribed so both authors and interviewees had the full access to the material. A thematic analysis was performed and centered on challenging events, the way emotions played a role and what would happen in the dialogue after something had happened. This was scaffolding on the elements from the study of 2010. The categories were made through item-centered analysis (Thagaard, 1998).

The research questions was how did they use their experiences and how did they reflect on learning from experiences in the cockpit?

4 RESULTS AND DISCUSSION

In the interviews the pilot talked a lot of characteristic of being a pilot. So we identified this as an important theme for analysis. First, we extract from the interviews the pilots' perception of characteristics of being a military pilot and how it might enable and limit the ability to learn and share. Further, we have extracted three incidents that could serve as illustrations and situations that can complement each other in understanding the experiences of the pilots, how they are dealt with, and the role of the group. The three situations can be termed as:

- A potential dangerous situation of coming out of course
- A near miss within the envelope with the instructor taking over
- An actual severe incident and the handling of it afterwards

The examples illustrate different situations but also how the group might shape the narratives. Narratives and the reshaping of narratives are important to understand safety.

4.1 *The pilots perception of characteristics of being a military pilot and how it might enable and limit the ability to learn and share*

The first theme is of interest since it is about characteristics of being a military pilot and how it might enable and limit the ability to learn and share. All the pilots refer to the systematic processes of sharing in a debrief. The debrief is mission oriented and performed after each flight, and it is structured to focus on technical and tactical aspects. It can also include more emotional aspects. First, experiences after a flight is shared within the flight or crew members. In addition, other personnel can be brought in to contribute, for instance in a near miss or incident. The severity of situations such as a shooting accident calls for others to be brought in, typically flight safety officers second in command of the squadron, and additional personnel. The pilots talk of early selection and intensive training particular in earlier part of the careers. They also talk strongly of a performance culture where they measure each other, or are measured within a group. The professional hierarchy is of importance. The date of receiving the wings are important, the number of hours flown are typically shown on patches (i.e. *"1000 hours in Fighting Falcon",* that is the F-16, courses and check-outs you have been through and experience from international operations). A check-out in aviation means that the pilot is testet and has "passed the exam" and are allowed i.e. to operate the 20 mm machine canon in an F-16. Other Check-outs are for instance leading other aircraft, lead a formation at night etc. The peers from flight training are important, since they are able to share experiences and bonds from the pilot training in the United States. A pilot will go through training and check-out at the squadron. In the current study, there are both fighter pilots and helicopter pilots. The fighter pilots operate alone in the aircraft, except when an instructor sit in the back in a tandem F-16B. Otherwise, they always operate in pairs of two, implying that the relational aspect is present at all times. The helicopter pilots fly minimum with a system operator, and in addition there might be rescue divers, nurses or gunners, depending on the mission and type of helicopter. The team aspects is always there. At the same time, they will always be evaluated individually. In the interviews, all pilots in the current study touch upon their status and reflect on their ranking. They are all reflecting on their position at the squadron, acknowledging that being a newcomer is demanding regarding speaking up. All informants point out that it is easier to speak up for a more senior pilot. However, the senior pilots like themselves play a crucial role in facilitating openness in for example a helicopter crew. The pilots should be aware of the hierarchy and make sure that the other crewmembers are free to speak out or to speak first because of the hierarchy. At the same time it is important to have established a position as a good pilot. Informant 4 express the positioning of the pilots in the following way:

"You do not need to be the best. But overall, you must avoid being the worst. However, it is not communicated, it can be expressed by who gets some courses or positions to a greater degree than others, but not communicated and stated very clearly, but we still know it"

Pilots lower in the hierarchy can be too focused on not getting blamed and pointing to others mistake that might put them self in a relative better position. The same was reported in a study in the Italian Air Force (Catino & Patriotta, 2013). Informant 1 thinks that the youngest pilots are not necessarily that concerned of their position since they are in a learning phase and that there are so much to learn at the start in the squadron. However, this can vary from person to person. For some pilots, flying a fighter is their main purpose in life. For others it can be viewed as a job, and other aspects of life are as important. Informant 1 reflects on the balance between learning and performance:

"Where is the boundary between learning and achievement, the limit of how bad you can present your own presentation for learning purposes? I think there is a limit there. I can contribute a lot for learning purposes, but at one point I need to protect myself".

Informant 2 remember commanders that had stepped out and announced within the squadron:

"If the boss could share his biggest mistakes, so can I. Nevertheless, I do not trust for sharing everything with everyone. We have examples of people in the hierarchy uttering something that makes us suspicious. The utterings do not seem very thought through". If you are unsure you learn for yourself or only share experiences among the closest".

Informant 3 reflects regarding facilitating debrief after the introduction of Holistic Debrief in the Norwegian helicopter detachment in Afghanistan in 2010 (Folland, 2009; Moldjord, 2016).

"A system operator approached me and expressed concern whether he had to share all my feelings with me". I said, "You do not need to if you do not want to. But you can if you like". The system operator seemed happy with the answer. But we, the older ones, need to do it, otherwise the youngest will not follow".

Informant 4 points to another reflection looking at his pilot career describing an environment that is very focused getting on goals, focusing on improvement and performance. Informant 4 reason that this lead to focusing on the past and the future, but less on the present.

4.2 *A potential dangerous situation of coming out of course*

The first situation is reported by the pilot of a Bell 412 SP flying with a system operator. The system operator had little experience and they were going to fly for a long distance at wintertime in Norway. The pilot saw this as a good opportunity to let the system operator have more experience. In addition, he did not choose to have a more experienced system operator to sit in the back seat. They were flying and they missed some spots, they needed to reroute and the weather was getting worse with lower visibility. Informant 2 explains:

"We start talking. I adjust the speed to the weather the conditions, I do it, and we wrongly navigate, I support him on the map while flying the route … I use him, asking him questions. I see this far. How long distance can you see? I further asked him what he is looking forward. I think he is too much focusing on a spot that he is concentrating on". Also his vocalization, how much/little he talks, what he talks of, what he is talking of, he is focusing far out"

The system operator identifies their position and the tension decreases. They continue flying the route, but the visibility in a valley gets below 800 meter. The pilot remembers he was no longer confident with the situation. What happens is interesting. There is calmness in the cockpit, however;

"I say that we will turn and land, and then, very sudden, he agrees wholeheartedly. We land in a farm field, and I shut down the system. At the same time I reflect that this is not something overly dramatic. It is dramatic within acceptable boundaries".

While the engine was shutting down, the pilot starts by asking the system operator what he thinks first. In the informant's view, it is important that the system operator is not overly influenced by the hierarchy. The system operator was very concerned that he did not find geographical features from the map such as a wire and a church. He is very concerned with his own mistake. Informant 2 underlines that we did not find it out. The informant also explained to the system operator:

"I explain to him that it was my feeling that it was not ok to fly further. Then I will raise my voice. I also say to him that I expect the same the other way. We have talked about this before, but now we have an actual shared experience of this".

They are invited in for a meal at a farmer's house and stay there for a while. They arrive at the destination one day later. It was a trip with a lot of learning. After landing at the final destination they debriefed the mission. The informant sums up the experiences:

"It's a matter that I was particularly pleased with when I recognized my activation (of fear). This is the activation we have talked about, written about. This I've experienced before, and it's ok. So it's not disturbing, what's going on with me now, it never turns in to high stress. This is information that means that I'm affected and I have to take some extra consideration, calm down. I have the opportunity to make some bigger margins now, so I can do that. In several situations on the same trip, I have been activated, but not done anything about it and put it away, more like an emotionally oriented coping strategy, (telling myself) this I have to take care of after landing"

We see an interesting example of the social interaction in the helicopter. We see that the pilot by opening up for questions invites and encourage the system operator to also have a saying. The pilot report that the system operator very quickly agree, assuming that he might have thought of the same. The professional hierarchy play an important role and by reducing the authority gradient (Rosness, 2001). Authority gradient was first defined in aviation when it was noted that pilots and copilots may not communicate effectively in stressful situations if there is a significant difference in their authority, experience, or perceived expertise (Crosby & Croskerry, 2004).

4.3 *A near miss within the envelope with the instructor taking over*

The next situation is told by informant 2 which was flying with a new fighter pilot. They were sitting in an F-16B tandem with the instructor (informant 2) sitting in the back seat. The student was about to be checked out on firing the 20 mm machine gun on ground target. The limit is 5 seconds to ditch the fighter aircraft down, aim, fire, and then you are at minimum distance. Then there are additional 4–5 seconds to pull up. In the first part of the training mission the pilot student has dropped bombs which is quite usual on that type of training missions. The shooting is always at the end of the mission. At that time the type of ammunition required the pilots to get near the target, meaning that this is very stressful for both student and instructor. The student had already tried shooting but fired too far out and then missed the target too much. The student aimed for a third time. Informant 1 explains further that on the next round the student was very concentrated and he was paying extra attention.

"I imagine he's going to come in and get closer, and he comes in too close, so I yelled at him that he should get off the target, which means he's going to pull the stick back, but he did not hear that because he was aiming, aiming and shooting while I was once again roared to him that he should get off. Then I pulled the stick back at the same time as him. I wanted him to become aware of what was going on, so I pulled the stick back"

This is explained by the informant as target fixation: The student was too focused on the target, aiming at the target and forgetting about assessing the distance. It was a very brief window to interfere. As instructors, we want them to succeed.

"He immediately understood what had happened and became, I think, perplexed. I took over the flying in order for him to get back. We talked briefly; however, we were running out of fuel, so there was not much else to do than landing".

After landing, they talked about the situation. The student was expressing his views. He said he was close, but further expressed that he was not too fixated on the target, that he had some degree of control. They agreed that they had been within the safety limit. They performed a debrief. This was his first mission shooting with the 20 mm machine canon. Grades are assigned. The student fails one or two points, and one of them is safety, meaning he has to do the mission all over again. At the same time the informant explains central elements in the debrief session:

"The student had a need to explain what kind of situational awareness he had in the situation, what he thought and why it happened. (First) I explain what I thought, what I did and why I did it. It's on that level. I say that I'm an instructor open for that I might have made a mistake, perhaps intervening too early or doing something that's not good (for the student). Then he's able to explain how he experienced the situation".

We see from the example that student pilot and the instructor have a different risk perception. The instructor pilot get an uneasey feeling and instruct the pilot student to pull up. When there is no reaction, the instructor pilot shout and pull the stick at the same time. We see an interesting process of making sense of the situation. They both agree that they were within the safety envelope. At the same time, the instructor do not pass the pilot student. And the pilot student needs to perform the shooting with the 20 mm machine canon again.

4.4 *An actual severe incident and the handling of it afterwards*

The last situation is from informant 4. The informant was not directly involved and is about how the situation was handled afterwards. During night time shooting, a pilot missed the targets and the

bullets hit the shooting observation tower. There were no injuries on personnel, but it was a potentially very dangerous situation with possible several fatalities. Informant 4's role was being part of the squadron command team, responsible for following up on the involved pilot, knowing what was going to happen to him afterwards and prepare him for that. Informant 4 was also a part of the command team responsible for facilitating the debrief setting for the involved person. In addition, he was responsible for informing the rest of the squadron and the involved families. The challenge was that the pilot reported a high degree of blame and guilt. The informant attributed a lot of his 'follow up' handling to what he had learned at the RNoAFA:

> *"The most central aspect in my knowledge is what I learned at the Air Force Academy and the focus on reactions and experience with feelings like fear, shame and guilt. Now I looked at the person carefully and looked for facial expressions to see and meet his needs. I avoided making him sit alone in a blood test session for hours, although the formal test was alright. But to have someone there to talk with a familiar face so trust became focus versus procedures. That was the most important knowledge that I brought with me. I steered consciously clear of such thing as "have you taken care of the video", "can I see your data card", all things that indicate that we must check in order to understand what the pilot did wrong, versus take care of the person. The others can handle the formal things".*

The informant reflected on what his perspectives might have been earlier, that typically would have been focus on the individual, little training, little flying lately, should have been better prepared, tried another round. The informant reported that they did focus on the circumstances. As illustrated in 3.2., the ammunition had changed, now allowing for firing at longer distances. However, the shooting range was built in a period when firing at closer range. In addition, light conditions, the arrangement of flashlights, different location of the towers that were hit could have been better designed. So informant 4 points to the root causes and seeing the pilot in the wider context;

> *"Looking at the system he was part of".*

This quote can be seen in relation of learning at the organizational level as well as the individual level (Reason, 1997). The informant told us how important it is to take care of the person behind the mistake, not just the role of the pilot. The most important thing is to get the pilot back in the cockpit, and then it is crucial to take care of the person. Also, it is important to avoid blaming and too much focus on procedures right after an accident. At the same time, this incident has an organizational perspective. The entire organization can learn from the error and how it was handled afterwards. Before one can focus on learning, one needs to calm down strong emotions such as fear, guilt, anger, and pride. Creating an arena where emotions can be expressed is important before focusing on learning. In such a situation, Holistic Debrief is very relevant (Moldjord and Fredriksen, 2017). This aligns with the practice of double loop learning (Argyris & Schön, 1978). The informant reflects on differences between the squadron and other environments he knows of and claim they are good at sharing mistakes. Not everything is fine, but they are sharing experiences after a mission. If you experience a close encounter with a wire, you raise up in the morning meeting and share it with the others.

5 CONCLUSION

We see from the interviews strong focus on performance and improvement. On the same time, there is a hierarchy to reflect on regardless of position. There is a need to protect one's position. However, the hierarchy can also be used positively to lead by example. In the current study, all the pilots had experiences and lead several missions, and three of them had management positions meaning they had significant positions in the squadron. They all reflected on how their position can be used to lead and facilitate learning. All acknowledged the challenge of the degree of openness and position in the hierarchy. They also report of something that is "private" and not shared widely. This is in accordance with the findings of Catino & Patriotta (2013). It is also a challenge to create a learning safety culture (Reason, 1997). The informants talk of what we would term a generative reporting culture. That would also be termed as double loop learning (Argyris & Schön, 1978). Some of the learning that involves blame is linked to a more internal learning or is only shared with peers. This is especially evident if it threatens one's overall position as a pilot or in the hierarchy. Being in a strong position means that you easier can lead by example. All pilots reported favorable learning of the course Effectiveness in the cockpit. Moreover, as we interpret it, there has been some transfer of knowledge from the course to practical military leadership. It seems important that operative leaders have the knowledge of addressing the balance between performance, caring and learning after an error event or a mistake. Furthermore, it is important that such experiences are lifted from an individual level to something that the

entire organization can learn from. Therefore the organization needs to create arenas where trust is established and there is enough openness to share important experiences, like a Holistic Debrief (Moldjord and Fredriksen, 2017).

5.1 *The need for further studies*

It is of great value to follow a group of officers over a period of time since long term studies are often missing within this research field. We therefore plan to conduct a third round of interviews which hopefully will be accepted by all the pilots. We will look more in detail into the topics covered here in the current paper. The analysis in this paper indicates the strong effect of the context. Further studies on how the processes are shaped and could be managed is of interest. Also, we want to further examine how a squadron can enable for more sharing and learning. We will look more into the organizational and management and leadership aspects of learning. The pilots in the current study has now moved on to administrative positions. That means that it might be easier for them to understand the environment on the inside, but still obtain an analytical distance to the subjects, all while examining sensemaking processes over time and seeing the interplay between professional hierarchy and sensemaking processes.

REFERENCES

Argyris, C. & Schön, D. 1978. *Organizational Learning. A Theory of Action Perspective*. San Francisco: Jossey- Bass.

Bennis, W.G. & Thomas, J.C. 2007. Leading for Lifetime: How Defining Moments shape the Leaders of Today and Tomorrow. Boston, Mass. Harvard Business Books.

Bruner, J. 1986. Actual Minds, Possible Words. Cambridge, M.A. Harvard University Press.

Catino, M., & Patriotta, G. 2013. Learning from Errors: Cognition, Emotions and Safety Culture in the Italian Air Force. Organization Studies, 34(4), 437–467.

Creswell, J.W. 1998. Qualitative Inquiry and Research Design: Choosing Among Five Traditions. California: SAGE Publications, Inc.

Cosby, K.S., & Croskerry, P. (2004). Profiles in patient safety: authority gradients in medical error. Academic Emergency Medicine, 11(12), 1341–1345.

Dewey, J. 1997. Experiences and Education. New York. Touchstone.

Dewey, J. 1961. Democracy and Education. An introduction to the philosophy of education. New York: The Macmillan Company.

Firing, K. & Moldjord, C. 2007. Tanker, følelser og handlinger i samspill- et grunnlag for personlig vekst i operative profesjoner [Thoughts, feelings and actions in interaction- a basis for personal growth in operative professions]. In: Moldjord, C. Arntzen, A., Firing, K. Solberg, O.A. & Laberg, J.C. (Eds.) 2007. Liv og Lære i operative miljøer. "Tøffe menn gråter". ["Tough Men Cry"] Life and Experiences in Operational Environments] Fagbokforlaget, Bergen.

Firing, K. & Lien, D.O. 2007. Pedagogisk Grunnlagstenkning og Veiledning. [Educational Basis-Thinking and Counselling]. In: Firing, K., Hellemsvik, K., & Haarberg, J. (eds.) 2007. Kryssild. Militært Lederskap i en Ny Tid [Crossfire. Military Leadership in New Times]. Trondheim: Tapir Academic Press. Folland, R. 2009. Holistic Debriefing: A Paradigm Shift in Leadership. A Research Report Submitted to the Faculty in Partial Fulfillment of the Graduation Requirements. Air Command and Staff College Air University, Maxwell Air Force Base, Alabama, April 2009.

Greenwood, D.J. & Lewin, M. 2007. Introduction to Action Research for Social Change. Second Edition. Sage. Thousands Oaks.

Moldjord, C. & Fredriksen, P.K. 2017. Debriefing-strategisk læringsverktøy i operative organisasjoner. [Debriefing-strategic learning tools in operational Organization] In: Heier, T. (ed.) 2017. Kompetanse forvaltning I Forsvaret [Competence management in the Armed Forces]. Bergen: Fagbokforlaget.

Moldjord, C. (2016). Coping with stress in Military and Operational Professions. Holistic Debriefing and Development of Trust in High Performance Teams. Doctoral theses, NTNU 2016:56.

Moldjord, C. 2007. Prolog- Læring og vekst i operative organisasjoner. [Prolog- Learning and development in operational organizations]. In: Moldjord, C. Arntzen, A., Firing, K. Solberg, O.A. & Laberg, J.C. (Eds.) 2007. Liv og Lære i operative miljøer. Tøffe menn gråter". ["Tough Men Cry" Life and Experiences in Operational Environments] Bergen: Fagbokforlaget.

Patriotta, G., & Brown, A.D. 2011. Sensemaking, metaphors and performance evaluation. Scandinavian Journal of Management, 27, 34–43.

Pennebaker, J.W. 1997. Opening Up. The Healing Power of Expressing Emotions. New York: Guildford Press.

Piaget, J. 1977. The Development of Thought. Equilibration of Cognitive Structures. New York: Viking Press.

Reason, J. 1997. Managing the Risks of Organizational Accidents. Ashgate.

Ron, N., Lipshitz, R., & Popper, M. 2006. How organizations learn: Post-flight reviews in an F-16 fighter squadron. Organization Studies, 27, 1069–1089.

Rosness, R. 2001. Slank og sårbar. Om verdien av organisatorisk redundans. [Lean and Vulnerable. On the Value of Organizational Redundancy], Trondheim: SINTEF-rapport STF38 A, 1413.

Skinner, B.F. 1953. Science and Human Behavior. New York: MacMillan.

Skjevdal, J., Solheim, J.A., & Henriksen, R.E. 1995. Håndbok i lederskap for Luftforsvaret. [Handbook in Leadership for the Norwegian Air Force]. Oslo: Luftforsvarsstaben.

Steiro, T.J., Moldjord, C., Fredriksen, P.K. & Firing, K. 2010. Tough Men Cry. Learning From Sharp End Military Operations. Paper Presented At ESREL, 2010, Rhodes Island, Greece.

Steiro, T.J. & Firing, K. 2009. Coaching and Mentoring in Military Training: An Educational Perspective. I.: Karlsdottir, R. & Kvalsund, R. (red.). Mentoring

og coaching i et læringsperspektiv [Mentoring and Coaching in a Learning Perspective]. Trondheim: Tapir Akademisk Forlag.
Thagaard, T. 1998. Systematikk og Innlevelse. En innføring i kvalitative metoder. Bergen: Fagbokforlaget.
Vygotsky, L.S. 1978. Mind in Society. The Development of Higher Psychological Processes. Cambridge, Massachusetts: Harvard University Press.
Weick, K.E., & Sutcliffe, K.M. 2007. Managing the unexpected: resilient performance in an age of uncertainty. San Francisco, CA: Jossey-Bass.
Weick, K.E., Sutcliffe, K.M., & Obstfeld, D. 2005. Organizing and the process of sensemaking. Organization Science, 16, 409–421.
Weick, K.E. 1987. Organizational culture as a source of high reliability. California Management Review, 29, 112–127.
Yin, R. 1994. The Case Study Anthology. Thousands Oaks, Sage Publications Inc.
Zhao, B. 2011. Learning from errors: The role of context, emotion, and personality. Journal of Organizational Behavior, 32, 435–463.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Applying elements of the STAMP method to the reorganization of the German nuclear waste management

Heinz-Peter Berg
Braunschweig, Deutschland, Germany

Stephan Griebel & Birgit Milius
Siemens AG Braunschweig, Deutschland, Germany

ABSTRACT: Over the last years, the structure of the legal entities to govern all issues of the German nuclear industries has changed significantly. What used to be just one company is now divided into four entities between which responsibilities were shared and new tasks allocated. For this actual example of comprehensive structural changes, in particular regarding conditioning, handling and interim/final storage of radioactive waste in Germany, we apply the STAMP (System-Theoretic Accident Model and Processes), approach in order to demonstrate how the STAMP structure can be used to monitor and master this change. The choice of the industry area of nuclear is solely motivated by scientific interest and excludes commercial aspects. Siemens is not active in the "nuclear waste" management industry.

1 MOTIVATION

Reorganisation is always a very complex task in any undertaking. When dividing existing organisations and establishing new ones with new stakeholders, formal and informal processes and tasks have to be considered and adapted where necessary. This becomes even more important when safety-critical aspects are involved. As a safety-critical area we will be focusing in this article on the Nuclear Waste Disposal in Germany. A very clear and structured approach is helpful to keep track of the relevant changes, but also to visualize those changes in a clear manner. The last aspect especially helps to inform all concerned parties and can be used as a starting point for the discussion and verification of the model.

The described difficulties can be modeled as a control problem. Instead of developing a completely new theory it seemed sensible to use an existing theory. The advantage of using an existing model framework is that a certain set of conditions exists on which the ongoing work can be based on. In our paper (Berg, Griebel, Milius 2017) we have shown that STAMP, a theory developed by Nancy Leveson, can be used to model organizations on a high level. We applied some elements of STAMP to the current situation in the German nuclear industry. We modeled the current organizations and the existing interactions and discussed some fictional examples. In this paper, we will elaborate the example further, detailing a certain part of the existing STAMP structure for the German nuclear industry, i.e. by clearly visualizing the structure and the processes using elements of STAMP we gain further insight

The paper is structured as follows:

In the first chapter, we explain the area of application that is the expected changes to the process of nuclear waste disposal. We will address the current situation, the planned changes the expected challenges. Afterwards, we will give a short overview of the basic concept of STAMP and explain our chosen modeling approach. In the third chapter, we will present today's situation in nuclear waste disposal as well as the future situation using the visualizing elements of STAMP. We will compare both structures and discuss interesting aspects.

2 NUCLEAR WASTE DISPOSAL IN GERMANY

2.1 *General*

In general, most radioactive waste in Germany comes from nuclear electricity production. However, it is also generated in hospitals from the use of radioactive material to diagnose and treat the sick and sterilize medical products, in the production of radiopharmaceuticals, at universities in conducting vital research in biology, chemistry and engineering.

These wastes must be safely managed at all stages prior to and including final safe disposal. Storage is an integral part of the waste management process.

In Germany, disposal in deep geological formations is intended for all types of radioactive waste. However, there are strong debates concerning various nuclear waste disposal options including direct deep geological disposal with and without waste retrievability, long-term interim storage, as well as new conditioning and treatment concepts such as partitioning and transmutation. Scientific research to provide input to the decision-making process is essential, as this is the only guarantee for reliable data and models required for any sound, scientifically-based safety assessment.

Decentralised storage facilities for spent fuel were licensed under nuclear law and constructed and commissioned at twelve sites with nuclear power plants. They are designed as dry storage facilities in which transport and storage casks loaded with spent fuel are emplaced. Storage casks for spent fuel in the three central storage facilities have the same properties as the storage casks in on-site storage facilities.

By means of conditioning of the radioactive waste, intermediate or final products shall be produced which fulfill the requirements on safe handling, storage and transport also for the period of extended interim storage. The radioactive waste is to be safely stored until it can be delivered to a facility of the Federation for disposal.

Currently, there are still eight operational nuclear power plants left in Germany. For one further nuclear power plant the authorisation for power operation expired on 31 December 2017 in accordance with the German Atomic Energy Act (Federal Law Gazette 2017c).

Moreover, there are currently three research reactors, three reactors for training purposes and one reactor for educational purposes in operation.

However, many nuclear facilities are already in the process of decommissioning as listed below:

- seventeen nuclear power plants in the process of decommissioning as at 30 April 2017,
- nine research reactors with an electric power of more than 1 MW permanently shut down, in the process of decommissioning, or decommissioning completed and released from nuclear regulatory control, as at 30 April 2017,
- twenty-nine research reactors with an electric power of less than 1 MW permanently shut down, in the process of decommissioning, or decommissioning completed and released from nuclear regulatory control, as at 30 April 2017,
- eight experimental and demonstration reactors in the process of decommissioning, or decommissioning completed and released from nuclear regulatory control, as at 30 April 2017,
- six commercial fuel cycle facilities in the process of decommissioning or decommissioning completed and released from nuclear regulatory control, as at 30 April 2017.

This underlines the necessity of urgent solutions.

2.2 *Explanation of current structure*

The Federal Ministry for Environment, Nature Conservation, Building and Nuclear Safety (BMUB) is the nuclear regulatory authority of the Federation. As defined in Article 73 of the Basic Law (Federal Law Gazette 2014), the Federation shall have exclusive legislative power with respect to "the production and utilisation of nuclear energy for peaceful purposes, the construction and operation of facilities serving such purposes, protection against hazards arising from the release of nuclear energy or from ionising radiation, and the disposal of radioactive substances".

The "Gesellschaft für Anlagen—und Reaktorsicherheit (GRS)" carries out research and analysis in its fields of competence, namely reactor safety and radioactive waste management, and supports the BMUB on technical issues.

The Nuclear Waste Management Commission (ESK) advises the BMUB in nuclear waste management issues (conditioning, interim storage and transports of radioactive material and waste, decommissioning and dismantling of nuclear installations, disposal in deep geological formations).

The Federal States carry out their tasks under nuclear law on behalf of the Federation (federal executive administration). Federal supervision extends to the legality and appropriateness of execution by the Land authorities. According to Article 85(3) of the Basic Law, these shall be subject to the instructions from the competent highest federal authority (BMUB).

In performing their activities, the Federal State authorities may consult technical expert organisations or individual experts according to § 20 of the Atomic Energy Act (Federal Law Gazette 2017c). Today, this is mainly ensured by the technical expert organisation (TÜV) for specific issues. With the involvement of experts, an examination on the safety-related issues is made which is independent of that of the applicant.

The Federal State Authority for Mining, Energy and Geology is the mining authority and is granting mining legal licenses for nuclear waste disposal in deep geological formations.

The Federal Office for Radiation Protection (BfS) is a subordinate authority of the BMUB in the area of radiation protection and nuclear safety is the BfS. The four technical departments of the BfS deal with the statutory tasks in the areas of environmental and industrial radiation protection, radiation biology, radiation medicine, nuclear fuel supply and waste management and nuclear safety. The BfS supports

the BMUB technically and scientifically, especially in the execution of supervision of legality and expediency, the preparation of legal and administrative procedures, and in intergovernmental cooperation. Moreover, the BfS is license holder for nuclear waste disposal facilities. Furthermore, the nuclear supervisor, for example responsible for monitoring compliance with waste acceptance requirements for the respective final disposal facility. These requirements result from the safety-analytical investigations that need to be complied with when waste packages to be disposed of will be delivered to the repository in future, is also the BfS by a self-monitoring section—directly tied to the vice president Of BfS.

The German Company for the Construction and Operation of Repositories (DBE) has been exploring, constructing and operating the German repository projects and mines in Gorleben, Morsleben and Salzgitter (Schacht Konrad) since 1979. After a legally enforceable plan approval decision has been in force since 2007, the Federation has decided to set up the Konrad mine as a repository. DBE is responsible for the comprehensive construction work.

The Asse GmbH is an operating company and as such responsible for all operational work in the Asse mine. Shareholder is 100% federal. The Asse GmbH implements the measures ordered by the BfS. The Asse-GmbH will also be responsible for the implementation of the decommissioning work and until then ensure operation in accordance with the requirements of nuclear legislation.

The Company for Nuclear Service (GNS) carries out services in the field of radioactive waste disposal and decommissioning of nuclear facilities and operates through several subsidiaries interim storage depots for spent fuel and radioactive waste. Further activities are related to the development of conditioning methods, including the development and qualification of the cask and emplacement systems, processing of the waste, loading of the casks, documentation of the containers, and control of delivery to the Konrad repository. GNS owns 75% of the DBE.

2.3 *Explanation of future structure*

On January 1, 2014, the Federal Act on the establishment of a Federal Office for Nuclear Waste Management (BfE) entered into force, so that BfE was formally founded on this day. However, the Federal Office for Radiation Protection (BfS) and its tasks remained unaffected at this time. By the Act on the Reorganization of the Organizational Structure in the Field of Final Disposal, BfE was renamed Federal Office for Nuclear Safety in Waste Management as of July 30, 2016 (Federal Gazette 2016a). The main reason for the renaming was the intention to differentiate the BfE more clearly against the Federal Company for Radioactive Waste Disposal (BGE).

Against this background, the BfS was divided in-to three organizations.

The Federal Office for Radiation Protection (BfS) is now responsible for the safety and protection of man and the environment against damage caused by ionizing and non-ionizing radiation (Federal Law Gazette 2016b). In the area of ionizing radiation, for example, X-ray diagnostics in medicine, safety in the handling of radioactive substances in nuclear technology, and the protection against increased natural radioactivity. Non-ionizing radiation workplaces include, inter alia, protection against ultraviolet radiation and the effects of mobile phones.

The Federal Office for the Safety of Nuclear Waste Management (BfE) advises the BMUB on nuclear waste disposal and safety issues for other nuclear facilities. In addition, BfE is bearer of the public participation in the process of determining the location in Germany for final disposal of high level radioactive waste. BfE has the task to check the safety of the disposal at all process steps. Thus, for the first time in Germany, BfE is now an independent regulatory, licensing and supervisory authority for the transport of radioactive waste and its interim and permanent storage.

Transitional provisions apply to the Konrad repository and the Morsleben repository for radioactive waste (ERAM), according to which the Federal States remain responsible for licensing until this responsibility is transferred to the BfE with the granting of the approval of commissioning by the nuclear supervisory authority for the Konrad repository or until the plan approval decision on decommissioning will be enforceable for the ERAM. A part of the workforce which is still planned for the BfE, mainly from the administrative sector, will still be involved in BfS by December 2017.

Furthermore, two additional Acts were recently set in force (Federal Gazette 2017 a and 2017b), partially in extension of the Act on Search and Selection of a Site for a Repository for Heat-Generating Radioactive Waste and for Amending Other Laws (Federal Law Gazette 2015).

The Federal Company for Radioactive Waste Disposal (BGE) is a state-owned company in which several tasks have been merged: operational tasks of site selection, construction and the operation of the repositories for high level nuclear waste as well as, e.g., for the Asse II salt mine site where in the past low and partially medium level radioactive waste has been stored. Moreover, the "Act for the Reorganization of Responsibility in Nuclear Waste Disposal", which came into effect in June 2017, has been redefined regarding the responsibility for the decommissioning and dismantling

of nuclear power plants and for the the interim storage of radioactive waste (Federal Law Gazette 2017b). Some sections of the BfS have been transferred to the BGE. The Asse GmbH and the DBE will also be incorporated in due time.

Moreover, the Company for Interim Storage (BGZ) has become in August 2017 a company of the Federal Government. From now on, the existing two central interim storage facilities will be part of the business of the BGZ. At the beginning of 2019, the twelve decentralized temporary storage facilities at the nuclear power plants—currently operated by the electric power utilities—will also become the responsibility of the BGZ, one year later additionally the twelve storage facilities with low- and medium radioactive waste resulting from the operation and dismantling of the German nuclear power plants.

3 WHAT IS STAMP?

STAMP is an acronym and stands for System-Theoretic Accident Model and Processes. The theory was developed by Nancy Leveson. The necessity for the model has arisen when it became clear that the traditional mechanisms to explain incidents were no longer valid. Typically, incidents were explained as linear chain of events, with a failure of one system at the beginning of the chain. However, due to the complexity of today's systems this does not need to be true. Every part of a complex system can work perfectly as specified but under certain preconditions it might lead nevertheless to an incident. For more examples of the general idea of this model, you can look at (Leveson and Stephanopoulos 2014). Leveson argues that in the future incidents should be considered as control problems. Actors and interactions need to be defined in a strict manner.

A typical STAMP control structure is shown in Figure 1. An example for a more complex situation can be found, e.g. in our paper (Berg, Griebel & Milius 2017).

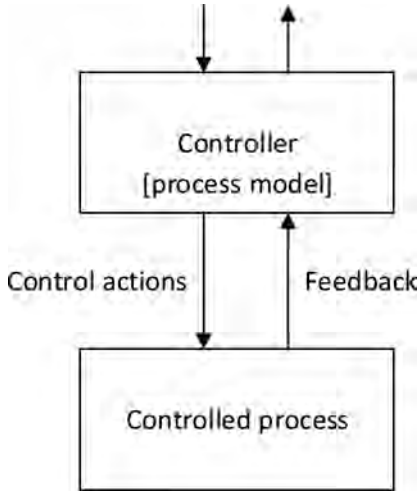


Figure 1. General description of the idea behind STAMP.

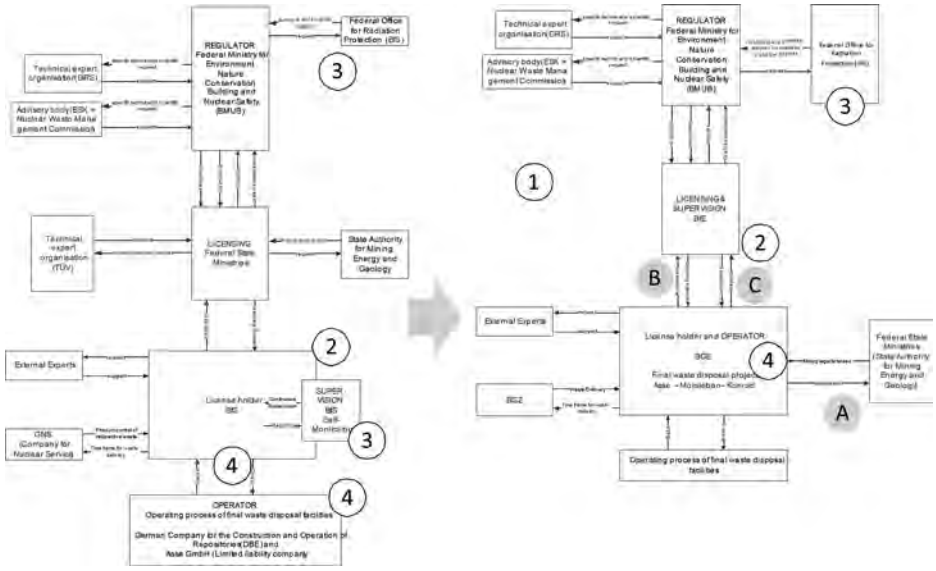


Figure 2. Comparison of the current (right) and future (left) STAMP structures; numbers indicated areas of change (1, 2, 3 and 4) discussed in the text; letters indicate the discussed control loops.

STAMP is the basis for different applications. (Thomas 2013) distinguishes CAST (Causal Analysis using System Theory) and STPA (System-Theoretic Process Analysis). For our work, STPA is the relevant component as it helps to identify the potentially hazardous control structures which can lead to failures in the process.

To identify controls which are potentially hazardous, four types can be distinguished:

- Control commands required for safety are not given
- Unsafe ones are given
- Potentially safe commands but given too early, too late
- Control action stops too soon or applied too long

The paper will not allow applying the issue of the controls in detail; however, we will highlight some examples where even on first look problems become obvious.

4 COMPARING AND ANALYSING THE TRANSFORMATION PROCESS

Using elements of STAMP the structure and control processes of the current and future situation are shown in Figures 3 and 4 at the end of this paper. Looking at how the elements of STAMP

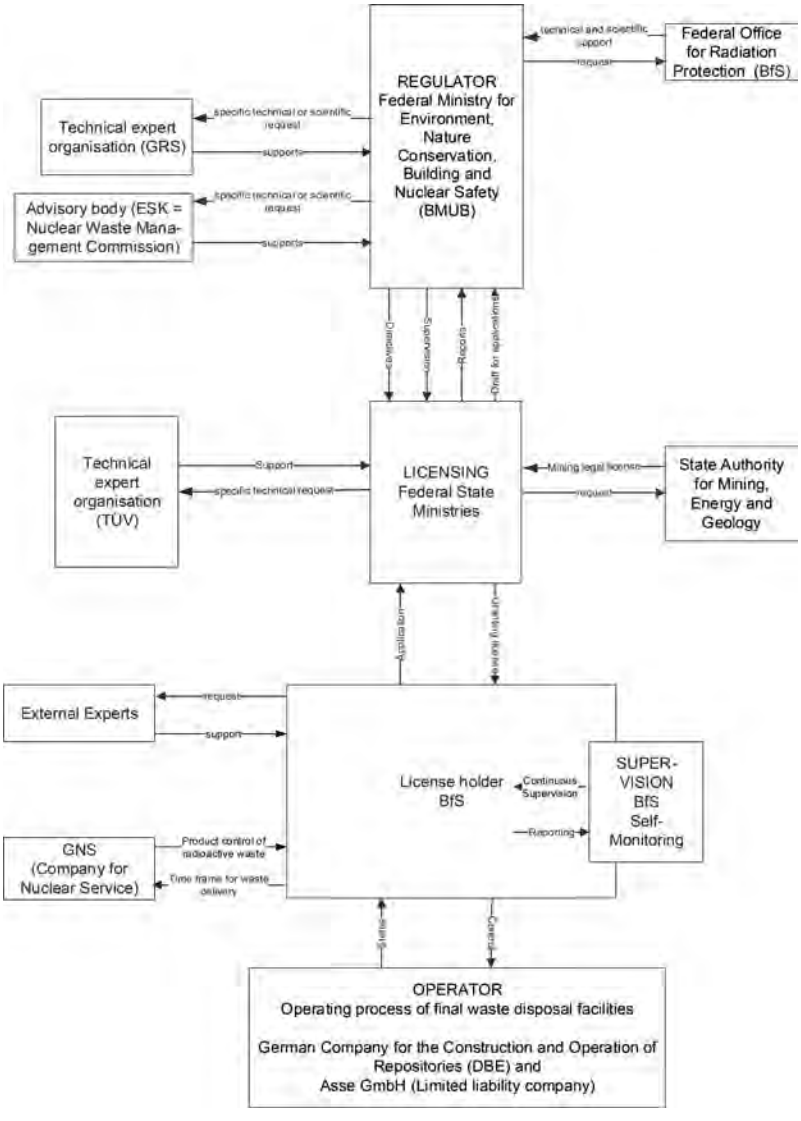


Figure 3. Current structure of German nuclear waste disposal process.

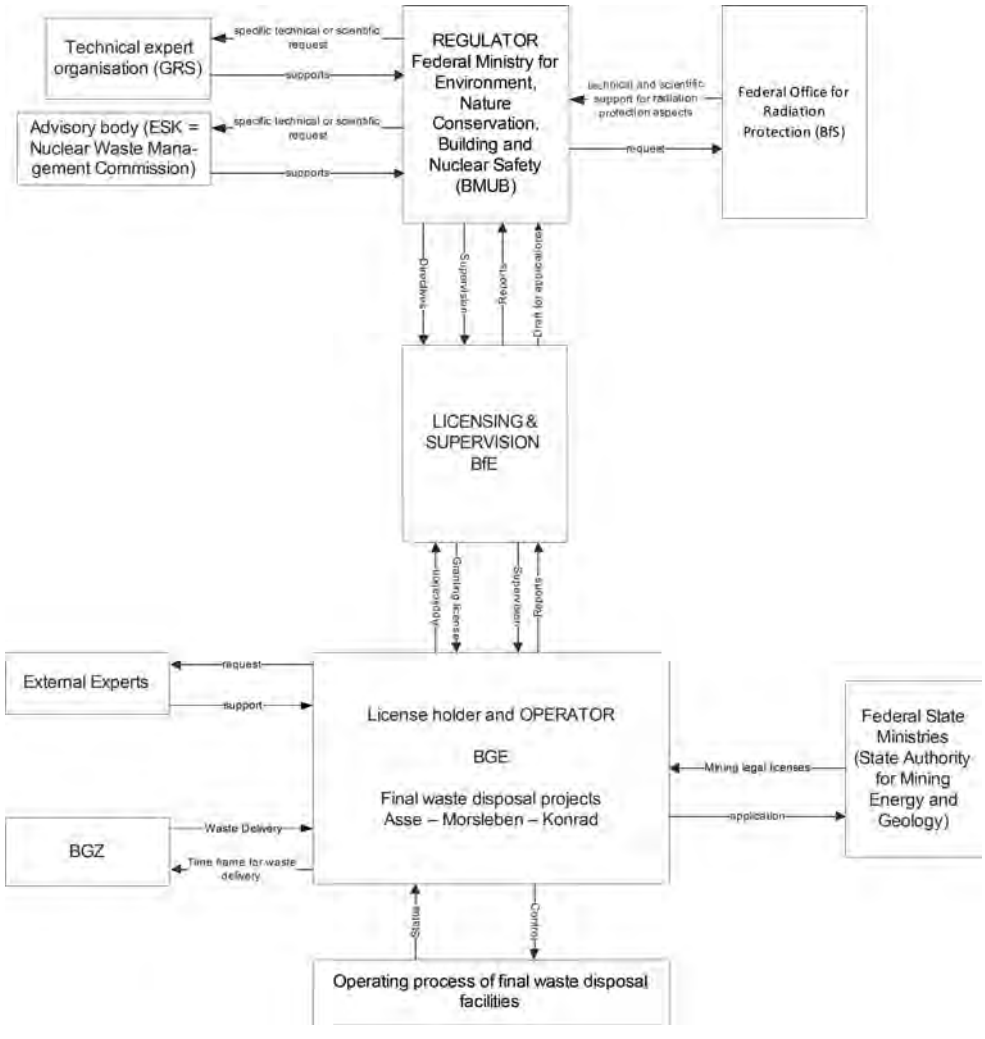


Figure 4. Future structure of German nuclear waste disposal process.

can contribute to a better understanding and evaluation of the change in the processes, the following aspects can be identified by looking at the structure (Figure 2).

1. The new process is more streamlined with a strong focus on the licensing and supervisory authority on the one hand and the license holder and operator on the other one.
2. The BfS as a self-monitoring organization of final waste repositories is replaced BfE which is not any more license holder and supervisory authority. Moreover, the new picture shows a clearer structure because in future all licensing and supervisory activities were performed by BfE whereas in past the Federal States were licensing authority and BfS operator and supervisory organization. The new legal basis also concentrates the responsibilities for licensing (in the past BfS) and supervision (in the past Federal States) of interim storage waste facilities not discussed in this paper in detail.
3. The general structure regarding the involvement of the BfS is now much clearer. Whereas in the old structure the BfS was having several, very different functions, this is now clarified and more focused on safety and protection of man and the environment against damage caused by ionizing and non-ionizing radiation.
4. The license holder and the operator were used to be separated from each other. This led to an additional need for control. Now they are subsumed under one roof. This would be a good

example to look deeper in the structure to see, if all aspects of the external control structure are now integrated in the internal processes of the license holder and operator.

Overall, all tasks of nuclear engineering disposal after dismantling have now been fully distributed to authorities (BfE) and federal private companies (BGE).

A central prerequisite for the credibility of the now responsible actors has been created by the reorganization of the responsibilities which leads to a clear functional description of the company's ownership as well as the operation in a private-law company on the one hand and the supervision and intensification of the public participation on the other hand.

Regarding the site selection of a repository for highly radioactive waste, BGE has improved structural preconditions for a rapid and well-founded presentation of results by the concentration of the project owner in the site search as well as all operator tasks.

BfE has the task of supervising the implementation of the site selection procedure in accordance with § 19 (1) to (4) of the Atomic Energy Act (Federal Law Gazette 2017c). It, therefore, supports the entire procedure from a scientific point of view and is the competent authority for the monitoring of the enforcement of the site selection procedure at all stages of the procedure (Federal Law Gazette 2017a).

5 ANALYSIS OF CONTROL LOOPS

The structured depiction of the control processes enables the identification of possible gaps in the control and information loops.

One example of a rather simple and clear loop consists of the requesting and granting of mining licenses (A). If we would apply here the different possibilities of hazardous controls, we could, e.g. discuss in detail what happens when a request is done at a wrong time or an application is granted to early. STAMP/STPA can help us here to better understand the process and guide the user to potentially hazardous situations which then can be counteracted by e.g. processes.

As for the area of the interdependence between the BfE and the BGE the situation is rather more complex: The loop regarding the application of license changes and the granting of revised licence seems a closed one (B). Yet when it comes to the issue of supervision by the BfE and providing reports by the BGE the causal relation would need more detailed analysis (C). One example for this fact is that supervision also contains inspections in the waste disposal facility, i.e. not only paper work is involved. This is different from the other reports&supervision loops in the structure, where only documents are transferred and checked.

This succinct analysis already shows that is important to clearly and comparably construct and describe the elements of the control loops. Without this structured analysis one might face the danger of leaving out pertinent aspects in the process of transforming and streamlining existing control and supervision structures.

6 CONCLUSION

In this paper we have applied some of the elements of STAMP to the reorganization of the German nuclear waste management of the German Nuclear Industry. We focused on the process of final waste disposal.

By structuring the process we have made a decisive leap towards

- Transparency (by providing a method of displaying lucidly the complex process)
- Elucidation of areas for improval
- Enabling checks if the yearned objectives have been attained.

Thus we have showed that even without going into much detail of the daily work, some areas of change became obvious and can be used as basis for further discussions. The identified changes can now be analyzed further to make sure that the responsibilities of changed ownership (partially spread and partially concentrated compared to the earlier structure) are taken care of in the new processes.

Furthermore it could be checked in future analysis, which effect envisaged changes would have on the overall safety of the processes and the site by highlighting the risks

REFERENCES

Act on Search and Selection of a Site for a Repository for Heat-Generating Radioactive Waste and for Amending Other Laws (Site Selection Act – StandAG) of 23 July 2013 (Federal Law Gazette I 2013, p. 2553–2564), last amended by Articel 309 on 31 August 2015, Federal Law Gazette I 2015, p. 1474.

Act on the Establishment of a Federal Office for Nuclear Waste Management of 23 July 2013, Federal Law Gazette I, pp. 2553, 2563), amended by Article 4 of the Act of 26 July 2016, Federal Law Gazette I 2016, p. 1843 (Federal Law Gazette 2016a).

Act on the Establishment of a Federal Office for Radiation Protection of 9 October 1989, last amended on 26 July 2016, Federal Law Gazette I 2016, p. 1845 (Federal Law Gazette 2016b).

Act for further Development of the Act on Search and Selection of a Site for a Repository for Heat-Generating Radioactive Waste and for Amending Other Laws of 5 May 2017, Federal Law Gazette I 2017, p. 1074–1102 (Federal Law Gazette 2017a).

Act on the Reorganisation of Responsibility in Nuclear Waste Disposal of 27 January 2017, Federal Law Gazette I 2017, p. 114–129. entering into force on 16 June 2017 (Federal Law Gazette 2017b).

Act on the Peaceful Utilisation of Atomic Energy and the Protection against its Hazards (Atomic Energy Act) of 23 December 1959, as amended and promulgated on 15 July 1985, last Amendment of 20 July 2017, Federal Law Gazette I 2017, p. 2808–2838 (Federal Law Gazette 2017c).

Basic Law for the Federal Republic of Germany including the amendment(s) to the Act by Article 1 of the Act of 23.12.2014, Federal Law Gazette I 2014, p. 2438.

Berg, H.-P., Griebel, S. and Milius, B., 2017. Managing change of safety-critical infrastructure via STAMP. Safety and Reliability – Theory and Applications, edited by Marko Čepin and Radim Briš, CRC Press Taylor & Francis Group.

Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (BMUB), Report by the Government of the Federal Republic of Germany for the Sixth Review Meeting of the Joint Convention on the Safety of Spent Fuel Management and on the Safety of Radioactive Waste Management in May 2018, Bonn, Germany, August 2017.

Leveson, N.G. and Stephanopoulos, G., 2014. A systemtheoretic, control-inspired view and approach to process safety. AiChE Journal Volume 60, Issue 1, January 2014.

Thomas, J., 2013. Basic STPA Tutorial. http://psas.scripts.mit.edu/home/wp-content/uploads/2013/04/Basic_STPA_Tutorial1.pdf.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Tourism industry facing crises: Setting the scene

C. Martin
Tourism Institute, Cannes, France
Mines ParisTech, PSL Paris University, France

F. Guarnieri
Mines ParisTech, PSL Paris University, France

F. Lamm
Auckland University of Technology, New Zealand

ABSTRACT: As the world's first tourist destination in 2015, with a target of 100 million visitors in 2030, the French tourism sector, is a national priority and a strategic challenge for the country. However, the dramatic events of 13 November 2015 (in Paris) and those of 14 July 2016 (in Nice) led to a reversal of the trend, and a substantial and worrying decline in activity. The events raise two key questions in terms of crisis management: The first is linked to the ability of regions and organizations to recover, as quickly as possible, their capacity to meet the needs of the public. The second, which is the direct consequence of the first, is the capacity of regions and tourism industries to re-establish their reputation, notwithstanding the fear that is generated by the threat

This paper presents a review of the management of major risks and the tourism sector at the international level and in France. It presents an ongoing qualitative study conducted in partnership with an association of event professionals on the management of crises on the French Riviera. The roles of the different actors in a crisis are studied from the perspective of organizational resilience.

1 INTRODUCTION

Whether resulting from natural or industrial causes, or due to terrorism, disasters and crises take a heavy toll on tourist activities. In a country like France, where tourism accounts for 7% of GDP, and represents 2 million direct and indirect jobs, a crisis can threatens the sector. Although in different ways and in different places, the terrorist attacks in Paris in 2015, and the attacks in Nice in July 2016 badly affected the sector's economy,

While crisis management is well-understood in conventional industrial activities (nuclear, oil and gas, chemistry, aeronautic…), it is clear that the academic literature contains few references to the tourism sector. Reasons for this oversight is that the sector is diverse, geographically dispersed with a high proportion of small businesses, thus making it a challenging topic to investigate. Having a good understanding the typology of crises and the consequences for the sector is, however, essential in order to prepare the resources that are required to recover the loss of patronage and business.

This paper reports an ongoing, longitudinal French study located in the tourism sector that identifies the key elements needed to successfully manage a crisis after a terrorist attack. Applying a systemic approach, the aim of the study is to increase our understanding of why and how regions and organizations become resilient after an acute crisis. We also argue that the findings of this study could inform the way in which other crises are managed. It should be noted that the study concentrates on crises associated with acts of terrorism. We begin, however, with presenting research on the perception of risk and tourism-related crisis management. We then introduce the concept of resilience applied to crisis management, and the leisure and business tourism sectors.

2 GLOBAL TOURISM, RISK PERCEPTION AND IMPACTS ON THE DESTINATION

The globalization of tourism has created a highly competitive environment. As a result, both the leisure and business tourism sectors are subject to fierce competition within and between destinations. Although France is one of the most popular destinations, aiming to welcome 120 million tourists in 2020, (Huchon, 2016), it has to compete with other destinations. Studies of terrorist crises

(2001 in New York, Hua hin and Phuket 2016 etc.) also show that there is a negative economic impact on destinations that have experienced a terrorist attack, (Enders, 1992, Fainstein, 2002, Beirman, 2003, Drakos, 2003, de Sausmarez, 2003). In this context, knowing the reasons for selecting a particular destination, (ie the level of risk associated with the chosen destination) is necessary.

Moreover, a number of studies highlight the fact that the level and perception of risks are key dimensions in the decision-making process. In a study of 290 young adults born in the United States, Lepp & Gibson (2003) report that there are seven risk factors associated with tourism: health, political instability, terrorism, unusual food, cultural barriers, political and religious dogmas and crime. However, the study found that men and women differ: women care more about health risks. A second difference is found in the type of destination: some look for a familiar destination, while others seek out novelty. The latter group are less sensitive to risk.

With regard to natural and technological risks, Chew & Jahari (2014) analyze the impact of per ceived risk on the image of a destination. Taking the case of the Fukushima disaster in Japan in 2011, where a natural disaster that combined a tsunami and earthquake caused a nuclear accident, the authors analyze the relationship between the image of the destination, perceived risk and the intention of returning to the destination. Their findings show that tourists' perception of risk is an important element in the destination's image and plays a key role in the choice of destination. In practice, although increased risk may play an inhibiting role in the choice of and return to a destination, the authors note that recent studies highlight repeat purchases despite a rise in the level of risk or a recent disaster. Acting on the image of the destination would therefore appear to help to mitigate the effect of perceived risk on the intention to return.

In sum, risk perception is a component of the decision-making process in the visit and revisit of a destination. The perceived risk associated with the destination, therefore, can play a role in mitigating the image of a destination.

3 CRISIS MANAGEMENT AND TOURISM

In order to understand risks and resilient organizations it is necessary to examine the literature on crisis management, with particular reference to the tourist sector. Aktas & Gunlu (2005) note that, although the concept of crisis management has existed as a research field since 1970, studies located in the tourism sector only began to emerge in the 1990s.

It is useful, however, to first distinguish between "disaster" and "crisis". Faulkner (2001) notes that the difference between the two is that a disaster is an external phenomena while a crisis is an internal phenomena and relates to the organization itself and management failures. Aktas &Gunlu (2005), taking Faulkner's example (2001), show that the Izmit earthquake in 1999 could be likened to a catastrophe on account of its unpredictability and external nature. On the other hand, the lack of respect for earthquake-resistant standards reflects organizational failures and, therefore, a crisis.

Atkas & Gunlu (2005) further explain that in practice, the unpredictable dimension of an unforeseen external event can degenerate into a if there are management failures. In this case, the disaster and the crisis can be subsumed. Here both share three elements: a trigger, damage, and a threat to the life or property.

Further, crises can be linked to events at different levels. At the micro level, a crisis can be limited to the organization (an industrial relations strike or epidemic that affects the workforce). Crises can also occur at the international level, moving from a micro to a macro level, for example a terrorist attack or a pandemic. Moreover, a crisis can be extend over many years (Ritchie, 2004, Ritchie & Campiranon, 2014). In this context, Both Ritchie (2004) and Laws and Prideaux (2006) argue that the complexity of crises suggests that there is a need to establish a typology in order to provide appropriate strategic management plans. Ritchie (2004) states that crisis management should be based on appropriate plans, trained staff, communication plans and established relations with the media. That is, crisis management is based on a capacity to mobilize resources and communicate at the right time. It also requires appropriate lobbying of decision-making bodies to rapidly allocate the funds that are needed to meet the needs of crisis management activities.

The ability to manage a crisis therefore depends on a number of factors that can prepare managers to cope, and also to communicate at the right time to contain the crisis and avoid changing the image of the destination. In addition, a significant part of crisis management is based on the creative ability of actors and in particular, it relies upon the support of actors, or even drawing from the group solutions that could not have been imagined. This observation is echoed by Atkas and Gundu (2005), who consider the implementation of strategic crisis management and post-crisis marketing plans are necessary for the revival of the destination. Finally, proactive and holistic management as well as an evaluation of each crisis and its management are essential to ensure that visitors return (Laws & Prideaux, 2006, Ritchie & Campiranon, 2014).

While a great deal of this discourse is located in the disciplines of marketing and economics, studies in other disciplines, such as resilience engineering and emergency engineering have provide another perspective.

4 RISK AND RESPONSE

There have been a number of scholars who have made a significant contribution to the research on risk and resilience. One of the most notable is Hollnagel (2007) who only studied the concept and the attributes of risk but also introduced the concept of *resilience engineering*. He defines resilience as a system's ability to adapt before, during or after changes or perturbations in order to continue a set of operations that are identified in expected or unexpected conditions (Hollnagel, et al. 2006). This key contribution to our understanding of crisis management is part of a rich field of research on the human and organizational factors. Others scholars have also made a contribution. For example, Perrow's (1981) work on complex organizations and risk in which he examines the organizational factors that underlie systemic and catastrophic accidents. His central argument is that accidents are normal in complex, socio-technical systems because they are by their very nature, susceptible to an elevated level risk. His work initiated critical thinking about industrial systems that are exclusively controlled by managed safety procedures.

Similarly, the work of the High Reliability Organization group at Berkeley University (see La Porte, 1996) have examined the resistance criteria in complex socio-technical systems. What brought the Berkeley colleagues together was their shared observation from three different disciplinary perspectives that the attention being paid to studies and cases of organizational failure was not matched by parallel studies of organizations that were operating safely and reliably in similar circumstances (Rochlin, 1996). They found that the role of technology and structure was not fixed, but varied from task to task. The organization had not one, but many overlapping cultures, tied together by a common purpose. Its performance was shaped not only by the skill and dedication of the operators, but also by intelligent and sensitive management. Weick (1995) also observed that in the event of a crisis, following procedures without begin able to innovate or rethink procedures can hamper action, or even end in a catastrophic scenario.

Taking the analysis to the next level, Hollnagel (2007) argues that it is critical for an organization to anticipate and learn from previous crises. Moreover, the resilience of organizations is linked to the ability of actors to adapt to changing conditions. Organizational resilience, according to Hollnagel (2007), is based on four central pillars: the ability of a system to respond to normal and abnormal conditions; its ability to control threats and performance in the short term; its ability to anticipate threats and opportunities in the long term; and its capacity to learn from the positives and negatives of past events.

Research on the management of major crises, including those in highly-regulated sectors, shows that comprehensive planning together with innovative solutions are necessary to deal with a crisis that is unexpected and has far-reaching, long-lasting consequences (Vogus & Sutcliffe, 2007; Burnard & Bhamra, 2011). The analysis of the Fukushima Daichi crisis is an example of the innovative capacity of one man and his team when faced with an unprecedented crisis that went far beyond any previously planned action (Funabashi & Kitazawa, 2012). Travadel, Martin and Guarnieri (2017) also highlight the need for this capacity. They introduce the notion of acting into extreme situation as a way of effectively responding to a long-lasting, hitherto unknown crisis.

In addition, Ritchie (2004) recommends scenario-based planning as a foundation for an appropriate response to a crisis. He notes that there is value in identifying and planning responses to a crisis but given what we know about complex systems, there will still be a need for some flexibility in order to support innovative responses.

Finally, it is crucial to not only link the capacity of government with sectoral institutions in order to generate a joint action and commitment but also necessary to understand the role of the various actors and their capacity to respond to crisis conditions (Glaesser, 2006).

With this in mind, the following section presents the main features of a longitudinal study looking at the response to the recent crises in the French tourism sector.

5 THREE TIERED RESPONSE SYSTEM

Over the past two years France has experienced several major terrorist crises that have had a significant impact on tourism activity. While the French government had predicted a significant growth in the tourist sector, the terrorist attacks of 12th November, 2015 brought the predicted growth in tourist numbers to a halt, and put a great deal of economic pressure on all the stakeholders in the sector. Paris was particularly affected. The *Huchon Report* (2016) (Governmental Report on Tourism) notes that in the days following the attacks, the various businesses operating in the sector suffered

significant reductions in visitor numbers and turnover. The report also noted that commercial activities in central Paris were particularly badly affected.

Despite efforts to revive tourist numbers, the end of 2015 and the whole of 2016 was still marred by more terrorist attacks and natural disasters, all of which continued to tarnish the image of France and in particular Paris and the Côte d'Azur, as a tourist destination. The terrorist attack of July 2016 on the *Promenade des Anglais* in Nice was one of the worst as it occurred in the middle of the summer season and targeted the second most popular tourist destination after Paris: the French Riviera.

Faced with the loss of its foreign clientele and a decline in hotel bookings, the sector asked the government to take measures to revive activity, similar to the actions that were taken after the 11th September terrorist attacks in New York. The *Huchon Report* (2016) argued that the response must be organized around a few, key measures. The first was to set up, at inter-government level, a *crisis unit* to revive activity by reassuring the public about the safety of the destination. The second was to provide *economic support* for the sector. The third was to coordinate crisis *communication and recovery actions*, with special attention on Paris as the country's flagship destination.

Based on the *Huchon Report's* recommendations, a three tiered, response system was put in place. At the national level, a recovery plan was developed focusing on coordinated action involving key stakeholders in both the public and private sectors with the aim of reviving tourist numbers in Paris and other key tourist destinations. In the South of France, the Regional Tourism Committee also established a recovery plan based on a communication strategy located centrally on the Côte d'Azur brand and uses existing well-established social networks. At the sector level, a communication plan for tourism professionals was created to reassure providers

As a result of these concerted efforts of the combined private and public stakeholders at the different levels, tourist numbers began to recover. In the French Riviera, one year after the attack of 14 July 2016, figures showed that foreign visitors were returning. Following a 10% fall in hotel bookings in the last quarter of 2016, figures for the summer of 2017 indicate that activity is recovering.

However, the tourist sector in the Côte d'Azur not only had to respond to the terrorist attacks, it also had to respond to climatic disasters. In particular, the storms of October 2015 badly affected the destination in which there was a heavy toll both human (20 people lost their lives) and economic (damage to tourist infrastructure). As the region tried to cope with the aftermath, those working tourist sector had to prepare for the arrival of 13,000 people attending a large international conference. These efforts by all the parties made the conference a success and were emblematic of the region's ability to respond to crisis and its resilience.

6 LONGITUDINAL STUDY

The aim of this qualitative study is to examine the tourism sector's processes and structures created in response to a series of crises in the Côte d'Azur. In particular, the study aims to understand the dynamics underlying the preand post-crises processes and structures. The focus of this two year study is on Provence Côte d'Azur Event Center, which represents 160 event management companies in the region (a total of 1.2 billion euros turnover).

Although Côte d'Azur Crisis Unit provided an initial impetus, the successful recovery of the tourist sector is highly dependent on the ability of key actors reassemble the affected areas of the sector during and after the crisis.

In practice, actors' ability to adjust to contextual conditions deserves to be studied in order to understand its modalities and reproducibility in other regions. The current study thus addresses, on the one hand, the organization of the actors during the management of the crisis. Was this organization, despite procedures, the result of management led by the institutional crisis unit, or was it an emerging organization of stakeholders and, possibly, the fruit of discussions and initiatives taken jointly by professional organizations, economic actors and institutional actors?

On the other hand, it addresses the innovative capacities of actors to respond, on a case-by-case basis, to the technical problems generated by the floods (loss of means of communication, transport, etc.). Are these capacities for innovation and adjustment part of a culture that is specific to the region, its actors and the experience of professionals?

The qualitative study will run for two years, and will involve professionals and actors that were involved in the crisis.

The plan is conduct 40 open-ended interviews with the various socio-economic actors who participated in the management of the crisis. So far, only exploratory interviews have been conducted with the governing bodies of companies involved in the event. A total of 12 informal interviews were carried out lasting an hour and a half that explore the role of actors in the management of the crisis, and report their perception of the event.

In addition, a systematic study of the regional press that covered events will be conducted, in

order to understand its role in the construction of the image that actors may have developed of the crisis, or even their role in the emergence of a crisis culture specific to the sector and the region.

The initial interview data indicates that recovery was driven by the desire to continue to provide a superior quality service while at the same time minimizing the crises. Statements like "the show must go on" and "putting on a brave face" were common. There was also a collective effort to portray the Côte d'Azur tourist sector as a dependable in spite of the previous crises.

Understanding the resilience of organizations—that is how they respond to a crisis—requires an understanding of the attitudes of the individuals involved. In this case it was the attitudes of the individual actors who lived through the crisis and who then have had to accommodate the effects of the crisis in their daily life. How are they able to create a world that encompasses the crisis? In practice, following Moreau (2017). Despite the tragedy, the world goes on. The way of thinking about the crisis in the tourism and events world, anchored in the mindset of the region, then becomes a condition for the commitment of actors and their capacity to innovate.

7 CONCLUSION

The tourism sector is considered to be a key economic player in France. However, it is exposed to a variety of crises and disasters at the micro, national and international levels. Risk perception is an important part of the tourist's decision process in both leisure and business tourism. Risk perception can also be mitigated by presenting a positive image of the destination and in particular how the sector was able to effectively respond and manage the crisis. The question here is how is that constructed and what are the key features?

Although an increasing number of studies have begun to appear in the literature in the past decade, there is still a lack of research on crisis management in the tourism sector. Those studies that are available all point to the need to implement a strategic plan at each stage of the crisis management. Moreover, our findings show that a three-tiered system is an effective way of providing a comprehensive response that incorporates input at all levels of the sector.

The literature also highlights the need to move beyond overly procedural plans but instead allow for innovative responses when required. That is, from a resilience perspective, it is necessary to allow individuals to be innovative and responsive to the environment. Our findings support the literature in that any crisis management approach must have a level of flexibility. By tapping into the mindsets of "the show must go on" and allowing individuals to be innovative, a more committed workforce will emerge, thus ensuring that the crisis management plans are deployed. But the key ingredient is to create a dynamic, sectoral environment that is responsive and resilient during and after crises.

REFERENCES

Atkas, G. & Gunlu, E. 2005. Crisis Management in tourist destination in Theobald, W.F. (Ed.). (2005). *Global tourism*. Routledge. (P. 440–456).

Beirman, D. 2003. Restoring tourism destinations in crisis: A strategic marketing approach. *CAUTHE 2003: Riding the Wave of Tourism and Hospitality Research*, 1146.

Burnard, K., & Bhamra, R. 2011. Organisational resilience: development of a conceptual framework for organizational responses. *International journal of Production Research,* 49(18), 5581–5599.

Chew, E.Y.T., & Jahari, S.A. 2014. Destination image as a mediator between perceived risks and revisit intention: A case of post-disaster Japan. *Tourism Management*, *40*, 382–393.

Enders, W., Sandler, T., & Parise, G.F. (1992). An econometric analysis of the impact of terrorism on tourism. *Kyklos*, *45*(4), 531–554.

Drakos, K., & Kutan, A.M. 2003. Regional effects of terrorism on tourism in three Mediterranean countries. *Journal of Conflict Resolution*, *47*(5), 621–641.

Fainstein, S.S. 2002. One year on. Reflections on September 11th and the 'War On Terrorism': regulating New York City's visitors in the aftermath of September 11th. *International Journal of Urban and Regional Research*, *26*(3), 591–595.

Faulkner, B. 2001. Towards a framework for tourism disaster management. *Tourism management*, *22*(2), 135–147.

Funabashi, Y., & Kitazawa, K. (2012). Fukushima in review: A complex disaster, a disastrous response. Bulletin of the Atomic Scientists, 68(2),9–21.

Hollnagel, E., Woods, D.D., & Leveson, N. 2007. *Resilience engineering: Concepts and precepts*. Ashgate Publishing.

Huchon, J.P. 2016. *Rapport au premier minister sur la destination France après les attentats*. Hôtel de Matignon..

Glaesser, D. (2006). *Crisis Management in the touriusm industry*. Routledge.

Travadel, S., Martin, C., Guarnieri, F., Decision Making on Trial: The Extreme Situation at Fukushima Daiichi. Vicki Bier. France. Routledge, Chapter 4 – p. 65–80, 2018, Risk in Extreme Environments: Preparing, Avoiding, Mitigating, and Managing.

Holt, M., Campbell, R.J., & Nikitin, M.B. 2012. *Fukushima nuclear disaster*. Congressional Research Service.

La Porte, T.R. (1996) High Reliability organizations: Unlikely, demanding and at risk. *Journal of contengencies and crisis management*, 4(2),83–92.

Moreau, Y. 2017. *Vivre avec les catastrophes*. Presses Universitaires de France.

Laws, E., & Prideaux, B. (2006). Crisis management: A suggested typology. *Journal of Travel & Tourism Marketing*, *19*(2–3), 1–8.

Lepp, A., & Gibson, H. 2003. Tourist roles, perceived risk and international tourism. *Annals of tourism research*, *30*(3), 606–624.
Perrow, C. 2011. *Normal accidents: Living with high risk technologies*. Princeton University Press.
Polhill, R.M. 1982. *Crotalaria in Africa and Madagascar*. Rotterdam: Balkema.
Ritchie, B.W. 2004. Chaos, crises and disasters: a strategic approach to crisis management in the tourism industry. *Tourism management*, *25*(6), 669–683.
Ritchie, B.W., & Campiranon, K. (Eds.). (2014). *Tourism crisis and disaster management in the Asia-Pacific* (Vol. 1).
Roberts, K.H. 1990. Managing high reliability organizations. *California Management Review*, 32(4), 101–113.
Sausmarez, N.D. (2003). Malaysia's response to the Asian financial crisis: implications for tourism and sectoral crisis management. *Journal of Travel & Tourism Marketing,* 15(4), 217–231.
Weick, K.E. (1995). *Sensemaking in organizations* (Vol. 3). Sage.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Validation of a gamified measure of safety behavior: The SBT

C.B.D. Burt, L. Crowe & K. Thomas
University of Canterbury, New Zealand

ABSTRACT: Safety behavior is defined by constructs such as safety compliance, safety participation, and risk-taking. A safety complaint employee follows safety rules and policies, and if they engage in safety participation, goes beyond their job to ensure everyone's safety. In contrast, risk taking and safety violation behaviors can cause accidents and system failures. The significance of safety behavior argues for its measurement in job applicants, and the use of the resulting data to select applicants in and out of high risk situations, and/or allocate to training programs. The development of the Safety Behavior Test (SBT) which is a gamified assessment tool operationalized within an animated work simulation environment is described. Participant's SBT score was correlated with data on their actual safety behavior provided by an independent source. Results indicate the SBT has good criterion-related validity, but this is influenced by computer-game playing experience. The advantages of using gamification to measure safety behavior are discussed.

1 INTRODUCTION

Research varies in the estimation of the influence of behavioral safety in accident causation. Some researchers have suggested as much as 90 percent of accidents can be shown to have a behavioral safety component as a contributing factor (see Hofmann et al. 2017 for a useful review). It is also clear that specific cohorts of employees, such as young employees (e.g., Salminen, 2004), or those new to a job can engage in less safe behaviors and more risky behaviors (e.g., Burt, 2015). Behavioral safety not only requires an individual to engage in safe behaviors, but also to avoid a spectrum of unsafe or risky behaviors, including routine rule-breaking (e.g., Darby et al. 2005). While behavioral safety is clearly an important factor in the overall safety performance of a system, organizations are somewhat restricted in their ability to manage it. Research has examined the influence on safety behavior of generating a positive safety culture (e.g., Cui et al. 2013, Diaz-cabera et al. 2007), and the impact on safety behavior of having a positive safety climate within work teams (e.g., Christian et al. 2009, Clarke, 2006). From this work it is clear that a positive safety culture and climate can help promote safety behaviors, and the avoidance of risky behavior.

A further option to manage safety behavior is to recruit people into the organization who are likely to behavior in a safe way. In order to do this organizations need to use measures during recruitment which allow for the prediction of individuals' on the job safety behavior. In examining the relationship between recruitment processes and safety, a number of studies have highlighted the need to ensure all employees have the skill, knowledge and training to complete their work in a safety manner (e.g., Ford & Wiggins, 2012, Grandell-Niemi et al. 2003, McMullan et al. 2010, Postlethwaite et al. 2009), suggesting that appropriate ability tests should be used during recruitment. A considerable body of work has also attempted to understand the link between personality and safety behavior, with the idea of an accident prone personality receiving research attention for almost 100 years (e.g., Dahlback, 1991, Green & Woods, 1919, Visser et al. 2007), along with the idea that risk-taking is a personality trait (e.g., Dahlback, 1990, Eysenck & Eysenck, 1977). More recently, Clarke and Robertson (2005) used meta-analysis to examine the criterion-related validity of the big five personality factors as predictors of accident involvement, and concluded that conscientiousness and agreeableness were valid and generalizable predictors of safety.

In addition to ability and personality testing, organizations may attempt to predict safety behavior using questions about safety behavior or the individual's past accident history in an application blank, or in an employment interview. While these measurement options have the potential to contribute to an organization's understanding of how a job applicant might behave in the future, they have serious limitations. Under these 'self-reporting' conditions, it is very obvious what is being measured, which makes responses susceptible to social desirability biases. Social desirability

bias refers to a phenomenon where participants over-report favorable opinions and behavior, while under-reporting those that are unfavorable, and is most common when the subject under investigation is considered to be sensitive by the respondent (Krumpal, 2013). Sensitive subjects are defined as those where there are potential costs or risks to the respondent for responding in a particular way, or to the collective population that the outcome of the question represents (Sieber & Stanely, 1988). Clearly, a focus on safety within a job application process would represent such a situation.

Given the importance of employee safety behavior, an ideal way to assess it would be to use a work sample approach where the job applicant can demonstrate their behavior. However, there are clear ethical reasons why job applicants' can not be placed into a risky situation to measure how they will respond. A way to avoid this ethical issue is to use a simulation, and as such this study reports on the development of a measure of safety behavior that uses a work simulation developed using the *gamification paradigm* which is rapidly growing in its application across a number of areas (e.g., Chen et al. 2015, Rodrigues, Costa & Oliveira, 2016, Singh, 2012). Gamification can be defined as the use of game design elements in non-game contexts, and is predominantly used to make real world activities more engaging (Deterding et al. 2011, Shrofeld, 2010). Organizations have also started applying gamification to their selection procedures (Chamorro-Premuzic, 2015). The popularity of gamification in the work place is likely due to the positive impact it has on engagement and motivation (Gagne & Deci, 2005, Harter et al. 2013). Within the area of safety behavior measurement during employee recruitment gamification has considerable potential to avoid measurement bias.

To be of value a gamified psychometric test must be valid, must accurately measure the construct intended, and be predictive of future performance related to that construct. Very few studies have investigated the implementation of gamified assessments, and even fewer have investigated the validity of gamified measures. Of those which do report on gamified assessments, validity results are mix. Some studies have reported the gamified measure examined was not valid (e.g., Jaffal & Wloka, 2015, Kim & Shute, 2015, Whetzel et al. 2012), while others report they were able to develop a valid gamified measure (e.g., Mislevy, Almond & Lukas, 2003, Shute,Ventura & Kim, 2013), supporting the use of a gamified measure as a way to observe authentic behaviors (Clarke, 2009, Clarke-Midura & Dede, 2010, Ketelhut et al. 2008, Shavelson et al. 1991). These mixed results may not necessarily reflect on gamified assessments as a whole, but the design of the measurement points within the specific gamified assessments which were examined in each study.

Gamified assessments contain two key design components, measurement and game design. Measurement design is developed first and contains the construct intended for measurement, and the design of measurement items for this construct. During this phase of gamification design the virtual environment is created and the measurement items are applied to decision points within the game (Halverson & Owen, 2014, Ifenthaler et al. 2012). Each decision point is recorded by a play log and summed to provide the user with a score or scores on the measured construct. Game design helps determine a game-based assessment's validity and ability to predict employee behaviors in terms of its similarity to a work-sample test (Gangadharbatla & Davis, 2016). Work samples have been consistently regarded as one of the most accurate and valid measures in predicting performance (Hunter & Hunter, 1984, Reilly & Warech, 1993, Roth et al. 2005). As noted it is difficult, if not impossible to use traditional work sample testing to measure safety constructs (e.g., to measure the constructs of safety compliance would require an individual to complete a potentially dangerous task), a gamified assessment can replicate a work-sample by requiring the individual to provide a performance based sample through a virtual rather than physical environment. In doing this safety behavior should be able to be assessed with no actual danger to the participant.

In summary, this study reports on the development and initial validation of a gamified measure of safety behavior: the *Safety Behavior Test* (*SBT*). The SBT was designed to be used during the employee recruitment process. The validation process involved correlating the SBT scores with data on safety behavior obtained from independent raters.

2 METHOD

2.1 *Design*

A concurrent criteria-related validity design was used. The participant completed the SBT followed by an Individual Characteristics Questionnaire (ICQ). The participant was then given an Acquaintance Questionnaire (AQ), and they nominated an individual (e.g., supervisor, co-worker) to complete it. The AQ provided the independent data on the participant's safety behavior against which the SBT score was validated. The ICQ data was used to determine if any of the variables measured showed adverse impact on the SBT score, and thus needed to be controlled for in the validation

analysis. The study was reviewed and approved by the University of Canterbury Human Ethics Committee, reference number HEC 2017/26.

2.2 *Sampling*

Haphazard sampling (Weisberg & Bowen, 1977) was used to obtain the SBT participants. The recruitment criteria were the participant had to be working (responses to a question on this indicated all participants were working), and had to agree that they had adequate vision to use a computer to complete the SBT. Acquaintances (e.g., supervisor, co-worker) for the collection of independent data on the participant's safety behavior were recruited by the SBT participant. To participate as an *acquaintance*, the individual had to have knowledge of their respective SBT participant's safety behavior. SBT participants and acquaintances each received a $10 petrol voucher for their participation.

2.3 *Participants*

The study involved 200 individuals, 100 participants who completed the SBT who had a mean age of 41.6 years (range 18–66 years), and 100 acquaintances with a mean age of 43.5 years (range 18–66 years). SBT participants reported an average total work tenure of 286.8 months (SD = 158), and an average job risk rating on a 100 point scale of 42.4 (SD = 28.3). Acquaintances indicate they had known their SBT participant for an average of 144.7 months (SD = 172.9), and using a 100 point scale rated the strength of their relationship with their SBT participant on average as 73.4 (SD = 22.1). These results suggests that acquaintances should have had ample opportunity to observe the SBT participant's safety behavior, and thus be able to provide valid ratings.

2.4 *Materials*

Data was collected using three sources: the SBT, the individual characteristics questionnaire, and the acquaintance questionnaire.

2.4.1 *Safety behavior test*

The SBT is a fully animated computer test of the point and click game genre, meaning that the test taker can point the cursor at an area on the screen and click in order to interact with the test environment. First person views are used in the SBT, thus the test taker experiences the test as if they were a character navigating the test environment. A number of gamification design elements providing feedback to the test taker are included in the SBT, including a timer shown in the top left corner of the screen which displays play time in seconds, a click counter also shown in the top left hand corner of the screen which displays the number of times the player has clicked the mouse, and feedback on some decisions, such as a red cross appearing if an incorrect choice is made within the game.

To protect the security of the SBT only a general description of it is provided here (note the testing procedure and pre-test instructions are described in the procedure section). The SBT requires the test taker to assume the role of a worker in a waste disposal company. The SBT starts with the animation opening a door and moving into an office scene. A sign on the desk reads "Press red button if unattended", which indicates that the participant needs to click the red button. Clicking the red button plays a narrative instruction: "*Hello forklift driver number 1. Sorry, I am up on level 6. It's good that you are here on time, there is only one job for you today. You have a shipment for disposal at the incinerator. The empty shipping container for the shipment is in loading dock C. A truck will take the loaded container to the incinerator as soon as you have finished loading it. I have already put the shipment items into the system, so when you get in a forklift the item list will be on the display screen. The new semi-automatic forklifts are working great, just click an item on the list and off you go to the relevant floor. Remember that control buttons appear when you need them. We have fixed the problem with the red right and left directional control arrows, and the central yellow stop button is working fine on all forklifts. Remember to load the items in the order shown on the list. The cloak room is nice and tidy this week, so let's keep it that way. Don't muck around as the transportation firm will charge us if they have to wait, but be careful. When you have got the order loaded come back here and let me know. If you would like me to repeat the instructions, just click the red button again*". After the participant had listened to the narrative (they can only play it twice) they make a number of decisions which allows them to navigate through the game/test by first entering the cloak room, followed by another room where they select a forklift. They then control the forklift to 5 different levels within the building. On each level they collect an item and move it to a container in the loading bay. During this process they encounter various hazards, and have safety related decisions to make. Once they have loaded all the required items into the container they exit via the supervisor's office where they complete a check list which has questions about the work they have just completed. Submitting the answers to these questions closes the test. During the SBT the player encounters 35 decision points. These decision points vary in terms of whether the decision has a safety aspect (e.g., following directional

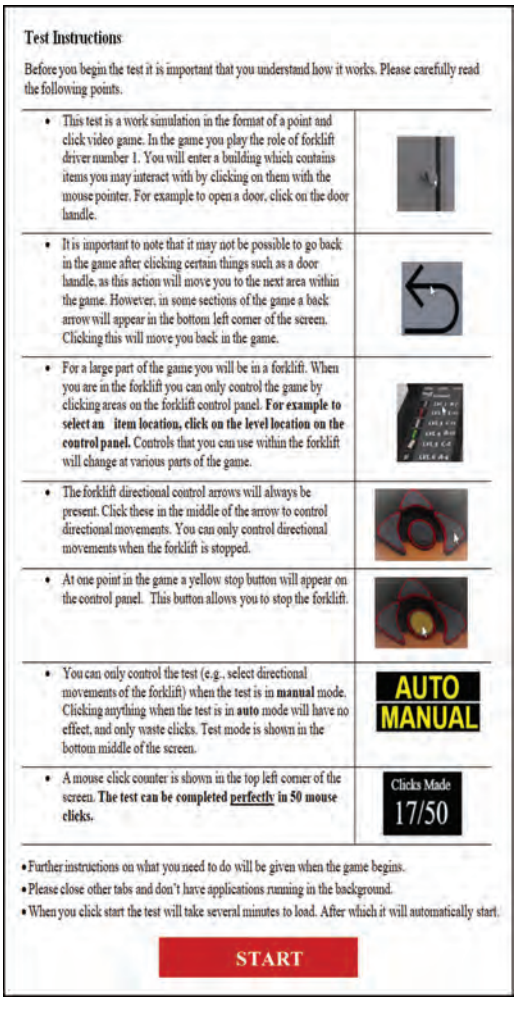


Figure 1. The SBT instruction page.

arrows on the floor while driving the forklift), or a risk aspect (e.g., avoiding driving over a hose in the forklift), or is simply a decision required to advance in the game (e.g., clicking a door handle to open a door). Safety and risk related decisions were used to form the SBT score. A safe decision was given 1 point and these points were summed to produce the score, with a possible SBT score range of 0 to 13. The SBT instruction page which participants were asked to read before taking the test is shown in Figure 1.

2.4.2 *Individual characteristics questionnaire*

Demographic information on age and gender, along with job tenure, experience driving a forklift, and computer game playing experience was collected. For the later variable, SBT participants' indicated how many *months* they had played computer games. Safety risk of the SBT participant's current job was assessed by placing a mark on a 100 point scale (0 = not risky at all, 100 = extremely risky).

2.4.3 *Acquaintance questionnaire*

The relationship the acquaintance has with their respective SBT participant was measured by how long the acquaintance had known the SBT participant for (months), and a rating of how well they know them on a 100 point scale where 0 = *Not very well at all* to 100 = *Extremely well*. Additionally, the SBT participants' safety behavior was rated by the acquaintance using scales measuring safety participation, safety compliance, safety voicing, safety consciousness, rule-bending and risk taking. The wording of scale items was adjusted from a first person format to suit third person acquaintance ratings. For example, "*I always use all the necessary safety equipment to do my job*" was adjusted to "*… *always uses all the necessary safety equipment to do their job*". Acquaintances were instructed that the *… referred to the individual that asked them to complete the questionnaire. Each scale item was responded to on a 5-point Likert scale anchored with 1 = strongly disagree and 5 = strongly agree. Scales were analyzed for internal consistency, and coefficient alphas are reported below. Scale scores were formed by summing the item ratings and dividing the sum by the number of items in the scale. Depending on the scale, a higher score is indicative of a higher level of safety behavior, or a higher level of risk-taking/rule-bending behavior.

Safety compliance and participation, were measured using a third person adapted version of Neal and Griffin's (2006) six item scale. Three items measured safety compliance, defined as core activities an employee needs to engage in to maintain workplace safety. The other three items measured safety participation, defined as behaviors which help to develop an environment that supports safety. An example items for safety participation is "*At work *… puts in extra effort to improve the safety of the workplace*". Coefficient alphas for safety compliance and participation are .87 and .83, respectively. Propensity for the SBT participant to breach workplace safety rules and procedures was measured using an adapted version of Chmiel's (2005) four-item bending the rules scale. An example item is, "*… *sometimes cuts corners if it makes the task easier*". A coefficient alpha of .87 was found for this scale. Safety consciousness and risk taking were measured using the 12-item safety consciousness and risk-taking scale developed by Westaby and Lee (2003). Safety consciousness is defined as "a positive attitude and awareness toward acting safely in general", and risk-taking is

defined as an "individual's willingness to engage in activities that knowingly have elements of physical danger" (Westaby & Lee, 2003, p. 228). An example item for safety consciousness is "*… *gets upset when seeing other people acting dangerously*", and an example for risk-taking is "*… *values having fun more than being safe*". The safety consciousness scale coefficient alpha was .82 and the risking-taking scale coefficient alpha was .82 after removal of one item. Acquaintance's perception of their respective SBT participant's safety voicing behaviors was measured using an adapted version of Tucker et al.'s (2008) five-item *safety voicing scale*. Safety voicing is measured as "any individual communication directed at improving safety conditions" (Tucker et al. 2008, P.319). An example item is "*… *makes suggestions about how safety could be improved*". The obtained safety voicing scale coefficient alpha was .87.

2.5 *Procedure*

SBT participants were tested individually in a quiet room. Participants read a consent sheet, and an information sheet which indicated they were being asked to take a safety test, complete a questionnaire and give a further questionnaire to an acquaintance who would be asked to complete questions about them. After consenting, participants were told "*Please read the instructions* (see Figure 1) *carefully, as you will only be able to see them once. Press the start button when you are ready to take the test. I will leave you to take the test privately, and will be waiting outside of the room for when you have finished. Please imagine that you have applied for a job. The test you are about to complete is being used to determine your suitability for the job. As a job applicant, try to do your best on the test*". A Lenovo ideapad 510–15ikb laptop which has a 15.6 inch screen was used to run the SBT.

After completing the SBT, the participant completed the *individual characteristics questionnaire*. The participant then received an unsealed envelope containing an *acquaintance questionnaire*. The SBT participant gave this envelop to their selected acquaintance, who completed the questionnaire and sealed it in the envelope which was then returned to the researchers.

3 RESULTS

Missing data for acquaintance rating scale items was replaced with the variable mean. Missing data for demographic questions was left as missing, and resulted in some n = variance across the analyses. SBT participants took an average 17.9 minutes (SD = 2.84, range 13.1 to 26.1) to complete the SBT. The average SBT score was 8.0 (SD = 2.72, range = 0 to 13, skew = –.454, kurtosis = –.067). Table 1 shows the comparison of SBT scores by gender, computer game player versus never played computer games, and previously driven a forklift versus never driven a forklift.

Inspection of Table 1 indicates that being a computer game player significantly increased the participant's SBT score. Thus computer game experience in the form of how many months the participant had played computer games (mean = 177.9, SD = 130.3) was used as a covariate in the criterion-related validity analysis. Correlational analysis using Pearson product moment correlations showed no significant relationship between the participant's age and SBT score (r = –.02, n = 99), total work tenure and SBT score (r = .02, n = 96), or rated safety risk of their current job and SBT score (mean = 42.4, SD = 28.3, r = .01, n = 98). Overall, the SBT score data shows a distribution suitable for validation analysis, and of the variables examined is only significantly impacted by computer game playing experience.

To examine the criterion-related validity of the SBT, several correlational analyses were performed. Table 2 shows the descriptive statistics for the independent safety behavior scale data from the acquaintance questionnaire and the correlation between each safety behavior scale scores and the SBT score. The first correlational analysis (column 5 of Table 2) used all 100 participants. The second analysis used the number of months the participant had been playing computer games as a covariate, and thus shows partial correlations. The partial correlation analysis resulted in a decrease in sample size (n = 30) due to the small number of computer game players in the total sample, and that only 30 of these reported their playing experience

Table 1. Comparison of SBT score by gender, computer game playing experience, and forklift driving experience.

	SBT Score Mean & (SD)	SBT Score Mean & (SD)	t-test
Gender	Male, n = 62 7.91 2.92	Female, n = 38 8.26 2.39	= –.611
Computer game player	Yes, n = 36 8.91 (2.61)	No, n = 62 7.66 (2.66)	= –2.263*
Driven a forklift	Yes, n = 57 7.96 (2.90)	No, n = 43 8.16 (2.49)	= .358

*P < .05.

Table 2. Safety behavior measure descriptive statistics and correlations with SBT scores.

Safety behavior measure	Mean (SD)	Skew (Kurtosis)	Correlation with SBT score N = 100	Partial correlation† with SBT score N = 30
Voice	3.67 (1.1)	–1.27 (1.88)	.02	.25
Compliance	4.03 (.95)	–1.33 (2.39)	.13	.42*
Participation	3.52 (1.2)	–.92 (.63)	.00	.22
Consciousness	3.63 (.78)	–.59 (.14)	.08	.42*
Rule bending	2.35 (2.2)	.08 (–.80)	–.20*	–.46**
Risk taking	2.06 (.92)	.82 (.90)	–.08	–.41*

†Controlling for months of computer game playing, *P < .05, **P < .01.

in months (mean = 177.9, SD = 130.3, range = 19 to 480). Inspection of Table 2 indicates SBT criterion-related validity evidence is shown when computer game playing experience is controlled for.

4 GENERAL DISCUSSION

The SBT scores across the 100 participants produced a distribution consistent with the SBT being a useful measurement tool. The overall mean SBT score was close to the score range mid-point, and the standard deviation and skew statistics are indicative of a normal distribution. These measurement characteristics are ideal for a measure which is attempting to scale job applicants or employees on their safety behavior. The overall average test time of approximately 18 minutes is also likely to be acceptable within the process of employee recruitment. Tests that run too long are said to be associated with response burden, where participants feel strained by the test experience and this could adversely impact their test score (Rolstad, et al. 2011). Of the variables examined, computer game experience showed an impact on SBT performance, and this is very evident in the criterion-related validity analysis shown in Table 2. The implications of this are discussed below.

The SBT was developed to help avoid bias associated with self-report measures of safety behavior. Using the assumption that behavior within a gamified work environment simulation would accurately reflect behavior in a 'real' situation, the SBT was designed to include decision points which could be scored as either safety or unsafety (risky). Individuals with a propensity to behave safely, not bend the rules or take risks should score higher on the SBT, and the correlations with the independent rating of safety behavior shown in Table 2 are consistent with the SBT being able to achieve this objective. However, a key factor in the SBT's measurement ability is the participant's experience with the gamification mode, their experience with computer game playing. When computer game experience is controlled for the partial correlations support the criterion-related validity of the SBT, and the size of the obtained partial correlations are sufficient to ensure that the SBT would have utility within an operational situation, such as employee recruitment.

The finding that computer game experience influenced SBT performance is consistent with other validation research into gamified tests (e.g., Kim & Shute, 2015). Kim and Shute (2015) examined the influence of computer game experience on participant's performance on their gamified physics test, and reported performance on the assessment was influenced by game playing experience. Participants identified as gamers were found to have had an advantage over non-games on achieving test points. A number of explanations for the influence of computer game experience on SBT performance need to be explored in future research. In completing the SBT it is possible that non-computer game players had to focus on learning how to control the game/test, made decision errors because of their unfamiliarity with the test mode, and this reduced their SBT score. Under such conditions their SBT behavior, as reflected in their SBT score, would not be expected to be associated with their acquaintance's ratings of their safety behavior. The influence of test mode unfamiliarity should be able to be easily addressed through the use of sufficient pre-test instruction given in the operational aspects of the test mode. The SBT instructions shown in Figure 1, while assumed to be sufficient, are static and further work on enhancing them using brief animation clips which include point and click control trials similar to those used in the SBT simulation environment need to be developed.

It is also possible that computer game players may have characteristics which are associated with behaving safely, and their SBT performance has less to do with game control familiarity, rather is a reflection of their tendency to behavior safely. Teng (2008) found that computer game players had higher scores on openness to experience, conscientiousness, and extraversion than non-computer game players. Geller (2004) suggested that being open to experience would make people more likely to accept and engage in new health and

safety related initiatives, while being conscientious would see people being more inherently interested in safety processes, and being extraverted would make it easier for people to use safety procedures that require communication between people. Furthermore, in a meta-analysis of personality traits and accident involvement, Clark and Robertson (2005) found low conscientiousness to be a valid predictor of occupational accident involvement.

There may also be a cognitive aspect associated with computer game playing which has relevance to safety behavior. Pillay (2002) revealed in an investigation of the cognitive processes engaged in by computer game players that children who played recreational computer games performed better on subsequent educational tasks than a control group of children that did not play computer games. Higher levels of intelligence have been shown to be associated with lower rates of accidents. Specifically, the information processing skills of being able to recall relevant information, quickly identify problematic situations, and react quickly to unforeseen situations are said to be key skills in accident prevention (Gottfredson, 2004).

While a number of issues need to be addressed in future research on the SBT, the evidence from this study suggests the SBT has the potential to be a useful tool for both employee recruitment and research on safety behavior. The merging of traditional psychometrics and animated game technology provides for measurement which might be less susceptible to bias than self report. In the case of safety behavior this may result in significant benefits to organizations through accident reduction, and help protect individuals from injury through the identification of those who could benefit greatly from instruction around safe work place behaviors.

REFERENCES

Burt, C.D.B. 2015. *New employee safety: Risk factor and management strategies*. Heilelberg: Springer International. Chamorro-Premuzic, T. 2015. Three emerging alternatives to traditional hiring methods. *Harvard Business Review*. June.

Chen, Y., Burton, T., Mihaela, V. & Whittinghill, D.M. 2015. Cogent: A case study of meaningful gamification in education with virtual currency. *International Journal or Emerging Technologies in Learning, 10*(1): 39–45.

Chmiel, N. 2005. Promoting health work: Self-reported minor injuries, work characteristics, and safety behaviour. In C. Krunka, P. Hoffmann, & A Bussing (Eds.), *Change and quality in human service work* (p 277–288). Mering: Rainer HamppVerlag.

Christian, M.S., Bradley, J.C., Wallace, J.C. & Burke, M.J. 2009. Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology*, 94(5): 1103–1127.

Clarke, J. 2009. *Studying the potential of virtual performance assessments for measuring student achievement in science.* Paper presented at the Annual Meeting of the American Educational Research Association (AERA), San Diego, CA.

Clarke, S. & Robertson, I.T. 2005. A meta-analytic review of the Big Five personality factors and accident involvement in occupational and non-occupational settings. *Journal of Occupational and Organizational Psychology*, 78, 355–376.

Clarke, S. 2006. The relationship between safety climate and safety performance: A meta-analytic review. *Journal of Occupational Health Psychology*, 11(4): 315–327.

Clarke-Midura, J., & Dede, C. 2010. Assessment, technology, and change. *Journal of Research on Technology in Education*, *42*(3): 309–328.

Cui, L., Fan, D. Fu, G., & Zhu, C.J. 2013. An integrative model of organizational safety behavior. *Journal of Safety Research*, 45, 37–46.

Dahlback, O. 1990. Personality and risk-taking. *Personality and Individual Differences*, 11, 1235–1242.

Dahlback, O. 1991. Accident-proneness and risk-taking. *Personality and Individual Differences*, 12, 79–85.

Darby, T.F., Pickup, L., & Wilson, T.R. 2005. "Safety culture in railway maintenance", *Safety Science*, 43, 39–60.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. 2011. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9–15). ACM.

Diaz-cabera, D. Hernandez-Fernaud, E., & Isla-Diaz, R. 2007. An evaluation of a new instrument to measure organizational safety culture values and practises. *Accident Analysis and Prevention*, 39, 1202–1211.

Eysenck, S.B.G. & Eysenck, H.J. 1977. The place of impulsiveness in a dimensions system of personality description. *British Journal of Social and Clinical Psychology*, 16, 57–68.

Ford, M.T. & Wiggins, B.K. 2012. Occupational-level interactions between physical hazards and cognitive ability and skill requirements in predicting incidence rates. *Journal of Occupational Health Psychology*, 17(3): 268–278.

Gagné, M., & Deci, E. L. 2005. Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26, 331–362.

Gangadharbatla H. & Davis, D.Z. (Ed.) 2016. *Emerging research and trends in gamification*. Information Science Reference.

Geller, S. 2004. The "big five" and you: How personality traits can affect behaviour. *Industrial Safety & Hygiene News, 38*(7): 12.

Gottfredson, L.S. 2004. Life, death, and intelligence. *Journal of Cognitive Education and Psychology*, *4*(1): 23–46.

Grandell-Niemi, H., Hupli, M., Leino-Kipli, H. & Puukka, P. 2003. Medication calculation skills of nurses in Finland. *Journal of Clinical Nursing*, 12, 519–528.

Greenwood, M., & Woods, H.M. 1919. A report on the incidence of industrial accidents with special refer-

ence to multiple accidents. *Report 4, Industrial Fatigue Research Board.*

Halverson R. & Owen E. 2014. Game-based assessment: an integrated model for capturing evidence of learning in play, *International Journal of Learning Technology*, 9(2): 111–138.

Harter, J.K., Schmidt, F.L., Agrawal, S., & Plowman, S.K. 2013. The relationship between engagement at work and organizational outcomes. *Gallup Poll Consulting University Press, Washington*.

Hofmann, D.A., Burke, M.J., & Zohar, D. 2017. 100 years of occupational safety research: From basic protections and work analysis to a multilevel view of workplace safety and risk. *Journal of Applied Psychology*, *102*(3): 375–388.

Hunter JE, Hunter RF. 1984. Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72–98.

Ifenthaler, D. Eseryel, & X. Ge 2012. Assessment in game-based learning: Foundations, innovations, and perspectives (pp. 59–81). New York, NY: Springer.

Jaffal, Y., & Wloka, D. 2015. Employing game analytics techniques in the psychometric measurement of game-based assessments with dynamic content. *Journal of e-Learning and Knowledge Society*, *11*(3), 101–115.

Ketelhut, D., Dede, C., Clarke, J., Nelson, B., & Bowman, C. 2008. Studying situated learning in a multi-user virtual environment. In E. Baker, J. Dickieson, W. Wulfeck, & H. O'Neil (Eds.), *Assessment of problem solving using simulations* (pp. 37–58). Mahweh, NJ: Erlbaum.

Kim, Y. J., & Shute, V. J. 2015. The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Computers & Education*, *87*, 340–356.

Krumpal, I. 2013. Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, *47*(4): 2025–2047.

McMullan, M., Jones, R. & Lea, S. 2010. Patient safety: Numerical skills and drug calculation abilities of nursing students and registered nurses. *Journal of Advanced Nursing*, 66(4): 891–899.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. 2003. A brief introduction to evidence-centered design. *ETS Research Report Series*, *2003*(1).

Neal, A., & Griffin, M.A. 2006. A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels. *Journal of Applied Psychology*, *91*(4): 946–953.

Pillay, H. 2002. An investigation of cognitive processes engaged in by recreational computer game players: Implications for skills of the future. *Journal of Research on Technology in Education*, *34*(3): 336–350.

Postlethwaite, B., Robbins, S., Rickerson, J. & McKinniss, T. 2009. The moderation of conscientiousness by cognitive ability when predicting workplace safety behavior. *Personality and Individual Differences*, 47, 711–716.

Reilly RR, Warech MA. 1993. The validity and fairness of alternatives to cognitive ability tests. In Wing L, Gifford B (Eds.), *Policy issues in employment te*sting. Boston: Kluwer.

Rodrigues, L.F., Costa, C.J., & Oliveira, A. 2016. Gamification: a framework for designing software in e-banking. *Computers in Human Behavior*, *62*, 620–634.

Rolstad, S., Adler, J., & Rydén, A. 2011. Response burden and questionnaire length: is shorter better? A review and meta-analysis. *Value in Health*, *14*(8): 1101–1108.

Roth, P. r., Bobko, P., & McFarland, L. A. 2005. A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology*, *58*(4): 1009–1037.

Salminen, S. 2004. Have young workers more injuries than older ones? An international literature review. *Journal of Safety Research* 35: 513–521.

Shavelson, R. J., Baxter, G. P., & Pine, J. 1991. Performance assessment in science. *Applied Measurement in Education, 4*(4): 347–362.

Shronfeld, E. 2010. *SCVNGR's Secret Game Mechanics Playdeck*. Retrieved August, 31,

Shute, V. J., Ventura, M., & Kim, Y. J. 2013. Assessment and learning of qualitative physics in newton's playground. *The Journal of Educational Research*, *106*(6): 423–430.

Singh, S.P. 2012. Gamification: A strategic tool for organizational effectiveness. *International Journal of Management*, *1*(1): 108–113.

Teng, C.I. 2008. Personality differences between online game players and nonplayers in a student sample. *CyberPsychology & Behavior*, *11*(2): 232–234.

Tucker, S., Chmiel, N., Turner, N., Hershcovis, M.S., & Stride, C. B. 2008. Perceived organizational support for safety and employee safety voice: the mediating role of coworker support for safety. *Journal of Occupational Health Psychology*, *13*(4): 319.

Visser, E., Pijl, Y.J., Stolk, R.P., Neeleman, J & Rosmalen, J.G.M. 2007. Accident proneness, does it exist? A review and meta-analysis. *Accident Analysis and Prevention*, 39, 556–564.

Weisberg, H.F., & Bowen, B.D. 1977. An introduction to survey research and data analysis. San Franciso: W.H. Freeman.

Westaby, J.D., & Lee, B.C. 2003. Antecedents of injury among youth in agricultural settings: A longitudinal examination of safety consciousness, dangerous risk taking, and safety knowledge. *Journal of Safety Research, 34*(3): 227–240.

Whetzel, D.L., McDaniel, M.A., & Pollack, J.M. 2012. Work simulations. *The handbook of work analysis: Methods, systems, applications and science of work measurement in organizations*, 401–418.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Professionalization in safety: A study of the professional context of a post master safety program's alumni

Wim van Wassenhove
CRC – Centre de Recherche sur les Risques et les Crises, MINES ParisTech, PSL Research University, France

Christian Foussard
Consultant Risk Management and Process Safety, Dubai Festival City, UAE

ABSTRACT: Our contribution to the ESREL conference will present the first results of a qualitative and quantitative study on the professionalization of alumni of a post master safety program. The post master Management of Industrial Risks has graduated 197 students since 2004. We addressed an online questionnaire to all the alumni, five in depth interviews will be realized with alumni and their work colleagues and a focus group discussion will be held. The online questionnaire will investigate their careers (company, position and wages) but also how safety is organized in their company, with whom they interact, what are their missions. For the job missions, we focus on their personal convictions, the importance given by their organization or company and the real time spent doing the job missions. We will also investigate the characteristics of their work conditions and what are, from their own point of view, the real contributions for safety in their organization. Finally, the knowledge and competencies they mobilize or want to mobilize doing their job will be assessed. Those findings intend to, firstly, dress a picture of the safety professional daily job, secondly, identify factors that shape the safety professional's role in the organization and, finally, elaborate a more adapted curriculum for the training and development of future safety professionals.

1 INTRODUCTION

The safety science is a rather young science. As it is a multidisciplinary science mixing social science with engineering science, education programs are also very young compared to the education of traditional professions (like medicine, law and applied sciences professions like engineers) who are depending on one traditional and identified discipline. Although INSHPO, the International Network of Safety and Health Practitioner Organizations (Pryor et al. 2015), considers that OHS (Occupational Health and Safety) is an emerging profession that is often not well defined, locally or globally, the safety profession in France can now be considered as a structured and recognized profession: several curriculums and descriptions of the safety or risk profession exist (APEC 2017, AMRAE 2013). This doesn't mean that before this period safety wasn't managed by professional people in organizations or companies, *Ecole des Mines de Paris* was founded by King Louis XVI in 1783 to ensure the management of the mines of the kingdom. Two issues in mining business were important at the moment, economy and safety. The school disappeared at the beginning of the French Revolution but was re-established by decree of the Committee of Public Safety in 1794. So engineers were trained to the traditional engineering disciplines (mathematics, physics, mechanics and chemistry) but also social sciences were developed and delivered, as economics and management.

The past years, a keen interest in «*professionalization in/of safety*» has emerged in the safety science community (Gilbert 2015). The FONCSI (Fondation pour une culture de sécurité industrielle) has initiated in 2015 a strategic analysis on "*skills and competencies for industrial safety*" in which several international researchers are involved (Bieder, Journé, and Laroche 2015, Bieder et al. 2018).

Professionalization in/of safety can be seen two ways: the professionalization of persons who will become experts in safety (professionalization in safety – PiS) and the integration of safety into the general professionalism of employees (Professionalization of Safety – PoS). Although Andrew Hale has worked since the eighties (Hale et al. 1986, Hale 1995, Booth et al. 1991, Hale et al. 2005, Hale and Ytrehus 2004) on the topic of the professionalization of the profession (PiS), this renewed interest gives birth to new models and tools like the Australian OHS Body of Knowledge (Paul and Pearse 2016). This article will present a study of

the professional context of a post master safety program's alumni.

The general research questions we want to tackle are the following: what is the career path of alumni of a post master program in risk management? Can we define their daily professional environment? Which knowledge and competencies are actually used? What is their position and role in an organization? What is the contribution of the post master program to their professionalism? How to develop better education in risk management to bridge the theory to the professional real life context? And in a very general way, what could be the safety professionals' contribution to safety?

In this article, we will introduce the context and the research methods with some early results. First, we will differentiate *professionalization of* and *professionalization in* safety. For information, we will use the concept of "*safety professional*" in this article, as "*safety practitioner*" is a person who is vocationally-educated where the safety professional is university-educated (or has attained a similar level of higher education) (Pryor et al. 2015).

2 PROFESSIONALIZATION OF SAFETY

Industry recently wonders if the amount of effort put into safety training of employees was worth it (Gilbert 2015). Although a lot of time and money is consecrated to safety trainings, accidents still occur. The professionalization of safety, the way how workers integrate safety issues in their daily work, is subject of attention of researchers. The concepts of ordinary and extra-ordinary safety have been developed by Claude Gilbert (Bieder et al. 2018). A new way of safety training is also a topic of interest. It is clear that the safety professional has a major role to play as he is the referee person for safety in the organization. Thus we see a clear link between professionalization « in » and « of » safety. This makes us wonder about the place of the safety professional in his organization. Gilbert states that the difficulties for good safety training program in a company (PoS) are that they have to fulfill internal performance obligations and requirements and external justifications (accountability).

3 PROFESSIONALIZATION IN SAFETY

There is significant literature about safety professionals, Andrew Hale has done a lot of work since the eighties (see above), professional associations have also contributed to the subject (Pryor et al. 2015) and even in the sixties, people (scientifics) have written pertinent insights concerning the safety professional and his role in organizations that are still very actual today (Harper et al. 1962). Scientific and professional courses have emerged since the seventies (Arezes and Swuste 2012). The last decennium, in France the curriculums who are dealing with risk management, safety or HSE (Health Safety and environment) grew in a significant way. Private organizations in France propose rankings of (business) education programs and since several years the category « risk management » exists (www.eduniversal-ranking.com). Last November, the author of this article for ESREL was solicited to review an article for Safety Science. Chinese contributors examined the professionalization of safety in China: « *Development of safety science in Chinese higher education* ». The last decade, we can observe a development of education of safety professionals in all industrialized countries.

Dekker investigated the role shaping factors about the safety profession (Provan et al. 2017). This work is interesting because it can give an idea of the importance on a safety professional in his organization and the way how he can influence, model or create safety in the organization. This is a topic that has to be investigated and taught to future safety professionals. This present study tends to bring some answers to that question also. The following part presents our study object, the post master degree industrial risk management.

4 POST MASTER DEGREE INDUSTRIAL RISK MANAGEMENT

In 1997, the top management of MINES ParisTech decided to launch a research department dedicated to risk and crisis with the mission of developing studies in cooperation with industrial partners and training activities. In 2002, the decision was taken to build a post-master program for training HSE professionals for the industry.

The post master degree ("*Mastère Spécialisé*" in French) Industrial Risk Management was launched in 2004 on demand of French industry and in collaboration with Tongji University of Shanghai (China) to form, besides French students, also Chinese students on risk management for French companies developing in China.

The design of the curriculum was achieved from an analysis of academic literature and current practices, in close cooperation with companies' representatives, in order to fit the contents of the training and the pedagogical methods to the needs of the industrial sector. The program has known several important changes in organization and content while its existence. In the actual form (promotion 2017–2018), it contains 500 hours of courses during one year (begin October to end September). The rest of the time, students realize

a professional mission with an industrial partner. The program forms future safety professionals in the occupational, environmental and industrial (process safety) field. The curriculum contains six main tracks:

- Safety regulations,
- Hazard and risk assessment,
- Safety Management Systems,
- Human and organizational aspects,
- Management and leadership aspects,
- Emergency and crisis management.

Lectures, exercises and practical work are provided and supervised by a faculty group composed of an equal number of academics and professionals.

Currently, 31 students attend the program, almost all of them alternate academic training and apprenticeship. This apprenticeship program consists in 5 weeks courses, 3 weeks enterprise, 3 weeks courses, 4 weeks enterprise, 3 weeks courses, 2 weeks enterprise, 1 week study trip, 5 weeks courses and 6 months enterprise followed by the defense in front of a jury of their professional project. The program in its current form contains six main tracks, a study trip (United Arab Emirates in 2017) and a professional conference, organized end of March by the students themselves.

The program has known an evolution from a «double degree or double competence» point of view to a program that forms «specialized» safety professionals. The students that apply for the program now have, for the great majority of them, already a Master degree in Health, Safety or Environment. The MS MRI makes them more "specialized", whereas in the past, students had degrees in chemistry, engineering, psychology or law. This is also the result of the increase in HSE Master Programs and the result of the demand of industry. Industry wants a student already trained in HSE for their professional missions and collaborations. This eliminates from the program the students who have no knowledge of risk and/or HSE.

The scientific transfer of HSE findings to the industry takes some time. HSE practitioners often lack the knowledge of the genesis and relations between HSE theories and methods, models and metaphors developed since World War 2 (Swuste, Gulijk, and Zwaard 2010, Swuste et al. 2014). It is also important to know the evolution of the HSE job. The integration of a HSE professionals' training program in a dedicated research laboratory (CRC—MINES ParisTech) facilitates this knowledge transfer and informs the future HSE professionals of the latest advances in the field. They will probably keep an intellectual curiosity all the rest of their career and hopefully will participate in co-developing tools using the latest safety models (Besnard et al. 2009).

Post master programs are known to be «*professionalizing*», the student will pass from the state of student to be a professional. Several characteristics of the MRI program tend to influence this transition: participation and courses of a lot of safety professionals, students pass a great deal of the program in the organization of their industrial partner, projects and case studies are real industry problems (with participation of industry), executive summaries are demanded on several topics, they work in teams, they do a lot of presentations, they learn to present their work to decision makers, the courses on "management and leadership" makes them aware of the importance of soft skills on the work floor.

The program has grown since 13 years to be more and more professionalizing. But finally, what is the contribution of the post master program to their professionalism? How to develop better education in risk management to bridge the theory to the professional real life context? For what kind of professional context do we prepare them? The next part will present the missions of a safety professional.

5 MISSIONS OF THE SAFETY PROFESSIONAL

Several studies on the missions of safety professionals have been done. Wybo and Van Wassenhove (Wybo and Van Wassenhove 2016) present the missions of safety professionals (Table 1), the needed skills to do the job and modeled the MRI program to correspond. They conclude by proposing requirements for a HSE professional training curriculum: students getting their degree in HSE must be fully operational and demonstrate their professionalism when they start their first job. So the curriculum must prepare them to their future work. From the literature review and the analysis of a HSE professional's job content, they identified the important matters our HSE professional's curriculum must address:

- Domains to teach:
  - Regulations and HSE management systems,
  - Hazard and Risk analysis in occupational health, system safety and environment,
  - Human and organizational matters,
  - Emergency and crisis management,
  - Communication, management and leadership,
- A strong implication of safety professionals,
- The use of realistic case studies,
- Interactions with industry practitioners:
  - "Field work" in industrial sites,
  - Annual conference for industry practitioners,
  - A long internship/presence in a company,

Table 1. Missions of safety professionals in literature (Wybo and Van Wassenhove 2016).

Job content topic	ENSHPO 2005	AMRAE 2013	ASSEF 2007	Wu 2011	INRS 2004	De Joy 1993	Hale 2004	Kohn 1991	Total
Advising management and decision makers	++		+			++	+		**6**
Definition of the missions and the organization of the safety management system	++	+	+			+	+		**6**
Risk management (hazard identification, evaluation and control)	+++	+	+++	+	++	+++	++	+++	**18**
Regulatory compliance	++		+	+			+	++	**7**
Diffusion of safety culture and culture change	++	+		+			++		**6**
Training and communication	++		+	+	+	++	+	++	**10**
Accident and incident investigation	++		+		+	+	+	++	**8**
Emergency and crisis management	++	+	+	+	+			++	**8**
Monitoring and reporting	++	+	++	+	+	+	+	++	**11**
Knowledge management	++			+	+				**4**
Insuring and costing risks	++	+		+					**4**

Nevertheless, questions still remain on other aspects of their professional context. What are the knowledge and the competencies they mobilize? What is their position and role in an organization? What is their vision of safety and what can be their contribution to safety? A first step is to look closer to what alumni of a post master program in risk management do once graduated. We developed a methodology to search for some first answers.

6 METHODOLOGY: GENERAL QUESTIONNAIRE

The post master program MRI has now about 200 alumni. To investigate the professional context of the safety professional, we dispose of exactly 197 persons to consult.

A similar study on the context of a profession has been conducted in the Netherlands (Corporaal et al. 2016). They investigate the Human Resources (HR) professional and they use the following methodology: several students (544 students participated to the study) have asked HR professionals (571 persons), their hierarchical boss (542 persons) and a line manager (553 persons) to fulfill a questionnaire. That questionnaire was used afterwards as guide to do in-depth interviews with those persons. This study is repeated yearly (!). Every four years, a synthesis is done and an education profile is created. This profile enables to adapt the curricula of the HR programs. The research themes are the missions of the HR professional, role and position of the HR professional, competencies needed for the HR professional and the relations with line management.

In the field that interests us, the OHS Professional Capability Framework is a framework for practice developed by the International Network of Safety and Health Practitioner Organizations (Pryor, Hale, and Hudson 2015). INSHPO differentiate competence and capability. The difference between competency and capability is that competency is about delivering the present based on the past, while capability is about imagining and being able to realize the future. Capability goes much further than competency, it's also about confidence and adaptability. For our study, we use "competency". With the first results, we will see and judge if there is a need to differentiate between competency and capability.

From the literature, personal work and observations, a general questionnaire was developed. This development was done with the support of co-researchers (Gilbert De Terssac, sociologist, research director CNRS) and was tested on some alumni. The remarks were integrated into the questionnaire. The questionnaire was proposed to the alumni through an on-line version. Each alumnus was personally contacted mid November 2017 by email with the request to take some time to answer the questions. The time to respond is set to two months. Several solicitations are done, the aim is to have at least 100 responses. To answer all questions, the participant has to take at least 30 minutes of his time and has to do some critical analysis of his work situation. Answering the questionnaire takes a real effort.

The questionnaire is structured as followed:

- questions about the identity
- career description
- job description
- benefits of the MRI program for the professional career
- comments

The job description is the most important for our research questions and is composed of twelve questions. Four questions are very important, we will present them here.

One question concerns the missions of the professional. From a synthesis from the literature, personal observations and discussions with safety professionals, fourteen missions are proposed (Table 2). For each mission, the respondent has to notify on a scale going from 1 (not important) to 10 (very important) the importance of the mission in his current job, according to several levels:

- according to his or her personnel conviction
- according to the organization in which he or she is working (his superiors)
- according to the real time he or she spent on the mission

In this way, we can identify discordances between the professional and his organization.

The second important question concerns the characteristics of the professional's job (Table 3). Sixteen descriptions of work characteristics are proposed. The professional is asked to score 1 (not agree) to 10 (totally agree) for each situation.

The third question is about the concept of safety. Twenty-two ideas are formulated about what contributes to safety (Table 4). The professional has to score (1 not important and 10 very important) the ideas according to his personal conviction and

Table 2. The fourteen missions of the safety professional.

Missions of the safety professional
1. Inform and advice the direction
2. Organize the safety management system
3. Follow up of regulations
4. Control compliance with regulations
5. Manage risks (identification/evaluation/treatment)
6. Develop safety culture
7. Inspections on the work floor, organizing information collection (bottom up)
8. Train and communicate about safety and risk
9. Analyze incidents and accidents
10. Manage crises and emergencies
11. Measure performance: monitoring and reporting
12. Give expertise for a specific type of risk
13. Manage the insurance and financial side of risks
14. Realize business continuity plans

Table 3. Job characteristics description.

Characteristics of the job
1. Presence of an administrative workload and an important "safety bureaucracy"
2. A big diversity of objects and subjects to handle
3. A lot of subjects to handle with urge
4. No time to treat subjects in depth
5. A lot of traveling
6. A lot of simultaneous missions with different deadlines
7. Recognition of the direction
8. Recognition of middle management and workers
9. Feeling of being isolated in an organization that is not much concerned by safety
10. Interruptions and adaptations of planning
11. Difficulties to conciliate safety prevention with regulatory compliance, requirements of inspection of the administration and the politics of the organization
12. Interactions with a lot of stakeholders
13. Lack of means
14. Necessity to adapt constantly the rules (theory) to the reality of the work floor (practice)
15. Frequent interactions with the work floor
16. Autonomy and liberty of initiative

Table 4. Safety "ideas".

Composantes de la sécurité
1. Engagement of the direction
2. Engagement of middle management
3. Compliance with regulations
4. Safety culture
5. Safe technical conception
6. Bottom up consultation
7. Consultation of administration and inspection
8. Consultation and information of the neighborhood living near the company
9. Presence of a Safety Management System
10. Safety training
11. Team spirit and cooperation on the work floor
12. Respect of procedures
13. Sharing of good practices
14. Sharing of information on accidents and incidents
15. Comprehension and modeling of hazards
16. Presence of a degree of liberty in interpreting procedures in function of the context
17. Analysis of root causes after an accident
18. Considering the variability of human performance by analyzing the work conditions
19. A good technical knowledge of the system by the stakeholders
20. Human error analysis
21. Give sense to work: organize debates on work situations
22. Organize crisis and emergency exercises

according to what he thinks is the conviction of his organization.

The fourth question (Table 5) concerns knowledge and competencies (skills or even capabilities) mobilized by the professional and the competencies that he or she would like to mobilize more. It should be noticed that knowledge and learning can be characterized at six levels. These levels are based on the well-known Bloom taxonomy (Murtonen, Gruber, and Lehtinen 2017) and go from 'simple' remembering to 'elaborated' evaluating. But we can consider the levels of analyzing, synthetizing and evaluating as hierarchical equal levels.

- Remembering (knowledge)
- Understanding (comprehension, translating, interpreting or extrapolating information)
- Applying (using principles or abstractions to solve novel or real-life problems)
- Analyzing (breaking down complex information or ideas into simpler parts to understand how the parts relate or are organized)
- Synthesizing/creating (creation of something that did not exist before)
- Evaluating (judging something against a given standard)

A safety professional who is fully competent is expected to operate minimum at level three (applying knowledge) for every knowledge category (Pryor, Hale, and Hudson 2015).

Skills can be categorized in three sections: personal skills (example: verbal communication), professional practice skills (example: problem solving and critical thinking) and professional technical skills (example: implements tools to assess risk). We chose to list in table four knowledge and skills but with differentiating neither the type of skills nor the taxonomy of Bloom. The last item of the list, 'ethics', is rather a personal value than a skill. Those aspects will be studied more in depth when doing interviews with selected people (see below).

The next paragraph will present some first results and a discussion about how to complete this methodology (the questionnaire).

Table 5. Knowledge and competencies.

Knowledge and competencies
1. Knowledge of regulations
2. Knowledge of hazards
3. Knowledge of the technical system
4. Knowledge of accident models and theories
5. Knowledge of risk analysis methods
6. Knowledge of business management: finance, human resources, organization, technologies, innovation…
7. Competency of applying risk analysis methods
8. Competency of « constructing » a risk analysis methodology adapted to the problem
9. Competency of crisis and emergency management
10. Competency of situation/activity/accident analysis
11. Competency of translation of politics into real actions
12. Competency of translation of accidents statistics into action plans
13. Competency of strategic planning
14. Competency of hierarchizing problems
15. Competency of looking for the right information and learning
16. Competency of critical analysis
17. Competency to propose solutions
18. Competency of written and spoken communication
19. Competency of synthetizing and adapting of wording to the public addressed
20. Competency of organizing personal work
21. Competency of listening
22. Competency of leadership
23. Competency of being exemplary
24. Competency of teamwork
25. Competency of working in an multicultural organization
26. Competency of using new communication technologies
27. Ethics

7 FIRST RESULTS, DISCUSSION AND PERSPECTIVES

The questionnaire is on-going and at this date, 61 alumni of the 197 persons contacted have responded and completed the questionnaire. To complete the questionnaire (the majority of questions are compulsory) at least 30 minutes and a great effort of reflecting on professional practices is needed. Another 50 respondents started but didn't finalize the questionnaire yet. The aim is to have about 70% of the alumni. Each alumnus will be contacted individually, and several times if necessary.

Several biases are present for this study. We will discuss the most important ones. We find that the youngest alumni completed in priority the questionnaire. The elder ones 'lost' probably the implication with their former school and are less motivated to respond. Also, several contact addresses of the elder alumni are obsolete. A work to trace them on Linked In is ongoing. The Chinese alumni are also difficult to contact. This could give a slightly distorted image because we will have only data from young professionals: the description of the professional context of a junior safety professional.

The curriculum of the program has been modified all along its existence. Several important modifications have been done: in 2010 an important content modification, in 2014 the introduction of alternate training and apprenticeship. The

respondents will evaluate a program that was not the same for all the alumni. We hope that the first results will give us insights on the importance on professionalization of the introduction of the alternate training and apprenticeship with an industrial partner in the curriculum.

A complementary methodology will allow us to gather more data. The first results will able us to organize focus groups with alumni to discuss several aspects or findings. A couple of monographies will be done also: a safety professional, his superior and his colleagues will be interviewed and observed in their daily work. It could be interesting to interview also line managers who are in contact with the safety professional.

The findings of this work will able us to:

- Have an idea of the careers of our alumni (company, wages, function,…)
- Have an idea of the professional context of our safety alumni (missions, job profile, organizational culture…)
- Have an idea of the knowledge and competencies they mobilize
- Have an idea of the impact of the MRI program on their professionalism

The perspectives of this work are of different nature. On one hand we continue to implement changes and modifications and we improve the program of MRI to prepare the students to the "real life" context of the safety professionals' job. This is managed by new pedagogical approaches: role plays, case studies, close interactions with safety professionals and collaboration of *Safety Scientifics* and *Safety Professionals* together on the same project. We discussed in a former paper (Wybo and Van Wassenhove 2016) the pedagogy used for education of safety professionals. The present study will be able to bring some data for improving the pedagogic approaches of the program.

In order to improve the program, we must absolutely question ourselves on the ways we evaluate the program. The four-level training evaluation model of Kirkpatrick (Gilibert and Gillet 2010) is widely known. We could think of integrating a yearly survey of alumni and/or interview and observation of alumni in work conditions in a general evaluation plan based on Kirkpatrick's model. Kirkpatrick proposes four levels of evaluation of learning: reaction (satisfaction of learners), learning (what knowledge do they acquire, behavior (do they use the new knowledge in work situation) and results (outcome for the organization). Paul Swuste (Swuste and Sillem 2018) is interested in the quality approach to evaluate a post academic course 'management of safety, health and environment'. Having a feedback of alumni is not enough to have a global evaluation of the program.

What is the goal to form and educate students in safety curricula? Teaching them theoretical risk analysis methods (HAZOP, FMECA, QRA...)? Preparing them to the real world problems when applying those approved methods? Maybe the role of an education program is also to form "new" professionals who are importing new (scientific) insights into the profession. They are the bridge between science and practice. Probably, it is all this and even more.

Last but not least, in our understanding of safety in organizations, the role of the safety professional is very important. David Provan (Provan, Dekker, and Rae 2017) did a literature review of the factors shaping the role of a safety professional (bureaucracy, influence and beliefs). This present study will certainly contribute to identify some factors. Being aware of those factors will surely enable future safety professionals to be more effective in their job: bringing more safety to our world. And this is exactly the fourth level of Kirkpatrick's evaluation model.

REFERENCES

AMRAE. 2013. "Referentiel Metier Risk Manager."

APEC. 2017. "Les Offres d'emploi Cadre Dans Le Domaine Des Risques Industriels." *Les Études de l'emploi Cadre*, February.

Arezes, Pedro M., and Paul Swuste. 2012. "Occupational Health and Safety Post-Graduation Courses in Europe: A General Overview." *Safety Science* 50 (3):433–42. https://doi.org/10.1016/j.ssci.2011.10.003.

Besnard, Denis, Damien Fabre, Wim Van Wassenhove, and Eduardo FA Runte. 2009. "An Account of Scientific Transfer to the Industry: The Co-Development of an Incident Analysis Tool." In *9th Conference of the European Sociology Association.*

Bieder, Corinne, Claude Gilbert, Benoît Journé, and Hervé Laroche, eds. 2018. *Beyond Safety Training. SpringerBriefs in Applied Sciences and Technology.* Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-65527-7.

Bieder, Corinne, Benoît Journé, and Hervé Laroche. 2015. "Analyse Stratégique Professionnalisation et Sécurité: Méthode & Pistes de Réflexion."

Booth, R.T., Andrew Richard Hale, and S. Dawson. 1991. "Identifying and Registering Safety Practitioners." *Safety Science* 14 (3):231–240.

Corporaal, S., M. Vos, T. Morssink, S. Peters, and N. van Dartel. 2016. "De HRM Praktijkmonitor." Samenwerkende lectoraten HRM, Holland.

Gilbert, Claude. 2015. "La Sécurité: Une Affaire de 'Professionnels'?" *Tribunes de La Sécurité—FONCSI,* no. N° 6 (June).

Gilibert, Daniel, and Isabelle Gillet. 2010. "Revue Des Modèles En Évaluation de Formation: Approches Conceptuelles Individuelles et Sociales." *Pratiques Psychologiques* 16 (3):217–238.

Hale, A.R. 1995. "Occupational Health and Safety Professionals and Management: Identity, Marriage, Servitude or Supervision?," no. 20:233–45.

Hale, A.R., G. Bianchi, G. Dudka, W. Hameister, R. Jones, P. Perttula, and I. Ytrehus. 2005. "Surveying the Role of Safety Professionals: Objectives, Methods and Early Results." *Safety Science Monitor* 9 (1). http://www.enshpo.eu/userfiles/Safety%20Science%20Monitor%202005(2).doc.

Hale, A.R., M. Piney, and R.J. Alesbury. 1986. "The Development of Occupational Hygiene and the Training of Health and Safety Professionals." *The Annals of Occupationnal Hygien* 30 (1).

Hale, A.R., and I. Ytrehus. 2004. "Changing Requirements for the Safety Profession: Roles and Tasks, Journal of Occupational Health and Safety." *Journal of Occupational Health and Safety,* no. 20:23–36.

Harper, George W., Thomas H. Rockwell, and D.A. Weaver. 1962. "The Safety Engineer – What Do We Expect of Him? – What Knowledge Does He Need? – How Do We Upgrade His Performance?" ASSE Journal VII (2).

Murtonen, Mari, Hans Gruber, and Erno Lehtinen. 2017. "The Return of Behaviourist Epistemology: A Review of Learning Outcomes Studies." *Educational Research Review* 22 (November):114–28. https://doi.org/10.1016/j.edurev.2017.08.001.

Paul, Gunther, and Warwick Pearse. 2016. "An International Benchmark for the Australian OHS Body of Knowledge (BoK)." *Safety Science* 81 (January): 13–24. https://doi.org/10.1016/j.ssci.2015.07.016.

Provan, David J., Sidney W.A. Dekker, and Andrew J. Rae. 2017. "Bureaucracy, Influence and Beliefs: A Literature Review of the Factors Shaping the Role of a Safety Professional." *Safety Science* 98 (October):98–112. https://doi.org/10.1016/j.ssci.2017.06.006.

Pryor, Pam, A.R. Hale, and D. Hudson. 2015. "The OHS Professional: A Framework for Practice – Role, Knowledge and Skills." International Network of Safety and Health Practitioner Organisations (INSHPO). Park Ridge, IL, USA.

Swuste, Paul, Coen van Gulijk, and Walter Zwaard. 2010. "Safety Metaphors and Theories, a Review of the Occupational Safety Literature of the US, UK and The Netherlands, till the First Part of the 20th Century." *Safety Science* 48 (8):1000–1018. https://doi.org/10.1016/j.ssci.2010.01.020.

Swuste, Paul, Coen van Gulijk, Walter Zwaard, and Yvette Oostendorp. 2014. "Occupational Safety Theories, Models and Metaphors in the Three Decades since World War II, in the United States, Britain and the Netherlands: A Literature Review." *Safety Science* 62 (February):16–27. https://doi.org/10.1016/j.ssci.2013.07.015.

Swuste, Paul, and Simone Sillem. 2018. "The Quality of the Post Academic Course 'management of Safety, Health and Environment (MoSHE) of Delft University of Technology." *Safety Science* 102 (February): 26–37. https://doi.org/10.1016/j.ssci.2017.09.026.

Wybo, Jean-Luc, and Wim Van Wassenhove. 2016. "Preparing Graduate Students to Be HSE Professionals." *Safety Science* 81 (January):25–34. https://doi.org/10.1016/j.ssci.2015.04.006.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Reversing the trend through collaboration in the petroleum industry

K. Skarholt & G.M. Lamvik
SINTEF Technology and Society, Trondheim, Norway

ABSTRACT: The purpose of this paper is to discuss the importance of bipartite and tripartite cooperation in the Norwegian petroleum industry and how it contributes to improve safety conditions. Lack of trust between the parties and pressure on the Norwegian model has great attention in this industry today, related to major cost reductions and downsizing the last 2–3 years. We discuss these challenges and how to re-build trust between the parties. The empirical material is mainly based on qualitative data from the ongoing four year RISKOP research project (Managing Risks in Offshore Operations). In addition, the data is based on a document study about the safety conditions and collaboration in the petroleum industry.

1 INTRODUCTION

In the Norwegian oil and gas industry there is a long tradition for employee participation at work, both at an individual level, and through formal bipartite and tripartite collaboration, inviting for employees' ideas and concerns about safety issues. A tripartite collaboration consists of the authorities, employers' association and worker unions, while bipartite collaboration is a local collaboration between employer/management and worker unions within a company (Levin et al., 2012).

Employee participation is about involvement in work-related matters, with the intentions to have influence on working conditions. The term employee voice is an expression of participation at work, and is defined as an employee's discretionary communication of ideas, suggestions, concerns, or opinions about work-related issues with the intent to improve organizational functioning (Morrison, 2011:375). The contribution of employee voice is to influence decisions and contribute to improvements in a company. According to safety, it is crucial that employees' opinions and voices are listened to, if not it may lead to dangerous safety conditions and accidents. However, the conditions for employee voice are challenged in periods with economic crisis (Farndale et al., 2011; Skarholt et al., 2017). Economic crisis and downsizing has led to a more fragile bipartite collaboration where the parties and authorities experience decrease in trust. Increased economic pressure has led to concerns about possible negative effects on safety and work environment. Figures from the latest study of trends in risk level in Norway's petroleum activity (RNNP 2016) from the Petroleum Safety Authority Norway (PSA), shows that more employees report higher work pressure and less influence on HSE than in previous RNNP studies. This negative development has led to campaigns and initiatives from both the PSA and the Ministry of Labor and Social Affairs in Norway.

Through the campaign "Reversing the trend", the PSA address what they see as a worrying development over the past years (PSA, 2016). How to actually improve bipartite and tripartite collaboration is one of the main issues in this initiative. According to PSA good collaboration between employers and employees has helped to boost the level of safety in Norway's petroleum industry. This interaction now appears to be under pressure. PSA stress the importance of employee participation in handling safety matters, stating that involvement of employees is a requirement in all phases of the petroleum sector for every issue which relates to safety. These rights and duties are to be practiced both directly by each individual worker and through union representatives and safety delegates. A good collaboration between the parties depends on mutual trust.

The Ministry of Labor and Social Affairs in Norway appointed a HSE committee representing authorities, employer organizations and employee unions in the petroleum sector in 2016, to discuss and consider the state- and development of HSE conditions in the Norwegian petroleum industry. The background for this was a negative safety trend in 2015 and 2016 with serious conditions and safety challenges. The report from this work (HSE committee—Ministry of Labor and Social Affairs, 2017) refers to that there have been major change processes with downsizing and restructuring that may be a challenge to the established collaboration between employers and employees. Indications

suggests that cooperation between the parties is more fragile compared to earlier, although disagreement prevails between the parties over how far this collaboration has come under pressure. A main conclusion in the report from 2017 is to keep and build mutual trust and respect to each other's role and responsibilities between the parties—to be able to take care of and develop the safety level in the petroleum industry.

The aim of this paper is to discuss the role of bipartite and tripartite collaboration due to develop and improve safe operations in the petroleum sector in Norway. We discuss how employee participation at work influence on safety conditions—as a mean of reducing risk of injuries and major accidents.

2 THEORETICAL BACKGROUND

2.1 *Collaboration, trust and safety*

The Norwegian oil industry, or rather the oil industry on the Norwegian Continental Shelf has been characterized by its social agreements grounded in the Norwegian model, producing alliances between all core stakeholders, thereby a "we" including the whole of the industry. The Norwegian model on a local company level is about the collaboration between management and union representatives. In Norway, this cooperation has been characterized by; trustful relations, willingness to collaborate for increased competitiveness, low level of conflicts at work (Levin et al., 2012). The voice of employees has thus to a large extent been listened to and have influenced over decisions made in the company.

Improved safety is something that all parties want for the industry. The Norwegian petroleum sector has been characterized by trustful bipartite and tripartite collaboration, and the participation among employees has been important for improving safety conditions. Important arenas for tripartite collaboraton is "Working Together for Safety" (Samarbeid for Sikkerhet/SfS) and Safety Forum (Sikkerhetsforum). The aim of their work is to increase safety in the petroleum industry and strengthen trust and cooperation among the actors of the industry. These arenas are central for cooperation among the parties in the industry and the authorities as regards health, safety and environment in the petroleum activities offshore and onshore. Here, the unions, the employers' organizations and the authorities have a significant influence on the safety agenda in this sector. One could say that this trust-based tripartite collaboration is a key cultural value related to how safety is maintained in the Norwegian petroleum industry. Trustful bipartite collaboration about safety matters have also been an important cornerstone of the safety regime. Bipartite collaboration is an integrated and critical part of the regulatory regime for HSE in the Norwegian petroleum sector (Rosness & Forseth, 2014). To report about dangerous safety situations and conditions is easier to obtain in a bipartite collaboration, where you can have an open relation between leader and employees/unions. The economic crisis in this sector has however put the collaboration between the parties under pressure. Indications from several actors claim that the Norwegian model with bipartite and tripartite collaboration is under pressure (PSA, 2016; Ministry of Labor and Social Affairs; Falkum et al., 2017). Safety Forum was established in 2000, based on distrust from unions to the employer association Norwegian Oil and Gas—stating that they constantly failed to include the employees in decisions concerning safety (Rosness & Forseth, 2014). Compared to the situation in 2000, we are seeing a similar development today with decrease of trust between the parties.

Trust in organizations has been studied in different ways to address positive outcomes on organizational phenomena, such as positive impact on safety culture and safety performance (Burns et al., 2006; Conchie et al., 2006; Reason, 1997). Trust also affect improved communication, knowledge sharing, commitment, and organizational learning (McEvily et al., 2003; Nonaka and Takeuchi, 1995).

Research has shown that the cultural aspects of work practice influence safety as much as technology and formal organization structures (Antonsen, 2009; Guldenmund, 2000; Haukelid, 2008). Also, the work from Tharaldsen (2011), addressing differences in safety climate and trust between UK and Norwegian sector, fits this picture.

In high reliability-organizations (HRO) organizational culture/safety culture influence on safety (Weick and Sutcliffe, 2007). The key aspect of building safety culture is the level of openness and trust and access to information that may indicate compromising of safety. Reason (1997) argues that the safety culture is based on an underlying element of trust, and research shows that high levels of trust in relationships contributes to high levels of safety in high risk enterprises (Conchie et al., 2006). Their study of safety performance in the offshore industry concluded that the impact on trust and distrust on safety performance is determined by the act of being trusted (or distrusted).

Trustful relations and openness requires the existence of a variety of channels, both formal and informal. Bipartite cooperation is a relation between managers and union representatives (and safety delegates) in a company. The Norwegian Working Environment Act from 1977 regulates

the rules for formal participation at work among union employee representatives, where they have influence through information- and discussion meetings with the line management. They thus take part in problem-solving in different matters at work, such as improvement of safety, change processes and similar.

A good bipartite collaboration demands good leadership, listening to the ideas and concerns from the employees. How the dialogue and trust is between the parties will influence on how the employees are involved and have influence on their work and safety matters at work. The presence of union representation at work contributes to increased individual employee participation (Trygstad & Hagen, 2007). The relations between managers and union representatives at work will influence on employee participation outside the formal bipartite cooperation, where an involving leadership style will strengthen open communication. To make individual employees actually speak up about safety concerns, leaders must invite to openness and involvement about safety among the employees. How managers respond to employees' opinions about safety improvements, will influence on the motivation and willingness to speak up. If they signal interest and willingness to act on employee voice, the employee's motivation to inform about safety concerns are enhanced (Detert & Burris 2007; Edmondson, 2003). Detert and Burris (2007) found that management openness and transformational leadership behaviour are consistently positively related to voice. To speak up involves sharing one's idea with someone who has the power to devote organisational attention or resources to the issue raised (French & Raven, 1959). To openly speak up about work environment and safety concerns at a work place without fear of being sanctioned/punished, is a prerequisite for reporting (Antonsen, 2009; Trygstad et al., 2014). Mutual trust between managers and employees also influence to which extent the employees will tend to keep silent or use their voice when safety concerns occurs (Skarholt et al., 2017).

3 METHODS

To highlight these issues we have drawn upon various sources. An important one is the project RISKOP (Managing Risks in Offshore Operations). This project has been run by Western Norway University of Applied Sciences in Haugesund, in collaboration with SINTEF and University of Stavanger, supported by the Norwegian Research Council and nine companies in the industry. In this project, we interviewed 16 managers from five different shipping companies. These shipping companies, that constitutes an important part of the petroleum cluster in Norway, are operating advanced vessels (e.g. supply vessels and anchor handling vessels) in the offshore petroleum industry, working for different oil and gas companies at the Norwegian Continental Shelf. Also, a broker, the Norwegian Maritime Directorate and trade unions were interviewed with topics pivoting around issues as: How the informants/shipping industry experience the crisis; What they actually do to meet and handle the situation, and; How the situation may affect safety operations offshore?

Besides the RISKOP project we have analysed some documents covering certain aspects of safety and collaboration at work. First, the Petroleum Safety Authority (PSA) and their campaign "Reversing the trend" (2016) has played an important basis for this topic. Second, in the extension of this campaign we have made use of the report "HSE in the petroleum industry" (HSE committee—Ministry of Labor and Social Affairs, 2017) to shed some light on the safety situation in the petroleum sector. Third, the survey "Participation Barometer" (Medbestemmelsesbarometeret) (Falkum et al., 2017) measure; "Is participation in Norwegian working life under pressure?" Moreover, participation as is described as main elements of leadership- and managerial practice is the focus. The survey is owned by six trade unions, covering private- (included the oil and gas industry) and public sector in Norway.

4 RESULTS

We present 1) results from the interviews with leaders in shipping companies, union leaders and broker—conducted in the RISKOP project, and 2) results from document analysis about the trend and development of safety in the Norwegian petroleum industry.

4.1 *Strong bipartite cooperation—to survive*

The economic crisis in the petroleum industry has led to fewer jobs in the offshore shipping industry, where the competition for jobs is tough in a marginal market. One of the companies we interviewed had reduced their staff with 400 employees, and feared further layoffs. Consequently, shipping companies have removed a considerable portion of the offshore fleet from the market. In December 2017, there are 134 vessels from the offshore fleet in layup, out of a total fleet of around 550 vessels. Ordinary Platform Service Vessels (PSV) were the largest category of these ships (61), while Anchor Handling Tug Supply (AHTS) were 42 vessels and the number of Seismic vessels were 14.

This layup activity has resulted in downsizing of personnel for oil companies, shipping companies and subcontractors. Another challenge for this industry, is the fewer long-term contracts compared to before the crisis. Today, most of the contracts are in the spot market, meaning short-term assignments with a maximum of one month, especially operations done by PSV's (Platform Supply Vessels) and anchor handling vessels. Leaders from shipping companies we interviewed accepted contracts they knew to be too low, not even covering basic running costs, solely to decelerate the decline of the company. When the market is weak and undergoing a crisis as today, there are many subcontractors that are weakened and not in a position to negotiate.

We find that one way to deal with this crisis is actually to fight for the survival of the work place—together in a bipartite cooperation. As a coping strategy in the offshore shipping industry, collective local organizational arrangements have been strengthened. Our material show proves of strong and solidarity constellations inside the shipping companies, e.g. close collaboration between managers and employee's in finding new solutions to handle the crisis in the industry. It seems that the major challenges in the offshore sector has made all the staff in the companies, to realize that they have a common interest in collaboration and find shared solutions. They realized that this is the time for dancing rather than boxing, to paraphrase the famous book by Huzzard et al. (2005).

Many of the shipping companies we interviewed pointed out that the unions and employees was willing to find ways to save money in the company, such as reducing salaries for a period. One of the shipping companies in our study reduced the wages by 29 percent among the crew from laid off vessels, so they could keep more of their staff and the competence in which they held. Cut in company specific bonuses was another instrument to reduce costs. Another example was change in the shift system; from four to five shifts, to be able to keep more of the employees working. This allowed for an extra shift, or crew onboard the ships, allowing extra leisure time at home. This way, the economic crisis made the cooperation between the management and the employees/union representatives really strong. As one of the leaders from a shipping company said; *"New solutions to survive is established because of the unions"*. Problems and crisis become like an outer enemy that may build strong alliances between employer and employees in an organization. Labor relations become a positive force, building trust and a good working environment. We see that this is what happens in the shipping industry. Other means of survival in the Norwegian shipping market has been mergers of shipping companies the last years.

4.2 *The Norwegian model under pressure*

On the other hand, there are many indications that the climate for collaboration between the parties has become worse in recent years. Increased economic pressure has led to more concerns about possible negative effects on safety in the petroleum industry in Norway. We have made analysis of documents describing the development and trend about safety in this industry. The analysis is mainly based on these documents; Reversing the trend (PSA, 2016), HSE in the petroleum industry (HSE committee—Ministry of Labour and Social Affairs, 2017) and The Participation Barometer (Falkum et al., 2017).

Five decades after the Norwegian oil adventure began, the petroleum sector faces important safety challenges. Trends are moving in the wrong direction in a number of areas (PSA, 2017). The development over the past two years have involved safety challenges and serious conditions: Figures from the 2016 study of trends in risk level in Norways's petroleum activity (RNNP) show an increase in serious hydrocarbon leaks and well control incidents. The major accident indicator is evaluated to be at a too high level (PSA, 2016). It is a reason to believe that this situation is affected by the economic crisis with restructuring and downsizing. Changes and demands for greater efficiency raise the level of conflicts.

To get safety development back on the right track, PSA have started a campaign; "Reversing the trend". PSA has put *collaboration* as one of the main issues in Reversing the trend: *"Collaboration between the various sides in the petroleum sector is under greater pressure, both between companies and unions and between them and the government. Such bi- and tri-partite interaction occupies a key place in Norwegian safety efforts."* PSA's concerns is that a weakened cooperation could include a poorer basis for important decisions by company managements, and weaker entrenchment with employees of important choices for the way forward. They are worried about that weaker employee participation may have negative consequences for safety in the petroleum industry. Numerous examples from major accidents in the petroleum industry show that information that could have prevented the accidents, was available, but was either not communicated or not acted upon. This indicates a safety culture lacking openness and trust for reporting and telling about dangerous situations. A key aspect of safety culture is the level of openness and access to information that may indicate compromising of safety. PSA's focus on bipartite cooperation is to emphasize the role industrial relations has on safety, where they address the responsibility of improvements towards the management

in oil and gas companies. PSA want to remind the management how safety can be taken care of and improved through formal collaboration between employer and employees. The voice of employees in decision-making is under pressure, and so the Norwegian model is under pressure.

A leader from one of the major trade unions in the petroleum sector we interviewed, stated that he/the union did not recommend employees to involve themselves in union activities at the moment, since it has become very troublesome to ask for leave for the individual representatives to involve themselves in trade union work. This opinion from such a strong voice in the industry can be seen as a barometer or an indication of a lack of trust or lack of collaboration in the offshore industry.

The report from the HSE committee—Ministry of Labor and Social Affairs (2017) "HSE in the petroleum industry" also emphasize collaboration between the parties—to enhance safety conditions in the petroleum industry: *"Bipartite and tripartite collaboration is an important cornerstone of the safety regime, and must be strengthened and further developed."* Participation and influence among unions/employees about safety in this industry has been given high priority for many years, and has influenced positively on safety results offshore and onshore. The Working Act law define the rights among employees to speak up and participate at work. There are both formal and informal arenas for cooperation between the employees' organizations and the employer organization, with collective agreements regulating the bipartite cooperation. The work group with participants for employer and employee organizations recommends to strengthen the collaboration between the parties in the future, but they disagree about how far the collaboration between the parties has come under pressure. They have different experiences and interpretations weather the cooperation and the Norwegian model is under pressure or not. Further, they conclude that the level of HSE and the working environment in the Norwegian petroleum sector is high. At the same time, safety challenges and serious conditions have arisen the past few years.

The aim of the "Participation Barometer" (Falkum et al., 2017) is to analyze the development of how employees experience their influence on work, and their perceptions of control/management, organization and leadership at work. This survey is conducted annually. According to Falkum et al. (2017), *"Employee participation and trustful relations influence on company development. It serves both employees and the company's interests at the same time"*. In the literature, leader performance/practice distinguishes between management and leadership; management means to lead through systems and routines, while leadership means to lead through dialogue and hands on relations with employees (Ladegård & Vabo, 2010). The study measure if participation at work (the Norwegian model) is under pressure or not, and shows the development over years. The analysis emphasizes the relations between leaders, union representatives and employees. The hypothesis is that leadership practice affects the relations and cooperation between leader, union representatives and employees to a large extent. The sample was totally 3053 respondents—from private sector and public sector (county council/municipality and state). According to the results, 42 percent of the respondents answer that Norwegian working life develops to be more authoritarian, while 12 percent answers a more democratic development. 28 percent of the respondents answered no change (status quo). In the oil and gas industry there is 56 percent of the respondents answering that work life is being more authoritarian, meaning reduced participation at work and high degree of control (management) and loyality. The analysis of the results from the survey are based on a model based on characteristics of ideal management and leadership performance/categories. And how management and leadership influence on; trust, restructuring, professional integrity and conflict handling in Norwegian work life in 2017.

The conclusions from this survey is that participation at work is the most widespread form of leadership in Norway, compared to standardization/management, despite of that many experience that working life has become more authoritarian than before.

5 DISCUSSION

5.1 *Reversing the trend—through trustful cooperation*

Based on our findings, the collaboration in the Norwegian petroleum industry is both weakened and strengthened during the economic crisis. We discuss the relationship between safety, collaboration, trust and leadership in the petroleum industry. How to improve the collaboration and re-build trustful relations—to collectively develop the safety level in this industry? On one side we find that the Norwegian model is under pressure and that bipartite collaboration needs to be improved, and the employee voice need to be heard. On the other hand, we find examples of a strengthened bipartite collaboration in the shipping companies we have studied—in the struggle of survival.

Traditionally, the Norwegian petroleum industry has been known for its high degree of safety

and trust—both in bipartite- and tripartite collaboration. However, the collaboration between the parties has been under greater pressure, both between companies and unions and between them and the government. Related to the negative trend and development with safety challenges and serious conditions in the petroleum industry over the past years, the safety has been a "hot topic" both from the authorities, oil and gas-companies, employer- and employee organizations, and researchers.

The Ministry of Labor and Social Affairs and authorities (PSA) has put effort in how to enhance safety conditions onshore and offshore at plants and installations—through collaboration. To strengthen collaboration between the parties, both bipartite- and tripar- tite cooperation has been one major goal and priority. PSA has through the campaign addressed cooperation, stressing the responsibility held by leaders in the oil and gas companies—to involve employees more actively in problem-solving about safety matters at work. If not, PSA are worried about that loss of employees' opinions in problem-solving may lead to poorer safety conditions.

The voice of employees is important—to build a safety culture characterized by open communication, where one could speak up about concerns and ideas of improvements. Cost reductions and downsizing in the petroleum industry may have influenced the organization culture in a negative way with less openness and increased fear of sanctions as response of reporting. As one of the union leaders we interviewed said; he would not recommend anyone to take a role as a union representative or safety delegate today because of the pressure on employees having such positions in a company. He experienced that the free voice of union representatives was not appreciated and listened to as it used to be.

How the leaders respond to employees' voice and how they deal with concerns will affect problem-solving about safety matters. If the leaders signal willingness to act on employee voice, the employees' motivation to speak up are enhanced (Detert and Burris, 2007; Edmondson, 2003). The relation between worker and manager will thus impact on the degree of employee voice and the employees' participation and influence on work and development of work. According to safety, trust is a key factor to get safety issues on the agenda and to inform the management about what the sharp end in the organization experience and know related to how to run the operations safely. Trust opens up for good communication, commitment and sharing of information and knowledge (McEvily et al., 2003). The workers are close to the operations and every day productions, with a hands on experience and knowledge about dangerous situations and possible incidents. Trust has positive impact on safety culture and safety performance in high-risk organizations (Reason, 1997; Conchie et al., 2006).

It is assumed that trust in organizations is beneficial for safety, (e.g. promotes open communication) and distrust is detrimental (e.g. leads to failed safety initiatives) (Conchie & Donald, 2007). What may be the consequences of distrust in tripartite and bipartite collaboration related to safety? According to Falkum et al. (2017), Norwegian work life are being more authoritarian, and the oil industry are going in that directions according the survey about participation at work. Authoritarian ledership style means less involvement and participation at work among the employees. So, their voice and opinions will not be listened to in the same manner compared to a work place characterized by a good formal bipartite collaboration, and the possibility for employees individually to bring their concerns up to their line manager (closest leader), trusting that it is safe to speak up without fear of sanctions.

Leadership practices affect the relations between managers, union representatives and employees to a large extent. Models of leadership practice inviting to participation at work, leads to higher agreements both in matters about restructuring processes and enhanced commitment towards company strategies and values in the organization (Falkum et al., 2017). They find that trust decrease with a management style (control), while trust decrease related to participation at work and trustful relation with nearest leader. According to safety, authoritarian management style may lead to poorer safety because of the problems of communication not build on trust. The problems associated with distrust or lack of trust are failed safety initiative and an absence of shared inter-group safety perceptions (Clarke, 1999). Reason (1997) argue that trust promotes open communication about safety (reporting culture) which enhances organization learning about accidents (informed culture). The main problem associated with under-reporting (or biased reporting) are the reductions in organizational learning and development of informed strategies to improve safety.

There are however bright spots regarding collaboration in the petroleum industry. The findings from the RISKOP project show how the collaboration between management and employees/union representatives in the offshore shipping industry have strengthened during the economic crisis. When human societies face an obstacle or an external enemy, they seem to seek collaboration and alternative solutions. One fruitful example of such a mechanism is described by Evans-Pritchard when he discusses the political system and decision making among the tribe Nuer in Sudan (Evans-Pritchard, 1940). The core term in this book is that this community consist of "a system of segmen-

tary opposition" and illustrates that local groups and communities can be united when it is a conflict on a higher level in the society. The Nuer society consist of potential of alliances and fissions, or as one informant told the anthropologist: "We fight against the Rengyan, but when either of us is fighting a third party we combine with them" (Evans-Pritchard 1940: 143).

Transferred to the Norwegian offshore sector, it gives sense that the local shipping company see a shared value in finding common solutions. When they are facing a crisis in the sector as they are at the moment, they see new and unusual internal arrangements. Both managers and employees will seek dancing rather than conflict, to survive as a company. So once again the metaphors "boxing and dancing" are useful to illustrate the strategic choices the staff of certain companies have, when it comes to strategic choices in the daily work during extraordinary times (Huzzard et al., 2004).

6 CONCLUSION

In this paper we have discussed the relation between cooperation, trust, and safety in the petroleum industry. The authorities (PSA), employee unions and employer organizations all want to improve safety in the oil industry. The economic crises with downsizing and cutting costs seems however to have changed the cooperation climate between the parties. Union representatives and employees experience that it is more difficult to tell- and report about dangerous situations that may lead to accident and unwanted incidents. This may lead to a safety culture not responding to dangerous situations at installations/plants offshore and onshore. We argue that the voice of employees must be listened to and taken notice of as an important instrument of improving safety.

On the other hand, in the RISKOP research project, studying safety at offshore operations in the shipping industry, we found examples of that bipartite cooperation actually was strengthened. In the struggle of existence in a marginal market, there has been collective initiatives between management and employee unions fighting for the company's survival and to keep as many jobs as possible. This way, the employees have gone a long way to find solutions, such as reduced pay in periods and abolish bonuses, and similar means of savings. Local alliances have become stronger—to fight the economic crisis.

REFERENCES

Antonsen, S. (2009). *Safety culture: theory, method and improvement*, Farnham, Ashgate.

Burns, C., Mearns, K. & McGeorge, P. (2006). Explicit and implicit trust within safety culture. *Risk Analysis* 26 (5), 1139–1150.

Clarke, S. (1999). Perceptions of organizational safety: Implications for the development of safety culture. *Journal of Organizational Behaviour,* 20, 185–198.

Conchie, S.M., Donald, I.J. & Taylor, P.J. (2006). Trust: Missing Piece(s) in the Safety in Puzzle. *Risk Analysis*, 26 (5).

Detert, J.R. & Burris, E.R. (2007). Leadership behaviour and employee voice: It the door really open*? The Academy of Management Journal*, 50(4), 869–884.

Edmondson, A.C. (2003). Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies*, 40, 1419–1452.

Evans-Pritchard, E.E. (1940). *The Nuer*, Oxford: Clarendon.

Falkum, E., Nordrik, B., Drange, I. & Wathne, C.T. (2017). Participation barometer 2017: Change in working life relations.

Farndale, E., van Ruiten, J., Kelliher, C. & Hope-Hailey, V. (2011). The influence of perceived employee voice on organizational commitment: An exchange perspective. *Human Resource Management* 50: 113–129.

French, J.R.P. & Raven, B. (1959). The bases of social power. In D.P. Cartwright (Ed.). *Studies in social power*, 150–167. Ann Arbor: University of Michigan.

Guldenmund, F.W. (2000). The Nature of Safety Culture: A Review of Theory and Research. *Safety Science*, 34, 1–14.

Haukelid, K. (2008). Theories of (Safety) Culture Revisited: An Anthropological Approach. *Safety Science*, 46:3, 413–426.

Huzzard, T., Gregory, D. & Scott, R. (eds.) (2004). *Boxing or Dancing?* Houndmills, Hampshire, UK: Palgrave Macmillan.

Ladegård. G. & Vabo, S.I. (2010). *Ledelse og styring.* Fagbokforlaget, Bergen.

Levin, M. et al. (2012). *Demokrati i arbeidslivet: Den norske samarbeidsmodellen som konkurransefortrinn*, Fagbokforlaget.

McEvily, B., Perrone, V., Zaheer, A. (2003). Trust as an Organizing Principle. *Organization Science* 14(1): 99–103.

Medbestemmelsesbarometeret 2017: Arbeidslivsrelasjoner i endring. FoU-resultat 2017:05, Arbeidsforskningsinsituttet ved Høgskolen i Oslo og Akershus.

Ministry of Labor and Social Affairs (2017). HSE in the petroleum industry (Helse, arbeidsmiljø og sikkerhet i petroleumsvirksomheten). Report 09/17.

Nonaka, I. & Takeuchi, H. (1995). *The Knowledge Creating Company*. Oxford University Press, New York.

Petroleum Safety Authorities (PSA) (2016). Reversing the trend.

Petroleum Safety Authorities (PSA) (2016). Trends in Risk Level in Norway's Petroleum Activity (RNNP).

Reason, J. (1997). *Managing the Risks of Organizational Accidents*. Aldershot: Ashgate.

Rosness, R. & Forseth, U. (2014). Boxing and dancing: Tripartite collaboration as an integral part of a regulatory regime. In P.H. Lindøe, M. Baram, O. Renn, Eds. *Risk Governance of Offshore Oil and Gas Operations,* New York: Cambridge University Press.

Skarholt, K., Lamvik, G., Antonsen, S., Røyrvik, R. & Jonassen, J.R. (2017). Crisis in the Norwegian petroleum industry: How does it affect safety conditions offshore? *Risk, Reliability and Safety: Innovating Theory and Practice* –Walls, Revie & Bedford (Eds), pp. 306, London: Taylor & Francis Group.

Tharaldsen, J.E. (2011). In Safety We Trust—Sikkerhet, risiko og tillit i offshore petroleumsindustri. PhD. Universitetet i Stavanger.

Trygstad, S.C., I.M. Hagen (2007). *Ledere i den norske modellen*. Oslo: Fafo. Faforapport 24.

Trygstad, S.C., Skivenes, M., Steen, J.H. & Ødegård, A.M. (2014). Evaluering av varslerbestemmelsene. Faforapport 2014:05.

Weick, K.E. & Sutcliffe, K.M. (2007). *Managing the Unexpected: Resilient performance in an age of uncertainty.* Jossey-Bass, San Fransisco.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Production and protection. Seafarers' handling of pressure in gemeinschaft and gesellshaft

K.V. Størkersen
NTNU Social Research, Norway

A. Laiou
National Technical University of Athens, Greece

T.O. Nævestad
Norwegian Centre for Transport Economics, Norway

G. Yannis
National Technical University of Athens, Greece

ABSTRACT: Seafarers experience conflicting objectives of production and protection in most operations. This study explores how seafarers deal with such pressures, through an analysis of interview data from 20 seafarers working on Norwegian- and Greek-controlled coastal and international passenger and cargo vessels of different sizes. Despite the various contexts, the results show similar conflicting objectives and pressure handling. The pressure is experienced differently, however, due to diverse organizational relations. Seafarers on the *large* vessels in large companies describe business-like relations (gesellshaft) and direct efficiency pressures from superiors. Seafarers on the *smaller* vessels in small companies contrastingly report of close relations (gemeinschaft), devotion to the company and thus an internal wish to be efficient.

1 INTRODUCTION

Seafarers, as personnel in other industries, have in recent decades experienced increasing work pressure. Fewer persons are to complete more tasks in less time (Österman and Hult, 2016), on shorter sea passages and rapid turnaround (Hetherington et al., 2006) without added resources (Lappalainen, 2016).

On top of the work pressure, seafarers need to take care of safety for themselves, the crew, vessel and cargo. Safety can be understood as the presence of organizational conditions making operations to be carried out without accidents or harm, in the short and long run (Kongsvik, 2013).

Conflicting goals of protection and production are present in all organizations (Reason, 1997). Production includes costs, work, and time pressure for the personnel. Protection is about making sure no one is harmed by production, and is related to competence, procedures, and material and immaterial support. Protective measures can also be viewed as pressure.

In this paper interviews of seafarers are analyzed to explore how they deal with pressures of production and protection. The seafarers work at Norwegian- and Greek-controlled coastal and international passenger and cargo vessels. We find that the seafarers across contexts handle the similar pressures similarly. The main difference involves how the pressure is experienced, and this seems to be defined by whether organizational relations are close (gemeinschaft, mainly on the small vessels, that usually are owned by small companies) or business-like (gesellshaft, mainly on the large vessels owned by large companies).

2 LITERATURE

2.1 *Research about conflicting conditions*

Operations are influenced by organizational structure and management, regulation and policymaking (Reason, 1997). Within this context, operational personnel constantly face conflicting goals.

Managers often value short-term financial criteria over safety, giving conflicting goals of production versus protection, or efficiency versus safety (Rasmussen, 1997, Reason, 1997). Production will generally be prioritized, since "production creates the resources that make protection possible"

(Reason, 1997). Vaughan (1997) shows how personnel often want thorough rule-complying operations, but that cost and time pressures slowly drive work practice away from the original quality ensured routines. Hollnagel (2009) describes an efficiency/thoroughness trade-off principle. Managers want efficiency, but if the personnel work quickly instead of thoroughly, lower safety might be a result, which paradoxically is not efficient. Likelihood "of failures grow[s] when production pressures do not allow sufficient time—and effort—to develop and maintain the precautions that normally keep failure at bay'' (Hollnagel, 2009). This efficiency paradox is also noted for seafarers by Fenstad et al. (2016). Seafarers are known to be efficient to help their company remain in business (Sampson et al., 2014). Personal injuries, violations and risk acceptance on board are related to work pressure and poor organizational safety culture (Nævestad et al., 2017). Crews' immaterial conditions, like time, concentration and competence, largely influence how safe they can work (Størkersen, 2017). Critical conditions are minimal resources, fast pace and accompanying lack of discretionary space, while regulation can moderate such pressures (ibid). Ferry personnel have several strategies on how to meet schedules rather than comply with rules:

> *The ability to keep the schedule and not canceling a departure, are associated with high competent navigators. Being delayed, or even worse, canceling a departure, may damage the navigator's reputation (Aalberg and Bye, 2017).*

Still, operational personnel are expected to comply, even though for example Hale and Swuste (1998), Bieder and Bourrier (2013) emphasize that safety is not assured by blind rule-following. Compliance with bad rules that do not fit the real-world situation can lead to accidents (Reason, 1997). Safe work vitally depends on personnel's skills and experience (Dekker, 2017), for example about which rules should be avoided. Formal rules are not viewed as a positive contribution to the traditional professional values of seafarers: "Good seamanship" belongs to a seafarer with practical and social abilities who maintains safe practices with professional judgment, without being told what to do (Antonsen, 2009a, Knudsen, 2009). Since rules usually define operations that everyone knows, vessel operations are rather performed using experience (Bhattacharya, 2012, Aalberg and Bye, 2017). Many companies implement safety management systems that are not tailored to specific vessels and activities. This makes procedures too numerous, detailed, and distanced from actual operations (Lappalainen, 2016, Bhattacharya, 2012). For some situations there are more than one procedure, or too few crewmembers to comply (Aalberg and Bye, 2017, Størkersen and Johansen, 2014). Seafarers are also required to perform documentation "essentially outside their primary functions of ensuring safe and efficient sailing" (Silos et al., 2012). It can lead to stress and exhaustion, particularly because it is viewed as unnecessary and disproportionate (Österman and Hult, 2016).

2.1.1 *Research about different types of relations*

The early sociologist Ferdinand Tönnies (lived 1855–1936) characterized relationships in different societies, applicable to maritime companies. *Gemeinshaft*, on the one hand, means close, personal relations with shared language, norms and values, based on feelings, habits and consciousness (Falk, 2000). *Gesellschaft*, on the other hand, define impersonal business-like relations characterized by strategic decisions and exchange of means (ibid).

Relations have also been a topic in safety research. Subordinate levels depend on trust and support from upper levels to be able to do their work safely. This can include care and concern (Jeffcott et al., 2006), personnel, equipment, leadership, time, rest, etc.

Vessel operations rely on a balanced relationship between shore management and crews (Xue et al., 2015), with effective communication (Bhattacharya, 2009) and a management that is committed to safety (Lappalainen, 2016). The safety level on each vessel depends on safety prioritization on board the vessel, in combination with seafarers' interactions with ship owners and regulators (Fenstad et al., 2016).

Most maritime studies report a lack of trust and communication inside organizations (Bhattacharya, 2009). The conclusion of Bhattacharya's (2009) double case study of vessels and ship owners from several countries is that managers and seafarers had fundamentally different understandings. Seafarers wanted to communicate as little as possible with shore-based management. Distant managers' top-down instructions about compliance bureaucratized the entire system. The personnel were offered only low-discretion roles, due to a lack of trust by managers. This is mainly what Oltedal (2011) found on Norwegian-owned tankers, leading her to urge managers to trust their highly skilled seafarers to adjust safety management systems. Employer engagement correlates with safety levels on vessels (Bhattacharya, 2012). Top management in poorly performing shipping companies have been found to be not committed to safety issues (Lappalainen, 2016).

Seafarers on short contracts are seen as particularly vulnerable, as they are in an asymmetrical relationship with their employers, which prevents them from speaking up for their labor rights

(Bhattacharya, 2009, Lappalainen, 2016). Seafarers on long contracts are reluctant to offend their managers since that can jeopardize their future plans and lives on the vessel (Xue et al., 2016). The dangers of a non-functioning relationship are described by Antonsen (2009a):

> *… asymmetrical power relations seem to influence on the decisions regarding when working conditions are to be considered safe enough. … The role of such asymmetries in safety-critical decisions should not be underestimated.*

Two companies studied by Xue et al. (2015) aimed to balance decision-making involvement but met limited success. Interviews with managers showed little tension between shore and vessels, but the personnel on four vessels had contrasting views. The seafarers had to follow management instructions, even though it compromised their decision making and even their safety. They felt obliged to maintain hectic sailing schedules and to accept prolonged working hours despite experiencing fatigue. The crews did not complain to management, as they saw that as useless, but sometimes they made decisions against management's wishes. Their contribution to safety management was weak overall. These conflicts in interests between management and vessel staff worsened safety practices on board.

3 METHOD

The data material consists of 18 qualitative in-depth research interviews with 20 seafarers from a range of ship-owning companies. The interviews were conducted in Greece and Norway (see Table 1).

The interviews give perspectives from different parts of maritime transport. They targeted seafarers of passenger and cargo ships, with coastal and international activity around Norway and Greece. These cargo vessels transport different types of gas, dry bulk cargo, general cargo, fodder for fish farms, or live fish.

The Greek material includes personnel from the passenger and cargo sectors, where the passenger vessels are Greek registered and operate in Greece, while the cargo vessels operate internationally and are registered both in Greece and countries with laxer regulation (called Flag of Convenience) and thus mostly foreign crewmembers. All of these vessels are rather large, usually have crews of some size (10–40 persons) and are owned by companies with many vessels.

In the Norwegian data material, however, there mainly are small vessels transporting cargo on the Norwegian coast. The vessels have Norwegian owners, and some are registered in Norway and carry only Norwegian personnel, while other have

Table 1. Information about the data material.

	Greece	Norway
Interviewees	10	10
Interviews	10	8
Background of interviewees	Crew members with professional experience between 3 and 30 years	Ship officers and educated navigators. Eight work as captains or mates on cargo vessels. One work in management. One is partly captain and partly ship-owner (common in Norwegian coastal cargo)
Nationality of interviewees	8 Greek, 2 Turkish	9 Norwegian and 1 Latvian officer
Gender	9 men, 1 woman	10 men
Contract length	4–7 months contracts—on the ship all the time. Unemployed after, but usually new contract and back on the ship after a month	Norwegians: Permanent contracts, working 4 weeks and staying at home 4 weeks. Foreigners in the crews: Often 4–8 months' contracts
Watch schedules	Cargo: Two shifts, commonly 4–4 (but in practice flexible). Passenger: One shift, and sleep at night.	One or two shifts. Two shifts commonly have 6–6 watchkeeping schedule, but in practice flexible.
Size of crews	10–40	6–15
Vessel type	Passenger ships with national routes (5 vessels) and cargo tankers with internet activity (5 vessels)	Cargo (7 vessels), mainly with coastal activity, some international activity
Registration of vessels	Greek and Flags of Convenience	Norwegian and Flags of Conv.
Data gathered	Spring 2017	Spring 2017

Flag of Convenience, a Norwegian captain and often Asian or Eastern European crew.

The seafarers volunteered to be interviewed after information about the project from the researchers through their companies to all their seafarers. In further studies one should work to include more voices from groups such as ratings and machine chiefs.

We conducted eight semi structured research interviews of 1–2 hours. The interviews were based on an interview guide constructed to explore safety culture and its relations to organizational and societal aspects. Among the subjects asked about, were conditions for work and rest (manning, watchkeeping, tasks, etc.), and perceptions of safety, leadership, team culture, safety management, safety regulation, and organizational and national values. In Greece, all interviews were face-to-face on board vessels. In Norway, six interviews were on phone, with one researcher talking to one ship officer. The other interviews were conducted on vessels, each with one researcher talking to two ship officers. One of these interviews were recorded and transcribed in verbatim. For all interviews, detailed and anonymized research notes were written. Categorization and pattern-analysis was performed manually. The quotes in Section 4 are direct citations from the interviews.

This study is not a comparison of the Greek and Norwegian maritime industry, since there are many groups and characteristics within the data collected in Greece and in Norway. It is a part of the SafeCulture project, funded by the Research Council of Norway, and undertaken by the Institute of Transport Economics (Norway), NTNU Social Research (Norway) and the National Technical University of Athens (Greece). The project's survey results show how work pressure and organizational safety culture are related to work, which is related to personal injuries (Nævestad et al., 2017, Nævestad et al., forthcoming).

4 RESULTS

Seafarers from many different groups are interviewed, and they describe many common features in how they deal with pressures of production and protection. Some conditions are special for certain groups. The differences are most evident between seafarers on large and small vessels, since the size of the vessel is connected to size of the company and activity, and to closer or more distant organizational relationships.

4.1 *Protection: Competence of the crew*

Most seafarers are aware that they are responsible for the safety of their shipmates, the vessel and the cargo. Many have great knowledge of and interest in company procedures, and national and international regulation and policymaking. They know their job by hearth, and have various opinions on the large changes derived from the implementation of electronic devices and equipment.

The seafarers tell that they always do the tasks as safe as possible—at the same time as being efficient. Most of the interviews indicate a pressure to go through with risky operations and to work while tired. Handling contradictory goals are talked about as a key characteristic of a good seafarer, but it differs how much of the decision-making is left to the seafarer.

On the *large* vessels, an efficiency pressure is sometimes stated directly to the seafarers from the company managements. This is described in interviews especially at the international and large vessels from large companies. It is not uncommon that officers order seafarers to work faster or under other conditions that they find dangerous, or that onshore management order navigators to take shortcuts to arrive in port on time. (Of course, many international seafarers say their company respect seafarers' judgement and do not force them to hurry up or push the ship into its limits.)

At most of the *smaller* vessels, however, the judgement or handling of conflicting goals is up to the seafarers. It is underlined as an internal criterion of being a good seafarer and employee. The pressure is not from management, but within each seafarer. They take responsibility for their company to stay in business, and thus indirectly for them to keep their job. The coastal seafarers agree that some operations cannot be accomplished, but their doubts and perception of pressure vary. Navigators on the small vessels have much decision-making power, and emphasize how they make their considerations and handle the pressure.

> *Sometimes you feel it. Maybe when you're approaching the quay, "will this work or not", but usually it works okay. You have to use your common sense, and know your limitations. You can lie at sea until the conditions are better. Even if someone stands at shore and waits, they just have to wait. But you do feel it. But in the end, you don't care, even though you think about it afterwards*. Captain, small bulk vessel

4.2 *Production pressure: Costs*

Maritime transport companies are in competition with other types of transportation, with each other and with vessels of different registration and conditions. Succeeding, buyers focus on price rates.

> *It's awful, just prices. It's nothing to ask about, just price and price and price. They don't look at what's in the dock, just as long as it floats it's okay*. Captain, small bulk vessel

Both small and large companies need to save costs, and a result is low manning, limited potential to buy new equipment, small time margins in routes or port calls, and so on. The seafarers on all vessels want quality in spare parts and other technology, for the sake of safety, but usually they need to cut costs.

> *What can I do with a Chinese spare part? I don't trust it but it's cheap. An employee behind a desk can't understand the difficulty or the danger*. Engineer, large cargo vessel

On *small* vessels, many seafarers see it as their responsibility to handle the economic production pressure. Mostly production can be performed as planned, but sometimes there is doubt whether or not one should start or continue an operation, for example because of bad weather. If they do not go through with operations they will miss out on essential earnings. In such situations seafarers themselves can make cost saving or profit their decision-criteria.

> *Yes, we can feel pressured to work. […] There are situations where we wouldn't have approached in that weather, but when we're already there we continue the operation. No one wants to make the decision to abort. It costs a lot to run this vessel*. Mate, coastal live fish carrier

Some conditions are truly different on the *large* vessels compared to the small vessels. There seem to be more cost-saving, more pressure from management, more sanctions, less discretionary space and less labor rights. Two interviewees mention that their equipment is of so poor quality that they have to buy new equipment on their own expenses. If they do not buy new equipment, they are not able to comply with procedures. They cannot risk being reported to the company for ignoring procedures, as this will affect their future in this job. Another seafarer tells about one time he got ill and did not get sufficient treatment, but he would not press charges to the company because that can spoil his reputation so he never can work on a ship again.

4.3 *Production pressure: Time*

Seafarers experience a pressure to work fast, sometimes under risky circumstances. Our interviewees especially feel the time pressure in situations related to port calls. They describe narrow time margins in all schedules, and too much work to keep the schedule. Vessels in large ports can be delayed by port authorities or logistics even if they get the work done in time themselves.

Time is a reason why seafarers experience a pressure to go through with operations that should have been stopped.

> *We take short cuts; we don't have manning to get everything formally right*. Captain, small general cargo vessel

Time pressure is common for every interviewee, but it varies where the pressure is perceived to come from. On a *large* vessel, an engineer mentioned that he felt terrible when he was given a few hours in order to fix a serious damage on board. On a *small* vessel, the seafarer with engineering tasks would typically not be given a deadline for repairing the damage, as they rather describe an internal pressure or a wish to fix the damage before planned departure from port. Time pressure limits the seafarers' possibility to rest and work safely.

4.4 *Production pressure: Much work, less sleep*

In addition to the production pressures of costs and time, seafarers experience a pressure of additional work and tasks.

On some *large* vessels, there usually are more than one shift on board, which is not common on *smaller* vessels. Deck personnel mostly rest during sailing, or in some rare periods of long inactivity. For both types, port calls prolong watch-keeping hours and gives no potential for rest until the vessel is back in clear waters (or anchored or docked). This results in limited discretionary space for all seafarers.

> *You've chosen an occupation and it's been like this since I started at sea. Since I started as deck boy. Everyone had to chip in when we loaded, and we could relax when the ship was at sea. It's a culture that It's not possible to change a culture that's been there forever. When the load's ready: «Oh, no, I have to sleep ten hours, I can't work», right. I won't make money and the company won't make money. Then I'd have to quit. I'd have to get home and stay on welfare, that's next*. Captain, small bulk vessel

Some vessels, both *large* and *small*, have sailing schedules with frequent port calls and short sailings. An engineer on a large vessel told us that he was continuously on duty for a long time because the ferry docked in many ports and there was no shift replacement. This made him feel weak and tired, but he accepted it as "how it is" for seafarers. A similar situation is common on some small vessels too.

> *Particularly on timber runs, some ports are close to each other. You get two-three hours on the pillow before it's up again. And we load for four-five hours and continue. Four-five hours to next port, and loading again. And maybe you have four ports like that after each other. Then you'll be tired when you're finished*. Captain, small bulk vessel

Organizational conditions contribute to lack of rest, like the amount of work compared to manning

level, watch-keeping schedules and sailing schedules. Even though ship-owning companies are in charge of this, sleepiness is mostly talked about as something that all seafarers experience and need to handle. They mostly blame the vast horizon view, darkness, or the weather.

The majority of the interviewees admit that it is easy to fall asleep on duty. On the *large* vessels, if their shift is on the bridge, they might ask for permission to leave and take a "power-nap" or just ask for a cup of coffee. On *small* vessels one usually consider and make such decisions for oneself.

Seafarers on *large* vessels also mention that in their valuable situations of rest, they still have to stay alert in case someone asks them a question. Especially electro-technicians and officers who have specific and special responsibilities are often asked to solve a problem. To stay alert, even off-duty, is an "unwritten law" on board on the international vessels in this study. Even though rest is a luxury on board, these interviewees point out that in case a superior demands your help, you must present yourself.

4.5 *Compliance and violations*

Protection equipment is essential and seafarers wear it as a habit and a necessity. They use gloves, googles and boots for their own safety, and not only to follow procedures.

All the interviewed seafarers report that it is compulsory to read and sign the vessel's safety management system and take part in drills. Still, the seafarers report that their system is violated on a daily basis.

> *For instance, it says you're to test the emergency radio every day. That's something you just don't bother*. Mate, small bulk vessel

Most of the work is done safely and according to procedures, and violations mostly happen because procedures do not fit the situation, the vessel do not have time or manning to comply, the seafarer do not know the procedure, or because of slips. Common slips are to forget to use a hard hat or life vest, but according to the seafarers this has decreased over time. As we have seen, «short cuts» or «calculated risks» to work efficient seems to be a regular part of work among all interviewees in the study. Through the stories in the interviews is evident that many procedures are neglected regularly among the coastal vessels. One of the interviewed seafarers notice that it is dangerous with too many procedures; Now no one has oversight, and some tasks might be neglected over a long time.

In general, it is common for seafarers on both large and small vessels to violate procedures to do the job more efficient.

At the *large* vessels, we are told it also happens that crewmembers are ordered by superiors to violate procedures. One interviewee tell he has been forced to pass alone through a tunnel under the holds of the ship, even though this involved risk of intoxication.

4.6 *Production pressure: Bureaucracy*

Seafarers on all the studied vessels underline that there is too much paperwork and bureaucracy every time they approach or leave a port. Many feel this as exhausting.

> *There is too much bureaucracy. Each country has regulations outside the IMO. There should be a list for when we arrive at the port. Not every country sends its own list and in many cases it is sent in the local language, not even in English*. Deck officer, large gas tanker

All interviewees talk negatively about their safety management systems. They are too complex, and with procedures that cannot be complied with. For example, the procedures for maintenance are seen as detailed and unfitting for especially *small* vessels. Gas tankers have additional regulations to follow. If a company owns gas tankers and also other types of vessels, procedures applied for tankers are usually implemented on the other vessels too.

> *The problem is that the ISM-system's too big and extensive for the ordinary man to take the trouble to get to know it. So it's usually the ship management that knows what it's about. This is an overstatement, because most [crewmembers] know the basics, but not more than that*. Mate, small bulk vessel

On the *large* gas tankers, it is told that foreigners sometimes quit because of the extended bureaucratic procedures.

> *The bad thing is that paperwork is harvested. For example, for each drill done, everyone must sign. In these 2 hours I lose, and I lose them every day, I would have learned a lot of things*. Deck crew, large gas tanker

4.7 *Production and protection: Social conditions*

In the interviews, it is described how the crews take care of another and do not let each other ignore safety measures. If someone finds themselves extremely tired, colleagues can replace them or change shifts. Inconsiderate actions are neither allowed nor forgiven. Interviewees often speak about themselves and their colleagues as one, as a crew or a team. One's safety depends on the others' safety and vice versa. This study has revealed a deep trust between many crewmembers. On *small* vessels this trust is often shared among all crewmembers.

From the *large* vessels, there are many stories about hierarchy and a gap between crew and ship management. Seafarers on large vessels follow and respect hierarchy on board. If something happens, they usually report to the next rank. If the situation is of minor importance it does not reach to higher officers or the captain. On cargo ships, it is reported that it depends on the atmosphere and the captain's attitude as a whole. A very strict captain is better to be avoided. In this study we have heard only a few stories about managers giving positive feedback for crewmembers' compliance with safety rules. Still, most interviewees tell that they always remind forgetful coworkers to wear their personal protective equipment, even if it is someone of higher rank. Safety is perceived as not a reason for misunderstanding or fight. A cadet with three years of experience gave a relevant example:

> *I saw the captain without his helmet. I felt weird, but I finally told him "captain you forgot it" and I gave it to him. The captain then praised me for this*. Cadet, large passenger ferry

On both small and large vessels, problems of behavior may occur when the "atmosphere" on board is not so warm and friendly. Such problems are usually confronted on board and without intervention of the company. Several interviewees informed us that problems can be the result of long contracts (of 6–7 months or seasonal) as nerves becomes tight when the crewmembers are on board for a long time. Long working periods on board are more common on large vessels.

> *You're always under pressure at work because you live in a prison. It's a small place because we live on the sea and beneath is the unknown. At most, we take a five-minute walk on the ship, we see the horizon, but we cannot take five steps*. Deck officer, large gas tanker

5 DISCUSSION

5.1 *Common pressure handling*

In line with earlier research of conflicting objectives (Rasmussen, 1997, Reason, 1997, Hollnagel, 2009), the seafarers in this study routinely handle pressures of production and protection, with many tasks and little time. Last section showed examples of how the financial situation and competition in transport are present in the seafarers' daily work. Cost pressure is related to time pressure and demands of efficiency. As core tasks are plenty, the added bureaucracy is not welcomed by the seafarers. Administration can lead to fatigue and increase risk (Silos et al., 2012, Österman and Hult, 2016). Loading and discharging situations are described as work intensive, including increased bureaucracy for each port, and no possibility for rest or going off duty in these situations. Most of the interviewed personnel, on smaller or large vessels, can rest on longer voyages. Vessels with frequent port calls—in coastal or international operations—describe a situation that is most difficult to handle, because of fatiguing pressure. Only when business is going slow, such seafarers have time to follow all safety procedures. Earlier research also has discussed how schedule and workload heavily influence the possibilities for safe work and rest (Størkersen, 2017, Sampson et al., 2014). In the present study, too, it is difficult to isolate which conditions are related to for example regulation or market.

All the interviewed seafarers describe how they handle pressures of production and protection with taking "short cuts", working a lot and resting little. This corresponds with seafarer traditions (Antonsen, 2009a, Knudsen, 2009, Bhattacharya, 2012, Aalberg and Bye, 2017). The norms onboard are strictly followed, as research of maritime safety culture report of (Antonsen, 2009b).

Our results show that conflicting goals of production and protection are constituted by a mix of conditions. These conditions stem from employers, market, and seafarers themselves. Pressures related to costs, time and work are evident for all seafarers in our study, but some aspects come out differently across the groups of interviewees.

5.2 *Maritime gemeinschaft and gesellschaft*

Two categories of seafarers seem to be divided between internal and external production pressure. Most seafarers on large vessels experience a pressure mainly from management, while seafarers on smaller vessels experience a pressure within themselves. Their different organizational conditions are related to Tönnies' types *gesellschaft* and *gemeinschaft*.

Many seafarers describe organizational relations corresponding with *gesellschaft*, with impersonal relations and strategic decisions (Falk, 2000). The seafarers describing their context like this, usually work at large vessels with crews and companies of size. Here, hierarchy is firm, and one are to do as told by superiors. Such relations between onshore management and vessel personnel are also described by Xue et al. (2016), Xue et al. (2015), Bhattacharya (2012). A common feature is that the seafarers have single contracts expiring when they leave the ship, and thus have to act strategic to be hired for the next voyage. Such seafarers experience explicit pressures from onshore management, and sometimes onboard management, and describe that they will not keep the job if they do not act upon the pressures. These seafarers' situation

is also related to a traditional *workers' collective* (Lysgaard, 1961), where subordinates oppose against work pressure through working as smooth and relaxing as possible (Rasmussen, 1997).

Other seafarers elaborate on an internal pressure, where they want to be efficient because they are responsible for their company's survival. Their relations with shipmates and management correspond with *gemeinshaft*'s close relations based on feelings (Falk, 2000). These seafarers typically work in small companies, and smaller vessels on the Norwegian coast (but not only Norwegian registered). In the interviews, they elaborate that they work fast and skip procedures in order to maintain earnings for their employers. They experience to be supported and trusted by the management. They value their autonomy and thus take a lot of responsibility, maybe beyond what management explicitly has stated or expected. This is also described in Norwegian coastal passenger transport (Aalberg and Bye, 2017, Størkersen and Johansen, 2014) and cargo transport and the aquaculture industry (Størkersen, 2012). According to earlier research, management's safety commitment is important for the safety on the vessels (Lappalainen, 2016, Bhattacharya, 2012). In the present study's interviews, the safety commitment of the management in these small companies are not elaborated on. It is possible that the managers are aware of the seafarers' internal pressure, and strategically let them prioritize production over protection.

Another uncertain factor in these results is whether the gesellschaft and gemeinschaft seafarers are different because of their organizations' or crews' sizes—or other conditions. For example, the seafarers in gesellschaft are all from vessels operating in and around Greece, while the gemeinschaft seafarers operate in Norway. There is need for more research to elaborate on the categories suggested in this study.

6 CONCLUDING REMARKS

This study adds to research results about a connection between safety and organizational relations. It also shows that traditional sociological literature of gemeinshaft/gesellshaft is valuable in safety research, since it gives a clear understanding of how different relations result in different safety practices.

All the interviewed seafarers describe how they handle pressures of production and protection with taking "short cuts", working much and resting less. The vital difference between the seafarers on large and smaller vessels is that on the large *gesellschaft* vessels formal structures and management explicitly favor production, while in *gesellschaft* seafarers experience to have support for protective measures, but still choose to favor production.

REFERENCES

Aalberg, A.L. & Bye, R.J. 2017. Violation enhancing conditions: A study of Norwegian car ferry workers' compliance of safety-related procedures. *In:* Cepin, M. & Bris, R. (eds.) *Safety and Reliability. Theory and Applications.* Contributions presented at the 27th European Safety and Reliability Conference (ESREL 2017, Portorož, Slovenia, June 18–22, 2017): CRP Press.

Antonsen, S. 2009a. The relationship between culture and safety on offshore supply vessels. *Safety Science,* 47, 1118–1128.

Antonsen, S. 2009b. *Safety culture: theory, method and improvement,* Farnham, United Kingdom, Ashgate.

Bhattacharya, S. 2009. *The impact of the ISM code on the management of occupational health and safety in the maritime industry.* PhD Doctoral dissertation, Cardiff University.

Bhattacharya, S. 2012. The effectiveness of the ISM Code: A qualitative enquiry. *Marine Policy,* 36, 528–535.

Bieder, C. & Bourrier, M. 2013. *Trapping safety into rules: How desirable or avoidable is proceduralization?,* Farnnham, United Kingdom, Ashgate.

Dekker, S. 2017. *The safety anarchist: Relying on human expertise and innovation, reducing bureaucracy and compliance,* London, United Kingdom, Routledge.

Falk, J. 2000. Ferdinand Tönnies. *In:* Andersen, H. & Kaspersen, L.B. (eds.) *Klassisk og moderne samfundsteori.* 2 ed. København, Denmark: Hans Reitzels Forlag.

Fenstad, J., Dahl, Ø. & Kongsvik, T.Ø. 2016. Shipboard safety: exploring organizational and regulatory factors. *Maritime Policy & Management,* 43, 552–568.

Hale, A.R. & Swuste, P.H.J.J. 1998. Safety rules: Procedural freedom or action constraint? *Safety Science,* 29, 163–177.

Hetherington, C., Flin, R. & Mearns, K. 2006. Safety in shipping: The human element. *Journal of safety research,* 37, 401–411.

Hollnagel, E. 2009. *The ETTO principle: efficiency-thoroughness trade-off: Why things that go right sometimes go wrong,* Farnham, United Kingdom, Ashgate.

Jeffcott, S., Pidgeon, N., Weyman, A. & Walls, J. 2006. Risk, trust, and safety culture in UK train operating companies. *Risk analysis,* 26, 1105–1121.

Knudsen, F. 2009. Paperwork at the service of safety? Workers' reluctance against written procedures exemplified by the concept of 'seamanship'. *Safety science,* 47, 295–303.

Kongsvik, T.Ø. 2013. *Sikkerhet i organisasjoner,* Oslo, Norway, Akademika.

Lappalainen, J. 2016. *Finnish maritime personnel's conceptions on safety management and safety culture.* Doctoral dissertation, University of Turku.

Lysgaard, S. 1961. *Arbeiderkollektivet: en studie i de underordnedes sosiologi,* Oslo, Universitetsforlaget.

Nævestad, T.-O., Størkersen, K.V., Laiou, A. & Yannis, G. 2017. Occupational Safety in Norwegian Maritime Transport: a Study of Respondents from Cargo and Passenger Vessels. *8th international congress on transportation research.* Thessaloniki, Greece.

Nævestad, T.-O., Størkersen, K.V., Laiou, A. & Yannis, G. Safety culture in maritime cargo transport in Norway and Greece: which factors predict unsafe maritime behaviours? 7th Transport Research Arena TRA 2018, April 16–19 2018 forthcoming Vienna, Austria.

Oltedal, H.A. 2011. *Safety culture and safety management within the Norwegian-controlled shipping industry: State of art, interrelationships, and influencing factors.* Doctoral dissertation, University of Stavanger.

Österman, C. & Hult, C. 2016. Administrative burdens and over-exertion in Swedish short sea shipping. *Maritime Policy & Management,* 43, 569–579.

Rasmussen, J. 1997. Risk management in a dynamic society: A modelling problem. *Safety Science,* 27, 183–213.

Reason, J. 1997. *Managing the risks of organizational accidents,* Aldershot, Ashgate.

Sampson, H., Walters, D., James, P. & Wadsworth, E. 2014. Making headway? Regulatory compliance in the shipping industry. *Social & Legal Studies,* 23, 383–402.

Silos, J.M., Piniella, F., Monedero, J. & Walliser, J. 2012. Trends in the global market for crews: A case study. *Marine Policy,* 36, 845–858.

Størkersen, K.V. & Johansen, J.P.K. 2014. No swans in sight. Analyzing the resilience in Norwegian water passenger transport. *In:* Steenbergen, R.D.J.M., Van Gelder, P.H.A.J.M., Miraglia, S. & Vrouwenvelder, A.C.V.M. (eds.) *Safety, Reliability and Risk Analysis: Beyond the Horizon.* London, United Kingdom: Taylor & Francis.

Størkersen, K.V. 2012. Fish first: Sharp end decision-making at Norwegian fish farms. *Safety Science,* 50, 2028–2034.

Størkersen, K.V. 2017. Coastal cargo work: How can safety shout instead of whisper when money talks? *In:* Cepin, M. & Bris, R. (eds.) *Safety and Reliability. Theory and Applications.* Contributions presented at the 27th European Safety and Reliability Conference (ESREL 2017, Portorož, Slovenia, June 18–22, 2017): CRC Press.

Vaughan, D. 1997. *The Challenger launch decision: Risky technology, culture, and deviance at NASA*, Chicago, Illinois, University of Chicago Press.

Xue, C., Tang, L. & Walters, D. 2016. Who is dominant? Occupational Health and Safety management in Chinese shipping. *Journal of Industrial Relations,* 59, 65–84.

Xue, C., Walters, D. & Tang, L. 2015. The Effectiveness of Health and Safety Management in Chinese Shipping: From the Perspective of a Shipmaster's Decisionmaking Power. *In:* Ao, S.I., Gelman, L., Hukins, D.W.L. & Korsunsky, A.M. (eds.) *Proceedings of the World Congress on Engineering.* Hongkong: Newswood Academic Publishing.

Human factors and human reliability

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Teamwork competence required across operational states: Findings from nuclear power plant operation

Ann Britt Skjerve & Lars Holmgren
Institute for Energy Technology, Norway

ABSTRACT: The tasks of Nuclear Power Plant (NPP) operators are highly interconnected, and operators need to engage in teamwork to ensure plant safety. Traditionally, teamwork-competence taxonomies for NPP operators do not distinguish among operational states. This study explored if differences exist among teamwork-competence requirements across the three main operational states in a NPP: normal operation, outage and emergencies. Data was collected from a north European NPP using observations, semi-structured interviews, and a questionnaire survey, and analyzed using a thematic-analysis approach. The study suggested that the teamwork competencies needed by NPP operators are similar, but not identical across the three operational states. The variations were suggested to be caused by a combination task differences and different impacts of three performance-shaping factors: time pressure, task complexity, and proactive attitude to safety. Based on the results, it was suggested that refresher training should be adjusted to increase resilience in teamwork in NPP operation.

1 INTRODUCTION

Nuclear Power Plants (NPPs) are key means for producing electricity in a range of countries today. NPPs are dynamic and highly complex production systems, and training of operational staff is one of the cornerstones for ensuring safe and efficient operation.

Competence can be defined as the "... ability to apply skills, knowledge and attitudes in order to perform an activity or a job to specified standards in an effective and efficient manner" (IAEA, 2002). Training of NPP operators addresses both technical and teamwork competencies (IAEA, 1996): The operators need technical competence to understand the design and functioning of the process system, and they need teamwork competence to be able to work in a team setting, due to the inter-dependability of their tasks. The technical competencies required of NPP operators is well established (e.g. U.S. NRC, 1998; 2007), and training of these is under continuous development within NPPs. The teamwork competencies required is less well-specified. This paper aims at contributing to the understanding of what teamwork competencies NPP operators need.

Teamwork can be defined as "... a distinguishable set of two or more people who interact dynamically, interdependently and adaptively toward a common goal" (Blickensderfer, Cannon-Bowers & Salas 1997, 250). There is general agreement that teamwork is a multi-dimensional concept, but there is no final agreement about the specific dimensions the concept encompasses.

In an NPP, teamwork is highly regulated by procedures and routines. Still, the level of details with which teamwork is regulated varies substantially. For example, in some cases it is specified exactly what information and operator should contribute, where as in other cases, it is merely stated that operators should contribute *all relevant* information. In addition, operators need teamwork competence to adapt performance to emerging situational characteristics, e.g., to the competence possessed by individual colleagues, the colleagues' level of workload, personal concerns, etc.

A teamwork-competence taxonomy is important to support the development of teamwork-training programs. A taxonomy facilitates identification of training needs, as well as documentation of what competencies a training program covers. Within the domain of NPP operation, various teamwork-competence taxonomies exist (e.g. Broberg, 2009; Crichton and Flin, 2004; IAEA, 1996; IAEA, 2001; O'Connor, O'Dea, Flin, and Belton, 2008; Skjerve, Kaarstad and Holmgren, 2013).

Traditionally, the teamwork-competence taxonomies are general in nature, covering the entire span of teamwork competencies needed by NPP operators to perform their tasks safety and efficiently. Still, based on the observation that the NPP operators' tasks are not identical across operational states, it was hypothesized that the teamwork-competence requirements might also not

be identical. If this hypothesis is true, traditional re-fresher training might not fully address all the teamwork competencies required by NPP operators, as it tends to focus on emergency scenarios.

The purpose of this study was to explore if differences exist between teamwork-competence requirements to NPP operators across the three main operational states in a NPP: normal operation, outage and emergencies.[1]

In this paper, the concept NPP operators, include the following roles on a shift: Shift Manager (SM), Reactor Operator (RO), Assistant Reactor Operator (ARO), Turbine Operator (TO), and Field Operator (FO). O'Connor et al. (2008) and Broberg (2009) both report that no differences were found between the teamwork-competence requirements to the two main groups of NPP operators: control-room operators and field operators. For this reason, there will be not distinctions between these two types of roles.

2 NUCLEAR POWER PLANT OPERATION

The main task of a NPP operator team is to ensure that plant safety is upheld. Overall, NPP operation can be decomposed into three operational states: normal operation, outages, and emergencies. The three operational states are defined below based on the tasks that are prototypically associated with each state.

Normal operation is the period when an NPP is producing electricity according to plan and is operated based on the requirements in the standard operating procedures, the technical specifications of the plant, the plant orders and/or other directives provided by the Operational Department. Normal operation may also be referred to as power operation. The overall task of an operator team is to ensure that the operational activities progress according to plan. The team's activities are largely based on routines, and involve monitoring plant parameters and intervening with planned adjustments and with immediate adjustment if necessary. When a shift begins, the first task is to engage in shift-handover: First, each position will have a semi-structured dialogue with his or her opposite on the departing shift to learn about ongoing and planned tasks and deviations (if any). Then all operators on the team will meet to jointly build a common understanding of the situation at hand, and decide how to proceed. Often, the SM will be away from the control-room for longer periods of time, leaving the RO in charge of the team. During normal operation, operator teams are required to perform a set of administrative activities, in addition to the operational activities. These involve, e.g., logging, preparing for upcoming tag-outs, refreshing knowledge, updating descriptions of plant systems, and job appraisal talks.

Outage is the period from when an NPP is brought to shut down for preventive maintenance, upgrades, and refueling until it has been started up again and is ready for production. During an outage, a plant is operated based on the standard operating procedures for shut-down and start-up, the technical specifications of the plant, the Outage Plan, and the outage direction documentation and plant orders. The overall task of an operator team is to ensure that the planned tasks are executed in accordance with the specifications in the Outage Plan and associated documentation to the extent this is possible. Team members' tasks are usually proceduralised, but often non-routine. Their taskload is high, and they need to engage in teamwork with staff members, whom they may not know well in advance (e.g. staff from the maintenance departments) and with external parties (e.g. various types of consultants). Also throughout an outage, time management is an issue of key concern. The conditions in the control-room will be non-normal during an outage compared to power operation: A high number of alarms will go on-and-off in unusual ways due to the various tests performed in the plant, and there tend to be more people present in the control-room. The RO and the TO each with their associated field operators may come to create what looks like two islands in the control-room, and it is important that the SM, who is offloaded at day time by an administrative support, contributes to ensure internal coordination in the operator team.

Emergency operation is the period in which an NPP is in a state described by the Safety Analysis Report (SAR) or in the plant specific Probabilistic Risk Assessment (PRA). In these situations, a plant is operated based on the emergency operation procedures, functional restoration guidelines and in extreme cases severe accident mitigation guidelines. The incidents and standard operation procedures may also be applied. The overall task of the operator team is to ensure that the plant is brought to a safe state. When an event occurs that has been addressed in SAR or PRAs (e.g. a rupture of a tube in the steam generator), task performance is heavily guided by procedures. When multiple failures (events) have occurred, the crew members will to a larger degree need to adapt the procedures to the characteristics of the situation. During emergencies, RO and ARO will typically be working together to execute the actions required

1. The paper is based on and a further elaboration of results reported in a 'limited distribution' work report by Skjerve & Holmgren (2016).

on the reactor side, whereas TO will execute the actions on the turbines, power and I&C supplies. The FOs will assist in the control-room or out in the plant. The SM will take a stand back position and survey the plant's behavior, including the critical safety functions, and coordinate the crew members' activity and plan ahead.

3 METHOD

The study was based on data collected in a PWR unit of a north European nuclear power plant. Data collection included 108 hours of observation in the control-room during normal operation and outages, distributed between two operator teams by the authors. Observations were further carried out across refresher training (i.e. regular training after the operators has been licensed to refresh and update competencies, comprising simulator and classroom sessions), including eight days of simulator runs addressing emergencies. In addition, data were obtained from 14 semi-structured interviews lasting in average 1.5 hours with plant personnel, and a questionnaire survey administered to 33 NPP operators. Method triangulation (Denzin, 1978) was applied to seek to increase the validity of the findings by cancelling out the limitations associated with each of the three methods.

Data was analyzed using a thematic analysis approach, i.e. a qualitative method that makes use of labelling and iterative restructuring of data, to identify patterns—or themes in the dataset. The analysis process was developed based on Braun and Clarke (2006). It contained four phases.

Phase 1: Familiarization with the dataset. This implied reading through notes from observations, the interview responses, and the responses to the questionnaire survey to obtain an overview of the content.

Phase 2: Generating initial codes: All data obtained, i.e. from observations, interviews, and questionnaire survey, was decomposed into segments. A segment was defined as an entity that described one aspect of the teamwork competence required by NPP operators as it emerged from the data collected. In all 136 segments were identified. Examples on segments include: "Insights into how adults learn" (Learning and Coaching); "Communicating via more information channels to increase the likelihood that a message is understood" (Communication), and "Team-orientation—expanded to unit, plant, and other entities of relevance for ensuring safe and efficient plant performance" (Attitudes).

Phase 3: Searching for themes: Establishing a taxonomy comprising a set of teamwork competence dimensions: First, each segment was assigned to one of the five categories in the taxonomy defined by O'Connor et al. (2008) based on whether the content of a segment. If a segment was judged not to fit into any of the categories, a new category was introduced and/or the definition of one of the existing categories was modified to accommodate a broader range of content. If possible, the segments were allocated one or more of the three operational states: normal operation, outages and/or emergency operation. If not, the segments were defined as common to all states.

Phase 4: The segments associated with each of the three operational states were then grouped across the teamwork-competence dimensions to identify if patterns emerged, which should help clarify the reason for potential differences.

4 RESULTS AND DISCUSSION

The teamwork-competence taxonomy established in analysis phase 3 comprised nine dimensions: Attitudes, communication, coordination, decision making, interpersonal competence, intrapersonal competence, leadership, learning and coaching, and situation awareness.

The distribution of segments across the nine dimensions can be seen in Table 1. The inter-rater reliability between the two authors showed a correspondence of 81%.

The teamwork-competence dimensions identified did not differ substantially from the dimensions identified in earlier studies addressing teamwork competencies in NPP operation. This was interpreted to support the validity of the taxonomy (see Table 2).

4.1 *Teamwork requirements across the teamwork-competence dimensions*

The results suggested that teamwork-competence requirements for NPP operators overall were highly similar: Competence associated with each dimension of teamwork was required in all three operational states. Still, a more detailed analysis showed that except for the teamwork-competence dimension *attitudes*, the specific competence aspects NPP operators were required to master showed some degree of variation across the operational states. Below the main findings are associated with each of the nine teamwork-competence dimensions are summarized:

4.1.1 *Situation awareness*

The task of building situation awareness is done based on somewhat different sources of

Table 1. Distribution of segments across the nine teamwork-competence dimensions.

Teamwork-competence dimensions	Total no. of segments	Segments common to normal operation outages and emergencies	Segments specific to normal operation outages or emergencies	Normal operation	Outages	Outages & Emergencies	Emergencies
Attitudes	8	8	0	0	0	0	0
Communication	13	8	5	0	2	0	3
Coordination	13	4	9	2	1	4	2
Decision making	13	7	6	1	0	3	2
Interpersonal competence	17	7	10	4	3	0	3
Intrapersonal competence	13	1	12	3	3	4	2
Leadership	25	8	17	5	4	0	8
Learning and coaching	16	4	12	7	2	1	2
Situation awareness	18	6	12	1	4	3	4
SUM	**136**	**53**	**83**	**23**	**19**	**15**	**26**

Table 2. Comparison of the taxonomy identified in the study with other teamwork-competence taxonomies.

Crichton & Flin (2004)	O'Connor et al. (2008)	Broberg (2009)	Skjerve (2013)	Present study
Situation assessment	Building situation awareness	Building situation awareness	Situation awareness—build and maintain an accurate and shared situation understanding	Situation Awareness—building and maintaining
Decision making	Team focused decision making	Decision-making Consultation	Decision making—team focused	Decision Making—team focused
Communication	Communication	Communication	Communication—sharing information and insights	Communication
Teamwork	Co-ordination	Planning	Coordination Back-up behaviour	Coordination
	Collaboration			Interpersonal competence
		Leadership	Leadership	Leadership
		Group climate	Attitudes—towards colleagues and the plant	Attitudes
			Personality fits	
Stress management				Intrapersonal competence
			Learning and refreshing competencies	Learning and coaching

information and under various workload levels across the operational states. For example, during normal operation the ability to establish accurate situation awareness involves teamwork-competence aspects associated with obtaining information from shift-handovers (i.e. semi-structured dialogues with colleagues), from various logs, etc. and to systematically assess these with colleagues to build a shared understanding. During outages and emergencies, NPP operators needs teamwork-competence aspects associated with obtaining information from colleagues about a dynamic situation on-the-fly, as well as competence aspects related to distinguishing between critical information and other types of information and addressing critical information in crew updates in a way all colleagues understand.

4.1.2 *Decision-making*

A range of teamwork-competence aspects associated with decision-making was shared across the operational states. This included, proactively determining how to verify the consequences of a decision and acknowledging and proactively addressing uncertainties together with team members. The differences found were mainly associated with the overall workload level, but also to some extent with concerns for ensuring the continuous learning of in the operator team. For example, during normal operation competence aspects associated with contributing to (depending on role) a more participatory decision-making process ensuring all understand the basis on which the decision should be made, aimed at jointly developing 'optimal solutions' were required. Whereas during emergencies, competence aspects associated with execution of a more authoritarian decision-making approach aimed at finding 'good enough' solutions were needed.

4.1.3 *Communication*

The communication tasks were basically the same across the three operational states. They involved, e.g., the use of "Three-way communication", adapting communication to the receiver(s)' competencies and active listening. Still, across the operational states the frequency with which communication tasks had to be executed varied, and thus the overall level of time pressure associated with task performance. This implied that the operators needed to master the communication competencies with substantially more *fluency* during outages and emergencies than during normal operation: the number of communication tasks was higher in these operational states, and the time available to identify and correct misunderstandings was more limited. Some communication tasks were further associated mainly with one operational state. During outages, e.g. the operators need to be prepared to communicate with consultants in English (a non-native language to the operators). Also during emergencies, there is a distinct need to uphold continuous communication among team mates during complex and/or stressful situations to promote collective sense-making processes and the provision of mutual support.

4.1.4 *Coordination*

The requirement to coordination competencies is essentially similar across operational states, in the sense that it covers a wide range of activities from performance-adaptation on-the-fly, engaging in backup behavior, to planning aimed at ensure coordination of future activity, which may be needed across the operational states. Still, the requirements to teamwork-competence aspects associated with coordination vary more than e.g. was the case for communication. The reason is that the content of coordination tasks prototypically associated with each operational state is more varied, and involves different teamwork-competence aspects. For example: During normal operation, it is necessary to continually to coordinate performance of operational tasks versus performance of administrative type of tasks; During outages, the need for carrying out Pre-Job Briefings is more pronounced than during normal operation, and will involve more staff, including external specialists; During emergencies coordinating activities to ensuring clear, precise and not least timely is a task of key importance.

4.1.5 *Interpersonal competence*

The inter-personal teamwork-competence aspects were to a large degree similar across the operational states, except they in general had to be mastered with greater *fluency* from normal operation, over outages, to emergencies. They comprised, e.g., building trust, mastering interactions, and recognizing the achievements of colleagues. The interpersonal teamwork-competence aspects were, however, suggested to serve different purposes during normal operation and emergencies: During normal operation, the overall purpose was to transform operators into a team and/or to strengthen the team spirit, whereas during emergencies the purpose was to uphold the operators' ability to function efficiently as a team under highly challenging conditions.

4.1.6 *Leadership*

This competence dimension was assessed to be useful for all operators, regardless of their particular role in the team, because all (with different degrees of likelihood) may end up in a situation, where they need to lead teammates. The teamwork-competence aspects required across the operational states varied, e.g., concerning the leadership style the operator should master: During normal operation, competencies associated with executing a more democratic type of leadership were needed, e.g. promoting team mates' motivation by involving them in decision-making processes and promoting learning processes. During outages, and especially during the acute part of emergencies, competence aspects associated with executing a more authoritarian type of leadership were needed, e.g. giving and meticulously following-up on orders.

4.1.7 *Attitudes*

Attitude requirements included, e.g., safety concerns pervade all thinking and decision-making processes, and conscientious and commitment to quality. For this dimension no variation was found. The attitudes identified were of key importance across all operational states.

4.1.8 *Intrapersonal competence*

This dimension contained a set of teamwork-competence aspects of generic importance for sustaining sound teamwork, such as the competence to monitor own ability to operate the plant safely and efficiently, and courage to speak-up when needing assistance to achieve these goals. Since intrapersonal competence was used to fulfill different purposes across the operational states, the teamwork-competence aspects associated with each state varied somewhat. For example, to uphold attention towards the plant processes during normal operation where 'little happened' over longer periods of time, teamwork-competence aspects associated with reducing the risk for complacency were needed. To uphold attention during outages and emergencies during prolonged periods with high workload and/or safety-critical situations, on the other hand, teamwork-competence aspects associated with preventing negative impacts of fatigue and/or of stressors on the task-performance process were required.

4.1.9 *Learning and coaching*

Learning and coaching activates may be carried out as an integrated part of task performance or as a dedicated activity. The teamwork-competence aspects implied include, e.g., coaching competence, the ability to give/receive and constructively use feedback, and techniques for self-improvement alone or with or assisted by other people.

Dedicated activities to promote learning are prototypically associated with lower workload periods during normal operation. The likelihood that such activities will take place seems to increase, if the operators find that continues competence improvement is important for the team.

During outages and emergencies, competence development may to a certain degree be an integrated part in the task-performance processes, involving teamwork-competence aspects associated with coaching.

However, dedicated learning activity in relation to outages and emergencies will usually be postponed to after the shift period is over and/or after the outage or event has been handled. At this time, a required teamwork-competencies aspect is the ability to address occurrence/events constructively in team setting, i.e. avoiding that the parties involved will be defensive and refuse to share and discuss actions, which may contain important lessons learned from the entire team.

4.2 *Why teamwork-competence requirements are not identical across operational states*

Exploratory analysis of the variations found in the requirements to teamwork-competence aspects across the three operational states, suggested that the dissimilarities might be caused by a combination of two influences: (1) differences among the operational tasks across the operational states, and (2) differences among the impact of performance-shaping factors on otherwise similar operational tasks across the operational states. These two potential causes for dissimilarity will be discussed below.

4.2.1 *Task differences*

The operational tasks that are prototypically associated with each of the operational states, as described in section 2, are not identical.

Shift-handover is a task that is prototypically associated with normal operation. This is, e.g., reflected in traditional refresher training where the hand-over process is substituted by a training instructor simply describe the plant state to the operator team. From an NPP operator's perspective, the shift-handover session in the beginning of a shift include, a semi-structured dialogue with the opposite on the departing team to obtain an accurate understanding of the plant state, including issues that need attention. Learning how to interact with the opposite to obtain the needed information is an important competence. It includes abilities to identify and constructively address potential omissions, misunderstandings and uncertainties in the information provided to build situation awareness. This type of competence is not addressed in dedicated training session following licensing.

The requirement to work with people from different professions and/or with whom the NPP operator is less familiar or unfamiliar is prototypically associated with outages. During outages extended workgroups may arise, which in addition to the NPP operator team consist of colleagues from other operator teams, maintenance personnel, contractor staff from external companies, etc.

In this setting, a key teamwork-competence aspect required is associated with promoting common ground between the diverse members of an extended workgroup. This includes the ability to present information in ways that are understandable to people with different professional background, and ensuring that the concerns of all parties are adequately brought forward and addressed. This type of teamwork-competence aspect is not addressed during training.

Handling of emergencies is highly proceduralised activity, especially in the first part of an event, which is traditionally the part that has been addressed in refresher training. In cases of multiple failures, the requirement to making situation assessment to understand how to proceed will increase. A teamwork-competence aspect that is particularly needed in this situation is the ability to uphold communication throughout periods of

uncertainty when operators tend to keep quiet and focus keep on making sense of the situation on their own. Emphasizing communication is important to promote the team's ability to build situation awareness and making sound decisions. Unless refresher training progresses into this type of situation, these skills may not be upheld.

4.2.2 *Performance-shaping factors*

The study points to three Performance-Shaping Factors (PSF) impacting the requirements to teamwork-competence aspects across the operational states: *time pressure, task complexity* and *proactive attitude to safety.* The influence of these PSFs implies that the performance of otherwise similar teamwork tasks will come to require partly different teamwork-competence aspects.

Time pressure implies that a task needs to be completed within a given time window. The time window is typically defined by constraints in the plant, e.g., the amount of break flow in a storage tank can secure. The impact of time pressure on task performance generally increases from normal operation over outages to emergencies. When time pressure is high, teamwork tasks should be mastered with a greater *fluency*. The ability to communicate concern to a team mate should, e.g., preferably be mastered effortlessly, as the time available for re-stating information and correcting misunderstandings is reduced.

As the level *of task complexity* increases, the more factors (parameters) and interdependencies an operator needs to address when performing a task. For this reason, task performance should preferably be thorough, highly systematic, and be ideally carried out without any time pressure. In situations with high task complexity, teamwork-competence aspects associated with the ability to lead and coordinate teamwork is particularly required, to help ensure that all parties involved in the task performance process will obtain accurate situation awareness, and thus a sound common basis for making decision about the course of actions needed. As for time pressure, task complexity tends to increase from normal operation, over outage to emergencies, providing the latter contain multiple failures. In situations with both *time pressure* and *task complexity*, there will be a need for mastering the teamwork-competence aspects associated with handling task complexity with more fluently.

The PSF *proactive attitude to safety* implies the conviction that it is important to establish the best possible basis for sound teamwork in future settings. This may be done by promoting learning processes, by coaching or encouraging team mates to engage in self-studies, e.g., by studying the background materials for given procedures, etc. If the proactive attitude to safety is deeply rooted in the operators, it will help overcome a tendency to perceive coaching and dedicated learning sessions as an "add on" to the normal wok practices. It will encourage the operator to perceive competence-promoting initiatives as an integrated and important aspect in task performance processes.

The impact of this PSF is most visible during normal operation, where operators may or may not priorities to engage in learning processes. It seems also to be visible in the extent to which operators are able to uphold a 'questioning attitude' while carrying out their work, e.g. reflected in the extent to which they critically review current work practices to protect against drifting.

4.2.3 *Are capturing variations in teamwork-competence requirements necessary?*

Even if the teamwork-competence requirements are not identical across the three operational states, they are highly similar. Has it any real impact on a training program if is based on a generic set of teamwork-competence aspect, rather than a set, which is decomposed across operational states? From a practical perspective, any of the three operational states may contain characteristics that from time to time may be warranted in one of the other operational states: During an outage, there may be intervals resembling operation, such as longer periods of time where 'little happens', and during normal operation, situations may arise where NPP operators need to collaborate with unfamiliar people with a difference professional background, etc. Since a certain level of overlap exists between the tasks that may arise across the three operational states, it should in principle be possible to uncover all teamwork-competence aspects required of NPP operators by studying any of the three operational states exclusively. However, this approach would be substantially less effective than studying the characteristics of each of the three operational states, as the situational characteristics, which traditionally are associated with any of the other two operational states, might likely be manifest only with highly irregular intervals in the given operational state.

Another way of answering the question is to explore the level of teamwork-competence aspects missed if leaving out one of the operational states from an analysis. This can be done based on the distribution of segments reported in Table 1. The exploratory analysis indicates that if data from normal operation is left out of an analysis, 23 teamwork-competence aspects required by NPP operators may be at risk for remaining hidden, because the need for these competence aspects is rare during outages and emergencies. This corresponds to 17% of the entire set of teamwork-competence

requirements identified in the present study. If an analysis does not include data from the outages, 14% of the teamwork-competence aspects (i.e. 19 segments) may be at risk for remaining hidden. If emergencies are left out of an analysis the corresponding figure is 19% (i.e. 26 segments). In addition, outages and emergency situations share a unique set of teamwork-competence aspects segments which together amounts to 11% (i.e. 15 segments) of the teamwork-competence aspects required.

Overall, the results indicate that it is useful to analyses each operational state when establishing requirements to teamwork competence for NPP operators.

4.3 *Implications for teamwork training*

The IAEA (1996) recommends that the Systematic Approach to Training (SAT) is used as a basis for developing training programs.

When preparing for teamwork training for NPP operators, it overall important to promote their ability to adapt performance to situational characteristics, including characteristics of team members, such as their role, current tasks and the type and level of competencies. Mutual performance adaption among team members, in combination with a clear understanding of the team's goals, will promote teamwork processes. Because of the multitude of requirement posed to teamwork-competence aspects in an NPP, it can be expected that operators, who master a wider repertoire of teamwork competencies will be better able to adapt teamwork processes, than operators, who have a more limited repertoire of teamwork competencies.

One potential use of identifying aspects of teamwork competence that are prototypically associated with particular operational states is to provide a mean for deepening operators' level of teamwork competencies. Expanding the scope of teamwork competencies addressed in refresher training will support the operators in developing and upholding an expanded repertoire of teamwork 'techniques' which can be flexibly applied depending on situational characteristics, reducing the risks for break-downs in teamwork (Skjerve, Holmgren & Widheden, 2015).

A teamwork-competence aspect it may be useful to address as a part of the classroom part of refresher training, despite it being prototypical associated with normal operation, is *team-building competence*, in particular teamwork-competence aspects associated with maintaining team members' ability to work together as a team, including upholding team spirit. A feeling of team efficacy and team spirit may promote the operators' ability to overcome challenges to teamwork, and thus contribute to resilient performance.

Similarly, another teamwork-competence aspect it may be useful to address as a part of the classroom part of refresher training or exercises, despite it being prototypically associated with outages, is *the ability to engage in teamwork with less familiar or unfamiliar parties with different professional backgrounds*. This type of training may be included as an element in emergencies exercises, comprising NPP operator teams and key positions in the technical-support center. It may be done, e.g., by asking participants to state their expectations to one another, describe why they have these expectations, and account for their concerns. This would contribute to resilience in the extended team by further strengthening team mates' ability to select information of relevance to team members and to communicate this information accurately, etc. This will promote building and maintaining situation awareness in the extended team.

Expanding the scope of refresher training by teamwork-competence aspects prototypically associated with normal operation and outages may further contribute to promote 'teamwork mode' awareness.

Being aware of the current 'teamwork mode' may increase the likelihood that they will consciously apply teamwork-competence aspects are prototypically associated with the given mode—despite the operational state they are currently in. For example, during a non-acute phase of an emergency with limited workload, an NPP operator may need to talk to various people from the maintenance to refine the teams' understanding of the situation. If the operator recognized the parallels these dialogues may have with the dialogues involved in a shift-hand over process, it could prompt the operator to remember applying similar 'techniques' (such as focus at distinguishing facts from interpretations, specifying what needs to be further examined after completing the dialogue to get a clear picture of the situation, etc.).

An operator's increased focus on how teamwork-competence aspects interplays, may promote meta-cognition about teamwork competencies. Meta-cognition may enable the operator to more readily identifying and developing solutions to teamwork challenges. Transfer of a message between two persons may, e.g., unsuccessful for a variety of reasons: A message may not be stated clearly, if may not formulated in a way the receiver understands, may be transferred at a time the receiver is unable to pay full attention to the message, the receiver may misinterpret the content due to lack of common ground, etc. Awareness that a message may not be transferred to the receiver for a variety of reasons, will allow the operator to

'troubleshoot' potential teamwork break-downs from a range of different angles, and thus increase the likelihood that a means to preventing teamwork break-down will be found.

Both the ability to transfer teamwork-competence aspects across operational states and to engage in meta-cognition about teamwork, may, thus, further contribute to resilience in teamwork.

5 CONCLUSION

The outcome of the study suggested that the teamwork competencies needed by NPP operators across the three operational states are similar, but not identical. The results indicated that unless requirements to teamwork competence are obtained from all operational states, there is a risk that important teamwork competencies will remain hidden. With respect to refresher training, this would imply that these competencies are not addressed and thus potentially that NPP operators will not maintain these competencies to the required standard.

Based on the results, it was suggested that resilience in teamwork could be strengthened if refresher training expanded its traditional focus on teamwork-competence aspects associated emergencies, to include aspects that are prototypically associated with normal operation and outages.

REFERENCES

Blickensderfer, E., Cannon-Bowers, J.A., Salas, E. 1997. Theoretical bases for team self-correction: Fostering shared mental models. *Advances in Interdisciplinary Studies of Work Teams*, 4: 249–279.

Braun, V. & Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2): 77–101.

Broberg, H. 2009. *Teamwork in Swedish nuclear power plant operations crews.* KSU dok.id 58365. Master's thesis in Ergonomics Université Paris Descartes.

Crichton, M.T. & Flin, R. 2004. Identifying and training non-technical skills of nuclear emergency response teams. *Annals of Nuclear Energy*, 31: 1317–1330.

Denzin, N.K. 1978. *The research act: A theoretical introduction to sociological methods*. New York: Praeger.

IAEA 1996. *Nuclear Power Plant Personnel Training and its Evaluation.* Technical Reports Series no. 380, Vienna, Austria: International Atomic Energy Agency.

IAEA 2001. *A systematic approach to human performance improvement in nuclear power plants: Training solutions*. IAEA-TECDOC-1204. Vienna, Austria: International Atomic Energy Agency.

IAEA 2002. *Recruitment, Qualification and Training of Personnel for Nuclear Power Plants.* IAEA Safety Standards Series. Safety Guide, No. NS-G-2.8, Vienna, Austria: International Atomic Energy Agency.

O'Connor, P., O'Dea, A., Flin, R. & Belton, S. 2008. Identifying the team skills required by nuclear power plant operations personnel. *International Journal of Industrial Ergonomics*, 28: 1028–1037.

Skjerve, A.B. & Holmgren, L., 2016. *An Investigation of Team-work Competence Requirements in Nuclear Power Plant Control-Room Crews across Operational States—a Field Study*. HWR-1107. Halden, Norway: OECD Halden Reactor Project.

Skjerve, A.B. Holmgren, L. & Widheden, B. 2015. Towards an Approach for Training Nuclear Power Plant Control-Room Crews in Handling Unforeseen Events. In: Luca Podofillini, Bruno Sudret, Božidar Stojadinović, Enrico Zio, Wolfgang Kröger (Eds), *Safety and Reliability of Complex Engineered Systems*: 3895–3902. London: Taylor & Francis Group.

Skjerve, A.B., Kaarstad, M. & Holmgren, L. 2013. Teamwork competence requirements in nuclear power plant control rooms. In: R.D.J.M. Steenbergen, P.H.A.J.M. van Gelder, S. Miraglia, A.C.W.M. Vrouwenvelder (Eds.), *Safety, Reliability and Risk Analysis: Beyond the Horizon*: 401–408. London, UK: Taylor and Francis Group.

U.S. NRC 1998. *Knowledge and Abilities. Catalog for Nuclear Power Plant Operators. Boiling Water Reactors.* Final Report. Rev. 2. NUREG-1123. Washington, DC: U.S. Nuclear Regulatory Commission.

U.S. NRC 2007. Knowledge and Abilities. Catalog for Nuclear Power Plant Operators. Pressurized Water Reactors. Final Report. Rev. 2, Supp. 1. NUREG-1122. Washington, DC: U.S. Nuclear Regulatory Commission.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Verification of HTC Vive deployment capabilities for ergonomic evaluations in virtual reality environments

Z. Tůma, L. Kotek, J. Kroupa, P. Blecha & F. Bradáč
Faculty of Mechanical Engineering, Brno University of Technology, Brno, Czech Republic

ABSTRACT: Musculoskeletal assessment of possible risks in the workplace of the future is difficult, because it is not possible to simply evaluate and predict the attitudes of workers. The virtual reality environment offers an initial insight into the future workplace. For better immersion in the environment. HTC Vive were selected. The ergonomic method selected for assessment is called RULA. Because of work in a virtual reality environment, it is necessary to measure in the real laboratory environment. As a reference device, the MS Kinect solution against occlusion has been selected. Measurements were performed in the laboratory according to a recently published article on RULA assessment using MS Kinect. The results show that HTC Vive can serve as an integral tool to assess a musculoskeletal risk in designing new workplaces.

1 INTRODUCTION

In every real production process there are risks that have the potential to endanger the operator or equipment. It is clear that the selected risks associated with the underlying process (e.g. the use of hazardous chemical substances) cannot be completely eliminated, but most of these risks can be significantly reduced. The human factor is one of the most important elements influencing the resulting safety. It is also a component that is often neglected under the prevailing emphasis on technical reliability. The results of major accident investigations consistently show that the human factor plays a very significant role in their development. Therefore, it is important to identify potential human errors and reduce the likelihood of failure of the human factor. The way how to deal with this problem is through an analysis of human factor reliability.

One of the elements of analysis of human factor reliability is also the analysis of the work load. This is both financially and time-consuming; the objective of the present article is to design an automated evaluation process using the virtual reality of HTC Vive and the Kinect system.

HTC vive and HMD are beginning to be widely used in a broad range of applications; such as in medicine (Pelargo et al. 2017) where it is used for education. In this case, it is deployed because of a large number of repetitive demonstrative operations in education and their adverse impact on the patient. It also emphasizes the predicative value in the cases of e.g. blood vessels. It is also deployed in treatment of various therapies. For example, in stress situations to determine if the patient does not suffer from height phobia (Bun et al. 2016). In the field of culture, for example, virtual tours of galleries (Choi et al. 2017) can be addressed. Also, in the field of fine art, primary applications have been made for fine arts and, from the point of view of artists involved in digital fine art, this technology is highly desirable. In engineering applications, we can see examples of the use of HMD technology such as the use of HMD in ship design (Pérez et al. 2015) or in the field of robotics students' education (Crespo et al. 2015). What all these articles have in common is the emphasis on the added value of virtual reality applications. In this case, a benefit is a greater clarity and thus a deeper understanding of the subject.

The present article focuses on ergonomics and a human factor in manufacturing systems and at production sites. There are many simulation programs available for modeling—Siemens Process Simulate, Siemens Classic Jack (Jack and Process Simulate Human 2017), Plant Simulation (Process Simulate 2017), IC: IDO (ESI Applications 2017) dealing with the ergonomic aspects of manual work in the early stages of design and planning of product manufacture. Jack and Process Simulate Human will allow for improvement of safety, performance and comfort of working environment with the use of digital human models. Work environments can be analyzed via virtual human models, while using a database of human body builds typical of the specified population. In our proposals, we can test a variety of human factors, including the

risk of injury, user comfort, accessibility, lines of the views, energy expenditure, fatigue limits and other important parameters. These products offer recommendations for more user-friendly designs throughout the entire design process to help save costs and time.

Key options include:
- Flexible human body builds that are anthropometrically and biomechanically accurate
- Supporting ergonomic analysis of total workforce using country-specific databases of workers and advanced anthropometric parameters
- A comprehensive set of ergonomic analytical tools
- An advanced positioning algorithm that can also analyze how the body responds to the force exerted in a particular direction
- Managing a wide range of workplace options that include work at different heights, stairs and ramps
- Views and analysis of field of view
- Envelope curves radius for fast workplace configuration
- Wide support for virtual motion capture technology including Microsoft Kinect® for Windows
- Support for virtual reality

A new platform in this field seems to be the use of the HTC Vive game console because it allows an observer to achieve a fully immersive insight into the virtual environment compared to the previous imaging headset systems. The HTC Vive console has the following parameters (Oculus Rift vs. HTC Vive: Prices are lower, but our favorite remains the same, 2017:

- Two displays panels with a resolution of 1,080 × 1,200 pixels
- Streaming video with 90 Hz refresh rate.
- Laser position sensors – 32 headset sensors and 48 sensors on the controls
- Microelectromechanical systems—gyroscope, accelerometer
- Minimum space for movement is 1.5 m × 1.5 m, with maximum of 4.5 m × 4.5 m.

It is also possible to use controls to interact with the virtual environment; in the case of HTC Vive, these are two controls captured using two infrared sensors.

The previous article (Plantard et al. 2017) dealt with the case of deploying MS Kinect in the assessment of the real workplace in both laboratory and real conditions. In this case, it would be obvious to use HTC Vive for further research for the above reasons.

It was interesting to assess how one is influenced by the deployment of the game console and whether it is possible to deploy this technology in the industry. For the assessment as in the previous article, a laboratory test should be performed by lifting a box of defined dimensions. This time, game console controls should be used to simulate the operator's hands, and the raised box should be virtual. The model case would fit into the overall concept of using virtual reality, i.e. to use the 3D model of the workplace prior to assembly. Using this technology, it would be possible to verify workplaces from an ergonomic point of view and a human factor without additional costs associated with building up of a dummy workplace, e.g. of cardboard engineering.

2 METHODS

As in the previous article, MS Kinect was used for validation of this method.

The test itself also took place under the RULA and REBA method, with 10 operators being involved in the measurement.

2.1 *Method RULA (Rapid Upper Limb Assessment)*

The RULA (RULA—Rapid Upper Limb Assessment 2017) method (Figure 1) was developed by ergonomists from the University of Nottingham and serves for a rapid and systematic assessment of the risk of damage to the musculoskeletal apparatus with respect to the upper limbs. In abroad, this method was used mainly for the assessment of upper limb disorders arising in the workplace and for the assessment of working postures.

2.2 *REBA method ("Rapid Entire Body Assessment")*

This method systematically evaluates musculoskeletal apparatus and is based on the RULA

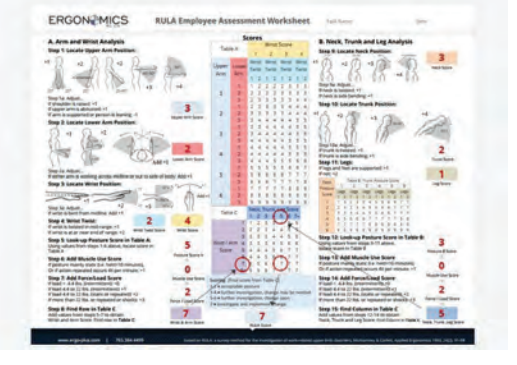


Figure 1. RULA method.

methodology. In abroad, the REBA method was used to assess ergonomic risks when working with imaging units and for risk assessment of healthcare staff. Both of these methods are a tool for postural analysis to evaluate biomechanical and postural loads of individual body parts. The body is divided into segments for individual scoring in relation to movement planes. Identifying of risk postures is very important for evaluation. These may be working postures that are physiologically unfavorable or occupied by the worker for most of the work shift. For RULA and REBA methods, the postures of individual parts of the body (arms, forearms, brace, neck, trunk and lower limbs) are scored with respect to the divergence from the neutral posture (Simulace výrobních procesů 2017). For each part of the body, the so-called basic postures are described to obtain a base score. It is a different range of flexions and extensions that are getting ascending points with increasing divergence from the neutral posture. There are also descriptions of postures to obtain additional points of the so-called variable score (e.g. rotation and bends sideways). The final assessment also includes the weight of the manipulated load (load-strength score) and the influence of the static postures at work (muscle score, activity score). In addition, REBA takes into account the impact of gripping techniques when handling the load (grip score) (Modern methods for the evaluation of ergonomic risks, 2017).

2.3 *Experimental procedure in laboratory conditions*

This section presents an experiment in simulated conditions. For this purpose, an experiment with 10 participants (age: 30 ± 6 years, height 1.78 ± 0.1 m, weight: 70 ± 10 kg) was performed. Each participant of experiment was equipped with the HTC Vive headset and handhelds for interaction simulating the hands. The movement of each participant was sensed (Figure 2), using two MS Kinect devices.

The reason for using the two MS Kinect devices to sense the participants was an insufficient sensing in singular postures when manipulating the virtual object, and also due to occlusion. The correction method described in (Planard et al. 2017) and (Diego-Mas et al. 2014) was also used to increase robustness.

The experiment itself consisted, as in (Plantard et al. 2017) in shiffting a 40 × 30 × 17 cm box between the starting and target positions (Figure 3). As in (Plantard et al. 2016), two target positions were defined as seen in Figure 4. The first placement of the target point was 1.7 m high and 0.35 m to the left and 0.5 m at the front. The second target position was 1.7 m high and 0.55 m to the left.

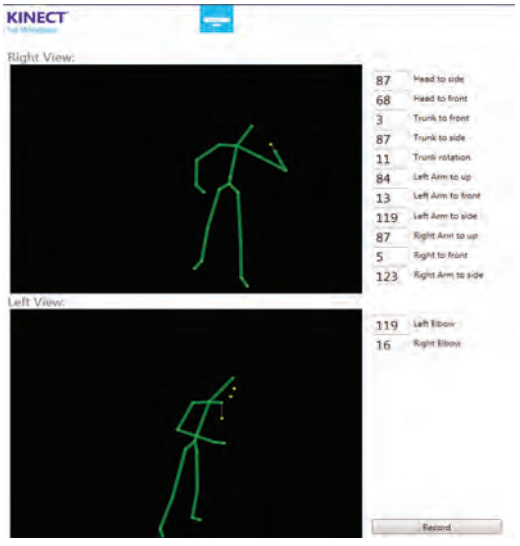


Figure 2. Record from cameras from two MS Kinect devices.

Figure 3. Course of experiment in laboratory conditions.

Figure 4. Positioning of target points in Unity3D environment with avatar.

Unlike in (Plantard et al. 2016), a virtual box of the same defined dimensions was used. The Unity 3D software was used to create the experimental environment.

2.4 *Data analysis*

In the laboratory experiment, we assessed the values of the joint angles calculated from the measured data from the Kinect device (using parallel scanning of two cameras) and compared them with the data evaluated by the expert estimate using the RULA principle.

3 RESULTS

Calculated RULA values and the values obtained by the expert estimate according to the RULA method:

The present article only evaluated the movement when lifting the object. The measured data was subjected to the Kolmogorov-Smirnov test for normality of the distribution of values. The data analysis has not shown normality.

Subsequently, the RMSE coefficient was calculated (Table 2) for each condition. Then we compared the resulting RULA score obtained from measurements with parallel use of two Kinect systems with a reference value obtained by expert estimate.

The results of the analysis show that the proposed RULA method of automatic measurement of RULA indicators provides relatively accurate results. Only the results for the Upper Arm score (left and right) are loaded with a major error; this could be reduced by changing the magnitude of field of view of the cameras. Other issues of measurement are a large number of singular points. These points could be eliminated using two MS Kinect devices scanning a moving person. However, this approach also has its disadvantages. An issue of measurement was the overshadowing of the limbs through the body. At this point, only one sensor was active. Therefore, in this case, measurements and data from the second sensor were not used, since it would put a significant error in the whole measurement.

Table 1. RULA score.

RULA	RULA—calculated	RULA—expert estimate
Upper arm score	2.35	2
Lower arm score	1.83	2
Score A Left (upper body)	2.79	3
RULA Grand Score Left	2.74	3
Upper arm score	2.31	2
Lower arm score	1.9	2
Score A Right (upper body)	2.78	3
RULA Grand Score Right	3.15	3
Neck score	2.51	3
Trunk score	2.24	2
Score B (neck, trunk and legs)	2.97	3

Table 2. Calculated RMSE values for each RULA indicator.

Score	RMSE
Upper arm score	1.05
Lower arm score	0.38
Score A Left (upper body)	0.83
RULA Grand Score Left	0.45
Upper arm score	0.98
Lower arm score	0.3
Score A Right (upper body)	0.7
RULA Grand Score Right	0.46
Neck score	0.75
Trunk score	0.44
Score B (neck, trunk and legs)	0.66

The advantage of this approach is to significantly reduce the cost of evaluation. Accuracy of the results is sufficient for practical measurements; however, for laboratory evaluation, it would be necessary to carry out further adjustments of the measuring procedure.

4 CONCLUSION

This article dealt with the idea of using the HTC Vive gaming console to analyze workloads. Comparison of the angles of movement is based to RULA method, within each body part, each angular assessment is divided into three sections and is tested for monotype actions. This method of automatic analysis has demonstrated its validity using assessment in a virtual environment.

Despite these constraints arising from performed measurement, HTC Vive can be considered as a promising tool for assessing workload analyses. Another indisputable advantage of the entire system is a very easy deployment in the industry, as it does not place high demands on software readiness of potential users. Another important role is also played by a good price value of the entire system. Another challenge will be to test the prepared solutions in assessing of the real production. The data from real-time measurement in traffic and the feedback from industry partners will provide us with a comprehensive picture of the use of these technologies. For a comprehensive examination of

the overall image of the production site design, further measurements including fine motor skills are required. This approach requires the deployment of other virtual reality technologies as well as other methods of assessing these movements.

ACKNOWLEDGMENTS

This work is an output of research and scientific activities of NETME Centre, regional R&D centre built with the financial support from the Operational Programme Research and Development for Innovations within the project NETME Centre (New Technologies for Mechanical Engineering), Reg. No. CZ.1.05/2.1.00/01.0002 and, in the follow-up sustainability stage, supported through NETME CENTRE PLUS (LO1202) by financial means from the Ministry of Education, Youth and Sports under the "National Sustainability Programme I".

REFERENCES

Bun, P., Gorski, F., Grajewski, D., Wichniarek, R. & Zawadzki, P. 2016, "Low—Cost Devices Used in Virtual Reality Exposure Therapy", *Procedia Computer Science*, pp. 445.

Choi, H. & Kim, S. 2017, "A content service deployment plan for metaverse museum exhibitions—Centering on the combination of beacons and HMDs", *International Journal of Information Management,* vol. 37, no. 1, pp. 1519–1527.

Crespo, R., García, R. & Quiroz, S. 2015, "Virtual Reality Application for Simulation and Off-line Programming of the Mitsubishi Movemaster RV-M1 Robot Integrated with the Oculus Rift to Improve Students Training", *Procedia Computer Science,* pp. 107.

Diego-Mas, J.A. & Alcaide-Marzal, J. 2014, "Using Kinect™ sensor in observational methods for assessing postures at work", *Applied Ergonomics*, vol. 45, no. 4, pp. 976–985.

ESI Applications. Available at: https://www.esi-group.com/cz/softwarova-reseni/virtualni-realita/icido/applications [Accessed December 12, 2017].

Jack and Process Simulate Human. Available at: (https://www.plm.automation.siemens.com/cz_cz/products/tecnomatix/manufacturing-simulation/human-ergonomics/jack.shtml [Accessed December 12, 2017].

Modern methods for the evaluation of ergonomic risks. Available at: http://www.bozpinfo.cz/josra/moderni-metody-v-hodnoceni-ergonomickych-rizik [Accessed December 12, 2017].

Oculus Rift vs. HTC Vive: Prices are lower, but our favorite remains the same. Available at: https://www.digitaltrends.com/virtual-reality/oculus-rift-vs-htc-vive/ [Accessed December 13, 2017].

Pelargos, P.E., Nagasawa, D.T., Lagman, C., Tenn, S., Demos, J.V., Lee, S.J., Bui, T.T., Barnette, N.E., Bhatt, N.S., Ung, N., Bari, A., Martin, N.A. & Yang, I. 2017, "Utilizing virtual and augmented reality for educational and clinical enhancements in neurosurgery", *Journal of Clinical Neuroscience*, vol. 35, pp. 1–4.

Pérez Fernández, R. & Alonso, V. 2015, "Virtual Reality in a shipbuilding environment", *Advances in Engineering Software,* vol. 81, no. C, pp. 30–40.

Plantard, P., Shum, H.P.H., Le Pierres, A.-. & Multon, F. 2017, "Validation of an ergonomic assessment method using Kinect data in real workplace conditions", *Applied Ergonomics,* vol. 65, pp. 562–569.

Process Simulate. Available at: https://www.plm.automation.siemens.com/en/products/tecnomatix/manufacturing-simulation/assembly/process-simulate.shtml [Accessed December 12, 2017].

RULA – Rapid Upper Limb Assessment. Available at: http://www.rula.co.uk/survey.html [Accessed December 12, 2017].

Simulace výrobních procesů. Available at: http://www.axiomtech.cz/24762-simulace-vyrobnich-procesu [Accessed December 12, 2017].

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Challenges with data for human reliability analysis

K. Laumann
Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

H. Blackman
Division of Research and Economic Development, Boise State University, Boise, Idaho, USA

M. Rasmussen
Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: A lack of empirical data is often been presented as a large challenge for HRA, which begs the question: why is this so difficult? HRA methods were not developed as objective quantitative test methods, but more as qualitative evaluation methods because objective data did not exist. Since HRA methods include substantial qualitative evaluation of the meaning of the elements in HRA methods, such as definitions of the performance shaping factors as well as their strength, these elements cannot be objective measured. This paper also discusses other challenges with collection data from event reports, literature reviews, experiments and databases. The conclusion in this paper is that a decision should be made about how we should look at HRA methods: as qualitative evaluation methods or objective quantitative test methods. Quantitative and qualitative methods have different approaches to evaluate the quality of the methods making it difficult to be something in between.

1 INTRODUCTION

Most of the Human Reliability Analysis (HRA) methods and techniques have been developed to estimate human reliability for tasks within Probabilistic Risk Analysis (PRA) in nuclear power plants. Human reliability analysis (HRA) was developed because of the lack of empirical data on human error probability. If data on the likelihood of human error on specific tasks existed, we would not need an HRA method. Williams (1992, p.20) said: "Therefore in cases where an assessor may have access to more specific and accurate task failure data, these should be used in preference to the HEART generic data-set." So if we had the data, the data should be used and not a HRA method.

HRA methods like; THERP (Swain & Guttmann, 1983), HEART (Williams, 1992), SPAR-H (Gertman,et al. 2005; Whaley, et al. 2011) and ATHEANA (Forester, et al. 2007), were developed as methods to support more qualitative expert judgements. The expert judgement provides gross estimates of failure probabilities for tasks defined by the PRA, when better data was missing. In the HEART manual (Williams, 1992, p. 4) states: "When considering system safety and reliability, engineers are generally concerned with gross changes in the probability of failure within system e.g. factors of 10, the proverbial order of magnitude. To be of value, therefore human reliability assessment techniques should be concerned with those factors, which are likely to produce probability of failure modification in excess of a factor of 3, and which, when cumulated, could produce significant changes in performance, and possible threaten system safety, operability and reliability."

Also HRA methods are often said to be simplified methods since it is often not enough resources in performing a comprehensive analysis which mean that the analyst is analyzing the most important influences on human reliability with limited time to read and understand guidelines and to perform the analysis.

Even though HRA analysis results in a quantitative likelihood for failure or success, HRA methods were not developed as positivistic quantitative test methods. HRA methods seem to be closer to a post positivistic research view. In positivism one assume that an objective reality exists, that this reality can be objective measured by scientific methods and that it is possible to develop scientific laws that can be generalized across settings (Guba & Lincoln, 1994). Within this approach, only quantitative research methods are used. In post positivism one assumes that an objective reality exists. However, this reality is complex and that

it can only be imperfectly apprehended and that we can never be sure that a true reality has actually been found (Guba & Lincoln, 1994). Within this approach, both qualitative and quantitative methods are used.

In spite of these characteristics of HRA methods described above, many authors claim that the biggest challenge in HRA is the lack of objective empirical data (for example; Boring et al. 2012; Hallbert et al. 2004; Kim, et al. 2015; Swain,1990; Williams, 1985). A question raised here is a question that was first presented in Laumann (in review) about what an HRA method actually is: Is an HRA method a qualitative evaluation method that leaves a lot up to the analyst qualitative evaluation and where mainly expert judgement was used to develop HRA? Or is an HRA method a quantitative test method where empirical tests where used to develop and test the methods? HRA today, is much more a qualitative evaluation method than a quantitative test method. Probably the words used about HRA methods such as "analysis" and "techniques" reflect the qualitative basis of HRA, rather than a quantitative.

In this paper, we will present challenges for HRA methods to be quantitative test methods and challenges with collecting quantitative HRA data. First, we will define challenges that exist for obtaining quantitative data for HRA that are the same challenges found with all kinds of quantitative methods. Then we will present challenges to obtain HRA data that exist with more specific methods such as; literature reviews, experiments designed to collect HRA data, event reports and databases. For simplicity, we have chosen to present example from SPAR-H (Gertman, et al. 2005; Whaley, et al. 2011) and HEART (Williams, 1992). However, the challenges for collecting empirical data presented exist for all HRA methods. SPAR-H and HEART were also chosen because these are methods where quantitative data collection have been much discussed.

2 DISCUSSION

2.1 *Challenges that exists for all kinds of data collections in HRA*

There are some demands for all kinds of quantitative test methods; they should be valid and reliable. HRA methods have a strong similarity to psychological test methods where human behavior is predicted from different "psychological" construct. HRA is also about predicting human behavior from constructs, since the elements in the methods such as Performance Shaping Factors (PSFs) or Error Producing Conditions (EPCs) are constructs, which are assumed to affect human behavior. In psychological test methods, validity is divided into different facets; content validity, construct validity, concurrent and predictive validity (Murphy & Davidshoffer, 2014). These demands will not be described here since they are well known and can be found in many textbooks as for example Murphy and Davidshoffer (2014).

A large challenge for HRA is content validity. Content validity is achieved by a) judgments and descriptions of the constructs and of the structure within these concepts and b) definitions and development of measurement scales to measure the constructs and their structure. If the content of the constructs, their structure, and how they should be measured (measurement scales) have not been clearly defined, then it is not possible to test the other aspects of validity such as construct validity, concurrent validity and predictive validity or reliability. Next, we will give some examples of content validity issues in two HRA methods SPAR-H and HEART. Laumann and Rasmussen have also presented challenges with content validity in SPAR-H in several papers (Laumann & Rasmussen, 2016; Rasmussen, Sandal & Laumann, 2015; Rasmussen & Laumann, in review).

The elements in SPAR-H (Gertman, et al. 2005; Whaley, et al. 2011) are: two nominal tasks with two nominal failure rates and eight performance shaping factors. The two nominal tasks in SPAR-H (Gertman,et al. 2005; Whaley, et al. 2011) are diagnosis and action. A task should be classified as diagnosis if it involves cognitive processing. An action involves limited cognitions. The separation between cognition and action within SPAR-H is very peculiar since probably nothing an operator does within an accident scenario is purely action without cognition. The diagnosis/action separation gives room for a much interpretations of what this actually means.

In the SPAR-H manual (Gertman et al. 2005), there are no specifications about what is meant by a task or at what task level the analysis should be performed. How a task is defined have a large effect on the result of the analysis or the probability for errors (for a discussion see Rasmussen & Laumann, 2017). SPAR-H leaves it up to the analyst to define at what task level SPAR-H should be applied and this gives much room for analyst choice and interpretation.

The elements in the HEART (Williams, 1992) method described in the user manual are 14 generic task types with a nominal human unreliability value and their suggested uncertainty bounds and 38 EPCs. For the EPCs the analyst should also assess the proportion of affects. Williams and Bell (2017) have recently reviewed HEART with a large literature review. Based on this review 32 out of the 38 original EPCs were kept, six of the EPCs were revised slightly and two new ones was incorporated in HEART.

It is difficult to find a definition of what a generic task actually is in HEART. It is said about generic tasks in HEART (Williams, 1992, p.8): "The first is the assumption that basic human reliability is dependent upon the generic nature of the task to be performed, i.e. for each task in life there is a basic probability of failure." So a generic task has something to do with the generic nature of the task which is not a very specific definition. Error producing conditions in HEART are defined as (Williams, 1992, p.1): 'Error producing conditions are factors that can affect human performance, making it less reliable than it would otherwise be.' The separation between what is a GTT and what is an EPC is not obvious in HEART, and some of the GTTs include elements that are very similar to the EPCs. This gives room for different interpretation by different analysts.

HEART defines that a task should be analyzed at the level that fits the GTT. How to analyze the proportion of affects or the strengths of the EPCs are not well defined in HEART, which gives much room for interpretation by the analyst.

To show an example of the difficulties with content validity, SPAR-H and HEARTs definitions of the PSF available time (SPAR-H) and time shortage (HEART) will be presented.

SPAR-H [Gertman et al. 2005, page 20] defines one of its PSFs available time as: "Available time refers to the amount of time that an operator or a crew has to diagnose and act upon an abnormal event. A shortage of time can affect the operator's ability to think clearly and consider alternatives. It may also affect the operator's ability to perform. Multipliers differ somewhat, depending on whether the activity is a diagnosis activity or an action."

The SPAR-H Step-by-Step (Whaley et al. 2014, page 2–4) give the following definitions of the levels for available time in SPAR-H:

Inadequate time—the time margin is negative because less time is available than is required.

Barely Adequate Time—the time margin is zero because the time available equals the time required

Nominal Time—there is a small time margin because the time available is slightly greater than the time required.

Extra Time—the time margin is greater than zero but less than the time required; the time available is greater than the time required

Expansive Time—the time margin exceeds the time required; the time available is much greater than the time required.

With these definitions, subjective evaluation depending on the characteristics of the tasks and the contexts are necessary to decide on a level for available time for a particular task. There is no way to objective define the level since there is no objective description of how much time one should assume to define the different levels. In addition, it is a question, what is the unit of analysis in SPAR-H. It has been claimed by the authors of SPAR-H that it is an analysis of the average operator. However, if there is barely adequate time for the average operator one might expect that there is too little time for the slower than average operators. How much failure one expects becomes circular with the definition of the unit and the unit of analysis is not well defined in SPAR-H.

HEART has a similar EPC, which is described as: "A shortage of time available for error detection and correction". HEART gives no advice on, how the analyst should go about analyzing this EPC. As shown under the discussion of SPAR-H, a lot of information needs to be clarified to analyze this EPC and since it is not available in the method, it is up to each analyst subjective judgement. Something that is peculiar with HEART is that the analyst is not instructed on how he/she should go about collecting information about the GTTs and the EPCs. The EPCs are defined by one sentence and then it is up to the analyst to interpret how this sentence fit their contexts/tasks, or what the sentence actually means, as for example—how much shortage of time should exist before the maximum predicted nominal amount should be chosen.

These characteristics with HRA methods (as SPAR-H and HEART) show that they are more qualitative evaluation methods than objective quantitative test methods and that they are far from being objective quantitative test methods. It is a question if we should expect methods that include so little definitions and descriptions (content validity) as SPAR-H and HEART to show interrater reliability. To obtain interrater reliability the concepts need to be precisely, defined. However, with a qualitative evaluation method view of the method, we will focus more on the analyst ability to predict correct error rates based on qualitative evaluations with use of a HRA method. With this approach, we might not expect high interrater reliability, but rather look at the quality of the data and the evaluations that the prediction is based on.

We claim that if HRA is going to be tested with quantitative methods they need to be improved and that the place to start is to develop good definitions of the content of the concepts included into the method. If good definitions exist, it might be possible to develop measurement scales for the PSFs/EPCs. If we have these measurement scales of the PSFs/EFCs it might be possible to predict how different levels of PSFs, such as for example complexity, affects performance.

However, it is a question, if this is possible. It could be that with the different elements included into HRA, it is impossible to be so well defined that quantitative measurements can be developed. For example, for time available, it could also be that the

tasks and contexts that are evaluated in HRA are so different that it is not possible to develop the exact meaning of the PSFs and the PSFs levels/strengths that counts for all kinds of contexts and tasks. If this is the case, the qualitative evaluation view of HRA methods might be better, but then also the qualitative part of the analysis need to be further developed.

2.2 *Challenges with data from psychological and human factors studies (literature reviews)*

HEART was developed based on human factor literature (Williams, 1992). Williams (1992) and Williams & Bell (2017) has done literature reviews to investigate studies that include EPCs and their maximum multipliers and nominal error rates on GTTs in experimental designs. Also, Laumann, Sandal & Rasmussen (Laumann & Rasmussen, 2016; Rasmussen, Sandal & Laumann, 2015; Rasmussen & Laumann, in review) have done literature reviews to investigate the meanings of the PSFs and how large effect the PSFs have on affecting human errors on tasks.

One challenge with using literature reviews on psychological and human factors studies to collect information for HRA is that these studies were not designed with the purpose of testing HRA methods and therefore it is difficult to transform the data to fit the HRA method. For example, literature studies usually only included one negative level for a PSF and this level is difficult to match with the level description for example in SPAR-H. It is also difficult to match the description of PSFs to the one manipulated PSFs in this kind of studies.

It is not obvious how the literature review to collect information on EPCs in HEART was done. Williams and Bell (2017) say that they have looked for the maximum multipliers of the EPCs. However, the human factors studies do usually not intent to manipulate the maximum multipliers of the EPCs. They usually just manipulate one level and it is often not described or discussed what this level actually is. In addition, the experiments usually intend to study why the PSFs/EPCs have an influence on performance rather than how much it affects performance. In the human factors literature there is also not developed measurement scales for the PSFs/EFCs. We have looked at some of the studies that are referred to by Williams & Bell (2017) and it is difficult to see that the maximum multiplier, in the meaning of 'the highest possible negative multipliers' for the PSFs/EFCs,' were the manipulations in these studies.

We are hopeful that Williams & Bell (2017) will present more from their literature review and discuss the evidence for the EPCs' maximum multipliers and nominal failure rates on the GTTs. In this way, other researchers can better understand the authors' arguments, and perhaps add to and relate this evidence to other methods. We do not think one should look at this data as objective evidence for an EPC/PSF but rather an evaluation done on the available evidence. Since it is an evaluation, it is important to understand the authors' arguments on for example, including or excluding experiments and the authors' argument about how similar the experimental manipulations found in these experiments are to the concepts in HEART.

2.3 *Challenges with performing new experiments to collect data that is relevant for HRA*

The unspecific definitions of the concepts in the HRA methods are also a large challenge for developing experiments since it is difficult to develop manipulations and measurements that fit with the HRA methods. An example of this is an experiment performed by Liu and Li (2014) where experimental data were compared to the multipliers in SPAR-H. In this experiment, one can see the difficulties the authors have in matching the definitions of the PSFs and the levels in their experiment to SPAR-H definitions and levels. For example, experience and training were defined as the 20 first trials as the negative level and the later 20 trials as the nominal level. This manipulation does not fit with the negative level description of experience and training given in SPAR-H, which is less than 6 months of relevant experience and training. It was difficult to develop an experiment manipulation that fits with this level description in SPAR-H. In this experiment, also complexity was manipulated, but the measurement of complexity measured the complexity of the procedures, and then it is a question if this should have been looked at as complexity or procedures. There were also questions on, how to match the manipulated levels to SPAR-H levels also for complexity and available time.

For the HEART EPCs, one should in experiments only manipulate the maximum strength of the EPCs, since these are the elements included into the method. However, usually, the maximum strength of an EPC, does not seem to be a meaningful experimental manipulation, if the "maximum multiplier" is interpret literary. For example, in an experiment on EPC 1 (Williams, 1992, p.22): 'Unfamiliarity with a situation which is potentially important but which only occurs infrequently or which is novel' one would give the participants no training on a completely new task. In this situation, one would have expected a human error probability for failure close to 1, and we might not need to test this because the result is too obvious.

For some of the GTTs in HEART and the nominal tasks in SPAR-H an obvious challenge for new experiments is the number of subjects needed, when errors is expected to occur in 1 of 100, 2 of 100 or 1 of 1000 subject.

There are many challenges with performing experiments that are relevant for HRA. One frequently mentioned challenge is that if these experiments should be done with actual operators in a simulated control room, the cost of the experiments are high and little data would be collected (Boring, 2012).

Another challenge that exists when performing experiments on PSFs (with both operators and other participants such as students) is that PSFs are difficult to completely control and without that control, it is difficult to measure the independent effect of the different nominal tasks and/or the PSFs/EPCs. We have seen in experiments that other PSFs than those manipulated often affect the results. For example, poor teamwork is a variable that is difficult to control, since this might exist within the crews before they come to the experiment, or develop during the experimental run. Other examples of PSFs that are difficult to control are; operators that are ill, hungry, stressed, fatigued or demotivated. The crews themselves might also increase the complexity of a scenario by some erroneous actions or by forgetting a procedural step. Then the manipulation for some crews might be different from the one the experimenter planned.

PSFs have a tendency to exist from before the experiments or occur during the experiments, they cannot be completely controlled, and some you might observe (e.g. poor teamwork) and other might be more hidden for the experimenter (e.g. fatigue, stress and illness).

As an example of this, the first author experienced that in an experiment at the Halden Reactor Project where we intended to manipulate available time and information load, but for some of the crews we also observed that poor teamwork also occurred. The poor teamwork and short available time combined had a very negative effect on performance for some of the crews (Laumann, Braarud & Svengren, 2005).

In addition, another challenge with simulator experiments with operators is that these simulators such as the simulator at the Halden Reactor Project, often are computer based simulators rather than analog simulators that the operators use at their own plant, making it difficult to know how much "the new interface for the operator" PSF interacts with the manipulated PSFs and how this affect error rates.

2.4 *Challenges with event report as a basis for HRA data*

One possible source of data in HRA is event reports. One problem with using event reports as a basis for HRA data is that the event reports only investigate and report when an error occurred and then we do not know if the PSFs are usually present when this task is done or if there were some particular PSFs that were present when the error occurred (Kim et al. 2015).

Another issue is that event reports are often written by operators that probably do not have much knowledge about PSFs/EFCs and how to investigate the presence of PSFs/EFCs. The event reports that we have seen have not been very specific about how they collected the data on the PSFs and how the data were interpreted. In addition, the strengths or levels of the PSFs are not defined in the event reports and much interpretation have to be done to decide on a specific level or strength.

Another problem with event reports might also be that the events occur so infrequently that they do not give much data for HRA (Boring et al. 2012).

It is also a problem for HRA that usually in the event report more than one PSF has occurred in the event, which makes it difficult to estimate the effects on orthogonal PSFs/EFCs, which is included in the HRA methods.

It is also a problem with event report that many organizations prefer to not be open about such matter as human errors and why they occur since this is regarded as sensitive information.

2.5 *Challenges with databases*

There have been several attempts to develop databases for HRA data, which included data from event reports, literature reviews, and/or from experiments or simulations. Examples of such databases are NUCLARR (Gertman et al, 1990). HERA (Halbert et al, and COREDATA (Gibson, Basra & Kirwan, 1999).

The challenges described for event reports, literature reviews and event reports are also challenges for databases because this is the information that is entered into the databases.

The general reason for a database is to organize data in some predefined ways. Also for databases, the unclear definitions in HRA are a large issue because when a data bases structure is developed the definitions, for example of PSFs/EPCs and PSFs/EPCs levels/strengths from one or more method, have to be used as template and the data then has to be interpreted based on these definitions in the database. Data from databases are never going to be better than the data included in the first place. To include quality data into a database, a good and precise structure and definitions, which were also used during the data-collection from either event reports or experiments is required.

Since the HRA methods, include so diverse definitions of the elements in the methods one structure in the databases for each method is necessary, which is very resource demanding.

In HRA, the structure and purpose of the databases are often not clearly specified and the

argument for them seems to be that one day, some analyst (a very smart analyst) could find a good way to analyze this data in a way that fits HRA. However if the developers of the database do not know more exactly how the database should be used and what is the purpose of it, the work invested in it might be useless. One might wonder if all the resources to develop HRA databases have been a good investment based on the amount of data relevant for HRA that has been provided so far.

3 CONCLUSION

HRA methods have been criticized for the lack of predictive data and validation of their results. However, is seems like HRA methods are criticized for not being something they never intended to be: quantitative test methods. It is not enough to say that HRA methods, are methods to estimate human reliability on tasks. Within HRA one should make a choice, whether we should look at HRA methods as a qualitative evaluation methods that gives gross and crude differences based on expert judgement or if HRA methods should be developed to be quantitative test methods. The inventors of SPAR-H and HEART seems sometimes to present their method as much more objective quantitative test methods than they actually are. Of course, this likely because they were developed to support probabilistic risk assessment where a quantitative result is required. ATHEANA went in another direction and defined an HRA method that is mainly a qualitative method, with an expert based quantification technique added at the back end. We think that also SPAR-H and HEART are mainly qualitative methods, requiring substantial analysts' judgement in order to produce the quantitative result.

To be a quantitative test method we need content validity and very good and specific definitions of the concepts the method includes, definitions of measurement scales for the concepts, definitions of who is the unit of the analysis, and definitions of how should a task be defined for that method. An important question is: Are these definitions possible within HRA? Maybe concepts such as PSFs and EPCs might be too difficult to be precisely defined, because the concepts include too much, and because they vary too much from context to context or from task to task. It might not be possible to develop definitions and measurement scales that can apply for all of the contexts and tasks where the HRA methods are used.

However, if we define HRA methods as qualitative evaluation methods, criteria for a good qualitative analysis should be developed and discussed. Qualitative research methods have other methods to evaluate the quality of the research than quantitative research methods. One paper by Laumann (in review) presents criteria for good qualitative analysis and discusses how these could be applied for HRA.

It might be that the definitions when a qualitative method is used do not need to be that specific and are more allowed to vary between contexts. However, even with a qualitative evaluation method view, as good as possible definitions and advice about how to perform the analyses, should be available.

This question about how we should look at HRA methods should be answered based on what we think about our data. Is it possible to precisely define the different elements in a HRA method that can be used across different contexts and tasks? If this is not possible, we have to collect HRA data with a qualitative approach.

After working for many years with HRA methods, definitions of PSFs and their levels and performing experiments within HRA, we doubt that it is possible to define and specify the PSFs/EFCs and measurement scales of the PSFs/EFCs enough that a quantitative approach is possible. An alternative for HRA then is to more focus on developing good qualitative methods for evaluation of PSFs/EPCs, PSF levels/EPC strengths, and error rates.

As HRA methods are today it would be a best to just admit that they are qualitative expert judgement methods trying to predict crude differences in performance, and that they are far from being objective test methods that can be empirical validated with quantitative methods. However, HRA methods are not good and systematic qualitative methods either and improvement in descriptions of how the qualitative analyses should be performed, are also needed. A qualitative method approach might demand lesser specification than a quantitative test approach.

One could argue, if HRA methods are not objective test methods why should they predict performance? There have been some studies to test the validity of HRA methods such as The international empirical study (Forester, et al. 2014) and the U.S HRA empirical study (Forester et al. 2014). These studies do not give an overall conclusion on how well HRA methods predict human errors. They give many and varied answer depending on the task, the HRA method and the analyst. However, in these studies one might wonder, what is actually tested? Is it the analysts' ability to use the HRA method to predict the likelihood of errors or is it the HRA method in itself that is validated? If one assume that the method was tested, the researcher should assure that the HRA method guideline was reliable followed by the analysts. This is not possible since some of the methods like SPAR-H and

HEART do not include complete and prescriptive descriptions of the qualitative parts of the analysis. In these studies, it seem to be the analysts' qualitative evaluation with use/help of the HRA method that was tested and not the method in itself.

HRA methods should not continue to be something in between qualitative and quantitative research methods, since then they are based neither on good qualitative research methods nor on good quantitative research methods. Qualitative and quantitative research methods have different assumptions about quality and have different ways to investigate the quality of the method or the quality of the research. A choice should be made within each method and the choice has to be made by the authors of the methods.

REFERENCES

Boring, R. et al. 2012 Microworlds, simulators, and simulation: Framework for a benchmark of human reliability data sources. In *Joint Probabilistic Safety Assessment and Management and European Safety and Reliability Conference,* 16B-Tu5-5.

Forester, J. et al. 2014. The International HRA Empirical Study. Lessons Learned from Comparing HRA Methods Predictions to HAMMLAB Simulator Data. *NUREG-2127*, US Nuclear Regulatory Commission, Washington, DC.

Forester, J. et al. 2014. The US HRA Empirical Study – Assessment of HRA Method Predictions against Operating Crew Performance on a US Nuclear Power Plant Simulator. *NUREG-2156,* US Nuclear Regulatory Commission, Washington, DC.

Forester J. et al. 2007 ATHEANA user's guide. *NUREG-1880*. 2007. Nuclear Regulatory Commission, Washington, DC: U.S.

Gertman, D. et al. 2005. *The SPAR-H human reliability analysis method*, NUREG/CR-6883. U.S Nuclear Regulatory Commission, Washington, DC, USA.

Gertman, D.I. et al. 1990. Nuclear Computerized library for assessing reactor reliability (NUCLARR), *NUREG/CR-4639*. Nuclear Regulatory Commission, Washington, DC, US.

Gibson, H. et al. 1999. Development of the CORE-Data database. *Safety & Reliability Journal*, 19, 6–20. Guba E.G, Lincoln Y.S. 1994, Competing paradigms in qualitative research. In N.K. Denzin & Y.S. Lincoln, (eds), *Handbook of Qualitative Research.* Thousand Oaks, CA: Sage; p.105–117.

Hallbert, B. et al. 2004. The use of empirical data sources in HRA. *Reliability Engineering and System Safety*, 82, 139–143.

Hallbert, B. et al. 2006. Human Event Repository and Analysis (HERA) System Overview. NUREG/CR6903. Nuclear Regulatory Commission, Washington, DC, US.

Kim, Y. et al. 2015. A statistical approch to estimating effects of performance shaping factors on human error probabilities of soft control. *Reliability Engineering and System Safety*, 142, 378–387.

Laumann, K. et al. 2005. The task complexity experiment 2003/2004, *HWR-758*. Institute for Energy Technology, Halden, Norway.

Laumann, K. In review. Criteria for qualitative methods in Human Reliability Analysis. *Reliability Engineering and System Safety*.

Laumann, K., & Rasmussen, M. 2016. Suggested improvements to the definitions of Standardized Plant Analysis of Risk-Human Reliability Analysis (SPAR-H) performance shaping factors, their levels and multipliers and the nominal tasks. *Reliability Engineering & System Safety*, 145, 287–300.

Liu, P. & Li, Z. 2014. Human error data collection and comparison with prediction by SPAR-H. Risk Analysis, 34, 1706–1719. DOI: 10.111/risa.12199.

Murphy, K.R. & Davidshofer, C.O. 2014. *Psychological Testing Principles and Applications*. Sixth edition. Person Education Limited, Essex, UK.

Rasmussen, M. & Laumann, In review. The evaluation of fatigue as a performance shaping factor in the Petro-HRA method *Reliability Engineering and System Safety.*

Rasmussen, M. & Laumann. 2017. The impact of decomposetion level in human reliability analysis quantification. L. Walls, M. Revie & T. Bedford (Eds.), *Risk, Reliability and Safety: Innovating Theory and Practice.* Taylor & Francis group, London, ISBN 978-1-138-02997-2.

Rasmussen, M. et al. 2015. Task complexity as a performance shaping factor: a review and recommendations in standardized plant analysis risk-human reliability analysis (SPAR-H) adaption. *Safety Science*, 76, 228–238.

Swain, A.D. 1990. Human reliability analysis: Need, status, trends and limitations. *Reliability Engineering & System Safety*, 29, 301–313.

Swain, D.A, & Guttmann H.E. 1983, Handbook of human reliability analysis with emphasis on nuclear power plant application *NUREG/CR-1278*, Washington, D.C, USA.

Whaley A.M. et al. 2011. The SPAR-H step-by-step guidance. *INL/EXT-10-18533, Rev 2*, Idaho Falls, USA.

Williams, J.C. & Bell, J.L. 2016. Consolidation of the human error assessment and reduction technique. L. Walls, M. Revie & T. Bedford (Eds.), *Risk, Reliability and Safety: Innovating Theory and Practice*. Taylor & Francis group, London, ISBN 978-1-138-02997-2.

Williams, J.C. 1985. Validation of human reliability assessment techniques, *Reliability Engineering*, 11, 149–162.

Williams J.C. 1992. *A user manual for the HEART*. Stockport, UK: DNV Technica Ltd.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Usability and user experience: Adaption and application for a railway related environment

Marc Burkhardt & Birgit Milius
Siemens AG Braunschweig, Germany

ABSTRACT: When developing consumer products, for a long time already usability is an issue and an integral part of the development process. For railway systems, usability is still a rather new area with new requirements. In our paper, we will discuss what makes usability and the newer research area of user experience for railways special and how certain features of the railway system, e.g. high requirements regarding safety will influence how usability can be included in the development process.

1 INTRODUCTION

For most industries, usability is a concept which is applied for years. In railways, at least in Germany, the application of usability methods and tools is still rather new. In our paper we give an overview about the general objectives of usability and user experience. We present an example for how usability methods are applied in the railway sector. In the later chapter, we discuss an aspect which makes usability in railways special. Railway applications have often safety requirements which need to be taken into account. This can lead to solutions which do not have a very good usability. We discuss this paradigm and its implications for the inclusion of usability in a railway systems development process.

2 USER EXPERIENCE AND USABILITY

Usability is defined in ISO 9241-11 as "*The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use*." This means that the technical systems should be designed in a way that a user can concentrate completely on the task at hand and is not distracted by the interface design. This means that a good usability can only be achieved when the resources and conditions of a human's perception and cognition are taken into account.

However, usability only looks at specific aspects of the human-machine-interaction. Especially in the last years the more holistic concept of user experience was developed and is researched. ISO 9241-210 defines user experience as "*a person's perceptions and responses that result from the use or* anticipated use of a product, system or service". User experience includes all the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments that occur before, during and after use. Even more so than usability, user experience is the result of the product, the user and the context/ environment.

To make a successful product, it is important to take both aspects into account. It would be a mistake to assume that because a product has a good usability the user experience is good. And also the other way around is not true. Just because a product is beautiful or aesthetically pleasing does not meant that it can be easily or intuitively used.

While it is obvious why a product which is used for pleasure, e.g. a game or a mobile phone should have a good usability and excellent user experience, some might argue that an emotional connection with a product is not necessary in a business setting. However, on a qualitative side, an emotional connection to a product or a system makes a person more relaxed and fosters a better immersion into work (Hassenzahl 2003, Vorderer 2005). Also on a quantitative side it can be shown that user experience is important. In Wright et al. (2007) is proven that personal well-being correlates with the performance in the job. Further research is needed but it can be concluded that especially in safety related industries this correlation can mean that satisfied and content users work more reliable.

Despite all the argumentation above, in practice the general ideas of usability are more often applied than those of user experience when developing new products. One of the reasons might be that we have a well founded set of methods and procedures to measure usability, whereas the same extended catalog of tools is missing for user experience.

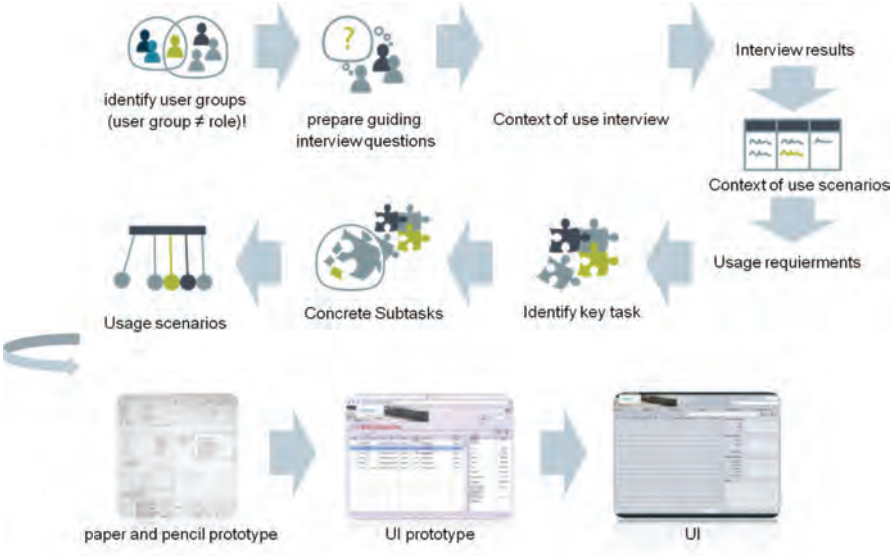


Figure 1. Usability design process at Siemens AG.

3 USABILITY IN RAILWAYS: AN EXAMPLE

In this chapter, we present an approach for designing a railway specific system.

Signaling systems in railways are very complex structures which need to be closely monitored to avoid major malfunctions which can lead to the disruption of railway traffic. Traditionally, the maintenance supervision of each system or system group was done separately, however, with the raising complexity and taking into account new and existing dependencies between such systems it became necessary to transfer all supervisory maintenance related tasks into one system. Such a system allows not only monitoring several systems, but also in cases of disruptions to coordinate all efforts regarding these systems with the aim of minimizing disruptions.

Traditionally, when developing such a user interface, functional requirements were the leading elements. Due to the new complexity, combined with usually less people being responsible, it becomes urgent to focus as much on functional aspects as it used to be but additionally to take human factors aspects into account to enable the user to fulfill the necessary tasks at a high quality. Also, the high reliability of new railway systems leads to very few failures. This has the effect that the procedures to recover from these failures are not as readily available as it would be if a person is confronted with them often. Therefore, a supervisory maintenance system needs to be able to provide detailed processes and a vast amount of information to enable and guide the user when solving problems and repairing systems.

At the beginning of most development processes in railways stands a detailed list of requirements. The complete list of requirements is describing the scope of the project and is used afterwards for verification and validation. The following list gives examples for how usability requirements were described.

- Req1: Identify, which user groups are interacting with the system and how they can be distinguished
- Req2: Identify on different level of detail, which information are need by which user/user group
- Req3: Identify, in which situations and in which environments/contexts will the system be used
- Req4: Identify, which systems with similar tasks/ in similar environments are used and known by the future user and which aspects of legacy systems[1] have to be taken into account
- Req5: Identify functions/tasks with safety relevance as these need special attention

The process as shown in Figure 1 was used to get a maximum of information from the users and

1. Legacy systems play a huge role in railways as signaling systems usually (at least in the past) had a very long life span. Even today, there are still mechanical interlockings in operation which date back to early last century.

to develop a User Interface (UI) with a very high usability. However, due to the fact how contracts in the area work, the application is not as straightforward as it looks.

Typically, the first step when deriving the user requirements is to identify the relevant user groups and develop the key questions which are necessary to derive all necessary information. However, especially when you develop a completely new system, the identification of the user groups is difficult. The (future) organizational structures and with this the functions and hierarchies are not necessarily known so assumptions have to be made. Experience has shown that for this, it is better for the development process to start with a more general description of the user groups and go into detail later than start with a too limited description. If the identified group of potential users is too small, danger is that important aspects are forgotten and consequently the User Interface (UI) cannot be used from all users with the same ease. Sometimes, it helps to think outside the box. In some projects, the user groups at the operators' side are generally known but the supplier has not the opportunity to talk to these people. Here it can help to interview people with indirect knowledge, e.g. from trainings to get a better understanding of the requirements.

During the interviews it is extremely important not to ask for specific functionalities but have people describe their daily tasks and how they are performed in the current setting. It is not the aim to listen to a general process description but rather to really understand the motivations behind actions and the needs to perform a task. This means that also short cuts or informal, typical processes should become known and are an important input in the development process. The approach becomes more difficult when safety related, process based tasks are to be discussed. If a usability study is done for, e.g. a traffic controller, the tasks of his jobs are described in detail in the rule book and in general it is not allowed to deviate from them. There is a danger that interviews will only show the officially allowed processes but potentially safety relevant changes are not talked about. Here, a longer field study is very helpful.

Like in most other IT industries the railway development process follows specific regulations. In railways e.g. the CENELEC norms DIN EN 50159 and DIN EN 50129 have very specific safety related requirements which in UI design have to be taken into account. In so called "context interviews" it is very important to understand all kind of input from of these regulations and their influence on the daily work. We recommend recording these interviews in audio and/or video. The experience shows that even if you have several well trained interviewers you could miss very important answers during the interview compared to a later "offline" analysis of audio or video recordings.

Figure 2. Iterative process when designing a UI as described in ISO 9241-210.

The next step in the process is to develop the real needs or demands from the system from the user's perspective. From this developed needs and demands the so called usage requirements will be generated. These set of usage requirements are the input to identify the key task and subtask of the users' daily work with the new system.

The goal of this separation of tasks is to find the concrete subtask a user wants or has to do and then develop usage scenarios out of these set of concrete sub tasks. This procedure is time consuming but definite worth it to do.

If you have your setup of usage scenarios it is time to start the first UI prototype design which then needs to be tested exceedingly. It is very much recommended to start the UI prototype design with a wire framework paper and pencil attempt. Start in simple black and white and concentrate on the content and not on the design. Use "placeholder" for parts which should be described in more detail later and concentrate on the current concrete subtask. Improve this paper and pencil design from subtask to subtask.

Even though we described the development process as straightforward, in reality it is iterative and will need several iterations until a final version of a UI is developed. The iterative process is visualized in Figure 2.

If you reach a first detailed design of your system start using design tools to develop a first digital version of your system design and improve this design from iteration step to iteration step. Colors and specific design elements should come up in the later iteration steps.

4 USABILITY FOR SAFETY-RELATED PRODUCTS: DIALOGUE CRITERIA

In the previous chapter we described the usability process as applied by the Siemens AG to a supervisory maintenance system. This process is of general

nature and only takes the special organizational requirements of the railway industry into account. Even though the process might not change significantly when safety aspects are to be taken into account, the contents of, e.g. interviews becomes more sensible. Also, at the latest when prototyping starts it will become necessary to balance good usability with the described safety-related processes.

As in any other industry, in safety related industries a good usability is necessary to help the personnel to work motivated and reliable. E.g., there is lots of research going on in the field of health engineering to provide safe, usable systems for health care. However, some of the aspects which are considered as making a product usable cannot be incorporated in a safety critical system. As an example, in this chapter, we have a look at the *seven dialogue criteria* from ISO and discuss, if and under which conditions these criteria can be used when assessing a safety relevant dialogue. As a railway example, we look specifically at the user interfaces for traffic operators.

Traffic operators usually work centralized and get all necessary information via screen. Typically, at least the tracks are shown, one can see which parts of the track are occupied and how the signals are set. Many more information is shown directly or via menus. The manipulation of the signals is done automatically, but they can also be handled manually. In these cases, the tasks of a traffic controller can be directly safety-critical. Besides information from the screens, traffic operators get or deliver information, e.g. by phone.

In the following paragraphs we discuss if the seven dialogue criteria can be used to evaluate the usability of a display.

The dialogue criteria are:

- Suitability for the task
- Conformity with user expectations
- Suitability for learning
- Suitability for individualization
- Self descriptiveness
- Controllability
- Error tolerance

An interactive system is suitable for a task, when it helps the user to complete the task. This means that functionality and dialogue are based on the task to be done. This is an aspect which should be applicable to safety systems. However, when dialogues are designed in rules and processes, this might not always be the case.

A dialogue conforms to user expectations when it behaves just as other dialogues in the given system. E.g. if a dialogue has to be confirmed, it should be confirmed every time in the same way (e.g. "confirm" and not sometimes "OK"). Also, it implies that the system behaves as the user expects it to do (SAP 2017a). This criterion can also be applied in a safety-related setting. As railways is an area where many aspects are standardised and it is an area which only develops very slowly, most people working in the area, e.g. most traffic controller, have the same professional background and their expectations to use the given systems are in general defined by the same technical systems. This might make it actually easier to define a new system, as the user group is very homogenous. This criterion can and should be applied when designing a display for a traffic controller.

A dialogue is suitable for learning when it helps the user to learn and understand the system. This aspect is less important when applied in safety-related settings, especially in railways. To work in a control centre, users had to undergo a rigid training. They are supposed to know the system very well and learning significant amounts of new aspects in the system would mean that the training is not good enough. In other areas, e.g. maintenance of systems, this aspect might be very sensible and can help to design systems which are actually making a task easier to achieve.

A dialogue is suitable for individualization when a user can adapt the man-machine-interaction and/or the display of information to his/her own preferences. Up to now, adaption of the man-machine-interface is not an option for traffic controllers. The reasons for this might be that the tradition exists to have the displays all looking the same. But there are also scenarios possible in which it is necessary that everyone gets the important information quickly, easily and reliably from the screen. This is the case, e.g. when in a difficult situation a second opinion is necessary. Today, it is not be possible to apply this criterion. For the future, research into the positive effect of individualization or changing general rules of cooperation in a train control center could result in the fact that the criterion is actually applicable.

Self descriptiveness means that it is obvious for a user at what point of the process he/she is working right now, what tasks are to be done and what options he/she has. In SAP (2017) it is explained that self-descriptiveness provides simplicity by reducing users' memory load. Users can retain their capacity for their tasks instead of bothering with the system. They can work more efficiently. For future developments this might actually be a game changer. The area controlled by just one controller gets larger all the time. There is the necessity to provide the controller with more and well structured information to always provide him/her with a well defined picture of the current situation. Also, railway displays contain many information, often in abbreviations. Applying the rule regarding self descriptiveness, this should be changed or the user should be well supported.

Controllability means that a user is able to start a dialogue as well as control its direction and

speed. In a safety related system sometimes it is almost always critical that dialogues are started and controlled by the system to provide on time information. The means, that the options for a traffic controller to influence the dialogue are limited. This might feel uncomfortable to the controller.

Error tolerance means that a user can make errors and can undo these with acceptable effort. Here, we have to distinguish between errors in operation in normal and degraded mode. In normal mode, errors can either not be made because the technical system prevents them or they are operational, then they can be redone as long as it is safe. There are situations, which are safe but lead to operational disturbances, e.g. when a train is led on a wrong track. In general, all measures to redo an operational error are governed by operational rules. In degraded mode, when technical systems are not or not fully available, an error can have directly safety critical effects. The display might show a critical situation but might not recognize it. Undoing such a safety-critical error is possible but in general time critical. Due to having accidents because of traffic operator errors (BadAibling 2016), it would be a very valuable research area to look at how a typical traffic controller display or complete system set up can be changed to at least help to recognize errors.

After having discussed the criteria one after the other, we will now look at a concrete example. A typical example in railways where the designed process does not fully comply with usability standards for dialogues is how safety critical actions are safeguarded. In these cases the so called "KF Bedienung (command release task)" has to be provided. After pushing the necessary buttons to, e.g. allow a train to pass a red signals, a window will pop up and ask the user if the planned action is correct and sensible. This has to be acknowledged after a givien time. In this time the user should take his time to be again aware of the complete situation on the tracks. The idea behind this is to make sure that such actions are not done lightly and by mistake. The authors are well aware that from a human factors point of view this process is flawed. However, for the sake of this paper, we only look at the design of the dialogue.

- Suitability for the task: To discuss the suitability of this dialogue for this task is an ongoing research questions. Due to several human factors, it can be argued the goal of this approach will not be reached. By designing the dialogue more intelligent as it is now, and taking into account how the brain usually works, the dialogue can be changed to be suitable for the task.
- Conformity with user expectations: As the traffic controllers are trained on these systems and the general processes are defined in rules and regulations the dialogue definitely confirms with the standards of the system and the expectations of the users.
- Suitability for learning: The dialogue not actually promotes a better learning or understanding of the system but focuses on repletion. The second dialogue can be confirmed (and probably often will be) without a deeper examination of the situation.
- Suitability for individualization: There is no individualization possible, as this type of dialogue is designed in the system.
- Self descriptiveness: The dialogue makes it clear what actions are expected from the controller.
- Controllability: The user does not have the chance to manipulate the dialogue, but he can stop the command.
- Error tolerance: There is little error tolerance. If the user actually made a mistake and is realising it, he/she cannot undo the wrongly given commando. Alternative processes are designed for these cases.

Overall, the general assessment of the usability of the dialogue regarding command release task is not very good. As it is known that there is a correlation between good usability and reliability/safety, the results of this assessment can be used to start a general discussion about how to design such processes in the future. The design process should take into account the basics of usability and combine them with the requirements of safety related, rule-based actions.

5 USABILITY AND THE SYSTEM LIFE CYCLE PROCESS

Another issue is the intervening of usability needs with the needs of the system life cycle. The system lifecycle is strictly structured in different phases in, e.g. several standards. In an ideal world the usability process for safety-critical systems will have an interaction with many of these system lifecycle phases. This was explained for general human factors issues in Milius (2017). However, as we do not have a very formalised process for human factors integration, it is easier to imagine an intervening of human factors and risk assessment. For usability, a very structured process already exists as was shown in chapter X of this paper. Combining both approaches is necessary for better products but also a difficult task.

6 CONCLUSION

In the previous chapter we stated the discussion regarding the integration of usability issues into the railway system development processes. We were focussing on usability rather than user experience

because the latter is still to new and provide several challenges before it can become a part of a formalized process. The given example has shown how usability is treated by Siemens Ag when a new, not directly safety related system is developed. Due to the fact that all tasks with a connection to safety operations are strongly regulated, even in the area of maintenance we have special aspects which distinguish railway usability from e.g. usability for a consumer product.

Another level of difficulty arrives when usability is applied to the design of safety-related systems. To show the possible effects, the generally acknowledged dialogue criteria were discussed and it was assessed if they can be applied to safety relevant functions. It is obvious that several of the criteria cannot be applied fully and/or directly. This was also shown using the command release tasks for safety-critical actions as an example. Here, lots of research needs to be done.

Usability is just a part of human factors. As was discussed in earlier papers (Milius 2017), recent incidents in railways show that even in highly automated systems humans still have to make safety-critical decisions. While the development in the past was focused heavily of replacing the human, newer research talks about effectively integrating the human in all necessary processes. The challenge will be that we start from a very different set-up as it used to be. As humans are more and more just supervisors, we have to make sure, that they still understand what's going on and we have to use modern technology to keep them informed and to provide help where necessary. This means that we have to integrate new functions in our new and existing systems. But it also means that we have to update existing processes to take the human, and usability, into account. Regarding general human factors integration, some ideas were already discussed in Milius (2017), but we need to focus on usability and life cycle as well. Last but not least, providing clear processes ho to integrate usability analysis in the general development process will lead to better acceptance.

REFERENCES

Bad Aibling 2016. https://en.wikipedia.org/wiki/Bad_Aibling_rail_accident. last checked at December 21, 2017. DIN EN 50129:2003-12; VDE 0831-129:2003-12.

Bahnanwendungen – Telekommunikationstechnik, Signaltechnik und Datenverarbeitungssysteme – Sicherheitsrelevante elektronische Systeme für Signaltechnik; Deutsche Fassung prEN 50129:2003. DIN EN 50159:2011-04; VDE 0831-159:2011-04.

Bahnanwendungen – Telekommunikationstechnik, Signaltechnik und Datenverarbeitungssysteme–Sicherheits relevante Kommunikation in Übertragungssystemen; Deutsche Fassung EN 50159:2010.

Hassenzahl, M. (2003). Attraktive Software – Was Gestalter von Computerspielen lernen können. In J. Machate & M. Burmester (Hrsg.), User Interface Tuning. Benutzungsschnittstellen menschlich gestalten (pp. 27–45). Frankfurt a. M.: Software & Support Verlag.

ISO 9241-11:1998. Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability.

ISO 9241-210:2010. Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems.

Milius, B. (2017). Human factors and their application in railways. In: Proceedings of the 27th European Safety and Reliability Conference (ESREL 2017). Safety and Reliability. Theory and Applications. CRC Press. ISBN 9781138629370.

SAP 2017. https://experience.sap.com/files/guidelines/simplification/principles/selfdescr.htm. Last checked December 21, 2017.

SAP 2017a. https://experience.sap.com/files/guidelines/simplification/ principles/consistency.htm.

Vorderer, P. (2005). Ernste Spiele. Retrieved June 30, 2017, from http://www.bpb.de/themen/8GJ1M3,0,0,Ernste_Spiele.html.

Wright, T.A., Cropanzano, R., & Bonett, D.G. (2007). The moderating role of employee positive well being on the relation between job satisfaction and job performance. *Journal of occupational health psychology*, 12(2), 93.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Human reliability analysis—accounting for human actions and external factors through the project life cycle

C. Morais
Institute for Risk and Uncertainty, University of Liverpool, UK
National Agency for Petroleum, Natural Gas and Biofuels (ANP), Brazil

R. Moura
National Agency for Petroleum, Natural Gas and Biofuels (ANP), Brazil
Institute for Risk and Uncertainty, University of Liverpool, UK

M. Beer
Institute for Risk and Reliability, Leibniz Universitat Hannover, Germany
Institute for Risk and Uncertainty, University of Liverpool, UK
School of Civil Engineering and Shanghai Institute of Disaster Prevention and Relief, Tongji University, China

E. Patelli
Institute for Risk and Uncertainty, University of Liverpool, UK

ABSTRACT: Airplanes, ships, nuclear power plants and chemical production plants (including oil & gas facilities) are examples of industries that depend upon the interaction between operators and machines. Consequently, to assess the risks of those systems, not only the reliability of the technological components has to be accounted for, but also the 'human model'. For this reason, engineers have been working together with psychologists and sociologists to understand cognitive functions and how the organisational context influences individual actions.

Human Reliability Analysis (HRA) identifies and analyses the causes, consequences and contributions of human performance (including failures) in complex sociotechnical systems. Generally, HRA research is concentrated in modelling workers' performance in the "sharp-end", assessing the ones directly involved in handling the system, especially operators. However, in theory, a reliability analysis can be applied to any kind of human action, including those from designers and managers.

This research will evaluate a way of conducting HRA in the design process, as previous research has demonstrated that design failure is the predominant contributor to human errors (Moura et al., 2016).

Bayesian Network (BN) – a systematic way of learning from experience and incorporating new evidence (deterministic or probabilistic) – is proposed to model the complex relationships within cognitive functions, organisational and technological factors. Conditional probability tables have been obtained from a dataset of major accidents from different industry sectors (Moura et al. 2017), using a classification scheme developed by Hollnagel (1998) for an HRA method called CREAM – Cognitive Reliability and Error Analysis Method.

The model allows to infer which factors most influence human performance in different scenarios. Also, we will discuss if the model can be applied to any human actions through the project life cycle—since the design phase to the operational phase, including their management.

1 INTRODUCTION

1.1 *Human Reliability Analysis (HRA)*

Human reliability analysis can have three objectives: identify human performance (as failures and their consequences), quantify the likelihood of failure (and error recovery) and to reduce or remediate those errors in the system (Kirwan, 1997).

The expected results of such study can be either qualitative or quantitative, depending on the industry sector best practice, data availability and regulatory requirements.

Quantitative results for HRA means giving the human performance a number, a probability of occurrence—the so-called Human Error Probability (HEP). This gives decision-makers the

opportunity to decrease the HEP to as low as practicable by tackling the factors that impact it, or to check if a certain risk criteria is met.

HRA research, practice and regulatory requirements are currently focused on operation and maintenance workers—called 'sharp-end' workers—those who actually interact with the processes (Reason, 1990; Hollnagel, 1998).

1.2 *HRA in design phase*

Can human reliability analysis be applied in other phases of an industrial project, such as design, construction, commissioning and decommissioning?

Theoretically speaking, it is possible: where there is human action, there is the possibility to model, analyse and measure performance (Hollnagel, 1998).

This research will focus on the design phase and design changes during other phases, as there is evidence from previous studies that design failure is the organisational factor that most triggers human failure actions (Moura et al., 2016).

One of the constraints of this approach is that human (engineers and managers) performance in the design stage has limited public data, preventing detailed task analyses.

However, it is also known that design failures identified in latter stages of the lifecycle (i.e. operational phase) are much more expensive to correct, compared to those detected during the design stage (Kohler and Moffatt, 2003).

Thus, it is believed that understanding engineers and managers performance during the design phase would have the potential to motivate improvements in organisational design procedures, based on overall accident patterns.

It is a trade-off between having perfect data but not sufficient resources to make design changes in the operational phase and having imperfect data but sufficient resources to improve the design in earlier stages of the lifecycle.

1.3 *Can human performance influence design?*

Design failure is often considered an organisational factor in HRAs, as the methods and assessors take into account that it influences human performance and not the opposite.

In contrast, there are studies, outside the safety and engineering community, showing that organisations are not an unmanned box of procedures, but individuals deciding whether using them, based on other factors like regulations, knowledge and resources (Rocha Fernandes et al., 2005). Besides, those individuals, usually middle and front-line managers, have a significant influence in all levels of the organisation, dictating and implementing the organisational strategy (Wooldridge et al., 2008; Purcell, 2007).

Another key aspect of this discussion is recognising the difference between 'managing the design' and 'managing design changes'. First, because they can occur in different phases of a project and thus managed by completely different team profiles.

Second, because decision-making in engineering practice can have two distinct meanings: 'design decisions' are the ones about the design itself (e.g. which equipment to choose in a system), while 'management decisions' are the ones about the team responsible for designing the system or issues that impact this team (Herrmann, 2015).

Each of these different concepts leads to a different kind of performance to analyse.

2 METHODOLOGY

2.1 *Classification scheme used*

The classification scheme is considered the collection of error modes (cognitive functions and human actions) and the Performance Shaping Factors (PSFs) which shapes the context that triggers each error mode.

To achieve the aim of this research, it is essential to use an HRA classification scheme that recognises cognitive functions, as both 'design decisions' and 'management decisions' cannot be evidenced only by actions described in most classification schemes from the first generation of HRAs.

For this reason, a classification scheme of the second generation of HRA was chosen (see Hollnagel, 1998, to understand the differences between the first and the second HRA generation).

From the publicly available ones, there are only two methods from the second generation that are considered useful to the Major Hazard Directorates of HSE, the UK safety regulator (Bell and Holroyd, 2009): CREAM (Hollnagel, 1998) and ATHEANA (Forester et al., 2007).

From these two choices, CREAM's (i.e. the Cognitive Reliability and Error Analysis Method) classification scheme was chosen to conduct this research, as it shows a clear distinction between causes and manifestations. This enables the application of the method in both directions: to analyse major accidents retrospectively and to predict events as a traditional HRA method. Therefore, this feature made it possible to use a pre-existent dataset from major industrial accidents (Moura et al., 2016) in the current work, as explained in the next section.

This classification scheme splits cognitive functions into two categories: analysis (the mental processes used when someone tries to understand a problem) and synthesis (the mental processes used

Table 1. Summary of errors of cognition used in CREAM.

		Observation missed
Analysis	Observation	False observation
		Wrong identification
	Interpretation	*Faulty diagnosis*
		Wrong reasoning
		Decision error
		Delayed interpretation
		Incorrect prediction
Synthesis	Planning	*Inadequate plan*
		Priority error

to solve the problem). Further, these are also split into subcategories, as summarized in Table 1.

One of the problems that may be argued against this choice is that most HRA performed in practice are the ones from the first generation (Zwirglmaier et al., 2015, Henderson and Embrey, 2012), such as THERP, HEART and SPAR-H. Also, according to CREAM's creator, all the PSFs presented at the classification scheme are still useful, apart from the cognitive reliability, that he considers a 'misleading oversimplification' (Hollnagel, 2012). According to him, "explaining human performance as based on 'cognitive processes' represents a myopic information processing view, and talking about the reliability of such processes is an artefact of the PRA/PSA mindset".

However, the current research is not using CREAM as an HRA method, limiting the discussion to the assessment of the HEPs, as a way to disclose possible improvements.

2.2 *Data used*

To generate Human Error Probabilities (HEP), or to validate HRA methods, different types of data are used. Kirwan (1997) classified them as: (i) real or operationally derived data (i.e. from incidents and near misses), (ii) simulator derived data, (iii) data from the psychological and ergonomics performance literature, (iv) expert judgement, (v) other techniques.

Data from real operation are considered the one with highest quality, but also the more difficult to obtain. That is because to achieve an absolute result for the HEP (number of observed errors by the number of opportunities for error) both the numerator and denominator of the equation should be assessed by the observation of each human action through an industry lifecycle. This is impractical, as one should count even the actions and errors that have not led to incidents.

For this reason, much research is being conducted using operationally derived data as near misses (i.e. events with the potential for undesirable consequences (CCPS, 2007) and accidents occurred in industrial installations.

Preischl and Hellmich (2013) used data from near misses, occurred on German nuclear power plants, to construct their model to estimate HEPs, in order to check validation of THERP handbook estimates. Groth and Mosleh (2012) have used the HERA database, from retrospective analyses of risk-significant events occurred on nuclear power plants, that contain at least one human error.

In the current research, it was decided to use a dataset derived from major accidents from different industrial sectors, not yet tested as model to estimate HEPs: the MATA-D – Multi-attribute Technological Accidents Dataset, built by Moura et al. (2016).

Differently from near misses reports, investigation reports of major accidents have the potential to uncover more PSFs that trigger a human error. This is because major accidents' investigations usually use several man-hours of an expert team (Moura et al. 2016)aiming to achieve an increased depth of analysis, eventually reaching root causes such as organizational issues (CCPS, 2007).

MATA-D has been derived from the analysis of 238 accident reports from different industrial sectors using the same classification scheme, with the intention to optimise the learning from cross-sector accidents. All the reports had evidence on the presence of organisational, technological and person-related factors, the PSFs described in Table 3. Also, nearly half of all the reports had indications about the cognitive functions and actions executed, described in Tables 1 and 2.

The MATA-D dataset is a table of 238 accidents by fifty-three parameters (thirty-nine factors, ten cognitive functions and four erroneous actions), where the number one represents the presence of a parameter in an accident report and the zero its absence.

For the reasons listed above, it has been decided to use the available MATA-D data to feed a model, in order to understand if the results could describe human performance in design and its management, instead of modelling the whole process, developing PSFs, collecting data and creating a new method from scratch.

2.3 *Modelling method used—Bayesian network*

The relationships between PSFs, cognitive functions and human erroneous actions described

Table 2. Erroneous actions used in the classification scheme.

Errors of execution			
Wrong time	*Wrong type*	*Wrong place*	Wrong object

Table 3. PSFs from CREAM classification scheme.

Organisational factors	Technological factors	Person related factors
Communication failure	*Equipment failure*	Memory failure
Missing information	Software fault	Fear
Maintenance failure	*Inadequate procedure*	*Distraction*
Inadequate quality control	Access limitations	Fatigue
Management problem	Ambiguous information	Performance Variability
Design failure	*Incomplete information*	Inattention
Inadequate task allocation	Access problems	Physiological stress
Social pressure	Mislabelling	Psychological stress
Insufficient skills		Functional impairment
Insufficient knowledge		*Cognitive style*
Temperature		*Cognitive bias*
Sound		
Humidity		
Illumination		
Other		
Adverse ambient conditions		
Excessive demand		
Inadequate workplace layout		
Inadequate team support		
Irregular working hours		

above were modelled into a Bayesian network (BN). BN is known as a systematic way of learning from experience and to incorporate new evidence (deterministic or probabilistic), and it was chosen due to the possibility of modelling those complex relationships within variables of different nature.

Mkrtchyan et al. (2015) had suggested that using BN, human reliability analysis also benefits from:

i. Its graphical formalism (Figure 1) of conditional probability equations (Equation 1). Using the visual representation of BN is a practical way of discussing the relations between factors, facilitating the communication between the multidisciplinary team that should be involved in an HRA meeting analysis, such as engineers, psychologists and sociologists.

$$P(C=c1 \mid A=a1,B=b1)\ P(C=c2 \mid A=a1,B=b1) = 1-P(C=c1 \mid A=a1,B=b1) \quad (1)$$

ii. A probabilistic representation of uncertainty, making it compatible with Probabilistic Safety Assessment.
iii. Combination of different sources of information: empirical sources as databases of events, theoretical models of human cognition and expert judgement.

The mathematical background of Bayesian networks was described by Tolo et al. (2014) as statistical models used to represent probability distributions, that can provide combined probability distribution associated to an accident, exploiting information about the existing conditional dependencies, e.g. between PSFs and cognitive functions.

BNs are represented by acyclic graphs, where nodes are connected to each other by arcs (Figure 1). Child nodes must have a causality relationship with each parent node.

For example, consider in Figure 2, the child node 'cognitive function'. The probability of its occurrence is conditioned to the occurrence of its parent nodes: organisation, technology and

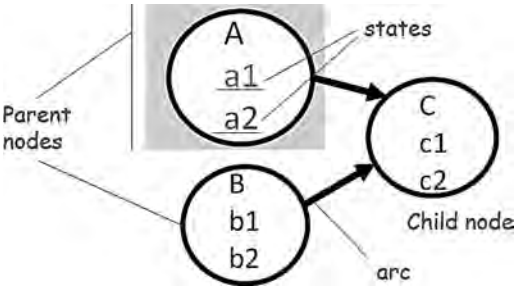


Figure 1. Directed acyclic graphs typical of a Bayesian network.

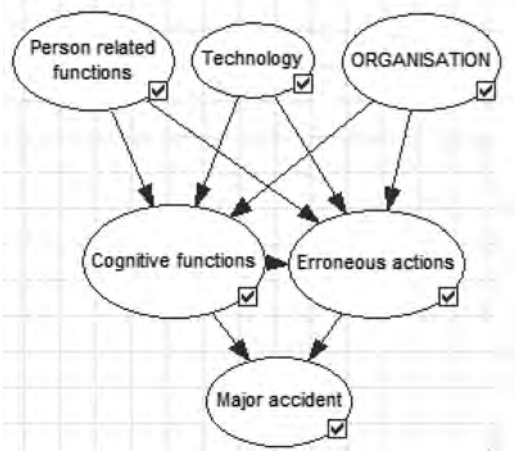


Figure 2. Example of a Bayesian network.

person-related functions. To have a proper causality, one has to know the answer to the question: what is the probability of occurring a cognitive function when the organisation, technology or person related factors occur altogether? What about when none of them occurs? And if only an organisational factor occurs, and no technology and person-related factor? All possible eight combinations from three parent nodes have to be answered, to establish a proper causality.

Generically speaking, the number of combinations a conditional probability table has to represent a child node is two (pair of combinations) to the power of the number of parent nodes (2^parent nodes).

This means a high number of combinations if all the factors of the CREAM methodology are considered. The implications of this issue are discussed in the next section.

3 MODEL

3.1 *Bayesian model of human reliability*

To build and test the human reliability model, it was used the summarised process represented by Figure 3. It was proposed by Mkrtchyan et al. (2015), through their review of HRA methods using BN models.

First, the nodes and their states were defined. Then, the structure, which means the links between the nodes. After the structure, comes the assessment of Conditional Probability Tables (CPT) for each node. Finally, a verification was conducted. The validation process will be conducted in a future work.

3.2 *Nodes and states*

The nodes used in the model are the sub-factors of CREAM classification scheme (Hollnagel, 1998), where the major factors are human, technology and organisation.

The states of the nodes will be 'presence' or 'absence' of the sub-factors observed during the investigation of major accidents.

The result of the MATA-D dataset, presented in the methodology section of this paper, has fed our model as a matrix of zeros and ones of 53 rows × 238 columns. At the dataset, the absence of a parameter (factor, cognitive function or action) is represented by the number zero and the presence of them in an accident represented by the number one.

Only the factors, cognitive functions and actions in *italic* in Tables 1 and 2 were used as nodes for the model. The reason is explained in the next section.

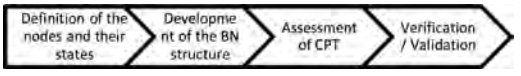


Figure 3. Process to build a BN model to HRA.

3.3 *Structure*

Basically, to create the structure of this BN model, parent nodes (organisational, technological and person-related factors) were linked by arrows to the child nodes (cognitive functions and human erroneous actions).

It would be that simple if there were no limitations from the algorithm used to build the model in Genie software. For the reason explained in section 2.2, the thirty-nine factors provided by the classification scheme would generate 549,755,813,888 combinations (two to the power of thirty-nine) – more that was supported by the BN software used.

The algorithm supports a large number of nodes, but not a large number of connections to one child node. Therefore, it was necessary to make assumptions to simplify the model structure.

To make assumptions about connections between nodes, one must have a clear understanding of the causal relationship that factors transmit to cognitive functions.

That is the reason why the most common way to simplify a model, for human reliability purpose, is using expert judgement, also known as expert elicitation (Mkrtchyan et al. 2015). However, it is also the stage where happens one of the most claimed disadvantages of using Bayesian networks for human reliability analysis: it is argued that experts can bring more uncertainties to a model due to their personal bias.

In an attempt to avoid this kind of uncertainty, the strategy was to let the data 'speak' for itself. This strategy has been already used by Groth and Mosleh's (2012) at their BN model, where they had introduced nodes of 'error context' to align certain combinations of PIFs that are more likely to produce human errors than the individual PSFs acting alone.

At the present work, the 'error context' was represented as the arcs of the BN model instead of the nodes. The context was imported from the treatment applied by Moura et al. (2017) to their dataset, to disclose common patterns and significant features among major accidents. They have used an artificial neural network approach to the dataset, a data mining process that translated the information into a graphical interface, the self-organising maps (SOM). Analysing them, one can perceive that the 238 accidents are allocated into four different regions, shaped by the clustering of accidents with a similar profile and, thus, a similar combination of factors, cognitive functions and actions.

Summing up, the model connections were proposed based on those SOM relations: factors that were in the cluster #1 were linked only to cognitive functions located on the same cluster, and the same process was repeated for all the clusters.

Simplifications to the network structure were applied not only to the connections but also to the number of nodes. Using previous research by Moura et al. (2017), the nodes were restricted to the factors, cognitive functions and actions considered significant for the dataset of major accidents by the self-organising maps algorithm.

Consequently, if a factor had represented negative or very low variations in the formation of one of the SOM clusters, it was interpreted that that factor was not significant to the causation pattern of major accidents and, consequently, it was not included in the Bayesian model presented in this paper. The considered nodes are presented in italic in Tables 1, 2 and 3, and in the nodes represented in Figure 4.

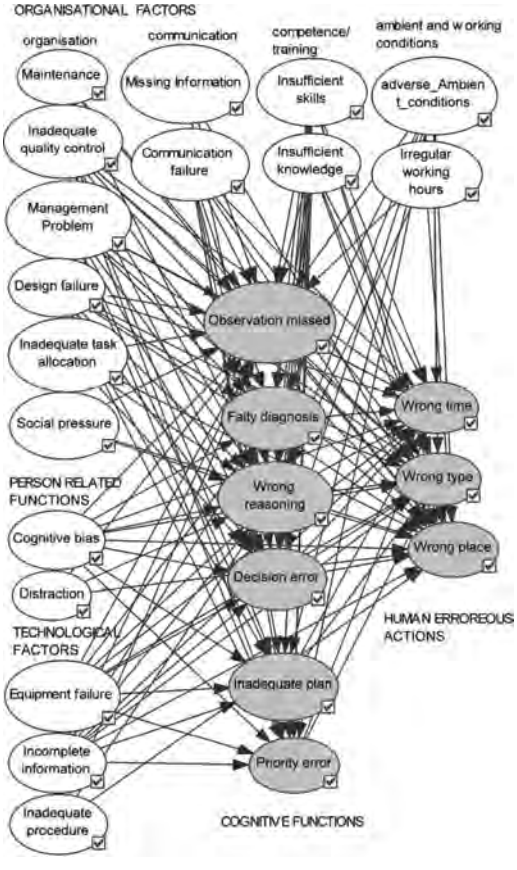


Figure 4. Bayesian model considered.

The model considers that cognitive functions are affected by each factor and that human erroneous actions are affected by the factors and by the cognitive functions. That is because the model assumes that workers have a mental process behind their actions (Hollnagel, 1998).

3.4 *Conditional Probability Tables (CPT)*

Conditional probability tables have been obtained from the dataset of major accidents from different industry sectors (Moura et al. 2017), using the CREAM classification scheme (Hollnagel, 1998).

After the simplification on the network, the higher number of combinations to a child node reached was the node 'wrong place', with nineteen parent nodes combinations, considering the possibility of occurring a 'wrong place action' influenced by sixteen factors and three cognitive functions. That means 524,288 combinations, and thus 524,288 probabilities of an action to occur or not.

The prior probabilities of the model were obtained by calculating how many times a specific combination occurred, divided by the total number of accidents of the dataset.

3.5 *Software used*

The accident dataset developed by Moura et al. (2016) has been originally built as a table of zeros and ones, that was uploaded to the BN software.

If the number of combinations was small, an Excel spreadsheet could be used to find the CPT and export to GeNIe software. However, as presented in the previous section, the dataset generated conditional probability tables of 524,288 probabilities for some child nodes. This row (or vector) of data extrapolates Excel software limits, and thus Matlab had to be used. Also, to optimise the data gathering, it was necessary some coding skills to create the Conditional Probability Tables—as 'filtering' the combinations in Excel would consume too much time.

The BN model was built in GeNIe Modeler for academic use (BayesFusion, LLC). The clustering algorithm embedded in the software was used to calculate the posterior probabilities, and the node type used was 'chance—general'.

Useful explanations of how to use the GeNIe software can be found in the manual provided by the developer and also by the authors of an evacuation time analysis of ships using BN (Sarshar et al., 2013).

4 RESULTS

After building the model, inserting the prior probabilities of parent and child nodes (through their

Table 4. Marginal probabilities of human performance.

Cognitive functions		Actions	
Observation		Wrong time	4.21e-04
Observation missed	8.12e-05	Wrong type	1.73e-04
Interpretation		Wrong place	4.64e-04
Faulty diagnosis	1.05e-03		
Wrong reasoning	1.05e-03		
Decision error	6.35e-04		
Planning			
Inadequate plan	1.22e-03		
Priority error	6.50e-04		

conditional probability tables), the marginal probability distributions were calculated using Genie software, as presented in Table 4:

Although a validation was not yet conducted, the order of magnitude of the HEPs (cognitive functions and erroneous actions) is consistent with HRA directives from the Oil & Gas industry (OGP, 2010), and HRA documents obtained at the website of the Environmental Protection Department of The Government of the Hong Kong (2017). A validation is needed to understand if the model is optimistic or pessimistic, according to validation criteria discribed by Kirwan (1997).

4.1 *Verification step*

To verify if the model behaves according to its specifications, some scenarios were created, changing the factors to its extremes. It means that each parent node was assumed to be 0 and 1 separately. In other words, each factor (organisational, personal and technological) was assumed to be absent or present in an industrial scenario.

To achieve that, after changing the factors, the posterior probabilities of the human performance nodes were calculated, updating the Bayesian network.

4.2 *Sensibility of human performance to each factor*

To infer which factors most influence human performance, the results from the verification process have been used.

When a factor node was set as present in a scenario (state 1 of the node), it was assumed that the variation caused to the posterior probability of a human performance node is the sensibility of this parameter to that change. Note that the parameters are represented by small probabilities, so changes in their marginal probability are also expected to be small. Thus, for better visualisation of the sensibility, the variation in percentage has been calculated.

As can be noticed in Table 5, there is a slight increase in the presence of missed observations when Inadequate Quality Control and Design Failure, both organizational factors, are present in a scenario. Moderated positive variation is also perceived when the technological factors equipment failure and inadequate procedure are present. Interpretation functions (faulty diagnosis, wrong reasoning and decision error) are the most influenced parameters by changes in organisational and technological factors.

Errors in design and equipment failures also increase errors in interpretation functions (mainly in wrong reasoning), but the results show a more accentuated positive variation at interpretation when changes of quality control, task allocation and knowledge (related to training) occurs.

While interpretation has its presence affected by training (knowledge), planning incidence seems to be more related to experience (skills).

Failures in equipment increase the possibility of poor planning in an accident scenario, but not as much as the inadequate quality control and design failure, both organisational factors.

Errors of execution (wrong time, type and place) are triggered by quality control, design failure, task allocation and, with fewer effect, equipment failure.

The results suggest that some factors do not make cognitive functions vary positively, possibly meaning that an error on cognitive functions would not occur with a scenario with errors in factors as maintenance, management, social pressure and irregular working hours.

5 CONCLUSION AND DISCUSSIONS

The Bayesian model proposed was built to serve as a tool to predict human performance in industrial scenarios, i.e. the human failure probability, with the potential to be applied in different sectors such as chemical (including oil & gas processing), nuclear and aviation.

The novelty of the present model is the use of a dataset of major accidents to define the basic aspects of a Bayesian network HRA model: nodes, states, structure and prior probabilities (conditional probability table). The model was developed as one of the objectives to achieve the aim of understanding human performance in the design phase.

Some discussions are developed below in the form of answers to questions proposed for this research.

5.1 *Can this model be used for HRA purposes?*

Not yet. All the steps of the model were executed, apart from the validation. The intention is to

Table 5, Sensibility of cognitive functions and erroneous actions to organisational, technological and personnal related factors.

	INTERPRETATION						PLANNING						ACTION					
OBSERVATION	Observation missed	Variation %	Faulty diagnosis	Variation %	Wrong reasoning	Variation %	Decision error	Variation %	Inadequate plan	Variation %	Priority error	Variation %	Wrong time	Variation %	Wrong type	Variation %	Wrong place	Variation %
Prior probabilty from dataset	8.1E-05		1.0E-03		1.05E-04		6.4E-04		1.2E-03		6.5E-04		4.2E-04		1.7E-04		4.6E-04	
ORGANISATIONAL FACTORS																		
Organisation																		
Communication failure = 100%	3.3E-05	−60	5.8E-04	−44	1.2E-06	−99	2.6E-05	−96	1.9E-04	−84	7.6E-05	−88	3.5E-07	−100	3.1E-10	−100	4.2E-05	−91
Missing Information = 100%	5.6E-05	−31	1.0E-03	0	7.0E-05	−33	6.4E-04	0	1.2E-03	0	6.5E-04	0	4.2E-04	0	1.7E-04	0	3.6E-05	−92
Maintenance = 100%	3.1E-05	−61	1.0E-03	0	5.9E-05	−44	6.4E-04	0	1.2E-03	0	6.5E-04	0	4.2E-04	0	1.7E-04	0	4.0E-04	−14
Inadequate quality control = 100%	9.4E-05	16	1.5E-03	46	1.6E-04	52	4.9E-04	−24	1.7E-03	39	8.5E-04	30	4.6E-04	9.0	1.8E-04	1.5	6.7E-04	44
Management problem = 100%	1.4E-06	−98	2.0E-05	−98	1.6E-06	−99	1.4E-05	−98	1.2E-04	−90	1.6E-04	−76	8.4E-05	−80	1.3E-04	−24	3.3E-05	−93
Desig failure = 100%	8.9E-05	9.3	1.4E-03	32	1.5E-04	41	6.4E-04	0.9	1.7E-03	38	6.9E-04	6	5.4E-04	28	1.2E-04	−29	5.2E-04	13
Inadequate task allocation = 100%	7.6E-05	−7	1.0E-03	0	1.6E-04	54	9.2E-04	45	1.2E-03	0	6.5E-04	0	4.2E-04	0	1.7E-04	0.37	6.2E-04	35
Social pressure = 100%	9.6E-06	−88	1.0E-03	0	8.2E-07	−99	6.4E-04	0	1.2E-03	0	6.5E-04	0	4.2E-04	0	1.7E-04	0	4.6E-05	−90
Competence/training																		
Insufficient skills = 100%	6.4E-05	−21	1.3E-03	23	7.3E-05	−31	1.6E-04	−75	1.5E-03	20	5.0E-04	−23	2.7E-04	−37	1.9E-04	8.8	2.4E-04	−48
Insufficient knowledge = 100%	6.1E-05	−25	1.6E-03	55	2.2E-04	108	2.4E-04	−62	9.7E-04	−20	4.2E-04	−36	3.1E-07	−100	3.9E-05	−78	1.8E-04	−60
Ambient and working conditions																		
Adverse ambiente conditions = 100%	8.8E-06	−89	0.00E+00	−100	0.0E+00	−100	6.5E-05	−90	4.4E-05	−96	9.5E-05	−85	3.1E-07	−100	4.1E-07	−100	4.0E-07	−100
Irregular working hours = 100%	1.5E-07	−100	1.0E-03	0	1.1E-07	−100	6.4E-04	0	1.2E-03	0	6.5E-04	0	4.2E-04	0	1.7E-04	0	7.8E-06	−98

PERSON RELATED FACTORS																				
Cognitive bias = 100%	1.0E-05	–87	4.4E-05	–96	1.2E-05	–88	5.9E-06	–99	5.8E-06	–100	3.8E-06	–99	2.8E-12	–100	1.6E-05	–91	2.1E-07	–100		
Distraction = 100%	4.8E-06	–94	1.0E-03	0	5.3E-07	–100	6.4E-04	0	1.2E-03	0	6.5E-04	0	4.2E-04	0	1.7E-04	0	6.3E-07	–100		
TECHNOLOGICAL FACTORS																				
Equipment failure = 100%	9.0E-05	11	1.0E-03	–2	1.6E-04	49	3.9E-04	–38	1.4E-03	15	2.3E-04	–64	4.4E-04	4.4	2.8E-05	–84	4.6E-04	0.1		
Inadequate procedure = 100%	9.0E-05	11	1.0E-03	0	7.1E-05	–33	6.4E-04	0	1.2E-03	0	6.5E-04	0	4.2E-04	0	1.7E-04	0	5.3E-04	14		
Incomplete information = 100%	4.7E-05	–42	1.1E-03	1	1.0E-05	–91	1.4E-04	–78	1.2E-04	–90	2.0E-04	–69	2.0E-04	–54	6.2E-06	–96	1.0E-04	–78		

validate the model against data from recent major accident that are not yet covered by the original dataset, to measure if the model describes other real operational data.

In future works, the dataset can be adapted to PSFs of classification schemes from other HRA methods.

5.2 *Is the model able to describe human actions through all the project life cycle?*

Although the verification step suggests the model is capable of inferring which factors most influence human performance, the model still interprets design as a PSF affecting human performance in other stages—and do not consider PSFs affecting designers. In addition, it is believed that some factors may influence design phase in a different way from the prior probability extracted from the dataset of major accidents

Although it is believed that this model cannot describe the design phase, it has the potential to describe changes in design during the operational phase. In fact, that is one of the uncertainties that need to be investigated in the dataset used, as some situations described as a design failure can be attributed to changes in the initial design during latter phases. This can also change part of the prior probabilities considered in this model.

Further development must consider the creation of new PSFs, with new organisational factors that should be considered during design. It seems that expert elicitation will have to be considered phase—as there is few public evidence of this process.

5.3 *Can this model be used to understand decision-makers performance during design?*

The model has the potential to describe front-line and middle managers' routine and emergency performance, during the operational phase. It is expected that it will give a better description of cognitive functions than actions, as a reflection of the decision-makers typical job description.

However, further investigation should be conducted to understand how specific factors affect different people in the organisation, specially for organisational factors for which results suggested that the impact in cognitive function is marginal, such as social pressure and irregular working hours. These factors, for instance, are reported in the literature (Thomas et al., 1999) to affect middle-managers in a different way, compared to sharp-end employees.

The importance of understanding managers' safety performance is part of the present research, as a way of investigating if improving their performance on safety issues has the potential to lead

industries to better organisational factors and fewer accidents. Managers are linked at the same time to top management and operational teams—having the opportunity to sell new ideas to top management and to promote strategic change to employees (Wooldridge et al., 2008; Purcell, 2007).

For this reason, it is recommended that further models consider all factors proposed by CREAM's classification scheme, instead of only using the significant ones according to previous research (Moura et al., 2017). Accounting for factors like 'excessive demand' and 'cognitive style' might give an improved model for managers' roles. With this purpose, different software and algorithms to calculate posterior probabilities have to be tested, to support more links between nodes. Therefore, Cossan (Patelli et al., 2018), a software for Uncertainty Quantification and its Bayesian network toolbox will be tested in the future.

REFERENCES

BayesFusion, LLC. GeNIe Modeler. http://www.bayesfusion.com/, Accessed: 30 November 2017.

Bell, J. and Holroyd, J., 2009. Review of human reliability assessment methods. Health and Safety Laboratory.

Center for chemical Process Safety – CCPS, Guidelines for Risk Based Process Safety. 2007.

Environmental Protection Department of The Government of the Hong Kong, Project file submitted for application for Environmental Impact Assessment. Chapter of Human Error Assessment & Reduction Technique. http://www.epd.gov. hk/eia/register/report/eiareport/eia_2242014/EIA/app/app12.10.pdf (Accessed: 11 December 2017).

Forester, J., Kolaczkowski, A., Cooper, S., Bley, D. and Lois, E., 2007. ATHEANA user's guide. NUREG-1880, NRC.

Groth, K.M. and Mosleh, A., 2012. Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model. Proceedings of the Institution of Mechanical Engineers, Part O: *Journal of Risk and Reliability*,226(4), pp.361–379.

Henderson, J. and Embrey, D., 2012. Quantifying human reliability in risk assessments. Petroleum review, 66(790).

Herrmann, J., 2015. Engineering Decision Making and Risk Management. New Jersey: John Wiley & Sons, Inc.

Hollnagel, E., 1998. Cognitive reliability and error analysis method (CREAM). Elsevier.

Hollnagel, 2016. Disclaimer about CREAM. http://erikhollnagel.com/ideas/cream.html (Accessed: 13 December 2017).

International Association of Oil & Gas Producers (IOGP), 2010. Human factors in QRA. Risk Assessment Data Directory, Report No. 434–5.

Kirwan, B., 1997. Validation of human reliability assessment techniques: part 1—validation issues. *Safety Science*, 27(1), pp.25–41.

Mkrtchyan, L., Podofillini, L. and Dang, V.N., 2015. Bayesian belief networks for human reliability analysis: A review of applications and gaps. *Reliability engineering & system safety*, 139, pp.1–16.

Moura, R.; Beer, M.; Patelli, E.; Lewis, J. & Knoll, F. Learning from past accidents to improve system design *Safety Science*, 2016, 84, 37–45

Moura, R.; Beer, M.; Patelli, E. & Lewis, J. Learning from major accidents: graphical representation and analysis of multi-attribute events to enhance risk communication Learning from Incidents Special Issue of the Safety Science Journal, 2017

Kohler, N. and Moffatt, S., 2003. Life-cycle analysis of the built environment. Industry and environment, 26(2–3), pp.17–21.

Patelli, E.; Tolo, S.; George-Williams, H.; Sadeghi, J.; Rocchetta, R.; Angelis, M.D. & Broggi, M. OpenCossan 2.0: an efficient computational toolbox for risk, reliability and resilience analysis Proceedings of the joint ICVRAM ISUMA UNCERTAINTIES conference, 2018 (submitted)

Preischl, W. and Hellmich, M., 2013. Human error probabilities from operational experience of German nuclear power plants. *Reliability Engineering & System Safety,* 109, pp.150–159.

Purcell, J. and Hutchinson, S., 2007. Front-line managers as agents in the HRM-performance causal chain: theory, analysis and evidence. *Human Resource management journal*, 17(1), pp.3–20.

Reason, J., 1990. Human Error. Cambridge University Press, Cambridge.

Rocha Fernandes, B.H., Mills, J.F. and Tereza L. Fleury, M., 2005. Resources that drive performance: an empirical investigation. *International Journal of Productivity and Performance Management,* 54(5/6), pp.340–354.

Sarshar, P., Granmo, O.C., Radianti, J. and Gonzalez, J.J., 2013, April. A Bayesian network model for evacuation time analysis during a ship fire. In Computational Intelligence in Dynamic and Uncertain Environments (CIDUE), 2013 IEEE Symposium on (pp. 100–107). IEEE.

Thomas, R. and Dunkerley, D., 1999. Careering downwards? Middle managers' experiences in the downsized organization. *British Journal of Management,* 10(2), pp.157–169.

Tolo, S.; Patelli, E. & Beer, M. Bayesian Network Approach for Risk Assessment of a Spent Nuclear Fuel Pond Vulnerability, Uncertainty, and Risk, American Society of Civil Engineers, 2014, 598–607

Wooldridge, B., Schmid, T. and Floyd, S.W., 2008. The middle management perspective on strategy process: Contributions, synthesis, and future research. *Journal of management*, 34(6), pp.1190–1221.

Zwirglmaier, K., Straub, D. and Groth, K.M., 2015. Framework for a Bayesian Network Version of IDHEAS. Safety and Reliability of Complex Engineered Systems: ESREL 2015, Taylor & Francis Group, London, pp. 3165–3172.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Risk assessment in military transport—human factor in estimation of risk

J. Ryczyński & M. Nowakowska
Tadeusz Kosciuszko Military University of Land Forces, Wroclaw, Poland

ABSTRACT: The article discusses the problems of conducting transport operations in the army. The original contribution of the authors in this field of research lies in the fact, that special attention has been paid to the Human Factor to estimate the risk of transport. Analysis of the impact that Human Factor, has on the risk of the transport depending on cultural aspects. The most important anthropological factors were identified in carrying out a successful military campaign in a culturally different country (Somalia, Iraq, Afghanistan). The authors present the results of pilot studies on the impact of the human factor on risk assessment in the implementation of transport operations in the army. The elements presented in the article are innovative, because the cultural and cultural factors were taken into account, conducting the analysis in the army.

1 INTRODUCTION

The design and implementation of military transport is a multilevel logistic process, aimed at preparing the cargo transport for safe transport along some particular segment of route. Regardless of of the mode of transport used to transport some cargo, opportunities and constraints exist for different modes of transport, so it is necessary to make a detailed analysis of the factors that need to be taken into consideration when planning for safe passage of cargo.

Analyzing the needs of the army, in the field of movement and transportation, specific features of the transport process should be taken into account. The primary feature is its intangible nature, which means that it is not possible to assess the transport process before implementation (Ryczyński and Smal, 2017). Specific features include a simultaneous realization and consumption of the transport service and the variability of quality standards, in particular for the movement of military forces including the mutual interaction between the provider and the recipient. No possibility of transport services congregation forces the user to anticipate the needs of transport in advance for military purposes.

Movement and transportation of troops and weapons systems is an issue arising mainly from the expeditionary nature of the modern armed forces and is associated with the need to movement forces in order to plan for their operational use.

The aim of the displacement of troops is to change the location of personnel, weapons and military equipment and means of supply caused by operational needs, logistics and training (Ryczyński and Smal, 2017).

Part of the military's movement and transportation involves the operation of one of the following subsystems: management, transport and transport networks. Movement and transportation is characterized by unpredictable nature referring to operational needs, the need to preserve the special conditions of caution and the need to prepare the means of transport to carry heavy and oversized and dangerous loads. Military transport is divided into due to its nature and purpose:

Operating, which includes transporting troops and their equipment.

Supply, the carriage of weapons, military equipment, munitions and materials.

Evacuation of unnecessary supply, failures and damaged of weapons as well as military equipment and packaging.

The proper choice of transport means it is one of the most important strategic decisions. This decision, depends on primarily on the operational criteria, the number of personnel provided for transportation, equipment and inventory quantities.

We have to also take into account the geographical location of the destination, distance, redeployment time, the possibility of using point and linear elements of the communication infrastructure and cost.

When planning a military movement and transportation we should be primarily aimed to achieve two functional objectives: ensure the timely

completion of tasks, and minimize costs at an acceptable level of effort.

2 RISK ASSESSMENT IN THE TRANSPORT—LITERATURE REVIEW

Currently, there is no single, unified and commonly used definition of the term of risk (Aven, 2012). We can even say that the basic concepts of risk are difficult to define and even more difficult to estimate (Heckmann et al, 2015).

In recent decades, we have observed that this term is used in many research areas, such as decision theory, management, emergency planning or critical-structure operations, including the efficiency of transport systems (Tubis and Werbińska-Wojciechowska, 2017). Historical trends in the development of the risk concept have been discussed, eg in (Aven, 2012 and Anderson. 1989). An overview of risk perspectives and discussions are provided, e.g. in (Aven, 2012, Aven et al., 2009, Aven and Krohn, 2014, Anderson, 1989). Aven distinguish between two types of approaches to risk analysis: relative frequency based approaches and Bayesian approaches. The first category includes both traditional methods of statistical inference as well as the so-called Frequency probability approach. Depending on the risk analysis method, the purpose of the analysis is different, and the results are presented in different ways.

This type of approach can be characterized as follows (Aven, 2012):

1. Relative frequency-based methods: The risk analysis consists in estimating certain basic probabilities of interpretations or frequencies. Such probabilities express the relative part of the time the event of interest, if the analyzed situation was hypothetically "repeated" an infinite number of times. The basic probabilities are unknown and are estimated in the risk analysis. These estimates are uncertain, because there may be large differences between estimates and valid (real) probabilities.
2. Bayesian methods: Probability is a measure of uncertainty about future events and results (consequences), seen through the eyes of an assessor and based on certain information and knowledge. Probability is a subjective measure of uncertainty.

These concepts are reflected in the literature dealing with the problem of risk estimation in transport processes. To effectively manage any organization (including companies providing transport services), a new concept—Enterprise Risk Management (ERM)—was introduced and promoted. One of the most popular definitions of the concept of corporate risk management (ERM) in literature is the one presented in the COSO II standard. In accordance with the COSO II standard, Enterprise Risk Management is defined as a process carried out by the management staff in order to identify potential events that may affect the reduction of the ability to achieve the entity's goals. According to COSO II, the organization's ERM system should be aimed at achieving the following four objectives:

1. Strategy: high-level goals, aligned with and supporting the organization's mission.
2. Operations: effective and efficient use of the organization's resources.
3. Reporting: reliability of the organization reporting system.
4. Compliance: organizational compliance with applicable laws and regulations.

Proper implementation of the ERM concept also affects the risk assessment processes performed in the selected company. Risk assessment is an indispensable and systematic process that is part of risk management, the purpose of which is to identify, assess risks and plan actions to address risks [34].

Referring to the previously presented definitions and taking into account the perspective of the process, the starting point for the risk assessment should be the identification of hazards that may cause the failure to achieve the goals of providing transport services. The holistic approach assumes that the area of analysis will cover various levels of the implemented process, eg technical elements, Human Factor as well as legal and organizational issues (Tubis and Werbińska-Wojciechowska, 2017). Risk assessment should be preceded by a process analysis that identifies potential adverse events. Risk assessment is usually only carried out at specific time points. Correct identification of the largest number of disturbing factors is the basis for a correct risk assessment.

3 RISK ASSESSMENT IN TRANSPORT DURING MILITARY OPERATIONS

The proposed methodology risk assessment in the transport of oversized loads consists of several stages (Fig. 1). The first step is, to diagnose the potential risk that accompanies transport, e.g. oversized loads. The next step is data collection and risk assessment. The diagnosis is necessary to determine the causes, factors that may cause difficulties in transport (Ryczyński and Smal, 2017). Evaluation of the results helps to choose a strategy for reducing the number of factors affecting the various types of risk and limit its negative impact. One of the most important element risk assess-

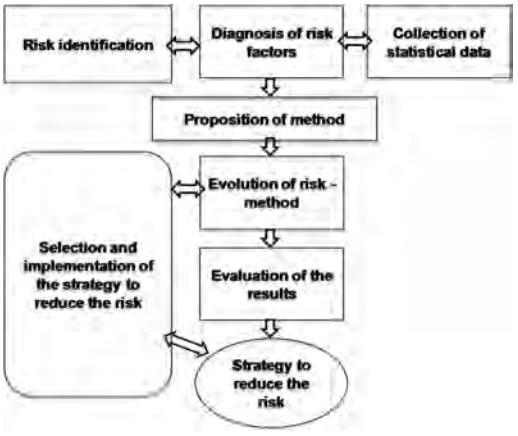


Figure 1. Methodology of risk assessment in military transport.

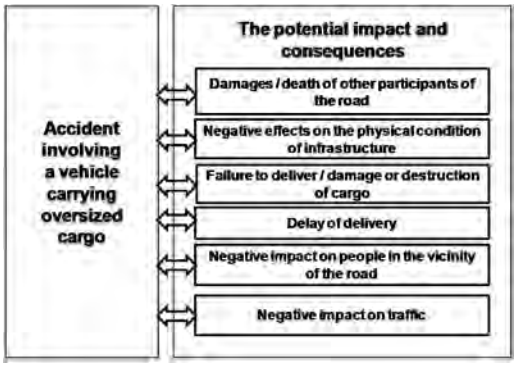


Figure 2. Dependence: accidents – potential consequences.

ment for the transport of oversized loads is the calculation of the probability of an accident.

Transportation of heavy oversized loads may have a negative impact on the infrastructure and the organization of traffic and generate, e.g.: the delay, damage to roads, damaged buildings located on the route of travel and the danger to other road users. To minimize the risk of an accident involving other vehicles to transport non-normative, it may be introduced ban on movement of other traffic on a particular stretch of road or along the entire length of the route. The diagram below shows the possible impact and the consequences of the traffic accident involving a vehicle carrying oversized loads (Fig. 2). The occurrence of an accident on the road can also affect vehicles normative.

Often the cause of the accident is incorrect actions of other road users, not adapted to the conditions prevailing on the road, the wrong choice of means of transport or incorrect load securing.

When modeling the risk for the oversized loads, the following data are used:

1. On the motorway network, including their physical characteristics.
2. On accidents involving heavy vehicles, including the frequency and severity of accidents.
3. On the flow of traffic.
4. On the loads, goods together with their characteristics.

These spatial data can be presented in GIS format. Information on highways, vehicles, accidents, traffic flow are readily available from various transport agencies. However, data on the transported loads, goods make bigger problem. Most of the available data in this area is insufficient to achieve the desired level of accuracy of risk assessment.

In practice, there are methods spreadsheet, which you can use to assess the degree of risk undertaken in the project. The classic model of risk calculation is as follows:

$$R = p \cdot C \tag{1}$$

where:

R – risk,

p – the probability of an accident,

C – the consequences of the accident.

The probability of an accident is dependent on many factors: the volume of traffic, time of transport operation (day or night), load of the vehicle, transport distance, the parameters of roads, road surface conditions, time of year, weather conditions, visibility, etc. The risk assessment process (Fig. 3) may contain more factors and deriving your situation, the consequences of the accident—but this is dependent on the availability of data on accidents.

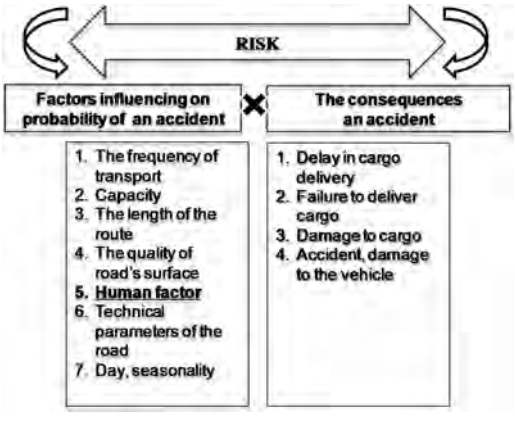


Figure 3. Risk assessment in transport: factors × consequences.

4 HUMAN FACTOR IN RISK ASSESSMENT

As shown in Figure 3, one of the most important element, is the Human Factor which affect the magnitude of risk in military transport. Based on expert research, the authors have determined that the Human Factor is more important during the realization of transport plans than, for example, the technical parameters of the roads during transportation.

A particularly important issue appears in matters of security and risk assessment related to the conduct of military activities. This is confirmed by the words of David Kilcullen (Killucen, 2010), an Australian anthropologist who has been developing new strategies for warfare for many years, including the use of cultural knowledge. Kilcullen claims that "knowing and understanding a foreign culture is a prerequisite for victory over an opponent".

What exactly is the Human Factor and how is it defined? We can define Human Factor as understanding human performance within a given system: trust, fear, decision-making, stress are crucial in so-called "golden hour" (Helsloot et al., 2004).

In industry, the Human Factor (also known as ergonomics) mean the study of how humans behave physically and psychologically in relation to particular environments, products, or services. Many large manufacturing companies have a Human Factors department or hire a consulting firm to study how any major new product would be accepted by the users in its design. A Human Factors specialist typically has an advanced academic degree in psychology or anthropology or has special training. The term usability is now sometimes used as an alternative to Human Factors like human error or human resource.

Today there are 2 different views on Human Factor as a cause of failure. The so-called "old view" means:

– human error is the cause of most accidents;
– the engineered systems in which people work are made to be basically safe;
– progress on safety can be made by protecting these systems from the unreliable human through selection, procedures, automation, training and discipline.

The other "new view" sees human error not as a case but as a symptom of failure:

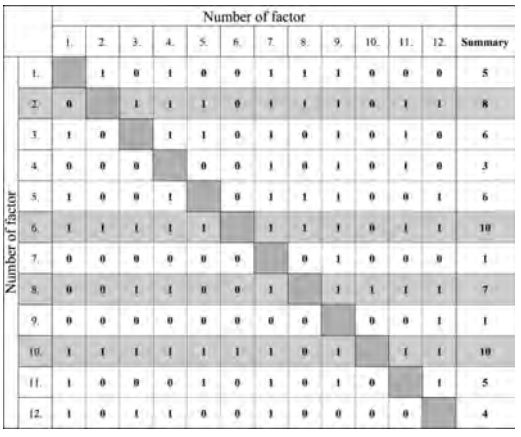
– human error is a symptom of a trouble deeper inside the system;
– safety is not inherent in the systems—people have to create safety;
– human error is systematically connected to features of people tools, tasks and operating environment.

Progress on safety comes from understanding and influencing these connections (Gambetti et al, 2012).

The Human Factor is very difficult to measure. Other component factors will be dominant in the determination of the Human Factor in the army and others in the case of testing according to the same methods in the civilian industry. Based on their own experience and using the knowledge of experts who deal with the problem of transport organization on a daily basis, the authors of the article have listed the 12 most important factors affecting the so-called Human Factor:

1. lack of communication – errors and disruptions in the information flow;
2. routine – certainty resulting from long-term practice combined with the loss of awareness of existing threats, caused by often repetitive activities and tedious work;
3. lack of knowledge – lack of clarity or certainty of understanding something, lack of language skills;
4. distraction – caused by distraction, confusion, mental chaos;
5. lack of cooperation in the team – inconsistent effort of a group of people caused by lack of a sense of community of purpose, fear of pointing management to mistakes made by others, inappropriate style of leadership or inappropriate communication;
6. fatigue – it is ignored, because until it is excessive, people do not realize it;
7. lack of resources – lack of tools, materials, outdated documentation, improper working conditions;
8. pressure – caused by the pressure of superiors or colleagues, lack of time, improper setting of tasks;
9. lack of assertiveness – lack of ability to refuse to perform a task resulting from lack of self-confidence, anxiety or complexes;
10. stress – nervousness caused by eg: time pressure, new methodology, change in the scope of tasks, competition or private factors;
11. carelessness – incorrect assessment of possible consequences of action caused by eg: pressure, lack of experience or lack of knowledge;
12. comfort (deviation) – acceptance by most people of deviations from the instructions as standards facilitating work.

The aim of the research was to indicate those factors that are dominant in the case of success or failure to complete the task. The described 12 factors were presented to a group of 35 military drivers (participants of foreign missions) and were asked to choose the most important factors. The comparison took place in such a way that participants



Number of factor	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	Summary
1.		1	0	1	0	0	1	1	1	0	0	0	5
2.	0		1	1	1	0	1	1	1	0	1	1	8
3.	1	0		1	1	0	1	0	1	0	1	0	6
4.	0	0	0		0	0	1	0	1	0	1	0	3
5.	1	0	0	1		0	1	1	1	0	0	1	6
6.	1	1	1	1	1		1	1	1	0	1	1	10
7.	0	0	0	0	0	0		0	1	0	0	0	1
8.	0	0	1	1	0	0	1		1	1	1	1	7
9.	0	0	0	0	0	0	0	0		0	0	1	1
10.	1	1	1	1	1	1	1	0	1		1	1	10
11.	1	0	0	0	1	0	1	0	1	0		1	5
12.	1	0	1	1	0	0	1	0	0	0	0		4

Figure 4. Result of research influence of chosen factors for success of complete the task.

compared all 12 factors with each other. The more important factor was rating 1 – a factor less important in a given pair – 0. The maximum value that a single indicator could have obtained was 11 points. Obtained results of tests are presented in Figure 4.

The studies carried out have shown that fatigue (10 points), stress (10 points), routine (8 points) and pressure (7 points) are of the greatest importance for success or failure in the implementation of the task. The least important was the lack of resources and the lack of assertiveness. It has been assumed that the lack of assertiveness will be the least important factor, because the army is a hierarchy institution and the soldiers carry out orders.

It should be noted that the presented results were made on a small research sample—the results of pilot studies. Currently, research is conducted on groups of 200 people, both civilian and military drivers. The completion of the research will allow for a more complete identification of the most important elements of the Human Factor in the process of estimating transport risk.

5 CULTURE ASPECTS OF HUMAN FACTOR

The research conducted by the authors allow to conclude that all available definitions completely omit the influence of cultural aspects on the success of the performance or non-performance of the task. The authors of the article, as participants of military foreign operations, observed this phenomenon during the mission in Iraq and Afghanistan.

The biggest problem in the course of military expeditionary operations in foreign countries is the mentality problems of the local population. The standard of behavior is a deliberate action consisting in violating applicable procedures, instructions, requirements or regulations. There is social acceptance for this type of behavior and they are not treated as an offense.

As one of the most representative example of the cultural aspect on the Human Factor, the Restore Hope operation in Somalia, can be mentioned. It was planned as a military and humanitarian mission under the auspices of the United Nations and was meant to provide transport services—providing food to the starving population of Somalia. One of the biggest problems of the mission in Somalia was a different understanding of the culture of this country by military components of individual countries, and thus different ways to accomplish mission objectives. Lack of understanding of the culture and history of Somalia was also characteristic of the United Nations, which worked on the assumptions of the mission. In particular, the clan system and the decentral nature of traditional political institutions in Somalia were not understood, which was manifested in the pressure on the UN that a particular clan would rule over the entire country. Such behavior, degrading the importance of the leaders of other clans, caused their reluctance to peace forces.

It should be emphasized that cultural problems were not only related to the relations between UNOSOM and Somali troops. Cultural differences influenced the effectiveness of UN force management and control. For example, soldiers of the Italian and French contingents ostentatiously ignored the orders of the Pakistani commander by following only the instructions coming from their own countries.

Based on the experience of foreign missions, the world's largest armies attach great importance to the issue of cultural differences.

Today, the American standard is *Operational Culture for the Warfighter and Applications for Operational Culture. Perspectives from the Field, as well as Through the Lens of Cultural Awareness: a Primer for US Armed Forces Deploying is Arab and Middle East Countries*. In January 2013 was published doctrinal document of the British forces JDN3/11 Decision Making and Problem Solving: Human and Organizational Factors. The document is devoted to the impact of a culturally conditioned Human Factor on the functioning of organizations such as the military, as well as on the ways of thinking and acting during various crises and military operations.

According to the mentioned documents, the culture can be "purchased" and not "biologically inherited" in the process of enculturation. It distinguishes members of one group from another and

is a lens through which members of a given group perceive and understand the world. Documents emphasize the importance of culture and man for the army "culture is part of a wider context in which military operations and everyday relations take place. Ignoring the importance of culture increases the risk of failure of a given company or mission and creates barriers to effective interaction. Developed cultural ability reduces these risks and creates: new opportunities that can contribute to the enemy's failure or success in negotiating with him; a better understanding of local communities and the foundations of a given conflict; fruitful cooperation with allies. (...) Understanding the value of culture in the whole spectrum of operations and military processes will result in increased situational awareness, better preparation of combat forces, better protection of forces and more effective tactical preparation. Culture is particularly important for security and stability where the human dimension is crucial".

In research on the Human Factor, the concept of intercultural awareness often appears—which is supposed to facilitate the functioning of culturally alien areas. An example is, the *Human Terrain System* (*HTS*), a new system of gaining knowledge about people and culture—operations in Iraq and Afghanistan revealed that the US army has no doctrine corresponding to the reality of operations. HTS operates on the basis of Human Terrain Teams, consisting of military, anthropologists, sociologists and linguists. Each team is recruited and trained to work in a specific region, where a military unit works, which the team is supposed to support. The system was criticized at the beginning of the 21st century as unethical from the point of view of anthropological research. However, the words of General David Petraeus are still quoted, saying that "knowledge of the culture of a given area can be just as important, and sometimes even more important, as knowledge of the geography of a given area. This observation is to draw attention to the fact that people's knowledge is, in many respects, a decisive factor, and that we must study the culture of a given area with the same attention we have so far devoted to the geography of the area" (Petraeus, 2006).

6 SUMMARY

When carrying out military transports, in addition to the behavior tests of drivers, it is necessary to create procedures according to which the organization of the entire transport system is to be organized. This system should be adapted to a specific cultural region. It must refer to all aspects of transport in which a human play role: communication, updated technical documentation or organization of teamwork. Components that need to be tested are:

1. defined main participants involved in transport, including the main coordinator responsible for the organization;
2. systematic characterization of the transport system of a given region and the socio-economic situation of the state from the residents' perspective;
3. indication of activities undertaken by residents and entities involved in the procedure;
4. examining the limitations, challenges and human potential in the region's transport system.

The conducted research shows communication problems at the local level, as well as problems with the lack of appropriate tools, lack of knowledge of customs, including stressful and motivating factors. Cultural research is proving necessary to understand the cultural environment of the planned transport operation. They also indicate the necessity of synergic cooperation between the army and the inhabitants, the government administration, and, in the case of missions, also non-governmental organizations. The aim of such action is to minimize and eliminate negative effects

REFERENCES

Anderson, E.L. 1989. Scientific trends in risk assessment re-search. *Toxicology and Industrial Health* 5(5), 777–790.

Aven, T. 2012. The risk concept – historical and recent development trends. *Reliability Engineering and System Safety* 99, 119–132.

Aven, T., Heide, B. 2009. Reliability and validity of risk analysis. *Reliability Engineering and System Safety* 94, 1862–1868.

Aven, T. 2013. Practical implications of the new risk perspectives. *Reliability Engineering and System Safety* 115, 136–145.

Aven, T. 2009. Perspectives on risk in a decision-making context – review and discussion. *Safety Science* 47, 798–806.

Aven, T., Kristensen, V.: Perspectives on risk: review and discussion of the basis for establishing a unified and holistic approach. *Reliability Engineering and System Safety* 90, 1–14.

Aven, T., Krohn, B.S. 2014. A new perspective on how to understand, assess and manage risk and the unforeseen. *Reliability Engineering and System Safety* 121, 1–10.

Gambetti, F., Casalli, A., Chisari, V. 2012. The Human Factor in Process Safety Management, *Chemical Engineering Transactions*, 26, p.279–280.

Heckmann, I., Comes, T., Nickel, S. 2015. A critical review on supply chain risk±definition, measure and modeling. *Omega* 52, 119–132.

Helsloot, I., Ruitenberg, A. 2004. Citizen Response to the Disasters; a Survey of Literature and Some Practical Implications, *Journal of Contingencies and Crisis Management*, 12(3), p. 98.
Kilcullen, D. 2010. *Counterinsurgency*, London, p. 4.
Petraeus, D. 2006. Observations from Soldiering in Iraq, *Military Review*. January–February 2006.
PN-ISO 31000:2010. Risk management – principles and guide-lines. Warsaw: PKN 2010.
Ryczyński, J. and Smal. T. 2017. Proposition of a Model for Risk Assessment in the Transport of the Oversized Loads in the Army. *2017 International Conference on Military Technology (ICMT)* [USB Proceedings]/ (eds.) Vaclav Krivanek, p-ISBN: 978-1-5386-1988-9, pp. 166–170.
Salmoni, B.A., Holmes-Eber, P. 2011. Operational Culture for the Warfighter and Applications for Operational Culture. Perspectives from the Field. Marine Corps University Press, Quantico, Virginia.
Tubis, A. and Werbińska-Wojciechowska, S. 2017. Operational risk assessment in road passenger transport companies performing at Polish market. Article prepared for conference European Safety and Reliability ESREL 2017, June 18–22, 2017, Portoroz, Slovenia
Tubis, A. and Werbińska-Wojciechowska S. 2017. The scope of the collected data for a holistic risk assessment performance in the road freight transport companies. Springer, cop. 2018. s. 450–463. Advances in Intelligent Systems and Computing, ISSN 2194-5357; vol. 582.
Wunderle, W.D. 2006. Through the Lens of Cultural Awareness: A Primer for US Armed Forces Deploying to Arab and Middle Eastern Countries. Military Bookshop.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Study on seafarers' emotion identification during watch-keeping using bridge simulation

S. Fan, J. Zhang & X. Yan
Intelligent Transport Systems Research Centre (ITSC), Wuhan University of Technology, Wuhan, China
National Engineering Research Centre for Water Transport Safety (WTSC), Wuhan, China

E. Blanco-Davis, Z. Yang & J. Wang
Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool, UK

ABSTRACT: Human factors present one of the essential contributors to maritime accidents, and seafarers' emotion is sensitive to working environment and information inaccessibility. The data collected from 11 experienced seafarers' Self-Assessment Manikin (SAM) questionnaires, is analysed to investigate the impact of their emotions during watch-keeping in a bridge simulator. SAM scale rating questionnaires are received separately after two sections, emotion calibration and recognition. The emotion is induced and identified in the calibration section. In the recognition, emotion is self-rated after the crew-qualified test and the Support Vector Machine (SVM) model is used for classification. The results indicate that SVM can effectively identify the emotions with a precision of 72.73%. Seafarers' emotion in maritime operations affects their behaviour and decision-making. The overall positive emotion identified by SAM rating does not mean positive effect on sailing, while negative emotion identified by SAM rating does not lead to negative behaviour.

1 INTRODUCTION

Nowadays, more than three-quarters of the world trade cargoes in terms of volume are accomplished by shipping (Grech et al., 2008). Shipping contributes to the increase of awareness on transport safety in maritime. Surveys (Ren et al., 2008) show that 75–96% of maritime accidents are fully or partially caused by human and organisational errors. There has been an overwhelming understanding since 1993 when the USCG reported that human errors had been the essential cause of approximately 80% of maritime accidents and near misses over the past decades (Grech et al., 2008). Moreover, out of nearly 62% of pollution and maritime accidents over the past years (Er and Celik, 2005), human factors result in 30% of deck officer error, 7% of shore-based personnel error, 2% of engine officer error, 8% of pilot error. It reveals that human factors are not the direct cause of the accident, but they are the beginning of a considerable incident or catastrophe by trigging following chain events.

In this regard, it is meaningful to investigate human factors in ship bridge as it is closer to the root causes of maritime accidents. One of the earliest initiatives was fired up by accidents caused by a typical radar-assisted collision (Grech et al., 2008). In 1956, the collision between the two passenger ships Andrea Doria and the Stockholm was one example. The root causes of the accident were related to the ship bridge. It is demonstrated that much attention should be paid to human factors and the bridge. Consequently, it causes some interest in the area of bridge design and cognition, both in Europe and the United States. Nowadays, the bridge has become more automated. Automation is often highlighted because it has been overwhelmingly understood that it would reduce the involvement of crew, so far as to reduce human workload and human errors, and increase efficiency. However, as demonstrated by the grounding of Royal Majesty (the Panamanian passenger ship grounded on Rose and Crown Shoal about 10 miles east of Nantucket Island, Massachusetts on June 10, 1995) as well as evidenced by other research results (Lutzhoft and Dekker, 2002), automation has a prospecting expectation of human work and safety, which cannot simply replace human work thoroughly. Fewer crew numbers do not lead to less workload. There exists increased mental workload affecting situation awareness (Aguiar et al., 2015). In this regard, automation in the bridge creates new error pathways, especially resulting from human errors, deficiencies in mission shifts, and postpone

chances to correct errors in the system further into the future. It is noteworthy that the bridge plays an essential role in the success and failure of navigation, as well as human errors research.

In the amendments of Seafarers' Training, Certification and Watchkeeping (STCW) Code in 1995, human error was classified in three major taxonomies. They are operational-based, management-based, and the combination of the two. For quantitative assessment of shipping accidents, Celik and Cebi (2009) generated a Human Factors Analysis and Classification System (HFACS) based on a Fuzzy Analytical Hierarchy Process (FAHP), to identify human errors in shipping accidents. In line with the HFACS, as well as Reason's Swiss Cheese Model and Hawkins' SHEL model, Chen et al. (2013) proposed HFACS for a Maritime Accidents (HFACS-MA) model to measure the Human and Organisational Factors (HOFs). Soner et al. (2015) combined Fuzzy Cognitive Mapping (FCM) and HFACS to generate a proactive model in fire prevention modelling onboard ships. Chauvin et al. (2013) found that most collisions were due to decision errors by a modified HFACS model in collisions reported by the Marine Accident and Investigation Branch (UK) and the Transportation Safety Board (Canada). Meanwhile, the accidents were also associated with poor visibility and evidence deficiencies of the socio-technical system (technical environment, the condition of operators, leadership level, and organisational level) (Chauvin et al., 2013). In that way, cognition and error in accidents attract research attention as well.

Quite a number of studies exist on human reliability to define human performance in accidents and estimate human failure probabilities (Yang et al., 2013, Yoshimura et al., 2015, Yang and Wang, 2012). Akyuz and Celik (2015) adopted Cognitive Reliability and Error Analysis Method (CREAM) to assess human reliability along with the cargo loading process, and Akhtar and Utne (2015) used it to study common patterns of interlinked fatigue factors. It was illustrated that "inattention", "inadequate procedures", "observation missed", and "communication failure" were related to fatigue factors that influenced the human cognitive processes in accidents. The bridge team should be trained to recognise fatigue and exercise caution related to the fatigue factors. Moreover, Hetherington et al. (2006) divided human factors into fatigue, stress, health, situation awareness, teamwork, decision-making, communication, automation, and safety cultural diversity.

Among them, the emotion factor of the crew is vulnerable to working space, inaccessible information sources, and communication. Their negative emotions are mainly related to irritability, tension, instability, depression, and burnout with periodic changes. From this perspective, Liu et al. (2016b) proposed an EEG (Electroencephalogram) system in bridge simulation to monitor officers' workload and pressure. It was one of the earliest studies on seafarer's psychological response using bridge simulators. However, the relation with psychological response and seafarers' performance was not demonstrated. For quantification of the crew emotion, this system also took into account monitoring emotion, emotional stress, and environmental stress (Liu et al., 2016a). It identified the emotion of cadets in the bridge simulator supervised by EEG, but not related to human errors neither. Geethanjali et al. (2017) detected and recognised the human emotion using Self-Assessment Manikin (SAM) rating. The statistical analysis revealed the emotion identification differences between several groups. Moreover, other studies on angry driving in road transportation(Yan et al., 2015, Wan et al., 2016, Yan et al., 2014, Zhang et al., 2014) have been conducted to find the emotional connection between humans and the safety. Hence, seafarers' emotion identification should be further studied by better incorporating psychological knowledge.

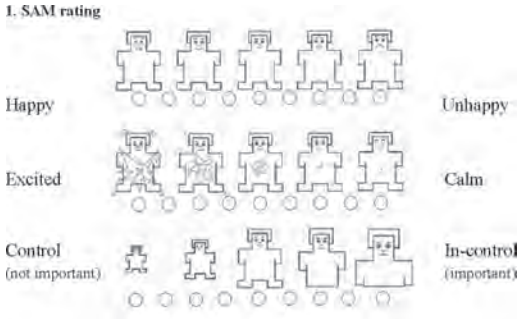
This study is conducted to identify the emotion in the bridge by SAM rating questionnaires, and classify the emotion in a Support Vector Machine model by use of bridge simulators. It would benefit the crew training aiming at navigational safety and improve the watch-keeping while sailing. The remainder of paper is organized as follows: In Section 2, the methodologies of this study are described, including SAM and SVM methods. The experiment design and procedures are illustrated in Section 3. The result and discussion are presented in Section 4. Finally, the conclusion is given in Section 5.

2 METHODOLOGY

2.1 *Self-Assessment Manikin (SAM)*

In this paper, the nine-point scale in Self-Assessment Manikin (SAM) (Bradley and Lang, 1994) (Bradley and Lang, 2007) is used to describe pleasure, arousal, and dominance in response to the stimuli. Figure 1 shows the questionnaire that the test subjects need to complete after the experiments, reflecting on their subjective feelings during the assessment.

The scoring measures the pleasure, arousal, and dominance associated with the stimuli. The first SAM scale is the happy/unhappy scale, which ranges from a smile to a frown. The second scale is the excited/calm scale, which ranges from left to right. The last dimension is the controlled/In-control dimension. The left end of the scale represents



2. Word rating

Joyful	Surprised	Satisfied	Protected
Angry	Fear	Unconcerned	Sad
Or give your own descriptive word:			

Figure 1. The questionnaire of emotion with SAM scale on a nine-point rating (Liu et al., 2016a).

the feeling of completely controlled and influenced. The right end of the scale is the feeling of completely controlling, important, and dominant.

The SAM methodology reveals the specific feature of a test subject's certain emotion, as the emotion is a subjective variable. This method quantifies the emotion in specific time and condition. The experiment is recorded by the audio for each test subject. After the qualified test, comments on the performance of seafarers from the experts is recorded by audio, and the test subjects are given a result of pass or not pass.

2.2 *Support Vector Machine (SVM)*

In this study, there are two sorts of emotion categories: positive emotion and negative emotion. The Support Vector Machine (SVM) is used to identify the emotion category for the tested seafarers. SVM is a supervised learning model with associated learning algorithms that analyse data used for classification and regression analysis. As shown in Figure 2, there are two kinds of sets, type "•" and type "▲". The SVM method finds an optimised hyperplane (e.g. a line defined by "w", "b"), calculating the w and b to maximise the margin while still separating the points (primal form). C is a cost function, which is $C = \max\left(0, 1 - y_i(\vec{w} \cdot \vec{x}_i - b)\right)$ (For data on the wrong side of the margin, C is proportional to the distance from the margin), $i=1,2,\ldots,l$, ξ_i is the smallest non-negative number satisfying $y_i(w \cdot x_i - b) \geq 1 - \xi_i$.

$$\min \frac{1}{l} \sum_{i=1}^{1} \xi_i + C\|w\|^2 \tag{1}$$

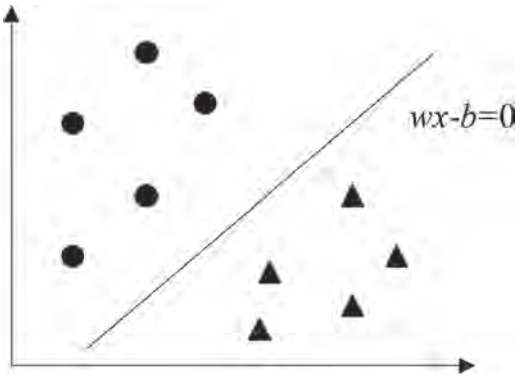


Figure 2. The support vector machine theory.

$$\text{s.t} \begin{cases} y_i(w \cdot x_i - b) \geq 1 - \xi_i \; i = 1,2,\cdots,l \\ \xi_i > 0 \end{cases} \tag{2}$$

In this study, these points are described in three dimensions illustrated in SAM as pleasure, arousal, and dominance. As the emotion is a subjective variable, the SVM uses the feature of a certain emotion in calibration to generate the classifier. Using the classifier training by SVM, emotion in the qualified test of seafarers is identified by the three-dimensional description questionnaire. Specifically, a $33 \cdot 3$ matrix is generated as input, among which "33" represents "$11 \cdot 3$" questionnaires, "3" defines three types of emotion dimensions. In addition, the output is a $33 \cdot 1$ matrix. The group of first 22 samples is the training set while the group of last 11 samples is the test set. After normalisation, the optimal parameters in the SVM is searched by cross-validation. The kernel function of the model is calculated. The result of identification of emotion taxonomy can be calculated.

3 EXPERIMENT

3.1 *Test subject selection*

Seafarers from different companies who were taking the captain and first officer qualification examinations were recruited to be involved in this study. There were 11 exams scheduled in two days. Each exam tested one participant who acted as a captain in a four-person exam group. All subjects were in good health without head injuries. They have 7.7 years of experience at sea on average, as they present a common emotional response during sailing when compared to beginners or cadets. The test subjects range from 26–38 years old, with

the average of 31.9 years old. They are all males. These seafarers attended these experiments as volunteers. The last also demonstrates that they could quit the experiments whenever they changed their minds. Based on this agreement, the calibration part of this study was conducted before the crew-qualified exam, and the test part was carried out after the whole exam. The test subjects were in bridge simulator room, while the staffs were in control room providing scenarios to subjects (Figure 3a, 3b).

3.2 *Stimuli selection*

The role of "captain" in four seafarers during the exam was selected as an independent sample. The rating of their perceived emotion for each stimulus presented uses a SAM scale. In view of this, IADS database was used as the stimulus with two categories (pleasant and unpleasant). It was presented for the first time, and all the test subjects in this study were not aware of the clips prior to the experiment.

(a) Test subjects in simulator room

(b) Staff in control room

Figure 3. The test subjects and staffs in control room.

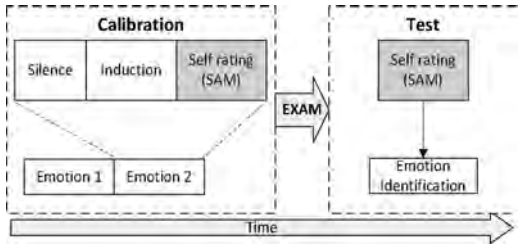


Figure 4. Experimental protocol.

3.3 *Experimental protocol*

The experiment was conducted by SAM scale rating questionnaires received separately after two sections, which are emotion calibration and recognition respectively. In calibration, two types of emotions were induced by the International Affective Digitized Sounds (IADS) methodology, developed by the National Institute of Mental Health Center for the Study of Human Emotion. Every test subject was given by listing to sound clips from IADS with eyes closed in case of blink interrupts. In the calibration part, emotion 1 began with 5 seconds silence to calm down, and 10 seconds for one category of emotion stimulus, and then the SAM rating was carried out. After that, another category of emotion 2 repeated. This section aimed to calibrate emotion for each subject. In other words, the specific feature or standard of personal emotion type was obtained.

In the test part, the subjects filled the questionnaires after at least 30 minutes' exam in the bridge simulator. Figure 4 demonstrates the process of the experiment.

3.4 *Statistical analysis*

As the subjective rating for the emotion questionnaire was not normally distributed, the nonparametric test is conducted. The sound clips of the IADS database are considered as independent variables; the pleasure, arousal, and dominance are considered as dependent variables. The correlation between valance and dominance within subjects is calculated using Spearman's correlation (Geethanjali et al., 2017).

4 RESULT AND DISCUSSION

4.1 *Descriptive statistics*

This study collects 22 (11×2) calibration questionnaires and 11 test questionnaires reflecting 11 seafarers' emotions. Table 1 demonstrates descriptive statistics for seafarers in this research, while Table 2

reveals the statistics in the IADS (2nd edition) database. There is correlation coefficient below in Table 3, and the correlation is significant at the 0.01 level. The clip sounds 105 represents negative emotion, while 220 represents positive emotion in this study. The letter "p", "a", and "d" represent "pleasure", "arousal", "dominance" respectively, and "t" means test emotion. The majority of the mean value of this study is approximately coincidental with the mean value of the IADS, except for the pleasure dimension in negative emotion. The distinction is also revealed in the questionnaires. Notes state that some of them do not make sense of the meaning of the first clip. Therefore, the neutral feeling with rating 5 is given by them.

4.2 *Emotion identification*

After collecting the emotion data from seafarers by SAM questionnaires, SVM is used to identify the emotion category during watch-keeping. Overall, 11 samples consist of 33 × 3 matrix of emotion description, and 33 × 1 matrix of emotion labels. The former 22 pieces are from the calibration part as a training set for SVM, and the later 11 pieces are from the test part as a test set. From these perspectives, the SVM model is proposed to find a hyperplane that divides the test set into two kinds of emotion categories. Figure 5 is the result of the test classification with the accuracy of 72.73%, where "1" represents negative emotion, and "2" means positive emotion. The kernel function of this model is calculated in the way that "$-t = 2$" represents kernel type is radial basis function: $\exp(-\gamma^*|x\text{-}x'|^2)$; "$-c$ = 776.0469" represents cost parameter C; "–g = 0.0068012" represents γ in kernel function.

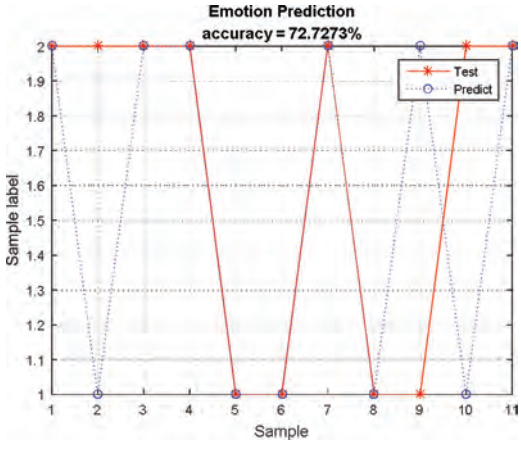


Figure 5. Emotion identification by using the SVM: Accuracy = 95.4545% (21/22) (training); Accuracy = 72.73% (8/11) (test).

Table 1. Statistics of seafarers in the research.

	Min.	Max.	Mean	SD
105p	1	9	4.82	2.601
105a	1	7	4.18	2.272
105d	1	8	5.18	2.523
220p	3	9	8.09	1.814
220a	1	8	5.27	2.195
220d	3	9	6.36	1.912
tp	3	9	5.73	1.679
ta	1	7	4.64	2.063
td	1	9	6.00	2.449

*p – pleasure, a – arousal, d – dominance.

Table 2. Statistics in the IADS (2nd Edition).

	Mean	Std. Deviation
105p	2.88	2.14
105a	6.40	2.13
105d	3.8	2.17
220p	7.28	1.91
220a	6.0	1.93
220d	5.99	1.88

Table 3. Correlation coefficient.

	Pleasure	Arousal	Dominance
Negative	1.000	–0.210	0.043
Positive	1.000	0.321	–0.068
Test	1.000	–0.480	–0.742

The emotion identifications by questionnaire from test subject and the SVM methodology are presented in Table 4, where "P" represents positive and "N" represents negative. More specifically, the self-rating emotions of subject 2 and 10 are positive emotions but were predicted as negative emotions. The self-rating emotion of subject 9 is negative while it was predicted as positive. All the others have the same results between self-rating and SVM.

4.3 *Human performance*

The comments on the examination for each test subject are further analysed to investigate if negative emotion identified by SAM scale questionnaire affects human errors and human performance. Meanwhile, the comments from experts as an inevitable process of the qualified exam are collected by audios. This part took place after the

Table 4. Comments from self-evaluation and third party.

ID	Emotion		Self-evaluation	Third party
	SR	SVM		
1	P	P	Untimely watch keeping in poor visibility Wrong operation sequence	Operate in incorrect sequence when stopping
			Too late to realise poor visibility	Unconcerned watch keeping
2	P	N	Speed control problem Inaccurate report in time	
			Anxious when collision	Not fulfil the Convention on the International
3	P	P	Wrong decision making (collision at ship body instead of bow)	Regulations for Preventing Collisions at Sea (COLREGs)
			Tension during ship encounter	Mistake sail with the current for sail against the current
4	P	P	Response too late Unfamiliar with navigation device	Not fulfil COLREGs Too panic when strand
			Speed control problem	Wrong decision making of the captain
5	N	N	Not enough communication Not stop timely	Manoeuvring inappropriate
6	N	N	Speed control problem Course deviation	Not enough communication Not enough cooperation not enough
			Late report in emergency	Wrong manoeuvring
7	P	P	Unconcerned Inappropriate manoeuvring	Too high speed Course deviation
8	N	N	Not familiar with rudder failure	Slow speed affect steering Failure to meet a contingency
9	N	P	Not switch on navigation lights when starting fog	Not on-time watch keeping Too large deflection angle
			Unfamiliar with navigation environment	Unfamiliar with navigation device
10	P	N	Not report the collision on time	Ignore environment when reporting Failed to fulfil COLREGs
11	P	P	Anxious when getting hurt	Speed control problem Irregular language

whole sessions, beginning with summarised comments from self-evaluation and third party, and ending in experts' comments.

According to the self-evaluation from the subjects and expert comments after the qualification exam, it is common to demonstrate that the human emotion emerging during watch-keeping affects ship manoeuvring, concentration, response to an emergency, and decision-making.

For example, test subject 1 was not able to concentrate on watch-keeping in poor visibility when sailing, which made him incapable of observing the crew onboard falling into the water. Moreover, a further step to eliminate the hazards they encountered was to stop in accurate and timely operation sequence. The test subjects 2 and 7 had the same result as unconcerned when encountering collision scenarios in poor visibility, resulting in a delayed report and operational problem. As a result, subject 2 reported inaccurately in collision scenario and subject 7 had course deviation. There was obvious anxiety when collision occurred as subject 3 demonstrated, causing not fulfilling COLREGs (International Regulations for Preventing Collisions at Sea). Subject 11 just became anxious when the crew got hurt, causing the irregular use of language and inappropriate manoeuvring. Subject 4 had tension emotion when the encounter happened and panic emotion during strand, which caused several mistakes, as shown in Table 4. In addition, subjects 4 and 10 had physiological problems because they were unfamiliar with the device. They were not fulfilling COLREGs.

From the above emotion problems existing in test subjects 1, 2, 3, 4, 7, 10, 11, all of them rated overall positive emotion after the sessions. However, the subjects who rated a negative emotion did not reveal apparent emotion interruption on performance. From this perspective, overall positive emotion identified by SAM rating does not mean positive effect on the sailing, while negative emotion identified by SAM rating does not lead to negative behaviour. Emotion rating through subjective judgement presents the overall feeling after the examination, whereas human errors occur at certain instant point. They are not matched well.

Moreover, the subjects may hide or ignore their true feelings when the mission is well done, and the questionnaires filled after the examination depend more on the result of the scenario completion than the process.

5 CONCLUSION

Seafarers' emotion exists when sailing. It emerges during watch-keeping and could jeopardise the performance and decision-making of seafarers. When an emergency happens, there are requests for the timely report and accurate operation on ships. This study utilised SAM rating scales of 11 test subjects to establish a classification model. The training model is studied by SVM classifier with an accuracy rate at 72.73%. The results concerning officers' emotion in a bridge simulator test reveal that seafarers' emotion in maritime operations affects their behaviour, as well as the possibility of errors in decision-making. In addition, there is no strong correlation between emotion modes identified and behavioural consequences. The overall positive emotion identified by SAM rating does not mean positive effect on sailing, while negative emotion identified by SAM rating does not lead to negative behaviour.

This study does not reveal the accurate in-time response, as the rating is done after the examination. Moreover, some seafarers may hide or ignore their true feelings in the questionnaire after the exam if emergency problems are solved properly in scenarios. Thus, the relations between emotion and human errors are complex, and need to be further analysed according to the real-time physiological responses.

Seafarers tend to be in vulnerable possession when manoeuvring in bridge simulator. Conducting the psychophysiology research in bridge simulator is a meaningful study on human factors or human errors in maritime operations. In addition, the bridge simulation benefits researches on human factors, especially for crew training requirement. From these perspectives, further studies will involve psychophysiological methods to measure seafarers' state and performance in real-time bridge simulation.

ACKNOWLEDGEMENTS

The research was supported by the National Science Foundation of China (NSFC) under grant No. 51609194, the National Key Technologies Research & Development Program (2017YFC080490, 2017 YFC0804904), the Fundamental Research Funds for the Central Universities (WUT: 2017IVA102, 2017IVB049, 2017IVB074), and the EU project RESET (H2020-MSCA-RISE-2016, 730888). The authors would like to thank the University of Florida for sharing the IADS database. The authors also appreciate the support of China MSA, the Bridge Simulator centre in Wuhan University of Technology, and all bridge officers participated in this experiment.

REFERENCES

Aguiar, Y.P.C., Vieira, M.D.Q., Galy-marie, E. & Santoni, C. 2015. Analysis of the User Behaviour When Interacting with Systems During Critical Situations. *In:* Mercantini, J.M. & Faucher, C. (eds.) *Risk and Cognition.* Berlin: Springer-Verlag Berlin.

Akhtar, M.J. & Utne, I.B. 2015. Common patterns in aggregated accident analysis charts from human fatigue-related groundings and collisions at sea. *Maritime Policy & Management,* 42, 186–206.

Akyuz, E. & Celik, M. 2015. Application of CREAM human reliability model to cargo loading process of LPG tankers. *Journal of Loss Prevention in the Process Industries,* 34, 39–48.

Bradley, M.M. & Lang, P.J. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry,* 25, 49–59.

Bradley, M.M. & Lang, P.J. 2007. The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. *Technical report B-3.* University of Florida, Gainesville, Fl.

Celik, M. & Cebi, S. 2009. Analytical HFACS for investigating human errors in shipping accidents. *Accident Analysis and Prevention,* 41, 66–75.

Chauvin, C., Lardjane, S., Morel, G., Clostermann, J.P. & Langard, B. 2013. Human and organisational factors in maritime accidents: Analysis of collisions at sea using the HFACS. *Accident Analysis and Prevention,* 59, 26–37.

Chen, S.T., Wall, A., Davies, P., Yang, Z.L., Wang, J. & Chou, Y.H. 2013. A Human and Organisational Factors (HOFs) analysis method for marine casualties using HFACS-Maritime Accidents (HFACS-MA). *Safety Science,* 60, 105–114.

Er, Z. & Celik, M. 2005. *Definitions of human factor analysis for the maritime safety management process.*

Geethanjali, B., Adalarasu, K., Hemapraba, A., Kumar, S.P. & Rajasekeran, R. 2017. Emotion analysis using SAM (Self-Assessment Manikin) scale. *Biomedical Research*.

Grech, M.R., Horberry, T. & Koester, T. 2008. Human factors in the maritime domain. CRC Press.

Hetherington, C., Flin, R. & Mearns, K. 2006. Safety in shipping: The human element. *Journal of Safety Research,* 37, 401–11.

Liu, Y., Hou, X., Sourina, O., Konovessis, D. & Krishnan, G. 2016a. *EEG-based human factors evaluation for maritime simulator-aided assessment: Proceedings of the 3rd International Conference on Maritime*

Technology and Engineering (MARTECH 2016, Lisbon, Portugal, 4–6 July 2016).

Liu, Y.S., Hou, X.Y., Sourina, O., Konovessis, D. & Krishnan, G. 2016b. *Human factor study for maritime simulator-based assessment of cadets,* New York, Amer Soc Mechanical Engineers.

Lutzhoft, M.H. & Dekker, S.W.A. 2002. On your watch: Automation on the bridge. *Journal of Navigation,* 55, 83–96.

Ren, J., Jenkinson, I., Wang, J., Xu, D.L. & Yang, J.B. 2008. A methodology to model causal relationships on offshore safety assessment focusing on human and organizational factors. *Journal of Safety Research,* 39, 87–100.

Soner, O., Asan, U. & Celik, M. 2015. Use of HFACS-FCM in fire prevention modelling on board ships. *Safety Science,* 77, 25–41.

Wan, P., Wu, C., Lin, Y. & Ma, X. 2016. A Recognition Model of Driving Anger Based on Physiological Features by ROC Curve Analysis.

Yan, L., Zheng, K., Wu, C., Wen, J., Zhu, D. & Shi, J. A Special Laboratory Method For Inducing Driving Anger: Based On The Virtual Scene. Transportation Research Board 94th Annual Meeting, 2015.

Yan, L., Zhu, D., Wu, C., Zhong, M. & Zheng, K. Ranking and Causal Relationship Analysis of Incentive Factors of Driving Anger: A Case Study from an On-Road Experiment in China. Cota International Conference of Transportation Professionals, 2014. 2534–2547.

Yang, Z.L., Bonsall, S., Wall, A., Wang, J. & Usman, M. 2013. A modified CREAM to human reliability quantification in marine engineering. *Ocean Engineering,* 58, 293–303.

Yang, Z.L. & Wang, J. 2012. *Quantitative retrospective analysis of CREAM in maritime operations.*

Yoshimura, K., Takemoto, T. & Mitomo, N. 2015. The Support for using the Cognitive Reliability and Error Analysis Method (CREAM) for Marine Accident Investigation. *2015 4th International Conference on Informatics, Electronics & Vision Iciev 15*.

Zhang, H., Yan, X., Wu, C. & Qiu, T.Z. 2014. Effect of Circadian Rhythms and Driving Duration on Fatigue Level and Driving Performance of Professional Drivers. *Transportation Research Record Journal of the Transportation Research Board,* 2402, 19–27.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Accounting for human failure in autonomous ship operations

M.A. Ramos, I.B. Utne & J.E. Vinnem
Department of Marine Technology, Norwegian University of Science and Technology, Trondheim, Norway

A. Mosleh
The John B. Garrick Institute for Risk Sciences, University of California in Los Angeles, Los Angeles, USA
Department of Marine Technology, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: Recently, numerous organizations have made progress in developing autonomous ships, motivated by, among other factors, the potential for increased safety. This applies especially to accidents involving human error, as autonomous ships would remove operator role from all or most operations. The reality is that although the human role is reduced, autonomous ships would still rely on operators for supervision, remote control, and involvement in case of a glitch or an unexpected situation. Thus, autonomous ships do not fully eliminate the possibility of human error. In this study, we assess the potential for human error in autonomous ship operations. We analyze an unmanned autonomous ship operation, and through a generic analysis of the interaction between operators working a Shore Control Centre (SCC) and system, we identify possible Human Failure Events (HFE). This provide a starting point for performing human reliability analysis of autonomous ships operation.

1 INTRODUCTION

Important projects have drawn attention to autonomous ships in recent years. The so-called MUNIN (Maritime Unmanned Ships through Intelligence in Network) is currently establishing a concept for an unmanned merchant ship. AAWA (Advanced Autonomous Waterborne Applications Initiative), in turn, investigates challenges in different scientific fields related to autonomous shipping operations (Laurinen, 2016). A third example is the DNV GL ReVolt project, which is a concept developed by DNV GL for an unmanned, zero-emission, shortsea vessel.

One of the motivations for using autonomous ships—common to all autonomous systems in general, concerns the potential for increased safety and reliability. Human error accounts for an important root cause or contributing factor of accidents in a diversity of industries and activities, e.g. 70 to 80% in aviation (Wiegmann and Shappell, 2012), over 80% in chemical and petrochemical industries (Kariuki and Löwe, 2007), over 90% in road traffic accidents (Treat et al., 1979). The maritime activity does not differ from those: the European Maritime Safety Agency points to human error as the triggering factor in 62% of incidents with EU registered ships from 2011 to 2016 (EMSA, 2014). Moreover, statistics on fatal accidents have ascertained that work on deck, for example mooring operations, is 5 to 16 times more dangerous than jobs ashore (Blanke, Henrique and Bang, 2017). Therefore, putting the human operator aside for all (or most part of) operation tasks is believed to avoid accidents (Laurinen, 2016). However, as highlighted by Rødseth and Tjora (2014), human error can still occur in autonomous ships operation.

Current projects on autonomous ships have different views on how they should operate in terms of crew onboard and autonomy level. The Munin project calls for a ship that would be unmanned only on the deep-sea part of the voyage, with a crew onboard for the departure from and approach to port. When unmanned, the ship would be autonomous, but monitored by operators in a Shore Control Center who can take over control of the ship in certain situations. The AAWA project, on the other hand, works with the concept of unmanned ships—i.e., there would be no crew onboard at any time of the voyage, and with dynamic levels of autonomy. This means the autonomy level approach would depend on the state of the vessel and mission being executed. In some cases, such as navigation in the open seas, the ship can be nearly fully autonomous whereas for some parts of the voyage it will require close supervision and decision making, or even full tele-operation from the human operator.

The two concepts above show that, although an autonomous ship would have less interference of a human operator, the human is still part of the operation: they would still rely on a human operator for one of the voyages phases (e.g. departure

and docking) or for taking over control in case there is a situation the autonomous system cannot resolve by itself. Therefore, the autonomous ship operations are not free of the possibility of accidents generated or aggravated by human error. The current literature, however, have not deeply focused on the human element when considering autonomous ships' safety. In fact, as pointed out by Parasuraman, Sheridan and Wickens (2000), for autonomous systems in general there is a voluminous technical literature on automation, but a still small (but growing) research base examining the human capabilities involved in work with automated systems.

The potential for human error in autonomous ships can be assessed through a Human Reliability Analysis (HRA). HRA is a technique to assess both quantitatively and qualitatively the human contribution to accidents. Swain and Guttman (1983) define human reliability as the probability that a person (1) correctly performs an action required by the system in a required time and (2) that this person does not perform any extraneous activity that can degrade the system. HRA is thus, in short, a method by which human reliability is estimated (Swain and Guttman, 1983; Swain, 1990).

To be able to perform an HRA it is essential to understand, first of all, how the operators will interact with the system. If autonomous ships will be a reality in years to come, ensuring they are safe and reliable is imperative, and the possibility of human errors cannot be minimized.

The current literature presents some relevant works on autonomous ships containing discussions on human factors topics, mostly pointing out the factors that could affect the operators' decision and actions (Ottesen, 2014; Rødseth and Tjora, 2014; Laurinen, 2016). A more general discussion on these factors, applied to all autonomous systems, can also be found (Parasuraman, Sheridan and Wickens, 2000; Chen, Haas and Barnes, 2007). In terms of identifying the possible human failure events (HFEs) in autonomous ships operations, however, the literature still falls short—and this paper aims to fill in this gap.

The identification and definition of HFEs is can be considered as the starting point of an HRA (Ekanem, 2013). Boring (2014) differentiates between two approaches for identifying HFEs. A top-down approach would start with the analysis of hardware faults and deducing human contributions to those faults, and is widely used in Probabilistic Risk Assessments in the nuclear industry. A bottom-up approach, on the other hand, would look at opportunities for human errors and then model them in terms of potential for affecting safety outcomes. This paper will adopt a bottom-up approach, performing a screening of the interactions between the operators and the system and the subsequent tasks in order to identify the HFEs. The present paper, hence, aims to analyze the interactions between the operators' and the system in the operation of autonomous ships, and identify the possible human failure events that derive from it.

The discussions of this paper are part of an ongoing research aiming to identify and model the risks arising from autonomous ships operation. The scope of this paper is limited to human actions and human failures during operation, under the assumption that the system would work as expected. Therefore, it does not cover system failures, which will be addressed in forthcoming papers by the authors.

Nonetheless, it is important to acknowledge that the possibility of human error is not restricted to the operation of the ships. Human error associated with autonomous ships can be related to design, construction and installation, testing and verification and maintenance, among other activities carried out by humans prior to the operation. This paper does not cover these tasks, as it focuses on the operation only, i.e., it considers that there would be no failures in all of these tasks previous to operation and navigation. Hence, the question it aims to answer is: given perfect design, maintenance, equipment and instruments behavior, could human actions affect safety during autonomous ships operation?

The paper is organized as follows: Section 2 presents the system description and the assumptions made in this study, Section 3 focuses on describing of the interactions between the operators' and the system and the possible Human Failure Events deriving from these interactions. Section 4 presents some concluding thoughts.

2 SYSTEM DESCRIPTION

As stated in Section 1, the ongoing projects on autonomous ships have different concepts in terms of the ship being manned or on the level of autonomy. Utne et al. (2017) use the following definition of autonomy (adjusted from National Institute of Standards and Technology (NIST) (2008)): "a system's or sub-system's own ability of integrated sensing, perceiving, analyzing, communicating, planning, decision-making, and acting, to achieve its goals as assigned by its human operator(s) through designed Human-Machine Interface (HMI)". From fully manual control to fully autonomous systems there can be distinct levels of autonomy (LoA), and the literature provides different proposals for these levels and its taxonomy. One of the oldest taxonomies is the one

proposed by Sheridan and Verplank (1978), with 10 LoA, where Level 1 corresponds to fully manual control and level 10 to fully autonomous control. A review of all proposals can be seen in Vagia et al. (2016).

From among the autonomous ship concepts indicated in Section 1, the analysis in this paper is based on the AAWA concept—unmanned ships and dynamic level of autonomy. A dynamic level of autonomy means that the autonomy level can change depending on the context of the voyage, e.g., one phase of the voyage is set to be fully autonomous (Level 10 in the Sheridan and Verplank taxonomy) but the operation encounters a small problem and give the operator a "veto" option before solving it autonomously (Level 6 of autonomy). The reason for choosing this concept are: (i) because it is unmanned, it offers the most different case study from ship operations nowadays, and (ii) because it has a dynamic autonomy level, it covers a different range of situations, from totally autonomous operation to tele-operated control.

Being an unmanned ship, the operators would be working onshore, and we assume they would be working in a Shore Control Center (SCC) as the one proposed in the Munin project. The Munin project website[1] offers a range of information and publications on the Shore Control Center. Essentially, the Shore Control Centre acts as a manned supervisory station for monitoring and remote controlling a fleet of autonomous ships. Most of the time the ships would operate autonomously, without the need for intervention from shore. When needed, though, the operators' in the SCC would provide assistance and may take over control of the ship (Porathe, 2013; Porathe, Prison and Man, 2014; MUNIN, 2016).

The voyage can be divided into four phases, in which the operators would have different possible levels of interaction with the system: Voyage Planning, Unmooring and maneuvering out of dock, Open Sea and Port approaching and docking. The following of these phases is based on the information stated in the AAWA whitepaper (Laurinen, 2016) for a general cargo vessel.

The first phase is the Voyage Planning, in which the operators assess/define certain conditions of the voyage. The operators' assessment makes use of systems that should be present in the ship, such as an automatic system for verifying the sea readiness before starting the voyage. Most of the systems can be checked remotely by the operator while in some areas (such as securing cargo) shore based crew can also be used to check that voyage can be started.

One of the conditions that have to be assessed by the operator previously to the voyage is the connectivity—some of the remote control or remote supervision modes might require a latency and bandwidth that exceeds the capability of the satellite systems in adverse weather conditions. The operator will have then to ensure that there is sufficient connectivity for the intended mission. If there is enough connectivity for the mission, the operator has then to define the primary operational strategy for each leg—autonomous or manual, considering the weather and environmental conditions. Note that manual operation, in this case, means remote operation from the SCC. Next, the operator defines the navigational and fallback strategies. Although the AAWA whitepaper does not describe what the operators take into account during "navigational strategies", we believe that in addition to predefined paths and waypoints it would also include considerations with maintenance (to verify when should the next maintenance of determined equipment be versus the length of the voyage), propulsion and fuel consumption. The fallback strategy, on the other hand, is a strategy executed if the ship experiences an unexpected situation that would require operator intervention. The fallback strategy could include: asking operator to take manual control, slow down and proceed to following waypoint, stop the vessel and stay in DP mode, navigate to previous waypoint, navigate back to preset safe location. The commands and their execution sequence is not same in all parts of the voyage. For example trying to maintain its position in the middle of a congested and narrow fairway in harsh weather might not be a feasible strategy.

It is important to bear in mind that, given dynamic autonomy, the definitions made in the voyage planning are not static, i.e., it can be that one leg was defined to be autonomous but due to external circumstances it goes manual. Moreover, the voyage plan as well as the fallback strategies can always be modified during the voyage using the satellite communication link.

The phase after voyage planning would be unmooring and maneuvering out of dock. The mooring systems can be fully or semi-automatic. A fully automatic mooring system would mean that the operation can be remote controlled or automatically executed by the autonomous vessel. A semi-automatic mooring, on the other hand, means that connection to the quay can be made automatically but the crew is needed to secure the docking. When the ship is maneuvered out of the congested harbor area, it can be controlled by the operator or it can use the dynamic positioning control computer and autonomous control system to reach the waypoint. Moreover, in some areas it could go directly to autonomous mode instead of starting with teleoperation or supervisory control.

[1] http://www.unmanned-ship.org/munin/

The third voyage phase, after maneuvering out of dock, is the open sea navigation. In autonomous mode the ship executes the voyage according to the defined plan, and the operator receives relevant status data such as ship's location, heading, speed, ETA to next waypoint (or area of closer supervision) and key information from the situational awareness systems as well as critical ship systems.. For situations where the autonomous navigation system's autonomous decision making threshold is exceeded, the operator is notified and can intervene. Therefore, the autonomy level is dynamically adjusted if the mission execution is not proceeding according to the original plan and the autonomous navigation system sees that adjustments are needed. AAWA differs between two different situations: one is a "veto" situation, in which for example the vessel is deviating from the planned course between the two waypoints but stays within specified margins the autonomous navigation system. In this case the system would notify the operator about planned evasion and give the operator a possibility to veto for a limited time. If modifications are needed, the operator can take the vessel in manual control. It can also be that the vessel would need to change the course in such a way that complete waypoint has to be replanned. In order to ensure that changes to the plan are made in a safe way operator confirmation will be requested. The autonomous navigation system will offer one or more alternatives of how the waypoint could be modified but the operator will finally make the decision how to continue the voyage.

A second case would be a "pan-pan" situation—when there is a complex scenario that the autonomous navigation system path planning and algorithms cannot unambiguously solve. Example of this could be if extremely large number of crafts or other objects are detected and the path planning algorithms are not capable to identify them and thereby the system cannot determine how the navigation should proceed. In this type of scenario the vessel will immediately send a "pan-pan" message to the operator indicating that it is in urgent need of assistance. The ship has a predefined set of fallback strategies (defined at the voyage planning phase) that it will start to execute in the planned order if user response is not received, and depending on the urgency, automatic fallback strategy execution can also be started immediately.

The last phase of the voyage is port approaching and docking. As the other phases, it can be remotely operated or autonomous. This phase together with open sea navigation and unmooring and maneuvering out of dock will be named "navigation phases".

The next section details the interactions between the operators and the system in each of these phases.

3 INTERACTIONS BETWEEN THE OPERATORS AND THE SYSTEM

This section discusses the interactions between operators and the system for each voyage phase of the autonomous ship described in the previous section. It explain the operators' main tasks and the possible decision/action paths they may take when accomplishing these tasks.

The outcomes of the operators' actions will be described in this paper as a "success" or a "failure" of that voyage phase. A successful operation is defined here as an operation that did not encounter any unexpected problem or an operation that did encounter a problem but successfully recovered from it, by operators' actions or autonomous solving. For instance, if during autonomous navigation in the open sea the ship faces a complex situation it cannot solve autonomously and it gives a "pan-pan" alert for the operators, they take over control in time and manage to bring the ship back to a safe status, this is a successful open sea voyage. An unsuccessful operation, on the other hand, is one that encounters a problem and does not recover from it. In the previous example, if the operators fail to respond to the "pan-pan" alert and the ship follows a fallback strategy that is inadequate, this would be a failure in the open sea voyage, leading to an incident.

Note that for voyage planning a failure will not itself cause an accident, but it will increase the probability of having a "veto" or "pan-pan" situations at the following phases. For example, if during voyage planning the operator decides for open sea voyage to be autonomous when the environmental conditions are not safe for the operation, there will be a higher chance that an unexpected situation during the voyage arises and the operator receives a "pan-pan" alert about it. Failures at the following phases, on the other hand, can cause accidents. These may, however, differ in terms of gravity: an accident when still in harbor is less probable to be of catastrophic consequences than during open sea voyage. Yet, the final events treated in this paper will be "success" and "failure", not distinguishing between the severities of this failure, such as collision, grounding, etc. This is illustrated in the general Event Sequence Diagram in Figure 1, where

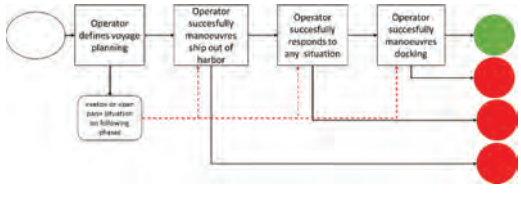


Figure 1. Interaction scheme for voyage planning.

the "success" outcome is represented by the green final event, which is reached if all voyage phases are successful, and the "failures" are represented by the red final events.

For the unmooring and maneuvering out of harbor phase it was considered a fully automatic system, i.e., the operation can be remote controlled or automatically executed by the autonomous vessel, depending on what was defined at the voyage planning phase. Moreover, in spite of the AAWA whitepaper describing the "veto" or "pan-pan" situations only during the open sea voyage, it was considered that they could also happen during unmooring and maneuvering out of harbor and port approaching and docking, when these are in autonomous mode. In this sense, the possible interactions between the operator and the system in the three voyage phases that follows voyage planning are similar: the operation can go manual or autonomous; in case it is autonomous it can i) operate as expected, ii) encounter a small problem and generate a "veto" situation, iii) encounter a more complex problem and generate a "pan-pan" situation. Moreover, the operators' possible responses to these situations are also similar in the three phases. Thus, what will be discussed below for unmooring and maneuvering out of harbor can be extended to the open sea navigation and port approaching and docking.

The operator's interactions with the system in the voyage planning phase are described in subsection 3.1, and for the following phases, exemplified by unmooring and maneuvering out of harbor operation, in sub-section 3.2.

3.1 *Voyage planning*

From the description in Section 2, it is possible to identify the operator's tasks and possible paths in the voyage planning, which are described in the tables below.

3.2 *Unmooring and maneuvering out of harbor*

The operator's tasks and possible paths in Unmooring and maneuvering out of harbor are described below, and can be extended for the open sea voyage and port approaching and docking phases.

I. Autonomous operation

When the operation is autonomous there can be a small problem that the vessel can solve autonomously, in which case the operator receives a "veto" alert (Table 5). If there is a significant problem, the vessel gives a "pan-pan" alert to the operator (Table 6). In that case, if the operator does not take over control the vessel follows the fallback strategy.

Table 1. Possible operator decisions for Task 1.

Task 1 (T1): Ensure there is sufficient connectivity			
I. If there is no sufficient connectivity, the operator can:		**II.** If there is sufficient connectivity, the operator can:	
T1_path 1: be wrong and believe there is sufficient connectivity, and the operation goes on	**T1_path 2:** be right about the connectivity level and cancel the voyage	**T1_path 3:** be wrong and believe the connectivity is not enough, and cancel the voyage	**T1_path 4:** be right and the operation goes on

Table 2. Possible operator decisions for Task 2.

Task 2 (T2): Define primary strategy for each leg (autonomous or manual)		
T2_path 1: Operator decides that the operation for one leg is autonomous when, due to weather and environmental conditions, it should be manual. For each phase:		
i. Unmooring and maneuvering out of harbor goes autonomous when it should be manual	**ii.** Operation in open sea goes autonomous when it should be manual	**iii**. Port approaching and docking goes autonomous when it should be manual
T2_path 2: Operator decides that one leg should be manual when it could be autonomous		
i. Unmooring and maneuvering out of harbor goes manual when it could be autonomous	**ii.** Operation in open sea goes manual when it could be autonomous	**iii.** Port approaching and docking goes manual when it could be autonomous
T2_path 3: Operator correctly decides that the operation for one leg is autonomous. For each phase:		
i. Unmooring and maneuvering out of harbor goes autonomous	**ii.** Operation in open sea goes autonomous	**iii**. Port approaching and docking goes autonomous
T2_path 4: Operator correctly decides that one leg should be manual:		
i. Unmooring and maneuvering out of harbor goes manual	**ii.** Operation in open sea goes manual	**iii.** Port approaching and docking goes manual

Table 3. Possible operator decisions for Task 3.

Task 3 (T3): Define navigational strategies for the autonomous operations	
T3_path 1: The operator defines an incorrect navigational strategy	**T3_path 2:** The operator decides for a good operational strategy

II. Manual unmooring

To aid in the visualization of these tasks and paths, these interactions are modeled through the schemes presented in Figure 2 for voyage planning

and Figure 3 for unmooring and maneuvering out of harbor.

As stated previously, these interactions were modeled not considering system failure yet—e.g. there is a "pan-pan" situation and the alert at the Shore Control Center fails, or the operator takes over manual control and the communication between the SCC and the vessel fails. It isolates, then, the human errors, considering no failure on other aspects of the operation.

From the interaction schemes above it is possible to identify the possible Human Failure Events that could lead or contribute to accidents in autonomous ships operation. Table 8 presents the HFEs involved in the voyage planning phase. Note that these failures would not cause an accident itself, but would contribute for having a "veto" or "pan-pan" situation in the following phases, as illustrated in Figure 1. Table 9 present the HFEs involved in

Table 4. Possible operator decisions for Task 4.

Task 4 (T4): Define fallback strategy for the autonomous operations	
T4_path 1: The operator defines an inadequate fallback strategy	**T4_path 2:** The operator defines an adequate fallback strategy

Table 5. Possible operator decisions for Task 5, after a "veto" alert is received during autonomous operation.

Task 5 (T5): Respond to "veto alert"	
T5_path 1: The operator does no respond to the veto alert	
T5_path 2: The operator responds to the alert and supervise the vessel solve the problem autonomously	
T5_path 2_1: The operator should have "veto" that operation and take over control because the autonomous solutions were not adequate	**T5_path 2_2:** The autonomous solution is adequate
T5_path 3: The operator responds to the alert and take over control of the vessel	
T5_path 3_1: The operator successfully operates the ships	**T5_path 3_2:** The operator fails when operating the ship

Table 6. Possible operator decisions for Task 6, after a "pan-pan" alert is received during autonomous operation.

Task 6 (T6): Respond to "pan-pan" alert	
T6_path 1: The operator does no respond to the "pan-pan" alert	
T6_path 2: The operator responds to the alert and supervise the vessel solve the problem autonomously following the fallback strategy	
T6_path 2_1: The operator should have taken over control of the ship because the fallback strategy is not adequate	**T6_path 2_2:** The fallback strategy is adequate
T6_path 3: The operator responds to the alert and take over control of the vessel	

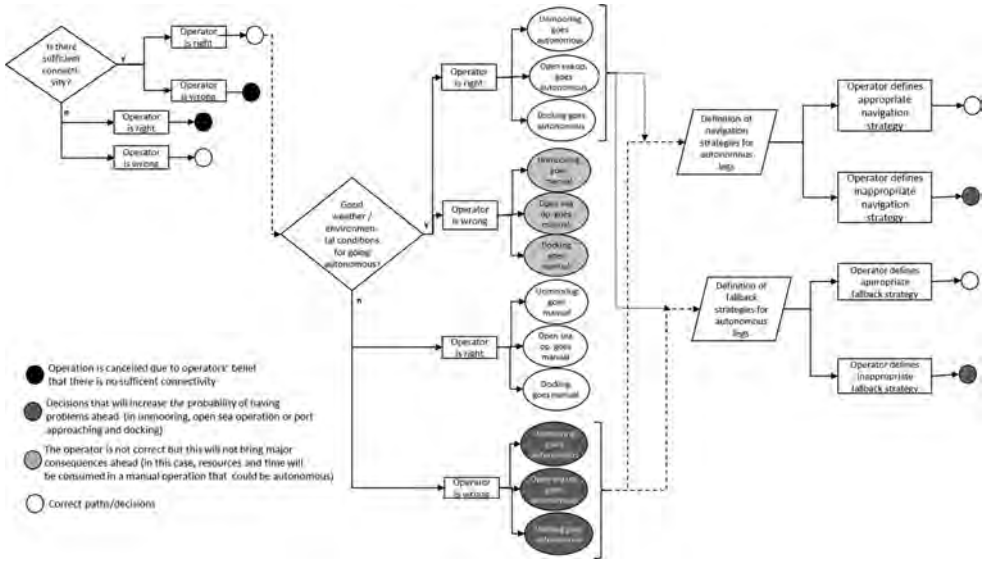


Figure 2. Interaction scheme for voyage planning.

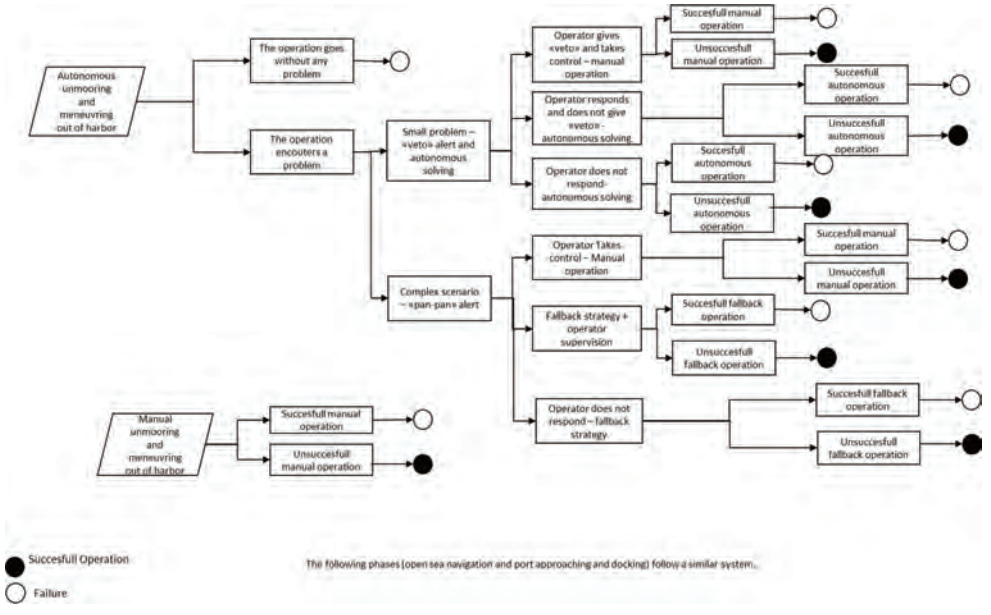


Figure 3. Interaction scheme for unmooring and maneuvering out of harbour.

Table 7. Possible operator decisions for Task 7, in manual operation.

Task 7 (T7): unmooring and maneuvering out of harbor by tele-operation	
T7_path 1: Operator successfully operates the ship	**T7_path 2:** Operator fails to operate the ship

the following phases (navigation phases). These are the HFEs that could lead to the "failure" final events in Figure 1.

The Human Failure Events presented in Table 8 and 9 are defined rather broadly, and can be decomposed, if needed, to identify sub-HFEs. In this sense, they are a general representation of what could go wrong, in terms of human failure, in the autonomous ships operation. They are a starting point that allows to analyze, for each HFE, the crew cognitive processes involved, in order to identify more specific Failure Modes that would lead to each HFE. Furthermore, for each Failure Mode it will be possible to identify and assess the factors that influence the operator's decisions and actions—the Performance Influencing Factors (PIFs).

In this sense, as stated in Section 1, the identification of Human Failure Events is the first step towards a solid Human Reliability Analysis.

The description of the operator-system interactions in autonomous ships and possible HFEs

Table 8. Human failure events in voyage planning phase.

Path	Human Failure Event	Description
T1 path 1	Failure to correctly assess connectivity level	The operator is wrong about the low level of connectivity. The operation goes on, and the low level of connectivity can lead to communication problems between the SCC and the ship.
T2 path 1	Failure to correctly define primary strategy	During definition of the primary strategy for each leg the operator believes the conditions are adequate for autonomous operation when, in that situation, it should be manual (tele-operated)
T3 path 1	Failure to define adequate navigational strategy	The operator defines an inadequate navigation strategy. This will increase the probability of having problems ahead and a "veto" or "pan-pan" situation
T4 path 1	Failure to define adequate fallback strategy	The operator defines an inadequate fallback strategy. In case there is a "pan-pan" situation the fallback strategy will be followed by the ship, if the operator does not take manual control of the ship

Table 9. Human failure events in navigation phases.

Path	Human Failure Event	Description
T5_path 1 **T6_path 1**	Failure to respond to an alert	The operator does no respond to an alert, which may be a "veto" alert or a "pan-pan" alert.
T5_path 3_2 **T6_path 3_2** **T7_path 2**	Failure to remote-operate the ship	The operator is manually operating the ship, which may be after a "veto" or a "pan-pan" alert or may be from the beginning of that operation, in case it was defined to be manual.
T5_path 2_1 **T6_path 2_1**	Failure to take over control of the ship when necessary	The operator trusts the autonomous solution or fallback strategy and does not take over control of the ship in a situation where this is needed

deriving from it demonstrates that there is still room for human failure in its operation. The assessment of human error, therefore, cannot be neglected or minimized when considering autonomous ships' safety and reliability.

4 CONCLUDING THOUGHTS

This paper demonstrates that although compared to conventional ships the human interaction is reduced in autonomous ships, the human still plays a role, with a potential for human error that has to be considered.

One of the concerns regarding autonomous ships' operation is the new risks they can pose, and how to asses them. Being a novel operation, the possible interactions between operator and autonomous ships are not yet clear, but this paper contributes to identifying these interactions and modeling it. The paper also identifies, at a high level, possible Human Failure Events deriving from these interactions.

Three HFEs deserve particular attention, for they can lead/contribute to an accident such as collision, grounding: a failure to respond to an alert, a failure to remotely operate the ship, and a failure to take over control of the ship when necessary. A deeper analysis of these events is needed to identify possible failure modes and the factors that can influence them. That analysis, in the context of an HRA, will make it possible to identify opportunities to reduce the likelihood of critical human failures.

Moreover, it can be a basis for discussing whether autonomy will indeed reduce the likelihood of accidents caused/aggravated by human error—and to determine the appropriate level of autonomy that can lead to a safer operation.

It is important to point that this paper approaches human actions focusing on their contribution to accidents. Actions of operators onboard, however, contribute also to avoiding accidents and/or to reducing the severity of their consequences. This aspect needs to be evaluated as well in further discussions on the shift from onboard to onshore operation.

REFERENCES

Blanke, M., Henrique, M. and Bang, J. 2017. A pre-analysis on autonomous ships. DTU Management Engineering

Boring, R.L. 2014. Top-Down and Bottom-Up Definitions of Human Failure Events in Human Reliability Analysis, in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE, p. 6.

Chen, J.Y.C., Haas, E.C. and Barnes, M.J. 2007. Human performance issues and user interface design for teleoperated robots, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 37(6), pp. 1231–1245. doi: 10.1109/TSMCC.2007.905819.

Ekanem, N. 2013. A Model-Based Human Reliabiliyt Analysis Methodology (PHOENIX Method). University of Maryland.

EMSA 2014. 'Annual overview of marine casualties and incidents 2014', p. 76. Available at: http://www.emsa.europa.eu/news-a-press-centre/external-news/item/2303-annual-overview-of-marine-casualties-and-incidents-2014.html.

Kariuki, S.G. and Löwe, K. 2007. Integrating human factors into process hazard analysis, Reliability Engineering and System Safety, 92(12), pp. 1764–1773. doi: 10.1016/j.ress.2007.01.002.

Laurinen, M. 2016. Remote and Autonomous Ships: The next steps, AAWA: Advanced Autonomous Waterborne Applications, p. 88. Available at: http://www.rolls-royce.com/~/media/Files/R/Rolls-Royce/documents/customers/marine/ship-intel/aawa-whitepaper-210616.pdf.

MUNIN 2016. Research in Maritime Autonomous Systems Project Results and Technology Potentials. Available at: http://www.unmanned-ship.org/munin/wp-content/uploads/2016/02/MUNIN-final-brochure.pdf.

National Institute of Standards and Technology (NIST) 2008. Autonomy Levels for Unmanned Systems (ALFUS). *NIST Special Publication 1011-I-2.0*. Gaithersburg.

Ottesen, A. (2014) Situation Awareness in Remote Operation of Autonomous Ships, pp. 1–12.

Parasuraman, R., Sheridan, T.B. and Wickens, C.D. 2000. A model for types and levels of human interaction with automation, *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 30(3), pp. 286–297*. doi: 10.1109/3468.844354.

Porathe, T. 2013. Maritime unmanned navigation through intelligence in networks: The MUNIN Project, 1*2th International Conference on Computer and IT Applications in the Maritime Industries*, (April), pp. 15–17. Available at: http://publications.lib.chalmers.se/publication/176214-maritime-unmanned-navigation-through-intelligence-in-networks-the-munin-project.

Porathe, T., Prison, J. and Man, Y. 2014. Situation Awareness in Remote Control Centers for Unmanned Ships. *in Proceedings of the Human Factors in Ship Design & Operation Conference*. London.

Rødseth, Ø.J. and Tjora, A. 2014. A risk based approach to the design of unmanned ship control systems, *Proceeding of the Conference on Maritime-Port Technology, pp. 153–162*.

Sheridan, T.B. and Verplank, W. 1978. Human and Computer Control of Undersea Teleoperators. Massachusetts Institute of Technology. Cambridge.

Swain, A. 1990. Human Reliability Analysis : Need, Status, Trends and Limitations, *Reliabity Engineering and System Safety, 29, pp. 301–313.*

Swain, A. and Guttman, H. 1983. Handbook of human reliability analysis with emphasis on nuclear power plant applications. Washington.

Treat, J.R. et al. 1979. Tri-level Study of the Causes of Traffic Accidents, Department of Transportation, United States of America.

Utne, I.B., Sørensen, A.J. and Schjølberg, I. 2017. Risk Management of Autonomous Marine Systems and Operations, P*roceedings of the 36th International Conference on Ocean, Offshore & Arctic Engineering, pp. 1–10.*

Vagia, M., Transeth, A.A. and Fjerdingen, S.A. 2016. A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed?, A*pplied Ergonomics. Elsevier, 53, pp. 190–202.* doi: 10.1016/j.apergo.2015.09.013.

Wiegmann, D. a. and Shappell, S. a. 2012. A human error approach to aviation accident analysis: The human factors analysis and classification system, Ashgate, pp. 1–182.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Human reliability analysis in NPP: A plant-specific sensitivity analysis considering dynamic operator actions versus accident management actions

D. Kancev, S. Heussen, J.U. Kluegel, P. Drinovac & T. Kozlik
NPP Goesgen-Daeniken AG, Daeniken, Switzerland

ABSTRACT: The human reliability analysis is a method by which, in general terms, the human impact to the safety and risk of a nuclear power plant operation can be modelled, quantified and analyzed. It is an indispensable element of the PSA process within the nuclear industry nowadays. The paper herein presents a sensitivity study of the human reliability analysis performed on a realistic nuclear power plant–specific probabilistic safety assessment model. The analysis is performed on a pre-selected set of post-imitator operator actions. The purpose of the study is to investigate the impact of these operator actions on the plant risk by altering their corresponding human error probabilities in a wide spectrum. The results direct the fact that the future effort should be focused on maintaining the current human reliability level, i.e. not letting it worsen, rather than improving it.

1 INTRODUCTION

The human reliability analysis (HRA) is an integral part of the probabilistic safety analysis (PSA) in the nuclear power plants (NPP) throughout the world. While there are those, who contend that it is the single most important element of the entire plant analysis, others contend that any attempt at rigorous evaluation of human reliability is merely an academic exercise in futility. Human beings are unpredictable, they do not obey the laws of physics, and it is impossible to collect relevant statistical evidence to understand or predict human behavior during complex "real world" situations. While it may be true that human behavior is not as statistically predictable as rolling dice, it is also true that the probabilistic framework of PSA acknowledges that all data are uncertain, and it rigorously accounts for these uncertainties. Lack of directly relevant experience and statistical evidence does not preclude a fundamental understanding of the motivations, physical and psychological factors, and external constraints that would most strongly influence human response in a specific situation.

The HRA has a multirole purpose: identifying critical human-system interactions, i.e. human actions; modelling and quantifying the associated human error probability (HEP); HRA optimization. Various preventive and mitigative operator actions (OA) are planned, trained and employed for various accident scenarios that are postulated in the nuclear industry.

There are various methods, models and databanks with estimated HEPs by the help of which qualitative and quantitative analysis of the reliability of OAs in the NPPs can be performed. The inherent variability of the human performance under different conditions and for different functions implies relatively wide uncertainty bounds for the HEP estimates. The uncertainty should be generally smaller for the routine tasks such as test, maintenance, normal control room operations and higher for the OAs as a response to an abnormal event (IAEA 1995, 1996, U.S. NRC 1983 a, b, c). The literature offers a wide spectrum of studies in regards to the various HRA qualitative and quantitative methods (Hannaman & Worledge 1988, Lydell 1992, Moieni et al. 1994, Kim & Seong 2006, Khalaquzzaman et al. 2010, Podofillini & Dang 2013), its coupling with the deterministic safety analysis (Cepin & Prosek 2008), as well as HRA uncertainty and sensitivity studies (Fujimoto et al. 1994, Cepin 2008, Bedford et al. 2013, Baraldi et al. 2015).

In that sense, the presented paper summarizes a HRA sensitivity study performed on a realistic NPP—specific PSA model. The NPP of interest is the Goesgen-Daeniken NPP (KKG) in Switzerland, a 3-loop PWR plant. The plant PSA model is prepared with the RISKMAN® analysis tool (ABS Consulting Ltd. 2008, 2015).

Firstly, all the Type C or post-initiator OAs modelled within the PSA model (ca. 30 different actions with their corresponding split fractions) are subjected to sensitivity analysis such that

their corresponding HEP values are being altered through a wide spectrum of values. Consequently, the quantitative effects on the PSA Level 1 (L1) and Level 2 (L2) risk measures – Core Damage Frequency (CDF) and Large Early Release Frequency (LERF) – as well as the qualitative effects regarding the general risk profile and the changes in the OAs importances are studied.

Secondly, the analysed OAs are then delineated into dynamic, post-initiator OAs on one side and accident management actions on the other side and are subjected to sensitivity analysis given the plant risk as well as their quantitative and qualitative impacts compared and discussed.

2 MODEL

2.1 *Preliminary considerations*

The PSA model herein considers three general groups of OAs:

Type A or pre-initiator system-specific OAs: This type of action involves routine activities such as restoring a component or flow path to normal after the completion of testing, inspection, maintenance, instrument calibration, etc. These system specific activities are typically performed by one or more individuals as part of their normal workday duties. They are not related directly to operator actions or equipment response during the plant transient after an initiating event. However, errors may leave important equipment disabled and require additional dynamic actions to restore it to service during the transient. These routine testing, maintenance, and surveillance actions are identified in each system analysis and are quantified as specific causes for equipment inoperability.

Type B OAs: The second type of human action that may be considered in a PSA is a personnel error that directly causes or contributes to an initiating event. These human errors are not quantified separately for most initiating events in contemporary PSAs because the available initiating event frequency data contain contributions from all causes, including human errors. Most of the initiating event frequencies for the KKG PSA are quantified from a combination of generic and plant-specific data. The initiating event databases do not differentiate among the specific causes for each type of event. Since the frequency of human errors is already included in the initiating event data, these errors are not quantified separately for any initiating events that are derived directly from generic and plant-specific experience.

Type C or post-initiator OAs: The third general type of human action in a PSA is a scenario-specific, directed mission activity. This action is an integral part of plant response to an initiating event. The operators must accomplish well-defined tasks for manual initiation, control, and alignment of plant emergency equipment or selected backup systems. These tasks are generally guided by the plant emergency response procedures. The available time window for successful response, the type of action that must be taken, and other factors that influence operator stress and confusion are determined by the type of event being evaluated and all preceding actions during a specific response scenario. These actions are incorporated as distinct decision points in the KKG PSA event tree models. This type of actions—the post-initiator OAs—are in the focus of the HRA sensitivity study herein. These type C post-initiator actions are being divided in two general groups: the immediate, dynamic OAs as a post-initiator accident OAs and the accident management actions (AMA). They are presented in the following table (Table 1).

In addition to the OAs encompassed by Table 1, the following three actions are also considered: L3 – late fire extinguishing system operation; L4 – fire extinguishing by fire brigade before ignition in adjacent room.

The part of the top event (TE) VU—related to the operator failure to clean debris after season event and thus rendering the main water intake unavailable—is also considered.

2.2 *Analysis and calculations*

The plant PSA model is prepared with the RISKMAN® analysis tool. It is a small fault tree (FT) – large event tree (ET) linking approach software. Each of the OAs summarized in Table 1 is modelled as a separate top event (TE) within the KKG PSA model. Each of these TEs are represented via multiple split fractions (SF), modelling different variants of the corresponding OAs under different boundary conditions. The failure probability of each OA is calculated as a logical OR between the cognitive/diagnosis and the implementation/manipulation failure probability. Failure probability density functions (PDF) – accounting for the HRA uncertainty—are assigned to each of the cognitive and the implementation part.

At this point, it is important to note that the dynamic OAs comprise not only the immediate, post-initiator, preventive measures but also recovery measures within these actions. Namely, one of the advantages of the SF modelling approach is that one can model different variations of the same OA that correspond to different boundary conditions. This, in turn, corresponds to different stages and/or different developments of a given accident.

This paper tends to present a HRA sensitivity study of the values of the L1 and L2 risk measures,

Table 1. Overview of the OAs used in the KKG PSA model.

Type			OA ID	Description
Accident management actions	Dynamic OAs		OALP	Start residual heat removing (RHR) cooling
			OCD	Start active cooldown
			ODP	Depressurize/cooldown
			OEFW	Start emergency feedwater (EFW)
			OPUD	Align demineralized water makeup to feedwater tank
			ORT	Reactor trip
			OSG	Isolate ruptured steam generator (SG)
		Level 1 PSA	OAMFT	Align feedwater tank
			OAMFW	Align fire truck
			OAMIS	Isolate large containment openings
			OAMLI	Isolate letdown line
			OAMPB	Depressurize vie primary pressure relief
			OAMPF	Align fire water to low head (LH) pump path
			OAMRW	Align flood tanks to operating high pressure injection (HPI) pump
			OAMSR	Open SG relief
			OAMTH	Align TH17/37 and VX01/02
			OAMFL	Injection and recirculation via TH17/37 after recovery of bus FL/FM
		Level 2 PSA	OAMIGA	H2 recombiners placed into service prior to vessel breach
			OAMIGB	H2 recombiners placed into service after to vessel breach
			OAMVA	Containment filtered venting system (CFVS) placed into service prior to vessel breach
			OAMVB	CFVS placed into service after to vessel breach
			OAMVC	CFVS scrubbing tank re-filled as necessary
			OAMVD	Isolate CFVS to prevent large release

i.e. CDF and LERF respectively, as a function of the altering values of the HEP for the dynamic OAs and the AMAs. By simple assignment of a scalar parameter (e.g. "*SENHRA*", "*SENHR3*") to the basic event (BE) calculation module where the OA failure probabilities are being calculated, one can alter all the wanted HEPs and hence, perform a sensitivity study. In that sense, the *SENHRA* parameter is part of the data module of the applied KKG PSA model. All the HEP related to all the dynamic OAs as well as AMAs within the KKG model are multiplied by the *SENHRA* parameter. In this way, one can conduct systematic sensitivity analyses by simply altering the value of the *SENHRA* parameter. Within the nominal model, i.e. within the base case, value of 1.0 is assigned to *SENHRA*. Further on, the *SENHRA3* parameter is assigned to the HEP values related only to the dynamic OAs in addition to the already assigned *SENHRA* parameter. In such a way, one can perform HRA sensitivity analysis only to the OA group, without altering the AMA group.

Two main cases within the sensitivity study are analysed:

i. altering the HEP values of all the dynamic OAs as well as all the AMAs together (via altering the parameter "*SENHRA*");
ii. altering the HEP values of the dynamic OAs only, while leaving the AMAs with their nominal HEP values (via altering the parameter "*SENHR3*").

Figure 1 and Figure 2 present the first of the above-mentioned cases—altering the HEP values of both the dynamic OAs as well as the AMAs defined within chapter 2.1. The effects of this sensitivity case show that the potential "worsening" of the HEP values would have much higher consequences on the CDF and LERF in terms of risk increase than the "improvement" of the HEP values might have on reducing this risk. In other words, by increasing the HEP values by a factor of 10, the CDF increases by factor 3 and the LERF by 80%. By worsening, i.e. increasing the HEP values by a factor of 100, the CDF increases by factor 50 and the LERF by factor 20. On the other side, by improving the HEP values by a factor of 10, the CDF reduces by merely 7.5% and the LERF by 7%.

By improving the HEP values by a factor of 100, the CDF reduces by merely 8% and the LERF by 7.6%.

Figure 3 and Figure 4 present the second of the above-mentioned cases—altering the HEP values of the dynamic OAs only, while leaving the AMAs with their nominal HEP values (via altering the parameter *SENHR3*).

The effects of this sensitivity case are qualitatively the same as the ones for the first sensitivity case. Namely, they show that the potential "worsening" of the HEP values would have much higher consequences on the CDF and LERF in terms of risk increase than the "improvement" of the HEP values might have on risk reduction. In addition to this, however, one can also see that if the two cases are compared on quantitative basis as well, they are relatively close to each other. This is especially true in the case of LERF (Figure 6). Figure 5 and Figure 6 address this comparison as separate presentation. Figure 5 presents the calculated CDF for both the above-mentioned cases. Figure 6

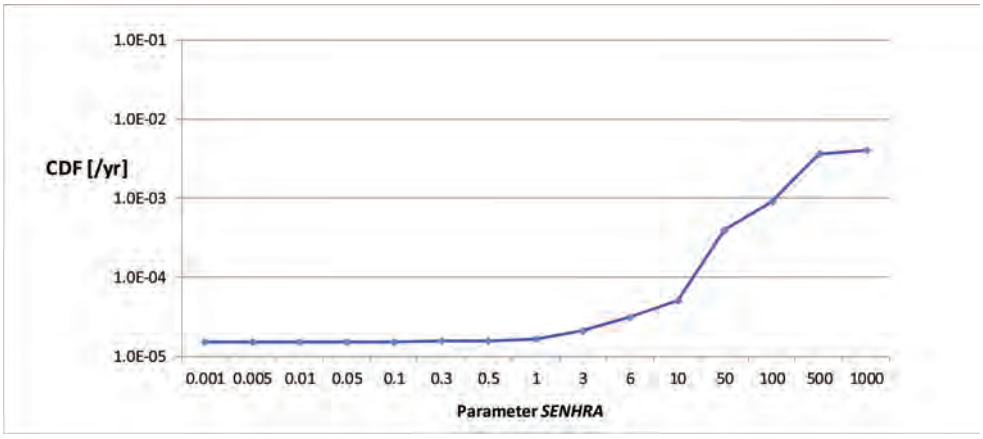


Figure 1. Sensitivity of the CDF as a function of the parameter *SENHRA*.

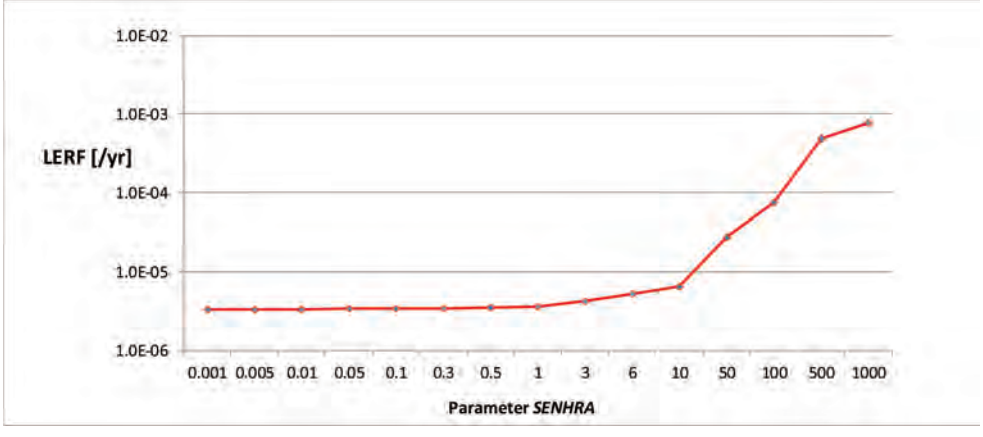


Figure 2. Sensitivity of the LERF as a function of the parameter *SENHRA*.

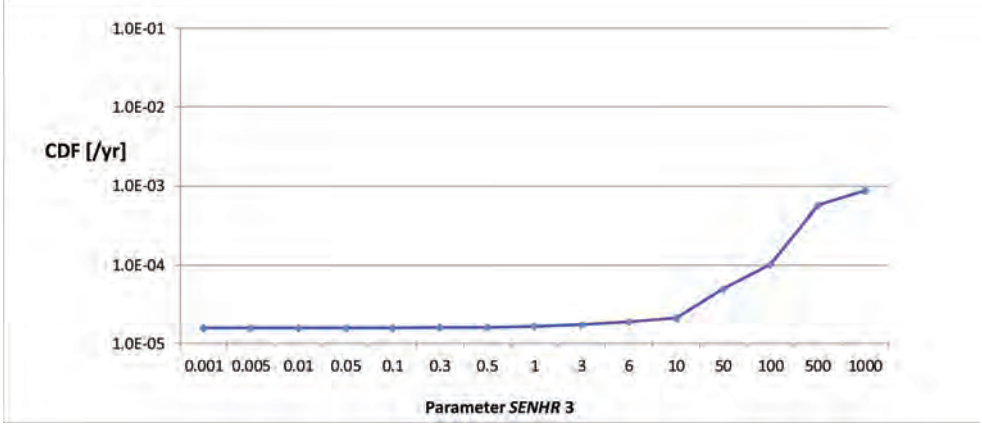


Figure 3. Sensitivity of the CDF as a function of the parameter *SENHR3*.

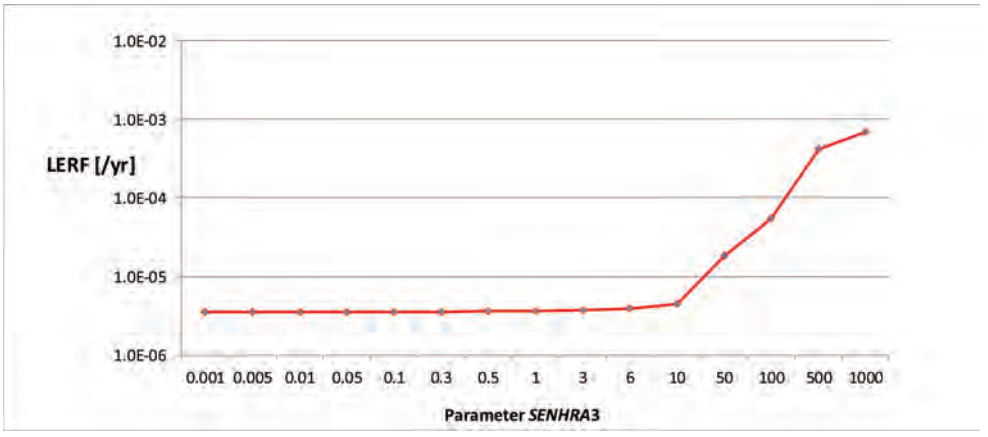


Figure 4. Sensitivity of the LERF as a function of the parameter *SENHR3*.

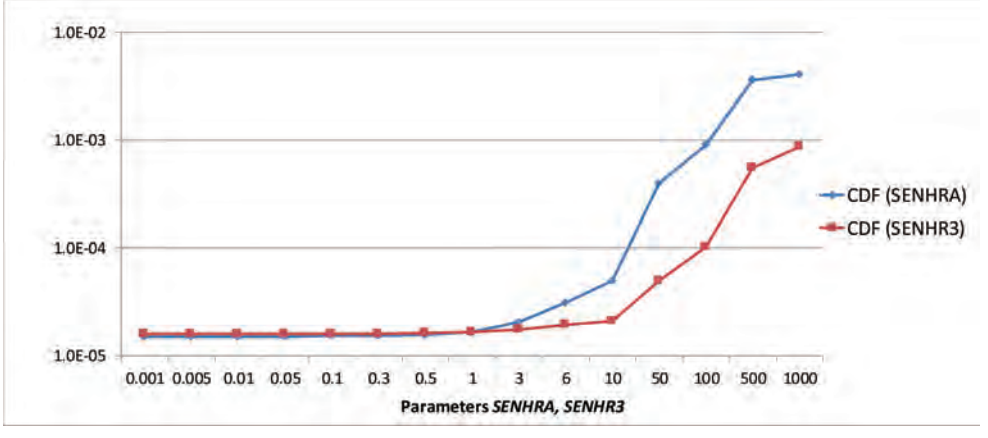


Figure 5. Comparison of the CDF between the two cases: CDF (*SENHRA*) vs. CDF (*SENHR3*).

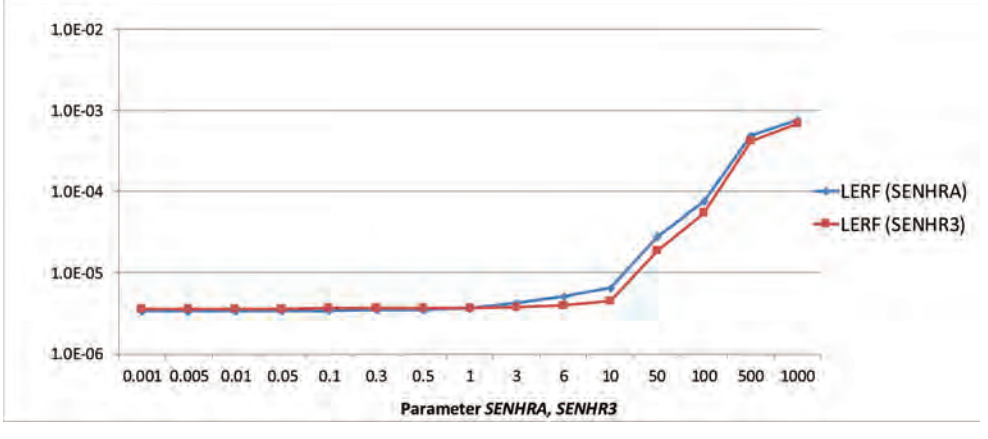


Figure 6. Comparison of the LERF between the two cases: LERF (*SENHRA*) vs. LERF (*SENHR3*).

presents the calculated LERF for both the above-mentioned cases.

In order to inspect the influence of the AMAs solely, a third case is established:

iii. altering the HEP values of the AMAs only, while leaving the dynamic OAs with their nominal HEP values (via altering the parameter "*SENHR1*");

Figure 7 and Figure 8 present this third case—altering the HEP values of the AMAs only, while leaving the dynamic OAs with their nominal HEP values (via altering the parameter *SENHR1*). By worsening the HEP values by a factor of 100, the CDF increases only by factor 2.5 and the LERF only by ca. 25%. By improving the HEP values by a factor of 100, the CDF and LERF reduce by ca. 6–7%. Thus, it is clear that the effect of the AMAs, seen in general as a group of personnel actions, is less than the one of the dynamic OAs presented in case ii.

When comparing Figure 3 with Figure 7 and Figure 4 with Figure 8, it can be concluded that the changes in the plant risk (CDF, LERF) due to altering (increasing) the HEP values are predominantly governed by the changes affecting the dynamic OAs solely when compared to the case where solely the AMA HEP values are being altered (increased).

Of course, the above stated is a general conclusion in a sense of comparing the personnel action

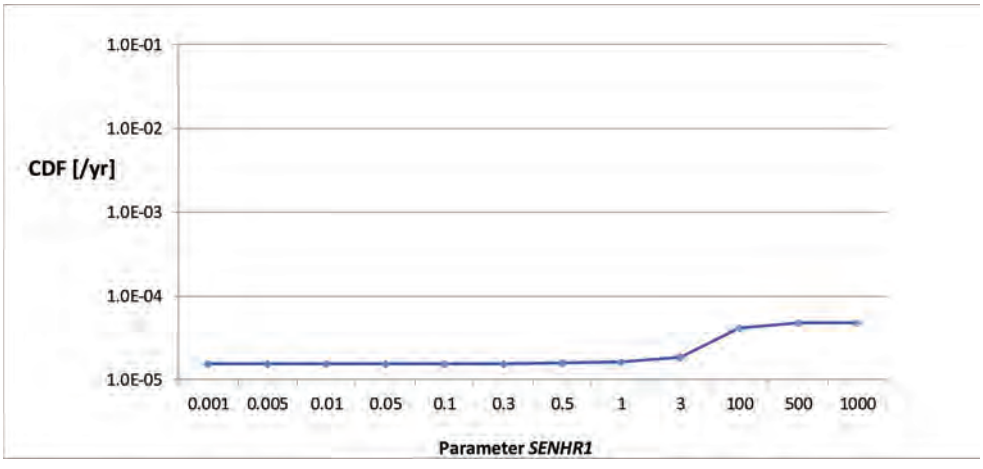


Figure 7. Sensitivity of the CDF as a function of the parameter *SENHR1*.

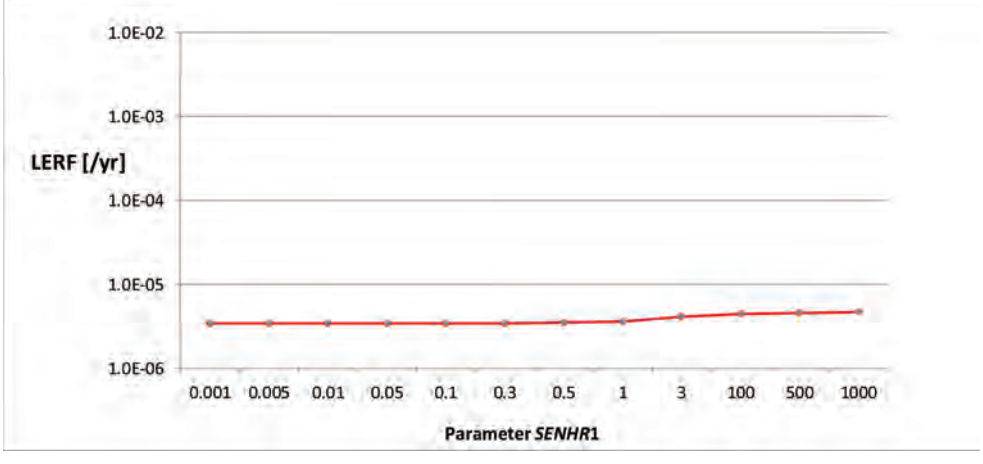


Figure 8. Sensitivity of the LERF as a function of the parameter *SENHR1*.

groups (dynamic OAs vs. AMAs) with each other. There are, of course, AMAs whose risk achievement worth (RAW) is higher than the RAW of some dynamic OAs.

3 DISCUSSION AND CONCLUSIONS

Probabilistic studies of risks show that the human factor can considerably contribute to overall risk. The potential for and mechanisms of human error to affect plant risk and safety is evaluated by the human reliability analysis. The HRA has quantitative and qualitative aspects, aimed at designing operator interfaces that will minimise operator error and provide for error detection and recovery capability. The objectives of HRA therefore, are to assure that potential effects on plant safety and reliability are analysed and that human actions that are important to plant risk are identified so that they can be addressed in both PSA and plant design.

The presented paper summarizes a HRA sensitivity study performed on a realistic NPP–specific PSA model. This sensitivity analysis is performed in order to investigate the role of the HEPs of two groups of OAs to the overall spectrum of plant risk.

Type C or post-initiator OAs are in the focus of the study. These type C post-initiator human actions are being divided in two general groups: the immediate, dynamic OAs as a post-initiator accident OAs and the accident management actions—AMAs. Each of the considered OAs is modelled as a separate TE within the KKG PSA model. Each of these TEs are represented via multiple SFs, modelling different variants of the corresponding OAs under different boundary conditions.

The paper presents a HRA sensitivity study of the values of the L1 and L2 risk measures, CDF and LERF respectively, as a function of the altering values of the HEP for the dynamic OAs and the AMAs.

Three cases were analysed:

i. altering the HEP values of all the dynamic OAs as well as all the AMAs together (via altering the parameter *SENHRA*);
ii. altering the HEP values of the dynamic OAs only, while leaving the AMAs with their nominal HEP values (via altering the parameter *SENHR3*);
iii. altering the HEP values of the AMAs only, while leaving the dynamic OAs with their nominal HEP values (via altering the parameter *SENHR1*).

The effects of the first sensitivity case i. show that the potential "worsening" of the HEP values would have much higher consequences on the CDF and LERF in terms of risk increase than the "improvement" of the HEP values might have on reducing the risk. In other words, by worsening the HEP values by a factor of 10, the CDF increases by factor 3 and the LERF by 80%. By worsening the HEP values by a factor of 100, the CDF increases by factor 50 and the LERF by factor 20. On the other side, by improving the HEP values by a factor of 10, the CDF reduces by merely 7.5% and the LERF by 7%. By improving the HEP values by a factor of 100, the CDF reduces by merely 8% and the LERF by 7.6%.

The effects of the second sensitivity case ii. are qualitatively the same as the ones for the first sensitivity case. In addition to this, however, one can also see that if the first and the second case are compared on quantitative basis as well, they are relatively close to each other. This is especially true in the case of LERF.

In order to inspect the influence of the AMAs solely, the third sensitivity case is run. By increasing, i.e. worsening the HEP values by a factor of 100, the CDF increases only by factor 2.5 and the LERF only by ca. 25%. By improving the HEP values by a factor of 100, the CDF and LERF reduce by ca. 6–7%. Thus, it is clear that the effect of the AMAs, seen in general as a group of personnel actions, is less than the one of the dynamic OAs presented in case ii.

The usability of the herein performed HRA sensitivity analysis is summarized through the conclusions derived below in text. The interplay of the two identified type C human action groups (dynamic OAs and AMAs) and their significance and contribution to plant risk is discussed.

From the results of the sensitivity study can be concluded that the changes in the plant risk (CDF, LERF) due to altering the HEP values are predominantly governed by the changes affecting the dynamic OAs when compared to the case where solely the AMA HEP values are altered. This coincides with the comparison of the RAW factors of the dynamic OAs vis-à-vis the ones for the AMAs. In general, the former group have higher RAW values than the latter.

Of course, the above stated is a general conclusion in a sense of comparing the personnel action groups (dynamic OAs vs. AMAs) with each other. There are, however, AMAs whose RAW is higher than the RAW of some dynamic OAs. Additionally, the effects of both the groups on the plant risk are not additive. Still and as already stated, the plant risk is much more sensitive to the increase of the HEP values of the dynamic OAs considered alone than to the increase of the HEP values of the AMAs considered alone.

Therefore, the general conclusion is that the focus should be put on maintaining the boundary

conditions for diagnosis and execution of the general group of all the personnel type C actions on such a level so that the current reliability of their performance would not be endangered. In other words, future effort should be focused on maintaining the current HEP values, i.e. not letting them get worse, rather than improving these HEP values. Further on and more specifically, when already delineating between the dynamic OAs on one side and the AMAs on the other, the focus should be set foremost to maintaining the HEP values of the dynamic OAs group, since their deteriorating effect on CDF, LERF can be much higher than the one of the AMAs.

REFERENCES

ABS Consulting Ltd. 2008. GPSA 2009 Gesamtdokumentation. R-1699596-1751.

ABS Consulting. 2015. GPSA 2015 Gösgen Probabilistic Safety Assessment. R-2129227-1853.

Baraldi, P. et al. 2015. Comparing the treatment of uncertainty in Bayesian networks and fuzzy expert systems used for a human reliability analysis application. *Reliability Engineering & System Safety* 138: 176–193.

Bedford, T. et al. 2013. Screening, sensitivity, and uncertainty for the CREAM method of Human Reliability Analysis. *Reliability Engineering & System Safety* 115: 100–110.

Cepin, M. & Prosek A. 2008. Success criteria time windows of operator actions using RELAP5/MOD3.3 within human reliability analysis. *Journal of Loss Prevention in the Process Industries* 21(3): 260–267.

Cepin, M. 2008. Importance of human contribution within the human reliability analysis (IJS-HRA). *Journal of Loss Prevention in the Process Industries* 21(3): 268–276.

Fujimoto, H. et al. 1994. Sensitivity study of human errors as a basis for human error reductions on new safety system design. *Reliability Engineering & System Safety* 45(1–2): 215–221.

Hannaman, G.W. & Worledge D.H. 1988. Some developments in human reliability analysis approaches and tools. *Reliability Engineering & System Safety* 22(1–4): 235–256.

IAEA. 1995. Human Reliability Analysis in Probabilistic Safety Assessment for Nuclear Power Plants. Safety Series No. 50-P-10, Vienna.

IAEA. 2016. Attributes of Full Scope Level 1 Probabilistic Safety Assessment (PSA) for Applications in Nuclear Power Plants. IAEA-TECDOC-1804, Vienna.

Khalaquzzaman, M. et al. 2010. Quantification of unavailability caused by random failures and maintenance human errors in nuclear power plants. *Reliability Engineering & System Safety* 44(1–2): 27–55.

Kim, M.C. & Seong P.H. 2006. A computational method for probabilistic safety assessment of I&C systems and human operators in nuclear power plants. *Reliability Engineering & System Safety* 91(5): 580–593.

Lydell, B.O.Y. 1992. Human reliability methodology. A discussion of the state of the art. *Reliability Engineering & System Safety* 36(1): 15–21.

Moieni, P. at al 1994. Advances in human reliability analysis methodology. Part I: frameworks, models and data. *Reliability Engineering & System Safety* 44(1–2): 27–55.

Podofillini, L. & Dang V.N. 2013. A Bayesian approach to treat expert-elicited probabilities in human reliability analysis model construction. *Reliability Engineering & System Safety* 117: 52–64.

U.S. NRC. 1983a. Handbook of Human Reliability Analysis with emphasis on Nuclear Power Plant Applications. NUREG/CR-1278.

U.S. NRC. 1983b. Accident Sequence Evaluation Program Human Reliability Analysis Procedure. NUREG/CR-4772.

U.S. NRC. 1983c. A Procedure for Conducting Human Reliability Analysis for Nuclear Power Plants. NUREG/CR-2254.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

The weighting method's impact on the weighting process in decision making problems

A. Tzioutziou & Y. Xenidis
Department of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

ABSTRACT: Assigning values to weights in a multi-criteria decision making problem is critical, since it introduces the decision maker's perception and preference over the importance and value of the decision problem's criteria and alternatives, respectively. This paper investigates theoretically and experimentally the impact of the applied weighting method on the decision maker's determination of weights. The research develops a methodology to evaluate the impact that a weighting method may have on the decision maker's attitude concerning the assignment of weighting values, comprising; a) a psychometric test revealing the decision maker's attitude against risk and ambiguity with a new modeling approach based on a psychometric function and b) an assignment of weighting values by the decision maker with different weighting methods. The results demonstrate that the expression and particularly the ranking of decision maker's attitudinal preference are affected by the weighting method and this impact is measurable through the proposed methodology.

1 INTRODUCTION

The principles of decision analysis set the framework wherein decision makers determine weights through the use of various methods and diverse approaches. According to the current theoretical approaches, decision analysis is characterized by two discrete principles, namely the normative and the descriptive (Barzilai, 2010; Riabacke, Danielson and Ekenberg, 2012; Aliev et al., 2016), whilst there is also the prescriptive principle, which constitutes a combination of them (Jia, Fischer and Dyer, 1998; Riabacke, Danielson and Ekenberg, 2012; Aliev et al., 2016). The decision analysis principles determine the characteristics of the method's input and process, imposing the use of numerical data or the combination of numerical data with verbal expressions.

Recently, great importance has been given to the cognitive process of weight determination, which includes three conceptual phases, namely the extraction of data, their representation and the interpretation (Riabacke, Danielson and Ekenberg, 2012). The two main classes of weight elicitation methods are the precise and the approximate methodological approaches (Jia, Fischer and Dyer, 1998; Bottomley and Doyle, 2001; Barzilai, 2010; Riabacke, Danielson and Ekenberg, 2012; Hafezalkotob, Hafezalkotob and Sayadi, 2016; Wang, Wang and Zhang, 2016; Podinovskaya and Podinovski, 2017). The literature review revealed that new weighting methods are constantly added to the existing ones, integrating contemporary mathematical analysis (Abidin, Rusli and Shariff, 2016), as well as complementary techniques for the determination of weights, like integration of weights and consensus building (Beliakov and James, 2015; Peng et al., 2015; Blagojevic et al., 2016).

The determination of attribute weights is essentially a problem of preference formation in different conditions, with many possible limitations. The concept of preference describes the degree that a subject desires a potential outcome (Hogarth, 2010). Its formation is related with multiple cognitive, emotional and even biological mechanisms, connected with cognition and behavioral traits (Hogarth and Einhorn, 1990; Weller, Levin and Bechara, 2010; Panno, Lauriola and Figner, 2013; Naili et al., 2015). The study of those mechanisms, along with the external conditions of risk and ambiguity from the fields of Psychology and Economic Theory could enhance the comprehension of the preference formation and launch a further challenging investigation; the prediction of preferences (Weller, Levin and Bechara, 2010; Retief et al., 2013; Brand et al., 2014; Roth and Voskort, 2014; Johansen and Rausand, 2015; Csermely and Rabas, 2016; Shao, Taisch and Ortega-Mier, 2016; Thomas, 2016; van Winsen et al., 2016; Jern, Lucas and Kemp, 2017).

Based on the above, there is evidence to connect the behavioral preferences with the process of weight elicitation. The paper investigates the determination of weights as an expression of preference and behavioral attitude, which can be described

and modeled with a psychometric approach and a new methodology. This new psychometric approach is presented in Section 2 and the new methodology for recognition of behavioral attitudes with the use of the psychometric function is discussed in Section 3. The cognitive process leads the decision maker to assign weight assessments, and this research examines experimentally this aspect of the weighting method in Section 4. Section 5 presents the conclusions of this work.

2 A NEW APPROACH FOR REVEALING BEHAVIORAL ATTITUDE

Previous works on the field offer the theoretical background to determine a new approach for the identification and classification of behavioral attitudes, based on the concept of psychometric function. The definition of this concept is presented in subsection 2.1, while the proposed approach for the use of the psychometric function for preference expression under risk or ambiguity is discussed in subsection 2.2.

2.1 *Definition of the psychometric function*

The psychophysic problem describes the formation of the mathematical operation which expresses the behavioral preference (Robert, 1985). The psychometric function that has been introduced in the context of the Signal Detection Theory offers an effective solution to the psychophysic problem, according to which the preferences are sensory responses that depend on a parameter of a specific stimulus for decision making (Gold and Ding, 2013; Hunter, 2017). The psychometric function can be applied in decision making problems with only two alternatives, also known as yes-no problems, in order to demonstrate the preference shift affected by the stimulus intensity (Gold and Ding, 2013) and reveal the behavioral attitude of the subject. The stimulus for making a choice is the probability of an event (Hunter, 2017) and especially for the decision making context, this stimulus becomes the gain probability or else the expected added value.

In practice, the psychometric function allows the simultaneous study of multiple preference variables and characteristics (Gold and Ding, 2013). In the context of the present paper the psychometric function takes the form of the Sigmoid function, which is defined by the formula $\frac{1}{1+e^{-(x-a)/b}}$, as shown in Figure 1 (Hunter, 2017). In the function variable a, determines the shift of the sigmoid and is interpreted as the preference's bias, whilst variable b, determines the steepness of the slope and characterizes the preference's sensitivity (Gold and Ding, 2013; Hunter, 2017).

2.2 *Modeling attitudinal preferences with the psychometric function*

In order to fit the psychometric function in choice problems affected by the phenomenon of risk and ambiguity aversion, it is essential to translate the preference expression into arithmetic terms. The phenomenon of aversion can be defined as a shift of preference, and the subject's response can be translated into values of the interval [0, 1], where a zero value reflects the risk or ambiguity aversion and a value of 1 reflects the risk or ambiguity seeking, respectively. The relevant tasks examine the subject's multiple responses to a specific problem, while the gain probabilities progressively increase and they lead to the shift of the preference, from 0 to 1, which can be represented both arithmetically and graphically. When there are several values of such responses from multiple tasks, the average of all the responses is extracted for every 10 hundredths of gain probability rise, so as to extract the average response pattern and produce the graph of Figure 2. This sensory response pattern represents graphically the shift of attitude for all the relevant expressions and also determines the appropriate

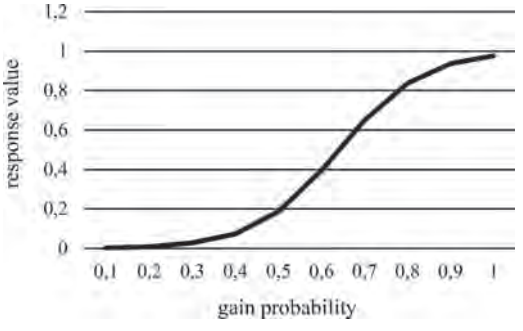


Figure 1. The graph of the psychometric function with $a = 0.642$ and $b = 0.096$, which reflects the attitude towards risk.

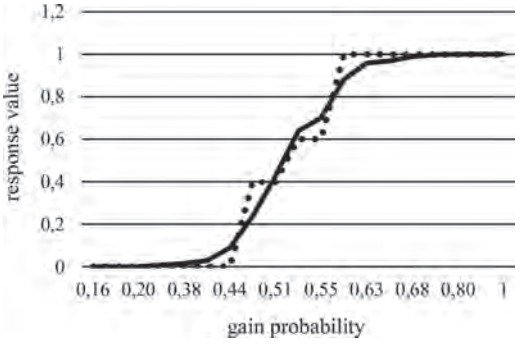


Figure 2. Fitting of the sigmoid function under ambiguity.

psychometric function which best reflects the subject's behavioral attitude. In particular, the sigmoid function is determined by the gain probability variable x and the output $f(x)$ is the value of the subject's response for the given probability, whilst the function fitting is estimated with non-linear regression and the least squares' method. Practically, the variables a and b are estimated accurately with the use of Solver Tool in Microsoft Excel 2013 and the psychometric function is defined, as shown in the graph of Figure 2.

3 A METHODOLOGY FOR BEHAVIORAL ATTITUDE IDENTIFICATION

The new approach for revealing the subjects' behavioral traits and attitudes was presented in Section 2. This research explores experimentally the utility and the significance of this approach with the methodology described in this Section, and particularly in the subsections 3.1, 3.2 and 3.3. The methodology includes the procedure of identifying attitudes under the conditions of risk and ambiguity, respectively, as presented in the subsections 3.4 and 3.5.

3.1 *The experiment*

The objective of the specific research is the investigation of the individual attitudinal preferences and their potential influence on the determination of weights in decision making. To meet this objective an experiment was realized in two phases. The initial phase included a personality test with control questions for the evaluation of the subjects' comprehension and consistency and a weighting test. The personality tests that were used included checklists of multiple values to two different problems of alternatives evaluations, i.e. a problem of a ballot choice and a problem of the best bet among a set of offered bets. The answers to these tests reflect the subject's respective psychometric function, which is either risk prone or risk averse (Holt and Laury, 2002; Csermely and Rabas, 2016; He, Veronesi and Engel 2017).

The evaluation of the responses led to the identification of the subjects who responded properly to the control questions and so their responses could be considered valid. These specific subjects were selected to participate in the second phase of the experiment, which included some explanatory questions about the subjects' personality and preferences in the conditions of risk and ambiguity, respectively, in order to validate and confirm their behavioral preferences and attitudes. The content of the second-phase tests was the same with the respective tests of the first phase.

3.2 *The sample*

The present research was addressed to 50 engineers with professional experience in decision making in the field of construction industry; 37 of them participated to the experiment providing a satisfactory response rate of 74%. The sample included engineers of different specialties, such as civil engineers, mechanical engineers, architects and surveyors, with various levels of working experience from two to 30 years, representing both the private and the public sector, and including individuals with diverse personalities, behaviors and cultures (Charness, Gneezy and Kuhn, 2013). Based on these features, the sample can be evaluated as satisfactory in the context of this research.

3.3 *Identification of behavioral attitudes*

The initial test was completed successfully by 37 subjects and after the evaluation of the results, 10 of the completed tests, which represent the 37.04% of the initial responses, were excluded from further processing because of incorrect responses to the control questions. Consequently, the number of valid responses in this phase is 27, which is sufficient for the estimation of the subject's preferences in risk and ambiguity and also allows the extraction of weighting tasks' results.

The first analyses of the personality test results lead to the determination and the optimization of each subject's psychometric functions for the shift of preferences in the conditions of risk and ambiguity, respectively. The determination of the sigmoid function requires the estimation of the characteristic values of variables a and b, which reveal the preference's switching point and also allow the identification and the categorization of the subject's behavioral attitudes. Moreover, the sigmoid function and its variables are uniquely determined for every subject in the environment of risk, as well as in the environment of ambiguity. The estimated values of these variables for the risk attitude, and also the corresponding values for the ambiguity attitude are presented in Table 1.

3.4 *Identification of risk attitudes*

The determination of risk attitude with the use of the psychometric function is based on the assumption that the switching point of the subject's preference can describe the behavioral attitude. According to this assumption, the critical switching point is observed when the gain probability is 60% (Holt and Laury, 2002; Csermely and Rabas, 2016; He et al., 2017).

In practice, this means that when the subject changes preference from risk avoidance to risk taking

Table 1. The values of the psychometric function's variables *a* and *b* for the subject's attitudes under risk and ambiguity, respectively.

		Variables for risk attitude		Variables for ambiguity attitude	
Id	SUBJECT	a	b	a	b
1	ANA	0.741	0.100	0.070	0.010
2	ARA	0.767	0.024	0.456	0.001
3	BOU	0.641	0.070	0.456	0.001
4	KOU	0.703	0.134	0.456	0.001
5	GOU	0.502	0.086	0.587	0.083
6	KOI	0.522	0.037	0.457	0.001
7	KOT	0.617	0.080	0.588	0.036
8	MPO	0.698	0.035	0.456	0.001
9	NEV	0.776	0.031	0.540	0.017
10	SMP	0.497	0.104	0.540	0.001
11	TSO	0.589	0.063	0.600	0.002
12	TZI	0.520	0.035	0.565	0.001
13	VAS	0.766	0.027	0.580	0.019
14	VEL	0.609	0.049	0.475	0.096
15	ZEL	0.689	0.089	0.551	0.029
16	ATSO	0.626	0.047	0.520	0.001
17	AC	0.635	0.113	0.456	0.001
18	NA	0.643	0.045	0.550	0.001
19	TSI	0.810	0.114	0.666	0.001
20	FAK	0.657	0.072	0.456	0.001
21	BIL	0.601	0.058	0.490	0.001
22	AGG	0.753	0.095	0.510	0,001
23	PET	0.604	0.053	0.510	0.001
24	THE	0.687	0.123	0.510	0.001
25	TOL	0.457	0.186	0.550	0.001
26	XAM	0.523	0.041	0.456	0.001
27	KAL	0.545	0.001	0.540	0.001

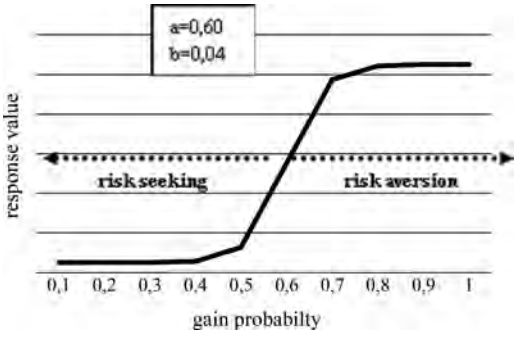


Figure 3. An example of the psychometric function with the critical switching point representing subject's risk attitude.

before the gain probability of 60%, this attitude is characterized as risk seeking. Accordingly, if this shift of preference occurs when the gain probability is greater than 60%, this attitude would be risk averse. In the psychometric function approach, the switching point can be directly identified because it is determined as the value of the characteristic variable *a*, as represented in Figure 3. The results of this analysis demonstrated that in the sample of 27 subjects, 19 of them, representing the 70.37%, are characterized as risk averse, while only 8 subjects, representing the 29.63% are characterized as risk prone.

In addition, it is important to examine the value of variable *b*, which expresses the sensitivity of subject's choice under risk. As the value of variable *b* increases, the sensitivity of the choices also increases and consequently, it is more likely to observe inconsistent responses. The value of variable *b* determines the different levels of sensitivity in the sample as shown in Figure 4 and particularly, in the total of 27, 11 subjects are characterized by *b* value smaller than 0,050 and therefore great consistency in their preferences, whilst 16 subjects demonstrate greater *b* value and three of them demonstrate remarkably great values, which indicate great sensitivity in preference expression.

3.5 *Identification of ambiguity attitudes*

The same analysis was conducted in order to extract results for subjects' attitudes in the condition of ambiguity. The responses' values are modelled with the psychometric function and the variables *a* and *b* are estimated. Based on the Ellsberg Paradox's assumptions about the inherent ambiguity aversion, the critical switching point of preference is set at 60% gain probability (Lauriola and Levin, 2001; Schneider and Nunez, 2015). In the sigmoid function approach, this fact can be translated to variable $a = 0.60$ and the subjects' behavioral attitudes can easily be identified as ambiguity averse or prone, as presented in Figure 5.

The majority of the sample, 25 subjects in total, confirms Ellsberg's assumptions about an inherent ambiguity avoidance, as they demonstrate *a* values between 0.50 and 0.60, while there are only two cases where the value of variable *a* is greater than 0.60, a fact that indicates an ambiguity seeking attitude, as shown in Figure 6. Moreover, the sensitivity of these preferences, as determined by the values of variable *b*, remains small with minor differentiations. The *b* values vary between 0.05 and 0.01, showing that the subjects express their preferences with high consistency and stability.

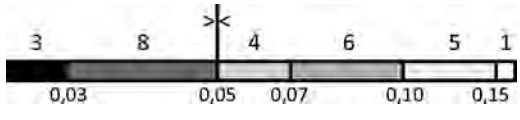


Figure 4. The distribution of the values of variable *b* for the risk attitude.

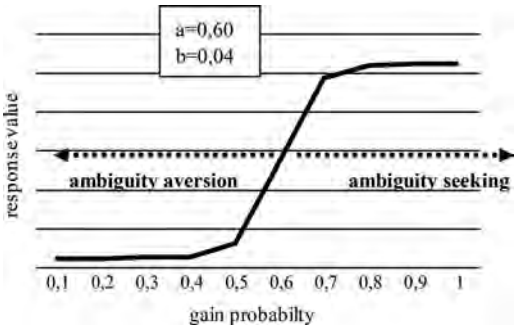


Figure 5. A psychometric function with the critical switching point representing a subject's ambiguity attitude.

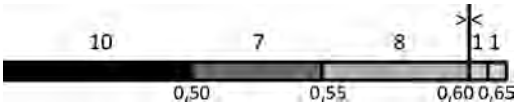


Figure 6. The distribution of the values of variable *a* for the ambiguity attitude.

4 FINDINGS ABOUT THE WEIGHTING METHOD'S IMPACT

The research aims to evaluate the impact that a weighting method may have on the decision maker's attitude concerning the assignment of weighting values. For this purpose, the proposed methodology employs three different weighting methods which are presented in subsection 4.1. The procedure of the analysis and the results are presented in detail in subsection 4.2.

4.1 *The examined weighting methods*

It is a fact that different weight elicitation methods lead to different weighting and consequently different decision results, a phenomenon which is mainly connected to the methodological approach and the cognitive process of weighting, rather than random response errors (Jia, Fischer and Dyer, 1998). The three basic weighting methods selected for this examination are the direct rating, the distance estimation and the pairwise comparison.

One of the most popular weighting methods is the direct rating, which depends on the decision makers' quantitative estimation of their preference. The method leads to a rank order of the alternatives in descending order, and also permits the formation of linear models for capturing that preference (Bottomley and Doyle, 2001). The direct weighting is implemented with the direct evaluation of the criteria or the alternatives with the application of a specific weighting scale. The cognitive approach of this method entails a preliminary examination of all the criteria, so that the subject discerns the significance of each element, and then proceeds to the final evaluation.

In an effort of disengagement from the direct and precise determination of weights, the method of distance estimation is used. According to this weighting process, the decision maker firstly determines the ideal situation and then uses this as reference base for the determination of weights. Every criterion or alternative is compared with the ideal situation, so as to determine the distance between the two situations, with a sense of proportion for all the rest situations. It has been observed by Jia, Fischer and Dyer (1998) that this method's results present significant differentiation which depends on the given distance scale.

The method of pairwise comparison offers a completely different approach for the elicitation of weights. This particular method depends on the concept of trade-off (Jia, Fischer and Dyer, 1998) and also offers the expression of reciprocal preference relations (Xu et al., 2015). In this case, the criteria are not examined generally, but in separate pairs and so the decision maker tends to rely more on intuition, rather than correlation.

4.2 *The examination of the weighting methods*

Initially, the purpose of this research is the examination of the consistency and reliability of each weighting method, as a correlation of the decision makers' behavioral attitudes. Each weighting task includes discrete sub-tasks, which are used to compare the assigned weights and extract the correlations, as a measure of each method's effectiveness. The correlation between direct rating, distance estimation and pairwise comparison has to depend on an approach that is not affected by each method's characteristics. For this reason, the present approach is based on the rank order of the alternatives and so the ranking correlation coefficient, well-known as the Spearman coefficient, is extracted. Additionally, the ranking correlations and methods' consistencies are related to the decision makers' behavioral attitude, so as to investigate any connection between them.

The first weighting test is formed with the method of direct rating. The specific weighting method demonstrates high correlation between its results, as Table 2 shows. The overall correlation coefficient is the average of the correlations of the quantitative and qualitative weightings, which is estimated at 0.922, with a minor variance 0.002. This high coefficient also shows the high consistency of the responses, expressed with the method of direct rating.

Table 2. The correlation coefficients and the compatibility degrees of the three weighting methods.

Id	SUBJECT	Direct rating	Distance estimation	Pairwise comparison
		Correlation coefficient	Correlation coefficient	Compatibility degree
1	ANA	0.946868	0.990680	0.93750
2	ARA	0.950439	0.959866	0.52083
3	BOU	0.930605	0.914056	0.79688
4	KOU	0.944028	0.923726	0.62500
5	GOU	0.971940	0.986747	0.75000
6	KOI	0.941935	0.927940	0.68750
7	KOT	0.946648	0.955168	0.59375
8	MPO	0.806452	0.875113	0.43750
9	NEV	0.880783	0.942050	0.50000
10	SMP	0.868462	0.834058	0.73438
11	TSO	0.848981	0.962911	0.54688
12	TZI	0.924662	0.920021	0.78125
13	VAS	0.942801	0.946782	0.73438
14	VEL	0.940413	0.936341	0.57813
15	ZEL	0.928976	0.981065	0.71875
16	ATSO	0.923407	0.635245	0.57813
17	AC	0.955210	0.927814	0.79690
18	NA	0.940804	0.938397	0.78125
19	TSI	0.931403	0.928515	0.65625
20	FAK	0.875011	0.905473	0.59375
21	BIL	0.945380	0.959723	1.00000
22	AGG	0.943035	0.924585	0.60938
23	PET	0.925352	0.965619	0.76565
24	THE	0.866520	0.850975	0.54688
25	TOL	0.950810	0.953618	0.59375
26	XAM	0.859410	0.863554	0.76565
27	KAL	0.924595	0.897692	0.81250
Averages		**0.921915**	**0.917263**	**0.66727**

Generally, referring to the subjects' attitudinal characteristics, both risk averse and risk prone subjects demonstrate high consistency degree for their weightings with this particular method. Specifically, the risk averse subjects demonstrate the same high consistency 0.922 and smaller variance 0.001, whilst the risk prone subjects are characterized by a slightly lower consistency coefficient 0.911 with variance 0.002. The consistency of weights and its relation with the decision makers' attitude is presented graphically in Figure 7.

The second method examined in this experiment is the distance estimation. In this weighting method, the correlation coefficient average and therefore the consistency of the responses is also high, at the level of 0.917, with slightly increased variance 0.009, compared to the previous method, as demonstrated in Table 2. The method of distance estimation helps the decision makers to determine weights with high consistency. The risk

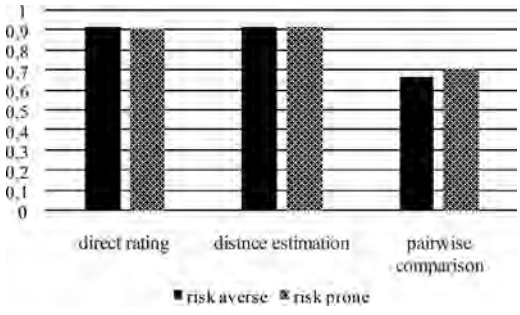


Figure 7. The consistency of responses for the three weighting methods.

averse decision makers express high consistency 0.919 and variance 0.006, while the risk prone decision makers seem equally consistent with consistency average 0.918 and variance 0.003.

These results partially confirm the observation of Jia, Fischer and Dyer (1998) for the increased effectiveness of distance estimation compared to the direct rating, only for the risk prone subjects, whose consistency increased slightly from 0.911 to 0.918, as presented in Figure 7.

The third weighting method examined is the pairwise comparison, where the decision maker's responses have the form of bits of a binary system, instead of a rank order of the alternatives. This is the reason for the determination of a different method for the correlation extraction. Every response with this weighting method was compared to the corresponding response and ranking with each one of the other two examined methods, in order to extract conclusions for the consistency of the method. When the response that is expressed with the pairwise comparison coincides with the response determined with one of the other two weighting methods, then this response is considered true and takes the value 1, whereas in any different case, with no indication of such coincidence, the response takes the value 0. For instance, when a decision maker assigns greater weight to the alternative with 70% gain probability, compared to the alternative with 50% gain probability, in a comparison between these specific alternatives, it is normally expected to select the alternative with 70% gain probability. This case can be described as compatibility between the responses of the direct rating and the pairwise comparison. In any different case, the coincidence cannot be achieved. With this procedure, all the responses of the pairwise comparison are tested for their coincidence with the other applied methods and finally, the average of the response compatibilities is estimated as the compatibility degree of Table 2, that reveals on the one hand the method's reliability, and on the other hand, the subject's consistency with this weighting method.

The compatibility degree average, hence, determines the consistency of weights, and simultaneously reflects the correlation with the two other weighting methods. It is obvious in Table 2 and in Figure 7 that the responses of the pairwise comparison demonstrate a severe lack of consistency, because with the specific weighting method more inconsistent answers occur, which entail, for most of the subjects, an unexpected risk and ambiguity seeking.

The correlation degree average for this method is significantly lower and estimated at 0.677, with high variance 0.031. The risk averse subjects seem to follow this trend, as they demonstrate a correlation average of 0.672 and variance 0.022. However, it is worth to mention that in this case, unlike previous observations, the risk prone decision makers seem slightly more consistent compared to the risk averse subjects, as they have a correlation average of 0.709 and variance 0.009.

5 CONCLUSIONS

This research focuses in the decision makers' determination of weights and features the cognitive process of weighting from the perspective of behavioral preference and attitude. Therefore, a new approach for revealing and classifying behavioral attitudes is proposed, based on the decision maker's preference expression with the use of the psychometric function. The proposed methodology allows the modeling, the recognition, and the classification of individual attitudinal preferences and the parallel examination of other crucial parameters, like sensitivity.

Moreover, the paper investigates the impact of the applied weighting method on the weighting process and consequently, on the determination of weights. A methodology is developed to evaluate the consistency and reliability of the weighting method, as well as the impact that it may have on the decision maker's attitude concerning the assignment of weighting values. The results demonstrate that the weighting methods affect the expression and the ranking of decision maker's attitudinal preference. In particular, the methods of direct rating and distance estimation demonstrate high consistency of responses, compared to the method of pairwise comparison that has significantly lower consistency. Additionally, risk averse subjects seem to respond more consistently with the two first methods, as opposed to the risk prone subjects, who respond better with the method of pairwise comparison. This evidence can be very helpful to the assignment of managers and decision analysts on projects and to the selection of the proper weighting method to apply to decision making problems.

REFERENCES

Abidin, M.Z., Rusli, R. and Shariff, A.M., 2016. Technique for Order Performance by Similarity to Ideal Solution (TOPSIS)-entropy Methodology for Inherent Safety Design Decision Making Tool. Procedia Engineering, [online] 148, pp. 1043–1050. http://dx.doi.org/10.1016/j.proeng. 2016.06.587.

Aliev, R.A., Pedrycz, W., Kreinovich, V. and Huseynov, O.H., 2016. The general theory of decisions. Information Sciences, [online] 327, pp. 125–148. http://dx.doi.org/10.1016/j.ins.2015.07.055.

Barzilai, J., 2010. Preference Function Modelling: The Mathematical Foundations of Decision Theory. In: M. Ehrgott, J.R. Figueira, και S. Greco, ed. 2010. Trends in Multiple Criteria Decision Analysis. [online] International Series in Operations Research and Management Science, Springer US. http://dx.doi.org/10.1007/978-1-4419-5904-1. Ch. 3

Beliakov, G. and James, S., 2015. Unifying approaches to consensus across different preference representations. Applied Soft Computing Journal, [online] 35, pp. 888–897. http://dx.doi.org/10.1016/j.asoc.2015.02.008.

Blagojevic, B., Srdjevic, B., Srdjevic, Z. and Zoranovic, T., 2016. Heuristic aggregation of individual judgments in AHP group decision making using simulated annealing algorithm. Information Sciences, [online] 330, pp. 260–273. http://dx.doi.org/10.1016/j.ins.2015.10.033.

Bottomley, P.A. and Doyle, J.R., 2001. A comparison of three weight elicitation methods: good, better, and best. Omega, [online] 29(6), pp. 553–560. http://dx.doi.org/10.1016/S0305-0483(01)00044-5.

Brand, M., Schiebener, J., Pertl, M.-T. and Delazer, M., 2014. Know the risk, take the win: How executive functions and probability processing influence advantageous decision making under risk conditions. Journal of Clinical and Experimental Neuropsychology, [online] 36(9), pp. 914–929. http://dx.doi.org/10.1080/13803395.2014.955783.

Charness, G., Gneezy, U. and Kuhn, M.A., 2013. Experimental methods: Extra-laboratory experiments-extending the reach of experimental economics. Journal of Economic Behavior & Organization, [online] 91, pp. 93–100. http://dx.doi.org/10.1016/j.jebo.2013.04.002.

Csermely, T. and Rabas, A., 2016. How to reveal people's preferences: Comparing time consistency and predictive power of multiple price list risk elicitation methods. J Risk Uncertain, [online] 53(53). http://dx.doi.org/10.1007/s11166-0-16-9247-6.

Gold, J.I. and Ding, L., 2013. How mechanisms of perceptual decision-making affect the psychometric function. Pro-gress in Neurobiology, [online] 103, pp. 98–114. http://dx.doi.org/10.1016/j.pneurobio.2012.05.008.

Hafezalkotob, A., Hafezalkotob, A. and Sayadi, M.K., 2016. Extension of MULTIMOORA method with interval numbers: An application in materials selection. Applied Mathematical Modelling, [online] 40(2), pp. 1372–1386. http://dx.doi.org/10.1016/j.apm.2015.07.019.

He, P., Veronesi, M. and Engel, S., 2017. Consistency of Risk Preference Measures: An Artefactual Field Experiment from Rural China. The Journal of Development Studies, [online] pp. 1–19. http://dx.doi.org/10.1080/00220388.2017.1336542.

Hogarth, R.M. and Einhorn, H.J., 1990. Venture Theory: A Model of Decision Weights. Management Science, 36(7), pp. 780–803. http://dx.doi.org/10.1287/mnsc.36.7.780.

Hogarth, R.M., 2010. Intuition: A Challenge for Psychological Research on Decision Making. Psychological Inquiry, [online] 21(4), pp. 338–353. http://dx.doi.org/10.1080/1047840X.2010.520260.

Holt, C. and Laury, S., 2002. Risk Aversion and Incentive Ef-fects. The American Economic Review, [online] (92), pp. 1644–1655. Available at: http://community.middleburh.edu/~jcarpent/EC499/Holt and Lauy (2002) AER.pdf [Accessed 25 Aug. 2017].

Hunter, D., 2017. Fitting a psychometric function. [online] Available at: http://davehunter.wp.st-andrews.ac.uk/2015/04/12/fitting-a-psychometric-function/ [Accessed 31 Aug. 2017].

Jern, A., Lucas, C.G. and Kemp, C., 2017. People learn other people's preferences through inverse decision-making. Cognition, [online] 168, pp. 46–64. http://dx.doi.org/10.1016/j.cognition.2017.06.017.

Jia, J., Fischer, G.W. and Dyer, J.S., 1998. Attribute weighting methods and decision quality in the presence of response error/: A simulation study. Journal of Behavioral Decision Making, 11(June 1997), pp.85–105. http://dx.doi.org/10.1002/(SICI)1099-0771(199806)11:2<85::AID-BDM282>3.0.CO;2-K.

Johansen, I.L. and Rausand, M., 2015. Ambiguity in risk assessment. Safety Science, [online] 80, pp. 243–251. http://dx.doi.org/10.1016/j.ssci.2015.07.028.

Lauriola, M. and Levin, I.P., 2001. Relating individual differ-ences in Attitude toward Ambiguity to risky choices. Jour-nal of Behavioral Decision Making, [online] 14(2), pp. 107–122. http://dx.doi.org/10.1002/bdm.368.

Naili, M., Boubetra, A., Tari, A., Bouguezza, Y. and Achroufene, A., 2015. Brain-inspired method for solving fuzzy multi-criteria decision making problems (BIFMCDM). Expert Systems with Applications, [online] 42(4), pp. 2173–2183. http://dx.doi.org/10.1016/j.eswa. 2014.07.047.

Panno, A., Lauriola, M. and Figner, B., 2013. Emotion regulation and risk taking: Predicting risky choice in deliberative decision making. Cognition & Emotion, [online] 27(2), pp. 326–334. http://dx.doi.org/10.1080/02699931.2012.707642.

Peng, J., Wang, J., Wu, X., Wang, J. and Chen, X., 2015. Multi-valued Neutrosophic Sets and Power Aggregation Operators with Their Applications in Multi-criteria Group Decision-making Problems. International Journal of Computational Intelligence Systems, [online] 8(2), pp. 345–363. http://dx.doi.org/10.1080/00207721.2016.1218975.

Podinovskaya, O.V. and Podinovski, V.V., 2017. Criteria importance theory for multicriterial decision making problems with a hierarchical structure. European Journal of Operational Research, [online] 258(3), pp. 983–992. http://dx.doi.org/10.1016/j.ejor.2016.09.038.

Retief, F., Morrison-Saunders, A., Geneletti, D. and Pope, J., 2013. Exploring the psychology of trade-off decision-making in environmental impact assessment. Impact Assessment and Project Appraisal, [online] 31(1), pp. 13–23. http://dx.doi.org/10.1080/14615517.2013.768007.

Riabacke, M., Danielson, M. and Ekenberg, L., 2012. State-of-the-Art Prescriptive Criteria Weight Elicitation. Advances in Decision Sciences, [online] 2012, pp. 1–24. http://dx.doi.org/10.1155/2012/276584.

Robert, F.S., 1985. Measurement Theory. 7th ed. [online] Ency-clopedia of mathematics and its applications. New York: Cambridge University Press. Available at: http://fitelson.org/roberts_measurement_theory.pdf.

Roth, B. and Voskort, A., 2014. Stereotypes and false consensus: How financial professionals predict risk preferences. Journal of Economic Behavior & Organization, [online] 107, pp. 553–565. http://dx.doi.org/10.1016/j.jebo.2014.05.006.

Schneider, M.A. and Nunez, M.A., 2015. A simple mean-dispersion model of ambiguity attitudes. Journal of Math-ematical Economics, [online] 58, pp. 25–31. http://dx.doi.org/10.1016/j.mateco.2015.03.002. Bell, J. & Holroyd, J. 2009. *Review of human reliability assessment methods*. London: Health and Safety Executive.

Shao, J., Taisch, M. and Ortega-Mier, M., 2016. A grey-DEcision-MAking Trial and Evaluation Laboratory (DEMATEL) analysis on the barriers between environmentally friendly products and consumers: practitioners' viewpoints on the European automobile industry. Journal of Cleaner Production, [online] 112, pp. 3185–3194. http://dx.doi.org/10.1016/j.jclepro.2015.10.113.

Thomas, P.J., 2016. Measuring risk-aversion: The challenge. Measurement, [online] 79, pp. 285–301. http://dx.doi.org/10.1016/j.measurement.2015.07.056.

van Winsen, F., de Mey, Y., Lauwers, L., Van Passel, S., Vancauteren, M. and Wauters, E., 2016. Determinants of risk behaviour: effects of perceived risks and risk attitude on farmer's adoption of risk management strategies. Journal of Risk Research, [online] 19(1), pp. 56–78. http://dx.doi.org/10.1080/13669877.2014.940597.

Wang, J., Wang, J. and Zhang, H., 2016. A likelihood-based TODIM approach based on multi-hesitant fuzzy linguistic information for evaluation in logistics outsourcing. Computers & Industrial Engineering, [online] 99, pp. 287–299. http://dx.doi.org/10.1016/j.cie.2016.07.023.

Weller, J.A., Levin, I.P. and Bechara, A., 2010. Do individual differences in Iowa Gambling Task performance predict adaptive decision making for risky gains and losses? Journal of Clinical and Experimental Neuropsychology, [online] 32(2), pp. 141–150. http://dx.doi.org/10.1080/13803390902881926.

Xu, Y., Chen, L., Li, K.W. and Wang, H., 2015. A chi-square method for priority derivation in group decision making with incomplete reciprocal preference relations. Information Sciences, [online] 306, pp. 166–179. http://dx.doi.org/10.1016/j.ins.2015.02.018.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A multi-discipline method to assess the human performance in manufacturing industry for safety and quality optimization

Lorenzo Comberti & Micaela Demichela
DISAT, Politecnico di Torino, Italy

Maria Chiara Leva
School of Food Science and Environmental Health, Dublin Institute of Technology (DIT), Dublin, Ireland

ABSTRACT: Nowadays the majority of organizations operating in manufacturing field recognize the importance of including the Human Factor contribution in the industrial process optimization (Hong et al. 2007). Technical measures and work organization procedures have been optimized in order to reduce the defects and waste generation but the Human Performance prediction still represents for Managers a difficult task to deal with.The prediction of the human performances of all workers involved in a production system would help Managers in better allocating the human resources. In order to reach this objective, a model to quantify the human capability of managing a complex task in a working context characterized by a set of physical, organizational and cognitive factors was designed.This paper presents the preliminary results of a three years industry/academia partnership project to assess the human performance in manufacturing plant. A multi-discipline approach involving both technical and individual factors was adopted.

1 INTRODUCTION

In manufacturing sector, the process optimization plays an important role to improve the production efficiency and economical profits.

Production is influenced by several factors such as: technology, organization, energy and workers performance.

In many cases, process optimization has been primary focused on technical measures and work organization procedures.

The Human Factor, despite the level of automatization in manufacturing industry is considerably increased and the standardization of working-procedures drives the working activity, still plays an important role on the efficiency of production system (Baines et al., 2005).

Human Factor has a strong influence on the occupational accident occurrence and defects generation.

Human Factor represents for Managers a difficult task to deal with even if most of organizations operating in manufacturing field recognize the importance of including the Human Factor (HF) contribution in the industrial process optimization (Hong 2007).

The Human Factor analysis has been approached differently in several areas.

Safety and Quality managers focused their attention to the deviation of human behavior from procedures. Miller (1987) analyzed a set of environmental, organizational and individual factors in relation to error-related outcomes. The ex-post events analysis approach has been used (Comberti et al., 2015) to identify causes of occupational accidents and defects with the aim of reducing their repetition.

Work Organization managers related the HF analysis to the ergonomic with the aim of calibrating and optimizing the task-time and reducing the operative risk task-related (Lin et al., 2001).

Many studies on Human Performance modeling suggest that the HF has to be approached as a complex system, where behavior, cognition, physiology and working condition deeply interact (Leva, 2016).

The knowledge of the relation between the human nature and the working condition of all workers involved in a production system would be crucial for the industrial Management.

Better allocating the human resources forward the different tasks will probably reduce the defects generation and the unsafe actions frequency.

In order to reach this objective it is necessary to model a system able to quantify and predict the human capability of managing a complex task in a context characterized by a set of physical, organizational and cognitive factors (Groth, 2012) in other words a model able to define and assess the Human Performance (HP).

Relevant researchs in this topic suggested which variables can be used to define the HP.

Baine & Benedettini (2007) suggested a multidisciplinary approach based on Sociology, Phycology and Engineering disciplines to be consistent on human nature representation. Eklund (1997) showed that Ergonomic has to be related to quality performances.

This paper presents the preliminary results of an industrial and academic project to develop a Human Performance assessment method for safety and quality optimization.

The aim of this work is to approach the HP modeling to facilitate the management of HF into the industrial improvement process.

The proposed model was developed on the basis of Straeter (2000) results.

It is based on the fundamental assumption that the HP can be represented as directly dependent from two macro-factors:

- Task Complexity (TC): that summarizes all factors contributing to the physical and mental requests to execute a given operative task, including work environmental factor.
- Human Capability (HC): that resumes the resources of workers under the real working condition. This factor represents both physical, mental and cognitive ability of the worker.

Section 2 of this paper presents the Conceptual Model of this project meanwhile section 3 shows the Operative Model deducted from the case study.

Section 3 gives an illustration of the project future development with a focus on the model validation. Conclusions will end the paper.

2 HP PROJECT DESIGN

This project has been managed in 4 steps as Figure 1 shows.

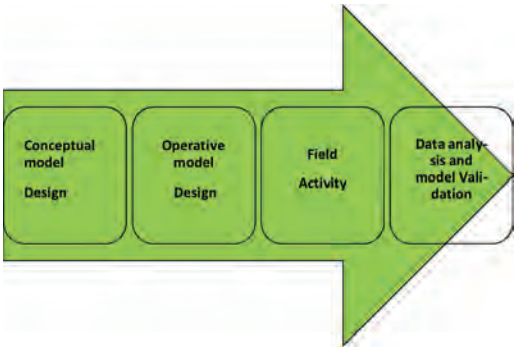


Figure 1. Project structure.

First step was focused on the "Conceptual Model" designing process.

Conceptual Model defines the variables and relations considered to the HP assessment.

Second step was characterized by the Operative model-design that represents the projection of Conceptual model into the industrial real life.

In other words each variables introduced into the Conceptual model have to be replaced by a measurable quantity into the Operative model.

Third step will be focused on HC and TC assessment with an intensive data field collection. This step will involves directly the workers of the plant with skill tests performed during the working activity.

In addition to this the descriptive parameters of TC will be collected with a deep analysis of working places.This step will be completed by a systematic interview of all workers involved. The interview will be structured on a set of questions related to individual motivation, risk-perception, working complexity perception.The information acquired with the survey will be used as a feedback for safety, work organization and quality improvements. On the basis of the results a validation or modification of the model will be done. This paper presents results related to the first and second steps of the project.

2.1 *Conceptual model*

The HP model represents the interaction between two macro factors: the Human Capability (HC) and the Task Complexity (TC).

Both factors can be analyzed with a wealth of methods for different purposes, such as data collection, task analysis (including cognitive task analysis), workload measurement, assessing situation awareness performance assessment (including team performance assessment), human error identification and interface evaluation methods (Stanton, 2004 and 2006).

In this work the proposed conceptual model of Human Performance is showed in Figure 2.

TC, as mentioned in the previous section, represents the total demand of resources asked to perform correctly a given task under certain work environmental condition.

TC is the result of the contribution of two main factors: Mental Workload (MW) and Physical Workload (PW), both associated to a single operative task.

PW factor is easily relatable to the physical, motion and postural efforts required to complete a given task.

Bad ergonomics combined with time pressure (coping with pace) have been estimated to cause about 50% of all quality deviations (Lin, 2001).

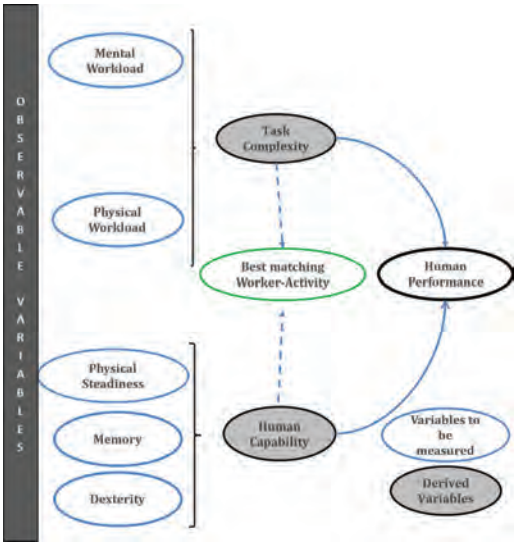


Figure 2. HP conceptual model.

Several studies demonstrates that high physical workload such as unkind postures can decrease the performance for discomfort (Erdinç, 2011).

A low variation, such as repetitive motions and static workload, was observed as additional cause for muscle fatigue (Punnett, 2000).

Other factors that may be included on PW modelling can be identified in the degree of rotation between high and low demanding tasks (Horton, 2012) and into the gender effects for the differences concerning discomfort and muscle fatigue in repetitive and static workload (Hunter, 2012).

In addition to the above mentioned factors, that are related to a specific operative task, other variables able to affect the PW are represented by environmental workload effects (Jung, 2001) which include: improper temperature, lighting, noise, vibration and exposure to chemical agents and physical agents as dust.

The physiological effects of these environmental factors, under industrial conditions, can contribute to an increase of the stress level and consequently to a loss of human performance (Grandejan, 1985).

MW was defined by Kahneman (1979) as "a factor directly related to the proportion of the mental capacity of an operator spends on task performance".

The MW assessment has been conducted in various research fields with both objectives and subjective measures such as: physiological activity under simple task normative condition (Kramer, 1991), cognitive performances, subjective analysis (Didomenico, 2008) and combined approach (Miyake, 2001).

All these studies have been performed in normative condition, with simple standardized tasks and under controlled environmental condition. This configuration is far away from industrial situation.

A relation between MW of assembly tasks and quality deviations was recently founded by Falck (2014). This work suggests that MW can be estimated trough the evaluation of the complexity of the task.

Operating in an industry plant it would be more suitable assessing the MW factor with a combination of subjective measurement and indirect task-related variable quantification, instead of approaching it with physiological measurement and cognitive normative test.

As a results of literature review and plant analysis a set of variables to TC definition was identified.

Figure 3 summarizes all variables selected to TC definition.

Human Capability (HC), as mentioned in the previous section, represents the total amount of resources that a worker is able to give for execute a given task under environmental working condition.

The HC factor is given by the contribution of several human skills that are all engaged in performing an operative task.

In particular the main Human skills that have been considered in relation to an assembly task are:

- Ability: skills like Precision, Manual Handling, Coordination are solicited continuously during a front line assembly work.
- Memory: remembering the sequence of operations and parts to complete correctly a given task can differ considerably.
- Physical: the ability of maintaining a constant performance during the shift and the ability of coping with pace.

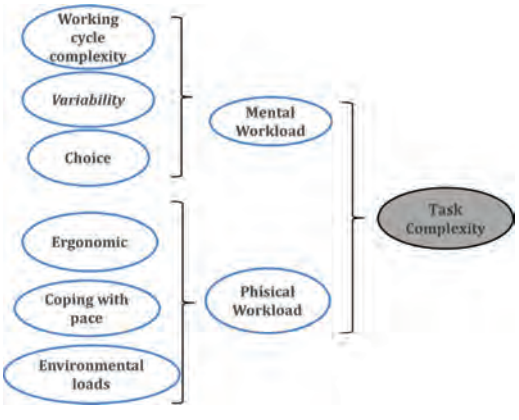


Figure 3. TC conceptual model.

2.2 *Operative model*

The Conceptual model defined in the previous section represent also the trace for the Operative model definition.

Operative model contains for each factor taken into account into the Conceptual model a set of observable and measurable variables.

The variables were selected after a field analysis performed in the beginning stages of the project with a participatory approach that involved both academic and industry professionals operating in the various management areas involved: Safety, Work Analysis, Quality, Work Organization.

The observable variables selected will be measured both in numerical and qualitative scales.

In order to allow the confrontation between variables with different nature and scale, all the variables will be harmonized in a common numerical scale.

TC factor will be estimated trough the assessment of observable variables that are showed in Figure 4 (Mental Work Load) and in Figure 5 (Task Complexity).

In the proposed representation of Figures 4 and 5 some variables are not used directly into the HP model but are compared with the results of the workers interviews previously mentioned.

HC factors will be estimated trough a set of measures, showed in Figure 6.

These measures will be obtained as a results of skill tests performed by workers during the real working activity.

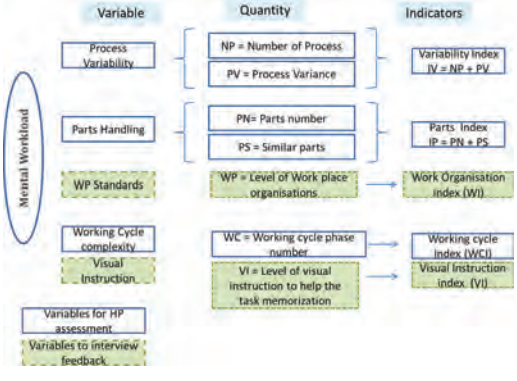


Figure 4. MW operative model.

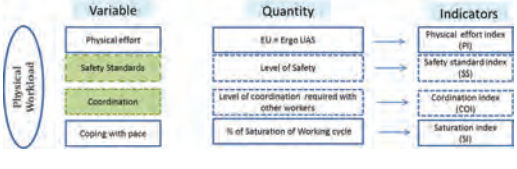


Figure 5. PW conceptual model.

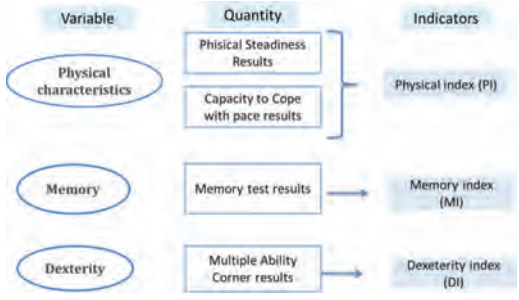


Figure 6. HC Operative model.

As an example the "Memory skill" will be tested recording the time spent by a worker to replicate a symbol sequence shortly showed.

Dexterity variable will be measured with 3 "ability tests" that simulate some typical operation asked into an assembly line.

The HC of each single workers will be assessed recording the time spent to complete all tests and recording the number of errors done.

In addition to these human skills, to model the Human Capability, it must be noticed that an important psycological aspects that can be described as "Motivation" can interact constructively or disprutively with this factor.

Motivation includes several psychological factors:

- Perception of task-risk;
- Perception of task complexity;
- Personal awarness;
- Job satisfaction\dissatisfaction

It is conceptually easy to consider that a severe mismatch between Task complexity and Perceived Task Complexity can facilitate the human error or unsafe act generation.

Information acquired with the interview will be used to estimate the level of motivation and perception of each workers.

3 PROJECT DEVELOPMENT AND MODEL VALIDATION

Nowadays the project ended the second step. On the basis of the Operative model in the next 6 months a field data collection will be done. This activity will involves 150 workers operating in 4 assembly lines.

The total number of working places can be approximately estimated in 70 units. The application of this model will imply the calculation of 170 Human Capability profiles and 70 Task Complexity profiles. The HP calculation will be done according the scheme showed in Figure 7.

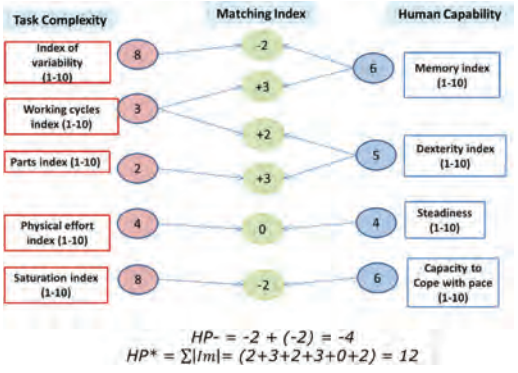


Figure 7. HP assessment.

Figure 7 summarize the generic scheme of calculation of HP between the TC of a working place characterized by 5 index (Variability, Working Cycles, Parts, Physical Efforts, Saturation) and the HC of a worker characterized by 4 index (Memory, Dexterity, Steadiness, Coping with pace).

This scheme of calculation is based on the operation "HC-index – TC-index" and leads to the definition of 6 matching indexs.

On the basis of the matching-index two Human Performance index are defined:

- HP-: represents the sum of all negatives matching index.
- HP*: represents the sum of all absolute values of matching index.

The assessment for each assembly line of HP— and HP* will allows a quantitative calculation of the potential Human Performance related to the matching workers-working places.

Changing the distribution of the workers will leads to a different HP estimation.

Minimizing HP-and HP* will implies the optimization of the distribution of the workers forward the working places on the basis of each individual human capability and each task complexity.

To validate this model a collaborative processes involving Quality and Production Managers in mutual learning processes will be adopted as suggested by Action Research method (Greenwood, 2006).

A set of 25 working places with relevant problems of quality will be selected.

HP results will be used to identify the best matching workers-working places and on the basis of that a new configuration of the line will be done.

A period of 3 months will be used to monitor the results of the new configuration workers-tasks and quality indicators will be collected.

The comparison of quality data ante and post configuration will allows the evaluation of the impact of the method.

This operation would leads to a reduction of human error related to a wrong matching worker-working place.

The HP assessment would be used as a sort of objective guideline to optimize the workers distribution into assembly lines.

The number of workers involved (more than 150) and working places analyzes will allows a statistical validation of the model.

4 CONCLUSION

Conclusively, the strengths of the proposed empirical approach with respect to the Human Performance assessment can be summarized as it follows:

- a model to HP definition as the ultimate product of the balance between the TC (driven by all the factors from the environment) and the operator characteristics (HC) was developed.
- The empirical based analysis will enhance the knowledge of the specific process operations at Managerial level, possibly highlighting latent drivers of Human Performance.
- This model was developed and will be tested in real operative condition and with a large number of workers directly involved. That represents a rare case of cooperation between academia and manufacturing. Results of the model application will be directly applied by industrial management with a measurable impact in term of process optimization.

REFERENCES

Baine T.S., Benedettini, O., 2007 Modelling human performance within manufacturing systems design: from a theoretical towards a practical framework, *Journal of Simulation, Volume* 1, pp 121–1306. National Center for Biotechnology Information.

Baines, T.S., Asch, R., Hadfield, L., Mason, J.P., Fletcher, S., Kay, J.M., 2005, Towards a theoretical framework for human performance modelling within manufacturing systems design. *Simulation Modelling Practice and Theory,* Vol. 13(6), pp. 451–524.

Comberti L., Baldissone, G., Bosca, S., Demichela, M., Murè, S., Petruni, A., Djapan, M., Cencetti, S. Comparison of two methodologies for occupational accidents pre-cursors data collection, 2015, *European Safety and Reliability Conference, ESREL* 2015, At Zurich, Switzerland.

DiDomenico, A., Nussbaum, M.A., 2008. "Interactive effects of physical and mental workload on subjective workload assessment. *International Journal of Industrial Ergonomics* 28, 977–983.

Eklund, J. 1997. "Ergonomics, Quality and Continuous Improvement conceptual and Empirical Relationships in an Industrial Context." *Ergonomics* 40: 982–1001.

Erdinç, O., Yeow. P.H.P. 2011. "Proving External Validity of Ergonomics and Quality Relationship through Review of Real-World Case Studies." *International Journal of Production Research* 49: 949–962.

Falck, A.C., R. Örtengren, and M. Rosenqvist. 2014. "Assembly Failures and Action Cost in Relation to Complexity Level and Assembly Ergonomics in Manual Assembly (Part 2)." *International Journal of Industrial Ergonomics* 44: 455–459.

Grandejan, E., 1985. *Fitting the Task to the Man—An Ergonomic Approach.* Taylor and Francis, London.

Greenwood, D.J., & Levin, M. (2006). Introduction to action research: Social research for social change. SAGE publications.

Groth K.M., Mosleh, A., 2012. A data-informed PIF hierarchy for model—based Human Reliability Analysis, *Reliability Engineering and System Safety* 108, pp.154–174.

Hong K., Nagaraja R., P. Iovenitti, M. Dunn, A sociotechnical Approach to Achieve Zero Defect Manufacturing of Complex Manual Assembly, 2007, *Human Factor and Ergonomics in Manufacturing,* Vol. 17 (2), pp. 137–148.

Horton, L.M., Nussbaum M.A., Agnew M.J., 2012. "Effects of Rotation Frequency and Task Order on Localised Muscle Fatigue and Performance during Repetitive Static Shoulder Exertions." *Ergonomics* 55: 1205–1217.

Hunter, S.K. 2014. "Sex Differences in Human Fatigability: Mechanisms and Insight to Physiological Responses." *Acta Physiologica* 210: 768–789.

Jung, H.S.,Jung HS. 2001."Establishment of overall workload assessment technique for various tasks and workplaces". *International Journal of Industrial Ergonomics* 28: 341–353.

Kahneman, D., Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, 47, pp. 263–291.

Kramer, A.F., 1991. Physiological metrics of mental workload: a review of recent progress. In: Damos, D.L. Ed., *Multiple task performance.* Taylor & Francis, London, pp. 279–328.

Leva, M.C., Ciarrapica Alunni, C., Demichela M., Allemandi, G., Adressing human performance in automotive industry: identifying main drivers of human reliability. *Irish Ergonomics Review 2016 Proceedings of the Irish Ergonomics society Annual Conference.*

Lin, L., C. g. Drury, and S.-W. Kim. 2001. "Ergonomics and Quality in Paced Assembly Lines." *Human Factors and Ergonomics in Manufacturing* 11: 377–382.

Miller, D.P., Swain, A.D., 1987, *Human error and human reliability.* Handbook Human Factor, Wiley-Interscience, New York.

Miyake, S., 2001. Multivariate workload evaluation combining physiological and subjective measures. *International Journal of Psychophysiology* 40, 233–238.

Punnett, L., L.J. Fine, W.M. Keyserling, g. D. Herrin, and D.B. Chaffin. 2000. "Shoulder Disorders and Postural Stress in Automobile Assembly Work." *Scandinavian Journal of Work, Environment & Health* 26: 283–291.

Stanton, N.A., (2004). Hierarchical Task Analysis: Developments, Applications and Extensions. *Applied Ergonomics,* 37, 55–79.

Stanton, N.A., et al., 2005. *Human factors methods: A practical guide for engineering and design.* Aldershot, UK: Ashgate.

Straeter O. 2000. Evaluation of Human Reliability on the basis of Operational Experience. GRS-170 Cologne (Germany) GRS.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Symptom-based approach for dynamic HRA and accident management

G.I. Petkov
Independent Consultant, Sofia, Bulgaria

ABSTRACT: The paper presents symptom-based approach for dynamic Human Reliability Assessment (HRA) and accident management through holistic context evaluation by the Performance Evaluation of Teamwork (PET) method. The macroscopic context evaluation procedure of the PET method gives opportunity for correct definitions of emerging issues, challenges, and possible solutions in the field of HRA for errors of commission, dependency analysis, multi-unit considerations, long time window scenarios, digitalized main control room, etc. In addition, the measuring the durations for recognition and disregard of symptoms, depending on various Performance Shaping Factors (PSFs), by utilization of well-developed stimulus-response models on the microscopic level, and extended use of simulator data could improve the quality of HRA, accident analysis and Probabilistic Safety Analysis (PSA) respectively.

1 INTRODUCTION

Human reliability assessment (HRA) is an applied science on the frontier between Probabilistic Safety Analysis (PSA), reliability theory, Deterministic Safety Analysis (DSA), complex system simulation, Human Factors (HF), Cognitive Systems Engineering (CSE), psychology and neuroscience. It is trying to apply what is known from sciences to reflect human interferences within Socio-Technical System (STS)[1] assess probability of Human Failure Events (HFEs), design and construct fault-tolerant and resilient system interactions between human, technology, organizations and environment.

With the increase in reliability of equipment in technical systems, the human who designs, manages and maintains them becomes a critical element. His/her reliability has to be explored and increased to match that of the machine. HRA has been emerging in the 60 s as a product of the development of the complex technical systems and technologies in the military, space and nuclear industry.

Human reliability is an irreplaceable element for the efficient and safe use of STS. The understanding and evaluation of safety is related to probabilistic risk assessment (PRA). Risk is characterized and includes at least two quantities:

- *magnitude (severity)* of the possible adverse consequence(s), and
- *probability of occurrence* of each consequence.

The development of nuclear technology is a priori related to ensuring its safety. Therefore, in the first systematic and comprehensive risk study, WASH-1400 (1975); HRA was included as an indisputable and important element.

There, HRA used the Technique for error-rate prediction (THERP) method (Swain & Guttman, 1983), developed by Alan Swain, and based on existing statistics and HF knowledge of that time, theoretical assumptions, and expert judgment. This method gives the *"thesis"* or the aim of "first generation" HRA methods—to obtain a Human Error Probability (HEP) of identified HFE by guessing, reasoning and weighting of internal and external factors influencing HEP. Thus, the scenario's severity is taking into account by expert judgment through multiplication of performance shaping factors (PSFs).

In every human action (HA), one may distinguish two sequential stages—Cognitive (Diagnosis or Decision-making) stage and Executive (Response or Manual) stage. For the second stage some reliable statistics does exist (e.g. THERP) or can be obtained. However, for the first cognitive stage refuted and recanted holographic-like[2] models are not available. This is due to the "unfinished business related to" HRA, which includes identification, specification and fitting of human cognition model to define the error potential and context of human action" (Dougherty, 1993). Hollnagel (1993) also emphasized that HFEs take

[1]STS consists of human, organization, technology and environment.

[2]Holographic-like behavior of STS is a system without any communication among separated processes (Townsend, 1984).

place in a cognitive system. He also argued that the detailed knowledge of HA objective context and its subjective image which exists in human mind is a basis for understanding of human performance: *"... any description of human actions must recognize that they occur in context" and in dynamics "on a second-by-second basis" and must take into account multidimensional dynamic context and "how the context influences actions" and a whole cognitive process* (Hollnagel, 1998).

When considering human performance reliability in accident conditions, the first dimension is time. This *temporal approach* used in THERP, e.g. Swain's Time Reliability Curves or its 'improved' version called the Human Cognitive Reliability (HCR) Correlation (Hannaman & Spurgin, 1984) is "virtually impervious to context". It should be complemented by *influential* or *contextual approach* to avoid "bareness in modeling" (Dougherty, 1993). He came up with the idea the change the "first generation" included HRA models such as THERP and HCR with the "second generation" HRA models (Dougherty, 1990). The *"antithesis"* or the aim of "second generation" HRA models was to take into account HA context with its specifics, severity, multidimensional dynamics and holistics for any individual, group mental or manual process.

However, the most of "first generation" HRA methods made a formalistic substitution of "influential factors", the THERP's holistic PSFs or their modifications, with contextual factors and it does not substantially alter HRA's outcomes. Moreover, difficulties arising from significant uncertainties in the quantification of each factor for HEP due to lack of data and an appropriate method to address uncertainties and dependencies between PSFs. This substitution exemplifies the Dougherty's observation and insight of that "the influential and contextual approaches may find themselves indistinguishable at the quantification stage because of the paucity of actual data" (Dougherty, 1993).

Since then the HRA community is discussing and substantiating some contentions and challenges of HRA, which caused much debate and launched the so-called "second generation" HRA. Some of HRA basics are already indisputable; others are the result of inertia in group thinking, misleading interpretation, judgment biases, business dependences and interests.

The following list shows HRA features that are valuable[3] for the accident analysis and whether they are implemented or not in the discussed second generation HRA methods (the features are in italic):

- *The response-related model is necessary*—not implemented in *Méthode d'Evaluation de la Réalisation des Missions Operateur* (MERMOS);
- *"…any description of human actions must recognize that they occur in context"*—implemented in *Cognitive Reliability and Error Analysis Method* (CREAM);
- *'extremely contextual'*—implemented in MERMOS;
- *situational and knowledge/mental models*—implemented in *A Technique for Human Event Analysis* (ATHEANA);
- *shift the problem from quantification of the operator behavior to the quantification of the error-forcing context*—implemented in CREAM, MERMOS, ATHEANA;
- *'error' identification*—implemented in CREAM, MERMOS, ATHEANA;
- *'error' quantification*—implemented in CREAM, MERMOS, ATHEANA;
- *'error' reduction*—implemented in CREAM;
- *accident context is a function of time "on a second-by-second basis"*—not implemented in CREAM;
- *individual dynamics of situation*—not implemented in MERMOS;
- *variability of performance is more important than how actions can fail*—not implemented in CREAM;
- *some combination between performance-related effects*—implemented in MERMOS;
- *context as a collection of weighted PSFs*—implemented in SLIM by domain expert judgments;
- *human performance limiting values*—implemented in MERMOS;
- *context could be different for each crew member*—not implemented in MERMOS;
- *systematic search process by series of lists of possible Human Failure Events, Unsafe Acts and error-forcing conditions (EFCs)*—implemented in ATHEANA;
- *identification of erroneous actions based on the accident* context and not to "subsume the errors identified with the event sequence mostly by historical means"—implemented in ATHEANA.

To overcome the above mentioned problems with subjective judgment of PSFs, holistic and dynamic human performance in STS, HFE identification, quantification, reduction, data-mining and measuring, a more realistic *symptom-based approach* for the statistical description of the STS context was proposed (Petkov and Furuta, 1998). The STS context quantification procedure is a first step of a *Performance Evaluation of Teamwork (PET)* HRA method that can implement the above features valuable for both HRA and accident analysis It relies on combinations of recognizable

[3] This means that the HRA method can assess the severity of the context, its potential for human error and thereby reduce the risk or optimize the action of an operator.

symptoms[4] for description of the variability of STS performance and gives controllable macro structures of main mental processes as cognition and communication (Petkov, 2000).

And finally, the HRA *"synthesis"*: to obtain a HEP of identified HFE in the STS context based on its holistic specific symptoms' recognition, severity & dynamics for any individual, group mental/manual response. A symptom's impact is reasoned, measured and weighted by internal and external "glocal" *(global & local)* PSFs for it.

2 SYSTEM APPROACH TO HUMAN PERFORMANCE

The HRA experience showed that often the approach to the study of human actions and failures is unilateral and biased, for example from the point of view of psychologists and engineers.

For example, the engineering modeling of human reliability, in similar way as equipment reliability, encounter grounded arguments by psychologists that it does not account for its activity and variability, and that man is cognitive, emotional, volitional, and not an item. On the other hand, the concerns of some of psychologists are unfounded that man cannot be simplified as one of the system components or cannot be evaluated as a number, i.e. neuroscience or HRA should not apply the universal ways for modeling and sharing scientific knowledge.

There are two approaches to human errors – "the person approach and the system approach." The first approach "focuses on the contribution of human errors on the system, their own psychological justification, accusations of forgetfulness, inattention, or moral weakness" (Reason, 2000). For example, "Over time, the role and importance of human contributions has grown for early PRAs, the contribution of humans has been set at about 15%, now the contribution has grown to 60% to 80%" (Spurgin, 2009).

The second approach focuses on the conditions, situations and context in which a person deliberately and conscientiously performs his/her actions to effectively manage the system and limit the consequences of the risk of its operation. The extreme statement is that "Human error is never the root cause."

Human performance needs to be considered as a variability of a whole STS where human interacts with technology, other humans, organizations and environment. The system approach is the more preferable and practical for context-based HRA. It is not important "who blundered, but how and why the defences failed" (Reason, 2000).

[4] A symptom is a measurable plant parameter that is available to the operator in the control room (IAEA, 2006).

3 SYMPTOM-BASED APPROACH TO STS STATISTICAL DESCRIPTION

The system approach means more components, more states, more interactions and information that have to be taken into account. How to analyze the system performance is the most important issue.

The enormous amounts of information and the use of digital technology, for information processing, have been changed individual (human) and group (social) mental processes and people behavior. The use of structural or functional decomposition, customary to PSA and CSE modeling, in HRA does not work because they break the most important dynamic interactions in the system and cannot explain its holistic behavior. The real-world decision-making processes need to be modeled in the context of the complex multi-agent and multi-level socio-technical system. The 'context' concept is the crucial in order to really make a difference in such multi-disciplinary models to control and coordinate the balance between enormous information and limited human cognitive capacities.

3.1 *Symptom-based approach in nuclear accident management*

The symptom-based approach is now usual for nuclear accident management. The IAEA NS-R-2 (2000) establishes the following requirements on accident management: "The training of operating personnel shall ensure their familiarity with the symptoms of accidents beyond the design basis and with the procedures for accident management." Later in IAEA SRS No. 48 (2006) "symptom/state based procedures" were justified and in IAEA NS-G-2.15 (2009) a "symptom-based approach" was also recommended: "2.14. The approach in accident management should be based on directly measurable plant parameters or parameters derived from these by simple calculations."

3.2 *Statistical entropy description of socio-technical system*

The STS dynamic interactions are manifested by interference of symptoms (*stimuli with meaning for human*) and the system context could be presented by them on the macro level. Of course, if we want to understand the root causes of human errors, we should searched in depth at micro level. But the macroscopic statistical description of the STS context would help to identify the dynamic and holistic nature of the system's behavior (Petkov et al., 2001).

The basic idea of the distinction between macro- and microscopic levels is to change the set of microscopic accessible states with equivalent

subsets of macroscopic states. A certain macroscopic state can be found in many microscopic accessible states. This idea follows the Shannon theorem (1948) regarding the entropy as the measure of information, and was the basis for the used, in the PET method, *an analogy of energy and information*. The mental process in the STS is described at each moment by its *microstate*. This is a specific *quantum state* that represents the most detailed possible STS description.

3.3 *Practicality of symptom-based STS description*

As the context characterization and analysis in the HRA is implemented with the assistance of the personnel, then it is more practical and natural, firstly, to describe STS context in symptoms used by personnel. Recognizing each symptom by operator is a mental process involving individual cognition, communication between the group of operators, decision-making, checking and recovery. Each symptom has its specific symptom-influencing factors (SIFs) that essentially coincide with the used holistic PSFs in HRA. But if all the symptoms into a given scenario are described with a common PSF then it leads to blur the specific influence of a PSF on a given symptom, masks the dependencies between the PSFs and increases the uncertainty of the results.

3.4 *Theoretical issues and limitations from unexplored mental processes*

In addition to the above, there are other deeper theoretical reasons for using the symptom-based approach, these are the unexplored mental processes.

Townsend (1984) emphasizes the need for a theory for connection of *holographic-like behavior* and separation of selective and non-selective influence:

- "*Systems based entirely on holographic-like behavior* without any communication among separated processes, are omitted."
- "The selective influence postulate is critical." The "empirical tests of selective influence will likely be tied closely to separation of selective from non-selective systems."
- *"Factors that influence* not only durations but also *outputs of processes* have not been investigated."
- "*The expectation* (the mean) *of a sum of random variables* (in this case, serial processing times) *is equal to the sum of the expectations, which is true for any set of random variables whether or not they are independent*."

If systems with holographic-like behavior are omitted from exploration of mental processes then the holistic PSFs-based approach for STS context quantification is very questionable! In symptom-based approach every symptom is recognized separately and context quantification is based on enumeration of STS accessible states.

Duration of symptom recognition is measured as a degree to which an operator's perception of the STS is compatible with the required action.

A model to establish a mathematical function that describes the relation f between the symptom x and the expected duration of the recognition ΔT is necessary: $\Delta T = f(t)$. A common simplification assumed for such functions is linear, thus we expect to see a relationship like:

$$\Delta \boldsymbol{T}_i = \Delta \boldsymbol{T}_{i,min} + \boldsymbol{t}.\boldsymbol{\beta}_{ij} \quad (1)$$

where β_{ij} is j-factor, $SIF_j\ (PSF_j)$, for i-symptom.

Since Townsend & Schweickert (1989) also emphasized "Although we view the test for additivity as one important strategy in an overall systematic approach to uncovering psychological processes, *we do not identify processes with additivity*," then this means that the influence of n factors could not have been considered simple by the expression as $\beta_i = \Sigma\beta_{ij}$, where $j = 1...n$. However, $\Delta T_{i,max}$ could be measured for given human performance context (STS accessible state), and the uncertainty of the results for HEP could also be reduced.

4 PET EXTENDED DEFINITIONS, HUMAN FAILURE EVENTS TAXONOMY, EVALUATION PROCEDURE AND MODELS

4.1 *Extended definition of dynamic STS context*

Petkov & Furuta (1998) have proposed a heuristic concept of Context Factors and Conditions (CFCs) as quantitative factors to indicate *"how symptoms (CFCs) influence context"* in order to study their effects on human performance—*"how context influences actions"* during the accident. As well as they consider the context to be a function of time and second argument of the HEP function, while first argument of the HEP is time, i.e. $HEP = f(t, context)$.

Petkov & Groudev (1999) proposed an indirect similarity between material ('transition temperature shifts') and mental processes ('human performance shifts') to measure dynamic deviations of symptoms (CFCs). The common ground of this study is the emergency context. The similarity is seen between the "reference" temperature for the pressurized thermal shock in materials and the "reference" context probability for the HFE. The latter is regarded as generalized evaluation of the context influence on HA. The basic assumption of

this study is that the stress both of materials and human is due to symptom deviations or parameter gradients (between reality and expected values) which determine their characteristics and behavior.

The interactions or *"STS performance shifts"* represent the dynamic STS context or processes in human-technology-organization-environment. Matching the object in situation is an approximation that could be identified and described as an *"image"* or as a *"signature"* of the object in the situation. The concept *"image"* is the imprint of the consciousness and sub-consciousness of the *"human performance shifts"* that affect the person's physical, physiological, psychological and psychosocial ability to make sense, perceive the *object (STS)* and to perform action in the STS context.

Generally, context is defined by psychologists as "a state of mind' or 'a set of internal or mental representations and operations rather than a set of external elements". In (Petkov, 2004) the context definition was extended as *"a common state of universe, mind and situation in their relation"—object-image-situation*. In (Petkov, 1999), the term *"context probability" (CP)* was coined as a measure for severity magnitude of error-forcing context. But this general definition of context is not so practical to be a measure. The practical definition of context is: *a statistical measure of the degree of the STS state randomness defined by the number of accessible states taking place in the STSs' ensemble*.

Since the man analyses and operates the machine most frequently by discrete actions, the combination (number, value and tendency) of symptoms (stimuli with meanings), which she/he manages to distinguish, is of greatest importance to him. As a measure for symptoms' influence the relative deviation of symptom image is proposed: $\Delta\phi/\phi_o = |\phi_o - \phi_s|/\phi_o$, where the indices denote the two types of values under interest (o – objective; s – subjective; ϕ–image).

4.2 *Extended definition of STS violated image of symptom*

Petkov & Petkov (2000) presented the Combinatorial Context Model (CCM) for context quantification of cognitive process based on deviations of the symptom image from object in situation. The CCM uses the following symptoms for context quantification: *event (E), parameter (P), action (A), resource (R), function (F), transition (T) & goal (G)*.

In order to demonstrate the crucial impact of violations (term "circumventions" is used in USA) on the post-accident context, a *Violation of Objective Kerbs* (VOK) method was also proposed there to account for the probability of aberrant circumstances (prior to or during the initiating event) in cognitive process. Following the Reason's qualitative definitions (1994) the quantitative definitions were proposed for the context quantification purposes:

Errors is probable when the differences between objective & subjective images of context symptom is not zero, $\phi_{sn}(t) \neq \phi_{on}(t)$, where zero-context is $|\phi_{on}(t) - \phi_{sn}(t)| \rightarrow 0/\min$.

Violation of Image of Symptom (VIS) occurs when the objective image of context symptom ϕ_n is changed from $\phi^1_{on}(t)$ to $\phi^2_{on}(t)$ due to any reason.

It should be emphasized that the quantitative definitions of *error* and *violation* thus formulated are dealt with not only in human performance context, but in the context of the whole STS. Therefore, they have not only a wider, but also more different meaning than these of Reason. For example, Reason (2000) defines "procedural violations" as an active failures that "have a direct and usually short-lived impact on the integrity of defenses."

In Reason (2000), the strategic decisions are "made by designers, builders, procedure writers, and top level management" refer to latent conditions that "can create longlasting holes or weaknesses in the defenses" and "may lie dormant within the system for many years before they combine with active failures and local triggers to create an accident opportunity."

In the PET method, the VIS is considered as a strategic decision and introduces of "inevitable 'resident pathogens' within the system," resulting in a resonant increase in severity of error-forcing context, i.e. the CP & HEP.

Taking into account the relativity of time when the strategic decision was made, seconds or years ago, we can conclude that the difference of the PET VIS is that it may be in apparent or latent conditions, but these conditions must lead to a sharp/resonant increase in the *severity of context, i.e. CP(t)*.

If we take into account the consequences of the PET VIS, it can be said that they are similar to those of *Error of Commission (EOC)*. However, such consequences can occur not only after actions but also in disturbing the objective image of other symptoms, i.e. EOCs are subset of the set of VISs.

For example, some VIS≠EOC during the Fukushima Daiichi NPP #1 accident are the following: VIS-E: *SBO – loss of AC power*; VIS-T: *Loss of DC power*; VIS-F: *Containment Failure*. These VIS aren't EOC, because the violated symptoms are E, T & F, but they dramatically increased the accident severity & C(t).

For example, some VIS≡EOC during Chernobyl NPP #4 accident are the following: VIS-A: *System for Emergency cooling system has turned off from the previous shift;* VIS-A: *The scram is disabled when two turbo-generators are turned off;* VIS-A: *Several manually operated rods are pulled out to increase power, they remain less than 8 of 15 required rods in the reactor core;* VIS-A: *The opera-*

tional reserve of reactivity reaches a value requiring the reactor shutdown, but it is not done on time.

4.3 *PET HFE taxonomy*

The explained concepts, approaches and definition extensions used by PET method are presented on the HFE taxonomy shown on Figure 1.

4.4 *Recursion of context and recognition*

Context probability (CP) is determined by the existing deviations of symptom images in human cognition and team communication processes.

Matching the object in situation is approximation of a subjective image to the objective one by comparison of certain symptoms (signals, symbols and signs) that could be identified and described as an image of the object in the situation. These symptoms are Event, Parameter, Action, Resource, Function, Transition, Goal and VIS.

Symptom recognition and cognitive context are iterative & recursive functions. In order to calculate CP, the durations of recognition for any symptom (as CFC) & cognitive disregard durations of VIS are needed. At the next step of iteration of cognitive process, we may use the new duration of symptom recognition based on previous calculation.

Duration of symptom recognition could be based on measurement or expert judgment about the type of the recognized symptom (skill-based, rule-based, and knowledge-based). The times for completion of cognitive process for 'skill-based', 'rule-based' & 'knowledge-based' symptoms would be in correlation 1:5:30 (Petkov, 2017).

A symptom (CFC), after its recognition, could be kept or removed from the operator's context model and his memory depending on the situation. Some unimportant symptoms could be disregarded after their recognition by analogy with hypothesis that Rodin (1987) labeled 'cognitive disregard'.

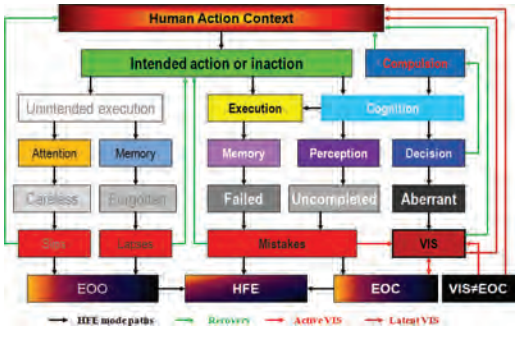


Figure 1. Human action context and HFE taxonomy.

The PET procedure for evaluation of context, cognition, communication and decision-making probabilities consists of eight steps (Petkov, 2018).

5 PET DIAGRAMS OF HRA PROCESS IN PSA

5.1 *PET scenario's timeline description, symptoms and violations of symptom's image qualification*

The PET scenario's timeline description of symptoms and violations of image of symptom qualification is presented on Figure 2. The following abbreviations are used: SAMG (Severe Accident Management Guideline); EOP (Emergency Operating Procedure); OED (Operational Experience Data); PIE (Postulated Initiating Event); TH (Thermo-Hydraulic Simulation); CS (Computer Simulator); FSS (Full Scope Simulator); normP, abnormP (Normal/Abnormal Procedure); OP/MP (Operational/Maintenance Procedure); PSA; DSA; HRA; VIS; S (Symptom/Stimulus): P, R, E, T, G, HA and F.

5.2 *PET screening, holistic and atomistic quantitative assessment and integration in PSA*

The PET screening, holistic & atomistic quantification and integration in PSA are presented on Figure 3. The following abbreviations are used: CP; EP (Error Probability); HEP; HA (Human Action/Task); HFE; RI (Risk Importance).

5.3 *Holistic dynamic profiles of CPs and EPs*

The PET diagram for obtaining holistic dynamic profiles of context and error probabilities is

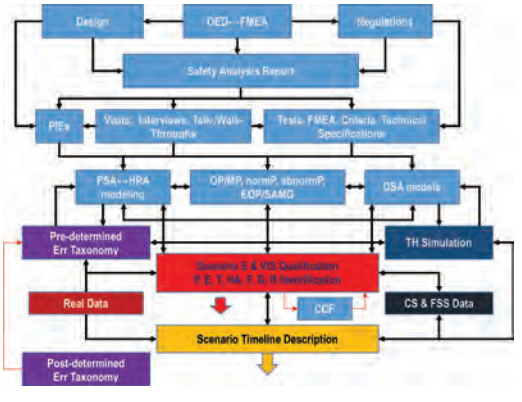


Figure 2. PET scenario's timeline description, symptoms and violations of symptom's image qualification.

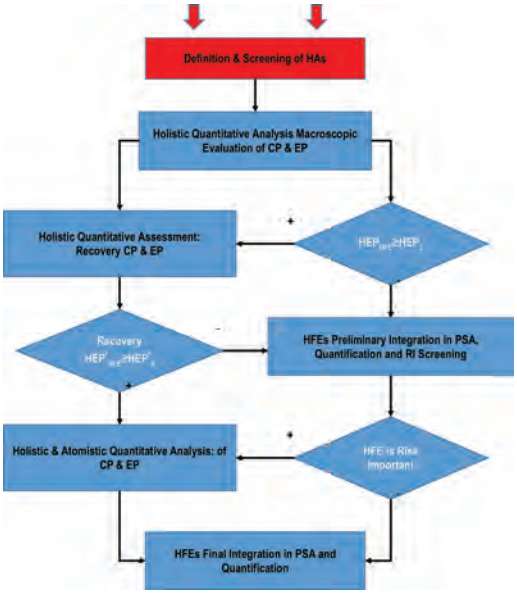


Figure 3. PET screening, holistic and atomistic quantitative assessment and integration in PSA.

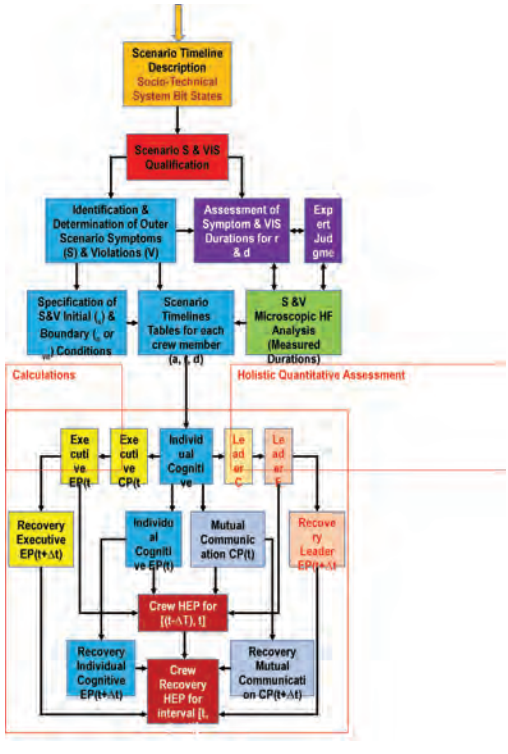


Figure 4. PET diagram for obtaining holistic dynamic profiles of context and error probabilities.

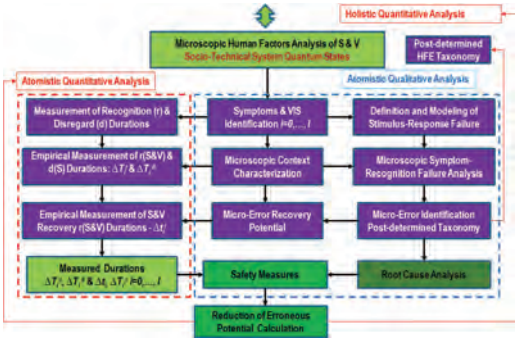


Figure 5. PET diagram for obtaining atomistic potential reduction and root causes.

presented on Figure 4. The following abbreviations are used: CP; EP (Error Probability); HEP; HA (Human Action or Task); HFE; a (appearing); r (recognition); d (disregard); ry (recovery); o (objective); S (subjective); vo (violated o).

5.4 *Atomistic error potential reduction and causes*

The PET diagram for obtaining atomistic potential reduction and root causes is presented on Figure 5.

6 HRA ADVANCED PRACTICE BY PET METHOD

HFEs are unexpected events in STS leading to unwanted outcomes. Therefore, the HEP has been changing in time before, during and after any HFE, and severity of STS context should be dynamic variable of error-producing potential. In nuclear accident conditions this dynamic and holistic measure could be defined as context probability, where for the most severity context—the probability is 1.

HRA methods try to predict a HEP in the "prevailing" context that means in a statistical average context of an average crew performance but only for some short time interval "prevailing" could exist during the accident. A static value (anchor) of HEP based on a judged average context of crew for identified task is calculated. The HEP is adjusted by multiplication of guessed values of PSFs taking into account the variability of all system components. Fuzzy logic of each HRA technique, tabulated and justified by its database is used to introduce PSFs into the HEP variation. Usually, a cited database is 'know-how' of the HRA method. It is verified and validated by the owner and concerned national regulator. But the structures and parameters of models and obtained data are not accessible for other users in order to check them and to repeat data-mining.

The benchmarking of the HRA methods is based on results for similar identified tasks, but not on models and experimental data.

Implicit, static and pseudo-holistic determination of context based on an anchor HEP and fuzzy PSFs values judged by expert, makes HRA methods superficial and insensitive to the STS models (structures and parameters), HFE symptoms and causes, and human performance processes.

The main reason for the HRA insensitiveness is the lack of models and data for a holographic-like behavior of the human interactions in a complex situation and multifactorial context. These models are substituted with expert judgments and multiplication of concurrent PSFs considered for specific task. This subjective way of HEP evaluation does not allow a systematic and multi-layered study of human and STS performance. Practical PET applications for retrospective and prospective HRA and accident analyses are shown in (Petkov, 2017).

7 CONTRIBUTIONS TO EXISTING HRA

The PET method allows dynamic evaluation of the individual CP and communication context probability (CCP) over time as gauge parameters.

Based on the dynamic individual CP and CCP, HEP and its constituents can be evaluated—cognitive, executive, recovery and decision-making by a group of people or organizations.

Average HEP that is usually evaluated by ther HRA methods is easy calculated as a mean over the time interval during which the human action is most likely to be performed.

The PET method provides an opportunity to generate HRA database based on simulations using thermo-hydraulic codes, computer-based, engineering or full-scope simulators.

It also allows to evaluate and optimize the individual and group performance of the main control room operators during real events or training.

The perspective to go in-depth and to move from a macroscopic to a microscopic context evaluation with the PET method would allow not only to scan and monitor the human error potential, but also to assess the importance of each symptom and PSFs influencing its recognition.

8 CONCLUSIONS

The PET, as a HRA method, applies a realistic procedure for dynamic symptom-based context evaluation of cognition and communication, and context-sensitive digraph models of cognition, communication and decision-making.

The intermediate use of symptoms for dynamic context evaluation gives better opportunities for the systematic identification, qualitative and quantitative interpretation of time-dependent HFEs during the accident and for improving of emergency response planning and/or severe accident management.

The PSA modeller is responsible for appropriate determination of the initiating event progression and needed actions based on accident reports, thermal-hydraulic simulations or full-scope simulator training exercises.

The detailed and qualitative data of accident reports are the best source for validation and verification of HRA methods.

The data use of thermal-hydraulic calculations & full-scope simulator are a valuable option for optimization of time to implement emergency measures and actions to ensure safety during the DBA and DEC/BDBA based on joint deterministic and probabilistic criteria.

The qualitative description and analysis of the accident scenario should be improved in order to improve the quantitative assessment of the context and the HFE probability. If important elements of description are missed then distortions occur in the evaluation. The HRA modeller is responsible for correct definition of symptoms, violated images of symptoms and determination of the durations of their manifestation, recognition and disregard, action and recovery implementation. It is preferably to measure them but guess or judgment could be acceptable as a first approximation.

The correct distribution of roles in modelling, limitation of expert guesses and possibility for experimental verification & validation supposes that PET could be applied with higher degree of confidence.

Insufficient exchange of information between designers, DSA, PSA and technologists may lead to violated context and increased risk.

The PET is a prospective emerging HRA method for dynamic, context-based, retrospective and prospective analysis and data-mining that could provide the PSA studies with HEPs for any specific action/task based on state-of-art simulations in order to avoid expert guesses.

REFERENCES

Dougherty, E.M., 1990, "HRA—Where Shouldst Thou Turn?" Reliability Engineering and System Safety, 29(3), 283–299.

Dougherty, Ed, 1993, *Context and Human Reliability Analysis*, Reliability Engineering and System Safety 41 (1993) 25–47.

Hannaman, G. W. and Spurgin, A. J., 1984, *"Systematic human action reliability procedure (SHARP),"* EPRI NP-3583, EPRI, Palo Alto.

Hollnagel, E., 1993, "*Human Reliability Analysis: Context and Control*," Elsevier Science Ltd., London.

Hollnagel, E., 1998, "*Cognitive Reliability and Error Analysis Method—CREAM*," Elsevier Science Ltd., London.

IAEA NS-G-2.15, 2009, Severe Accident Management Programs for Nuclear Power Plants, IAEA Safety Standards Series No. NS-G-2.15, IAEA, Vienna.

IAEA NS-R-2, 2000, Safety of NPPs: Operation, IAEA Safety Standards Series No. NS-R-2, IAEA, Vienna.

IAEA Safety Reports Series No. 48, 2006, Development and Review of Plant Specific Emergency Operating Procedures, IAEA, Vienna.

Petkov, G. & Petkov, S., 2000, Towards Defense-in-Depth of Human-Machine Interaction, Proceedings of the PSAM-V Conference, Osaka, Japan, Universal Academy Press, Inc., ISBN 4-946443-64-9, 2000.

Petkov, G. and Groudev, P., 1999, *Correlation between Human and Material Shocks In Symptom-Based Emergency Procedures*, ASME-PVP Conference, ASME-PVP, Vol.392, pp. 25–35, 1999, Boston, USA.

Petkov, G., 2004, 'Dealing with Dynamic Aspects of Operators' Performance,' Proceedings of the FLINS'2004 Conference, Duinse Polders, Belgium, September 1–3, 2004), World Scientific, ISBN 981-238-873-7, pp.677–683.

Petkov, G., 2017, Team performance comparison in core-melt units of Fukushima Daiichi NPS based on dynamic context quantification of accident, PSAM Topical Conference 2017, Munich, Germany, June 7–9, 2017, GRS.

Petkov, G., 2018, Enhancing time-dependent criteria for design-basis and design extension conditions based on human performance context evaluation in ATWS events, *ASME J of Nuclear Rad Sci,* (2018); doi: 10.1115/1.4039000.

Petkov, G.I., 1999, Networked Risk Images in Human Action Context, CT'99 Conference, (San Francisco, CA, USA, August 11–14, 1999) http://www.cogtech.org/CT99.

Petkov G., 2000, Identification and Fitting of Human Cognition Model, Proceedings of the ASME-PVP Conference, ASME-PVP, Vol. 400, pp. 35–48, Seattle, USA.

Petkov G., Antao P. and Guedes Soares C., 2001, 'Context Quantification of Individual Performance in Accidents,' Proceedings of ESREL'2001, Vol. 3, Torino, Italy, 16–20 September 2001, pp.1851–1858.

Petkov G., Furuta K., 1998, Application of PN-Based Method for Identification and Classification of Human Actions in NPP TLRS, Proceedings of PSAM4, Vol.2, pp.1136–1141, New York City, USA.

Reason J., 2000, Human error: models and management. *BMJ: British Medical Journal*. 2000; 320(7237): 768–770.

Spurgin, A. J., 2009. Human Reliability Assessment Theory and Practice. CRC Press.

Swain, A. D. and Guttman, H. E., 1983, Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications (NUREG/CR-1278), Washington, DC, NRC.

Townsend, J.T. & Schweickert, R., 1989, Toward the Trichotomy Method of Reaction Times: Laying the Foundation of Stochastic Mental Networks, Journal of Mathematical Psychology, 33, 309–327.

Townsend, J.T., 1984, *Uncovering Mental Processes with Factorial Experiments*, Journal of Mathematical Psychology, 28, 363–400.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Data learning and expert judgment in a bayesian belief network for offshore decommissioning risk assessment

M.L. Fam
Lloyd's Register Global Technology Centre, Singapore
Nanyang Technological University, Singapore
Royal Institute of Technology, Stockholm, Sweden

X.H. He
Lloyd's Register Global Technology Centre, Singapore

P. Hilber
Royal Institute of Technology, Stockholm, Sweden

L.S. Ong
Nanyang Technological University, Singapore

D. Konovessis
Singapore Institute of Technology, Singapore

H.K. Tan
Lloyd's Register Global Technology Centre, Singapore

ABSTRACT: Decommissioning of offshore facilities involve changing risk profiles at different decommissioning phases. Bayesian Belief networks (BBNs) are used as part of the proposed risk assessment method to capture the multiple interactions of a decommissioning activity. The Bayesian Belief network is structured from the data learning of an accident database and a modification of the BBN nodes to incorporate human factors and barrier performance modelling. The analysis covers one case study of one area of decommissioning operations by extrapolating well workover data to well plugging and abandonment. Initial analysis from well workover data, of a 5-node BBN provided insights on two different levels of severity of an accident, the "Accident" and "Incident" level, and on its respective profiles of the initiating events and the investigation-reported human causes. The initial results demonstrate that the data learnt from the database can be used to structure the BBN, and give insights on how human factors pertaining to well activities can be modelled, and that the relative frequencies can act as initial data input for the proposed nodes. It is also proposed that the integrated treatment of various sources of information (database and expert judgement) through a BBN model can support the risk assessment of a dynamic situation such as offshore decommissioning.

1 INTRODUCTION

Decommissioning of offshore facilities takes place in different phases which can include the warm suspension phase, the cold suspension phases and the removal phase. Some examples of the warm suspension phase activities are pipeline decontamination and sectional removal or well plugging and abandonment. Well plugging and abandonment involves multiple tasks at the same time, and the assessment of location specific risks such as reservoir profile, and presence of gas deposits from its drilling history. Location-specific historical information can be used to model risks more specifically. Bayesian Belief Network (BBN) is defined as an acyclic graphical network that is capable of representing qualitative and quantitative relationships between factors of interest defined in it. The nature of the relationships can be defined by whether there exists a (i) causal (ii) functional or (iii) statistical relationship. The BBN model structure can be derived by experts' judgement or statistically, or through a combination of both. The proposed model in this paper is structured from the data learning of an accident database and the BBN nodes are modified to incorporate human

factors and barrier performance modelling. The analysis considers one case study of well workover data extrapolated to well plugging and abandonment activities. The information fed into the BBN consists of relative frequencies from data learning from the database and expert-elicited where there are data gaps. The desired output of the model is a probabilistic risk profile of the potential consequences of a particular decommissioning activity, and one that has considered and can trace different sources of information—generic historical data, location-specific information and expert-judgement, instead of only historical data in most existing methods.

2 PROPOSED STATISTICAL MODEL FOR STRUCTURE LEARNING

2.1 *Data types*

The most comprehensive offshore safety database is the World Offshore Accident Database (WOAD) (DNV GL, 2017) that has been collecting data since 1976 by the Norwegian company DNV GL. The WOAD database provides principle information such as accident causes and its chain of consecutive events (for e.g. a dropped object resulting in fire and oil spill), location of accident facility, year of accident etc. The WOAD database has also been built on merging information from existing databases such as the Offshore Blowout Database (SINTEF, Norway), MAIB accident database (UK Marine Accident Investigation Branch) and the COIN/ORION database (UK HSE—offshore Safety Division).

The data in WOAD appears as a discrete, 'count' nature, each time there is an accident reported, it is updated. Also, some factors have order in its states, such as the severity of events, where "Accident" is more severe than a "Near Miss".

2.2 *Statistical model—generalised linear models*

The proposed model should be able to accept the updating of information and would be able to present the risk picture as per the most updated snapshot. The Beta and Dirichlet density function allows the quantification of the prior data (belief) and updating new information (beliefs). For analysis where only binary variables are involved, such as in a Fault Tree/ Event Tree where there are only 'True' or 'False' status, or the location of vessels such as 'In the port' or 'At sea', the Beta density function can be used. The beta density function f with parameters a, b, $N = a + b$, where a and b are real numbers > 0, is (Neapolitan, 2004, p. 306):

$$\begin{aligned} p(f) &= \frac{\Gamma(N)}{\Gamma(a)\Gamma(b)} f^{a-1}(1-f)^{b-1} \\ &= beta(f;a,b),\ 0 \le f \le 1 \end{aligned} \tag{1}$$

With updated information, t and $M = s + t$, and considering d as the dataset, the updated beta density function, with respect to obtaining the probability of having dataset d can be written as (Neapolitan, 2004, p. 306):

$$\begin{aligned} P(d) &= \frac{f^s(1-f)^t\, p(f)}{E(F^s((1-F)^t)} \\ &= \frac{\Gamma(N+M)}{\Gamma(a+s)\Gamma(b+t)} f^{a+s-1}(1-f)^{b+t-1} \end{aligned} \tag{2}$$

For analysis of multinomial variables, the Dirichlet density function can be used to provide equal counts for all statuses to be studied. For example, the severity of a consequence can be 'Near miss', 'Injury/Accident' or 'Fatality'. The Beta density function is, in fact, a special case of the more generic Dirichlet density function. Similarly, considering d as the dataset, the probability of having the dataset d by assuming all statuses are Dirichlet-distributed is shown below (Neapolitan, 2004, p. 306,386):

$$\begin{aligned} P(d) &= E\left(\prod_{k=1}^{r} F_k^{S_k}\right) \\ &= \frac{\Gamma(N)}{\Gamma(N+M)} \prod_{k=1}^{r} \frac{\Gamma(a_k+s_k)}{\Gamma(a_k)} \end{aligned} \tag{3}$$

Offshore safety data is based mainly on response and explanatory variables of non-metric, discrete forms, such as operation sections on a platform (Drilling and/or Production area) or initiating events of accidents (Falling Load, Leak etc). A common response variable would refer to consequence status differing in levels of severity, and in this case, the order of the status is important, such as 'Fatality' having more severe considerations than a 'Near Miss'. A statistical model suitable for offshore safety data should be required to be able to handle multivariate data or non-metric format, and hence the classical regression models are unsuitable. The proposed statistical model able to handle ordered, categorical data is a log-linear model that falls under the classification of a Generalised Linear Model (GLM) (Agresti, 2002, p. 125). A GLM extend ordinary regression models to encompass non-normal response distributions and modelling functions of the mean and has three components (similar to classical regression models)

consisting of a random component (response variable), a systematic component (explanatory variables) and a link function that transforms the mean to the natural parameter. The categorical data is arranged in a table form with its frequency information, and the log-linear modelling involves fitting models to the cell count in the cross-tabulation of categorical variables to derive the association and interaction patterns among variables.

For a table of response variable y and explanatory variable x (or two categorical responses), in a table with row i and column j, the cell probabilities are π_{ij} and the expected frequencies are $\mu_{ij} = n\pi_{ij}$.

$$log\mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y \tag{4}$$

where λ_i^X refers to the row effect and λ_j^Y refers to the column effect.

Assuming there are interaction effects, then a saturated model with statistically dependent variables would be (Agresti, 2002, p. 316):

$$log\mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \tag{5}$$

where λ_{ij}^{XY} refers to deviations from independence.

When extrapolating the assessment from a two factor to a three factor contingency table, the following summarised log-linear models (see Table 1) illustrate the combination of potential interactions.

The modelling process begins with a saturated model, i.e. with the highest order interaction between the variables, before systematically and sequentially removing a higher-order interaction term so that model complexity is reduced without any significant loss in accuracy. The removal stops at a point where any removal leads to a poor fit of the data. The threshold of removal is considered by comparing the test model against the saturated model based on the difference of its deviances.

Table 1. Loglinear models for three-dimensional tables.

Loglinear Model	Symbol
$log\mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$	(X,Y,Z)
$log\mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$	(XY,Z)
$log\mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}$	(XY,YZ)
$log\mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ}$	(XY,YZ,XZ)
$log\mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ} + \lambda_{ijk}^{XYZ}$	(XYZ)

Deviance is the likelihood-ratio statistics for testing the null hypothesis that the simplified model would hold against the saturated model. For a particular GLM for observations $y = (y_1, \ldots, y_N)$, let $L(\mu; y)$ denote the log-likelihood function expressed in terms of the means $\mu = (\mu_1, \ldots, \mu_N)$. Let $L(\hat{\mu}; y)$ denote the maximum of the log-likelihood for the model, i.e. a saturated model, which can provide a baseline for comparison with other model fits. In this saturated model, $\hat{\mu} = y$, and the deviance, D, of a Poisson GLM is defined to be (Agresti, 2002, p. 119):

$$\begin{aligned} &D(y; \hat{\mu}) \\ &= -2log\frac{\max likelihood\ for\ model}{\max likelihood\ for\ saturated\ model} \\ &= -2\left[L(\hat{\mu}; y) - L(y; y) \right] \end{aligned} \tag{6}$$

Consider two models, M_0 with fitted values $\hat{\mu}_o$ and a saturated model M_1 with fitted values $\hat{\mu}_1$. Since M_0 has lesser interactions considered than M_1, a smaller set of parameter values satisfies M_0 as compared to M_1. Maximizing the log likelihood over a smaller space cannot yield a larger maximum, thus $\left[L(\hat{\mu}_o; y) \le L(\hat{\mu}_1; y) \right]$. Continuing from equation (6) (Agresti, 2002, p. 141), and assuming that model M_1 holds, the likelihood-ratio test of the hypothesis that M_0 holds uses the test statistic (Agresti, 2002, p. 141):

$$\begin{aligned} &D(y; \hat{\mu}_o) - D(y; \hat{\mu}_1) \\ &= -2\left[L(\hat{\mu}_o; y) - L(y; y) \right] - \{-2\left[L(\hat{\mu}_1; y) - L(y; y) \right] \\ &= -2\left[L(\hat{\mu}_o; y) - L(\hat{\mu}_1; y) \right] \end{aligned} \tag{7}$$

Under regularity conditions, this difference has approximately a chi-squared null distribution with degrees of freedom equal to the difference between the numbers of parameters in the two models in comparison. The difference of the models' deviances follows a χ^2 null distribution (Agresti, 2002, p. 142).

Before determining the factors with the loglinear model, it is proposed to use a two variables dependency test based on the Pearson χ^2 model in order to shortlist the factors for consideration of a 3-variable analysis. The Pearson's χ^2 test compares the frequencies observed in certain categories to the frequencies that might be expected in the same categories by chance, with respect to the degrees of freedom given by the (number of rows -1) multiplied by (number of columns -1)(Field, Miles, & Field, 2012, pp. 802–803).

Similar to regression, the residual is simply the error between what the model predicts (the

expected frequency) and the data actually observed (the observed frequency):

$$standardised\ residual = \frac{observed_{ij} - model_{ij}}{\sqrt{model_{ij}}} \quad (8)$$

Through determining statistical dependencies between factors, the nodes of the BBNs are then established, and from which the base structure of a BBN can be defined.

3 CASE STUDY—WELL WORKOVER TO WELL P&A

3.1 *Well workover*

Well workover data from the UKHSE database refers to operations in which a well is re-entered for any purpose, for example, maintenance related activities' of replacing retrievable downhole safety valves, or malfunctioned electrical submersible pumps or worn out tubings. Before any well workover, the well must be killed, which requires the removal of the well head, flowline, packers and lifting the tubing hanger from the casing head, before putting a column of heavy fluid into a well bore to prevent the flow of reservoir fluids without the need for pressure control equipment at the surface (which have been removed, a part of the well kill process). Such an operation usually requires a drilling rig to be involved.

While this data is strictly not well abandonment and plugging work required for a decommissioning process, well plugging and abandonment work includes removing casing and other downhole equipment, thus suggesting the similarities with wireline operation. The risks of well killing can also be comparable to a well plugging and abandonment procedure which involves cement being injected to plug the well, while in the well kill event; heavy fluid is injected into the well.

The factors in WOAD database, relating to Well Workover activities are: (i) Schedule quarter (ii) Human causes (iii) Initiating event (iv) Evacuation and (v) Severity of accident.

Thus the five factors from Table 2 will be analysed to understand its dependency relationships.

Table 2. Table of factors to be analysed in the Bayesian belief networks (DNV GL, 2017).

Factors	Description
Schedule Quarter	Each work schedule is divided into 12 hours, and the quarters are 3 hour session each.
Human Causes	In investigation reports that have concluded the cause of the incident to be human related, four major causes have been identified based on the incident reports: (i) unsafe act carried out without any procedures, (ii) procedures deemed to be unsafe (iii) improper design and (iv) unsafe act carried out against procedures.
Initiating Event	The initiating events were pre-defined in the WOAD database and are identified as follows: (i) falling load or dropped object (ii) release of fluid or gas (iii) well problem without blow out (iv) blow out (v) out of position (vi) fire
Evacuation	Indicates whether the emergency procedures of leaving the facility is (i) not required (ii) required and successful (iii) required and not successful
Severity of Accident	There are four levels of severity and are classified below: • Accident (A) has recorded fatalities and severe injuries. • Incident (I) refers to low degree of damage, but repairs/replacements are required or for events causing minor injuries to personnel or health injuries. • Near-Miss (N) refers to events that might have developed into an accidental situation. No damage and no repairs required. • Unsignificant (U), refers to events with minor consequences. No damage, no repairs required. Other inclusions are small spills of crude oil and chemicals, and very minor personnel injuries, i.e. "lost time incidents".

3.2 *Results of well workover data analysis from two-variable and three-variable dependency analysis*

Dependency analysis between two variables is performed for all possible combinations in order to generate an initial list of dependency relationships. The analysis is based on the Pearson's χ^2 test which was elaborated in Section 2.2. With the chi-squared value, and the corresponding number of degree of freedoms, the P-value can be obtained. The confidence interval then plays an important role on the significance test on the hypothesis of independence. The most commonly set confidence interval is at 95%, i.e. a P-value exceeding 0.05. For this data set (see Table 3), the confidence interval is set at 95%, and if the obtained P-value from the dependency analysis is less than 0.05, this implies a dependency between the two factors being investigated.

Table 3. Snapshot of 'cleaned' data to be processed in the software R for the dependency analysis between 5 factors: (i) Schedule quarter (ii) Human causes (iii) Initiating event (iv) Evacuation and (v) Severity of accident (not all factors are shown in below).

No.	Schedule_ qtr	Accident_ Category	Human_Cause
1	3	Incident	Unsafe act/No Procedure
2	4	Near_miss	Improper design
…	…	…	…
99	5	Incident	Unsafe Procedure
100	2	Incident	Unsafe Procedure

Within the five factors, a 2×2 frequency table and its ten combinations of the pair-wise comparison of the five factors, has been extrapolated from the data source, and the Pearson's χ^2 test ran on R to calculate the P-value. The results tabulated in Table 4 only reflect the results where the P-Value has been found to be less than 0.05 and indicates a "dependent' relationship. Other independent results have been omitted since they do not have an impact on how the BBN can be structured. The pairs of variables thus provide the background for the three variable independency analyses.

The modelling of a three variable dependency begins with a saturated model for (A,B,C) which consist the individual effects from each factor alone, and the secondary interactions AB, BC, AC as well as the highest order which is the effects of ABC. Backwards eliminated is initiated from the saturated model, by removing the highest order interaction. The model with the highest order removed, is compared with the saturated model by considering the differences in the 'likelihood', 'degrees of freedom', and in the 'P-value'. The 'P-value' demonstrates the difference between the models, and if the value < 0.001, it represents a highly significant 'P-value difference' and that removing the higher order interaction has made a significantly worse fit to the data (Field et al., 2012). In other words, the higher order interaction is a significant factor to ensuring a good fit of the model.

If the removal of the three-way interaction shows a 'P-value delta' greater than 0.001, then the combination of the two-way interaction is compared against the next highest saturated model—which would refer to the three-way interaction.

Based on the results from the two-way dependency analysis, the following three-way dependency analyses have been grouped (see Table 5).

From the analysis from R (R Core Team, 2017) and summarised in Table 6, it can be observed that the model considering the interactions (Human Causes: Initiating Event + Initiating Event: Evacuation)

Table 4. Dependency relationships between two variables.

Combinations	X^2 value	DoF*	P-Value
Accident Severity, Initiating Event	47.98	21	0.000692
Accident Severity, Evacuation	18.748	6	0.00461
Initiating Event, Human Causes	41.47	21	0.004903
Initiating Event, Evacuation	42.131	14	0.0001178
Imitating Event, Schedule Quarter	71.101	28	0.00001304
Human causes, Evacuation	14.332	6	0.02614

DoF* = Degrees of Freedom.

Table 5. Proposed three-way dependency analysis.

Results from two-way dependency analysis	Proposed three way dependency analysis
• Initiating Event—Human Causes • Initiating event—Evacuation • Initiating event—Schedule Quarter • Accident Severity—Initiating Event • Accident Severity—Evacuation	1. Initiating Event, Human Causes, Evacuation 2. Initiating Event, Human Causes, Schedule Quarter 3. Initiating Event, Evacuation, Schedule Quarter 4. Human Causes, Evacuation, Schedule Quarter 5. Initiating Events, Number of Chains of Events, Nature of Events

Table 6. Summary table of conditional independency analysis.

Model	DoF	Likelihood Ratio	Deviance (diff)	Diff in DoF	P-Value (diff)
Human_ Cause (H) * Initiating_ Event (IE)* Evacuation (E)	0	–	–	–	–
H:IE + H:E + IE:E	42	0.332754	0.332754	42	**1**
H:IE + H:E	56	24.68428	24.35152	14	0.0415
H:IE + IE:E	48	1.549424	1.216669	6	**0.9760**
H:E + IE:E	63	21.71695	21.38419	21	0.4357
H + IE + E	83	67.52035	67.18759	41	0.0060

has a P-value delta of 0.9760 which is much higher than the P-value delta of the remaining models. The dropping of the interaction (Human Causes: Evacuation) still produces a good-enough fit for the model based on the P-value delta difference of 0.9760 compared to the P-value difference of the saturated model (Human Causes: Initiating Event + Human Causes: Evacuation + Initiating Event: Evacuation. The interaction (Human Causes: Initiating Event + Initiating Event: Evacuation) thus suggests a conditionally independent relationship between Human Causes and Evacuation given the occurrence of the Initiating Event, which can be noted as **I**(Human Causes, Evacuation | Initiating Event).

Similar operations have been performed for the remaining three-way combinations numbered from 2 to 5 in Table 4. As an example, the second three-way combination in Table 4 identified as (Initiating Event – IE, Human Causes – H, Schedule Quarter – S) is also investigated in the same manner, in that it considers the saturated model of H*I*S which consists of the following interactions of (H + IE + S + H:IE + H:S + IE:S + H:IE:S), followed by comparing the performance of the (H:IE + H:S + IE:S) model with the higher order interaction term H:IE:S removed, and subsequently proceeding towards the simplest independent model of just H + IE + S. The best fitting model, not necessarily the most complex, is then chosen based on the P-value differences in a similar process as documented in Table 6, which referred to the three-way dependency analysis of the factors (Human Causes: Initiating Event + Initiating Event: Evacuation). Having investigated the relationships between the factors and in creating the nodes, the next step is to put together the node network (see Figure 1), and its corresponding joint probability distribution through Bayesian Belief Networks (BBNs).

In order to analyse the data in a BBN, the data obtained from the database needs to be transformed into probabilities and/or conditional probabilities. For any node, the probabilities can be represented by the Dirichlet Density Function, which can be looked at as a ratio of a status, against the total number of statuses for that event. If an event has a parent event, then the conditional probability distribution should be used. The resulting BBN can then represent the relative frequencies of the occurrence of events as in Figure 2.

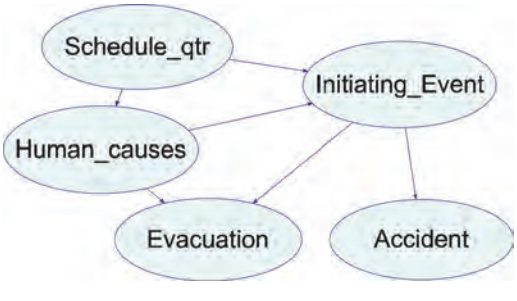


Figure 1. Snapshot of proposed BBN structure using GeNie (BayesFusion, 2018).

4 EXPERT ELICITATION TO BBN STRUCTURE AND CONDITIONAL PROBABILITY DATA

4.1 *Risk profile categorised by severity of accidents*

The BBN also allows for a sensitivity analysis to identify the types of initiating events associating with the 'Incident' level of severity of event (see Figure 3): which reflects that 'Falling Load/ Dropped Objects" make up the majority of the initiating event, with the "Release of fluids/gas" the next most common initiating event, and that the "Well problems leading to no blow outs" is the least common initiating event. It is also interesting to

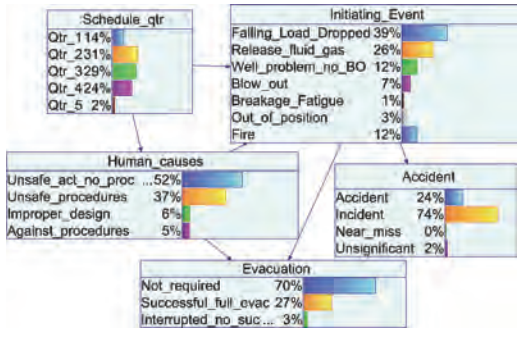


Figure 2. Snapshot of BBN with distribution of respective factors based on the data from WOAD.

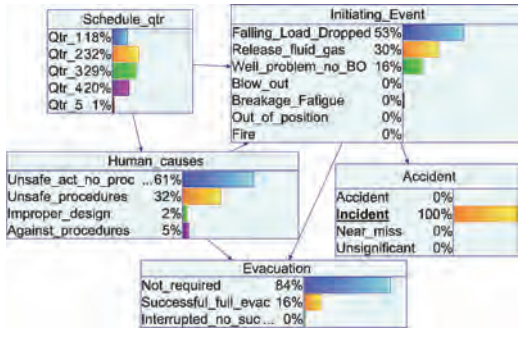


Figure 3. Risk profile when "Incident" in the Severity node is reflected as 100% (observed to have happened).

identify that unsafe act carried out stemming from the lack of procedures is the most common documented human cause in the investigation reports.

In terms of "Accident" level of the severity of incidents (see Figure 4), "Fire" as a triggering event makes up the majority of the initiating event. The next most common is a "Blow-out" followed by "out of position" and lastly the "Release of fluid or gas". It is also interesting to identify that unsafe procedures is the most common documented human cause in the investigation reports. It can be noted that the documented human cause between different levels of severities of an incident reflects the trends in the procedural level, at the "Incident" level, most of the time no procedures were in place, while in the "Accident level", procedures were in place but needed to be improved. In both cases, human action against procedures makes up a small percentage.

4.2 *Combining human factors, location specific risks and equipment reliability in the model*

The Petro-HRA guidelines published in 2017 (Taylor, 2017) adopted the SPAR-H methodology for assessing human error probabilities in the petroleum industry based on a nominal HEP that is adjusted by performance shaping factors. The diversity of the activities in a petroleum facility makes it difficult to list the performance shaping factors, or define what constitutes a "missing/misleading" human machine interface design, especially since decommissioning activities constitute a wide range from well P&A to structural removal. Strand et al (Strand & Lundteigen, 2016) further defined the crucial or "good" factors in a human-machine interface for drilling operations, such as having the ability to interpret wellbore flowrates, in situ pressures along the wellbores etc, and with each factor bearing different weights, indicating that some factors have a greater effect than

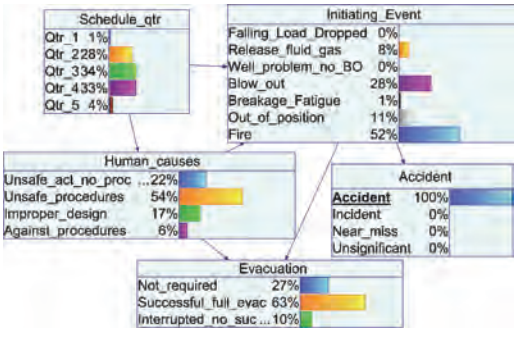


Figure 4. Risk profile when "Accident" in the Severity node is reflected as 100% (observed to have happened).

others. This exact ranking and its respective weight ratios proposed by Strand et al is also a reflection of an expert judgement as a source of information, where it is the other mechanism proposed in this paper; the first method being to obtain information from the database. It is proposed that the relationships learnt and the relative frequencies obtained from the database can be used to supplement the HRA model by Strand et al (Strand & Lundteigen, 2016).

The proposed BBN (Figure 5) consist of several parts: the nodes in light blue are information extracted from the data learning part as in Section 3.2. The "Severity of Accident" node has been expanded to include more consequences, including the state of safe operations, which is not originally in the accident database as it only recorded accidents. The green-shade of the node "Kick Frequency during drilling operations" represent learning from information as well, but not from an accident database, rather, from the record keeping of the operating history of the facility. The motivation behind such an information node is to incorporate location-specific risks. A plugging and abandonment activity will also be susceptible to kicks from the well, and often it is due to the reservoir profile of a well, hence in this case, location specific historical data would help to assess the risks more dynamically.

The relative frequencies for the "Human Cause" node has been removed and replaced with a "Procedures_PSF", in which PSF stands for Performance Shaping Factors, and the data learning part indicated that the relative frequencies for 'Not_ Available" or "Incomplete" would 0.52 and 0.37, and this could be the information for the initial conditional probability.

The next part of the BBN structure is indicated in yellow, which are safety barriers in relation to the identified, and shortlisted Initiating Event (a well kick), which are "Kick Detection" or the action to "Kill Well". These information can be obtained from the failure probability data from reliability databases, as in the approach for Bhandhari et al (Bhandari, Abbassi, Garaniya, & Khan, 2015) where "Kick Detection" failure probability was 0.01, and Well Kill operations had a failure probability of 0.25.

The third part of the BBN structure is indicated in pink, and refer to human factors that may influence how well a kick can be detected, or how the killing of a well can be conducted with success. The human factors were identified as specific to a well drilling operation by Strand et al (Strand & Lundteigen, 2016), such as how the presence of certain features would assign the Human Machine Interface PSF to be nominal or good. In Strand et al's human factor framework, they had based it on the

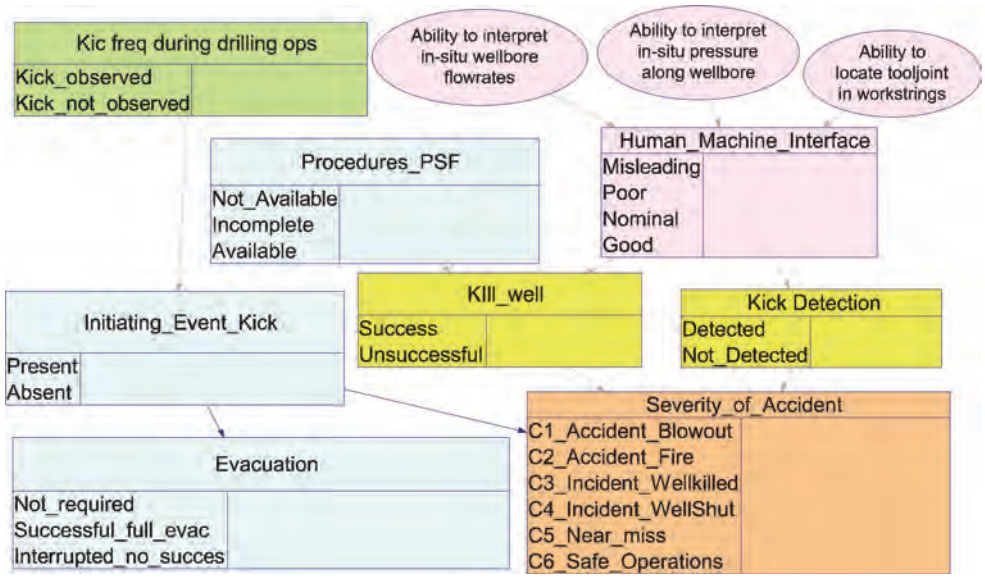


Figure 5. Proposed BBN with structure/information learned from database, and expert judgement input.

BORA-Release methodology (Sklet, Vinnem, & Aven, 2006) where weights have been assigned to each contributing item to a PSF. The weights in nodes leading to the HMI PSF (the BORA-Release methodology) are derived from expert judgement. The discussion of expert judgement input for the pink nodes, and arguably the data gaps for the orange node for the output node of interest could allow the risk to be computed from a combination of historical data (with updating mechanism, each time the database is updated), expert judgement in the human factors part of the model, or to fill in data gaps in other parts of the model.

4.3 *Expert judgment and elicitation of information in nodes of BBN*

A literature review conducted by Mkrtchyan et al (Mkrtchyan, Podofillini, & Dang, 2016) identified eleven methods of the assessment of conditional probability information, in which the methods vary in terms of whether a probability elicitation is direct or indirect, or whether it allowed multiple expert aggregation.

No expert judgment and elicitation have been carried out on the proposed BBN model yet, however, the method proposed should be one that is used to model human reliability as one of the nodes is a human factor node, and as pointed out by Podofillini et al (Mkrtchyan, Podofillini, & Dang, 2015), the conditional probability information building methodology needs to meet two criteria of allowing combination effects such as error-forcing effects where two factors in consideration are working on the same polarity, or the opposite end of the combination effect such as a compensation effect from a poor factor in combination with a good factor.

5 CONCLUSION

From the initial modelling of the BBNs derived from accident data, the results show potential in using industry data for accidents having an overall risk profile of the industry. The model provided insights on the "Accident" and "Incident" level of a severity of an accident, and its corresponding profile of the initiating events and the investigation-reported human causes. This forms the base structure of adapting the initial BBN model to include also barrier performance, such as the performance of kick detection, and potentially include human factors analysis for high operator involvement work, such as well plugging and abandonment and heavy lifting. This analysis also meets one aspect of the QRA framework requirements in demonstrating how the performance of barriers changes with different conditions. The model indicates the source of historical data and expert judgement, and is able to keep track of how data used in the model affects the results. Existing risk assessment methods are often based on generic failure statistics alone, which may not be reflective of an existing risk picture of a location. The model combining generic failure statistic, location-specific historical

data, and expert judgement of a particular work task could provide a more updated risk picture.

6 FUTURE WORK

The scope for future work is to expand the model to include the effects of expert judgment in the model, sensitivity analysis of how robust the model is, information learning from other databases or information sources, and subsequently expanding the analysis to other decommissioning activities.

REFERENCES

Agresti, A. (2002). *Categorical Data Analysis*. *Ophthalmology*. Hoboken, New Jersey: John Wiley & Sons.

BayesFusion. (2018). GeNie. http://www.bayesfusion.com/. Retrieved from http://www.bayesfusion.com/.

Bhandari, J., Abbassi, R., Garaniya, V., & Khan, F. (2015). Risk analysis of deepwater drilling operations using Bayesian network. *Journal of Loss Prevention in the Process Industries*, *38*, 11–23. https://doi.org/10.1016/j.jlp.2015.08.004.

DNV GL. (2017). World Offshore Accident Database. Retrieved from https://www.dnvgl.com/services/world-offshore-accident-database-woad-1747.

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics using R* (2nd ed.). Singapore: SAGE Publications Asia-Pacific Pte Ltd.

Mkrtchyan, L., Podofillini, L., & Dang, V.N. (2015). Bayesian belief networks for human reliability analysis: A review of applications and gaps. *Reliability Engineering & System Safety*, *139*, 1–16. https://doi.org/10.1016/j.ress.2015.02.006.

Mkrtchyan, L., Podofillini, L., & Dang, V.N. (2016). Methods for building Conditional Probability Tables of Bayesian Belief Networks from limited judgment: An evaluation for Human Reliability Application. *Reliability Engineering and System Safety*, *151*, 93–112. https://doi.org/10.1016/j.ress.2016.01.004.

Neapolitan, R.E. (2004). *Learning Bayesian networks*. Prentice Hall.

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.r-project.org/.

Sklet, S., Vinnem, J.E., & Aven, T. (2006). Barrier and operational risk analysis of hydrocarbon releases (BORA-Release). Part II: Results from a case study. *Journal of Hazardous Materials*, *137*(2), 692–708. https://doi.org/10.1016/j.jhazmat.2006.03.027.

Strand, G.-O., & Lundteigen, M.A. (2016). Human factors modelling in offshore drilling operations. *Journal of Loss Prevention in the Process Industries*, *43*, 654–667. https://doi.org/10.1016/j.jlp.2016.06.013.

Taylor, C. (2017). *The Petro-HRA Guideline*. Halden: Institute for Energy Technology.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Urban avalanche search and rescue operations in Longyearbyen: A study of public-private cooperation

S.M. Tengesdal
Civil Protection, County Governor of Sogn og Fjordane, Leikanger, Norway

B.I. Kruke
Centre for Risk Management and Societal Safety, University of Stavanger, Stavanger, Norway

ABSTRACT: The two urban avalanches in Longyearbyen, Svalbard, in 2015 and 2017 were unexpected and fast developing, with trying conditions and consequences both for people buried by the snow, and for people coming to their rescue. The fatality rate increases rapidly as the minutes pass if people are buried by the snow. Thus, successful rescue if buried by an avalanche depends on a prompt response. The aim of this paper is to discuss the role of the local population in urban avalanche search and rescue. Empirical data was collected from 8 months of fieldwork in Longyearbyen, including interviews with representatives of public and private organisations and the local population, as well as literature surveys on the population's involvement in acute crises, reports and newspaper articles. People are often present when crises strike. Many of them have the knowledge, training and equipment to be a valuable resource in avalanche search and rescue.

1 INTRODUCTION

19 December 2015 an avalanche hit urban areas in Longyearbyen, a small town situated on the Norwegian archipelago of Svalbard in the Arctic. The avalanche resulted in two fatalities and destroyed eleven houses below the mountain Sukkertoppen. Survivors and neighbours immediately commenced search and rescue activities. Over 150 unorganized volunteers stood shoulder to shoulder with the professional emergency responders in a joint effort to search for fellow citizens buried by the snow. Seven people were excavated and rescued (Indreiten and Svarstad 2016). Another urban avalanche hit Longyearbyen 21 February 2017, in almost the same area. This time there were no fatalities. However, the snow masses resulted in major damages to an apartment building. The people who lived there assisted each other in the evacuation. Once again, a large number of unorganized volunteers stood in line ready to come to the rescue of possible friends and neighbours buried by the snow. The local police, in close cooperation with the local Red Cross Society, quickly clarified that none were buried by the snow. Thus, further assistance from the unorganized volunteers was not required this time (Tengesdal 2017).

In this paper we aim to discuss the role of the population in urban avalanche search and rescue. We will use the term «population» synonymously with the term «unorganized volunteers». This includes the efforts of survivors and people who are randomly present or nearby the avalanche area. Organized volunteers, such as members of the Red Cross Society, are not included in these groups of volunteers.

We will start with a brief introduction to some important avalanche characteristics. Then we will present the conceptual frameworks, with an emphasis on crisis characteristics, crisis management and the population's behaviour and capacities in crises. After a brief methodological presentation we will present our empirical findings from the two urban avalanches. The findings will be discussed in relation to the conceptual frameworks before we conclude with some remarks on the population's contribution in search and rescue in urban avalanches.

2 AN INTRODUCTION TO AVALANCHES

In a historical perspective, the risk of avalanches was mainly a problem for people in settlements and along roads and railways located in avalanche prone terrain. This picture started to change in the 1970s, as a result of a growing interest in outdoor activities in steep terrain. Nowadays, the majority of the fatalities due to avalanches are related to winter recreation activities (McClung and Schaerer 2006). However, avalanches in countries like Norway, Iceland, USA, Switzerland, Austria

and France clearly demonstrate that we have to take into consideration that avalanches may represent a major risk also for people, buildings and infrastructure in urban areas.

To get caught in an avalanche is associated with great danger. The probability of survival are approximately 90 per cent if a victim, completely buried by the snow, is excavated within 15 minutes, unless the victim is not already dead due to trauma (killed by debris in the snow). After 15–45 minutes, the chances of survival decrease rapidly due to oxygen deficiency. After 45 minutes some 75 per cent of those excavated die, mostly due to hypothermia. The level of moisture in the snow and the victim's clothing are both crucial factors for survival after 90 minutes. A victim excavated at this point in time may die due to the reflow syndrome, a fatal condition caused by hypothermia and cold blood flowing from the extremities to the heart (Landrø 2007).

It is quite common to distinguish between two main types of avalanches; loose snow avalanches and slab avalanches. A loose snow avalanche is a chain reaction initiated at a specific point, and expands both in width and depth as it continues downwards (Landrø 2007). The initiating point in a loose snow avalanche consists of a small layer of surface snow that loosens, while it is a thin and weak layer of snow at a deeper level that fails in a slab avalanche. A slab avalanche is further characterized by a large and coherent mass of snow that loosens and splits into blocks as it moves downwards (McClung and Schaerer 2006). This type of avalanche reaches high speed in a short period of time, and the width can be several hundred meters, within seconds (Landrø 2007). A loose snow avalanche has a pear shaped pattern, while a slab avalanche usually has a rectangular shape (McClung and Schaerer 2006). 99 per cent of the avalanche accidents occur due to slab avalanches (Landrø 2007). The 2015 and 2017 urban avalanches in Longyearbyen were slab avalanches.

As we have already discussed, the chances of survival after 15 minutes are relatively poor if a victim is completely buried by the snow in an avalanche. Therefore, the best avalanche rescue strategy is to avoid to get caught in an avalanche in the first place (Fredston and Fesler 1986). In most situations, successful avalanche rescue depends on a prompt response from people nearby (Kruke 2016). Statistically, professional emergency responders arrive *too* late in the avalanche area to save lives. The mobilisation and deployment of professional emergency responders might take hours. Consequently, people moving in avalanche prone terrain should have the basic knowledge and training, both to avoid getting caught in an avalanche, and to perform buddy search and rescue if needed (Brattlien 2017, Landrø 2007, McClung and Schaerer 2006). Basic avalanche equipment, such as avalanche beacons, probes and shovels, is crucial to perform a speedy and reliable search and rescue operation (Brattlien 2017).

3 CONCEPTUAL FRAMEWORKS

After this brief introduction to some central avalanche characteristics, we now turn to the conceptual frameworks, with an emphasis on crisis characteristics, crisis management and the population's behaviour and capacities in crises.

3.1 *Crisis characteristics*

Rosenthal and colleagues define a crisis as «a serious threat to the basic structures or the fundamental values and norms of a system, which under time and pressure and highly uncertain circumstances necessitates making critical decisions» (1989: 10). This definition highlights crisis characteristics such as threat, uncertainty, time pressure and the need for critical decision-making. In addition to these crisis features, we need to look further into some defining characteristics of crises. 't Hart and Boin (2001) and Gundel (2005) present two different approaches to the classification of crises. 't Hart and Boin (2001) classify crises based on their speed of development and termination patterns, while Gundel (2005) differentiates between crises based on the degree to which they are predictable and influenceable. It is fair to assume that unexpected and fast developing crisis will be particularly difficult for professional responders to manage in the initial phase. Thus, the population present will often be left to manage the initial acute phase of these crises themselves.

It is quite common to distinguish between different phases of a crisis. A number of researchers

Figure 1. Crisis phases as a circular process (Kruke 2015).

Table 1. Some fundamental differences between the military model and the problem-solving model (Dynes 1993).

The military model	The problem-solving model
The initial response in the incident area is characterized by chaos	The initial response in the incident area is characterized by a degree of continuity
Command systems are needed to re-establish control	The response must be coordinated to reach an overall degree of cooperation between the actors present
Existing social structures have limited capacities	Existing social structures perform crisis management
Centralized approach	Decentralized approach

have suggested ways of doing this, e.g. (Turner 1976, Kruke 2015, Olson 2000). The model below outlines three basic phases, which are usually inherent in most phase descriptions in the crisis literature (Kruke 2015).

We will put emphasis on the acute phase of the crisis in this paper. This is the phase where lives might be at stake and responders come to the aid of people in acute need. This phase, when the population is conducting the response, prior to the arrival of the professional emergency responders, is also termed the *golden hour* (Helsloot and Ruitenberg, 2004). It is important to notice that the response, or the acute crisis management, and to what extent it succeeds, has a strong connection with the prevention and preparedness activities in the pre-crisis phase (Engen, Kruke et al. 2016; Kruke 2015).

3.2 *Crisis management*

Dynes (1993) argues that the dominant approach to crisis planning and management is based on a centralized command and control model, derived from military analogies. He claims that this model is based on inadequate assumptions about social behaviour in crises (ibidem). The military model is built on the notion that crises create social chaos. Hence, command and control must be established by professional emergency responders to eliminate such chaos (Dynes 1993). In this perspective, it is assumed that much of the civilian population become passive, helpless and incapable of making reasonable decisions in crises (Dynes 1994a). Dynes (1993) also suggests a decentralized problem-solving model as an alternative approach, derived from research on social behaviour in crises. The problem-solving model suggests that the focus should be on continuity rather than chaos, coordination rather than command and cooperation rather than control. The model takes the stand that civilians are fully capable of making reasonable decisions in a crisis, and that they will not be stunned, passive or irresponsible (Dynes 1994a). The model assumes that a crisis represents a set of problems that also have to be dealt with by the resources that are already present in the local community (Dynes 1993). Problem-solving capacities are inherent in all individuals and social structures. Thus, such capacities must be utilized in an effective manner when a crisis strikes us (Dynes 1994a).

These models are based on a totally different view of the population's behaviour and capacities in crises.

3.3 *The population's behaviour and capacities*

Panic and helplessness are two viable myths about people's behaviour in acute crises. What is panic? The tabloid newspapers, and the public discourse, often label the reactions of victims as panic. Many researchers have contributed to our understanding of the degree to which we panic in crisis situations, for example (Fritz and Marks 1954, Quarantelli 1954, Helsloot and Ruitenberg 2004, Ripley 2008, Kruke 2015). However, there is no complete consensus about what panic really is. The term is used to describe many human reactions in stressful situations, such as an excessive alertness and/or uncontrolled fear that leads to unwise action, or loss of judgment. A common use of the term is that panic is a very excited emotional state of acute fear, and that it results in uncontrolled escape (Quarantelli 1953). Panic can also be characterized as a form of irrational behaviour (Fritz and Marks 1954, Quarantelli 1999). There are some conditions that may lead to panic behaviour (Fritz and Marks 1954, Perry and Lindell 2003):

- Understanding of an immediate and serious threat
- Flight options are limited and disappearing.

Even though there are examples of panic in acute and life-threatening situations, there are also examples of such situations were panic do not occur (Quarantelli 1954, Dynes 1993, Helsloot and Ruitenberg 2004, Ripley 2008, Kruke 2015). This research concludes that citizens provide a lot of assistance, even in situations where they put their own life at stake, trying to rescue fellow citizens.

Some people become helpless in acute crises, mainly due to shock leading to paralysis and passivity (Helsloot and Ruitenberg 2004, Kruke 2015) or a disaster or crisis syndrome (Perry and Lindell 1978, Tierney, Lindell et al. 2001), described as a dull state of being. Passivity may also result in a collective disclaimer, called the «bystander effect» (spectator effect) (Darley and Latanté 1968), where no one takes responsibility immediately after an

event because everyone thinks someone else will do it.

Another string of research is related to altruism, a stand that helpless and panic behaviour are myths, and that most people both want to and also contribute to saving lives in the midst of a crisis. Altruism may be understood as a moral commitment people have to serve others and to put the interests of others in front of their own (Comte 1973), as a form of solidarity during crises (Tierney, Lindell et al. 2001). Altruism can be manifested as an individual, collective or situational commitment (Dynes 1994b). The society's collective and institutional resources may in some crises be inadequate for a prompt and reliable response (ibidem). To cope with these unexpected, acute and uncertain circumstances, individual altruism may result in responses supplementing traditional response structures (ibidem).

4 METHODS

This paper mainly draws on empirical findings from 8 months of fieldwork in Longyearbyen, from mid-September 2016 to end of May 2017. The primary data collection methods were semi-structured interviews, field conversations, observations and participant observations. Semi-structured interviews were conducted with the local population and response actors in Longyearbyen, including: The Police, Longyearbyen Red Cross Society, Longyearbyen Hospital, Lufttransport AS[1], Longyearbyen Fire and Rescue Agency, and the Longyearbyen Community Council. Additionally, 40 field conversations were conducted with the local population. Empirical data was further collected from three public meetings in conjunction with urban avalanches in Longyearbyen. Furthermore, one of the authors took part in the local population's response to the urban avalanche 21 February 2017. Empirical data was also collected from document analysis, primarily public documents, newspaper articles, and avalanche literature and statistics.

5 EMPIRICAL FINDINGS

5.1 *The urban avalanche in 2015 and the local population's reactions in the acute phase*

A slab avalanche hit urban areas in Longyearbyen Saturday 19 December 2015 at 10.23 am (DSB 2016). According to an informant from the Red Cross Society survivors and neighbours immediately commenced search and rescue activities. Professional emergency responders were also in the avalanche area within minutes. Shortly after the avalanche, a private citizen posted a message on a local community Facebook group to encourage people to pick up their shovels and meet in the avalanche area. The Longyearbyen Community Council posted a similar message on their public Facebook page. In a short period of time a huge mass of unorganized volunteers lined up to participate in the excavation efforts. There are many examples of this initial response.

Five people sat in their kitchen eating breakfast when the avalanche hit their house. All of them were buried by the snow. One of them managed to get out of the snow masses by himself. He then managed to locate and excavate his wife and his eight-week-old daughter. He continued to search after the two missing persons, and dug in the snow masses with a wok pan, while he wore a sweater, sweatpants and socks. This was one of the first impressions meeting the police officers upon their arrival in the avalanche area.

Another informant lived close by the houses that were hit by the avalanche. He noticed that the lights indoors and outdoors started to flicker. Suddenly, everything went dark. He immediately understood what had happened when his wife shouted that several houses had collapsed. He grabbed his headlight and shovel, and ran out. Meanwhile, his wife called the police and went from house to house to alert the neighbours. She also brought the children in the area to safety. The informant came home with two survivors and provided them with warm clothes, blankets and food. As soon as medical personnel arrived, he went out again to continue search and rescue efforts.

Yet another informant was on his way to work when he noticed blue lights and sirens nearby the town centre. He met some people outside the shopping mall who told him about the avalanche, and that they were on their way to offer their assistance. He instantly joined them in the excavation efforts.

None of the informants from the professional emergency responders observed people affected by panic or shock. However, a small number of the informants noticed a few passive spectators. A doctor working at the hospital told about a man who stood nearby when the avalanche hit. He had some trouble coping with the situation and seemed to be rather confused. One of the informants from the local population described an acquaintance who allegedly was struck by panic. As she stood on the stairs outside her home smoking a cigarette a house started to move and hit the house next to her. According to the informant, the woman then

[1]Lufttransport AS is a private company operating the Governor of Svalbard's two search and rescue helicopters.

panicked and ran away, and was taken care of by others later on.

A number of informants told about people who felt helpless as a consequence of not being able to participate in the search and rescue operations in the acute phase of the crisis. One informant said that he suffered for months because of not being involved in the response. He had no choice but to stay at work when the avalanche hit. Another informant told about some friends unable to get out of their house, due to the snow masses. Tt made them feel extremely indignant, as they had a strong desire to help their fellow citizens in the avalanche area.

5.2 *The urban avalanche in 2017*

In the morning hours Tuesday 21 February 2017 the Longyearbyen Community Council informed the public on their website about an increased avalanche hazard. The message reassured that the settlement was not considered to be at risk. A couple of hours later, a slab avalanche hit the settlement and destroyed an apartment building (Tengesdal 2017). The people at home managed to assist each other in the evacuation out through the windows. Since also this avalanche hit nearby the town centre, the professional emergency responders were not far away. The police quickly confirmed that none of the people living in the area were missing nor seriously injured. They still decided to maintain the search, in case other citizens had been in the area (Malmo 2017).

5.3 *The professional emergency responders' perspectives on the local population contribution*

All of the informants interviewed characterized the urban avalanche in 2015 as a totally unexpected event. However, a couple discussed the bad weather conditions the previous night and the resulting huge snowdrifts on Sukkertoppen, just uphill from their house. While looking up the hill through the window, they agreed that this could not possibly be good. Soon after, the avalanche hit their house (Palm 2016). Both of them survived the avalanche.

The informants described the avalanche area as rather chaotic. Many referred to houses that were moved and destroyed. However, the informants underlined that the avalanche area was quiet and that the level of stress was low. A professional rescue worker from Lufttransport AS said: «What you see is chaos, but there is no panic». All of the informants expressed that the local population was a crucial resource in the search and rescue operation. An informant from the Red Cross Society said that «the help we got from the unorganized volunteers made it possible to quickly locate and excavate the victims». A police officer further explained why the local population was an important resource in this response operation: «Time is of the essence when it comes to avalanches. The chances of survival decrease rapidly. The important thing in this situation was to gather a lot of people who could dig».

An informant from the Red Cross Society explained that intelligence is of the utmost importance to be able to locate victims, both in backcountry and urban avalanches. In line with him, Genswein and Harvey (2002) point out that intelligence in a backcountry avalanche is gathered by looking for tracks into the avalanche area, an initial primary search for people or their belongings, and a more thorough systematic search in the avalanche area. In an urban avalanche, the snow masses are so polluted with debris that the conventional search strategies are no longer effective. Instead, intelligence is gathered by talking to survivors and neighbours. The professional rescue workers from Lufttransport AS underlined that they were totally dependent on the local population during the initial search and rescue phase. They especially pointed to the population's efforts in intelligence gathering and excavation of people buried by the snow. One of them claimed that «without the information and knowledge provided by the local population, it would have been like searching for the needle in the famous haystack». The local population also delivered different kinds of rescue equipment to the rescuers in the avalanche area, such as shovels, probes, jerven bags[2], chainsaws etc. In addition, people offered warm clothes, food and drinks to the rescue workers. None of this assistance was ever requested, according to an informant from the Longyearbyen Fire and Rescue Agency.

One of the informants from the Red Cross Society said that the professional emergency responders in Longyearbyen do not have the capacity to manage a crisis of this scale on their own. Several informants also mentioned that rapid assistance from mainland Norway in the acute phase of a fast developing crisis is not likely, due to the geographical distance. Tough weather conditions may delay assistance from outside even longer. This makes the local population a crucial resource in a crisis like the urban avalanche in 2015. Almost all of the informants described that the local population has proved to be a valuable resource in countless backcountry avalanches near Longyearbyen. People often come to the rescue on their snowmobiles, loaded with avalanche equipment. Depending on the situation, they initiate search and rescue operations

[2]A jerven bag is a protective poncho and tarpaulin.

themselves, or offer their assistance to the professional responders.

The vast majority of the informants from the professional emergency organisations expect the local population to be a resource in avalanche search and rescue. They pointed out that many of the citizens have the knowledge, training and equipment to be a resource. They also pointed to the fact that most of the local population is young and employable, and that they typically have a great interest in outdoor activities. The children in one of the kindergartens actually conduct search training when they play with avalanche beacons searching for candy buried together with an avalanche transmitter.

Some informants added that Longyearbyen is a rather small town where people know their neighbours, which makes all affected when an urban avalanche strikes the town. One of the rescue workers from Lufttransport AS referred to the message on Facebook posted by a citizen (mentioned earlier), and added that «the local population know their role in an event like this. Most of us have avalanche beacons, probes and shovels. Meet up»! A doctor working at the hospital said that «history has shown that the local population represents a difference when crises strike in Svalbard. It is a resource we can count on».

The authorities did not request any assistance from the local population after the avalanche in 2017. Nevertheless, a large number of unorganized volunteers, men and women, youths and adults, stood in line nearby the avalanche area with avalanche beacons, probes and shovels, ready to assist if needed. Everyone behaved calmly and focused, and there was no sign of irrational behaviour. After a relatively short period of time, the police informed people waiting to stand down, as none were missing and the risk of a second avalanche was considered to be high (Tengesdal 2017).

6 DISCUSSION

6.1 *Unexpected and fast developing crises*

The urban avalanches took most citizens and authorities in Longyearbyen completely by surprise, even though some of the citizens were worried due to the snow drifts uphill on Sukkertoppen in 2015. Also in 2017 some people were worried. But, they were reassured by the Longyearbyen Community Council that urban areas were safe. However, it is important to notice that the exact location and time of an avalanche cannot be predicted (Brattlien 2017). When the avalanches were triggered (natural causes), it was only a matter of seconds before they hit urban areas in Longyearbyen. Thus, with reference to the crisis typologies presented by 't Hart and Boin (2001) and Gundel (2005), it can be argued that the urban avalanches in Longyearbyen in 2015 and 2017 were unexpected and fast developing crises. It may also be noted that the avalanche in 2015 has some legal aftereffects influencing the termination of the event[3].

6.2 *The local population's response*

There are some factors that point to the need for the local population's response in avalanches. First of all, time is of the essence in such crises. The statistics tell us that the chances of survival following an avalanche decrease rapidly after just 15 minutes, if a victim is completely buried by the snow (Landrø 2007). Consequently, the time and pressure that Rosenthal et al. (1989) stress in their crisis definition are extreme. In general, it will often take some time before the professional emergency responders arrive at the scene in an acute crisis (Kruke 2012, 2015). Statistically, they arrive *too* late to save lives (McClung and Schaerer 2006, Landrø 2007). This means that the victims themselves, and the people who are randomly present, are the ones who need to take the first shift pending the arrival of the professional emergency responders, in the *golden hour* (Helsloot and Ruitenberg 2004). This golden hour is crucial when the difference between life and death is a race against time. As we have discussed, it was the victims and the people randomly present who initiated search and rescue in the urban avalanche in 2015. Also, in the urban avalanche in 2017, the victims themselves helped each other to safety. In other words, it was the victims and the people randomly present who took the responsibility to manage the initial acute phase of these crises. Therefore, they can be considered as *first responders* (Kruke 2015).

Another factor is related to the professional emergency responders' lack of capacities. The *golden hour* in these crises do not last more than a few minutes. The urban avalanche in 2015 clearly demonstrated that the professional emergency responders in Longyearbyen lack the capacity to manage a crisis of this magnitude on their own. Therefore, altruistic behaviour is necessary in such situations as a supplement to the traditional response structures (Dynes 1994b). The professional emergency responders were totally dependent on the continuity in operations (Dynes 1993) provided by the local population in the acute phase of the avalanches. The inclusion of the population

[3]The parents of a little girl dying in the avalanche may go to court to get compensation due to accusations of inadequate urban planning.

made it possible to quickly evacuate people, and to locate and rescue people buried by the snow. The local population has proved to be an important resource in many search and rescue operations following avalanches in Longyearbyen and surrounding areas. They often initiate search and rescue activities themselves, or offer their assistance to the professional emergency responders. The local authorities did not encourage the population to participate in the urban avalanche in 2017, as was the case in 2015. Still, approximately 150 unorganized volunteers met up next to the avalanche area (Tengesdal 2017). The fact that they actually met up unrequested, strengthen the view that the local population is an asset in avalanche search and rescue.

6.3 *The local population's behaviour*

An informant from the local population told about a woman who allegedly panicked and ran away when a house crashed into the house next to her. Can this behaviour actually be characterized as panic? Quarantelli (1953) describes panic as an excessive alertness or fear that leads to an unwise action. Panic can further be regarded as a form of irrational behaviour (Fritz and Marks 1954, Quarantelli 1999). In this specific situation, it is fair to assume that her decision to run away from houses *in motion* is a rather wise and rational modus operandi. She brought herself to safety. Generally, it is possible that panic occurs in acute crises, but we know that this form of irrational behaviour is rare (Helsloot and Ruitenberg 2004). The rest of the informants did not witness any examples of panic among the local population following the urban avalanches in 2015 and 2017. The doctor's story about the man who stood nearby when the avalanche hit in 2015 is a possible example of shock. He had some difficulties coping with the situation and seemed to be rather confused. It is not unlikely that he was paralysed (Helsloot and Ruitenberg 2004, Kruke 2015) as a result of watching the enormous slab avalanche hitting the settlement.

There are several examples of people who actually felt helpless, as a consequence of *not* being able to participate in the search and rescue operation in 2015. The fact that people felt helpless as a result of *not* being able to participate in the acute response, can be regarded as a considerable contrast to the widespread assumption that we turn into helpless human beings when crises strike. This study clearly shows that the population has a strong desire to help others in need. This may be rooted in a moral commitment (Comte 1973) and a form of solidarity (Tierney, Lindell et al. 2001). The spectator effect (Darley and Latanté 1968) was minimal in the urban avalanches in 2015 and 2017. The informants have very few examples of passive spectators in the avalanche area in 2015. There were also very few passive spectators observed in 2017 (Tengesdal 2017).

6.4 *The military model versus the problem-solving model*

The responses to the 2015 avalanche indicate that many of the citizens in Longyearbyen possess both the knowledge and training to be a valuable resource in avalanche search and rescue. The empirical findings also show that the professional emergency responders are aware of the local population's capacities, and that they actually count on the local population to be a resource in avalanches rescue when needed. This is largely in line with the problem-solving model presented by Dynes (1993) stressing the need for effective utilization of existing problem-solving capacities in the community (Dynes 1994a). The population quickly initiated search and rescue operations following the avalanche in 2015, displaying a sort of continuity (Dynes 1993) witnessed by the professional emergency responders upon their arrival in the avalanche area. The local authorities actively encouraged the population to participate in the excavation efforts in the urban avalanche in 2015 and coordinated the joint efforts between the actors present. Clearly, the authorities and the professional emergency responders didn't expect social chaos, which is an inherent assumption in the military model (Dynes 1993). Instead, they expected the population to be a resource, as the problem-solving model strongly suggests (ibidem).

Instead of establishing command and control structures, the authorities and the professional emergency responders focused on coordination and cooperation among public and private response structures in the avalanche area in 2015. It is fair to assume that the public-private cooperation was decisive for the outcome of the avalanche in 2015. Victims and neighbours immediately initiated search and rescue, and a large number of unorganized volunteers came to assist the professional responders. As a result, the victims were quickly located and excavated.

6.5 *Characteristics of the local population*

Can we expect the population to be a resource in urban avalanche search and rescue elsewhere, or is there something unique about Longyearbyen? Many social researchers take the stand that most citizens are rational actors who often provide lifesaving assistance in acute crises and demanding situations where lives are at stake (Quarantelli 1954, Dynes 1993, Helsloot and Ruitenberg 2004, Ripley 2008, Kruke 2015). Therefore, it is fair to assume that also citizens elsewhere would have

done their best to rescue their neighbours. Even so, there seem to be some aspects that particularly make the population in Longyearbyen a resource in avalanche search and rescue.

Firstly, the ones who come to the rescue possess the essential equipment that is needed to perform a speedy and reliable avalanche search and rescue operation. Secondly, many of them have both the knowledge and training to be a valuable resource. Thirdly, the population is young and employable. Fourthly, many citizens in Longyearbyen have a great interest in outdoor activities. Consequently, many of the citizens are also physically fit for an avalanche search and rescue operation. This represents a set of characteristics that presumably is quite unique for the population in Longyearbyen. In addition to this, it is likely that the rural location is influencing the population's expectations to help each other. Rescue resources from outside the archipelago cannot be expected in the acute phase of an avalanche. The citizens know that *they* are the ones who must provide the assistance. The sense of unity among citizens is strong, making the whole community affected when an urban avalanche strikes. Although these characteristics of the population in Longyearbyen make them a valuable response resource in an unexpected and fast developing urban avalanche, it can hardly be argued that they apply exclusively for the population in Longyearbyen.

7 CONCLUSIONS

The aim of this paper has been to discuss the role of the local population in urban avalanche search and rescue. The cooperation between the professional emergency responders and the local population was decisive for the outcome of the unexpected and fast developing urban avalanche in 2015. The urban avalanches in 2015 and 2017 show that the local population are the ones who are present when crises strike, and they perform search and rescue in the acute phase. The immediate response from survivors and neighbours, and the inclusion of unorganized volunteers that came to the rescue, made it possible to locate and rescue seven victims buried by the snow. This was possible due to an altruistic behaviour of the local population, that the citizens possessed the required equipment, knowledge and training, and that the local authorities and the professional emergency responders understood the value of including the population in the search and rescue operation. The responses to the avalanches in 2015 and 2017 show that the widespread assumptions about panic and helpless behaviour are mostly myths. The local population are first responders and have a strong desire to help their fellow citizens in need.

REFERENCES

Brattlien, K. (2017). *The Little Avalanche Book* (own translation). Oslo: Fri Flyt AS.

Comte, A. (1973). *System of Positive Polity*. Translated by R. Congreve and H.D. Hutton. (Original published in 1854). New York: Burt Franklin.

Darley, J.M. and B. Latanté (1968). Bystander Intervention in Emergencies: Diffusion of Responsibility. *Journal of Personality and Social Psychology 6(4)*: 377–383.

DSB (2016). *Report on the avalanche in Longyearbyen 19th December 2015*. DSB: Tønsberg, Norway.

Dynes, R.R. (1993). Disaster reduction: The importance of adequate assumptions about social organization. *Sociological Spectrum 13 (1)*: 175–191.

Dynes, R.R. (1994a). Community Emergency Planning: False Assumptions and Inappropriate Analogies. *International Journal of Mass Emergencies and Disasters 12(2)*: 141–158.

Dynes, R.R. (1994b). *Situational Altruism: Toward an Explanation of Pathologies in Disaster Assistance*. World Congress of Sociology. Bielefeld, Germany, University of Delaware Disaster Research Center. Preliminary paper #201.

Engen, O.A., B.I. Kruke, P.H. Lindøe, K.H. Olsen, O.E. Olsen & K.A. Pettersen (2016). *Perspectives on societal safety and security* (own translation). Oslo, Cappelen Damm.

Fredston, J., D. Fesler, K. Birkeland & D. Chabot (1986). *Snow Sense – A guide to Evaluating Snow and Avalanche Hazard*. Anchorage: Alaska Mountain Safety Centre.

Fritz, C.E. & E.S. Marks (1954). The NORC studies of human behavior in disaster. *Journal of Social Issues 10*: 26–41.

Genswein, M. & S. Harvey (2002). *Statistical Analyses on Multiple Burial Situations and Search Strategies for Multiple Burials.* International Snow Science Workshop, Penticton, British Columbia.

Gundel, S. (2005). Towards a New Typology of Crises. *Journal of Contingencies and Crisis Management 13(3)*: 106–115.

Hart, 't P. & A. Boin (2001). Between Crises and Normalcy: The Long Shadow of Post-crisis Politics. In Rosenthal, U., A. Boin, & L. Comfort (eds.) (2001). *Managing Crises: Threats, Dilemmas, Opportunities*. Springfield, Charles: Thomas Publisher ltd.

Helsloot, I. and A. Ruitenberg (2004). Citizen Response to Disasters: a Survey of Literature and Some Practical Implications. *Journal of Contingencies and Crisis Management 12(3)*: 98–111.

Indreiten, M. and C. Svarstad (2016). *The Longyearbyen Fatal Avalanche Accident 19th December 2015. Svalbard – Lessons Learned From Avalanche Rescue Inside A Settlement*. Svalbard: The University Centre in Svalbard.

Kruke, B.I. (2012). Societal safety and crisis management: Relevance for 22 July 2011. 22 July-Commission Paper. *22 July Commission 7(12)*.

Kruke, B.I. (2015). *Planning for crisis response: the case of the population contribution.* ESREL Safety and Reliability of Complex Engineered Systems: 177–185.

Kruke, B.I. (2016). *The population contribution in crisis management: A case of uncertainty and resilience.* Risk, Reliability and Safety: Innovating Theory and Practice. R.B.E. Walls, Taylor & Francis Group, London. Safety and Reliability of Complex Engineered Systems: 2160–2167.

Landrø, M. (2007). *Avalanche danger – avalanches, risk, rescue* (own translation). Oslo: Fri Flyt AS.

Malmo, O.J. (2017). *Public meeting after the urban avalanche in Longyearbyen.* Longyearbyen: Longyearbyen Kulturhus 21. February.

McClung, D. & P. Schaerer (2006). *The Avalanche Handbook*. Third edition. Seattle: The Mountaineers books.

Olson, R.S. (2000). Towards a politics of disaster: Losses, values, agendas and blame. *International Journal of Mass Emergencies and Disasters 18(2)*: 265–287.

Palm, E. (2016). *The story of Anne* (own translation). Longyearbyen: Svalbardposten.

Perry, R.W. & M. Lindell (2003). Understanding Citizen Response to Disasters with Implications for Terrorism. *Journal of Contingencies and Crisis Management 11(2)*: 49–60.

Perry, R.W. and M. Lindell (1978). The Psychological Consequences of Natural Disasters: a review of research on American communities. *Mass Emergencies 3*: 105–115.

Quarantelli, E.L. (1953). *A study of panic: its nature, types, and conditions*. Chicago, University of Chicago, National Opinion Research Center. Master Thesis.

Quarantelli, E.L. (1954). The Nature and Conditions of Panic. *American Journal of Sociology 60*: 267–275.

Quarantelli, E.L. (1999). *The Sociology of Panic.* Delaware: University of Delaware Disaster Research Center. Preliminary paper 283.

Ripley, A. (2008). *The Unthinkable: Who survives when disaster strikes – and why*. New York: Crown Publishers.

Tengesdal, S.M. (2017). *We were totally dependent on the help of the local population: A qualitative study of the role of the local population in Longyearbyen in the management of avalanches and avalanche exercises.* (Master thesis). University of Stavanger, Stavanger, Norway.

Tierney, K., Lindell, M.K., & Perry, R.W. (2001). Facing the Unexpected: Disaster Preparedness and Response in the United States. Washington, D.C.: Joseph Henry Press.

Turner, B.A. (1976). The organizational and interorganizational development of disasters. *Administrative Science Quarterly 21*: 378–397.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

At least as safe as manned shipping? Autonomous shipping, safety and "human error"

T. Porathe & Å. Hoem
Norwegian University of Science and Technology, Trondheim, Norway

Ø. Rødseth & K. Fjørtoft
SINTEF Ocean, Trondheim, Norway

S.O. Johnsen
SINTEF Technology and Society, Trondheim, Norway

ABSTRACT: A paradigm shift is presently underway in the shipping industry promising safer, greener and more efficient ship traffic with unmanned, autonomous vessels. In this article, we will look at some of these promises. The expression "autonomous" and "unmanned" are often used interchangeably. We will therefore start out by suggesting a taxonomy of automation and manning of these ships. We will then go on examining the promise of safety. An hypotheses of increased safety is often brought forward and we know from various studies that the number of maritime accidents that involves what is called "human error" ranges from some 70–90 percent. If we replace the human with automation, can we then reduce the number of accidents? And is there a potential for new types of accidents to appear? Risk assessment will be a valuable tool, but will only reach as long as to the "known unknowns".

1 INTRODUCTION

The shipping industry are about to enter a new epoch. The story started in the 1800 when mechanized power was introduced and the vessels moved from propulsion by sail to propulsion by steam. The next stage came in the early 1900's when the diesel engine enabled more efficient and reliable ship services, analogous to the introduction of mass production on shore. In the 1970's the computerized control of ships was introduced. Now we are about to go a step further where cyber physical systems and autonomy, as part of "Shipping 4.0" (Rødseth 2017), will form a new gravity.

1.1 *The first autonomous ship accident*

We will start this article by a fictive illustration: It was an unusually warm to be in the end of October. The water in the strait was completely calm and mirrored the sky and the setting afternoon sun. In the Vessel Traffic Service (VTS) tower under the bridge the operator followed a lone kayak with his binoculars. It seemed like the kayaker was a child and not very proficient in his or her paddling and the kayak only slowly worked its way across the sound. The timing for crossing was not the best, the operator thought. He had an outbound oil tanker due in a few minutes and the autonomous *Yara* shuttle was to pass in the other direction soon after. The tanker was already approaching from the far side of the bridge sounding her horn to let the kayaker know she was approaching the 200 meters wide strait, something that probably did not make the situation better for the child in the kayak, the VTS operator thought. From the other side the autonomous shuttle was visible inbound on a westerly course with her 6 knots. He expected her to slow down any minute as her sensors detected the kayak in the sound.

Suddenly two water scooters appeared from nowhere, criss-crossing over the strait and around the kayak at some thirty or forty knots. The VTS operator could hear the roar from their engines all the way into the VTS tower. The surplus water shot up like a fountain from the back of the scooters and their wakes brought the water into turmoil around the kayak. In his binoculars, the VTS operator saw the child in the kayak letting go of his paddle and waving his arms to signal the scooters. Suddenly the kayak flipped over and the boy disappeared into the water. The scooters shot off towards the far side and the operator could see the head of the boy reappear on the surface beside the overturned kayak. He was right in the way of the tanker. The operator quickly grabbed the VHF receiver and called the tanker.

"*Tarnfjord, Tarnfjord* this is Brevik VTS on channel 16. Have you seen the overturned kayak ahead of you?"

"Brevik VTS, this is *Tarnfjord.* Rodger that. We are slowing down and holding to port. We should manage to avoid the kayak. But we cannot reverse. And we will have close call with *Yara.*"

"OK, *Tarnfjord*, thank you for that," the VTS operator replied, and continued immediately to call the shuttle, "Yara remote control, Yara remote control, are you following what is happening in the Brevik strait?"

He turned and looked at the shuttle and could see that she had not slowed down as he had expected. Both of the ships were now only a few hundred meters from the overturned kayak under the bridge.

"*Yara* remote control, *Yara* remote control, this is Brevik VTS on channel 16. Please respond *Yara*."

He took up his binoculars and saw that the tanker was slowly turning. The shuttle was now only some 100 meters from the overturned kayak and the turning tanker and still showed no sign of slowing down.

The radio crackled. "Brevik VTS, this is *Yara*. Did you call me? I had a coffee break."

"Thank, you, *Yara*," the operator quickly replied. "Stop immediately; can't you see the kayak in front of you?"

"No, the sun is completely blinding both my cameras and on the radar I only see the bridge" the remote operator answered, and then he shouted "What the hell is the tanker doing!"

We will not know how this incident ended as it is pure fiction and the Yara shuttle will not start to traffic the Brevik strait in southern Norway until 2021 (she will be manned in 2019, remote controlled in 2020, before attempting to go autonomous 2021). Nevertheless, the situation could be plausible. Kayaks, scooters and other leisure crafts will be close companions to autonomous ships in Scandinavian waters summertime. Cameras and radars can be deceive, as was shown in the Tesla car accident in 2017 (Lambert 2017; NTSB 2017). Bridges may obscure radar detection of objects underneath. Objects coming and leaving like the two scooters may confuse the artificial intelligence of collision avoidance systems, and LIDAR (Light Imaging, Detection, And Ranging) is only useful at close range, closer than the stopping distance. Finally, the human backup may have gone for a cup of coffee.

The fictional incident above is, maybe unfairly, attributed to the planned autonomous Yara-Birkeland container feeder (Kongsberg Maritime 2017). This unmanned, autonomous vessel, taking 120 containers on a fully electric propulsion system, will replace some 20 000 trucks taking the same amount of containers on the road today. There is an economic as well as environmental gain to be made. Doing this autonomously and unmanned will be a challenge. So let us start by looking at that.

1.2 *Ambiguity in definitions*

The concepts of unmanned and autonomous when used on ships are ambiguous. The ship bridge may be unmanned, perhaps in periods, but crew may still be on board, ready to take control when needed. A ship can also be remotely controlled from a shore station via highly redundant and high capacity communication links. Is this ship unmanned or autonomous? A dynamic positioning (DP) system on a ship will automatically control the position and perhaps the heading of the ship, but most DP systems will rely on an operator to handle any errors, e.g. in sensors, that occur during the operation. Is the DP automatic or autonomous?

Furthermore, to what ship functions do unmanned or autonomous apply? In (Rødseth & Tjora 2017), eight main functional groups are identified, including, e.g. navigation, engine control, cargo monitoring and onboard safety functions. In the following text, we will refer to typical bridge functions, but in a truly autonomous ship, all shipboard functions must be automated to some degree and the degree of autonomy may be different for each function.

Finally, the degree of autonomy will be different during the ship's voyage. Tighter supervision and perhaps continuous remote control will be necessary during berthing while a high degree of autonomy is normally desired during the deep-sea passage.

This ambiguity is reflected in many existing definitions of "autonomy levels". In (Vagia et al. 2016), 12 different "levels of autonomy" are examined and even more have become available as autonomy levels have been extended to ships (Rødseth & Nordahl 2017). One reason for the numerous definitions is that autonomy must be defined along several axes and with a strong focus on the operational profile at hand. The idea of autonomy is very context dependent.

1.3 *Three axes of autonomy*

For ships, we propose to characterize autonomy along three axes (Rødseth & Nordahl 2017).

One axis is the *complexity* of the intended operation. Is the ship operating in sheltered or open seas, what are the likely weather or visibility impacts, how much other traffic is there, how complex is the sailing routes in terms of shallows, turns and obstacles, and so on. We propose to capture the complexity in the operational design domain (ODD) as explained in the next section.

Table 1. List of autonomous ship operation types.

	Continuously manned bridge	Unmanned bridge, crew on board	Unmanned bridge, no crew on board
Operator controlled	Direct control	Remote control	Remote control
Automatic	Automatic control	Automatic control	Automatic control
Partly autonomous	Partly autonomous	Partly autonomous	Partly autonomous
Constrained autonomy		Constrained autonomy	Constrained autonomy
Full autonomy			Full autonomy

The second axis is the *manning level*. The ship can have a continuously manned bridge, but still have a high degree of autonomy in automated object detection and collision avoidance. One can foresee ships with enough autonomy to allow the crew to go to bed at night, when sailing in open waters and fair weather. Ships can also be remotely controlled, with hardly any "real" autonomy at all. On the other end of the axis, one may see ships with no crew and no remote monitoring at all: they are fully autonomous. The manning level is dealt with in Table 1.

The third axis is the operational autonomy, how the necessary operations to satisfy requirements of the ODD are divided between human and machines. We propose to capture this aspect by diving the Dynamic Navigation Tasks (DNT) into two parts: One part that requires human intervention to be executed (Operator Exclusive DNT) and one that can be handled by the automation systems (Control System DNT).

1.4 *A proposed taxonomy*

To simplify the definition of autonomous and unmanned, we will start with a concept borrowed from the US car industry and its definition of terminology for autonomous cars (SAE 2016). This is called the "Operational Design Domain" (ODD) which is the operational conditions that limits when and where a specific autonomous car can be used. The corresponding capabilities of the car and its control systems is the "Dynamic Driving Task" (DDT). The concept also includes the "DDT Fallback" which is procedures and safety guards that are built into the vehicle and control systems for handling situations when the ODD is exceeded. The DDT Fallback will bring the system to a "minimal risk condition" (SAE 2016). For a ship, we suggest renaming DDT to the "Dynamic Navigation Task" (DNT).

Most autonomous or unmanned ships are expected to have a "backup" operator somewhere on board or on shore, so that situations that cannot be handled by automatic functions can be safely handed over to the operator. This can be illustrated by dividing the DNT into two regions:

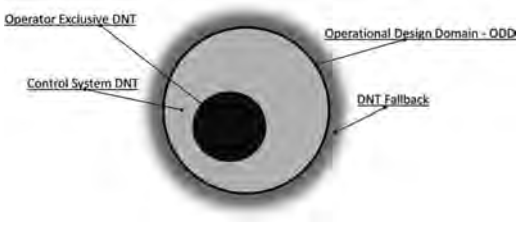


Figure 1. The operational design domain and dynamic navigation task.

The "Operator Exclusive DNT" where the operator is needed to resolve problems that the automation cannot handle and the "Control System DNT" which represents the unassisted capabilities of the automatic systems. The complete concept is illustrated in Figure 1.

A proposed set of definitions for autonomous merchant ships (Rødseth, Nordahl 2017) indicates that four distinct levels of autonomy may be needed and are probably sufficient. These levels are defined independently of the human operator being located on board the ship or in a remote location:

1. *Operator controlled (AL0-1):* The DNT is fully handled by the operator. Systems may provide decision support or very limited automatic control, e.g. as in an auto pilot or track pilot. This is the current situation on today's ships.
2. *Automatic (AL2):* The ship systems can operate without human intervention for a very specific function, typically as a DP system works today. An operator is required to handle all deviations from expected operational parameters. This autonomy level is probably appropriate for automatic berthing or other situations where very accurate control is needed and where less deterministic and autonomous problem handling is unwanted.
3. *Partly autonomous (AL3):* The ship can perform certain tasks in the DNT autonomously, e.g. transiting open sea in fair weather. This can, e.g. be used to have a periodically unmanned bridge.
4. *Constrained autonomous (AL 4):* The ship can operate autonomously within most or all of the

DNT, but it has clear limits to what actions it can take by itself, e.g. maximum speed and track deviations. If the ship needs to exceed these limits, e.g. due to anti-collision manoeuvres, the operator has to be called to change limits or to remotely control it until constrained operations can resume.

5. *Fully autonomous:* The ship systems can perform all its DNT tasks without human intervention. There are no operational limits beyond those defined by the OOD.

Constrained autonomy is the most likely type of autonomy for fully unmanned ships with shore supervision. It enables the ship to solve all "standard" problems by itself while reducing system complexity by having an operator available for the more complex situations. It also gives a high degree of operational determinism due to the operational envelope it cannot exceed without human acceptance. Fully autonomous is the necessary level for autonomous ships that have no remote supervisor. This will in many cases require very complex control systems and is not very likely level for ships in the near future.

The levels can be characterized by having different ratios between the operator exclusive DNT (black) and the control system DNT (grey), as illustrated in Figure 2. One may validly argue that the levels between automatic and constrained autonomy should be the same class as they both have operator and control system DNTs. However, it is useful to differentiate between them since they are likely to be used in different context during the voyage.

Dependent on autonomy level and the operator being available on the ship or on shore, one can de-fine the matrix in Table 1. The shaded cells represent operations where one will require a manned shore control center to handle deviations from operator DNT fast enough. The empty cells represent types that are not very relevant, although possible.

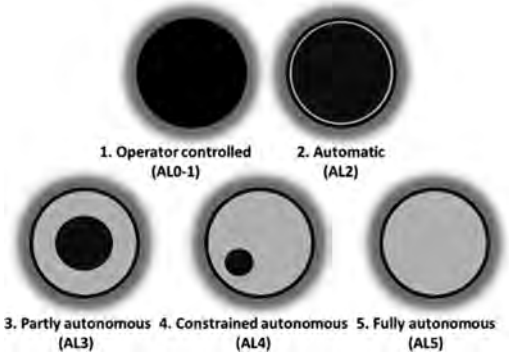


Figure 2. Five levels of autonomy.

The level of autonomy will vary over the ship's different functions such as engine control, cargo monitoring and navigation functions. It will also vary during the ship's voyage. This may be result of, e.g. using an unmanned bridge during night and open sea passage or by having different modes in different phases of the voyage, e.g. using remote control during port approach and automatic control during berthing.

2 AUTOMATION

Going back to the concept of ODD and DNT, one may argue that most incidents occurring with automated systems may be of the following types:

1. *Errors in control system DNT (CS-DNT):* These are purely technical errors that occur in the automation systems and associated sensors. It may be caused by technical system malfunctions or by design errors in system designs or configurations.
2. *Errors in operator exclusive DNT (OE-DNT):* These are human operational errors that may have been caused by, e.g. fatigue or low situation awareness which, in turn, may have been caused by bad technical systems. However, the incident is directly attributed to a human operational error.
3. *Transition from CS-DNT to OE-DNT:* This is a critical issue as the transition both has a timing aspect and must be fast enough and a situation awareness aspect as the human must understand the background for the transition to make the correct decisions.
4. *Operator intervention in CS-DNT:* There are also examples of incidents that have been caused by operators intervening in automated processes when they should have left the automation system alone.
5. *Transition from OE-DNT to CS-DNT:* This is probably a less common type, but it may be challenging to make sure that the automatic control system is activated at the right time and with the right parameters settings.
6. *Transition to DNT Fallback:* When to activate the DNT Fallback is also a critical issue. The DNT Fallback is not necessarily a "fail to safe" control as ships do not have a generally safe state. It is a "minimal risk condition" (SAE 2016). Thus, there is an inherent risk in going from OE-DNT or CS-DNT to DNT Fallback and it is a challenge to define the proper conditions for doing so, particularly when a human is in the control loop.

While this classification seems most relevant for autonomous ships, it is also applicable to manned ships with automation or decision support components. In particular, the transitions between automatic and human control in current automated systems will be a good indication of how this problem will develop when more autonomy is added in the system.

In the following, we will discuss known benefits and shortcomings of today's manned operation with automation and see how that can be applied to autonomous ships.

3 SAFETY, HUMANS AND AUTOMATION

If autonomous unmanned ships are to become a success they have to prove successful in several areas, and safety is one of them. Thus, the first thing we might ask is how safe is then manned shipping?

3.1 *At least as safe as manned shipping*

In a study by Oxford University on British data from 1976 to 1995, the seafaring job is ranked as the second most dangerous occupation in Britain—after being a fisher (Roberts 2002). This is however not usually because ships are sinking, but because of occupational hazards like slips, trips, and falls on a moving platform full of heavy gear and a hazardous environment. In this sense, we might conclude that already removing humans from this hazardous environment has a safety benefit.

However, if we by safety think of the safety of the ship we can say that shipping is very safe and is becoming even safer every year. Just to provide a background we can note that in the three years between 1833 and 1835, on average 563 ships per year were reported wrecked or lost in United Kingdom alone (Crosbie 2006). Today the total number of tankers, bulk carriers, containerships and multipurpose ships (over 100 Gross Tons) in the world fleet has risen from about 12,000 in 1996 to some 33,000 in 2016 (Clarkson 2017). During the same time, the number of ships totally lost per year (ships over 500 Gross Tons) declined from 225 in the year 1980, to 150 in 1996 and 33 in 2016 (total losses as reported in Lloyds List – IUMI 2016) – and this worldwide.

If we look at ship accidents broken down into different causes, we can see that between 2012 and 2016 50% of ships totally lost did this because of weather. Some 20% grounded, 10% was lost because of fire or explosion, 5% by collision, and 10% by machine failure. (Total Losses, all vessel types over 500 Gross Tons – IUMI 2017)

As we can note from the above, there is no mentioning of any losses due to "human error". This is because the statistics often chose a single, simple cause of the accident, but if we drill down looking for a root cause we often find "human error" on one level or another in almost all cases. Dhillon (2007) compiled the following statistics:

A study of 6091 major accident claims associated with all classes of commercial ships, revealed that 62% of the claims were attributable to "human error".

"Human error" contributes to 84–88% of tanker accidents.

"Human error" contributes to 79% of towing vessel groundings.

Over 80% of marine accidents are caused or influenced by human and organization factors.

"Human error" contributes to 89–96% of ship collisions.

A Dutch study of 100 marine casualties found that "human error" contributed to 96 of the 100 accidents. (For detailed references see Dhillon 2007, p. 2)

Let us illustrate how "human error" can be a part of almost all accidents. Let us briefly look at the recent collision accident between the general cargo ship *Daroja* and the oil bunker barge *Erin Wood* that took place in Scottish waters in 2015 (MAIB 2016). In August 2015 the two vessels collided off the east coast of Scotland. It was a nice summer afternoon with light wind and no sea state. The two vessels were both north bound but with crossing courses which brought them closer and closer together for almost two hours without any one of the two bridge officers apparently noticing the other ship until too late. Visibility was excellent, radar and AIS tracking was available on both bridges. The UK Maritime Accident Investigation Board concluded that "*Daroja* and *Erin Wood* collided because a proper lookout was not being kept on either vessel." (MAIB 2016, p. 40) This accident would appear in the aforementioned statistics as a "collision", but the underlying root cause was "improper lookout", which would classify it as "human error".

A variety of taxonomies for "human error" has been proposed. One example is the simple dichotomy between "errors of omission" and "errors of commission" (Wickens et al., 2013). "Errors of omission" mean: not doing anything when something should have been done, as the watch keepers above. "Error of commission", on the other hand, means: doing the wrong thing.

A more elaborated taxonomy developed by Norman (1988) and Reason (1990) involves "mistakes," "slips" and "lapses."

"Mistakes," are when the operator has not fully understood the situation and acts intentionally.

"Slips," on the other hand, are when the intention is right but the action is carried out wrong.

Maybe the wrong button is pressed although the intention was to press the right one. Because humans monitor their own actions, slips are often noticed and corrected before any harm has been done.

"Lapses," finally, are a failure of making any action at all, i.e. an error of omission. Often they are lapses of memory, forgetfulness. Humans forget, we become distracted or think about other things. This is all part of the human condition. Maybe the two watch keepers in the accident above was thinking about other things and forgot to monitor their systems and look out of the window? "Lapses" are sometimes easy to prevent by technical solutions like automation.

One may ask how come there was no warning issued to make the two watch officers aware of the pending danger. Radar systems on both ships as well as the AIS tracks in the electronic chart systems could theoretically extrapolate the courses of the vessels to a collision point. In addition, systems on land that gather AIS data could have made the same calculation. Why is it that available data is not used to the benefit of safety when possible? Why was there no warning and why did not the systems automatically make a small course or speed change to stay out of the close quarter situation? It is because automation is a controversial issue. Warnings are often turned off by operators, because of many false alarms.

3.2 *Why automation can make ships safer*

A large part of the robustness of the shipping industry demonstrated by the constant decline in shipping accidents has to do with automation. The error prone and difficult position fixing, previously done by manual methods like dead reckoning, or sun heights and bearings to landmarks, when sun, stars and land was in sight, has now been replaced by satellite based navigation systems with very high reliability. Manual steering which in old days caused large course errors has been replaced by auto pilots or even track pilots which can follow a pre-programmed path with an accuracy of a few meters- or even centimetres when augmentation systems are used. Just to mention a few areas of marine automation.

The reason automation is safer is that they address human shortcomings like:

Fatigue: Humans are day animals. We are designed to be active by day and sleep by night. Our whole cognitive system is designed for work by day. Even if augmented by technical means, our decision making is crippled during night, even if we are accustomed to shift work by night. A larger degree of accidents happen during night. (e.g. Wagstaff & Sigstad Lie 2011)

Attention span: The ability to focus and sustain attention on a task is crucial for the achievement of one's goals. Although *attention span* is a complex concept and measures depend on a lot of different thing, most researchers agree that the time span humans need to concentrate to handle tasks without being distracted is limited, e.g. 10–20 minutes in healthy teenagers and adults (Wilson & Korn 2017).

Information overload: Overload can be of many kinds. Too much to do, and too little time to do it. Too much information that needs to be considered presented in an unintegrated way at the same time. It boils down to limits of the human *working memory*. Miller in 1956 famously stated that humans at the most could handle 5–9 information chunks at one time. But, *underload* can also be a problem. During a conference in 2014 a British maritime accident investigator mentioned a new type of *boredom-induced accidents*. Evidence of the so-called Yerkes-Dodson law (first proved on mice in 1908) show that human performance describes an inverted U-shaped curve when plotted against arousal (or stress) so as low arousal also may lead to low performance and elevated arousal lead to higher performance to a certain point when performance declines with higher stress (cognitive tunnelling).

Normality bias: This is a form of denial 70% humans revert to when facing events of disaster, as a result of which they underestimate the possibility of the disaster actually happening and its potential results (Omer & Alon 1994).

We could go on stating human shortcomings in this way for many pages, however we think the point is made: automation can make ships safer.

3.3 *Why automation can make ships less safe*

In the everlasting strive to make life easier, humans have automated tasks that are tedious, dangerous, dirty, boring, etc. However, a paradox in automation is that it has often been the easiest tasks that has been possible to automate. In complex and ambiguous situation, the human has had to step in to resolve the ambiguity and finish the task.

Automation needs to be programmed and can therefore only solve simple or *complicated* problems. By "complicated", we here mean that there is a finite solution space that can be parsed by computers. In reality, many real world problems are *complex* in the sense that they have an infinite solution space due to many unknown factors and interrelationships. For such problems, it is not even theoretically possible to program to solve all possible situations (possibly leaving machine or deep learning aside).

The dynamic maritime environment with sea and current, weather, topography, manned and autonomous ships is such a complex environment and will for a very long time need a human to step in and resolve problems out of the range of automation. As we have seen above, there is relatively good statistics on "human error", however there are almost no statistics on "human recoveries", where humans has stepped in and saved a situation caused by e.g. technical malfunction.

An illustration of such a recovery can be fetched from an incident in1991.

In this incident a product tanker loaded with 20 000 metric tons of gasoil was under way through the narrows of a winding Scandinavian archipelago. In a bend in the fairway she had a routine meeting with one of the large ferries trafficking the area. The ferry had almost 1000 passengers and crew onboard. As the tanker applied starboard rudder to negotiate the bend in the fairway, the captain noticed that the rudder instead turned to port and a port turn was commenced a few hundred meters in front of the oncoming ferry. The captain immediately reversed the engine, but realizing that he would not be able to prevent the turn, he called the ferry on the VHF saying they had a breakdown on the steering engine and asked for "green-to-green" (starboard side to starboard side) meeting. The ferry responded promptly, but by making a starboard 360 degree turn and the ships passed each other on parallel courses with 20–30 meter between. The accident investigation board calculated that if the action from the ferry had been delayed 30–60 seconds a collision with the ferry running into the amidships section of the tanker in a right angle would have been impossible to avoid (SHK 1992). The consequences can only be imagined.

The accident investigation concludes that it was the decisive actions by the captains of the two ships that avoided a possible catastrophe. One may wonder what would have happened if one or both of the ships had been autonomous. Remember also the pilot of the airliner that landed on Hudson River in 2009, and who, by acting against protocol and procedures, miraculously saved the lives of passengers onboard (NTSB 2010).So, on one hand we have incidents due to human error that can be avoided with automation, on the other hand we have incidents that is now avoided with humans, but will happen when no humans are onboard. But new technology also opens for new types of accidents.

These relationships are described in Figure 3.

Automation of human processes (middle circle, Figure 3) are expected to significantly reduce the number of incidents happening in shipping today, but one must also assume that a number of

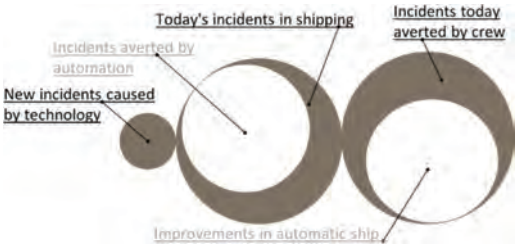


Figure 3. Remaining incidents in the autonomous ship after automating human processes.

potential incidents are averted by the crew's actions and it is not clear if improved automation can match these numbers. Finally, one must also assume that some new types of incidents will occur as a result of the introduction of new technology (far left). The net result is the remaining grey areas and the question is if this will be low enough for societal acceptance of the new ship types.

Thus, while the assumption is that the net result of automation will be lesser accidents and incidents, this remains to be shown. Within commercial air industry, automation has improved safety, (e.g. Billings 1997; Pritchett 2009; Wiener 1988). Can we assume that the same is true for the shipping domain? One way of dealing with this is through *risk analysis*.

3.4 *Risk analysis*

Risk analysis can be "broadly defined to include risk assessment, risk characterization, risk communication, risk management, and policy relating to risk, and risks of concern to individuals, to public- and private-sector organizations, and to society at a local, regional, national, or global level" (SRA 2012). In this paper's context, we look at risk analysis as risk assessment where risk is defined as the combination of the frequency and the severity of the outcome of an accident (IMO 2002).

The expected frequency of accidents must often be derived from an assumed accident probability, as statistical significant data on frequencies are impossible to find. Obviously, this particularly applies to new technology or ship types as in autonomous ships. The probabilities are difficult to determine in themselves and, in addition, the strength of knowledge used to establish the probabilities need to be addressed. In autonomous systems the strength of knowledge is generally low due to lack of experience and the complexity of the autonomous marine system.

The prevalent strategy to the increased (socio-technical) complexity, lack of coherence, and speed of change in contemporary systems, science and

the discipline of risk management, is to incorporate uncertainty, ambiguity, and the knowledge dimension per se in the risk measure (Paltrinieri et al. 2016). This is done through risk analysis of potential accident scenarios that we eventually are aware off and can manage. This is emergent research and there is not much hard knowledge in the area, although some papers have been published, e.g. (Utne et al. 2017) and (Rødseth & Tjora 2014).

The second paper is mainly a preliminary hazard identification (HazId) study based on use cases and ship function breakdowns. It suggests a framework for doing HazId in the unknown environment of the autonomous ship based on assumptions on what can happen and how this influences on the different functions the ship systems have to provide. The first paper argues for a more holistic approach to risk management, including dynamic risk assessments during the autonomous voyage.

This paper will not go further into this area, but it is important to point out that determining the complete risk level for the autonomous ship will be very challenging. As was illustrated in Figure 3, there are more new issues that have to be taken into consideration and for at least two of these we do not have any statistics that can be used in estimates of probabilities. Although, e.g. HazId may be able to identify the hazards and accident consequences, we are still left with very uncertain probabilities and the limitation to the known knowns and known unknowns.

Within safety science, the concept of "human error" are seldom used after 1990's since it has been seen that "human error" is not a cause but a result of other factors such as poor design, poor planning, poor procedures, etc. (Dekker 2006). Instead the concept of "human variabity" from Resilience Theory is often used (Hollnagel, Woods & Leveson 2006). Human variability that sometimes might lead to "human errors" but maybe more often to "miraculous recovery". Positive actions and successful recoveries are usually not recorded, as mentioned in Leveson (1995, p. 94); where an U.S. Air Force study showed 659 crew recoveries in 681 inflight emergencies; with only 10 pilot errors.

4 CONCLUSION

It seems to be generally accepted that automation has the potential to decrease accidents that are due to human variability.

However, automation has the potential of creating accidents in itself, e.g. through transitions between automatic and manual control and the human having to rapidly assess the situation and make the right decisions.

Automation also sometimes creates problems by reducing the work load of the human, inducing boredom and by that further increasing the time needed to do a correct assessment.

With constrained autonomy being the most likely form of ship autonomy, one needs to investigate if these issues actually can increase the probability of some accident types compared to conventional manned ships.

Also, autonomy will create new types of accidents, as suggested by the illustration in the beginning of the paper This is partly due to accidents that was before averted by the human crew and partly due to introduction of new technology and corresponding new accident types. These types of accidents are very challenging to include in the risk analysis as we lack statistical evidence for their probability.

To address the new risk picture, one probably need new types and extensive use of human centred risk analysis. Also, one needs to consider the development and use of dynamic risk assessment systems during autonomous voyages, as well as other real time tools that can be used on the ship or in the shore control centre.

ACKNOWLEDGEMENT

This work is supported by the SAREPTA (Safety, autonomy, remote control and operations of industrial transport systems) project, which is financed by Norwegian Research Council with Grant No. 267860.

REFERENCES

Bilings, C.E. 1997. *Aviation automation: the search for a human-centered approach.* Mahwah, N.J.: Lawrence Erlbaum Associates Publishers.

Crosbie, J.W. 2006. Lookout Versus Lights: Some Sidelights on the Dark History of Navigation Lights. *The Journal of Navigation*, 59, 1–7.

Dekker, S. 2009. *The field guide to human error*. Bedford, UK: CRC Press.

Dhillon, B.S. 2007. *Human Reliability and Error in Transportation Systems*. London: Springer.

Hollnagel, E., Woods, D.D. & Leveson, N.C. (eds.) 2006. *Resilience engineering: Concepts and precepts*. Aldershot, UK: Ashgate.

IMO MSC/Circ.102/MEPC/Circ.392. 2002. *Guidelines for Formal Safety Assessment (FSA) for use in the IMO Rule-Making Process. As amended.* London: IMO.

Kongsberg Maritime. 2017. Autonomous ship project, key facts about YARA Birkeland https://www.km.kongsberg.com/ks/web/nokbg0240.nsf/AllWeb/4B8113B707 A50 A4FC125811D00407045?OpenDocument. [Accessed 2017–11–14].

Lambert, F. 2016. Understanding the fatal Tesla accident on Autopilot and the NHTSA probe. https://electrek.co/2016/07/01/understanding-fatal-tesla-accident-autopilot-nhtsa-probe/ [Accessed 2017–12–14].

Miller, G. 1956. The Magical Number Seven, Plus or Minus Two: Some limits on our capacity for processing information. *Psychological Review*. 63 (2): 81–97.

Norman, D.A. 1988. *The Psychology of Everyday Things*. New York: Basic Books.

NTSB, National Transport Safety Board. 2010. Aircraft Accident Report: Loss of Thrust in Both Engines After Encountering a Flock of Birds and Subsequent Ditching on the Hudson River: US Airways Flight 1549. https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1003.pdf [Accessed 2018–02–09].

NTSB, National Transport Safety Board. 2017. Collision Between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida May 7, 2016. NTSB/HAR-17/02, PB2017-102600.

Omer, H. & Alon, N. 1994. The continuity principle: A unified approach to disaster and trauma. *American Journal of Community Psychology*, 22: 273.

Paltrinieri, N. & Khan, F. (eds.). 2016. Dynamic Risk Analysis in the Chemical and Petroleum Industry: Evolution and Interaction with Parallel Disciplines in the Perspective of Industrial Application. Butterworth-Heinemann.

Pritchett, A.R. 2009. Aviation Automation: General Perspectives and Specific Guidance for the Design of Modes and Alerts. *Reviews of Human Factors and Ergonomics. Vol 5*, Issue 1, 2009.

Reason, J. 1990. *Human error*. Cambridge University Press.

Rødseth ØJ. & Nordahl H. (eds.). 2017. Definition for autonomous merchant ships. Version 1.0, October 10. 2017. Norwegian Forum for Autonomous Ships. http://nfas.autonomous-ship.org/resources-en.html. [Accessed 2017–12–10].

Rødseth, Ø.J. 2017. Towards Shipping 4.0. Proceedings of Smart Ship Technology. Royal Institution of Naval Architects. ISBN 978-1-909024-63-2.

Rødseth, Ø.J., & Tjora, A. 2014. A risk based approach to the design of unmanned ship control systems. *Maritime-Port Technology and Development,* 2014.

SAE. 2016. SAE J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems, Revision September 2016, SAE International.

SHK, Statens Haverikommision. 1992. Near collision between MT Tarnfjord and RoPax Wellamo. Statens Haverikommision, RAPPORT S 1992:1 Ärende S-06/91. Stockholm: SHK.

Utne, I.B., Sørensen, A.J., & Schjølberg, I. 2017. Risk Management of Autonomous Marine Systems and Operations. In ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering. American Society of Mechanical Engineers.

Vagia, M., Transeth, A.A. & Fjerdingen, S.A. 2016. A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied ergonomics* 53: 190–202.

Wagstaff, A.S. & Sigstad Lie J.A. 2011. Shift and night work and long working hours – a systematic review of safety implications. *Scandinavian Journal of Work Environment and Health,* 2011;37(3):173–185.

Wickens, C.D., Hollands, J.G., Banbury, S. & Parasuraman, R. 2013. *Engineering Psychology and Human Performance* (3rd ed.). New York: Pearson.

Wiener, E.L. 1988. Cockpit automation. *Human factors in aviation* (A89–34431 14–54). San Diego, CA: Academic Press, p. 433–461.

Wilson, K. & Korn, J.H. 2007. Attention During Lectures: Beyond Ten Minutes. *Teaching of Psychology*. 34 (2): 85–89.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A computerized procedure system framework for U.S. utilities

R. Lew
University of Idaho, Idaho, USA

R.L. Boring & T.A. Ulrich
Idaho National Laboratory, Idaho, USA

ABSTRACT: The United States has a fleet of light water reactors that continue to operate using paper-based procedures. Here we review existing implementations of computerized procedure systems, the potential benefits, as well as their caveats. We also review U.S. regulatory requirements for computerized procedure systems in an effort to identify barriers to adoption. We conjecture that process control procedures, especially the formalized procedures used by U.S. utilities, can be viewed as E-Type Systems from a software engineering evolution perspective. After presenting this argument, we discuss corollaries from treating procedures as E-Type Systems. Based on these analyses, a strategy was formulated to aid utilities in transitioning to computerized procedure systems. And lastly, missing tooling to aid utilities in cost-effective, timely transitions is discussed.

1 INTRODUCTION

The U.S. Department of Energy's (DOE) Gateway for Accelerated Innovation in Nuclear (GAIN) program partners nuclear technology industry partners with DOE expertise and resources. As part of GAIN, Idaho National Laboratory is teaming with the simulator vendor GSE Systems to develop a Computer Based Procedure (CBP) Engine for Nuclear Power Plant (NPP) Main Control Room (MCR) operations. This platform would enable critical research related to CBPs with full-scope simulators specifically to support paper to digital procedure transitions for U.S. utilities seeking to realize the benefits of CBPs.

The aging U.S. fleet of light water nuclear power plants (NPPs) is of paramount importance to the nation's overall base energy production capabilities. Each nuclear power plant has been operating for multiple decades and has already undergone numerous engineering changes. Approximately 74 operating plants have filed license extensions to operate beyond their original 40 year operating license. To operate beyond 40 years, many plants are undergoing modernization efforts of ensure the plants can be operated in a safe and cost effective manner. These changes result in frequent amendments to procedures. The procedures have evolved substantially from when the plant was first commissioned. The U.S. nuclear industry, in addition to being highly proceduralized with operations and maintenance in and out of the control room, is also one of the most stringently regulated energy industries.

Plants were originally commissioned with paper procedures and, despite their shortcomings, have stood the test of time and are still in use today. It is well documented and often cited that human reliability and error are the leading cause for plant events. When the contributors of human error are examined, a U.S. Nuclear Regulatory Commission report (1995) reveals that procedure problems have been cited as contributing to 69% of event reports. Following paper procedures is challenging because:

- Operators must identify the correct procedure and have the most up-to-date version of the procedure. The use of paper procedures can result in an excess number of procedures, often with poor classification schemes (O'Hara et al., 2000).
- Operators must determine the correct path through the procedure and complete the procedure without skipping steps or following steps out of order. Operators must manually keep their place in procedures (Fink et al., 2009).
- Procedure following imposes high memory demands on operators because of the need to track multiple pieces of information tied to disparate pieces of equipment.
- In emergency events, each operator might be performing multiple nested procedures simultaneously. It is in these stressful environments that humans have a high potential for decision making errors, such as making incorrect conclusions (Converse, 1995; Kim, et al. 2011).
- Operators must determine plant state from control board indications or approved (qualified) plant computer systems. This requires frequent searching for information or requests for information from reactor operators (ROs). ROs must determine

the correct instrumentation and control (I&C), and correctly perceive and report the correct value. Senior reactor operators (SROs) often are completely reliant on the RO for critical pieces of information.
- Procedures also require continuous actions to monitor plant variables, and operators must take action if a specified threshold is reached. These continuous actions divide the operator's attention across the board.
- Decisions must be based on the current plant state. The plant state changes as operators progress through a procedure. Normal operating conditions can be different depending on the plant state.
- Procedure following requires engaging in critical actions within specified time intervals. Operators often must keep track of several time intervals simultaneously.

Computer-based systems for nuclear power are precedented with over 30 years of research and multiple use cases. Several systems have shown that computerized procedures can aid operators with many of these challenges as illustrated in the following section.

2 RESEARCH AND IMPLEMENTATIONS OF COMPUTERIZED PROCEDURE SYSTEMS

2.1 *Korea advanced reactors*

KOPEC Nuclear Power operations for Shin Kori Units 3 and 4, which are APR1400 units (Unit 4 is not yet online), rely heavily on computerized procedure systems in a completely different paradigm from conventional operations (Hong, Lee, and Hwang, 2009). With conventional (U.S.) operations the operator follows a paper procedure and references the boards to assess the current state of the plant. Computerized procedures allow the relevant information to be provided within the context of the procedure. However, with individual workstations displaying CBPs, operators communicate less with each other, leading to poorer team situation awareness (Kim, et al. 2011). Operations becomes much more centric to the computerized procedure system, and many of the inefficiencies associated with seeking information and verbally communicating between operators are removed.

With traditional procedures, information is coupled primarily through the reactor operators. With CBPs the communication between operators is loosely coupled.

The Korea Electric Power Research Institute has developed a computerized procedure system that in addition to displaying and executing procedures has functionality for writing and editing procedures, and assessing the validity of logic and plant parameters through a testing framework that integrates with an engineering framework (Hong, Lee, and Hwang, 2009).

2.2 *The OECD Halden reactor project*

The Halden Man-Machine Laboratory (HAMMLAB) of the OECD Halden Reactor Project (HRP) has conducted pioneering research related to computerized systems for nuclear control. In 1985, work on developing computerized procedures began, which culminated in the Computerized Procedure Manual (COPMA) and later COPMA-II. A 1995 experimental study (Converse) found that operators committed half as many errors with COPMA-II compared to using paper procedures. This is even more significant when considering that the participants had only a few hours of training and experience using COPMA-II, although it should be noted that crews responded faster with paper procedures. In 2000, the third iteration of COPMA, based on web technologies, was released (IFE, 2017).

While OECD Halden Reactor Project is perhaps most well-known for ecological interface design and information rich displays, much effort has focused on how task-based displays can be designed to convey the most important procedure relevant information, as well as identifying how task-based displays can integrate procedures with process displays (Braseth, et al., 2009). This philosophy is advantageous because the layout of the information is kept in a standard format and the procedure is simultaneously visible to all operators, resulting in an improved shared situation awareness.

2.3 *Westinghouse*

Westinghouse Electric Corporation began developing CBPs for commercial applications in the early 1990s. Their COMPRO system was highlighted as having a graphical user interface and using relational database technologies. COMPRO has been deployed in Beznau, Switzerland, and Temelin, Czech Republic. Westinghouse's AP1000 will incorporate a computerized procedure system that will also provide a diverse set of procedure views including graphical flowcharts, a textual view and a dynamic logical view (Lipner, Mundy, Franusich, 2006) while retaining much of the COMPRO "DNA" in philosophy.

In the early 1990s Westinghouse collaborated with Army/NASA Ames Research Center to apply the Man-machine Integrated Design and Analysis System (MIDAS), a tool for computerizing the cockpits of advanced commercial and military aircraft, to a computer-based procedure system for main control room operations. This collaboration resulted in an impressive proof-of-concept demonstration of how CBPs could substantially reduce operator memory demands and improve situational awareness (Hoecker, et al. 1994). The caveat being that incorporating MIDAS to translate PBPs to CBPs is a fairly significant refactoring requiring detailed and extensive collaboration between operations and human factors practitioners.

2.4 *EdF N4 series reactors*

The Électricité de France (EdF) N4 series of reactors in France use CBPs. Dien, Montmayeul, and Beltranda (1991) describe the philosophy of developing the N4's original computerized procedures from a human factors perspective. They are careful to note the computerized procedures are not intended to guide the operator actions and the operators should always have the final decision on the actions that are performed in the plant. In operations, they recognized the necessity for operators to compensate for inadequacies in procedures, as well as handling conflicting or ambiguous conflicts between procedures. Procedures can also result in operators looping through a procedure multiple times before finding the correct indication; the EdF approach recognizes that human operators may be able to implement more economical or practical control strategies than those prescribed by procedures.

3 U.S. ADOPTION OF COMPUTERIZED PROCEDURES

Thus far, we have reviewed existing implementations of computer-based procedures and procedural systems. Despite several decades of commercial implementations and several new systems in modernized plants, adoption of CBPs in U.S. main control rooms is almost non-existent in spite of the fact procedure related problems have been implicated in 69% of event reports.

Le Blanc, Oxstand, and Waicosky (2012) surveyed U.S. nuclear utilities about their plans for implementing CBPs in the field, and the perceived barriers of implementation. Utilities are hesitant to be first. Being first in the industry usually entails higher costs. Utilities also need to justify the capital investment, and to do so CBPs need to be demonstrably better than paper-based counterparts.

While there are mixed results, the available evidence suggests that tangible benefits exist:

- Lin et al. (2016) found with a full-scope advanced Main Control Room simulator that operators with CBPs had significantly better task monitoring awareness, and higher task performance while operators completed emergency response procedures. Their results suggest that reduced communication between team members is a side effect of having higher situation awareness. Team members were observed to make verbal inquiries when their situation awareness was low.
- CBPs offload some of the complexity of procedure following, leading to fewer procedure deviations.
- CBPs can be organized by task so that switching between procedures is seamless (Le Blanc, Oxstrand, and Joe, 2015).
- Some of the work performed by operators can be offloaded to a CBP system, e.g. continuous monitoring of a plant parameter.
- Recovery from human error is faster with CBPs compared to paper (O'Hara, 2000).
- CBPs are even more beneficial with multiple failures.
- Dynamic presentation of information is possible. The information does not need to adhere to the standard *one sensor, one-indicator* format. Data can be tailored by clustering and aggregation, through summaries, and graphical presentations (Dien, Montmayeul, and Beltranda, 2009).
- Tasks that would normally require multiple procedures can be combined and ordered in a logical sequence.
- Visual representations of procedural flow paths can result in better awareness of the plant's current state and were the procedure is headed (NRC, 1995).

Why then, is there not greater adoption of CBPs, and how can adoption be encouraged? This question is difficult to answer with the available information, but we can provide some conjecture on the issue. Procedures, to state the obvious, tie to every sub-system of the plant. Undertaking an adoption of a CBP system is a large commitment from an organizational perspective. The benefits of a successful transition from paper to CBPs is clear, but plants also run the risk of disrupting the functional aspects of current procedures and procedural systems. The existing methodologies require completely disassembling and refactoring existing procedures into computerized procedure systems. This might work for new plants and new reactor designs, but could be seen as risky for the currently operating Generation II plants. Procedures are living documents that have been continually maintained over their multi-decade lifespans. Plants are modernizing control systems, and procedures must be amended to stay accurate with plant systems. They contain human years of operating experience and implicit functional characteristics that could be lost if haphazardly transferred to computer-based procedure systems.

4 SYSTEMS OF PROCEDURES

Here we suggest that process control procedures can be viewed from a software engineering perspective. The logic of a procedure is analogous to computer code. Instead of the code being compiled by a computer program, the code is interpreted and executed by a human. Lehman (1980) wrote a seminal article on software engineering evolution. Lehman describes three classes of software programs. *S-Type Systems* are programs whose function is defined by a specification. Programs in this category are provably correct if they meet the specification. The logic

blocks within a conventional distributed control system (DCS) is a prime example of S-Type Systems.

The second class of programs are *P-Type Systems* focusing on real world Problem solutions. P-Type Systems incorporate feedback loops comparing the measured state of the world to desired outcomes. Control systems are P-Type Systems. A controller may monitor a single or multiple parameters and produce one or more control signals to change the state of the system. A vendor will provide a detailed specification of a DCS's inputs and outputs along with descriptions of how the inputs should map to the outputs. The functionality of the DCS can be verified over the range of anticipated operating conditions. However, system control and safety outside of normal operating conditions is intractable. Efforts can be made to make systems more robust through strict quality control, redundancy, incorporation of passive safety, and more resilient through advanced control schemes and artificial intelligence, but ultimately closed form solutions for every possible contingency will not exist. Arguably this is why it is still a good idea to have highly trained and knowledgeable human operators in the control room and human eyes and ears about the plant. Humans can be resilient when hardware and control systems are not.

Lastly, Lehman describes *E-Type Systems* or embedded programs. Embedded programs are named so because they are part of the overall system. E-Type Systems are described as mechanizing a human or societal activity. Human decision making inherently involves some level of prediction as well as uncertainty, intuition, opinion, and judgement. The validity of E-Type Systems is in the human domain, on whether human assessments of its effectiveness for the intended application.

Here we suggest that the procedures used to control nuclear power plant processes are E-Type Systems and as such should be treated distinctly from the S and P countertypes. First, the human operators, maintenance workers, and engineers are the embedded aspect of energy production process. A plant on its own would be incapable of producing power and as such comprises a sociotechnical system. Even if the plant configuration remained static, the constraints, demands, and requirements are constantly evolving. For example, plants incur high downtime costs and are always striving for increasing their capacity factor. Second, the proliferation of renewables presents load shaping requirements for plants. Third, an aging workforce presents challenges to preserving worker knowledge and skills needed to maintain legacy systems. Fourth, consider events like Fukushima Daiichi, or Ukrainian cyber-events and the influence they have on operations, upgrades, and maintenance.

The distinction between S/P-Type Systems and E-Type Systems is important because they convey distinct requirements for their respective life-cycles. A control block that meets its functional specification can operate for decades with virtually no maintenance. E-Type Systems on the other hand are inherently change prone. Operations and procedures change because of the drivers that from a pure control systems perspective are purely exogenous. Procedures need to be adaptable and computer procedure systems need to support the authoring and continued maintenance of procedures as much as the actual running of procedures. If they do not change, Lehman suggests, their utility diminishes over time because their environment is evolving. To remain useful, they must change and have a tendency to become more complex, and Lehman postulates a law of "Continuing Change" to describe this phenomena.

Unfortunately, continual change is also associated with increased complexity (2nd law). Anecdotally, paper based procedures are becoming more complex to fill administrative gaps and remain technically accurate as systems change. Lehman's 6th law of continuing growth describes how the functional content of an E-type system must continually increase to maintain user satisfaction. As an example, procedures have become embedded with notable user experience, which is a function never originally intended for procedures. While the increased complexity captures relevant process related information it also increases the cognitive difficulty of carrying out procedures.

From this perceptive, computer-based procedures could be vital in managing the increasing complexity associated with E-Type Systems. Lehman's 7th law predicts declining quality if an E-Type system is not rigorously maintained and adapted to changing operational environments.

If procedures are more akin to E-Type systems than P-Type systems, then maintaining the ability of procedures to *evolve,* to adapt and continuously modify procedures is of paramount importance. By the conclusion of this document we formulate a strategy to aid utilities in transitioning from paper to CBP systems. The previous survey of computerized procedures suggests that the technology exists, and it is likely that the hurdles to adoption are not technologically based. In the context of modernization, computerized procedure systems are likely not expensive relative to costs associated with modernizing instrumentation and controls. If not technical or financial, then perhaps regulatory considerations are the barrier to entry. In the next section we briefly review regulatory considerations for CBPs.

5 U.S. NUCLEAR REGULATORY COMMISION GUIDANCE

Early work by the U.S. Nuclear Regulatory Commission (NRC) concluded that CBPs were a desirable

goal but the implementation and adoption needed to justify the use of CBPs over paper procedures (O'Hara et al., 2000).

NUREG/CR-6634 provides a detailed technical basis for CBP systems. The technical basis was developed by deconstructing the hierarchical influence of human behavior on plant performance in order to identify how human errors cause unsafe circumstances. NUREG/CR-6634 expresses some valid concern regarding whether CBPs could lead to operators becoming disoriented amongst tabs, or losing situation awareness due to keyhole attention affects, or out-of-the-loop loss of situation awareness. Automation of control actions can also be detrimental because it increases task complexity and can add confusion to diagnosing the root cause of a disturbance (Andresen, et al. 2003). CBPs can use plant process data to guide operators; however, if these systems are disrupted, the CBP may not function, or even provide incorrect assessments and guidance. Operators may become too trusting of the guidance provided by a CBP and induce complacently to follow CBP directions.

The Electric Power Research Institute (EPRI) has guidance for developing Utility Requirements Documents. According to O'Hara (2000), the guidance is based on EdF's CBP experience. EPRI's high level guidance covers human factors criteria describing the allowable presentation formats for procedures, as well as stipulating operators should have final authority in regard to control actions. It also has technical requirements for CBPs to verify operator decisions, provide logging capabilities, and redundancy in the case of loss of CBPs. The EPRI requirements include a validation of each operating procedure using the plant's simulator and performance model. Plants would also need to make sure that alternative procedures are available in the event of loss of CBPs. Utility requirement documents would need to validate and verify CBPs with plant simulations.

The NRC does not consider paper procedures to be a Human System Interface (HSI); however, computerized procedures would be displayed through an HSI would need to meet the human factors requirements of NUREG-0700 as well as the NUREG/CR–6634. For first adopters, the CBP would be considered an *unproven* HSI technology until there has been at least three years of documented satisfactory service in a light water reactor (O'Hara 2000).

6 CATEGORIES OF COMPUTER BASED PROCEDURES

IEEE Standard 1789–2011 describes three types of CBPs along a continuum of increased automation. The least automated are Type 1 and are sometimes referred to as E-procedures in the literature. These are electronic versions of standard PBPs. They provide the ability to view the procedure in an electronic format and provide navigation links between procedures. The second category is Type 2 procedures. These link to plant data in real time to provide process relevant information embedded in the procedure, perform logical assertions based on the procedure and available data, and provide links to process displays and soft-controls that reside in separate systems. Lastly, Type 3 procedures have embedded soft controls and procedure-based automation that carries out a full procedure or several sub-steps of a procedure.

7 STRATEGY FOR TRANSITIONING PAPER PROCEDURES TO COMPUTERIZED PROCEDURE SYSTEMS

Based on the identified constraints and evolving nature of procedures it may be strategically advantageous to transition to a computer based format that is similar in organization and layout to existing procedures. The goal is to provide a process to transition the existing corpus of PBPs into a computerized support system without substantively changing the organization, or format and layout of the procedures. The strategy is to avoid major refactoring that is costly, time consuming, and also has the potential of losing critical information contained within existing PBPs. Plants have finite human resources to put towards the numerous modernization efforts. Modernization schedules can easily run a decade or more into the future. The industry has a need for innovative solutions that can help change occur faster, and our thinking is centered on what can be done in a time effective manner, while offering a high value proposition, and satisfying or exceeding regulatory constraints. Such a strategy would avoid having to completely dissemble, reorganize, and refactor existing procedures, as well as ease the process of transcribing to computerized representations. Maintaining resemblances to paper countertypes also eases training and verification and validation activities necessary for regulation. Therefore, a solution should maintain a textual presentation in the conventional two column format.

Metaphorically this approach would be described as not shooting for the moon, but rather planning a course towards the moon. Based on our proposition that procedures are E-Type Systems it is vitally important that utilities maintain the ability to quickly and adeptly author and edit procedures without continued support from a vendor.

Strategically translating procedures should be thought of as an editing process instead of a complete refactoring of how procedures are performed and constructed. The editing process would likely be iterative, and would contain the following steps:

1. Multidisciplinary review of the existing procedures to identify and remedy shortcomings in the original procedure to prevent them from getting carried forward to a computerized format
2. Provide Type 2 computerized procedure enhancements like live plant parameters, monitoring continuous actions (Choi and Park, 2012), decision support aids for integrating multiple variables, and real-time trends for predefined sets of parameters.
3. Software engineering routinely uses code quality metrics to assess portions of code that may be ambiguous or difficult to interpret. Once a plant's procedures are represented in a form interpretable by a computerized procedure system it becomes possible to develop automated routines for finding indictors of poor quality procedures based on step complexity (Park, Kim, and Ha, 2003) or human reliability analysis performance shaping factors and unique failure modes relevant to computerized procedures (Boring et al., 2011).
4. Technical review of information in the procedure.
5. Plant simulator validation of the procedure.

None of the procedure authoring systems are intended to ease the translation from existing paper implementations. Here we have identified a gap in existing technology capabilities that we aim to fill. We envision a what-you-see-is-what-you-mean (WYSIWYM) authoring system based on Markdown and Jinja templating. Jinja templating allows for context processors to be added to a backend that would, at rendering, insert of live values or handle plant dependencies logically, branching when they are rendered on the display. Most importantly, the representation in the procedure (template) would have simple syntax expressions.

We are currently developing a prototype computerized procedure system intended to conduct the full-scope simulator research needed to validate the engine and the strategy. The goal would be to transition the research framework to a commercial platform for computerized procedures.

REFERENCES

Andresen. G, Braarud, P.Ø. Bye, A., Øwre, F. (2003). Experiments in the Halden Human-Machine Laboratory (HAMMLAB) investigating human reliability issues. *Eurosafe*.

Boring, R.L., Gertman, D.I., and Le Blanc, K. (2011). Human reliability analysis for computerized procedures. *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting*, 1720–1724.

Braseth, A. O, Wing, C.N., Svengren, H., Veland, Ø, Hurlen, L. Kvalem, J. (2009). Lessons learned from Halden Project research on human system interfaces. *Nuclear Engineering and Technology, 41* (3), 215–224.

Converse, S.A. (1995). Evaluation of the Computerized Procedures Manual II (COPMA II). NURGE/CR-6398.

Choi, S.Y., Park, J. (2012). Operator behaviors observed in following emergency operating procedures under a simulated emergency. *Nuclear Engineering and Technology, 44* (4), 379–386.

Dien, Y., Montmayel, R., Beltranda G. (1991). Allowing for Human Factors in Computerized Procedure Design. *Proceedings of the Human Factors and Ergonomics Society 35th Annual Meeting.*

Fink, R., Killian, C., Hanes, L., & Naser, J. (2009). Guidelines for the design and implementation of computerized procedures. *Nuclear News, 52*(3), 85–88, 90.

Hoecker, D., Corker, K, Roth, E., Lipner, M, & Bunzo, M. (1994). Man-machine design and analysis system (MIDAS) applied to a computer-based procedure-aiding system, *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting,* 195–199.

Hong, J.-H., Lee, M.-S., Hwang, D.-H. (2009). Computerized procedure system for the APR1400 simulator. *Nuclear Engineering and Design 239* (12), 3092–3104.

Institute for Energy Technologies (2017). Development of COPMA within the Halden project. Retrieved from https://www.ife.no/en/ife/departments/system-and-interface-design/projects/copma/copmafiles/copma-history.

Kim, J., Kim, J., Park J., Jang, S.C., Chin, Y.C. (2011). Some empirical insights on diagnostic performance of the operating crew in a computer-based-advanced control room. *Human Factors and Ergonomics in Manufacturing 21* (4), 379–396.

Le Blanc, K. Oxstrand, J., Joe, J. (2015). Requirements for control room computer-based procedures for use in hybrid control rooms. INL/EXT-15-35284 Revision 0.

Le Blanc, K., Oxstrand J., Waicosky, T. (2012). Requirements for computer based-procedures for nuclear power plant field operators: Results from a qualitative Study. INL/CON-12-25691 PREPRINT.

Lehman, M.M. (1980). Programs, Life Cycles, and Laws of Software Evolution. Proceedings of the IEEE, 68 (9), 1060–1076.

Lin, C.J., Hsieh, T.-L., Yang, C.W., Huang, R.-J. (2016). The impact of computer-based procedures on team performance, communication, and situation awareness. *International Journal of Industrial Ergonomics, 51*, 21–29.

Lipner, M.H., Mundy, R.A., Franusich, M.D. (2006). Dynamic computer based procedures system for the AP1000. *NPIC/HMIT*.

O'Hara, J.M., Higgins, J., Stubler, W. (2000). Computerization of nuclear power plant emergency operating procedures. BNL-NUREG-67216.

O'Hara, J., Higgens, J.C., Stubler, J., Kramer, J. (2000). Computer-based Procedure Systems: Technical Basis and Human Factors Review Guidance (Technical Report No. NUREG/CR-6634).

NRC (1995). Evaluation of the computerized procedures manual II (COPMAI I)

Park, J., Jung, W., Kim, J., Ha, J. (2003) The step complexity measure—Its meaning and applications. *Journal of the Korean Nuclear Society, 35* (1) 80–90.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Task level errors for human error prediction in GOMS-HRA

R.L. Boring & T.A. Ulrich
Idaho National Laboratory, Idaho, USA

M. Rasmussen
NTNU Social Research, Trondheim, Norway

ABSTRACT: Goals-Operators-Methods-Selection rules (GOMS) was originally developed as a task analytic tool for modeling behavioral primitives in human users of human-computer interfaces. GOMS was recently adapted for Human Reliability Analysis (HRA), producing the GOMS-HRA method. The GOMS-HRA method provides a taxonomy of task level primitives in human activities that correspond to human error probabilities and task timing. The GOMS-HRA method has been used in computation-based HRA (CoBHRA), due to its calibration to the subtask level of human performance, the optimal decomposition level for dynamic risk modeling. GOMS-HRA has also been linked to procedures, and it is possible to map procedure steps (called procedure level primitives) to task level primitives. This paper introduces another important development to the GOMS-HRA framework—the task level errors, which represent the use of the GOMS-HRA taxonomy for predicting human error types. While many HRA methods map task types or task primitives to error rates, the prediction of error is often a generic error type. In reality, each task level primitive has predilections to certain types of errors. To model human error dynamically requires the determination of the types of errors that can occur.

1 A FRAMEWORK FOR DYNAMIC HUMAN RELIABILITY ANALYSIS

1.1 *Human Unimodel for Nuclear Technology to Enhance Reliability (HUNTER)*

Researchers at Idaho National Laboratory set out to create a simplified implementation of dynamic Human Reliability Analysis (HRA), also known as computation-based HRA (CoBHRA) due to consideration of modeling beyond temporal dynamics. The goal of this effort is to demonstrate proofs of concept of modeling HRA dynamically by adapting existing static HRA approaches to make them dynamic, rather than develop an entirely new dynamic HRA method. Because of the complexity of modeling human behavior and cognition, creating a new dynamic HRA method could quickly approach the complexity of artificial intelligence production systems. Thus, this simplified CoBHRA approach should borrow from current simplified, worksheet-based approaches to HRA, seeking to find ways to automate the analysis tasks. Further, the approach should integrate with other modeling tools like thermal hydraulic codes that are used for nuclear power plant simulations. This new HRA framework came to be called the Human Unimodel for Nuclear Technology to Enhance Reliability (HUNTER; Boring et al., 2016). While the initial implementation of HUNTER is focused on adapting static approaches to HRA, the framework is flexible and can in the future readily incorporate other models including dynamic HRA methods that have been developed in parallel at other institutions.

Central to the first implementation of HUNTER is quantification of the human error probability (HEP). Quantification in many HRA methods occurs in three stages (Swain and Guttman, 1983):

- *Nominal HEP*—the generic or default error rate for particular task types
- *Basic HEP*—the nominal HEP modified by contextual factors that increase or decrease errors
- *Conditional HEP*—the basic HEP modified to account for dependence and recovery

Some HRA methods like the Technique for Human Error Rate Prediction (THERP; Swain and Guttman, 1983) or Human Error Assessment and Reduction Technique (HEART; Williams and Bell, 2017) possess a large number of generic task types with a wide range of nominal HEPs, while simplified methods like the Standardized Plant Analysis Risk-Human (SPAR-H; Gertman et al., 2005) contain only two nominal task types. In most HRA methods, the basic HEP is calculated using performance shaping factors (PSFs), which simply

serve as multipliers on the nominal HEP. Finally, for the conditional HEP, dependence suggests that error begets error, and a sequence of human tasks may have increased overall error rates once an initial human error occurs. In contrast, recovery restores the success path and serves to decrease the HEP.

1.2 *Stages of HUNTER Quantification*

1.2.1 *Nominal human error probability*

To address nominal HEPs, HUNTER presently uses an approach called GOMS-HRA (Boring and Rasmussen, 2016). GOMS stands for Goals-Operators-Method-Selection rules, and serves as a task analytic approach to evaluate human-system interactions (Card et al., 1983). Despite many implementations of GOMS for user interface applications, GOMS has not historically been used for purposes of HRA. GOMS-HRA aligns its "operators" (i.e., operations or tasks) to an extended version of the Systematic Human Error Reduction and Prediction Approach (SHERPA) taxonomy (Stanton et al., 2013; Boring and Rasmussen, 2016), as depicted in Table 1. Here the GOMS operators are called *Task Level Primitives* (TLPs), and encompass the spectrum of operations in and beyond the control room. The TLPs are calibrated to scenarios in THERP to produce nominal HEPs as shown in Table 2.

Note that the GOMS-HRA TLPs have recently been mapped to *Procedure Level Primitives* (PLPs; Boring et al., 2017; Ulrich et al., 2017). PLPs represent simple procedure steps derived from a standardized list of operator commands. A common procedure step like CHECK VALVE readily maps to C_C, a single TLP. Other procedure steps may require mapping to multiple TLPs. The PLPs provide a lookup table of common procedure steps to their respective TLPs. This lookup table allows easy analysis reuse of commonly occurring operator actions. Moreover, it allows quick extraction of operator actions into an HRA model format, provided there is good procedure following by the crews.

Table 1. GOMS operators used to define task level primitives.

Primitive	Description
A_C	Performing required physical actions on the control boards
A_F	Performing required physical actions in the field
C_C	Looking for required information on the control boards
C_F	Looking for required information in the field
R_C	Obtaining required information on the control boards
R_F	Obtaining required information in the field
I_P	Producing verbal or written instructions
I_R	Receiving verbal or written instructions
S_S	Selecting or setting a value on the control boards
S_F	Selecting or setting a value in the field
D_P	Making a decision based on procedures
D_W	Making a decision without available procedures
W	Waiting

Table 2. HEPs associated with each task level primitive.

Operator	Nominal HEP	THERP Source*
A_C	0.001	20–12 (3)
A_F	0.008	20–13 (4)
C_C	0.001	20–9 (3)
C_F	0.01	20–14 (4)
R_C	0.001	20–9 (3)
R_F	0.01	20–14 (4)
I_P	0.003	20–5 (1)
I_R	0.001	20–8 (1)
S_C	0.001	20–12 (9)
S_F	0.008	20–13 (4)
D_P	0.001	20–3 (4)
D_W	0.01	20–1 (4)
W	n/a	n/a

*Corresponds to THERP Table values from Chapter 20.

1.2.2 *Basic human error probability*

SPAR-H (Gertman et al., 2005) is a simplified HRA method that features eight PSFs that serve as multipliers on the nominal HEP. The eight PSFs have different influence levels and corresponding multipliers that are selected by the analyst according to whether the PSF has a positive effect (i.e., decreasing the HEP) or negative effect (i.e., increasing the HEP). In the HUNTER implementation, SPAR-H PSFs are auto-calculated (Rasmussen and Boring, 2016; Boring et al., 2017) where possible from available plant parameters. This process works for so-called *external PSFs*, which are situational factors that influence the operators' performance. Internal PSFs—factors intrinsic to the operators—must still be assigned by the analyst or coded to change in particular contexts (e.g., the PSF for *Fitness for Duty*, which represents fatigue, may degrade after a long-duration event but not usually in response to particular plant parameters). The auto-calculated SPAR-H PSFs modify the basic HEPs from GOMS-HRA in the current implementation of HUNTER.

1.2.3 *Conditional human error probability*

HUNTER does not yet include an explicit model for conditional HEPs beyond the dependence

model that THERP and SPAR-H share in common. It is important to note that dynamic HRA modeling occurs at the subtask level. The human failure events (HFEs) used in static HRA represent hardware or process failures that may be influenced by human actions or inactions. These HFEs are often defined at the hardware component or system level, which does not meaningfully specify the spectrum of human activities associated with that hardware. Thus, an HFE may encompass a single step of a procedure or may require a whole procedure. The HFE level of modeling may require considerable task aggregation and expertise by analysts. In creating a virtual operator model in HUNTER, the focus necessarily falls not on the hardware but on the tasks the virtual operator must undertake, which will potentially have an effect on hardware process performance. In HUNTER and other dynamic HRA implementations, human activities are modeled at the subtask level, corresponding to activities surrounding each procedure step or equivalent for those activities that don't feature formal written procedures. This level of analysis is common in task analysis approaches used in human factors engineering, because it represents the most fundamental conscious level of action or cognition of the human (Rasmussen and Laumann, 2017).

Dependence in HRA is the notion that error begets error—that once an error has been committed, it may prime subsequent errors unless there is a clear break in activities. Contemporary thinking frames dependence in terms of mindset (Whaley et al., 2007), which is to say once a human error occurs (e.g., misdiagnosis of a plant upset), that error will propagate until something breaks the chain of operations (e.g., realization of the misdiagnosis). Dependence occurs because the HEP of downstream human activities is related to the first human error and will generally be larger than the calculated basic HEP. Most human error dependence modeling occurs at the task level—between HFEs—which may not readily translate into how dependence should be treated for subtasks—within HFEs. Work is ongoing in HUNTER (e.g., Boring, 2015) to develop better subtask dependence modeling.

2 A COGNITIVE MODEL OF TASK LEVEL PRIMITIVES

The GOMS-HRA TLPs represent a continuum of human cognition. Classical cognitive psychology, heavily influenced by information processing theory associated with computer technology, stressed three phases of cognition (Proctor and Vu, 2006):

Input → *process* → *output*,

which corresponds roughly to the following human activities:

Sensation/perception → *cognition* → *behavior*.

This framework to understand human cognitive functioning is echoed in several contemporary models. For example, Endsley's (2005) situation awareness (SA) model delineates the input and cognition phase into three separate phases: perception (Level 1 SA), comprehension (Level 2 SA), and Projection (Level 3) SA: which culminate in a decision and corresponding action:

perception → *comprehension* → *projection* → *decision* → *action*.

It should be noted that the modality of information and action can incorporate communication, whereby communication is a perceptual input or expressed action.

Cacciabue and Hollnagel (1995) make a distinction between *microcognition* and *macrocognition*. This distinction originally referred to the degree of context involved in decisions. Simplified cognition—microcognition—in carefully controlled laboratory experiments is removed of the context required in real-world complex cognition—macrocognition. While this distinction remains useful for framing the validity of human factors research, the concept of macrocognition has been generalized to represent high-level cognition like decision making in a real-world context. The distinction is helpful in understanding the focus of the GOMS-HRA task-level and procedure-level primitives. TLPs are microcognitive in that they represent the most basic level of human activities. PLPs are aggregated at a higher level to represent a series of cognitive functions and actions corresponding to a procedure step. It might be argued that the HFEs used as the unit of analysis in most HRAs represent a further macrocognitive grouping of multiple steps.

Recent work (Whaley et al., 2016) funded by the U.S. Nuclear Regulatory Commission (NRC) has identified several macrocognitive functions as the basis of the Integrated Decision-tree Human Event Analysis System (IDHEAS) HRA method (Xing et al., 2016). These macrocognitive functions are derived from research by Patterson and Hoffman (2012) and build on earlier functions identified by Klein et al. (2003). The IDHEAS macrocognitive functions include:

detection and noticing → *understanding and sensemaking* → *decision making* → *action* → *teamwork*,

whereby *teamwork* may be seen as an overarching function rather than as a discrete tail-end activity.

Table 3. Failures according to macrocognitive functions in Whaley et al. (2016).

Detecting/ Noticing	Understanding/ Sensemaking	Decision Making	Action	Teamwork
• Cues/information not attended to	• Incorrect data	• Incorrect goals or priorities	• Failure to execute desired action	• Failure of team communication
• Cues/information not perceived	• Incorrect frame	• Incorrect pattern matching	• Execute desired action incorrectly	• Failure in leadership/ supervision
• Cues/information misperceived	• Incorrect integration of data, frames, or data with a frame	• Incorrect mental simulation or evaluation of options		

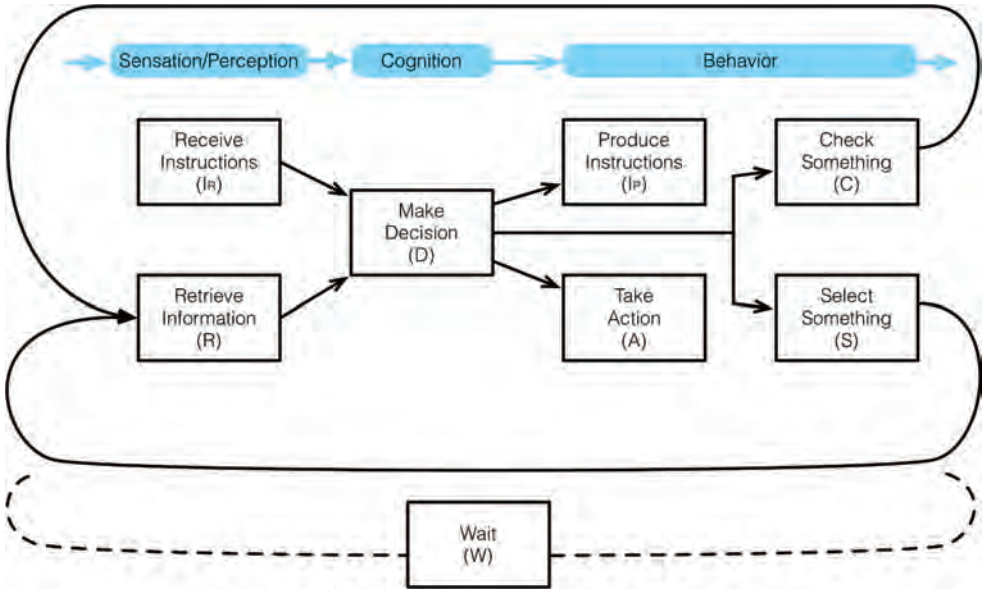


Figure 1. Cognitive model of GOMS-HRA task level primitives.

The IDHEAS macrocognitive functions are linked to proximate causes, which are error manifestations. The IDHEAS proximate causes are listed in Table 3.

Each of these proximate causes may be triggered for different reasons or cognitive mechanisms. For example, under the *detecting/noticing* macrocognitive function, if cues or information are not perceived, this could be caused by low salience of the cues or information, inability to maintain vigilance, inattention, mismatches between expected and actual cues (including biases toward particular information), or overloads of working memory. In turn, these cognitive mechanisms are influenced by PSFs (called performance influencing factors in IDHEAS).

The TLPs in GOMS-HRA as depicted in Table 1 are not ordered in a particular cognitive flow. Figure 1 represents a cognitive model of the TLPs. The distinction between control room and field operations (as depicted with subscripted *C* and *F* in Table 1) is not included in Figure 1, because the location of operations affects the context and corresponding error rates but not the function of the underlying cognition. The basic cognitive framework of GOMS-HRA mirrors a simple information processing model, with core functions related to sensation and perception (represented by *R*), cognitive (represented by *D*), and Action (represented by *A*) activities. As a pragmatic model based on observable operational functions, GOMS-HRA precludes a detailed delineation between *detecting/noticing* and *understanding/sensemaking*. In fact, the original SHERPA taxonomy (Stanton et al., 2013), upon which the TLPs are based, does not include the relatively unobservable decision making (*D*) operator.

Table 4. Crosswalk of IDHEAS macrocognitive functions with GOMS-HRA task level primitives.

IDHEAS macrocognitive function	GOMS-HRA A	GOMS-HRA C	GOMS-HRA R	GOMS-HRA I	GOMS-HRA S	GOMS-HRA D	GOMS-HRA W
Detecting/ Noticing		✓	✓	✓ (I_R)			
Understanding/ Sensemaking		✓	✓	✓ (I_R)	✓		
Decision-Making						✓	
Action	✓	✓		✓ (I_P)	✓		
Teamwork				✓ (I_P)			

GOMS-HRA includes several common operational functions that impact sensation and perception on the input side and actions on the output side. Communication in the form of verbal or written instructions (*I*) is both a cognitive input (I_R) and output (I_P). Checking information (*C*) is an active instance of looking for information that encompasses both an overt action (*A*) and a retrieval (*R*). Selecting or setting a value (*S*) and producing a verbal or written instruction (I_P) are both special instances of an action (*A*). Waiting (*W*) in GOMS-HRA is an overriding category that is used to indicate delays in tasking or periods of inactivity due to monitoring activities. Waiting does not inherently consist of a cognitive function, but it has implications across all GOMS-HRA operators.

The TLPs can be mapped to the IDHEAS macrocognitive functions as depicted in Table 4. Several of the TLPs span multiple macrocognitive functions but demonstrate a good match between concepts.

Table 5. Definition of key terms in GOMS-HRA.

Term	Abbreviation	Definition
Task Level Primitive	TLP	A basic human operation occurring at the subtask level. Multiple operations are typically required to achieve specific actions and goals.
Procedure Level Primitive	PLP	A human activity occurring at the procedure step level. Often, multiple task level primitives will be required to achieve a procedure level primitive activity.
Task Level Error	TLE	A nominal human error associated with a task level primitive. Each task level primitive is associated with multiple possible task level errors.

3 TASK LEVEL ERRORS

While it is helpful to automate many of the analytic processes in dynamic HRA as described earlier in this paper and in Rasmussen et al. (2017), such a model falls short on a key area: anticipating the types of errors that will likely occur for each TLP. The HUNTER framework outlined demonstrates a way to extract TLPs from procedure steps and a method to modify their influence using auto-calculated PSFs. Each TLP has particular error proclivities, which might be described as deviation paths from the normal path (U.S. Nuclear Regulatory Commission, 2000). Based on the proximate causes identified in the IDHEAS framework (see Table 3), here we map particular errors to the TLPs. These are called *task level errors* (TLEs). To avoid possible terminological confusion due to unique terms introduced through GOMS-HRA, key terms are summarized in Table 5.

A mapping of TLPs to TLEs is found in Table 6. As noted, the TLEs are derived from the proximate causes in Whaley et al. (2016). However, the mapping is not one-to-one. As depicted in Figure 1, some TLPs span multiple macrocognitive functions. Checking (*C*), for example, includes both an active operation of looking for information and the subsequent information input when the information is found. GOMS-HRA does not explicitly model teamwork, but much of the function of teamwork is covered in the Information Production (I_P) and Information Received (I_R) operations, which also cover both written and spoken communications.

The TLEs do not distinguish between errors of omission and errors of commission. An error of commission occurs when a person performs an unintended action. For example, an operator

Table 6. Task level errors in GOMS-HRA.

TLP	Task level errors
A	TLE-A1: Failure to execute desired action
	TLE-A2: Execute desired action incorrectly
C	TLE-C1: Wrong information checked
	TLE-C2: Information missed
	TLE-C3: Information misinterpreted
	TLE-C4: Failure to check information
R	TLE-R1: Information not attended to
	TLE-R2: Information not perceived
	TLE-R3: Information misinterpreted
I_P	TLE-IP1: Failure to produce desired communication
	TLE-IP2: Failure to produce correct communication
I_R	TLE-IR1: Communication not attended to
	TLE-IR2: Communication not perceived
	TLE-IR3: Communication misinterpreted
S	TLE-S1: Failure to select
	TLE-S2: Selection make incorrectly
D	TLE-D1: Incorrect goals or priorities
	TLE-D2: Incorrect use of information
	TLE-D3: Incorrect mental model
W	TLE-W1: Incorrect inaction
	TLE-W2: Waiting too long
	TLE-W3: Waiting too short

might accidentally activate a pump. In a plant, such actions can be particularly troublesome, because they may shift the plant beyond its standard operating range and may confound the normal procedural sequence by triggering anomalies beyond hardware failures. As noted in Whaley et al. (2016), errors of commission manifest at the action phase. The thought by an operator to activate a pump (perhaps due to an incorrect situation assessment) is caused by faulty decision making, but it will have no effect until that decision is put into action. As such, the TLEs that are mapped to Action—namely *A*, I_P, *C*, and *S*—include provision for errors of commission. One critique against the macrocognitive functions like those found in IDHEAS is that they focus primarily on cognition to the exclusion of behavioral or activity functions. The GOMS-HRA TLPs attempt to redress some of this shortcoming, which is also reflected in an expanded series of behavior operations as depicted in Figure 1.

4 CURRENT AND FUTURE APPLICATION OF TASK LEVEL ERRORS

The task level errors in GOMS-HRA provide two important additions to HRA: (1) The human error probability will vary depending on the type of error. Although the same overt failure may manifest from different causes, it is not the same error and should not be treated as a single probability value. This paper introduces a crosswalk of task level primitives to task level errors that can provide a more precise measure of human error and accompanying probabilities. (2) The task level errors bound the types of decision errors that can occur. This part of GOMS-HRA benefits decision trees, where branch points in operator decisions can have markedly different outcomes on event progression. Task level errors related to decisions can guide the likelihood of particular decision paths. Work still remains to specify such decision making and branching in HUNTER.

TLEs present an opportunity to model different errors that may occur for each subtask that is modeled in GOMS-HRA. Each TLE is an opportunity for error. As such, a human failure event is not the occurrence of a single error probability but rather the composite of multiple error types across multiple subtasks. The aggregation of these subtasks and co-likely error types presents a challenge for HRA aggregation. One way to link subtasks is through dependence. Dynamic dependence is far more complicated than the simple dependence used in static HRA methods, and the approach is not yet mature (Boring, 2015). Likewise, calculating dynamic HEPs can be orders of magnitude more complex than their static HRA counterparts. To date, HUNTER has not resolved the subtask-by-error type aggregation issue. There remains work to be done to determine suitable equations and then to calibrate them to HEPs for HFEs that have been generated by human analysts. In bringing subtask aggregation into HUNTER, there is the promise for greater consistency in the level of human task decomposition in HRA. Previous research (Poucet, 1989) has shown that not controlling for the level of decomposition results in spurious estimates in HRA. GOMS-HRA, as an implementation of HUNTER, aims to map the layers of decomposition to the appropriate HEP, enabling true dynamic autocalculation of human error.

5 DISCLAIMER

The opinions expressed in this paper are entirely those of the author and do not represent official position. This work of authorship was prepared as an account of work sponsored by Idaho National Laboratory, an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees makes any warranty, express or implied, or

assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately-owned rights. Idaho National Laboratory is a multi-program laboratory operated by Battelle Energy Alliance LLC, for the United States Department of Energy under Contract DE-AC07-05ID14517. This research was funded through the Laboratory Directed Research and Development program at Idaho National Laboratory.

REFERENCES

Boring, R.L. (2015). A dynamic approach to modeling dependence between human failure events. *Proceedings of the 2015 European Safety and Reliability (ESREL) Conference,* pp. 2845–2851.

Boring, R., Mandelli, D., Rasmussen, M., Herberger, S., Ulrich, T., Groth, K., & Smith, C. (2016). Human Unimodel for Nuclear Technology to Enhance Reliability (HUNTER): A framework for computational-based human reliability analysis. *13th International Conference on Probabilistic Safety Assessment and Management (PSAM 13),* Paper A-531, pp. 1–7.

Boring, R.L., and Rasmussen, M. (2016). GOMS-HRA: A Method for Treating Subtasks in Dynamic Human Reliability Analysis. *Proceedings of the European Safety and Reliability Conference*, 956–963.

Boring, R., Rasmussen, M., Smith, C., Mandelli, D., and Ewing, S. (2017). Dynamicizing the SPAR-H Method: A Simplified Approach to Computation-Based Human Reliability Analysis. *Proceedings of Probabilistic Safety Assessment Conference,* 1024–1031.

Boring, R., Rasmussen, M., Ulrich, T., Ewing, S., and Mandelli, D. (2017). Task and Procedure Level Primitives for Modeling Human Error. *Advances in Intelligent Systems and Computing, 589,* 30–40.

Cacciabue, P.C., & Hollnagel, E. (1995). Simulation of cognition: Applications. In J.-M. Hoc, P.C. Cacciabue & E. Hollnagel (Eds.), *Expertise and technology: Cognition & human–computer cooperation.* (pp. 55–73). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Card, S., Moran, T., and Newell, A. (1983). *The Psychology of Human-Computer Interaction.* Hillsdale: Lawrence Erlbaum Associates.

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32–64.

Gertman, D., Blackman, H., Marble, J., Byers, J., and Smith, C. (2005). *The SPAR-H Human Reliability Analysis Method, NUREG/CR–6883.* Washington, DC: U.S. Nuclear Regulatory Commission.

Klein, G.A., Ross, K.G., Moon, B.M., Klein, D.E., Hoffman, R.R., and Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems, 18*, 81–85.

Patterson, E.S., & Hoffman, R.R. (2012). Visualization Framework of Macrocognition Functions. *Cognition, Technology & Work, 14,* 221–227.

Poucet, A. (Ed.) (1989). *Human Factors Reliability Benchmark Exercise, Synthesis Report, HF-RBE.* Ispra: Joint Research Center of the Commission of the European Communities.

Proctor, R.W. & Vu K.P.L. (2006). The cognitive revolution at age 50: has the promise of the information processing approach been fulfilled? *Journal of Human-Computer Interaction, 23*, 253–284.

Rasmussen, M., and Boring, R.L. (2016). The Implementation of Complexity in Computation-Based Human Reliability Analysis. *Proceedings of the European Safety and Reliability Conference*, 972–977.

Rasmussen, M., Boring, R.L., Ulrich, T., & Ewing, S. (2017). The virtual human reliability analyst. *Advances in Intelligent Systems and Computing, 589,* 250–260.

Rasmussen, M., & Laumann, K. (2017). Decomposition level of quantification in human reliability analysis. *In Risk, Reliability and Safety: Innovating Theory and Practice – Proceedings of the 26th European Safety and Reliability Conference, ESREL 2016*, 997–1002.

Stanton, N.A., Salmon, P.M., Rafferty, L.A., Walker, G.H., and Baber, C. (2013). *Human Factors Methods: A Practical Guide for Engineering and Design, Second Edition.* Aldershot, UK: Ashgate Publishing Co.

Swain, A.D., and Guttman, H.E. (1983). *Handbook of human reliability analysis with emphasis on nuclear power plant applications. Final report, NUREG/CR-1278.* Washington, DC: US Nuclear Regulatory Commission.

Ulrich, T., Boring, R., Ewing, S., and Rasmussen, M. (2017). Operator Timing of Task Level Primitives for Use in Computation-Based Human Reliability Analysis. *Advances in Intelligent Systems and Computing, 589,* 41–49.

U.S. Nuclear Regulatory Commission. (2000). *Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA), NUREG-1624, Rev. 1.* Washington, DC: U.S. Nuclear Regulatory Commission.

Whaley, A.M., Boring, R.L., Blackman, H.S., McCabe, P.H., & Hallbert, B.P. (2007). Lessons learned from dependency usage in HERA: Implications for THERP-related HRA methods. *Official Proceedings of the Joint 8th IEEE Conference on Human Factors and Power Plants and the 13th Annual Workshop on Human Performance/Root Cause/Trending/Operating Experience/Self Assessment*, 322–327.

Whaley, A.M., Xing, J., Boring, R.L., Hendrickson, S.M.L., Joes, J.C., Le Blanc, K., and Morrow, S.L. (2016). *Cognitive Basis for Human Reliability Analysis, NUREG-2114.* Washington, DC: U.S. Nuclear Regulatory Commission.

Williams, J.C., and Bell, J.L. (2016). Consolidation of human error assessment and reduction technique. *Proceedings of European Safety and Reliability Conference*, 883–889.

Xing, J., Parry, G., Presley, M., Forester, J., Hendrickson, S., and Dang, V. (2016). *An Integrated Human Event Analysis Systems (IDHEAS) for Nuclear Power Plant Internal Events At-Power Application, NUREG-2199, Vol. 1.* Washington, DC: U.S. Nuclear Regulatory Commission.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A computational cognitive modeling approach to human performance assessment in nuclear power plants

Y. Zhao & C. Smidts
Nuclear Engineering Program, Department of Mechanical and Aerospace Engineering, The Ohio State University, Columbus, USA

ABSTRACT: A computational modeling approach to human performance assessment in nuclear power plants is proposed. The approach takes into consideration three main processes involved in a human's cognitive process: information perception, reasoning and response. Both the saliency and activation levels of a sensor signal in the control room are used to determine whether the signal can be actually perceived by the operator. The corresponding nodes in the operator's knowledge base are activated by the perceived signals and the activation is propagated in the knowledge base to reach diagnoses and develop responses. Uncertainties in these processes are dealt with using sampling methods. The interplay between an operator's stress and fatigue levels and the cognitive process are considered. A case study derived from the Three Mile Island accident is conducted to demonstrate the proposed approach.

1 INTRODUCTION

In complex technological systems where operator intervention is needed, such as nuclear power plants, chemical plants, and air traffic control centers, operator performance plays a central role in maintaining the normal operation of these systems and in mitigating consequences in case of incidents. However, high operator performance is challenged by a number of factors, for instance the increasing complexity of modern technological systems and the stricter requirement for the operator's knowledge in the systems. As a result, human performance assessment becomes necessary and has drawn increasing attention since the 1970s. The assessment not only serves as an integral part of the overall risk assessment of the systems (e.g. probabilistic risk assessment for nuclear power plants), but also provides insights into the ways of improving human performance (e.g. optimization of the human-machine interface).

To assess human performance, conventional methods, for instance the Technique for Human Error Rate Prediction (THERP) (Swain and Guttmann, 1983) and the Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H) (Gertman et al., 2005), use performance shaping factors (PSFs), for instance stress and experience, to represent the context that influences human performance and to adjust nominal human error probabilities. The number of PSFs varies between different methods, ranging from just one single factor (e.g. time available) up to 50 or more in some current methods (Boring, 2010). In these methods, expert judgment is extensively used to assign the levels of PSFs and to determine the relationships between PSFs and human error probabilities. These lead to variations in the results of different methods and even of the same method but conducted by different individuals. Besides, these methods deal with human performance assessment at the macroscopic level, which means that they cannot provide the users detailed information about how an operator commits an error. To address this limitation, novel methods based on a representation of human cognition were introduced (Chang and Mosleh, 2007a; Li, 2013; Smidts et al., 1997; Zhao and Smidts, 2017). These methods take explicitly human cognitive processes into consideration, for instance information perception, reasoning and decision making. The most prominent advantage of this type of methods is that they are able to provide more detailed interpretation of a human error.

In this paper, we further extend the capability of existing methods based on human cognition for human reliability assessment by proposing a computational cognitive modeling approach to human performance assessment in the context of nuclear power plants. In this approach, three main processes of human cognition, information perception, reasoning and response (Endsley, 1995), are modeled explicitly. The main contributions of our research include: 1) development of an approach to modeling information perception based on the saliency of and the attention paid by the operator

to a piece of information; 2) application of activation propagation in modeling the reasoning and response processes of an operator; 3) development of an approach to modeling the interplay between two significant PSFs, stress and fatigue, and the cognitive process; 4) development of an approach to modeling the uncertainties in the cognitive process based on sampling techniques; 5) integration of the pieces of the cognitive process into a framework. The proposed method exhibits several advantages compared with conventional methods of human performance assessment. In addition to the capability of providing microscopic interpretations of human errors, it is able to model the effect of task dependence on human performance, and to model the uncertainties in the human cognitive process based on simulation. The proposed approach is illustrated by the operator error in identifying the failure in the Pilot-Operated Relief Valve (PORV) in the Three Mile Island accident.

The paper is organized as follows. Section 2 provides a detailed introduction of the approach. In Section 3, the case study is introduced and the results are discussed. The paper is concluded in Section 4.

2 METHODOLOGY

2.1 *Framework*

The framework of the approach is shown in Figure 1. It consists of three main modules: information perception, reasoning and response, and PSFs.

In a nuclear power plant, an operator is faced with a large number of sensor signals in the control room and uses them to infer the dynamics and state of the plant. The first module, that is the module of information perception, is used to determine which signals can be perceived by the operator. The process of information perception is influenced not only by the signals themselves, but also by certain PSFs and the operator's mental state, as shown by the dashed lines in Figure 1.

The signals that are actually perceived then enter the operator's knowledge base and initiate the reasoning process. The knowledge base is represented as a network of nodes representing for instance specific systems or phenomena. The reasoning process is modeled by activating the nodes in the knowledge base corresponding to the perceived signals, as illustrated by the light grey node in Figure 1, and propagating the activation through the knowledge base. At the end of this process, the mental state of the operator is updated, for instance a system failure may be identified as illustrated by the medium grey node in Figure 1.

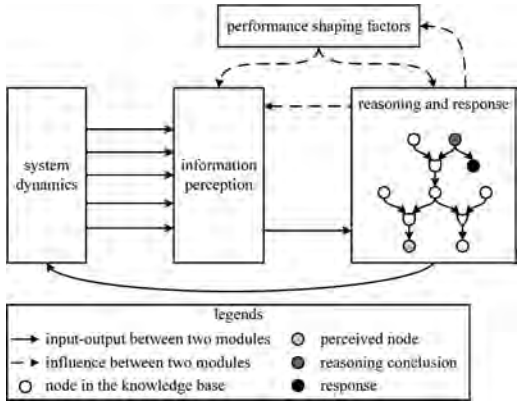


Figure 1. The framework of the approach.

An updated evaluation of the plant state will direct the operator to take corresponding actions, as illustrated by the black node in Figure 1. The developed actions are then executed, which in turn influences the plant's dynamics and state. System dynamics, information perception, reasoning, and response form a cycle, which represents the interaction between the nuclear power plant and the cognitive process of an operator.

It needs to be noted that there are uncertainties in the information perception, reasoning, and response processes as to whether a sensor signal can be perceived and whether a node in the knowledge base can be retrieved. These uncertainties are modeled by simulating the processes with sampling methods.

As shown in Figure 1, certain PSFs (i.e. stress and fatigue in our model) influence the cognitive process and the PSFs themselves are influenced by the reasoning and response processes. The interplay between them is modeled in the proposed approach.

The elements in the framework are introduced in further detail in the following sections.

2.2 *Information perception*

In a nuclear power plant, the data from the sensors plays a central role in the operator's situation awareness. An operator is usually faced with a large amount of information from the sensors (e.g. pressure, temperature, flow, water level). However, there is a limit in a normal human's capacity of perceiving the information around him/her. So only a sub-set of the available information can be perceived by the operator (Endsley et al., 2003). There are also uncertainties in information perception as to which signal or signals can be perceived by the operator (Chang and Mosleh, 2007b). These

two factors lead to the fact that some information important for situation awareness may be ignored by the operator, which may then result in human errors. As a result, information perception is an important element in the cognitive model.

The process of information perception is influenced by several factors. Research in the field of cognitive science shows that the perception of a piece of information depends on two factors: the saliency of the signal and the attention paid by the human to the signal (Corbetta and Shulman, 2002; Rybak et al., 1998; Theeuwes, 2010; Wolfe, 1994). The first factor corresponds to the stimulus-driven process of information perception, and the second factor corresponds to the goal-directed process. This finding is used as the basis of modeling the process of information perception in our research.

To be specific, the saliency of a signal is represented as the average of the contributions from three sources: the signal's variation, whether the signal is an alarm, and the signal's importance. The variation measures the change of a signal with time. The more significant the change, the more likely it is that the signal successfully draws the operator's attention. Whether a signal is an alarm or not indicates whether the signal has exceeded a specified threshold. An alarm is more likely to attract the operator's attention compared with a regular signal. The importance here refers to the accumulated emphasis on a signal through training or practice. For instance, the primary system pressure in a Pressurized Water Reactor (PWR) is usually emphasized more often compared to the position of a non-safety related valve during the training of an operator. As a result, the importance of the first signal is usually higher in the operator's mind and more easily draws the attention of an operator. The attention paid by an operator to a signal is represented by the activation level of the node that corresponds to the signal in the operator's knowledge base. As will be introduced in Section 2.3, the activation levels of the nodes in the operator's knowledge base are updated as the reasoning and response processes are ongoing. A node with a high activation level means that the node relates more to the operator's current cognitive activity and the operator pays more attention to the node. Finally, the overall activation level of a signal can be calculated with Equation 1 below:

$$A = \alpha S + \beta G = \alpha\left(\frac{1}{3}(V + Al + I)\right) + \beta G \tag{1}$$

where A is the overall activation level, S is the saliency, G is the goal related contribution, V is the variation, Al represents whether the signal is an alarm or not, I is the importance, α is the weight for the stimulus-driven contribution, and β is the weight for the goal-directed contribution.

With the overall activation level of a signal, the probability that the signal is perceived is calculated with Equation 2 below (Anderson et al., 2004):

$$p = \frac{1}{1 + e^{-(A-\tau)/s}} \tag{2}$$

where τ is the activation level threshold, and s is the noise in information perception. Both parameters are greater than 0.

From Equation 2, it is easy to see that the higher A is, the higher the probability p is. It is also easy to see that when A is greater than τ, p is greater than 0.5, and in this case the smaller s is, the higher p is, which means that the operator's attention is more focused. In contrast, when A is smaller than τ, p is smaller than 0.5, and in this case the higher s is, the higher p is, which means that the operator's attention is more dispersed. Based on the probability, sampling methods can be used to determine whether the signal can be actually perceived.

2.3 *An operator's knowledge base*

The reasoning and response processes are implemented in the operator's knowledge base. Before introducing these processes, a brief description of the representation of the knowledge base is provided in this section.

In our method, an operator's knowledge base is represented as a network (Zhao et al., 2017; Zhao and Smidts, 2017). In the network, the nodes are used to represent their counterparts in the physical world. They are classified into four categories: system, process, action, and state. System nodes represent systems or components in a nuclear power plant, for instance the reactor coolant piping and the relief valve in the pressurizer in a PWR. Process nodes represent the physical variables or phenomena in a nuclear power plant, for instance the pressure level in the primary system and hydrogen ignition in the containment. Action nodes represent the possible operator actions during the progression of an incident. State nodes here are used to represent the possible states of system and process nodes, for instance "open" and "closed" states for a relief valve and "high" and "normal" states for primary system pressure level.

In the network, the relationships between the nodes are modeled with three types of links: affiliation, AND gate, and OR gate. Affiliation links are used to represent the relationships between state nodes and the corresponding system or process nodes. There are no directions on affiliation links. AND and OR gates are used to represent the

cause-effect relationships between state nodes, and between state nodes and action nodes. The second class of cause-effect relationships is meant to model the fact that operator actions are guided by specific states of systems or process variables and the actions influence the states in turn. For instance, when the operator realizes that the reactor coolant piping is in the "break" state, he/she will take corresponding actions such as "shut down reactor".

In addition to the nodes and links in the network, three other elements are tied to some nodes. Specifically, beliefs, which are greater than 0, are tied to state nodes to represent the operator's belief in each state. Activation levels, which are from 0 to an upper limit A_{upp} are tied to all the nodes to represent the operator's attention on the nodes. Observability is tied to state nodes to indicate whether a state is observable or not to the operator. Observability has two values: 1 if the node is observable, and 0 otherwise.

An example knowledge base is shown in Figure 2.

2.4 *Reasoning and response*

Information perception plays an important role in the operator's situation awareness and response development. However, perception of the right information does not guarantee a successful diagnosis of the situation and an appropriate response.

Errors may be introduced in the reasoning and response processes. The operator's thought may be stuck at one point in either process. This means that the operator cannot identify the failure of a system, or cannot develop the appropriate response even if the operator has identified the failure.

To cover these possible errors, a method of modeling the reasoning and response processes based on activation propagation (Anderson, 1983; Anderson et al., 2004) and updating of belief is proposed. Based on the updated beliefs in the nodes, diagnoses will be reached and responses will be developed. The modeling method is illustrated in Figure 3 with the same example in Figure 2.

In the first step (see s1 in Figure 3), the node in the operator's knowledge base corresponding to the perceived signal is activated, and the activation level of this node is raised to an upper limit A_{upp}. In this example, low pressure in the primary system is observed, and the activated node is marked in light grey in Figure 3. It needs to be noted that only if the perceived information is inconsistent with the operator's belief, further inference is then made. This is modeled in a way as follows. First, for a perceived node, the ratio of the belief in the node to the summation of the beliefs in all the state nodes of the corresponding process node (i.e. "primary system pressure" in this example) is calculated. This ratio is then compared with a threshold ω (e.g. 0.8) to determine whether the perceived information is consistent with the operator's belief or not. If the ratio is greater than ω, the reasoning process is not implemented because the perceived information is normal, at least in the operator's mind. Otherwise, further inference is made, and the belief in the perceived node is raised by a predefined value ΔB.

In the second step (see s2 in Figure 3), the activation of the perceived node is propagated from bottom (i.e. the perceived node) up to the causal state nodes to infer the possible system failures that gave rise to the erroneous situation. In the example in Figure 3, the activation propagates from the "low" state of "primary system pressure" to two causal state nodes, "standby" of "injection system" and "break" of "reactor coolant system" through the AND gate. In this case, activation propagation and belief updating are carried out based on the following procedure. First, Equation 2 is used to calculate the probability of a causal node being retrieved, and sampling methods are applied to determine whether

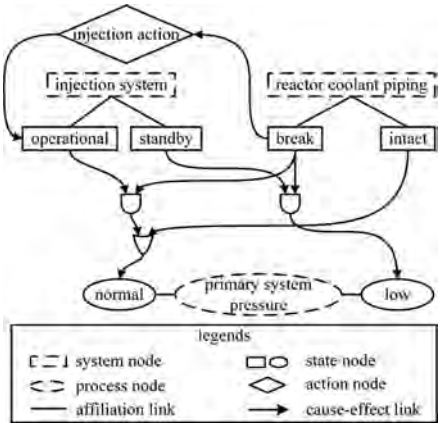


Figure 2. An example knowledge base.

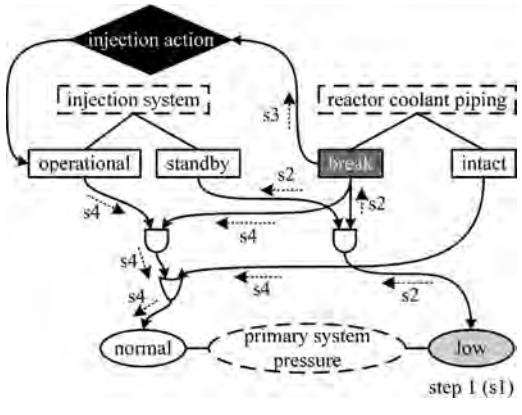


Figure 3. Illustration of the modeling of the reasoning and response processes.

the causal node is actually retrieved. A in Equation 2 now is the activation level of a causal node, for instance "standby" or "break" in this example. Then, for each retrieved causal node, its activation level is raised to the upper limit A_{upp} and its belief level is raised by ΔB. If no node is retrieved, the reasoning process stops at this point, which means that the operator's thinking process is stuck at the perceived node. For the cases of OR gate and multiple gates between the perceived node and its causal nodes, similar procedures are followed. Usually, there are several layers of nodes between the perceived node and system failure nodes rather than direct links as in the example in Figure 3. In this case, activation propagation and belief updating can be carried out step by step from bottom up following the same procedure. The result of this step is also the result of the reasoning process. It outputs the diagnosis of possible system failures (e.g. "break" in medium grey in Figure 3).

In the third step (see s3 in Figure 3), if the ratio of the belief in the failure state of a system is higher than a threshold Th_A, the operator will take corresponding actions, for instance "injection action" in the example in Figure 3. This step is also subject to errors of omission. So Equation 2 and sampling methods are used again to determine whether the operator will develop the response. In this calculation, A in Equation 2 is the activation level of an action node, for instance "injection action" in this example. If the action node is retrieved, its activation level is raised to the upper limit A_{upp}. An action will affect the state of the corresponding system (e.g. "operational" of "injection system"). The activation level of the affected state is then raised to A_{upp} and the belief in this node is raised by ΔB.

Situation awareness and response development have been modeled in the first three steps. In the last step (see s4 in Figure 3), activation propagation is implemented from top (i.e. identified system failures and executed actions) down to process nodes. This represents the fact that the operator will pay more attention to the system response related to the executed actions and the identified system failures, and those signals that represent the system response are more likely to be perceived in the next time step. Specifically, the process starts from the activated state nodes (i.e. "break" and "operational" in this example), and ends at the last state node (e.g. "normal" in this example). It is implemented following slightly different procedures as in the bottom-up process. For the example in Figure 3, as both "operational" and "break" have been activated, their activation is propagated to the OR gate through the AND gate. In addition to propagating the activation to effect node (i.e. "normal"), activation propagation to the other causal nodes of the effect node is also implemented, because all these nodes will draw the operator's attention. Equation 2 and sampling methods are used to determine whether a node can be retrieved or not. For the retrieved nodes, their activation levels are raised to A_{upp}. The change in the process variables of the nuclear power plant is usually delayed for some time after operator actions are taken, so the operator's belief in the states of these process variables are not updated. As a result, belief updating is not implemented in the top-down process.

In addition to activation increase resulting from the propagation from other nodes, the activation levels of the nodes in the knowledge base are also subject to decay (Endsley et al., 2003), which is calculated with Equation 3:

$$A_i(t+\Delta t) = A_i(t+\Delta t)e^{-\lambda \Delta t} \tag{3}$$

where Δt is the time duration between two discrete time steps, and λ is the decay constant.

2.5 *Integration of stress and fatigue into the model*

All the processes of information perception, reasoning and response are influenced by certain PSFs. On the other hand, the PSFs are also influenced by the operator's mental state, which is updated by the three processes. In our research, the interplay between two significant PSFs (i.e. stress and fatigue) and the three processes are modeled.

With respect to the effects of stress and fatigue on human performance, the major effects of the stress level can be summarized as: 1) people under high stress tend to focus on what they are focused on and ignore the other information (Endsley, 1995); 2) people under high stress tend to make premature decisions (Endsley, 1995); and 3) high stress influences human performance through decrements in working memory retrieval (Weerda et al., 2010). These effects are modeled by adjusting the parameters in the model accordingly following Equations 4 through 7 below:

$$\frac{\alpha}{\beta} = \frac{\alpha^*}{\beta^*} \cdot \ stress \tag{4}$$

$$s = s^* \cdot \frac{1}{stress} \tag{5}$$

$$Th_A = Th_A^* \times \frac{1}{stress} \tag{6}$$

$$\lambda = \lambda^* \cdot \ stress \tag{7}$$

where α and β are the weights in Equation 1, s is the noise parameter in Equation 2, Th_A is the threshold for action development, λ is the decay constant in Equation 3, and *stress* is the stress level. The parameters with a superscript of star denote the parameters

before they are adjusted by stress. Equations 4 and 5, 6, and 7 correspond to the three effects respectively.

The major effects of the fatigue level can be summarized as: 1) people with high fatigue are easily distracted by irrelevant subjects (Boksem et al., 2005); and 2) people with high fatigue tend to commit errors more easily because of the decrement of the activation level (Li, 2013). These effects are modeled by adjusting the parameters following Equations 8 through 10 below:

$$\frac{\alpha}{\beta} = \frac{\alpha^*}{\beta^*} \cdot \frac{1}{fatigue} \tag{8}$$

$$s = s^* \cdot fatigue \tag{9}$$

$$\lambda = \lambda^* \cdot fatigue \tag{10}$$

where α, β, s, and λ are the same parameters as in Equations 4, 5, and 7, and *fatigue* is the fatigue level of the operator. Equations 8 and 9, and 10 correspond to the two effects respectively.

As to the effects of the cognitive process on PSFs, the stress level is updated every time step with Equation 11:

$$stress = \frac{1}{N_s} \sum_{i}^{N_s} R_i \tag{11}$$

where N_s is the number of system and process nodes related to negative consequences, for instance "reactor coolant piping" in Figure 3, and R_i is the ratio of the belief in the state node corresponding to negative consequences, for instance "break" in Figure 3, to the summation of the beliefs in all the state nodes of system or process node *i*.

The fatigue level is updated every time step with Equation 12:

$$fatigue = \alpha_1 \left(\frac{1}{N_f N} \sum_{j=1}^{N_f} \sum_{i=1}^{N} A_i\left(t - j\Delta t\right) \right) + \alpha_2 \min\left\{1, \frac{T}{T^*}\right\} \tag{12}$$

where N_f is the number of past time steps from the current time point, N is the number of nodes in the knowledge base, Δt is the time duration between two time steps, T is the time the operator has spent on the accident management, T^* is the threshold for the time period that leads to the exhaustion of an operator, and α_1 and α_2 are weights. This calculation represents the fact that the fatigue level of an operator can be contributed by two sources, one is the attention the operator has paid on the tasks, and another one is the time duration the operator has been involved in the tasks.

3 CASE STUDY

A simple case study is conducted with the proposed computational modeling approach. Results are obtained and discussed.

3.1 *Case introduction*

The case studied is derived from the Three Mile Island accident. To be specific, it refers to the operators' failing to identify the failure in the Pilot-Operated Relief Valve (PORV) at the top of the pressurizer which caused the valve to stick open during the accident and aggregated the situation (Rogovin, 1979). During the accident progression, there were several instrument readings relevant to this failure available to the operator in the control room. Some were misleading but the others provided useful evidence of this failure. Unfortunately, the operator paid too much attention to the misleading readings and ignored the useful readings for over 2 hours. The misleading instrument readings include high water level in the pressurizer which in fact was caused by steam voids and gas in the reactor coolant system, and closed indicator of the PORV which was only circumstantial evidence of the actual state of the PORV. Useful instrument readings include low pressure of the reactor coolant system because of the release of coolant from the stuck-open PORV, and high temperature downstream of the PORV which was also caused by the release of high-temperature coolant from the PORV.

To provide a clear illustration of the essence of the proposed approach, the case study is simplified in the following ways. Firstly, the task in the case study is defined as to identify the failure in the PORV mechanism. This means we are just focused on the information perception and reasoning processes. The actions following this diagnosis are not considered. Secondly, only one time step is considered, and the dynamical environment around the operator which needs to be represented by multiple time steps is not considered. This means that the information perception and reasoning processes are implemented one time. Thirdly, the operator's knowledge base about this situation is simplified as in Figure 4, which only consists of the aforementioned relevant instrument readings and the PORV. Six nodes without any notation are added in Figure 4 to represent the distracting effects of irrelevant readings. It is assumed that perception of these nodes has no contribution to the completion of the task.

In the case study, "primary system pressure", "PORV position", "temperature downstream of PORV", "pressurizer water level", and the six irrelevant nodes are observable. The instrument readings indicate that the states of the relevant nodes are "low", "closed", "high", and "high" respectively. Normally, only when the ratio of the belief in the

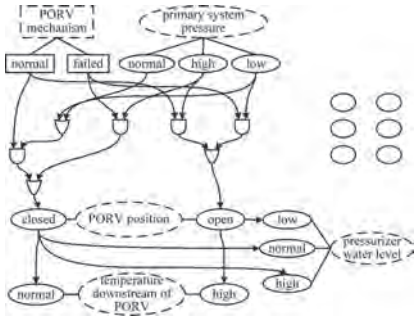


Figure 4. The knowledge base in the PORV case study.

"failed" state to the summation of the beliefs in all the state nodes of "PORV mechanism" is above the action threshold Th_A, the task is seen as success. However, the cognitive process is implemented for only one time step in the case study. This means that the belief in the "failed" state may be increased but may not be increased to be above Th_A. As a result, the success criterion is relaxed to be the activation of the "failed" state. This means that once the "failed" state is reached in the reasoning process, the task is thought to be successful. The values of the parameters are as follows: τ, the initial value of A for each node, and s in Equation 2 are 0.4, 0.5, and 1 respectively. The initial belief in each state is 1 and the default belief ratio ω is 0.8. These values are just used to illustrate the calculation, and do not necessarily reflect the real situation.

3.2 *Results*

The information perception and reasoning processes for the case study are simulated 1000 times. The perceived nodes and activation propagation paths in the knowledge base are recorded at each time. The results are analyzed from the following two aspects.

First, the probability of success is calculated with Equation 13:

$$p_S = \frac{N_S}{N_T} \tag{13}$$

where N_S is the number of successes, and N_T is the number of simulations which is 1000 in the case study. The simulation results show that the probability of success is 0.39.

Second, statistics on the combinations of perceiving useful or useless information and success or failure of the task are obtained. Here perceiving useful information means that at least one of the two useful signals are perceived. Perceiving useless information means that none of the two useful signals are perceived. Useless information refers to the misleading signals and the distracting instrument readings. The probabilities of success condition on perceiving useful information, success condition on perceiving useless information, failure conditioned on perceiving useful information, and failure conditioned on perceiving useless information are 0.91, 0.09, 0.83, and 0.17 respectively. From the result, we can see that perceiving useful information does not guarantee the success of the task, and perceiving useless information does not necessarily lead to the failure of the task. But perceiving useful information does help increase the success probability, as can be seen from the comparison between 0.91 and 0.83. The small increase of the probability of success also suggests that the failure of the task is dominated by the unsuccessful retrieval of information in the reasoning process and by the high complexity of the situation under study.

Two examples, one for success and another for failure, are explained to illustrate the reasoning process. In the example for success, "closed" state of "PORV state", "high" state of "temperature downstream of PORV", and the six irrelevant signals are perceived by the operator. However, except for the "closed" state, the activation propagation from the other signals is stuck at the initial points. The "closed" state leads to the activation of "failed" state of "PORV mechanism" and "normal" state of "pressurizer pressure". In the example for failure, "high" of "pressurizer water level", "low" of "primary system pressure", and two irrelevant signals are perceived. Perception of "high" of "pressurizer water level" leads to the activation of "closed" of "PORV state", then to the activation of "normal" of "primary system pressure" and "normal" of "PORV mechanism". The activation propagation from all the other signals is stuck at the initial points.

3.3 *Discussion*

From the case study, it is easy to see that the proposed computational cognitive modeling approach is able to provide the probability of success or failure in an easy way from the simulation results. In the calculation, no expert opinion is needed. Another more prominent strength of the proposed approach is that the simulation results provide the basis for detailed investigation of an operator's cognitive process in a specific situation from different perspectives, for instance the relation between failure in the task and perceived signals in the case study.

However, significant improvements are still needed to enable the proposed approach for practical applications and to further enhance the capabilities of the approach. First, a better way of representing the knowledge base needs to be developed to cover more complex structures of an operator's knowledge in one field. Second, a systematic method of determining the parameters in the cognitive model through experimental results

or the results in existing research is needed. Third, the proposed approach needs to be validated by comparison with experimental results.

4 CONCLUSIONS

In this paper, a computational cognitive modeling approach to human performance assessment in nuclear power plants is proposed. The three main processes in a human's cognitive process, that is information perception, reasoning and response development, are modeled. The interplay between two significant PSFs (i.e. stress and fatigue) and the cognitive process is considered. Sampling methods are used to simulate the possible outcomes of an operator's cognitive process. The simulation results enable the probability of success or failure of an operator performing a specific task to be calculated in an easy way. In addition, the simulation results provide the basis for detailed investigation of the cognitive process. With the proposed approach, the effect of task dependence on human performance is considered implicitly based on the use of activation levels for the nodes. A case study derived from the Three Mile Island accident is conducted with the proposed approach. Although the case study is simplified and the results are preliminary, it demonstrates the power of the approach. Future research for further improvements of the proposed approach is pointed out.

ACKNOWLEDGMENT

This research was prepared under award NRC-HQ-60-15-G-0002 from the US Nuclear Regulatory Commission. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the view of the US Nuclear Regulatory Commission.

REFERENCES

Anderson, J.R., 1983. A spreading activation theory of memory. J. Verbal Learn. Verbal Behav. 22, 261–295. https://doi.org/10.1016/S0022–5371(83)90201–3.

Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y., 2004. An integrated theory of the mind. Psychol. Rev. 111, 1036–1060. https://doi.org/10.1037/0033–295X.111.4.1036.

Boksem, M.A.S., Meijman, T.F., Lorist, M.M., 2005. Effects of mental fatigue on attention: An ERP study. Cogn. Brain Res. 25, 107–116. https://doi.org/10.1016/j.cogbrainres.2005.04.011.

Boring, R.L., 2010. How Many Performance Shaping Factors Are Necessary for Human Reliability Analysis? (No. INL/CON-10-18620). Idaho National Laboratory (INL).

Chang, Y.H.J., Mosleh, A., 2007a. Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 1: Overview of the IDAC Model. *Reliab. Eng. Syst. Saf.* 92, 997–1013. https://doi.org/10.1016/j.ress.2006.05.014.

Chang, Y.H.J., Mosleh, A., 2007b. Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 3: IDAC operator response model. *Reliab. Eng. Syst. Saf.* 92, 1041–1060. https://doi.org/10.1016/j.ress.2006.05.013.

Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3, 201. https://doi.org/10.1038/nrn755.

Endsley, M.R., 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Hum. Factors* 37, 32–64. https://doi.org/10.1518/001872095779049543.

Endsley, M.R., Bolté, B., Jones, D.G., 2003. Designing for situtation awareness: an approach to user-centered design. Taylor & Francis, London; New York.

Gertman, D., Blackman, H., Marble, J., Byers, J., Smith, C., 2005. The SPAR-H human reliability analysis method (No. NUREG/CR-6883). U.S. Nuclear Regulatory Commission, Washington, DC.

Li, Y., 2013. Modeling and simulation of operator knowledge-based behavior (Ph.D.). University of Maryland, College Park, United States -- Maryland.

Rogovin, M., 1979. Three Mile Island: A Report to the Commissioners and to the Public (No. NUREG/CR-1250(Vol.1)). Nuclear Regulatory Commission, Washington, DC (USA).

Rybak, I.A., Gusakova, V.I., Golovan, A.V., Podladchikova, L.N., Shevtsova, N.A., 1998. A model of attention-guided visual perception and recognition. Vision Res. 38, 2387–2400. https://doi.org/10.1016/S0042–6989(98)00020–0.

Smidts, C., Shen, S.H., Mosleh, A., 1997. The IDA cognitive model for the analysis of nuclear power plant operator response under accident conditions. Part I: problem solving and decision making model. *Reliab. Eng. Syst. Saf. 5*5, 51–71. https://doi.org/10.1016/S0951–8320(96)00104–4.

Swain, A.D., Guttmann, H.E., 1983. Handbook of Human-Reliability Analysis with Emphasis on Nuclear Power Plant Applications. Final Report (No. NUREG/CR-1278; SAND-80–0200). Sandia National Labs., Albuquerque, NM (USA).

Theeuwes, J., 2010. Top–down and bottom–up control of visual selection. *Acta Psychol. (Amst.)* 135, 77–99. https://doi.org/10.1016/j.actpsy.2010.02.006.

Weerda, R., Muehlhan, M., Wolf, O.T., Thiel, C.M., 2010. Effects of acute psychosocial stress on working memory related brain activity in men. Hum. Brain Mapp. 31, 1418–1429. https://doi.org/10.1002/hbm.20945.

Wolfe, J.M., 1994. Guided Search 2.0 A revised model of visual search. Psychon. Bull. Rev. 1, 202–238. https://doi.org/10.3758/BF03200774.

Zhao, Y., Altman, K., Chaudhury, K., Anandika, M., Smidts, C., 2017. A Systematic Method to Build a Knowledge Base to be Used in a Human Reliability Analysis Model, in: Advances in Intelligent Systems and Computing. Presented at the International Conference on Applied Human Factors and Ergonomics, Springer, Cham, pp. 50–60. https://doi.org/10.1007/978-3-319-60645-3_6.

Zhao, Y., Smidts, C., 2017. A Dynamic Mechanistic Model of Human Response Proposed for Human Reliability Analysis, in: Advances in Intelligent Systems and Computing. Presented at the International Conference on Applied Human Factors and Ergonomics, Springer, Cham, pp. 261–270. https://doi.org/10.1007/978-3-319-60645-3_26.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Towards a framework for assurance of autonomous navigation systems in the maritime industry

A. Brandsæter
DNV GL, Høvik, Norway
University of Oslo, Oslo, Norway

K.E. Knutsen
DNV GL, Høvik, Norway

ABSTRACT: We discuss potential assurance frameworks for autonomous navigation systems in the maritime industry, with emphasis on testing and verification of the system's perception performance and capacities. Ongoing research in this field has revealed profound challenges related to artificial situation awareness and machine perception specific to the marine environment. The lack of a clear and transparent framework and methodologies to assure the safety associated with the usage of such solutions, have been identified as key barriers for the implementation of autonomous navigation solutions at scale. Because the machine perception and situational awareness algorithms are expected to be partly or fully based on machine learning algorithms, including deep learning, whose functional reasoning is challenging or even impossible to understand and predict, the verification of such systems is fundamentally different from a traditional verification process based on physical understanding and theory. We review several methods for testing autonomous navigation systems, proposed and used mainly in the automotive industry, and discuss how these methods can be adapted, combined and applied to form a framework for assurance of autonomy in the maritime industry.

1 INTRODUCTION

Autonomous transport on land, in the air and at sea has been coined the technology trend with the highest potential to disrupt the transport sector in the future. It has the potential for making transport solutions more cost effective, safe and environmentally friendly, but also to disrupt entire business models and value chains associated with the mode of transport. Given the disruptive potential of this technology trend, increasing research efforts are being invested to realize the technological solutions.

1.1 *The autonomy revolution*

Technologies and methods for autonomous systems is a very active area of research both in the industry and in academia. However, the majority of the research being done for autonomous vehicle navigation is focused around the automotive industry. The amount of test data for such vehicles is becoming abundant and is considered an important contributor to the current state of the art in the research field. Major advances in object detection, classification and image analysis have been made in recent years, with extensive use of artificial intelligence related technologies such as feature extraction, artificial neural networks, deep learning models such as Convolutional Neural Networks (CNNs), gradient-based and derivative-based matching approaches (see for example (Hofmann 2013, Rout 2013, MathWorks 2017c)). Research is needed to identify if and how the algorithms, methods and sensors, developed for the automotive industry, can be utilized in the maritime domain.

1.2 *Opportunities in the maritime industry*

Several studies have shown that human error contributes to a majority of marine casualties (Rothblum 2000, Harrald et al. 1998). However, automated systems and autonomy can also introduce new challenges, and existing challenges might be amplified (Lützhöft & Dekker 2002). Nevertheless, we expect that if the interaction between the humans and machines are treated carefully, with thorough testing and verification, autonomy can contribute significantly to increase safety in many maritime operations.

Unmanned ships will enable optimization of energy efficiency due to changes in design

constraints and freeing of space, previously used to accommodate crew. In addition, more hydrodynamic and aerodynamic designs may in turn lead to less fuel consumption and reduced emissions. Furthermore, autonomous ships might be able to compete with road transportation and contribute to reduced emission from road transportation as well as reduced road wear and tear.

If autonomous ships are successfully implemented, it will most probably enable fundamentally new types of ship transportation operations, such as for example single container shipment (Woodgate 2017); extremely slow speed transportation with very low emissions (Tvete 2017); container feeder to replace road transport (Kongsberg 2017); and unmanned patrol ships (Fingas 2017).

Several demonstrators have already proven that it is feasible for a transport solution to be operated by sensors and software either partially or fully based on deep learning algorithms (Huval et al. 2015, Ackerman 2017). Among others, a company Drive.ai, has an ambition to use deep learning fully from sensory input to decision making, while others usually use deep learning in parts of the system, e.g. situational awareness, while relying on traditional control system logic in other parts of the system (Huval et al. 2015, Ackerman 2017, Muoio 2016). Nevertheless, the solutions are yet to be deployed at scale. One of the reasons for the lack of deployments is that the solutions are still not proven to be sufficiently safe.

1.3 *Early rule development as an enabler for innovation*

A key element required to keep the autonomous system safe, is the ability of the system to achieve situational awareness. Situational awareness algorithms are usually partly or fully based on machine learning algorithms whose functional reasoning are challenging or even impossible to understand and predict. Hence, the verification of such a system is fundamentally different from a traditional verification process based on physical understanding and theory. The machine learning algorithms are data driven, and completely dependent on the quality of the training data. Therefore, verification will likely be carried out by a combination of testing, simulations and benchmarking against real and synthetic data sets. Furthermore, adaptive methods, where data are automatically collected and used to retrain the system, will also be considered.

For a manned system, awareness is achieved by the human operator by using his or her senses and perceptive abilities to interpret instrument signals and input from surroundings. An unmanned ship should use a priori information, such as maps, combined with sensor readings to make observations relative to the environment, and use software to perceive the situation based on this input. This digital perception will be used as input to a decision-making algorithm. In turn, this controls the actuators of the vessel which are effectuating the decision made. For the autonomous system to make safe decisions, the situational awareness must be sufficiently accurate for all feasible situations and conditions which the vessel may encounter.

System functional and performance requirements necessary to obtain a required safety level of an automated situational awareness system should be established as early as possible, as this will offer the technology providers a standard to be met by their solutions. If requirements are not set before or early in the technology development phase, developers risk spending significant efforts and money on developing solutions which in the end may not meet the safety standard. However, establishing such a standard is difficult when no solutions exist to evaluate the standard against.

In addition to a standard for required system and component performance, tools are needed for verifying that the technology meets the requirements set in the standard. For a situational awareness system, this will include verifying that the sensors adequately detect objects affecting the safety of the vessel and its surroundings under various conditions, and that the perception algorithm can use this information together with other a priori information to adequately understand the situation.

1.4 *Focus of this study*

In this paper, we discuss rules and regulations related to autonomous navigation systems in a maritime context, with focus on autonomous perception and situational awareness. However, we believe that a framework for approval developed for autonomous applications will also be applicable to other systems that are based on machine learning algorithms and artificial intelligence.

The remainder of the paper is structured as follows. In section 2, we propose and describe a range of recommended practices and tools that can be applied to test and validate the ability, performance and robustness of safety critical systems which decisions are based on data-driven methods. These practices and tools originate partly from traditional statistical analysis and are suggested and applied for testing and assurance of autonomy in the automotive industry. In section 3, we discuss challenges related to machine perception that are unique or particularly pronounced in the maritime domain, and suggest how the recommended practices and tools should be used and possibly adapted to suit the maritime domain. Furthermore, we present a

possible scope for assurance framework, and discuss potential implications of autonomy such as for example operational dependent requirements. We also describe the IMO guidelines for approval of alternatives and equivalents. We conclude in section 4.

2 TOOLS AND RECOMMENDED PRACTICES FOR ASSURANCE

In the following, we propose and describe a different recommended practices and tools that can all be applied to test and validate the ability, performance and robustness of safety critical systems which decisions are based on data-driven methods. See Figure 1 for an overview of the methods.

2.1 *Confusion matrix*

It is not obvious how to measure and evaluate the performance of autonomous navigation systems. If two systems provides divergent predictions or decision, it is difficult to define and quantify which reaction was most correct. In classification problems, the results are often presented in a confusion matrix, where the predicted class is compared with the actual or the true class. With two classes, for example object detection with two objects, it is straight forward to define the confusion matrix, which inhere the number of

- True Positives (TP), hits
- True Negatives (TN), correct rejections
- False Positives (FP), false alarms, Type I errors
- False Negatives (FN), misses, Type II errors

When more classes are needed, defining the criteria for performance evaluation becomes more challenging. For example, how should we quantify the performance of a perception system which correctly detects a vessel, but misclassifies it as a ferry? And how should this be compared to misclassifying it as a kayak? Or what if the system is not even able to recognize it as an object?

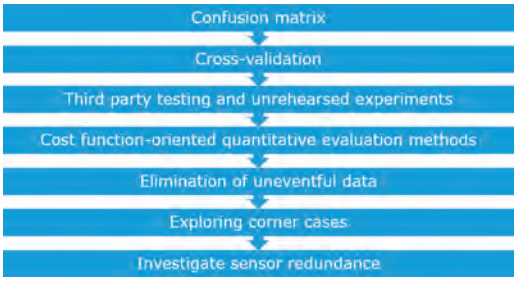


Figure 1. Proposed components in an assurance framework for safety critical systems.

To be able to make the above-mentioned comparison, we are fully dependent on a correctly labelled dataset of the ground truth. An autonomous ship will likely be equipped with multiple sensors, including multiple daylight and IR cameras in various directions, radars with different settings, in addition to Automatic Identification System (AIS) and other satellite data, etc. The labelling process should take all these sources into account when labelling the data. For example, if an object is not visible in the video stream due to thick fog or other difficult weather conditions, but we know the objects position from AIS data or other sources, the object will be labelled in the ground truth data set. All relevant information should be correctly labelled, in all datasets.

Data collection, and especially data annotation or labelling, are surprisingly time consuming and costly tasks for vehicle classification (Schöning et al. 2015, Chen and Ellis 2014). However, various tools and methods for semi-automatic ground truth labelling on video streams designed for the automotive industry are available, such as for example (MathWorks 2017b, MathWorks 2017a, Cuevas et al. 2015, Lopez-Villa et al. 2015, Schöning et al. 2015). Another approach is to crowd source the data annotation like Mighty AI has done in automotive, where they have developed a mobile app in which users may annotate images manually and get paid for it, whereupon Mighty AI makes a business out of selling annotated datasets [https://mty.ai/]. The available solutions should be explored, and if necessary adapted for our use in a maritime context.

2.2 *Cross-validation*

It is well known that when we evaluate predictions from a statistical model on the dataset used to train the model, our accuracy estimates tend to be over-optimistic (Arlot & Celisse 2010). To build robust and accurate models we ideally want to use all data available. The same applies to testing; we want to test our models in all situations, not only on a subset. Cross-validation introduces various methods of repetitively splitting the data $\mathcal{D}$ into two exclusive parts $\mathcal{D}_t$ and; $\mathcal{D}_v$ where one part $\mathcal{D}_t$ is used to train the model, and the other $\mathcal{D}_v$ is reserved for validation.

A range of different splitting techniques can be applied, providing different cross-validation estimates. See for example Arlot & Celisse 2010, Kohavi 1995 for a brief overview of the most common splitting techniques.

One of the most widely used splitting technique is called K-fold cross-validation, which in its

standard form splits the original dataset $\mathcal{D}$ into K subsets (folds), $\mathcal{D}_1$, …, $\mathcal{D}_K$, as described in (Arlot and Celisse 2010, Brandsæter and Vanem 2016). For each $k \in 1,2,\ldots K$ the models are trained on $\mathcal{D}_t = \mathcal{D} \backslash \mathcal{D}_K$, and tested on $\mathcal{D}_k$. To make sure that the results are not strongly dependent on how the folds are selected, we repeatedly run the K-fold cross-validation with new selections. The sets are often chosen to be mutually exclusive with equal size.

2.3 *Extensive testing*

The standard approach for assurance of autonomous navigation in the automotive industry is extensive testing (Pei et al. 2017a, Zhao and Peng 2017, Waymo 2016, Fei-Fei 2010), where large amounts of real world data from ordinary operation is gathered and manually labelled, and data on driver performance, behaviour, environment, driving context and other factors that were associated with critical incidents, near misses and crashes are analysed and used in evaluating the system performance (Zhao and Peng 2017).

Simulated real-world data is also sometimes used to massively increase the amount of data (Madrigal 2017, Zhao and Peng 2017), but usually this is completely unguided, and due to the large input space of real-world scenarios, none of these approaches can hope to cover more than a tiny fraction (if any at all) of all possible corner cases (Pei et al. 2017a). Here, a corner case is defined as an unusual, but far from impossible, scenario. In particular, if each individual parameter, such as temperature, fog, daylight, driving speed, number of other vehicles involved, etc. are well within the normal range for that parameter, but still the combined scenario is unusual. As an example, again from the automotive industry, a Tesla in autopilot recently crashed into a trailer because the autopilot system failed to recognize the trailer as an obstacle due to its white color against a brightly lit sky and the high ride height (Lambert 2016). Such corner cases were not part of Waymos (Googles) or Teslas test set (Pei et al. 2017a) and thus never showed up during testing.

2.4 *Third party testing and unrehearsed experiments*

In 2003 the Defense Advanced Research Projects Agency (DARPA) announced the first Grand Challenge with the goal of developing vehicles capable of autonomously navigating desert trails and roads at high speeds. In Krotkov et al. 2007, the conduct of six evaluation experiments for the DARPA PerceptOR program is described. Key distinctions of the testing methodology include conduct of the experiments by an independent third party, and the use of unrehearsed experiments that provide little advance knowledge of and access to the test courses. The article also presents quantified, objective performance metrics for the systems evaluated. Furthermore, it includes blind experiments that do not allow the system operators to see the test courses until all tests are completed.

The test environment and the test content are described in detail; however, the evaluation approach are not thoroughly discussed (Sun et al. 2011).

2.5 *Cost function-oriented quantitative evaluation methods*

Wei and Dolan 2009 claims that most teams in the 2007 DARPA Urban Challenge preferred to avoid difficult manoeuvres in high-density traffic by stopping and waiting for a clear opening instead of interacting with it and operating the vehicle and human drivers. To encounter this, researchers at Beijing Institute of Technology, propose a design method for a scientific and comprehensive test and evaluation system for autonomous ground vehicles competitions, to better guide and regulate the development of autonomous ground vehicles. The evaluation method proposed by Sun et al. 2014, Sun et al. 2011 aims to evaluate the quality of completion with a cost function-oriented quantitative evaluation method. This evaluation method can presumably evaluate the overall technical performance and individual technical performance of autonomous ground vehicles. A complete test system that includes the test contents, the test environment, and the test methods to meet the demands of testing for autonomous ground vehicles is developed, and a fuzzy evaluation method is combined with an analytic hierarchy process to solve fuzzy and hard-to-quantify problems (Sun et al. 2014).

2.6 *Elimination of uneventful data*

Recently, a new approach to testing autonomous cars was proposed by researchers affiliated with the University of Michigans Mcity connected and automated vehicle center. Zhao and Peng 2017 presents an accelerated evaluation process which aims to eliminate the uneventful driving activity, and filter out only the potentially dangerous driving situations where an automated vehicle needs to respond, creating a faster, less expensive testing program. It is claimed that this approach can reduce the amount of testing needed by a factor of 300 to 100,000.

Four methodologies that form the basis of the accelerated evaluation process are listed (Zhao and Peng 2017):

1. Evaluate how frequently a significant driving event happens on the road, and stripe out the more common, uneventful safe driving situations.
2. Use importance sampling to statistically increase the number of critical driving events in a way that still accurately reflects real-world driving situations.
3. Construct a formula that accurately distils those critical events, tests the formula, and apply it to further reduce the amount of testing required.
4. Analyse interactions between human-driven vehicles and robotic vehicles and optimize the random occurrences of significant driving events in the most complex scenarios.

2.7 *Exploring corner cases in deep learning systems*

In Pei et al. 2017a, Tian et al. 2017, Pei et al. 2017b prepared by researchers at Colombia University, Lehigh University and University of Virginia, a method for automated whitebox testing of deep learning systems is proposed. Deep Learning (DL) has made tremendous progress, achieving or surpassing human-level performance for a diverse set of tasks including image classification (He et al. 2016, Simonyan and Zisserman 2014), which has led to widespread adoption and deployment of DL in security- and safety-critical systems such as self-driving cars (Bojarski et al. 2016). Unfortunately, DL systems, despite their impressive capabilities, often demonstrate unexpected or incorrect behaviours in corner cases for several reasons such as biased training data, overfitting, and underfitting of the models (Pei et al. 2017a).

The proposed method aims to identify erroneous behaviours of a DL system without manual labelling/checking, by jointly maximizing a joint objective function combining a metric called neural coverage, and differential behaviour between multiple tested methods. The objective function is maximized by changing the input variable x, under some physical constraints. For example, an input image can be rotated or scaled differently, brightness and contrast can be changed, and rain and fog can be added to the input image.

With *differential behaviour* we mean that when different deep neural networks (DNNs) are tested, the same input will be classified into different classes by the different DNNs. The aim is to maximize the probability that a randomly selected DNN provides an output that differs from the output of the other DNNs. Suppose we have N different DNNs, then each DNN has its own function model $F_k : x \rightarrow y$ for, $k \in 1 \ldots N$, where x and y are the input and output values respectively. If $F_k[c]$ is the class probability that the output of the k-th neural network c is, and the j-th neural network is randomly chosen, the objective function (which will be maximized) is formulated as

$$obj_1(x) = \sum_{k \neq j} F_k(x)[c] - \lambda_1 \cdot F_j(x)[c] \quad (1)$$

where λ_1 is a parameter to balance the objective terms between the DNNs that maintain the same class outputs as before ($F_{k \neq j}$) and the DNN that produce different class outputs (F_j).

Neural coverage is a measure of how many rules in a DNN are exercised by a set of inputs. The neuron coverage of a set of test inputs is defined as the ratio of the number of unique activated neurons for all test inputs and the total number of neurons in the DNN.

To maximize the neural coverage, Pei et al. 2017a and Tian et al. 2017 propose to iteratively pick inactivated neurons and modify the input such that output of that neuron goes above a predefined threshold t. Hence, for a given neuron, n we maximize the following function

$$obj_2(x) = G_n(x) \text{ such that } G_n(x) > t \quad (2)$$

where G_n is the output value of neural n.

The neural coverage and the differential behaviour is jointly maximized, by slightly changing the input values using gradient ascent. The joint objective function is

$$obj_{joint}(x) = \sum_{k \neq j} F_k(x)[c] - \lambda_1 \cdot F_j(x)[c] + G_n(x) \quad (3)$$

By changing the input variables x to maximize this joint objective function, the paper claims that the method finds thousands of erroneous behaviours in fifteen state-of-the-art DNNs trained on five real-world datasets. Hence, new corner cases are explored and different types of erroneous behaviours are uncovered. In addition, test inputs generated by the proposed method can be used to retrain the corresponding deep learning model to improve classification accuracy, and also identify potentially polluted training data (Pei et al. 2017a).

2.8 *Demonstrate need for sensor redundancy*

Inpired by Pei et al. 2017a, as introduced above, we propose to invoke differential behaviour by repeatedly exclude one information source from the sensor fusion machinery. In addition to revealing differential behaviour, we believe this method will be useful to demonstrate the importance of sensor redundancy. If differential behaviour often occurs when a specific information source is removed

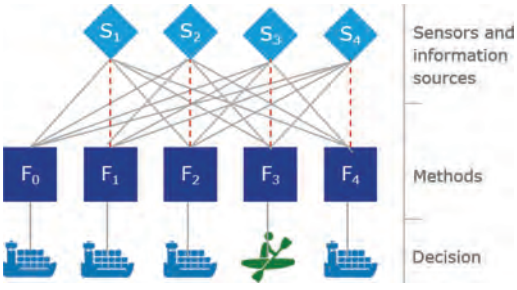


Figure 2. Illustrating how we invoke differential behaviour by repeatedly excluding one information source from the sensor fusion machinery.

from the set of explanatory variables, it can indicate that redundancy of this information is needed to achieve adequate robustness.

To illustrate the idea, we consider a method which fuses four information sources: S_1 a daylight camera; S_2 an IR camera; S_3 a radar; and S_4 AIS (Automatic Identification System). F_0 is the standard method which uses all information sources, while the methods F_k for $k > 0$ cannot use the information from information source S_k. The goal is to change the input variable x in way to invoke differential behaviour as illustrated in Figure 2, where the output of method, F_3 which does not take information source S_k into account, diverges from the other methods.

In the same way as in section 2.7, we let $F_k[c]$ be the class probability that the output of the k-th method is c. Now the k-th method is the method where information source k is excluded as an explanatory variable. In addition, we propose to include method F_0 which includes all variables. Now the objective function (which will be maximized) is formulated as

$$obj_3(x) = \sum_{k \neq j} F_k(x)[c] - \lambda_3 \cdot F_j(x)[c] \quad (4)$$

where j is randomly chosen, and λ_3 is a parameter to balance the objective terms between the method that maintain the same class outputs as before ($F_{k \neq j}$) and the method that produce different class outputs (F_j).

3 ASSURANCE IN THE MARITIME DOMAIN

The assurance of systems which safety is dependent on the accuracy and reliability of data driven models needs to be thoroughly tested. In this chapter, we present challenges related to machine perception that are unique or particularly pronounced in the maritime domain. We describe potential requirements, and discuss how the recommended practices and tools should be used and possibly adapted to form a framework for assurance in the maritime domain.

3.1 *Important technical challenges in the maritime domain*

One of the major differences, relevant for autonomous navigation, between the automotive and the maritime industry is machine perception. Machine perception, also referred to as artificial or digital perception, is the process where information from sensing, maps, satellite data and the vessel condition, are transformed into situation awareness (see Fig. 3).

The requirements of a machine perception system in the maritime industry will most likely concern both what should be *detected*, such as object types, sizes, distances, reflexibilities, etc.; and what should be *classified*, such as ship types, number of ships, seamarks, complexity, etc. The requirements should be evaluated under various external conditions such as weather and daylight.

Several technical challenges, particularly prominent in the maritime domain, remain open as described by for example (Prasad et al. 2016, Prasad et al. 2017):

- *Vessel movements effect on sensors:* For sensors like video cameras which are mounted on-board ships, the unpredictable motion of the ship complicates the object detection.
- *Background subtraction:* The water background is dynamic due to waves. Hence, background learning methods which recognizes background when a pixel stays constant for at least some time, fails. Also, waves and foam are often misinterpreted as foreground objects when using standard methods.
- *Weather and illumination conditions:* The maritime environment is exposed to a variety of different weather and illumination conditions such as fog, rainfall, clouds, bright sunlight, twilight

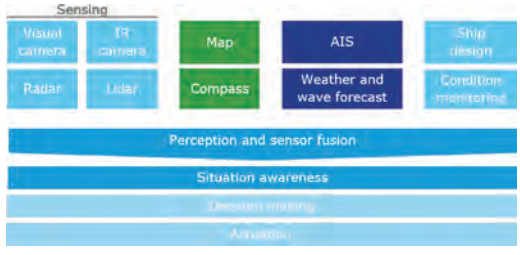


Figure 3. Key components in autonomous navigation in the maritime industry.

and night. The different solar angles pose significant challenges with speckle and glint which makes it difficult to distinguish background and foreground.
- *Insufficient training data from the maritime domain:* Very limited work has been carried out to develop object classification algorithms for objects relevant to the maritime environment. The objects of interest include other ships, leisure boats, kayaks, land marks, buoys, ice bergs, etc.
- *Uneventful sailing:* On ocean going ships especially, a very large fraction of the collected data from a voyage are uneventful, hence a very large portion of the corner cases are left unproven.

3.2 *Operational specific requirements*

Operational specific requirements are not considered in current class rules. We foresee that this might change in the future, especially for ships with autonomous navigation systems, as the operation will be embedded into the technology rather than being the responsibility of the human operator. For example, if the ships perception is limited due to fog, the permitted speed might be lowered, or the ship might be denied access to specific geographical areas, until the weather conditions improve. This decision might also be based on ship type, cargo, manoeuvring capabilities, etc.

3.3 *Triple modular redundancy*

The tools presented in 2.7 and 2.8 above, both search for differential behaviour from multiple algorithms or sensor selections, using a majority organ (or voting circuit). This concept was first described by Von Neumann 1956. Today, the concept is often referred to as *triple modular redundancy* and is perhaps most widely used in space and aeronautics applications (Wu et al. 2017, Yeh 1996), where reliability requirements sometimes are very high. Using the majority vote out of three (or more) methods ensures that a single failure will not cause a system failure. We believe this concept is highly relevant for autonomous navigation, as well as other black box AI algorithms, and believe the use of this should be required, in some form, to ensure sufficient system reliability and robustness.

3.4 *Approval of alternatives and equivalents*

According to the International Maritime Organizations guidelines for the approval of alternatives and equivalents (IMO Maritime Safety Committee 2013), the approval of an alternative and/or equivalent design can be performed by comparing the alternative design to existing designs to demonstrate that the design has an equivalent level of safety. Hence, the approval of autonomous systems used in shipping, including everything from smaller automated tasks to fully autonomously navigated ships, will be based on the equivalence principle: The autonomous functionality must make the operation safer or at least as safe as the conventional operation.

3.5 *Automatic assessment of human perception ability*

To enable the comparison of human and autonomous perception, evaluation metrics and measures should be defined, and performance thresholds should be identified. To achieve this, the human perception in representative real ship operations has to be studied. Research in the field of human errors have shown that a large number of investigated maritime accidents are related to loss of situation awareness (Grech et al. 2002).

It should be noted that the perception ability is not necessarily the ambition. We know that the perception performance can be influenced by many factors such as for example stress, distractions, monotony, boredom, etc. (Horrey et al. 2017, Brodsky and Slor 2013, Schwebel et al. 2012), but our aim is to measure the perception performance in practice.

Simulation tools might be applied to provide more extensive data sets, to complement the data collected from real operation. In addition to increasing the data set, the simulation tool offers the possibilities to create controlled situations, and explore changes in weather, rotated objects, etc. as well as the possibility to explore potentially dangerous situations. Another advantage with the simulated data is that it is pre-labelled, and one will therefore avoid spending time and effort to establish the ground truth on the simulated datasets.

4 CONCLUSIONS

A framework and tentative guidelines for assurance of autonomous systems in the maritime industry are proposed and discussed, with additional focus on the perception and situation awareness functionality. Because vital parts of the autonomous systems, such as the machine perception and situational awareness algorithms, are expected to be partly or fully based on machine learning algorithms, including deep learning, whose functional reasoning is challenging or even impossible to understand and predict, we believe the assurance of such systems are fundamentally different from a traditional assurance process based on physical understanding and theory. Hence, we believe

new guidelines, framework and methodologies are needed.

We propose and describe a range of recommended practices and tools that can be applied to test and validate the ability, performance and robustness of safety critical systems which decisions are based on data-driven methods. We discuss challenges related to machine perception that are unique or particularly pronounced in the maritime domain, and suggest how the recommended practices and tools should be used and possibly adapted to constitute an assurance framework for autonomous navigation in the maritime domain. Furthermore, we discuss potential implications of autonomy such as for example operational dependent requirements. We also discuss the assurance framework for autonomous systems relative to the IMO guidelines for approval of alternatives and equivalents.

ACKNOWLEDGEMENT

We greatly acknowledge the feedback and support from Per Ove Husøy (Kongsberg Digital), Erlend Vågsholm (Kongsberg Maritime) and Jørgen Ernstsen (University College of Southeast Norway).

REFERENCES

Ackerman, E. (2017). How drive.ai is mastering autonomous driving with deep learning. https://spectrum:ieee:org. Date: 10.5.2017.

Arlot, S. & A. Celisse (2010). A survey of cross-validation procedures for model selection. *Statist. Surv. 4*, 40–79.

Bojarski, M., D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L.D. Jackel, M. Monfort, U. Muller, J. Zhang, et al. (2016). End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.

Brandsæter, A. & E. Vanem (2016). Ship speed prediction based on full scale sensor measurements of shaft thrust and environmental conditions. *Submitted.*

Brodsky, W. & Z. Slor (2013). Background music as a risk factor for distraction among young-novice drivers. *Accident Analysis & Prevention 59,* 382–393.

Chen, Z. & T. Ellis (2014). Semi-automatic annotation samples for vehicle type classification in urban environments. *IET Intelligent Transport Systems 9*(3), 240–249.

Cuevas, C., E.M. Yáñez, & N. García (2015). Tool for semiautomatic labeling of moving objects in video sequences: Tslab. *Sensors 15*(7), 15159–15178.

Fei-Fei, L. (2010). Imagenet: crowdsourcing, benchmarking & other cool things. In *CMU VASC Seminar,* Volume 16, pp. 18–25.

Fingas, J. (2017). Rolls-royce unveils plans for an autonomous patrol ship. https://www.engadget.com/2017/09/12/rolls-royce-autonomous-patrol-ship/. Retrieved 10.10.2017.

Grech, M.R., T. Horberry, & A. Smith (2002). Human error in maritime operations: Analyses of accident reports using the leximancer tool. In *Proceedings of the human factors and ergonomics society annual meeting,* Volume 46, pp. 1718–1721. Sage Publications Sage CA: Los Angeles, CA.

Harrald, J.R., T. Mazzuchi, J. Spahn, R.V. Dorp, J. Merrick, S. Shrestha, & M. Grabowski (1998). Using system simulation to model the impact of human error in a maritime system. *Safety Science 30*(1), 235–247.

He, K., X. Zhang, S. Ren, & J. Sun (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 770–778.

Hofmann, P. (2013). Object detection and tracking with side cameras and radar in an automotive context. Master's thesis, Institute of Computer Science of Freie Universitt Berlin.

Horrey, W.J., M.F. Lesch, A. Garabet, L. Simmons, & R. Maikala (2017). Distraction and task engagement: how interesting and boring information impact driving performance and subjective and physiological responses. *Applied ergonomics 58,* 342–348.

Huval, B., T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, R. Cheng-Yue, et al. (2015). An empirical evaluation of deep learning on highway driving. *arXiv preprint arXiv:1504.01716.*

IMO Maritime Safety Committee (2013). Guidelines for the approval of alternatives and equivalents as provided for in various imo instruments (24 june 2013 ed.). *IM Organization, Ed. London: International Maritime Organization.*

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence—Volume 2,* IJCAI'95, San Francisco, CA, USA, pp. 1137–1143. Morgan Kaufmann Publishers Inc.

Kongsberg (2017). Autonomous ship project, key facts about yara birkeland. https://www.km.kongsberg.com/. Retrieved 8.10.2017.

Krotkov, E., S. Fish, L. Jackel, B. McBride, M. Perschbacher, & J. Pippine (2007). The darpa perceptor evaluation experiments. *Autonomous Robots 22*(1), 19–35.

Lambert, F. (2016). Understanding the fatal tesla accident on autopilot and the nhtsa probe. *Electrek, July.*

Lopez-Villa, J., H. Insuasti-Ceballos, S. Molina-Giraldo, A. Alvarez-Meza, & G. Castellanos-Dominguez (2015). A novel tool for ground truth data generation for video-based object classification. In *Signal Processing, Images and Computer Vision (STSIVA), 2015 20th Symposium on,* pp. 1–6. IEEE.

Lützhöft, M. & S.W. Dekker (2002). On your watch: automation on the bridge. *The Journal of Navigation 55*(1), 83–96.

Madrigal, A.C. (2017). Inside waymos secret world for training self-driving cars. https://www.theatlantic.com/technology/archive/2017/08/inside-waymos-secret-testing-and-simulation-facilities/537648/.

MathWorks (2017a). Automated driving system toolbox for matlab. https://se.mathworks.com/products/automated-driving.html. Retrieved: 29.9.2017.

MathWorks (2017b). Computer vision system toolbox for matlab. https://se.mathworks.com/products/computer-vision.html. Retrieved: 29.9.2017.

MathWorks (2017c). Object recognition methods in computer vision. https://in.mathworks.com/discovery/object-recognition.html. Retrieved: 25.9.2017.

Muoio, D. (2016). A start-up born out of stanford just entered the driverless car race with a radical approach. http://www.businessinsider.com/driveai-using-deep-learning-for-its-autonomous-cars-2016-8?r=US&IR=T&IR=T. Retreived: 30.8.2016.

Pei, K., Y. Cao, J. Yang, & S. Jana (2017a). Deepxplore: Automated whitebox testing of deep learning systems. *arXiv preprint arXiv:1705.06640*.

Pei, K., Y. Cao, J. Yang, & S. Jana (2017b). Towards practical verification of machine learning: The case of computer vision systems. *arXiv preprint arXiv:1712.01785*.

Prasad, D.K., C.K. Prasath, D. Rajan, L. Rachmawati, E. Rajabaly, & C. Quek (2016). Challenges in video based object detection in maritime scenario using computer vision. *arXivpreprint arXiv:1608.01079*.

Prasad, D.K., D. Rajan, L. Rachmawati, E. Rajabally, & C. Quek (2017). Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. *IEEE Transactions on Intelligent Transportation Systems*.

Rothblum, A.M. (2000). Human error and marine safety. In *National Safety Council Congress and Expo, Orlando, FL*.

Rout, R.K. (2013). *A survey on object detection and tracking algorithms*. Ph.D. thesis, The department of Computer Science and Engineering of National Institute of Technology Rourkela.

Schöning, J., P. Faion, & G. Heidemann (2015). Semi-automatic ground truth annotation in videos: An interactive tool for polygon-based object annotation and segmentation. In *Proceedings of the 8th International Conference on Knowledge Capture,* pp. 17. ACM.

Schwebel, D.C., D. Stavrinos, K.W. Byington, T. Davis, E.E. ONeal, & D. De Jong (2012). Distraction and pedestrian safety: how talking on the phone, texting, and listening to music impact crossing the street. *Accident Analysis & Prevention 45*, 266–271.

Simonyan, K. & A. Zisserman (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sun, Y., G. Xiong, W. Song, J. Gong, & H. Chen (2014). Test and evaluation of autonomous ground vehicles. *Advances in Mechanical Engineering 6,* 681326.

Sun, Y., G.M. Xiong, H.Y. Chen, S.B. Wu, J.W. Gong, & Y. Jiang (2011). A cost function-oriented quantitative evaluation method for unmanned ground vehicles. In *Advanced Materials Research,* Volume 301, pp. 701–706. Trans Tech Publ.

Tian, Y., K. Pei, S. Jana, & B. Ray (2017). Deeptest: Automated testing of deep-neural-network-driven autonomous cars. *arXiv preprint arXiv:1708.08559*.

Tvete, H.A. (2017). The revolt, a new inspirational ship concept. https://www.dnvgl.com/technology-innovation/revolt/index.html. DNV GL, Retreived: 8.10.2017.

Von Neumann, J. (1956). Probabilistic logics and the synthesis of reliable organisms from unreliable components. *Automata studies 34*, 43–98.

Waymo (2016). Report on autonomous mode disengagements for waymo self-driving vehicles in california. Technical report, Waymo.

Wei, J. & J.M. Dolan (2009). A robust autonomous freeway driving algorithm. In *Intelligent Vehicles Symposium, 2009 IEEE,* pp. 1015–1020. IEEE.

Woodgate, E. (2017). Students design autonomous containers to disrupt sea freight of aquaculture products. https://www.dnvgl.com/. DNV GL Press release, 15.8.2017.

Wu, C.-H., T.-J. Chen, T.-Y. Hsu, S.-H. Tsai, & H.-P. Chang (2017). Design of applying flexray-bus to federated architecture for triple redundant reliable uav flight control system. In *Dependable and Secure Computing, 2017 IEEE Conference on,* pp. 73–78. IEEE.

Yeh, Y.C. (1996). Triple-triple redundant 777 primary flight computer. In *Aerospace Applications Conference, 1996. Proceedings., 1996 IEEE,* Volume 1, pp. 293–307. IEEE.

Zhao, D. & H. Peng (2017). From the lab to the street: Solving the challenge of accelerating automated vehicle testing. *arXiv preprint arXiv:1707.04792*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Subjective assessment of risk among urban work travel cyclists

A.-M. Kummeneje & T. Rundmo
Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: The results of the current research are based on two questionnaire studies of risk perception and worry about being involved in an accident when cycling. The data for Study 1 were collected through a questionnaire survey (n = 291) distributed in collaboration with the Norwegian Cyclists' Association. Study 2 was based on survey data collected in a representative sample of the Norwegian public (n = 2000). The results of the first study showed significant associations between risk perception, worry, and risk tolerance on the one hand and cycling frequency during winter on the other hand. Risk perception was not as important for cycling frequency during the other seasons of the year. Study 2 showed that previous accident experience was associated with risk perception and worry.

1 INTRODUCTION

The risk of being involved in an accident is greater among cyclists compared with users of other travel modes (Høye et al., 2012). However, cycling is a more pro-environmental travel mode than motorized travel modes. Hence, Norwegian transport policy and the Norwegian authorities give high priority to increasing the number of cyclists and frequency of trips made by bicycle, especially in urban areas. Therefore, it is important to examine factors that may be associated with the frequency of bicycle use. Risk exposure may be a particular challenge during harsh winter conditions in Norway, and due to the risk factors, priority should be given to examining risk perception among cyclists. Consequently, the aim of the study on which this paper is based was to investigate how urban work travel cyclists perceive risk when cycling, and to investigate the association between perceived risk and the decision to cycle. An additional objective was to investigate whether experiences of accidents are associated with cyclists' perceptions of risk.

2 THEORETICAL FRAMEWORK

2.1 *Risk perception and worry*

In accordance with Breakwell (2007), studies carried out recent years have used assessments of probability and judgments of the severity of consequences as indicators of perceived risk (Moen & Rundmo, 2006, Nordfjaern et al., 2014, Roche-Cerasi et al., 2013, Rundmo & Moen, 2006). Consequently, in the current study perception of risk was measured using the two factors 'assessments of probability' and 'judgments of the severity of consequences' of an accident occurring when cycling. According to Sjöberg, Moen, and Rundmo (2004), risk perception can be defined as people's cognitive assessment of risk on these two dimensions.

Cognitive models have dominated risk perception and decision-making research, but recently affective processes have received increased attention (Breakwell, 2007). The risk-as-feeling approach highlights the role of anticipatory and anticipated emotions for individuals perception of risk (Loewenstein et al., 2001). Anticipatory emotions are immediate visceral reactions to risk, such as worry, fear, anxiety, and dread. Anticipated emotions are those that the individual expects to feel as a consequence of their decision. There are two types of anticipatory emotions: integral emotions and incidental emotions. Integral emotions are caused by the decision problem itself, whereas incidental emotions are caused by other factors, such as mood (Loewenstein & Lerner, 2003). In this paper, we conceive worry as an anticipatory emotion and integral to the decision problem, which implies that worry is defined as a feeling that emerges as a reaction to the individual's cognitive assessment of risk.

Risk perception and worry are primarily of interest because they may be related to people's behavioural choices. According to the risk-as-feeling approach, behaviour is influenced by the interplay between cognitive evaluations of risk and feelings. Further, emotions often produce behavioural responses that differ from the individual's cognitive assessment of the best course of action. According to (Loewenstein et al., 2001), it is apparent that

when such differences occur, behaviour is driven by emotional reactions, not cognitive assessments.

2.2 *Risk tolerance and priority given to safety*

It is not only important to study risk perception but also how risk is tolerated by individuals. Individuals may differ in their thresholds for the level of risk they find acceptable. In his classic study, Starr (1969) aimed to answer the question 'How safe is safe enough?' He measured the level of risk that individuals found acceptable for different activities, and found that the risks involved in activities that were voluntary and perceived as beneficial were tolerated more than the risks involved in other activities. In a more recent study by Fischhoff et al. (1978), respondents were asked to judge the acceptable level of risk associated with different activities or technologies. The researchers found that risk was less tolerated when the activities were associated with dread. Fischhoff et al. (1978) also found that higher risk levels were tolerated in voluntary activities with well-known and immediate consequences.

To date, few studies have measured risk tolerance among cyclists. Parkin et al. (2007) developed a model based on a risk threshold and provided a measure of acceptability for different cycling routes. The model shows how different infrastructure reduces the perceived risk and makes the route acceptable to cyclists. In Parkin et al.'s (2007) study, the model showed that young and old people considered cycling less acceptable than did people in the age group 35–44 years, and that males considered cycling more acceptable than did females.

The terms 'acceptance' and 'tolerance' are often used synonymously. However, Sjöberg (1999) argues that risk tolerance and risk acceptance are two separate concepts. According Sjöberg (1999, p. 131), 'risks are not accepted but tolerated'. One may be aware of the risk and choose to tolerate it, even if one does not accept it. It is more accurate to talk about accepting, for example, cycling as an activity that generates risk, but the risk in itself is not accepted but tolerated. In our study, we investigated risk tolerance.

Several studies have investigated demands for safety priority or risk mitigation related to transport mode use (Nordfjaern & Rundmo, 2010, Rundmo & Moen, 2006, Simsekoglu et al., 2015, Sjöberg, 1999), and some of these studies have included cycling (Nordfjaern & Rundmo, 2010). To date, no studies have investigated how demands for safety priority relate to transport mode use. Nordfjaern and Rundmo (2010) found that in Norway, the demand for risk mitigation and the priorities related to transport safety increased significantly between 2004 and 2008. Demands for risk mitigation can be defined as public demands for decision-makers to reduce specific sources of transport risks (Moen & Rundmo, 2004). In this paper, we define the demand for safety priority as demands from the public directed towards decision-makers to prioritize traffic safety for cyclists. In the following, we examine how both risk tolerance and priority given to safety are associated with cycling.

2.3 *Risk perception and worry among cyclists*

The study of risk perception among cyclists has received little attention to date (Chaurand & Delhomme, 2013). Much of the research in this field has been conducted mainly to aid engineers and planners wanting to design, improve, and prioritize roads and intersections to cater for cyclists. In the majority of these studies, cyclists were asked to rate their overall risk perception of a route shown in video clips or simulations, or described in surveys. In one recent study, mental mapping was used to study cyclists' risk perceptions by drawing their regular route with different colours according to their perception of safety and risk (Manton et al., 2016). Common to the studies of risk perception among cyclists is that the focus is on the perception of either the road infrastructure or the traffic (Lawson et al., 2013). Some researchers have investigated cyclists' risk perception using specific roads (Llorca et al., 2017) or crossings and roundabouts (Moller & Hels, 2008).

Several previous studies have examined risk perception in relation to travel mode use. Moen & Rundmo (2006) and Oltedal & Rundmo (2007) included cycling with other travel modes when investigating perceptions of risk. Both pairs of researchers found that the probability of being involved in an accident was higher when cycling compared with when using other travel modes. However, the severity of the consequences when cycling was judged as small. Another interesting finding was that respondents reported they were more worried about experiencing an accident when cycling than when they used other travel modes.

Further, researchers have found gender differences in perceptions of risk and worry among cyclists. Lawson et al. (2013) studied gender differences in the perception of safety among cyclists, and found that both males and females more often described cycling as less safe than driving, and that older women perceived cycling as less safe than did younger women. Manton et al. (2016) found that females more often perceived their regular cycling routes as dangerous than did males perceive their own routes. Moen & Rundmo (2006) found that, in contrast to women, men scored lower on their perceptions of probability, on expected consequences,

and on worry. They also found that individuals below the age of 25 years regarded the consequences as least serious and were least worried when travelling by private modes of transport, including bicycles, than were individuals aged 25 years or over. The same age group (i.e. individuals under 25 years) perceived the probability of being involved in an accident as highest. Individuals with a university degree perceived the risk as lower than did others and were less worried about being involved in an accident. Other studied predictors of risk perception and worry are age (Hermand et al., 1999, Sjöberg, 1998) and level of education (Sjöberg, 2000).

2.4 *Aims of the study*

The specific aims of the current study were as follows:

1. To examine differences in work travel cyclists' risk perception, worry, risk tolerance, and priority given to safety when cycling in winter and summer conditions
2. To examine whether risk perception, worry, risk tolerance, and priority given to safety were associated with the cycling frequency during winter
3. To compare the role of risk perception, worry, risk tolerance, and priority given to safety, for cycling frequency during all four seasons
4. To examine the association between accident experience, and risk perception and worry.

3 DATA AND METHODS

3.1 *Samples*

The results of the current study are based on two questionnaire studies about risk perception and travel behaviour among Norwegian cyclists. The data collection for both studies was carried out in spring 2017.

3.1.1 *Study 1*

The data for Study 1 were collected through a self-completion questionnaire survey (n = 291) carried out among work travel cyclists in Trondheim Municipality, Norway. The survey was distributed in collaboration with the Norwegian Cyclists' Association (Syklistenes Landsforening). All respondents cycled on a daily basis and were members of an Internet-based group for cyclists. We invited all 2240 members to participate in the study, and the response rate was 13 percent. Respondents who never used their bicycle for travelling to work or university, or during their workday were not defined as work travel cyclists and were excluded from the sample (28 respondents).

The percentages of females and males in the sample were 35 percent and 65 percent respectively. They were all in the age range 20–67 years (mean = 42.63, standard deviation = 11.30). A total of 69 percent of the respondents reported they had more than three years of university education, 19 percent had three years or less of university education, and 12 percent had received their highest level of education at upper secondary school. A total of 91 percent reported their main occupation as employed, and the remaining 9 percent were students. The respondents reported that they used their bicycle to travel to and/or from work (94%), for their work (13%) and/or to and/or from university (10%). A total of 3 percent of the respondents reported that they did not have a driving license.

Figure 1 shows how often the work travel cyclists used their bicycle during the different seasons. The cyclists reported that they cycled less during winter compared with in the other seasons. During winter, 48 percent cycled five or more times per week, compared with 67 percent during autumn, 70 percent during spring, and 76 percent during summer. A total of 8 percent reported that they never cycled during winter. None of the respondents reported that they never cycled during the other seasons.

3.1.2 *Study 2*

Study 2 was based on a telephone questionnaire survey (n = 2000) carried out among a representative

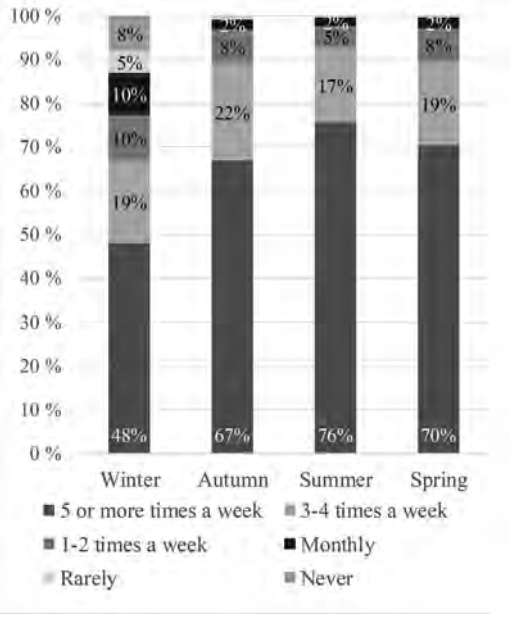


Figure 1. Cycling frequency (%) during winter, spring, summer, and autumn (n = 263).

sample of the Norwegian population in the age range 15–88 years (mean = 45.38, standard deviation = 17.56). The response rate was 27 percent. In the sample, 57 percent were males and 43 percent were females. A total of 9 percent of the respondents had primary or secondary school education as their highest completed education level. A total of 34 percent had upper secondary school as their highest completed education level. A high proportion of the sample (57%) had a higher education level from college or university. A total of 62 percent reported that they were employed or self-employed and 10 percent were students. A total of 10 percent of the responds reported that they did not have a driving license. In the sample, 8 percent (n = 164) had experienced an accident as a cyclist during the last two years. Two-thirds (66%) of the accidents were single accidents (i.e. accidents not involving other road users), and 28 percent of the cyclists involved in an accident needed medical treatment afterwards.

3.2 *Questionnaires and measurement instruments*

In both Study 1 and Study 2, the respondents were asked to evaluate the probability of an adverse event when cycling and the severity of its consequences, and their degree of worry about being in an adverse event when cycling in winter and summer conditions. The questionnaires also contained questions about age, gender, highest level of completed education, employment status, and possession of a driving licence. In addition, the questionnaire used in Study 1 contained questions about risk tolerance, safety priority, and cycling frequency during the four seasons. The survey in Study 2 contained questions about accident experiences as a cyclist during the last two years.

To measure risk perception, the respondents were asked to assess their probability of experiencing an accident or injury, and to judge the severity of the consequences if such an event were to occur. The scale for measuring the probability assessments was a five-point evaluation scale ranging from 'not at all probable' to 'very probable'. For the judgement of severity of the consequences, the scale ranged from 'not at all serious' to 'very serious'. The respondents were further asked to rate how worried they were about being involved in an accident when cycling, and the measurement scale ranged from 'not at all worried' to 'very worried'.

3.2.1 *Study 1*

To measure risk tolerance, the respondents were asked the following question: 'To what extent do you tolerate being exposed to risk when cycling?' The five-point evaluation scale ranged from 'tolerate the risk absolutely' to 'do not tolerate any risk'. To measure priority given to safety, the respondents were asked to assess the following question: 'How important do you think it is that the authorities prioritize measures to improve safety for cyclists?' The five-point scale ranged from 'not at all important' to 'very important'. The respondents were asked how often they cycled each season (winter, spring, summer, and autumn). For this measurement, a six-point evaluation scale was applied: 5 or more times per week, 3–4 times per week, 1–2 times per week, Monthly, Rarely, and Never. This measure has been found appropriate in previous studies of bicycle use in Norway (Kummeneje & Tretvik, 2015, Tretvik, 2015).

3.2.2 *Study 2*

To measure cyclists' experiences of accidents, the respondents were asked whether they had been involved in an accident, including a single accident, as a cyclist during the last two years. If they reported that they had been in an accident, they were further asked if other road users were involved and whether they had needed medical help after the accident.

3.3 *Statistical procedures*

3.3.1 *Study 1*

Paired sample t-tests were used to investigate differences in risk perception (probability and consequences of being in an accident), worry, risk tolerance, and priority given to safety, between cycling in winter and summer conditions. A hierarchical regression analysis was used to predict the amount of cycling done in all seasons.

3.3.2 *Study 2*

MANCOVAs were used to investigate differences in risk perception (probability and consequences of being in an accident) and worry between respondents who had experienced an accident while cycling during the last two years and respondents who did not have the same experience. The results were controlled for gender, age, and education level.

4 RESULTS

4.1 *Risk perception relating to cycling in winter and summer conditions*

The paired sampled t-tests showed significant differences in the respondents' assessment of risk during winter and summer cycling conditions. This was the case for the subjective assessments of the probability of an accident ($t = 5.722$, $p < .001$) and for how worried they were about being involved in an accident ($t = 6.597$, $p < .001$). The respondents' perceived greater risks for cycling in winter

Table 1. Differences in risk perception, worry, risk tolerance, and priority given to safety when cycling in winter and summer conditions (n = 263).

	Winter cycling conditions		Summer cycling conditions		t-value (Sig. 2-tailed)
	Mean	SD	Mean	SD	
Probability	2.94	1.090	2.64	1.014	5.722***
Consequence	3.21	.967	3.24	.944	–.724
Worry	2.60	1.212	2.23	1.015	6.597***
Risk tolerance	2.62	1.083	2.47	1.025	3.320***
Safety priority	4.51	.744	4.56	.622	–1.643

*p < .05, **p < .01, ***p < .001.

conditions compared with cycling in summer conditions. However, it is interesting to note that there were no significant seasonal differences in the respondents' judgements of the severity of the consequences if an accident were to occur. Further, there were significant differences in risk tolerance (t = 3.320, p < .001). With regard to risk tolerance, the larger the mean value was, the less the tolerance for risk was. Table 1 shows that the respondents tolerated less risk when cycling in winter than when cycling in summer conditions. Further, the results revealed that there were no differences in priority given to safety when cycling in winter conditions compared with when cycling in summer conditions (Table 1).

The standard deviations for all variables were relatively high and there were variations in the respondents' perceptions of risk. This was the case for both cycling in summer conditions and cycling in winter conditions.

4.2 *Predictors for cycling frequency during winter*

Table 2 shows the results of a hierarchical multiple regression analysis that were used to predict cycling frequency during the winter season. The independent variables were entered in five blocks: demographics, risk tolerance, priority given to safety, risk perception, and worry. Table 2 shows only the two final steps of the analysis. Demographics were entered as controlling variables in the analysis. Gender and age were found to be significant predictor variables. Female respondents cycled less than male respondents during winter, and the older the cyclists were, the more they cycled during winter. Educational level did not seem to be associated with whether the respondents cycled during winter.

In total, the predictor variables explained an acceptable percentage of variance (adjusted R^2 = .33). The results showed that risk tolerance, when cycling in both winter and summer conditions, were the most important predictors of cycling frequency during the winter. The more the cyclists tolerated exposure to risk when cycling in winter conditions, the more they cycled during the winter. By contrast, the less the cyclists tolerated exposure to risk when cycling in summer conditions, the more they cycled during the winter. The inclusion of priority given to safety improved the model and the explained variance improved significantly, but the variables did not separately significantly influence bicycle use. When risk perception was included in the model as the fourth block of variables, the explained variance improved significantly (Table 2). Both the judgements of probability and the severity of consequences when cycling in winter conditions were found important predictors for cycling frequency during winter. The higher the cyclists perceived the probability of being involved in an accident and the more serious the consequence of an accident was perceived, the less they cycled during winter. The perceived severity of consequences was more strongly correlated with cycling frequency than the probability assessment. Additionally, worry related to cycling in winter conditions was found to be an important predictor of cycling frequency. The more worried

Table 2. Dimensions of cycling frequency during winter, showing only the two final steps of the analysis (n = 263).

	Standardized beta coefficient	
	Model 4	Model 5
Block 1: Demographics		
Gender (male = 0, female = 1)	–.14*	–.13*
Age	.14	.12*
Education level	.08	.11*
Block 2: Risk tolerance		
Risk tolerance, winter conditions	–.49***	–.39***
Risk tolerance, summer conditions	.28	.21*
Block 3: Safety priority		
Safety priority, winter conditions	.15	.11
Safety priority, summer conditions	.04	.08
Block 4: Risk perception		
Probability, winter conditions	–.16*	–.03
Probability, summer conditions	.09	.04
Consequence, winter conditions	–.23*	–.18
Consequence, summer conditions	.16	.15
Block 5: Worry		
Worry, winter conditions		.32**
Worry, summer conditions		.12
Adjusted R^2	.30	.33
F Change	4.216**	6.031**

*p < .05, **p < .01, ***p < .001.

the cyclists were, the less they used their bicycle during the winter.

When worry was included in to the model as the last step in the analysis, the risk perception predictors (probability and severity of consequences) lost prediction power. This may indicate that risk perception has an indirect effect on cycling frequency during winter.

4.3 *Predictors of cycling frequency during all seasons of the year*

Our next step was to predict and compare seasonal differences in cycling. A total of four hierarchical multiple regression analyses were carried out, and the results are summarized in Table 3 and Table 4. The same group of predictors as used previously to predict cycling frequency during the winter were used to predict cycling frequency during all the seasons of the year.

Table 3. Predictors of cycling frequency in winter and summer (n = 263).

	Winter		Summer	
	Adj. R^2	F Change	Adj. R^2	F Change
Block 1: Demographics	.10	9.938***	.02	2.567
Block 2: Risk tolerance	.25	24.964***	.03	2.731
Block 3: Safety priority	.26	3.663*	.03	.580
Block 4: Risk perception	.30	4.216**	.03	.846
Block 5: Worry	.33	6.031**	.03	1.824

*p < .05, **p < .01, ***p < .001.

Table 4. Predictors of cycling frequency in spring and autumn (n = 263).

	Spring		Autumn	
	Adj. R^2	F Change	Adj. R^2	F Change
Block 1: Demographics	.00	1.043	.02	2.244
Block 2: Risk tolerance	.06	8.864***	.05	24.714*
Block 3: Safety priority	.07	2.322	.04	.594
Block 4: Risk perception	.10	2.514*	.10	5.5154***
Block 5: Worry	.11	3.342*	.11	1.117

*p < .05, **p < .01, ***p < .001.

The model explained the largest amount of the variance in cycling frequency during winter (adjusted $R^2 = .33$). The model was least successful in explaining cycling frequency during summer (adjusted $R^2 = .03$). Therefore, we did not find the model a good fit for predicting cycling frequency during summer.

The model explained an identical amount of variance in cycling frequency during spring (adjusted $R^2 = .11$) and autumn (adjusted $R^2 = .11$). For both seasons, risk tolerance and risk perception were significantly associated with frequency of cycling. In addition, worry was significantly associated with frequency of cycling during spring but not autumn.

To summarize, the results showed that risk perception was significantly associated with cycling frequency during winter. However, perceived risk was not strongly associated with cycling frequency during the other seasons. Additionally, worry was found significant for cycling frequency during winter and spring, but not summer and autumn.

4.4 *The role of accident experience in risk perception and worry*

In Study 2, we conducted a MANCOVA to examine whether the group of individuals that had experienced a bicycle accident during the last two years perceived the risk differently from the other individuals in the sample (Table 5), including both their perceived probability being in an accident and the severity of consequences if an accident should occur. In addition, we wanted to investigate differences between these groups with respect to worry about being involved in an accident. The results were controlled for differences in gender, age, and education level. The results showed differences between the two groups for cycling in summer conditions (Wilks' $\lambda = .990$, $F = 6.792$, $p < .001$) and cycling in winter conditions (Wilks'

Table 5. The role of accident experience in risk perception and worry when cycling in summer conditions (n = 2000).

	Accident experience		No accidents last two years		
	Mean	SD	Mean	SD	F (Sig.)
Probability	2.44	1.175	2.16	1.053	10.238***
Consequence	3.28	1.222	3.27	1.253	.084
Worry	2.22	1.191	1.95	1.081	14.233***

*p < .05, **p < .01, ***p < .001.

$\lambda = .995$, $F = 3.232$, $p < .05$). In the following, we only present the results of investigations of differences in risk perception and worry for cycling in summer condition due to the fact that most cycling accidents happen during summer.

As shown in Table 5, individuals that had experienced an accident as a cyclist perceived their probability of being in an accident as higher than did the other individuals. They also tended to be more worried about being involved in an accident when cycling. There were no differences in the perceived severity of consequences between the two groups.

In addition, we conducted two separate MANCOVAs. One tested whether the experience of an accident that involved other road-users would influence the perception of risk or worry, and the other tested whether the need for medical help after an accident influenced the perception of risk or worry. There were no significant correlations.

5 DISCUSSION

The results showed seasonal differences in work travel cyclists' perceptions of risk, worry, and risk tolerance when cycling. The risk of being involved in an accident was perceived as higher and the cyclists tended to be more worried about such incidents and tolerated less risk when cycling in winter conditions compared with cycling in summer condition. Primarily, the probability of being involved in an accident was perceived as higher when cycling in winter conditions. There were no differences in the perceived severity of consequences. With darkness and icy and snowy roads in Norway in winter, it is natural that cycling in winter may be perceived as a bigger challenge than in summer, and the probability of being involved in an accident in winter was perceived as increased. One reason why the consequences were not perceived as increased might have been that the type of accidents the cyclists imagined they could be involved in did not differ with winter and summer conditions.

We found that, on average, the respondents cycled less during the winter season. Further, the results showed that risk perception, worry, and risk tolerance influenced cycling frequency during the winter season. The assessment of risk was less important for cycling frequency during the other seasons of the year. One explanation for this is that risk in general is perceived as very low when cycling in summer condition and that a cyclist has to experience that a risk is over their tolerance threshold before it will influence their behaviour.

In common with other researchers (Hermand et al., 1999, Lawson et al., 2013, Manton et al., 2016, Moen & Rundmo, 2006, Parkin et al., 2007, Sjöberg, 1998, 2000), we found that demographic variables were associated with risk perception, risk tolerance, and worry. Compared with males, females tended to tolerate risk less, and they were more worried about involvement in an accident and perceived the risk as higher.

The results of our study contribute to an understanding of why work travel cyclists in Norway cycle less during the winter than in other seasons of the year. From a pro-environmental perspective, it is important that people who use bicycles for their daily commutes and work-related travel do not change to motorized modes of travel during the winter season. Campaigns that aim to increase the number of cyclists may be ineffective if they do not take into account that the risk of being involved in an accident is perceived differently during the different seasons of the year.

The results from Study 2 showed that accident experience can influence risk perception and worry when cycling. Primarily, the probability of being in an accident was perceived as higher for who had experienced an accident compared with persons who had not experienced an accident when cycling. The perceived severity of consequences did not differ between individuals who had experienced an accident and those who had not, and in general the consequences were perceived as serious. This may have been due to the fact that the cycling accidents reported in the media are often the most serious ones.

The response rate in Study 1 was relatively low and can be regarded as a limitation of our study. However, low response rates do not necessary constitute a methodological problem. Rather, this is only the case if the sample is overall non-representative of the target population (Krosnick, 1999). We did not have any information about the non-respondents, and we cannot draw any conclusions about whether the respondents and the non-respondents differed significantly. However, the target group of the study was work travel cyclists who cycled on a daily base. We assume that those who cycled often also visited the web page, from where the respondents were recruited, more often than other cyclists.

6 CONCLUSION

Study 1 showed seasonal differences in how work travel cyclists perceived the risk and how worried they were about being involved in an accident as well as how they tolerated being exposed to the risk. As expected, cyclists perceived the risk as higher, were less tolerant of risk, and were more worried about being involved in an accident when cycling in winter conditions compared with cycling in summer conditions. Further, the results showed

that risk perception, worry, and risk tolerance influenced cycling frequency during winter. The poor prediction of the model for cycling frequency during spring, summer, and autumn may have been due to the fact that the majority of the respondents perceived the risk as low during these seasons. Lastly, Study 2 showed that previous experience of an accident was associated with risk perception and worry.

REFERENCES

Breakwell, G.M. (2007). *The psychology of risk*. Cambridge: Cambridge University Press.

Chaurand, N., & Delhomme, P. (2013). Cyclists and drivers in road interactions: A comparison of perceived crash risk. *Accident Analysis and Prevention, 50*, 1176–1184. doi:10.1016/j.aap.2012.09.005.

Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., & Combs, B. (1978). How Safe Is Safe Enough—Psychometric Study of Attitudes Towards Technological Risks and Benefits. *Policy Sciences, 9*(2), 127–152.

Hermand, D., Mullet, E., & Rompteaux, L. (1999). Societal risk perception among children, adolescents, adults, and elderly people. *Journal of Adult Development, 6*(2), 137–143. doi:10.1023/A:1021676909857.

Høye, A., Elvik, R., Sørensen, M.W.J., & Vaa, T. (2012). *Trafikksikkerhetshåndboken* (4rd Ed.). Oslo: Transportøkonomisk institutt.

Krosnick, J.A. (1999). Survey research. *Annual Review of Psychology, 50*, 537–567. doi:10.1146/annurev.psych.50.1.537.

Kummeneje, A.-M., & Tretvik, T. (2015). *Sykkelbyundersøkelsen i Region sør 2015*. Trondheim, Norway: SINTEF A27221.

Lawson, A.R., Pakrashi, V., Ghosh, B., & Szeto, W.Y. (2013). Perception of safety of cyclists in Dublin City. *Accident Analysis and Prevention, 50*, 499–511. doi:10.1016/j.aap.2012.05.029.

Llorca, C., Angel-Domenech, A., Agustin-Gomez, F., & Garcia, A. (2017). Motor vehicles overtaking cyclists on two-lane rural roads: Analysis on speed and lateral clearance. *Safety Science, 92*, 302–310. doi:10.1016/j.ssci.2015.11.005.

Loewenstein, G.F., & Lerner, J.S. (2003). The role of affect in decition making. In R.J. Davidson, K.R. Scherer, & H.H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 619–642). Oxford, New York: Oxford University Press.

Loewenstein, G.F., Weber, E.U., Hsee, C.K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127*(2), 267–286. doi:10.1037//0033-2909.127.2.267.

Manton, R., Rau, H., Fahy, F., Sheahan, J., & Clifford, E. (2016). Using mental mapping to unpack perceived cycling risk. *Accident Analysis and Prevention, 88*, 138–149. doi:10.1016/j.aap.2015.12.017.

Moen, B.-E., & Rundmo, T. (2004). *Explaining demand for risk mitigation*. Trondheim, Norway: Rotunde Publ.

Moen, B.-E., & Rundmo, T. (2006). Perception of Transport Risk in the Norwegian Public. *Risk Management, 8*(1), 43–60. doi:10.1057/palgrave.rm.8250003.

Moller, M., & Hels, T. (2008). Cyclists' perception of risk in roundabouts. *Accident Analysis and Prevention, 40*(3), 1055–1062. doi:10.1016/j.aap.2007.10.013.

Nordfjaern, T., & Rundmo, T. (2010). Differences in risk perception, priorities, worry and demand for risk mitigation in transport among Norwegians in 2004 and 2008. *Safety Science, 48*(3), 357–364. doi:10.1016/j.ssci.2009.10.001.

Nordfjaern, T., Simsekoglu, O., Lind, H.B., Jorgensen, S.H., & Rundmo, T. (2014). Transport priorities, risk perception and worry associated with mode use and preferences among Norwegian commuters. *Accident Analysis and Prevention, 72*, 391–400.

Oltedal, S., & Rundmo, T. (2007). Using cluster analysis to test the cultural theory of risk perception. *Transportation Research Part F-Traffic Psychology and Behaviour, 10*(3), 254–262.

Parkin, J., Wardman, M., & Page, M. (2007). Models of perceived cycling risk and route acceptability. *Accident Analysis and Prevention, 39*(2), 364–371. doi:10.1016/j.aap.2006.08.007.

Roche-Cerasi, I., Rundmo, T., Sigurdson, J.F., & Moe, D. (2013). Transport mode preferences, risk perception and worry in a Norwegian urban population. *Accident Analysis and Prevention, 50*, 698–704. doi:10.1016/j.aap.2012.06.020.

Rundmo, T., & Moen, B.-E. (2006). Risk perception and demand for risk mitigation in transport: A comparison of lay people, politicians and experts. *Journal of Risk Research, 9*(6), 623–640. doi:10.1080/13669870600813811.

Simsekoglu, O., Nordfjaern, T., & Rundmo, T. (2015). The role of attitudes, transport priorities, and car use habit for travel mode use and intentions to use public transportation in an urban Norwegian public. *Transport Policy, 42*, 113–120. doi:10.1016/j.tranpol.2015.05.019.

Sjöberg, L. (1998). Worry and risk perception. *Risk Analysis, 18*(1), 85–93. doi:10.1111/j.1539-6924.1998.tb00918.x.

Sjöberg, L. (1999). Consequences of perceived risk: Demand for mitigation. *Journal of Risk Research, 2*(2), 129–149. doi:10.1080/136698799376899.

Sjöberg, L. (2000). Factors in risk perception. *Risk Analysis, 20*(1), 1–11. doi:10.1111/0272-4332.00001.

Sjöberg, L., Moen, B.-E., & Rundmo, T. (2004). *Explaining risk perception: An evaluation of the psychometric paradigm in risk perception research* (T. Rundmo Ed.). Trondheim: Rotunde publikasjoner.

Starr, C. (1969). Social benefit versus technological risk. *Science, 165*(3899), 1232–1238.

Tretvik, T. (2015). *Sykkelundersøkelse 2015 Osloområdet*. Trondheim, Norway: SINTEF A27141.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Applying an operational safety barrier framework in a major oil and gas field development project

J.T. Ludvigsen
Statoil ASA, Oslo, Norway

K. van de Merwe
DNV GL, Høvik, Norway

E. Klemsdal le-Borgne & T. Teigen
Statoil ASA, Oslo, Norway

ABSTRACT: This paper addresses Operational Barrier Elements (OBE) to strengthen the management of Major Accident Hazards (MAH) in major oil and gas field development project. OBE are defined as safety critical tasks that contribute to prevent and mitigate major accident hazards. 87 OBEs were identified, distributed across four platforms and 15 safety performance standards. The majority of the OBEs are related to gas detection, emergency shut-down, active fire protection, and well integrity. The primary output of the activity is Performance Requirements (PR) for each OBE, and recommendations for strengthening human performance through Performance Influencing Factors (PIF). OBEs were integrated in the field wide safety strategies, and the Operational Readiness team incorporated the OBEs in System descriptions and Operational Procedures (SO). Even though no critical deficiencies in design or barriers were identified, systematic integration of OBEs in engineering and operations will strengthen the fields robustness to withstand major accidents hazards in operations.

1 INTRODUCTION

1.1 *Background*

In recent years there has been a renewed focus on the concept of safety barriers, including safety critical tasks, and its role to prevent and mitigate major accident hazards on the Norwegian Continental Shelf. Partly, this is due to the Macondo accident where human and organizational factors contributed to escalation of the accident (CSB, 2016). In volume 3 of the Macondo well incident investigation report, CSB assert that the safety barrier concept must extend beyond physical safeguard, and that solutions to technical failures cannot prevent future incidents without giving equal attention to failures of less visible, non-physical barriers and support systems (CSB, 2016, pp. 19).

The Norwegian Petroleum Safety Authority (PSA) reinforces these conclusion by describing the principles for barrier management in the petroleum industry (2013; 2017). Here, safety barrier management is defined as the systematic effort to ensure that barriers are in place to provide protection of hazards and accident situations (PSA, 2017).

Traditionally, major accidents risk on offshore petroleum installations is reduced by striving for inherently safe design, automation systems and fail-safe principles. PSA (2017), however, states that the design of safe and robust installations is not sufficient to protect against failures and hazard and that accident will continue to happen. It is postulated that most major accidents are significantly influenced by human actions, both in cause and consequence. Hence, since the safety of petroleum installations are dependent on reliable human performance, safety barrier management must incorporate operational and organizational aspects, in addition to the mere technical ones (PSA 2017).

Barriers and controls mechanisms are often visualized as a bow-tie diagram, with a top event in the center of the model, with the hazards and preventive barrier on the left-hand side, and barrier for mitigation of the event on the right-hand side (Ruijter & Guldenmund, 2014). In Leva, Angel, Plot & Gattuso, (2013) the bow-tie model is applied to assess a scenario where human actions are the main barrier to accidental conditions (overfilling a storage vessel).

The concept of operational barriers in the petroleum industry are addressed in several recent publications. In addition to the PSA, the Norwegian Shipowner Association's (NSA) report on good practices for barrier management for the rig industry (NSA, 2014) provide definitions and framework that are in line with the methodology outlined in this paper. Similarly, Hauge & Øien, (2016) provide practical guidelines for barrier management in the petroleum industry, in accordance to the recommendations in PSA (2013).

In a recent paper on human factors in barrier management, McLeod (2017) describes three generic types of barriers against threats, and their order of importance, or expected strength:

- Engineered controls, which are physical, technical and automated barriers
- Organizational control, including governing systems and procedures, team organization etc.
- Human Controls, which are the individual factors such as training, competence etc.

According to McLeod (2017), human performance is central in maintaining and ensuring the integrity of barriers. Instead of considering human error as a hazard, the focus should be on improving the resilience of barriers, to be able to withstand human errors that degrade the barriers.

1.2 *Safety critical tasks*

Task analysis is the study of what a person is required to do, in terms of actions and mental processes, to achieve a goal (Kirwan & Ainsworth, 1992). For SCTA, Energy Institute (EI) extends this to tasks which contributes to MAHs in positive or negative ways including: initiating events; prevention and detection; control and mitigation, and emergency response. EI list the following steps of SCTA:

1. Identify main MAHs
2. Identify safety critical tasks
3. Understand tasks
4. Represent critical tasks
5. Identify human failures and PIF
6. Determine safety measures

According to OGP (2011), a safety critical task inventory should be established for projects involving major accident hazard potential, e.g. to summarize all human tasks which are identified a safety barrier.

Human performance of safety critical tasks is shaped by a range of factors, which, depending on the operational context, can drive performance both in a negative and positive direction. Systematic management of such PIFs is a key to ensure that Operational Barrier Elements (OBE) are effective.

Different methodologies apply different sets of PIFs. As a basis for establishing a standard framework of PIFs, various methods were reviewed. PIFs generally can be classified into four main categories: According to Boring & Blackman (2007, p. 177), a PIF is an aspect of the human's individual characteristics, environment, organization, or task that specifically decrements or improves human performance, thus respectively increasing or decreasing the "likelihood of human error" (Boring & Blackman, 2007, p. 177).

In the context of operational barriers or safety critical tasks, PIFs are the factors which have a significant effect on barrier element performance. In this context, the definition refers to individual, workplace or other contextual factors which have significant effect on an operator or crews of operator's performance. A task based approach to OBE enables application of Human Reliability Analysis (HRA) to assess the qualitative or quantitative reliability of Safety Critical Tasks. According to Bye et al (2017), HRA can be applied to assess and understand human actions as barriers in major accidents.

Quantification of Human Error Probability can provide valuable input to decision making, i.e. whether to implement design measures such as technical systems to eliminate risk, or whether the risk can be considered acceptable, and operational measures are sufficient. Petro-HRA applies a list eight Performance Shaping Factors (PSF), which are modified for a petroleum context based on SPAR-H (U.S. Nuclear Regulatory Commission, 2005). According to Bye et al. (2017) only the PSFs that have been shown in general psychological literature and in other HRA methods to have a substantial effect on human performance when performing control room tasks are included in Petro-HRA. This includes:

- Task factors: task complexity, time, procedures
- Individual factors: threat stress, experience/ training
- Organizational factors: teamwork, attitudes,
- Environmental factors: physical working environment.

In the framework described in this paper, the PIFs defined in Bye et al. (2017) was selected for assessment of the safety critical tasks. In engineering, the first step of strengthening task performance is through task design, e.g. removing the task entirely by automating a manual task, increasing level of automation or introducing additional technical barrier elements. Other measures to improve human reliability is adapting the physical configuration, conditions and surroundings to human cognition and anthropometrics, e.g. through human centered design (ISO 11064-1).

1.3 *Definitions*

Requirements to operational barriers are outlined in Statoil governing documentation, and chapter 5 of PSA management regulation. Safety barriers shall be effective in all situations, and the role to maintain its intended safety functions must be well-known across the organization PSA, (2017).

This paper applies definitions provided in the Statoil internal report Definitions and guidelines for non-technical barriers (Gould, Sklet & Ludvigsen, 2015), which aimed to standardize the definition and application of operational barrier elements in safety barrier management in Statoil.

In this report**, an operational barrier element** is defined as the safety-critical tasks performed by a person, or team of personnel, which realizes one or several barrier functions.

A technical barrier element is defined as engineered equipment, systems and structures which realize one or several barrier functions.

A barrier function consists of the range of technical systems, structures, personnel and tasks, which are required to fulfill ("realize") the barrier function. These are referred to as barrier elements, and can be defined as technical or operational measures which alone or together realize one or several barrier functions.

Performance Requirements (PRs) shall be established both for technical and operational barrier elements. Performance requirements for operational barrier elements include a description of how the task should be conducted, who is responsible for performing a task (the formal role), and if possible, when (time criterion).

Safety Critical Tasks (SCT) are the physical actions or activities by which human performance contributes positively or negatively to major accident risk, through either:

- Initiation of events;
- Detection and prevention;
- Control and mitigation; or,
- Emergency response

A similar definition is provided by the Energy Institute's report on Safety Critical Task analysis (2011). However, the SCT term is broader and covers a wider range of tasks than just OBEs. Some tasks can be critical because of their indirect influence on barrier performance. This typically refers to inspection, testing and maintenance of technical barrier elements.

This paper aims to capture SCTs that can be defined as OBEs, i.e. tasks that have a direct and real-time role in realizing a barrier function related to preventing or mitigating major accident hazards in an accident sequence are included. The work aimed to strengthen barrier performance by systematically addressing operational elements throughout engineering, to enable barrier management in operations.

OBE includes task required for realization of on-demand barrier functions, where human performance directly contribute to the availability or integrity of a barrier function, as maintaining the primary barrier during well operations, crane operations, and containment during maintenance, etc.

Barrier functions may be gas detection, Emergency Shut Down (ESD), blowdown, etc., where human action may prevent an accident, or where an omission or error of commission may directly contribute to the unavailability or degradation of a barrier function.

In this context, major accidents are accidents leading to multiple fatalities, major environmental harm or loss of an asset. The term "hazard" is a potential source of harm. The term "defined situations of hazard and accident" (DSHA) refers to a selection of hazardous and accidental events that will be used for the dimensioning of the emergency preparedness for the activity.

1.4 *Limitations*

Process disturbances or minor deviations during normal operations are excluded, because they are controlled by inherent safety measures or by control mechanisms that prevent escalation.

Tasks which have an indirect influence on the barrier function are excluded, and for which routine work and operational safeguards are an established part of planned maintenance or safe work process. This may be task such as:

- Routine inspections to check condition
- Maintenance, calibration and testing
- Inhibitions of safety systems as part of planned operations
- Work permits and safe job analysis
- Purely administrative tasks, such as handovers, issuing work orders etc.

The framework does not address barriers that protect against risks related to working environment, personal injury, security, or production regularity. Topics such as strategic management of change, organizational learning, and safety culture are all considered important, are also excluded from the framework described in the paper. This is in line with guidelines stipulate in Gould et al. (2015).

2 CASE STUDY

2.1 *Project description*

The project is one of the five largest oil and gas fields on the NCS. The resources are estimated to be between 2.0–3.0 billion barrels of oil equivalents,

and peak production will reach 660,000 barrels daily. The field will be operated by electrical power generated onshore, reducing offshore emissions of climate gases by 80%–90% compared to installations utilizing gas turbines.

The field development consists of four interconnected topside platforms, including a Living Quarter with more than 500 beds, Drilling Platform with utility module and integrated drilling facilities, Process Platform including 3 stage separation process and gas compression equipment, and a Riser Platform with oil and gas export risers.

The engineering of the different platforms is performed by different engineering teams at separate locations. The field has several safety critical interfaces, such as facility for converting power from shore, gas export pipeline to an onshore plant and oil export pipeline to an onshore oil refinery.

The entire field will be controlled from the Control Centre, including Central Control Room and Emergence Control Centre located at the Living Quarter. The size, complexity and value of the field make high demands on safety and productivity. Production start for Phase One is planned for late 2019.

2.2 *Scope*

This paper describes the effort put into operationalizing the theoretical framework presented above through identifying, analyzing and defining operational barriers in a major oil & gas field development project. The goal of mapping and assessing OBEs was to define the factors that shape human performance, and to strengthen these factors through engineering and operational readiness. This will enable a systematic framework for management of operational safety barriers in operation.

3 METHODOLOGY

3.1 *Mapping and assessment of operational barrier elements*

The methodology for mapping and assessment of OBE is described in a project specific Terms of reference (ToR) document. The purpose of the ToR is to establish guiding principles for the work, and ensure standardization across different platforms and contractors. The method is divided into two main steps, summarized in the section below.

3.2 *Identification and assessment of safety critical tasks*

Data required for identification and assessment of Safety Critical Tasks (SCT) was primarily collected by means of document reviews and workshops with Operations and technical disciplines (Energy Institute, 2011).

First, Safety Critical Tasks were identified by reviewing project information and documentation, and attending workshops (HAZOPs, LOPAs, etc). Experience from other installations and projects was also gathered as input to the SCTA inventory.

A set of meetings and workshops were arranged to collect and analyze the data, task identification and screening. The meetings followed the activities described in the project Terms of Reference, and mostly involved representatives from Statoil Operations and relevant technical disciplines on each platform.

Second, a screening of the SCTA in terms of risk and relevance, and where it could be consisted as critical for realizing a barrier function. Screening criteria consisted of:

- Barrier functions that are depended on the task performance
- The consequence of human failure
- The severity of the MAH
- Task familiarity and complexity

The outcome of the screening was a list of proposed Operational Barrier Elements. The aim of the screening analysis was to reduce the number of tasks such that further work would be limited to the ones significantly contributing to the risk picture. For the tasks identified as highly critical against a set of pre-defined criteria, more detailed task and failure analysis was conducted:

- OBE task analysis; provide task description and assessment of Performance Influencing Factors
- OBE failure analysis; objective is to identify and mitigate risk associated with human failures
- Assessment of Performance Influencing Factors
- Definition of design requirements

Although the definition of major accident refers to safety, environment and asset losses, only OBEs which have a role in managing MAHs with an impact to safety and environment were considered.

A set of criticality levels were used to identify tasks where the barrier function itself is automatic, but where the operator(s) has a back-up role in monitoring and correcting system performance (if necessary).

For example, the task "Manual cancellation of Emergency Depressurization DP in case of gas leak in the flare system/ KO drum" was considered as highly critical. One of the concerns about this task was whether it was applicable to all or just a specific set of scenarios involving gas leaks in the flare system or Knock-Out drum. Discussions focused on whether the Control Room Operator would be able to perform this task safely and reliably, and if there will be enough time available for decision-making and action. To be able to address

these questions in a systematic manner a set of different scenarios was assessed.

3.3 *Definition of performance requirements and performance influencing factors*

This step consisted of definition of performance requirements to each OBE, and providing recommendations about how to improve design and operations to optimize relevant PIFs.

PSA (2017) states that definition of performance requirements is an essential part of establishing effective barrier management. Establishing PRs to OBEs is an integral part of this. Performance Requirements (PRs) are verifiable requirements related to barrier elements to ensure that the intended barrier function is effective (PSA, 2017). For each OBE, the following performance requirements are defined:

1. Coarse description of how the task should be conducted
2. Who is responsible for the task (formal role)
3. When the task need to be conducted
4. If feasible: time available for task performance
5. If feasible: criterion for successful task performance

PRs were established as part of a series of OBE tasks analysis workshops relevant technical disciplines and operations. The agenda of the workshops was structured according to a standard list of Safety Performance Standards (PS), and the OBEs related to these. Participants of the workshops included process technicians (end users), technical safety, process and instrument system responsible and platform management. The information from the OBE analysis was structured according to the following parameters:

- Unique ID
- Reference to relevant safety strategy
- Platform and Areas
- Major Accident Hazard
- Barrier Function
- Performance Standard
- Operational Barrier Element
- Performance Requirement

The main output from this stage included draft Performance Requirement for each OBE, and Design and Operational recommendation, proposed for strengthening the PIFs.

3.4 *Quality assessment and alignment meetings*

Since the OBE framework adapted by the project was new, alignment with key stakeholders in the project was necessary. End-user representatives participated during all phases of the work, however alignment meeting with engineering and operational management was required to ensure mutual understanding of purpose, scope and method, agree upon the quality of the results, and how the results best should be applied by operational readiness.

4 RESULTS

The results section is divided into four sections, including; (1) Operational Barrier Elements and Performance recommendations; (2) inclusion in safety barrier strategies; (3) implementation of recommendations in design; and (4) incorporation of results in Operational Readiness.

4.1 *Operational barrier elements and performance requirements*

87 OBEs were identified for the field development project, and distributed across 15 safety performance standards (see Figure 1 below). Almost half of the OBEs were categorized as generic, indicating that they are applicable to the entire field, not to specific platforms. The drilling platform has highest recorded frequency of OBEs (22), followed by the riser platform (11), process platform (5), and Living Quarter (4) (see Figure 1 below).

Figure 2 illustrates how OBEs are distributed across performance standards. The PSs with

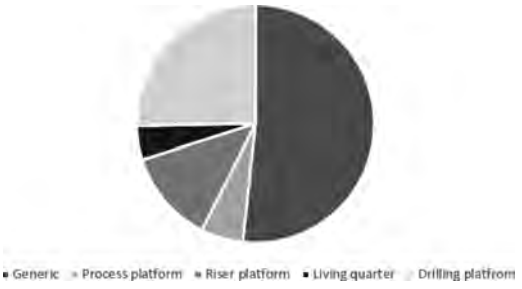


Figure 1. Distribution of operational barrier elements per platform.

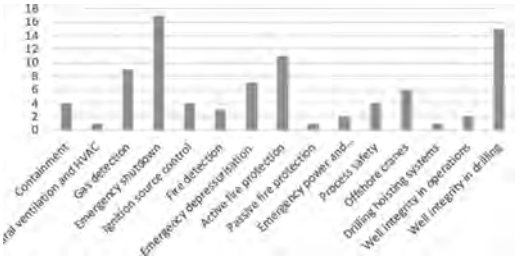


Figure 2. Distribution of operational barrier elements per safety performance standard.

Table 1. Example A: Respond to hydrocarbon leakage from oil export riser.

MAH	DSHA 01: Hydrocarbon Leaks
PS	P1 – Containment
Barrier function	Reduce/limit hydrocarbon leaks from risers or pipelines
OBE	Respond to hydrocarbon leakage from oil export riser (air gap area) or pipeline.
PR	Onshore site has a dedicated system for leak detection on the oil export pipeline going from JS. Onshore CCR is responsible for notifying the JS CCR in case a leak is suspected. JS CCR shall investigate and confirm that a leak is occurring. In case of a confirmed leak, JS CCR shall manually shut down the oil export pumps, close the riser ESDV, and confirm that pressure has dropped. Onshore CCR shall be informed of all mitigating actions.
Design measures	HMI system 21: Trykktransmitter olje eksport <tag> Priority 1 alarm Text: "Olje eksport fra pumper A/B/C til rørledningen"
SO	Included in system description for system 30

Table 2. Example B: Respond to H2 alarm from battery room.

MAH	DSHA 01: Hydrocarbon Leaks
PS	PS 3 – Gas detection
Barrier function	Prevent ignition of H2 gas
OBE	Respond to H2 alarm from battery rooms.
PR	Upon H2 gas detection in battery rooms, the CRO shall electrically isolate the area using the MEI function on the CAP panel. The CRO shall communicate with the Electrical team to ensure the affected room remains unoccupied until the H2 gas has been dispersed via the HVAC and the H2 alarm has been cleared.
Design measures	New HMI symbol H2 gas Priority 1 alarm Alarm text: "Informer over PA Følg instruksen i DFU"
SO	Included in system description for system 84 and 85

highest number of OBEs are Emergency Shutdown (17), Well integrity in drilling (15), Active Fire protection (11), Gas Detection (9), Emergency depressurization (7) and Offshore Cranes (6).

Two examples of how the results are structured are provided below. These data are extracted from the OBE master database.

4.2 *Inclusion of OBE in safety barrier strategies*

During engineering, five separate safety strategies were established, i.e. one for each facility/platform, and one for the entire field. The safety strategies have a standard organization, included common definitions of barrier terminology. A purpose of the safety strategies is to outline the role of safety barriers to be implemented to manage the risks that have been identified for the project. The strategy is used by Operations to understand how risks of major accidents are managed on the field and for each specific area. OBEs and associated Performance Requirements are written into the project safety strategies, established during engineering. The OBEs are related to the relevant safety barrier functions and technical barrier elements.

4.3 *Implementation of recommendation in design*

To ensure traceability and inclusion in design, the main recommendation was included in action follow-up system to be formally handled as part of engineering quality system.

A total of 81 design measures were implemented. It should be noted that these measures are in addition to the Human Factors Engineering process performed in the project (ISO 11064). Most of the findings were related to enabling situation awareness by means of alarm priority and alarm descriptions, operator station HMI and Large Screen Display design. In total, 45 measures were included on HMI (large screen and operators displays), and 38 measures were implemented in the configuration alarm system (alarm priority, alarm text or operator guidance, etc.). This included input to HMI for platform crane cabin and drillers cabin.

The PS that have the highest number of measures are ESD, (23) active fire protection (15), and gas detection (13), see Figure 4 below.

4.4 *Operational readiness*

An essential goal for mapping and assessing of OBE was that the results should be adopted by

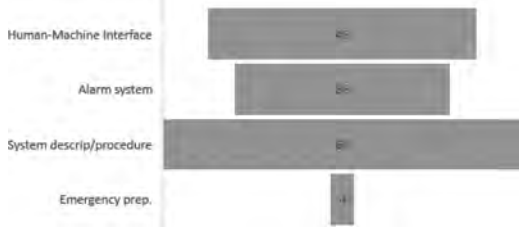


Figure 3. Record of recommendations implemented in design and operational readiness.

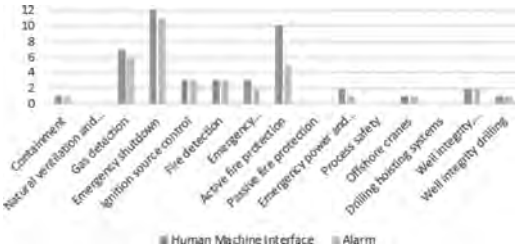


Figure 4. Record of recommendations implemented in HMI and alarm system per performance standard.

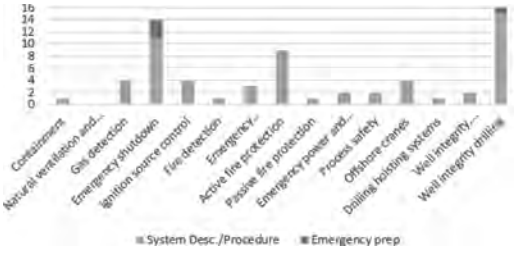


Figure 5. Record of reccomendations implemented in system descriptions & procedures, and emergency preparedness plan.

operational readiness for use in operations. The outcome of mapping and assessment of OBE included a set of recommendations for improvement of human performance through operational readiness.

When the results were finalized, the information was documented in a master OBE database, with additional parameters:

- Reference to existing work process or procedure for a generic OBE
- Actions
- Status of action, i.e. implemented in SO, design etc.

The master database is shared by engineering and operations. The results from the master list is maintained by engineering, and implemented in operational documentation by Operational Readiness.

These are measures such as inclusion of OBE in system description and operational procedures, referred to as SO documents. These documents are organized according to system number, so the results were converted from PSs to systems.

In total, 64 measures were implemented in System Descriptions and Operational Procedures (SO).

The results in the master database was reviewed as part of developing the Emergency Preparedness Plan. Four elements from the database were used as direct input to these plans. One of the purposes of the SO documentation is to provide as a basis for operator training, which is the next step of the project.

5 CONCLUSION

This paper described the systematic integration of operational barrier elements in design and barrier management to strengthening an oil and gas field's robustness to withstand major accidents hazards. The results show how the barrier framework was successfully applied in an engineering context. The results presented indicate how the project captured the criticality of the human role in preventing and mitigating major accident hazards, and ensuring efficient detection, initiate or control effective emergency shutdown, and realize sufficient active fire-fighting means. Also, the human role as a continuous operational barrier in ensuring sufficient well integrity in drilling, and safe operations of offshore cranes is assessed and described.

An essential part of a control room operator role is to monitor and control that safety and automatic systems perform on demand, and act accordingly if these systems fail. It was decided to include the most critical of these control tasks, because they provide valuable input to system design, procedures and other PIFs. Also, it gives an improved understanding of the interaction between technical and operational barrier elements across different accident scenarios.

By limiting OBE to mere cases where the barrier function is entirely dependent on a direct, physical human action, we would not capture the factual human role in protecting and mitigating against MAH, and the dependency between technical and operational barrier elements during accident.

Applying this framework early in detailed engineering allowed for providing input to design. Throughout engineering, adjustments have been made to HMI, alarm system, but also the physical design of the facility to either remove/reduce human intervention or to optimize human reliability. Furthermore, applying the framework in engineering allowed a systematic focus on safety critical tasks in developing procedures, emergency preparedness plans and training. Inclusion of OBE in safety strategies help to improve risk awareness in operations, as OBE and PRs shall be known by the management and the roles responsible for realizing the OBEs.

A central concept in human factors engineering is the use of iterative design cycles. Likewise, it is beneficial to allow for several iterations when working with operational barriers. For example, the project would have benefited from starting earlier than detailed design (e.g. Front-End Engineering and Design), however at a cost of available details to work with. The OBE work for this

project started in Detailed Design and applied several iterations to come to sufficient maturity before starting developing procedures and training.

Currently, the project has focused on identifying and defining OBE and their corresponding PR, facilitating operator reliability through design adjustments, procedures and training. The next steps in the barrier framework are to set up assurance and verification activities to ensure barrier functions remain intact under daily operations and to ensure that the performance of the barriers is working as intended (NSA, 2014).

The framework described in this paper demonstrated the benefits of including assessment of safety critical tasks and PIFs in an engineering context. Second, the performance requirements for OBEs have provided operational readiness with risk-informed input to development of system descriptions, procedures and training program of the operational phase. Third, the framework provides a better understanding of the mutual dependency between operational barrier elements and technical barrier elements in realizing barrier functions, thus contribute to strengthen the field's safety barrier management.

ACKNOWLEDGEMENTS

We would like to acknowledge the contributions made by Sondre Øie and Barnaby Annan for their technical knowledge and developing the project specific methodology.

REFERENCES

Boring, R.L., Blackman, H.S. 2007. The origins of the SPAR-H method's performance shaping factor multipliers. *Official Proceedings of the Joint 8th IEEE Conference on Human Factors and Power Plants and the 13th Annual Workshop on Human Performance/Root Cause/Trending/Operating Experience/Self Assessment*: 177–184.

Energy Institute. 2011. *Guidance on human factors safety critical task analysis.* Retrieved from http://publishing.energyinst.org/special-offers/free-to-download/human-and-organisational-factors/guidance-on-human-factors-safety-critical-task-analysis.

Gould, K., Sklet, S., & Ludvigsen. J.T. 2015. *Definitions and guidelines for non-technical barriers*. (Statoil Internal Technical Report).

Hauge, K. & K. Øien 2016. *Guidance for barrier management in the petroleum industry* (Report No. A27623). SINTEF Trondheim. Retrieved from https://www.researchgate.net/publication/309319877.

Institute for Energy Technology. 2017. *The Petro-HRA Guideline.* (Report no. IFE/HR/E-2017/001). Retrieved from https://www.ife.no/no/publications/2017/mto/the-petro-hra-guideline.

International Association of Oil and Gas Producers (IOGP). 2011. *Human factors engineering in projects*. (Report No. 454). Retrieved from https://humanfactors101.files.wordpress.com/2016/02/human-factors-engineering-in-projects.pdf.

International Organization for Standardization. 2000. *ISO 11064-1. Ergonomic design of control centres. Part 1: Principles for the design of control centres.* Retrieved from https://www.iso.org/standard/19042.html.

Kirwan, B. & Ainsworth, L. 1992. *A Guide to Task Analysis*. Taylor and Francis, London.

Leva, M.C., Angel, C., Plot, E., & Gattuso, M. 2013. When the human factor is at the core of the safety barrier. *Chemical Engineering Transactions*, 33: 439–444.

McLeod, R.W. 2017. Human Factors in Barrier Management: Hard Truths and Challenges. *Process Safety and Environment Protection*, 110: 31–42.

Norwegian Shipowners' Association (NSA). 2014. *Barrier Management in Operation for the Rig Industry—Good Practices*. (Report No. 2013–1622). Retrieved from https://www.rederi.no/DownloadFile/?file=12349.

Petroleum Safety Authority (PSA). 2016. *Regulations relating to management and the duty to provide information in the petroleum activities and at certain onshore facilities (the management regulations)*. Retrieved from http://www.ptil.no/getfile.php/ 1341748/Regelverket/Styringsforskriften_e.pdf.

Petroleum Safety Authority (PSA). 2013. *Principles for barrier management in the petroleum industry*. Retrieved from http://www.ptil.no/getfile.php/1319891/PDF/Barrierenotatet%202013%20engelsk%20april.pdf.

Petroleum Safety Authority (PSA). 2017. Principles for barrier management in the petroleum industry. Barrier Memorandum 2017. Retrieved from http://www.ptil.no/getfile.php/1344810/PDF/BARRIERS%20memorandum%202017%20eng.pdf.

Ruijter, A.de. & Guldenmund, F. 2016. The bowtie method: A review. *Safety Science,* 88: 211–218.

U.S. Chemical Safety and Hazards Investigation Board (CSB). 2016. *Investigation report volume 3 drilling rig explosion and fire at the Macondo well*. (Report NO. 2010-10-I-OS 4/17/2016). Retrieved from http://www.csb.gov/assets/1/7/Macondo_Vol3_Final_staff_report.pdf.

U.S. Nuclear Regulatory Commission. 2005. *The SPAR-H Human Reliability Analysis Method*. (Report NO.NUREG/CR-6883). Retrieved from https://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr6883/cr6883.pdf.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Naturalistic decision making in process control: The guidance-expertise model and the model of resilience in situation

S. Massaiu
Institute for Energy Technology, Halden, Norway

ABSTRACT: The defining elements of naturalistic decision-making, such as proficient decision makers, ill-defined goals, uncertainty, high stakes, tools, and teamwork, are clearly present in process control. However, the domain is still heavily anchored in normative approaches for design, analysis and evaluation of human-technology systems that make unrealistic assumptions about the operators. The paper presents two naturalistic decision making models for process control developed in the nuclear power production sector and based on extensive observations of control room emergency operation in high-fidelity, human-in-the-loop simulators: the Guidance-Expertise Model (GEM) and the Model of Resilience in Situation (MRS). Unlike better-known naturalistic decision-making models, the GEM and MRS models recognize the central role that operating procedures and other organizational prescriptions play in process control decision making, elaborating on aspects that so far have received little attention in the naturalistic decision making community.

1 INTRODUCTION

1.1 *Process control and naturalistic decision making*

Control centers operators in process industries supervise systems that are extensively automated and intervene in case of malfunctions. Control centers are technological environments in which the operators' interactions with the system are mediated by human-machine interfaces and supported by decision aids ranging from pen and paper to intelligent expert-systems. The single most important decision support aid used by control centers operators is represented by the operating procedures. In the nuclear energy sector, for instance, when an emergency arises the control room operators respond by immediately opening the emergency operating procedures and implementing these until the reactor is brought to a safe state. In this environment understanding decision-making requires understanding how the operators interact with the procedures. The second distinguishing aspect of decision-making in process control industries is the collective nature of the decisions. Most of the times the decisions are made by groups, often called operating crews. Even when single operators are the decision makers, the high level of proceduralization of their work implies some sort of deferred relation with other actors like procedure designers, trainers, or management, who sets expectations on the decisions and behavior of the front-line operators. The most critical decisions in process control centers occur during incidents and accents. In such conditions the decision landscape typically include elements of time pressure, stress, ill-defined or conflicting goals, uncertainty about conditions, and high stakes. In other words, process control is a good illustration of all defining elements of naturalistic decision-making (Lipshitz, Klein, Orasanu, & Salas, 2001). At the same time the central role of operating procedures (and other organizational prescriptions) for process control is an aspect that sets it apart from other, more studied, naturalistic settings and makes widely-known naturalistic decision making models less readily applicable to the sector.

1.2 *The prevalence of normative approaches in process control*

Although process control decision-making is clearly a naturalistic setting, normative approaches still dominate system analysis, evaluation and design in the sector. This is likely a legacy from when these industries viewed their technical core areas as sufficient for designing productive and safe systems. The nuclear industry, for example, did not pay attention to human and organizational factors before the Three Mile Island and Chernobyl accidents, assuming that reactor physics, thermodynamics and other technical factors were solely responsible for designing safe plants (Moray & Huey, 1988). In normative approaches the operators are assigned the role of executors of procedures. As the procedures incorporate the rational

benchmark for how to behave in different situations, they define what the operators are required to do in terms of actions, and even cognitive processes, in order to achieve the system's goals. The situations for which the procedures are made are seen as predictable conditions in which there are limited ways of performing the tasks correctly. Although these assumptions may have been appropriate for workers of the first industrial revolution, they are inadequate for automated and computerized production processes (Vicente, 1999). These are characterized by external disturbances (unanticipated faults, automation failures) and other forms of uncertainty (degraded indicators) for which the procedures do not apply and in which the operators are required to adapt to moment-by-moment changes in conditions by generating appropriate responses based on their conceptual understanding of the work domain. In such cases there are no standards of correct performance and no predefined correct decisions, if not after the fact. Research from anthropology, activity theory and naturalistic decision making has show that "workers' actions frequently do not—and indeed, should not—always follow these normative prescriptions" (id., p. 62). From the point of view of this paper the main problem of relying on normative approaches to decision making in process control is that the assumptions they make on the operators are not realistic and therefore of limited use for the design, analysis and evaluation of systems in which human and organizational factors play a critical role.

2 NATURALISTIC MODELS OF PROCESS CONTROL DECISION-MAKING

This section describes two naturalistic decision making models for process control work. They are both based on extensive observations of nuclear power plant control room emergency operation in high-fidelity, human-in-the-loop simulators. This section provides a description of the models' theoretical foils, concepts, purposes and applications. Their limitations will be discussed also taking into account their level of maturity and intended use.

2.1 *The guidance-expertise model of procedure following*

The Guidance-Expertise Model of Procedure Following (GEM) is a methodology to describe and predict nuclear power plant operating crews' behavior in emergency operation (Massaiu, Hildebrandt, & Bone, 2011; Massaiu & Bones, 2011). The model, based on the framework of cognitive system engineering (Rasmussen, Pejtersen-Mark, & Goodstein, 1994) assumes

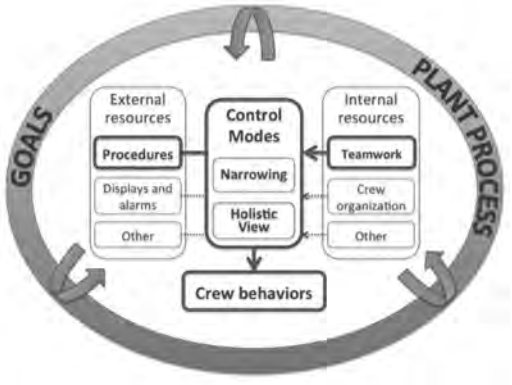


Figure 1. Conceptual diagram for the Guidance-Expertise Model of procedure following (GEM).

that macrocognitive processes, or control modes, determine decisions in proceduralized operational environments (Figure 1). In cognitive system engineering control is defined as the way the next action at any given point in time is chosen. Cognitive control refers to the organization of cognitive functions (e.g. monitoring, evaluating) and processes (e.g. heuristics, externalization of memory storage) in a situation (Hollnagel, 1998). Control modes in the language of activity theory are different orientations towards the object of activity that result in different ways of acting (Kaptelinin & Nardi, 2006). Basically, the control modes are different ways of thinking in a situation that determine behavior.

The GEM model borrows the language of 'situated course-of-action' theory (Theureau, Jeffroy, & Vermersch, 2000) and defines three control modes: Narrowing, Holistic View and Persisting Narrowing. During narrowing the operator's horizon is limited by what is referenced by the procedures. When the operators read the procedure in this mode their attention is focused, the situations are classified by their structural-mechanistic features, and are mapped into pre-planned methods of action. During narrowing, information is not actively searched for except for what is required by the procedure. The pre-planned methods for actions are incorporated in the emergency procedures as well as previously trained. Narrowing in a broader sense is the cognitive control mode in which the operators let the procedures guide them and do not try to figure out ad-hoc plans for dealing with the situation.

Holistic view occurs when the operators interpret the procedure steps relative to the situation dynamics, which include the activities of automated systems and of other people. As such it is analogous to the concepts of situation awareness and sense-making. In holistic view information is actively searched

for in the control panels and through communication with other crewmembers to create an interpretation of the situation in a functional way, by considering the process as an integral whole. Holistic view takes into account larger time windows: the present is explained as the effect of previous events (diagnosis), the future evolution is represented to evaluate if the planned course of action is appropriate (Lipshitz et al., 2001). Holistic view includes metacognition (thinking about thinking) which will manifest itself in activities such as reconsidering the course of actions, determining whether to act outside the procedure while following them, redirecting procedure paths, detecting strains in teamwork and making adjustments.

Switching from one control mode to the other can be challenging in several ways, as the two modes require different cognitive effort and different configurations of cognitive functions, including how much is memorized and consciously represented. Changing from holistic view to narrowing implies difficulties in establishing the necessary local focus and attention as well as the right procedures progression pace. It can also challenge the capacity to re-construct an uncompleted course of action from the exact point it was left. Low-level errors, like slips and lapses, might occur as a result. Yet, most performance difficulties occur when the required control mode is holistic view but the crew struggle in establishing it, thereby continuing in narrowing mode. A typical example is when the crew starts engaging in problem solving behavior (e.g., discussion of transfer points, evaluation of novel events) but ends up reverting to literal procedural adherence. In such cases sustained periods of narrowing impede the achievement of a holistic view (i.e., a level of situation awareness adequate to develop an autonomous plan of action). This state is termed '*persistent narrowing*' and is considered a third control mode. The longer the narrowing continues, the higher the risk of losing global situation understanding. For self-paced processes, like emergency operation in nuclear power plants, the more the crew lags behind the process the more it will be pushed into a reactive mode and the more difficult it will be to achieve holistic view. Persistent narrowing ends when the crew is able to constructively generate a solution strategy, even if this is not an effective one.

In order to make predictions regarding crew behavior in emergency situations, the GEM model relates the control modes to aspects of the situation on one side and to aspects of crew expertise on the other. Regularities among situations, control modes and expertise for specific operational settings are derived from empirical human-in-the-loop simulations. The model outcome behaviors are task-independent behaviors that nuclear power plant operating crews exhibit in emergencies (Table 1).

Table 1. The 16 outcome behaviors included in the GEM model are task-independent behaviors that nuclear power plant operating crews exhibit in emergencies.

#	Outcome behaviour
1	Slow progression by meticulous procedure following
2	Slow reaction to *recently discovered* information
3	No reaction to *important* information received
4	No/slow action to unexpected event
5	Literal step following rather than purpose
6	Incorrect procedural transition
7	Cue explained away
8	Notes/warning/foldout pages ignored
9	No priority between concurrent goals
10	Priority given to minor goal/most recent deviation
11	Autonomous decision avoided
12	Successful step execution
13	Incomplete step execution
14	Inference to previous condition not made
15	Sub-step skipped
16	Stuck in procedure step

According to the model, two formal features of the emergency procedures are the most important environmental aspects to consider. Procedural features are identified as being either 'loose' or 'detailed', with the understanding that there is variance of the degree. Broadly speaking, when the procedures provide meticulous step-by-step direction they are pronounced to be "detailed", otherwise as "loose" (e.g., evaluation of trends, adjustment and control actions, navigational decisions).

Empirically observed regularities between procedural features and control modes should help predicting the crews' procedure progression in accidental conditions. According to the model the procedures-behavior pairings are in fact determined by the control modes, which in turn are determined by the crew expertise. In GEM expertise is measured by teamwork indicators through a classification scheme based on both generic and nuclear-specific process control teamwork literature (Braarud & Johansson, 2010; Klein, 1999; Norros, 2004; Salas et al., 2005; O'Connor et al., 2008). The model considers 8 teamwork dimensions: monitoring progress, communicating intents, communicating interpretations, looking for same cues, reconciling viewpoints, adapting, backing-up, team monitoring & flexibility.

The GEM model has thus three elements: (1) structural features of the procedures, (2) control modes, and (3) the crew expertise.

2.1.1 *Application of the GEM model*

The GEM methodology was tested by Massaiu and Bones (2011) in a retrospective analysis of

74 critical decisions by four Nuclear Power Plant (NPP) operating crews' in simulated Steam Generator Tube Rupture (SGTR) events (Massaiu & Bones, 2011). The four crews that exhibited most operational difficulties (the models' outcome behaviors) were selected to test whether the model could describe the observed performances and help estimating the likelihood of the outcome behaviors given observable aspects of the context of action (i.e., the procedural features) and the crew cognitive control modes.

The analysis showed different patterns of outcome behaviors, control modes and procedural conditions as well as different patterns between the crew expertise (as measured by indicators of teamwork proficiency) and observed control modes. For instance, almost all the times a crew transitioned to a wrong procedure (outcome behavior 6 in Table 1) it was with 'loose' procedural guidance and in 'persistent narrowing' control mode. Different teamwork characteristics were associated with the three control modes. The preliminary results showed that nearly all instances of positive teamwork were observed under the 'holistic view' control mode, and that all negative teamwork dimensions were observed when the crews exhibited 'persistent narrowing'. No positive teamwork indictors and some negative indicators were observed when the crews were in 'narrowing' mode (and to a lesser degree than when in 'persistent narrowing').

2.1.2 *Evaluation of the GEM methodology*

The GEM model can be used as classification system for analysis of observed team decision-making and behavior. Its main benefit is the possibility of evaluating the likelihood of outcome behaviors given observable features of the environment and to measured team expertise. The intended use of the methodology is for cognitive simulation and predictive task analysis. However the methodology has been tested on a small data set only and a number of challenges remain to be solved. These are the main ones:

1. The set of outcome behavior is not a complete and not-overlapping set.
2. Further aspects of the guidance system should be included (e.g., the crew operating policies and training).
3. The model is currently limited to two environmental features: the procedural features and the crew expertise (which determines the control mode). Although these are recognized as the most important factors for crew performance in emergency, the inclusion of other structural features of task and environment in the model is likely necessary for improving its predictive accuracy.

2.1.3 *The model of resilience in situation*

The Model of Resilience in Situation underlies the human reliability method MERMOS (Pierre Le Bot, Cara, & Bieder, 1999). Its primary application has been in the context of predictive risk analysis, but it has proved a valid tool for retrospective accident analysis in the nuclear (Le Bot, 2004) and medical (Le Bot, 2008) fields. Recently it has been proposed as a way to analyze human and organizational factors in a High Reliability Organizations perspective for supporting design of risk-critical systems (Le Bot & Pesme, 2014).

The MRS explains how operating teams in emergency organizations make decisions during the course of an accident. The model is based on Jean-Daniel Reynaud's theory of social regulation (Reynaud, 1989), a sociological theory that understands social relations (particularly in the working environment) as social regulations, that is, the social production of formal and informal rules governing the behavior of groups.

In the MRS the object of analysis is the Emergency Operating System (EOS), the ensemble of control room operating crew, the human-machine interface, and the operating procedures. The MRS is about team decision-making mediated by technology and procedures and thus consistent with Edward Hutchins' distributed cognition paradigm (Hutchins, 1995; Hutchins & Klausen, 1998) in assuming that cognitive resources are not only in the operators' heads but also in the procedures and the interface.

The MRS can be seen as constituted by three interrelated building blocks: (1) a description of the dynamics of emergency operation (Figure 2); (2) the functions that the EOS fulfills during emergency (Figure 3), and (3) the stable characteristics, or features, of the emergency operating system (see Table 3 for an example) (Massaiu & Braarud, 2013).

The dynamics of emergency operation are described by cycles of stability, ruptures, and new stability phases (Figure 2). During a stable

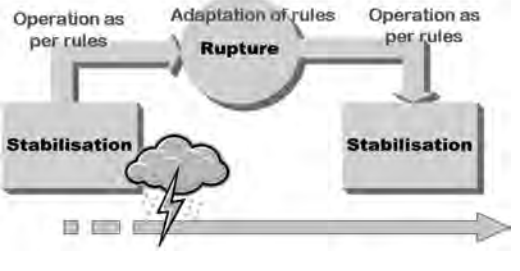


Figure 2. The dynamics of emergency operation according to the model of resilience in situation.

Figure 3. The Model of Resilience in Situation (MRS) identifies five functions of the Emergency Operating System (EOS): 'Execution' and 'Control' ensure the system's robustness, 'Verification' and 'Reconfiguration' the system's adaptation, while 'Information selection and sharing' is a cross-cutting function enabling the other functions.

Table 2. The MRS model specifies 'sub-functions' and 'details' for the main functions of the model depicted in Figure 3. Here are the sub-functions and details for 'Control' (of rule execution).

Sub-function	Detail
Understand goals and priorities	Understand timing of tasks (when to do, when to get info, time lags, urgency)
Allocate resources (cognitive, material, human)	Distribute tasks/Position operators in CR
	Understand task allocation
Continuous monitoring of expected plant responses	
Small deviations detections and adjustments	Keep focus on task and process
Recovery (of individual errors)	Team monitoring, communicate significant actions
	Consult and peer check before performing significant actions (feed forward control to avoid need for recovery)
Concentrate on current plan	Avoid distractions: Do not respond to all incoming information/requests
	Attention on procedure following, read notes, read foldout, referenced parameters
Resist external demands (for resources)	Keep focus on priorities
Reach plan goals	Ensure goals achieved
	Completing pending procedures/steps
Manage dynamics (e.g. Concurrent goals)	Manage multiple/parallel tasks (procedures)
	Manage interruptions and deferred tasks (including continuous EOPs' steps and conditions)

Table 3. The emergency operating system 'Features' are the structural elements that determine the systems' capacity to perform its functions (Figure 3). Here the EOS sub-features and indicators for the category 'Procedures'.

Sub-feature	Indicators
Monitoring/ re-evaluation loops	Re-evaluate procedure appropriateness
	Re-evaluate procedure optimality
	Continuously/periodically re-evaluate priorities
Writable/ bookmarks	To aid memory
Redundant information sources	Look for extra information to validate itself
	Look for extra information to assess reliability of cues
Overview/status trees	Counter fixation on current plan
	Takes into account simultaneous influences
	Easy to look ahead/browse

operating phase, called the stabilization phase, the system follows the effective rules that it has set itself, typically the operating procedures, allowing the attainment of its objectives and avoiding the continual demands made by the dense flow of information (for instance, several hundred alarms in a nuclear power plant control room). However, this organizational inertia, protecting the actors from unexpected demands, must be counterbalanced by permanent redundant verification (or monitoring), i.e., constantly checking that the rules applied are appropriate to the situation (for example, that the procedure in effect is adequate). A rupture occurs when the active rules become inappropriate and the operating system has to be reconfigured so that it has new effective rules. This can happen for two reasons: (1) the objectives may have been reached in compliance with the applied rules; or (2) the rules are not longer adequate due to (2a) errors in rule implementation necessitating a reconfiguration (re-planning) that is more than mere error recovery, or (2b) when the team recognizes that conditions existed or have newly arisen for which the rules in effect are not adequate. In these cases, the verification of the of rule's inadequacy should trigger a "rupture" of the operation so that the system reconfigures itself with new effective rules. (Figure 2).

It should be noted that during emergencies at nuclear power plants the rupture phases may last minutes while the stability phases may last hours.

The second building block of the MRS model is the description of the functions of the operating system (Figure 3).

There are two main functions that define the resilience of the EOS: "Robustness" and

"Adaptation". Adaptation is accomplished by the functions described above of verification (i.e., verifying that the plans are good for the situation) and reconfiguration (the capacity to timely produce plans that fit changed conditions). Robustness is defined by Execution and Control. Execution is defined as "acting on the process given the effective operating rules". It includes object discrimination (selecting the right control out of similar ones and the right mode in multi-mode displays) and situation discrimination (acting differently in different plant operating modes). "Control" consists in a permanent monitoring of the consistency of actions and effective operating rules (are the rules well applied?). Control is about the execution of the rule in effect, is the function that ensures that the rule is being implemented as intended. Effective control requires continuous monitoring of process and staff, detection of deviations, rapid adjustments, and management of concurrent demands and interruptions.

In order to perform these functions the system has also to select and share information from the environment. Information Selection is then defined as a "common function" needed by Control, Verification and Reconfiguration. Teamwork is treated as a set of processes (e.g., cooperation, team situation understanding) used in performing EOS functions. Therefore, teamwork functions are not represented as independent functions but are 'built-in' in the other functions.

The third building block of the MRS model is constituted by the "emergency operating system features", i.e., stable characteristics of the system that allow it to perform its functions. Features are identified for the personnel (e.g., staffing, training, safety culture), the human-system interface (e.g., displays, alarm logs) and the procedures (e.g., symptom based) elements of the system. The EOS features determine the systems' capacity to perform its functions. Different configurations of personnel, HSI, and procedures will produce different capabilities with regard to the various EOS functions. For instance, an operating crew with authoritarian line of command will likely facilitate execution and control functions, but might counter effective reconfiguration. The MRS model organizes the features under the following categories: Team, Prescriptions, Formal communications, Human-Machine Interface (HMI), Training, and Procedures (see for instance the features for Procedures in Table 3 below). These six categories include sub-features, that is, specific indicators that evaluate their contribution to the fulfillment of the EOS functions. For example, the HMI feature includes the "Control Room Layout" sub-feature to evaluate whether the HMI provides "visibility of other operators" and "visibility of others' actions" (i.e. "does the control room layout allow the operators to see each other and their actions?"). Another example is the Team feature's sub-feature "Supervisory function", that evaluates among others the system capability of "Monitoring others actions" and "Searching redundant information" (i.e. "does the supervisor monitor operators actions and search for redundant information?").

The MRS model specifies the influences of the EOS features on the EOS functions. The result is a complete matrix of influences from the features' indicators to the sub-functions' details.

2.1.4 *Evaluation of the MRS methodology*

The Model of Resilience in Situation is the theoretical backbone of the human reliability method MERMOS, and in this form it has been applied in the French nuclear industry for more than a decade. The decision-making model presented in this paper has received a more limited application and testing, but it nonetheless has proved capable of capturing the essentials aspects of the decision-making processes followed by nuclear control room crews responding to simulated accidents in full scope simulators. These were detailed, minute-by-minute analyses of teams of professional operators performing in realistic conditions. Compared to other naturalistic decision-making models the MRS model treats decision processes that span over relatively long time windows and include several decision points, as it is necessary for dealing with emergency operation in nuclear power plants. A second innovative aspect, also strongly dependent on the model's domain of origin, is the importance reserved to the technology and the organizational environment in which the decision makers operate. These are furthermore teams rather than individuals so that teamwork aspects become central. Finally, the model lends itself to predictive applications (trough the MERMOS human reliability method), retrospective accident analysis, for verification and validation purposes (by providing an overall framework that can be used as basis for performance-based evaluation of human-machine systems), and as an observation protocol for on-line recording and classification.

The main limitation of the MRS model is that, beside the applications for human reliability which is at an industrial stage, the methodology needs further refinement and testing (Massaiu & Braarud, 2012).

3 CONCLUSION

Normative approaches are still the preferred option for analysis, design and evaluation of human-technology systems in process control industries.

This is partly due to the field strong technical tradition, one that assumes that the core technical disciplines are sufficient for achieving safe and productive systems, but also to field specificities (like the prominence of operating procedures) that make well-known naturalistic decision making approaches less readily applicable.

This paper has presented two decision-making models specifically developed in process control settings. The models are informed by extensive empirical observations and have been tested and implemented at varying degrees for different applications. The two models contribute to the naturalistic decision making discipline at large by tackling the not so well-studied aspect of team decision making *with* operating procedures.

REFERENCES

Braarud, P. O., & Johansson, B. (2010). *Team Cognition in a Complex Accident Scenario* (No. HWR-955). Halden, Norway: OECD Halden Reactor Project.

Hollnagel, E. (1998). *Cognitive Reliability and Error Analysis Method*. Elsevier Science.

Hutchins, E. (1995). *Cognition in the Wild* (New edition). The MIT Press.

Hutchins, E., & Klausen, T. (1998). Distributed Cognition in an airline cockpit. In Y. Engeström and D. Middleton (Eds.), *Cognition and Communication at work*. Cambridge University Press.

Kaptelinin, V., & Nardi, B. A. (2006). *Acting with technology: activity theory and interaction design*. MIT Press.

Le Bot, P. (2004). Human reliability data, human error and accident models—illustration through the Three Mile Island accident analysis. *Reliability Engineering & System Safety*, *83*(2), 153–167.

Le Bot, P. (2008). Analysis of the Scottish case. In *Remaining Sensitive to the possibility of Failure* (Vol. 1). Ashgate Publishing.

Le Bot, P., Cara, F., & Bieder, C. (1999). MERMOS, a second generation HRA method: what it does and doesn't do. In *Proceedings of the international topical meeting on Probabilistic Safety Assessment (PSA'99)* (Vol. 2, pp. 852–880). Washington DC, USA.

Le Bot, P., & Pesme, H. (2014). Organising Human and Organisational Reliability. In *12th Probabilistic Safety Assessment and Management Conference*. Honolulu, Hawaii.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, *14*(5), 331–352.

Massaiu, S., & Bones, A. (2011). *Emergency procedures and crew behavior: A Retrospective test of the Guidance-Expertise Model* (No. HWR-995). Halden, Norway: Halden Reactor Project.

Massaiu, S., & Braarud, P. Ø. (2012). *Emergency Operating Systems Profiling: Proposals for Developing the Model of Resilience in Situation and for Classifying EOS Features* (Internal report No. IFE/HR/F-2012/1541). Halden, Norway: Institute for Energy Technology.

Massaiu, S., & Braarud, P. Ø. (2013). Including Organizational and Teamwork Factors in HRA: the EOS Approach. Presented at the EHPG 2013, Storefjell Resort Hotel, Gol, Norway: Halden Reactor Project.

Massaiu, S., Hildebrandt, M., & Bone, A. (2011). The guidance-expertise model: Modeling team decision making with emergency procedures. Presented at the International Conference on Naturalistic Decision Making, 10 (NDM 2011), Orlando.

Moray, N. P., & Huey, B. M. (1988). *Human factors research and nuclear safety*. National Academies.

Norros, L. (2004). *Acting under uncertainty: The Core-Task Analysis in ecological study of work*. Helsinki, Finland: VTT Technical Research Centre of Finland.

O'Connor, P., O'Dea, A., Flin, R., & Belton, S. (2008). Identifying the team skills required by nuclear power plant operations personnel. *International Journal of Industrial Ergonomics*, *38*(11–12), 1028–1037.

Rasmussen, J., Pejtersen-Mark, A., & Goodstein, L. P. (1994). *Cognitive systems engineering*. Wiley.

Reynaud, J.-D. (1989). *Les règles du jeu: l'action collective et la régulation sociale*. Colin.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in Teamwork? *Small Group Research*, *36*(5), 555.

Theureau, J., Jeffroy, F., & Vermersch, P. (2000). Controling a nuclear reactor in accidental situations with symptombased computerized procedures: a semiological & phenomenological analysis. *CSEPC 2000 Proceedings*, 22–25.

Vicente, K. J. (1999). *Cognitive work analysis: toward safe, productive, and healthy computer-based work.* Routledge.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Sensemaking and resilience in safety-critical situations: A literature review

S.S. Kilskar
SINTEF, Trondheim, Norway

B.-E. Danielsen
CIRiS NTNU Samfunnsforskning, Trondheim, Norway
NTNU, Trondheim, Norway

S.O. Johnsen
SINTEF, Trondheim, Norway
NTNU, Trondheim, Norway

ABSTRACT: Recent accidents and near-accidents, such as the capsizing of the anchor handling vessel Bourbon Dolphin in 2007 and the unintended list of the drilling rig Scarabeo 8 in 2012, underline the need for addressing sensemaking in safety-critical situations within the maritime domain. This paper is a literature review to answer the research question: *What are the characteristics of sensemaking and resilience in safety-critical situations?* The aim was to establish more knowledge on sensemaking in safety-critical situations and the relationship between sensemaking and resilience. The majority of authors provide definitions based on Weick's work on sensemaking, describing sensemaking as a social process, involving the extracting of cues and enactment to create meaning to events retrospectively. Few authors provide descriptions that characterise sensemaking in safety-critical situations. There is a lack of literature regarding sensemaking in safety-critical situations in the maritime domain that addresses the issues of training and human-machine interactions.

1 INTRODUCTION AND OBJECTIVE

1.1 *Background*

This literature review is part of a research project (SMACS) that addresses the issue of sensemaking in safety-critical situations within the maritime domain. The aim of sensemaking processes in an organisation is to provide meaning to an event or situation in a given context. In such situations, sensemaking can be a source of resilience, in that it enables a person or a crew to "bounce back" when put under stress. Hence, the review not only focuses on sensemaking in safety-critical situations, but also on how the literature describe the relationship between sensemaking and resilience. This paper describes the search strategy and presents the results from the literature review.

1.2 *Purpose and research question*

The purpose of the review was to answer the research question: *What are the characteristics of sensemaking and resilience in safety-critical situations?* This was done by establishing a knowledge base on sensemaking in safety-critical situations, and by exploring the relationship between sensemaking and resilience. In addition, we wanted to examine whether this literature addresses sensemaking in relation to training or Human-Machine Interaction (HMI). This review is not specific to the maritime domain, but it was of interest to be able to later narrow it down to maritime operations.

1.3 *Concepts and definitions*

Sensemaking and *resilience* are two central concepts in this study, both of which have been approached within different theoretical frameworks. In the following, we provide definitions and explanations for these terms, as well as for our use of the term *safety-critical*.

1.3.1 *Sensemaking*

Sensemaking has been of interest in the on-going research project, since the concept supports the idea that human actors in safety-critical operations and their actions are dependent on the whole socio-technical systems consisting of organisational, technological and human factors.

The concept of *sensemaking* started to emerge in organisational literature in the late 1960s (Maitlis &

Christianson, 2014), but was made prominent by Karl E. Weick in 1995 with his seminal book *Sensemaking in Organizations*. In this work, Weick summarised the sensemaking research up to that point and presented seven key properties of sensemaking; 1) grounded in identity construction, 2) retrospective, 3) enactive, 4) social, 5) ongoing, 6) focused on and by extracted cues, and 7) driven by plausibility rather than accuracy. Sensemaking has since been the subject of considerable research and there is an extensive variation in how the term is defined in the organisational literature (Maitlis & Christianson, 2014).

Weick et al. (2005) describe sensemaking as "a sequence in which people concerned with identity in the social context of other actors engage ongoing circumstances from which they extract cues and make plausible sense retrospectively, while enacting more or less order into those ongoing circumstances" (p. 409). Maitlis & Christianson (2014) developed a definition of sensemaking rooted in recurrent themes found in their literature review: "A process, prompted by violated expectations, that involves attending to and bracketing cues in the environment, creating intersubjective meaning through cycles of interpretation and action, and thereby enacting a more ordered environment from which further cues can be drawn" (p. 67). There are several factors that can influence sensemaking. Sandberg & Tsoukas (2015) found from their literature review that context, language, identity, cognitive frameworks, emotion, politics and technology constitute the main factors.

Sensemaking is thus a *process* triggered by ambiguous events that interrupt an ongoing activity and make individuals question what is going on. Individuals will extract cues from the environment that are interpreted and they act on those interpretations and revise them through the consequences of their actions. This is an ongoing cycle and according to Weick (1995) sensemaking never starts or stops as people are always in the middle of things. The events that trigger sensemaking can range from unplanned to planned events and from minor to major events (Sandberg & Tsoukas, 2015).

Sensemaking has been described as an individual cognitive process that has to do with interpretation and development of mental models (Elsbach et al., 2005). However, Weick (1995) described sensemaking as a social process where people actively shape each other's meanings, and argued that even individual sensemaking is influenced by the actual, imagined or implied presence of others.

The concept has traditionally been seen as a retrospective activity that occurs as people look back on action that has already taken place (Maitlis & Christianson, 2014). Weick (1995) argued that people can know what they are doing only after they have done it. The notion of prospective sensemaking (i.e. consideration of impact of actions) has long been a part of the literature (Gioia et al., 1994). In recent years prospective or future-oriented sensemaking has gained more attention (e.g. Gephart et al., 2010; Rosness, Evjemo, Haavik & Wærø, 2016).

1.3.2 *Resilience*

The commonly used definition of safety has been "freedom from unacceptable risk". In resilience engineering safety is defined as the ability to succeed under varying conditions (Hollnagel et al., 2011). Resilience Engineering is concerned with understanding the normal functioning of socio-technical systems and how they perform under varying conditions. Thus, performance variability is not a threat that should be avoided by the use of constraining means; in complex socio-technical systems variability is considered normal and necessary. In this view it is equally important to study things that go right as things that go wrong, with the aim to reinforce the variability that leads to positive outcomes (Hollnagel et al., 2011). Ibid define *resilience* as "the intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances, so that it can sustain required operations under both expected and unexpected conditions" (p. 275).

According to Hollnagel et al. (2011), there are four corner-stones that characterise resilient systems: 1) the ability to respond to events, 2) to monitor ongoing developments, 3) to anticipate future threats and opportunities, and 4) to learn from past failures and successes alike.

1.3.3 *Safety-critical*

In this paper, we use the term *safety-critical situation* or *safety-critical operation* to denote situations or operations that, if they go wrong, have a large potential for causing harm to people, property or environment.

2 METHODOLOGY

A literature search was conducted to establish a knowledge base on sensemaking in safety-critical situations, as well as the relationship between sensemaking and resilience. Literature was obtained through Boolean searches of the following interdisciplinary databases: Scopus, Web of Science, Google Scholar and Oria. Based on the objective of the study, the keywords *sensemaking*, *resilience* and *safety-critical* were selected as the most relevant. In addition, some of the searches in the abstract databases, Scopus and Web of Science, included the keyword *maritime*. To capture variations in these keywords, more specific search terms were used as shown in Table 1. Different combinations of the terms in Table 1 were used due to different search approaches in the various databases.

Table 1. Search terms.

Keyword	Search terms
Sensemaking	sensemaking; sense-making; sense making
Resilience	resilience; resiliency; resilient
Safety-critical	safety-critical; safety-critical situation(s); safety-critical operation(s); safety critical; safety critical situation(s); safety critical operation(s); high-risk; high-risk situation(s); high-risk operations(s); high risk; high risk situation(s); high risk operation(s); hazardous; hazardous situation(s); hazardous operation(s)
Maritime	maritime; at sea; boat; vessel; offshore

Broad search terms (e.g. "high-risk") were used when searching the abstract databases, Scopus and Web of Science, whereas searches in Google Scholar and Oria were conducted using more specific terms (e.g. "high-risk situation").

To avoid an excessive amount of search results, general searches in Google Scholar covered all three keywords of sensemaking, resilience and safety-critical. When searching the abstract databases, and when using the "all in title" function in Google Scholar, search terms related to two of the three keywords were used. To be included, the documents either had to address sensemaking in the context of safety-critical situations or operations, or discuss a relationship between sensemaking and resilience. For this reason, documents discussing sensemaking in other contexts were

Table 2. Overview of the literature.

Author(s)	Year	Publication type	Topic
Weick	1993	Journal paper	Disruptions of sensemaking
Gephard	1997	Journal paper	Quantitative sensemaking during crises
Beunza & Stark	2004	Working paper	Organisational resilience in a Wall Street Trading Room After 9/11
Furniss et al.	2009	Workshop paper	Reflection in the control room during safety-critical work
Baran & Scott	2010	Journal paper	A grounded theory of leadership and sensemaking
Bergström	2012	PhD Thesis	Aspects of organisational resilience in escalating situations
Grøtan & Størseth	2012	Conference paper	Integration of organisational resilience into safety management
Hayes	2012	Journal paper	Operator competence and capacity in complex hazardous activities
Lundberg et al.	2012	Journal paper	Resilience in sensemaking and control of emergency response
Sanne	2012	Journal paper	Learning from adverse events in the nuclear power industry
Hutter & Kuhlicke	2013	Journal paper	Understanding resilience in the context of planning and institutions
Rankin	2013	Licentiate's thesis	Adaptive performance and resilience in high-risk work
Rantatalo	2013	Thesis	Sensemaking and organising in the policing of high-risk situations
Rankin et al.	2014	Journal paper	A framework for analysing adaptations in high-risk work
Busby & Collins	2014	Journal paper	Sensemaking about risk control in offshore hydrocarbons production
Haavik	2014	Journal paper	The nature of sociotechnical work in safety-critical operations
Norros et al.	2014	Journal paper	Operators' orientations to procedure guidance in NPP process control
Saleh et al.	2014	Journal paper	Safety diagnosability and observability of hazards in design
van den Heuvel et al.	2014	Journal paper	Police strategies for resilient decision-making and action implementation
Barton et al.	2015	Journal paper	Contextualised engagement in wildland firefighting
Dahlberg	2015	Journal paper	Exploration of resilience and complexity
Grøtan & van der Vorm	2015	Symposium paper	Conceptual approach to operational and managerial training of resilience
Hunte et al.	2015	Symposium paper	Dialogic sensemaking as a resource for safety and resilience
van der Beek & Schraagen	2015	Journal paper	Adaptability and performance in teams to enhance resilience
Danielsson	2016	Journal paper	Cross-sectorial collaboration in a potentially dangerous situation
Jahn	2016	Journal paper	Adapting safety rules in high reliability contexts
Hoffman & Hancock	2017	Journal paper	How to measure resilience
Landman et al.	2017	Journal paper	A conceptual model for pilot's ability to deal with unexpected events
Lofquist et al.	2017	Journal paper	Why different sub-cultures interpret safety rule gaps in different ways
Siegel & Schraagen	2017	Journal paper	Making resilience-related knowledge explicit through team reflection
Takeda et al.	2017	Journal paper	Developing resilience in disaster management promoting sensemaking
Teo et al.	2017	Journal paper	How leaders utilise relationships to activate resilience during crisis
Favarò & Saleh	2018	Journal paper	Temporal logic for safety supervisory control and hazard monitoring

excluded. So were documents that primarily discuss resilience and that do not relate to the notion of sensemaking. We did not include books in this literature review, and a few papers were excluded as we requested, but did not receive, the full-text.

Going through the identified documents, we found some key references that we included in this review as background (i.e. not included in Table 2).

3 FINDINGS FROM THE LITERATURE REVIEW

The literature search resulted in 33 documents that were included in the review. See Table 2 for the complete, chronologically listed, literature overview. As can be seen from the table, the reviewed literature includes 25 articles published in peer-reviewed scientific journals, four papers presented at international conferences, workshops or symposiums, three theses and one working paper.

No inclusion criteria were applied regarding publication year. The results clearly indicate that the use of the term sensemaking in the context of safety-critical situations, or in relation to the term resilience, is relatively recent. Except from the papers by Weick (1993), Gephart (1997) and Beunza & Stark (2004), the rest of the included literature was published in the ten-year period from 2009 to 2018.

In addition to the 33 publications in Table 2, we have also included often cited key research, among others Weick (1995) and Endsley et al. (2003).

The following chapters describe how this literature use the term sensemaking; how it characterises sensemaking in the context of safety-critical situations; and how it describes the relationship between sensemaking and resilience. In addition, we describe the few issues we have found of sensemaking in relation to training, human-machine interface, the maritime domain and design/development.

3.1 *The use of the term 'sensemaking'*

As outlined in chapter 1.3.1, the concept of sensemaking does not have one single definition. Weick (1995) stated that "(…) people can make sense of everything. This makes life easy for people who study sensemaking in the sense that their phenomenon is everywhere" (p. 49). For this reason, we started the review by taking a closer look at how the various authors of the included literature use the term.

In 1993, Weick reanalysed the Mann Gulch fire disaster in Montana in which 13 firefighters died. Here he provided analyses of sensemaking as a generic phenomenon, explaining that "the basic idea of sensemaking is that reality is an ongoing accomplishment that emerges from efforts to create order and make retrospective sense of what occurs" (p. 635). He further uses the example of Mann Gulch to argue that sensemaking is about contextual rationality and that it is "built out of vague questions, muddy answers and negotiated agreements that attempt to reduce confusion" (Weick, 1993, p. 636).

As described by Maitlis & Christianson (2014), the term sensemaking is often used without any associated definition from the literature, and when definitions are provided there are a variety of meanings asserted to it. Correspondingly, in our study we found that several authors include sensemaking as a general notion and do not provide any associated definition (e.g. Favarò & Saleh, 2018; Saleh et al., 2014; Sanne, 2012). Some provide references to the work of others, but without reproducing the actual definition (e.g. Grøtan & van der Vorm, 2015; Jahn, 2016).

However, most of the papers in this review provide definitions or references based on Weick's work on sensemaking, describing sensemaking as a social process, involving the extracting of cues and enactment to create meaning to events. These include, among others, the work of Baran & Scott (2010), Danielsson (2016), Hayes (2012) and Tekeda, Jones & Helms (2017).

Some describe sensemaking as a more cognitive process and refer to Klein's macro-cognitive/data-frame model (Hoffman & Hancock, 2017; Siegel & Schraagen, 2017), whereas others refer to the work of both Weick & Klein (e.g. Landman et al., 2017; Norros et al., 2014; Rankin et al. 2014).

Some authors describe sensemaking as a process building and supporting situational awareness (Lundberg et al., 2012; van den Heuvel et al., 2014). Situational awareness being a tactical (short term issue) while sensemaking is a broader strategic concept (long range issue) creating and supporting understanding. In his paper on sensework, Haavik (2014) uses the definitions of Weick to explain how sensemaking is a theoretical, generic framework that addresses mental processes and aspects of work. Alternatively, a few papers focus on the Cynefin sensemaking framework described in the work by Kurtz & Snowden (2003) (Dahlberg, 2015; Grøtan & Størseth, 2012).

The concept of sensemaking has traditionally, and in accordance with the work of Karl E. Weick, been described as retrospective in the sense that we make sense of our actions and experience after they have occurred. Most of the literature in this review uses the notion of sensemaking accordingly, often referring to Weick when doing so (e.g. Baran & Scott, 2010; Rantatalo, 2013; Teo et al., 2017).

However, a few of the authors use the term in a more future-oriented sense. As an example, Barton et al. (2015) introduce the term of proactive leader sensemaking, arguing that leaders in particular play a critical role in creating and maintaining a context for actively managing uncertain contexts.

3.2 *Characteristics of sensemaking in the context of safety-critical situations*

There are several factors that can influence sensemaking; context, language, identity, cognitive frameworks, emotion, politics and technology (Sandberg & Tsoukas, 2015). Thus, in the context of a safety-critical situation there might be characteristics of sensemaking other than or more prominent than the characteristics of everyday sensemaking. For instance, it might be expected that strong negative emotions like stress and fear would be influential on sensemaking in such circumstances. However, the literature found in this review did not discuss these characteristics explicitly.

In his analysis of the Mann Gulch fire disaster Weick (1993) describes the disaster "was produced by the interrelated collapse of sensemaking and structure". The smokejumpers expected to find a fire that they would have control over within the next morning. This positive illusion prohibited them from making sense of the cues in their environment contradicting this expectation. Weick describes unclear roles, identity issues and in the end the intense emotion of panic that led to the disintegration of the group and to the primitive tendency to flight. Unfortunately, this response was too simple to match the complexity of the fire and 13 men lost their lives.

After completing a field study of 80 interviews, Busby & Collins (2014) categorised the many ways of acting through which informants made sense of the risk control task. The authors provide explanations to each of their 32 categories, but elaborate on the five more commonly used. These are 1) being circumscribed (constrained, realistic, moderate), 2) being engaged (closely involved, concerned), 3) being resolute (rapid, and consistent in acting) 4) being socialised (social outcomes and systems of social obligation), and 5) being solicitous (seeks opinion and external references). The authors use their qualitative findings to suggest that the sensemaking of organisational members is simultaneously optimistic and pessimistic about the capacities of social organisation to manage risk. This balance is, however, not of individual sensemaking and is not a deliberate choice.

Lundberg et al. (2012) explored a model for describing and studying resilience in management of safety-critical/irregular events. The model was based on changes in the ongoing process, the actors sensemaking and control functions and the technology used for sensemaking and control. The model helped to identify resilience building processes and sources of resilience emergency responses.

Several other authors describe the characteristics of sensemaking by describing it in terms of how it relates to the concept of resilience. This is the topic of the following chapter.

3.3 *The link between sensemaking and resilience*

As described in the introduction, one goal was to establish a knowledge base on the relationship between the two concepts of sensemaking and resilience.

In his reanalyses of the Mann Gulch fire disaster, Weick (1993) states that the disaster was produced by the interrelated collapse of sensemaking and structure. Weick mentioned the importance of nonstop talk as a critical source of coordination in complex systems. He proposes four potential sources of resilience that "make groups less vulnerable to disruptions of sensemaking" (p. 628). These include improvisation and bricolage, virtual role systems, the attitude of wisdom, and norms of respectful interaction. Thus, from this perspective, resilience is core to maintaining sensemaking in critical situations.

Other authors argue that sensemaking is an important source to achieving resilience. Through their review of disaster management literature, along with illustrative examples from global disasters, Takeda et al. (2017) highlight the importance of resilience in disaster management. They argue that heedful interrelating and sensemaking are two of the central tenets of resilience research and that a greater attention to resilience in the disaster management process could be achieved through a focus on the development of sensemaking and heedful interrelating. Takeda et al. (2017) conclude that future research is needed to further understand resilience and sensemaking.

Rankin (2013) draws lines between Hollnagel's four central abilities to characterise a resilient system and the sensemaking capabilities of seeking information, ascribing meaning and action as described by Grøtan et al. (2008). In a paper from the same year, Rankin and her co-workers present a framework for analysing adaptations in high-risk work (Rankin et al., 2014). Here they explain how *sensemaking variety* "includes the ability to process information and revise it as the world changes, given contextual constraints and the experience and knowledge of the individuals involved" (p. 84). Thus, sensemaking is important for adaptive behaviour, which in turn is a prerequisite for resilience. They focus on the importance of

observing sharp-end adaptations as critical to identify system brittleness and resilience.

Others suggest that sensemaking plays an important role in accomplishing tasks that facilitates organisational resilience, especially when the sensemaking is carried out by leaders (Teo et al., 2017). According to Hunte et al. (2015), "shared (social) sensemaking creates and nourishes common awareness and understanding of the 'operating point', and in so doing facilitates coordination and safer performance. This is an essential condition for the emergence of safety and resilience" (p. 1). Similarly, van der Beek & Schraagen (2015) list sensemaking, or situation assessment, as one of several team resilience abilities. However, they do not provide a thorough discussion on sensemaking as such.

The sensemaking perspective is also used in the literature as a means to analyse or explain resilience. According to Bergström (2012) "the development of a theoretical framework for analysing organisational resilience in escalating situations needs to relate to the explanatory potential of sensemaking theory" (p. 8). To enhance a dynamic understanding of resilience, Hutter & Kuhlicke (2013) analyse its elusive character from a sensemaking perspective. In their paper, resilience is understood as a "content of sensemaking processes in the context of a crisis" (p. 294). The authors state that the work of Weick in 'sensemaking in organisations' on the one hand and 'resilience' on the other is only loosely coupled, and they connect the two a bit more explicitly for planning research about resilience. To understand how groups, organisations and networks make sense of resilience in the context of a crisis, one should consider the four processes of committing to resilience, expecting resilience, arguing about resilience and manipulating with resilience. These are referred to as sensemaking processes in planning research and practice (Hutter & Kuhlicke, 2013).

In their paper, Lundberg et al. (2012) study resilience in the context of sensemaking and control in emergency management of irregular emergencies, and proposes an emergency management analysis model. The model unifies and complements existing models by explicitly modelling resilience factors and the actors 'sensemaking and control functions and technologies' variety. Other authors using sensemaking theory to describe aspects of resilience, is Siegel & Schraagen (2017). In their article on team reflection, they make an attempt to use reflection and the data-frame theory of sensemaking to show the relationship between knowledge and resilience.

Finally, Hoffman & Hancock (2017) aim to promote a discussion on how to measure resilience. They explain how sensemaking provide information to the work system about whether and when the system needs to change its understanding of problem situations, and further argue that this means that "adaptive and resilient sensemaking requires mechanisms for recognizing anomalies and situations that mandate change" (p. 571).

3.4 *Sensemaking as a basis for change, innovation, creativity and design*

The literature has indicated how sensemaking supports innovation, design and creativity. Sensemaking has often been limited to an organisational context, seldom discussing issues such as system design. Saleh et al. (2014), point out that safety science seems to have drifted from the engineering and design side of system safety towards organisational and social sciences or refinement of probabilistic models. To improve safety and resilience in safety-critical operations, we must have a broad based approach involving the sociotechnical system, and also how cues and prospective sensemaking can be enabled from the design phase on.

It is also mentioned that sensemaking is a key process for learning in organisations, teams and individuals (Maitlis & Christianson, 2014) – one challenge is to use new information rather than engage in sensemaking based on prior beliefs. Sensemaking is also concerned with new meanings that can underpin new ways of organizing, understanding and design. When sensemaking of organisational members are impacted, the participants are motivated to change their own roles and practices. This has especially been supported by looking at interpretations and actions through the process of action research, Greenwood & Levin (2006).

3.5 *Topics partly covered in the literature review*

Accident reports have shown that insufficient training and poor HMI may impair sensemaking processes and thus lead to incidents and accidents. After the incident at Scarabeo 8 the investigation report attributed the incident to insufficient training of control room personnel and weaknesses in the control room's human-machine interface (Ptil, 2012). As discussed by Endsley et al. (2003) human-machine-interface is a key factor shaping operator performance, via concepts like sensemaking and situation awareness. We thus expected some of this literature to addresses sensemaking in relation to training or HMI. However, we did not find much relevant literature through our review. In the following, we have summarised our findings related to training, HMI and sensemaking in safety-critical situations within the maritime domain.

Not many of the reviewed documents look at ways of training to improve sensemaking. However, Takeda et al. (2017) focus on building capacity for individual actors to interrelate in a heedful manner. In Saleh et al. (2014) it is mentioned that the ability to diagnose hazardous states provides one way to improve operators' sensemaking and situational awareness after an adverse event. It is synergetic with organisational factors in support of accident prevention, particular safety training, that can be shaped by including off-nominal conditions. Rantatalo (2013) describes how observations that were carried out were targeted joint police management training in the setting of full-scale simulated scenario, arguing that "from sensemaking and organisational reliability perspectives, high-strain situations like that described above offer a possibility to observe interaction patterns during incident management in a realistic setting" (p. 55). One of the conclusions drawn in the paper by Landman et al. (2017) about dealing with unexpected events on the flight deck, is that interventions should focus on "increasing pilot reframing skills (e.g. through the use of unpredictability in training scenarios)" (p. 1161). The authors propose a conceptual model for explaining pilot performance in surprising and startling situations; a model that can be used to design experiments and training simulations. However, several of the reviewed papers discuss training that is aimed at enhancing resilience (Bergström, 2012; Grøtan & van der Vorm, 2015; van der Beek & Schraagen, 2015).

Siegel & Schraagen (2017), describe processes and HMI tools to make boundaries explicit in railway operations. In the maritime sector the ability to handle demanding operations safely is increasingly dependent on ICT-based control systems, e.g. dynamic positioning and ballasting. Hence, the impact of HMI on sensemaking is an important topic for our project. However, few of the documents covered by this literature review address this issue. Relevant, but brief, discussions on such interaction are made in the papers by Dahlberg (2015), Landman et al. (2017) and Sanne (2012).

Furthermore, almost none of the identified publications on sensemaking in safety-critical situations are related to maritime operations. A couple of the papers discuss sensemaking in the context of offshore oil and gas production (Busby & Collins, 2014; Hayes, 2012). The doctoral thesis by Bergström (2012) is the only publication in our review that is related to navigation and shipping, although not specifically concerning sensemaking.

3.6 *Limitation of the review*

Our findings should be considered in light of the limitation in the search terms. In addition to the search terms related to the keyword *safety-critical*, we could have included *surprising*, *emergency*, etc. Also, we could have obtained interesting findings had the review included papers on *situational awareness*. However, we chose to keep this review focused on the terms of specific interest for our project.

4 DISCUSSION AND CONCLUSIONS

The current literature review aimed at describing how the selected literature use the term sensemaking; how it characterises sensemaking in the context of safety-critical situations; and how it describes the relationship between sensemaking and resilience.

We found that several authors use the term sensemaking without providing a definition, and those who do refer to Weick's work describing sensemaking as a social process, involving extracting cues and enactment to create meaning to events retrospectively.

Sensemaking and resilience were found to be described as related in the reviewed literature. Sensemaking creates the context for being resilient; at the same time sources of resilience, such as redundancy (i.e. redundant clues), help to make sense of the situation. Lundberg et al. (2012) have suggested a model for resilient sensemaking, exploring changes in the ongoing process, the actors sensemaking and control functions and the technology used for sensemaking and control. However, little is written on the issue of sensemaking in safety-critical situations that also concern aspects of training, human-machine interaction or the maritime domain; thus, we a see a need to increase our knowledge in these areas by observation studies and targeted literature reviews.

Sensemaking is seen as a long term strategic process, creating understanding. There has been a discussion whether sensemaking is something that happens inside the individuals' heads or if it is a social construct. In our further work, we would like to explore sensemaking as a social construct, impacted by organisations, technology and human factors. Also, sensemaking has been described as both a retrospective and a prospective process. We would like to build on the research of prospective sensemaking to understand how to build resilience through future actions.

Sensemaking is accomplished through perceiving cues, creating meaning/learning (trough interpretations and actions) as discussed in Maitlis & Christianson (2014), thus it is dependent on responsibilities, procedures, training and technology (such as human machine interactions). However, unexpected or safety-critical situations

do not necessarily trigger sensemaking; it happens when there is a discrepancy from what one expects. The expectations are influenced by the amount of experience, training and the degree of questioning attitude, i.e. in line with existing group norms or organisational culture.

Discrepancies must also be supported by design, i.e. having redundant systems that can reveal discrepancies, and by training to ensure a questioning attitude. This is in line with sensemaking in HRO – High Reliability Organisations, where practices such as "preoccupation with failure", "reluctance to simplify" and "sensitivity to operations" support the explorations of cues and interpretations (Weick & Sutcliffe, 2011).

Further work in the project will focus on how to strengthen the loop of perceiving cues, creating interpretations, assert meaning to events taking actions. Further, how to improve the design of interfaces between automation (i.e. HMI and procedures) and how to train to facilitate sensemaking and resilience. Through this work, we aim to contribute to improve the ability to handle safety-critical situations in demanding maritime operations.

REFERENCES

Baran, B.E. & Scott, C.W. (2010). Organizing ambiguity: A grounded theory of leadership and sensemaking within dangerous contexts. *Military Psychology, 22* (Suppl 1), 42–69.

Barton, M.A., Sutcliffe, K.M., Vogus, T. J. & DeWitt, T. (2015). Performing under uncertainty: Contextualized engagement in wildland firefighting. *Journal of Contingencies and Crisis Management, 23*(2), 74–83.

Bergström, J. (2012). *Escalation: explorative studies of high-risk situations from the theoretical perspectives of complexity and joint cognitive systems.* Lund University (Media-Tryck).

Busby, J.S. & Collins, A.M. (2014). Organizational Sensemaking About Risk Controls: The Case of Offshore Hydrocarbons Production. *Risk Analysis, 34*(9), 1738–1752.

Dahlberg, R. (2015). Resilience and complexity conjoining the discourses of two contested concepts. *Culture Unbound, 7*(3), 541–557.

Danielsson, E. (2016). Following Routines: A Challenge in Cross-Sectorial Collaboration. *Journal of Contingencies and Crisis Management, 24*(1), 36–45.

Endsley, M. R., Bolte, B., & Jones, D. G. (2003) Designing for situation awareness: an approach to user-centered design. *Boca Raton, FL: CRC Press.*

Elsbach, K.D., Barr, P.S. & Hargadon, A.B. (2005). Identifying Situated Cognition in Organizations. *Organization Science, 16*(4), 422–433.

Favarò, F.M. & Saleh, J.H. (2018). Application of temporal logic for safety supervisory control and model-based hazard monitoring. *Reliability Engineering & System Safety, 169*, 166–178.

Gephart, R. (1997). Hazardous measures: An interpretive textual analysis of quantitative sensemaking during crises. *Journal of Organizational Behavior*, 583–622.

Gephart, R., Topal, C. & Zhang, Z. (2010). Future-oriented sensemaking: Temporalities and institutional legitimation. In T. Hernes & S. Maitlis (Eds.), *Process, sensemaking, and organizing* (pp. 275–312). New York: Oxford U.P.

Gioia, D.A., Thomas, J.B., Clark, S.M., & Chittipeddi, K. (1994). Symbolism and strategic change in academia: The dynamics of sensemaking and influence. *Organization Science*, 5(3), 363–383.

Greenwood, D.J. & Levin, M. (2006). *Introduction to action research: Social research for social change*. SAGE publications.

Grøtan, T.O. & Størseth, F. (2012). Integrated safety management based on organizational resilience. In C. Berenguer, A. Grall, & C. G. Soares (Eds.), *Advances in Safety, Reliability and Risk Management: ESREL 2011*, pp. 1732–1740.

Grøtan, T.O., Størseth, F., Rø, M.H. & Skjerve, A.B. (2008). *Resilience, Adaptation and Improvisation– increasing resilience by organising for successful improvisation*. Paper presented at the 3rd Symposium on Resilience Engineering, Antibes, Juan-Les-Pins, France.

Grøtan, T.O. & Van der Vorm, J. (2015). *Training for Operational Resilience Capabilities*. Paper presented at the 6th Symposium on Resilience Engineering, Lisbon, Portugal.

Haavik, T.K. (2014). Sensework. *Computer Supported Cooperative Work (CSCW)*, 23(3), 269–298.

Hayes, J. (2012). Operator competence and capacity–Lessons from the Montara blowout. *Safety Science, 50*(3), 563–574.

Hoffman, R.R. & Hancock, P. (2017). Measuring resilience. *Human Factors, 59*(4), 564–581.

Hollnagel, E., Pariès, J., Woods, D.D. & Wreathall, J. (2011). *Resilience engineering in practice: A guidebook*: Ashgate Publishing, Ltd.

Hunte, G.S., Schubert, C.C. & Wears, R.L. (2015). *Dialogic sensemaking as a resource for safety and resilience*. Paper presented at the 6th Symposium on Resilience Engineering, Lisbon, Portugal.

Hutter, G., & Kuhlicke, C. (2013). Resilience, Talk and Action: Exploring the Meanings of Resilience in the Context of Planning and Institutions. *Planning Practice and Research, 28*(3), 294–306.

Jahn, J.L.S. (2016). Adapting Safety Rules in a High Reliability Context. *Management Communication Quarterly, 30*(3), 362–389.

Kurtz, C.F. & Snowden, D.J. (2003). The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM systems journal, 42*(3), 462–483.

Landman, A., Groen, E.L., van Paassen, M.M.R., Bronkhorst, A.W. & Mulder, M. (2017). Dealing with Unexpected Events on the Flight Deck: A Conceptual Model of Startle and Surprise. *Hum Factors, 59*(8), 1161–1172.

Lundberg, J., Törnqvist, E. & Tehrani, S.N. (2012). Resilience in sensemaking and control of emergency response. *International Journal of Emergency Management, 8*(2).

Maitlis, S. & Christianson, M. (2014). Sensemaking in Organizations: Taking Stock and Moving Forward. *The Academy of Management Annals, 8*(1), 57–125.

Norros, L., Liinasuo, M. & Savioja, P. (2014). Operators' orientations to procedure guidance in NPP process control. *Cognition, Technology and Work, 16*(4), 487–499.

Ptil (2012). Investigation Scarabeo 8 (Gransking Saipem – Ballasthendelse Scarabeo 8, 4.9.2012) from www.Ptil.no.

Rankin, A. (2013). *Resilience in High Risk Work: Analysing Adaptive Performance.* (Licentiate's Thesis), Linköping University Electronic Press,

Rankin, A., Lundberg, J., Woltjer, R., Rollenhagen, C. & Hollnagel, E. (2014). Resilience in Everyday Operations. *Journal of Cognitive Engineering and Decision Making, 8*(1), 78–97.

Rantatalo, O. (2013). *Sensemaking and organising in the policing of high risk situations: Focusing the Swedish Police National Counter-Terrorist Unit*. Umeå Universitet,

Rosness, R., Evjemo, T.E., Haavik, T. & Wærø, I. (2016). Prospective sensemaking in the operating theatre. *Cognition, Technology & Work, 18*(1), 53–69.

Saleh, J. H., Haga, R. A., Favarò, F. M., & Bakolas, E. (2014). Texas City refinery accident: Case study in breakdown of defense-in-depth and violation of the safety–diagnosability principle in design. *Engineering Failure Analysis*, *36*, 121–133.

Sandberg, J. & Tsoukas, H. (2015). Making sense of the sensemaking perspective: Its constituents, limitations, and opportunities for further development. *Journal of Organizational Behavior, 36*(S1), 6–32.

Sanne, J.M. (2012). Learning from adverse events in the nuclear power industry: Organizational learning, policy making and normalization. *Technology in Society, 34*(3), 239–250.

Siegel, A.W. & Schraagen, J.M. (2017). Team reflection makes resilience-related knowledge explicit through collaborative sensemaking: observation study at a rail post. *Cognition, Technology & Work, 19*(1), 127–142.

Takeda, M., Jones, R. & Helms, M.M. (2017). Promoting sense-making in volatile environments: Developing resilience in disaster management. *Journal of Human Behavior in the Social Environment, 27*(8), 791–805.

Teo, W.L., Lee, M. & Lim, W.-S. (2017). The relational activation of resilience model: How leadership activates resilience in an organizational crisis. *Journal of Contingencies and Crisis Management, 25*(3), 136–147.

van den Heuvel, C., Alison, L. & Power, N. (2014). Coping with uncertainty: Police strategies for resilient decision-making and action implementation. *Cognition, Technology & Work, 16*(1), 25–45.

van der Beek, D. & Schraagen, J.M. (2015). ADAPTER: Analysing and developing adaptability and performance in teams to enhance resilience. *Reliability Engineering & System Safety, 141*, 33–44.

Weick, K.E. (1993). The collapse of sensemaking in organizations: The Mann Gulch disaster. *Administrative science quarterly*, 628–652.

Weick, K.E. (1995). *Sensemaking in organizations*: Sage.

Weick, K.E., Sutcliffe, K.M. & Obstfeld, D. (2005). Organizing and the Process of Sensemaking. *Organization Science, 16*(4), 409–421.

Weick, K.E. & Sutcliffe, K.M. (2011). Managing the unexpected: Resilient performance in an age of uncertainty (Vol. 8). John Wiley & Sons.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Task complexity, and operators' capabilities as predictor of human error: Modeling framework and an example of application

M.C. Leva, A. Caimo & R. Duane
Dublin Institute of Technology, Ireland

M. Demichela & L. Comberti
Politecnico di Torino, Italy

ABSTRACT: This paper presents the initial framework adopted to assess human error in assembly tasks at a large manufacturing company in Ireland. The model to characterize and predict human error presented in this paper is linked conceptually to the model introduced by Rasch (1980), where the probability of a specified outcome is modelled as a logistic function of the difference between the person capacity and item difficulty. The model needs to be modified to take into account an outcome that is not dichotomous and feed into the interaction between two macro factors: (a) Task complexity: that summarises all factors contributing to physical and mental workload requirements for execution of a given operative task & (b) Human capability: that considered the skills, training and experience of the people facing the tasks, representing a synthesis of their physical and cognitive abilities to verify whether or not they are matching the task requirements. Task complexity can be evaluated as a mathematical construct considering the compound effects of Mental Workload Demands and Physical Workload Demands associated to an operator task. Similarly, operator capability can be estimated on the basis of the operators' set of cognitive capabilities and physical conditions. A linear regression model was used to fit a dataset collected in R. The estimation of task complexity and operator skills was used to estimate human performance in a Poisson regression model. The preliminary results suggest that both elements are significant in predicting error occurrence.

1 INTRODUCTION

1.1 *Scope of work and background*

This paper presents the initial framework adopted to assess human error in assembly tasks at a large manufacturing company in Ireland [1].

The aim of this study was to carry out an observational, empirical study on the existing human errors in the dispatching department, find a way to model the issue and if possible propose approaches to reduce and eliminate errors and variations in the end product. The company dispatches technology goods to national and international customers and the focus of the project was the assembly of goods for dispatch. Operators prepare the goods at workstations along conveyor lines, however at these conveyors inefficiencies and inaccuracies relating to human performance were identified. Two primary workstations were selected for inclusion in the dispatching unit based on their recorded error rates. Conditions vary and fluctuate at workstations, which may increase the probability of making mistakes, including the complexity and number of the activities, environmental conditions and the quality of the product. An understanding of both the human nature (characteristics, feelings, and behavioural traits) and the impact of the features of the workstation on human nature (typology of activities, working load, anxiety induced, environmental factors etc.) was required to holistically determine the performance shaping factors for the workstations under examination. The focus is on the role of operator's capability to complete tasks and the means to reduce human errors whilst retraining product quality. Changes were proposed for the assembly lines at the dispatching stations, including changes in the procedures and training to employ an understanding of human performance and improvements to safety, with an overall beneficial impact on both productivity and quality.

The researcher conducted a task analysis of the critical activities completed by operators when packing out the variety of product units at two primary workstations. Questionnaires were prepared examining the skills requirements, skills rating of operators, mental workload requirements, physical workload requirements, perceived task complexity and motivation. Finally, the implementation of an applied model Task Execution Reliability Model (TERM) was used to identify the main fac-

tors affecting human performance for this settings. Three methods were used to inform the research:

1. Firstly, an examination of performance shaping factors in the literature to inform a set of specific questionnaires.
2. Secondly, the collection and analysis of the data from the questionnaires completed by operators, technicians, supervisors, group leaders and process engineers in the manufacturing facility familiar with the work undertaken at the workstations under examination.
3. Thirdly, focus group sessions were run discussing possible participatory redesign for process and procedures at the workstations
4. Finally the data from the questionnaire was also used to predict task complexity and error occurrences using two different types of regression models.

2 MODELLING HUMAN ERROR

2.1 *Human error in manufacturing*

Human nature can be shaped and driven by factors including individual characteristics, personal issues, physical and psychological conditions (Tooby & Cosmides, 1990). These factors interact with each other and may determine the output and productivity of the performance of the individual. Human performance is unavoidably susceptible to human error, as humans are not infallible and the occurrence of errors must be expected (Karl & Karl, 2012). Humans are often capable of recognising errors and rectifying such errors before any serious or critical consequences occur (Sheridan, 2008). With this in mind, human performance can be accepted and understood as the definitive product of the balance between task complexity and capability (Morgeson et al, 2010).

When the capabilities and limitations of humans are understood, incorporated and acknowledged, Harris (2006) argues that benefits can include increased efficiency and improved safety performance. Individual employee's competencies may be challenged by fatigue, stressors and unpredictability, whilst competencies may benefit from skills, training and a clear comprehension of the task (Miller and & Parasuraman, 2007, Jo et al, 2012, Kostina et al, 2012). The capabilities of the operator and the physical skills required for the task must be taken into consideration when reviewing tasks and the errors associated with them (Harris, 2006). A balance between workload, both physical and mental, ought to be reached to reduce human errors among competent operators (Miller and & Parasuraman, 2007).

2.2 *The TERM model: Task execution reliability model*

The model used is linked conceptually to the model introduced by Rasch (1980) to analyse correct or incorrect execution of a task as a function of the trade-off between (a) the respondent's abilities, attitudes or personality traits and (b) the item difficulty. In the Rasch model, the probability of a specified outcome (e.g. right/wrong results) is modelled as a logistic function of the difference between the person and item difficulty parameter.

The mathematical form of the model is provided in equation (1).

$$\Pr\left(X_{ni}=1\right)=e^{\beta_n-\delta_i}\Big/1+e^{\beta_n-\delta_i} \quad (1)$$

Let X_{ni} be a dichotomous random variable with binary values where, for example, $X_{ni} = 1$ denotes a correct response and an $X_{ni} = 0$ an incorrect response to a given assessment item. In the Rasch model for dichotomous data, the probability of the outcome is given by:
where β_n the ability of person n and δ_i the difficulty of item i.

The model needs to be radically enhanced to take into account an assessment of performance that is not dichotomous and feed into the interaction between two macro factors:

- Task Complexity (TC): summarising all factors contributing to physical and mental workload requirements for execution of a given operative task.
- Human Capability (HC): summarising the skills, training and experience of the people facing the tasks, representing a synthesis of their physical and cognitive abilities to verify whether or not they match the task requirements.

Task complexity can be evaluated as a mathematical construct considering also the compound effects of two main factors: "Mental Workload Demands" (MW) and, where relevant, "Physical Workload Demands" (PW), both associated to an operator task. Recent sensorised EEG experimental studies have shown that the simultaneous executions of tasks, whether physical or cognitive, tends to increase cognitive demands for the human brain (Mijović, 2017).

Similarly then, operator capability should be estimated on the basis of the operators' set of cognitive capabilities and physical conditions. A regression model was used to fit a dataset collected in R. The model and the preliminary results are discussed in chapter 3 of the present paper.

3 THE CASE STUDY: SUMMARY OF THE DATA COLLECTED AND THE TERM MODEL AS APPLIED

3.1 *The case study and the data collection plan*

The setting and focus of this study is a large electronic manufacturing facility in the south of Ireland, which prepares and distributes technology goods to both national and global customers. In the dispatching unit of the facility, operators are provided with work stations and conveyors to prepare the products for dispatch and shipment through pack out procedures. The aim of this study was to carry out an observational, empirical study on the existing human errors in the dispatching department of the facility in a subsequent phase the study also lead to the identification of suitable approaches to reduce and/or eliminate such errors.

Two primary workstations were the focus of the assessment of the project, namely the conveyor line and another packaging workstation called the POD cell. To examine these workstations, an overview of the existing error rate at the conveyor line was required to be used as a benchmark against other workstations in the facility and to identify any possible improvements. As a means of comparison, the error rates for nine control workstations from within the manufacturing facility were acquired to facilitate data analysis and interpretation.

Error rates for both the control and non-control workstations were calculated in the same manner. Records were filtered from 1st December 2016 to 31st March 2017 for all workstations to retrieve the information for the calculations. This four months timeframe was deemed adequate due to the large number of products passing through the workstations. We considered only errors classified as stemming from a human related cause.

The human error rates were calculated using the following formula:

Number of Human Errors/the opportunity for error (2)

where the number of human errors were the errors recorded or captured due to a human cause

While the opportunities for error were the total output at the workstation i.e. number of processed units

For the pod and the conveyor, to attain the number of human errors, data relating to MWDs (missing, wrong or damaged) goods was collected within the four month period from 01/12/16 to 31/03/17. The MWDs originate from customer complaints or returned goods following disparities from the sales orders or damaged goods. MWDs can be slow information to capture, due to the possible time lapse between the shipment of an order, the start of use of the product by the customer and the identification of an error. MWDs may be reported some months after a product was shipped, however due to the nature of the timeframe selected, it was deemed appropriate that by the completion of the project, the number of MWDs recorded for that time frame would be sufficient. The opportunity for error was derived from the total output at the workstations within the four months period from the beginning of December 2016 to the end of March 2017.

For the control workstations, the numbers of human errors were retrieved from an online software platform within the four months period outlined above. The platform is used to record both the total output at the workstations and the number of errors recorded. The platform records errors with varying root causes through a classification system, many of which are not of a human nature. Twenty-seven classifications were deemed suitable for inclusion for the human errors recorded.

In the control workstations, when an error has occurred, the operator or technician is forced to input an error report at the time of the error occurring detailing the source of the error i.e. human, equipment, technical. The process cannot continue until an error report has been submitted. Due to this, the error reports recorded in the system can be regarded as representative of the total number of errors occurring during the timeframe. When an error is recorded, users are prompted to categorise the error under a variety of descriptions. The categories can include aspects of technology or equipment failure, and not all were relevant for inclusion in the error rate calculation.

3.2 *The observation and questionnaire protocol used for the wider case study*

Members of staff who work closely with the workstations involved in the project and the control workstations were invited to complete questionnaires to assess their opinions relating to:

- The importance of skills at different workstations
- Skills rating of individual operators
- Job satisfaction/motivation
- Mental workload requirements
- Physical workload requirements
- Perceived task complexity

Two questionnaires were prepared with one for supervisors, group leaders and process engineers, and a second questionnaire for operators and

technicians. Questionnaires were broken up in this fashion in order to capture observable variables from the supervisors/management and the individual subjective opinions of the operators. There was a difference in the type and volume of questions in the questionnaires, as the supervisor/group leader questionnaires asked two different things:

- Asked supervisors role participants to rate the skills of operators under their supervision
- Asked all participants to rate the skills requirement to complete work at the workstations

The questionnaires were completed by the employees of all eleven workstations and their supervisors leading to a total of 149 employees completing the questionnaire (100% response rate).

Participants were asked to rate their answers on a 10-point Likert Scale, with one meaning low and ten meaning high. Questionnaires were used to measure the mental and physical workload, worker skills, job satisfaction (motivation) and the perceived task complexity for operators, supervisors, group leaders and process engineers. As different duties and tasks require certain skills (e.g. manual skills, memory), practical training and underpinning knowledge, the questionnaire was designed to capture information relating to the following areas:

Mental Workload Requirements

- Need to cope with pace
- Variance of product
- Recognition requirements
- Load due to quality of coordination
- Requirement for training/experience
- Requirements for human machine interface (HMI)

Physical Workload Requirements

- Ergonomic score (REBA Assessment)
- Dexterity requirements/manual skills
- Adherence to procedure
- Reliance on automation

Job Satisfaction/Motivation

- Motivation e.g. satisfaction, meaningfulness

Worker Skills

- Memory
- Decision-making
- Recognition
- Coordination/communication—teamwork
- Coping with pace
- Experience
- Dexterity/manual skills
- Physical resilience
- Adherence to procedure

Perceived Task Complexity

- How mentally demanding are the tasks
- How physically demanding are the tasks
- How complex is this task

The error rate for all eleven workstations has been calculated and is outlined in Table 1

Data collection involved a rich integration of data from many sources, acquired observationally or through documented information. There were four primary sources of data:

1. The questionnaires outlined above. The data collected would facilitate the assessment of the relationship between the task complexity (mental workload requirements, physical workload requirements) and the worker capability (cognitive skills, physical skills).
2. Focus groups were conducted to understand the process and procedures at the workstations and aspects of the workstations that would benefit from redesign.
3. Key Performance Indicators (KPI's) were gathered for information relating to:

Table 1. Error rate dataset collected for each workstation.

Workstation No	No of human errors	Opportunity for errors i.e. total output	No of operators	Error per 1000pc
1 Pod	0	747	19	0.01
2 Conveyor	14	8,913	19	1.5
3 Control A	3	12,055	19	0.2
4 Control B	1	1,359	19	0.7
5 Control C	44	221	2	203.6
6 Control D	93	221	2	425
7 Control E	28	3,971	5	7
8 Control F	81	3,971	7	20.3
9 Control G	368	5,402	5	68.1
10 Control H	107	5,402	0.019	19.8
11 Control I	133	5,402	0.0246	24.6

- The actual time at the workstation (productivity KPI)
- The number of quality issues due to human error (quality KPI)

4. Error rates for the workstations were formulated to provide insight into the rate of human error and its resulting quality effects on the workstation end products.

Videos/Pictures

In order to capture and assess information regarding the routine activities and work patterns of staff in the facility, video recordings and photographs were taken as an observational method of data collection. The videos were used to:

– Measure the amount of time the entire task took to complete
– Measure the amount of time an aspect of the task took to complete e.g. closing with sellotape
– Compare the procedure completed to the actual projected procedure for the completion of actions
– Task analysis using Video TimerPro software to break down the tasks required of the operator to complete
– The photographs and video recordings were used to:
– Provide a basis for the Ergonomic Risk Assessment method used
– Compare comparable tasks completed at alternate work stations

For the first part of the regression model, an assessment of task complexity was conducted. The data gathered was evaluated on the basis of Task Complexity with a linear regression model. In order to complete this evaluation, a task complexity index was applied, namely:

Task Complexity index = a (Memory req.) + b (recognition req) + c (coordination req.) + d (cope with pace req) + e (Experience req) + f (Resilience req.) + g (adherence to procedure req.)

The Correlation matrix obtained for the element used for the regression to evaluate task complexity obtained in the statistical software R are shown in Figure 1.

Figure 2 reports the preliminary results of the linear regression model used to predict task complexity in R.

The model indicates that the parameters used to estimate task complexity in the linear regression are quite significant. They predict task complexity with a Standard error of 0.2991 on 36 degrees of freedom. The adjusted R squared obtained is 0.93996 and the F statistics on 36 Degrees of freedom is 96.52, with a p value of 2.2 e-16. Therefore

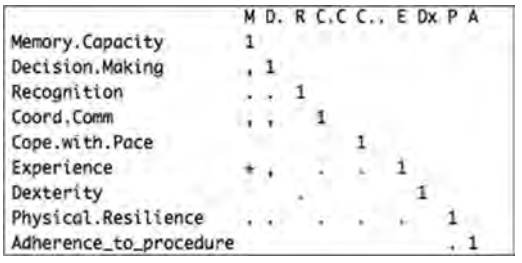


Figure 1. Correlation matrix evaluated for the element used for the regression to evaluate task complexity.

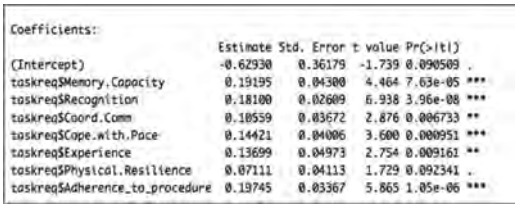


Figure 2. Results obtained from R to evaluate the relevance for the coefficient used to estimate task complexity.

the linear regression model to estimate task complexity seems to deliver significant results.

For the second part of the model, an estimation of the error occurrence of each workstation considering task complexity and operator capability was conducted. The use of the Rasch model with the dataset gathered was not possible as for the Rasch model the output needed to be a binary success or failure for each individual task. This was a type of data which was not able to be collected. Due to this, a generalised linear regression with a Poisson model, which was still based on the assumption that Human Performance can be represented as directly dependent from two macrofactors of task complexity and human capability, was used (see formula 3).

$$\lambda_i = e^{\beta_0+\beta_1 x_1-\beta_2 x_2+\varepsilon_i} = e^{\eta_i} \rightarrow log\lambda_i = \eta_i \quad (3)$$

where λ_i is the amount of error recorded, x_1 is task complexity and x_2 is operator skill level/capacity. The results obtained in R suggest that both elements are significant in predicting error occurrence, as shown in Figure 3.

The likelihood ratio test results confirmed the meaningfulness of the significance for the parameter chosen for estimating the error rate with this model, as shown in Figure 4.

However the limited data set and that the estimates of skill rating were gathered done using a subjective rating. Therefore the model could be

Table 2. Summary of data collected and revised for each workstation used in the regression model.

Id workstation	Average skills recorded	Task complexity	Errors_on_10000 parts
1 Pod	6.45	7.4	1
2 Conveyor	6.45	7.28	15
3 Control A	6.45	6.8	2
4 Control B	6.45	6.8	7
5 Control C	7.18	8	2036
6 Control D	7.23	9	4250
7 Control E	7.09	7.57	70
8 Control F	7.86	7.33	203
9 Control G	5.33	6.33	681
10 Control H	6.27	6.4	190
11 Control I	7.83	6.88	246

```
Call:
glm(formula = errorate$errors_on_10000parts ~ errorate$average.skills +
    (errorate$task.complexity), family = poisson, data = errorate)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-31.533  -21.480   -1.975    8.522   31.609

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               -2.41163    0.19002  -12.69   <2e-16 ***
errorate$average.skills   -0.47110    0.03457  -13.63   <2e-16 ***
errorate$task.complexity   1.57751    0.01710   92.24   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3. Results of the analysis run in R for the generalised Poisson linear model.

```
Model 1: errorate$error.rate ~ errorate$average.skills + errorate$task.complexity
Model 2: errorate$error.rate ~ 1
  #Df LogLik Df Chisq Pr(>Chisq)
1   4 33.667
2   2 28.272 -2 10.79   0.004539 **
```

Figure 4. Results of the analysis run in R for the likelihood ratio test for the generalised Poisson linear model.

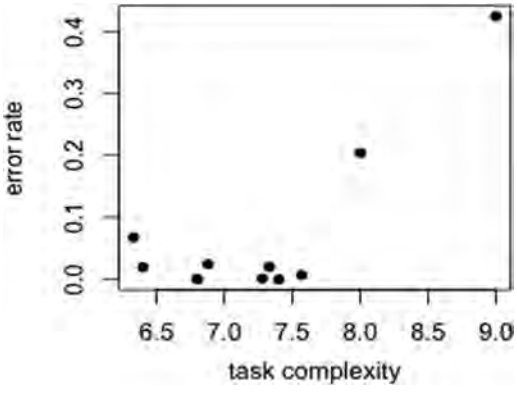


Figure 5. Plotting of the expected error rate calculated in respect to task complexity.

improved if a more extensive data collection campaign and a more objective estimation for skill rating is to be achieved.

Figure 5 provides a graphical representation of the plotting of the expected error rate calculated in respect to task complexity.

4 CONCLUSIONS AND WAY FORWARD

Following this study a focus group and some observations study were performed suggesting that a reorganisation of work practices between the original conveyor line and the new pod cell design served to improve overall human performance in the facility. This has been demonstrated through the reduction in the number of human errors reported for the workstations during the four month timeframe of the project.

The data formed the basis of an empirically based, cross-verified model of human performance that can be used to provide objective feedback to users increasing their awareness of risks related to their own human characteristics and impact the design of safety critical systems and current approaches for vocational training. For the manufacturing facility involved in the project, further developments may include engaging operators in all elements of a process, induction testing to match operator's capabilities to task most suited to them and orientation of workstations to facilitate operators considering human error and ergonomics principles.

Human error in the manufacturing facility prior to an intervention or examination of human performance contributed to the occurrence of a large number of errors resulting in financial costs and productivity losses for the organisation. The reorientation of work practices at work stations, considering the role of human error and ergonomic principles, has allowed for a reduction in

the incidence of human related errors across the workstations examined.

The results may be limited by the four month time frame for which human errors were considered. However results shown that task complexity can be significantly predicted starting from the variables observed in the case study.

The TERM model used (the Poisson generalised linear regression) also suggests that both task complexity and operator's skill are valid predictors of error occurrence in a workstation. It is maybe also possible that while task complexity increases a corresponding linear increase in worker skills and capability is not able to sufficiently compensate for the increased complexity.

REFERENCES

Harris, D. (2006). The influence of human factors on operational efficiency. *Aircraft Engineering and Aerospace Technology*, 78(1), pp. 20–25.

Jo, S., Myung, R. and Yoon, D. (2012). Quantitative prediction of mental workload with the ACT-R cognitive architecture. *International Journal of Industrial Ergonomics*, 42(4), pp. 359–370.

Karl, R. and Karl, M. (2012). Adverse Events: Root Causes and Latent Factors. *Surgical Clinics of North America*, 92(1), pp. 89–100.

Kostina, M., Karaulova, T., Sahno, J. and Maleki, M. (2012). Reliability estimation for manufacturing processes. *Journal of Achievements in Materials and Manufacturing Engineering.*

Leva, M.C., Ciarapica Alunni, C., Demichela, M. & Allemandi, G. (2016) Addressing human performance in automotive industry: identifying main drivers of human reliability. Irish Ergonomic Society Conference 2016.

Mijovic B. 2017. How hard is walking. blog https://blog.mbraintrain.com/@boggisha (last accessed 20/09/2017).

Miller, C. and Parasuraman, R. (2007). Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. Human Factors, 49(1), pp. 57–75.

Morgeson, F., DeRue, D. and Karam, E. (2010). Leadership in Teams: A Functional Approach to Understanding Leadership Structures and Processes. *Journal of Management*, 36(1), pp. 5–39.

Rasch G. (1980) *Probabilistic Model for some Intelligence and Attainment Tests*. University of Chicago Press. Chicago.

Sheridan, T. (2008). Risk, Human Error, and System Resilience: Fundamental Ideas. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3).

Tooby, J. and Cosmides, L. (1990). On the Universality of Human Nature and the Uniqueness of the Individual: The Role of Genetics and Adaptation. *Journal of Personality*, 58(1), pp. 17–67.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Bayesian aggregation of expert judgment data for quantification of human failure probabilities for radiotherapy

L. Podofillini, D. Pandya, F. Emert, A.J. Lomax & V.N. Dang
Paul Scherrer Institute, Villigen, Switzerland

G. Sansavini
Polytechnic of Zürich, ETH, Switzerland

ABSTRACT: The paper deals with the quantification of probabilities for human failures in the radiotherapy domain. The probabilities are used as input for the development of a Human Reliability Analysis (HRA) method specific for radiotherapy. Quantification is based on expert judgment, in view of the lack of relevant data. A Bayesian aggregation model is used to aggregate the judgments collected during elicitation sessions with domain experts. A qualitative scale is first used; then the judgments are interpreted as information on the order of magnitude of the error likelihood and aggregated under the Bayesian scheme. Besides for the specific domain of interest, this work is relevant for novel HRA applications outside typical domains, for which the need to incorporate expert judgment in traceable and defendable ways is key.

1 INTRODUCTION

Human failures are important contributors to near misses, incidents, and accidents in radiotherapy (WHO 2008), as in many other domains. Efforts are undertaken to systematically address the potential for failures and continuously improve the patient treatment process, e.g. Huq et al. (2016). In this context, the Risk and Human Reliability research group at the Paul Scherrer Institute (Switzerland), in collaboration with the institute's Center for Proton Therapy, is developing a method to support Human Reliability Analysis (HRA), specific for external beam radiotherapy. Previous work by the authors identified the personnel tasks critical to patient safety and possibly influencing factors (Pandya et al. 2017). Current work is addressing the quantification of the corresponding human failure probabilities.

In particular, the present paper focusses on the quantification of the failure probabilities for representative tasks, given a set of Performance Influencing Factors (PIFs). Given the shortage of directly usable experience data, the quantification resorts to expert judgment. The paper presents the application of a Bayesian aggregation model (Podofillini and Dang, 2013) to the judgments collected during elicitation sessions with domain experts. To avoid direct elicitation of probability values, the experts are asked to provide their judgments on a qualitative scale. The judgments are then interpreted as information on the order of magnitude of the error likelihood and aggregated under the Bayesian scheme. The paper presents the results of the aggregation. The application shows the ability of the aggregation approach to formally represent the variability of the experts' estimates.

Besides for the radiotherapy domain, the work presented in this paper is relevant for the various efforts recently done to extend HRA methods for application beyond their most typical applications, i.e. nuclear power plant operation. Lack of relevant data is a major issue for these novel applications (Bye et al. 2017, NASA 2012, Gibson 2012, Mkrtchyan et al. 2015, NUREG 2016) and methods to elicit expert judgment in a formal and defendable way are needed along with specific data collection initiatives.

The paper is organized as follows. The next Section provides the background on the HRA method under development, for which probability values are sought for in this paper. Section 3 presents the design of the elicitation sessions and the concepts underlying the Bayesian approach for processing and aggregation of the judgments. Section 4 presents the application to two Decision Trees part of the framework of the HRA method under development. Concluding remarks close the paper.

2 BACKGROUND INFORMATION

The framework for the HRA method consists of eighteen decision trees, one for each failure mode

corresponding to a different Generic Task Type (GTT, Table 1). The concept behind the GTTs is taken from the Human Error Assessment and Reduction Technique (HEART, Williams 2017), and is intended to define a set of task types, each with similar characteristics as it relates to the factors influencing performance and to the corresponding failure probabilities. The definition of the GTTs and of the influencing factors is based on GTTs- Performance influencing Factors (PIFs) structures developed in (Pandya et al. 2017): these structures link each GTT to the set of PIFs that influence the failure probability. These structures have been developed via a systematic and traceable process which, for each GTT, progressively identifies the involved cognitive functions, their failure modes and causes, failure mechanisms and PIFs.

The DT framework is well suited to represent the cause-based influences on failures identified by the GTT-PIF structures: the DTs identify the causes possibly leading to the GTT failure; in a similar way as done in other HRA methods (e.g. NUREG 2016, Moieni et al. 1994), each decision tree addresses a failure mode, with branching points representing the effects of PIFs. Two examples of DTs are presented in Figure 1. The decision trees develop from eight branching points, e.g. "Problematic interface", "Information content unclear", "Low vigilance due to expectations". As shown in Figure 1, each DT includes a subset of the eight branching points, three or four in most cases. The same branching point heading may appear across different DTs, e.g. "Problematic interface" in Figure 1; however, the influence of the branch on the failure probability may not necessarily be the same. This aspect will be returned to in the result Section 4. Each branch point is specified in terms of negative conditions: if any of the negative conditions is verified, then the lower branch applies. Example negative conditions are given in Table 2. The presentation of the development of the DTs from the GTT-PIF framework and of the negative conditions for each branching point is outside the scope of the present paper and will be presented in a separate publication (Pandya et al., working paper).

Table 1. Set of Generic Task Types (GTTs) and corresponding failure modes identified in Pandya et al. (2017). DTs are developed for each failure mode of the GTTs.

#	GTT	Failure mode
1	Identification of patient or patient related items	Patient information incorrectly matched
		Identification check not performed (decision based)
		Failure to execute desired action
2	Quality Check	Deviation from requirement not recognized
		Inappropriate understanding of underlying principles
		Check not performed (decision based)
		Execute desired action incorrectly
		Failure to execute desired action
		Coordination failure
3	Complex interaction with software or tool	Misinterpretation of data
		Execute desired action incorrectly
		Mismatch or inconsistency not recognized
4	Simple interaction with software or tool	Execute desired action incorrectly
		Failure to execute desired action
5	Iterative determination of optimum parameters	Misinterpretation of information
		Inappropriate decision on strategy selection
6	Verbal communication	Communication failure
		Not communicated (decision based)

To assess the failure probability of a specific radiotherapy task, an analyst would have to select the applicable DTs based on the relevant type of task and failure mode. Then, for each branching point, the analyst would have to select the appropriate branch based on the negative conditions proposed for each branch, in a similar way as done with other HRA methods involving DTs, e.g. NUREG (2016), Moieni et al. (1994).

The present paper focuses on the quantification of the DTs, i.e. on the assessment of the failure probabilities in correspondence of each path defined by the combination of the branching points.

3 EXPERT JUDGMENT ELICITATION AND AGGREGATION APPROACH

3.1 *Expert judgment elicitation*

As mentioned in the Introduction, due to the lack of relevant data, quantification is made via expert judgment. In particular, the expert elicitation sessions were designed with two aims. First, to support the identification of the negative conditions underlying each branch point. Second, to assess the impact of each branching point on the failure probability. Only the effects of single branch points were addressed (i.e. determining failure probabilities 1, 2, and 4 in Figure 1, top part). The combination effects will be addressed in future work.

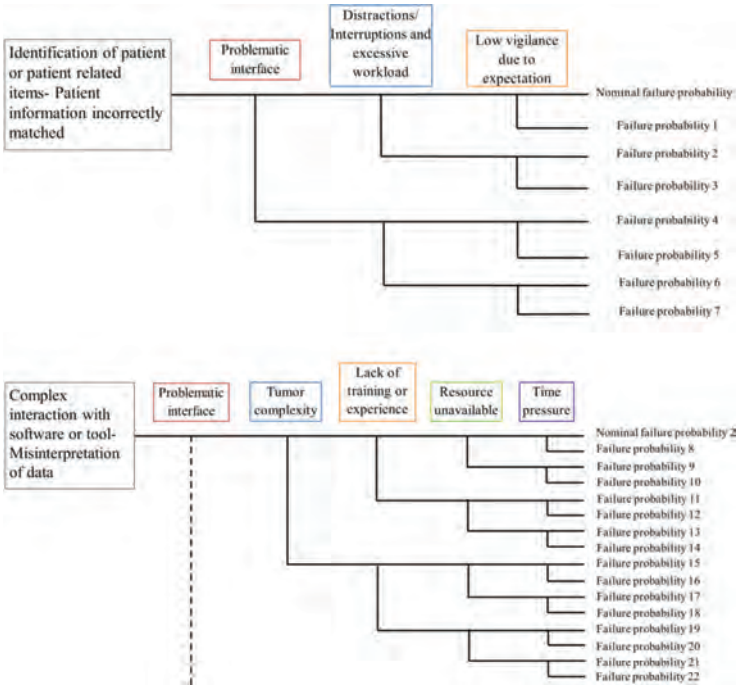


Figure 1. Two examples of decision tree; above: GTT "Identification of patient and patient related items", Failure mode "Patient information incorrectly matched"; below: GTT "GTT: Complex interaction with software or tool", Failure mode "Misinterpretation of data". The focus of the present paper is on quantification of the failure probabilities at each tree branch (only single branch effects elicited).

Table 2. Examples of negative conditions for two branching points in two different DTs.

DT	Branch point	Negative conditions
GTT: Identification of patient or patient related items; Failure mode: patient information incorrectly matched	Problematic interface	The written values look alike *(e.g. 111, 117)* The value on the label or file not easily readable There is no ID number on the patient item
GTT: Complex interaction with software or tool; Failure mode: Misinterpretation of data	Lack of adequate training or experience	Lack of familiarity with the tumor case Lack of training or experience on treating special tumor locations (e.g. close to multiple artefacts) Lack of experience or training to distinguish healthy and tumor tissues

Six failure scenarios were developed for the elicitation, Table 3 gives two examples. The idea is to elicit the impact of the branching point on these failure scenarios, which would then be representative of the overall GTT. The selection aimed at addressing the largest set of GTT failure modes, as well as prioritizing failure scenarios with the most critical consequences on patient safety. As shown for the examples in Table 3, each failure scenario is associated to a different GTT. Indeed as again shown in Table 3 and by the negative conditions in Table 2, the elicitation of the branching point impact was conducted on specific tasks and situations. This has been made to help experts to contextualize their judgments to the real tasks they perform and link their assessments to the daily experience. Alternatively, judgments may have been elicited directly for the GTTs and branching point categories. The former approach was chosen to avoid that experts would need to deal with abstract categories such as GTTs and the branch point labels. The focus of this paper is on the part of the elicitation session aimed at eliciting the impact of each branching point on the failure probability. The details of the overall elicitation design and its results will be presented in a different paper (Pandya et al., working paper).

Indeed, the elicitation addressed directly only part of the DTs required for quantification, i.e. six out of the eighteen from Table 1. However, some branching points may be thought of having very similar impact across different DTs; therefore the results from the elicitation for one DT may be used for others. It was assessed that the selected tasks may allow to quantify about two thirds of the whole set of branching points (recall only single branch point effects are considered here). Indeed future work may address the quantification of the remaining DTs and develop an approach to address multiple branch points as well.

Twelve experts were interviewed: medical physicists, medical doctors, dosimetrists and radiation technologists. Each expert dealt with tasks part of his/her daily job. Three tasks were elicited at most per expert. Each expert took part in the exercise alone. For each of the assigned tasks and each of the negative conditions corresponding to the branching points, the experts were asked to assess the impact of the negative condition on the failure probability when performing the task. The impact is elicited on a qualitative scale (Table 4), to avoid the known shortcomings of directly eliciting probability values, see e.g. Meyer and Booker (2001), Tversky and Kahneman (1974).

3.2 *Aggregation of expert assessments*

The approach to process the expert assessments has been as follows. The qualitative scale in Table 4 is first anchored to quantitative values, as shown in Table 5. The basis for the anchoring is the scale

Table 3. Example of failure situations used to elicit the impact of the branch points on the failure probability.

Failure situations	Failure mode	Generic task type
Failure to identify correct ID from control document on the bite-block, couch or file etc. such that incorrect item is picked up	Patient information incorrectly matched	Identification of patient or patient related items
Draw suboptimal (incorrect or incomplete) contours around volumes of interest for every slice due to misunderstanding of the data	Misinterpretation of data	Complex interaction with software or tool

Table 4. Qualitative scale used to elicit impact of negative conditions on the personnel tasks.

Impact	Descriptor	Meaning
Low impact	Failure is not expected to happen, although I see how it could happen.	Given the negative condition, the desired task is still so easy that it is inconceivable that any personnel would fail if they were to experience this condition.
Moderate impact	Failures happen occasionally/ sometimes with such conditions	Given the negative condition, the desired task becomes moderately difficult that it is possible so that personnel would occasionally/ sometimes fail if they were to experience this condition.
High impact	Failures happen often with such conditions	Given the negative condition, the desired task becomes highly difficult that is expected so that personnel would often fail if they were to experience this condition.
Extreme impact	Failure is almost unavoidable	Failure is almost unavoidable. Almost all personnel would not be able to perform the desired task.

Table 5. Anchoring of the qualitative impact scale to probability values (adapted from NUREG 2007).

Impact	Order of magnitude of failure probability
Low	1e-3
Moderate	1e-2
High	1e-1
Extreme	1

presented as part of the ATHEANA human reliability analysis method (NUREG 2007). Note the values on the scale are to be interpreted as reference orders of magnitude for the failure probability value. The value representative for low impact (1e-3) is confirmed in the studies by Wahi et al. (2008) and Salinas et al. (2013), from which it can be inferred that nominal error rates in patient identification and data entry in healthcare lie around 1e-3 and 3e-3. These are interpreted as lower bounds for error probabilities for the sector. Low impact of the branching point is not expected to change the order of magnitude of the probability so that the reference lower bound value still remains in the same order of magnitude.

The scale allows converting each assessment by the experts into a statement on the order of magnitude where the probability value would lie. It is interpreted as evidence of the relevant order of magnitude, and used to update the belief on that quantity in a Bayesian framework.

The process for aggregation of the judgments comprises two steps. First, for each negative condition, the judgments by the experts are aggregated: a distribution of the applicable probability for each condition is obtained. Then, these distributions are themselves aggregated into the final distribution of the corresponding branch. The aggregation approach is based on the Bayesian model presented in Podofillini & Dang (2013). The model represents the human error probability as an inherently variable quantity, resulting from the inherent variability of people performance as well as of the specific manifestations of the type of tasks and of the influencing factors. More specifically, the combination of GTTs and branch point conditions envelop specific tasks and specific performance conditions that are assumed to be characterized by inherently different failure probability values. The elicitation carried out in this work addressed specific manifestations of the combination (see Table 3 and Table 2): the Bayesian model is intended to consider the expert assessments on these manifestations (a specific task affected by specific negative performance conditions) and determine the original variability distribution of interest. Mathematically, the failure probability is assumed to be lognormally distributed, with unknown median to be determined based on the expert input. The error factor (square root of 95th and 5th percentile) is assumed to be known, equal to 3. The latter assumption of known error factor is not a requirement of the approach, but largely simplifies the calculations and decreases the amount of data required to be elicited. It is indeed a typically used and accepted value in HRA. The prior distribution of the median is assumed uniform for the four orders of magnitude in Table 5 (all impact levels are equally likely).

For the first part of the aggregation process, the expert assessments are used to update the degree of belief on the correct order of magnitude for the median of the probability distribution. The model requires as well assumptions on the confidence that the experts would be able to provide the correct value of the probability. The confidence is expressed in terms of a conditional probability that, given the real order of magnitude of the probability is one of the four in Table 5, the experts would assess the correct one or be off by one or more orders of magnitude. This conditional probability can be defined to model biases and dependence across experts, indeed provided that adequate information on the distribution is available (these are not modeled in the present work). In this work, it is assumed that experts have about 80% probability to provide the correct order of magnitude, 10% of being one order of magnitude off, 5% of being two or more orders of magnitude off. The exact values of these probabilities depend on the position of the interval with respect to the lower and upper bounds to have them normalized to a probability distribution. These values have been assumed by the authors of the paper; they appear to represent reasonable assumptions on the ability of the experts to provide correct estimates in this context. It is anyway important to mention that as more than a few experts are available (say five or more), the specific assumptions on the confidence to each expert do not play a significant role anymore in the final probability distribution. The output of this step is a distribution of the degree of belief on which of the levels in Table 5 represents the real value of the median of the probability distribution, for each negative condition possibly affecting each branch point.

The second part of the aggregation entails combining the degrees of belief obtained for each negative condition underlying each branch points. As the negative conditions are assumed equally likely, the final distribution is simply obtained as the average distribution across the negative conditions. In particular, for each of the levels in Table 5, the final degree of belief is the average degree of belief across the corresponding negative conditions. Applications of the approach will be presented in the next section.

4 AGGREGATION OF EXPERT ASSESSMENT: RESULTS AND DISCUSSION

This paper presents the result obtained for two DTs:

- GTT "Identification of patient and patient related items", failure mode "Patient information incorrectly matched";
- GTT "GTT: Complex interaction with software or tool", failure mode "Misinterpretation of data".

These are the two DTs shown in Figure 1. The results obtained from the whole elicitation are planned to be presented in Pandya et al. (Working paper).

The first part of this Section presents an overview of the aggregated results for the two DTs. The aim is to discuss how the quantitative results relate to the justification provided by the experts on their assessments. In other words, the goal is to check if the different values of error probability reflect in corresponding differences in the assessments by the experts. The second part of the Section provides details on how the expert assessments are aggregated.

Figure 2 shows the aggregated results from the expert assessments. The largest impact on failure probability corresponds to the branch point "Lack of training or experience" branching point affecting GTT "Complex interaction with software/tool". This is a complex task, related to defining an optimal therapy plan and requiring knowledge and expertise. As shown by the expert assessments, the influence of inadequacies in this respect can have high impact on the failure probability. Indeed, the resulting median probability is around 0.01, corresponding to the "high" impact level on the adopted scale. Branching point "Time pressure" was also assessed among the most influencing ones: it was felt that the need to complete the task with urgency would highly impact the quality of the therapy plan. On the other hand, two branching points were assessed to have generally low impact. In particular, interface issues (branching point "Problematic interface") were not felt to affect much the performance when identifying patients: identification of patients is made with diverse means; besides checking the patient ID, identification is checked verbally (calling patient name) and by the patient picture. Additionally the interface to deal with is extremely simple so that there is little possibility for confusion. Also, the complexity, in shape, size and location of the tumor was not felt to increase much the probability of errors in the development of the therapy plan. Typically, complex tumor cases are discussed in larger groups and the treatment details are thoroughly defined.

It is interesting to see in Figure 2 how the same type of branching point may affect GTTs differently. For example, the assessed probability for "Problematic interface" affecting GTT "Complex interaction with software/tool" is more than one order of magnitude larger than when affecting GTT "Identification of patient or patient-related tools". Again, this is the result of the expert opinions on the importance of the respective influences. The reasoning underlying the low impact according to the experts of the branch point on the latter GTT has already been discussed. On the other hand, the former GTT involves complex interactions with multiple software interfaces, dialog boxes, figures, etc: the impact of interface issues for this task was considered to have important effects on the failure probability.

The length of the error bars reflects differences in the expert assessments both for each negative condition and across the different conditions. The larger the bar, the larger the differences. This aspect will be returned to later in this Section.

Figures 3 and 4 show how the expert assessments are progressively processed to obtain the final probability distributions presented in Figure 2. The left side of the figures gives the assessments of the experts provided on the scale for each of the negative conditions. The middle shows the distribution results aggregated across the experts for each negative condition. The right side gives the final distribution, aggregating across the conditions. Specifically, Figure 3 addresses the effect of branching point "Problematic interface" on GTT

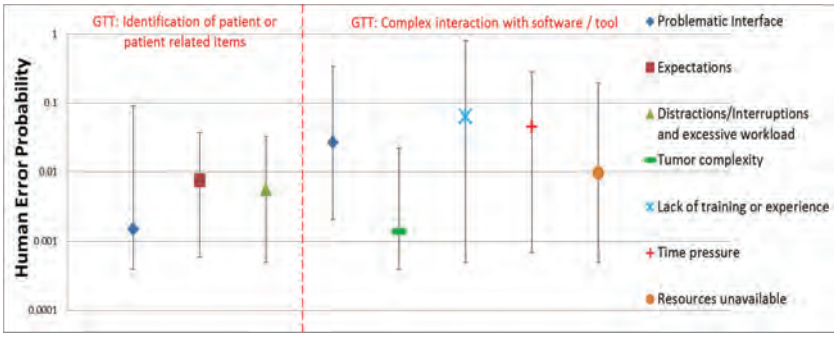


Figure 2. Aggregated results for the two considered GTTs; Symbols identify the median failure probability affected by each branch point (acting one at a time, presented on the right of the figure), error bars the 5th and the 95th percentiles.

"Identification of patient or patient-related items". It is interesting to see in the Figure how different assessments from the experts result in different distributions. For the first negative conditions, three experts provided the assessments of "low" (thus corresponding to an error probability of about 0.001) and two of "moderate". Correspondingly, the aggregated distribution in the middle of Figure 3 presents larger degree of belief for the latter level on the scale compared to the former. Degrees of belief for the other levels are in practice negligible. For the second condition, there is strong agreement for low impact: the peak in the degree of belief for the latter value is accordingly higher than in the previous case. A very different situation is present for the last condition, where the three experts provided three different assessments. The effect to spread the degree of belief for the latter condition is evident.

Another interesting case is presented in Figure 4, related to the effect of "Lack of adequate training or experience" for the GTT "Complex interaction with software/tool". There is general consistency across the expert on the effect of each condition, as reflected in the distributions in the middle part of the figure. The large span in the expert assessments, from "low" to "extreme", results then in the large spread in the final aggregated distribution on the right.

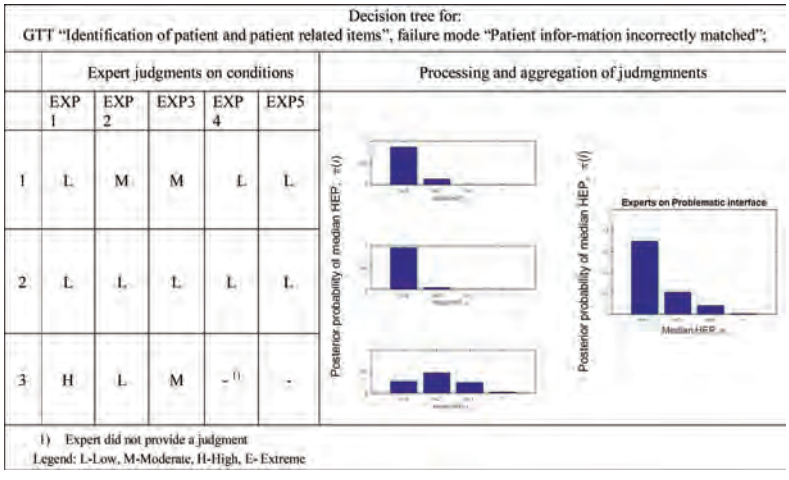


Figure 3. Processing and aggregation of judgments (GTT "Identification of patient and patient related items", failure mode "Patient information incorrectly matched"): Left: judgements from experts, Middle: expert-aggregated posterior distribution of median HEP for each condition, Right: posterior probability distribution of median HEP for the branch point.

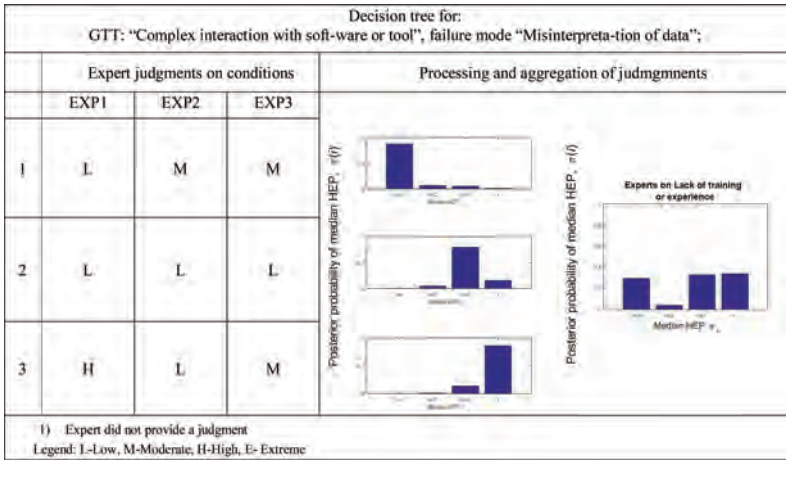


Figure 4. Processing and aggregation of judgments (GTT: "Complex interaction with software or tool", failure mode "Misinterpretation of data") Left: judgements from experts, Middle: expert-aggregated posterior distribution of median HEP for each condition, Right: posterior probability distribution of median HEP for the branch point.

5 CONCLUSIONS

The paper has presented the work performed to quantify the human failure probabilities to be used as input to a novel HRA method. Elicitation sessions were designed, with the following two main features. First, information on probabilities is asked to experts on a qualitative scale, with the goal of getting evidence on the order of magnitude for the probability. Second, specific situations are presented to the expert, i.e. specific failure scenarios influences by specific negative conditions. The latter feature was incorporated to avoid that the expert deal with abstract categories such as tasks types and influencing factors.

The expert statements are processed and aggregated to determine degrees of belief on the correct values of the failure probability. The latte is assumed as an inherently variable quantity so that the main parameter of its distribution is the subject of the elicitation. The Bayesian model used to aggregate the assessments was found to represent well the differences in the experts statements, providing a credible approach to process the expert input.

As a next step of the work, comparison of the obtained values with values from existing HRA methods is envisioned. Indeed, although HRA methods are sector-specific, some of the underlying data can be thought of being general, e.g. data regarding dealing with indicators, simple execution tasks. This comparison may provide some validation to the elicitation process.

With broader perspective, future work will apply the developed HRA method to hypothetical accident scenarios at the institute's Center for Proton Therapy.

ACKNOWLEDGMENT

This work is funded by CROSS (a PSI interdepartmental funding initiative) and the Paul Scherrer Institute's Center for Proton Therapy. The authors would also like to thank the personnel working at PSI's CPT for their cooperation and support. The authors acknowledge the support from the Future Resilient Systems program at the Singapore-ETH Centre, established between the Swiss Federal Institute of Technology in Zurich (ETH Zurich) and Singapore's National Research Foundation (FI 370074011).

REFERENCES

Bye et al. 2017. *The Petro-HRA Guideline*. IFE/HR/E-2017/001, Institute for energy technology, Halden, Norway.

Gibson, H. 2012. *Railway Action Reliability Assessment user manual*. UK Rail Safety and Standards Board Ltd.

Huq, M.S., Fraass, B.A., Dunscombe, P.B., Gibbons, Jr. J.P., Ibbott, G.S., Mundt, A.J., Mutic, S., Palta, J.R., Rath, F., Thomadsen, B.R., Williamson, J.F. and Yorke, E.D. 2016. The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management, *Medical physics* 43 (7).

López-Garrigós, M., Asencio, A., Lugo, J., Gutiérrez, M., Flors, L., Leiva-Salinas, C. 2013. Alert value reporting: a new strategy for patient safety. Salinas, M., *Clin Biochem.* 2013 46(3): 245–9.

Meyer, M.A. & Booker, J.M, 2001. *Eliciting and Analyzing Expert Judgment: A Practical Guide*. Society for industrial and applied mathematics.

Mkrtchyan, L., Podofillini, L., Dang, V.N. 2015. Bayesian belief Networks for Human reliability analysis: a review of applications and gaps. *Reliability Engineering & System Safety* 139:1–16.

Moieni et al., 1994. Advances in Human Reliability Analysis Methodology. Part I: Frameworks, Models and Data. *Reliability Engineering and System Safety* 44:27–55.

NASA, 2012. *Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners*. NASA/SP-2011-3421, December 2011.

NUREG 2007. Forester et al. *ATHEANA User's Guide*. NUREG-1880, US Nuclear Regulatory Commission, Washington DC, USA.

NUREG, 2016. NUREG-2199, Xing et al. 2016. *IDHEAS – A New Approach for Human Reliability Analysis, An Integrated Decision-Tree Human Event Analysis System (IDHEAS) Method for NPP internal at-power operation*. US Nuclear Regulatory Commission, Washington DC, USA.

Pandya, D., Podofillini, L., Emert, F., Lomax, A.J., Dang, V.N. 2017. Developing the foundations of a cognition-based human reliability analysis model via mapping task types and performance-influencing factors: Application to radiotherapy. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*. Available online: https://doi.org/10.1177/1748006X17731903.

Pandya, D., Podofillini, L., Emert, F., Lomax, A.J., Dang, V.N., Sansavini, G. 2017. Quantification of a human reliability analysis method from expert judgment for radiotherapy applications. Working Paper.

Podofillini, L. & Dang, V.N. 2013. A Bayesian Approach to Treat Expert-Elicited Probabilities in Human Reliability Analysis Model Construction. *Reliability Engineering & System Safety*. Volume 117, September 2013, Pages 52–64.

Tversky, A. & Kahneman, D. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science*. Vol. 185, Issue 4157: 1124–1131.

Wahi, M.M., Parks, D.V., Skeate R.C., Goldin, S.B., 2008. Reducing Errors from the Electronic Transcription of Data Collected on Paper Forms: A Research Data Case Study. *J Am Med Inform Assoc.* 15(3): 386–389.

Williams, J.C. (2015) HEART—A Proposed Method for Achieving High Reliability in Process Operation by Means of Human Factors Engineering Technology. *Safety and Reliability* 35(3) (reprint of orginal publication from 1983).

WHO 2008. *Radiotherapy risk profile*. World Health Organization (WHO) press.

Maintenance modeling and applications

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

An optimal maintenance policy based on partial information

R. Ahmadi
School of Mathematics, Iran University of Science and Technology, Narmak, Tehran, Iran

S. Wu
Kent Business School University of Kent, Canterbury, UK

ABSTRACT: This paper proposes an integrated model for maintenance scheduling of parallel systems whose failures are detected by inspections. A common characteristic of such systems is that the system failures are detected only by inspections and the failure of a component may not cause its system to fail. As such, the failure may not be immediately detected and the random (disruption) time at which the number of failed components reaches a certain predefined number d may therefore be unknown. For such systems, scheduling maintenance policy is a difficult task, which is tackled in this paper. The main issue considered here is to get an estimate of the disruption time on the basis of inspection point process observations in the framework of filtering theorem. The paper develops a unified cost structure to jointly optimise inspection frequency and replacement time for the system when the lifetime distribution of a component follows the Weibull distribution. Numerical results are provided to show the application of the proposed model. In addition, a sensitivity analysis is performed to examine the effect of maintenance parameters on the model.

1 INTRODUCTION

This paper proposes an approach to the joint determination of optimal inspection and replacement policies for m-component parallel systems subject to non-self announcing failures. The interest in such systems is natural as they provide a redundant approach to improving system reliability and availability. Nuclear reactor safety systems, emergency core cooling systems, fire detectors, and protective devise are good examples of parallel systems used in the real world. A common feature of such systems is that the failure of a component may not cause its system to fail and the system failures are detected only by inspections. As such, the failure may not be immediately detected and can be regarded as a hidden failure. Consequently, the number of failed components may be unknown at a given time. This raises a thought-provoking question: how can the failure time of d^{th} $(d = 1,2,\cdots,m-1)$ components (called disruption time) and the reliability and the maintenance cost be respectively estimated and analyzed to handle such a lack of data scenario while providing effective decision making? This paper attempts to answer those questions. The approach depends on the identification of the disruption time based on the inspection point process observations. The estimated disruption time contributes to scheduling a maintenance policy for an m-component parallel system. More generally, given partial information, the approach explored here can deal with two basic problems: how to inspect and when to stop operating the system and carrying out a replacement in order to detect the system failure and minimize some maintenance cost.

Although some maintenance models consider joint inspection and maintenance policies for systems whose state are detected only by inspections (He, Maillart, & Prokopyev 2015, Tambe, Mohite, & Kulkarni 2013, Tsaia, Sheu, & Zhangc 2017), there is only a few research handling such a lack of data while providing effective decision making. For instance, using the filtering theorem argument, Ahmadi & Wu (2017) propose a novel technique to estimate the disruption time, aiming at revealing the true state of the system. The technique also helps to deliver a warning, or an alert on approaching a disruption. This alert is the signal that an inspection or a preventive maintenance has to be performed. As such, the proposed technique augments the detection of failure. Further advantages of their model enhancing its applicability include modeling inspection through the modulated Poisson process and introducing a state-dependent cost structure. Both modeling approaches make the model superior to those (Berrade & Scarf 2012, He, Maillart, & Prokopyev 2015, Liu, Wang, Peng, &

Zhao 2015, Bjarnason & Teghipour 2016, Babishin & Teghipour 2016, Tsaia, Sheu, & Zhangc 2017) in which the inspection frequency and/or cost parameters do not respond to the variation of the system state. The structure also allows aperiodic inspections that is often more useful and realistic than the periodic policy, since it is more adaptive to deteriorating systems and typically leads to policies with lower costs.

The approach in this paper differs through the use of a general degradation model motivated by the extension of the model in the earlier paper by Ahmadi & Wu (2017). Furthermore, it encompasses and examines some characteristics which have not been addressed or studied in isolation.

2 ASSUMPTION AND MODEL DEVELOPMENT

2.1 *Model assumptions*

The following assumptions are made. (a) The paralell system consists of m components. (b) The failure of the system can only be detected by inspections ("hidden or non-self announcing failures"). (c) The inspection intensity process is assumed to follow a modulated Poisson process. (d) The only available information is given by the inspection point process observations (observation filtration). (e) The system is replaced at periodic times $\{t_r, 2t_r, \cdots\}$ (f) Inspections do not impact on the failure characteristics of the system.

2.2 *Model development*

2.2.1 *Modelling degradation*

Here we describe a stochastic model of degradation. For this discussion, we assume that all random variables are defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider a multicomponent parallel system consisting of m components whose lifetimes are independent and identically distributed random variables. The system is subject to random failure which can only be detected through inspection. The system state is characterized by a two unobservable states: a normal state and a degraded state. The transition time from the steady state to the degraded state called "disruption time" is defined as the first time the total number of failed components reach a predetermined threshold d:

$$\tau_d = \inf\{t : Y(t) = d\}; d = 1,2,\cdots,m-1 \tag{1}$$

where τ_d is the disruption time and $Y(t)$ is a stochastic process counting the total number of failed components up to time t. Denote

$$X_d(t) = \begin{cases} 0, & \text{if} \quad \tau_d > t, \\ 1, & \text{if} \quad \tau_d \le t. \end{cases} \tag{2}$$

From equations (1) and (2), one can see that the disruption time at which the system state, $X_d(t)$, jumps from 0 to 1 depends on the threshold's value d: a smaller value of d may result in earlier disruption.

2.2.2 *Modelling inspections*

The inspection modeling approach to detect the system failure is similar to that of Ahmadi & Newby (2011) and Ahmadi & Wu (2017) assuming that inter-inspection times conform to a modulated Poisson process. Specifically, let $N(t)$ be a modulated Poisson process such that $N(t)$ with the associated time points of inspections, $T_1 < T_2 < \cdots$, depicts the total number of arrivals up to time t. In other words, according to Aven and Jensen (Aven & Jensen 1998), $N(t)$ admits a smooth semimartingale (SSM) with the $\mathcal{F}$-intensity λ_t and the $\mathcal{F}$-martingale M_t:

$$N(t) = \int_0^t \lambda_s ds + M_t$$

where $t \in \mathbb{R}^+, M \in \mathcal{M}$ and $\mathcal{M}$ denotes the class of martingales adapted to the filtration $\mathcal{F}_t$. As noted, the inspection intensity process λ_t is modulated by the stochastic process $X_d(t)$ such that

$$\lambda_t = \lambda_{X_d(t)} = \lambda_0 + (\lambda_1 - \lambda_0) X_d(s), \quad \forall t \ge 0, \tag{3}$$

where λ_i satisfying

$$\lambda_0 < \lambda_1 < \infty, \tag{4}$$

denotes the rate of arrivals when the state of $X_d(t)$ is i (i = 0,1). Here, $X_d(t)$ influences and modulates the arrival rate of the Poisson process. The reader is referred to Özekici (Ozekici 1996) for more details about the modulated Poisson process. Equations (1)–(3) indicate that changes in the threshold's value d induce changes in both the disruption time and the inspection intensity: as d decreases, the state process $X_d(t)$ jumps sooner from the normal state to the degraded state which implies an increase in the number of inspections. It is evident that this modeling technique, in an elegant way, not only allows the inspection intensity responds to the variation of the system state, but also ensures (see equation (4)) the system upon approaching disruption is inspected more frequently which properly results in more certain detection of failure. Since the model including $N(t)$ is driven by the unobservable state process $X_d(t)$,

the first main problem investigated here is to get an estimate of the unobservable state $X_d(t)$ given the inspection point process observations up to time t, $\mathcal{F}_t^N = \sigma\{N(s): 0 \le s \le t\}$, that is

$$\phi_d(t) = \mathbf{E}\left[X_d(t) \middle| \mathcal{F}_t^N \right] = \mathbf{P}\left[\tau_d \le t \middle| \mathcal{F}_t^N \right]. \qquad (5)$$

The measure (5) not only contributes to detecting the disruption time τ_d at which the state process $X_d(t)$ jumps from the normal state to the degraded state (or, the estimate of components lifetime), but also allows the estimation of the intensity process $\lambda_{X_d(t)}$ by projection on the observed history. That is,

$$\hat{\lambda}_t = \mathbb{E}\left[\lambda_{X_d(t)} \middle| \mathcal{F}_t^N \right] = \mathbb{E}\left[\lambda_0 + (\lambda_1 - \lambda_0) X_d(t) \middle| \mathcal{F}_t^N \right].$$

The function $\hat{\lambda}_t$ reflecting the reaction of the maintenance crew is called inspection alert function. Indeed, given that the probability of the disruption detection at time t is $\phi_d(t)$, the inspection intensity that is at the discretion of the maintenance crew should be $\hat{\lambda}_t = \lambda_0 + (\lambda_1 - \lambda_0)\phi_d(t)$. The situation may be regarded as a case of condition-based maintenance in a sense that the function (5) provides some information (alert) regarding the state of the disruption. Through the inspection alert function $\hat{\lambda}_t$, the crew will use the information to perform inspections in order to detect system failures. Intuitively, this alert is the signal that an inspection has to be performed.

2.3 *Modelling disruption*

This section aims to provide a solution to the disruption alert function $\phi_d(t)$ by setting in the filtering theorem framework (Bremaud 1981). The solution technique used here is similar to that of models proposed by Ahmadi & Newby (2011).

2.3.1 *Complete information-based disruption model*

We have to detect the random time τ_d at which the inspection intensity $\lambda_{X_d(t)}$ increases: $\lambda_0 \mapsto \lambda_1$. For this, let $F(t) = \int_0^t f(u)du$ be the cumulative distribution of components lifetime. Using the semi-martingale argument (Aven & Jensen 1998), one can note that the increasing right-continuous state process $X_d(t) = I(\tau_d \le t)$ with the associated maintenance parameter d admits the following semi-martingale representation:

$$X_d(t) = \int_0^t \tilde{q}_{01}(s)\big(1 - X_d(s)\big)ds + m_t,$$

where m_t is an $\mathcal{F}_t$ -martingale and

$$\tilde{q}_{01}(t) = \begin{cases} \dfrac{f_d(t)}{1 - F_d(t)}, & \text{if } F_d(t) < 1; \\ 0, & \text{otherwise,} \end{cases}$$

denotes the transition rate of the state process $X_d(t)$ from normal state (0) to the degraded state (1) with the disruption distribution function

$$F_d(t) = \mathbf{P}(\tau_d \le t) = \sum_{k=d}^{m} \binom{m}{k} [F(t)]^k [1 - F(t)]^{m-k}.$$

Using the fact that the density function of the disruption time is

$$f_d(t) = \frac{m!}{(d-1)!(m-d)!} [F(t)]^{m-1} [1 - F(t)]^{m-d} f(t),$$

the transition rate of the state process $X_d(t)$ can be formulated as

$$\tilde{q}_{01}(t) = \frac{\binom{m}{d-1, m-d} q_{01}(t)}{\sum_{k=0}^{d-1} \binom{m}{k} [\exp(Q_{01}(t)) - 1]^{k-d+1}} \qquad (6)$$

where $q_{01}(t) = \frac{f(t)}{\bar{F}(t)}$ denotes the failure rate of the system and

$$Q_{01}(t) = \int_0^t q_{01}(s)ds.$$

2.3.2 *Partial information-based disruption model*

Since the disruption time at which the process $X_d(t)$ jumps from 0 to 1 is not directly observed, in the filtering theorem framework (Bremaud 1981), we get an estimate of the state process $X_d(t) = I(\tau_d \le t)$ based on the observed history F_t^N. More precisely, the problem is that of computing $\phi_d(t)$ given in (5) and getting an $\mathcal{F}^N$-adapted estimate of the inspection intensity process

$$\lambda_t = \lambda_{X_d(t)} = \lambda_0 + (\lambda_1 - \lambda_0) X_d(t), \qquad (7)$$

that is to say, before the disruption time τ_d, $\lambda_t = \lambda_0$, and after τ_d, $\lambda_t = \lambda_1$.

By projection on the observed history $\mathcal{F}_t^N$ one can show that

$$\phi_d(t) = \int_0^t [\tilde{q}_{01}(s) + (\lambda_1 - \lambda_0)\phi_d(s)]\big(1 - \phi_d(s)\big)ds + \sum_{n \ge 1} \frac{(\lambda_1 - \lambda_0)\phi_d(T_n^-)\big(1 - \phi_d(T_n^-)\big)}{\lambda_0 + (\lambda_1 - \lambda_0)\phi_d(T_n^-)} I(T_n \le t),$$

or, equivalently, between the jumps $\left(t \in [T_n, T_{n+1})\right)$,

$$\bar{\phi}_d(t,n) = \bar{\phi}_d(t)\, I(T_n \le t < T_{n+1}) = \bar{\phi}_d(T_n) - \int_{T_n}^{t} (\tilde{q}_{01}(s) + \bar{\lambda}\phi_d(s,n))\big(1 - \phi_d(s,n)\big)\, ds; \quad (8)$$

at the jumps,

$$\phi_d(T_n) = \frac{\lambda_1 \phi_d(T_n^-)}{\lambda_0 + \bar{\lambda}\phi_d(T_n^-)}, \quad (9)$$

where $\phi_d(T_n^-)$ denote the left limit of $\phi_d(\cdot)$ at time T_n, $\bar{\phi}(\cdot) = 1 - \phi(\cdot)$ and $\bar{\lambda} = \lambda_1 - \lambda_0$. From (8) one can see that for $t \in [T_n, T_{n+1})$

i. the sojourn time distribution in normal state is a decreasing function of the transition rate of the state process $X_d(t)$ that is $\tilde{q}_{01}$ and $\bar{\lambda}$. The latter can be realized as a measure whose scale reflects the reaction intensity of the maintenance crew to perform inspections upon switching of $X_d(t)$ from normal state to the degraded state.
ii. the disruption time distribution $\phi_d(t,n)$ satisfies the differential equation (7) with the initial condition (9):

$$\phi_d'(t,n) = \left(\tilde{q}_{01}(t) + \bar{\lambda}\phi_d(t,n)\right)\left(1 - \phi_d(t,n)\right) > 0, \quad (10)$$

where the positive derivative implies increasing trajectories of $\phi_d(t,n)$ between the jumps.
iii. Let $\bar{\lambda} = 0$. This assumption intuitively implies that the maintenance crew (inspection intensity) does respond to the variation of the state process $X_d(t) : 0 \mapsto 1$. From (ii) an explicit solution to the sojourn time distribution of $X_d(t)$ can be given by

$$\bar{\phi}_d(t,n) = \exp\left(-\int_0^t \tilde{q}_{01}(s)\, ds\right).$$

From above argument one can see that by projection on the observed history $\mathcal{F}_t^N$ for $t \in [T_n, T_{n+1})\,(n \ge 0)$ and the use of the disruption time distribution $\phi_d(t,n)$, an $\mathcal{F}_t^N$-adapted estimate of the inspection intensity (7) can be given by

$$\hat{\lambda}_t = \mathbb{E}[\lambda_{X_d(t)} \mid \mathcal{F}_t^N] = \lambda_0 + \bar{\lambda}\phi_d(t,n). \quad (11)$$

To gain a better understanding of the impact of the disruption alert function $\phi_d(t,n)$ and the inspection alert function $\hat{\lambda}_t$ on both the performance of inspections and failures, let the measure of alertness on approaching a disruption, $\phi_d(t,n)$, tend to 1. From (11) we get $\hat{\lambda}_t \to \lambda_1$ $(\lambda_1 > \lambda_0)$ that indicates more frequent inspections to prevent the system failure.

2.4 *Mean time to disruption*

The measure $\bar{\phi}_d(t,n)$ can contribute to determination of the mean time to disruption m_d:

$$m_d = \int_0^\infty \bar{\phi}_d(t,n)\, dt.$$

The mean time to disruption can provide some information regarding the disruption time at which $X_d(t)$ jumps from normal state to the degraded state. The crew will use this insight to perform maintenance in order to timely detect the components failure of the system.

From equations (8) and (9) one can note that the solution of the differential equation (10) based on the initial condition (9) depends on the estimation of inspection times T_n $(n \ge 1)$. In Section 2.5 we devise an inspection scheduling function based on the $\mathcal{F}^N$-adapted inspection intensity. We will see the explored scheduling function incorporating the disruption time distribution $\phi_d(t,n)$ enables us (i) to provide a systematic approach to the determination of non-periodic inspection times and (ii) to estimate the disruption time distribution $\phi_d(t,n)$ over inter-arrival inspection times $[T_n, T_{n+1})\,(n \ge 1)$.

2.5 *Scheduling inspections*

Let $\varphi_n(v)$ for $v \in [0, V_{n+1})$ be the distribution function of the $(n+1)^{\text{th}}$ inter-inspection time $V_{n+1} = T_{n+1} - T_n\,(n \ge 0)$ adapted to the observed information $\mathcal{F}^N$. Then for $t \in [T_n, T_{n+1})$ we have

$$\bar{\varphi}_n(v) = \mathbf{P}(V_{n+1} \ge v \mid \hat{\lambda}_t) = \exp\left(-\int_{T_n}^{T_n+v} \hat{\lambda}_s\, ds\right),$$

where $\bar{\varphi}_n(v) = 1 - \varphi_n(v)$. Since $\hat{\lambda}(t,n) = \lambda_0 + (\lambda_1 - \lambda_0)\phi_d(t,n)$, $t \in [T_n, T_{n+1})$, $\bar{\varphi}_n(v)$ can be expressed as

$$\bar{\varphi}_n(v) = \exp(-\lambda_0 v) \times \exp\left(-(\lambda_1 - \lambda_0)\int_{T_n}^{T_n+v} \phi_d(s,n)\, ds\right). \quad (12)$$

If $\mu_n = \mathbb{E}[V_{n+1} \mid \hat{\lambda}_t]\,(n = 0,1,2,\cdots)$ denote the $(n+1)^{\text{th}}$ expected time between inspections, using the inter-inspection time distribution (12), an $\mathcal{F}^N$-adapted estimate of inter-inspection times can be given by

$$\mu_n = \int_0^\infty \bar{\varphi}_n(t)\, dt \quad (13)$$

This provides a sequence of inspection times $\eta_n = \sum_{k=0}^{n-1} \mu_k\,(n = 1,2,\cdots)$.

2.6 *Example: Examining the model*

To examine the response of the model to threshold's values d = 1,2, consider a 3-component

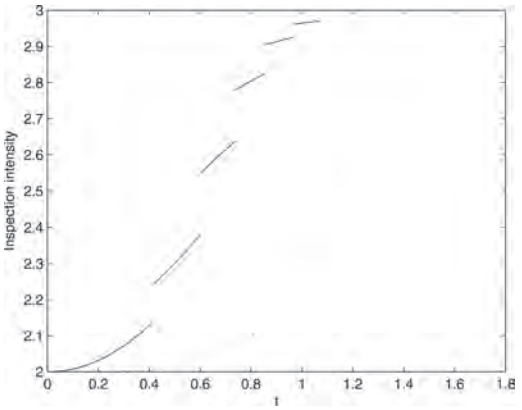


Figure 1. Inspection intensity $\hat{\lambda}_t$ for different $d \in \{1,2\}$.

parallel system whose components lifetime conforms to a Weibull distribution with the failure rate

$$q_{01}(t) = \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}. \tag{14}$$

In addition let $(\alpha,\beta) = (2,2)$ and $(\lambda_0,\lambda_1) = (2,3)$. Given known parameters, by using the equation (10), an expression for the transition rate $\tilde{q}_{01}$ corresponding to $d = 1,2$ can be given by

$$\tilde{q}_{01} = 1.5t, \qquad \tilde{q}_{01} = \frac{3t \times (exp(0.25t^2) - 1)}{3exp(0.25t^2) - 2}. \tag{15}$$

Figure 1 reveals that the inspection intensity $\hat{\lambda}_t$ in terms of the sojourn time distribution $\phi_d(t,n)$ emerging as solution of the differential equation (10) strongly depends on the threshold's value d: as the threshold's value d decreases to 1 the inspection intensity occurs more frequently. In addition, Figure 1 indicates that the inspection intensity as a measure of alertness reflects the reaction of the maintenance crew. In the sense that they deliver an alert regarding a possibly approaching disruption and this alert is used to make inspections. Thus, the alert is a signal that an inspection has to be performed to detect a failure.

3 COST MODEL

This section aims to jointly determine an optimal replacement policy and an optimal inspection frequency with respect to a state-dependent cost-reward model. The cost structure used here is similar to that of the model proposed by Ahmadi & Wu (2017).

3.1 *Average cycle costs*

A cycle is comprised of a sequence of inspections that ends with replacement scheduled at periodic times $\{t_r, 2t_r, \cdots\}$. Repair and maintenance actions in a cycle incur costs that include: (i) inspection costs to detect the system failures, (ii) a penalty cost associated with undetected failures and (iii) a periodic replacement cost made after every t_r units of time. More precisely, each inspection at time t incurs a time-dependent cost $c_t = ct\ (c > 0)$. Undetected failures within inter-inspection times generate a state-dependent penalty cost per unit time

$$C_{X_d(t)} = C_0 + (C_1 - C_0)X_d(t),$$

$(C_1 > C_0)$. This implies that as the system state shifts to more degraded state $X_d(t): 0 \mapsto 1$, the penalty cost increases from C_0 to C_1. In addition, we assume that the replacement of the system in different states incurs different costs. In other words, the replacement of the system at time t is measured by a state-dependent cost

$$k_{X_d(t)} = k_0 + (k_1 - k_0)X_d(t),$$

$(k_1 > k_0)$. It is noted, as the state process $X_d(t)$ moves from the normal state $(X_d(t) = 0)$ to the degraded state $(X_d(t) = 1)$ the replacement cost increases $k_0 \mapsto k_1$. With respect to the above cost structure, the total cost up to the periodic replacement time t_r termed by $C_d(t_r)$ can be expressed as

$$C_d(t_r) = \underbrace{\int_0^{t_r} c_s dN(s)}_{\text{Inspection cost}} + \underbrace{\int_0^{t_r} C_{X_d(s)} ds}_{\text{Penalty cost}} + \underbrace{k_{X_d(t_r)}}_{\text{Replacement cost}}.$$

The last cost is incurred as the system is found in the state $X_d(t_r)$ at replacement time t_r. Since $\lambda_{X_d(t)}$ is an $\mathcal{F}$-intensity of $N(t)$ and both the $C_{X_d(s)}$ and $k_{X_d(t)}$ are $\mathcal{F}$-measurable, an $\mathcal{F}$-adapted total cost $\mathcal{C}_d(t,n), t \in [T_n, T_{n+1})\ (n \geq 0)$, can be expressed as

$$\mathcal{C}_d(t_r,n) = \mathbb{E}[C_d(t_r) \mid \mathcal{F}_{t_r}] = k_0 + \int_0^{t_r} c_d(s,n)ds,$$

where

$$c_d(t,n) = \left[C_{\lambda_0}(t) + \left(C_{\lambda_1}(t) - C_{\lambda_0}(t)\right)X_d(t)\right] + (k_1 - k_0)dX_d(t), \tag{16}$$

is the total cost rate and $C_{\lambda_i}(t) = C_i + \lambda_i c_t$ for $i = 0,1$.

3.2 *Partial information-based cost model*

Since the term $c_d(t,n)$ depends on the state indicator $X_d(t)$ not directly observed, a projection

on the observation filtration $\mathcal{F}^N$ is needed. As described in Section 2.3.2 such a projection from the $\mathcal{F}$-level to the $\mathcal{F}^N$-level leads to the following expression;

$$\hat{c}_d(t,n) = \Big[C_{\lambda_0}(t) + \big(C_{\lambda_1}(t) - C_{\lambda_0}(t) \big) \phi_d(t,n) \Big] + \big(k_1 - k_0 \big) d\phi_d(t,n). \tag{17}$$

Thus, an $\mathcal{F}^N$-adapted estimate of the expected total cost can be given by

$$\hat{\mathcal{C}}_d(t_r,n) = k_0 + \int_0^{t_r} \hat{c}_d(s,n)\, ds, \tag{18}$$

The integrand $\hat{c}_d(s,n)$ is the conditional expectation of the cost rate at time s given the observations up to time s. In addition, by plugging the derivative of $\phi_d(t,n)$ (see equation (10)) into equation (17), in terms of the sojourn time distribution $\bar{\phi}_d(t,n)$, an $\mathcal{F}^N$-adapted estimate of the cost rate can be given by

$$\hat{c}_d(t,n) = -k_{10}\bar{\lambda}\bar{\phi}_d^2(t,n) + \big(\bar{\lambda}(k_{10} - ct) + k_{10}\tilde{q}_{01}(t) - C_{10} \big) \bar{\phi}_d(t,n) + C_{\lambda_1}(t)$$

where $k_{10} = k_1 - k_0$ and $C_{10} = C_1 - C_0$.

Continuing the example we illustrate an evolution of the maintenance cost rate for both alert parameters d = 1,2 (see Figure 2). One can note that lower threshold value d causes an earlier alert on approaching a disruption which makes inspections more frequent, increases penalty cost and so incurring more maintenance cost.

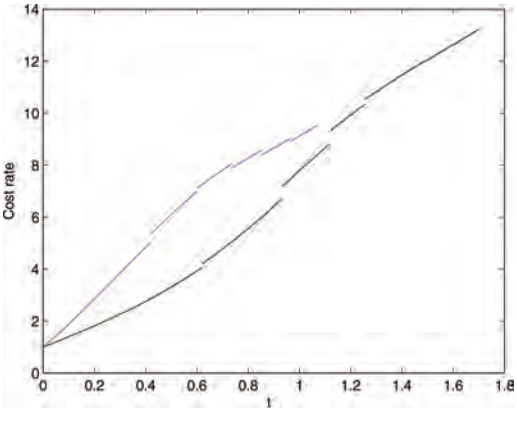


Figure 2. $\mathcal{F}^N$-adapted cost rate for different $d \in \{1,2\}$. Blue and black color correspond to threshold values $d = 1$ and $d = 2$ respectively.

3.2.1 *Long-run average cost rate*

The main objective is to minimize the long-run average cost rate by optimizing the periodic replacement time t_r. To this end, let $\psi_d(n,t_r)$ be the long-run average cost per unit time. Since the sequence of replacement times $\{t_r, 2t_r, \cdots\}$ forms a regenerative process, the time between two consecutive replacements is a renewal cycle. Therefore, by the renewal reward theorem (Ross 1970), the long-run average cost rate is the average per cycle cost divided by the cycle length t_r given by

$$\psi_d(t_r,n) = \frac{\hat{\mathcal{C}}_d(t_r,n)}{t_r},$$

with expression for $\hat{\mathcal{C}}_d(t_r,n)$ provided in (18). We set out to solve the optimization problem

$$t_r^* = \underset{t_r \in \mathbb{R}_+}{\arg\min} \big\{ \psi_d(t_r,n) \big\},$$

along with determining the optimal inspection frequency n^* subject to the optimal replacement time t_r^*.

4 NUMERICAL RESULTS

The aim of the proposed model is to derive a cost-optimal inspection and maintenance policy for a three-component parallel system subject to non-self announcing failures. For this, let the components lifetime be modelled by the Weibull distribution with the failure rate (14). Main numerical results are based on the known degradation parameters $(\alpha,\beta) = (2,2)$ and the inspection parameters $(\lambda_0,\lambda_1) = (2,3)$. The choice for cost parameters are $(C_0,C_1) = (1,3), (k_0,k_1) = (3,6)$ and $c = 2$.

4.1 *Optimal maintenance policy*

To investigate the effect of the alert parameter d on the model, optimal solutions for different threshold values $d \in \{1,2\}$ are obtained (see Table 1 and Figure 2). As noted above, the alert parameter d reflects the attitude of the decision maker towards inspection: changes in d induce changes in the inspection intensity upon disruption time at which dth component of the system experiences failure. Given inspection parameters $(\lambda_0,\lambda_1) = (2,3)$, results reveal insignificant impact of d on the optimal replacement time. On the other hand, the optimal maintenance cost and the optimal frequency of inspection range over several orders of magnitude as the threshold varies. Subject to the results given in Table 1 and Table 2, the sequence of inspections and maintenance actions

Table 1. Optimal parameters and mean time to disruption m_d for different $d \in \{1,2\}$ given $\lambda_1 = 3$.

d	n^*	m_1	t_r^*	m_2	$\psi_d(t_r^*, n^*)$
1	**4**	0.6316	**0.8553**	1.1593	**8.7575**
2	**1**	0.6316	**0.9156**	1.1593	**6.7237**

Table 2. Expected inspection times for different $d \in \{1,2\}$ given $\lambda_1 = 3$.

d	η_1	η_2	η_3	η_4	η_5	η_6
1	0.414	0.6	0.736	0.853	0.936	1.07
2	0.62	0.93	1.115	1.254	1.373	1.484

corresponding to the threshold values $d = 1,2$ are respectively scheduled as follows:

- starting from the normal state, inspections with intensity $\lambda_0 = 2$ are scheduled at times η_i $(i = 0,1)$. Upon failure of the first component at (disruption) time $m_1 = 0.6316$ inspections are considered more often $(\lambda_0 \mapsto \lambda_1)$. Following disruption, inspections to detect the system failures are scheduled at η_i $(i = 2,3)$. To restrain the frequency of inspections and avoid a costly strategy, the model suggests the replacement of the system at $t_r^* = 0.8553$ before disruption time $m_2 = 1.1593$ at which the second component of the system experiences failure. This process incurs the optimal cost $\psi_d(t_r^*, n^*) = 8.7575$.
- starting from the normal state, the first inspection is scheduled at $\eta_0 = 0.62$ just before the first disruption at $m_1 = 0.6316$ and a planned replacement is recommended during inter-disruption times $[0.6316, 1.1593]$ at $t_r^* = 0.9156$. This incurs the optimal cost $\psi_d(t_r^*, n^*) = 6.7237$.

4.2 *Attitude to maintenance*

The inspection model adapts itself to the decision maker's attitudes to maintenance (the value of d) by moving the inspection parameter $\lambda_0 \mapsto \lambda_1$. Increasing the threshold value $d : 1 \mapsto 2$ affects both the inspection intensity and the transition rate of the state process $(\tilde{q}_{01}(t))$ from normal state to more degraded state (see equation (15)) which gives higher value for the mean time to disruption $m_d : 0.6316 \mapsto 1.1593$. The higher threshold d reduces the frequency of inspection and penalty cost which results in reducing overall maintenance cost. As illustrated by Table 1 and Figure 3, the threshold value $d = 2$ leads to more effective strategy as the disruption time $m_2 = 1.1593$ (failure time of the second component) is preceded by

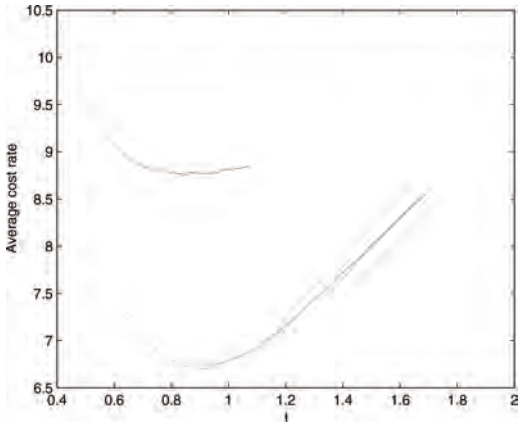


Figure 3. Average cost rate for different $d \in \{1,2\}$ given $(\lambda_0, \lambda_1) = (2,3)$. Blue and black color correspond to threshold values $d = 1$ and $d = 2$ respectively.

Table 3. Optimal parameters and mean time to disruption m_d for different $d \in \{1,2\}$ given $\lambda_1 = 4$.

d	n^*	m_1	t_r^*	m_2	$\psi_d(t_r^*, n^*)$
1	**3**	0.5978	**0.7145**	1.0808	**9.4972**
2	**1**	0.5978	**0.8609**	1.0808	**6.8923**

the replacement time $t_r^* = 0.9156$. On the other hand, the model based on the maintenance threshold $d = 1$ penalizes a costly strategy which favors not only an early replacement, but more frequent inspections (see Table 1 and Table 2).

4.3 *Level of maintenance*

The response of the model to maintenance level is examined with inspection parameter $\lambda_1 \in \{3,4\}$. Given $\lambda_1 = 4$ for $d \in \{1,2\}$ the optimal parameters, the mean time to disruption and inspection times for both threshold values are shown in Table 3 and Table 4. Also, an evolution of average cost rate is illustrated by Figure 4. In both cases the optimal cost increases with λ_1 implying an increase in the amount of maintenance undertaken on the system. On the other hand, with increasing λ_1 or $\bar{\lambda}$ more alert on a possibly approaching disruption is discovered. As discussed in Section 2.3.2, part (ii), this causes a reduction in sojourn time of $X_d(t)$ in normal state hence decreasing disruption time (see Table 3). In addition, given $d = 1$, with increasing inspection intensity $\lambda_1 : 3 \mapsto 4$ inspections will be considered more often but t_r^* decreases to restrain the frequency of inspections. But, in the case $d = 2$, due to insignificant effect of λ_1 on the inspection

Table 4. Expected inspection times for different $d \in \{1,2\}$ given $\lambda_1 = 4$.

d	η_1	η_2	η_3	η_4	η_5	η_6
1	0.405	0.583	0.713	0.827	0.935	1.041
2	0.61	0.904	1.076	1.207	1.322	1.431

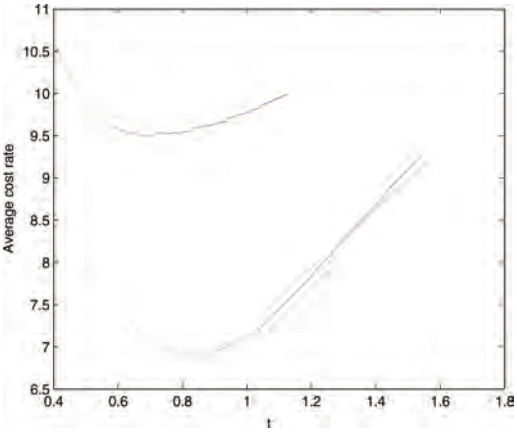


Figure 4. Average cost rate for different $d \in \{1,2\}$ given $(\lambda_0, \lambda_1) = (2,4)$. Blue and black color correspond to threshold values $d = 1$ and $d = 2$ respectively.

times, the optimal solutions (n^*, t_r^*) and the resulting optimal average cost are not particularly sensitive to the inspection parameter λ_1.

5 CONCLUSIONS

The main issue discussed here is to detect the random time of change of the unobservable state (disruption time) based on partial information for systems subject to hidden failures. The approach to estimating the disruption time distribution, which rests on an associated maintenance alert parameter d, helps in two ways: firstly, it delivers an alert or a signal on approaching the components failure of the system and to control the inspection frequency in order to timely detect the system failures. Secondly, it provides insight to perform a replacement in order to avoid typically more costly system failures. The other problem investigated here is the construction and the estimation of a unified cost model based on the available information. The estimated cost model as a measure of policy contributes to the joint determination of an optimal inspection frequency and an optimal replacement time for systems with partial information. The main examples considered are that of a Weibull degradation model for a three-component parallel safety system with the alert parameter *d*. The response of the model to both the threshold and the inspection parameter is examined. Although the numerical results did not reveal a strong influence of the threshold's value *d* on optimal replacement time, but its impact on other parameters including inspection times was more significant. Also, the results provide sensible and realistic inspection policies for such systems and gives insight into the effect of applying various inspection policies.

REFERENCES

Ahmadi, R. & M. Newby (2011). Maintenance scheduling of a manufacturing system subject to deterioration. *Reliability Engineering and System Safety 96(10)*, 1411–1420.

Ahmadi, R. & S. Wu (2017). A novel data-driven approach to optimizing replacement policy. *Reliability Engineering and System Safety 167*, 506–516.

Aven, T. & U. Jensen (1998). *Stochastic Models in Reliability*. New York: Springer.

Babishin, V. & S. Teghipour (2016). Joint optimal maintenance and inspection for a k-out-of-n system. *Int J Adv Manuf Technol. 87(5)*, 1739–1749.

Berrade, C. & P. Scarf (2012). Optimal replacement in the proportional hazard model. *Eur. J. Oper. Res*. 218(3), 360–367.

Bjarnason, E. & S. Teghipour (2016). Periodic inspection frequency and inventory policies for a k-out-of-n system. *IEEE Transactions 48(7)*, 638–650.

Bremaud, P. (1981). *Point processes and queues.* New York: Springer.

He, K., L. Maillart, & O. Prokopyev (2015). Scheduling preventive maintenance as a function of an imperfect inspection interval. *IEEE Trans. Rel. 64(3)*, 983–997.

Liu, X., W. Wang, R. Peng, & F. Zhao (2015). A delay-time-based inspection model for parallel systems. *J. Risk and Reliability 229(6)*, 556–567.

Ozekici, S. (1996). *Complex systems in random environments, In: Ozeciki S (ed) Reliability and maintenance of complex systems. NATO ASI Series Vol. F154.* New York: Springer-Berlin Heidelberg.

Ross, S. (1970). *Applied probability models with optimization applications*. San Francisco: Holden-Day.

Tambe, P., S. Mohite, & M. Kulkarni (2013). Optimisation of opportunistic maintenance of a multi-component system considering the effect of failures on quality and production schedule: a case study. *Int J Adv Manuf Technol. 69(5)*, 1743–1756.

Tsaia, H., S. Sheu, & Z. Zhangc (2017). A trivariate optimal replacement policy for a deteriorating system based on cumulative damage and inspections. *Reliability Engineering and System Safety 160*, 74–88.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Two imperfect repair models for a gamma deteriorating system: A comparison

Sophie Mercier
Univ Pau & Pays Adour/CNRS, IPRA-LMAP, 64000 PAU, France

I.T. Castro
Department of Mathematics, University of Extremadura, Spain

ABSTRACT: A system is considered, which is deteriorating over time according to a non-homogeneous gamma process. The point of the presentation is to propose and compare two models of imperfect repairs for the system. For sake of simplicity, only periodic (and instantaneous) repairs are here envisioned. The first model, called the Arithmetic Reduction of Deterioration of order 1 (ARD1), assumes that a repair removes a given proportion of the degradation accumulated by the system from the last maintenance action. The second model, called Arithmetic Reduction of Age of order 1 (ARA1), refers to the virtual age models proposed by Kijima (1989) and further studied by Doyen & Gaudoin (2004) in the context of recurrent events: the ARA1 model assumes that a repair reduces the age accumulated by the system since the last maintenance action, in a given proportion. An ARD1 repair hence lowers the deterioration level, without rejuvenating the system. On the contrary, by an ARA1 repair, the system is put back to the exact situation where it was some time before, which entails the lowering of both its deterioration level and (virtual) age. The two models may hence correspond to different maintenance actions in an applicative context. This presentation focuses on the comparison between the two models, from a probabilistic point of view (moments and stochastic ordering). An application in a maintenance optimization context is also provided, for illustration purpose. A specific case is analyzed, where the two repair models provide identical expected deterioration levels at maintenance times ("equivalent" case). The comparison results can help understanding which among the two models is the best adapted in an applicative context.

1 INTRODUCTION

Many systems suffer a physical degradation before they fail. This degradation is a complex process as it depends of many factors (material, stress loads, temperature, …). To mitigate the effect of the system degradation and to extend the system lifetime, a large volume of maintenance models have been proposed in the literature, with different maintenance actions. Most of these models are limited to perfect repairs (Huynh et al. 2014, Caballé et al. 2015, Hong et al. 2014 among others). However, imperfect maintenance actions describe more realistic situations than perfect repairs. Some advances have been made to include imperfect repairs in a degrading system. Alaswad & Xiang (2017) classified the impact of the maintenance actions over the maintained system into three types. The first type assumes that the maintenance actions return the system to a previous stage of deterioration. In the second type, the imperfect maintenance reduces the degradation level of the maintained system (Castro & Mercier 2016). The third approach is based on the idea that the maintenance action changes the rate of degradation of the system (Zhang et al. 2015). However, as Zhang et al. (2015) claimed, the issue of treating imperfect maintenance in the context of degrading systems remains widely open nowadays.

Following the spirit showed in (Mercier & Castro 2013) and (Castro & Mercier 2016), two models of imperfect repair are here proposed and compared for a system accumulating deterioration over time. The first model, called the Arithmetic Reduction of Deterioration of order 1 (ARD1), assumes that the repair removes the proportion ρ of the degradation accumulated by the system from the last maintenance action (with $\rho \in (0, 1)$). The second model, called Arithmetic Reduction of Age of order 1 (ARA1), refers to the virtual age models proposed by Kijima (1989) and further studied by Doyen & Gaudoin (2004) in the context of recurrent events: the ARA1 model assumes that the repair removes the proportion ρ of the age accumulated by the system since the last maintenance action. An ARD1 repair hence lowers the deterioration level, without rejuvenating the system. On the contrary, by an ARA1 repair, the system is put

back to the exact situation where it was some time before, which entails the lowering of both its deterioration level and (virtual) age. The two models may hence correspond to different maintenance actions in an applicative context.

For a better understanding of the differences between the two models, this paper focuses on their comparison, from a probabilistic point of view. Assuming that the degradation of the system is modeled by a non homogeneous gamma process, stochastic comparisons of both location and spread of the two resulting processes are given. Moreover, a specific case is analyzed, where the two models provide identical expected deterioration levels at repair times ("equivalent" case). Finally, an illustration of the two models is provided, by including them in a global maintenance policy, with replacement of the system when it is too deteriorated.

The paper is organized as follows: The two models of imperfect repair are described in Section 2. The comparison results are given in Section 3 (including the "equivalent" case). Section 4 deals with the application to the global maintenance strategy and concluding remarks are provided in Section 5, together with possible extensions.

2 THE TWO MODELS OF IMPERFECT REPAIRS

2.1 *The intrinsic deterioration and notations*

For $a, b > 0$, let us first recall that the gamma distribution $\Gamma(a, b)$ with parameters (a, b) admits the following p.d.f. (probability distribution function):

$$f_{a,b}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \forall x > 0.$$

The corresponding mean and variance are $\frac{a}{b}$ and $\frac{a}{b^2}$, respectively.

A system is considered with degradation modeled by a non homogeneous gamma process $(X_t)_{t\geq 0}$ with parameters $A(\cdot)$ and b, where $A(\cdot): \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and non-decreasing with $A(0) = 0$, and $b > 0$. We recall that $(X_t)_{t\geq 0}$ is a process with independent increments such that $X_0 = 0$ almost surely and such that each increment $X_{t+s} - X_t$ is gamma distributed $\Gamma(A(t+s) - A(t), b)$ for all $s, t > 0$.

The system is periodically and instantaneously repaired each T units of time. For modeling purpose, we set $X^{(i)}, i \in \mathbb{N}^*$ to be i.i.d. copies of $X = (X_t)_{t\geq 0}$, where $X^{(i)}$ describes the evolution of the deterioration level between the i-th and $(i + 1)$-th maintenance actions. For each imperfect repair model, the maintenance efficiency is measured by an Euclidian parameter $\rho \in (0, 1)$.

2.2 *First model: Arithmetic Reduction of Deterioration of order 1 (ARD1)*

In this model, the maintenance action instantaneously removes the proportion ρ of the degradation accumulated by the system from the last maintenance action (or from the origin). We set $(Y_t)_{t\geq 0}$ be the process that describes the degradation level of the maintained system under this model of repair.

The ARD1 model is sketched in Figure 1 for $\rho = 0.5$. It is developed as follows: At the beginning, the system deteriorates according to $X^{(1)}$. It is first maintained at time T, where the proportion ρ

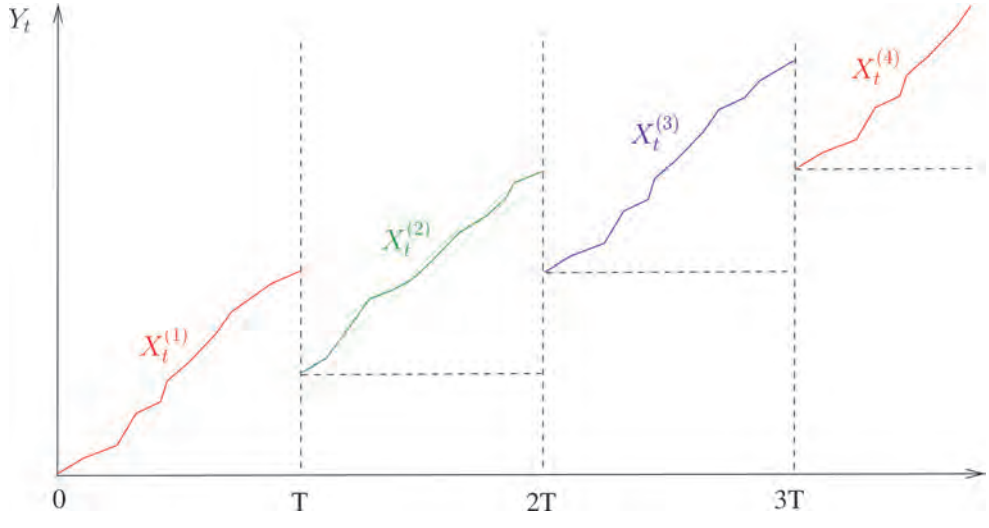


Figure 1. Illustration of the ARD1 policy for $\rho = 0.5$.

of the accumulated deterioration since the origin is removed. This provides:

$$Y_t = X_t^{(1)} \quad t < T, \qquad Y_T = (1-\rho) X_T^{(1)}.$$

Between T and $2T$, the system deteriorates according to $X^{(2)}$. The age of the system is unchanged at time T and we have

$$Y_t = Y_T + \left(X_t^{(2)} - X_T^{(2)}\right)$$

for all $T \le t < 2T$. At the second maintenance time $2T$, the proportion ρ of the degradation accumulated between T and $2T$ is removed, which provides:

$$Y_{2T} = Y_T + (1-\rho)\left(X_{2T}^{(2)} - X_T^{(2)}\right).$$

More generally, we have:

$$Y_t = Y_{nT} + \left(X_t^{(n+1)} - X_{nT}^{(n+1)}\right)$$

for all $nT \le t < (n+1)T$, where $X_t^{(n+1)} - X_{nT}^{(n+1)}$ is gamma distributed $\Gamma\left(A(t) - A(nT), b\right)$ and

$$Y_{(n+1)T} = Y_{nT} + (1-\rho)\left(X_{(n+1)T}^{(n+1)} - X_{nT}^{(n+1)}\right).$$

This provides

$$Y_{nT} = (1-\rho)\sum_{i=1}^{n}\left(X_{iT}^{(i)} - X_{(i-1)T}^{(i)}\right)$$

with $\Gamma(A(nT), \frac{b}{1-\rho})$ as distribution.

When $\rho \to 1^-$, then $Y_{nT} \to 0^+$ and the system is renewed at time nT (As Good As New repair: AGAN). When $\rho \to 0^+$, the repair is ineffective and it is As Bad As Old (ABAO).

Except for the case $\rho \to 0^+$, if $t \bmod T \neq 0$, Y_t is the sum of two independent and gamma distributed random variables (r.v.s) with different scale parameters, and it is not gamma distributed. Its expectation and variance are given by:

$$\mathbb{E}(Y_t) = \frac{A(t) - \rho A(nT)}{b}, \tag{1}$$

$$var(Y_t) = \frac{A(t) - \rho(2-\rho) A(nT)}{b^2}, \tag{2}$$

for $nT \le t < (n+1)T$.

It is easy to check that $\mathbb{E}(Y_t)$ and $var(Y_t)$ are decreasing with respect to ρ, that is, the more efficient the repair, the smaller the expectation and variance of the deterioration level. In this way, the expectation and variance are minimal when $\rho \to 1^-$ that is when the repair is AGAN, and maximal when $\rho \to 0^+$ that is when the repair is ABAO.

2.3 *Second model: Arithmetic Reduction of (virtual) Age of order 1 (ARA1)*

As told in the introduction, the ARA1 model is based on the notion of virtual age, which is reduced by each maintenance action: each repair removes the proportion ρ of the age accumulated by the system since the last maintenance action (or from the origin). This means that, at each maintenance action, the system goes back into its past: the deterioration level is hence reduced (just as for the ARD1 model) but the system is also rejuvenated at the same time (it becomes younger). In case of an increasing deterioration rate ($A(\cdot)$ convex), the rate of deterioration is hence reduced by the repair together with the deterioration level. We set $(Z_t)_{t\ge 0}$ be the process that describes the degradation level of the maintained system under this model of repair.

The ARA1 model is sketched in Figure 2 for ρ = 0.5. It is developed as follows: At the first maintenance time T, the (virtual) age of the system is suddenly reduced of ρT, so that it becomes $T - \rho T = (1 - \rho)T$. This provides:

$$Z_t = X_T^{(1)} \text{ for } t < T, \qquad Z_T = X_{(1-\rho)T}^{(1)}.$$

For $T \le t < 2T$, the age of the system is $(1-\rho)T + (t-T) = t - \rho T$. We get:

$$Z_t = Z_T + \left(X_{t-\rho T}^{(2)} - X_{(1-\rho)T}^{(2)}\right).$$

At time $2T^-$ (just before the repair), the age of the system is $2T - \rho T$ which is reduced of ρT at time $2T$. The age hence is $2(1-\rho)T$ at time $2T$. This provides:

$$Z_{2T} = Z_T + \left(X_{2(1-\rho)T}^{(2)} - X_{(1-\rho)T}^{(2)}\right).$$

More generally, for $nT \le t < (n+1)T$, the virtual age of the system at time t is $t - \rho nT$ (which is just the same as for an ARA1 model for recurrent events, see (Doyen & Gaudoin 2004)). We obtain:

$$Z_t = Z_{nT} + \left(X_{t-\rho nT}^{(n+1)} - X_{(1-\rho)nT}^{(n+1)}\right),$$

for all $nT \le t < (n+1)T$, where $X_{t-\rho nT}^{(n+1)} - X_{(1-\rho)nT}^{(n+1)}$ is gamma distributed $\Gamma(A(t-\rho nT) - A((1-\rho)nT), b)$ and

$$Z_{(n+1)T} = Z_{nT} + \left(X_{(1-\rho)(n+1)T}^{(n+1)} - X_{(1-\rho)nT}^{(n+1)}\right).$$

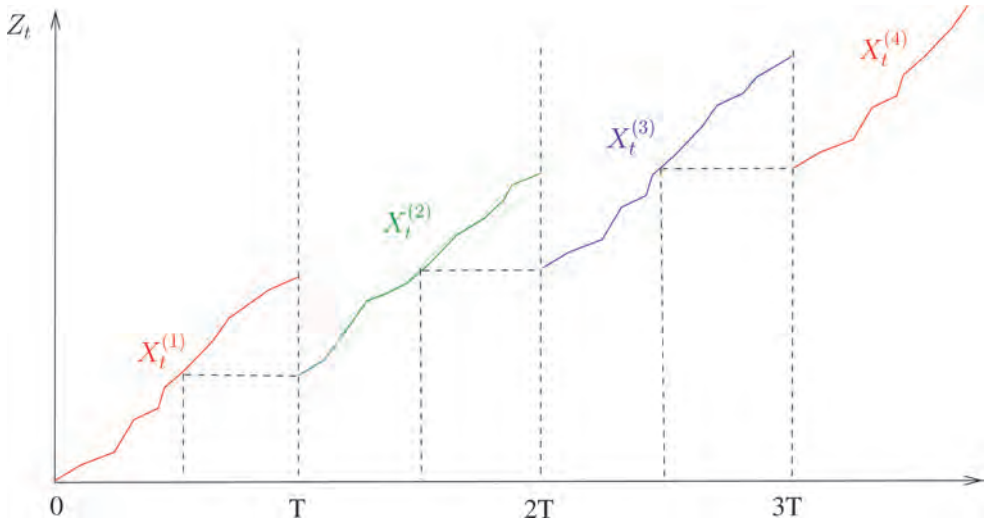


Figure 2. Illustration of the ARA1 policy for $\rho = 0.5$.

Hence

$$Z_{nT} = \sum_{i=1}^{n} \left(X^{(i)}_{(1-\rho)iT} - X^{(i)}_{(1-\rho)(i-1)T} \right)$$

and it is gamma distributed $\Gamma\left(A\left((1-\rho)nT\right), b\right)$. Here, Z_t is the sum of two independent gamma distributed r.v.s which share the same scale parameter b and it is gamma distributed $\Gamma\left(A\left(t-\rho nT\right), b\right)$. Also:

$$\mathbb{E}\left(Z_t\right) = \frac{A\left(t-\rho nT\right)}{b}, \quad var\left(Z_t\right) = \frac{A\left(t-\rho nT\right)}{b^2} \qquad (3)$$

for all $nT \le t < (n+1)T$.

Here again, it is easy to check that $\mathbb{E}\left(Z_t\right)$ and $var\left(Z_t\right)$ are decreasing with respect to ρ. Also, the cases $\rho \to 1^-$ and $\rho \to 0^+$ correspond to AGAN and ABAO repairs, respectively.

3 COMPARISON RESULTS

3.1 *Comparison of the moments*

We now come to the main object of the paper, which is the comparison between the two models of imperfect repairs. Note that, in an applicative context, there is no reason why the estimated repair efficiency should be the same when the impact of the maintenance is modeled by an ARD1 or ARA1 model. Considering two different efficiency parameters $\rho_i, i = 1,2$, our point here is to compare $Y_t^{(1)}$ and $Z_t^{(2)}$, where exponent $^{(i)}$ refers to ρ_i for i = 1, 2. We begin with the comparison of their respective means and variances.

Due to the reduced size of the paper, results on the comparison of moments are provided only in case of a power-law shape function for the gamma process. The interested reader will find general results in an extended version of the paper (Mercier and Castro, submitted), as well as proofs for all the results of the paper.

Proposition 1. *Let* $A(t) = \alpha t^{\beta}$ *with* α, $\beta > 0$. *Assume* $\beta \le 1$.

1. *We have*

$$\mathbb{E}\left(Z_t^{(2)}\right) \ge \mathbb{E}\left(Y_t^{(1)}\right), \forall t > 0$$

if and only if $(1-\rho_2)^{\beta} \ge (1-\rho_1)$.

2. *We have*

$$Var\left(Z_t^{(2)}\right) \ge Var\left(Y_t^{(1)}\right), \forall t > 0$$

if and only if $(1-\rho_2)^{\beta} \ge (1-\rho_1)^2$.

If $\beta \ge 1$, *all results are valid with reversed inequalities.*

For illustration purpose, we consider $A(t) = t^3$, $b = 1$, $T = 1$ and $\rho_2 = 0.5$. As a first case, we take $\rho_1 = 0.3$ so that the conditions of the previous proposition are fulfilled for both expectation and variance. As expected, Figure 3 shows that $\mathbb{E}\left(Y_t^{(1)}\right)$ and $\mathbb{E}\left(Z_t^{(2)}\right)$ are ordered in the same way on the whole real line, the same for the variance. As a second case, we take $\rho_1 = 0.95$ so that the conditions are *not* fulfilled neither for expectation and variance. Figure 4 shows that $\mathbb{E}\left(Y_t^{(1)}\right)$ and $\mathbb{E}\left(Z_t^{(2)}\right)$ are not

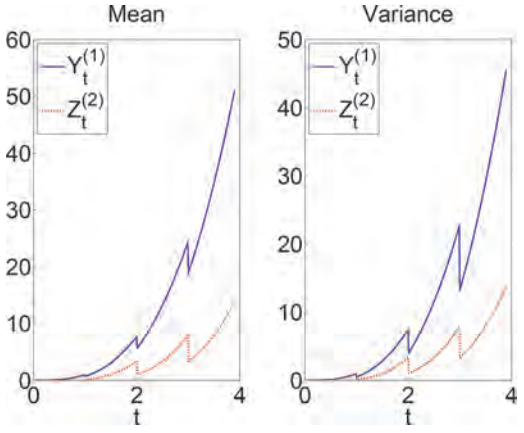


Figure 3. Mean and variance of $Y_t^{(1)}$ and $Z_t^{(2)}$ with respect to t for $A(t) = t^3$, $b = 1$, $\rho_1 = 0.95$, $\rho_2 = 0.5$, $T = 1$.

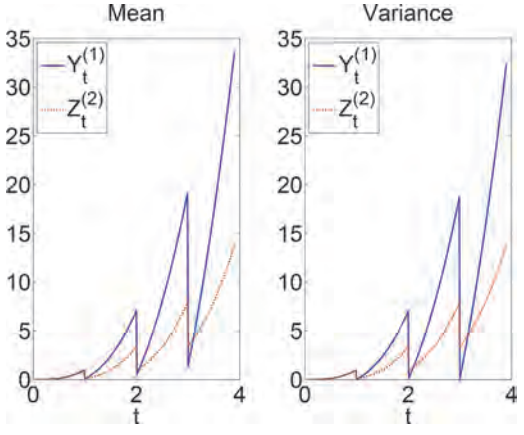


Figure 4. Mean and variance of $Y_t^{(1)}$ and $Z_t^{(2)}$ with respect to t for $A(t) = t^3$, $b = 1$, $\rho_1 = 0.95$, $\rho_2 = 0.5$, $T = 1$.

ordered in the same way on the whole real line, the same for the variance.

3.2 *Technical reminders*

Before providing stochastic comparison results between $Y_t^{(1)}$ and $Z_t^{(2)}$, we here recall a few definitions and results from the literature.

As a first step, let us recall that given two non negative random variables X and Y with probability distribution functions f_X and f_Y, and survival functions $\bar{F}_X$ and $\bar{F}_Y$, respectively, then:

1. X is said to be smaller than Y in the usual stochastic order $\left(X \prec_{sto} Y\right)$ if $\bar{F}_X \le \bar{F}_Y$;
2. X is said to be smaller than Y in the likelihood ratio order $\left(X \prec_{lr} Y\right)$ if $\frac{f_Y}{f_X}$ is non-decreasing;
3. X is said to be smaller than Y in the convex (concave) order $\left(X \prec_{cx(cv)} Y\right)$ if $\mathbb{E}\left(\varphi(X)\right) \le \mathbb{E}\left(\varphi(Y)\right)$ for all convex functions φ (provided the expectations exist);
4. X is said to be smaller than Y in the increasing convex (concave) order $\left(X \prec_{icx(icv)} Y\right)$ if $\mathbb{E}\left(\varphi(X)\right) \le \mathbb{E}\left(\varphi(Y)\right)$ for all increasing convex (concave) functions φ (provided the expectations exist).

The usual stochastic order and the likelihood ratio order compare the locations of random variables whereas increasing convex (concave) orders also compare their variability: $X \prec_{icx} \left(\prec_{icv}\right) Y$ roughly means that $\mathbb{E}(X) \le \mathbb{E}(Y)$ (location condition) plus the fact that X is less (more) "variable" than Y, in a stochastic sense. Also, $X \prec_{cx} \left(\prec_{cv}\right) Y$ is equivalent to $X \prec_{icx} \left(\prec_{icv}\right) Y$ plus $\mathbb{E}(X) = \mathbb{E}(Y)$.

The likelihood ratio order is known to imply the usual stochastic order, which itself implies both increasing convex and concave orders, see Müller & Stoyan 2002 or Shaked & Shanthikumar 2007 for more details.

We finally review known results on the comparison of gamma distributions in the following lemma, see (Müller & Stoyan 2002, p. 62) for instance.

Lemma 1. *Let X and Y be gamma distributed random variables with parameters (a_1, b_1) and (a_2, b_2), respectively, where $a_i, b_i > 0$ for $i = 1, 2$. Then:*

1. *If $a_1 \le a_2$ and $b_1 \ge b_2$, then $X \prec_{lr} Y$;*
2. *If $a_1 \ge a_2$ and $a_1 / b_1 \le a_2 / b_2$, then $X \prec_{icx} Y$;*
3. *If $a_1 \le a_2, b_1 \le b_2$ and $a_1 / b_1 \le a_2 / b_2$, then $X \prec_{icv} Y$.*

The previous lemma is the basis for deriving the stochastic comparison results between the different imperfect repair models given in the next subsection.

3.3 *Stochastic comparison results*

As a first step, the influence of the efficiency parameter ρ on the deterioration level is studied for each imperfect repair model.

Proposition 2. *We have:*

1. *Y_{nT} decreases with respect to ρ in the sense of the likelihood order: If $\rho_1 < \rho_2$, then $Y_{nT}^{(2)} \prec_{lr} Y_{nT}^{(1)}$,*
2. *Y_t decreases with respect to ρ in the sense of both increasing convex and concave orders: If $\rho_1 < \rho_2$, then $Y_t^{(2)} \prec_{icx} Y_t^{(1)}$ and $Y_t^{(2)} \prec_{icv} Y_t^{(1)}$;*
3. *Z_t decreases with respect to ρ for the likelihood ratio order (and hence also for both increasing convex and concave orders): If $\rho_1 < \rho_2$, then $Z_t^{(2)} \prec_{lr} Z_t^{(1)}$.*

In each case, we can see that as expected, the more efficient the maintenance action is (namely

the larger ρ is), the smaller the deterioration level is. Note however that, based on the fact that the likelihood ratio order is stronger than both increasing convex and concave orders, the results for the ARA1 model are stronger than for the ARD1 one. (Counter-examples can be found, which show that that $Y_t^{(1)}$ and $Y_t^{(2)}$ are not comparable for the likelihood ratio order in a general setting).

We now come to the stochastic comparison between the two models.

Theorem 1. *If $A(\cdot)$ is concave and*

$$A\left((1-\rho_2)t\right) \geq (1-\rho_1)A(t) \text{ for all } t \tag{4}$$

(which is true if $\rho_1 \geq \rho_2$), then $Y_t^{(1)} \prec_{icx} Z_t^{(2)}$ for all $t \geq 0$.

If $A(t)$ is convex with a reversed inequality in (4), *then $Z_t^{(2)} \prec_{icv} Y_t^{(1)}$ for all $t \geq 0$.*

As a specific case, if $\rho_1 = \rho_2 = \rho$, then $Y_t \prec_{icx} Z_t$ for all $t \geq 0$ if $A(\cdot)$ is concave, and $Z_t \prec_{icv} Y_t$ for all $t \geq 0$ if $A(\cdot)$ is convex. The concavity/convexity of the shape function $A(\cdot)$ of the underlying gamma process hence deeply infers on the comparison results between the two models: When $A(\cdot)$ is convex, the rejuvenation included in a ARA1 model leads to a lower deterioration level than for an ARD1 model.

Specific results are now summarized in the classical case of a homogeneous gamma process for both moments and stochastic comparisons.

Corollary 1. *Assume that $A(t) = \alpha t$ for all $t \geq 0$, where $\alpha > 0$. We have the following results:*

1. *If $\rho_1 \geq \rho_2$, then $Y_t^{(1)} \prec_{icx} Z_t^{(2)}$ and hence $\mathbb{E}\left(Y_t^{(1)}\right) \leq \mathbb{E}\left(Z_t^{(2)}\right)$ for all $t \geq 0$;*
2. *If $\rho_1 \leq \rho_2$, then $Z_t^{(2)} \prec_{icv} Y_t^{(1)}$ and hence $\mathbb{E}\left(Z_t^{(2)}\right) \leq \mathbb{E}\left(Y_t^{(1)}\right)$ for all $t \geq 0$;*
3. *If $\rho_1 = \rho_2$, then $Y_t^{(1)} \prec_{ex} Z_t^{(2)}$ and $Z_t^{(2)} \prec_{cv} Y_t^{(1)}$ and hence $\mathbb{E}\left(Z_t^{(2)}\right) = \mathbb{E}\left(Y_t^{(1)}\right)$ for all $t \geq 0$;*
4. *$Var\left(Z_t^{(2)}\right) \geq Var\left(Y_t^{(1)}\right)$ for all $t > 0$ if and only if $1 - \rho_2 \geq (1-\rho_1)^2$.*

3.4 *Mostly equivalent imperfect repair models*

In an applied context, parameters ρ_1 and ρ_2 for ARD1 and ARA1 models, respectively, will be estimated from feedback data, which will be typically gathered at maintenance times iT, $i \geq 1$. As a consequence, we can expect that the estimated parameters $\hat{\rho}_1$ and $\hat{\rho}_2$ should be such that the corresponding expected deterioration levels should be very similar at maintenance times, namely such that $\mathbb{E}\left(Y_{iT}^{(1)}\right) \simeq \mathbb{E}\left(Z_{iT}^{(2)}\right)$ for all $i \in \mathbb{N}^*$. Equivalently, the estimated parameters should be such that

$$\frac{(1-\hat{\rho}_1)A(iT)}{b} \simeq \frac{A\left((1-\hat{\rho}_2)iT\right)}{b}, \text{ for all } i \geq 1.$$

There hence is a specific interest for the applications to compare the ARD1 and ARA1 models under the condition

$$(1-\rho_1)A(iT) = A\left((1-\rho_2)iT\right) \text{ for all } i \geq 1, \tag{5}$$

which will lead to equivalent deterioration levels at maintenance times. However, the previous requirement (5) does not seem to have a solution for a general shape function $A(\cdot)$. We hence restrict the study to the power-law case $A(t) = \alpha t^{\beta}$ (with α, $\beta > 0$), for which (5) is just equivalent to

$$1 - \rho_1 = (1-\rho_2)^{\beta}.$$

A homogeneous gamma process corresponds to a power-law shape function with $\beta = 1$. Then the equivalent case just means that the two models share the same efficiency parameters $(\rho_1 = \rho_2)$.

In case of a general power-law shape function, note that the "equivalent" case has a similar spirit to that detailed in Doyen & Gaudoin (2004, Property 4), where the authors match the minimal wear intensities of two imperfect repair models for recurrent events, based on the reduction of either virtual age or failure intensity.

The results for the equivalent case are summarized in the following proposition.

Proposition 3. *Assume that $A(t) = \alpha t^{\beta}$ (with $\alpha, \beta > 0$) and that $1 - \rho_1 = (1-\rho_2)^{\beta}$. Then:*

1. *$Z_{nT}^{(2)} \prec_{cv} Y_{nT}^{(1)}$ and $Y_{nT}^{(1)} \prec_{cx} Z_{nT}^{(2)}$ (which both entail that $var\left(Z_{nT}^{(2)}\right) \geq var\left(Y_{nT}^{(1)}\right)$) for all $n \geq 1$.*
2. *If $\beta \leq (\geq) 1$, then $Y_t^{(1)} \prec_{icx} (\succ_{icv}) Z_t^{(2)}$ (which entails that $\mathbb{E}\left(Y_t^{(1)}\right) \leq (\geq) \mathbb{E}\left(Z_t^{(2)}\right)$) for all $t \geq 0$.*
3. *If $\beta \leq 1$, then $Var\left(Z_t^{(2)}\right) \geq Var\left(Y_t^{(1)}\right)$ for all $t \geq 0$.*

4 APPLICATION

For illustration purpose of the previous results, the system is now supposed to provide some reward per unit time, which decreases when the deterioration level of the system increases. Based on classical functions used in the insurance literature (Rolski et al. 1998), we assume that the reward function is of the shape

$$g(x) = (b_1 - k_1 e^{\alpha_1 x}) 1_{\{0 \leq x \leq c\}} + (b_2 - k_2 e^{\alpha_2 x}) 1_{\{c < x\}}, \tag{6}$$

with $b_1, b_2, \alpha_1, \alpha_2, k_1, k_2, c > 0$. The reward function is supposed to be continuous and positive on $(0, c)$, which implies that

$$b_2 - k_2 e^{\alpha_2 c} = b_1 - k_1 e^{\alpha_1 c} > 0. \tag{7}$$

Also, we assume that $\alpha_1 \le \alpha_2$ and $k_1 \le k_2$ so that level c appears as a critical level, from which the system becomes less performing.

With the previous assumptions, it is easy to check that g is a concave function and that $g(x) > 0$ if and only if $x < L = \frac{\ln(b_2/k_2)}{\alpha_2}$. Level L hence appears as a critical threshold, from where the unitary reward becomes negative.

An example of reward function is plotted in Figure 5 with parameters $\alpha_1 = 0.1, b_1 = 11$ monetary units (m.u.), $\alpha_2 = 0.25, k_1 = 1$ (m.u.), $k_2 = 1$ (m.u.), $c = 4$, and b_2 obtained through (7). With this dataset $L \simeq 10.01$.

The system is assumed to be preventively repaired each T units of time according to an ARD1 or ARA1 model, with efficiency parameters ρ_1 and ρ_2, respectively. The accumulated reward on a time interval [0, t] hence is

$$R^{(1)}_{ARD1}(t) = \mathbb{E}\left(\int_0^t g\left(Y_s^{(1)}\right) ds\right) = \int_0^t \mathbb{E}\left(g\left(Y_s^{(1)}\right)\right) ds$$

with a similar expression for the accumulated reward $R^{(2)}_{ARA1}(t)$ in the ARA1 case.

Considering for instance the equivalent case ($A(t) = \alpha t^{\beta}$ and $1 - \rho_1 = (1 - \rho_2)^{\beta}$) with $\beta \le 1$, we easily derive from point 2 in Proposition 3 that

$$\mathbb{E}\left(g\left(Y_s^{(1)}\right)\right) \ge \mathbb{E}\left(g\left(Z_s^{(2)}\right)\right)$$

for all $s \in [0,t]$, using the fact that $-g$ is an increasing convex function. This immediately entails that

$$R^{(1)}_{ARD1}(t) \ge R^{(2)}_{ARA1}(t) \text{ for all } t \ge 0.$$

As a consequence, even if the "equivalent" case provides similar expected levels at maintenance times, they do not provide the same accumulated rewards on some given time interval. They may hence lead to different decisions in an applied context, for instance for optimizing maintenance strategies.

To better illustrate this difference, we now consider a preventive maintenance strategy, which we next optimize. The assumptions are the following:

- The unitary reward of the system per unit time is given by the reward function g provided in (6). We recall that $L = \frac{\ln(b_2/k_2)}{\alpha_2}$ stands for a "critical" level, from where the reward becomes negative.
- The system is preventively repaired each T units of time according to an ARD1 or ARA1 model (respective efficiency parameters: ρ_1 and ρ_2), up to the first maintenance time KT where the deterioration level is observed to be beyond a given preventive threshold M (with $M < L$). In that case, instead of an imperfect repair (ARA1 or ARD1), a replacement is performed at time KT with cost C_c m.u. when the level is beyond L (corrective replacement) and C_p m.u. when it is between M and L (preventive replacement).

The successive (corrective or preventive) replacements of the system appear as the points of a renewal process, and the long time (operating) profit rate per unit time is given by

$$C_{ARD}(T,M) = \frac{1}{\mathbb{E}(K)T}\left[\mathbb{E}\left(\int_0^{KT} g\left(Y_s^{(1)}\right) ds\right) - C_r\left(\mathbb{E}(K) - 1\right) - C_p\ \mathbb{P}\left(M \le Y_{KT}^{(1)} < L\right) - C_c\ \mathbb{P}\left(L \le Y_{KT}^{(1)}\right)\right]$$

for the ARD1 model with a similar expression for the ARA1 model ($C_{ARA}(T, M)$). Due to the complexity of the model, there is not hope here to find conditions that ensure the dominance of one of the two functions $C_{ARD}(T, M)$ or $C_{ARA}(T, M)$ on the other. The comparison is hence made on a numerical example.

The degradation is modeled by a homogeneous gamma process with parameters $A(t) = 1.3t$ and $b = 0.8$. The parameters of the reward function g are $\alpha_1 = 0.4$, $\alpha_2 = 0.5$, $b_1 = 800$ m.u., $k_1 = 1.05$ m.u., $k_2 = 1.07$ m.u., $c = 8$, which implies $b_2 \simeq 832.66$ m.u. and $L \simeq 13.31$. The repair efficiencies of the ARD1/ARA1 repairs are $\rho_1 = \rho_2 = 0.9$, which corresponds to an equivalent case (homogeneous gamma process and $\rho_1 = \rho_2$). Their common cost is $C_r = 200$ m.u.. The cost of a preventive replacement is $C_p = 1000$ m.u. whereas itis $C_c = 1300$ m.u. for a corrective one. Figures 6a and 6b show the profit rates for the maintained system under ARD1 and ARA1 repairs, respectively. These figures have been computed considering a grid of 10 points for T from 1.14 to 4 and a grid of 13 points for M from 1

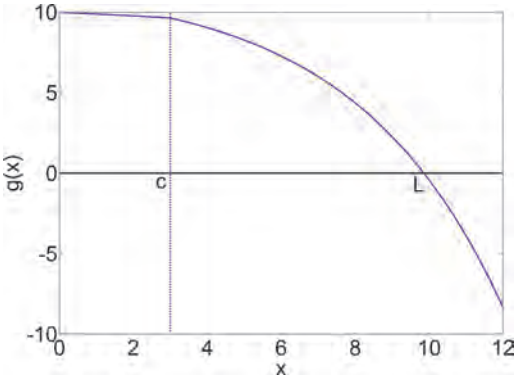
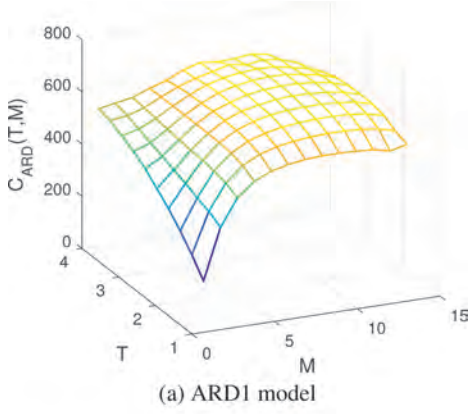
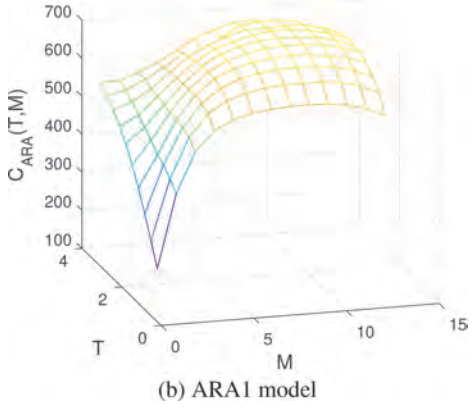


Figure 5. The reward function g for $\alpha_1 = 0.1$, $b_1 = 11$ m.u., $\alpha_2 = 0.25$, $k_1 = 1$ m.u., $k_2 = 1$ m.u., $c = 4$.



(a) ARD1 model



(b) ARA1 model

Figure 6. Operational profit rates, $\rho_1 = \rho_2 = 0.9$.

to L and 10000 simulations for each pair of points. For both ARD1 and ARA1 cases, the optimal maintenance strategy corresponds to $T^{opt} \simeq 3.05$. However, the corresponding optimal maintenance levels are $M^{opt}_{ARD} \simeq 9.21$ and $M^{opt}_{ARA} \simeq 10.24$, and the optimal unitary rewards $C^{opt}_{ARD} \simeq 673.94$ and $C^{opt}_{ARA} \simeq 684.34$ m.u. per unit time, respectively. There hence is a difference of about 10% between the two optimal preventive levels, which may entail inappropriate decision in an applicative context, when considering one imperfect repair model or the other, whereas the effective behavior of the maintained system is closer to the other one.

5 CONCLUSIONS AND FURTHER EXTENSIONS

Two imperfect repair models for a degrading system have been proposed and compared in this paper. One is based on the reduction of the single deterioration level at maintenance times (ARD1), the other one on the reduction of both deterioration level and age (ARA1). They hence correspond to different maintenance actions, in an applicative context. The comparison of both location and spread have been made in terms of moments and stochastic ordering of the two resulting processes. It has been seen that the concavity/convexity of the shape function of the underlying gamma wear process plays a central role in the comparison results. This corresponds to intuition as a convex shape function (for instance) induces some increasingness property in the rate of deterioration over time. Hence, rejuvenating the system as in an ARA1 one action will decrease the rate of deterioration together with the deterioration level.

As for the stochastic comparison results, the paper focuses on the likelihood ratio order, which is well-known in reliability theory, but also on the (increasing/decreasing) convex/concave orders, which come from the insurance literature and seem a littleless common in papers devoted to reliability theory. Clearly, other stochastic orders might be considered such as Laplace transform order or Excess Wealth order for instance. Other questions of interest concern the comparison of remaining lifetimes, considering the system as failed (or too degraded) when its deterioration level is beyond a fixed failure (critical) threshold. From a theoretical point of view, this seems a difficult issue in a general setting. One could then look at partial results on specific models.

Furthermore, the paper made an attempt for comparing the two types of imperfect repair including them in a global maintenance policy. Clearly, this subject requires further investigation for a better understanding of the practical consequences on the optimal policy of choosing one model of imperfect repair or the other, and typically, this would require a numerical study at a larger scale. Of course, it would also be of interest to consider other types of global maintenance policy (among the numerous ones developed in the literature, e.g. see van Noortwijk (2009)), including one model of imperfect repair or the other.

Finally, a gamma deterioration model has been assumed in this paper. The stochastic comparison results between the two imperfect repair models deeply rely on the comparison properties of the gamma distribution, as summed up in Lemma 1. As noted by the referee, a question of interest would be to revisit the comparison between the two imperfect repair models considering other deterioration models (such as Wiener processes with trend or other subordinators).

REFERENCES

Alaswad, S. & Y. Xiang (2017). A review on condition-based maintenance optimization models for stochastically

deteriorating system. *Reliability Engineering & System Safety 15*, 54–63.

Caballé, N.C., I.T. Castro, C.J. Pérez, & J.M. Lanza-Gutiérrez (2015). A condition-based maintenance of a dependent degradation-threshold-shock model in a system with multiple degradation processes. *Reliability Engineering & System Safety 134*, 98–109.

Castro, I.T. & S. Mercier (2016). Performance measures for a deteriorating system subject to imperfect maintenance and delayed repairs. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 1–22.

Doyen, L. & O. Gaudoin (2004). Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering & System Safety 84(1)*, 45–56.

Hong, H.P., W. Zhou, S. Zhang, & W. Ye (2014). Optimal condition-based maintenance decisions for systems with dependent stochastic degradation of components. *Reliability Engineering & System Safety 121*, 276–288.

Huynh, K.T., I.T. Castro, A. Barros, & C. Bérenguer (2014). On the use of mean residual life as a condition index for condition-based maintenance decision-making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems 44(7)*, 877–893.

Kijima, M. (1989). Some results for repairable systems with general repair. *Journal of Applied Probability 26*(1), 89–102.

Mercier, S. & I.T. Castro (2013). On the modelling of imperfect repairs for a continuously monitored gamma wear process through age reduction. *Journal of Applied Probability 50*(4), 1057–1076.

Mercier, S. & I.T. Castro (submitted). Stochastic comparisons of imperfect maintenance models for a gamma deteriorating system.

Müller, A. & D. Stoyan (2002). *Comparison methods for stochastic models and risks*. John Wiley & Sons.

Rolski, J., H. Schmidli, V. Schmidt, & J. Teugels (1998). *Stochastic Processes for Insurance and Finance*. John Wiley & Sons.

Shaked, M. & J.G. Shanthikumar (2007). *Stochastic Orders*. Springer.

van Noortwijk, J. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety 94(1)*, 2–21.

Zhang, M., G. Olivier, & X. Min (2015). Degradation-based maintenance decision using stochastic filtering for systems under imperfect maintenance. *European Journal of Operational Research 245*, 531–541.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A predictive approach to jointly schedule missions and maintenances for a deteriorating vehicle

E. Robert
University of Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, Grenoble, France
Volvo Group Trucks Technology, Saint-Priest, France

C. Bérenguer
University of Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, Grenoble, France

K. Bouvard & H. Tedie
Volvo Group Trucks Technology, Saint-Priest, France

R. Lesobre
Renault Trucks Defense, Versailles, France

ABSTRACT: Both for the manufacturer and the user of industrial vehicles, optimizing simultaneously the maintenance and missions schedule at the fleet level becomes a necessity to improve the profitability. Lots of researches have been realized to optimize either the preventive maintenance schedule or the production planning. Integrating both activities in the same schedule has become a new hotspot. Few researches have been led to schedule both maintenance operations and missions for a fleet of systems which deteriorate over time in a comprehensive predictive approach. As a first step to reach this objective, a single vehicle is considered and we propose a method to jointly optimize its predictive maintenance and its mission planning using the remaining useful life and deterioration information. The vehicle has a set of missions to complete. The aim is to group the missions in blocks and interpose maintenance operations between these blocks. However, the vehicle deteriorates over time and each mission impacts differently the deterioration according to its severity. A stochastic process is used to model the deterioration phenomenon and its parameters are changed for each mission. To obtain the best arrangement between missions and maintenance operations, a genetic algorithm is used. A criterion to minimize the maintenance cost is defined, which takes into account preventive and corrective maintenance costs associated with the missions failure probabilities. The genetic algorithm enables to approach the optimal solution in a reasonable time and avoid considering every possible arrangement.

1 INTRODUCTION

Many researchers have developed methods to optimize the preventive maintenance schedule for multi-component systems (Bouvard et al. 2011) without considering production constraints but taking into account the system availability on a fixed time period (Lesobre 2015). From these previous achievements, a new problematic has arisen aiming at jointly scheduling production and maintenance. Two strategies to solve this problem can be identified.

The sequential strategy consists in scheduling one of the two activities, maintenance or production, and using this schedule as an unavailability additional constraint to solve the joint scheduling. Benbouzid et al. (2003) develop a sequential process to schedule production activities and maintenance in flowshop workshops. The production schedule is generated using heuristic methods. Then, by considering the production tasks order as a constraint, the periodic maintenance operations are added to the initial schedule. This period is determined to reach a trade-off between the maintenance cost and the risk of reducing the machine availability. The same procedure is applied to integrate preventive maintenance in the production scheduling for a single machine to minimize the total expected weighted completion time to do all the jobs (Cassady & Kutanoglu 2005). The exact method, for small size problems, considers all the job sequence possibilities. Then, for each possibility, all the feasible Preventive Maintenance (PM) decisions sets are tested. The job sequence-PM decisions minimizing the objective function is the optimal solution. For larger problems, a two-steps

heuristic method is developed. The first step selects the job sequence having the shortest total weighted completion time. The weigh is an importance factor for each job. The second step identifies the PM decisions minimizing the total expected weighted completion time for the job-sequence established in the first step.

The integrated strategy approach aims at scheduling simultaneously maintenance and production activities. Yalaoui et al. (2014) use a linear programming model to minimize the total production cost. The production is divided in cycles and a preventive maintenance is completed at each cycle beginning to restore the production lines capacities. Each maintenance cycle cost depends on the preventive and corrective maintenance cost. The corrective one is estimated thanks to the failure rate. This exact method is satisfactory for reduced size problem. The following research works apply heuristic methods to solve the integrated scheduling problem. Feng et al. (2016) apply a genetic algorithm to minimize the costs induced by jobs tardiness, corrective and preventive maintenance action. Ladj et al. (2016) suggest an approach hybridizing genetic algorithms and artificial immune systems. A deterministic deterioration model is considered as each job deteriorates the machine of a fixed value and it is assumed that no accidental failure occurs during the time horizon. The maintenance actions are triggered when the deterioration level oversteps a failure threshold. These approaches are single-objective optimization methods but multi-objective methods exist. Da et al. (2016) choose to minimize both the maintenance cost rate and makespan.

A comparison between the sequential approach and a new integrated strategy to minimize the manufacturing system cost for a single-unit system has been drawn by Li et al. (2010). The maintenance operations are imperfect and an improvement factor is defined to quantify each operation effectiveness. The integrated approach uses a heuristic method to find the optimal schedule and enables to save about 12% of the manufacturing costs with respect to the sequential approach. The terms sequential or integrated only refer to the way of solving the scheduling problem. In both cases, the result is a single simultaneous schedule for production and maintenance.

In the previous contributions, the deterioration modelling considers either the machine age or a deterministic model where each job deterioration is exactly known. Li et al. (2010) do not consider the failure risk or any reliability based constraints to obtain the joint schedule.

The objective of this paper is to develop a new integrated strategy to schedule both missions and maintenance for a deteriorating vehicle using deterioration information. Firstly, the joint scheduling problem is defined. Then, the adopted approach to solve it is explained. The models for the vehicle deterioration evolution and for the missions impact on the deterioration are developed and the optimization method is exposed. Finally, performances and sensitivity studies are presented on application examples to illustrate the algorithm behaviour.

2 PROBLEM FORMULATION

A single vehicle has a set of missions to complete. It is modelled as a single-unit system which deteriorates over time according to its activity. The vehicle operates in missions with different severity levels, characterized not only by their duration but also by several environment parameters as road condition, topography. The vehicle usage modelling is then adapted and takes into account the mission severity level.

The objective of the present work is to jointly optimize the vehicle predictive maintenance and its mission planning using the deterioration information (Figure 1). The preventive maintenance operations have to be well planned to prevent immobilizing failures, maximize the vehicle availability and not disturb the missions progression. According to these conditions, the schedule is defined as a

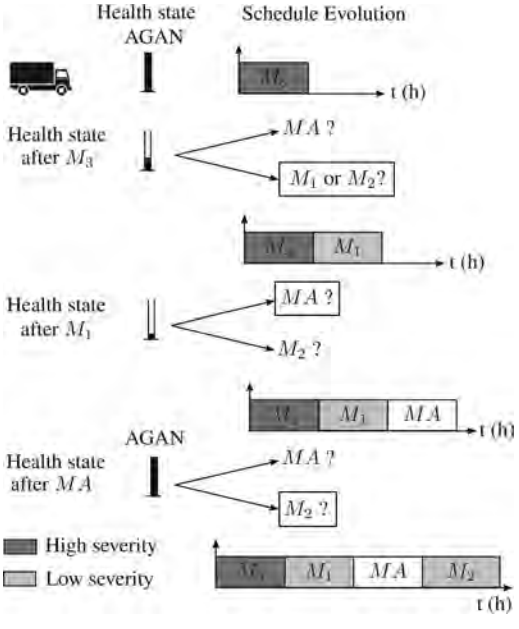


Figure 1. Framework to schedule the three missions M_1, M_2 and M_3 and the maintenance actions MA according to the health state. Their durations and deterioration severities are respectively defined by the boxes lengths and colors.

series of mission blocks interrupted by maintenance operations to restore the vehicle deterioration state to an As Good As New (AGAN) state.

The schedule optimization is based on the total maintenance cost and aims at maximizing the vehicle activity between two consecutive preventive maintenance operations while considering its health state. The total maintenance cost is then based on the preventive maintenance costs C_{pt}, for which each operation costs C_0, and the costs C_{ct} associated with corrective maintenance. To estimate C_{ct}, the missions failure risks and the corrective maintenance cost C_f for each operation are considered.

3 RESOLUTION APPROACH

3.1 *Vehicle deterioration model*

3.1.1 *Deterioration evolution*

The vehicle is characterized by a global health indicator. Its deterioration evolution is modelled by a stationary Gamma process. The choice of this stochastic continuous deterioration process is motivated by its ability to represent observable and gradual deterioration phenomena in industrial systems (Lesobre 2015, Van Noortwijk 2009). A Gamma process $X(t)$, $t > 0$ is defined by its shape and scale parameters respectively denoted α and β. A failure occurs when the cumulated deterioration $X(t)$ exceeds the failure threshold L. The distribution function F followed by the time to failure T is then given by Eq. 1. The failure probability for a mission whose duration is equal to t_m is equal to $F(t_m)$.

$$F(t) = P(T \leq t) = \frac{\Gamma(\alpha t, L\beta)}{\Gamma(\alpha t)}, \quad (1)$$

where $\Gamma(.)$ is the Gamma function and $\Gamma(\alpha,L\beta)$ is the incomplete Gamma function defined by $\Gamma(\alpha t, L\beta) = \int_{L\beta}^{\infty} u^{\alpha t-1} e^{-u} du$.

3.1.2 *Missions impact on the deterioration*

The missions correspond to the deliveries the truck has to complete. They are described by their durations and they affect the vehicle deterioration evolution. Indeed, the vehicle evolves in a dynamic environment influencing its deterioration. The environment evolution results from changes in the missions characteristics during the vehicle lifetime. The deterioration-threshold failure model used in this work allows one to integrate the mission characteristics through the modification of the deterioration parameters. It is thus assumed that these changes have a time-related impact on the deterioration process, modelled by a change in the deterioration speed. The deterioration is then modelled by a Gamma process with varying parameters. Each mission is associated with a pair of parameters.

Thanks to the deterioration model, the failure probabilities can be estimated based on the distribution function followed by the time to failure. An optimization method to jointly schedule mission and maintenance can then be developed.

3.2 *Optimization method for joint scheduling*

A genetic algorithm based approach is developed to find the optimal joint schedule for maintenance and missions minimizing the total maintenance cost. The optimization criterion definition is one of the major point to discuss.

3.2.1 *Criteria definition*

The schedule is composed of mission blocks separated by preventive maintenance operations. The optimization criterion is then based on the maintenance cost associated with the schedule. This criterion relies on two elements:

- The preventive maintenance cost C_0 corresponding to the maintenance operations scheduled at each block end.
- The corrective maintenance cost C_f related to the failure occurring within each mission block.

An optimal balance is to be found between preventive and corrective maintenance i.e. the number of blocks and the blocks filling. To determine the possible mission number in the blocks, the deterioration process is used to estimate the failure probabilityfor a block. This failure probability considers the parameters characterizing the missions in the block.

For p missions in a block, p environment changes, representing the missions severity levels, occur. Figure 2 presents an example with 3 missions and 3 environment changes. The only known information are that the deterioration level is equal to 0 at the block beginning. No information on

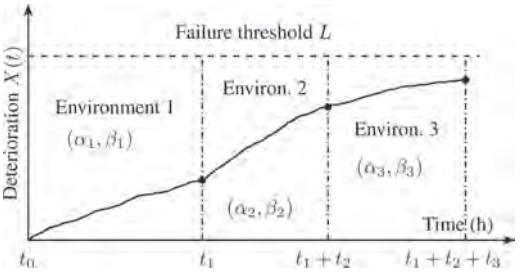


Figure 2. Deterioration evolution in a block with 3 missions.

its deterioration evolution are available when the vehicle is completing the schedule.

The cumulated deterioration between t_0 and $t_1+t_2+t_3$ corresponds to the deterioration increments sum between t_0 and t_1, t_1 and $t_1 + t_2$ and between $t_1 + t_2$ and $t_1 + t_2 + t_3$. The increments probability densities of the missions are respectively denoted f_1, f_2 and f_3. The increments deterioration density for the block is the increments densities convolution for the missions in the block. The more numerous the environment changes are, the more complex the reliability expression is to compute (Khoury 2012). Based on the Gamma process properties and the environment changes, an approximation for the block deterioration process can be given by an equivalent Gamma process $Ga(\alpha_e, \beta_e)$. Its average value and variance are respectively weighted mean values of the average values and variances related to the Gamma processes of the missions. The different weights are the proportions of the block duration spent in each environment. For Figure 2, the equivalent Gamma process parameters are defined as follows:

$$\frac{\alpha_e}{\beta_e}=\frac{1}{T}\sum_{i=1}^{p} t_i \frac{\alpha_i}{\beta_i} \quad \text{and} \quad \frac{\alpha_e}{\beta_e^2}=\frac{1}{T}\sum_{i=1}^{p} t_i \frac{\alpha_i}{\beta_i^2}, \tag{2}$$

where $p = 3$ and $T=\sum_{j=1}^{p} t_j$. Note that if all the Gamma processes related to the missions have the same scale parameter β and the same duration, the resulting convolution density is exactly the one for a Gamma process $Ga\left(\sum_{i=1}^{p}\alpha_i,\beta\right)$.

The block failure probability is obtained based on the equivalent Gamma process. It is the probability that the cumulated deterioration exceeds the failure threshold L, knowing that $X(t_0)$, the deterioration at t_0, is equal to 0.

$$P(X(T)-X(t_0)\geq L)=\frac{\Gamma(\alpha_e(T-t_0),L\beta_e)}{\Gamma(\alpha_e(T-t_0))} \tag{3}$$

The criterion C_1 (Eq. 4) is defined to estimate the maintenance costs for a joint schedule π composed of N_b blocks. It considers only one failure by block.

$$C_1(\pi)=\sum_{k=1}^{N_b}(C_0+C_f\mathbb{P}_f(k)), \tag{4}$$

where $\mathbb{P}_f(k)$ is the probability to have one failure in the block k as explained in Eq. 3. Variations can then be defined to accept multi failures in the blocks. We define another criterion C_2 which considers that in a block composed of m missions, no more than m failures can occur. Only one failure by mission could happen. Considering multi failure in a block corresponds to estimate the expected failure occurrences in a block.

The replacement process is such that once the deterioration exceeds the failure threshold L, a corrective maintenance is completed and the deterioration level is back to 0. Based on the equivalent Gamma process, the deterioration evolution in the block can be characterized. The probability to have two failures in the block defined in Figure 2 can then be estimated by $P(X(T) - X(t_0) \geq 2L)$. The principle is the same as the replacement process but without the deterioration level reset after a failure. The criterion C_2 for a joint schedule π composed of N_b blocks is:

$$C_2(\pi)=\sum_{k=1}^{N_b}\left(C_0+C_f\sum_{i=1}^{N_m(k)}\mathbb{P}_{fk}(D\geq iL)\right), \tag{5}$$

where $N_m(k)$ is the number of missions in the block k, D the deterioration level and $\mathbb{P}_{f_k}(D\geq iL)$ the probability to exceed the failure threshold L for the i^{th} time in the block k.

The two criteria C_1, C_2 are used as optimization criteria for the genetic algorithm developed to solve the joint scheduling problem.

3.2.2 *A genetic algorithm based method*

A genetic algorithm (GA) is developed to solve the joint scheduling problem for missions and maintenance using a maintenance cost based criterion defined in the section 3.2.1. The choice of this heuristic method is justified by its adaptability regarding the individuals and operators definitions as well as its performances to reach a good solution in a satisfying computation time. The genetic algorithm principle is based on the hybrid genetic-immune algorithm proposed by Ladj et al. (2016). The operators are adapted to be applied in a random deterioration case as Ladj et al. (2016) develop it for a deterministic deterioration case. Its general main principle is illustrated in Figure 3. The different stages are explained in the following paragraphs.

A parameter is defined to condition the block filling. As the deterioration evolution is stochastic, the exact deterioration level reached at a mission block end cannot be known. However, a failure probability can be estimated according to the missions in the block. This parameter acts on the maximum admissible failure probability for a block. It is more likely to dispatch a vehicle on a mission block whose failure probability is not too high to avoid at best corrective maintenance. This parameter is denoted $\mathbb{P}_{f_{max}}$ and its values is between 0 and 1. The closer to 1 it is, the more flexible the genetic algorithm is to generate individuals. Acting on this parameter guides the genetic algorithm towards the best joint schedule.

Individual representation: The individuals correspond to candidate schedules for both maintenance

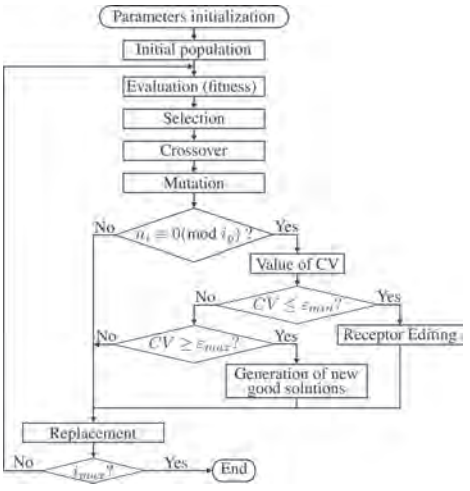


Figure 3. Genetic algorithm principle.

and missions. In the GA, the genotype is obtained by sequencing the mission set into different blocks. In a block, the missions are supposed to be completed one after the other. The preventive maintenance operations occurs at each block end. For instance, in case of a problem with $n = 6$ missions to schedule, a possible candidate schedule is $\pi = \{(6,2)(5,3,4)(1)\}$.

Initial population: The initial population is composed of N_{pop} individuals. 60% of them are randomly generated. Each mission is randomly put in a block while respecting the block filling condition. For the remaining individuals generation, special techniques are applied. 20% of the population is generated using the First Fit (FF) method (Coffman et al. (1996)). This method is applied on random missions sequences to form candidate solutions by assigning the missions to the blocks. The last 20% is generated using to heuristic methods called First Fit Decreasing (FFD) and Best Fit Decreasing (BFD) (Coffman et al. (1996)). Then, block permutations are performed on these solutions to generate other individuals. Note that the individuals generated by these techniques still respect the block filling condition.

Evaluation: The evaluation stage consists in applying the fitness function *Fit* to evaluate the quality of an individual in the population. The fitness function is related to the maintenance cost based criteria defined in section 3.2.1. Two fitness functions for a individual π are defined in Eq. 6. The best individuals are the ones for which the maintenance costs are the lowest.

$$Fit_1(\pi) = \frac{1}{C_1(\pi)} \quad \text{and} \quad Fit_2(\pi) = \frac{1}{C_2(\pi)} \tag{6}$$

Selection: The selection operator is a 2-tournament selection operator (Michalewicz 1996). It randomly chooses two individuals in the current population and selects the fittest one. When the selected individuals number reaches N_{pop}, the selection stage is over.

Crossover: The crossover operator crosses two parent individuals to obtain new offspring individuals. A crossover probability P_{cross} defines whether or not the parents will procreate. The principle is to randomly select a pair of parents. Both parent blocks are listed and a random selection is done to copy non-overlapping blocks to both offspring individuals so that no missions are duplicated. Then, for each offspring, the remaining missions are randomly added to blocks, either an existing one while respecting the block filling condition or a new one. This principle is inspired from the crossover operator developed by Rohlfshagen and Bullinaria (2010).

Mutation: The mutation operator is based on Swap mutation (Michalewicz 1996) and consists in exchanging two randomly selected missions from two different blocks while respecting the block filling condition. This swapping occurs with a probability P_{mut}.

Population dispersion: This stage periodically evaluates the total population dispersion which is of great importance to avoid converging towards a local optimal solution. The total population includes the parents and their children with their possible mutations and contains $2N_{pop}$ individuals. As explained by Ladj et al. (2016), diversification and intensification are two major issues when building effective search algorithms. Diversification refers to the capacity to explore different regions of the search space while intensification refers to the ability to generate high fitted solutions in those regions. A balance between these two notions has to be obtained. When the iteration number n_i is a multiple of the iteration period i_p, the population dispersion is estimated through the Coefficient of Variation (CV), also known as the Relative Standard Deviation (RSD) (Ladj et al. 2016). CV considers the fitness value of each schedule π_i in the total population *Pop* (Eq. 9). According to its value, different decisions are adopted to enhance either the diversification or the intensification.

$$CV = \frac{\sigma}{\mu} \cdot 100\% \tag{7}$$

$$\text{with} \quad \mu = \frac{1}{2N_{pop}} \sum_{\pi_i \in Pop} \frac{1}{Fit(\pi_i)} \tag{8}$$

$$\text{and} \quad \sigma = \sqrt{\frac{1}{2N_{pop}} \sum_{\pi_i \in Pop} \frac{1}{(Fit(\pi_i) - \mu)^2}} \tag{9}$$

Receptor editing: When the population dispersion CV is lower than a minimum dispersion threshold ε_{min}, individuals are very similar and focused on a limited search space region. The receptor editing operator aims at reducing the risk of premature convergence (Ladj et al. 2016). A part of the least fitted individuals (α% of the population) are eliminated and replaced by random new individuals to possibly explore new search space regions.

Generation of new good solutions: When the population dispersion CV is higher than a maximum dispersion threshold ε_{max}, individuals cover distinct search space regions. To promote the most promising regions, a part of the least fitted individuals are replaced by mutations of the most fitted individuals (α% of the population).

Replacement: Individuals for the next generation are selected among the total population formed by the parents and the children. A part of the least fitted individuals (β%) is directly added to the next generation while the fittest members among parents and children complete the next generation until it contains N_{pop} individuals.

Termination condition: The genetic algorithm terminates after i_{max} iterations and a joint schedule for missions and maintenance operations minimizing the corrective and preventive maintenance costs can be obtained. The results are illustrated through application examples.

4 APPLICATION EXAMPLES

4.1 *Simulated data*

Two datasets of $n = 6$ missions represent two scenarii (Tables 1 & 2). The difference between them comes from the mission failure probabilities range. For the dataset A, the maximum mission failure probability is 1% while for the dataset B, it is 19%.

The dataset A further emphasizes the influence of the missions variances changes while the dataset B better illustrates the sensitivity studies for the impact of the ratio between the preventive and corrective maintenance costs and for the effect of $\mathbb{P}_{f_{max}}$.

Table 1. Dataset A.

Missions	Durations (h)	α_m	β_m	Failure probabilities
1	21	0.13	0.1	0.002
2	21	0.18	0.1	0.009
3	8	0.4	0.1	0.004
4	8	0.33	0.1	0.002
5	2	1.33	0.1	0.002
6	3	1.32	0.1	0.01

Table 2. Dataset B.

Missions	Durations (h)	α_m	β_m	Failure probabilities
1	9	0.85	0.1	0.189
2	4	1.85	0.1	0.163
3	3	1.34	0.1	0.011
4	6	0.93	0.1	0.048
5	6	0.90	0.1	0.041
6	2	3.81	0.1	0.184

Table 3. Parameters definition.

Parameter	Value	Parameter	Value
C_0	1000	C_f	3000
L	100%	N_{pop}	30
P_{cross}	0.7	P_{mut}	0.1
i_p	4	i_{max}	100
α	20%	β	20%
ε_{min}	10	ε_{max}	60

Table 4. Performance results.

	Dataset A		Dataset B	
Criterion	C_1	C_2	C_1	C_2
C_m	3787.2	3788.3	7898.4	7905
GA: T_c (s)	2.34	2.90	3.02	3.24
EM: T_c (s)	7.79	7.76	3.47	3.48

The preventive and corrective maintenance costs C_0 and C_f, the failure threshold L and the genetic algorithm parameters have to be initialized (Table 3). The maximum failure probability $\mathbb{P}_{f_{max}}$ is fixed at 0.95 for all the studies except for the one on the variance changes (4.3.3) for which $\mathbb{P}_{f_{max}}$ is equal to 0.1.

4.2 *Performances analysis*

The performance analysis is realized for both datasets and both criteria defined in section 3.2.1. It is performed on a computer with Intel© Xeon© CPU E3-1240 v5 @ 3.50 GHz and 16.0 GB RAM. For each dataset, 1000 realizations are generated to compare the maintenance cost and computation time results between the Genetic Algorithm (GA) and an Exact Method (EM). The exact method considers all the possible schedules respecting the block filling condition and their associated criterion values to find the schedule minimizing the criterion. The maintenance cost and the computation time are respectively denoted C_m and T_c. The study results are explained in Table 4.

Dataset A: The optimal joint schedule obtained with the exact method is the same for both criteria. This schedule is $\pi^{12}_{opt_1} = \{(1,3)(2,5)(4,6)\}$. The computation time gains with the genetic algorithm are quite significant: 70% forthe criterion C_1 and 62.6% for the criterion C_2.

Dataset B: The optimal joint schedule obtained with the exact method differs according to the chosen criterion. For the criterion C_1, the optimal schedule is $\pi^1_{opt_2} = \{(1)(2)(4)(6)(3,5)\}$ while for the criterion C_2, it is $\pi^2_{opt_2} = \{(1)(2)(3)(4)(5)(6)\}$. The computation time gains are lower than with the dataset A. They are respectively of 13.1% with C_1 and 7% with C_2. This difference comes from the number of feasible schedules when respecting the block filling condition. Indeed, the feasible schedule number for the datasets A and B are respectively 199 and 80. It explains why the exact method is faster for the dataset B.

For all the 1000 realizations with both datasets and criteria, the genetic algorithm converges towards the same schedule as the one obtained with the exact method. The only differences that can occur are permutations between the blocks or permutations of missions in the same block. As the criteria are computed based on the equivalent deterioration process for each block, the mission order does not have an influence on the block failure probabilities. The maintenance costs for the criterion C_2 are slightly higher because multi failures are considered in the blocks. It increases the associated corrective maintenance cost.

The genetic algorithm computation time for the criterion C_2 has increased with respect to the computation time for the criterion C_1. For the datasets A and B, the increases are about 24% and 7.4%. As the criterion C_2 considers multi failures probabilities, there are more calculations to do when computing the fitness values for the candidate schedules.

The results show the interest of using a genetic algorithm based method instead of an exact method to converge towards an optimal joint schedule for missions and maintenance while saving computation time. After showing the genetic algorithm performances, it is essential to study its sensitivity to several parameters to analyse its behaviour.

4.3 *Sensitivity study*

This section studies the influence of some parameters on the genetic algorithm behaviour, on the obtained joint schedule and on its associated performance. For each study, the dataset is selected to illustrate at best its behaviour.

4.3.1 *Impact of the ratio R_c*

The first study is interested in the impact of R_c, the ratio between the preventive and corrective maintenance costs respectively denoted C_0 and C_f. It is realized with the dataset B. R_c varies between 0.1 and 1 with a step fixed at 0.1 and the value for the preventive maintenance cost C_0 is the one defined in Table 3.

For both criteria, the block number in the schedule decreases when R_c decreases, i.e. when the corrective maintenance cost decreases (Table 5). When the corrective cost is very high with respect to the preventive one, the best schedule favours a high block number because grouping missions increases the failure risk in each block and leads to a significant maintenance cost inflation. On the contrary, when the two maintenance costs are quite similar, grouping the missions in few blocks is less expensive with respect to the maintenance costs.

Note that the reduction of the number of blocks is faster for the criterion C_1. When R_c is equal to 0.5, the optimal schedule for the criterion C_1 is composed of 4 blocks while the one for the criterion C_2 has 5 blocks.

4.3.2 *Effect of the variations of $\mathbb{P}_{f_{max}}$*

This section studies the effect of the maximum admissible failure probability for a block on the optimal schedule computation. This probability aims at guiding the genetic algorithm to converge faster towards the optimal solution. The smaller the probability is, the stricter the constraint on the block filling is.

This part is illustrated for the dataset B with the criterion C_1. Table 6 shows that when $\mathbb{P}_{f_{max}}$ increases, the number of blocks decreases. For most

Table 5. Optimal schedule when R_c varies (Criterion C_1).

R_c	Maintenance cost	Optimal schedule
0.1	12350	{(2)(6)(4)(1)(5)(3)}
0.2	9175	{(6)(2)(3)(1)(4)(5)}
0.3	8116.7	{(4)(3)(5)(2)(1)(6)}
0.4	7415.4	{(4)(1)(6)(5,3(2)}
0.5	6901.5	{(2)(1)(6)(5,4,3)}
0.6	6345.4	{(2)(6,1)(3,5,4)}
0.7	5867.4	{(5,3,4)(1,6)(2)}
0.8	5509	{(1,6)(5,4,3)(2)}
0.9	5230.2	{(1,6)(2)(5,3,4)}
1	5007.2	{(5,3,4)(2)(1,6)}

Table 6. Optimal schedule when $\mathbb{P}_{f_{max}}$ varies (Criterion C_1).

$\mathbb{P}_{f_{max}}$	Maintenance cost	Optimal schedule
0.2	7905	{(3)(4)(5)(6)(1)(2)}
0.3	7905	{(6)(5)(2)(1)(4)(3)}
⟦0.4;0.9⟧	7898.4	{(4)(2)(1)(3,5)(6)}
1	4000	{(1,5,6,4,3,2)}

of $\mathbb{P}_{f_{max}}$ values, the optimal schedule is composed of 5 blocks. When $\mathbb{P}_{f_{max}}$ is equal to 1, the optimal schedule groups all the missions in the same block because it is less expensive for the maintenance.

As expected, when the block filling constraint is relaxed, the optimal schedule is composed of less blocks. Note that when $\mathbb{P}_{f_{max}}$ is equal to 1, the optimal schedule not necessarily counts one block. For the dataset A, it is composed of 3 blocks for both criteria.

4.3.3 *Influence of missions variance changes*

In this part, the influence of the missions variance changes are studied to evaluate the genetic algorithm behaviour. For this specific study, the block filling condition $\mathbb{P}_{f_{max}}$ is fixed at 0.1. It is to be sure to illustrate only the effects due to the variance changes.

The initial dataset is the dataset A. As $\mathbb{P}_{f_{max}}$ is reduced with respect to the part 4.2, the optimal joint schedule differs. With both criteria, it is $\pi_{opt} = \{(2)(6)(5,1)(4,3)\}$. Based on this dataset missions, three datasets are generated. The missions durations are identical but the deterioration processes parameters are modified so that the increments expected value for each mission remains the same from one dataset to another. For each mission, the variance of the deterioration increment is either increased or decreased with respect to the initial variance (Tables 7 & 8). $\mathbb{P}$ denotes the mission failure probability.

Table 7. Datasets when increasing the variance by 2 or 5.

	Variance × 2			Variance × 5		
Missions	α_m	β_m	$\mathbb{P}$	α_m	β_m	$\mathbb{P}$
1	0.07	0.05	0.02	0.03	0.02	0.05
2	0.09	0.05	0.04	0.04	0.02	0.09
3	0.20	0.05	0.02	0.08	0.02	0.07
4	0.16	0.05	0.01	0.07	0.02	0.05
5	0.66	0.05	0.01	0.27	0.02	0.05
6	0.66	0.05	0.04	0.26	0.02	0.09

Table 8. Dataset when decreasing the variance by 2.

	Variance/2		
Missions	α_m	β_m	$\mathbb{P}$
1	0.27	0.2	4.2×10^{-5}
2	0.37	0.2	5.8×10^{-4}
3	0.79	0.2	1.1×10^{-4}
4	0.66	0.2	2.5×10^{-5}
5	2.65	0.2	2.7×10^{-5}
6	2.64	0.2	7.2×10^{-4}

When the variances are increased by a factor 2 and 5, the optimal joint schedules are respectively $\pi_{i_2} = \{(5,4)(6)(1)(3)(2)\}$ and $\pi_{i_5} = \{(5)(6)(1)(2)(4)(3)\}$. Increasing the variances increases the number of blocks in the optimal joint schedule. Indeed, increasing the variances also increases the probability to have a failure for each mission. It is then harder to group the missions into blocks. On the contrary, when the variances are reduced by a factor 2, the optimal joint schedule is $\pi_{d_2} = \{(4,6)(3,1)(2,5)\}$. The number of blocks is reduced. If the variances continue to decrease, the optimal schedule remains π_{d_2} because reducing more the block number lead to failure probabilities exceeding $\mathbb{P}_{f_{max}}$. The results are explained by the fact that the failure uncertainty increases when the variance increases.

4.4 *Larger size problems*

The higher the missions number to schedule is, the harder it becomes to compare the genetic algorithm results with the exact method results owing to the exact method computation time. The number of feasible schedules becomes too numerous. When considering $n = 9$ missions with $\mathbb{P}_{f_{max}}$ equal to 0.95, the exact method needs more than 11 days to reach the optimal schedule. With the genetic algorithm, the average computation times for the criteria C_1 and C_2 are respectively 2.63s and 3.77s. Note that when using C_1, 16% of the 1000 realizations do not reach the optimal solution and the maintenance cost deviation represents 1.45% of the optimal maintenance cost. Considering several failures by blocks gives a better maintenance cost estimation and a more coherent cost surface from one iteration to another. The convergence is then improved and when using C_2, all the schedules are identical to the optimal one.

To improve the genetic algorithm convergence, the population size N_{pop} and the maximum iteration number i_{max} can be increased, but the convergence is not guaranteed. But it will be to the detriment of the computation time. Indeed, increasing the population size and the iteration number also increase the computation time.

5 CONCLUSION

A static method based on a genetic algorithm is proposed to schedule missions and preventive maintenance for a vehicle by optimizing a maintenance cost based criterion. Two criteria consider either one or multi failures in the blocks and include the vehicle deterioration model, evolving due to the different mission severity. Its performances and sensitivity are described through

application examples. The genetic algorithm converges towards the optimal schedule in a satisfying computation time.

The obtained results are promising and offer improvement perspectives. Dynamic information, such as the deterioration level or the failure occurrences, could be integrated to update the schedule and reduce the maintenance costs even more. New missions could be added during the schedule completion. These different points will be further investigated to evolve towards a dynamic scheduling method.

REFERENCES

Benbouzid, F., Y. Bessadi, S. A. Guebli, C. Varnier, & N. Zerhouni (2003). Résolution du problème de l'ordonnancement conjoint maintenance/production par la stratégie séquentielle. In *e Conférence Francophone de Modélisation et de SIMulation, Toulouse, France*, Volume 2, pp. 627–633.

Bouvard, K., S. Artus, C. Bérenguer, & V. Cocquempot (2011). Condition-based dynamic maintenance operations planning & grouping. Application to commercial heavy vehicles. *Reliab. Eng. Syst. Saf. 96*(6), 601–610.

Cassady, C. R. & E. Kutanoglu (2005). Integrating preventive maintenance planning and production scheduling for a single machine. *IEEE Transactions on Reliability 54*(2), 304–309.

Coffman, E., M. Garey, & D. Johnson (1996). Approximation algorithms for bin packing: A survey. *Approximation Algorithms for NP-Hard Problems*, 46–93.

Da, W., H. Feng, & E. Pan (2016). Integrated preventive maintenance and production scheduling optimization on uniform parallel machines with deterioration effect. In *IEEE International Conference on Industrial Engineering and Engineering Management*, pp. 951–955.

Feng, H., W. Da, & L. Xi (2016). Joint optimization of flowshop sequence-dependent manufacturing cell scheduling and preventive maintenance. In *IEEE International Conference on Industrial Engineering and Engineering Management*, pp. 946–950.

Khoury, E. (2012). *Modélisation de la durée de vie résiduelle et maintenance prédictive: application à des véhicules industriels [In french]*. Ph. D. thesis, Université de Technologie de Troyes.

Ladj, A., F. Benbouzid-Si Tayed, & C. Varnier (2016). An integrated prognostic based hybrid genetic-immune algorithm for scheduling jobs and predictive maintenance. In *IEEE Congress on Evolutionary Computation*, pp. 2083–2089.

Lesobre, R. (2015). *Modélisation et optimisation de la maintenance et de la surveillance des systèmes multi-composants—Applications à la maintenance et à la conception de véhicules industriels [In french]*. Ph.D. thesis, Université Grenoble Alpes.

Li, H., M. Li, Q. Liu, & S. Li (2010). Integrated optimization research on preventive maintenance planning and production scheduling. In *IEEE International Conference on Management and Service Science*, pp. 1–5.

Michalewicz, Z. (1996). *Genetic Algorithms + data Structures = Evolution Programs*. Springer series Artificial Intelligence. Berlin; New York: Springer.

Rohlfshagen, P. & J.A. Bullinaria (2010). Nature inspired genetic algorithms for hard packing problems. *Annals of Operations Research 179*(1), 393–419.

Van Noortwijk, J. (2009). A survey of the application of gamma processes in maintenance. *Reliab. Eng. Syst. Saf. 94*(1), 2–21.

Yalaoui, A., K. Chaabi, & F. Yalaoui (2014). Integrated production planning and preventive maintenance in deteriorating production systems. *Information Sciences*, *278*, 841–861.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A concept for a holistic risk-based operation and maintenance strategy for wind turbines

Christian T. Geiss & Christian U. Grosse
Technical University of Munich, Munich, Germany

ABSTRACT: Operational wind turbines are highly loaded structures. Especially the high amount of load cycles-up to 10^9 –, leads to the conclusion that these renewable energy structures are in need of a suitable structural health monitoring and management strategy. In times of energy transition from conventional power plants to renewable power plants a safe and reliable operation of these assets is a premise. The paper presented here introduces an approach on determining risk-based operation and maintenance strategies for wind turbines. The approach combines Bayesian decision analysis with the concepts known in structural reliability analysis. Information from sensors can be used in probabilistic models to update the incomplete knowledge of the state of nature. They are a cost-effective risk-reduction measure and can be used for uncertainty quantification and reduction of uncertainties. Therefore the combination of the described methods is able to optimize maintenance and inspection activities according to the associated costs.

1 INTRODUCTION

A design lifetime of 20 years is generally assumed for modern wind turbines. The extension of service life beyond those 20 years has to be justifiable from a technical, safety, and economic point of view. Especially important are all load-transferring components that are relevant for the structural integrity of the wind turbine and the control and protection system.

The supporting structure of a wind energy systems is a large capital expenditure (CAPEX), on average between 20 to 30% of the overall capital expenditures. During the service life of wind energy systems, the structural supporting structure is particularly important and at risk. While drive train, generator and rotor blade systems already are monitored for structural health, this is not the case for the structural supporting structure.

As shown in the failure statistic in Figure 1 failures in the supporting structure cause the highest and most expensive downtime events. From both a technological and economical point of view it is worthwhile to use the supporting structure of a wind turbine systems as long as safety and reliability levels can be fulfilled at reasonable cost (Geiss 2014).

As shown in Figure 4, currently about 1.000 wind turbine systems already have exceeded their 20 year designed service life. The overall capacity of the German wind energy installation already in its second half of service life is 41%.

The described situation stresses the need for new maintenance strategies and life cycle assessments

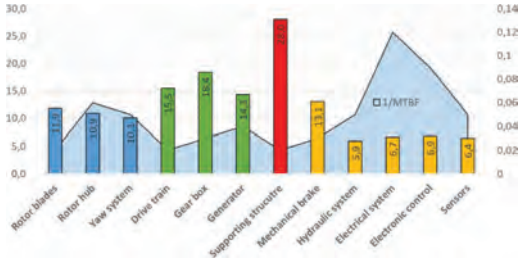


Figure 1. Wind turbine damage statistic (WInD-Pool 2017).

of onshore wind turbines. Especially challenging for existing wind turbine structures is that the original design documents cannot often be used as a reference for a remaining useful lifetime analysis.

According to the DNV GL guideline (DNV GL AS 2016) there exist four methods of lifetime extension analysis:

1. Lifetime extension inspection
2. Simplified approach
3. Detailed approach
4. Probabilistic approach

The assessment shall always be based on a combination of an analytical part and a practical part. The analytical part incorporates an assessment based on new or additional calculations for the wind turbine, considering the site-specific conditions. The lifetime calculation of the analytical part should be supplemented with relevant field experience of the wind

turbine model concerning weak points, known failures or retrofits. Figure 2 gives an overview of the different life cycle phases of a wind turbine system.

The probabilistic approach integrates the use of stochastic methods in the assessment of structural integrity. Alternatively, rather than use of deterministic values in the simplified and detailed approach, the probabilistic approach uses appropriate probability distributions to characterize the uncertainty in models and model inputs. The stochastic parameters—such as probability distribution types, expected values, coefficients of variation, or correlation coefficients—in the limit state formulations have to justify that they do not introduce errors into the analyses. A structural reliability analysis (SRA) in the following course of actions has to be carried out according to DNV GL:

1. Selection of a target reliability level
2. Identification of failure modes in the system
3. Development of limit state functions (g-functions) for each failure mode based on engineering theory
4. Quantification of the deterministic and stochastic variables within the limit state function and their correlations
5. Use of appropriate methods (e.g. first order reliability method—FORM) to compute reliability indices or probabilities of failure for the structural components
6. Comparison of the computed component reliability with target reliability level for each component
7. Analysis of results using sensitivity analysis

The aleatory and epistemic uncertainty in mathematic models and input parameters are to be described by appropriate probability distributions, and the nature of uncertainty has to be clarified. Any type of measurement can be used to update the probability models. If aero-elastic models or component resistance are not available, generic load and resistance models can be used. Based on the probabilistic approach, risk-based inspection methods may be developed.

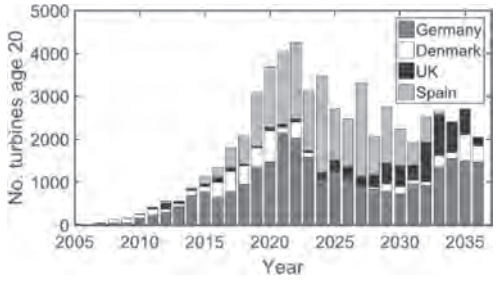


Figure 2. Wind turbine age structure (Ziegler et al. 2018).

The two most relevant deterioration mechanisms for a wind turbine supporting structure are corrosion mechanisms and fatigue crack growth mechanisms. The current research approach focusses on the fatigue crack-growth mechanism, as the most critical deterioration mechanism of the structural system.

Theoretical models exist that describe the specific fatigue mechanisms causing wear of a structural element. However, these models inherit implied model uncertainties, which can be reduced by in-field inspection data. Non-destructive testing (NDT) methods are therefore effective risk mitigation measures for existing structures.

2 DETERIORATION MODELLING

All deterioration mechanisms are time-dependent and consequently all reliability problems in fatigue are time dependent (JCSS 2002). A failure event of a deteriorating structure can be modelled as a first passage problem. The limit state function is then also a function of time. Failure occurs when the limit state function becomes negative for the first time, given that it was positive at $t = 0$. In that case, the probability of failure between time $t = 0$ and time $t = T$ can be expressed by the following equation:

$$p_F(T) = 1 - P(g(\underline{X}(t)) > 0 \tag{1}$$

For most deterioration processes the problem is simplified by the fact that damage is monotonically increasing with time. If the modelled deterioration problem has a fixed damage limit—failure occurs when damage reaches a constant limit—the deterioration problem can be solved as a time-independent problem. The time variable t is then a simple parameter of the model and deterioration is treated as a monotonously increasing process.

E.g., if failure has not occurred at time t_1, failure has also not occurred at time $t < t_1$. For the definition of a failure rate of the modelled system, several definitions are possible. In this case the annual failure probability of the modelled system

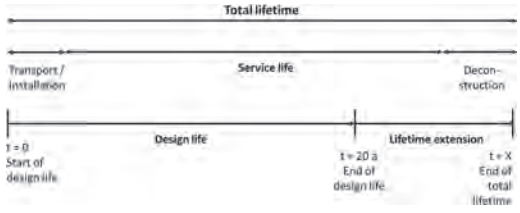


Figure 3. Life cycle phases of wind turbines.

is best fitting. This circumstance enables to evaluate the reliability problem at fixed time intervals $t = t_1; t_2 \quad ; t_n$.

$$t = t_{i-1} + 1yr \tag{2}$$

Consequently, the annual probability of failure in year t_i can be expressed as follows:

$$\Delta p_F(t_i) \approx \frac{p_F(t_i) - p_F(t_{i-1})}{1 - p_F(t_{i-1})} \tag{3}$$

Very typically this is a SN fatigue modelling problem. Failure is defined to occur when the accumulated damage has reached Δ or commonly defined as $\Delta = 1$.

In practical engineering application the computation of probabilities is done with Monte-Carlo-simulations and FORM-algorithms, e.g., (Strurel 2017).

3 STEEL FATIGUE PROCESS

Generally, fatigue arises at points of local stress concentrations; so-called hot spots can be represented by welds, cut-outs and a wide variety of mechanical connection types—e.g., bolts. Due to inhomogeneities, welds are especially relevant as local hot spots and stress concentrations. State of the art engineering theory uses specific steel fatigue models for the description of the fatigue process caused by fluctuating stresses. Fundamentally the steel fatigue models are sub-divided into S-N models—based on experiments- and fracture mechanic models.

A classic formulation is represented by the Basquin equation, which describes a linear relationship between lnN_F and $ln\Delta S$.

$$N_F = C_1 \cdot \Delta S^{-m_1} \tag{4}$$

C_1 and m_1 are material parameters and are defined by experiments.

The theory of linear damage accumulation goes back to Palmgren (Palmgren 1924). The Palmgren-Miner law is based on the assumption of linear and interaction-free damage accumulation. The damage accumulated after N cycles is independent of the order in which the stress cycles occur. The damage increment for each cycle has the stress range ΔS_i.

$$\Delta D_i = \frac{1}{N_{F,i}} \tag{5}$$

$N_{F,i}$ is the number of cycles to failure for ΔS_i as given by the associates SN-curve. The total accumulated damage after N cycles can be described as follows:

$$D_{tot} = \sum_{i=1}^{N} \Delta D_i \tag{6}$$

Fatigue failure is reached when D_{tot} reaches a specific value, here expressed as Δ. Generally, Δ is modelled with the mean value 1. Subsequently, the SN fatigue limit state function can be described:

$$g_{SN} = \Delta - D_{tot} \tag{7}$$

The particular stress ranges and number of stress cycles are commonly application specific. Considering the distribution those two parameters can often be approximated by a Weibull or Rayleigh distribution. The Weibull distribution is a good description for natural processes related to dynamic response of elastic system—e.g. marine structures (Almar-Naess 1984) or wind turbines (Ziegler, Muskulus 2016a).

Generally, the described empirical fatigue design laws and models inherit obviously many uncertainties. The three general uncertainties in the empirical fatigue design approach are:

- Uncertainty of the fatigue model (SN-curve)
- Uncertainty of the fatigue resistance (uncertainty on the applied SN curve)
- Uncertainty about the loading and monitoring sensors

An applicable limit state function for the fatigue design state of a steel structure can be found in (Straub 2014) and is expressed as follows:

$$g_{SN} = \Delta - \nu TE\left[\Delta D_i\right] \tag{8}$$

The fracture mechanic approach is opposed to the SN-model approach. The theory implies the initiation and gradual development and propagation of small cracks in the microstructure due to cyclic loads. In practice, the crack propagation mechanism is often combined with corrosion mechanisms.

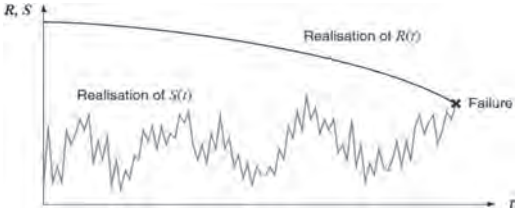


Figure 4. Stress and resistance of a structural system in time.

Table 1. Generic steel SN-fatigue model (Straub 2014).

Param.	Dimension	Distrib.	Mean	COV
Δ	–	LogN	1	0,3
v	yr^{-1}	Det	10^7	
T_{SL}	yr	Det	40	
$k_{\Delta S}$	Nmm^{-2}	Det	7,448	
B_S	–	LogN	1	0,25
$\lambda_{\Delta S}$	–	Det	0,9	
C_1	$(Nmm^{-2})^{m1}$	Lognormal	4,48e12	0,51
m_1	–	Det	3	
m_2	–	Det	5	
N_q	–	Det	10^7	
N_0	–	Det	∞	
d	mm	Det	16	

($FDF = 2$).

The fracture mechanic fatigue models introduce a stress intensity factor, which is derived from load intensity and crack length, and is the predominant control of crack propagation. Fatigue crack growth models are usually based on the linear elastic fracture mechanic theory.

The evolution or lifetime of a crack can be divided in three stages:

- Initiation → Number of cycles spent in that phase → N_I
- Propagation → Number of cycles spent in that phase → N_P
- Failure → Total number of cycles N_F → analogous to the SN-approach

$$N_F = N_I + N_P \tag{9}$$

For the purpose of deterioration control, an innovative concept for wind turbine structures is to combine the SN-model approach with the fracture mechanics approach to a hybrid probabilistic model, which describe the fatigue crack dimension at any time during the service life of a wind turbine.

The SN-model approach represents all the design assumptions that have been made to design the structural element resistant to fatigue damage. On the other hand, the fracture mechanics approach is basically suitable for in-service deterioration control of structural elements, crack width and crack length can be measured by non-destructive testing methods—e.g. ultrasound or magnaflux.

In every model calibration, a measure of how well a particular calibration fits has to be introduced. The SN-model gives information on whether a hot spot has failed or survived, whereas the FM-model gives the crack dimension after any number of cycles. Therefore, the FM-model has to be calibrated to the SN-model after the number of cycles to failure the critical crack size is reached in the FM-model.

The parameters to which the model should be fitted are random. In engineering theory there exist already a few calibration algorithms as state of the art.

Considering the described facts Straub (Straub 2014) developed a hybrid solution for a calibration of the SN-model to the FM model. The probability distribution of $F_{N_F}(N)$ is equivalent to the probability of failure p_f as a function of the number of cycles.

$$p_F(N) = F_{N_F}(N) \tag{10}$$

The actual calibration is performed by a least-square fitting in β-space.

$$\beta = -\Phi^{-1}(p_f) \tag{11}$$

A minimization with respect to the parameters of the fracture mechanics model $x_1 \ldots x_N$ is carried out.

$$\min_{x_1 \ldots x_N} \sum_{t=1}^{T_{SL}} \left(\beta_{SN}(t) - \beta_{FM}(t; x_1 \ldots x_N)\right)^2 \tag{12}$$

with

- $\beta_{SN}(t)$ as the reliability at time t using the SN-model
- $\beta_{FM}(t; x_1 \ldots x_N)$ as the reliability at time t using the FM model

The evaluation of reliability indexes is performed by a FORM or SORM algorithm. The choice of parameters to be calibrated depends on the applied FM-model. Generally, two parameters have to be calibrated. First, the crack growth rate parameter, which has a large influence on crack growth, and the second parameter for which least information is available, but influences the structural reliability.

A solution for the one-dimensional crack growth model can be expressed as follows:

$$C_P \cdot \Delta S^{m_{fm}} \cdot (N - N_I) = \int_{a_0}^{a} \frac{dz}{Y_G(z)^{m_{fm}} (z \cdot \pi)^{m_{fm}}} \tag{13}$$

It follows:

$$a(N) = \left(a_0^{(2-m_{f_m})/2} + \frac{2 - m_{f_m}}{2} C_P \left(Y_G \pi^{1/2} \Delta S\right)^{m_{f_m}} (N - N_I) \right)^{2/(2-m_{f_m})}, m_{f_m} \neq 2 \tag{14}$$

Ziegler (Ziegler, Muskulus. 2016b) is also conducting research in analyzing the application of risk-based inspection methods for deterioration control of offshore supporting structures.

4 CONCRETE FATIGUE PROCESS

In the field and onshore a tremendous amount of hybrid supporting structures currently exists, which combine a pre-stressed concrete part as lower tower section with a conical steel part as tower top; a holistic research approach also has to consider fatigue performance indicators of concrete elements, which can be measured by NDT techniques in field.

Thiele (Thiele 2016) conducted a study on suitable fatigue performance and damage indicators for concrete specimen in the framework of his doctoral thesis. A main finding relevant for the research here is, that the E-Modulus tends to fit as a fatigue performance indicator for concrete structures. Besides the E-Modulus Thiele investigated deformations, cycle counts, and micro cracks as potential fatigue performance indicators. Ultrasound amplitudes showed a high sensitivity to the damage evolution in the concrete specimen. In both cases scattering in the measurement campaign was relatively low compared to other measurement principles, e.g. sound emission measurements.

Ultra sound measurement has the advantage as an in-service inspection technique since measurements can be carried out while the structural element is in service without a special load slope or any other special preparations. However, it must be assured that the coupling of the ultrasound sensors is appropriate.

In a current research project—called MISTRAL-WIND – at the Chair of Non-Destructive Testing at the Technical University of Munich, more sophisticated experiments considering those findings will be run in January 2018 (Geiss, et al. 2017).

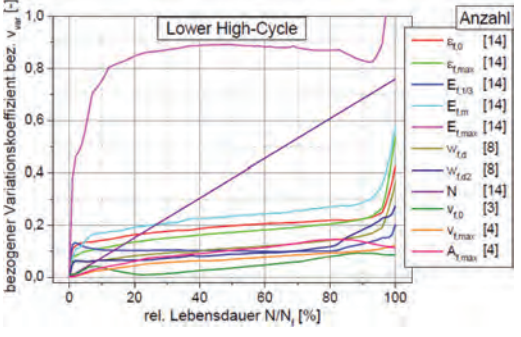


Figure 5. Variation coefficient of several physical fatigue damage indicators over the relative life time of a concrete specimen (Thiele 2016).

5 RISK-BASED INSPECTION PLANNING

The concept of risk based inspection planning inherits the primary goal of quantifying the effect of inspections on the risk condition of a component and thus enables cost optimal inspection planning. Given the fact that design deterioration laws represent imperfect knowledge considering the component's in-service deterioration process, risk-based inspection planning is a suitable method for deterioration control. Madsen showed one of the first applications of the concept (Madsen, Krenk, Link 1986).

The method of uncertainty quantification has two main objectives, first the quantitative characterization of uncertainties, and second the reduction of uncertainties. Inspections can reduce uncertainties and update the incomplete knowledge of the structural state, which can be described as an epistemistic uncertainty. In many applications in structural asset management, inspections can be a cost-effective risk reduction measure. Empirical statistics for structural systems are rarely available since every structure is more or less unique. Therefore professional experience with structural failures is scarce. A first approach in this direction is represented by the WInD-Pool project (WIndPool 2017). Qualitative estimations of the probability of failure are not suitable. Eventually risk-based inspection planning has to seek the equilibrium between a desired level of reliability and the cost optimal allocation of maintenance activities in a holistic life cycle asset management framework.

6 RISK-BASED INSPECTION ALGORITHM

In many practical situations, the conditional probability of specific events is of interest, meaning the probability of occurrence of event E2 given the occurrence of Event E1. This classical probabilistic dependency is generally handled with Bayes' rule (Faber 2007):

$$P(E_2 \mid E_1) = \frac{P\left(E_1 \cap E_2\right)}{P\left(E_1\right)} = \frac{P(E_1 \mid E_2)P\left(E_2\right)}{P\left(E_1\right)} \qquad (15)$$

From this basic equation, one can derive basic and important definitions for the RBI framework:

- $P(E_1|E_2)$ is the likelihood measure for the amount of information on E_2 gained by knowledge of E_1. The likelihood measure is typically used to describe the quality of an inspection.

- $P(E_2|E_1)$ is known as the posterior probability of occurrence of E_2 or its updated occurrence probability.
- $P(E_2)$ is the prior probability of Event E_2, prior to the knowledge of E_1.

In the probabilistic RBI framework different inspection outcomes or results are possible. Those different inspections results will trigger different maintenance actions, which are also described by limit state functions.

As relevant inspection outcomes in the RBI framework the following outcomes can be defined:

- Event of indication of a defect I
- Event of detection of a defect D
- Event of false indication FI
- Event of a defect measurement with a measured size s_m

7 INSPECTION MODELLING

The number of inspection performance models today is limited, because round-robin tests for empirical models are expensive and empirical models for one application cannot be transferred to other applications. Straub and Faber (Straub, Faber. 2002a, Straub, Faber. 2002b, Straub, Faber. 2003) developed quantitative models for a risk-based inspection framework using Bayesian updating techniques.

The performance of a non-destructive inspection is dependent on many parameters.

- Defect size and geometry
- Defect orientation
- Environmental condition
- Sensitivity of the inspection method
- Sensitivity of the sensors
- Sensor placement
- Accessibility of the structure tested
- Inspector performance
- etc.

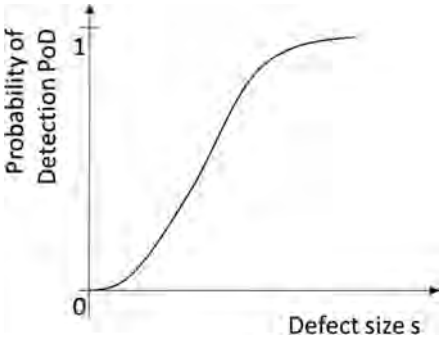


Figure 6. Probability of detection vs. defect size.

The probability of detection (POD) is the mean rate of success when the specific inspection technique is performed and is exposed to various sources of uncertainties. One and two dimensional POD models are typically used for the analysis.

Because the POD-function is a monotonically increasing function, the probability of detecting a crack smaller than or equal to the size s is $POD(s)$. $POD(s)$ becomes 1 for very large crack sizes.

The classical approach of the POD formulation has some shortcomings. The POD-function must be a distribution function, which is not always the case in reality. Furthermore it is difficult to integrate the two dimensional POD-functions.

Uncertainties inherited in the inspection performance models are:

- Variability due to scatter in the response signal (aleatory uncertainty)
- Statistical uncertainty due to limited set of trials in experimentally determined POD/POI models (epistemic)
- Model uncertainty due to empirical nature of the parametric model (epistemic)

Straub and Faber (see citations above) analyzed the aleatory and epistemic uncertainty sources in such models in more detail. Jüngert and Kurz (Kurz, et al. 2011) conducted a research study concerning the POD performance of ultrasound NDT tests. POD performance test represent a considerable bottleneck in bringing probabilistic approaches in practical and reliable applications.

8 DECISION ANALYSIS

The ultimate goal in the decision analysis process is the identification of optimal decisions on maintenance actions for deteriorating structures. The decision environment is subjected to uncertainty under the following aspects:

- Uncertainty on the state of the system; state of deterioration
- Uncertainty on the performance of the inspection; probability of detection (POD)
- Uncertainty on the performance of repair actions
- Uncertainty on the consequences of failures

One method of evaluating such decision problems is through the deployment of so-called decision trees. Each path of the decision tree is assigned with a utility value and its probability characteristics. Considering the point in time and informational characteristics of the decision problem three basic situations can be defined.

1. The prior analysis represents a decision analysis with given information. At this state, the utility

function and the probabilities of the various states of nature corresponding to the different consequences have been defined. The decision analysis is reduced to the computation of the expected utilities and finding the optimal point of the optimization problem.
2. The posterior decision analysis represents a decision analysis problem with additional information on the state of nature. If additional information becomes available—e.g. through inspections—the probability structure in the decision problem can be updated. The probability update is carried out using Bayes' rule.
3. The pre-posterior analysis represents a decision analysis situation dealing with unknown information. The decision maker has the possibility to buy additional information through an experiment. If the cost of this information is small in comparison to the potential value of information, the experiment should be performed. If several experiments are potentially suitable, the decision maker has to choose the experiment yielding the overall largest utility for his decision problem.

As a first step a set of possible events will be defined, E_1 to E_n. Those events can be compared with different outcomes of a game. Events with a large index are preferred over other events with a lower index. The decision maker can choose different actions also called lotteries. Each action will lead to probabilities of occurrence for different events. E.g., action a will lead to Event E_1 with the probability $p_1^{(a)}$ and to Event E_2 with the probability $p_2^{(a)}$. Action b will lead to event E_1 with the probability $p_1^{(b)}$ and so forth.

All probabilities of occurrence have to fulfill the basic condition: $p_1^i + p_2^i + p_n^i = 1$.

The utility index is used to express specific preferences of the decisionmaker, in such a way that one decision is preferred to another if the expected utility of the former is larger than the utility of the latter. A utility index u can be assigned to the different basic events E_1 to E_n.

The final optimization criterion is the expected cost criterion. The utility index u is linearly assigned to monetary for the considered range of events. The indirect costs associated with the failure events and repair and inspection are included in the probabilistic modelling. Thus, all consequences of an event have to be expressed in monetary terms, which can also introduce weakness in the approach due to lack of reliable information on monetary consequences,.

The pre-posterior decision analysis has the intention to identify the optimal decision on possible inspection actions, enabling optimal maintenance planning.

Concerning the inspection actions it has to be determined:

- Where to inspect → location of inspection
- What to inspect → indicator of the system state
- How to inspect → what kind of inspection technique
- When to inspect → time of inspection

The inspected costs of an inspection strategy have to be determined. Initially, the number of decision tree branches to consider is evaluated:

$$n_b = n_a^{n_{Insp}} + \sum_{i=0}^{n_{Insp}} n_a^i \tag{16}$$

9 INSPECTION COST MODEL

The inspection cost model integrates the following cost parameters

- Expected costs of failure C_F
- Cost of inspection as a function of inspection technique e_i at time t
- Cost of repair C_R
- Interest rate r

The total expected costs during the service life period T_{SL} is computed as the summation of the expected failure costs, the expected inspection costs and the expected repair costs:

$$\begin{aligned} E\left[C_T\left(\underline{e},d,T_{SL}\right)\right] &= E\left[C_F\left(\underline{e},d,T_{SL}\right)\right] \\ &+ E\left[C_I\left(\underline{e},d,T_{SL}\right)\right] + E\left[C_R\left(\underline{e},d,T_{SL}\right)\right] \end{aligned} \tag{17}$$

The decision rule has a significant influence on the cost function. A short compilation of possible maintenance decision rules can look like:

1. Repair all defects indicated at the inspection
2. After indication perform a measurement and repair only cracks deeper than a_R
3. Etc.

For the optimization procedure a maximum annually probability of failure is allowed Δp_F^{max}:

$$\begin{aligned} &min_{e,d} E\left[C_T\left(\underline{e},d,T_{SL}\right)\right] s.t. \\ &\Delta p_F\left(\underline{e},d,t\right) \le \Delta p_F^{max} \; t = 0,\ldots T_{SL} \end{aligned} \tag{18}$$

However, this optimization procedure inherits to major restrictions:

- The minimum analysis period is one year
- The calculation of total service life is prohibitive

To level out the latter restriction, two simplification approaches are possible, the constant threshold approach and the equidistant inspection time approach (Straub 2004).

Applying the constant threshold approach, the optimization parameter is the annual probability of failure $\Delta p_F{}^T$ and inspection is always performed in the year before $\Delta p_F{}^T$ is exceeded. It follows as new optimization problem:

$$min_{e,d,\Delta p_F{}^T} E\left[C_T\left(\underline{e}, \Delta p_F{}^T, d, T_{SL} \right) \right] s.t. \Delta p_F{}^T \leq \Delta p_F^{max} \tag{19}$$

In that way an RBI-based approach can be deployed a holistic and effect structural health monitoring and management strategy, which combines design information with realistic in situ performance measurements using non-destructive testing techniques.

10 OUTLOOK

The author's future research will focus on integrating the RBI-approach into the overall asset management strategy of wind turbine systems as well as developing, validating and maturing the approach in in situ and laboratory tests. Laboratory experiments will be undertaken on broadening the understanding of applicable fatigue performance indicators of concrete structures under the influence of fatigue loading, which can be measured with ultrasound based NDT techniques.

Furthermore, validated methods and concepts can be broadened to other civil engineering structures that have a high demand of cost-optimal allocation of maintenance activities, such as bridges.

Last but not least it has to be considered how far such models are capable of integrating combined effects of deterioration mechanisms, e.g., corrosion and fatigue deterioration and their interaction.

ACKNOWLEDGEMENT

This research was partly funded by the Federal Ministry for Economic Affairs and Energy. Furthermore, the help of Prof. Grosse as co-author and Prof. Glaser from UC Berkeley through the preparation of this article is gratefully acknowledged.

REFERENCES

Almar-Naess, A. (ed). 1984. Fatigue Handbook of Offshore Steel Structures. *Tapir Publishers.* Trondheim.

Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit. 2015. Erhöhung der Verfügbarkeit von Windkraftanlagen. Berlin.

DNV GL AS. 2016. Lifetime extension of wind turbines. DNVGL-ST-0262.

Faber, M.H. 2007. Risk and Safety in Civil Engineering. Lecture Notes. *Swiss Federal Institute of Technology*. Zurich.

Geiss et al. 2017. The Mistralwind Project – Towards A Remaining Useful Lifetime Analysis And Holistic Asset Management Approach For More Sustainability Of Wind Turbine Structures*, International Workshop on Structural Health Monitoring 2017*. Stanford.

Geiss, C.T. 2014. Economic Aspects of Prognostics and Health Management Systems in the Wind Industry. *European Conference of the Prognostics and Health Management Society*. Nantes.

Geiss, Guder. 2017. Reliability-centered Asset Management of Wind Turbines – A Holistic Approach for a Sustainable and Cost-optimal Maintenance Strategy. *International Conference on System Reliability and Safety.* Milan.

JCSS. 2002. Probabilistic Model Code. *Joint Commitee on Structural Safety*, http://www.jcss.byg.dtu.dk/Publications/Probabilistic_Model_Code.aspx.

Kurz, et al. 2011. Experimentelle POD Bestimmung mittels Ultraschall Phased Array zur Einbeziehung zerstörungsfreier Prüfungen in probabilistischen Versagensanalysen. *DGZfP-Jahrestagung.* Berlin.

Madsen, Krenk, Link. 1986. Methods of Stuctural Safety. *Prentice Hall.* New Jersey.

Palmgren, A. 1924. Die Lebensdauer von Kugellagern. *Zeitschrift des Vereins deutscher Ingenieure.* 68(14): 339–341.

Straub, D. 2004. Generic Approaches to Risk Based Inspection Planning for Steel Structures. *Swiss Federal Institute of Technology*. Zurich.

Straub, Faber. 2002a. On the Relation between Inspection Quality and Quantity. *Proceedings of the European-American Workshop on Reliability of NDE*. Berlin.

Straub, Faber. 2002b. System Effects in Generic Risk Based Inspection Planning. Journal of Offshore Mechanics and Arctic Engineering. Oslo.

Straub, Faber. 2003. Modeling Dependency in Inspection Performance. Applications of Statistics and Probability in Civil Engineering. *Der Kiureghian, Madanat & Pestana:* 1123–1130.

Strurel. 2017. Structural reliability analysis program system. Comrel 9.5. *RCP & ERACONS.* Munich.

Thiele, M. 2016. Experimentelle Untersuchung und Analyse der Schädigungsevolution in Beton unter hochzyklischen Ermüdungsbeanspruchungen. BAM-Dissertationsreihe. Band 140. Berlin.

WInD-Pool 2017, Windenergie-Informations-Daten-Pool https://wind-pool.iwes.fraunhofer.de/.

Ziegler, Muskulus. 2016a. Lifetime Extension of Offshore Wind Monopiles – Assessment Process and Relevance of Fatigue Crack Inspection. *European Academy of Wind Energy*. Trondheim.

Ziegler, Muskulus. 2016b. Fatigue reassessment for lifetime extension of offshore wind monopile substructures. *Journal of Physics (753).*

Ziegler et al. 2018. "Lifetime Extension of Onshore Wind Turbines: A Review Covering Germany, Spain, Denmark, and the UK." Renewable and Sustainable Energy Reviews 82.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Optimising the maintenance strategy for a multi-AGV system using genetic algorithms

R.D. Yan, S.J. Dunnett & L.M. Jackson
Loughborough University, Loughborough, Leicestershire, UK

ABSTRACT: Automated Guided Vehicles (AGVs) are playing increasingly vital roles in a variety of applications in modern society, such as intelligent transportation in warehouses and material distribution in automated production lines. They improve production efficiency, save labour cost, and bring significant economic benefit to end users. However, to utilise these potential benefits is highly dependent on the reliability and availability of the AGVs. In other words, an effective maintenance strategy is critical in the application of AGVs. The research activity reported in this paper is to realise an effective maintenance strategy for a multi-AGV system by the approach of Genetic Algorithms (GA). To facilitate the research, an automated material distribution system consisting of three AGVs is considered in this paper for methodology development. The movement of every AGV in the multi-AGV system, and the corrective and periodic preventive maintenances of failed AGVs are modelled using the approach of Coloured Petri Nets (CPNs). Then, a GA is adopted for optimising the maintenance and associated design and operation of the multi-AGV system. From this research, it is disclosed that both the location selection of the maintenance site and the maintenance strategies that are adopted for AGV maintenance have significant influences on the efficiency, cost, and productivity of a multi-AGV system.

1 INTRODUCTION

AGVs are increasingly used in modern society attributed to their high efficiency, accuracy, low cost and therefore significant economic benefit (Tuan, 2006).

However, with the emerging trend in modern society for AGVs designed for more complex tasks they have become larger and larger in size, where the reliability and maintenance issues in recent years are receiving increasing concern. However, to the author's best knowledge, so far there has not been sufficient research being conducted in this area except a few preliminary researches (Vis, 2006). For example, three major hazards, i.e. collision, tilting over and falling, have been identified during the operation of AGVs (Trenkle, 2013); as well as a combined Markovian model and a neural network were applied to maximise the reliability of AGVs and minimise their repair cost at the same time (Fazlollahtabar, 2013). Little research has been conducted to deal with the maintenance issue of failed AGVs except using a control method for enhancing the failure control management of both loaded and unloaded AGVs in an underground transportation system (Ebben, 2001). For this reason, the purpose of this research is to fill this technology gap through developing an optimal maintenance strategy for a typical multi-AGV system. At present, preventive and corrective maintenance are two basic strategies that are popularly adopted in engineering practice (Smith et al., 1973). In the past decades, there have been a number of research studies conducted to optimise the maintenance strategies dedicated to various kinds of industrial applications. For example, a simulation was carried out to evaluate the performance of manufacturing production lines with different maintenance policies (Lei, 2010); The maintenance cost and availability of an aircraft system was optimized using a mathematical replacement model (Fornlöf, 2016) and so on.

In this paper, both preventive and corrective maintenance strategies dedicated to a multi-AGV system are studied by the combined use of a Coloured Petri nets (CPN) simulation model and a specifically designed Genetic Algorithm (GA) model.

The remaining part of the paper is organised as follows. A brief description of the multi-AGV system considered in the paper is given first in Section 2; the potential of the CPN in describing the paths, routing and maintenance issues of the AGVs is explored in Section 3; the maintenance strategy of the multi-AGV system is optimised with the aid of GA in Section 4; and the paper is finally ended with a few key research conclusions in Section 5.

Table 1. Presumed time duration of every phase.

Phase	Phase length (hours)
1	0.02
2	0.2
3	0.02
4	0.2
5	0.02
6	0.2

2 CONFIGURATION OF THE MULTI-AGV SYSTEM

The AGV transport system described in (Yan, 2017) is also considered in this research. However, instead of considering a single AGV, a more complicated transport system consisting of three AGVs will be investigated in this paper. This will allow the investigation of the interactions between different AGVs and the impact of the failure of either one or more AGVs on the operation of the others in the same transport system. In addition, it is worth noting that in this research the subsystems of the individual AGVs are assumed to fail 12 times every year, as cited in (Yan, 2017). The mission of the AGVs is divided into six phases, namely (1) mission allocation and route optimization, (2) dispatch to station, (3) loading of item, (4) travelling to storage, (5) unloading and (6) travelling back to base. In order to facilitate the research, the time duration of every phase is presumed and listed in Table 1 for demonstration purpose. They would be different when the AGV is requested to deliver different types of missions.

In the model, it is assumed that the AGV will be taken away from the system immediately to prevent deadlock and conflicts as long as it fails, so that the downtime of the system due to AGV failures can be minimised. To meet such a need, it is essential to optimise the location of the maintenance site in the system to enable the recycle vehicle (the vehicle collecting the failed AGV) to reach and recycle the failed AGV in the shortest time.

3 SIMULATION MODELLING

3.1 *Coloured Petri Nets (CPNs)*

Attributed to the unique efficiency and cost-effectiveness features, modelling has been identified as one of the most important approaches to improve the design and operation of a system. In particular, a Petri net (PN) is regarded as one of the most economic and effective tools to model AGV systems (Wu, 1999; Nishi, 2010). The concept of a PN was developed by Petri (1962), which is a direct bipartite graph. It consists of four types of symbols, i.e. circles, rectangles, arrows and tokens, as shown in Figure 1. Where, circles represent the places, which may be conditions or states (e.g. mission failure, phase failure, or component failure); rectangles represent the transitions, more abstractly actions, or events which cause the change of condition or state; Arrows connect places and transitions; Tokens are small marks that gives dynamic properties of the PN. They move via transitions if the enabling condition is satisfied. It provides an intuitive graphical representation of a system and allows flexible description of events.

What Figure 1 shows is an example explaining how the tokens move through a net. From Figure 1a, it is seen that there are two inputs and one output place connected to a timed transition with a time delay t. The input places have arcs with weights 2 and 1, respectively. Once the transition is enabled after the time delay, t, the arc weight number of tokens will be taken out from the corresponding input place to fulfil the transition after the time delay t associated with the transition. For the example, as Figure 1b shows, one more token will appear in the output place.

However, conventional PN methods are found inefficient in describing complex systems or describing a system that is designed to carry out complex tasks or missions (Jensen, 2015). To address this issue, a more advanced PN method, namely Coloured Petri nets (CPN), was proposed by Rene (1994). In comparison with the conventional PN, each individual token in the CPN is designed with a specific colour, which either has different identities or carries different information. Therefore, they are more informative than those present in the conventional PN.

3.2 *System modelling*

In a multi-AGV system, every AGV need to be distinguishable as they may be located at different positions in the transport system and may fail at different times. In view of the powerful capability of CPNs in describing the kind of complex situations (Wu, 2002; Aized, 2009), CPN is employed in the following research.

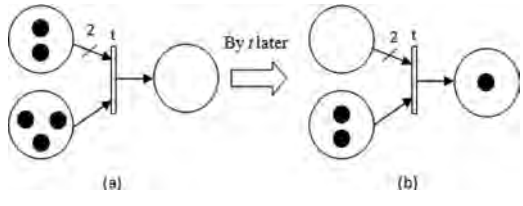


Figure 1. Enabling and switching of transition, (a) before enabling transition, (b) after enabling transition.

To correctly describe the operation and maintenance activities in a multi-AGV system, three types of CPN models are purposely developed and detailed below:

1. Path Petri nets (PPN) – for describing the layout configuration of the system;
2. Corrective maintenance Petri nets (CMPN) – for defining the corrective maintenance of failed AGVs in the system;
3. Periodic maintenance Petri nets (PMPN) – for defining the periodic maintenance of all AGVs in the system.

Herein, the CMPN and PMPN share the AGV failure information and feed their responses into PPN.

3.2.1 *PPN*

A three-AGV dispatching system is considered in this paper. It consists of 1 AGV base, 1 pickup station, 1storage site, 1 maintenance site, and a number of transport paths. The base is for storing and recharging the AGVs; the pickup station is the place where items are collected; and storage is the destination for unloading the items. All these places are assumed to have sufficient space for parking multiple AGVs. To demonstrate the significant influence of layout configuration on the efficiency of recycling failed AGVs from a multi-AGV system, three different layout configurations are considered, as shown in Figure 2. Where, MS indicates the location of the maintenance site.

From Figure 2, it is seen that different layout configurations are distinguished by the different locations of maintenance site and the extra paths for recycling failed AGVs. For example, in Figure 2a the maintenance site shares the same space with the base; in Figure 2b the maintenance site is located between the base and the storage. In addition, an extra path between the pickup station and the maintenance site is designed to prevent deadlock caused by the breakdown of AGVs; in Figure 2c the maintenance site situates at the centre of the system. Accordingly, three extra paths are designed to assure its accessibility to the AGVs that could fail at anywhere of the system. Based on the aforementioned designs, the PPN models for these three different layout configurations can be readily constructed by defining the movement directions of the AGVs. For example, the PPN for the configuration in Figure 2b is shown in Figure 3, where only one direction of movement is enabled, and the dotted arrows represent the information flows coming from other CPNs. The tokens in the figure represent AGVs. Once the required action is completed a token from other CPN enables the corresponding transitions.

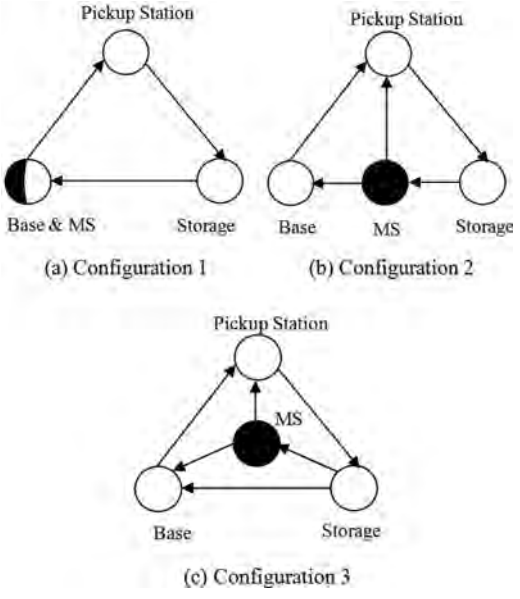


Figure 2. Three layout configurations.

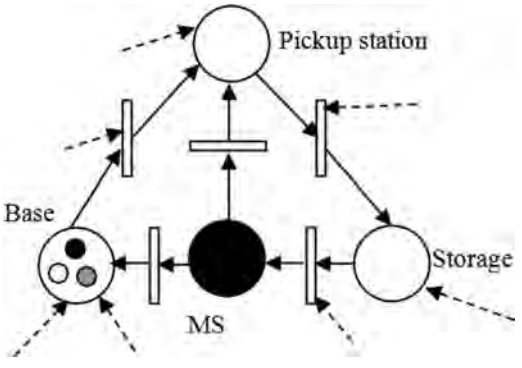


Figure 3. PPN for the configuration in Figure 2b.

Then, the AGV token with the same colour can move to the place of the next station.

3.2.2 *Corrective Maintenance Petri Nets (CMPN)*

Once the failed AGVs are recycled and towed back to maintenance site, the corrective maintenance will be implemented immediately if the maintenance engineers are available to work on the failed AGV. However, once the maintenance engineers are unavailable, the failed AGVs will have to queue. On completing the corrective maintenance, the recovered AGV will be assumed having a perfect condition as a brand new one does. In the meantime, the maintenance engineer who undertakes

the repair of this AGV will be released. They will become available to undertake the repair of other failed AGVs. In the model, a normal distribution function is employed to describe the repair time of the failed AGVs. The developed CMPN model is shown in Figure 4. Once a token exists in both 'Failed AGVs recycled' and 'Available engineering' places, the token will be produced in 'under repair' place. Following the repair process, the AGV will be back to the healthy state. This will be indicated by a token produced in the 'Up' place.

3.2.3 *Periodic Maintenance Petri Nets (PMPN)*
A PMPN model that considers periodic maintenance has been developed and is shown in Figure 5. Likewise, in this model the recovered AGVs are assumed having perfect health condition as a new one does.

It is worth noting that in the model shown in Figure 5 the three transitions with different colours indicate the failure time of the three AGVs in the system. For the simplification, a simple corrective maintenance policy is taken in this research, i.e. all AGVs in the system will receive periodic maintenance in spite of their actual health condition. Moreover, the corrective maintenance will last only for 2 days regardless of the actual condition of the AGVs. For example, in Figure 5 it is assumed there are *m* AGVs in healthy condition and n AGVs in faulty condition. The healthy AGVs are in 'Up' place and faulty AGVs are in 'Failed AGVs recycled' place. Regardless the actual health status, all AGVs will receive periodic maintenance. Therefore, there will be *m+n* tokens in 'Periodic Maintenance' place. Accordingly, all AGVs in the system will not start to work until the 2-day period of corrective maintenance expires. On the expiry of the period of corrective maintenance, all AGVs in the system are assumed to have perfect health condition as a brand new one does.

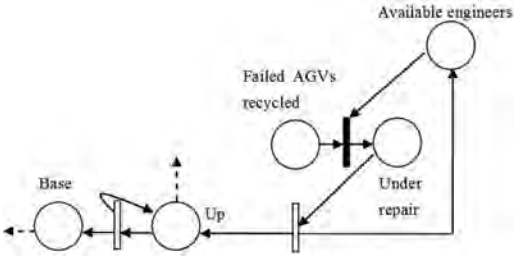


Figure 4. CMPN model.

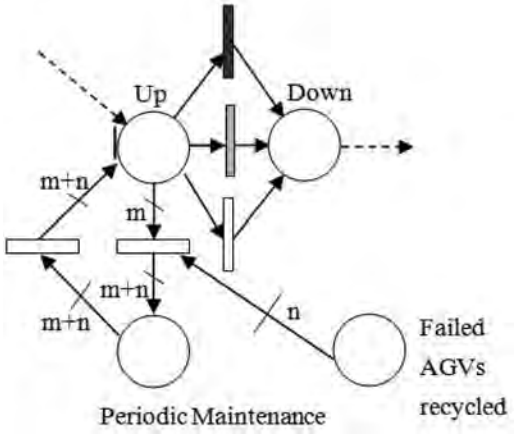


Figure 5. PMPN model.

3.3 *Simulation results*

By integrating the above CPN models, a more comprehensive model can be readily obtained, which not only considers the specialities of the layout configuration but also considers the maintenance processes of the AGVs. In order to verify the model and investigate the influences of different maintenance strategies on the operational performance of a multi-AGV system, an algorithm has been developed to simulate the comprehensive model, the input variables of which include the failure rate and repair rate of all AGVs, the time taken to perform periodic maintenance, and phase lengths that are required by the AGVs to deliver assigned tasks.

Firstly, the influence of different layout configurations on the recycle time of failed AGVs is investigated. In the layout configurations described in Figure 2b and c, separate maintenance sites are designed. Such a design significantly reduces the risk of conflict and deadlock and therefore improves the efficiency of the recycle process, although with the cost of extra space and extra routes to enable the operation of such a design. The simulations considering all three types of layout configurations are performed and the corresponding recycle time calculation results are listed in Table 2. From the table, it is found that when the maintenance site is placed in the centre (see Figure 2c), the recycle time will be the minimum.

Subsequently, the influence of different maintenance strategies on the performance of the multi-AGV system is investigated. Assume the operation time of the system is 10 hours per day, the corresponding simulation results obtained for the layout configuration illustrated in Figure 2b are listed in Table 3. In the table, the number of missions completed is employed as a criterion for performance assessment.

From Table 3, it is found that if without applying any maintenance strategy within the period of

Table 2. Recycle time.

Location indicated by	Recycle time (hours)	Extra space (unit)	Length of extra route required (unit)
Figure 2a	0.132	0	0
Figure 2b	0.128	1	$\sqrt{3}/2$
Figure 2c	0.101	1	$3\sqrt{3}/4$

Table 3. Number of completed missions.

T	P	N1	N2
7 days	0.03	11518	11840
1 month	3.93	12840	15264
3 months	36.32	9372	15972
6 months	77.34	6084	16142
12 months	98.06	3280	16234

Note: T—Time interval of periodic maintenance; P—Percentage of AGVs failed within the time interval if there is no maintenance (%); N1—Number of missions completed per year with periodic but without corrective maintenance; N2—Number of missions completed per year after taking both periodic and corrective maintenance.

12 months, 98% of AGVs will fail after completing 3280 missions. This fully highlights the added value and the necessities of conducting appropriate maintenance to the AGVs during their service life. Moreover, the larger values of N2 than the corresponding values of N1 prove that the corrective maintenance can actually enhance the performance of the multi-AGV system, i.e. the corrective maintenance can help to keep long-term high efficiency of the system, although it could cause extra financial and labour costs.

4 OPTIMISATION OF MAINTENANCE STRATEGY

4.1 *Genetic algorithm*

The results obtained from the CPN simulations can be used as factors for optimising the maintenance strategy of the AGV system. Since the resultant optimal maintenance strategy is desired to lead to a cost effective and time efficient operation of the multi-AGV system, the optimisation considered in this research becomes a typical multi-objective optimisation problem. In the paper, Genetic Algorithm (GA) is employed to carry out the optimisation. Nowadays, the GA has been regarded as one of the most popular tools to solve this kind of multi-objective optimisation problem attributed to its powerful capability of conducting optimisation in a global range regardless of initial conditions and other derivative factors. Inspired by the biological evolution of living species, GA was first introduced by John Holland in 1970s (Holland, 1975). GA has been well applied for solving the scheduling and dispatching problems to AGV systems. For example, Reddy and Rao applied GA to minimise the makespan, mean flow time and mean tardiness at the same time (2006). A GA based simulation approach was proposed to find the optimal dispatching rules in complex environments (Chang et al., 2013).

To implement the GA optimisation, an initial population of individuals (also known as chromosomes consisting of genes) will be generated. The fitness of each chromosome is evaluated subject to the predefined objective functions. By selecting pairs of parents in the population, new chromosomes or children can be generated. This is known as crossover. The chromosomes with the higher fitness are more likely to be selected so that their genes can be passed on with higher probability. A mutation might also be involved to prevent early convergence of the solution. Through repeating such a process, the chromosomes with larger fitness values can be obtained until an optimal solution is reached.

4.2 *Fitness functions*

Following this idea, a GA program is developed in the research to optimise the multi-AGV system. The flowchart of the GA program is shown in Figure 6. The parameters used in the calculation are listed in Table 4. The following two objective functions are defined to optimise the system design:

Objective function 1: The maximum number of missions completed within a given time

$$Mission = \max\left(N_m \cdot N_p - T_{rc} \cdot N_f / T_m\right) \tag{1}$$

Objective function 2: The minimum cost for completing the missions

$$Cost = \min\begin{pmatrix} N_a \cdot N_p \cdot C_p + N_f \cdot C_c + N_e \cdot C_e + \\ C_{ms} + N_m \cdot C_a + L_r \cdot C_r + C_{lm} \end{pmatrix} \tag{2}$$

where $N_P = 365/T$ is based on the assumption that there are 365 days in a year.

Based on the aforementioned two objective functions, a fitness function is developed as

$$fitness = \frac{Mission}{Cost} \tag{3}$$

The maintenance strategy is optimised subject to:

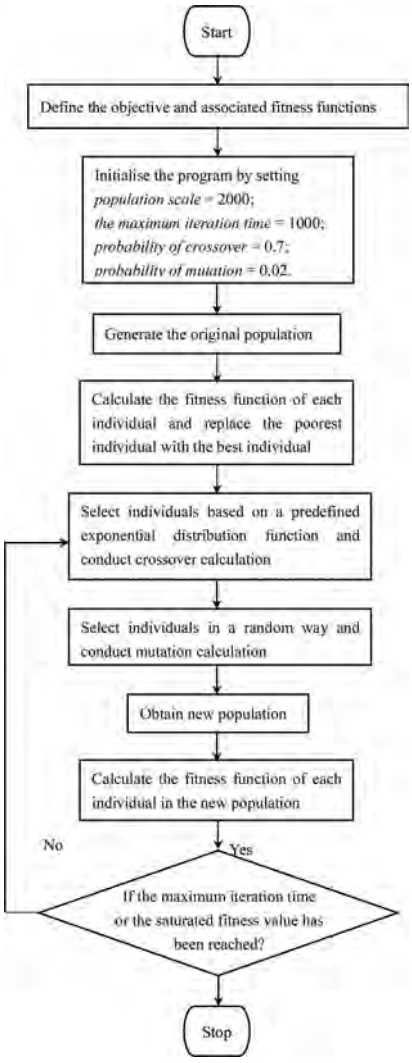


Figure 6. Flowchart of the GA based optimisation program.

$$number\ of\ mission\ completed \geq 10000 \quad (4)$$

$$probability\ of\ all\ AGVs\ failed \leq 0.1 \quad (5)$$

where, equation (4) means that the number of missions required to complete within one year is not less than 10,000; equation (5) means that the probability of all AGVs fail is not larger than 10%.

4.3 *Selection*

In the program, an exponential function is specifically defined for simulating the 'survival of the fittest' principle in natural evolutionary process. For the i-th individual, its probability P_i being selected for participating in GA crossover calculation can be expressed as:

$$P_i = \frac{e^{w(f_i - f_{\min})}}{\sum_{j=1}^{N} e^{w(f_j - f_{\min})}} \times 100\% \quad (6)$$

Table 4. Parameters used in GA program.

Parameters	Symbol	Value
Number of AGVs	N_a	3
Operation cost of an AGV to complete a single mission	C_a	8
Business costs of maintenance site per year	C_{ms}	10000 – with corrective maintenance 5000 – without corrective maintenance
Land cost for maintenance site per year	C_{lm}	1000 – Share site with AGV base 5000 – Separate site
Number of missions competed per year	N_m	See the values of N1 and N2 in Table 3
Time interval of periodic maintenance	T	See the values of T in Table 3
Periodic maintenance cost per AGV	C_p	400
Recycle time	T_{rc}	See the values of recycle time in Table 2
Average time to complete a mission	T_m	0.66
No. of maintenance engineers on site	N_e	1
Cost of one Engineer in a year	C_e	25000
Total number of failures occurring in the system with corrective maintenance per year	N_f	14 (results obtained using PN)
Average cost for conducting corrective maintenance of an AGV failure	C_c	200
Extra route length	L_r	See the values of Length of extra route required in Table 2
Cost of per unit length extra route	C_r	1000

where N denotes the size of population scale; f_i is the fitness of i-th individual; f_{min} is the fitness of the poorest individual; and w is a constant for controlling the efficiency of population evolution. It is worth noting that the larger the value of w, the more efficient the evolution tends to be. But it should be aware that too large a value of w would lead to risk of failure to obtain global optimisation results. In this research, w is taken to be 100.

4.4 *Coding*

It should be noted that there are three major factors, namely the period of periodic maintenance, the system configurations and the adoption of corrective maintenance. Their values are obtained from the CPN simulations. Hence the variation ranges of these factors need to be controlled by constraints. These three parameters are coded into binary numbers and then connected together to create a single 'chromosome'.

4.5 *Crossover operator and mutation operation*

The crossover operation is applied to two randomly selected chromosomes with the crossover rate of 0.7. A one-point crossover is adopted with an illustrative example shown in Figure 7.

The alternating position can be chosen at any point within the chromosomes. By combing two sets of genes from both parent generations, an offspring chromosome can be produced. The operation of mutation illustrated in Figure 8, maintains genetic diversity of the population and prevents the solutions trapping to the local best. It is also implanted with a fixed mutation rate of 0.02.

4.6 *GA results and discussion*

Using the developed GA program the location of maintenance site and the maintenance strategies of the multi-AGV system are optimised through integrating the two objective functions into one fitness function, i.e. the unit mission cost shown in equation (3). To illustrate the effectiveness of the GA optimisation, a numerical example has been taken. By applying the parameters defined in Table 4 to the program, the population starts to evolve gradually. The resultant variation tendency of average fitness against the number of evolution times is shown in Figure 9.

From Figure 9, it is found that after the population is evolved for about 200 times, the average fitness reaches a saturated value. That means the optimal design of the multi-AGV system is achieved through 200 times of evolution calculations. The optimised results are listed in Table 5.

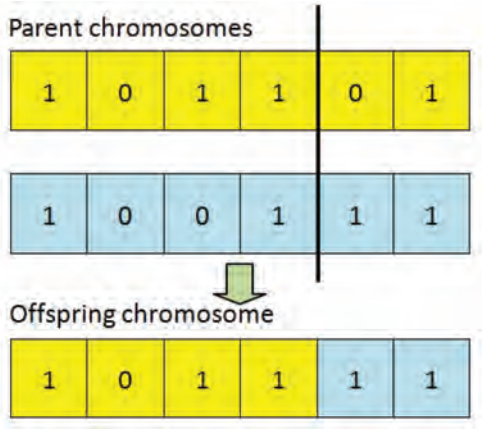


Figure 7. One-point crossover operator.

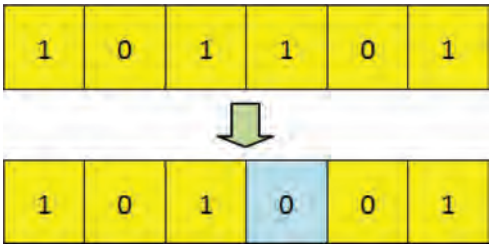


Figure 8. Mutation operator.

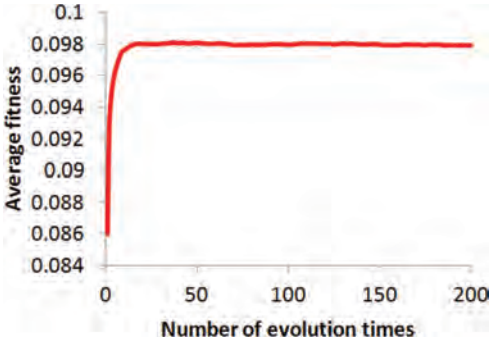


Figure 9. Evolution of GA population.

Table 5. Optimal results obtained from GA.

With corrective maintenance?	Yes
Location of maintenance site	In the AGV base
Time interval of periodic maintenance	12 months
Total cost (£)	164872
Mission completed per year	16231

From Table 5, it is found that

- The corrective maintenance is indeed essential for maintaining the long-term high efficiency of a multi-AGV system;
- Arrange the maintenance site to share the same place with the AGV base will save the cost on land, therefore result in the minimum unit mission cost;
- The positive influence of periodic maintenance on improving the performance of the system cannot be demonstrated if the ageing issue of the AGVs is not taken into account in the optimisation.

5 CONCLUSIONS

In order to develop a feasible and efficient approach to optimising the design, operation, and maintenance of a multi-AGV system, the CPN simulation models and the GA-based optimisation approach are developed in this research. From the research results described above, the following conclusions can be drawn:

1. The combined use of CPN and GA has been demonstrated an effective approach to assessing the performance of multi-AGV systems;
2. This hybrid approach enables the prediction to the optimal time interval of periodic maintenance and the assessment of the influence of correct maintenance on system efficiency;
3. The optimisation of the location of maintenance site and maintenance strategies can be skilfully converted to be a simple single objective optimisation problem with the fitness function of unit mission cost;
4. The corrective maintenance is an effective measure to maintain the long-term high efficiency of the system, although it may lead to extra maintenance costs.

Future work of this research will focus on dealing with more complex AGV systems.

ACKNOWLEDGEMENT

The work reported in this paper aligns to the working being researched as part of the EPSRC grant EP1K01413711.

REFERENCES

Aized, T. 2009. Modelling and performance maximization of an integrated automated guided vehicle system using coloured Petri net and response surface methods, *Computers and Industrial Engineering* 57(3): 822–831.

Chang, X. & Dong, M. & Yang, D. 2013. Multi-objective real-time dispatching for integrated delivery in a Fab using GA based simulation optimization, *Journal of Manufacturing Systems* 32(4): 741–751.

Ebben, M. 2001, *Logistic control in automated transportation networks*, University of Twente, Enschede, PhD thesis.

Fazlollahtabar, H. & Jalali Naini, S.G. 2013. Adapted Markovian model to control reliability assessment in multiple AGV manufacturing system, *Scientia Iranica* 20(06): 2224–2237.

Fornlöf V. & Galar D. & Syberfeldt, A. & Almgren, T. 2016. RUL estimation and maintenance optimization for aircraft engines: a system of system approach, *International Journal of System Assurance Engineering and Management* 7(4): 450–461.

Holland, JH. 1975. *Adaptation in natural and artificial system: an introduction with application to biology, control and artificial intelligence*: The University of Michigan Press.

Jensen, K. 2015. *Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use*; Springer: Heidelberg, Germany.

Lei, Y. & Liu, J. & Ni, J. & Lee, J. 2010. Production line simulation using STPN for maintenance scheduling, Journal of Intelligent Manufacturing 21(2): 213–221.

Nishi, T. & Maeno, R. 2010. Petri net decomposition approach to optimization of route planning problems for AGV systems, *IEEE Transactions on Automation Science and Engineering* 7(3): 523–537.

Petri, C.A. 1962. *Kommunikation mit automaten,* PhD thesis.

Reddy, B. & Rao, C. 2006. A hybrid multi-objective GA for simultaneous scheduling of machines and AGVs in FMS, International *Journal of Advanced Manufacturing Technology* 31(5–6): 602–613.

Rene, D. & Alla, H. 1994. Petri Nets for Modeling of Dynamic Systems A Survey, *Automatica* 30(2): 175–202.

Smith, D. & Babb, A. 1973. *Maintainability engineering*, Pitman Publishing.

Trenkle, A. & Seibold, Z. & Stoll, T. 2013. Safety Requirements and Safety Functions for Decentralized Controlled Autonomous Systems, in 2013 *XXIV International Conference on Information, Communication and Automation Technologies (ICAT)*.

Tuan, Le-Anh & M.B.M. De Koster RSM. 2006. A review of design and control of automated guided vehicle, *European Journal of Operational Research* 171(1): 1–23.

Vis I.F. 2006. Survey of research in the design and control of automated guided vehicle, *European Journal of Operational Research* 170(3): 677–709.

Wu N. 1999. Necessary and sufficient conditions for deadlock-free operation in flexible manufacturing systems using a colored Petri net model, *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* 29(2): 192–204.

Wu, N. & Zeng, W. 2002. Deadlock avoidance in an automated guidance vehicle system using a coloured Petri net model, International *Journal of Production Research* 40(1): 223–238.

Yan, R. & Jackson, L.M. & Dunnett, S.J. 2017. Automated guided vehicle mission reliability modelling using a combined fault tree and Petri net approach. The International *Journal of Advanced Manufacturing Technology* 92 (5–8): 1825–1837.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A modelling methodology for the assessment of preventive maintenance on a compressor drive system

Y. Zhang, A. Barros & A. Rauzy
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technolgy, Trondheim, Norway

E. Lunde
Research and Technology, Statoil ASA, Trondheim, Norway

ABSTRACT: Huge rotary machines are commonly used in oil and gas processing plants for separation, compression and boosting. Their reliability is of high importance to avoid operation downtime and production loss. In this paper, we present a modelling methodology, based on the AltaRica 3.0 modeling language and stochastic simulation, to assess the average production level of a compressor drive system. This system consists of six trains, where each of them contributes to one sixth of the total production capacity. It runs under two operation modes (full and reduced capacity) corresponding to seasonal demand periods (winter and summer). The problem at stake is to design a model at system level that captures the various degradation processes, monitoring policies, and maintenance rules involved in the system under study. The aging of units is represented by means of multiple degradation levels. Given units information provided by monitoring and inspection, preventive and corrective maintenance interventions are decided locally to each unit. Performance indicators such as the cumulative production and production loss over a certain mission time can then be assessed. This paper contributes to the development of engineering models for maintenance assessment based on framework and patterns designed to architecting some typical oil and gas systems.

1 INTRODUCTION

Huge rotary machines are commonly used in oil and gas processing plants for separation, compression and boosting. Ensuring a high reliability of these machines is of primary importance to avoid operation downtime and production loss. As of today, this high reliability is achieved thanks to robust-by-design machines, and by rigorous preventive maintenance policies (DNV-GL-RP-002 2014, API-RP-7L 2006). Nevertheless, rigorous preventive maintenance comes with a high cost. This is the reason why the oil and gas industry is currently moving to the so-called condition-based approach (Gustavsson and Eriksen 2005, Markeset et al. 2013). In order to deploy a condition-based maintenance policy, one needs to assess its potential benefit and risk over more traditional approaches.

In this paper, we present the results of a preliminary study we made on a compressor drive system. This system consists of six trains, where each of them contributes to one sixth of the total production capacity. The system runs under two operation modes (full and reduced capacity) corresponding to two distinct seasonal demand periods (winter and summer). Maintenance interventions have to be scheduled during the low demand season where some of the trains can be stopped while still fulfilling the demand.

We present here the modelling methodology we used, together with modelling patterns and simulation experiments. This methodology relies on the AltaRica 3.0 language Rauzy (2008) and Prosvirnova (2014) and stochastic simulation. It aims at developing models that make it possible to answer questions like "Will the system survive in the coming winter without loss of production? How much can we gain/lose by changing inspection interval of this component or group of components?" and so on. The key performance indicator is the expected production loss due to failures and maintenance operations over a given time period.

The object-oriented language AltaRica 3.0 makes it possible to handle the various modelling challenges at stake: It provides mechanisms to represent faithfully the various degradation processes, monitoring policies, and maintenance rules involved in the system under study. It makes it possible to represent implicitly very large state spaces;

It facilitates information propagation through the network of components and therefore the calculation of key performance indicators; It supports the reproduction of basic patterns in order to develop models of large systems by assembling seamlessly models of components; …

The main purpose of this paper is to present this modelling framework and to illustrate its application. We report here the results of a number of experiments we performed on the model in order to run what-if scenarios. We show the various possibilities to refine and optimize inspection/maintenance policies as well as to asses their impact on the production.

The use of this study is not limited to onshore installations with given maintenance problems. It helps to improve the knowledge and decisions making with a tool for future subsea installations, as there will be more and more seabed compression with equivalent and even more complex maintenance problems.

The remainder of this paper is organized as follows. Section 2 presents state of the art of Preventive Maintenance PM in compressor drive system. Section 3 describes the use case of compressor train system and its assumptions. Section 4 introduces the modelling methodology and design of components and system. Section 5 shows numerical experiments for assessing various inspection and maintenance policies. Section 6 concludes the work and discusses future research.

2 STATE OF THE ART

Compression is a common practice in Q&G processing technology. Due to the complex configuration of a compression drive system, the control over such a system turns to be of paramount importance. This section is thus dedicated to the state of art investigation on compressor drive system which is normally composed of VSD, pump, compressor, valves and so on. Findings from for instance (Andersen et al. 2006, Eriksson and Staver 2010, Eriksson and Konstantinos 2014) and interviews with industrial partners are summarized as a preparation for our model.

2.1 *Health indicator and degradation modelling*

A typical compressor drive system is arranged with multiple trains in parallel where each train consists of a number of components (e.g. VSD, gear, compressor, valve etc.). Each of them is subject to different aging processes (e.g. fouling, wear, harmonics, unbalance) which could be revealed in several condition monitoring sources as described previously. An option to aggregate these sources to formulate a single health indicator is Technical Condition Index (TCI) (Nystad 2008, Nystad & Rasmussen 2010). It is defined to represent the degree of degradation with respect to production availability. The overall TCI value of a compressor, for instance, is aggregated from bottom-level TCI for each degradation symptom revealed by its monitoring source. Its aging measurement includes but not limited to: vibration, seal wear, bearing temperature, compressor efficiency and so on. Each of them is assigned a weight based on expert judgement about its criticality, and it is further merged upwards in the hierarchy for the total TCI of the compressor. Besides, previous maintenance actions and their impact (mainly missing or incomplete working log) can be included in the model as left truncated censoring data. Given all these inputs, a parametric hazard regression model Weibull PHM model is applied in Nystad & Rasmussen (2010) to estimate parameters in the degradation model by maximum likelihood method. Once the parameters are available, the expected Remaining Useful Lifetime (RUL) given a specific time point is tractable with certain confidence.

In this paper, we rely on an alternative solution proposed by Moholt (2016). The health indicator is a discrete variable with a finite number of possible values from new to fail. It is used by guidelines to assess the health of an equipment in a more qualitative way. Such a simplified model is aligned with the amount of data that are currently available for the case under study in this paper. Moreover, it is more adapted to our modelling framework based on discrete states.

2.2 *Intervention and maintenance modelling*

Often maintenance programs are established by project teams by each companies. They propose alternative solutions, demonstrate the advantages/disadvantages for the maintenance strategies and discuss an optimal solution with manufactures. Most components are under calendar-based maintenance together with condition-based maintenance program.

For calendar-based maintenance, the frequency, content, duration and required preparations for maintenances vary in the user manual by equipment type, size and application. The components are analysed in design phase and assigned maintenance levels (L1-L4) according to their criticality (ABB 2013). Periodic maintenance is implemented according the plan.

For condition-based maintenance, there is short term preventive maintenance plan scheduled for the next low demand season based on the input from condition monitoring and inspection. Such decisions are mainly based on experience and

expert judgement. Notice that for both maintenance interventions, there is no possibility to react immediately on any degradations/failures in the systems due to the time for preparing maintenance kits.

A risk based simulation approach (RBI) has been applied to develop maintenance strategy for subsea systems in Ormen Lange field (Gustavsson & Eriksen 2005). The lifetime scenario of the system is represented by discrete degradation/failure events happen for each component. The cost model quantifies intervention cost and production loss due to unexpected failures for the mission time. Then a variety of maintenance strategies are fed into the model to calculate desired performance indicators (e.g. average lost production, average repair cost) and help to assess the impact of different maintenance strategies for the subsea systems. However, it is not clear how this model is implemented technically and whether the modelling language and simulation tool are open to external users.

In our paper, we illustrate our method to model maintenance planning on a system-size compressor drive system. The model presented in this paper is authorized by integrated modelling environment AltaRicaWizard in the framework of OpenAlta Rica project. We show the possibilities to develop various maintenance strategies in the model, and demonstrate the loss/gain of these alternatives.

3 USE CASE: COMPRESSOR TRAIN

3.1 *System description*

We focus on 6 electrical trains that are used to compress the gas. Each of them contributes to one sixth of the total production capacity. Each compressor drive system consists of, from the left: a Variable Speed Drive (VSD), a Motor (M), a Gear box (G) and the compressor (C), see Figure 1.

During winter (6 months), full capacity is required and all the 6 trains are supposed to be used. During the summer, only part of the total capacity is required, for instance 1/2. Then only 3 trains are needed. The switch from full capacity to reduced one and then backwards is operated once a year (e.g. 01/10 and 01/04). At system level, once one train is failed, it is revealed naturally by production capacity. Similarly, at train level if one of the component is degrading, it reduces production capacity immediately.

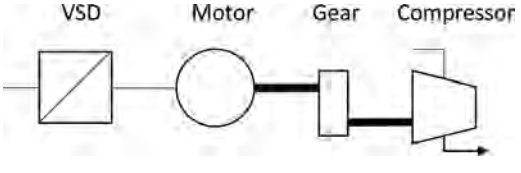


Figure 1. Diagram of one compressor train.

3.2 *Assumptions of each unit*

3.2.1 *VSD and gear*

VSD and Gear are much more reliable than motor and compressor. They have only binary states: Working (W) and Failed (F) (Figure 2). The failure time follows exponential law with low failure rate.

Both components are continuously monitored. In addition, they are easy to repair so the maintenance time is short. Only corrective but no preventive maintenance is planned on these units. The intervention can occur at any time during operation of the system.

Concerning production, it depends on working state of the unit and production flow plug into it.

3.2.2 *Compressor*

Compressor is subject to indirect continuous monitoring on the degradation process. It has four states: Working (W), Degraded 1 (D1), Degraded 2 (D2) and Failed (F). The time to change between these states follows exponential laws where the rate is respectively λ_{d1}, λ_{d2} and λ_f (Figure 3). In W state, the compressor runs at full capacity 100%; in D2 state, the compressor has fouling and it consumes more power to maintain the full production capacity; in D2 state, fouling is accumulated and compressor capacity decreases below an acceptable threshold 80%; in F state, the compressor cannot operate any longer and its capacity sinks to 0%.

When compressor reaches its D1 state, a preventive minor maintenance (e.g. cleaning) is arranged; when it reaches D2/F state, a spare part is ordered and the preventive/corrective major maintenance (e.g. replacement) will be implemented. The duration of minor and major maintenance is respectively δ_{mn} and δ_{mj}. Delay time to prepare corresponding maintenance interventions is ρ_{mn} and ρ_{mj}.

The actual production of the compressor depends on its degradation level, the operation phase and also the input production passing to it.

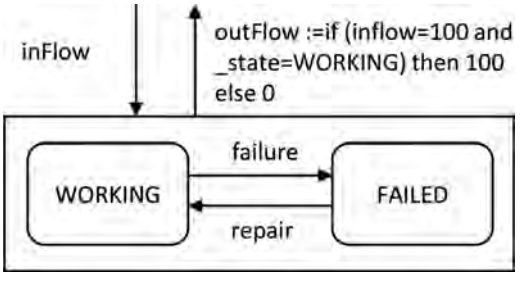


Figure 2. Automaton for VSD and Gear.

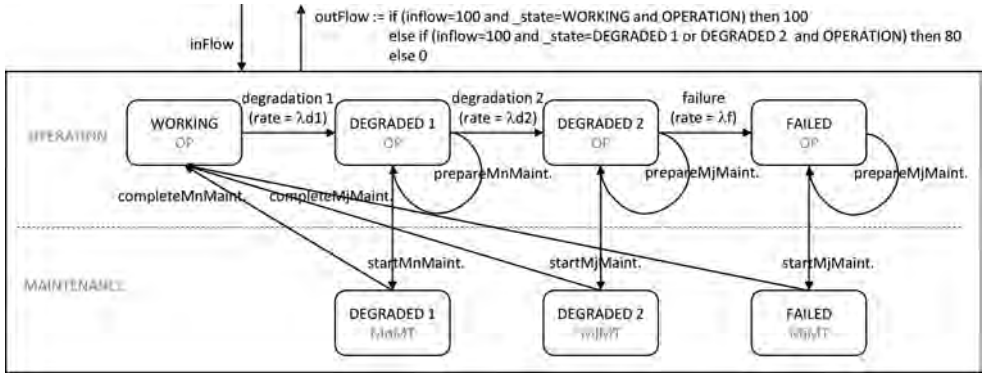


Figure 3. Automaton for compressor.

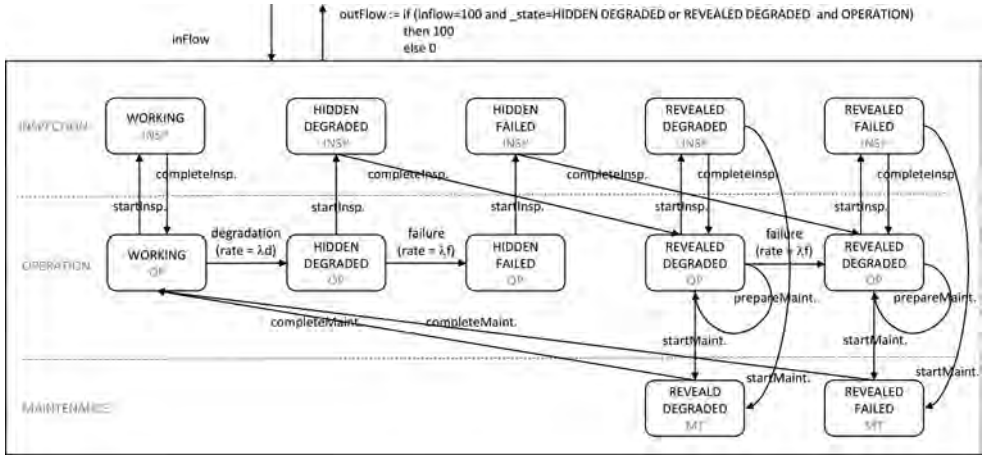


Figure 4. Automaton for motor.

3.2.3 *Motor*

Motor is under periodic inspection. It has six states: Working (W), Hidden Degraded (HD), Revealed Degraded (RD), Hidden Failed (HF), Revealed Failed (RF) (Figure 4). A periodic inspection is perfect and it reveals all the hidden degradations and failures. The inspection occurs every τ and it lasts for Δ unit time. After detection of degradation (RD), the condition can deteriorate further to a failure (RF).

There could be discrepancy between observed state and actual state of the unit. However, the maintenance planning is made based on observed state of the unit. In W/HD/HF, no action is taken; in RD/RF, a spare unit is ordered and the proactive/corrective maintenance is implemented which lasts for δ. The delay to prepare the intervention is ρ.

The actual production of the motor depends on its working state, the operation phase together with the upstream production passing to it.

As a first study of the compressor drive system, we simplify the model by saying that we have enough maintenance teams. We do not consider system-level diagnosis neither maintenance planning. For the mentioned units, any maintenance and inspection stops the production of the intervened train, but do not interfere behaviours of the units on the same line or the other trains. All the components are independent of each other and interventions are decided locally.

4 MODELLING METHODOLOGY

The modelling formalism that we use here is Alta Rica 3.0 and its underlying mathematical framework, Guarded Transitions Systems (GTS). For formal presentation see (Rauzy 2008, Prosvirnova 2014). The formalism can handle complex systems with four main Modules as explained in (Zhang et al. 2017):

1. description of individual behaviours of units
2. description of the actual state of the system
3. diagnosis on the state of the system
4. description of maintenance planning and actions

In our case, we simplify the situation by saying that components are independent and maintenance decisions are made local. Therefore, there is no system-level diagnosis netiher decision making. We only use part of the framework (Module 1, 2) while Module 4 can be further embedded into Module 1. The model constituting two parts: modelling units and modelling system are explained as follows:

4.1 *Modelling units*

The first part describes behaviours of each unit. The purpose is to translate finite state automata as Figures 2, 3 and 4 into AltaRica language. GTS of each unit describes indigenous events that happen locally by itself, for example degradation and failure. Meanwhile, it can embed foreign interventions that posed on top when the local condition or clock reaches certain trigging point, for instance start maintenance and start inspection. To clarify the mechanism, we use motor (Figure 4) as an example. The AltaRica code implementing the GTS for the motor is given in Figure 12.

The class of a generic inspected unit is designed and named as `Motor` (line 5). The motor can be in one of the following states: Working (W), Hidden Degraded (HD), Revealed Degraded (RD), Hidden Failed (HF) and Revealed Failed (RF) (line 1). It experiences three alternating stages in its life cycle: Operation (OP), Inspection (INSP) and Maintenance (MAINT) (line 2). In addition, it runs under winter and summer profile (line 3). Line 10–16 assign names to events and declare their duration by keyword `delay`.

The definition of a `transition` (line 20–33) starts with the name assigned to it, then comes several pre-conditions of the transition with are connected by logic operators `and,or`. After a transition sign comes the final effect of the event, e.g. state or phase values are modified. Under such mechanism, the fireable events and their transitions mimic the behaviour of the component as time elapses.

Figure 5. A map of maintenance optimization scenarios.

According to the assumptions, motor degrades and fails only in operation phase. Therefore the state changes are preconditioned with `_phase==OP` (line 20–22). The initiation of an inspection follows fixed interval. Upon completion, a hidden degradation or failure is detected and meanwhile a clock starts to count the time for maintenance preparation (line 26–29). The counting, however, does not stop further degradation of the motor, so the event `prepareMaint` is defined by state indicator `_clock` independent of motor `_state` (line 30). Once the preparation is ready, all the maintenance resources are ready for use but we have to check that it is the low demand season (e.g. summer) to launch an actual campaign. If the `_season==SUMMER`, the maintenance starts immediately and the completion brings state back to working and resets clock again (line 31–33).

The assertion part (line 35–36) describes how to update flow variables after each transition firing. The production of the motor is 100% when it is in operation and not failed, otherwise it produces nothing.

In summary, classes of generic automata can be designed for each unit (VSD, Gear, Motor, Compressor) and season demand following the same routine.

4.2 *Modelling system*

The codes in section 4.1 defines actually several classes of continuously monitored (VSD, Gear, Compressor) and periodically inspected units (Motor) that can be reused for many times. Instantiations of these classes provide the basic elements for constructing a class or prototype of a bigger system. For instance, one train of a CDS can be described in Figure 10.

The code declares a class of `Train` which consists of two instances of the basic `RepairableUnit`, one instance of `COMPRESSOR` and one instance of `MOTOR` (line 2–5). Similarly, for a complete CDS with 6 identical trains, the class `Train` can be recalled for 6 times (`Train T1, Train T2,...`) to instantiate the real structure. The production of one train is determined by input and output flows passing through the series of the units. Namely, the input of a unit is plugged, by the equation `sign:=`, into the output of the precedent unit (line 9–13). Therefore, the equation means that the production of one train depends on the upstream production and availability of each unit in series.

4.3 *Performance indicators*

In practice, maintenances have a cost, just as shutdowns of systems due to failures. It is however very difficult to get realistic figures for these costs, because they depend on too many factors. More obviously, there is a discrepancy between the production expectation and the actual production of the system throughout the mission time. Here, we consider the average production and the loss of the system (per unit time) as relevant performance indicators. The GTS representation of these indicators are shown in Figure 11.

The defined observers are real numbers. Potential capacity of the CDS is normalized summation of each train (line 8–9). It is then compared with the actual demand according to the season. If the potential capacity satisfies the demand, then there is no production loss; otherwise the production loss equals the demand minus potential capacity of the CDS.

5 NUMERICAL EXPERIMENT

The mentioned modelling methodology is applied to the use case of CDS. A set of illustrative data (Table 1) is fed into the model. The mission time is 87600 h (10 years). We set the maximum production capacity for all six working trains to 100 per hour. For 10 years that would add up to $8.76e+6$, and with nominal seasonal production profile in summer (*0.5) we get $4.38e+6$. We assume that at the beginning of simulation, the production capacity of each unit is 100% per hour and the operation starts from winter.

The GTS model can thus be able to run a variety of what-if scenarios and the result can be assessed by the stochastic simulator embedded in AltaRicaWizard. The map in Figure 5 shows possible experiments that we can simulate with the model. Due to scientific focus of this paper, only selected experiments are discussed in the following subsections.

Table 1. Input parameters for each unit.

Component	VSD	Motor	Gear	Compressor
λ_{d1}		$1.0e-6$		$1.0e-6$
λ_{d2}				$1.0e-5$
λ_f	$1.0e-7$	$1.0e-5$	$1.0e-7$	$1.5e-4$
ρ_{mn}		4380		2190
ρ_{mj}				4380
δ_{mn}	6	365	6	182.5
δ_{mj}				365
τ		730		
Δ		12		

5.1 *Inspection interval*

Figure 6 plots 1 million realizations of the process for accumulated production and loss versus different inspection interval values $730 \leq \tau \leq 11680$ (1 to 16 months). When inspection interval is less than 4 months, the production loss increases dramatically. This is because with too frequent inspections, operations are dominated by unnecessary shutdowns to check still functioning components. After this point, the total production approaches a stable high level and it reaches the peak value when τ is around 12 months. Notice that the curve is almost flat when τ is from 4 to 16 months. It implies that the total production may be not so sensitive to the inspection interval in such range given our assumptions and input parameters.

5.2 *Time horizon*

Figure 7 shows Monte Carlo simulation for accumulated production and production loss from 1 to 10 years with 1 million realizations at each year. As time elapses, the production and loss increase almost linearly. It quantifies the provided overall production and revenue reduction over a period of mission time and thus gives decision makers a look-ahead horizon of the production profile of the whole CDS system.

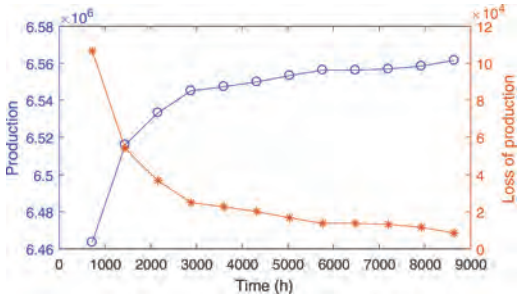


Figure 6. Production and loss versus inspection interval τ.

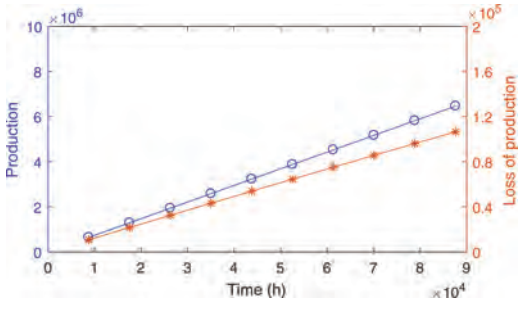


Figure 7. Production and loss versus mission time.

5.3 *Maintenance delay*

Figure 8 presents consequence of reducing minor PM preparation time (delay). Here, we assume that delay of major preventive/corrective maintenance interventions remain as it was ρ_{mj} and the changes only apply to ρ_{mn} of inspected components. From the figure we see that as minor preventive maintenance delay decreases from 10 to 1 year, the total production increases around 0.045%. The production loss with (ρ_{mn} < 87600) and without minor PM (ρ_{mn} = 87600) does not differ much. The link between instant reaction and increased production is obvious, but the gain of having maintenance resource immediately ready can be questionable. This is relevant for deciding an optimal spare parts strategy, when the cost of contracting maintenance service and storing spare parts has to be evaluated against the benefits of income.

5.4 *Seaon profile*

Figure 8 presents consequence of extending low demand period (i.e. summer) from 1 to 12 months. This time corresponds to the window when maintenance intervention is allowed. As the window extends, the system spends more time in reduced operation mode, and thus the total production decreases. Yet, production loss is reducing rather than increasing because the potential loss is compensated by relaxed requirement of production in the extended summer period. As such, we quantify the effect of a changing season profile on the accumulated production of the system.

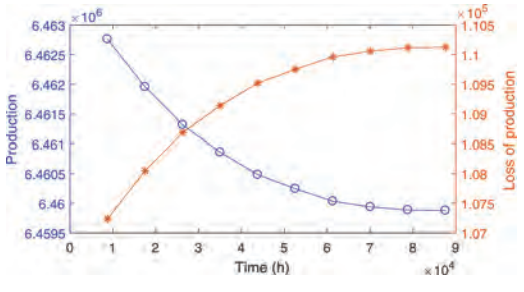


Figure 8. Production and loss versus maintenance delay.

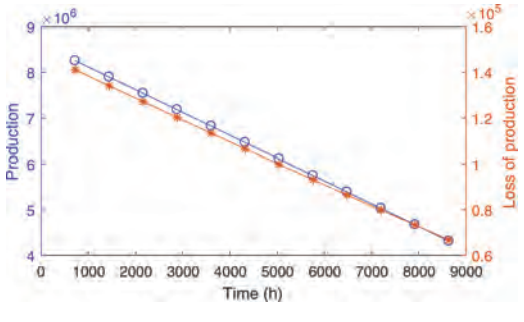


Figure 9. Production and loss versus low demand duration.

6 CONCLUSIONS

In this paper, we present a preventive maintenance model on a compressor drive system. The model relies on the formal modelling formalism AltaRica 3.0 and its mathematical framework guarded transitions systems. They provide mechanisms to represent state and transitions of local units, enable composition of multi-unit systems and allow information flows (e.g. production, observed condition) circulating through them. We illustrate how our modelling methodology handle the various modelling challenges in our CDS case, like multiple unit types, huge state space, monitoring together with inspection policies, multi-level maintenance actions and so on. We pose on top of the model different maintenance and inspection policies and perform numerical experiments using Monte Carlo simulations. The simulator, as a decision support tool, demonstrates how the health of components in the system affect the kind of intervention decisions need to make now/soon. The model calculates accumulated production and its loss in a certain mission time. It can provide practitioners motivation with respect to, for instance, intensifying/relieving work on condition monitoring so that interventions become efficient and necessary.

However, this work is very preliminary and there are several directions that can be further investigated.

Direction 1. Degradation profile We assume that components deterioration follow exponential laws with its respective degradation or failure rate lambda. However, if the plant data is available and tractable for more accurate estimation of component degradation, we may introduce multiple degradation states, or fit the data with other laws (e.g. Weibull, empirical distribution). In addition, if the operation mode changes, we may need to consider different degradation profiles (e.g. failure rate) under normal production, increased production, decreased production.

Direction 2. Season profile The current season profile is simplified with only two modes and constant production requirement for each mode. However, there could be fluctuations of production requirement given that it follows a mean within each season. For instance, in summer the production demands is 52% for 20 days, 65% for 40 days, and then 43% for 7 days and so on. Such

fluctuations in the seasonally demand profile can be introduced if a general abstraction of time dependent plant data is available.

Direction 3. Monitoring policy In practice, the inspection dates cannot be freely fixed, because they depend on many factors external to system such as the availability of the maintenance crew. Inspection may be destructive to the component. Certain inspections may take some predictable time and the system may be partly or totally shutdown during inspections. Meanwhile, other inspections do not interrupt operation. In addition, as we have 6 trains in parallel, the inspection can be conducted for all at the same time or only certain pieces at a time. These situations can be further introduced to the model to make it more realistic as implemented in O&G industry.

Direction 4. Maintenance policy It is important to sustain production all the mission time, especially in summer period when window is open for interventions. Therefore, in the occurrence of several degradations or failures in a system, we may need to decide when to react and which ones to intervene in priority. For instance, we can decide to maintain when we lose 3 compressors, or 2 compressors plus 1 motor, or wait for more failures and so on. When there are 1 working, 3 degraded, and 2 failed trains, we can decide to repair the 2 failed trains first and then repair the 3 degraded ones later to ensure highest possible production capacity. System level maintenance planning can be considered in these cases.

ACKNOWLEDGEMENT

This work was carried out as a part of SUBPRO, a Research-based Innovation Centre within Subsea Production and Processing. The authors gratefully acknowledge the financial support from SUBPRO, which is financed by the Research Council of Norway, major industry partners and NTNU.

REFERENCES

ABB (2013). Maintenance for motors and generators. IFEA Høstarrangement Borregaard.

Andersen, T.M., L. Thuestad, & T.A. Thorstensen (2006). Rapid: A new approach for improved regularity and decreased maintenance costs. In *OnePetro*, pp. 1–3. Offshore Technology Conference.

API-RP-7 L (2006). Procedures for inspection, maintenance, repair, and remanufacture of drilling equipment. Technical report, American Petroleum Institute, USA.

DNV-GL-RP-002 (2014). Integrity management of subsea production systems. Technical report, Det Norske Veritas (Norway) and Germanischer Lloyd (Germany), Norway.

Eriksson, K.G. & A. Konstantinos (2014). Subsea processing systems: Optimising the maintenance, maximising the production. In *OnePetro*. Offshore Technology Conference.

Eriksson, K.G. & K.O. Staver (2010). Large scale condition monitoring for a subsea gas compression system pilot. In *OnePetro*. Offshore Technology Conference.

Gustavsson, F. & R.H. Eriksen (2005). Developing an intervention, maintenance and repair strategy for ormen lange. In *OnePetro*. Society of Petroleum Engineers.

Markeset, T., J. Moreno-Trejo, & R. Kumar (2013). Maintenance of subsea petroleum production systems: a case study. *Journal of Quality in Maintenance Engineering 19*(2), 128–143.

Moholt, K. (2016). Insulation quality measured in partial discharge. Internal presentation at NTNU, Trondheim.

Nystad, B.H. (2008). *Technical Condition Indexes and Remaining Useful Life of Aggregated Systems*. Ph.D. thesis, Norwegian University of Science and Technology.

Nystad, B.H. & M. Rasmussen (2010). Remaining useful life of natural gas export compressors. *Journal of Quality in Maintenance Engineering 16*(2), 129–143.

Prosvirnova, T. (2014, November). *AltaRica 3.0: a Model-Based Approach for Safety Analyses*. Thèse de doctorat, École Polytechnique, Palaiseau, France.

Rauzy, A. (2008). Guarded transition systems: a new states/events formalism for reliability studies. *Journal of Risk and Reliability 222*(4), 495–505.

Zhang, Y., A. Barros, & A. Rauzy (2017). Assessment of operational performance of multi-unit production systems using altarica 3.0: a case study.

ANNEX

```
class Train
    RepairableUnit VSD;
    RepairableUnit Gear;
    COMPRESSOR Compressor;
    MOTOR Motor;
    Real inflow (reset = 100);
    Real outflow (reset = 100);
    assertion
        VSD.inflow := inflow;
        Motor.inflow := VSD.outflow;
        Gear.inflow := Motor.outflow;
        Compressor.inflow :=Gear.outflow;
        outflow := Compressor.outflow;
    ...
end
```

Figure 10. The AltaRica 3.0 code implementing the GTS pictured Figure 1.

```
block Plant
    ...
    Real capacity (reset = 100);
    observer Real Production = production;
    observer Real ProductionLoss = C.demand - production;
    assertion
        ...
        capacity := (T1.outflow + T2.outflow + T3.outflow + T4.outflow +
                     T5.outflow + T6.outflow)/6.0;
        production := if C.demand<capacity then C.demand else capacity;
end
```

Figure 11. The AltaRica 3.0 code implementing the GTS for performance indicators.

```
domain State{W, HD, RD, HF, RF}
domain Phase {OP, MAINT, INSP}
domain Season {WINTER, SUMMER}
domain Clock {STB, CALL, READY}
class MOTOR
    State _state (init = WORKING);
    Phase _phase (init = OPERATION);
    Season _season (reset = WINTER);
    Clock _clock (init = STB);
    event degradation (delay = exponential(lambda_d));
    event failure (delay = exponential(lambda_f));
    event startInsp (delay = Dirac(tau), policy = MEMORY);
    event completeInsp (delay = Dirac(Delta));
    event prepareMaint (delay = Dirac(rho));
    event startMaint (delay = Dirac(0));
    event completeMaint (delay = Dirac(delta));
    Real inflow (reset = 100);
    Real outflow (reset = 100);
    transition
        degradation: _state==W and _phase==OP -> _state := HD;
        failure: _state==HD and _phase==OP -> _state := HF;
        failure: _state==RD and _phase==OP -> _state := RF;
        startInsp: _phase==OP -> _phase := TEST;
        completeInsp: (_state==W or _state==RD or _state==RF) and
                      _phase==INSP -> _phase:=OPERATION;
        completeInsp: _state==HD and _phase==INSP -> {_phase:=OP;_state:=RD;
                      _clock:=CALL;}
        completeInsp: _state==HF and _phase==INSP -> {_phase:=OP;_state:=RF;
                      _clock:=CALL;}
        prepareMaint: _clock==CALL -> _clock := READY;
        startMaint: (_state==RD or _state==RF) and _phase!=MAINT and
                    _clock==READY and _season==SUMMER -> _phase := MAINT;
        completeMaint: _phase==MAINT -> {_state:=W; _phase:=OP;_clock:=STB;}
    assertion
        outflow := if inflow==100 and ((_state==W or _state==HD or _state==RD) and
                   _phase==OP) then 100 else 0;
end
```

Figure 12. The AltaRica 3.0 code implementing the GTS pictured Figure 4.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

A maintenance time estimated method based on virtual reality

Juan Wu, Dong Zhou & Pengyan Liu
State Key Laboratory of Virtual Reality Technology and System, Beihang University, Beijing, P.R. China

ABSTRACT: Traditional method of maintenance time estimation is based on the time regular obtained from large statistical data. However, this method requires physical prototypes, which is much more difficult during design and development stage. Moreover, to obtain accurate data, a large amount of experiments need to be conducted to calculate maintenance time and eliminate the error. A method is proposed in this paper, which uses maintenance platform based on virtual reality to avoid the necessity of physical prototype. In this method, based on maintenance simulation analysis under virtual reality environment, Methods Time Measurement (MTM) and compensation-based time prediction are integrated. Then, according to difference between the actual repair time and the estimated time from virtual simulation, time compensation model is built considering the factors of proficiency, fatigue and maintenance environment. Finally, a real case is used to validate the model. This method could give suggestions of the evaluation of quantitative maintenance factors in the early design stage of the product.

1 INTRODUCTION

Maintenance time is an important quantitative parameter in the analysis of maintenance. However, it is usually hard to get in the development and manufacturing stages. Traditional maintenance time estimation mainly through time-accumulated method (Griswold 1970) or similar product comparative time analysis (Pliska et al. 1987), which are based on massive data. Physical prototypes are necessary in traditional methods, while in the early stage of development and manufacturing it is hard to meet this necessity. Moreover, to obtain accurate data, a large amount of experiments need to be conducted to calculate maintenance time and eliminate the error.

Given these issues above, the main problems that traditional methods faced are the lack of physical prototypes and the large amount of experiments.

While the application of virtual reality in maintenance can fix these problems. Virtual prototypes are used in virtual maintenance, which avoids the use of physical prototype. And the analyst can make virtual simulation which contains all the motions of a maintenance task. This reduces the number of the experiments.

The application of virtual maintenance has been increased in early stage of development and manufacturing during the last few decades. Real-time immersive virtual environments, such as the CAVE (Cruz-Neira et al. 1993) and the Workbench (Cutler et al. 1997) have been used to evaluate the maintainability of virtual prototypes. A novel assembly optimization framework based on genetic algorithm was proposed in 2009 (Christiand et al. 2009).

Except of the use of virtual environment, a lot of studies analyze virtual simulation which consists of a series of virtual human motions. And many methods have been proposed based on these motions, including work force, methods time measurement, modular arrangement of predetermined time standard and so on (Genaidy et al. 1989, Kanai et al. 1996, Genaidy et al. 1990, Laurig et al. 1985, Dossett 1992, Laringa et al. 2002, Fischer et al. 1991, Wygant et al., 1993, Hoffmann et al. 1993). These methods classify different types of human motion and give the relevant motion-time principles. Except for those methods, maintenance time estimated based on system maintenance work procedure through Monte Carlo simulation (LIU et al, 2014). Similarly, maintenance time predicted through structural complexity metric model was proposed in 2014 (Owensby et al, 2014). While the average repair time could be also estimated by failure rate (Shen et al, 2017). While in this paper, author uses methods time measurement as the basic method to estimate maintenance time and gives some compensation principle.

The rest of the paper is organized as follow. Section 2 shows the overview of methods time measurement. Section 3 discusses the details of the compensation principle. Section 4 presents a case study for the implementation of the methodology. While section 5 gives a conclusion of the study.

2 OVERVIEW OF METHODS TIME MEASUREMENT

The maintenance process is consists of a series of human motions. While in methods time measurement, these motion are divided into three parts, including human work, posture adjustment and hand operations (Geng et al. 2014). The following paragraphs will discuss each part in detail.

2.1 *Human walk*

Human walk simply means the walk motion in the virtual maintenance, such as approaching the object. And the time of human walk is decided by walking distance and the weight of the object in the hand if the operator need to carry something to somewhere.

2.2 *Posture adjustment*

Posture adjustment contains 5 parts, which are leg adjustment, position, pick, place and collision. Leg adjustment is the movement of leg, while pick and place are the motion of human arm and hand. Position is the combination of leg and upper limbs, while collision is the movement of all human limbs.

2.3 *Hand operations*

In this part, hand operations contain only hand motion with tools. Bare-handed operation belongs to posture adjustment. Connector operation and maintenance tool operation are the two section of hand operations.

3 TIME COMPENSATION METHODOLOGY

As shown in Figure 1, the overall framework consists of 2 parts: human motion time and compensation time. Human motion time is estimated by methods time measurement that have been discussed in Section 2, while compensation time will be fully discovered in Section 3.

In the study of field maintenance work, the author have found that the proficiency and the fatigue of the operator have great impact on maintenance time. While other qualitative maintainability parameters, such as visual accessibility, do have some influence on maintenance time; but they are not important as the factors of proficiency and fatigue are. Therefore, the author takes proficiency and fatigue as two main compensation part, while other qualitative maintainability parameters are integrated into maintenance environmental factor.

$$T = T_0 \cdot \alpha_i \cdot \left(1 + \beta_j\right) \cdot \gamma_k \tag{1}$$

where T = final estimated time; T_0 = human motion time measured by methods time measurement; α_i = proficiency ratio; β_j = fatigue rate; γ_k = maintenance environmental ratio.

3.1 *Proficiency*

Proficiency refers to the degree to which the operator does the maintenance work. A high proficiency means that the operator is skilled at the maintenance work including the maintenance process, tools used during maintaining, which are operating objects, where to put dissembling items and other maintenance matters. An operator with high proficiency would start his work in order and finish them one by one. While without high proficiency, a maintainer would need more time to look at the maintenance manual to see what to do next, which would waste a lot of time.

This paper classifies proficiency in three levels. Below are detailed rules describing the degree to proficiency in each level:

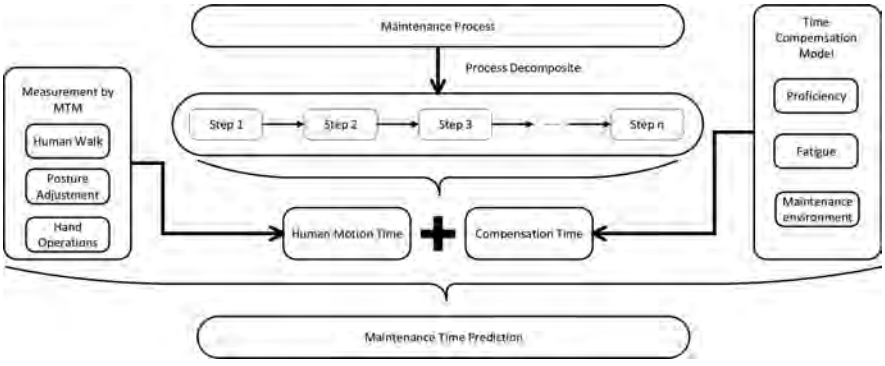


Figure 1. Framework of the time compensation methodology.

Level 1 – The operator knows the whole maintenance process including steps, tools, objects and he masters this operation.
Level 2 – The operator knows the whole maintenance process, but he have never done it or he only have done it once before this operation. He may need to look at the maintenance manual when he forgets what to do next, but it is not often.
Level 3 – The operator doesn't know how to maintain the object. He needs to look at the maintenance manual while operating.

The relationship between proficiency level and its ratio is shown in Table 1.

3.2 *Fatigue*

Fatigue is common especially in long time maintenance. After long time maintaining, operators would feel tired, which causes the decline of mental concentration. Less concentration results in the increase in maintenance time. Fatigue would also cause the decrease in moving speed including hand moving, walking and turning.

The detailed rules are shown as follow:

Level 1 – The operator doesn't feel tired. He is in the same mood as beginning. Therefore, his motion would not be affected.
Level 2 – The operator feels a little tired. His motions are also affected by his fatigue. He may take a little break or slow down his operation, but it wouldn't be long. His efficiency may also reduce due to tiredness.
Level 3 – The operator feels tired. He need to take a break or he would make mistakes. In this mood, he should have a rest or something dangerous may occur.

The relationship between fatigue level and its rate is shown in Table 2.

Table 1. Proficiency ratio.

Proficiency level	Ratio
Level 1	α_1
Level 2	α_2
Level 3	α_3

Table 2. Fatigue rate.

Fatigue level	Rate
Level 1	β_1
Level 2	β_2
Level 3	β_3

3.3 *Maintenance environment*

Due to the limit of virtual maintenance environment, it is hard to simulate the working environment same with the actual one in the aspect of temperature, humidity, illumination condition. But it still has other maintenance environmental factors that will influence maintenance time.

Visual accessibility means the extent or visibility that can be seen from the operator's current location. A good visual accessibility refers to that during maintaining, the operator could directly see the working area, objects and tools. A bad vision could result in the increase of maintenance time because the maintainer would need more time to search for tools or others.

Operating space refers to area of an operator touching objects and getting tools. A good operating space could provide the maintainer a comfortable working place. While a bad operating space would increase human motion time.

Those two maintenance environmental factor could be examined in virtual maintenance which are shown in Figure 2. The upper one shows the operator's visual accessibility, while the other presents the operating space.

In this paper, the author presents three rules of maintenance environment:

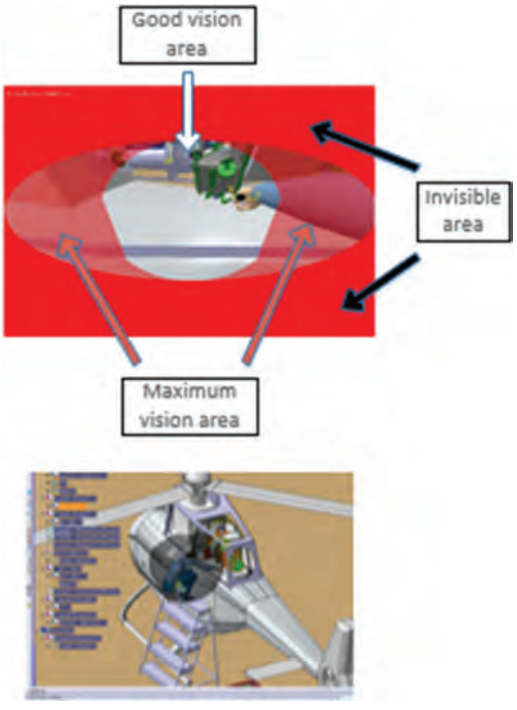


Figure 2. Visual accessibility and operating space in virtual maintenance.

Table 3. Maintenance environmental ratio.

Maintenance environmental level	Ratio
Level 1	γ_1
Level 2	γ_2
Level 3	γ_3

Level 1 – The operator is in a natural posture, which means he could easily get what he want or touch what he want to operate, and at the same time all maintenance objects and tools are right in the his sight.

Level 2 – Either the operator is in a narrow space that he could not be so easy to get or touch what he want, or he doesn't have a good vision that some are out of his sight, which means he needs to turn to a certain angle to touch the object or the tool.

Level 3 – The operator is in a narrow space and most or all maintenance objects and tools are out of his view.

The relationship between maintenance environmental level and its ratio is shown in Table 3.

3.4 *Time estimation*

Before doing the final calculating, the analyst should check the integrity of the virtual simulation. If there is any simplification, the analyst should take that into calculation. After checking, the analyst could choosing different ratio according the different situation and the final value is the result of this time estimation.

4 CASE STUDY

Using the methodology above, the author uses the maintenance of an electric fan as example. The virtual model of the fan is shown as Figure 3. While the process of its maintenance is presented in Table 4. The author uses DELMIA (Digital Enterprise Lean Manufacturing Interactive Application) as the virtual maintenance platform, where the simulation is made exactly based on steps showing in Table 4. While in the simulation, time is estimated by PTS, which refers to Predetermined Time Standards. And the actual time is collected through several surveys.

Table 4 show the detailed process of this maintenance work. And the time of each task is listed in Table 5, which including data from methods time measurement, proposed method and actual work. Time of methods time measurement only consider human motion, while proposed method not only

Figure 3. Virtual model of a fan.

Table 4. Maintenance process.

Number	Detailed task	Component of motion
1	Pick up the tool	Pick
2	Remove 2 nuts on the two sides	Pick, screw × 2 and place
3	Put down the tool and take out the body part of the fan	Pick, place
4	Pick up another tool	Pick
5	Remove 3 nuts on the upper side	Arm raise, screw × 3, arm fall
6	Put down the tool	Place
7	Remove the front cover of the body part	Pick, place

Table 5. Maintenance time calculation.

Motion	Methods time measurement	Proposed method	Actual time
Pick	1.1	1.1	2.5
Screw 1	12.4	18.6	17.8
Pick	1.1	1.1	1.1
Place	1.4	1.4	1
Screw 2	12.6	18.6	18
Place	1.4	1.4	1.3
Pick	1.1	1.1	2.7
Arm raise	0.5	0.5	0.9
Screw 3	12.2	18.6	18.3
Screw 4	12.2	17.0	14.2
Screw 5	12.8	17.0	23
Arm fall	0.5	0.5	2.1
Place	1.4	1.4	2.5
Pick	1.1	1.1	1.7
Place	1.4	1.4	1.6
Total	73.2	100.8	108.7

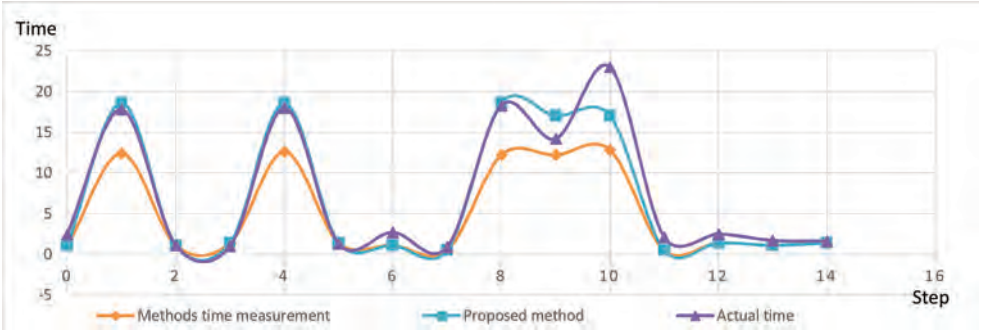


Figure 4. Comparison between two methods.

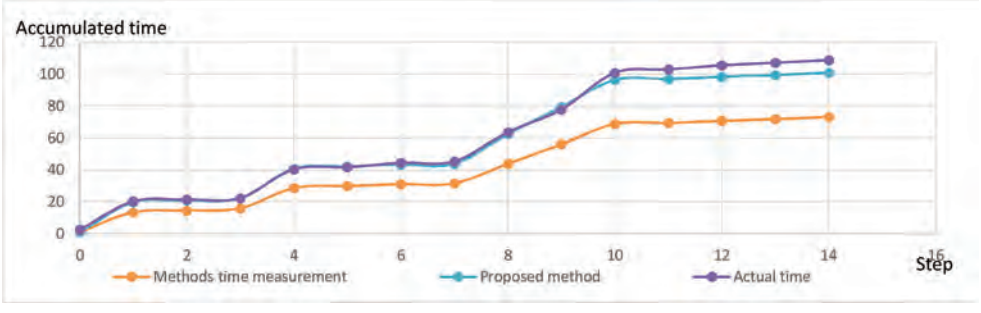


Figure 5. Accumulated time comparison.

uses human motion, but also time compensation rules.

As can be seen from Figure 4 and Figure 5, the orange line stands for methods time measurement, while the blue one represents proposed methods and the actual result shows in purple line. It is clearly that compared to methods time measurement (the orange line), time estimated by proposed method is closer to practice. And result also shows that proficiency is the main factor that influences maintenance, which fits in with the actual survey result the author have gotten in field maintenance.

In real maintenance, operators are often unfamiliar with the maintenance process because the failure rate of the equipment is extremely low, especially in the field of aeronautics and astronautics. Therefore, proficiency is the key factor in time estimation.

5 CONCLUSION

This paper presents a methodology for maintenance time estimation based on virtual reality, where methods time measurement and time compensation model are discussed. The advantages of this methodology is that it provides a time estimated way without physical prototype and it is more accurate than methods time measurement. By analyzing the comparison result, the author finds that proficiency is much more important in maintenance time estimation, and by actual survey the author also finds that qualitative maintainability parameters may not play an important role in time prediction. The author must admit that they do have some influence, but they do not have as much as I have thought before, which is only the author's opinion.

Future work will focus on the importance of qualitative maintainability and the simplification of the proposed method. Moreover, with the development of virtual reality, more methods will be found based on methods time measurement and the proposed method, which may be more simple and useful in real life.

REFERENCES

Christiand, Yoon, J., & Kumar, P. 2009. A novel optimal assembly algorithm for haptic interface applications of a virtual maintenance system. *Journal of Mechanical Science & Technology, 23(1),* 183–194.

Cruz-Neira, C., Sandin, D.J., & Defanti, T.A. 1993. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. *Conference on Computer Graphics and Interactive Techniques* (pp. 135–142). ACM.

Cutler, L.D., & Hanrahan, P. 1997. Two-handed direct manipulation on the responsive workbench. *Symposium on Interactive 3d Graphics, Si3d '97, Providence, Ri, Usa, April* (pp. 107–114). DBLP.

Dossett, R. 1992. Computer application of a natural-language predetermined motion time system. *Computers & Industrial Engineering, 23*(1–4), 319–322.

Fischer, & Wygant. 1991. A computerized system to measure repetitive motion stress on the lower back. *Computers & Industrial Engineering, 21*(1–4), 613–616.

Griswold, G.H. 1970. Maintainability prediction and demonstration technique development of prediction, Techniques 1.

Geng, J., Lv, C., Zhou, D., Li, Y., & Wang, Z. 2014. Compensation-based methodology for maintenance time prediction in a virtual environment. *Simulation Modelling Practice & Theory, 47*, 92–109.

Genaidy, A.M., Agrawal, A., & Mital, A. 1990. Computerized predetermined motion-time systems in manufacturing industries. *Computers & Industrial Engineering, 18(4),* 571–584.

Genaidy, A.M., Mital, A., & Obeidat, M. 1989. The validity of predetermined motion time systems in setting production standards for industrial tasks. *International Journal of Industrial Ergonomics*, 3(3), 249–263.

Hoffmann, E.R., Macdonald, W.A., & Almond, G.C. 1993. Quantification of the cognitive difficulty of mail sorting. *International Journal of Industrial Ergonomics, 11*(2), 83–98.

Kanai, S., Takahashi, H., & Makino, H. 1996. Aspen: computer-aided assembly sequence planning and evaluation system based on predetermined time standard. *CIRP Annals—Manufacturing Technology, 45(1),* 35–39.

Laring, J., Forsman, M., Kadefors, R., & Örtengren, R. 2002. Mtm-based ergonomic workload analysis. *International Journal of Industrial Ergonomics, 30*(3), 135–148.

Laurig, W., Kühn, F.M., & Schoo, K.C. 1985. An approach to assessing motor workload in assembly tasks by the use of predetermined-motion-time systems. *Applied Ergonomics, 16*(2), 119–125.

LIU-Duan, Jian-Bo, H.U., Xiao-Kai, G.E., Zhang, L., & Wang, X.W. 2014. Monte carlo simulation of maintenance time based on system maintenance work procedure. *Fire Control & Command Control, 39(7)*, 119–123.

Owensby, J.E., & Summers, J.D. 2014. Assembly time estimation: assembly mate based structural complexity metric predictive modeling. *Journal of Computing & Information Science in Engineering, 14*(1), 011004.

Pliska, T.F., Jew, F.L., & Angus, J.E. 1978. Maintainability prediction and analysis study. revision a. *Maintainability Prediction & Analysis Study Revision A*.

Shen, G.X., Zeng, W.B., Zhang, Y.Z., Wu, M.K., & Zheng, Y.B. 2017. Determination of average maintenance time of cnc machine tools under minimum failure rate. *Jilin Daxue Xuebao, 47*(5), 1519–1526.

Wygant, R.M., White, B.E., & Hunt, D. 1993. Combining ergonomics and work measurement for job analysis. *Computers & Industrial Engineering, 25*(1–4), 423–426.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Assessing the impact of operational context variables on rolling stock reliability. A real case study

J. Izquierdo
IK4-Ikerlan Technology Centre, Operations and Maintenance Technologies Area, Gipuzkoa, Spain
Department of Industrial Management I, School of Engineering, University of Seville, Sevilla, Spain

A. Crespo
Department of Industrial Management I, School of Engineering, University of Seville, Sevilla, Spain

J. Uribetxebarria
IK4-Ikerlan Technology Centre, Operations and Maintenance Technologies Area, Gipuzkoa, Spain

A. Erguido
IK4-Ikerlan Technology Centre, Operations and Maintenance Technologies Area, Gipuzkoa, Spain
Department of Industrial Management I, School of Engineering, University of Seville, Sevilla, Spain

ABSTRACT: The Original Equipment Manufacturers (OEMs) nowadays face the need of establishing an optimized maintenance plan from the design stage of the assets. Up to date, the production-centered business model has limited their after-sales maintenance strategy, and accordingly their knowledge about the assets operational behaviour. Furthermore, the added difficulty of the assets operating in different contexts (increasing the variability of their behavioural patterns) contributes to the misalignment of the maintenance plan with the assets' actual needs. Therefore the purpose of this paper is to propose a methodology based on the proportional hazards model for assessing the behaviour of the assets and the influence of the different operational context variables in their reliability. This methodology aims to provide support information to a better customization of the maintenance plans in the offer stage. Likewise, the proposed methodology has been verified and validated through a real case study with data provided by a leading company in the railway sector.

1 INTRODUCTION

Traditionally Original Equipment Manufacturers (OEMs) have focused their efforts on asset production and sale, however, the highly demanding market has imposed the need of maintaining their competitiveness levels through the after sales service, which might become a profitable source of income. Since the business approach of the OEMs has shifted to this new paradigm, it can be identified in the literature a growing body of publications recognizing the importance of optimizing the maintenance plans.

Nevertheless, the OEMs have not achieved a desirable optimization level in their after sales service. To date, because of these inefficient maintenance plans, the after sales service has incurred into avoidable costs and excessive resources consumption. This becomes a major issue when the maintenance strategy is designed at the offer stage of the asset, where the customization of maintenance actions and frequencies is difficult due to the lack of available information to characterize the asset failure mechanisms. Furthermore, the operational environment of products can change, requiring reliable operation in unfamiliar circumstances, or uncertain asset's operators' behaviour (de Rocquigny et al. 2008).

Nowadays, to face the mentioned problem, the OEMs follow a holistic approach supported by information provided by the suppliers and by information regarding similar assets' patterns. The customization degree of the resulting maintenance plans does not properly integrate information regarding the operational context in which the assets will perform. Thus, in order to avoid the misalignment of expected and actual assets' maintenance needs, it is crucial to develop capabilities and tools to assess the influence that the different operational context variables have in the behaviour of the asset. To this aim, a novel methodology is developed and presented in this paper, with its

corresponding application to a real case study in the railway sector.

The Proportional Hazards Model (PHM) plays a key role in the proposed methodology by being a valid approach to relate survival analysis and the impact of operational context variables. The main advantage of the model is the possibility of considering the impact of more than one variable simultaneously (Tang et al. 2014). The PHM was first introduced by Cox (1972) in a seminar paper in the Royal Statistical Society, not only has it been proposed for medical studies but for reliability analysis by Cox himself and by other authors in the literature (Bendell et al. 1986, Wightman & Bendell 1986, Lawless 1983).

A critical review of the existing literature in the beginnings of this new methodology can be seen in Bendell (1985) where the complexities of the model are enumerated along with the proper way of applying it to reliability studies. Baxter et al. (1988) presented it as a powerful tool for examining reliability data-sets where the failure data may be inhomogeneous due to the presence of risk factors. Through the literature various examples of its applications can be found, in these applications, it is used as an exploratory tool in Bendell et al. (1986) and in Wightman & Bendell (1986); but also as a reliability prediction model for, among others, rail diesel engines (Jardine et al. 1989), marine and aircraft turbines (Jardine et al. 1987), traction transformers (Lin et al. 2016), vitrified clay pipes (Xie et al. 2017), electromagnetic relays (Li et al. 2015) and mobile handsets (Tiwari & Roy 2013).

As stated by Li et al. (2010), the PHM lays a mathematical foundation for predicting the failure occurrence and developing optimal maintenance policy. Thus, in this paper the application of the PHM to a real case study in the railway sector is presented and discussed, having the obtained results been validated by a leading company in the sector. More precisely, several components of the Heating Ventilating and Air Conditioning (HVAC) System have been analyzed through the study of a database that gathered failure information from different operational contexts. The intention is to provide a case study in the application of the PHM and its potential for the maintenance plan design at the beginning of the asset's lifecycle, where no information to characterize its behaviour is available.

The structure of this paper is as follows. Firstly the HVAC system and the several aspects to be taken into account regarding its maintenance are presented, besides the failure data to be analyzed is also discussed in section 2. In sections 3 and 4 the Cox model, along with its main equations, and the proposed methodology are explained correspondingly. Once the main concepts are introduced, in section 5 the application to a real case study is performed, where several components of the HVAC systems with their individual resulting PHMs are analyzed. Finally, in section 6, the conclusions and benefits resulting from the application of the proposed methodology to the real case study are presented.

2 HVAC SYSTEM

The railway sector is a growing industry that provides good mobility for a reasonable price. Transportation systems' complexity is increasing, as well as its number of users; hence it is necessary both, to improve their availability and to guarantee high levels of security and comfort whilst ensuring reasonable maintenance costs (Foulliaron et al. 2014). There exists a trend towards a cost-efficient and performance-based system, this trend is highly influenced by political, economic and environmental motivations (Umiliacchi et al. 2011). One of the main pillars within the trend is the asset management discipline, and the importance of a proper asset management strategy is supported by the vast amount of literature focused on optimizing the maintenance plans for rolling stock.

As previously stated, in order to optimize the maintenance actions, customization to operational context is required. This paper discusses the application of PHM to the HVAC system in the railway sector, which has been chosen because it plays a key role in comfort levels of the passengers. The HVAC system in a train, metro or tramway provides an air flow in order to maintain proper and comfortable room environmental conditions. Among the functionalities of the HVAC system several comfort-related aspects are taken into account, these functionalities include (1) temperature adjustment, (2) refreshing of room air, (3) filtration of the outside air, (4) ensuring proper noise levels and (5) avoiding pressure waves (in high-speed trains).

To regulate the inside comfort environment, it is necessary to take into account that the setpoint temperature depends on the outside temperature in order to avoid excessive thermal shock. However this is not the only parameter to be considered when designing an HVAC system; the indoor relative humidity level should be maintained below 60% on average, and it is important to remark that passengers are heat and humidity sources. Accordingly, HVAC system should be designed to maintain desirable levels of inside temperature and humidity; the technical solution adopted should endure the outside conditions, as well as the coach maximum occupancy.

As stated in Bendell (1985), the proportional hazards modeling techniques for reliability analysis are useful when the analyzed data correspond

to failures of non-repairable items whose times to failure are completely independent and equally distributed. Following these concerns, for the case study three different components of the HVAC system have been selected: the control panel, the electronic control (as a subcomponent of the control panel) and the compressor unit.

The database has been provided by a leading company in the sector, it includes the work orders of maintenance services for HVAC systems in rail projects working in different European cities. The full study considered the HVAC systems running from 1st January 2015 to 31st August 2017, since the quality of the data is more consistent in this period. From the maintenance work orders record, the Time Between Failures (TBFs) has been calculated, and then the information regarding the operational context has been associated with each of the TBFs.

3 PROPORTIONAL HAZARDS MODEL

The PHM assumes that the hazard function of an asset decomposes into the product of a baseline hazard function and an exponential term incorporating the effects of the explanatory operational context's variables (covariates) (Cox 1972). The most extended PHM is given by Equation 1,

$$h(t, X_1 \ldots, X_p) = h_0(t) \exp\left(\sum_{j=1}^{k} \beta_j X_j \right) \quad (1)$$

where $h(t, X_1, \ldots, X_k)$ represents the hazard rate at time t for an asset with covariates $(X_1, \ldots, X_k)$; $\mathbf{X} = (X_1, \ldots, X_k)$ is the vector containing the values of the explanatory variables where $X_i (\forall i = 1,2,\ldots,k)$ can be either a naturally variable or an indicator variable; $h_o(t)$ represents the baseline hazard function, it would be the hazard function for the null vector $\mathbf{X} = \mathbf{0} \in \mathbb{R}^k$ – it can be parametric following certain distribution or of a unspecified form; and $\beta = (\beta_1, \ldots, \beta_k)$ is the vector of the parameters of the model which describe the effects of each of the covariates.

Given two identical assets operating in two different operational contexts, that would be $\mathbf{X} = (X_1, \ldots, X_k)$ and $\mathrm{X}' = (\mathrm{X}'_1,\ldots,\mathrm{X}'_k)$, being the X' one with a higher risk; it is defined the Hazard Ratio (HR) between the two of them as:

$$HR = \frac{h(t,\mathbf{X}')}{h(t,\mathbf{X})} = \frac{h_0(t)\exp\left(\sum_{j=1}^{k} \beta_j X'_j\right)}{h_0(t)\exp\left(\sum_{j=1}^{k} \beta_j X_j\right)} \quad (2)$$

$$HR = \exp\left(\sum_{j=1}^{k} \beta_j \left(X'_j - X_j \right) \right) \quad (3)$$

It can be observed in Equation 3, that the Hazard Ratio is independent of time t, being that the property which gives the model its name of Proportional Hazards Model.

A special case of the HR would be when comparing an asset operating in an environment where the covariates take mean values $\overline{\mathbf{X}} = (\overline{X}_1 \ldots, \overline{X}_k)$ with an asset operating in a context $\mathbf{X_i} = (X_{i1}, \ldots, X_{ik})$; given this HR let the hazard function of the asset operating in $\mathbf{X_i}$ be expressed by Equation 4,

$$h(t,\mathbf{X_i}) = h_{\overline{\mathbf{X}}}(t) exp\left(\sum_{j=1}^{k} \beta_j \left(X_{ij} - \overline{X}_j \right) \right) \quad (4)$$

where the hazard function is decomposed into a baseline hazard function in the mean values of the covariates $(h_{\overline{\mathbf{X}}}(t))$, and into the exponential part where the deviations from the mean value of each covariate are taken into account, instead of the covariates values themselves.

The baseline hazard function in the mean values of the covariates can be estimated in the model and it can also be either an unspecified form function, or a function following certain distribution. In this paper it has been fitted to follow a two-parameter Weibull distribution, thus the obtained model would be expressed by Equation 5,

$$h(t,\mathbf{X_i}) = \frac{\gamma}{\alpha} \cdot \left(\frac{t}{\alpha} \right)^{\gamma - 1} \cdot \exp\left(\sum_{j=1}^{k} \beta_j \left(X_{ij} - \overline{X}_j \right) \right) \quad (5)$$

where:
α is the characteristic life (scale parameter).
γ is the shape parameter.
T is the time.

4 METHODOLOGY & PROPORTIONAL HAZARDS MODEL

The proposed methodology for assessing the impact of operational context variables is shown in the flowchart of Figure 1. It consists of two modules, the first one is oriented to data treatment and preparation, and the second one mainly comprehends the model fit and tests of the PHM. In the next paragraphs, a detailed explanation of both is presented.

4.1 *Module 1. Data treatment and preparation*

In this first module, the work is directed towards two focus of interest, the identification of the operational context variables and the treatment of the maintenance data. On the one hand, for identifying the variables that affect the reliability of the

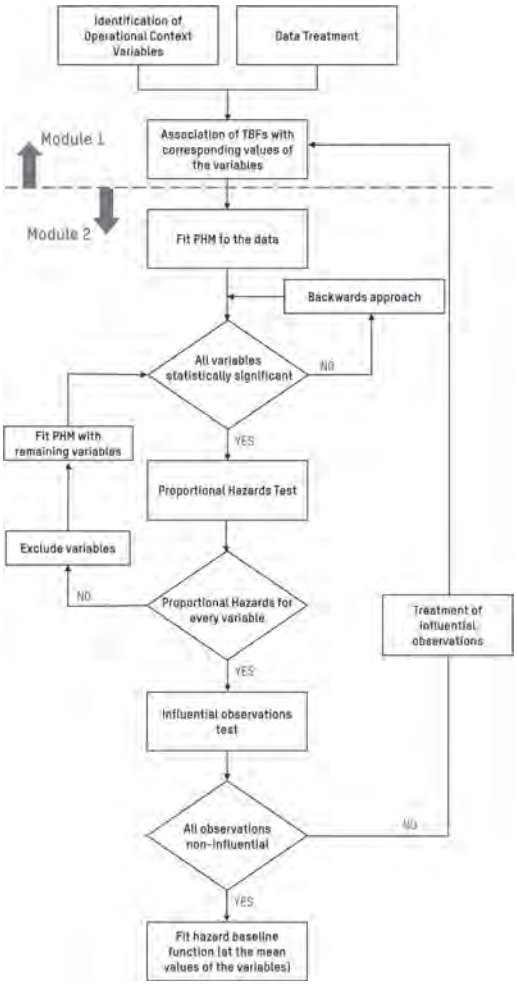


Figure 1. Flowchart of the proposed methodology.

asset it is essential to cooperate with maintenance workers in order to translate their know-how into valuable feedback that helps in the integration of every environmental aspect influencing the failure mechanisms. On the other hand, following the general framework in a reliability improvement study proposed by Louit et al. (2009), the work orders are treated to calculate the TBFs taking into consideration the diverse issues such as scarce data, censoring (only right is contemplated) or data pooling. Likewise, it is important to check and minimize as much as possible potential human errors and noise, such as double imputed work orders or non-failure related work orders.

Having calculated the TBFs, as well as the variables that might influence the reliability, the linkage between both should be performed. To this aim, it is necessary to define a way of associating a value of the variables to every TBF; nonetheless, the procedure to follow for each of the variables should be ad-hoc designed depending on the available information and the characteristics of the variable.

4.2 *Module 2. Valid PHM fit*

With the data set containing the failure times associated with the appropriate context variable value, it is the next step to seek for a valid PHM fit using the method of partial likelihood. In this paper's case study the fit has been performed with the Cox Proportional Hazards regression in R, but there are many software solutions for this regression. The way to proceed should be to try to fit a PHM that includes all the identified covariates, however, it is very likely that many of them will not result influential with enough statistical significance. Significance levels have to be arbitrary set for the null hypothesis contrast where every $\beta_i = 0 (\forall i = 1,2,\ldots,k)$, in the case study a p-value <0.1 was considered significant.

Moreover, it also exists the possibility of having identified collinear variables, i.e. covariates which can be linearly predicted from any of the others. When in the results appears either a not statistically significant or a collinear covariate, they are eliminated following the backward approach proposed by Bendell et al. (1986) until a PHM is reached, where each variable is influential with a p-value lower than the selected one, and no collinearity is detected.

Already the model contains the proper covariates, it is required to check the proportional hazards hypothesis under which the model has been fit. To this aim, a test and a graphical analysis for each significant variable are performed based on the Schoenfeld residuals obtained from the regression. If a covariate performs well in neither the analytical test, nor the graphical one, it needs to be excluded from the PHM model and the regression performed again with the remaining covariates.

When the proportional hazards hypothesis has been tested and validated, the next phase is to spot any possible influential observations that may have biased the estimation of the regression parameters. In case any is identified, it is examined to consider how it should be treated before proceeding in the fit of a more accurate PHM.

Since by default $h_{\bar{x}}(t)$ is a free distribution function estimated in the regression, the last step of the methodology consists of fitting it to a suitable probability distribution. So the resulting Proportional Hazards Model will be a completely parametric function (see Equation 5), that can be decomposed in two terms:

- A hazard baseline function following certain distribution, in the case study a two-parameter

Weibull, in the mean values of the covariates. Thus it will depend on the parameters of the chosen distribution and it will depend on the time (t) as well.

- An exponential function that integrates the information regarding the operational context, since the deviation from the mean of each influential variable is multiplied by its corresponding estimated parameter.

5 REAL CASE STUDY RESULTS & DISCUSSION

The system to study is the HVAC system integrated into different projects of the railway sector, however, the PHM is not applied to the system as a whole unit but to selected subsystems. In order to select the subsystems several criteria have been followed but firstly a Failure Mode and Effect Analysis has been performed to fully understand the failures mechanisms of the HVAC system. Having formalized the knowledge about HVAC failure modes, an approach based on the CIB-framework proposed by Waeyenbergh and Pintelon (2004), is applied. To identify both the Most Important Systems (MIS) and the Most Critical Components (MCCs), the know-how of the company workers is documented and implemented in the Maintenance Cost Matrix in Figure 2.

Every MIS, at different indenture levels, is included in the matrix where in the vertical axis Corrective Costs (CC) are represented and the horizontal axis corresponds to preventive costs (PC). However this cost classification is based on a Pareto qualitative approach since for each component both have been calculated by Equations 6 and 7,

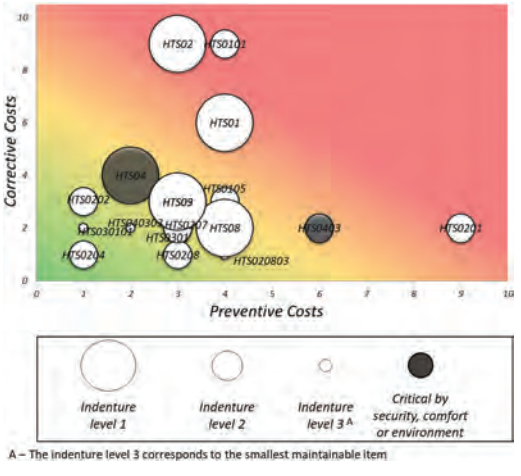


Figure 2. Maintenance cost matrix.

$$CC = PI_{CorrectiveCost} \cdot PI_{FailureFrequency} \quad (6)$$

$$PC = PI_{PreventiveCost} \cdot PI_{PreventiveFrequency} \quad (7)$$

where PI is the Pareto Index being: 1 for low, 2 for medium and 3 for high.

In the Matrix of Figure 2, four MIS can be seen in the red area of the matrix representing high maintenance costs. It also shows two MIS whose criticality is due to security, comfort or environmental issues. These six MIS are the ones that are categorized as MCC and thus are the ones that have been analyzed by the PHM. The identified MCCs are shown in Table 1, with their reference, description and indenture level.

Once the items to be analyzed have been defined, their corresponding failure data have been extracted from the database and treated as proposed in the methodology (see subsection 4).

The next step is to identify what are the influential variables that are going to be integrated into the analysis. By cooperating with the company, 11 covariates have been selected as possible risk factors, some of them are defined by the project specifications and the others are exogenous uncertainty variables. In Table 2, the criteria for selecting every variable are explained.

When applying the proposed methodology only three out of the six MCCs provide a valid PHM model, the control panel (HTS01), the control electronics (HTS0101) and the compressor module (HTS04). The evaporator module (HTS02) can be adjusted to a PHM where the maximum temperature and the relative humidity have influence in the hazard function, however, they do not pass the proportional hazards test.

For neither the HTS0403 nor the HTS0201, a valid PHM can be adjusted; this is due to data scarcity. Both of them have been selected because their preventive maintenance cost was high (bottom-right position in the matrix). As a matter of fact, is this over-maintenance that causes failures absence, and therefore data scarcity.

In every case, the integration of the same 7 variables was problematic due to collinearity problems, these variables correspond to X_1, X_2, X_3, X_4, X_6, X_7

Table 1. Identified MCCs.

Reference	Indenture Lv.	Description
HTS01	1	Control panel
HTS0101	2	Control Electronics
HTS02	1	Evaporator Module
HTS0201	2	Air filter
HTS04	1	Compressor Module
HTS0403	2	Compressor

Table 2. Covariates.

Variable	X_i	Selection criteria
Number of passengers	X_1	People is a source of humidity and heat so the company believes that the annual average of passenger for every project has to be included since it might affect the reliability
Refrigerant	X_2	The kind of refrigerant the HVAC system of every project is using modify the pressure it works
Number of stops	X_3	The bigger the number of stops of every project is the more the HVAC system is exposed to heat loss or gain and therefore the more it has to work.
Intensity of use	X_4	For every project the contractual available trains working is different as well as the fleet to provide that availability and therefore the HVAC system work load differ from every project
Relative humidity	X_5	As one of the functions of the HVAC system is to provide comfortable room humidity levels the outside atmospheric humidity will condition its work load
Maximum distance between stops	X_6	The distance between stops its equivalent to the time the HVAC is working and its capability to reach working permanent regime. To incorporate the deviation of this variable the maximum and the minimum distances for every project are also taken into account.
Minimum distance between stops	X_7	
Average distance between stops	X_8	
Maximum temperature	X_9	As the main function of the HVAC system is to provide comfortable room temperature the outside temperatures will condition its work load. To incorporate the deviation of this variable the maximum and the minimum temperatures is also taken into account
Mean temperature	X_{10}	
Minimum temperature	X_{11}	

and X_8. Most of them are project specifications and therefore the value of one of them determines the value of the others since it corresponds to a certain project. As a result of the variables demeanor, it is impossible to distinguish in the regression the individual effects of every one of them.

Due to this problem, it has been only possible to fit into the models the uncertainty variables corresponding to the relative humidity and the temperature, which are X_5, X_9, X_{10} and X_{11}.

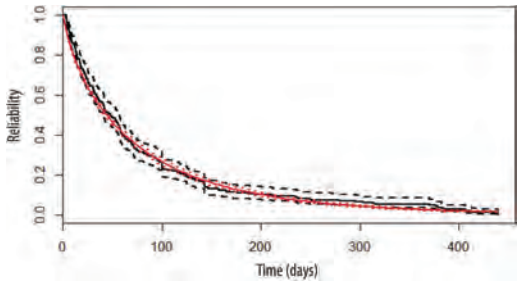


Figure 3. HTS01 survival baseline fit (Weibull, 95% CI).

5.1 *HTS01: Control panel*

In the analysis of this MCC, two covariates are found to be statistically significant, with a default p-value of 0.01 for the hypothesis contrasts. These two covariates are the relative humidity and the minimum temperature (X_5 and X_{11}) with coefficients $\beta_5 = -0.008$ and $\beta_{11} = -0.041$. Both of the coefficients are lower than zero meaning that the higher the variables' values are, the lower the asset's failure rate.

Due to the resulting values of the coefficients, a unitary increase in the relative humidity (at a constant minimum temperature) provides a 0.8% lower failure rate; and a unitary increase in the minimum temperature (at a constant relative humidity) provides a 4.02% lower failure rate.

With a valid PHM, the fit of the baseline function in the mean values of the two covariates has been performed. A Weibull Distribution has been fitted to the non-parametric baseline function, it can be seen in Figure 3 with the 95% confidence intervals for both functions. These results have been contrasted with the collaborating company and there is one major conclusion about the component and its resulting PHM.

It is worth to highlight how the higher the minimum temperature is the better the component performs. After discussion with the HVAC technicians, a likely explanation for this phenomenon has been found. The technicians point out that there is not forced ventilation for the electronic components, thus their temperature is not controlled and the outside temperature will have more influence on their behaviour. It also has been long noticed how the failure rate of the control panel increases when there are important temperature fluctuations

within the day. Thus it is believed that the modeled positive impact of the minimum temperature increase, reflects the observed negative impact of temperature fluctuation. This discussion has lead to future conjoint works.

5.2 *HTS0101: Control electronics*

This component is a subsystem of the HTS01 previously studied, and it is reasonable to think that its resulting PHM will share some features with the one from the HTS01 analysis. As it was thought, they do share properties; the minimum temperature is the only variable that allows fitting a valid Cox model and therefore a single-covariate PHM is obtained. The coefficient associated with this variable is $\beta_{11} = -0.045$ and it provides a 4.44% lower failure rate by a unitary increase of the minimum temperature.

It can be seen that the effect of the minimum temperature is defined by a negative coefficient, meaning that the higher the minimum temperature, the better reliability the asset will have. The technicians have validated this result through the same reasoning applied to HTS01. The fit of its undefined baseline function, in the mean value of the covariate, to a Weibull distribution can be seen in Figure 4.

It is important to notice that the effect of the minimum temperature is almost equal in HTS01 and HTS0101. However not only is this coefficient similar, but the fits to Weibull distributions result in very similar shape parameters. It is also important to mention that the scale parameter of HTS01 is smaller which is reasonable since HTS01 includes HTS0101 whose scale parameter is over 65% higher.

Therefore the control panel and the control electronics are both affected by the same risk factor, and their behaviour is also similar because of their shape parameters. However, the reliability of the HTS01 is lower because it depends on the failure rate of HTS0101, and on other components' failure with a serial arrangement as well.

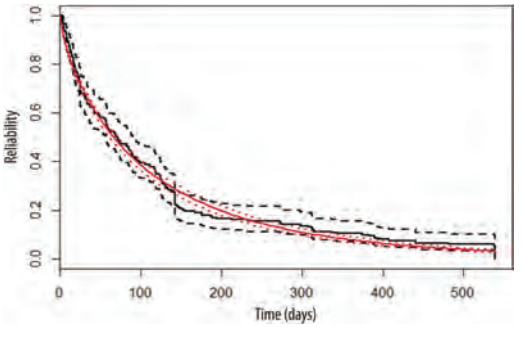


Figure 4. HTS0101 survival baseline fit (Weibull, 95% CI).

5.3 *HTS04: Compressor module*

The number of data available for this MCC was not very large, in spite of this inconvenience, a proper Proportional Hazards Model has been found to describe the behaviour of the component with enough statistical significance. The influential variables for this MCC are the relative humidity and the mean temperature, with their corresponding coefficients equal to $\beta_5 = 0.0547$ and $\beta_{11} = 0.1316$. In this case, both of the covariates are associated with positive coefficients meaning that the higher values of the covariates, the higher risk of failure. Being more specific, if the relative humidity increases one unit (at a constant mean temperature) it provides a 5.62% increase of the hazard function; and if the unitary increase happens in the mean temperature (at a constant relative humidity value), the increase in the hazard function equals 14.07%. Once the PHM has been adjusted and validated, the hazard function in the mean values of the covariates is fitted as proposed in the methodology. It can be seen in Figure 5, and it is important to notice how the fitted curve is not as accurate as in the previous MCC because of the data scarceness.

It is remarkable the major effect of the mean temperature, it reveals that it is an important risk factor to take into account when designing the maintenance strategy of this critical component. The collaborating company validates this results, and they have been dealing with a problem that backs up the results here obtained.

6 BENEFITS & CONCLUDING REMARKS

In this paper, a failure rate model for different components of the HVAC system is proposed considering different risk factors of the operational context. The main contribution of this paper is the PHM approach following an established methodology that set a number of steps for a proper application of the Cox model. It has been demonstrated

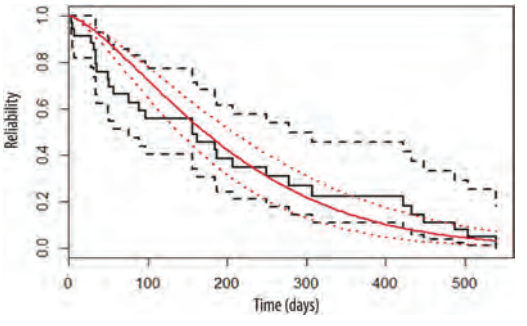


Figure 5. HTS04 survival baseline fit (Weibull, 95% CI).

that as thought, the temperature and the relative humidity influence the behaviour of the asset.

In the last step of the methodology, by fitting the baseline hazard function in the mean values of the covariates, a Weibull PHM is obtained. The obtained failure rate model consists of a completely parametric equation (see Equation 5). The parametric equation has three important advantages:

- In the offer stage it allows to know the asset reliability before it starts to operate, and therefore it is possible to develop an optimized maintenance plan that better fits the reality of the asset operation improving the efficiency and effectiveness of the maintenance activities. The customized reliability enables a more accurate Life Cycle Cost (LCC) analysis that may be a competitive advantage in tender processes, setting the company a step forward their competitors.
- From the after-sales service's point of view, one of the most significant advantages is the improvement of the management of an assets fleet in which every one of them is working in different operational conditions and shows a different failure mechanism. This improvement in the management will result in higher availability rates, and accordingly a revenue from the decreased costs due to non-availability, and in savings derived from the enhanced customization of maintenance frequencies. It is remarkable that not only is the OEM making profit of the new reliability model, but the assets users as well.
- Depending on the arrangement of the components of the asset, it is possible that in certain operational contexts the criticality of the components will increase due to a higher failure rate caused by one or more variables of the context. The proposed model allows to foresee this criticality increase and to anticipate its consequences by identifying the components in which the maintenance service should focus its efforts.

As a future work, it is necessary to study a way of modeling the impact of project specifications, since it is not possible to integrate their information in the proposed PHM due to collinearity problems.

It is interesting to remark that some uncertainty variables, such as the number of passengers, could be integrated into the model if a more detailed record of the variable were available.

The work here presented can serve as basis approach to future predictions of HVAC systems reliability, and also as a tool to integrate external information into the assets managers decision-making process. The authors reckon that by further work it will be possible to extrapolate the methodology here proposed to reliability modeling for several assets, and it will enable a more efficient asset management by a better customized and accurate maintenance plan design in the offer stage.

REFERENCES

Baxter, M. et al. (1988). Proportional hazards modelling of transmission equipment failures. *Reliability Engineering & System Safety 21*(2), 129–144.

Bendell, A. (1985). Proportional hazards modelling in reliability assessment. *Reliability Engineering 11*(3), 175–183.

Bendell, A. et al. (1986). Proportional hazards modeling in reliability analysis. An application to brake discs on high speed trains. *Quality and reliability engineering international 2*(1), 45–52.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological) 34*(2), 187–220.

de Rocquigny, E. et al. (2008). *Uncertainty in industrial practice: a guide to quantitative uncertainty management*. John Wiley & Sons.

Foulliaron, J. et al. (2014). A specific dynamic Bayesian network for a prognosis based maintenance strategy. In *The Second International Conference on Railway Technology: Research, Development and Maintenance*, pp. 18p.

Jardine, A. et al. (1987). Application of the Weibull proportional hazards model to aircraft and marine engine failure data. *Quality and reliability engineering international 3*(2), 77–82.

Jardine, A. et al. (1989). Proportional hazards analysis of diesel engine failure data. *Quality and Reliability Engineering International 5*(3), 207–216.

Lawless, J. (1983). Statistical methods in reliability. *Technometrics 25*(4), 305–316.

Li, L. et al. (2015). Cox-proportional hazards modeling in reliability analysis. A study of electromagnetic relays data. *IEEE Transactions on Components, Packaging and Manufacturing Technology 5*(11), 1582–1589.

Li, Z. et al. (2010). Change detection in the cox proportional hazards models from different reliability data. *Quality and Reliability Engineering International 26*(7), 677–689.

Lin, S. et al. (2016, Oct). A failure rate model for traction transformer based on phm considering multiple factors. In *2016 Prognostics and System Health Management Conference (PHM-Chengdu)*, pp. 1–6.

Louit, D.M. et al. (2009). A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data. *Reliability Engineering & System Safety 94*(10), 1618–1628.

Tang, Z. et al. (2014). Analysis of significant factors on cable failure using the cox proportional hazard model. *IEEE Transactions on Power Delivery 29*(2), 951–957.

Tiwari, A. & D. Roy (2013). Estimation of reliability of mobile handsets using cox-proportional hazard model. *Microelectronics Reliability 53*(3), 481–487.

Umiliacchi, P. et al. (2011). Predictive maintenance of railway subsystems using an ontology based modelling approach. *In Proceedings of 9th world conference on railway research, May*, pp. 22–26.

Waeyenbergh, G. & L. Pintelon (2004). Maintenance concept development: a case study. *International Journal of Production Economics 89*(3), 395–405.

Wightman, D. & A. Bendell (1986). The practical application of proportional hazards modelling. *Reliability engineering 15*(1), 29–53.

Xie, Q. et al. (2017). Cox proportional hazards modeling of blockage risk in vitrified clay wastewater pipes. *Urban Water Journal 14*(7), 669–675.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Opportunistic maintenance strategy for a train fleet under safety constraints and inter-system dependencies

H. Ghamlouch & A. Grall
ICD/LM2S – UMR CNRS 6281, Université de Technologie de Troyes, Troyes, France

ABSTRACT: The aim of this work is to propose a modeling framework to call into question the maintenance regulations of a train fleet and improve an enforced maintenance policy without disrupting its main structure. The considered fleet is constituted of identical trains. A fixed periodic preventive maintenance planning is imposed by the manufacturers for safety and operational quality requirements. Each overhaul includes a set of pre-defined tasks such as inspection, minimal repair activities and preventive replacement of some components. During inspection if the degradation level of a component exceeds a primary replacement threshold it is precautionary replaced. In case of failure of at least one component when the train is in operation, it is immediately replaced. In this paper we propose the introduction of opportunistic maintenance activities to improve the periodic maintenance strategy while operating within the constraints of imposed planning. Opportunistic maintenance can include early revisions or precipitate preventive replacements of some component (with respect to scheduled preventive replacement or their primary replacement threshold). Dynamic maintenance task can be applied additionally during corrective activities or pre-scheduled revisions. Dependence matrices are used in order to consider stochastic dependencies between the system's components. Economic dependence is basically derived from a tree-like set-up model. A specific replacement threshold dedicated to opportunistic preventive replacement as well as a flexibility degree of revision dates are introduced. Multiple constraint are also considered for optimization: availability, work hours and number of simultaneously revised systems. A comparison between the periodic maintenance strategy with and without opportunistic tasks is presented.

1 INTRODUCTION

In transportation industry, production, operation and maintenance costs along with system's reliability and safety can be very decisive in the business success and continuity. In this context, management tools adopted by manufacturing and operators companies are continuously enhanced in order to increase systems' reliability as well as passengers safety and comfort. These tools specially allows to determine operation scheduling methods and maintenance strategies.

The optimization of maintenance policies in manufacturing organization is not a recent subject for researchers (Barlow and Hunter 1960). A first review of issues and results concerning maintenance scheduling problems and how crucial this later can be on the productivity and competitiveness of both manufacturing and service organization was pointed out by Paz and Leigh 1994. Ever since, maintenance strategies development and optimization have been considered in different domains, for instance: energy production systems (Sikos and Klemes 2010), (Shafiee 2015), transportation (Fritzsche et al. 2014, Liden 2015), networks (Morcous and Lounis 2005), infrastructures (Rios-Mercado and Borraz-Sánchez 2015), etc.

Maintenance strategies are usually classified in two categories: 1) corrective maintenance where maintenance is held when the system fails and 2) preventive maintenance where maintenance is held in a precautional way in order to prevent system failure. This later is usually adopted for systems with severe failure consequences. In a preventive maintenance strategy the systems components are maintained or replaced according to specific rules that usually consider system's age, deterioration level, operation time, etc. Depending on the system nature, different criteria can be considered for maintenance policy assessment and optimization such as mean cost rate, availability, downtime or combined economic and reliability criteria. A detailed survey of maintenance policies and deterioration modeling as well as optimization criteria is given by Wang (2002). A recent review of past and current research on optimal maintenance policy selection issues in manufacturing domain is given by Ding and Kamarudding (2011). Maintenance mathematical models as well as numerical algorithms for optimization were respectively reviewed by Vasili et al. (2011) and Alrabghi and Tiwari (2015).

Most of maintenance optimization studies in railway field have been focused on railway infrastructure maintenance Liden (2015), Bianchini Ciampoli

et al. (2017). This later can consume very large budgets has numerous challenging planning problems. Besides of the railway infrastructure, the main part of railway system is the rolling stock. Time spent in the maintenance of rolling stock may disrupt the operation and can cause financial losses as well as customers dissatisfaction. A major challenge for trains operators is to ensure required train service with limited rolling stock units. An operator should also respect the requirements of the trains manufacturer and security norms. This can be achieved if the correct maintenance strategy is adopted.

Despite the increasing efficiency of sensors and on line monitoring devices most of the maintenance strategies recommended by rolling stock manufacturers and adopted by railway operators are based on traditional periodic scheduling of maintenance operations, Eisenberger and Fink (2017). Classical periodic preventive maintenance strategies often lead to incorrect maintenance work, unnecessary maintenance tasks and often revert to corrective maintenance or breakdown maintenance (Rezvanizaniani et al. 2008). Different levels of inspections are defined and scheduled in a periodic manner according to a specified number of operation time or kilometers Sriskandarajah et al. (1998). Significant improvement in operation and maintenance cost could be obtained by implementing condition-based or predictive maintenance strategies which allow to anticipate system failure (Giacco et al. 2014). In complex systems the interaction between different components plays a major role in system deterioration and failure prediction and should be considered when setting the decision rules. Dependencies can offer more opportunities to group maintenance activities and reduce maintenance costs. Maintenance optimization for multi-component systems with internal dependencies is an important subject in maintenance optimization scholars (Nicolai and Dekker 2007). The main aim of this paper is to investigate the possibility to take advantage of opportunities to improve a traditional maintenance policy based on regulatory obligations.

This paper considers constraints of manufacturer and security instructions which are controlling the existing maintenance planning and cannot be modified easily. An opportunistic maintenance strategy is proposed for tram trains considering system dependencies. This work is a part of DIADEM project (Dynamic Diagnosis and Predictive Maintenance for Train Onboard Systems) which handles the case of Rennes's tram fleet. The company KEOLIS which is operating the Rennes's tram fleet is currently adopting a systematic preventive maintenance strategy that is scheduled according to manufacturer instruction and security norms. Tram fleet system description as well as the existing actual maintenance strategy are detailed in section 3. The approach proposed to add flexibility is presented in section 4. First the system and maintenance model are presented with strict maintenance rules. Then two new parameters are introduced: 1) the flexibility degree which allows the operator to modify the scheduled systematic replacement activities and 2) the secondary replacement threshold which allows the operator to anticipate replacement activities when it is beneficial. Finally, empirical results from numerical simulation are given and discussed.

2 COMPLEX SYSTEMS AND DEPENDENCIES

In order to manage system health and maintenance scheduling for multi-component systems a good understanding of components interactions and dependencies is essential. According to Thomas (1986) components dependencies can be classified as economic, structural or stochastic dependencies. Economic dependence implies that grouping maintenance activities causes variation of the maintenance cost. Stochastic dependence means that the conditions (or status) of two or more components are related either in a deterministic or probabilistic way. In case of structural dependences the maintenance of one component requires the maintenance or at least dismantling of others. In this section a brief review of system dependencies and their modeling in maintenance optimization is presented.

2.1 *Economic dependence*

Economic dependence means that the cost of a joint maintenance on several components is not equal to the sum of their individual maintenance costs.

In this paper we consider positive economic dependence which usually refers to economies of scale. It takes into account the reduction of setup activities and costs when several components are maintained simultaneously (Papadakis and Kleindorfer 2005). The economies of scale have been introduced to maintenance optimization in multiple forms. A single setup model is considered by Schouten et al. (1998) in order to evaluate preventive and opportunistic maintenance policies for traffic control lights. "Single setup" means that for simultaneous maintenance of two or more components a specific setup cost is paid only once. The single setup model is commonly used in maintenance optimization, see for example (Castanier et al. 2005), (Scarf and Deara 2003) and (Budai et al. 2004). Different setup activities can also be considered, leading to multiple setup models. In this case dif-

ferent components may or not share one or more setup activity. This model has been used by van Dijkhuizen (2000) where a tree-like setup structure was developed in order to reduce maintenance cost for an industrial production system. Another form of positive dependence is the downtime opportunity. Some components failures can be considered as an opportunities for preventive maintenance e.g. because they cause a system shutdown. The downtime required for corrective maintenance allows preventive maintenance and/or inspection in shared time which reduces the total downtime costs. This model is basically used in transportation field because of the nature of the system operation and maintenance (Sheu and Jhang 1997, Higgins 1998 and Sriskandarajah et al. 1998).

2.2 *Stochastic dependence*

Stochastic dependence occurs as the result of physical or operational interaction between components. The status of one component may influence the deterioration level or the lifetime of others. Stochastic dependence can be classified into three different types (Murthy & Nguyen 1985). The first type (type I) is when the failure on component i may introduce complete failure of component j with a certain probability p. In this case the failure of component i is called natural failure, while the failure of component j is considered as induced failure. This model has been considered for maintenance e.g. in (Scarf & Deara 2003) and (Jhang & Sheu 2000). Type II is defined when the failure of component i can induce a failure of component j with probability p but failure of component j induces a direct failure of component i. In case of deteriorating components, failure of component i acts as a shock to component j increasing its deterioration level with a random amount. Component j fails when the total deterioration level exceeds a failure level. This model has been e.g. used in context of maintenance optimization in Satow & Osaki (2003). The third type of stochastic dependence considers that failure of each components acts like a shock on other components, inducing a random increase on their failure rate or deterioration levels. It was first used by Özekici (1988).

2.3 *Structural dependence*

Structural dependence between multiple components implies that the maintenance of these components should always be applied simultaneously. Related components can not be maintained individually. For systems with structural dependence both opportunistic maintenance policy and preventive maintenance policy seems to be advantageous (Nicolai & Dekker 2007).

3 DIADEM PROJECT AND THE CASE OF RENNES' TRAM TRAIN

DIADEM project focuses on the development of diagnostic and prognostic tools for sensitive components of railway rolling stocks. The final aim is to develop new decision-making tools for maintenance management. This paper focuses on one collateral objective which is to question an actual current maintenance policies and to propose improvements while preserving the same rules. A fleet of 12 identical trains is considered. The tram fleet is operating in an alternative way in order to ensure the requested service availability of $P = 8$ out of $N = 12$ trains. During stand-by phase the non operating trains can be held in the main station in order to cover any additional load, for example in the case of a train failure, or are brought to the workshop in order to undergo maintenance activities.

The operation of different units (trains) is controlled by a special management tool which accords daily services to the train fleet according to three activity profiles (intense, medium or low activity profile). Note that different operation modes for different trains implies that, some trains will undergo faster deterioration process than the others. Thus, preventive maintenance scheduling and probability of failure are distributed with respect to the time interval which prevents the overcrowding of the maintenance workshop.

3.1 *The applied maintenance strategy*

The actual maintenance strategies adopted by the operator of Rennes' tram fleet is described hereafter. A sequence of periodic revisions is imposed by the manufacturer for safety and operational quality requirements.

Additionally, online monitoring system is available on each train and connected to a control room in the workshop. The monitoring system notifies the operator of any malfunction in any of the train's monitored equipment. No matter of how serious or dangerous the malfunctioning can be or not, the operation of the faulty train is immediately interrupted. The train is temporary replaced and directed to the workshop for corrective maintenance. Revisions and preventive maintenance activities are not allowed during the corrective maintenance period. No matter how short is the time before the expiry of the next revision period, the faulty train should be returned to operation mode directly after the application of corrective maintenance.

3.2 *System components and subsystems*

The considered tram system consists of 12 identical trains. Each train is composed of two cars

connected by a coupling system. Referring to trains operational and maintenance manuals given by the manufacturer, trains subsystems are defined according to their operational function. Thus a single subsystem may include several components that are physical distributed in one or both cars.

Nine types of embedded subsystems are considered. Each subsystem consists of multiple components which can be physically distributed in different locations of the trains. 16 different types of components are recorded. The list of components and subsystem under consideration as well as their quantities and other characteristics are presented in Table 1.

The lifetime law of the components are derived from both manufacturing maintenance manuals and the results of the first step of DIADEM project where historical data on trains maintenance activities and failures are collected. For each component type a set of useful information including replacement cost and the total lifetime probability parameters (considered to follow a Weibull function) is available.

In the next section a description of inter-system dependencies and their modeling is presented.

4 THE PROPOSED OPPORTUNISTIC MAINTENANCE APPROACH

As explained in previous section, the adopted maintenance strategy is based on the application of periodic preventive maintenance activities that are strictly scheduled and are coupled with corrective maintenance activities in case of failure. Maintenance activities are performed in the company's maintenance workshop with limited capacities in terms of space and manpower.

According to the company records, and under the current maintenance and operation policies, the tram system ensures a global availability of 0.98. Nevertheless, improvements can be reported on the current maintenance strategy in order to reduce operation costs and because of future requirements on transportation system capacities and availability.

In this paper we show that the consideration of inter-system dependencies and the addition of some flexibility degrees to the preventive maintenance schedule may introduce gains on maintenance costs without harming system reliability and availability. Our proposed approach relies on the addition of opportunistic maintenance actions to the mandatory mixed maintenance strategy. Opportunistic maintenance requires that earlier revisions or precipitate preventive replacements may be permitted under specific rules.

In order to cover the continuity of transportation service at the required frequency, 8 trains out of 12 must be operating simultaneously at any time. Non-operating trains are put on stand-by mode or sent to the workshop for maintenance operations. Note that the operator's workshop dispose of exactly $n = 4$ tracks where non-operating trains can be received. As for today, limitation on financial, technical and human resources of the operator company have not caused any obstruction on tram service and trains availability.

4.1 *The modeling of system operation and dependencies*

Trains operation scheduling is devoted to a special software that uses a selection algorithm designed for this purpose. According to this algorithm, the running schedule of different trains is established so that the dates of systematic maintenance interventions would be distributed in order to avoid the overcrowding of the maintenance workshop.

Because of their physical locations and functional interactions, different components from

Table 1. List of trains' subsystems and components.

N	Subsystem	Component	Components	Additional information
1	Traction	– Inverter – Motor	1	
2	Bearings	– Wheels – Guide frame – Sensors Antennas	4	2 in each car
3	Brake system	–Pads and discs – Control Unit	8	4 in each car
4	Doors	– Body – Motor	8	4 in each car
5	Pneumatic system	– Compressed air generator – Consumer	1	
6	Low voltage system	– Batteries – Converter	1	Distributed between 2 cars
7	Coupling system	–	1	Connecting the 2 cars
8	Body	–	1	
9	Automatic control unit	–	1	Situated in car A

different subsystems may present several types of dependencies. Noting that components with structural dependence are considered as a single part, we are only interested by the modeling of stochastic and economic dependencies.

Economic dependence: Setup levels tree
Economic dependence is the natural result of the physical installation of different components. According to system and components description given by the manufacturer manual as well as technical experience of the maintenance team, the setup level of each component is deduced as follows.

We first consider that each train is composed of two main blocks represented by car A and car B. Then each car is divided into smaller block that includes different components and equipment from different subsystems. Block definition and components grouping are done according to their position inside the car. Thus, components of different subsystem may belong to the same block while the components of one subsystem are not necessary belonging to a single block. One simple example is the position of pads and iscs from the braking system which are distributed into 4 blocks, along the two cars. However, pads and discs of one car are installed on one side and the other of its bearing system.

The division of systems components into blocks facilitates the definition of setup activities required for reaching each component. Components requesting the same setup activities in order to be reached and maintained are said to be on the same setup level. Different setup levels can be related according to a sequential order. However the division of setup levels is not straightforward. Setup levels are divided according to a tree-like model. The root of setup tree is considered as the simple activity of bringing the train into the workshop. Disassembling activities of different blocks and components are then classified into their correct position inside the setup levels tree. For each setup level basic setup time and costs are then assigned. Information about the setup levels required by each component are also included in a setup information table. During revision or corrective maintenance activities and in order to reach to specified component, setup activities of its associated setup levels must be applied.

The effect of economic dependence appears when two components having a shared setup level or belonging to the same block are jointly maintained. In this case, to reach both components during the maintenance session, common setup activities should be applied only once. Thus, the setup cost and downtime time of the concerned train can be reduced.

Stochastic dependence matrix
Depending on their physical or operational relationship, stochastic dependences between components may exist or not. The considered dependence are as follows: after failure of one component a shocks occur on the health condition of dependent components. The shock is considered as a random amount of additional damage which follow a normal distribution. Hence a stochastic dependence matrix is proposed to gather cross dependences between the components of the system. The element *(i,j)* of the stochastic matrix represents the effect of failure of a component of type *i* on the component of type *j*. This element consists of four parameters:

- The first parameter is defined as a boolean variable that determines either or not the failure of a component *i* will affected the status of other components of type *j*.
- The second parameter is also defined as a boolean variable that specifies whether the stochastic relation between components of types *i* and *j* is conditioned by the fact that both components belongs to same block or car.
- Parameters 3 and 4 represent the mean and variance of the normal distribution that quantifies the additional damage caused by the failure of component *i* on the deterioration level on the component of type *j*.

One can notice that the definition of the stochastic matrix allows the representation of different stochastic dependence types for different components by simply changing the parameters of its element. The parameters of the stochastic matrix in our case study are estimated according to component interaction represented in the manufacturer manual as well as historical records of components failure.

4.2 *Flexibility degree and secondary replacement parameters*

As described in section 3.1, the currently adopted maintenance strategy is a combination of preventive and corrective strategies. Preventive maintenance activities are grouped into 3 types of systematic revisions which are periodically scheduled for each train according to its rolling distance. At each revision the specified train should be directed to the maintenance workshop and a list of different inspections, cleaning, minimal maintenance activities and components replacements should be applied. For each type of revision a different list of maintenance tasks is defined. Maintenance tasks are carefully selected according to the manufacturer's specifications, components lifetime estimation as well as feedback from the workshop technicians experience. During revision, replacement of non-defected components should be applied if the inspection activities reveal that the deterioration level of this component exceed a predefined replacement threshold (which we will call the pri-

mary threshold). Each revision type requires a minimal revision cost and minimal downtime of the system which are both registered in association of the maintenance tasks list. The total cost of a revision is the sum on minimal revision cost setup costs and the eventual cost of parts replacement.

According to the actual maintenance strategy, corrective maintenance is applied to the faulty train immediately after the detection of faulty behavior. However revisions and preventive maintenance activities are not allowed during corrective maintenance period. Thus, the total corrective maintenance cost includes setup costs, replacement costs as well as any penalty costs in the case of availability loss.

Opportunistic maintenance

An opportunistic maintenance policy is proposed in order to take better advantage of corrective downtime and setup of the system. Early revisions and preventive activities will be allowed according to specified rules. This can be done by adding two new parameters to the maintenance decision rule:

- Flexibility degree: it is applied to revisions periods and allows the operator to perform early revisions. When a faulty train is directed to the workshop for corrective maintenance, the technical team is required to perform the next planned revision if the expiry date of the revision period is close enough.
- Secondary replacement threshold: during the corrective maintenance of a given component, technical team is allowed to inspect and replace other components on the same setup level if their deterioration level exceeds a secondary replacement threshold. The application of this rule may introduce savings on setup costs and also prevents future failure of the system.

In next sections, the addition of these two parameters and the application of the two additional rules is described and discussed.

4.3 *Simulated operation and maintenance: An empirical example*

In this step we are going to test the new opportunistic maintenance approach for the case of Rennes's tram system using numerical and simulation methods. Simulation are based on the following system description: (i) Train fleet consist of 12 trains; (ii) The maintenance workshop dispose of 4 maintenance tracks; (iii) 8 trains must be in service mode continuously in order to ensure the required load.

As explained previously, three types of periodic revisions are defined. Trains are directed to the workshop for revision when their running distance reach one of the periodic revision distance. Running and maintenance of different trains are scheduled by specific software in order to ensure the continuity of transport service and correct scheduling of required revisions. If the load is not handled because of a low number of available trains, a penalty charge is imposed. Penalty costs are calculated per day according the number of lacking trains.

Different data and parameters are required, especially about dependencies and revisions.

- About maintenance baseline: for each type of revision a revision periodicity, maintenance operations as well as minimal cost and duration are imposed.
- About trains states: information about the last revision of each type, running distance, total time spent in the workshop, time and cost spent for all maintenance activities are registered as well as all historical information about failures.
- About components: the type of the component, the block and car it belongs to and its deterioration level are recorded.
- About dependencies: the Stochastic dependence matrix is given.
- About maintenance costs: the setup matrix includes information about duration and cost of each setup level.

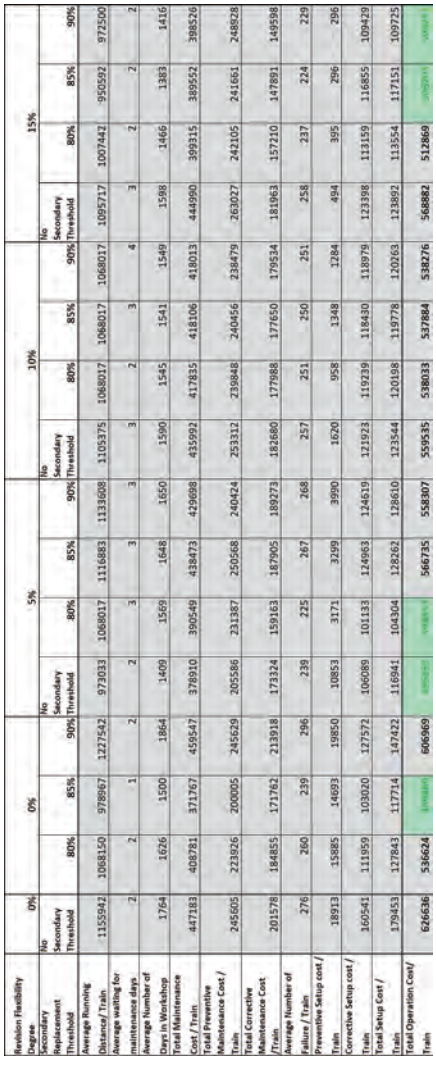
According to the maintenance strategy, two different cases are compared. The first case present the imposed maintenance policy as it is currently adopted by the train operator. The second one takes advantage of opportunistic maintenance as described in section 4.2 by introducing flexibility degree and secondary replacement threshold. In both cases trains operation and revisions are scheduled using the same method. When the monitoring system detects a malfunctioning on one of the trains, the operation of this train is interrupted and the train is directed to the workshop for corrective maintenance. Components deterioration and lifetime probabilistic rules are also the same. The two algorithms differ in the following:

First results

Simulation of trains operation and maintenance is held for 60 years. And the empirical results regarding preventive and corrective maintenance as well as setup total costs of the train fleet are given in Table 1. The first column of this table represent the case of initial maintenance policy without opportunistic maintenance. The rest of the table presents the case where opportunistic maintenance actions are introduced with different flexibility degrees and secondary replacement threshold. Note that secondary replacement threshold is represented in % with respect to the primary replacement threshold. Results in this table show the advantage of opportunistic maintenance policies in mixed maintenance

strategy for the case of Rennes' tram fleet. The major advantage is due to savings on setup costs which leads to a smaller cost for operation per train.

The best result is found with no flexibility degree on revisions periods but introducing a secondary replacement threshold of 85% of the primary replacement threshold. Secondary replacement threshold identifies when non-defected components can be replaced during corrective maintenance of other components. This preventive maintenance activity introduce no cost on preventive setup cost because it benefits from corrective setup activities to reach all related components.



Revision Flexibility Degree	0%					5%					10%					15%				
Secondary Replacement Threshold	No Secondary Threshold	80%	85%	90%	No Secondary Threshold	80%	85%	90%	No Secondary Threshold	80%	85%	90%	No Secondary Threshold	80%	85%	90%	No Secondary Threshold	80%	85%	90%
Average Running Distance/ Train	1155942	1068150	978967	1227542	973033	1068017	1116883	1133608	1105375	1068017	1068017	1068017	1095717	1007442	950592	972500				
Average waiting for maintenance days	2	2	1	2	2	3	3	3	3	2	3	4	3	2	2	2				
Average Number of Days in Workshop	1764	1626	1500	1864	1409	1569	1648	1650	1590	1545	1541	1549	1598	1466	1383	1416				
Total Maintenance Cost / Train	447183	408781	371767	459547	378910	390549	438473	429698	435992	417835	418106	418013	444990	399315	389552	398526				
Total Preventive Maintenance Cost / Train	245605	223926	200005	245629	205586	231387	250568	240424	253312	236848	240456	238479	263027	242105	241661	248828				
Total Corrective Maintenance Cost /Train	201578	184855	171762	213918	173324	159163	187905	189273	182680	177988	177650	179534	181963	157210	147891	149598				
Average Number of Failure / Train	276	260	239	296	239	225	267	268	257	251	250	251	258	237	224	229				
Preventive Setup cost / Train	18913	15885	14093	19850	10853	3171	3299	3990	1620	958	1348	1284	494	395	296	296				
Corrective Setup cost / Train	160541	111959	103020	127572	106089	101133	124963	124619	121923	119239	118430	118979	123398	113159	116855	109429				
Total Setup Cost / Train	179453	127843	117714	147422	116941	104304	128262	128610	123544	120198	119778	120263	123892	113554	117151	109725				
Total Operation Cost/ Train	626636	536624	[illegible]	606969	[illegible]	[illegible]	566735	558307	559535	538033	537884	538276	568882	512869	[illegible]	[illegible]				

Figure 1. Operational information and costs for one train simulated with mixed and opportunistic maintenance strategy using different maintenance secondary threshold and flexibility degrees.

A very close result is found for the opportunistic maintenance with 5% of flexibility degree and secondary replacement thresholds of 80% and 85%. This is also due to savings on setup cost and failure costs. Looking at unavailability durations and average failure rate of trains we also notice that the addition of opportunistic maintenance policy does not introduce any harm to the system availability. However, results in this table suggest that the relation between economic gains and additional decision parameters (flexibility degree and secondary threshold) is not linear nor straightforward. Numerical optimization methods should be implemented for each special case in order to choose the best flexibility parameters to be used.

5 CONCLUSION

Maintenance strategy and optimization is a major concern of operators company in transportation industry. Due to security and safety reasons, the flexibility with maintenance activity planning is very low on the considered fleet which is derived from Rennes' tram train fleet. Additional constraints are due to high request on availability and limitation on maintenance facilities and labor. Trains are maintained according to a strictly scheduled preventive maintenance planning which involves three types of periodic preventive revisions. This paper considers new degrees of freedom in the maintenance policy which are related to additional opportunistic maintenance actions based on dependencies. A dependence matrix and a setup "tree like" model are considered in-order to characterize stochastic and economic dependencies between different components of trains. Two "flexibility parameters" are introduced in the maintenance decision rule. The first one is the flexibility degree which allows the operator to perform earlier revisions. And the second one referred as "secondary maintenance degree" allows the operator to perform preventive replacement of non-defective parts during corrective maintenance of other components.

The effect of introducing flexibility parameters into maintenance strategy has been tested numerically. The results show the interest of opportunistic maintenance and suggest that it could be very advantageous in terms of maintenance cost savings without harming system availability. However, the choice of flexibility parameters is not obvious. Advanced numerical methods should be used in order to optimize the decision variables. The effects of different failures on system security and passengers safety have not been considered in this study. A fixed penalty cost is used to quantify the unavailability cost caused by the failure of any of the train's components. Further studies may focus

on types and effects of system failures in order to take them into consideration in maintenance optimization process.

REFERENCES

Alrabghi, A. & A. Tiwari (2015). State of the art in simulation-based optimisation for maintenance systems. *Computers & Industrial Engineering 82* (Supplement C), 167–182.

Barlow, R. & L. Hunter (1960). Optimum Preventive Maintenance Policies. *Operations Research 8*(1), 90–100.

Bianchini Ciampoli, L., F. Tosti, A. Calvi, A. Alani, & A. Benedetto (2017). Efficient practices in railway ballast maintenance and quality assessment using GPR. *Transport Infrastructure and Systems: Proceedings of the AIIT International Congress on Transport Infrastructure and Systems—CRC Press 419–424*.

Budai, G., D. Huisman, & R. Dekker (2004). *Scheduling preventive railway maintenance activities*, Volume 5. 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583.

Castanier, B., A. Grall, & C. Bérenguer (2005). A condition-based maintenance policy with non-periodic inspections for a two-unit series system. *Reliability Engineering and System Safety 87*(1), 109–120.

Ding, S.-H. & S. Kamarudding (2011). Maintenance policy optimization—literature review and directions. *The international journal of advanced manufacturing technology 76*(5–8), 1263–1283.

Eisenberger, D. & O. Fink (2017, January). Assessment of maintenance strategies for railway vehicles using Petrinets. *Transportation Research Procedia 27*, 205–214.

Fritzsche, R., J. N. D. Gupta, & R. Lasch (2014). Optimal prognostic distance to minimize total maintenance cost: The case of the airline industry. *International Journal of Production Economics 151*(Supplement C), 76–88.

Giacco, G. L., D. Carillo, A. D'Ariano, D. Pacciarelli, & Á. G. Marín (2014). Short-term Rail Rolling Stock Rostering and Maintenance Scheduling. *Transportation Research Procedia 3*(Supplement C), 651–659.

Higgins, A. (1998, October). Scheduling of railway track maintenance activities and crews. *Journal of the Operational Research Society 49*(10), 1026–1033.

Jhang, J.-P. & S.-H. Sheu (2000). Optimal age and block replacement policies for a multi-component system with failure interaction. *International Journal of Systems Science 31*(5), 593–603.

Liden, T. (2015). Railway Infrastructure Maintenance—A Survey of Planning Problems and Conducted Research. *Transportation Research Procedia 10*, 574–583.

Morcous, G. & Z. Lounis (2005). Maintenance optimization of infrastructure networks using genetic algorithms. *Automation in Construction 14*(1), 129–142.

Murthy, D.N. P. & D. G. Nguyen (1985). Study of a multicomponent system with failure interaction. *European Journal of Operational Research 21*(3), 330–338.

Nicolai, R. & R. Dekker (2007). *Optimal Maintenance of Multi-Component Systems—A review*. Complex System Maintenance Handbook, London Springer.

Özekici, S. (1988). Optimal Periodic Replacement of Multicomponent Reliability Systems. *Operations Research 36*(4), 542–552.

Papadakis, I. S. & P. R. Kleindorfer (2005, January). Optimizing infrastructure network maintenance when benefits are interdependent. *OR Spectrum 27*(1), 63–84.

Paz, N. & W. Leigh (1994). Maintenance Scheduling: Issues, Results and Research Needs. *International Journal of Operations & Production Management 14*(8), 47–49.

Rezvanizaniani, S. M., M. Valibeigloo, M. Asghari, J. Barabady, & U. Kumani (2008). Reliability Centered Maintenance for rolling stock: A case study in coaches #x2019; wheel sets of passenger trains of Iranian railway. *2008 IEEE International Conference on Industrial Engineering and Engineering Management*, 516–520.

Rios-Mercado, R. Z. & C. Borraz-Sánchez (2015). Optimization problems in natural gas transportation systems: A state-of-the-art review. *Applied Energy 147* (Supplement C), 536–555.

Satow, T. & S. Osaki (2003). Optimal replacement policies for a two-unit system with shock damage interaction. *Computers & Mathematics with Applications 46*(7), 1129–1138.

Scarf, P. A. & M. Deara (2003). Block replacement policies for a two-component system with failure dependence. *Naval Research Logistics (NRL) 50*(1), 70–87.

Schouten, F. V. D. D., B. V. Vlijmen, & S. V. D. Wael (1998). Replacement Policies fro traffic control signals. *IMA Journal of Management Mathematics 9*(4), 325–346.

Shafiee, M. (2015). Maintenance logistics organization for offshore wind energy: Current progress and future perspectives. *Renewable Energy 77*, 182–193.

Sheu, S.-H. & J.-P. Jhang (1997). A generalized group maintenance policy. *European Journal of Operational Research 96*(2), 232–247.

Sikos, L. & J. Klemes (2010). Reliability, availability and maintenance optimisation of heat exchanger networks. *Applied Thermal Engineering 30*(1), 63–69.

Sriskandarajah, C., A. K. S. Jardine, & C. K. Chan (1998). Maintenance scheduling of rolling stock using a genetic algorithm. *Journal of the Operational Research Society 49*(11), 1130–1145.

Stengos, D. & L. C. Thomas (1980). The blast furnaces problem. *European Journal of Operational Research 4*(5), 330–336.

Thomas, L. C. (1986). A survey of maintenance and replacement models for maintainability and reliability of multi-item systems. *Reliability Engineering 16*(4), 297–309.

van Dijkhuizen, G. (2000). Maintenance Grouping in Multi-Step Multi-Component Production Systems. In M. Ben-Daya, S. O. Duffuaa, and A. Raouf (Eds.), *Maintenance, Modeling and Optimization*, pp. 283–306. Boston, MA: Springer US. DOI: 10.1007/978-1-4615-4329-9_12.

Vasili, M., T. S. Hong, N. Ismail, & M. Vasili (2011). Maintenance optimization models: a review and analysis. *Optimization 1*(2).

Wang, H. (2002). A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research 139*(3), 469–489.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Alternative Weibull analysis for road markings: An EM approach

Maxime Redondin & Nadège Faul
VEDECOM Institute, Versailles, France

Laurent Bouillaut & Allou Samé
Université Paris-Est, Grettia (IFSTTAR), Marne-la-Vallée, France

Dimitri Daucher
Université Paris-Est, Lepsis (IFSTTAR), Marne-la-Vallée, France

ABSTRACT: The quality and reliability of road infrastructure and its equipment play a major role in road safety. This is especially true for autonomous car traffic guided mainly by a GPS system that is, unfortunately, neither precise nor reliable. In order to improve the guidance systems, one option could be to equip the vehicle with a camera reading road markings. Such solution require maintenance strategies guaranteeing markings' perceptibility to the human eye or the autonomous car camera. Currently, the retroreflection luminance of markings is measured for evaluating marking degradation. An important remaining step is a life time analysis depending on the inspection strategy. Since the exact failure time isn't generally observed, feedback database contain many censured data: the left-censure corresponding to a marking failing before the first inspection, the interval-censure that corresponds to markings failing between two inspections, and the right-censure corresponding to a marking that never fails. In the literature, a Weibull analysis was proposed to estimate the markings reliable distributions using the Maximum Likelihood through the Newton-Raphson method. Facing with censored data, this approach couldn't be computed without introducing strong bias in the reliability estimation. For generic interval-censored data, Pradhan and Kundu proposed an alternative, based on the EM algorithm. In our study an extension of the EM algorithm processing left and right censures is proposed. Finally, this algorithm is applicable for all kind of observations, whatever the censure nature. After introducing this EM extension, the paper focuses on the fact that computations are simpler than the Newton-Raphson methods and censored-data are directly estimated. The French National Road 4 markings case is considered to illustrate the proposed approach. Moreover, the proposed algorithm being generic, its application is, of course, not limited to our road marking case study.

1 INTRODUCTION

The quality and reliability of road infrastructure and its equipment play a major role in road safety. This is especially true for autonomous car traffic. Currently, autonomous vehicles are guided mainly by a GPS system. Unfortunately, GPS systems are neither precise nor reliable. For example, GPS signals couldn't work in urban canyons or tunnels. In order to improve autonomous vehicle guidance systems, one option would be to equip the vehicle with a camera able to "read" road markings. However, this solution requires a maintenance strategy guaranteeing that road markings remain perceptible to a human eye or an autonomous car camera.

According to both the AFNOR rules (AFNOR 2009) and the available inspection devices, the retroreflection luminance of markings is the only measure used for evaluating marking degradation. A retroreflective marking reflects light from a vehicle headlight back in the direction of the driver. For waterborne markings, the retroreflective property is guaranteed by glass spheres mixed into the paint during application. The retroreflection luminance is measured in millicandela per square meter and by lux ($mcd/m^2/lx$). A minimum threshold of 150 $mcd/m^2/lx$ is required for a new marking (AFNOR 2009).

Several decay models for retroreflective marking exist in the current literature which mainly calculate retroreflective luminance based on a regression model. For example, Lu (1995) proposed an exponential regression model function of age of markings, Abboud & Bowman (2002) developed exponential models as a function of the Annual Average Daily Traffic (AADT) and the age of markings, Sarasua, Clarke, & Davis (2003) calculated the difference in reflectivity over time,

Sitzabee et al. (2009) proposed the most complete decay multilinear model as a function of time, the initial retroreflection, the AADT, the lateral locations of markings and marking color. All decay models have a common weakness: they are difficult to apply directly to a given road network. For example, consider a single road. The road is not systematically maintained in its entirety. For safety reasons, road managers maintain only specific areas at a time. This is especially true for road surface maintenance.

In a previous study, a clustering approach able to segment a road network according to past inspections was proposed (Redondin et al. 2017). Each cluster was interpreted with respect to a specific area of the road network and admits its own retroreflective luminance evolution over time. This fact leads to each cluster having its own optimum maintenance strategy. An important remaining step is to do a life time analysis. This work could confirms the necessity of one maintenance strategy by cluster and is the first step to develop any maintenance model.

Currently, road markings are monthly or yearly inspected by a retroreflectometer. Thus, such a periodical approach isn't able to determine the exact failure time for any marking. Moreover, in feedback database, three kinds of censure are observed: the left-censure corresponding to a marking failing before the first inspection, the interval-censure that corresponds to markings failing between two inspections, and the right-censure corresponding to a marking that never fails.

A Weibull analysis was proposed by Sathya narayanan et al. (2008), which consists of estimating the markings reliable function using a Weibull distribution. Its parameters are estimated according to the Maximum Likelihood Estimator (MLE). In attendance of censored-data, the likelihood function depends on the probability density, the cumulative distribution and the reliable function of a Weibull distribution. For interval-censored data, Pradhan and Kundu (2014) found instances where the Newton-Raphson does not compute. An alternative, based on the EM algorithm (McLachlan & Krishnan 2008) has been proposed.

Section 2 proposed an extension of the EM algorithm processing left and right censures. The proposed algorithm is applicable for all observation vectors, independently of the nature of the censure. After introducing this EM extension, this paper will focus on a simpler computations than the Newton-Raphson methods (Bain & Englehardt 1975). Morever, censored-data are directly estimated. Then, the proposed approach could be applied for other cases studies and are not restricted to roadmarkings considered in this paper.

As in, the French National Road 4 inspection database is considered to illustrate the proposed approach and its use for analysing maintenance cycles. Two applications are proposed: on the one hand, the first take into account the whole maintenance cycle and the EM Algorithm produces one global Weibull analysis. On the other hand, the second segements first the cycle according to a clustering approach (Redondin et al. 2017). Each cluster admits its own Weibull analysis taking into account different decay profiles.

2 IMPROVED EM ALGORITHM

2.1 *Censored Weibull distribution*

In this paper, a Weibull distribution $\mathcal{W}(\alpha,\beta)$ is defined by its associated probability density function (1) where $\alpha > 0$ and $\beta > 0$ are respectively the associated scale and shape parameters.

$$f(t) = \begin{cases} \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} e^{-\left(\frac{t}{\alpha}\right)^{\beta}} & \text{if } t > 0 \\ 0 & \text{if } t \leq 0 \end{cases} \quad (1)$$

The observed data $(t_1,...,t_n)$ is assumed independent identically distributed. In this paper, an observed failure is a failure established according to a given inspection strategy. $\forall i = 1,...,n$, t_i is assumed to be the first observed failure moment of i. Three censorship cases are assumed.

- If a failure has occurred before the first inspection, then it is called left-censored.
- If a failure has occurred between two inspections, then it is called interval-censored.
- If a failure hasn't been observed, then it is called right-censored.

If t_i is interval-censored, then its associated interval is $[l_i, r_i]$. The interval is clearly defined according to the current inspection strategy. A censor detector is associated for each observation as follows $\forall i = 1,..,n$.

$$\delta_i = \begin{cases} 0 & \text{if } t_i \text{ is uncensored} \\ 1 & \text{if } t_i \text{ is left-censored} \\ 2 & \text{if } t_i \text{ is interval-censored} \\ 3 & \text{if } t_i \text{ is right-censored} \end{cases} \quad (2)$$

From now, the observed data is subdivided according to the censor detector.

$$\mathcal{T} = \left\{ t \in \left((t_1, \delta_1),...,(t_n, \delta_n) \right) / \delta_i = 0 \right\} \quad (3)$$

$$\mathcal{X} = \left\{ x \in \left((t_1, \delta_1), ..., (t_n, \delta_n) \right) / \delta_i = 1 \right\} \quad (4)$$

$$\mathcal{Y} = \left\{ y \in \left((t_1, \delta_1), ..., (t_n, \delta_n) \right) / \delta_i = 2 \right\} \quad (5)$$

$$\mathcal{Z} = \left\{ z \in \left((t_1, \delta_1), ..., (t_n, \delta_n) \right) / \delta_i = 3 \right\} \quad (6)$$

In the case of an uncensored Weibull distribution, the MLE is the couple (α, β) which maximizes the associated likelihood function (7). This couple annuls also symultaneously the two partial derivates of the log-likelihood function. For the specific case of a Weibull distribution, the solution of this non-linear equations system is currently done by a Newton-Raphson approach (Bain & Englehardt 1975).

$$L(\alpha, \beta) = \prod_{i=1}^{n} f(t_i) \quad (7)$$

The likelihood function associated to a censored Weibull distribution (8) depends on the reliable function S for interval and right censorship cases and it also depends on the cumulative function $F = 1 - S$ for the left censorship case.

$$L(\alpha, \beta) = \prod_{t \in \mathcal{T}} f(t) \prod_{x \in \mathcal{X}} F(x) \prod_{y \in \mathcal{Y}} R(r) - R(l) \prod_{z \in \mathcal{Z}} R(z) \quad (8)$$

Reminder: let's take W a Weibull distribution $\mathcal{W}(\alpha, \beta)$, the reliable function (9) is the probability that the time of failure is later than some specified time $t > 0$.

$$R(t) = P(W > t) = e^{-\left(\frac{t}{\alpha}\right)^{\beta}} \quad (9)$$

For the road markings study, Sathyanarayanan et al (Sathyanarayanan, Shankar, & Donnell 2008) chose this classic approach for a Weibull analysis. In a breast cancer case study and restricted only to interval-censored data, Pradhan and Kundu (Pradhan & Kundu 2014) showed different examples where this approach does not compute. Specifically, the Newton-Raphson approach isn't able to estimate (α, β). An alternative based on a EM Algorithm is proposed. This paper proposed an extension of the algorithm, processing left and right censures.

The EM algorithm interprets censored data like missing data to estimate and computes iteratively into two steps:

1. The Expected Step estimates censored data according to a given Weibull distribution $\mathcal{W}(\alpha, \beta)$.
2. The Maximization Step estimates a Weibull distribution $\mathcal{W}(\alpha, \beta)$ by the MLE according to both uncensored and completed data.

The algorithm computes until one distribution converges. This point is detailed in section 2.4. First, sections 2.2 and 2.3 present the extended algorithm through one given iteration.

2.2 *Improved expected step*

Let's take W a Weibull distribution $\mathcal{W}(\alpha, \beta)$. To simplify the EM formalism, the substitution $\alpha \leftarrow 1/\alpha^{\beta}$ is suggested by Pradhan and Kundu (Pradhan & Kundu 2014). According to that, the density function is done by (10) in this section.

$$f(t) = \begin{cases} \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}} & \text{if } t > 0 \\ 0 & \text{if } t \leq 0 \end{cases} \quad (10)$$

The Expected Step calculates the estimated likelihood function defined as the maximized expectation of the likelihood function (7) conditioned to uncensored data (11).

$$L_c(\alpha, \beta) = E[L(\alpha, \beta) \mid \mathcal{T}] \quad (11)$$

This conditional expectation leads to produce three estimators $\forall x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}$. Each one is a conditional expectation adapted for one specific censure: (12) and (14) estimates respectively the left and right censures and (13) is proposed by Pradhan and Kundu (Pradhan & Kundu 2014) for the interval-censures case.

$$\hat{x} = E[W \mid W < x] = \frac{\alpha\beta \int_0^x t^{\beta} e^{-\alpha t^{\beta}} \, dt}{1 - e^{-\alpha x^{\beta}}} \quad (12)$$

$$\hat{y} = E[W \mid l < W < r] = \frac{\alpha\beta \int_l^r t^{\beta} e^{-\alpha t^{\beta}} \, dt}{e^{-\alpha l^{\beta}} - e^{-\alpha r^{\beta}}} \quad (13)$$

$$\hat{z} = E\left[W \mid W > z \right] = \frac{\alpha\beta \int_z^{+\infty} t^{\beta} e^{-\alpha t^{\beta}} \, dt}{e^{-\alpha z^{\beta}}} \quad (14)$$

Finally, the completed likelihood function associated to a censored Weibull distribution is defined by (15).

$$L_c(\alpha, \beta) = \prod_{t \in \mathcal{T}} f(t) \prod_{x \in \mathcal{X}} f(\hat{x}) \prod_{y \in \mathcal{Y}} f(\hat{y}) \prod_{z \in \mathcal{Z}} f(\hat{z}) \quad (15)$$

2.3 *Improved maximization step*

According to the MLE formalism, the couple (α,β) annuls simultaneously the two partial derivates of the completed log-likelihood function.

$$\partial_\alpha L_c = \frac{n}{\alpha} - \sum_{t\in\mathcal{T}} t^\beta - \sum_{x\in\mathcal{X}} \hat{x}^\beta - \sum_{y\in\mathcal{Y}} \hat{y}^\beta - \sum_{z\in\mathcal{Z}} \hat{z}^\beta \tag{16}$$

$$\partial_\beta L_c = \frac{n}{\beta} + \sum_{t\in\mathcal{T}} \ln\, t + \sum_{x\in\mathcal{X}} \ln\, \hat{x} + \sum_{y\in\mathcal{Y}} \ln\, \hat{y} + \sum_{z\in\mathcal{Z}} \ln\, \hat{z} - \alpha\left[\sum_{t\in\mathcal{T}} t^\beta \ln\, t + \sum_{x\in\mathcal{X}} \hat{x}^\beta \ln\, \hat{x} + \sum_{y\in\mathcal{Y}} \hat{y}^\beta \ln\, \hat{y} + \sum_{z\in\mathcal{Z}} \hat{z}^\beta \ln\, \hat{z} \right] \tag{17}$$

This non-linear equations system doesn't admit an obvious solution. Restricted to interval-censored data, Pradhan and Kundu (Pradhan & Kundu 2014) solved this non-linear equations system by a fixed point approach. This choice is due to an estimation of α done by the equation $\partial_\alpha L_c = 0$. Processing left and right censures, the extending estimator is:

$$\alpha = \frac{n}{\sum_{t\in\mathcal{T}} t^\beta + \sum_{x\in\mathcal{X}} \hat{x}^\beta + \sum_{y\in\mathcal{Y}} \hat{y}^\beta + \sum_{z\in\mathcal{Z}} \hat{z}^\beta} \tag{18}$$

This estimator (18) is completely dependent on β: an estimation of β is enough. Furthermore replacing α by (18) in (17), a function $g(\beta)$ as $\partial_\beta L_c = g(\beta) - \beta$ is extracted (19). To simplify, let's take $\mathcal{S} = \mathcal{X} \cup \mathcal{Y} \cup \mathcal{Z}$.

$$g(\beta) = \frac{1}{\dfrac{\sum_{t\in\mathcal{T}} t^\beta \ln t + \sum_{s\in\mathcal{S}} \hat{s}^\beta \ln \hat{s}}{\sum_{t\in\mathcal{T}} t^\beta + \sum_{s\in\mathcal{S}} \hat{s}^\beta} + \dfrac{\sum_{t\in\mathcal{T}} \ln t + \sum_{s\in\mathcal{S}} \ln \hat{s}}{n}} \tag{19}$$

The fixed point of g is the point β as $g(\beta) = \beta$ also defined as the convergence point of the sequence (20). The initial value is arbitrary. This point is detailed at section 2.4

$$(u_n)_{n\geq 0} = \begin{cases} u_0 > 0 & \text{arbitrary} \\ u_n = g(u_{n-1}) & n > 0 \end{cases} \tag{20}$$

To conclude, α is finally deduced from (18).

2.4 *Iteration process and convergence*

The Weibull distribution estimated at the algorithm iteration $k > 0$ is denoted $\mathcal{W}(\alpha_k, \beta_k)$. The initial distribution $\mathcal{W}(\alpha_0, \beta_0)$ is arbitrary. In this paper, α_0 and β_0 are the MLE where the censorships not taken into account.

Sections 2.2 and 2.3 show that one algorithm iteration could be reduced to an estimation of β done by the convergence point of (20) depending on the function g (19) conditioned by estimated data (12–14). As α_k is directly defined both by β_k and (18), the next sequence $(\beta_k)_{k\geq 0}$ is clear-defined as:

$$(\beta_k)_{k\geq 0} = \begin{cases} \beta_0 > 0 \\ \beta_k = \lim\limits_{n\to+\infty} \begin{cases} u_0 = \beta_{k-1} \\ u_n = g(u_{n-1}) \end{cases} \end{cases} \tag{21}$$

$|\beta_k - \beta_{k-1}| < 10^{-4}$ is the selected stopping criterion. Finally, the EM Algorithm convergence is defined both by the convergence of the sequence $(\beta_k)_{k\geq 0}$ and (18):

$$\begin{cases} \beta = \lim\limits_{k\to+\infty} \beta_k \\ \alpha = \dfrac{n}{\sum_{t\in\mathcal{T}} t^\beta + \sum_{x\in\mathcal{X}} \hat{x}^\beta + \sum_{y\in\mathcal{Y}} \hat{y}^\beta + \sum_{z\in\mathcal{Z}} \hat{z}^\beta} \end{cases} \tag{22}$$

Again, the stopping criterion is $|\beta_k - \beta_{k-1}| < 10^{-4}$. To conclude, the substitution $\alpha \leftarrow 1/\sqrt[\beta]{\alpha}$ returns the current Weibull distribution (1).

3 MAINTENANCE CYCLE 2008–2012

3.1 *Presentation of the French National Road 4*

The broken centerline of the French National Road 4 (NR4) is considered to illustrate the proposed approach. The NR4 runs between Paris and Strasbourg. Since 2007, the section of this road between Courgivaux and Vauclerc (~102 km) has been managed by the DIR Est, which inspects this section of the road once a year. Inspections are organized in September in collaboration with CEREMA Est. The selected retroreflectometer is an Ecodyn.

The maintenance cycle selected is composed of 73 measures annotated with a PR. To simplify, each measure is interpreted as one marking. Markings laid in March 2008 and replaced in March 2012. The marking material chosen is supposed be the same. The cycle is localized around three cities: Courgivaux, Sommesous and Vitry-le-François. The direction heading toward Vauclerc is chosen.

Four inspections are available: 6, 18, 30 and 42 months (after March 2008). The retroreflection luminance of a given marking i at the inspection point t is denoted $RL_t(i) \in \mathbb{N}^*$. If $RL_t(i) \leq 150\ mcd/m^2/lx$ then the marking i is failed at t. Let's take τ be the first time when a given marking is observed failing.

$$\tau = \min\{t \in \{6,18,30,42\} / RL_t(i) \leq 150\} \tag{23}$$

According to τ, t_i is the first time when the marking $i = 1,...,73$ is observed failing. If a marking i hasn't been observed, then the marking is right-censored and $\tau = \varnothing$. In this situation, $t_i = 42+$ tentatively.

$$t_i = \begin{cases} \min \tau & \text{if } \tau = \varnothing \\ 42+ & \text{else} \end{cases} \quad (24)$$

The adapted censor detector is deduced (25). From now on, if $t_i = 42+$ them $t_i = 42$. This formalism assumes three censorship intervals ([6,18], [18,30] and [30,42]) and t_i isn't an uncensored observation. Finally, the observed data is finally defined as $\left((t_1,\delta_1),...,(t_{73},\delta_{73})\right)$.

$$\delta_i = \begin{cases} 1 & \text{if } t_i = 6. \\ 2 & \text{if } t_i = 18,30,42. \\ 3 & \text{if } t_i = 42+. \end{cases} \quad (25)$$

According to inspection campaigns, the monitoring of the retroreflective luminance is presented by the Figure 1. All censor case are presented: two failures are observed during the 6th month (September 2008), a group of markings which never failed until the 42nd month (September 2011) exists and the majority of failure is interval-censored. At least two decay models are also presented.

Two approaches could be produced: a global and a clustering Weibull analysis. The global analysis consists in selecting the whole monitoring and estimating one global Weibull distribution. This approach is interesting if the observed data is small. This situation could happen for two reasons. Firstly, current retroreflectometer like the Ecodyn produce generally one average measure every 100 m. This fact reduces considerably the data set size. Secondly, current maintenance strategy could concern only one specific area. The clustering analysis consists in segmenting first the whole monitoring and estimating one Weibull distribution by cluster. This approach is interesting to produce one specific maintenance strategy by cluster over time, but the observed data must be large and must present a diversified censorship case.

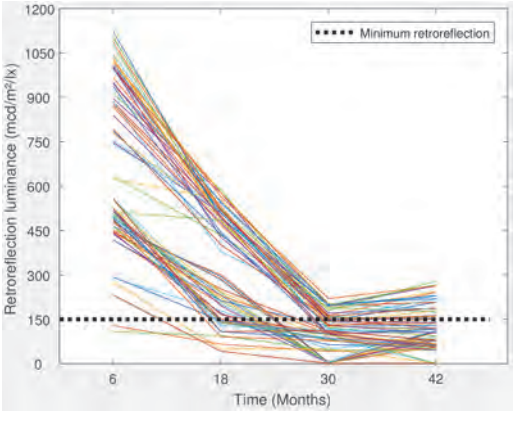


Figure 1. Maintenance cycle on the NR4 brocken centerlines between September 2008 and September 2011.

3.2 *Global Weibull analysis*

The Table 1 presents the Weibull analysis. The first column indicates the observed data associated to the censor indicator. For example, (6,1) corresponds to failures observed during the 6th month (September 2008) and the left-censored case. The second column indicates the number of observation: for example, there are 2 left-censored data and they correspond to 3% of the observed data. The last column indicates the failure time (in month) estimated by the EM Algorith. For example, according to the Weibull distribution $\mathcal{W}(31.65,2.20)$, left-censored data are estimated at 4.12 months (July 2008).

Table 1 confirms observations made on the monitoring. An important case of interval-censored data (74%) is presented. Particularly the main failure is emerged between 18 and 30 months (53%). Right-censored data are the second most important case (23%). Only two markings are left-censored.

The EM Algorithm converges after 12 iterations to the Weibull distribution $\mathcal{W}(31.65,2.20)$. According to this distribution, several estimated failures are proposed. Interval-censored failures are estimated 12.94, 24.01, 35.44 months (March 2009–2010, February 2011). This period corresponds to the average of each interval. Left and right censored failures are respectively estimated to 4.12 months (July 2008) and 50.41 months (May 2012). The maintenance campaign in March 2012 is finally warranted.

Finally, the EM Algorithm is able to produce one Weibull analysis adapted to the whole monitoring.

Table 1. Weibull analysis of the 2008–2012 maintenance cycle.

Observed failure	Numbers	Estimated failure $\mathcal{W}(31.65,2.20)$
(6,1)	2 (3%)	4.12
([6,18],2)	13 (18%)	12.94
([18,30],2)	39 (53%)	24.01
([30,42],2)	2 (3%)	35.44
(42,3)	17 (23%)	50.41

3.3 *Clustering Weibull analysis*

Based on an Agglomerative Hierarchical Clustering, the clustering process proposed in the previous ESREL conference could be restricted to the maintenance cycle (Redondin, Bouillaut, Daucher, & Faul 2017). Two clusters are proposed: Sommesous and other cities. Figure 2 distinguishes clusters on the monitoring. Sommesous (red) represents markings which admit a strong retroreflection luminance, a fast decay between 6 and 30 months and finally a stationary decay until the 42 Months (September 2011). Other cities (blue) present a fast decay between 6 and 18 months and a slow decay until the 42 Months.

The Table 2 presents the Weibull analysis by clusters. The structure is similar to Table 1 excepted the added first column to indicate both the cluster and the Weibull distribution estimated. The EM Algorithm proposed the Sommesous Weibull distribution $\mathcal{W}(41.35,3.5)$ and the Other Weibull distribution $\mathcal{W}(18.83,6.61)$ respectively after 19 and 38 iterations.

Sommesous markings failed mainly between 18 and 30 months and focused all right-censored data. Their failure are respectively estimated at 24.95 months (March 2010) and 49.84 months (April 2012). Other markings failed mainly between 6 and 18 months or 18 and 30 months and focused all left-censored data. Their failures are respectively estimated at 15.23 months (June 2009), 20.14 months (November 2009) and 5.21 months (August 2008).

The clustering process clearly separated the monitoring into two decay profiles. Sommesous markings had a strong retroreflectivity and failed either within 30th month or after the 42nd month while Others markings had a weaker retroreflectivity and failed before the 30th month. Finally, the EM Algorithm is also able to produce one specific Weibull analysis by cluster.

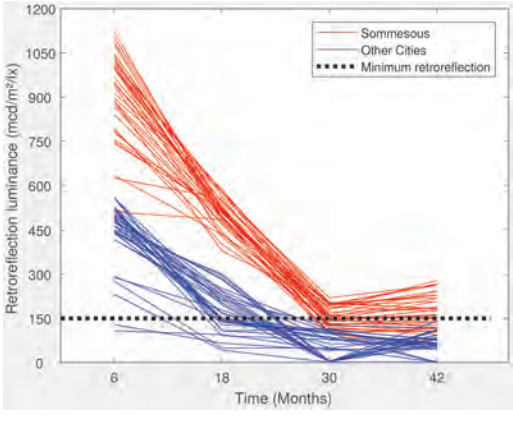


Figure 2. Maintenance cycle 2008–2012 by clusters.

Table 2. Weibull analysis of the maintenance cycle 2008–2012 by clusters.

Cluster	Observed failure	Numbers	Estimated failure
Sommesous	([18, 30], 2)	18 (25%)	24.95
	([30, 42], 2)	2 (3%)	35.11
$\mathcal{W}(41.35,3.5)$	(42,3)	17 (23%)	49.84
Other Cities	(6,1)	2 (3%)	5.21
	([6, 18], 2)	13 (18%)	15.23
$\mathcal{W}(18.83,6.61)$	([18, 30], 2)	21 (29%)	20.14

3.4 *Comparison*

Estimated failures between the global and the clustering models are equivalent to one month. Markings failed between 18 and 30 months in Other Cities are the main difference. The global model estimated the failure in March 2010 and the clustering model estimated the failure in November 2009. The difference is due both to the EM Algorithm computation and the current inspection strategy. First, the September 2010 inspection observed that 53% of markings failed since September 2009 and the associated interval-censored data is ([18, 30], 2) for all markings. Second, the global approach ignores different decay profiles. Therefore in this case, the EM-Algorithm produced one global estimation.

The Figure 3 compares reliable functions according to the global model and the two clusters. The global reliable function (green) underestimates Sommesous over time (–0.12 in average). Other Cities is slightly underestimated the 14th first months (–0.03 in average) and overestimated next months (+0.12 in average) in particular after the 23rd months. Finally, the global reliable function is an average estimation.

The global model is a compromise estimation according to different observations: two left-censored observations, important failure between 18 and 30 months, the quarter of observations is right-censored... The clustering model suggests first several clusters according to the decay profile and each one admit its own Weibull distribution. Therefore the Sommesous Weibull distribution is estimated according to two facts: without failures before the 18th month and all right-censored observations.

The clustering model admits two main disadvantages. First, the clustering process tends to isolate markings which admit one specific censure. For example, all right-censored observation could be gathered into a third cluster extracted from

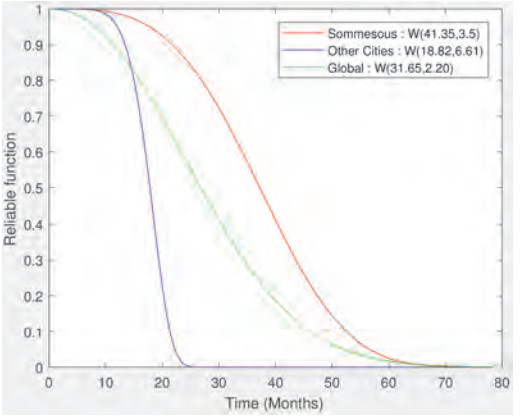


Figure 3. Associated reliable functions according to the maintenance 2008–2012.

Sommesous. In this situation, the MLE is based on the only observation (42, 3) and doesn't compute. The second problem is each cluster admits its own Weibull analysis. The model admits a total of 57 EM Algorithm iterations, whereas the global model admits only 12.

However, the global reliable function concludes to a poor compromise: Sommesous and Other markings reliable are respectively underestimated and overestimated.

Finally, the clustering model has done a more reliable Weibull analysis but the clustering process and the number of EM Algorithm iteration should be monitored.

4 CONCLUSIONS

The current literature presents a first Weibull analysis for road markings (Sathyanarayanan, Shankar, & Donnell 2008). Censored Weibull distributions are estimated by the current approach based on the MLE upgraded by a Newton-Raphson approach. Restricted to interval-censored data, Pradhan and Kundu (Pradhan & Kundu 2014) discoved several limits to this method and proposed a first alternative based on a EM Algorithm.

The introduced extended EM Algorithm is a credible alternative to the MLE of a censored Weibull distribution. This is specially true on the censored data management. Indeed, the EM Algorithm replaces censored data by a failure iteratively estimated at once the Weibull distribution. Furthermore, the EM Algorithm is upgraded by a fixed point approach. This method is simpler than the Newton-Raphson. Indeed, the fixed point depends only to uncensored data and estimated data and doesn't need any upstream verification.

The introduced algorithm isn't limited to our case study. Pradhan and Kundu themselves presented a breast cancer case study for example. The extension accepts a greater variety of situations: uncensored and right-censored data, no-uncensored data, no-interval-censored data. From the moment the decay monitoring over the time and the maximum decay level are both clear, the EM Algorithm is able to produce and rank all censored data.

The EM Algorithm is able to estimate several Weibull distribution in our road markings study. Two approach is proposed. The global model is interesting in the case where the observed data is small. However, the estimated Weibull distribution is an average compromise between different decay profiles. In the NR4 case, the global reliable function overestimates Other markings and underestimates Sommesous markings. The clustering analysis segments first the whole monitoring and estimates one Weibull distribution by cluster. This situation is interesting if the observed data is large and could distinguishes different decay profiles. In the NR4 case, reliable functions are more reliable.

This paper shows that the clutering approach is more reliable that the global model. However, this approach needs a monitoring of the clustering process. Indeed, the current process tends to isolate one specific censure by cluster. Consequently, the MLE could be based on only one observed data and it formalism cannot compute. An alternative based on a mixture model also produces by an EM Algorithm is currently investigated. This approach could estimate directly the optimum segmentation and associated it in one mixture Weibull distribution.

Finally, the Weibull analysis completed by an EM approach is adapted in a road markings study. Furthermore, for given sections of the road network, reliable functions are able to indicate directly each section admits a premature aging for example. These facts lead to the development of a probabilistic opportunistic maintenance model adapted to a whole road network or a segmented road network.

ACKNOWLEDGMENT

VEDECOM and IFSTTAR thank the DIR Est and the CEREMA for inspection data on the National Road 4.

REFERENCES

Abboud, N. & B. Bowman (2002). Cost and longevity-based scheduling of paint and thermoplastic striping. *Transportation Research Record 1794*.

AFNOR (2009). Nf en 1436+a1 – road marking materials—road marking performance for road users. *AFNOR Editions*.

Bain, L. & M. Englehardt (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, Inc.

Lu, J. (1995). Performance of traffic markings in cold regions. *University of Alaska Fairbanks—Transportation Research Center (Report No. INE/TRC 95.03)*.

McLachlan, G. & T. Krishnan (2008). *The EM Algorithm and Extensions*. John Wiley and Sons.

Pradhan, B. & D. Kundu (2014). Analysis of interval-censored data with Weibull lifetime distribution. *Sankhya B : The Indian Journal of Statistics 76*, 120–139.

Redondin, M., L. Bouillaut, D. Daucher, & N. Faul (2017). Temporal clustering for retroreflective markings. *European Safety and Reliability Conference 2017 (Portorz, Slovenia) (76)*, 7p.

Sarasua, W.A., D.B. Clarke, & W.J. Davis (2003). Evaluation of interstate pavement marking retroreflectivity. *South Carolina Department of Transportation (Report No. FHWA-SC-03-01)*.

Sathyanarayanan, S., V. Shankar, & E.T. Donnell (2008). Pavement marking retroreflectivity inspection data: A Weibull analysis. *Journal of the Transportation Research Board. 2055*, 63–70.

Sitzabee,W., J. Hummer, & W. Rasdorf (2009). Pavement marking degredation modeling and analysis. *Journal of infrastructure systems 15(3)*, 190–199.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Time-dependent unavailability model integrating on demand-caused and standby-related failures addressing positive and negative effects of testing and maintenance

P. Martorell, S. Martorell, I. Martón & S. Carlos
Department of Chemical and Nuclear Engineering, Universitat Politècnica de València, Valencia, Spain

A.I. Sánchez
Department of Statistics and Operational Research, Universitat Politècnica de València, Valencia, Spain

ABSTRACT: In recent years, many authors have proposed alternative approaches to modelling the effect of ageing and test and maintenance activities. Some authors have proposed approaches to modelling the unavailability of safety-related components associated with standby-related failures that explicitly addresses all aspects of the effect of ageing, maintenance effectiveness and test efficiency. Recently, other authors have proposed a new reliability model for the demand failure probability that explicitly addresses all aspects of the effect of demand-induced stress, maintenance effectiveness and test efficiency. In this context, the paper presents a whole time-dependent unavailability model for a safety component, regarding time-dependent unreliability contributions for each failure mode addressing ageing, degradation stress by demand and effectiveness of maintenance and testing as well as the downtime contributions related to maintenance and testing activities. An application case of a motor-operated valve of a pressurized water reactor nuclear power plant is included. A set of sensitivity cases are presented to show how this model can be used, for example, in the planning of maintenance and surveillance test activities with the aim of minimizing equipment unavailability.

1 INTRODUCTION

The safety of Nuclear Power Plants (NPPs) depends on the availability of safety-related components that are normally on standby and only operate in the case of a true demand. These components typically have two main types of failure modes that contribute to the probability of failure on demand: by demand-caused and standby-related failure.

Both are generally associated with constant values in a standard Probabilistic Risk Assessment (PRA) models. However, both failure modes are often affected by degradation such as demand-related stress and ageing, which cause the component to degrade with chronological time and ultimately to fail. Maintenance and test activities are performed to control degradation and the unreliability and unavailability of such components, although this has both positive and negative effects.

Initial studies reported in Kim et al. (1994) already provided a well-organized foundation to account for ageing as well as positive and adverse effects of testing components for both by demand-caused and standby-related failure modes. However, this model does not take into account the positive effect of maintenance activities as a function of their effectiveness in managing component degradation due to demand-induced stress and ageing.

In recent years, many authors have proposed alternative approaches to modelling the effect of ageing and test and maintenance activities (Kančev et al 2016, Shin et al. 2015, Volkanovski 2012). Martón et al. (2015) proposes an approach to modelling the unavailability of safety-related components associated with standby-related failures that explicitly addresses all aspects of the effect of ageing, maintenance effectiveness and test efficiency. Recently, Martorell et al. (2017) have proposed a new reliability model for the demand failure probability that explicitly addresses all aspects of the effect of demand-induced stress, maintenance effectiveness and test efficiency.

This paper presents a whole time-dependent unavailability model for a safety component, regarding time-dependent unreliability contributions developed by Martón et al. (2015) and Martorell et al. (2017) for each failure mode as well as the downtime contributions related to maintenance and testing activities.

An application case of a motor-operated valve of a nuclear power plant is included. A set

of sensitivity cases are presented to show how this model can be used, for example, in the planning of maintenance and surveillance test activities with the aim of minimizing equipment unavailability.

2 UNRELIABILITY MODELS UNDER IMPERFECT MAINTENANCE

In this paper the model of demand and failure probability standby failure rate, proposed by Martorell et al. (2017) and Martón et al. (2015), respectively, have been joined to develop a whole time-dependent reliability model taking into account the related preventive maintenance and testing activities performed.

On the one hand, demand failure probability of a component normally in standby depends on the number of demands on the component, which depends of performing the planned and unplanned surveillance and functional test, operational demands and test after corrective actions. So, surveillance testing not only introduce a positive effect, but also and adverse one, which is usually compensate by performing maintenance activities to eliminate or reduce the accumulated degradation.

On the other hand, standby failure rate depends on its age, which is a function of the chronological time elapsed since its installation and the effectiveness of the maintenance activities performed on it.

In both models, in order to introduce the effect of preventive maintenance, two imperfect maintenance models are considered: Proportional Age Reduction (PAR) and Proportional Age Setback (PAS).

In the PAR approach, each maintenance activity is assumed to reduce proportionally, in a factor ε, only the component degradation gained from the previous maintenance, while the rest remains unaffected, where ε represents the maintenance effectiveness that ranges in the interval [0,1]. Nevertheless, in the PAS approach, each maintenance activity is assumed to reduce proportionally, in a factor ε, the degradation that the component has immediately before it enters maintenance.

2.1 *Demand failure probability addressing maintenance effectiveness*

The time-dependent demand failure probability could formulated in terms of a time-dependent degradation function, $f(t)$, as follows:

$$\rho(t) = \rho_0 + \rho_0 * f(t) \tag{1}$$

where ρ_0 is the residual demand failure probability and $f(t)$, assuming that the degradation factor is the same for all types of demands and is equal to p_1, $f(t)$ can be formulated as follows:

$$f(t) = p_1 * n(t) \tag{2}$$

where $n(t)$ should include the number of demands performed up to time t. When only test-induced stress is considered, $n(t) = [t/T]$, where T represents the test interval and [x] the floor function that gives the largest integer less than or equal to x.

A time-dependent demand failure probability model that addresses the demand-induced stress and the effect of *m-1* maintenance activities can be formulated for the period *m* as follows:

$$\rho_m(f) = \rho_0 + \rho_0 . f_m(t) \tag{3}$$

being ρ_0 the residual demand failure probability and $f_m(t)$ the degradation function.

This equation can be particularized for $t = t_m = m \cdot M$, immediately after performing maintenance number *m*, and the PAR model, to obtain the formulation of the time-dependent demand failure probability immediately after maintenance *m* (P. Martorell *et al.*, 2017):

$$\rho_m(t = t_m) = \rho_0 + \rho_0 . p_1 \cdot \frac{M}{T} \cdot (1 - \varepsilon_D) . m \tag{4}$$

where p_1 is the degradation factor associated with demand failures, T is the test interval, M is the preventive maintenance interval, m is the number of maintenances performed and ε_D is the preventive maintenance effectiveness associated with demand failures.

Analogously, the time-dependent demand failure probability immediately after maintenance *m* for PAS model can be expressed as follow (P. Martorell *et al.*, 2017):

$$\rho_m(t = t_m) = \rho_0 + \rho_0 . p_1 \cdot \frac{M}{T} \cdot \frac{(1 - \varepsilon_D)}{\varepsilon_D} \cdot \left\{ 1 - (1 - \varepsilon_D)^m \right\} \tag{5}$$

2.2 *Standby failure rate addressing maintenance effectiveness*

In this paper, the linear ageing failure rate model has been considered to model the reliability model of standby—related failures. This model assumes that the failure rate has a linear behaviour with component age departing form an initial reliability, inherently by design, which can be expressed as (Martorell et al., 1999):

$$\lambda(t) = \lambda_0 + \alpha.w(t) \quad (6)$$

where t represents the chronological time, α is the linear ageing factor and $w(t)$ is the component's age after maintenance $m+1$. In Equation 6 component inherent failure rate by design is given by the term λ_0, which represents random failures, while the second term $\alpha \cdot w(t)$ represents the degradation of the equipment failure rate due to equipment ageing, which is counterbalanced by the effectiveness of the maintenance policy.

This equation can be particularized for $t=t_m = m \cdot M$, immediately after performing maintenance number m, and the PAR model, to obtain the formulation of the time-dependent standby failure rate immediately after maintenance m (Martón et al., 2015):

$$\lambda_m(t = t_m) = \lambda_0 + \mathrm{M} \cdot (1 - \varepsilon_s).m \quad (7)$$

where, ε_S is the preventive maintenance effectiveness associated with standby failures.

Analogously, the time-dependent standby failure rate immediately after maintenance m for PAS model can be expressed as follow (Martón et al., 2015):

$$\lambda_m(t = t_m) = \lambda_0 + \mathrm{M} \cdot \frac{(1 - \varepsilon_s)}{\varepsilon_s} \cdot \left\{1 - (1 - \varepsilon_s)^m\right\} \quad (8)$$

3 UNAVAILABILITY MODELLING

The average unreliability contribution to the unavailability of a component normally in standby over its renewal period can be formulated as follows (Martón et al. 2015, Martorell et al. 2017):

$$u_R = u_{R,S} + u_{R,D} \quad (9)$$

where the $u_{R,S}$ is the standby-related unreliability contribution and $u_{R,D}$ is the demand-caused unreliability contribution.

On one hand, adopting the PAS model to represent the behavior of the imperfect maintenance for the standby-related failures of the component according to the results in the previous section, $u_{R,S}$ is given by (Martón et al, 2015):

$$u_{R,S} \approx \frac{1}{2}\left(\lambda_0 + \frac{1}{2}\alpha M\left(\frac{2 - \varepsilon_s}{\varepsilon_s}\right)\right)T \quad (10)$$

On the other hand, adopting the PAS model to represent the behavior of the imperfect maintenance for demand caused failures of the component according to the results in the previous section, $u_{R,D}$ is given by (P. Martorell *et al.*, 2017):

$$u_{R,D} = \rho_0 + \frac{1}{2}\rho_0 . p_1 . \frac{M}{T} . \left(\frac{2 - \varepsilon_D}{\varepsilon_D}\right) \quad (11)$$

In accordance with Martón (2015), the averaged unavailability of a component is the sum of the average unreliability contributions and the unavailability contributions due to detected downtimes for performing testing and maintenance activities with the plant at power, which can be formulated as follows:

$$u = u_R + u_T + u_M + u_C + u_O \quad (12)$$

where u_T represents the unavailability contribution due to testing, u_M is the unavailability contribution due to performing preventive maintenance, u_C is the unavailability contribution due to performing corrective maintenance conditional to detecting a failure during a previous test, and u_O is the contribution due to replacement of the equipment, if any.

For sake of simplicity, the last two contributions, u_C and u_O, are not included as contributions of total average unavailability due to both are negligible as compared with the downtime effect of preventive maintenance and testing activities.

Thus, the downtime contributions considered in this paper can be evaluated using the following equations (Martón et al. 2015):

$$u_T = \frac{\tau}{T} \quad (13)$$

$$u_M = \frac{\sigma}{M} \quad (14)$$

where τ is the downtime for testing and σ is the downtime for preventive maintenance.

Thus, the averaged unavailability of the component is given by:

$$u = u_{R,S} + u_{R,D} + u_T + u_M \quad (15)$$

Substituting Equations 10, 11, 12 and 14 into Equation 15 yields the following formulation of the total average unavailability of a component:

$$u = \frac{1}{2}\left(\lambda_0 + \frac{1}{2}\alpha M\left(\frac{2 - \varepsilon_s}{\varepsilon_s}\right)\right)T + \rho_0 + \frac{1}{2}\rho_0 p_1 \frac{M}{T}\left(\frac{2 - \varepsilon_D}{\varepsilon_D}\right) + \frac{\tau}{T} + \frac{\sigma}{M} \quad (16)$$

4 CASE STUDY

In this section, an example of application of the model is presented that focuses on a Motor-Operated Valve (MOV) of a nuclear power plant.

Table 1 and Table 2 shows reliability parameters estimated obtained from the plant operational data, i.e. historical failure, maintenance and test data. The imperfect maintenance reliability model that best fit the plant data for by demand caused failures and standby-related failures is the PAS model in both cases. These results are extracted from (Martorell et al. 2017).

The estimates obtained are used to predict the performance of the MOV as a function of test and maintenance intervals. In particular, the MOV average unreliability contribution of each failure mode and the total MOV unavailability are computed and plotted as a function of maintenance and test intervals for a 10 years horizon.

Figure 1 shows the evolution of $u_{R,S}$ and $u_{R,D}$ as a function of the test interval, regarding different preventive maintenance intervals for a 10 years horizon renewal period. It can be seen that $u_{R,S}$ increases significantly for high T and M values.

Table 1. Parameters of the reliability model of standby related failures under PAS model.

Parameter	Value	[Units]
λ_0	5.860E-06	$\mathbf{h^{-1}}$
α	3.424E-10	$\mathbf{h^{-2}}$
ε_S	0.716	–

Table 2. Parameters of the reliability model of demand caused failures under PAS model.

Parameter	Value	[Units]
ρ_0	6.420E-03	–
p_1	5.415E-3	–
ε_D	0.886	–

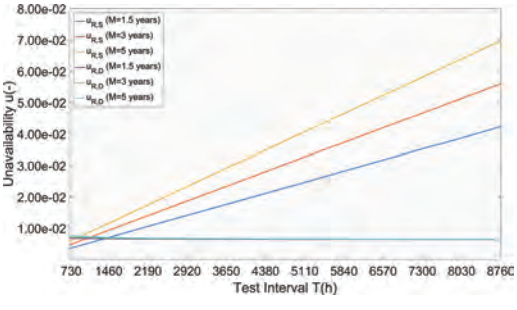


Figure 1. $u_{R,S}$ and $u_{R,D}$ as a function of the test interval for different maintenance periods.

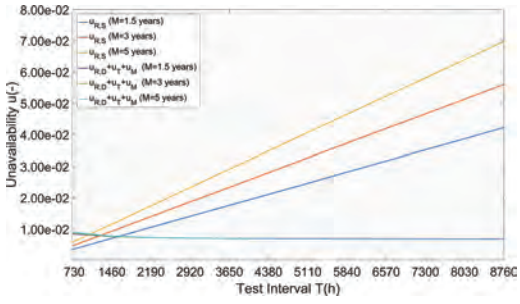


Figure 2. $u_{R,S}$ and $u_{R,D} + u_T + u_M$ as a function of the test interval for different maintenance periods.

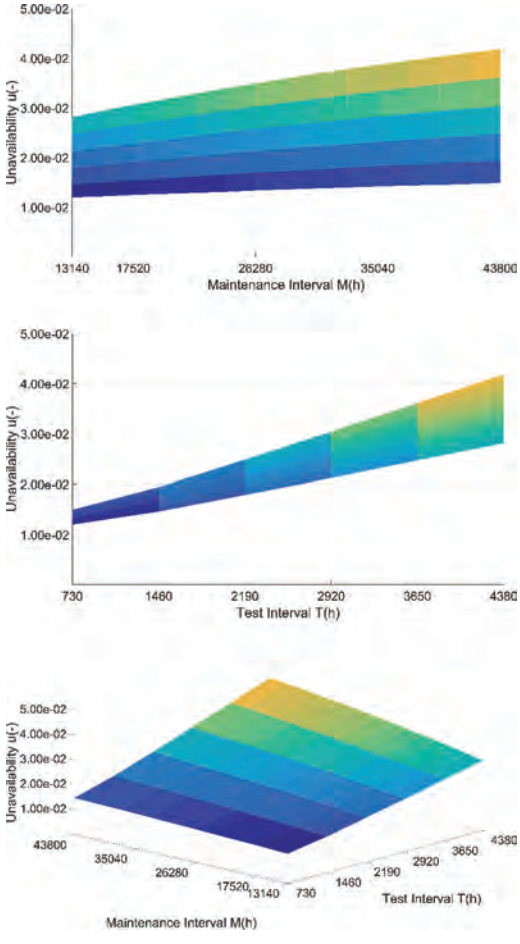


Figure 3. Unavailability for different maintenance and test intervals under PAS model.

Nevertheless, the effect of maintenance is positive for both unreliability contributions. Moreover, an increase on test frequency between maintenances, i.e. low T values, have a very negative effect on $u_{R,D}$ for very low T intervals.

Figure 2 shows the evolution of $u_{R,S}$ versus $u_{R,D} + u_T + u_M$ as a function of the test interval considering different preventive maintenance intervals for a 10 years horizon renewal period. The term $u_{R,S}$ allows quantifying the benefit of developing test and maintenance activities on the total component unavailability while the sum of contributions $u_{R,D} + u_T + u_M$ represents their negative effect.

The last study involves the analysis of the total average unavailability of the component as a function of the couple $\{M, T\}$ for a 10 years horizon, which is shown in Figure 3. The highest values of u are reached adopting the highest maintenance and test intervals. The main contributor to the total unavailability, u, see Equation 15, is the standby-related unreliability contribution given by Equation 10 as it can be seen in Figure 2. This explains the direct and proportional dependence between u and T and M. Nevertheless, the sum of the demand unreliability contribution and downtime effects considered, i.e. downtime effect of preventive maintenance and testing activities, become more relevant for very low T values. This fact is appreciated in Figure 2 too.

5 CONCLUDING REMARKS

This paper presents a whole time-dependent unavailability model for a safety component, regarding time-dependent unreliability contributions for by demand caused failures and standby-related failures as well as the downtime contributions related to maintenance and testing activities. Regarding imperfect maintenance effects, two models are considered: Proportional Age Reduction (PAR) and Proportional Age Setback (PAS).

A practical and realistic case study is included facing the parameters estimation of a typical motor-operated valve in a nuclear power plant. Equipment RAM is quantified based on the best model fitted to make it clear the impact of such an estimation in a testing and maintenance-planning context.

Thus, the results of a whole time-dependent unavailability model may help to plan in a more efficient way the test and maintenance program, which should provide appropriate balance among the different contributions to the unavailability of the MOV, with the aim of minimizing its unavailability assuring a low level of unreliability.

ACKNOWLEDGEMENTS

The authors are grateful to the Spanish Ministry of Science and Innovation for the financial support received (Research Project ENE2016-80401-R) and the doctoral scholarship awarded (BES-2014-067602).

REFERENCES

Kančev, D., Gjorgiev, B., Volkanovski, A. and Čepin, M. (2016) 'Time-dependent unavailability of equipment in an ageing NPP: Sensitivity study of a developed model', *Reliability Engineering & System Safety*, 148, pp. 11–20. doi: 10.1016/j.ress.2015.11.014.

Kim, I. S., Martorell, S. A., Vesely, W. E. and Samanta, P. K. (1994) 'Risk analysis of surveillance requirements including their adverse effects', *Reliability Engineering & System Safety*, 45(3), pp. 225–234. doi: 10.1016/0951-8320(94)90139-2.

Martón, I., Sánchez, A. I. and Martorell, S. (2015) 'Ageing PSA incorporating effectiveness of maintenance and testing', *Reliability Engineering & System Safety*, 139, pp. 131–140. doi: 10.1016/j.ress.2015.03.022.

Martorell, P., Martón, I., Sánchez, A. I. and Martorell, S. (2017) 'Unavailability model for demand-caused failures of safety components addressing degradation by demand-induced stress, maintenance effectiveness and test efficiency', *Reliability Engineering & System Safety*. doi: 10.1016/j.ress.2017.05.044.

Martorell, S., Martorell, P., Sanchez, A. I., Mullor, R. and Martón, I. (2017) 'Parameter estimation of a reliability model of demand-caused and stand-by related failures of safety components exposed to degradation by demand stress and ageing that undergo imperfect maintenance', *Mathematical Problems in Engineering*, In press.

Martorell, S., Sanchez, A. and Serradell, V. (1999) 'Age-dependent reliability model considering effects of maintenance and working conditions', *Reliability Engineering & System Safety*, 64(1), pp. 19–31. doi: 10.1016/S0951-8320(98)00050-7.

Shin, S. M., Jeon, I. S. and Kang, H. G. (2015) 'Surveillance test and monitoring strategy for the availability improvement of standby equipment using age-dependent model', *Reliability Engineering & System Safety*, 135, pp. 100–106. doi: 10.1016/j.ress.2014.11.001.

Volkanovski, A. (2012) 'Method for assessment of ageing based on PSA results', *Nuclear Engineering and Design*, 246, pp. 141–146. doi: 10.1016/j.nucengdes.2011.06.037.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Modelling demand-caused failures. Estimation procedure

R. Mullor
Universitat d'Alacant, Alicante, Spain

A.I. Sánchez, P. Martorell & S. Martorell
Universitat Politècnica de València, Valencia, Spain

ABSTRACT: In the literature related to reliability and maintainability analysis on reparable equipments, a great number of studies have been published about modelling and analyzing their failure times. However, there are no so much studies which analyze the behavior of these equipments when failures happen at demand. In this case, models are completely different since the failure distribution is discrete. In this work, the model is performed from the real behavior of equipments without any a priori probability distribution. From this model, discrete probability function and cumulative distribution function, needed for the subsequent parameters estimation, are obtained. By combination of these functions and imperfect maintenance models, likelihood function for reliability analysis is constructed to jointly estimate its parameters, that is, the failure at demand probability and the maintenance effectiveness, and their variability. Then, an application case performs the estimation procedure applied to a database of a safety equipment of a Nuclear Power Plant, where the obtained estimations together with their variability can be used to plan optimal test and maintenance intervals.

1 INTRODUCTION

From the end of the last century, studies related to improvement of safety in industrial plants in general, and in nuclear power plants in particular, have acquired special relevance in order to optimize both, the safety of these plants and the resources spent for this purpose. In this context, analyzing reliability through a failure model that adequately represents the behavior of safety equipments is an essential task to which many authors have been devoted (Martorell et al. 1999, Busacca et al. 2001, Nakamura et al. 2004). First problem to solve in these kind of studies is modeling the behavior of these equipments and, consequently, estimating the parameters of such models for a later exploitation of results, fundamentally in terms of optimizing some objectives (Shin et al. 1996, Mullor et al. 2006). Although initially models used were excessively simple, exponential distribution with null or perfect maintenance, the classic Bad As Old (BAO) and Good As New (GAN) models, in order to improve the fit of the model to reality, more complex models, both in terms of failure distribution and maintenance models, have been considered. Weibull, Gamma, linear… failure distributions are currently used since they allow a better fit to reality, simultaneously, imperfect maintenance models, as Proportional Age-Setback (PAS) or Proportional Age Reduction (PAR), which better reproduce the influence of maintenance activities in the components aging process are combined with the previous failure distributions obtaining like this, models more adequate to the objectives.

Even though a high degree of utility in the exploitation of the information obtained from the combination of previous models has been achieved in several applications (Lapa et al. 2000, Mullor et al. 2007), the standby condition in most of safety equipments to which this methodology is applied requires periodic verification of their availability to ensure a high probability of their correctly operation when they are called for performing their task. Mentioned verification of availability generates another problem, each such verification of the state of safety components degrades the said state due to the stress generated by the test, so, in addition to the previous age-dependent failure model for standby equipments, the probability of failure at demand generated by checking their status must be also analyzed. The demand-induced stress has been modeled by an stochastic degradation jumps (Yang et al. 2017) considering that random shocks occur according to a Non Homogeneous Poisson Process, or addressing the effects of test strategies on the probability of failure at demand for safety instrumented systems (Torres-Echeverría et al. 2009).

However, most of these models are not close enough to the true behavior of safety equipments in terms of tests degradation. So this paper focuses

on modeling the probability of failures related to the influence of tests and maintenance activities without assuming any a priori distribution, that is, constructing the probability and the cumulative functions of a discrete probability distribution directly by observed data, that combined with the imperfect maintenance models PAS and PAR provide the needed elements to construct the likelihood function which allows us to jointly estimate the probability of failure due to tests and the effectiveness of each maintenance activity, through the well known Maximum Likelihood Estimation (MLE) method based on the Nelder-Mead Simplex (Nelder & Mead 1965).

Finally, a practical and realistic case study of a typical motor-operated valve in a nuclear power plant is presented. Estimations of probability failure at demand and maintenance effectiveness are obtained through the new discrete model presented adapting its expressions to the MLE functions.

The rest of this paper is organized as follows: Section 2 presents the discrete failure at demand model. Section 3 describes the parameter estimation method in these kind of reliability models. The above methodology is applied in Section 4 to the observed data from a set of High-Pressure Injection System (HPIS) motor valves of a NPP. Finally, Sections 5 and 6 present the conclusions and references.

2 DEMAND-CAUSED FAILURE MODEL

The construction of this model is, in essence, simple, but somewhat more complex in its formalization. We start from a null initial failure probability, which means that the probability of failure at demand is zero until the first test. When this first test is performed, the probability of demand failure becomes p, and remains constant until the instant of the next test in which increases, again, by p, becoming from that moment $2p$, and so on until the test immediately before the first maintenance, in which has been reached a failure probability of

$$p_1^- = kp \tag{1}$$

where k is the number of test performed in the first maintenance interval. Next action will be the first maintenance in which previous probability of failure will be reduced, depending on its effectiveness, ε, up to

$$p_1^+ = p_1^-(1-\varepsilon) = kp(1-\varepsilon) \tag{2}$$

and the same process is repeated from this point. Starting, in this case, from p_1^+, each test increases the failure probability in a constant p, until reaching, in the test immediately before the second maintenance, a failure probability of

$$p_2^- = p_1^+ + kp = kp(1-\varepsilon) + kp \tag{3}$$

At this instant, the reduction of failure probability after the second, and following, maintenance will depend on the imperfect maintenance model applied.

In the case of implementing the PAR model, where the reduction affects only the failure probability increased in the last maintenance interval, we obtain

$$p_2^+ = kp(1-\varepsilon) + kp(1-\varepsilon) = 2kp(1-\varepsilon) \tag{4}$$

In the case of selecting PAS as maintenance model, the reduction affects the total failure probability reached before performing the maintenance, so this probability, after the said maintenance, will became

$$p_2^+ = \left(kp(1-\varepsilon) + kp\right)(1-\varepsilon) \tag{5}$$

Apart from simplifications or different representations for subsequent treatment and applications, the previous one is the evolution of the failure probability at any instant of the useful life period of the analyzed component, that is represented in Figure 1.

Then, the construction of probability functions, that determine the likelihood functions which allow as estimating the objective parameters under models that really represent the behavior of the equipment in terms of the occurrence of failures at demand, are showed.

2.1 *Proportional age reduction model*

If a PAR model is considered, the maintenance only reduces the increase of probability from the previous maintenance, we would obtain that, for

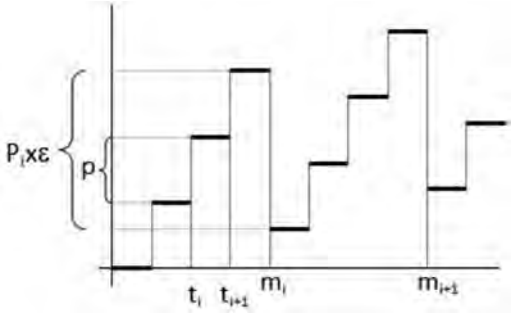


Figure 1. Evolution of the failure probability vs time.

any maintenance m, the failure probability before and later applying it will be

$$\begin{aligned}\rho_m^- &= (m-1)kp(1-\varepsilon)+kp \\ \rho_m^+ &= mkp(1-\varepsilon)\end{aligned} \tag{6}$$

Since, for the construction of the likelihood function we will only need the probability function, $p(t)$, at failure times and the cumulative distribution function, $F(t)$, at failure and maintenance times, from the previous expressions, and denoting by $k_{f,m}$ the instant, the number of test in maintenance m, in which the failure is observed, the probability function in said failure will be

$$p\left(t_{f,m}\right) = (m-1)kp(1-\varepsilon)+k_{f,m}p \tag{7}$$

and adding the probability functions in the test and maintenances until the maintenance m we obtain the cumulative distribution function in each maintenance

$$F\left(\tau_m\right) = m\frac{k(k+1)}{2}p+\left((k+1)\frac{m(m+1)}{2}-km\right)kp(1-\varepsilon) \tag{8}$$

finally, the cumulative distribution function in each failure is obtained adding to the cumulative distribution function in the previous maintenance, the probability function in each test until the failure, applied in the maintenance interval where the said failure is observed

$$\begin{aligned}F\left(t_{f,m}\right) &= F\left(\tau_{m-1}\right)+k_{f,m}(m-1)kp(1-\varepsilon) \\ &\quad +\frac{k_{f,m}\left(k_{f,m}+1\right)}{2}p\end{aligned} \tag{9}$$

2.2 *Proportional age-setback model*

If the imperfect maintenance model is the PAS one, which reduces the failure probability that each component has accumulated until the time before that maintenance activity is performed, we find, obviously, the same situation in the first maintenance, but this will not happen in the following ones. The failure probability before the second maintenance will continue to be the same than in the PAR model, but after then it will be

$$\rho_2^+ = \left(kp(1-\varepsilon)+kp\right)(1-\varepsilon) = kp\left((1-\varepsilon)+(1-\varepsilon)^2\right) \tag{10}$$

and, by generalizing the process again, we obtain the failure probabilities before and after the *m-th* maintenance, which will be given by

$$\begin{aligned}\rho_m^- &= kp\left(1+(1-\varepsilon)+(1-\varepsilon)^2+\ldots+(1-\varepsilon)^{m-1}\right) \\ &= kp\sum_{i=0}^{m-1}(1-\varepsilon)^i \\ \rho_m^+ &= kp\left((1-\varepsilon)+(1-\epsilon)^2+\ldots+(1-\varepsilon)^m\right) \\ &= kp\sum_{i=1}^{m}(1-\varepsilon)^i\end{aligned} \tag{11}$$

Although, again, what we want to obtain are probabilities of failure and their cumulated sums or distribution function, we can get, proceeding in a similar way to the previous model, the equivalent expressions

$$p\left(t_{f,m}\right) = kp\sum_{i=1}^{m-1}(1-\varepsilon)^i + k_{f,m}p \tag{12}$$

$$F\left(\tau_m\right) = kp\left(\sum_{i=1}^{m}((k+1)(i-1)+1)(1-\varepsilon)^{m-i+1} + \frac{mk(k+1)}{2}\right) \tag{13}$$

$$\begin{aligned}F\left(t_{f,m}\right) &= F\left(\tau_m\right)+k_{f,m}kp\sum_{i=1}^{m-1}(1-\varepsilon)^i \\ &\quad +\frac{k_{f,m}(k_{f,m}+1)}{2}p\end{aligned} \tag{14}$$

It should be noted that, unlike what it happens with continuous models of failure distributions, these expressions, and their subsequent contribution to the likelihood function, do not depend on the instant in which the failure is observed but on the number of tests and maintenance, and type of maintenance, which have been carried out until the observed failure. Finally, it should be pointed that, although these expressions seem very complex, in fact they are, when substituting k, $k_{f,m}$ and m by their values in a particular problem as in our application case, they become much clearer and very useful for their implementation.

3 ESTIMATION PROCEDURE

The construction of the likelihood function for proposed reliability models under imperfect maintenance is presented below. For a given model and a set of observed data, the likelihood function, L, is the product of probabilities that the observed data will occur as a function of the parameters that must to be estimated

$$L\left(\xi \mid observed\ data\right) = \prod_{events} p_i \tag{15}$$

This expression can be applied to the age dependent failure models formulated previously by considering the probability function as probabilities of failures and the reliability function to

model probabilities after maintenances, both normalized on the time interval in which the events are observed, obtaining in this way the general expression for the likelihood function in reliability models with imperfect maintenance that will be given by

$$L(\xi \mid observed\ data) = \prod_{failures} h(t) \cdot \prod_{maintenances} R(t) \tag{16}$$

which applied to discrete models becomes

$$L(\xi \mid observed\ data) = \prod_{failures} \frac{p(t)}{1-F(t)} \cdot \prod_{maintenances} (1-R(t)) \tag{17}$$

From these expressions, the maximum likelihood estimation (MLE) method provide estimates of the parameters of reliability and maintenance models. The maximum likelihood estimations of these parameters are those values that maximize the likelihood function, that is, that maximize the probability that the observed events occur. Although sometimes direct methods can be applied to obtain the solution, this is not usually the situation in reliability models. In our applications, to maximize the likelihood function of each model, we use the Nelder-Mead Simplex method.

Let r_m be the number of failures observed in the maintenance interval m, which occur at instants $t_{m_1}, t_{m_2}, \cdots$ and let τ_m be the instant when maintenance m is performed, with $m \in \{1, 2, \ldots, M\}$. The likelihood function under imperfect preventive maintenance will be given by

$$L(\xi) = \prod_{m=1}^{M} \left[\prod_{j=1}^{r_m} h_m\left(t_{m,j}\right) \cdot R_m\left(\tau_m\right) \cdot R_M\left(\tau^*\right) \right] \tag{18}$$

where ξ is the vector of unknown parameters, M is the number of maintenances performed in the observation period τ^*, with $h_m(t)$ and $R(t)$ being the failure rate and the reliability function in the maintenance interval m, and $R(\tau^*)$ the reliability function in the censor time τ^*. Since the logarithm is an increasing function, the likelihood function and its logarithm achieve their maximum at the same value of the objective parameters vector. So, in order to simplifying the computational process, we will use the log likelihood function given by

$$log\ L(\xi) = \sum_{m=1}^{M} \sum_{j=1}^{r_m} \log(h_m(t_{m,j})) - \sum_{m=1}^{M} H\left(\tau_m\right) - H_{m+1}\left(\tau^*\right) \tag{19}$$

This expression will be useful when we have a probability distribution of failures that provides in closed form the failure rate and the cumulated failure rate, $H(t)$, as in the case of the continuous distributions. However, in the above presented discrete model we do not dispose of failure rate and cumulated failure rate functions in closed form, so we construct the likelihood function, and its logarithm, from the probability function and the cumulated distribution function, or one minus it, the reliability function, $R(t)$, in each failure and maintenance times

$$log\ L(\xi) = \sum_{m=1}^{M} \sum_{j=1}^{r_m} log\left(\frac{p_m\left(t_{m,j}\right)}{R\left(t_{m,j}\right)} \right) + \sum_{m=1}^{M} log\left(R_m\left(\tau_m\right)\right) + log\left(R_{M+1}\left(\tau^*\right)\right) \tag{20}$$

Again, by replacing the probability and cumulate functions obtained in the previous section for PAR and PAS maintenance models, the corresponding log likelihood functions are obtained.

Maximizing these functions by applying Nelder-Mead simplex method provide the maximum likelihood estimations of the parameters of each age dependent failure model. The optimization process itself provides, in addition to the vector of punctual estimations, information about its variability through the Fisher information matrix, which is defined as the opposite of the matrix of partial second derivates. So if the parameters vector is of dimension n, the information matrix will be given by

$$I(\xi) = \begin{bmatrix} \frac{\partial^2 \log L(\xi)}{\partial \xi_1^2} & \cdots & \frac{\partial^2 \log L(\xi)}{\partial \xi_1 \partial \xi_i} & \cdots & \frac{\partial^2 \log L(\xi)}{\partial \xi_1 \partial \xi_n} \\ \vdots & \ddots & \vdots & & \vdots \\ \frac{\partial^2 \log L(\xi)}{\partial \xi_i \partial \xi_1} & \cdots & \frac{\partial^2 \log L(\xi)}{\partial \xi_i^2} & \cdots & \frac{\partial^2 \log L(\xi)}{\partial \xi_i \partial \xi_n} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial^2 \log L(\xi)}{\partial \xi_n \partial \xi_1} & \cdots & \frac{\partial^2 \log L(\xi)}{\partial \xi_n \partial \xi_i} & \cdots & \frac{\partial^2 \log L(\xi)}{\partial \xi_n^2} \end{bmatrix} \tag{21}$$

then, we can obtain, for each model, the variance-covariance matrix as the inverse of the information matrix divided by the sample size.

To conclude this section, we must comment that an inconvenient of the application of the Nelder-Mead simplex method to the optimization of our log likelihood functions to estimate the objective parameters is that this is an optimization method without restrictions. Obviously, the parameters in our models have a limited space of definition and these restrictions must be introduced in some way in the optimization process. Fortunately, the restrictions of the parameters can be included through transformations of themselves. These

restrictions are $0 \le p, \varepsilon \le 1$, so the parameters, in the optimization process are defined as $p = \exp(x_1)/(1 + \exp(x_1))$ and $\varepsilon = \exp(x_2)/(1 + \exp(x_2))$, which ensures that solutions will be obtained within the domain of the decision variables. In addition, since the method only provides local optimum, its search must be repeated several times from different initial points, selected in a random way, to ensure obtaining a global optimum.

4 APPLICATION CASE

The application case presented below concerns the analysis of data collected from a High-Pressure Injection System (HPIS) consisting of a set of motor operated valves, normally in stand-by, reason why they must be checked periodically which causes some deterioration that increases the probability of failure at demand. Given the set of available data, failures, tests and maintenance activities, the objective of this application is the joint estimation of effectiveness of maintenance, ε, and probability of failure at demand, p, under the assumption of imperfect maintenance models PAR or PAS.

The available database presents the monitoring of the described equipment during 5100 days, in which 195 tests and 9 maintenances have been applied, and 3 failures have been observed. Since more information is not available, and this is usually the real situation, we assume that both maintenance and testing are time uniformly distributed so that all intervals between tests and maintenances have the same length, what means that finally we have tests every 25 days and maintenances every 525 days, which provides 9 maintenance activities in 14 years and 20 tests in each maintenance interval.

Once constructed the logarithm of the likelihood function presented in Section 3, the sum of the logarithm of probability function divide by reliability function in failure times and the logarithm of this reliability function in maintenance and censor times, Maximum Likelihood Estimation is performed through the Nelder-Mead Simplex method obtaining the results, about parameter estimations and their variability, showed in Table 1.

The variance-covariance matrix for the PAR and PAS models is given respectively by:

Table 1. Parameter estimation results.

Parameter	PAR model	PAS model
$\hat{p}$	4.63E-4	4.19E-4
$\hat{\varepsilon}$	0.9824	0.9056
L	0.024	0.005

$$PAR \qquad \widehat{\Sigma^2} = \begin{pmatrix} 8.24E-10 & 1.04E-7 \\ 1.04E-7 & 9.48E-5 \end{pmatrix}$$

$$PAS \qquad \widehat{\Sigma^2} = \begin{pmatrix} 1.23E-9 & 1.31E-6 \\ 1.31E-6 & 2.81E-3 \end{pmatrix}$$

The results of this application, together with those obtained in applications of these models to other databases, provide interesting conclusions. Firstly, and focusing in the presented case, we obtain the estimation of both, failure probability at demand and maintenance effectiveness. We can realize that results do not differ too much between both models, which is due to the fact that, since the maintenance effectiveness is in both cases close to one, the two models are very similar, providing very close estimations. However, we can conclude that the model which better represent the behavior of this equipment is the PAR one, since the value of its likelihood function in the obtained optimum is greater than in the PAS one, which means, under the same number of parameters, as it happens, that the probability of occurrence of the observed data is greater in the PAR model than in the PAS one. Additionally, when obtaining together with punctual estimations, its variance-covariance matrix that provides, in addition to their covariance, the standard deviation of the estimation of each parameter as the square root of its main diagonal, we can construct confidence intervals for each parameter, which provides us with a very usefulness information about their variability for further applications.

$$PAR \qquad CI_{95\%}(p) = (4.07E-4, 5.19E-4)$$
$$CI_{95\%}(\varepsilon) = (0.9633, 1)$$

$$PAS \qquad CI_{95\%}(p) = (3.50E-4, 4.88E-4)$$
$$CI_{95\%}(\varepsilon) = (0.8018, 1)$$

We also note that confidence intervals are narrower in the PAR model, that is, their estimates show less variability, which is consistent with the previous selection as the most appropriate model. Finally, it must be pointed that, in general, the maintenance effectiveness in PAR model is greater than in the PAS one, as it happens in our application. Since PAS maintenance provides better results, it reaches the same state as the PAR one with less effort.

5 CONCLUSIONS

This paper presents the construction of a model which really represent the behavior of an equipment which requires of periodic test, which deteriorate at each realization. This model allows us to

jointly estimate, via Maximum Likelihood Estimation, both the failure probability at demand and the maintenance effectiveness, when imperfect maintenance is, as really occurs, incorporated to the age dependent model.

Results obtained in the application case, which are in accordance with those usually verified in its context, validate the usefulness of this model and expand the field of the optimization of efforts aimed at improving the reliability of this kind of equipment.

ACKNOWLEDGEMENTS

The authors are grateful to the Spanish Ministry of Science and Innovation for the financial support received (Research Project ENE2016-80401-R) and the doctoral scholarship awarded (BES-2014-067602).

REFERENCES

Busacca P, Marseguerra M, Zio E. Multiobjective optimization by genetic algorithms: application to safety systems. *Reliability Engineering and System Safety,* 2001; 72: 59–74.

Lapa C, Pereira C, Mol A. Maximization of a nuclear system availability through maintenance scheduling optimization using genetic algorithm. *Nuclear Engineering and Design,* 2000; 196: 219–231.

Martorell S, Sanchez A, Serradell V. Age-dependent reliability model considering effects of maintenance and working conditions. *Reliability Engineering and System Safety,* 1999; 64(1): 19–31.

Mullor R, Sanchez A, Martinez N, Martorell S. Motor-operated valve maintenance optimization considering multiple failure modes and imperfect maintenance models. *Proceedings of ESREL* 2007.

Mullor R, Sanchez A, Martinez N. Parameters estimation under preventive imperfect maintenance. *Proceedings of ESREL* 2006.

Nakamura M, Katafuchi T, Hatazaki H. Decision for maintenance-intervals of equipment in thermal power stations, based on few data. *IEEE Transactions on Reliability,* 2004; 50(4): 360–364.

Nelder JA, Mead R. A simplex method for function optimization. *Computer Journal,* 1965; 308–313.

Shin I, Lim TJ, Lie CH. Estimating parameters of intensity function and maintenance effect for repairable unit. *Reliability Engineering and System Safety,* 1996; 54: 1–10.

Torres-Echevarria AC, Martorell S, Thompson HA. Modeling and optimization of proof testing policies for safety instrumented systems. *Reliability Engineering and System Safety,* 2009; 94: 838–854.

Yang L, Ma X, Zhao, Y. A condition-based maintenance model for a three-state system subject to degradation and environmental shocks. *Comput Ind Eng,* 105 (2017), pp. 210–222.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Data-driven and risk-based decision support for maintenance planning on electrical power grid systems

N.J. Edwin
Safetec Nordic AS, Trondheim, Norway

H. Mjølnerød
Wiseline AS, Trondheim, Norway

B.A. Gran
NTNU/Wiseline AS, Trondheim, Norway

ABSTRACT: Planning maintenance over a large, distributed and ageing power infrastructure is a challenge. Interruptions in power supply cost large amounts of money in terms of undelivered energy and therefore maintaining the infrastructure to deliver at all times is imperative. This paper presents a novel method developed to model the risk of power line infrastructure. The user is presented with a visual risk picture and the associated expected costs based on the user's maintenance strategy of choice. Visual risk presentation overlaid on map data allows the decision-maker to understand the risk picture and make appropriate investment decisions based on the same. The developed model is currently being used in-house at various grid operators in Norway. Users can test out the effect different maintenance strategies and its effect on the risk picture.

1 INTRODUCTION

1.1 *Background*

Every year there are thousands of planned and unplanned interruptions on the electrical power grid in Norway. These disconnections result in undelivered energy that cost the society millions each year. The year 2016 saw a total of 25,777 incidents totaling a loss of 8,239 MWh undelivered energy (Statnett 2017). Assuming an average interruption of 1.3 hours (Kjølle 2011), the cost of undelivered energy is approximately 30.7 NOK per kWh. This equals 253.2 million NOK, or over 32 million USD.

Norway as a country is characterized by large energy differences from region to region. Some areas are under producers and others, over-producers (Fornybar.no). In general, transporting energy over large distances is imperative in ensuring a robust power distribution network in the country.

Planning maintenance of such a large power distribution infrastructure is a challenge. A large portion of the infrastructure is aging, pushing for significant re-investment costs from year to year. In such a situation, the challenge is identifying where efforts should be focused and where money should be invested.

Furthermore, expanding infrastructure within renewables, creates challenges with respect to resource availability and utilization. This places more pressure on power grid companies to improve strategies while simultaneously reducing costs.

1.2 *Objective*

Paragraph §2–13 in the regulations for electrical power distribution (FEF 2006) states that the power distribution grid owner shall have an overview of their infrastructure such that they can decide where to focus maintenance effects and which areas require more focus/investment that others. This is with regard to reputation, Health Safety and Environment (HSE), technical condition and of course reliability of power supply. RENblad ((Norwegian: Rasjonell Elekstrisk Virksomhet)) claims that todays' choice of maintenance strategy often results in components being replaced much earlier than they must. Therefore, risk-based maintenance is a pre-requisite to enable better maintenance decisions. This is the objective of the approach and model developed.

This paper presents a simulation model developed, and results from a case study. Furthermore, preliminary feedback from pilot users are also discussed. The developed risk model quantifies the technical, HSE and economic-risk condition of a given power line and puts this in a financial context.

The approach simulates line degradation over multiple years and models the effect of different maintenance strategies. The result is a risk picture and cost associated with a chosen maintenance strategy.

2 STATE-OF-THE-ART

2.1 *Maintenance practice in the industry*

Moving from a reactive to a proactive maintenance approach is a demanding change for a power company. Grid owners systematically document results from physical inspections on their power lines from year to year. Certain operators perform limited visual inspections from the ground level yearly. More detailed inspections examining each utility pole for internal damage, rotting etc., is performed at 10-year inspection intervals. The data collected from these inspections is used to a very limited extent in planning maintenance. Deciding on a maintenance plan purely based on "age of infrastructure" and "total number of findings from inspections" is not optimal. Such a strategy is not risk-based and can result in significant investments made on infrastructure that has a very limited risk connected to power supply reliability.

Maintenance and reinvestment is a necessity to maintain the functionality of the power distribution grid. As components age and are exposed to external factors such as wind, lightning, vegetation, etc., their functionality may weaken or fail. Furthermore, parts of the power grid are so old that lines are associated with the need for reinvestments. Driving maintenance versus reinvestment decisions is a fine balance between economy and risk. "*When is it economically beneficial to replace/reinvest in infrastructure versus performing periodic maintenance on the same?*". Recent trends (Nordgård and Samdal 2009) point towards developing strategies by looking at the balancing of cost effectiveness with other risks, such as economy, HSE, reputation, supply reliability etc. Thus, bringing the advent of risk-based maintenance.

A well-known technique for the same is Reliability Centered Maintenance (RCM) (IEC 60300-3-11). RCM is primarily a qualitative approach which involves a systematic consideration of system functions, where they may fail and a consideration of safety vs. economy to decide on an appropriate maintenance strategy, e.g. preventive maintenance vs. corrective run-to-failure maintenance (Moubray 1997). The outcome of applying RCM on a set of systems including a number of components, on which reliability measurements can be done, would be a maintenance program pointing to a set of components or sub-systems to be maintained at different timeslots.

In Norway, the RENblad is a set of guidelines universally followed by most electric power distribution companies. These guidelines cover standardized material and methods for the power distribution business. One such guidebook is the "Distribution Network – Maintenance Strategy" (Norwegian: Distribusjonsnett – Vedlikeholdsstrategi"). This guidebook focuses on "condition-based maintenance management".

2.2 *Available decision-support methods and tools*

The available decision support methods and tools, vary with the size and ambitions of the power companies. The project "Smarter Asset Management using Big dAta" (SAMBA) is a 3-year industry innovation project headed by Statnett, partially founded by the Norwegian Research Council and involving several partners such as SINTEF Energy Research, GE Grid Solutions, ABB and IBM. Their ambition is to add value by activating more online, automatically collected condition and system operation data. These data will be combined with maintenance data collected on site, and then be used for estimating asset's condition, probability of failure and remaining lifetime. Some research questions are:

- What is the condition of the components now?
- How will the condition of a component develop over time?
- Do I do the right maintenance at the right time?
- When should a component be replaced?
- What is the consequence of a component break?

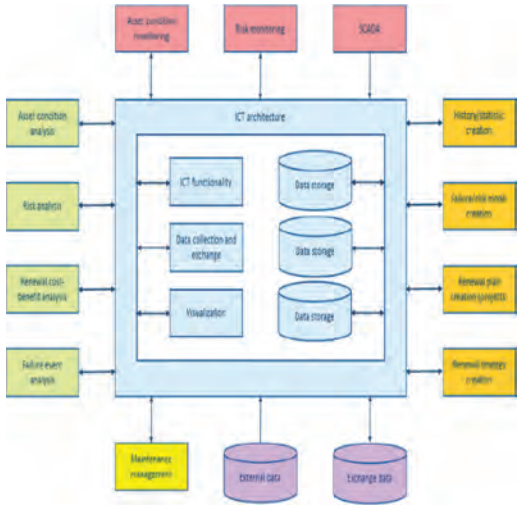


Figure 1. The SAMBA approach (Downloaded from Statnett: www.statnett.no/Global/Smarte%20Nett.pdf).

The SAMBA approach is described by the following figure:

Connected Drone is another project approved by the Norwegian Research Council and The Norwegian Water Resources and Energy Directorate (NVE). The goal of the project is to create an efficient and secure system for inspection of power lines using drones. By utilizing deep learning to analyze data obtained from sensors mounted on drones, the project seeks to give power grid utilities insights into the state of their infrastructure. The system will be delivered as an integrated, total solution and is specially adapted to the monitoring and control of power grids. The project is managed and led by eSmart Systems in close collaboration with 12 network companies and several technology partners. The first commercially available product—Connected Drone and The Intelligent Assistant, is capable of analyzing massive amounts of image data.

In the other end of the scale one can find the small net grid companies where the maintenance strategy is based on yearlong collection of experience. Through the human capital in the companies they have learned what causes degradation of their specific net, and they know the status of their components.

In between there are the big group of power companies which do their inspections, record and assess the data through Excel sheets, and perform both reactive repairs and proactive maintenance. One often-used strategy seems to be that they inspect 10% of their net (Guidelines requires a full inspection at least every 10 years). Thereafter they do repair or replacement on those components that fail according to the maintenance strategy (RENblad).

2.3 *Risk modelling and decision support*

Operational safety has over the years received more and more attention in many industries. Within the nuclear field and in the offshore industry large accidents such as Three Mile Island (US, 1979), Piper Alpha (UK, 1988) and Texas City (US, 2005) has contributed to this attention. One reason for this is that the focus in the future for the offshore industry is on operation of existing installations, extending the operational life of some of these installations as well as tying in new subsea installations.

As a way forward to understand, model and predict large accidents many risk models or barrier models has been developed, for example the BORA (Vinnem et al. 2003), OTS (Sklet et al. 2010) and Risk OMT model (Gran et al. 2012), or through master thesis works such as "Activity-based Modelling for Operational Decision-support" (Edwin 2015). Common for these models and works are that they build a model containing of target objects and background indicators.

3 WISELINE MODEL

3.1 *A pole and a line*

The Wiseline model has at its basic element an electric pole. The pole itself can be constructed by timber or by steel (as shown in Figure 2) and consist of a number of components with the mean to stabilize the pole, carry the power line, or provide as a barrier against different unwanted incidents. One such is lightning (as shown in Figure 2), thereby the need of isolators to avoid high current to spread. For a timber pole the top-hat is a component to protect the top of a pole from water impregnation and consequent decay.

At a specific geographic spot, there might be a single pole, or a number of poles. The number of poles at a single location depends on for example the need to stabilize the line when the line shifts direction. Another reason for more poles at a single location is the number of power lines the line carries. In Wiseline we call this location a "*pole site*".

Finally, a row of such pole sites forms a *line*. Typically, a line runs from one switch to another switch, i.e. representing a part of the line which can be isolated during for example maintenance.

Figure 2. An electric pole (Photo: wiseline.no).

3.2 *Inspection data*

Collection of data on a set of lines are done through inspection. A grid owner is required to do a complete inspection of every pole every 10 years. This involves an inspection of all parts of a pole, both at earth level and in the air. This is typically performed by professional inspectors using advanced tools and methods.

In between, at more frequent intervals, simpler inspections are recommended and performed. Here the poles are only inspected from a distance, and the observer can for example fly along with a helicopter or travel along using a skidoo, small engine or on foot. These methods are now being replaced by the use of drones with camera. This allows the observer to analyse the photos afterwards or also apply intelligent computer image processing methods to automatically register any abnormalities. Finally, inspections are also performed every time there is an outage. In these cases, both an inspection as well as a repair must take place.

There are in total about 30–40 inspection points for a given pole site. Some components have only one inspection point, e.g. if the top hat is intact, damaged or missing. While for others there are several inspection points, e.g. for a timber pole the following inspection points are recorded—external rot, internal rot, woodpecker holes, pole damage etc. The format of inspection data varies depending on the company that has performed the inspection. Each inspection is mostly recorded in terms of a textual description fitting with the damage categories listed in the REN-blad.

Deviations per indicator are associated with a character running from 1 to 5 (or 0 to 4), where 1 is "no finding", also equivalent to "as good as new" and 5 corresponds to "non-functioning". For this exercise of translating inspection data to character scales, input from the Energi Norge handbook (Energi Norge 2011) is used. For instance, if a "top-hat" at the top of the utility pole is missing, it translates to a 5 on the character scale. On the other hand, the character setting on a slightly bent or misoriented pole may be nuanced from between 2 to 5 depending on the degree of damage. The inspections records of such grades are used as input in the Wiseline model for each pole.

3.3 *Risk model*

In the Wiseline model each grade is translated to a value between 0.0 and 1.0, where 1 is the best. A grade 1 is always translated to the value 1.0. However, a grade 5 is translated differently depending upon the role of the component. If the component is critical it may be given a value 0.01, while if it is not critical it may be given a value 0.9. This can be illustrated by the component "warning sign". It provides no barrier in terms of delivering power, therefore a missing sign (with grade 5) will in this case be translated to 0.9. On the other hand, it provides a barrier in terms of HSE and will be translated to 0.01. Note, the numbers 0.9 and 0.01 are here only illustrative.

In real cases, the values to be used in the translation are set by the grid owner to reflect their interpretation of the importance of the inspections. The risk model is developed for three top-events – "technical condition", "HSE" and "power supply ability". Dependent on the top-event being modelled, the various components and their corresponding indicators are weighted differently in the underlying risk model. For instance, the "power transmission" subfunction will be weighted with a higher importance for the "power supply ability" model as compared to in the "HSE" model. Similarly, the high-voltage signage on utility poles will be weighted higher for the HSE model in comparison to the other two risk models. Criticality tables are established for the various components, as shown in Table 1.

The combination of the values within each group is done by a weighting function. Thereafter the groups are combined up to a risk value for the pole, the risk values at poles are combined to a risk value for the pole site (see illustration in Fig. 3). Finally, the risk values for pole sites are combined to a risk value for the line. This combination can be viewed upon as a factor model as applied in various risk models. In the current model the poles are handled independently, but obviously further work on the model will also take into account that the state of on pole will have influence on the state of its neighboring poles.

3.4 *A set of risk values*

The outcome of applying the Wiseline model is a set of risk values for each pole and for each line,

Table 1. Example—Component criticality specification.

Function/Sub function	Component	Technical condition	HSE	Power supply ability
Earthing/Protection	Voltage Signage	Medium	Important	Insignificant
Structural	Common Brace	Medium	Very Important	Very Important
Transmission	Insulator	Medium	Important	Important

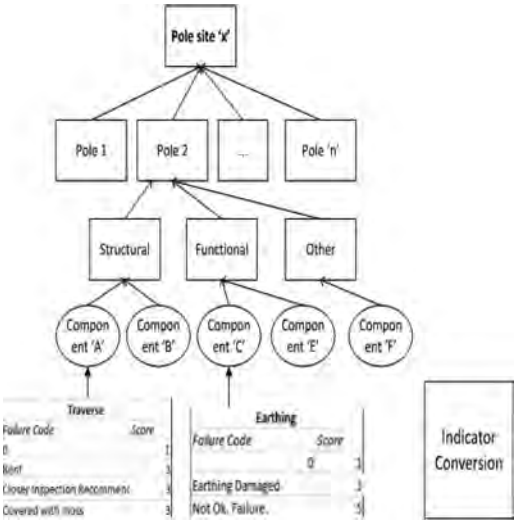


Figure 3. Combining the translated inspection values on components up to a risk value for a pole and a pole site.

Table 2. Risk values for a demo-line.

Risk factor	Risk value
Technical state	0.98
Power ability	0.99
HMS	0.96

as shown in Table 2 for demo line. The set of risk values contains of one risk value for the "technical condition" (where all components are treated equal in the combination), one for the "ability to deliver power" and one for "the HMS". For some clients, with classification of poles and lines according to, the set is extended with a fourth risk value related to the economical consequence.

The same is done for each pole. This can be used to visualize for example the HMS risk value along a line, as shown in Figure 4. Here also the translated component values for some components are included, as well as a red and yellow line representing different acceptance levels. The two lines at the bottom represents height above see and vegetation at the spots (here flat and equal vegetation all along).

3.5 *Predicting the future*

The next step in the Wiseline method is to do predictions of the future. This is done by allowing each component to degrade from its current inspected state according the reliability figure yielding for that type of component. Monte Carlo simulation is used to simulate the degradation of component condition over time and calculate the associated costs connected to the chosen maintenance strategy.

The developed risk model provides the user with significant flexibility in choosing a maintenance strategy. Examples of strategies include:

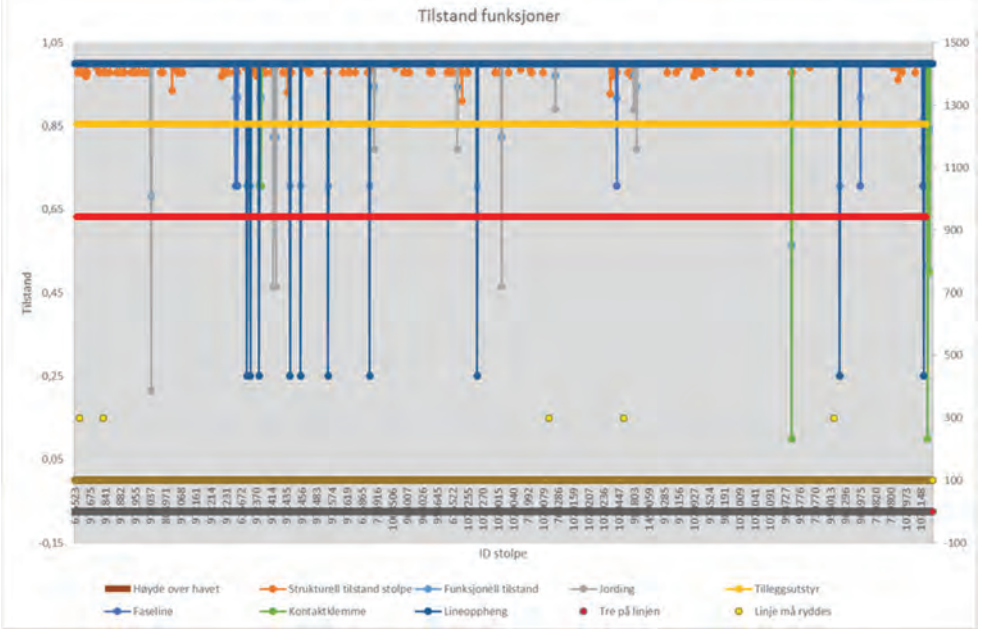


Figure 4. The risk value for HMS for the poles along the demo line. The red and yellow line represents different acceptance levels.

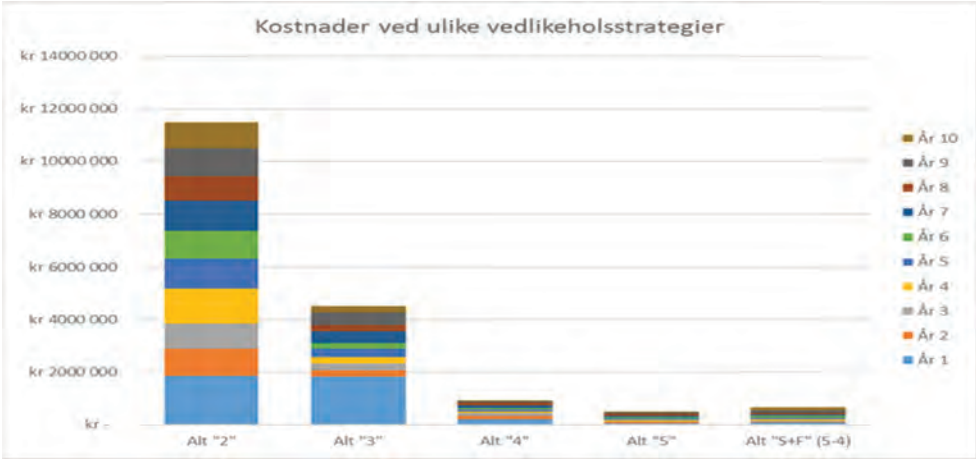


Figure 5. The cost per year for 5 different maintenance strategies.

1. No maintenance. Run-to-failure.
2. Preventive maintenance (Fix when component condition > = indicator score 'x')
3. Overhaul selected utility poles based on specified criteria at t = 'x' and run-to failure thereafter.

In the current Wiseline model the same reliability model is applied for a component independent of its age, type, producer, location etc. In the future this can be exchanged with more specific models. One such model is under development by Sintef, power grid companies and inspection companies to model how fast a timber pole will rot, and thereby provide an estimate also of the rest life time. In another project, Statnett and Sintef (Solheim et al, 2016) aims at establishing wind dependent failure rates for overhead transmission lines using reanalysis data and a Bayesian updating scheme.

3.6 *Estimating the cost of a maintenance strategy*

The final step in the Wiseline method is to associate each maintenance strategy with an investment cost. This investment cost includes both the cost of the component repaired or replaced, and the labor cost. These costs are specific for each grid owner.

The result is that decisions can be made by balancing the costs towards the acceptable risk for the different top events. In Figure 5 the cost per year for 5 different maintenance strategies are shown for a demo-case. Visual risk presentation overlaid on map data allows the decision-maker to understand the risk picture and make appropriate investment decisions based on the same. A demonstration of this shown in Figure 6 for the client Vesterålskraft.

4 CASE STUDY

To demonstrate the functionality of the developed risk model, anonymized inspection data from power lines across different Norwegian electric power distribution owners were used in the following. The data input basis included inspection points from a total of more than 300 power lines, where each line included from 10 to up to 300 utility poles. These include power voltage distribution lines right from 230V to 22kV. As discussed in Section 3, input inspection data was first mapped to the indicator set in the risk model. As the inspection data were collected in different years, the next step was to degrade each component from its current inspected state according the reliability figure yielding for that type of component up to the predicted state in 2017.

The Wiseline model was then run using the 2017 data as the starting-point for the simulations. An outline of the results is shown in Table 3. Here the results for HSE is shown, first for the calculated 2017 value, then the predicted risk value applying two different maintenance strategies: maintenance when grade 4 is predicted (VG4) and maintenance when grade 3 is predicted (VG3). Here the cost estimates are left out for simplification. However, only by observing at the lines in the table, one sees that there is a low correlation between the number of deviations and the predicted risk values. Furthermore, one also sees that the risk value does not correlate with the age of the poles either.

Figure 6. The risk value displayed on a map (Vesterålskraft).

Table 3. Outline of the results for the case study.

Line	KV	# Poles	Inspected	Pole year (min)	Pole year (max)	Deviations	HSE (2017)	HSE (2027) VG4	HSE (2027) VG3
Wiseline003	22	11	aug. 14	1963	1964	8	94.59	98.71	98.23
Wiseline006	22	6	aug. 14	1962	1985	9	94.98	99.14	99.49
Wiseline192	22	8	jun. 16	1990	1992	4	96.34	99.81	96.71
Wiseline002	22	34	aug. 14	1955	1994	22	96.53	99.35	99.82
Wiseline045	0.23	19	aug. 14	1950	2000	11	96.59	99.03	96.78
Wiseline025	0.23	12	aug. 14	1954	1978	2	97.30	99.11	99.65
Wiseline026	0.23	14	aug. 14	?	1968	9	97.42	99.23	99.58
Wiseline337	0.23	27	aug. 11	1900	1959	38	97.48	98.81	99.01
Wiseline339	0.23	15	jun. 11	1900	1990	38	97.51	98.83	99.63
Wiseline097	22	17	okt. 15	1955	2008	17	97.55	98.77	98.24
Wiseline141	22	1	jun. 16	1957	1957	3	97.58	95.88	99.76
Wiseline336	22	36	aug. 11	1974	1974	32	97.65	98.37	99.56
Wiseline046	0.23	16	aug. 14	?	2010	1	97.71	99.02	99.30
Wiseline159	22	1	jun. 16	1960	1960	2	97.79	97.79	99.34
Wiseline301	0.23	34	nov. 13	1900	2004	80	97.83	97.98	99.49

In two clients cases the degradation has been run up to 2027. At this stage, the end-user can select appropriate risk acceptance criteria for each of the risk models. By looking at the lines that do not meet the risk acceptance criteria in 2027, the clients get a proposal for which lines to address first. For the same line the associated cost by maintenance is calculated, so that a risk-based approach can be applied for the decision.

5 DISCUSSION

5.1 *Application for decision-support*

As the Wiseline model now is implemented, it applies a number of simplifications in the modelling. One such simplification is to apply the same reliability model independent of both geography and age of a component. Furthermore, the risk

model applies a limited set of component criticality values. In the next versions of the Wiseline model the intention is to extend the different parts, and also to include the effect of causes for deviation, such as lightning, wind, trees, woodpeckers etc. In its current state, the Wiseline model runs by applying Excel, and transferring the data from a client into the model, and from the model into result visualization. The plan is to address these aspects through a development and innovation project together with partners from the beginning of 2018.

5.2 *Limitations*

Two limitations in the model is of course the quality of the inspection data, and errors in the translation of the inspection data into the input data. The more manual steps there are in this process, the more influence it can have. On the other hand, when the inspection data is professional collected and exported directly from an inspection data base, like the database of NordiConsult, this problem is minimized.

Another limitation lays in the interpretation of the risk factors. The technical state, where all indicators are treated equal, is a construct. The HSE factor should probably be split into inspector/skilled laborer and third party (animals and general public). Finally, the "ability to deliver power" is not the same as "causing an outage" and do not say much about how long an outage will be. Thereby the Wiseline model do not say much about expected costs due to outage.

6 CONCLUSION

Planning maintenance over a large, distributed and ageing power infrastructure is a challenge. This paper has presented a novel method developed to model the risk of power line infrastructure. The model uses inspection data, reliability models for the aging of components and associated maintenance costs to present a risk picture and the associated expected costs based on the user's maintenance strategy of choice. Visual risk presentation overlaid on map data allows the decision-maker to understand the risk picture and make appropriate investment decisions based on the same. The developed model is currently being used in-house at various grid operators in Norway. Here the Wiseline model has demonstrated that there is no single maintenance strategy that fits all cases. What the Wiseline model do is to enable the grid operators to do risk-based decisions.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contribtion of our partners Safetec Nordic AS and NordiConsult AS. We also specially acknowledge the support of our clients Vesterålskraft Nett AS, NTE Nett AS, Troms Kraft Nett AS and Vokks Nett AS.

REFERENCES

Energi Norge. Tilstandskontroll av kraftnett—Samleutgave, 2011.

FEF 2006, Forskrift om Elektriske Forsyningsanlegg med veiledning. 1. utgave–3. opplag. Lysaker Pronorm AS, 2010.

Fornybar.no, Overforing og lagring av energi/kraftoverforing. http://www.fornybar.no/overforing-og-lagring-av-energi/kraftoverforing (accessed 9.26.17).

Gran, B.A., Bye, R., Nyheim, O.M., Okstad, E.H., Seljelid, J., Sklet, S., Vatn, J. & Vinnem, J.E. Evaluation of the risk model of maintenance work on major process equipment on offshore petroleum installations. Journal of Loss Prevention in the Process Industries, Volume 25, Issue 3, May 2012, Pages 582–593, 2012.

IEC 60300-3-11, Dependability management—Part 3–11: Application guide—Reliability centred maintenance, 2009.

Kjølle, G., KILE-satsene og hva de dekker, 2011.

Moubray, J., Reliability-centered maintenance. Industrial Press Inc, 1997

REN, https://www.ren.no/liste-over-alle-renblad (accessed 11.23.17).

Sklet, S., Ringstad, A.J., Steen, S.A., Tronstad, L., Haugen, S. Seljelid, J., Kongsvik, T. and Wærø, I.; Monitoring of Human and Organizational Factors Influencing Risk of Major Accidents. *SPE International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production,* Rio de Janeiro, Brazil, 12–14 April, 2010.

Solheim, Ø.R., Trötscher, T., Kjølle, G. Wind dependent failure rates for overhead transmission lines using reanalysis data and a Bayesian updating scheme. Presented at the 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Beijing, 2016.

Statnett, Årsstatistikk 2016 1–22kV. www.statnett.no, (accessed 9.26.17), 2017.

Vinnem, J.E. et al.; Risk assessment for offshore installations in the operational phase, presented at the ESREL 2003, Maastrict, 16–18 June, 2003.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

The real-time reliability evaluation and sequential inspection decision based on Wiener process

Senyang Bai, Zhijun Cheng, Yong Yang & Bo Guo
College of Systems Engineering, National University of Defense Technology, Changsha, Hunan, P.R. China

ABSTRACT: Due to the complexity and high reliability requirements of aerospace equipment, it's a significant challenge to establish the efficiency maintenance management in terms of little failure data or even no failure data in the actual operation or storage environment. This paper investigates the issue of real-time reliability and sequential inspection policy for aerospace equipment based on a Wiener process-based degradation model. Firstly, the stochastic characteristics of the equipment's degradation is described by the Wiener process. The product-to-product variability and the temporal uncertainty of the degradation can be characterized by the random drift parameters. Secondly, the expression of reliability distribution is obtained with close form by use of the first-hitting time theory. An adaptive method is proposed to evaluate the unknown parameters with the optimal smoothing algorithm (RTS) and the Expectation Maximization algorithm (EM). Once the new degradation information is available, the parameters should be updated with Bayesian equation. Moreover, the historical information of the same type product can also be integrated in the selection of the initial parameters to ensure the convergence in the iterative updating. Thirdly, a sequential inspection model is discussed to determine the optimal intervals to satisfy the requirements for the real-time reliability at the certain time. Finally, an example of fatigue crack growth in an aerospace aluminum alloy is given to illustrate the validity of the proposed method.

1 INTRODUCTION

The majority of aviation products are high reliability products, such as aircraft engines, satellites, gyroscopes, which need to maintain a certain degree of reliability during long storage or operation process. Due to a variety of factors (such as temperature stress, fatigue, corrosion, etc.), their internal material properties will change, including the corrosion of mechanical parts, rubber materials, aging and so on. It results in a reduction in the reliability of the product. Therefore, to improve the reliability and minimize the times of inspection or maintenance, it is of vital significance to optimize the inspection and maintenance policy.

The Wiener process-based degradation model is a statistical data driven method, it has been extensively utilized in modeling degradation paths both academically and practically. Its biggest advancement is that the distribution of failure time can be formulated as an inverse Gaussian distribution. Ray (2002) and Ye (2015) et al. successfully applied Wiener process to fatigue crack growth analysis. Gebraeel (2005) et al. described the degradation path of rotating element bearing by Wiener process. Aiming at the effect of the non-stationary feature and the delay of state detection caused by discrete inspection on long-run operation cost, Zhang (2016) proposed an optimal inspection-based maintenance policy for three-state mechanical components subject to competing failure modes. A maintenance optimization model for mission-oriented systems based on Wiener degradation is proposed by Guo (2013). These all demonstrate that Wiener process model can achieve satisfactory result by using large sample data. However, most aviation products are costly and high reliability, it is almost impossible to obtain enough data of the failure life through the life test or accelerated life test, and even "zero failure" phenomenon may occur. Therefore, we will consider establishing the degradation model based on Wiener process by using the small sample data, so as to be more suitable for aviation products.

Besides, in order to ensure the normal operation of products and preventive maintenance, variation in reliability under different conditions should be considered. Real-time reliability evaluation is ideal because the factors that affect reliability change continuously with time for dynamic systems. Yan (2016) developed a two-phase Wiener degradation model to evaluate the real-time reliability of devices. Xu et al. (2008) proposed a real-time reliability prediction method for a dynamic system that suffers from a hidden degradation process. Wang et al. (2014) investigated the issue of real-time

reliability evaluation based on a general Wiener process-based degradation model. Faghih-Roohi et al (2014) developed a dynamic model for availability assessment of multi-state weighted k-out-of-n systems using the universal generating function and Markov process. Each method has its advantages and disadvantages. For example, Markov models have state space explosion problems (Tanrioven (2004)). The major limitation of stochastic simulation is that the monitoring value has to be a scalar. However, there are usually many monitoring values exist in real situations. The hidden degradation process identification is unsuitable for a situation in which the characteristics of the degradation processes are time varying, or the path function of the fault process is nonlinear. Therefore the temporal uncertainty of the degradation will be characterized by the random drift parameters in the model of this paper, which is defined as a random walk model, thus we can improve the degradation model.

It is a significant challenge for engineers to define the appropriate inspection interval in term of uncertainty in product deterioration and environment change. Fewer inspections will lead to lower reliability, and frequent inspection will lead to higher cost, so the optimal inspection policy should be set up and tradeoff between the reliability and operation cost. For the long-storage products, Feng (2011) proposed a sequential inspection method based on Weibull distribution, and the sequential inspection time is confirmed based on storage reliability request. Cui (2004) developed a sequential inspection policy for multiple systems under availability requirement. And considering the optimization of alarm threshold, a sequential inspection scheme is determined by Jiang (2010). For systems subject to stochastic deterioration, Zhu (2017) proposed a sequential condition-based maintenance policy. The above sequential inspection methods are based on the constraint of the degradation threshold, availability, cost and so on to optimize the maintenance and inspection decision. But for the aviation products studied in the paper, real-time reliability is more necessary to be assessed. Therefore we will focus on improving the related sequential inspection and maintenance policy by using the real-time reliability model, which will ensure the high reliability of the aviation products.

The aim of this paper is to find the optimal maintenance policy for the aviation products. The model of this paper can makes full use of the mathematical properties of Wiener process, and the uncertainty of stochastic model parameter estimation is also taken into consideration. To ensure the convergence of parameters in the iterative updating, the historical information of the same type product is integrated in the selection of the initial parameters. Moreover, a variety of algorithms are used to iteratively update the model parameters, which realize the real-time updating of parameter estimation and the adaptive prediction of remaining life. Combined with the requirement of real-time reliability, the sequential inspection policy is determined for the aviation product, thus reducing the number of inspection and maintenance effectively.

The remainder of this paper is organized as follows. Section 2 describes the model assumptions and maintenance policy. In Section 3, the real-time reliability model of the product is established and the unknown parameters are estimated. Section 4 presents an optimal model for sequential inspection. In section 5, a numerical example of fatigue crack growth in an aerospace aluminum alloy is given to illustrate the validity of the proposed method. Conclusions and some future works are drawn in Section 6.

2 MODEL DESCRIPTION

2.1 *Model assumptions*

1. It is assumed that the performance of the product is recovered as new after replacement

Nomenclature	
μ	drift coefficient
μ_0	initial drift coefficient
σ	diffusion coefficient
w_0	preventive maintenance threshold
w	failure threshold
η	the random walk parameter of μ
a_0	the mean of normal distribution about μ_0
D_0	the variance of normal distribution about μ_0
$D_{k\|k}$	the updated variance of μ
$D_{k\|k-1}$	the conditional variance of μ
$\boldsymbol{\theta}$	set of the model parameter$(a_0, D_0, Q_0, \sigma_0^2)$
R_s	requirement of the real-time reliability for the product (constant)
Δt_{ik}	inspection interval
t_i	time of the i-th maintenance
t_{ik}	time of k-th inspection after i-th maintenance
x_{ik}	degradation data of k-th inspection after i-th maintenance
$x_{i(0:k)}$	set of degradation data after i-th maintenance
l_{ik}	remaining life of the profuct at the time t_{ik}
$p(\mu_k \| X_{0:k})$	the conditional PDF of μ_k for the given $X_{0:k}$
$f_{T\|x_k}(l \mid x_k)$	the PDF of the remaining life when the inspection data is x_k
$f_{T_{ik}\|X_{0:k}}(l_{ik} \mid X_{0:k})$	the PDF of the remaining life for the product when the inspection data is $X_{0:k}$
$R_s(t \mid x_k)$	real-time reliability of the product at time t
$R_s(l_{ik} \mid X_{0:k})$	real-time reliability of the remaining life for the product

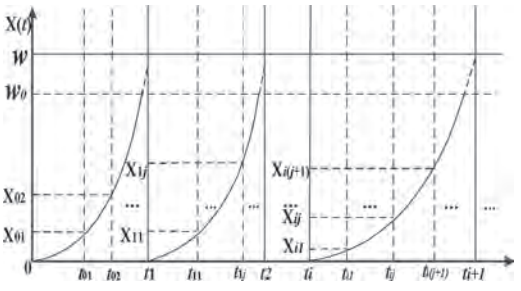


Figure 1. Inspection and maintenance cycle diagram for the product.

2. According to the actual requirement of engineering, R_S is generally set to be a constant.
3. Failures are observed only by inspection, and if degeneration data of inspection exceed the preventive maintenance threshold w_0, the maintenance should be carried out and the component will be replaced.
4. The inspection time and replacement time are negligible.

2.2 *Deterioration and maintenance process*

In this paper, a metal material of an important component from the aviation product is taken as the research object. It is found that the degradation processes of many aviation products performance are random and uncertain. The Wiener process can well describe the time uncertainty of the degradation process, and it is also easier to deal with the error of data measurement. Due to the long service life and high cost of maintenance for aviation products, periodic inspection or real-time monitoring is generally required. If the performance degradation is found to exceed the preventive maintenance threshold w_0, the component will be replaced immediately. The diagram of inspection and maintenance cycle is shown in Figure 1.

The Wiener process is used to establish the degradation model of aviation products. $R_S(t \mid x_k)$ is defined as the real-time reliability at time t. And according to the definition of real-time reliability, the value of $R_S(t \mid x_k)$ would gradually decreases from 1 after each inspection as time goes on. And the process can be divided into three steps based on the sequential inspection policy. Firstly, the real-time reliability model is established based on Wiener process; then, the model parameters are estimated by using the degradation data $x_{i(0:k)}$, so the PDF expression $f_{L_k|X_{0:k}}(l_{ik}|X_{0:k})$ of the remaining life for the product can be obtained, furthermore the expression $R_S(l_{ik} \mid X_{0:k})$ of the real-time reliability can be derived. Finally, the inspection interval Δt_{ik} is determined to satisfy the requirement of R_S, so the next inspection time is $t_{i(k+1)} = t_{ik} + \Delta t_{ik}$.

3 REAL-TIME RELIABILITY

If the aviation product is known to have not failed at the current moment, a conditional probability can be used to express its reliability level. When new degradation data is obtained, the real-time reliability is further updated.

3.1 *The model of real-time reliability*

For the actual operation or storage of aviation products, the degradation process is described by the Wiener process, so we can get the following expression by using the Markov property of $\{B(t), t \geq 0\}$:

$$X(t) = x_k + \mu(t - t_k) + \sigma B(t - t_k) \tag{1}$$

It is assumed that $x_k < w$ at the time t_k, otherwise, the device is considered invalid. In this study, the definition of product life is based on the first-hitting time theory for random process proposed by Lee (2007), then the PDF of the product life can be obtained:

$$f_{T|x_k}(t \mid x_k) = \frac{w - x_k}{\sqrt{2\pi(t - t_k)^3 \sigma^2}} \exp\left\{-\frac{(w - x_k - \mu(t - t_k))^2}{2\sigma^2(t - t_k)}\right\} \tag{2}$$

Due to the small sample of the degradation data for the aviation product, it is important to integrate all the data from each inspection into degradation modeling. In this study, μ is defined as a random walk model, which can be updated as the new inspection data is obtained, that is $\mu_k = \mu_{k-1} + \eta$, and $\eta \sim N(0, Q)$. Then based on the state-space model framework, the degradation process of the aviation product can be constructed by a linear state-space model as:

$$\mu_k = \mu_{k-1} + \eta \tag{3}$$

$$x_k = x_{k-1} + \mu_{k-1}(t_k - t_{k-1}) + \sigma\varepsilon_k \tag{4}$$

where $t_0 = 0$, $x_0 = 0$, $\varepsilon_k \sim N(0, t_k - t_{k-1})$ and $t_k - t_{k-1}$ is the variance of ε_k based on the characteristics of Brownian motion.

It is assumed that the initial drift coefficient follows $\mu_0 \sim N(a_0, D_0)$, and the drift coefficient μ can be regarded as an implicit "state" estimated from the inspection data $X_{0:k} = \{x_0, x_1, x_2, ..., x_k\}$. Therefore, the above state-space model establishes the relationship between the drift coefficient and all inspection data from the beginning to the present time t_k. In equation (3), once the new degradation data is obtained, the probability distribution of μ_k can be calculated from all the inspection data by using recursive filtering method. And the posterior estimation expectation of μ_k can be defined as $\hat{\mu}_k = E(\mu_k \mid X_{0:k})$, the corresponding variance

can be defined as $D_{k|k} = \mathrm{var}(\mu_k \mid X_{0:k})$. In order to obtain $\hat{\mu}_k$ and $D_{k|k}$, we need to get the PDF of μ_k for a given $X_{0:k}$, which denoted by $p(\mu_k \mid X_{0:k})$, and $p(\mu_k \mid X_{0:k})$ can be recursively calculated from $p(\mu_{k-1} \mid X_{0:k-1})$ using Bayesian rules. According to the stochastic filter theory and the strong tracking filtering technique, the conditional PDF of μ_k can be obtained:

$$p(\mu_k \mid X_{0:k}) = \frac{1}{\sqrt{2\pi D_{k|k}}} \exp\left\{-\frac{(\mu_k - \hat{\mu}_k)^2}{2D_{k|k}}\right\} \tag{5}$$

According to the first-hitting time theory of random process, we can use equation (6) to measure the real-time reliability of the product:

$$R(t; X_{0:k}, \boldsymbol{\theta}_k) = P(X(\tau) < w, \forall\, \tau \in [t_k, t] \mid \boldsymbol{\theta}_k, X_{0:k}, x(t_j) < w, j = 1,2,\cdots,k) \tag{6}$$

where $\boldsymbol{\theta}_k = \left(a_k, D_k, Q_k, {\sigma_k}^2\right)^T$. In order to get the expression of $R(t; X_{0:k}, \boldsymbol{\theta}_k)$ the definition of the remaining life for the product is given firstly:

$$L_k = \inf\{l_k : X(l_k + t_k) \geq w \mid X_{0:k}, x(t_j) < w, j = 1,2,\cdots,k\} \tag{7}$$

So $R(t, X_{0:k}, \boldsymbol{\theta}_k)$ can be expressed as follows:

$$R(t; X_{0:k}, \boldsymbol{\theta}_k) = P(L_k > t - t_k) \tag{8}$$

According to the literature (Si (2016), Peng (2009)), and combined with the state-space models of equation (3) and equation (4), the PDF of remaining life L_k can be obtained as follows:

$$f_{L_k|X_{0:k}}(l_k \mid X_{0:k}) = \frac{w - x_k}{\sqrt{2\pi {l_k}^3 (D_{k|k} l_k + \sigma^2)}} \times \exp\left\{-\frac{(w - x_k - \hat{\mu}_k l_k)^2}{2l_k(D_{k|k} l_k + \sigma^2)}\right\}, l_k > 0 \tag{9}$$

Then when $t - t_k = l_k$, the real-time reliability of the remaining life l_k can be obtained:

$$\begin{aligned} R_S(l_k \mid X_{0:k}) &= \int_{l_k}^{\infty} f_{L_k|X_{0:k}}(l_k \mid X_{0:k}) dl_k \\ &= \Phi\left(\frac{w - x_k - \hat{\mu}_k l_k}{\sqrt{D_{k|k} {l_k}^2 + \sigma^2 l_k}}\right) \\ &\quad - \exp\left\{\frac{2\hat{\mu}_k (w - x_k)}{\sigma^2} + \frac{2D_{k|k}(w - x_k)^2}{\sigma^4}\right\} \\ &\quad \times \Phi\left(-\frac{2D_{k|k}(w - x_k) l_k + \sigma^2(\hat{\mu}_k l_k + w - x_k)}{\sigma^2 \sqrt{D_{k|k} {l_k}^2 + \sigma^2 l_k}}\right) \end{aligned} \tag{10}$$

The parameters a_k, D_k, Q_k and ${\sigma_k}^2$ of the random degradation model are unknown in equation (8), (9) and (10), which need to be estimated based on the inspection data $X_{0:k}$. Therefore, Section 3.2 mainly describes the adaptive estimation method of unknown parameters.

3.2 *The adaptive estimation method of parameters*

In order to sequentially evaluate the unknown parameters, it is necessary to use all the inspection data of the product from the start of the operation or storage to the current time. Once the new inspection data is acquired, the previous parameters can be updated recursively. The unknown parameter vector is denoted as $\boldsymbol{\theta} = (a_0, D_0, Q, \sigma^2)^T$. When new degradation data x_k is obtained, the maximum likelihood estimation method is used to estimate $\boldsymbol{\theta}$. The log-likelihood function based on $X_{0:k}$ can be expressed as:

$$L_k(\boldsymbol{\theta}) = \log[p(X_{0:k} \mid \boldsymbol{\theta})] \tag{11}$$

where $p(X_{0:k} \mid \boldsymbol{\theta})$ is a joint PDF of degradation data $X_{0:k}$. Then, based on the likelihood function of the degradation data $X_{0:k}$, the maximum likelihood estimation value $\hat{\boldsymbol{\theta}}_k$ of $\boldsymbol{\theta}$ can be obtained by maximizing the likelihood function, which can be expressed as:

$$\hat{\boldsymbol{\theta}}_k = \arg\max_{\theta} L_k(\boldsymbol{\theta}) \tag{12}$$

The EM algorithm can be used to approximate the maximum likelihood estimation value of the parameter by maximizing the likelihood function $p(X_{0:k}, \boldsymbol{U}_k \mid \boldsymbol{\theta})$ $\left(\boldsymbol{U}_k = \left(\mu_0, \mu_1, ..., \mu_k\right)\right)$. According to the relationship between $p(\mu_k \mid X_{0:k})$ and $p(X_{0:k}, \boldsymbol{U}_k \mid \boldsymbol{\theta})$, $L_k(\boldsymbol{\theta})$ can be divided into two parts, namely:

$$L_k(\boldsymbol{\theta}) = \ell_k(\boldsymbol{\theta}) - \log p(U_k \mid X_{0:k}, \boldsymbol{\theta}) \tag{13}$$

where

$$\ell_k(\boldsymbol{\theta}) = \log p(X_{0:k}, U_k \mid \boldsymbol{\theta}) \tag{14}$$

According to the adaptive estimation method proposed by Si (2016), the unknown parameters of the model can be divided into the following steps:

- Step 1: The joint log-likelihood function in the EM algorithm can be expressed as:

$$\begin{aligned} \ell_k(\boldsymbol{\theta}) &= \log p(\mu_0 \mid \boldsymbol{\theta}) + \log \prod_{j=1}^{k} p(\mu_j \mid \mu_{j-1}, \boldsymbol{\theta}) \\ &\quad + \log \prod_{j=1}^{k} p(x_j \mid \mu_{j-1}, \boldsymbol{\theta}) \end{aligned} \tag{15}$$

As a result, the conditional expectation $E[\ell_k(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_k^i)]$ of $\ell_k(\boldsymbol{\theta})$ is as follows:

$$E[\ell_k(\boldsymbol{\theta}\mid\hat{\boldsymbol{\theta}}_k^i)] = E_{U_k|X_{0:k},\hat{\theta}_k^i}[\ell_k(\boldsymbol{\theta})]$$
$$= \frac{1}{2}E_{U_k|X_{0:k},\hat{\theta}_k^i}[-\log D_0 - (\mu_0 - a_0)^2 D_0 - \sum_{j=1}^{k}(\log Q + (\mu_j - \mu_{j-1})^2 Q) - \sum_{j=1}^{k}(\log\sigma^2 + (x_j - x_{j-1} - \mu_{j-1}(t_j - t_{j-1}))^2 / (\sigma^2(t_j - t_{j-1})))] \tag{16}$$

- Step 2: The Rauch-Tung-Striebel (RTS) optimal smoothing algorithm is used to calculate the conditional expectations $E_{U_k|X_{0:k},\hat{\theta}_k^i}(\mu_j)$, $E_{U_k|X_{0:k},\hat{\theta}_k^i}(\mu_j^2)$ and $E_{U_k|X_{0:k},\hat{\theta}_k^i}(\mu_j\mu_{j-1})$:
- Step 3: Substituting the results of $E_{U_k|X_{0:k},\hat{\theta}_k^i}(\mu_j)$, $E_{U_k|X_{0:k},\hat{\theta}_k^i}(\mu_j^2)$ and $E_{U_k|X_{0:k},\hat{\theta}_k^i}(\mu_j\mu_{j-1})$ into equation (16), the specific expression of $E[\ell_k(\boldsymbol{\theta}\mid\boldsymbol{\theta}_k^i)]$ can be obtained. This completes the E step in the EM algorithm.
- Step 4: Followed by the M step, according to the literature (Si (2016)), the parameter $\hat{\theta}_k^{(i+1)}$ estimated from the i+1 step iteration in M step is the global unique optimal solution of the equation(12), there is $\hat{\theta}_k^{(i+1)} = (a_{0k}^{(i+1)}, D_{0k}^{(i+1)}, Q_k^{(i+1)}, (\sigma^2)_k^{(i+1)})^T$. Specific proof of the process and convergence analysis of the algorithm can be found in the literature (Si (2016)).

Thus the parameter estimation value $\hat{\boldsymbol{\theta}}_k = (a_{0k}^{(i+1)}, D_{0k}^{(i+1)}, Q_k^{(i+1)}, (\sigma^2)_k^{(i+1)})^T$ is obtained, which will be used as the model parameter of the real-time reliability in the next inspection time.

4 DETERMINATION OF SEQUENTIAL INSPECTION INTERVALS

In order to make the model parameters estimation converge faster and obtain higher accuracy, we can use the historical data of similar aviation products to estimate the approximate initialization parameters $\boldsymbol{\theta}_0 = (a_0, D_0, Q_0, \sigma_0^2)^T$. For an important component of aviation product, it is generally necessary to set the real-time reliability R_S (constant) according to the task requirement, so as to ensure the reliability of the product before inspection or maintenance.

4.1 *Determination of the initial inspection time t_{01} and t_{02} for the new product*

In order to estimate the parameter $\hat{\theta}_k$ more accurately, and make parameter estimation converge more quickly, two state information of product are required at least. Conservatively, the interval for the first inspection is set to be the same as the second inspection considering the product safety and reliability, i.e. t_{01}=$t_{02}/2$. Since the real-time reliability should be guaranteed before the next inspection, there is following expression based on equation (10):

$$R_S(l_k \mid X_{0:k}) = \Phi\left(\frac{w - x_k - \hat{\mu}_k l_k}{\sqrt{D_{k|k}{l_k}^2 + \sigma^2 l_k}}\right) - \exp\left\{\frac{2\hat{\mu}_k(w - x_k)}{\sigma^2} + \frac{2D_{k|k}(w - x_k)^2}{\sigma^4}\right\} \times \Phi\left(-\frac{2D_{k|k}(w - x_k)l_k + \sigma^2(\hat{\mu}_k l_k + w - x_k)}{\sigma^2\sqrt{D_{k|k}{l_k}^2 + \sigma^2 l_k}}\right) \geq R_S \tag{17}$$

Thus, the second inspection time t_{02} satisfies the following formula:

$$\Phi\left(\frac{w - x_0 - \hat{\mu}_0 t_{02}}{\sqrt{D_0{t_{02}}^2 + \sigma_0^2 t_{02}}}\right) - \exp\left\{\frac{2\hat{\mu}_0(w - x_0)}{\sigma_0^2} + \frac{2D_0(w - x_0)^2}{\sigma_0^4}\right\} \times \Phi\left(-\frac{2D_0(w - x_0)t_{02} + \sigma_0^2(\hat{\mu}_0 t_{02} + w - x_0)}{\sigma_0^2\sqrt{D_0{t_{02}}^2 + \sigma_0^2 t_{02}}}\right) \geq R_S \tag{18}$$

Based on the monotonicity of the distribution function, it is easy to get the solution $\hat{t}_{02}$ of the unary equation by using MATLAB, then:

$$t_{02} \leq \hat{t}_{02}$$

According to engineering practice, the second inspection time can be set as t_{02}= $\hat{t}_{02}$, so the first inspection time is t_{01}=$t_{02}/2$=$\hat{t}_{02}/2$. then the product can be respectively inspected at t_{01} and t_{02} to obtain new degradation data x_{01} and x_{02}.

4.2 *Determination of the inspection time $t_{i(k+1)}(i \geq 0, k \geq 0)$*

After obtaining the product degradation data x_{01} and x_{02}, the new parameter $\hat{\boldsymbol{\theta}}_{\mathbf{01}}$ can be determined by using the adaptive estimation method to integrate x_{01} and $\boldsymbol{\theta}_0$. Similarly, the parameter $\hat{\boldsymbol{\theta}}_{\mathbf{02}}$ is adaptively estimated by integrating the degradation data x_{01}, x_{02} and parameter $\hat{\boldsymbol{\theta}}_{\mathbf{01}}$, then the inspection interval $\Delta\hat{t}_{02}$ can be determined according to the requirement of R_S, so the third inspection time is $t_{03} = t_{02} + \Delta\hat{t}_{02}$.

If the degradation data of the next inspection still does not exceed the preventive maintenance threshold w_0, the new inspection data w_{0k} and parameter $\hat{\boldsymbol{\theta}}_{\mathbf{0(k-1)}}$ will be integrated to estimate the parameter $\hat{\boldsymbol{\theta}}_{\mathbf{0k}}$ recursively, then the interval after each inspection can be determined. Similarly, the next inspection time will meet the following condition:

$$t_{0(k+1)} = t_{0k} + \Delta\hat{t}_{0k} \tag{19}$$

For further consideration, if the degradation data x_{0n} exceeds the preventive maintenance threshold w_0 during the inspection, this important component will be replaced. Then all the degradation data $x_{(i-1)(0:n)}$ of inspections after the $(i-1)$-th $(i \geq 1)$ maintenance is used to estimate the parameter $\hat{\boldsymbol{\theta}}_{i0}$ adaptively, and $\hat{\boldsymbol{\theta}}_{i0}$ will be taken as the initial parameter of the model after the i-th maintenance. Similarly, all the inspection data $x_{i(0:k)}$ is used to estimate the model parameters $\hat{\boldsymbol{\theta}}_{ik}$ adaptively, then the interval Δt_{ik} can be determined, and the interval Δt_{ik} satisfies the following condition:

$$\begin{aligned}&\Phi\left(\frac{w-x_{ik}-\hat{\mu}_{ik}\Delta t_{ik}}{\sqrt{D_{ik}\Delta t_{ik}{}^2+\sigma_{ik}^2\Delta t_{ik}}}\right)\\&-\exp\left\{\frac{2\hat{\mu}_{ik}(w-x_{ik})}{\sigma_{ik}^2}+\frac{2D_{ik}(w-x_{ik})^2}{\sigma_{ik}^4}\right\}\\&\times\Phi\left(-\frac{2D_{ik}(w-x_{ik})\Delta t_{ik}+\sigma_{ik}^2(\hat{\mu}_{ik}\Delta t_{ik}+w-x_{ik})}{\sigma_{ik}^2\sqrt{D_{ik}\Delta t_{ik}{}^2+\sigma_{ik}^2\Delta t_{ik}}}\right)\\&\geq R_S\end{aligned} \tag{20}$$

Similarly, $\Delta\hat{t}_{ik}$ can be easily obtained by using the MATLAB, so there is:

$$t_{i(k+1)} = t_{ik} + \Delta\hat{t}_{ik} \tag{21}$$

5 NUMERICAL EXAMPLE

In this section, a certain type of aluminum alloy material (Meeker (1998)) commonly used in aviation products is taken as an example, the quality of this material is generally evaluated by the length of its fatigue crack. The initial crack length of the material is 0.9 inches, and the failure threshold is set to be $w = 1.6$ inches, the preventive maintenance threshold is set to be $w_0 = 1.58$ inches.

5.1 *Evaluation of the real-time reliability model*

In the experiment, firstly the approximate initial parameter $\hat{\boldsymbol{\theta}}_0 = (a_0, D_0, Q_0, \sigma_0^2)^T = (6.5, 0.5, 0.5, 0.1)^T$ is obtained by integrating the fatigue crack growth data of the same type product. In order to verify the effectiveness of the model proposed in this paper, we select the crack length information corresponding to time $\boldsymbol{t}_{1:28}$ as the degradation data. And for each inspection time, the degradation information $\boldsymbol{X}_{1:h}$ before the inspection time is set as the history degradation information, thus the parameter of degradation model can be updated by using the proposed method, then the future crack length of the material can be predicted by the formula $\hat{x}_{k+1} = x_k + \hat{\mu}_k(t_{k+1} - t_k)$ step by step. Finally, the result is compared with the real curve

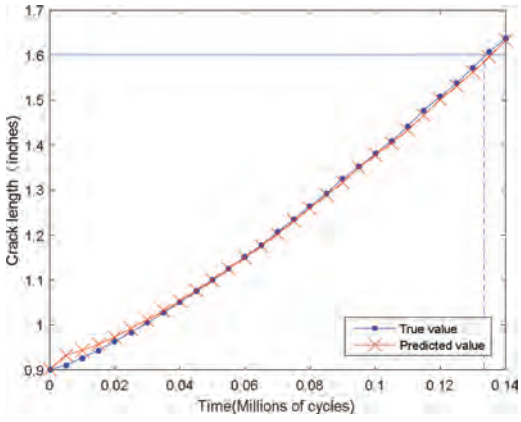


Figure 2. The comparison plot of crack growth curve.

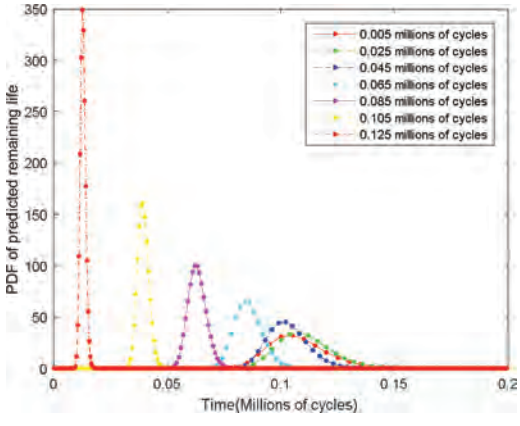


Figure 3. PDF of predicted remaining life at seven different inspection points.

of fatigue crack growth for the aluminum alloy. The comparison plot is shown in Figure 2.

According to the result of the experiment, the theoretical failure time of crack length reaches the failure threshold is about 0.13343 million cycles, and the mean square error between the predicted value and the true value is: $\mathbf{MSE} = 6.5575e\text{-}05$.

Figure 3 gives a PDF plot of the predicted remaining life at seven different inspection points. The curves gradually shift left and become sharper as data accumulates, which means that the uncertainty of the remaining life prediction is decreasing as more and more data are used to estimate the parameters of the model. According to Fig. 2 and Fig. 3, it can be seen that the model proposed in this paper fits very well with the real model.

5.2 *Determination of sequential inspection intervals*

The calculation process of sequential inspection interval is given as follows: Firstly, R_S is set to be

Table 1. The calculation results of the sequential inspection interval ($R_S = 0.9$).

The serial number	The initial state ($t_0 = \mathbf{0}$)	The new product						Replace as new (t_1)	After one maintenance				Replace a new (t_2)	After two maintenances			
		t01	t02	t03	t04	t05	t06		t11	t12	t13	t14		t21	t22	t23	…
t_{ik}/millions of cycles	0.0000	0.0429	0.0858	0.1171	0.1274	0.1309	0.1322	0.1322	0.2328	0.2575	0.2629	0.2645	0.2645	0.3725	0.3941	0.3960	…
xik/inches	0.9000	1.0647	1.2970	1.4851	1.5541	1.5792	1.5825	0.9000	1.3860	1.5417	1.5783	1.5885	0.9000	1.4294	1.5744	1.5814	…
a_{ik}	6.5000	6.3122	6.2132	6.0897	5.9436	5.8073	5.7091	5.7091	5.6880	5.6158	5.5307	5.5032	5.5032	5.4906	5.4340	5.9370	…
D_{ik}	0.5000	0.4647	0.4403	0.4070	0.3668	0.3302	0.3099	0.3099	0.3025	0.2685	0.2316	0.2211	0.2211	0.2164	0.1867	0.1671	…
Q_{ik}	0.5000	4.2608	2.9779	2.0715	1.5555	3.9689	12.0634	12.0634	1.4950	1.1045	4.4839	9.9541	9.9541	0.8480	1.1483	10.9867	…
$(\hat{\sigma}^2)$ *ik*	0.1000	0.2829	0.1780	0.1211	0.0809	0.0594	0.0575	0.0575	0.1048	0.0659	0.0426	0.0357	0.0357	0.0609	0.0539	0.0386	…
$\Delta\hat{t}_{ik}$ / millions of cycles	0.0429	0.0429	0.0312	0.0103	0.0035	0.0013	—	—	0.10007	0.0246	0.0054	0.0016	—	0.1080	0.0216	0.0019	—	…
Remarks	The degradation data exceeds the preventive maintenance threshold 1.58 inches at the sixth inspection, which needs to be replaced.							The degradation data exceeds the preventive maintenance threshold at the fourth inspection, which needs to be replaced.					The degradation data exceeds the preventive maintenance threshold at the third inspection, which needs to be replaced.				…

$R_S = 0.9$, and $t_{02} = 0.0858$ can be obtained by using the method proposed in Section 4.1, so $t_{01} = t_{02}/2 = 0.0429$; Secondly, when $i = 0$, $k = 1, 2$, $\hat{\theta}_{01} = (a_{01}, D_{01}, Q_{01}, \sigma_{01}^2)^T$ and $\hat{\theta}_{02} = (a_{02}, D_{02}, Q_{02}, \sigma_{02}^2)^T$ are respectively estimated by the adaptive estimation method. Thirdly, the inspection interval $\Delta\hat{t}_{ik}$ is determined according to the requirement of real-time reliability, i.e. $t_{i(k+1)} = t_{ik} + \Delta\hat{t}_{ik}$; when $i = 0, k = 2$, t_{03} is obtained; when $i = 0, k = 3$, t_{04} is obtained. If the crack length exceeds the preventive maintenance threshold, the product will be replaced immediately. In the second operating phase, the last parameter $\hat{\boldsymbol{\theta}}_{06}$ of the first phase is taken as the initial parameter, and the parameters are recursively updated, then the time of each inspection and maintenance can be determined, the following inspection time will be obtained in the same way. The inspection time t_{ik} determined by the proposed sequential method is shown in the second row of Table 1, and the third row shows the degradation data x_{ik} of each inspection. Finally, the following results can be obtained from the numerical example:

1. If the aluminum alloy material is not be inspected, the theoretical service life of the material will be about 0.10837 million cycles when R_S = 0.9. However, if the sequential inspection method is used, the effective service life of the material is about 0.132 million cycles, which is close to the theoretical failure time of 0.13343 million cycles, the service life of aviation products is extended effectively.
2. As can be seen from Table 1, the sequential inspection interval is not equal spacing, the number of inspection is reduced from 6 times to 4 times after one maintenance, and reduced to 3 times after two maintenances. As the data accumulated, the number of inspection can be further reduced. Therefore, the sequential inspection method proposed in this paper, not only can effectively reduce the number of inspection, but also can improve the efficiency of inspection and maintenance while ensuring the reliability of aviation products.
3. It can be seen that the convergence rate of model parameters is very fast in the case of small sample data, and the prediction accuracy also meets the actual requirements.
4. If $R_S = 0.8$, the theoretical service life of the material is about 0.11482 million cycles. When the rest conditions remain unchanged, using the similar calculation procedure above, we can obtain the following inspection times: The initial new product {0.0463, 0.0926, 0.1230, 0.1304, 0.1322}; After one maintenance {0.2401, 0.2605, 0.2656}; After two maintenances {0.3770, 0.3966}. According to the calculation results, the effective service life

of the material is about 0.133 million cycles. Moreover, the number of inspections required for the new product is reduced from 6 to 5 times, and reduced from 4 to 3 times for the product after one maintenance. This shows that sequential inspection time, the number of inspections and effective service life are closely related with the requirement of real-time reliability.

6 CONCLUSIONS

In this paper, a sequential inspection method for aviation products based on Wiener process and real-time reliability requirement has been proposed. It is known that in the actual operation process of aviation products, the mechanism of performance degradation is complex, and the degradation process shows randomness. The sequential inspection method can adaptively update the relevant parameters of degradation model in real time, so as to determine the time of each inspection and maintenance more accurately. The method can not only ensure the reliability of the product, but also avoid the problems of over-inspection or under-inspection which may be caused by the traditional equal-pitch cycle inspection. And the number of inspection data is also required less by the sequential inspection method. Moreover, the convergence rate of relevant parameters become faster by integrating the historical information of similar products, and the model has good robustness, so as to ensure the reliability requirement and the prediction accuracy of degradation information for the aviation products, which is very suitable for the determination of inspection and maintenance intervals of the new small-sample product.

The research of this paper is carried out under the condition of considering the perfect maintenance for the product. In some cases, the performance of the product as a whole is affected by other non-replaceable parts and some environmental factors of the system, the performance cannot recover as new after maintenance. Therefore, the further study will consider the inspection method for the product in the case of imperfect maintenance.

ACKNOWLEDGEMENT

This research is supported by the project of Natural Science Foundation of China (with granted number 61573370).

REFERENCES

Cui, L.R., Loh, H.T., & Xie, M. (2004). Sequential inspection strategy for multiple systems under availability requirement. European Journal of Operational Research, 155(1), 170–177.

Faghih-Roohi, S., Xie, M., Ng, K.M., & Yam, R.C.M. (2014). Dynamic availability assessment and optimal component design of multi-state weighted k-out-of-n systems. Reliability Engineering & System Safety, 123(4), 57–62.

Feng, J. 2011. Sequential inspection method for long-storage-products based on storage degradation mechanism. Journal of Aerospace Power, 26(3), 611–616.

Gebraeel N, Lawley MA, Li R, et al. (2005). Residual-life distributions from component degradation signals: a bayesian approach. Iie Transactions, 37(6), 543–557.

Guo, C., Wang, W., Guo, B., & Si, X. (2013). A maintenance optimization model for mission-oriented systems based on wiener degradation. Reliability Engineering & System Safety, 111(3), 183–194.

Jiang, R. (2010). Optimization of alarm threshold and sequential inspection scheme. Reliability Engineering & System Safety, 95(3), 208–215.

Lee, M.L.T., & Whitmore, G.A. (2007). Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary. Statistical Science, 21(4), 501–513.

Meeker, W.Q., & Escobar, L. 1998. Statistical methods for reliability data /. Wiley.

Peng, C.Y., & Tseng, S.T. 2009. Mis-specification analysis of linear degradation models. IEEE Transactions on Reliability, 58(3), 444–455.

Ray, A., & Tangirala, S. (2002). Stochastic modeling of fatigue crack dynamics for on-line failure prognostics. IEEE Transactions on Control Systems Technology, 4(4), 443–451.

Si, X.S., Hu, C.H. 2016. Data-driven remaining useful life prediction theory and applications for equipment. Beijing: National Defense Industry Press.

Tanrioven, M., Wu, Q.H., Turner, D.R., Kocatepe, C., & Wang, J. (2004). A new approach to real-time reliability analysis of transmission system using fuzzy markov model. International Journal of Electrical Power & Energy Systems, 26(10), 821–832.

Wang, X., Jiang, P., Guo, B., & Cheng, Z. (2014). Real-time reliability evaluation with a general wiener process-based degradation model. Quality & Reliability Engineering International, 30(2), 205–220.

Xu, Z., Ji, Y., & Zhou, D. (2008). Real-time reliability prediction for a dynamic system based on the hidden degradation process identification. IEEE Transactions on Reliability, 57(2), 230–242.

Yan, W.A., Song, B.W., Duan, G.L., & Shi, Y.M. (2016). Real-time reliability evaluation of two-phase wiener degradation process. Communications in Statistics, 46(1), 176–188.

Ye, Z.S., Chen, N., & Shen, Y. (2015). A new class of wiener process models for degradation analysis. Reliability Engineering & System Safety, 139, 58–67.

Zhang, J., Huang, X., Fang, Y., Zhou, J., Zhang, H., & Li, J. (2016). Optimal inspection-based preventive maintenance policy for three-state mechanical components under competing failure modes. Reliability Engineering & System Safety, 152, 95–103.

Zhu, Z., Xiang, Y., Alaswad, S., & Cassady, C.R. (2017). A sequential inspection and replacement policy for degradation-based systems. Reliability and Maintainability Symposium. IEEE.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Optimal burn-in for repairable products sold with two-dimensional warranty considering preventive maintenance

X.P. Li & Z.X. Liu
Tianjin University, Tianjin, China

Y.K. Wang
Tianjin Chengjian University, Tianjin, China

Y.L. Liu
Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: For repairable products sold with two-dimensional warranty, although burn-in and Preventive Maintenance (PM) actions result in extra costs, they both can be effective approaches to reduce number of warranty claims and servicing cost by early detecting defects and reliability improvement. Longer burn-in or higher burn-in usage rate can remove more defects, but it accelerates the product degradation and incurs more wear-out failures, which in turn need to be alleviated by PMs in the warranty period. This article aims to propose a new warranty performance-based model to minimize the warranty claims and find out the optimal burn-in and PM decisions for two-dimensional warranted products. More specifically, the proposed model subsumes a special case under one-dimensional warranty, allows different failure modes—i.e. defects and wear-out failures, and takes usage heterogeneity into consideration. We find that, it is always reasonable to carry out a burn-in on products. If wear-out failures are more (less) sensitive to product usage rate than defects, the burn-in duration should be extended (increased with the upper limit of PM degree); The optimal burn-in usage rate should be increased with the upper limit of PM degree (conducted at conditions as harsh as possible) and decrease with the upper limit of detecting degree; the optimal PM should always be set at its upper limit.

1 INTRODUCTION

Burn-in as an important production process, where products are operated under the actual working stress for a period. Those weak items and premature failures (e.g. from manufacturing and assembly errors) can be revealed, and then the general product reliability is improved (Chan and Meeker, 1999; Blischke et al, 2011).

Products are sold with warranties. Currently, most of burn-in models focus on the cases of one-dimensional warranties. Leemis and Beneke (1990) have highlighted in their review paper that some researchers had investigated relations between burn-in and warranty policies. According to Nguyen and Murthy (1982), the optimal burn-in duration is determined by the trade-off between the reduction in the warranty cost and the increase in the cost with burn-in. Sheu and Chien (2005) extended the above models by studying two types of failures. Similarly, Kar and Nachlas (1997), and Wu et al. (2007) also have proposed models to determine the optimal burn-in duration for products sold with warranties. They considered different types of warranty policies.

After burn-in, preventive maintenance (PM) is an effective way to reduce warranty claims. Unlike many warranty policies which require restoring a failed product without any schedules, PM is commonly scheduled, aiming to control the wear-out degradation and reduce likelihood of failures. Shafiee et al. (2011, 2013) have paid attention to the effects of PMs on the failure rate and the optimal burn-in duration.

A warranty can be one-dimensional, meaning that only calendar time is considered in determining the warranty period, or two-dimensional, where usages are often evaluated. However, all the above burn-in models are limited to one-dimensional warranted products. Ye et al. (2013) contribute to deal with burn-in under two-dimensional warranties, but this work does not consider PM in the burn-in decisions.

Therefore, this paper focuses on the burn-in and PM decisions for two-dimensional warranted products. Firstly, since the optimal burn-in

decisions should be made by a trade-off between the increase in wear-out failures and the decrease in defects during warranty, we were wondering whether a burn-in should be applied to two-dimensional warranted products. If the answer is yes, the next question is how to set the burn-in duration and burn-in usage rate. Furthermore, we will explore how PM influence the burn-in decision.

The rest of this paper is organized as follows. The modeling assumptions are listed in Section 2. In Section 3, we investigate how to model failures under PM actions and introduce the warranty performance-based model to minimize the warranty claims and help make optimal decisions. Then numerical examples are presented to illustrate the applicability of the proposed model in Section 4. Conclusions and some extensions are given in Section 5.

2 MODEL ASSUMPTIONS

Firstly, all the item failures during the warranty coverage are statistically independent and minimally repaired. Each item is assumed to be repairable and sold with a non-renewing free repair warranty policy. In addition, each failure will incur a warranty claim, and all warranty claims are valid.

We assume that there are two types of failures, the wear-out failures and the defects. The wear-out failures, also known as normal failures, occur over time, while the defects can be revealed in the early period of lifetime. The two failures are independent.

Effects of usage rate on both types of failures are modeled using Accelerated Failure Time (AFT) model (Baik and Murthy 2008; Shahanaghi et al. 2013). We also assume that the failure rate functions are convex, in order to ensure that the failure rate curve of products is close to the bathtub curve as much as possible.

After burn-in, the product is released to the market with a non-renewing free repair warranty (FRW).

Like most of related papers (Iskandar et al. 2005; Wang et al. 2015), we confine the two-dimensional warranty region to be a rectangle one in which the horizontal axis represents age and the vertical axis represents usage. The warranty region is a rectangle $[0, W_0) \cdot [0, U_0)$. The warranty expires when the item reaches either the age or usage limit.

The time of manufacturer implementing repair or PMs programs is negligible and assumed to be zero. And all the PMs are assumed to be performed with fixed time interval τ. In addition, defects can be detected during PMs.

3 MODELING AND ANALYSIS

In this section, we firstly investigate how to model wear-out failures and defects. Then performance-based model is introduced to minimize the warranty claims and help make optimal decisions.

3.1 *Modeling of failures*

The usage rate over time is constant for one customer, but varies across customers. Let R be the random usage rate, and $G(r)$ and $g(r)$ represent cumulative distribution function (CDF) and probability density function (PDF) of R respectively.

Under a two-dimensional warranty, we model the wear-out with a counting process characterized by an intensity function dependent on both age and usage. We introduce $v^w(t|r)$ as the corresponding virtual age under the condition of usage rate r. Considering the item designed for certain nominal usage rate r_0, according to the AFT model, if the usage time is t and the specific usage rate is r, then $v^w(t|r) = t\left(\frac{r}{r_0}\right)^{\gamma}$, where γ is accelerated coefficient due to the wear-out.

Conditional on a specific customer $R = r$, the wear-out process is a Non-homogeneous Poisson process (NHPP) with a non-decreasing rate of occurrence of failure $\lambda^w(v^w)$. Note that v^w is the function of t, so we can achieve the actual failure rate function of an item due to wear-out failures for any customer with usage rate r as $\lambda^w\left(v^w(t|r)\right)\left(\frac{r}{r_0}\right)^{\gamma}$ according to the AFT model.

The number of defects is modeled by a random variable M with a probability mass function $\pi(\cdot)$. All defects are assumed to be independent and identically distributed. For any specific defect, the virtual age is $v^i(t|r) = t\left(\frac{r}{r_0}\right)^{\eta}$, where η is accelerated coefficients due to the defects. And its first time to failure conforms to a CDF $F^d(\cdot)$. Therefore, the number of defects can be described by a binomial process with M trials, each with success probability given by $F^d(\cdot)$.

3.2 *Failures within burn-in*

It is limited to set the values of burn-in durations and usage rates, since longer duration and higher rates can decrease defect on the one hand, but intensify wear-out on the other hand. We use $\overline{t_b}$ and $\underline{t_b}$ as the upper bound and lower bound of Burn-in duration t_b, and $\overline{r_b}$ and $\underline{r_b}$ as upper bound and lower bound of burn-in usage rate r_b.

For an item arriving at the end of burn-in, we let v_0^w and v_0^i be the virtual age of wear-out failure process and potential defects, respectively. The time lag between the end of burn-in and the beginning

of product's running into operation is ignored. Therefore, after burn-in under the usage rate r_b and duration t_b, we can obtain that $v_0^w = t_b\left(\frac{r_b}{r_0}\right)^\gamma$ and $v_0^i = t_b\left(\frac{r_b}{r_0}\right)^\eta$.

The expected number of failures of wear-out failures within the burn-in can be given by,

$$E^w\left[N_b\left(t_b, r_b\right)\right] = \int_0^{t_b}\left[\lambda^w\left(v^w\left(t|r_b\right)\right)\left(\frac{r_b}{r_0}\right)^\gamma\right]dt \qquad (1)$$

For any specific defect, the probability that it is detected within burn-in is $F^d\left(t\left(\frac{r}{r_0}\right)^\eta\right)$. Conditional on M, the number of defects occurs during burn-in follows a binomial distribution with mean value as follow,

$$E^i\left[N_b\left(t_b, r_b\right)\right] = E\left(M\right)\times F^d\left(t_b\left(\frac{r_b}{r_0}\right)^\eta\right)$$

Therefore, the expected number of failures within burn-in can be given by,

$$E\left[N_b\left(t_b, r_b\right)\right] = E^w\left[N_b\left(t_b, r_b\right)\right] + E^i\left[N_b\left(t_b, r_b\right)\right]$$

3.3 *Failures within warranty period under PM*

Different usage rates mean different usage patterns which lead to different termination of warranties due to its age limits or usage limits. For any specific usage rate, only two cases exist, which is illustrated in Fig. 1.

From the Fig. 1, we know that, for a specific usage rate r, define r_0 as $\frac{U_0}{W_0}$, if $r > r_0$, then warranty will end at the time $\frac{U_0}{r}$. If $r \le r_0$, then warranty will end at the time W_0. Therefore, we need to investigate these two cases.

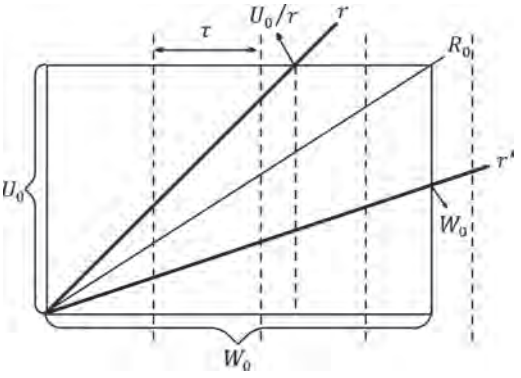


Figure 1. Warranty period under PM.

Within the warranty coverage, though minimal repair is free of charge, the customer has already experienced loss incurred by failures. Besides, the wear-out failure rate of product becomes higher, so that the manufacturer tends to suffer more repair cost. Therefore, PMs are needed, not only to improve product reliability, and reduce repair costs.

We let periodical imperfect PMs are implemented with fixed interval τ, and their effectiveness is depicted by a maintenance degree δ_w for wear-out failures and a detect degree δ_d for potential defects, both of which are between 0 and 1. $\delta_w = 1$ means that wear-out failure rate of the product is back to original level by a perfect PM and then the product is "as good as new", while $\delta_w = 0$ means that no action is carried out. Also, $\delta_d = 1$ means that any defects can be found during PM with probability of 1, while $\delta_d = 0$ means that PM actions cannot find any defects.

3.3.1 *Wear-out failures*

For wear-out failures, there are at least two classes of imperfect PM models in previous literatures. The readers can refer to Doyen and Gaudoin (2004) as a good source of reference. We use the one that PM actions result in a rejuvenation of the product through effectively reducing the virtual age (Doyen and Gaudoin, 2004). We let $\lambda_k^W\left(t|r\right)$ be conditional wear-out failure rate of the item after the kth PM action, and v_k^w be the virtual age of wear-out failure process after the kth PM action. And we already know that $v_0^w = t_b\left(\frac{r_b}{r_0}\right)^\gamma$, then the virtual age of wear-out failure process after the 1st PM action is

$$v_1^w = \left[v_0^w + \tau\cdot\left(\frac{r}{r_0}\right)^\gamma\right]\left(1-\delta_w\right)$$

where r is specific usage rate of a certain customer. Then we have,

$$v_k^w = \left[v_{k-1}^w + \tau\cdot\left(\frac{r}{r_0}\right)^\gamma\right]\left(1-\delta_w\right)$$

When $\delta_w = 1$, v_k^w is always returned to 0. When $\delta_w \ne 1$, divided by $\left(1-\delta_w\right)^k$ on both sides of the equation, so $\frac{v_k^w}{\left(1-\delta_w\right)^k} = \frac{v_{k-1}^w}{\left(1-\delta_w\right)^{k-1}} + \tau\cdot\left(\frac{r}{r_0}\right)^\gamma\left(1-\delta_w\right)^{1-k}$.

Let $S_k = \frac{v_k^w}{\left(1-\delta_w\right)^k}$, then we have $S_k = S_{k-1} + \tau\cdot\left(\frac{r}{r_0}\right)^\gamma\left(1-\delta_w\right)^{1-k}$, add up these equations from S_1 to S_k, so

$$S_k = v_0^w + \tau \cdot \left(\frac{r}{r_0}\right)^{\gamma} \left(1-\delta_w\right) \left(\frac{1-\left(1-\delta_w\right)^k}{\delta_w \left(1-\delta_w\right)^k}\right)$$

Using $S_k = \frac{v_k^w}{(1-\delta)^k}$ to replace it, then we have the general term of v_k^w as follow,

$$v_k^w = \left(1-\delta_w\right)^k v_0^w + \tau \cdot \left(\frac{1-\delta_w}{\delta_w}\right)\left(1-\left(1-\delta_w\right)^k\right)\left(\frac{r}{r_0}\right)^{\gamma}$$

Then we have the wear-out failure rate $\lambda_k^W(t|r)$ of the item after the kth PM action as follow,

$$\lambda_k^W(t|r) = \lambda^w\left(v_k^w + t \cdot \left(\frac{r}{r_0}\right)^{\gamma}\right)\left(\frac{r}{r_0}\right)^{\gamma}$$

Note that v_k^w is still the function of t. Specifically, when $\delta_w = 0$, $\lambda_k^W(t|r)$ is given by,

$$\lambda_k^w(t|r) = \lambda^w\left(v_0^w + k\tau\left(\frac{r}{r_0}\right)^{\gamma} + t\left(\frac{r}{r_0}\right)^{\gamma}\right)\left(\frac{r}{r_0}\right)^{\gamma} \tag{2}$$

For case $r > r_0$, the warranty for the product ceases when the product usage reaches U_0, which corresponds to the age $\frac{U_0}{r}$. And we let the number of PMs within this warranty region be $n_{r>r_0}$, estimated by $\max_j\left\{j \times \tau \le \frac{U_0}{r}\right\}$, where $j \in N$. Therefore, we have the number of PMs performed within warranty when $r > r_0$, namely $n_{r>r_0} = arg\max_j\left\{j \times \tau \le \frac{U_0}{r}\right\}$. Then the expected failures of wear-out failures during this period are given by,

$$E\left[N^w\left(t_b, r_b | r > r_0\right)\right] = \sum_{k=0}^{n_{r>r_0}-1} \int_0^{\tau} \lambda_k^W\left(t|r\right) dt + \int_0^{\frac{U_0}{r}-\tau n_{r>r_0}} \lambda_{n_{r>r_0}}^W\left(t|r\right) dt \tag{3}$$

Here the first term is expected wear-out failures before the last PM within the warranty. The second term is expected wear-out failures between the last PM and the end of warranty. Since the meanings of terms hereafter are similar, we will not repeat.

For the case $r \le r_0$, the warranty for the product terminates when the product age reaches W_0. Similarly, we let the number of PMs within this warranty region be $n_{r\le r_0}$, estimated by $\max\left\{j \times \tau \le W_0\right\}$, where $j \in N$. Therefore, we have the number of PM actions performed within warranty when $r \le r_0$, namely $n_{r\le r_0} = arg\max\left\{j \times \tau \le W_0\right\}$. And then the expected failures during this period are given by,

$$E\left[N^w\left(t_b, r_b | r \le r_0\right)\right] = \sum_{k=0}^{n_{r\le r_0}-1} \int_0^{\tau} \lambda_k^W\left(t|r\right) dt + \int_0^{W_0-\tau n_{r\le r_0}} \lambda_{n_{r\le r_0}}^W\left(t|r\right) dt \tag{4}$$

We use $E\left[N^w\left(t_b, r_b|r\right)\right]$ to denote the expected wear-out failures within warranty period of any specific customer, which means $E\left[N^w\left(t_b, r_b|r\right)\right]$ should be either $E\left[N^w\left(t_b, r_b|r > r_0\right)\right]$ or $E\left[N^w\left(t_b, r_b|r \le r_0\right)\right]$ according to the usage rate r.

Also, we use $E\left(n_{BP}|r\right)$ to denote the number of PMs within this warranty region, which means $E\left(n_{BP}|r\right)$ should be either $n_{r>r_0}$ or $n_{r\le r_0}$ according to the usage rate r. These declarations will be useful for our subsequent modelling analysis.

3.3.2 *Defects*

Consider a specific remaining defect, the virtual age of this defect after burn-in is $t_b\left(\frac{r_b}{r_0}\right)^{\eta}$. After the burn-in, the remaining defects can either be found during operation or detected by PM actions. The distribution for the time to find this remaining defect conditional on specific r before 1st PM action can be given by

$$G_0^d(t|r) = \frac{F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta} + \tau\left(\frac{r}{r_0}\right)^{\eta}\right) - F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}\right)}{1 - F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}\right)} \tag{5}$$

Suppose that a total of K defects remains. Given M, K follows a binomial distribution $bi\left(M, 1 - F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}\right)\right)$.

Conditional on K, the number of remaining defects before the 1st PM actions follows the binomial distribution $bi\left(K, G_0^d(t|r)\right)$. Then we can deduce the probability of remaining defects K that can be found from the end of burn-in to 1st PM action as follow,

$$F_0^d(t|r) = \left(1 - F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}\right)\right) G_0^d(t|r) = F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta} + \tau\left(\frac{r}{r_0}\right)^{\eta}\right) - F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}\right) \tag{6}$$

A defect may be detected during the 1st PM. Using the similar deduction of $F_0^d(t|r)$, we can achieve that the probability of remaining defects K that can be detected in the 1st PM with detect degree δ_d is given by,

$$P_1^d(\tau|r)=\left(1-F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}\right)\right)\left(1-G_0^d(\tau|r)\right)\delta_d$$
$$=\left[1-F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}+\tau\left(\frac{r}{r_0}\right)^{\eta}\right)\right]\delta_d \quad (7)$$

The rest as follows can be done as the same manner. Then the probability of remaining defects K that can be detected between kst and (k + 1)st PM action, and during the kst PM action respectively can be given by,

$$F_k^d(\tau|r)=\left[F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}+(k+1)\tau\left(\frac{r}{r_0}\right)^{\eta}\right)-F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}+k\tau\left(\frac{r}{r_0}\right)^{\eta}\right)\right]\left(1-\delta_d\right)^k \quad (8)$$

$$P_k^d(\tau|r)=\left[1-F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}+k\tau\left(\frac{r}{r_0}\right)^{\eta}\right)\right]\delta_d\left(1-\delta_d\right)^{k-1} \quad (9)$$

Therefore, considering the cases of warranty region in Fig. 1, taking expectation with respect to remaining defects K, we can find that the expected number of remaining defects found during the operation and during PMs,

$$E\left[N_{op}^d(t_b,r_b|r>r_0)\right]=\sum_{k=0}^{n_{r>r_0}-1}\left[E(M)F_k^d(\tau|r)\right]+E(M)\left[F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}+W_0\left(\frac{r}{r_0}\right)^{\eta}\right)-F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}+n_{r\le r_0}\cdot\tau\left(\frac{r}{r_0}\right)^{\eta}\right)\right]\left(1-\delta_d\right)^{n_{r\le r_0}} \quad (10)$$

$$E\left[N_{op}^d(t_b,r_b|r\le r_0)\right]=\sum_{k=0}^{n_{r\le r_0}-1}\left[E(M)F_k^d(\tau|r)\right]+E(M)\left[F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}+W_0\left(\frac{r}{r_0}\right)^{\eta}\right)-F^d\left(t_b\left(\frac{r_b}{r_0}\right)^{\eta}+n_{r\le r_0}\cdot\tau\left(\frac{r}{r_0}\right)^{\eta}\right)\right]\left(1-\delta_d\right)^{n_{r\le r_0}} \quad (11)$$

$$E\left[N_{pm}^d(t_b,r_b|r>r_0)\right]=\sum_{k=1}^{n_{r>r_0}}\left[E(M)P_k^d(\tau|r)\right] \quad (12)$$

$$E\left[N_{pm}^d(t_b,r_b|r\le r_0)\right]=\sum_{k=1}^{n_{r\le r_0}}\left[E(M)P_k^d(\tau|r)\right] \quad (13)$$

We use $E\left[N_{op}^d(t_b,r_b|r)\right]$ to denote the expected number of defects within warranty period of any specific customer, which means $E\left[N_{op}^d(t_b,r_b|r)\right]$ should be either $E\left[N_{op}^d(t_b,r_b|r>r_0)\right]$ or $E\left[N_{op}^d(t_b,r_b|r\le r_0)\right]$ according to the usage rate r. Also, we use $E\left[N_{pm}^d(t_b,r_b|r)\right]$ to denote the expected number of detects found during PM actions, which means $E\left[N_{pm}^d(t_b,r_b|r)\right]$ should be either $E\left[N_{pm}^d(t_b,r_b|r>r_0)\right]$ or $E\left[N_{pm}^d(t_b,r_b|r\le r_0)\right]$ according to the usage rate r.

3.4 *Performance-based model*

The warranty performance-based model involves simultaneously selecting burn-in duration t_b and usage rate r_b to minimize the expected failures during the warranty period. For a customer with specific usage rate r, we have the expected warranty claims per unit as follows,

$$E\left[N(t_b,r_b|r)\right]=E\left[N^w(t_b,r_b|r)\right]+E\left[N_{op}^d(t_b,r_b|r)\right] \quad (14)$$

As mentioned earlier, we consider that the usage rate is constant for one customer but varies across different customers. Since we let R be its random usage rate, and $G(r)$ and $g(r)$ represent CDF and PDF of R respectively, the final expected number of failures is obtained by un-conditioning, namely by taking the expectation of $E\left[N_W(t_b,r_b|r)\right]$ with respect to the usage rate,

$$E\left[N(t_b,r_b)\right]=\int_0^{\infty}E\left[N(t_b,r_b|r)\right]dG(r) \quad (15)$$

Then the performance-based model can be given by,

$$[t_b^*,r_b^*]=argmin\ E\left[N(t_b,r_b)\right]$$
$$s.t.\quad t_b\in\left[\underline{t_b},\overline{t_b}\right]$$
$$r_b\in\left[\underline{r_b},\overline{r_b}\right]$$

where t_b^* and r_b^* are optimal burn-in duration and burn-in usage rate.

In this performance-based model, obviously the larger the maintenance degree δ_w and the detect degree δ_d are, the smaller the expected number of failures within warranty period is, which means if they were decision variables, they would always be set at the max. In practice, some PM actions are paid by customers, which means manufactures can ignore the maintenance cost and only take into consideration their own possible limitation to implement the PM actions. Therefore, we don't make the mainte-

nance degree δ_w and the detect degree δ_d as the decision variable in this performance-based model.

Proposition 1 Suppose that $[\, t_b^*, r_b^* \,]$ is the optimal burn-in setting for the performance-based model. When the accelerated coefficients of wear-out failures and defects follow the relationship $\gamma \geq \eta$, (i) the optimal burn-in duration t_b^* is always t_b, (ii) the optimal burn-in usage rate r_b^* is increasing with the maintenance degree δ_w, and (iii) decreasing with the detect degree δ_d.

Proof of Proposition 1 Suppose that $[\, t_b^*, r_b^* \,]$ is the optimal burn-in setting for the performance-based model. We use the contradiction to prove $t_b^* = \bar{t}_b$, if $\gamma > \eta$.

Suppose $t_b^* < \bar{t}_b$, because the virtual age $t_b \left(\frac{r_b}{r_0}\right)^{\eta}$ is increasing with t_b and r_b, and also decreasing with t_b and r_b, we can always find r_b' with $r_b' < r_b^*$ such that,

$$t_b^* \left(\frac{r_b^*}{r_0}\right)^{\eta} = \bar{t}_b \left(\frac{r_b'}{r_0}\right)^{\eta} \tag{16}$$

It means that the virtual age of defects is the same under the cases of $[\, t_b^*, r_b^* \,]$ and $[\, \bar{t}_b, r_b' \,]$ Therefore, as $E\left[N_{op}^d(t_b, r_b)\right]$ is the expected number of defects within warranty coverage, we obviously have the same expected numbers under these two cases, namely,

$$E\left[\, N_{op}^d(t_b^*, r_b^*) \,\right] = E\left[\, N_{op}^d(\bar{t}_b, r_b') \,\right]$$

Based on (16), we have,

$$\bar{t}_b \left(\frac{r_b'}{r_0}\right)^{\gamma} = t_b^* \left(\frac{r_b^*}{r_0}\right)^{\eta} \left(\frac{r_b'}{r_0}\right)^{-\eta} \left(\frac{r_b'}{r_0}\right)^{\gamma}$$
$$= t_b^* \left(\frac{r_b^*}{r_0}\right)^{\gamma} \left(\frac{r_b'}{r_b^*}\right)^{\gamma-\eta}$$

Because $\gamma > \eta$ and $r_b' < r_b^* \left(\frac{r_b'}{r_b^*}\right)^{\gamma-\eta} < 1$, then we have,

$$\bar{t}_b \left(\frac{r_b'}{r_0}\right)^{\gamma} < t_b^* \left(\frac{r_b^*}{r_0}\right)^{\gamma}$$

Therefore, the virtual age of wear-out failures of case $[\, t_b^*, r_b^* \,]$ is bigger than that of case $[\bar{t}_b, r_b']$. As $E\left[N^w(t_b, r_b)\right]$ is the expected number of wear-out failures during operations. Since the failure rate is increasing, we have,

$$E\left[N^w(\bar{t}_b, r_b')\right] < E\left[N^w(\, t_b^*, r_b^*)\right]$$

Because $E\left[\, N(t_b, r_b) \,\right] = E\left[\, N^w(t_b, r_b) \,\right] + E\left[\, N_{op}^d(t_b, r_b) \,\right], E\left[\, N(\bar{t}_b, r_b') \,\right] < E\left[\, N(\, t_b^*, r_b^*) \,\right]$

Therefore, the case $[\bar{t}_b, r_b']$ is better than $[\, t_b^*, r_b^* \,]$, which obviously contradicts our assumption. Then we have $t_b^* = \bar{t}_b$. This completes the proof of (i).

Now we set t_b as a constant $\bar{t}_b$, and give the proof of (ii).

When $r_b = r_b^*$, based on principle of optimization, we obviously have,

$$\left.\frac{dE\left[N^w(\bar{t}_b, r_b)\right]}{dr_b}\right|_{r_b = r_b^*} = -\left.\frac{dE\left[N_{op}^d(\bar{t}_b, r_b)\right]}{dr_b}\right|_{r_b = r_b^*}$$

Since δ exists in $\frac{dE\left[N^w(\bar{t}_b, r_b)\right]}{dr_b}$, we let $Z^w(\delta, r_b) = \frac{dE\left[N^w(\bar{t}_b, r_b)\right]}{dr_b}$ and $Z^d(\delta, r_b) = -\frac{dE\left[N_{op}^d(\bar{t}_b, r_b)\right]}{dr_b}$.

Suppose that when $\delta = \delta'$, optimal burn-in usage rate $r_b^* = r_b'$, We use the contradiction to prove (ii). When $\delta' < \delta''$, suppose $r_b' \geq r_b''$.

Lemma 1 If $f(x, y)$ is convex in x for each $y \in A$, and $\omega(y) \geq 0$ for each $y \in A$, then the function ϕ defined as,

$$\phi(x) = \int_A \omega(y) f(x, y)\, dy$$

is convex in x. (Boyd and Vandenberghe 2004).

Based on (3), we can find $\int_0^{\frac{U_0}{r} - \tau \cdot n_{r>\eta_0}} \lambda_{n_{r>\eta_0}}^W(t \mid r)\, dt$ conforms to Lemma 1 since $\lambda_{n_{r>\eta_0}}^W(t|r)$ is a convex function which is our assumption. Therefore, when we set $t_b = \bar{t}_b$ as a constant, we can easily achieve that $\int_0^{\frac{U_0}{r} - \tau \cdot n_{r>\eta_0}} \lambda_{n_{r>\eta_0}}^W(t|r)\, dt$ is convex function in r_b. Then based on (14) and (15), we can find $\int g(r) \left(\int_0^{\frac{U_0}{r} - \tau \cdot n_{r>\eta_0}} \lambda_{n_{r>\eta_0}}^W(t|r)\, dt \right) dr$ also conforms to Lemma 1 since $g(r) \geq 0$ and $\int_0^{\frac{U_0}{r} - \tau \cdot n_{r>\eta_0}} \lambda_{n_{r>\eta_0}}^W(t \mid r)\, dt$ is convex function in r_b. Therefore, we can easily achieve that the whole term is convex in r_b. The other terms can all be deduced in this way.

Therefore, we can find that $\frac{\partial Z^w(\delta, r_b)}{\partial r_b} = \frac{\partial^2 E\left[N^w(\bar{t}_b, r_b)\right]}{\partial r_b^2} \geq 0$. Also, $\frac{\partial Z^d(\delta, r_b)}{\partial r_b} < 0$. It means for specific δ'', $Z^w(\delta'', r_b'') \leq Z^w(\delta'', r_b')$ and $Z^i(\delta'', r_b'') \geq Z^i(\delta'', r_b')$, if $r_b' \geq r_b''$.

Note that δ also exists in $Z^d(\delta, r_b)$, we can easily achieve $\frac{\partial Z^d(\delta, r_b)}{\partial \delta} > 0$, then we have $Z^d(\delta', r_b') < Z^d(\delta'', r_b')$, if $\delta' < \delta''$. In addition, for the expected numbers of wear-out failures $E\left[N^w(t_b, r_b)\right]$ within warranty coverage, when we set $t_b = \bar{t}_b$ as a constant, obviously for achieving the same expected numbers of wear-out failures within warranty coverage, the deeper the mainte-

nance degree δ is, the larger the burn-in usage rate is, which conforms to the property of strictly sub-modular function. Therefore, we can easily find that $\frac{\partial Z^w(\delta,r_b)}{\partial \delta} = \frac{\partial^2 E\left[N^w(t_b,r_b)\right]}{\partial r_b \partial \delta} < 0$. The rigorous mathematical deduction of this term is very similar with the proof of (i), so we omit it.

Then we can find that for specific r_b', $Z^w(\delta',r_b') > Z^w(\delta'',r_b')$ and $Z^d(\delta',r_b') < Z^d(\delta'',r_b')$, if $\delta' < \delta''$. Since r_b' is the optimal value, we have $Z^d(\delta',r_b') = Z^w(\delta',r_b')$, then,

$$Z^d(\delta'',r_b') > Z^d(\delta',r_b') = Z^w(\delta',r_b') > Z^w(\delta'',r_b')$$

Therefore, based on $Z^w(\delta'',r_b'') \le Z^w(\delta'',r_b')$ and $Z^d(\delta'',r_b'') \ge Z^d(\delta'',r_b')$, we have,

$$\begin{aligned} &Z^d(\delta'',r_b'') \ge Z^d(\delta'',r_b') > Z^d(\delta',r_b') \\ &= Z^w(\delta',r_b') > Z^w(\delta'',r_b') \ge Z^w(\delta'',r_b') \end{aligned}$$

It means $Z^d(\delta'',r_b'')$ is strictly bigger than $Z^w(\delta'',r_b'')$, which obviously contradicts the principle that $Z^d(\delta'',r_b'')$ should be equal to $Z^w(\delta'',r_b'')$, if r_b'' is the optimal value. Therefore, r_b'' cannot be the optimal value of r_b, if $r_b' \ge r_b''$. Then we have the conclusion of (ii).

The proof of (iii) can be deduced in the way of proof of (ii).

These complete the proof of Proposition 1.

Proposition 2 Suppose that $[t_b^*, r_b^*]$ is the optimal burn-in setting for the performance-based model. When the accelerated coefficients of wear-out failures and defects follow the relationship $\gamma < \eta$, then(i) the optimal burn-in usage rate r_b^* is always $\underline{r_b}$, (ii) the optimal burn-in duration t_b^* is increasing with the maintenance degree δ_w, and (iii) decreasing with the detect degree δ_d.

Similar with the proof of proposition 1, proposition 2 can be proved. These propositions indicate that if for a certain product of which wear-out failures are more sensitive than defects to the usage rate, the burn-in duration should be conducted at as long as possible. From the proof of proposition, we know that this is because if the burn-in duration increases, for achieving the same virtual age of the failures result from defects under the two burn-in duration settings, the virtual age of the wear-out failures decreases which leads to less failures during warranty. When the burn-in duration is set to be a constant at its upper limit, that the optimal burn-in usage rate is increasing with the maintenance degree, meaning that if manufacturers implement deeper maintenance, they should set burn-in conditions harsher than before. Meanwhile, if they are able to detect more possible defects during PMs, they should set burn-in conditions milder. Conversely, if defects are more sensitive than wear-out failures to the usage rate, the burn-in should be conducted at conditions as harsh as possible, and the optimal burn-in duration increases with the maintenance degree and decreases with the defecting degree. These two converse propositions cover all the possible cases, so that manufacturers can easily select the burn-in decisions according to the features of any specific product or component. In addition, based on these propositions, the proposition 3 below can be easily set up:

Proposition 3 In performance-based model, implementing burn-in is always better than nothing.

4 NUMERICAL EXAMPLES

In this section, numerical examples are presented to illustrate the applicability of the proposed model above.

For the performance-based model, we use an example to show how to reduce the warranty claims by choosing the optimal burn-in duration t_b and usage rate r_b.

Suppose that the product under consideration is certain automobile component covered by a two-dimensional FRW. The following settings are adopted into our performance-based model,

We calculate the expected failures within the warranty region which is a rectangle of 3 years and 100,000 km of usage. Therefore, we let $W_0 = 3, U_0 = 10$.

We assume that the nominal usage rate is $r_0 = 2 \cdot 10^4$ km/y. And the customer usage rate follows uniform (0.4, 4.2), which means the least usage rate is $0.4 \cdot 10^4$ km/y, and the highest usage rate is $4 \cdot 10^4$ km/y.

Wear-out failures follow a Weibull distribution with increasing failure rate, which is $\lambda^w(t) = \frac{\beta_w}{\alpha_w}\left(\frac{t}{\alpha_w}\right)^{\beta_w - 1}$, where $\beta_w > 2$ (as we assume the failure rate functions of two type failures are convex). The number of defects M follows $Poisson(\mu)$. Defect failure them follows $exp(\theta)$.

Other parameters for this model are given in Table 1.

Under different maintenance degrees and detect degrees of PM actions, we perform a grid search for the minimization of numerical examples based on performance model.

Table 1. Parameter settings for performance-based model.

	$[\underline{t_b}, \overline{t_b}]$	$[\underline{r_b}, \overline{r_b}]$	r_0	τ	α_w	β_w	μ	θ
Value	[0,0.5]	[0.2,5]	2	0.5	2.5	3.5	0.5	0.25

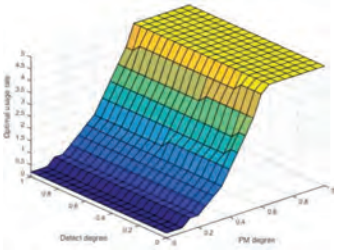


Figure 1a. Optimal usage rate under different PM and detect degree.

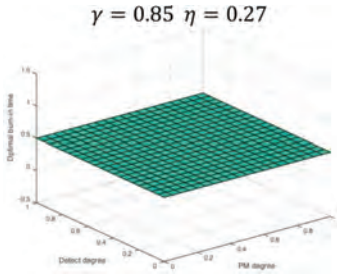


Figure 1b. Optimal burn-in time under different PM and detect degree.

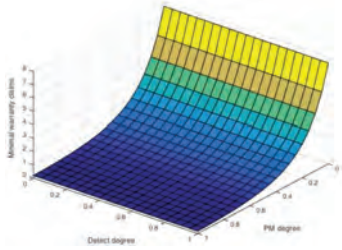


Figure 1c. Minimal warranty claims under different PM and detect degree.

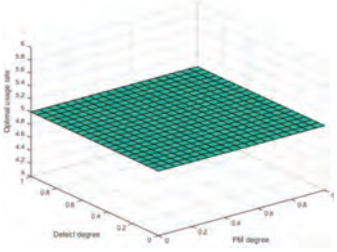


Figure 2a. Optimal usage rate under different PM and detect degree.

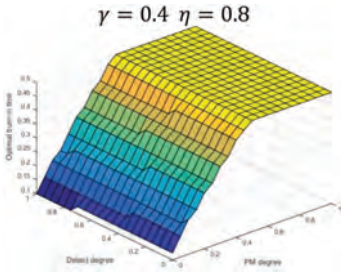


Figure 2b. Optimal burn-in time under different PM and detect degree.

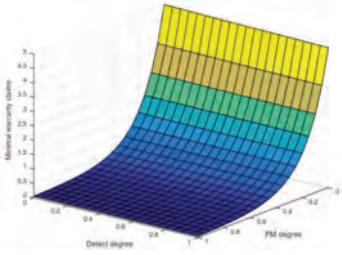


Figure 2c. Minimal warranty claims under different PM and detect degree.

From Fig. 1(a, b, c), we can see that if $\gamma > \eta$, then the optimal burn-in duration is always set to be upper bound $\overline{t_b}$, while optimal burn-in usage rate increases from r_b to $\overline{r_b}$ with the degree of PM actions and decreases with the detecting degree. Conversely, Fig. 2(a, b, c) shows that if $\gamma < \eta$, then the optimal burn-in usage rate stays permanently at the upper bound $\overline{r_b}$, while the optimal burn-in duration increases from t_b to $\overline{t_b}$ with the degree of PMs and decreases with the detecting degree. These curves conform to the proposition 1 and 2. In addition, we also find that as the maintenance or the detecting degree increases, the minimal warranty claims in all examples decrease, verifying that the larger the maintenance degree or the detecting degree increases, the better performance the warranty contract is. Therefore, manufacturers only have to consider their own limited capabilities to implement PM actions.

5 CONCLUSION

In this article, we have studied the optimal strategies of burn-in under PM actions for two-dimensional warranted products.

To help manufacturers to improve the warranty contract performance, we have investigated a performance-based model to help manufacturers to make optimal decisions on burn-in under PMs. In our model, the burn-in duration and burn-in usage rate are set as the decision variables to minimize the expected warranty claims within the warranty coverage. Through computing specific numerical examples, we demonstrate the validity of the model. These findings could help manufacturers to promote the warranty performance in burn-in and warranty management.

ACKNOWLEDGMENTS

This research is supported by the National Science Foundation of China (No. 71171142 &71532008).

REFERENCES

Baik J., Murthy D.N.P., 2008. Reliability assessment based on two-dimensional warranty data and an accelerated failure time model. International Journal of Reliability and Safety, 2(3), 190–208.

Blischke W.R., Karim M.R., Murthy D.N.P., 2011. Warranty data collection and analysis[M]. Springer Science & Business Media.

Boyd S., Vandenberghe L., 2004. Convex optimization[M]. Cambridge university press.

Chan V., Meeker W.Q., 1999. A failure-time model for infant-mortality and wearout failure modes. IEEE Transactions on Reliability, 48(4), 377–387.
Doyen L., Gaudoin O., 2004. Classes of imperfect repair models based on reduction of failure intensity or virtual age. Reliability Engineering & System Safety, 84(1), 45–56.
Iskandar B.P., Murthy D.N.P., Jack N., 2005. A new repair–replace strategy for items sold with a two-dimensional warranty. Computers & Operations Research, 32(3), 669–682.
Kar T.R., Nachlas J.A., 1997. Coordinated warranty and burn-in strategies. IEEE Transactions on Reliability, 46(4), 512–518.
Leemis L.M., Beneke M., 1990. Burn-in models and methods: a review. IIE transactions, 22(2), 172–180.
Nguyen D.G., Murthy D.N.P., 1982. Optimal burn-in time to minimize cost for products sold under warranty. IIE Transactions, 14, 167–174.
Shafiee M., Asgharizadeh E., 2011. Optimal burn-in time and imperfect maintenance strategy for a warranted product with bathtub shaped failure rate. International Journal of Collaborative Enterprise, 2(4), 263–274.
Shafiee M., Finkelstein M., Zuo M.J., 2013. Optimal burn-in and preventive maintenance warranty strategies with time-dependent maintenance costs. IIE Transactions, 45(9), 1024–1033.
Shahanaghi K., Noorossana R., Jalali-Naini S.G., et al., 2013. Failure modeling and optimizing preventive maintenance strategy during two-dimensional extended warranty contracts. Engineering Failure Analysis, 28, 90–102.
Sheu S.H., Chien Y.H., 2005. Optimal burn-in time to minimize the cost for general repairable products sold under warranty. European Journal of Operational Research, 163(2), 445–461.
Wang Y.K., Liu Z.X., Liu Y.L., 2015. Optimal preventive maintenance strategy for repairable items under two-dimensional warranty. Reliability Engineering & System Safety, 142, 326–333.
Wu C.C., Chou C.Y., Huang C., 2007. Optimal burn-in time and warranty length under fully renewing combination free replacement and pro-rata warranty. Reliability Engineering & System Safety, 92(7), 914–920.
Ye Z.S., Murthy D.N.P., Xie M., et al., 2013. Optimal burn-in for repairable products sold with a two-dimensional warranty. IIE Transactions, 45(2), 164–176.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Evaluation method of maintenance operation space based on virtual reality

Pengyan Liu, Dong Zhou, Ziyue Guo, Juan Wu & Yuan Li
Beihang University, Beijing, P.R. China

ABSTRACT: In order to ensure the excellent maintainability of the product, it is necessary to give full consideration to the operation space of the maintenance personnel at design time. The traditional evaluation method of the operation space is that gives qualitative evaluation results and puts forward suggestions based on the virtual maintenance simulation animation and the digital prototype while comparing with rules of maintainability design by experts. The evaluation result usually depends on the expert level with subjective tendency. There is still a lack of widely used quantitative evaluation methods. This paper analyzes the main attitude of maintenance personnel in maintenance activities, evaluates their comfort level, and then evaluates the operation space. First of all, through the analysis of the characteristics of maintenance activities, the article draws the main factors of human comfort evaluation. Based on the artificial potential field theory commonly used in robot obstacle avoidance path planning, a reachability potential field that can be used to describe the upper limb posture of maintenance personnel is established and a criterion for evaluating upper limb comfort is proposed. Then, combined with the virtual maintenance technology, using the virtual maintenance simulation platform, together with a comprehensive analysis of tool rotation angle, operation space ratio and the hand swept volume of maintenance posture, the comfort of the maintenance staff's hand gesture is analyzed. After that, the above two parts are combined to comprehensively evaluate the comfort degree of the human body posture and give the evaluation criteria. And a new operation space evaluation method for the key maintenance steps in the maintenance operation process is established. In the end, the maintenance and disassembly process of an engine is taken as an example to verify the validity of this method.

1 INTRODUCTION

With the rapid development of industrial technology and the ever-increasing demands of manufacturers on the machining accuracy, product maintenance activities have become more and more complicated. At present, most design manufacturers have realized that at the same time, ensure the high reliability and maintainability of the products can guarantee in maximum availability throughout the product life cycle and achieve the greatest economic benefits (Sobral & Guedes, 2016). The quality of maintainability determines the procedures and duration of maintenance activities. In addition, excellent maintainability will reduce maintenance errors and avoid safety accidents. Maintenance reachability of the product can directly affect the size of the workload and maintenance posture of maintenance activities. Poor maintenance reachability will lead to fatigue for service personnel, increase the probability of occupational disease and further increase the risk of maintenance errors. Stader, for example, confirmed this by conducting interviews and ergonomics analyzes of general aircraft maintenance tasks (Stader, 2013).

Since the maintenance reachability of a product is an inherent quality characteristic of the product itself, it is important to start with the design phase to improve maintenance reachability. The maintenance reachability of the product means that when the maintenance work is carried out, the degree of difficulty to the system, equipment, and parts of the machine can be seen, touched, inspected, adjusted, disassembled, or other maintenance activities. (Krause & Jager, 1988).The maintenance reachability includes three evaluation indexes, sight line reachability, physical reachability and operation space (Zeng, 2007).

This paper mainly studies the evaluation method of the maintenance operation space, which is the actual operation space for the maintenance personnel to repair the product fault unit. At the design stage, enough maintenance operation space must be reserved to avoid collision in maintenance process and ensure the convenience, safety and comfort of maintenance personnel (Zhou et al, 2011). Designing a suitable maintenance operation space

will enable the maintenance personnel to observe and maintain the equipment more conveniently. Even if it takes a long time to maintain a certain working attitude, the reduction of maintenance personnel fatigue sensation and discomfort can be achieved as much as possible as well as the maintenance staff's safety, health and comfort, so as to improve the overall product maintenance efficiency and reduce maintenance error probability caused by personnel fatigue. At present, the judgement of maintenance operation space in the design stage is mainly based on expert experience and visual effect of virtual maintenance animation presentation (Peng, 2010). Evaluation results usually rely on the level of experts, with strong subjectivity. There is still a lack of a widely used quantitative assessment method. Therefore, this paper presents a calculation method for the quantitative evaluation of maintenance operation space during the design phase.

The rest of this paper is organized as follows: Section 2 proposes the concept of reachability potential field. Section 3 combined with the previous analysis of hand activities, puts forward the evaluation criteria for maintenance space. Section 4 is a case study. Section 5 is conclusion.

2 REACHABILITY POTENTIAL FIELD

2.1 *The proposed process of reachability potential field*

In 1986, Khatib first proposed the artificial potential field method (Weerakoon, 2015), and applied it in the field of robot obstacle avoidance. The basic idea of this method is to construct a repulsive potential field around the obstacle and construct a gravitational potential field around the target point, similar to the electromagnetic field in physics, as shown in Figure 1. The controlled object is under repulsion and gravitation in the composite field composed of forgoing two kinds of potential fields. The combined forces of repulsion and gravitation direct the movement of the controlled object and search for the obstacle avoidance path without collision. In the maintenance process, maintenance personnel in a narrow operation space, bypassing the obstacles, reach the location of the maintenance object. The process is similar to robotic obstacle avoidance, so the concept of artificial potential field is introduced into the evaluation system of maintenance operation space, meanwhile proposes the concept of reachability potential field.

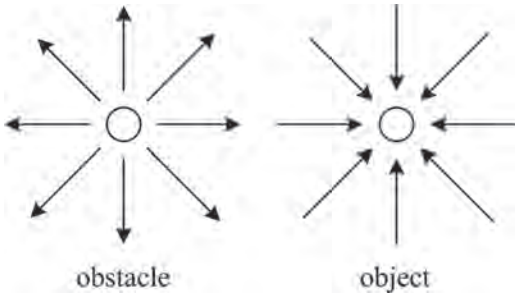


Figure 1.

2.2 *Gravitational potential field*

Since the maintenance staff mainly rely on the upper limb to complete the maintenance action, the following definition is made that the gravitational potential field is mainly related to the distance between the maintenance staff's joint and the object to be repaired. The greater the distance, the greater the potential energy value of the joint; the smaller the distance, the smaller the potential energy value of the joint. The function of the gravitational potential field is:

$$U_g(q)=\begin{cases}\frac{1}{2}\eta\rho^2\left(q,q_g\right), & 0\le\rho\left(q,q_g\right)\le\rho_0\\ 0, & \rho\left(q,q_g\right)\ge\rho_0\end{cases} \tag{1}$$

where η is the proportional gain coefficient, $\rho(q, q_g)$ is a vector representing the Euclidean distance $|q\text{-}q_g|$ between the joint position q and the object position q_g which to be repaired, and the vector direction is from the position of the joint to the position of the object to be repaired. When maintenance personnel are far away from the object to be repaired, it makes no sense to consider the comfort of the operation space during maintenance. Therefore, it can directly consider whether the maintenance object can be reached. To simplify the formula, this paper only considers the situation when the distance between maintenance staff and the object to be repaired is close. As the maintenance action is mainly completed by the human upper limb, define ρ_0 as the maximum length of the serviceman's shoulder joint to the tip of the ipsilateral middle finger, as shown in Figure 2.

Define when the joint of the maintenance personnel is just an arm away from the object to be repaired, the reachability gravitational potential of the joint is 1; and when the joint of maintenance personnel completely wraps the object to be repaired, the reachability gravitational potential of the joint is 0. That is, when $|q\text{-}q_g| = \rho_0$, $U_g(q) = 1$; and when $|q\text{-}q_g| = 0$, $U_g(q) = 0$.

$\eta=\frac{2}{\rho_0^2}$ can be obtained.

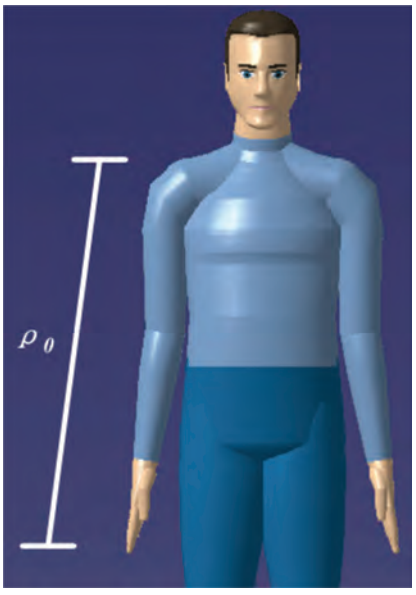


Figure 2. Shoulder joint to the ipsilateral middle finger fingertips maximum length.

2.3 *Repulsion potential field*

Only in the case of limited maintenance operation space, can the analyzation and verification be meaningful. Therefore, the obstructions in the vicinity of the object to be repaired always exist. The factor that determines the repulsion force field of obstacle is the distance between a certain joint of the maintenance staff and the obstacle. When the joint is beyond the influence range of the obstacle, its potential energy value is zero. After the joint enters the influence range of the obstacle, the greater the distance between the two is, the smaller the potential energy value of the joint will be, and the smaller the distance, the greater the potential energy value of the joint will be. The repulsive potential field potential function is:

$$U_r(q) = \begin{cases} \frac{1}{2}k\left(\frac{1}{\rho(q,q_r)} - \frac{1}{\rho_0}\right)^2, & 0 \le \rho(q,q_r) \le \rho_0 \\ 0, & \rho(q,q_r) \ge \rho_0 \end{cases} \quad (2)$$

where k is a positive proportional coefficient. $\rho(q, q_r)$ is a vector whose size is the distance $\rho(q, q_r)$ between the joint and the obstacle, pointing in the direction from the obstacle to the joint. ρ_0 is a constant that represents the maximum distance the obstacle affects the joint. Only when the obstacle is within reach of the upper limbs of the maintenance personnel will the maintenance activities be affected. Therefore, define that the maximum influence distance of the obstacle is the length ρ_0 of maintenance staff shoulder to the ipsilateral middle finger fingertip maximum length.

Define when the maintenance personnel just enter the edge of the obstacle range, the maintenance staff's joint repulsion potential energy is 0. When the maintenance staff completely wrapped obstacles, the maintenance staff's joint repulsion potential energy is 1. That is, when $|q\text{-}q_r| = \rho_0$, $U_r(q) = 0$; and when $|q\text{-}q_r| = 0$, $U_r(q) = 1$.

$k = 2\rho_0^2$ can be obtained.

3 MAINTENANCE POSTURE COMFORT EVALUATION

3.1 *Upper limb comfort evaluation*

Human upper limb mainly includes four parts of hand knuckles, wrist, elbow and shoulder joints. Knuckle and wrist joint comfort will be discussed in Section 3.2. This part mainly analyzes the reachability potential energy and relative comfort of the human elbow and shoulder joints.

Upper limb comfort is mainly determined by two factors, one is the upper limb stretch, and the other is whether there are obstacles around the upper limbs. The upper arm's stretch is mainly determined by the size of the gravitational potential field. The impact of obstacles mainly depends on the size of the repulsive potential field. Since each obstacle in the maintenance process affects the maintenance activities respectively without counteraction, it makes no sense to superimpose repulsion potential fields. In the calculation, engineers only need to examine the distance from the joint of the most obstructions, namely the maximum repulsion potential field.

Define that ω is the evaluation value of upper limb comfort obtained from the theory of reachability potential field.

$$\omega = \omega_g \cdot \omega_r \quad (3)$$

The reachability gravitational potential suffered by the shoulder joint measures the stretch of the human upper limb during maintenance activities. According to ergonomics (Xie & Huang, 2009), people tend to feel tired when they bend over. The optimal range and normal range of human arm activity are respectively 0.59 and 0.78 times of maximum range when the upper body is in the upright position.

When $\rho(q,q_g) \le 0.59\rho_0$, that is, $0 \le U_g(q) \le 0.35$, the upper limbs' stretching posture is the most comfortable, and $\omega_g = 1$.

When $0.59\rho_0 \le \rho(q,q_g) \le 0.78\rho_0$, that is, when $0.35 \le U_g(q) \le 0.61$, the extension posture of the upper limbs is normal, and $\omega_g = 0.61$.

When $0.78\rho_0 \le \rho(q,q_g) \le \rho_0$, that is, when $0.61 \le U_g(q) \le 1$, the posture of the upper extremity is the worst, and $\omega_g = 0.35$.

The reachability gravitational potential energy size of elbow joint and the repulsive potential energy size caused by the recent obstacles are compared to measure the influence of the obstacles around the upper limbs during maintenance activities.

When the maintenance posture is determined, the elbow's reachability gravitational potential energy size is

$$U_g(q) = \left(\frac{\rho(q,q_g)}{\rho_0} \right)^2 \tag{4}$$

Among them, $\rho(q,q_g)$ is the longest distance from the elbow to the fingertip in this maintenance posture.

The elbow reachability repulsive potential energy size is

$$U_r(q) = \left(\frac{\rho_0}{\rho(q,q_r)} - 1 \right)^2 \tag{5}$$

Among them, $\rho(q,q_r)$ is the shortest distance between the elbow joint and the nearest obstacle in this maintenance attitude.

When $U_g(q) > U_r(q)$, it is considered that the elbow joint is subjected to a larger gravitational potential field and a smaller repulsive potential field, and the maintenance posture is comfortable. Define that $\omega_r = 1$.

When $U_g(q) \le U_r(q)$, it is considered that the elbow joint is subjected to a small gravitational potential field and a large repulsion force potential field, which is obstructed by obstacles during the maintenance activities. The maintenance posture is uncomfortable. Define that $\omega_r = 0.5$.

The evaluation index of the comfortableness of the upper limb of the maintenance personnel can be obtained according to the Equation 3.

3.2 *Hand swept comfort index*

Since most of the maintenance operations are completed by the maintenance staff's hands, this part of the measurement method focuses on the hand comfort of the maintenance personnel. Maintenance operations are divided into three basic maintenance activity units (MAUs): screw, twist and translate (Zhou et al, 2011). Specific actions shown in Figure 3, Figure 4, Figure 5. Screw is the hand act of tightening or loosing a screw with a screwdriver. Twist is the hand act when the nut is installed or dismantled by a wrench. Translate is the parallel movement of human hands without any posture change. DELMIA (Digital Enterprise Lean Manufacturing Interaction Application) is a digital manufacturing application made by the French Dassault Systemes company. In the product design stage, designers can make maintenance simulation animation with it, to perform virtual maintenance and verify the maintainability of the product. With DELMIA, we can establish the swept volume for each maintenance activity unit. There are two types of swept volumes: free swept volume (V_{fsv}) and constraint swept volume (V_{csv}).

Figure 3. Screw sweep volume.

Figure 4. Twist sweep volume.

Figure 5. Translate sweep volume.

Under completely freedom circumstance, free swept volume is the maximum swept volume of the hand movement range when taking a comfortable attitude to perform maintenance tasks. The maximum or maximum angle at which the human hand can move comfortably can be obtained by consulting ergonomic data. The constraint swept volume is the swept volume determined by the actual hand movements of the maintenance personnel in the virtual maintenance animation, subject to the constraints of the machine parts, which can be obtained by DELMIA.

The restricted sweep volume is equal to or less than the free-sweep volume. The wrist maximum angle of motion is 180° in the screwed state

through the ergonomics. So define that the sizes of sweep volume of the wrist when screw 180° is the screw free sweep volume. When the person is in a free state, the twist angle is generally 120.

We can see through the body efficacy that the maximum wrist motion range is 180°, so the definition of the wrist swept volume of 180° twisting volume of free sweep. When the person is in a free state, the rotation angle is generally 120°. Define the swept volume at 120° twist as a free-sweep volume.

Define the sweep comfort index P_v, as shown in Equation 6. Based on the ergonomic data to establish a quantitative evaluation standard for P_v -based operation space quantitative evaluation criteria (Zhou et al, 2011). The evaluation criteria of different maintenance activity units are shown in Table 1.

$$P_v = V_{csv} / V_{fsv} \tag{6}$$

3.3 *The total comfort evaluation*

Maintenance staff's hand movements are important in themselves, and comfort in the upper limbs also significantly affects comfort. Furthermore, a complete set of maintenance actions is a combination of a series of operations. Therefore, in order to combine the comfort of the upper limb with the comfort of the hand, total comfort is defined.

$$s = \sum_{i=1}^{n} \omega_i P_{vi} \tag{7}$$

where s is the total maintenance operation space scores. The set of maintenance actions is made up of i maintenance activity units, ω_i is the upper limbs comfort score for an action, and P_{vi} is the hand swept volume ratio for that action. The higher the value of s, the higher the comfort level.

$$0 \le \omega_i \le 1$$

$$0 \le P_{vi} \le 1$$

3.4 *Evaluation criteria*

Suppose a set of maintenance actions have x screwing action, y twisting action, and z translation action. According to the criteria in Table 1, the maintenance operation space involved in the whole set of actions is defined as excellent when the evaluation of each maintenance activity unit is excellent. When the evaluation of each maintenance activity unit is bad, define maintenance operation space involved in the entire operation as bad. When the evaluations of each maintenance activity units consist of all levels of excellent, normal and bad, the maintenance operation space involved in the entire operation is defined as normal. In summary, the maintenance of operation space evaluation criteria as shown in Table 2.

Table 1. The evaluation criteria of hand.

Maintenance activity unit	Evaluation criteria	P_v
Screw	Good	>0.8
	Normal	0.5–0.8
	Bad	<0.5
Twist	Good	>0.75
	Normal	0.25–0.75
	Bad	<0.25
Translate	Good	>0.9
	Normal	0.7–0.9
	Bad	<0.7

Table 2. The evaluation criteria of operation space.

Evaluation criteria	s
Good	$s \ge 0.8x+0.75y+0.9z$
Normal	$0.5x + 0.25y + 0.7z < s < 0.8x + 0.75y + 0.9z$
Bad	$s \le 0.5x + 0.25y + 0.7z$

3.5 *The scope of application*

Operation space determines the maintenance staff's physical posture. By evaluating the comfort of the maintenance personnel posture, the design of the operation space can be obtained and quantitatively evaluated. The evaluation of the operation space is not based on the entire maintenance process, but rather on the key actions in the maintenance process. Throughout the maintenance process, visual inspection and qualitative analysis by engineers shall confirm that most of the operation gestures are comfortable. Quantitative analysis, which can be used to help improve design, is only necessary in the key steps that engineers have questions.

4 APPLICATION EXAMPLES

This case is the quantitative evaluation of the maintenance operation space of the disassemble a certain type of passenger aircraft APU hexagonal nut. The purpose is to use the proposed method to quantitatively evaluate maintenance operation space to demonstrate the flexibility and effectiveness of the method. The APU virtual maintenance operation animation screenshot shown in Figure 6. There are six orange hex nuts on the APU. Take nuts 1, 2, and 3 as typical cases. The positions of the three nuts are shown in Figure 7.

When disassembling the three nuts, remove the No. 1 nut and the right elbow is closest to the No. 4 thick round pipe. When disassembling the No. 2 and No. 3 nuts, the right elbow joint is closest to the No. 5 thin round pipe. A large space around nuts

Figure 6. Virtual maintenance process screenshots.

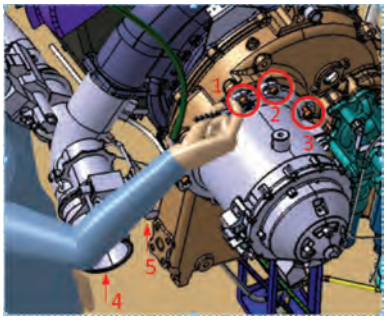


Figure 7. APU nut position diagram.

Table 3. Case study results.

	No. 1 nut	No. 2 nut	No. 3 nut
V_{csv}	1.152	1.152	0.576
V_{fsv}	1.152	1.152	1.152
P_{vi}	1	1	0.5
ρ_0(mm)	733	733	733
$\rho(q, q_g)$ (mm)	569	604	676
$\rho(q, q_r)$ (mm)	280	263	255
$U_g(q)$	0.6025826	0.6789940	0.8505218
$U_r(q)$	2.6174617	3.193627	3.513787
ω_g	0.61	0.35	0.35
ω_r	0.5	0.5	0.5
s	0.305	0.175	0.175
Level	Normal	Bad	Bad

1, 2, and maintenance personnel can reach 120° hand rotation. No. 3 nut is relatively narrow, and maintenance personnel can only reach 60 degrees rotation hand. According to Equation 4, Equation 5, Equation 7 gives the following Table 3.

5 CONCLUSION

When evaluating the comfort of the operation space during the design phase, the traditional method is to use DELMIA, Jack and other simulation software to make the maintenance simulation animation and give the evaluation results and suggestions for improvement through the animation effects. This paper presents a new method of quantitative evaluation of maintenance operation space in Section 2 and Section 3. Maintenance activities are mainly concentrated in the human upper limbs, so the evaluation of the operation space should be taken into account the hand movements and arm movements. Reference to the concept of artificial potential in the research of robot obstacle avoidance, the concept of reachability potential energy is proposed and used to analyze and evaluate the upper limb posture of maintenance personnel in maintenance activities. At the same time, combined with the predecessors' use of swept volume to evaluate the maintenance of hand space, get the overall evaluation of maintenance attitude data. Later, make a comparison with the evaluation criteria proposed in Section 3.4. Finally, the paper concludes whether that operation space is qualified or not by assessing the comfort of the maintenance posture.

In the evaluation of human upper limb comfort, this article considers only the single obstacle nearest. In subsequent studies, the effects of multiple obstacles acting simultaneously on the upper limbs can be considered.

REFERENCES

Krause G S, Jager R. 1988. CAD techniques for im proved maintainability design. *Reliability and Maintainability Symposium*, 122–126.

Peng G., Yu H., Liu X., et al. 2010. A desktop virtual reality-based integrated system for complex product maintainability design and verification. *Assembly Automation,* 30(4): 333–344.

Sobral, J., Guedes, C. 2016. Repairable items inventory optimization based on maintenance data and risk criteria. *Risk, Reliability and Safety: Innovating Theory and Practice*.

Stader, Sally A. 2013. Ergonomic Evaluation of Aircraft Wing Recovering Tasks in General Aviation Maintenance. *Proceedings of the Human Factors and Ergonomics Society 57th Annual Meeting,* 57(1): 1249–1253.

Weerakoon T., Ishii K., Nassiraei AA F. 2015. An artificial potential field based mobile robot navigation method to prevent from deadlock. *Journal of Artificial Intelligence & Soft Computing Research*, 5(3):189–203.

Xie Q., Huang Y. 2009. Ergonomics. 2nd ed. *China Building Industry Press.* (in chinese)

Zeng, Y. 2007. Research on analysis and evaluation of reachability based on maintainability design. *Graduate School of National University of Defense Technology. Changsha, Hunan, P.R.China* (in Chinese)

Zhou D., Jia X., Kang L., et al. 2011. Using the swept volume to verify maintenance space in virtual environment. *Assembly Automation,* 34(2): 192–203.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Maintenance resources allocation for the profit maximization of a park of identical systems

W. Zhu
School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an, China

B. Castanier
Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS), Université d' Angers, Angers, France

ABSTRACT: For the park (system of systems) consisting of a set of identical systems, the mission of each time unit is shared by all the survival systems which could be overexploited to achieve the global park objective. This overexploitation is stressful for each individual system and increases its respective degradation. This leads to increase the probability of failure of the system before the next planned maintenance. Otherwise, the system can be subject to operational constraints such as the reduction of exploitation because of an excessive degradation. Such constraint could affect the overall objective. We propose in this study to analyze the problem of the maintenance resource allocation on a park of n identical systems for ensuring a given production goal on a two successive maintenance period. Each system is degrading due to cumulative load and can be totally or partially renewed only during planned maintenance. We propose to construct a simulation-based model for the profit assessment of the whole park on a given time horizon for different maintenance allocation policies given the different assumptions described above.

1 INTRODUCTION

For the park which consists of n identical system, which referred as "system of systems" in this study, the load of each mission is shared by the survival systems. During the mission phase, the random failure of systems keep changing the load charging state of the survival systems (Park 2010, Ye et al. 2014, Zhang et al. 2017). Meanwhile, the maintenance planning and completion of the systems are affected by the random system failures and realistic conditions, such as the maintenance resources, the maintenance windows and production plan (Gundegjerde et al. 2015, Abdollahzadeh et al. 2016). Nowadays, the study of system degradation has extended from two-state to multi-state (Peng et al. 2017), (Dao & Zuo 2017b, Dao & Zuo 2017a). The performance levels of each system are corresponding to different stage of system degradation. For the system of systems, the mission load is usually achieved by redistributing the remaining load on the survival systems, meanwhile the total cost of maintaining the system state and performance is preferable to be as low as possible. Taking offshore wind farm as an example, the electricity production of the whole wind farm should be in accordance with the contract (Hawker & McMillan 2015). The manager of the wind farm has to control and distribute the production demand according to the contract and the actual state of each wind turbine. The maintenance resources, such as the maintenance time, the maintenance team and spare parts of the wind farm are limited due to the economic concerns and the offshore environment, which makes the maintenance policy of offshore farm more complex (Martin et al. 2016). Hence, how to balance the control strategy and the maintenance policy is the common challenge for the system of systems (Irawan et al. 2017, Santos et al. 2015).

Ye et al. (2014) proposed a load-sharing industrial system where the operator allocates load to balance the level of the degradation condition of all parallel components to achieve system performance, where a simple replacement policy is conducted according to the cumulative work load. Zhu et al. (2011) conducted a cost-based selective maintenance decision-making on a machine line for selecting the optimal machine group under limited maintenance duration. Zuo (2017b) addressed a selective maintenance problem for multi-state systems where each component can be in one of multiple working levels and several maintenance actions are optional. The dependency of the system is presented by multiple hierarchical levels and dependence groups. Further Dao and Zuo (2017a) studied the selective maintenance problem on multi-state series systems sharing in variable loading conditions in the next mission with the aim of

maximizing the expected system reliability in the next mission within available resources. Santos et al. (2015) presented a simulation method to study how the variation of failure and repair models, vessels logistic times, weather windows and waiting times affect a wind turbine performance, hence to identify the factors which most influence the turbines performance.

Based on the previous industrial practice of offshore wind farm and maintenance practices of complex system, the main goal of the present paper is to focus on a specific case of system of systems and to propose a simulation-based optimal maintenance resource allocation rules in the context of offshore wind farm. Besides, control rules based on the system condition and remaining production of mission are included in this study. By setting the different assumptions for the elaboration of the assessment model, we will conduct a series numerical simulations based on the degradation model, control and maintenance decision rules with the proposed algorithms. Our study contributes on the follow facts: (1) The study is based on the framework of "system of systems" under random working condition. (2) The control rules which regulate the system degradation and production rate are considered. (3) The maintenance crew transfer time from one-to-one site is introduced, which is the practical issue for offshore wind farm. The remainder of this paper is organized as follows: Section 2 describes the degradation model of single wind turbine based on cumulative load and the performance of the wind farm. The assumptions of maintenance policy is introduced as well. Section 3 describes the production decision rule and integrates the maintenance decision considering both the production and maintenance resource allocation. The simulation algorithm and numerical results are given in section 4. The contributions, limits and future works of this paper are discussed in conclusion.

2 PERFORMANCE MODEL

The objective of this section is to present the performance model of the wind farm with n Wind Turbines (WTs) from period to period. The performance is defined by an average cumulated amount of energy objective P_τ produced over a period τ. Each of the wind turbines will contribute to reach this production objective. A wind turbine is subject to degradation. In the next paragraph, the degradation model for one WT is presented.

2.1 *Performance model for one wind turbine*

We assume that a WT is subject to continuous random cumulative degradation $X(t)$ from 0 to a failure threshold x_F. $X(t)$ is a function of the cumulative load of the WT. The instantaneous load at time t is a function of the rotation speed of the blades given a wind speed, w_t and a WT rotation control parameter, ρ_t. Rotation can be controlled by both the pitch control and a brake. Braking should increase degradation. Finally, the degradation rate is given by:

$$\lim_{dt\to 0}\frac{X(t+dt)-X(t)}{dt} = \frac{\alpha(\rho_t,w_t)}{\beta} = \frac{a_1\cdot\rho\cdot w + a_2\cdot(1-\rho)}{\beta} \quad (1)$$

Let remark that the blades rotation speed is assumed to be directly proportional to the wind speed $\omega = a\rho\cdot w$ and the degradation rate be constant if $w_t = w$ and $\rho_t = \rho$. Another constant can be directly introduced for modeling degradation for idle states, specifically due to the lack of wind.

Under such assumptions, the degradation over a time period $(0,t_k)$ can be modeled by a non-stationary Gamma process where the degradation increments over $(t_i,t_{i+1}), i\in\{1,\ldots,k\}$ are gamma distributed. Some assumptions on a_1 and a_2 can be done to ensure some of empirical degradation properties.

The failure $t_f = \min\{t>0 \mid X_t \geq x_F\}$ leads to an immediate stop of the WT.

We assume a constant production rate of one WT per unit of time given by $f_P(\omega_t) = f_P(\rho_t,w_t)$ where ω_t is the instantaneous rotation speed as a function of ρ_t and the wind w_t. The production rate of a WT has to be evaluated at each changing times, wind transition or change in the ρ-value.

2.2 *Maintenance assumptions*

A periodic maintenance policy is implemented every τ period. No maintenance is allowed out of this maintenance period (block replacement model). This maintenance consists in visiting a certain number of WTs and renew them as much as possible. Other maintenance such as minimal maintenance are not considered in this model. The total maintenance capacity is here assumed to be fixed and not all of the failed or degraded WTs will be maintained in the good state. The maintenance capacity is related to a limited number of maintenance resources and is defined by a given maintenance duration D_M and a given renewal efficiency rate per unit of time e_m. Hence, the maximum allowed degradation reducing is $D_M\cdot e_m$. Moreover, some of maintenance crew transfer time from one-to-one WT is introduced. It is assumed this crew tranfer time is constant and denoted d_{ij}. Finally, if n_i over the n WTs are visited during a maintenance period, the maximum maintenance capacity is $\left(D_M - n_i\cdot d_{ij}\right)\cdot e_m$.

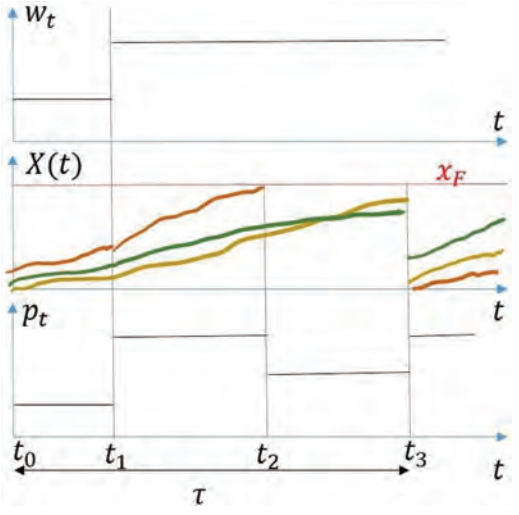


Figure 1. Sketch of the wind farm over a period.

2.3 *Illustration of the performance model*

Figure 1 sketches the evolution of the production rate, p_t, and the WTs degradation, $X_t(t)$, given the wind behavior, w_t, and a constant rotation control parameter ρ on the whole τ period. Here, the wind is modelled by a Continuous Time Markov Chain. Such a discrete model is motivated by, first, the wind speed forecasts are not so precise for a middle-term time horizon and only average values would be necessary. Second, a precise effect of the wind on degradation can be a challenge and should be estimated for average wind levels. Note that the associated transition rates and probabilities can be functions of the seasons. For the sake of simplicity, no variations are considered in this paper. When the wind increases at t_1, both the production rate p_{t_1} which is the sum of the functioning WTs and the WTs degradation rates increase. When a WT fails and stops at t_2, the others continue to produce but the production rate of the farm is decreasing until the next maintenance at t_3. In this example, the 3 WTs are repaired and the sum of the repair level is therefore $\Delta_M = \left(D_M - 3d_{ij}\right)\cdot e_m$. All of the times t_0, t_1, t_2 and t_3 define a change in the wind farm environment.

3 DECISION PROCESS

The decision rules are twofold: production and maintenance which should be combined for ensuring the production objective.

3.1 *Production decision rule*

We propose here to focus on the production decision rules which can be captured in the blade rotation speed control parameter ρ (controlled by the pitch). At time t, these decision rules are a function of the production objective P_τ over the time period τ, the cumulative production at time t, $t < \tau$, $P(t)$, the current wind speed w_t, the wind forecasts over the end of the time period, $Ew(t)$, the current state vector of the WTs, $X(t) = \left(X_1(t), X_2(t), \ldots, X_n(t)\right)$ and finally the forecasts in the production from t to τ, $EP\left(t; X(t), w_t, Ew(t), \rho_t\right)$.

For the sake of simplicity, the decision in ρ at time t will be defined in order to satisfy the production objective given a constant forecast, i.e. without considering the future state of the other WTs on the farm. Because no additional information occurs between two successive changes, the ρ_t decision will be considered as constant. Finally, updates in the decision would only occur at any change. If a WT fails, if it is possible, the production rate of all the remaining WTs has to be increased to ensure the same yield and so ρ_t should be greater.

Let n_F the number of failed WTs. We can approximate the expected energy production for the $n - n_F$ survival WTs given the ρ_t variable, the wind forecasts and the vector $x(t) = \left(x_1(t), \ldots, x_n(t)\right)$ of the known WTs states by:

$$EP\left(t; X(t), w_t, Ew(t), \rho_t\right) = \sum_{i=1}^{n} 1_{\{x_i(t) < x_F\}} \times \cdots$$
$$F(\tau - t \mid x_i(t), Ew(t), \rho_t) \cdot f_P\left(\rho, \frac{Ew(t)}{\tau - t}\right) \cdot (\tau - t) \quad (2)$$

where $F(\cdot \mid x, w, \rho)$ is the conditional survival distribution and $f_P(\cdot,\cdot)$ the WT production rate function per unit of time. Let remark that the ρ_t is the same for all the survivals and is considered unchanged until the end of the period. Then, if the production criterion is P_τ, ρ_t should be chosen to ensure the missing production $P_\tau - P(t)$ and thus ρ_t is the solution of:

$$EP\left(t; X(t), w_t, Ew(t), \rho_t\right) = P_\tau - P(t) \quad (3)$$

If there is no solution, then $\rho_t = 1$ which ensures the maximum of the potential production.

3.2 *The decision variable as a function of the maintenance*

ρ_t should be defined to ensure both the current and the future periods production objectives. If the maintenance resources are limited, then it will not be possible to maintain the whole WTs in a perfect good state. Denote $X_p = D_M \cdot e_m$ the maximum cumulative degradation level the maintenance crew can save in a fixed maintenance period. Some maintenance allocation rules will be discussed later in the paper.

Let define X_{min} that is the minimum amount of degradation of the whole WTs to ensure the P_τ production over a complete period. In this paper, we define the farm state $Y(t) = \sum_{i=1}^{n} x_i(t)$. At time t, ρ_t should be updated to respect:

$$\Pr(Y(\tau) \le X_p + X_{\min} \mid x(t), Ew(t), \rho_t) < \epsilon_1 \quad (4)$$

where ϵ_1 is a decision parameter decision variable which represents the risk aversion of the decision-maker faced the non-respect of the production in a period. Numerical experiments should be conducted to highlight the influence on such a parameter as a function of the X_p maintenance capacity. Note that if there is no solution, it means that the current state farm $Y(t)$ is already greater than $X_p + X_{min}$. A lot of decision alternatives can be designed:

1. Increase the maintenance resources, X_p. Penalty costs should be added. It could be done with extra-time (and so decrease the next production period) or extra-resources.
2. Decrease the next maintenance period to ensure the respect of this maintenance specification next maintenance time
3. This maintenance specification can be evaluated at any time rather only at any operational change.
4. This can motivate the definition of a more flexible control rule for each of the WTs as a function of the individual degradation states. This would be closer to the load repartition problem.

3.3 *Maintenance allocation decision rules*

At the end of the production cycle, the overall farm system is $Y(\tau)$. A maintenance is specified by:

- D_M its duration;
- d_{ij} the time for a crew to move from one WT to another;
- e_m the efficiency of the maintenance per unit of time, i.e. the number of degradation points the maintenance can rehabilitated in one time unit (this degradation point is a function of the production point);
- costs and penalty costs which could be integrated.

If δ_i denotes the indicator of visiting the WT_i, the maintenance problem is then defined as the following optimization problem:

$$\max_{(\delta_i, t_i)} \sum_{i=1}^{n} \delta_i (d_{ij} + t_i) \quad (5)$$

subject to the following constraints:

$$\begin{cases} \sum_{i=1}^{n} \delta_i (d_{ij} + t_i) \le D_M + d_{ij} \\ e_m t_i \le x_i(\tau), \forall i \in \{1, \ldots, n\} \\ \sum_{i=1}^{n} (x_i(\tau) - e_m t_i) \le X_{min} \end{cases} \quad (6)$$

This optimization problem is a mixed-integer linear problem and it will not be treated in this paper. We propose to discuss two different allocation rules in the next section.

4 NUMERICAL EXPERIMENTS

We propose in this section to compare two maintenance resource allocation strategies. The main question here is should the maintenance favor the number of WTs in operation versus a limited number of operating WTs but in a better state. Before presenting and analyzing the different results, we propose to briefly introduce the simulation model with additional assumptions and the data.

4.1 *Simulation model and data*

We propose to analyze the performance of the model based on a Monte Carlo approach for the simulation of the production wind farm over a τ period. For ensuring the steady-state in the simulation process, a long-term time horizon with thousands τ periods will be considered. Average quantities will be then analyzed for conclusions. We will consider that, for the first period of this long-term horizon, all of the $n = 5$ wind turbines are new $x(0) = (0,0,0,0,0)$.

The decision framework requires the definition of P_τ, the production objective for the wind farm over τ. In the wind turbine industry, a WT can be considered available approximatively 70% of its life cycle. We define P_τ as the 0.7 of the maximum of the expected total production of n operating WTs over the whole production cycle given an average wind speed, $Ew(0)$ without any rotation control, $\rho = 1$. We have:

$$P_\tau = 0.7 \cdot n \cdot \left[f_P(1, Ew(0)/\tau) \cdot \tau \right] \quad (7)$$

The evaluation of X_{min}, the minimum farm state at the beginning of the period for ensuring the production with a given probability ϵ_1, is the solution of the following equation:

$$Pr\left(\bar{P}_0(X_{min}) \ge P_\tau\right) > \epsilon_1 \quad (8)$$

where $\bar{P}_0(X_{min})$ is the expected production over τ which is defined as a Gamma-distributed variable. The shape function of this Gamma variable is function of the wind forecast, $\bar{W}$, and an average

rotation control level ρ_0 over the period. In our paper, we have considered that the wind is modeled by a CTMC and the forecast can be directly evaluated. ρ_0 can be optimized from the simulation model. In this paper, we consider ρ_0 as a decision parameter.

The simulation algorithm is iterative from 0 to τ and can be summarized by the following steps, at each changing time $t \in (0, \tau)$:

1. Simulate the current wind condition (speed and next change time) and evaluate the forecast over the remaining time to the end of the period;
2. Find ρ_t, the minimum of the solutions of Equations (3) and (4);
3. Simulate the WTs degradation. For each of the WTs in operation:
   a. Simulate the remaining time as a function of $x_i(t)$ the current degradation level, the wind estimates and ρ_t;
   b. Identify the change time which is the minimum between the wind change, the remaining times and the end of the period
   c. Simulate the corresponding degradations for every WTs in operation.

The data used in our numerical experiments are fixed and presented in Table 1.

Let remind that ρ_0 and ϵ_1 are two decision parameters. Their relevance will be analyzed in the next paragraphs. Given these data, the production objective P_τ equals 14312 units and the Mean Time to Failure, in unit of time, for each WT without any maintenance is:

$$MTTF_i = \int_0^\infty Pr(X_i(t) < x_F)dt = 487 \quad (9)$$

4.2 *Policy 1: Maximization of the degradation rehabilitation levels*

Maximizing the degradation rehabilitation levels is here equivalent to restrict the crew move between WTs. Finally, the allocation rule for Policy 1 is:

Table 1. Data for the numerical experiments.

General	τ		WT yield factor	
	200		$r_p = 10$	
Degradation	a_1	a_2	β	x_F
	0.02	0.006	1.25	20
Maintenance operation	D_M	e_m	d_{ij}	
	10	4	1	
Constraints	Production		Maintenance	
	$\bar{\rho}_p = 0.7$		ρ_0	ϵ_1

Algorithm 1 Policy 1

Require: $x(\tau) = (x_1(\tau), \cdots, x_n(0)); D_M; e_m; d_{ij}$
  Sort the WTs from the most to the less degraded;
  $t_r \leftarrow 0$
  while $t_r < D_M$ **do**
    Renew to the best condition as possible the most degraded WT
    Update the repair time t_r (maintenance + transport)
  end while

Estimations of different performance indicators through 1000 runs are presented in Table 2: $\bar{P}(\tau)$ the energy produced in a cycle, $\bar{Y}_0$ the wind farm level after the maintenance, *rho* the mean of the rotation control on a cycle and $\rho == \rho_m$ the percentage of time ρ is defined by the maintenance constraint in a cycle.

These results are obtained for different values of the decision parameters ρ_0 and ϵ_1. From these experiments, we can measure the low effect of the ρ_0 parameter on the proposed indicators. This can be partially explained because of the choice of different linear functions for $\alpha(\cdot,\cdot)$ and $f_P(\cdot,\cdot)$. The relaxation of the maintenance constraint, ϵ_1, leads to some intuitive results such as more energy production in average because the rotation is mainly controlled by the production objective but the wind farm is more degraded in average. Another is, because the maintenance resource are here very restricted, some of the 5 WTs remain failed after the maintenance.

4.3 *Policy 2: Maximization of the operating WTs number*

We propose to conduct the same experiments when the objective of the maintenance is to ensure a maximum available WTs at the beginning of the production cycle. Policy 2 algorithm is now to find the threshold x_0 which verifies $x_i(0) \geq x_0, \forall i \in \{1, \cdots, n\}$ and $Y(\tau) - Y(0) = (D_M - (n_m - 1) \cdot d_{ij}) \cdot e_m$ where n_m is the number of maintained wind turbines. Table 3 presents the numerical results.

Table 2. Performance indicators for Policy 1.

Parameters		Performance indicators			
ρ_0	ϵ_1	$\bar{P}(\tau)$	$\bar{Y}_0$	$\bar{\rho}$	$\rho == \rho_m$
0.9	0.95	14362	34.2	82%	94%
	0.9	14555	32.9	79%	85%
	0.8	14713	32.5	80%	60%
	0.6	15155	37.5	84%	39%
	0.4	15131	36.9	85%	18%
0.8	0.8	14560	35.8	80%	73%
0.6	0.8	14633	35.0	82%	76%

Table 3. Performance indicators for Policy 2.

Parameters		Performance indicators			
ρ_0	ϵ_1	$\bar{P}(\tau)$	$\bar{Y}_0$	$\bar{\rho}$	$\rho == \rho_m$
0.9	0.95	13545	74.5	96%	99%
	0.9	13394	74.5	96%	99%
	0.8	13422	73.9	96%	99%
	0.6	13590	74.6	96%	98%
	0.4	13470	74.0	96%	99%
0.8	0.8	13438	74.4	96%	99%

In this case, the experiments do not allow to highlight the impact of the decision variables. In fact, the maintenance resource is too low for such a policy and the losses because of the transportation of the maintenance crew from one wind turbine to another are too high. After a maintenance, the wind farm state Y_0 remains very degraded. The mean ρ remains very high for ensuring the production given the fact that a lot of WTs would fail before the end of the cycle. The maintenance constraint cannot minimize these number of failures. Extra maintenance resource is required.

Finally, from these two numerical analysis, we can conclude the dominance of Policy 1 versus Policy 2 according to the proposed assumptions.

5 CONCLUSION AND PERSPECTIVES

In this paper, we have proposed a decision-making framework for the production and maintenance management of a wind farm which is defined as a system of systems. Based on this problem, we have introduced the problem of maintenance allocation when the resource are shared and limited. Two maintenance allocation policies have been presented. We have also proposed a rule for managing the production rate to guarantee both the production objective and the degradation level according to the maintenance resource. This is clearly a contribution in the maintenance optimization topic and especially for the study of imperfect block replacement policies.

This paper should be seen as a preliminary work. We have voluntary chosen some simple assumptions that seem to be realistic to us in an industrial context. Nevertheless, some of them could be seen too restrictive (the linearity of the production rate function or a constant degradation rate for given operating conditions, e.g.) and other more explicit (what means the degradation level for a complex system such as a wind turbine). At this stage, the current problem could be modeled as a Markov Decision Problem. Such a model could allow to identify some structural properties for the optimal control rules. Other perspectives can be drawn for, e.g., the production rate rule (rotation control ρ). In this paper, the updates in the ρ value are driven by stationary states in terms of wind forecasts and in ρ (no change from now to the end of the production cycle). To increase profitability at a minimum of risk, variability and uncertainty should be integrated in the decision framework.

REFERENCES

Abdollahzadeh, H., K. Atashgar, & M. Abbasi (2016). Multi-objective opportunistic maintenance optimization of a wind farm considering limited number of maintenance groups. *Renew. Energ. 88*, 247–261.

Dao, C. & M. Zuo (2017a). Optimal selective maintenance for multi-state systems in variable loading conditions. *Reliab. Eng. Syst. Safe. 166*, 171–180.

Dao, C. & M. Zuo (2017b). Selective maintenance of multi-state systems with structural dependence. *Reliab. Eng. Syst. Safe. 159*, 184–195.

Gundegjerde, C., I. Halvorsen, E. Halvorsen-Weare, L. Hvattum, & L. Nonås (2015). A stochastic fleet size and mix model for maintenance operations at offshore wind farms. *Transport. Res. C-Emer. 52*, 74–92.

Hawker, G. & D. McMillan (2015). The impact of maintenance contract arrangements on the yield of offshore wind power plants. *Proc IMechE Part O: J Risk and Reliability 229(5)*, 394–402.

Irawan, C., D. Ouelhadj, D. Jones, M. Stålhane, & I. Sperstad (2017). Optimisation of maintenance routing and scheduling for offshore wind farms. *Eur. J. Oper. Res. 256*, 76–89.

Martin, R., I. Lazakis, S. Barbouchi, & L. Johanning (2016). Sensitivity analysis of offshore wind farm operation and maintenance cost and availability. *Renew. Energ. 85*, 1226–1236.

Park, C. (2010). Parameter estimation for the reliability of load-sharing systems. *IIE Trans. 42*, 753–765.

Peng, R., H. Xiao, & H. Liu (2017). Reliability of multi-state systems with a performance sharing group of limited size. *Reliab. Eng. Syst. Safe. 166*, 164–170.

Santos, F., A. Teixeira, & C. Soares (2015). Modelling and simulation of the operation and maintenance of offshore wind turbines. *Proc IMechE Part O: J Risk and Reliability 229(5)*, 385–393.

Ye, Z., M. Revie, & L. Walls (2014). A load sharing system reliability model with managed component degradation. *IEEE T. Reliab. 63(3)*, 721–730.

Zhang, N., M. Fouladirad, & A. Barros (2017). Maintenance analysis of a two-component load-sharing system. *Reliab. Eng. Syst. Safe. 167*, 67–74.

Zhu, H., F. Liu, X. Shao, Q. Liu, & Y. Deng (2011). A costbased selective maintenance decision-making method for machining line. *Quality and Reliability Engineering International. 27(2)*, 191–201.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Risk-based maintenance backlog

Harald Rødseth
Norwegian University of Science and Technology (NTNU), Trondheim, Norway

ABSTRACT: A relevant issue in manufacturing and production seems to be "silo"-organisations and "silo"-planning with lack of coordination between departments. Integrated Planning (IPL) is a concept that aims to cope with this "silo"-problem. With the ground-breaking potentials from Industry 4.0 it should be expected that the advancement of IPL will speed up in development and implementation in companies. To manage IPL sound Key Performance Indicators (KPIs) must be implemented and established in the company. A promising indicator for IPL is Maintenance Backlog (MB). A strength of this indicator is the capability to be modelled with Risk OMT (Risk modelling—Integration of Organisational, human and Technical factors). It remains to investigate how MB can be modelled to a Quantitative Risk Analysis (QRA). The main objective of this article is to develop a model of MB in QRA. In particular the article demonstrates a case study of a production system where both Fault Tree Analysis (FTA), and Event Tree Analysis (ETA) is modelled. The article discusses the demonstration results and evaluate how potentials in Industry 4.0 can support QRA.

1 INTRODUCTION

The Oil & Gas (O&G) industry has experienced challenges with the demand of increasing oil production, lowering operating costs and life extension (Ramstad et al., 2010). These challenges have among other efforts resulted in the concept Integrated Operations (IO) and is described to be a new way of doing business (Rosendahl and Hepsø, 2013). This concept has further resulted in the Center for Integrated Operations in the petroleum industry (IO Center) where one important issue is to go from "silo organisations" towards integrated operation of all the relevant organisations (IO Center, 2012). When transferring the IO principle into the planning domain, leads us to the concept *Integrated Planning* (IPL) (Ramstad et al., 2010). In the planning domain the "silo" problem is also present.

The problem with "silo" planning in O&G industry is lack of coordination across domains and organisations (Rosendahl and Hepsø, 2013). In particular, lack of IPL results in limited resources, system failures and unscheduled maintenance (Wahl and Sleire, 2009). It is also argued that other disciplines such as drilling may affect maintenance (Sleire and Wahl, 2008). The maintenance backlog (MB) is according to Øien and Schjølberg (2009) and Meland et al. (2009) to be represented in the ageing phase for O&G facility and should be controlled at that stage. In addition, the Petroleum Safety Authority (PSA) Norway measures MB systematically for Oil companies at the Norwegian Continental Shelf (Petroleumstilsynet, 2012). In fact, according to PSA, maintenance critical backlog is regarded as a potential for major accidents.

A case study of indicators related to IPL that are applied in the O&G industry has been performed (Wahl and Sleire, 2009). Plan attainment was one type of indicator and can be related to maintenance backlog. More research for improving indicators for plan performance is concluded (Wahl and Sleire, 2009). However it is not clear in this case study how these indicators are modelled to a Quantitative Risk Analysis (QRA).

In risk modelling, the Risk OMT (Risk modelling—Integration of Organisational, human and Technical factors) model has been developed by Vinnem et al. (2012) and evaluated though a case study (Gran et al., 2012). In this model, a Bayesian belief network is applied to structure two levels of Risk Influencing Factors (RIFs) connected to the basic events in QRA. Also, the principles for updating the risk picture with a QRA-basis have been demonstrated (Vatn, 2014). The Risk OMT seems to be a promising model for a dynamic risk barometer based on indicators (Paltrinieri et al., 2017, Paltrinieri et al., 2014).

Due to different views of the term "maintenance backlog" and how it is modelled and the relevance for IPL, a novel model for MB of physical assets has been recently developed and structured in a framework for IPL (Rødseth and Schjølberg, 2017). Furthermore, the Risk OMT was tested for MB in a reliability model, demonstrating that the risk aspect is included for MB. In the Risk OMT

model proposed by Rødseth and Schjølberg (2017) the RIF structure adjusted the level of MB after evaluating the RIF of *people* such as the skills to the craft technicians and the RIF of *tools* they use in maintenance planning. It remains however to investigate how MB can be evaluated as a RIF itself and connected to the QRA.

With the potentials within Industry 4.0 it would be expected that enterprises establish Cyber Physical Systems (CPS) where the physical world and the virtual world are converging (Kagermann et al., 2013, Monostori, 2014).

To implement the potentials from Industry 4.0 an architecture for CPS should be established in the organisation (Lee et al., 2015). Nevertheless, more effort is needed to investigate more in detail how Risk OMT can be adapted to such an architecture.

The main objective of this article is to develop a model of MB in QRA. To achieve this main objective following sub-objectives have been outlined:

1. Develop a general model that connects MB with QRA.
2. Test the model with a case example.
3. Propose how the model can be improved with support from the potentials in Industry 4.0.

The remainder of this article is structured as follows: Section 2 presents the CPS as a potential in Industry 4.0, Section 3 presents an example case with a corresponding risk model developed in Section 4. The results from the example case is presented in Section 5. Further, Section 6 provides a discussion of how the results can be related to CPS where concluding remarks are made in Section 7.

2 CPS AS A POTENTIAL IN INDUSTRY 4.0

With the trend of digitalizing manufacturing, Industry 4.0 offers several promising technologies. An essential element in Industry 4.0 is convergence of the physical world and the virtual world represented in CPS (Kagermann et al., 2013). This enables network resources, information, physical assets and people to create Internet of Things (IoT) and Internet of Services.

Maintenance clearly positions in Industry 4.0 where both predictive and remote maintenance provides value creation in enterprises in terms of improved asset utilization and reduced maintenance costs (McKinsey & Company, 2015). For maintenance, the 5C architecture seems to be promising as a CPS architecture (Lee et al., 2015, Lee et al., 2017). This architecture has also been tested for manufacturing (Lee et al., 2017) and process industry (Rødseth et al., 2016). The 5C architecture forms a pyramid with following levels:

- Level 1: *Connection.* Data collection from e.g. sensors connected to machines.
- Level 2: *Conversion.* Data converted into useful information, e.g. calculations of vibration data.
- Level 3: *Cyber.* The information is connected to internet where advanced analytics can take place in terms of e.g. fleet analytics.
- Level 4: *Cognition.* To support a decision, visual interfaces such as dashboards and key performance indicators are necessary.
- Level 5: *Configuration.* Decision-making is supported through e.g. "digital advices" for conducting maintenance activities.

This architecture has also been proposed to be a sound structure for the maintenance model Deep Digital Maintenance (DDM) (Rødseth et al., 2017). DDM comprises the module planning where it remains to elaborate how MB can improve the planning function for DDM.

3 DESCRIPTION OF SYSTEM AND AN EXAMPLE CASE

In this article a heat exchanger with a barrier system is outlined in Figure 1 as an example. The system is a heat exchanger with a barrier system. The heat exchanger receives gas from two processes, process 1 and 2. If there is a leakage from the heat exchanger both valve 1 and valve 2 must close in order to avoid further leakages. Figure 2 further

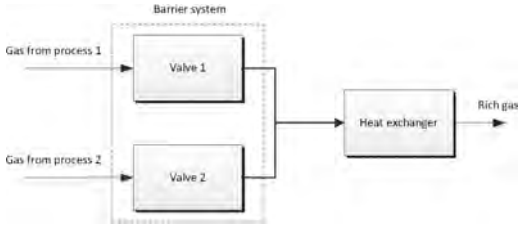


Figure 1. Illustration of a heat exchanger with a barrier system.

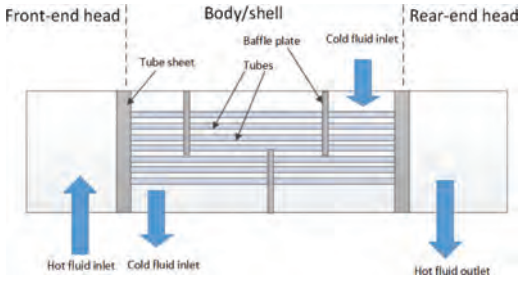


Figure 2. Illustration of a heat exchanger, adapted from (Utne et al., 2012).

describes the heat exchanger system in details (Utne et al., 2012).

This heat exchanger is a single-pass tube, baffled single-pass shell, shell-and-tube heat exchanger designed to provide counter flow conditions. In addition, this type of heat exchanger is one of the most commonly used in offshore oil and gas processing plants. In the case study we will consider an example case of the tubes in the heat exchanger where the model developed is tested with example data. The tubes constitute a bundle, meaning that no single tube can be solely replaced. If there is leakage from a tube it will be plugged. However, this will decrease the efficiency of the heat exchanger and at some point the whole bundle will be changed. This will then set the efficiency back to 100% of the design efficiency.

For the heat exchanger it is assumed that leakage only occurs from tubes and is located in the shell section. For the barrier system, only valve 1 is of interest. For the heat exchanger and the valves there exists a specific maintenance programme coordinated by the maintenance management.

4 RISK MODELLING

The core of the RISK OMT is modelling RIFs and how these affect the operational barriers. In this paper the RIFs are identified in the tasks in the maintenance programme and affect both the barriers (valves) and the production facility (heat exchanger). A RIF is defined by (Øien, 2001) to be *"an aspect (event/condition) of a system or an activity that affects the risk level of this system/activity"*. Further a RIF is a theoretical variable that can be measured.

The risk picture is illustrated in Figure 3. The initiating event in the QRA is leakage from shell and tube heat exchanger. This leakage is due to either leakage from front-end head section, rear-end head section or shell section.

In this article the shell section denoted as basic event 3 is further studied. It is assumed that the frequency of leakage from front-end head section and rear-end head section is negligible. When the initiating event occurs the barrier system shall shut down the gas supply. Both valve 1 and valve 2 must function in order to avoid a safety critical event.

The worst scenario in the QRA occurs with an impact of both a safety critical and economic critical consequence shown with the dotted circle in Figure 3. For this scenario the annual expected frequency is of interest.

Figure 4 presents the FTA for the initiating event with leakage from the shell and tube heat exchanger, while Figure 5 presents the barrier system structured with FTA. The barrier system comprises two shutdown valves where both must fail.

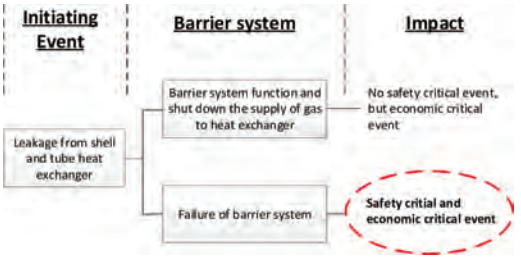


Figure 3. The total risk picture in QRA.

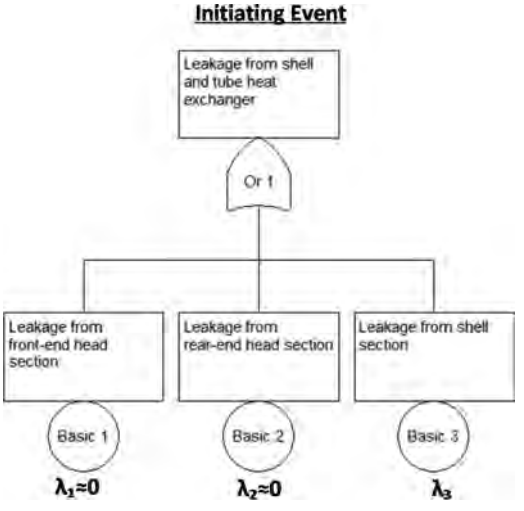


Figure 4. FTA of initialing event.

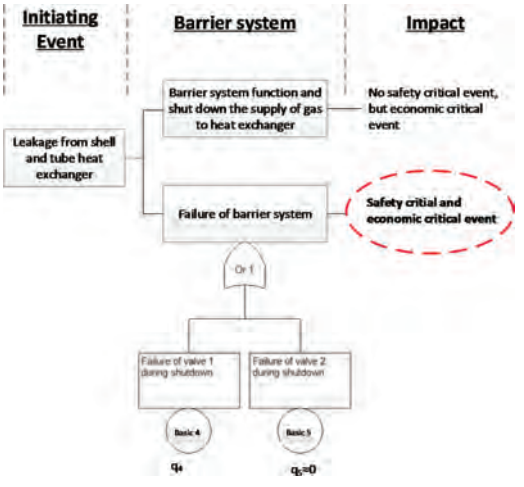


Figure 5. FTA of barrier system.

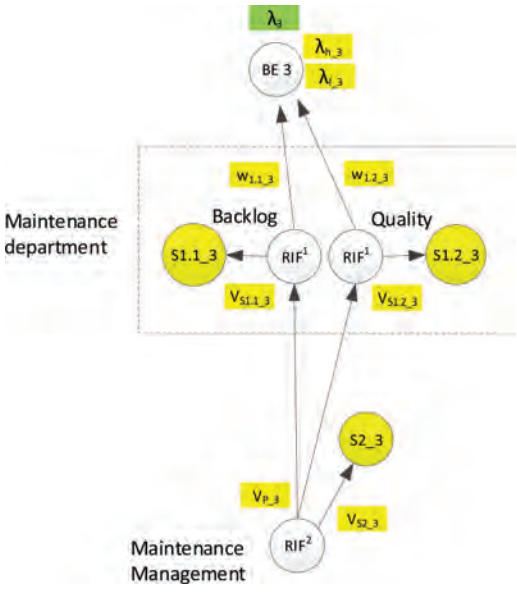


Figure 6. RIF structure that connects to QRA.

Figure 6 presents the RIF structure which is connected to the basic event 3 in the FTA from Figure 4. The same design of the RIF structure is also connected to the basic event 4 in the FTA from Figure 5. For each RIF, a score, S from A-F is observed and a variance, Vs of the score is evaluated.

The variance of the score reflects on how accurate the observation is to the "real" RIF. The RIF structure consists of two levels where there is a structural dependency Vp between level 2 parent RIF and level 1 child RIFs. At level 1 the RIFs are weighted with weight w based on expert judgment. The RIF structure is further described in this section with elaboration of Risk OMT.

4.1 *Level 2 RIF*

Maintenance management is defined as (CEN, 2010) "*all activities of the management that determine the maintenance objectives, strategies and responsibilities, and implementation of them by such means as maintenance planning, maintenance control, and the improvement of maintenance activities and economics*". This level is on the organisational level and affects the operative level 1.

4.2 *Level 1 RIFs*

4.2.1 *Maintenance backlog*

Maintenance backlog is defined as (Rødseth and Schjølberg, 2017) "*the amount of unfulfilled demands at a given point of time in explicit reference to predefined standards to be achieved. The demands comprise both demands for the technical condition itself and demand in meeting the planned due dates in the work orders. Furthermore, maintenance backlog can be expressed in functional (non-monetary) or monetary terms and it refers to single components, sub-assets or to the whole asset*".

Table 1. Score and evaluation criteria.

Score	Evaluation criteria
A	«Best case» score
B	
C	«Normal case» score
D	
E	
F	«Worst case» score

It is proposed by Rødseth and Schjølberg (2017) that maintenance backlog can have a financial perspective that is both based on work orders and the technical condition of the facility based on the understanding of maintenance backlog from petroleum authorities (Petroleumstilsynet, 2016) and road authorities (Weninger-Vycudil et al., 2009) respectively. When providing a score of maintenance backlog, both the deviation of expected wok completed from the work orders and the technical condition should be evaluated.

Table 1 presents the score and evaluation criteria for maintenance backlog.

4.2.2 *Maintenance quality*

Maintenance quality is an aggregation of all factors that affects the quality of the service provided in the maintenance task. A comprehensive list of factors that can be related to maintenance quality has been outlined (Rødseth and Schjølberg, 2017). The list of relevant factors for maintenance quality can also be based on the findings from audits (Øien et al., 2010):

1. Classification
2. Documentation
3. Use of classification
4. Competence
5. Maintenance efficiency evaluation

The evaluation of the score, variance and weight of this RIF is based on questionnaires and interviews for the assessing the "soft" issues like competence and maintenance efficiency evaluation. For the more "hard" issues like classification the observation of the score is based on direct observations in the organisation.

4.3 *Standard calculation of two level 1 RIFs and one level 2 RIF*

When calculating the basic event with the RIF structure presented in Figure 6 a mathematical approach is needed. This approach has also been presented by (Rødseth and Schjølberg, 2017). It is here included for completeness. Further details and foundations for this approach and the formulas are elaborated by (Vatn, 2013). The approach comprises 6 stages for calculating basic event q_i and can be summarized as follows:

1. Perform an expert judgement and evaluate the score of each RIF in the range of A-F, the variances, the weights w_i and values for q.
2. Map the scores into values in the interval [0,1] with following scores:
   A = [1/12], B = [3/12], C = [5/12], D = [7/12], E = [9/12], F = [11/12]. The range in the vector is then: [1/12, 3/12, 5/12, 7/12, 9/12, 11/12].
3. Calculate the posterior distribution of parents RIF based on following formulas where Jeffreys prior is used with the prior parameters $\alpha_0 = \beta_0 = 0.5$:

$$\alpha = \alpha_0 + s^2(1-s) \Big/ V_s \tag{1}$$

$$\beta = \beta_0 + s(1-s)^2 \Big/ V_s \tag{2}$$

4. Calculate the prior distributions α_0 and β_0 of child RIFs based on following equations:

$$\beta_0 = \left(\frac{p(1-p)}{V_p} - 1 \right)(1-p) \tag{3}$$

$$\alpha_0 = \frac{p\beta_0}{(1-p)} \tag{4}$$

5. Calculate the weighted sum for level 1 RIFs and the expected probability for each possible combination, *i*. All the combinations are distributed in a list.
6. Apply the law of total probability and calculate the basic event with following formula of qi:

$$q = \sum_p \left[\sum_r q_L * \left(\frac{q_H}{q_L} \right)^{\sum_j w_j * r_j} * p_{\boldsymbol{R}}(\boldsymbol{r} \mid \boldsymbol{P} = \boldsymbol{p}) \right] * p(\mathrm{p}) \tag{5}$$

4.4 *Modelling the basic events*

For the example case we assume that we have two independent RIF structures that affects the initial event and the barrier system in the QRA. In this case example we have two outsourced maintenance organisations that are specialised in maintenance for heat exchangers and valves. Further it is assumed that these organisations are independent from each other.

When λ_3 for basic event 3 is calculated, the frequency can also be calculated. Since the heat exchanger is regarded as a repairable unit following formula is used:

$$q_i \approx \lambda_i * MTTR_i \tag{6}$$

From basic event 3, the frequency of initiating event, IE, is of interest. Since we have one OR gate and neglect basic event 1 and 2, the frequency of the initial event can be calculated as follows:

$$\lambda_{IE} \approx \frac{q_1}{MTTR_1} + \frac{q_2}{MTTR_2} + \frac{q_3}{MTTR_3} \approx \frac{q_3}{MTTR_3} \tag{7}$$

5 RESULT

5.1 *Input data*

The input data is shown in Table 2 and Table 3, partly based on data from the OREDA handbook (Sintef and Oreda, 2009). The failure rates for the heat exchanger are collected from OREDA data base. The high and low values for λ_3 and q_4 is in accordance with the Risk OMT model.

5.2 *Result data*

The result from the example case is structured in Table 4. For the scenario, the frequency is calculated.

Table 2. Input date for basic event 3.

Parameter	Value
$S_{1.1\_3}$	D = 0.58333
$S_{1.2\_3}$	B = 0.25000
S_{2_3}	C = 0.41667
$w_{1.1\_3}$	0.3
$w_{1.2\_3}$	0.7
$VS_{1.1\_3}$	0.01
$VS_{1.2\_3}$	0.04
VS_{2_3}	0.04
VP_3	0.0025
$MTTR_{\_3}$ (hours)	3.0
λ_{l_3} (/hours) from (Sintef and Oreda, 2009)	$0.39*10^{-6}$
λ_{h_3} (/hours) from (Sintef and Oreda, 2009)	$23.87*10^{-6}$

Table 3. Input data for basic event 4.

Parameter	Value
$S_{1.1\_4}$	C = 0.41667
$S_{1.2\_4}$	C = 0.41667
S_{2_4}	C = 0.41667
$w_{1.1\_4}$	0.3
$w_{1.2\_4}$	0.7
$VS_{1.1\_4}$	0.01
$VS_{1.2\_4}$	0.04
VS_{2_4}	0.04
$VP_{\_4}$	0.0025
q_{h_4} (/hours)	10^{-3}
q_{l_4} (/hours)	10^{-4}

Table 4. Changes in QRA.

Initiating event, (/hours)	Barrier system, (/hours)	Frequency in QRA (/year)
$\lambda_3 = \lambda_{IE}$ $= 2.9500*10^{-6}$	q4 = 0.0030	$F_2 = \lambda_{IE}*q4$ $= \mathbf{7.75*10^{-5}}$

6 ADAPTING RISK OMT WITH CPS

The objective of this article was to develop a model of MB in QRA. With a model from Risk OMT, a structure of RIF with two levels was developed and connected to a case example of a heat exchanger. Improved maintenance quality can to some extent compensate for poor maintenance backlog. Still the organisation should strive to reduce MB in order to improve the overall risk picture of QRA of the plant.

To improve the implementation and application of the model of MB, an essential evaluation would be to what degree could this model be automated. The motivation of this automation is due to the huge amount of technical objects in a plant that is maintenance significant. Obviously some evaluations such as maintenance quality should remain manually performed by experts, but the evaluation of MB should have potential in being conducted more automatically.

Figure 7 proposes how an CPS architecture could be constructed, inspired by the 5C model from (Lee et al., 2015). At level 1 relevant data for maintenance backlog is captured in real-time from both the computerized maintenance management system and condition monitoring systems. At Level 2 the values for MB is calculated both from CMMS and condition monitoring in monetary terms in accordance to (Rødseth and Schjølberg, 2017). At level 3 the score, S is automatically evaluated based on comparing

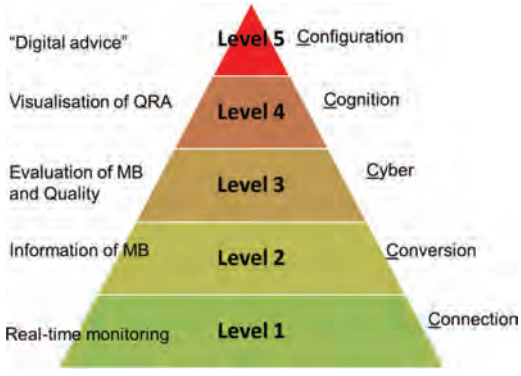


Figure 7. CPS architecture proposed for Risk OMT, inspired by the 5C model from (Lee et al., 2015).

with earlier measurements of MB. In level 4 the QRA will be visualized as a risk picture. Finally in level 5 a digital advice is provided to propose reduction of maintenance backlog if necessary.

7 CONCLUDING REMARKS

It is concluded in this article that the proposed model of MB in QRA should be further developed, in particular with respect to decision criteria for providing the score. In addition, a CPS architecture for industry should be established with respect to Risk OMT modelling and include it in the maintenance model DDM. Further studies of this model should also be performed not only in the O&G industry, but also in industry branches such as the maritime industry, manufacturing, process industry and the railway industry.

REFERENCES

CEN (2010) EN 13306: Maintenance—Maintenance terminology.

Gran, B.A., Bye, R., Nyheim, O.M., Okstad, E.H., Seljelid, J., Sklet, S., Vatn, J. & Vinnem, J.E. (2012) Evaluation of the Risk OMT model for maintenance work on major offshore process equipment. *Journal of Loss Prevention in the Process Industries,* 25, 582–593.

IO CENTER (2012) Annual Report 2012 – Center for Integrated Operations in the petroleum industry.

Kagermann, H., Wahlster, W. & Helbig, J. (2013) Recommendations for implementing the strategic initiative INDUSTRIE 4.0.

Lee, J., Bagheri, B. & Kao, H.A. (2015) A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters,* 3, 18–23.

Lee, J., Jin, C. & Bagheri, B. (2017) Cyber physical systems for predictive production systems. *Production Engineering,* 11, 155–165.

McKinsey & Company (2015) Industry 4.0 – How to navigate digitization of the manufacturing sector.

Meland, O., Schjølberg, P. & Øien, K. (2009) *Vedlikehold for aldrende innretninger: en utredning,* Trondheim, SINTEF.

Monostori, L. (2014) Cyber-physical Production Systems: Roots, Expectations and R&D Challenges. *Procedia CIRP,* 17, 9–13.

Øien, K. (2001) Risk indicators as a tool for risk control. *Reliability Engineering & System Safety,* 74, 129–145.

Øien, K. & Schjølberg, P. (2009) Kartlegging av konsekvensene for vedlikeholdsstyring ved aldring og levetidsforlengelse. Trondheim, Norway, SINTEF.

Øien, K., Schjølberg, P., Meland, O., Leto, S. & Spilde, H. (2010) Correct Maintenance Prevents Major Accidents. *Maintworld: maintenance & asset management.* Helsinki, KP-Media Oy.

Paltrinieri, N., Landucci, G. & Rossi, P.S. (2017) Real-time data for risk assessment in the offshore oil & gas industry. *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering – OMAE.*

Paltrinieri, N., Scarponi, G.E., Khan, F. & Hauge, S. (2014) Addressing dynamic risk in the petroleum industry by means of innovative analysis solutions. *Chemical Engineering Transactions.*

Petroleumstilsynet (2012) Hovedrapport – Utviklingstrekk 2012 – Norsk Sokkel. *RNNP: Risikonivå i norsk petroleumsvirksomhet.*

Petroleumstilsynet (2016) Hovedrapport – Utviklingstrekk 2016 – Norsk Sokkel. *RNNP: Risikonivå i norsk petroleumsvirksomhet.*

Ramstad, L.S., Halvorsen, K. & Wahl, A.M. (2010) Improved Coordination with Integrated Planning: Organisational Capabilities. *Society of Petroleum Engineers*.

Rosendahl, T. & Hepsø, V. (2013) *Integrated operations in the oil and gas industry,* Hershey, Pa., Business Science Reference.

Rødseth, H. & Schjølberg, P. (2017) Maintenance backlog for improving integrated planning. *Journal of Quality in Maintenance Engineering,* 23, 195–225.

Rødseth, H., Schjølberg, P. & Larsen, L.T. (2016) Industrie 4.0 – A new trend in predictive maintenance and maintenance management. *EuroMaintenance 2016 – Conference Proceedings.* Artion Conferences & Events.

Rødseth, H., Schjølberg, P. & Marhaug, A. (2017) Deep digital maintenance. *Advances in Manufacturing*.

Sintef & Oreda (2009) *OREDA: offshore reliability data handbook: Vol. 1: Topside equipment*, Trondheim, OREDA Participants.

Sleire, H. & Wahl, A.M. (2008) Integrated Planning: One road to reach Integrated Operations. *SPE Bergen 2008.* Bergen, Norway.

Utne, I.B., Brurok, T. & Rødseth, H. (2012) A structured approach to improved condition monitoring. *Journal of Loss Prevention in the Process Industries,* 25, 478–488.

Vatn, J. (2013) Risk_OMT – Hybrid approach. NTNU.

Vatn, J. (2014) Principles for dynamic updating and visualization of the risk picture with a QRA-basis. *Safety, Reliability and Risk Analysis: Beyond the Horizon—Proceedings of the European Safety and Reliability Conference, ESREL 2013*.

Vinnem, J.E., Bye, R., Gran, B.A., Kongsvik, T., Nyheim, O.M., Okstad, E.H., Seljelid, J. & Vatn, J. (2012) Risk modelling of maintenance work on major process equipment on offshore petroleum installations. *Journal of Loss Prevention in the Process Industries,* 25, 274–292.

Wahl, A.M. & Sleire, H. (2009) Measuring Performance in Offshore Maintenance. *Condition Monitoring and Diagnostics Engineering Management COMADEM 2009*.

Weninger-Vycudil, A., Litzka, J., Schiffmann, F., Lindenmann, H.P., Haberl, J., Scazziga, I., Rodriguez, M., Hueppi, A. & Jamnik, J. (2009) Maintenance Backlog estimation and Use. *Final Report.* ERA-NET ROAD (ENR).

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Reliability-based maintenance optimization for the leased equipment with deterioration depending on age and usage

Lijun Shang, Shubin Si, Zhiqiang Cai & Xianzhi Wang
School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an, Shaanxi, China

ABSTRACT: In this paper, we first consider a two-dimensional lease region that is composed of age limit and usage limit, and then from the leaser's perspective develop maintenance policies so that the total cost in the lease region can be reduced. Deterioration process of the equipment is simultaneously dependent on age and usage. When reliability drops to a threshold, called reliability threshold, Preventive Maintenance (PM) is carried out to improve reliability. This repeated PM action is ceased until the associated remaining lease limits terminate. Under the situation that the lessee's usage rate is single usage rate class, a unified maintenance policy is used to maintain reliability of the equipment. Under the situation that the lessee's usage rate is classified, customized PM integrating with usage rate class is used to maintain reliability. We derive the leaser's total cost models that are respectively related to a benchmark lease contract with the unified maintenance policy and the proposed lease contract with customized PM. By means of Root Mean Square (RMS), lease contract selection is performed. An example experiment is provided to verify the proposed methodologies. As showed in the example experiment, optimal values can be searched by minimizing the total cost and the proposed lease contract is beneficial for both the leaser and the lessee.

1 INTRODUCTION

Due to the limits of ability to maintain reliability and the lack of capital, some users (called lessees) buy only the right to use and don't buy an equipment from its owner (called leaser). The lessee uses the equipment to accomplish tasks or missions by paying rent for the leaser; and the leaser is often responsible for maintaining reliability of his equipment in a period, called lease period. In the lease period, reliability size of the equipment not only has effect on the leaser' maintenance cost but also influences the lessee's task progress. Specially mentioning, the lessee may incur some loss if the task is not accomplished on time due to a longer time to recover failure of the equipment. In this setting, the lessee usually requires a penalty cost for the leaser so that her loss is compensated. For reducing penalty cost and maintenance cost in the lease period, the leaser usually performs preventive maintenance (PM) in the lease period.

For purpose of maintenance cost saving and penalty cost saving, various PM policies have been proposed by academic researchers and industry practitioners. Jaturonnatee et al. (2006) developed a sequential PM scheme using the reduction in failure rate. Yeh et al. (2009) proposed more effective PM policies than PM policy in Jaturonnatee et al. (2006). Other PM policies related to leased equipment have offered in Pongpech & Murthy (2006), Zhou et al. (2014) Mabrouk et al. (2016), Hajej et al. (2016) and Hung et al. (2017).

Although the mentioned-above literature studied various types of PM policies related to the equipment, their implicit assumption is that aging of the equipment is uniquely characterized by the time (age)-dependent random distribution. In real life, the deterioration of some equipment is affected not just by age, but it is influenced by accumulative usage. For example, accumulative mileage of automobile can increase the probability of a failure. Considering this fact, Iskandar & Husniah (2017) recently developed a PM policy for the equipment with a two-dimensional lease region. However, this literature ignored a fact that all usage rates of individual lessee are not same. According to difference of usage rate, in practice, it is more popular with the lessee for the leaser to perform customized PM integrating with usage rate class in the two-dimensional lease region. Besides, PM in this literature was performed periodically. For the equipment with deterioration depending on age and usage, it is well known that its reliability sharply and nonlinearly declines with respect to age and accumulative usage or its deterioration is accelerated with respect to age and accumulative usage. Obviously, periodic PM actions can't be applied to effectively and in real time improve the sharply and nonlinearly declining reliability since its interval is a constant.

In this paper, we consider a two-dimensional lease region with customized PM, which is used as a lease limit of the equipment with deterioration depending on age and usage. Because the leaser

can't obtain usage rates of all individual lessees when lease contract takes effect, customized PM can't be used to maintain reliability of the equipment. In order to use customized PM in the two-dimensional lease region with the unknown usage rates, we divide the two-dimensional lease region into multiple disjoint regions and then propose a flexible lease contract with customized PM. According to repair records or/and electronic monitoring records in the first two-dimensional lease region, the leaser can obtain usage rates of all individual lessees. This acquisition is convenient to usage of customized PM in the remaining lease limits. Besides, since reliability sharply declines with respect to age and accumulative usage, the customized PM considered in this paper is based on reliability threshold, as a decision variable. That is, PM is performed depending on whether reliability declines to reliability threshold. As well known, lease termination of weighty lessee is decided by usage limit and lease termination of light lessee is decided by age limit. To measure effect of customized PM, some assume that customized PM for each weighty lessee class reduces age and customized PM for each light lessee class reduces usage. For illustrating the performance of the proposed lease contract, moreover, we compare the total cost models under different lease contracts by means of Root Mean Square (RMS) technology.

The structure of this paper is listed as follows. In Section 2, a benchmark lease contract and the lease contract proposed in this paper are defined. In Section 3, we derive the leaser's total cost models respectively related to two lease contracts. Section4 presents a numerical experiment to illustrate the considered lease contracts. Finally, Section 5 concludes this paper.

2 MODEL FORMULATION

2.1 *Assumptions*

The following assumptions are made in this paper.

- The usage rate of individual lessee is fixed through the whole life cycle;
- Usage rates after the first two-dimensional lease region are obtained and they are classified into $\boldsymbol{m}$ ($m>1$) classes.

2.2 *Lease contract definition*

2.2.1 *Benchmark lease contract*

In real life, minimal repair frequently is used to remove failure. In this paper, we call the two-dimensional lease contract where all failures are removed by minimal repair as the benchmark lease contract. Under the benchmark lease contract, since minimal repair is unique method to sustain running, we call minimal repair as the unified maintenance policy. Due to the unknown usage rate in the two-dimensional lease region Ω, age and accumulative usage at the two-dimensional lease region Ω termination are respectively expressed as

$$T_{\Omega}(t)=\begin{cases}\varpi, & if\ \ r\le r^*\\ u/r, & if\ \ r>r^*\end{cases} \text{ and } X_{\Omega}(t)=\begin{cases}r\varpi, & if\ \ r\le r^*\\ u, & if\ \ r>r^*\end{cases}.$$

2.2.2 *The proposed lease contract*

Let ϖ be maximal time limit, u be maximal usage limit and r^* be average usage rate. Then $r^*=u/\varpi$ and the two-dimensional lease region can be defined by the set $\Omega=(\varpi,u)$. For the sake of analytical tractability, we define the first two-dimensional lease region as $\Omega_1=\{(\omega_1,u_1)\,|\,0\le u_1<\varpi r_{\min} \text{ and } 0\le \omega_1<u/r_{\max}\}$ where $r_{\min}$ is minimal usage; $r_{\max}$ is maximal usage; and $u_1=r^*\varpi_1$. Obviously, the first two-dimensional lease region Ω_1 is a part of the two-dimensional lease region Ω, which can be depicted in Figure 1.

As showed in Figure 1, the two-dimensional lease region Ω is classified three regions. According to repair records in the first two-dimensional lease region Ω_1, the leaser can conclude usage rate of individual lessee and further classify usage rate for them. So the leaser can obtain the following information when the equipment with the unknown usage rate class goes through the first two-dimensional lease region Ω_1, as follows: a) the usage rate class can be known; and b) the remaining lease limits, namely, the remaining time limit $\varpi-\varpi_1>0$ and the remaining usage limit $u-u_1>0$, can be known.

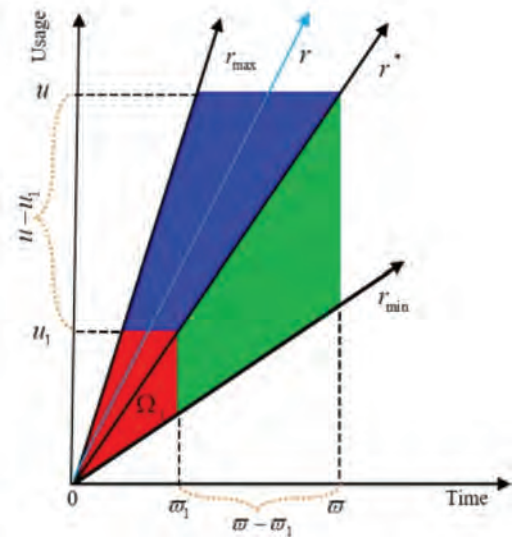


Figure 1. Description of lease region.

Further, age and accumulative usage at the first two-dimensional lease region Ω_1 termination are respectively given by

$$T_{\Omega_1}(t)=\begin{cases}\varpi_1, & if\ r\le r^*\\ u_1/r, & if\ r>r^*\end{cases} \text{ and } X_{\Omega_1}(t)=\begin{cases}r\varpi_1, & if\ r\le r^*\\ u_1, & if\ r>r^*\end{cases}.$$

Reliability-centered (or based) PM policy is a nice maintenance policy, which is frequently applied in reliability maintenance field (see Niu et al., 2010, Zhou et al., 2007, Lin et al., 2015). From the leaser's perspective, here we propose a two-dimensional lease contract with customized PM by integrating reliability-centered PM policy with usage rate classes. For this lease contract, when reliability in the remaining lease limits declines to a threshold, called reliability threshold, then PM is performed to improve reliability. This repeated PM action is ceased until the associated remaining lease limit expires. That is, customized PM is applied to keep reliability no lower than reliability threshold, in the remaining lease limits. For the given remaining lease limits, a smaller reliability threshold must result in a lower number of PMs, and vice versa. Under the proposed lease contract, therefore, both reliability threshold and number of PMs are considered as decision variables.

3 MODEL ANALYSIS

Failure rate (function), as an increasing function, is generally used to characterize aging. In this paper, we assume that aging is modeled by a failure rate given by $\gamma(x|r)$ where r is a random variable following $G(r)$. The total cost under each lease contract is the sum of the expected maintenance cost, the expected penalty cost and the expected depreciation cost, i.e.,

$$\begin{aligned}\textit{The total cost} = {} & \textit{maintenance cost} + \textit{penalty cost}\\ & + \textit{depreciation cost.}\end{aligned} \tag{1}$$

By (1), next, we will derive the associated total cost under the case that c_m is minimal repair cost.

3.1 *Cost model under the unified maintenance*

Under the unified maintenance policy, the equipment is not classified. That is, all equipment have a same usage rate class $r\in[r_{\min},r_{\max}]$. We call this same usage rate class as single usage rate class in this paper.

3.1.1 *The expected maintenance cost*

The expected maintenance cost of single usage rate class can be obtained as

$$MC_b^s = c_m\left(\int_{r^*}^{r_{\max}}\int_0^{u/r}\gamma(x|r)\,\mathrm{d}x\mathrm{d}G(r)+\int_{r_{\min}}^{r^*}\int_0^{w}\gamma(x|r)\,\mathrm{d}x\mathrm{d}G(r)\right), \tag{2}$$

3.1.2 *The expected penalty cost*

From the lessee's perspective, essentially, recovery of the equipment failure can cause revenue loss. Hamidi *et al.* (2016) formulated the revenue rate as the product of an increasing function in r and a decreasing function in t. Similarly, we also use this product to describe the revenue rate of single class, as

$$\mathbb{R}_s(t\,|\,r)=\frac{\mathbb{R}_{\max}}{r_{\max}}r\left(1-\frac{t}{L}\right), \tag{3}$$

where $\mathbb{R}_{\max}$ be the potential revenue rate that can be generated by the new equipment if its capacity is fully used (i.e., $\mathbb{R}_s(0\,|\,r_{\max})$); and L is life cycle of the equipment.

If $r\le r^*$, then the expected revenue rate of single usage rate class in the interval $(0,\varpi]$ is given by $\mathbb{R}_s=\int_{r_{\min}}^{r^*}\int_0^{\varpi}\mathbb{R}_s(t\,|\,r)\mathrm{d}t\mathrm{d}G(r)/\varpi$. If $r>r^*$, then the expected revenue rate of single usage rate class in the interval $(0,u/r]$ is given by $\mathbb{R}_s^u=\int_{r^*}^{r_{\max}}r\int_0^{\varpi}\mathbb{R}_s(t\,|\,r)\mathrm{d}t\mathrm{d}G(r)/u$. Further, the expected penalty cost of single usage rate class in the two-dimensional lease region Ω can be expressed as

$$\begin{aligned}PC_b^s = \bar{H}(\vartheta)\cdot\Big(&\mathbb{R}_s^u\int_{r^*}^{r_{\max}}\int_0^{u/r}\gamma(x|r)\,\mathrm{d}t\mathrm{d}G(r)\\ &+\mathbb{R}_s\int_{r_{\min}}^{r^*}\int_0^{\varpi}\gamma(x|r)\,\mathrm{d}t\mathrm{d}G(r)\Big),\end{aligned} \tag{4}$$

where the function $\bar{H}(\vartheta)$ is the expected excess of recovery time to failure.

3.1.3 *The expected depreciation cost*

The equipment in the two-dimensional lease region Ω faster deteriorates with respect to age and accumulative usage. This process makes the equipment depreciate. In order to measure depreciation value of the equipment at the two-dimensional lease region Ω termination, we offer a measure model by modifying residual value model in Hamidi *et al.* (2016), as

$$DC_s(r,t)=\theta_1X^2(t)+\theta_2T^2(t) \tag{5}$$

where $\theta_1>0;\ \theta_2>0;\ \theta_2=(V_0-V_L)/\varpi^2-\theta_1r_{i+1}^2;$ V_0 is purchase price; and V_L is residual value at the end of the life cycle L when the equipment is used at the maximum usage rate.

If $r\le r^*$, then the expected depreciation cost of single usage rate class at the two-dimensional lease region Ω termination is given by

$DC_b^l = \theta_1 \varpi^2 \int_{r^*}^{r_{\max}} r^2 \mathrm{d}G(r) + \varpi^2$. If $r > r^*$, then the expected depreciation cost of single usage rate class at the two-dimensional lease region Ω termination is given by $DC_b^w = \theta_1 u^2 + \theta_2 \int_{r_{\min}}^{r^*} (u/r)^2 \mathrm{d}G(r)$. Thus, the expected depreciation cost of single usage rate class at the two-dimensional lease region Ω termination is given by

$$\begin{aligned} DC_b^s &= DC_b^w + DC_b^l \\ &= \theta_1 \left(u^2 + \varpi^2 \int_{r^*}^{r_{\max}} r^2 \mathrm{d}G(r) \right) + \\ &\quad \theta_2 \left(\int_{r_{\min}}^{r^*} (u/r)^2 \mathrm{d}G(r) + \varpi^2 \right). \end{aligned} \tag{6}$$

3.1.4 *The total cost model*

By (1), finally, the total cost of single usage rate class in the lease Ω can be obtained as

$$\begin{aligned} TC_b^s &= MC_b^s + PC_b^s + DC_b^s \\ &= \left(c_m + \bar{H}(\vartheta) \mathbb{R}_s^w \right) \int_{r^*}^{r_{\max}} \int_0^{u/r} \gamma(x|r) \mathrm{d}x \mathrm{d}G(r) + \\ &\quad \left(c_m + \bar{H}(\vartheta) \mathbb{R}_s^l \right) \int_{r_{\min}}^{r^*} \int_0^{w} \gamma(x|r) \mathrm{d}x \mathrm{d}G(r) + \\ &\quad \theta_1 \left(u^2 + \varpi^2 \int_{r^*}^{r_{\max}} r^2 \mathrm{d}G(r) \right) + \\ &\quad \theta_2 \left(\int_{r_{\min}}^{r^*} (u/r)^2 \mathrm{d}G(r) + \varpi^2 \right). \end{aligned} \tag{7}$$

3.2 *Cost model under the customized PM*

As indicated in the definition of the proposed lease, customized PM is used to maintain equipment reliability in the remaining lease limits. Next, so we derive the expected total costs in the remaining lease limits. We state that the expected total cost of light lessee is derived in this subsection, and the expected total cost of weighty lessee will be presented in Appendix A.

We use the reduction in age as a PM effect. As well known, the relationship between reliability and failure rate is a one-to-one mapping. Let R_i be reliability threshold corresponding to the i^{th} usage rate class (similarly hereinafter) $r \in (r_i, r_{i+1}]$, then R_i can be expressed as

$$R_i = \exp\left\{ -\int_{r_i}^{r_{i+1}} \left(\int_{W_{k-1}^i}^{W_{k-1}^i + T_k^i} \gamma(\varpi_1 + S_{k-1}^i + x|r) \mathrm{d}x \right) \mathrm{d}G(r) \right\}, \tag{8}$$

where $H_k^i(-\ln R_i)$ is inverse function of $\boldsymbol{H}_k^i(T_k^i) = \int_{r_i}^{r_{i+1}} \left(\int_{\Gamma_{k-1}^i}^{\Gamma_{k-1}^i + T_k^i} \gamma(\varpi_1 + S_{k-1}^i + x|r) \mathrm{d}x \right) \mathrm{d}G(r)$; T_i^k is a time interval between the $(k-1)^{\text{th}}$ PM and the k^{th} PM; $S_{k-1}^i = \sum_{j=0}^{k-1} \alpha_j^i T_j^i$; $\Gamma_{k-1}^i = \sum_{j=0}^{k-1} T_j^i$; $1 \geq \alpha_1^i \geq \alpha_2^i \geq \cdots \geq 0$; and $T_0^i = S_0^i = \Gamma_0^i = 0$.

Since lease of light lessee terminates only when age reaches age limit, customized PM of light lessee is only confined to the remaining time limit $\varpi - \varpi_1$. Next, so we derive the total cost in the remaining time limit $\varpi - \varpi_1$.

3.2.1 *The expected maintenance cost*

Since each failure between two successive PMs is removed by minimal repair, the expected maintenance cost in the k^{th} PM interval $(0, T_k^i]$ is minimal repair cost in the k^{th} PM interval $(0, T_k^i]$. So, it can be calculated as

$$C_k^i(T_k^i) = c_m \int_{r_i}^{r_{i+1}} \left(\int_{\Gamma_{k-1}^i}^{\Gamma_{k-1}^i + T_k^i} \gamma(\varpi_1 + S_{k-1}^i + x|r) \mathrm{d}x \right) \mathrm{d}G(r). \tag{9}$$

After the last PM, i.e., the n_i^{th}, minimal repair cost of the i^{th} rate class is given by

$$C_{n_i}^i = c_m \int_{r_i}^{r_{i+1}} \left(\int_{\Gamma_{n_i}^i}^{\varpi - \varpi_1} \gamma(\varpi_1 + S_{n_i}^i + x|r) \mathrm{d}x \right) \mathrm{d}G(r), \tag{10}$$

where $n_i = \{ n_i > 0 | \Gamma_{n_i}^i \leq \varpi - \varpi_1 \}$.

Let the increasing function $C_{p_i}^k(\beta_k^i T_k^i)$ be PM cost resulted from the reduction $\beta_k^i T_k^i$, where $\beta_k^i = 1 - \alpha_k^i$. The expected maintenance cost in the remaining time limit $\varpi - \varpi_1$ is equal to the sum of minimal repair cost $\sum_{k=1}^{n_i - 1} C_k^i(T_k^i)$ resulted from the first $n_i - 1$ PM intervals, the minimal repair cost $C_{n_i}^i$ after the n_i^{th} PM and the PM cost $\sum_{k=1}^{n_i} C_{p_i}^k(\beta_k^i T_k^i)$ resulted from n_i PMs. And it can be mathematically expressed as

$$\begin{aligned} MC_i(R_i, n_i) &= \sum_{k=1}^{n_i - 1} C_k^i(T_k^i) + C_{n_i}^i + \sum_{k=1}^{n_i} C_{p_i}^k(\beta_k^i T_k^i) \\ &= c_m \left(\sum_{k=1}^{n_i - 1} \int_{r_i}^{r_{i+1}} \left(\int_{\Gamma_{k-1}^i}^{\Gamma_{k-1}^i + T_k^i} \gamma(\varpi_1 + S_{k-1}^i + x|r) \mathrm{d}x \right) \right. \\ &\quad \left. \mathrm{d}G(r) + \int_{r_i}^{r_{i+1}} \left(\int_{\Gamma_{n_i}^i}^{\varpi - \varpi_1} \gamma(\varpi_1 + S_{n_i}^i + x|r) \mathrm{d}x \right) \mathrm{d}G(r) \right) \\ &\quad + \sum_{k=1}^{n_i} C_{p_i}^k(\beta_k^i T_k^i). \end{aligned} \tag{11}$$

3.2.2 *The expected penalty cost*

We modify the revenue rate in (3) as

$$\mathbb{R}_p^i(t|r) = \frac{\mathbb{R}_{i+1}}{r_{i+1}} r \left(1 - \frac{t}{L}\right), \tag{12}$$

where $\mathbb{R}_{i+1}$ be the potential revenue rate that can be generated by the new equipment with the usage rate $r \in (r_i, r_{i+1}]$ if its capacity is fully used (i.e., $\mathbb{R}_p^i(0 | r_{i+1})$).

The expected revenue rate in the k^{th} PM interval $(0, T_k^i]$ is given by $\mathbb{R}_k^i = (\int_{r_i}^{r_{i+1}} \int_{\Gamma_{k-1}^i}^{\Gamma_{k-1}^i + T_k^i} \mathbb{R}_p^i(\varpi_1 + S_{k-1}^i + x \mid r) \mathrm{d}x \mathrm{d}G(r)) / T_k^i$. And the expected revenue rate after the n_i^{th} PM is given by $\mathbb{R}_{n_i}^i = \int_{r_i}^{r_{i+1}} \int_{\Gamma_{n_i}^i}^{\varpi - \varpi_1} \mathbb{R}_p^i(\varpi - \varpi_1 + S_{n_i}^i + x \mid r) \, \mathrm{d}x \mathrm{d}G(r) / (\varpi - \varpi_1 - \Gamma_{n_i}^i)$. Further, the expected penalty cost in the remaining time limit $\varpi - \varpi_1$ can be expressed as

$$\begin{aligned} PC_s(R_i, n_i) = & \bar{H}(\vartheta) \\ & \left(\sum_{k=1}^{n_i - 1} \mathbb{R}_k^i \int_{r_i}^{r_{i+1}} \left(\int_{\Gamma_{k-1}^i}^{\Gamma_{k-1}^i + T_k^i} \gamma(\varpi_1 + S_{k-1}^i + x \mid r) \mathrm{d}x \right) \mathrm{d}G(r) \right. \\ & \left. + \mathbb{R}_{n_i}^i \int_{r_i}^{r_{i+1}} \left(\int_{\Gamma_{n_i}^i}^{\varpi - \varpi_1} \gamma(\varpi_1 + S_{n_i}^i + x \mid r) \mathrm{d}x \right) \mathrm{d}G(r) \right). \end{aligned} \tag{13}$$

3.2.3 *The expected depreciation cost*

When the two-dimensional lease region Ω terminates, accumulative usage and age of the equipment is respectively $X_\Omega(t) = r\varpi$ and $T_\Omega(t) = \varpi - \sum_{k=1}^{n_i} \beta_k^i T_k^i$. In this setting, by (5), the expected depreciation cost is given by

$$DC_i^\Omega(R_i, n_i) = \theta_1 \varpi^2 \int_{r_i}^{r_{i+1}} r^2 \mathrm{d}G(r) + \theta_2 (\varpi - \sum_{k=1}^{n_i} \beta_k^i T_k^i)^2. \tag{14}$$

Similarly, the expected depreciation cost at the first two-dimensional lease region Ω_1 termination is given by

$$DC_i^{\Omega_1} = \theta_1 \omega_1^2 \int_{r_i}^{r_{i+1}} r^2 \mathrm{d}G(r) + \theta_2 \omega_1^2. \tag{15}$$

Thus, the expected depreciation cost at the remaining time limit $\varpi - \varpi_1$ termination is obtained as

$$\begin{aligned} DC_i(R_i, n_i) &= DC_i^\Omega(R_i, n_i) - DC_i^{\Omega_1} \\ &= \theta_1 (\varpi^2 - \omega_1^2) \int_{r_i}^{r_{i+1}} r^2 \mathrm{d}G(r) + \\ & \theta_2 \left((\varpi - \sum_{k=1}^{n_i} \beta_k^i T_k^i)^2 - \omega_1^2 \right). \end{aligned} \tag{16}$$

3.2.4 *The total cost model*

By (1), finally, the total cost in the remaining time limit $\varpi - \varpi_1$ can obtained as

$$\begin{aligned} TC_i(R_i, n_i) &= MC_i(R_i, n_i) + PC_s(R_i, n_i) + DC_i(R_i, n_i) \\ &= \left(c_m + \bar{G}(\vartheta) \mathbb{R}_{n_i}^i \right) \int_{r_i}^{r_{i+1}} \int_{\Gamma_{n_i}^i}^{\varpi - \varpi_1} \gamma(\varpi_1 + S_{n_i}^i + x \mid r) \\ & \mathrm{d}x \mathrm{d}G(r) + \sum_{k=1}^{n_i} C_{p_i}^k (\beta_k^i T_k^i) + \theta_1 (\varpi^2 - \omega_1^2) \\ & \int_{r_i}^{r_{i+1}} r^2 \mathrm{d}G(r) + \theta_2 \left((\varpi - \sum_{k=1}^{n_i} \beta_k^i T_k^i)^2 - \omega_1^2 \right). \end{aligned} \tag{17}$$

Our target is to seek the optimal reliability threshold R_i^* and the optimal number n_i^* of PMs so that the total cost in (17) is minimized. As mentioned earlier, for a given remaining time limit $\varpi - \varpi_1$, if the reliability threshold R_i is smaller, then number n_i of PMs is lower; if the reliability threshold R_i is bigger, then the number n_i of PMs is greater. These monotonous regularities mean that once the optimal reliability threshold R_i^* is determined, then the optimal number n_i^* of PMs can be uniquely determined. Since expressions of $G(r)$ and $\gamma(\varpi_1 + S_{k-1}^i + x \mid r)$ are uncertain, obtaining analytical solution is clearly difficult and so the optimal values (R_i^*, n_i^*) need to be searched numerically by masterly designing reliability in the remaining lease limits.

3.3 *Improvement evaluation and contract selection*

How to select lease contract is an important problem for the leaser. In practice, a key approach of lease contract selection is to compare the total costs under different lease contracts. In this subsection, we consider a qualitative analysis method to select lease contract.

As mentioned in Assumption, usage rates are classified into m classes. In the remaining lease limits (including the remaining time limit and the remaining usage limit), finally, the minimizing total cost of m usage rate classes can be obtained as

$$TC_2^T(\boldsymbol{R}^*, \boldsymbol{N}^*) = \sum_{i=1}^{m} TC_i(R_i^*, n_i^*), \tag{18}$$

where $\boldsymbol{R}^*$ is a vector composed of the optimal reliability thresholds and it is given by $\boldsymbol{R}^* = \{R_1^*, R_2^*, \cdots, R_m^*\}$; and $\boldsymbol{N}^*$ is a vector composed of the optimal number of PMs and it is given by $\boldsymbol{N}^* = \{n_1^*, n_2^*, \cdots, n_m^*\}$.

The total cost in (7) is the total cost of single usage rate class in the two-dimensional lease region Ω, whereas the total cost in (18) is the minimizing total cost of m usage rate classes in the remaining lease limits. This means that they are not equivalent, and so they are not compared directly. In order to compare them, we must transform the minimizing total cost of m usage rate classes into the total cost of single usage rate class. By root mean square (RMS), the total cost of m usage rate classes in the remaining lease limits can be approximately transformed as

$$TC_2^s(\boldsymbol{R}^*, \boldsymbol{N}^*) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \left(TC_i(R_i^*, n_i^*) \right)^2}. \tag{19}$$

Besides, by (7), the total cost of single usage rate class in the first two-dimensional lease region Ω_1 can be obtained as

$$TC_1^s = \left(c_m + \bar{H}(\vartheta)\Re_s^w\right)\int_{r^*}^{r_{\max}}\int_0^{u_1/r}\gamma(x|r)\,\mathrm{d}x\mathrm{d}G(r) + \left(c_m + \bar{H}(\vartheta)\Re_s^l\right)\int_{r_{\min}}^{r^*}\int_0^{\varpi_1}\gamma(x|r)\,\mathrm{d}x\mathrm{d}G(r) + \theta_1\left(u_1^2 + \varpi_1^2\int_{r^*}^{r_{\max}} r^2\mathrm{d}G(r)\right) + \theta_2\left(\int_{r_{\min}}^{r^*}(u_1/r)^2\mathrm{d}G(r) + \varpi_1\right), \quad (20)$$

where $\Re_s^w(=\int_{r^*}^{r_{\max}} r\int_0^{\varpi_1}\mathbb{R}_s(t\,|\,r)\mathrm{d}t\mathrm{d}G(r)/u_1)$ is the expected revenue rate of single usage rate class in the interval $(0,u_1/r]$; $\Re_s^l(=\int_{r_{\min}}^{r^*}\int_0^{\varpi_1}\mathbb{R}_s(t\,|\,r)\mathrm{d}t\mathrm{d}G(r)/\varpi_1)$ is the expected revenue rate of single usage rate class in $(0,\varpi_1]$.

Under the proposed lease contract, further, the total cost of single usage rate class in the whole two-dimensional lease region Ω can be calculated as

$$TC_p^s(\boldsymbol{R}^*,\boldsymbol{N}^*) = TC_1^s + TC_2^s(\boldsymbol{R}^*,\boldsymbol{N}^*). \quad (21)$$

Clearly, the total cost in (21) is the total cost of single usage rate class in the two-dimensional lease region, and it is equivalent to the total cost in (7). This equivalency is facilitated to select lease contract by comparing the total costs. We neglect the error produced by RMS and consider two performance measures, as follows.

If $TC_b^s > TC_p^s(\boldsymbol{R}^*,\boldsymbol{N}^*)$ or $TC_b^s < W_p^s(\boldsymbol{R}^*,\boldsymbol{N}^*)$, then the leaser should select the proposed lease contract with optimal customized PM or the benchmark lease contract to maintain reliability of the equipment in the two-dimensional lease region Ω. Under this situation, the relative improvement of the total cost can be measured as

$$\xi_p = \left|\frac{TC_p^s(\boldsymbol{R}^*,\boldsymbol{N}^*) - TC_b^s}{TC_p^s(\boldsymbol{R}^*,\boldsymbol{N}^*)}\right|\times 100\% \quad \text{or} \quad \left|\frac{TC_b^s - TC_p^s(\boldsymbol{R}^*,\boldsymbol{N}^*)}{TC_b^s}\right|\times 100\%. \quad (22)$$

4 NUMERICAL EXPERIMENT

The lease region considered by us possesses the time limit and usage limit. This is consistent with vehicle warranty in real life. Huang *et al.* (2017) recently utilized the data collected from a car dealer to demonstrate the effectiveness of the proposed policy. Similarly, the intensity function is used as a failure rate that reflects the interdependence among the age and usage, and it is given by $\gamma(x|r) = 0.1 + 0.4r + (1.5x + 1.5r)x$. Also, we assume that the random usage rate r is subject to the uniform distribution $G(r)$ with the maximum $r_{\max} = 8$ and the minimum $r_{\min} = 0.5$. Shang *et al.* (2016) used a quadratic function to model the PM cost function regarding maintenance degree. We use power function to model PM cost and it is given by $C_{p_k}^k(\bullet) = c(\bullet)^\rho$, where c, $\rho > 0$ and dot • is the reduction in age or usage due to PM.

We specify $\varpi = 2$ (years) and $u = 4$ (10^4 kilometers), then the average usage rate $r^* = 2$ and further the two-dimensional lease region is given by Ω = (2,4). The first two-dimensional lease region includes the following three cases, as $\Omega_1 = (0.3, 0.6)$, (0.4, 0.8) or (0.45,0.9). We classify all lessees into two classes, light lessee and weighty lessee. In practice, revenue rate is increasing with usage rate. Due to this fact, we specify the potential revenue rate as \$12500. Some parameters are constants throughout the whole numerical experiment, which are listed in the Table 1. In order to analyze sensitivity, we present some figures, which are used to describe variation tendency related to parametric variation.

For light lessee, since reliability at the first two-dimensional lease region Ω_1(=(0.45,0.9)) is equal to 0.8875, we take step length as 0.1 and obtain Table 2. As indicated in Table 2, the optimal values (i.e., the optimal reliability threshold and the optimal number of PMs) are existed for light lessee. The existence of the optimal values is visually and vividly displayed in Figure 2. For weighty lessee, besides, reliability at the first two-dimensional lease region Ω_1 is equal to 0.6553. Since it is less than reliability 0.8875 of light lessee, we take a lower step length as 0.05 and obtain Table 3. From the Table 3, we can see that the optimal values are existed for weighty lessee. Similarly, this existence is also visually and vividly displayed in Figure 3. In order to illustrate the performance of the considered two-dimensional lease contracts, we plot Figure 4, where step length is 0.05. From Figure 4, we can obtain the following results.

a. The total cost under the proposed lease (PL) contract are increasing and tends to be equal to total cost under the benchmark lease (BL) contract when the area of the first two-dimensional lease region gets bigger.
b. The relative improvement of the total cost under PL contract is decreasing when the area of the first total region gets bigger.

The bigger area of the first two-dimensional lease region Ω_1 produces the smaller the remaining lease limits. Further, the smaller the remaining lease limits leads to the lower PM count. Furthermore, the

Table 1. Parameter settings.

$\alpha_k^i = \rho$	L	c	c_m	$\bar{H}(\vartheta)$	V_0	V_L	θ_1	θ_2
0.5	10	100	25	2×10^{-4}	2×10^4	2×10^3	1	0.8

Table 2. The existence of optimal value for light lessee.

Values	Reliability threshold							
	0.8875	0.7875	0.6875	0.5875	0.4875	0.3875	0.2875	0.1875
N	1	1	1	1	1	2	5	17
TC_p^s	2137.6	828.3	519	389.4	318.5	229	284.1	398.9

Table 3. The existence of optimal value for weighty lessee.

Values	Reliability threshold						
	0.6553	0.6053	0.5553	0.5053	0.4553	0.4053	0.3053
N	1	2	3	4	6	10	21
TC_p^s	289.8022	199.2240	221.4222	248.8676	290.5435	355.2092	469.8107

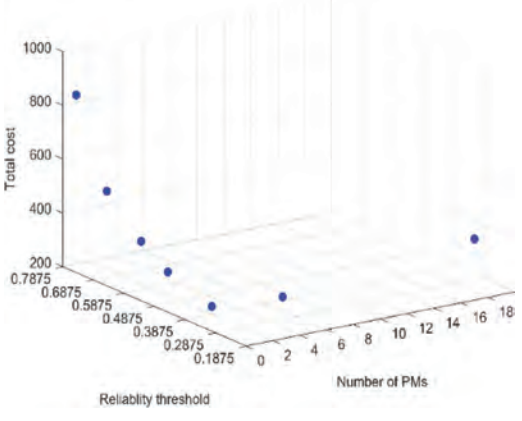


Figure 2. Total cost versus decision variables for light lessee.

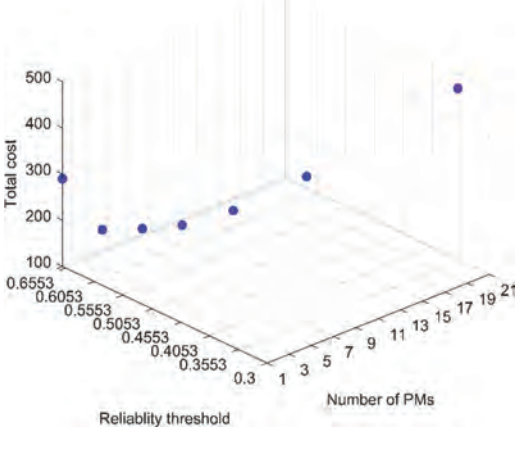


Figure 3. Total cost versus decision variables for weighty lessee.

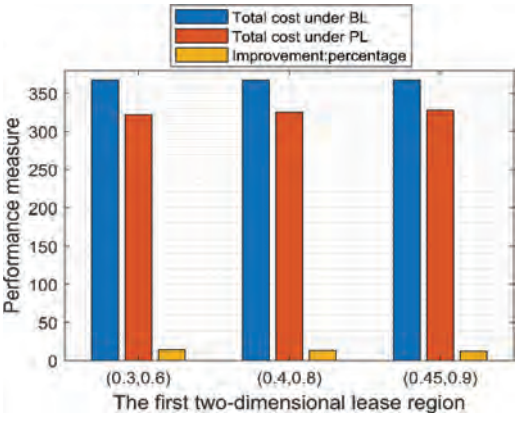


Figure 4. Total cost versus the first two-dimension.

lower PM count leads to the lower minimal repair cost reduction and the greater depreciation cost, thus the results presented in a) is obvious. Since the total costs under the two types of contracts tends to be same when the area of the first region Ω_1 approaches the area of the two-dimensional lease region Ω, the relative improvement of the total cost under PL contract is decreasing. Therefore, the results presented in b) is obtained. These changes illustrate the performance of PL contract is superior to that of BL contract. Compared with BL contract, conclusively, the leaser should use PL contract as a reliability guarantee in the two-dimensional lease region, which is more beneficial since the leaser incurs the lower total cost. Compared with BL contract, additionally, the lessee may be also inclined to PL contract since PMs in the PL contract can keep a higher reliability which can slow down deterioration and indirectly reduce number of failures and increase revenue rate.

5 CONCLUSIONS

In this paper, the two-dimensional lease contract in which all failures in the two-dimensional lease region were removed by minimal repair was taken as a benchmark lease contract. From the leaser's perspective, we proposed a two-dimensional lease contract with customized PM by dividing the two-dimensional lease region. We specified that minimal repair was used to remove failure in the first two-dimensional lease region and customized PM integrating with usage rate class was adopted to improve reliability in the remaining lease limits so that maintenance cost, penalty cost and depreciation cost can be reduced simultaneously. From viewpoint of generality, the proposed two-dimensional lease contract can be translated into benchmark lease contract by adjusting parameters. From viewpoint of management, the proposed lease contract is beneficial for both the leaser and the lessee since the total cost for the leaser and the failure counts for the lessee are less that under the benchmark lease contract. These have been illustrated by the numerical results.

The main contribution of this paper is that the proposed two-dimensional contract with the divided regions is a solution of the leaser how to use customized PM to maintain reliability, under the case that usage rate classes in the two-dimensional lease region can't be known when the lease contract takes effect. This method still can be applied to maintain reliability in the two-dimensional warranty region although it was considered from viewpoint of the two-dimensional lease contract.

ACKNOWLEDGEMENTS

This paper is funded by the National Natural Science Foundation of China (Nos. 71771186, 71471147, 71631001) and the 111 Project (No. B13044).

REFERENCES

Hajej, Z., Rezg, N. & Ali, G. 2016. An optimal production/maintenance strategy under lease contract with warranty periods. *Journal of Quality in Maintenance Engineering* 22(1): 35–50.

Hamidi, M., Liao, H. & Szidarovszky, F. 2016. Non-cooperative and cooperative game theoretic models for usage-based lease contracts. *European Journal of Operational Research* 255: 163–174.

Huang, Y.-S., Huang, C.-D. & Ho J.-Y. 2017. A customized two-dimensional extended total with preventive maintenance. *European Journal of Operational Research* 257: 971–978.

Hung, W.-H., Tsai, T.-R. & Chang, Y.-C. 2017. Periodical preventive maintenance contract for leased equipment with random failure penalties. *Computers & Industrial Engineering* 113: 437–444.

Iskandar, B.P. & Husniah, H. 2017. Optimal preventive maintenance for a two dimensional lease contract. *Computers & Industrial Engineering* 113: 693–703.

Jaturonnatee J., Murthy, D.N.P. & Boondiskulchok R. 2006. Optimal preventive maintenance of leased equipment with corrective minimal repairs. *European Journal of Operational Research* 174: 201–215.

Lin, Z.-L., Huang, Y.-S. & Fang, C.-C. 2015. Non-periodic preventive maintenance with reliability thresholds for complex repairable systems. *Reliability Engineering and System Safety* 136: 145–156.

Mabrouk, B., Chelbi, A. & Radhoui, M. 2016. Optimal imperfect maintenance strategy for leased equipment. *International Journal of Production Economics* 178: 57–64.

Niu, G., Yang, B.-S. & Pecht, M. 2010. Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance. *Reliability Engineering and System Safety* 95: 786–796.

Pongpech, J. & Murthy, D.N.P. 2006. Optimal periodic preventive maintenance policy for leased equipment. *Reliability Engineering & System Safety* 91: 772–777.

Shang, L., Si, S. & Cai, Z. 2016. Optimal maintenance–replacement policy of products with competing failures after expiry of the warranty. *Computers and Industrial Engineering* 98: 68–77.

Wang, Y., Liu, Y., Liu, Z. & Li, X. 2017. On reliability improvement program for second-hand products sold with a two-dimensional warranty. *Reliability Engineering and System Safety* 167:452–463.

Yeh, R.H., Kao, K.C. & Chang, W.I. 2009. Optimal preventive maintenance policy for leased equipment using failure rate reduction. *Computers & Industrial Engineering* 57(1): 304–309.

Zhao. X., Li. Y., Xi, L. & Lee, J. 2015. Multi-phase preventive maintenance policy for leased equipment. *International Journal of Production Research* 53(15): 4528–4537.

Zhou, X., Xi, L. & Lee, J. 2007. Reliability-centered predictive maintenance scheduling for a continuously monitored system subject to degradation. *Reliability Engineering and System Safety* 92: 530–534.

APPENDIX A

For weighty lessee, the lease expires only when usage reaches the usage limit u, so we use the reduction in usage to measure effect of customized PM.

The reliability threshold R_i associated with the i^{th} usage rate class $r \in (r_i, r_{i+1}]$ can be expressed as

$$R_i = \exp\{-\int_{r_i}^{r_{i+1}} \left(\int_{U_{k-1}^i/r}^{(U_{k-1}^i+u_k^i)/r} \gamma((u_1+Z_{k-1}^i)/r+x|r)\,\mathrm{d}x \right) \mathrm{d}G(r)\}, \tag{23}$$

where $J_k^i(-\ln R_i)$ is inverse function of $\pi_k^i(u_k^i) = \int_{r_i}^{r_{i+1}} \left(\int_{U_{k-1}^i/r}^{(U_{k-1}^i+u_k^i)/r} \gamma((u_1+Z_{k-1}^i)/r+x|r)\,\mathrm{d}x \right) \mathrm{d}G(r)$; $(u_k^i = J_k^i(-\ln R_i))$ is an usage interval between the $(k-1)^{\text{th}}$ PM and the k^{th} PM; $Z_{k-1}^i = \sum_{l=0}^{k-1} \alpha_l^i u_l^i$; $U_{k-1}^i = \sum_{l=0}^{k-1} u_l^i$; and $u_0^i = U_0^i = Z_0^i = 0$.

The expected maintenance cost

The expected maintenance cost in the k^{th} PM interval $(0,\ u_i^k/r]$ can be calculated as

$$C_i^k(u_i^k/r) = c_m \int_{r_i}^{r_{i+1}} \left(\int_{U_{k-1}^i/r}^{(U_{k-1}^i+u_k^i)/r} \gamma((u_1+Z_{k-1}^i)/r+x|r)\,\mathrm{d}x \right) \mathrm{d}G(r). \tag{24}$$

After the n_i PM, the expected maintenance cost is

$$C_i^{n_i} = c_m \int_{r_i}^{r_{i+1}} \left(\int_{U_{n_i}^i/r}^{(u-u_1)/r} \gamma((u_1+Z_{n_i}^i)/r+x|r)\,\mathrm{d}x \right) \mathrm{d}G(r), \tag{25}$$

where $n_i = \{n_i > 0 | U_{n_i}^i \le u - u_1\}$.

The expected maintenance cost in the remaining usage limit $u-u_1$ is expressed as

$$\begin{aligned} MC_i(R_i,n_i) &= \sum_{k=1}^{n_i-1} C_i^k(u_i^k/r) + C_{n_i}^i + \sum_{k=1}^{n_i} C_{p_i}^k(\beta_k^i u_k^i) \\ &= \sum_{k=1}^{n_i-1} c_m \int_{r_i}^{r_{i+1}} \left(\int_{U_{k-1}^i/r}^{(U_{k-1}^i+u_k^i)/r} \gamma((u_1+Z_{k-1}^i)/r+x|r)\,\mathrm{d}x \right) \mathrm{d}G(r) \\ &\quad + c_m \int_{r_i}^{r_{i+1}} \left(\int_{U_{n_i}^i/r}^{(u-u_1)/r} \gamma((u_1+Z_{n_i}^i)/r+x|r)\,\mathrm{d}x \right) \mathrm{d}G(r) + \sum_{k=1}^{n_i} \left(C_{p_i}^k(\beta_k^i u_k^i) \right). \end{aligned} \tag{26}$$

The expected penalty cost

The expected revenue rate in the k^{th} PM interval $(0,\ u_i^k/r]$ is given by $\mathbb{R}_k^i = \int_{r_i}^{r_{i+1}} r \int_{U_{k-1}^i/r}^{(U_{k-1}^i+u_k^i)/r} \mathbb{R}_p^i((u_1+Z_{k-1}^i)/r+x\,|\,r)\mathrm{d}x\mathrm{d}G(r)/u_k^i$.

And the expected revenue rate after the n_i^{th} PM is given by $\mathbb{R}_{n_i}^i = \int_{r_i}^{r_{i+1}} r \int_{Z_{n_i}^i/r}^{(u-u_1)/r} \mathbb{R}_p^i((u-u_1-Z_{n_i}^i)/r+x\,|\,r)\mathrm{d}x\mathrm{d}G(r)$

Thus, the expected penalty cost in the remaining $/(u-u_1-Z_{n_i}^i)$. usage limit $u-u_1$ can be expressed as

$$\begin{aligned} PC_i(R_i,n_i) &= \bar{H}(\vartheta) \cdot \\ &\left(\sum_{k=1}^{n_i-1} \mathbb{R}_k^i \int_{r_i}^{r_{i+1}} r \int_{U_{k-1}^i/r}^{(U_{k-1}^i+u_k^i)/r} \mathbb{R}_p^i((u_1+Z_{k-1}^i)/r+x\,|\,r)\mathrm{d}x\mathrm{d}G(r) \right. \\ &\left. + \int_{r_i}^{r_{i+1}} r \int_{Z_{n_i}^i/r}^{(u-u_1)/r} \mathbb{R}_p^i((u-u_1-Z_{n_i}^i)/r+x\,|\,r)\mathrm{d}x\mathrm{d}G(r) \right). \end{aligned} \tag{27}$$

The expected depreciation cost

Accumulative usage and age at the lease region Ω termination are respectively $X_\Omega(t) = u - \sum_{k=1}^{n_i} \beta_k^i u_k^i$ and $T_\Omega(t) = u/r$. By (5), so the expected depreciation cost is

$$DC_i^\Omega(R_i,n_i) = \theta_1(u - \sum_{k=1}^{n_i} \beta_k^i u_k^i)^2 + \theta_2 \int_{r_i}^{r_{i+1}} (u/r)^2 \mathrm{d}G(r). \tag{28}$$

Similarly, the expected depreciation cost at the lease region Ω_1 termination is

$$DC_i^{\Omega_1} = \theta_1 u_1^2 + \theta_2 \int_{r_i}^{r_{i+1}} (u_1/r)^2 \mathrm{d}G(r). \tag{29}$$

Thus, the expected depreciation cost at the remaining usage limit $u-u_1$ termination is

$$DC_i(R_i,n_i) = DC_i^\Omega(R_i,n_i) - DC_i^{\Omega_1}. \tag{30}$$

The total cost model

By (1), the total cost in the remaining usage limit $u-u_1$ can be obtained as

$$\begin{aligned} TC_i(R_i,n_i) &= MC_i(R_i,n_i) + PC_s(R_i,n_i) + DC_i(R_i,n_i) = \\ &\sum_{k=1}^{n_i-1} \left(c_m + \bar{H}(\vartheta)\mathbb{R}_k^i \right) \int_{r_i}^{r_{i+1}} \left(\int_{U_{k-1}^i/r}^{(U_{k-1}^i+u_k^i)/r} \gamma((u_1+Z_{k-1}^i)/r+x|r)\,\mathrm{d}x \right) \mathrm{d}G(r) \\ &+ \left(c_m + \bar{H}(\vartheta)\mathbb{R}_{n_i}^i \right) \int_{r_i}^{r_{i+1}} \left(\int_{U_{n_i}^i/r}^{(u-u_1)/r} \gamma((u_1+Z_{n_i}^i)/r+x|r)\,\mathrm{d}x \right) + \mathrm{d}G(r) \\ &\sum_{k=1}^{n_i} \left(C_{p_i}^k(\beta_k^i u_k^i) \right) + \theta_1 \left((u - \sum_{k=1}^{n_i} \beta_k^i u_k^i)^2 - u_1^2 \right) + \\ &\theta_2 \int_{r_i}^{r_{i+1}} \left((u/r)^2 - (u_1/r)^2 \right) \mathrm{d}G(r). \end{aligned} \tag{31}$$

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A fuzzy evaluation method based on fuzzy consistency matrix for evaluating maintenance design program: Case study on heavy vehicle systems

X.J. Yi
China North Vehicle Research Institute, Beijing, China

Y.H. Lai
College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing, China

P. Hou & H.N. Mu
Beijing Institute of Technology, Beijing, China

ABSTRACT: This paper takes heavy vehicle systems as case study to propose a fuzzy evaluation method based on fuzzy consistency matrix for evaluation maintenance design program, and formulate its process, which are determining the evaluation index, developing the hierarchical structure model, constructing fuzzy complementary matrix, operating the single hierarchical arrangement, operating the weight total ordering, and evaluating the level of maintenance design program. In order to illustrate the reasonable and feasible of the proposed evaluation method, its weight is compared with the results by other methods: Analytic hierarchy process, Fuzzy analytic hierarchy process based on consistency matrix, Fuzzy analytic hierarchy process based on triangular fuzzy number. Furthermore, the evaluation result is compared with those of the evaluation method based on cloud model. All in all, this paper proposes a new evaluation method for evaluating maintenance design program of repairable systems.

1 INTRODUCTION

The availability of product is the key factor to determine the stability of product performance. The availability of the product is determined by the reliability and maintainability [1]. By carrying out the reliability design of the product, the failure rate of the product can be reduced. Similarly, the maintenance design [2–3] in the product design stage is needed to achieve economic, convenient and effective maintenance. At present, the maintainability design has been paid attention to by the designers. The key factors affecting the maintenance level are as follows: simplification, standardization and interchangeability, detection, repairability, mistake proofing, safety, accessibility, human element engineering, maintenance time. The important part of the maintenance design is the evaluation of the maintenance design scheme with the comprehensive consideration of the above factors [4–5]. At the same time, the evaluation results can provide the basis for the analysis of the maintainability of the product and the improvement of its maintenance level. For this kind of evaluation, analytic hierarchy process, cloud model algorithm and so on are often used. There are two outstanding issues in the evaluation of maintenance design scheme by such evaluation methods: (1) The results obtained by these evaluation techniques can be affected by the subjective factors of experts. Only the influence of the subjective factors turn weak, can the evaluation results become more reasonable, (2) The expert scoring matrix is widely used in this kind of evaluation method, it is necessary to keep the evaluation intention of experts without changing.

Thus, based on the above problems, taking heavy vehicles as the research object, a fuzzy evaluation method based on consistency matrix for evaluating maintenance design program is presented. The main contributions of this paper are as follows:

1. As an appropriate scale [6–7] can soften the impact of subjective factors, the comparisons between the following scales are carried out: 1~9 scale, the deformation scales 9/9~9/1 and 10/10~18/2, and exponential scales $9^{0/9}$~$9^{8/9}$, $2^{0/2}$~$2^{8/2}$, $e^{0/4}$~$e^{8/4}$. The results can reduce the trouble about the selection of scale.
2. A fuzzy evaluation method based on consistency matrix for evaluating maintenance design program is proposed, and the corresponding evaluation process is also developed. Naturally,

the problem of consistency of expert opinions and scoring matrix is solved.

The remainder of the paper is organized as follows. A fuzzy evaluation method based on consistency matrix for evaluating maintenance design program is proposed in section 2. Section 3 illustrates a practical example, which is a heavy vehicle, based on the proposed method. Section 4 conducts result analysis with the method based on the cloud model. Section 5 provides some conclusions on the findings of the research.

2 THE IMPROVED FUZZY EVALUATION METHOD BASED ON CONSISTENCY MATRIX

For the maintainability initiative design, the maintenance evaluation is necessary and an improved evaluation method based on consistency matrix is proposed. The method combines evaluation with the maintainability design content and weaken the subjective influence of expert. At the same time, a simple and reliable maintainability evaluation process is given for design personnel to provide the basis for maintainability design work actively.

2.1 *Scale analysis*

Scale is the tool for expert to evaluate, and the existing scale can be divided into three categories: three scales, five scales and nine scales. The nine scales have the highest accuracy, strong psychological foundation, and the widest application. The nine scales are mainly: 1~9 scale, the deformation scales 9/9~9/1 and 10/10~18/2, and exponential scales $9^{0/9}$~$9^{8/9}$, $2^{0/2}$~$2^{8/2}$, $e^{0/4}$~$e^{8/4}$. In addition, the exponential scales have higher requirements on readability, availability, and normalization. There are so many kinds of nine scales and large differences between them, so the applications are suitable after comparisons.

For convenience, we define the natural language description and the corresponding scale level as shown in Table 1.

The relationship between the scale value and the natural language description can be connected, and the result is shown as in Table 2.

The relative characteristics of different scales are shown in Table 3.

From Table 3, some conclusion can be obtained:

1. All of the scales can keep the rank property under a single criterion, but the rank property under multiple criteria is unclear.
2. From the uniformity, memory, and perception, 1~9 scale is the best, but its consistency is just ordinary. Therefore, 1~9 scale is only applicable in the case of that the requirement about the evaluation accuracy is not high. The consistency of $1\sim9^2$ scale is not qualified and other indexes

Table 1. The corresponding relationship between scale level and natural language description.

Scale level	1	3	5
Natural language description	E—equal	S—slight	O—obvious
Scale level	7	9	2,4,6,8
Natural language description	I—intense	U—ultra	intermediate value

Table 2. The corresponding relationship between scale value and natural language description.

Scale	E	S	O
1~9	1	3	5
$1\sim\sqrt{9}$	1	$\sqrt{3}$	$\sqrt{5}$
$1\sim9^2$	1	9	25
$\frac{9}{9}\sim\frac{9}{1}$	9 / 9 (1)	9 / 7 (1.286)	9 / 5 (1.8)
$\frac{10}{10}\sim\frac{18}{2}$	10 / 10 (1)	12 / 8 (1.5)	14 / 6 (2.33)
$9^{0/9}\sim9^{8/9}$	$9^{(0)}$ (1)	$9^{(1/9)}$ (1.277)	$9^{(3/9)}$ (2.08)
$2^{0/2}\sim2^{8/2}$	$2^{0/2}$ (1)	$2^{2/2}$ (2)	$2^{4/2}$ (4)
$e^{0/4}\sim e^{8/4}$	$e^{0/4}$ (1)	$e^{2/4}$ (1.649)	$e^{4/4}$ (2.718)
Scale	I	U	Formula
1~9	7	9	k
$1\sim\sqrt{9}$	$\sqrt{7}$	3	$\sqrt{k}$
$1\sim9^2$	49	81	k^2
$\frac{9}{9}\sim\frac{9}{1}$	9 / 3 (3)	9 / 1 (9)	$\frac{9}{10-k}$
$\frac{10}{10}\sim\frac{18}{2}$	16 / 4 (4)	18 / 2 (9)	$\frac{9+k}{11-k}$
$9^{0/9}\sim9^{8/9}$	$9^{(6/9)}$ (4.327)	$9^{(9/9)}$ (9)	$9^{(\frac{k-1}{9})}$
$2^{0/2}\sim2^{8/2}$	$2^{6/2}$ (8)	$2^{8/2}$ (16)	$2^{\frac{k-1}{2}}$
$e^{0/4}\sim e^{8/4}$	$e^{6/4}$ (4.482)	$e^{8/4}$ (7.39)	$e^{\frac{k-1}{4}}$

Table 3. The relative properties of different scales.

Characteristics	1~9	$1\sim\sqrt{9}$	$1\sim9^2$	$\frac{9}{9}\sim\frac{9}{1}$
Retentivity (Single criterion) [8]	Correct	Correct	Correct	Correct
Consistency	Ordinary	Ordinary	Nonconformity	Ordinary
Uniformity	Excellent	Excellent	Poor	Ordinary
Memory	Excellent	Ordinary	Ordinary	Poor
Perception	Excellent	Poor	Ordinary	Poor
Weight fitting [8]	Ordinary	Good	Ordinary	Ordinary
Total evaluation	Good	Ordinary	Poor	Good
Characteristics	$\frac{10}{10}\sim\frac{18}{2}$	$9^{0/9}\sim9^{8/9}$	$2^{0/2}\sim2^{8/2}$	$e^{0/4}\sim e^{8/4}$
Retentivity (Single criterion) [8]	Correct	Correct	Correct	Correct
Consistency	Ordinary	Excellent	Excellent	Excellent
Uniformity	Ordinary	Good	Ordinary	Good
Memory	Poor	Poor	Poor	Poor
Perception	Poor	Poor	Poor	Poor
Weight fitting [8]	Good	Excellent	Excellent	Excellent
Total evaluation	Good	Excellent	Good	Excellent

are not good, so it is not recommended. The perception of $1\sim\sqrt{9}$ scale is poor, and the calculation is troublesome, and can be used when calculating conditions permit. The uniformity of 9/9~9/1 scale and 10/10~18/2 scale is poor, and they may be considered as appropriate.

3. Although the memory and perception of exponential scales $9^{0/9}\sim9^{8/9}$, $2^{0/2}\sim2^{8/2}$, $e^{0/4}\sim e^{8/4}$ are poor, their consistency and weight fitting is good. They are suitable for the requirement of high accuracy, and can be complementary with 1~9 scale.

2.2 *Improved evaluation method based on consistency matrix*

1. Determining the indexes of the object in an evaluation

As the evaluation object has multiple attributes, it is necessary to determine the indexes to guide the evaluation direction. The evaluation result must be achieved with these indexes under the same conditions.

2. Establishment of hierarchical model

The evaluation target is difficult to be evaluated directly, so the object is usually decomposed into different parts. These parts can indirectly represent the object, the same as the evaluation result. According to the relationship between the parts, as well as the subordinate relations, a multi-level analysis model can be obtained. The factors with similar character are divided in the same level, i.e. they are the parallel relationship. Naturally, the factors are part of another factor or dominated, they are divided in the lower layer. The hierarchical model can be shown as in Figure 1.

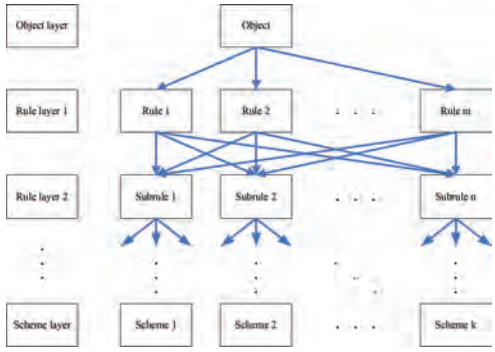


Figure 1. Hierarchical structure diagram.

Although there is no limit on the number of elements per layer in a rule layer, it is generally not more than 9. Psychological research shows that people can have reasonable judgments when the number of object is within nine. When there are more than 9 elements in the same level, the number of elements in the same layer can be reduced by increasing the number of layer.

3. Establishment of consistency matrix

When experts start to give estimation scale for every comparison, it is best to choose the memorable and perceptive scales, for example 1~9 scale. Meanwhile, the effect of scale is that translate the qualitative results of experts into quantitative value and the retentivity and consistency of scale

must meet the requirements to avoid the quantitative result deviate from the expert's intention. Last but not least, taking into account the evaluation habits of experts, the uniform and weight fitting indexes are also important. Above all, experts give estimation scale using 1~9 scale and 1~9 scale will be translated correspondingly into $e^{0/4}$~$e^{8/4}$ scale to form the consistency matrix.

4. Hierarchical ranking

For every layer, the weight value of all elements can be calculated according to the consistency matrix. After the eigenvector corresponding to the eigenvalue of maximum in a consistency matrix is obtained, the weight value of the elements in the same layer can be got by the normalization of eigenvector.

5. Total ranking

By step (4), the weight value of elements in each layer can be obtained. The following assumptions are given: the weight of m elements in k-1_{th} layer relative to the object is $w^{(k-1)}=\left(w_1^{(k-1)},w_2^{(k-1)},\ldots,w_m^{(k-1)}\right)$; the weight of n elements in k layer subject to the j_{th} element in k-1_{th} layer is $p_j^{(k)}=\left(p_{j1}^{(k)},p_{j2}^{(k)},\ \ ,p_{jn}^{(k)}\right)$ and the weight of elements in k layer irrelative to the j_{th} element in k-1_{th} layer is zero; $P^{(k)}=\left(p_1^{(k)},p_2^{(k)},\ \ ,p_m^{(k)}\right)^T$ represents the weight of all elements in k_{th} layer subject to k-1_{th} layer. The calculation formulas are shown as in Eqn. (1) and (2).

$$w^{(k)}=\left(w_1^{(k)},w_2^{(k)},\ldots,w_n^{(k)}\right)^T=w^{(k-1)}\ P^{(k)} \tag{1}$$

$$w_i^{(k)}=\sum_{j=1}^{m}p_{ij}^{(k)}w_j^{(k-1)},\ i=1,2,\ldots,n \tag{2}$$

All the weight values are summarized and unified from top to bottom to get the total weight. The total weight can represent the priority of all the schemes.

6. Set up a set of fuzzy comments

For the evaluation of maintainability design content, a set of fuzzy comments is needed for designers. The set of fuzzy comments is usually divided into five levels, V = {v1, v2, v3, v4, v5} = {excellent, good, ordinary, poor, bad}.

7. Establishment of fuzzy evaluation matrix

Through questionnaire survey of designers, fuzzy evaluation matrix is constructed for calculating the evaluation levels of units. Based on step (6), the fuzzy evaluation matrix E of all factors in the lowest level can be counted as shown in Eqn. (3).

$$TE=\begin{bmatrix}E_1\\ \vdots\\ E_n\end{bmatrix}=\begin{bmatrix}e_{11} & \cdots & e_{15}\\ \vdots & \ddots & \vdots\\ e_{n1} & \cdots & e_{n5}\end{bmatrix} \tag{3}$$

where, e_{ij} represents the fuzzy evaluation of i_{th} factor, which subject to the j_{th} evaluation level, given by designers.

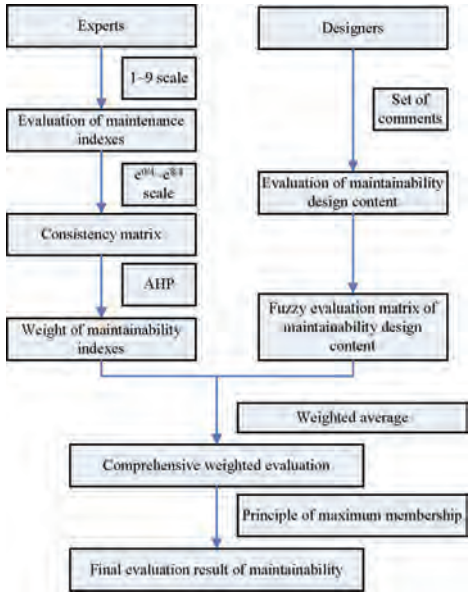


Figure 2. The procedure of proposed fuzzy evaluation method based on consistency matrix.

Combined with the weight values above, comprehensive weighted evaluation [9] can be calculated from bottom to top, i.e. membership matrix. The membership matrix can be calculated as shown in Eqn. (4).

$$B=f\left(w^{(k)}E\right)=\left(w_1^{(k)},w_2^{(k)},\cdots,w_n^{(k)}\right)\begin{bmatrix}e_{11} & e_{12} & \cdots & e_{1m}\\ e_{21} & e_{22} & \cdots & e_{2m}\\ \vdots & \vdots & \ddots & \vdots\\ e_{n1} & e_{n2} & \cdots & e_{nm}\end{bmatrix}=\left(b_1,b_2,\cdots,b_m\right) \tag{4}$$

where, b_i indicates that the membership of evaluated object subject to v_i degree.

8. Final evaluation result

Based on the above analysis results, the final evaluation result of system can be obtained based on the principle of maximum membership. Meanwhile, the weak links can be found and provide basis for the corresponding correction strategies.

The procedure of proposed fuzzy evaluation method based on consistency matrix is shown in Figure 2.

3 CASE STUDY

Taking heavy vehicles as example, the maintenance evaluation based on the proposed evaluation method is conducted. Specific implementation steps are as follows:

1. Selection of evaluation indexes for heavy vehicles

The key factors affecting the maintenance level are as follows: simplification (A), standardization and interchangeability (B), detection (C), repairability (D), mistake proofing (E), safety (F), accessibility (G), human element engineering (H), maintenance time (I). All the above evaluation indexes can be accepted by relevant personnel and departments. The first seven indexes are qualitative properties, and the last two indexes are quantitative properties. The corresponding detailed maintainability design content can be listed simply: A1, A2, A3, B1, B2, B3, B4, C1, C2, C3, D1, D2, D3, E1, E2, E3, F1, F2, F3, G1, G2, G3, H1, H2, H3, I1, I2, I3.

2. Establishment of hierarchical model

The object is the maintainability of heavy vehicle system, and the maintainability indexes are listed in rule layer. The lowest layer is the corresponding detailed maintainability design content. The hierarchical model is conducted as in Figure. 3.

3. Consistency matrix

Several experts provide the evaluations about the elements in rule layer 1 by 1~9 scale. The evaluation of the first expert can be shown as in Table 4.

According to the proposed method and the evaluation results of experts, the corresponding consistency matrix can be obtained, as shown in Table 5.

4. Order ranking

According to the Table 5, the weight of maintainability indexes can be obtained and shown in Table 6.

For every layer, the weight value of all elements can be calculated according to the consistency matrix. For the same elements, different weight can be obtained by different expert. Taking the average of different weight, the final weight of maintainability indexes are shown in Table 7.

5. Total ranking

All the weight values are summarized and unified from top to bottom to get the total weight. The total weight is shown in Table 8.

6. Set up a set of fuzzy comments

For the evaluation of maintainability design content, the set of fuzzy comments V = {v1, v2, v3, v4, v5} = {excellent, good, ordinary, poor, bad} is used.

7. Establishment of fuzzy evaluation matrix

Through questionnaire survey of designers, the fuzzy evaluation matrix *E* is shown as in Table 9.

8. Final evaluation result.

Based on the above analysis results and Eqn. (4), the membership result of all elements in rule layer 1 and system are shown in Table 10. Based on the principle of maximum membership, the final evaluation result is also shown in Table 10.

We can see from Table 10 that the membership of system subject to ordinary evaluation as high as 46.3%. It is indicated that the maintainability of heavy vehicle is still insufficient and need to be improved urgently. At the same time, according to

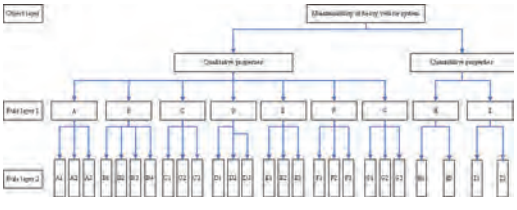


Figure 3. The hierarchical model of maintainability of heavy vehicle system.

Table 4. The evaluation of the first expert.

	1	2	3	4	5	6	7	8	9
1	1	1	1/2	1	5	1	1	1	1
2	1	1	1/2	1	5	1	1	1	1
3	2	2	1	1	8	2	2	1	1
4	1	1	1	1	6	1	2	1	1
5	1/5	1/5	1/8	1/6	1	1/5	1/4	1/7	1/6
6	1	1	1/2	1	5	1	1	1	1
7	1	1	1/2	1/2	4	1	1	1/2	1/2
8	1	1	1	1	7	1	2	1	1
9	1	1	1	1	6	1	2	1	1

Table 5. Consistency matrix

	1	2	3	4	5	6	7	8	9
1	1	$e^{0/4}$	$1/e^{1/4}$	$e^{0/4}$	$e^{4/4}$	$e^{0/4}$	$e^{0/4}$	$e^{0/4}$	$e^{0/4}$
2	$e^{0/4}$	1	$1/e^{1/4}$	$e^{0/4}$	$e^{4/4}$	$e^{0/4}$	$e^{0/4}$	$e^{0/4}$	$e^{0/4}$
3	$e^{1/4}$	$e^{1/4}$	1	$e^{0/4}$	$e^{7/4}$	$e^{1/4}$	$e1^{/4}$	$e^{0/4}$	$e^{0/4}$
4	$e^{0/4}$	$e^{0/4}$	$e^{0/4}$	1	$e^{5/4}$	$e^{0/4}$	$e^{1/4}$	$e^{0/4}$	$e^{0/4}$
5	$1/e^{0/4}$	$1/e^{0/4}$	$1/e^{7/4}$	$1/^{5/4}$	1	$1/e^{4/4}$	$1/e^{3/4}$	$1/e^{6/4}$	$1/e^{5/4}$
6	$e^{0/4}$	$e^{0/4}$	$1/e^{1/4}$	$e^{0/4}$	$e^{4/4}$	$1/e^{1/4}$	$e^{0/4}$	$e^{0/4}$	$e^{0/4}$
7	$e^{0/4}$	$e^{0/4}$	$1/e^{1/4}$	1/2	$e^{3/4}$	$e^{0/4}$	1	$1/e^{1/4}$	$1/e^{1/4}$
8	$e^{0/4}$	$e^{0/4}$	$e^{0/4}$	$e^{0/4}$	$e^{6/4}$	$e^{0/4}$	$e^{1/4}$	1	$e^{0/4}$
9	$e^{0/4}$	$e^{0/4}$	$e^{0/4}$	$e^{0/4}$	$e^{5/4}$	$e^{0/4}$	$e^{1/4}$	$e^{0/4}$	1

Table 6. Weight of maintainability indexes.

W_A	W_B	W_C	W_D	W_E
0.114	0.114	0.144	0.124	0.037
W_F	W_G	W_H	W_I	
0.114	0.102	0.128	0.124	

Table 7. Final weight of maintainability indexes.

W_A	W_B	W_C	W_D	W_E
0.1	0.095	0.129	0.132	0.065
W_F	W_G	W_H	W_I	
0.103	0.09	0.141	0.143	

Table 8. The total weight of all elements.

Rule layer 1	Weight	Rule layer 2	Weight	Total weight
		A1	0.347	0.035
Simplification	0.145	A2	0.332	0.033
		A3	0.321	0.032
		B1	0.249	0.024
Standardization and interchangeability	0.133	B2	0.243	0.023
		B3	0.258	0.025
		B4	0.249	0.024
		C1	0.391	0.051
Detection	0.091	C2	0.304	0.039
		C3	0.304	0.039
		D1	0.333	0.044
Repairability	0.074	D2	0.333	0.044
		D3	0.333	0.044
		E1	0.347	0.023
Mistake proofing	0.146	E2	0.332	0.022
		E3	0.321	0.021
		F1	0.392	0.04
Safety	0.145	F2	0.392	0.04
		F3	0.216	0.022
		G1	0.43	0.039
Accessibility	0.063	G2	0.447	0.04
		G3	0.123	0.011
Human element engineering	0.121	H1	0.558	0.079
		H2	0.442	0.062
Maintenance time	0.083	I1	0.36	0.052
		I2	0.64	0.092

Table 9. Fuzzy evaluation matrix *E*.

Rule layer 2	Excellent	Good	Ordinary	Poor	Bad
A1	1	0	0	0	0
A2	1	0	0	0	0
A3	0.8	0.2	0	0	0
B1	0.8	0.2	0	0	0
B2	0.8	0.2	0	0	0
B3	1	0	0	0	0
B4	1	0	0	0	0
C1	0	0.2	0.6	0.2	0
C2	0	0	0.8	0.2	0
C3	0	0	0.8	0.2	0
D1	0	0	0.8	0.2	0
D2	0	0	0.6	0.4	0
D3	0	0	0.6	0.4	0
E1	1	0	0	0	0
E2	1	0	0	0	0
E3	0.8	0.2	0	0	0
F1	1	0	0	0	0
F2	1	0	0	0	0
F3	0	0	1	0	0
G1	0	0	1	0	0
G2	0	0	1	0	0
G3	0	0	0	0.4	0.6
H1	0.4	0.4	0.2	0	0
H2	0	0	1	0	0
I1	0	0	0.2	0.8	0
I2	0	0	1	0	0

Table 10. The membership and evaluation level of elements in rule layer1 and object layer.

Rule layer 1	Excellent	Good	Ordinary	Poor	Bad	Evaluation level
Simplification	0.094	0.006	0	0	0	Excellent
Standardization and interchangeability	0.086	0.009	0	0	0	Excellent
Detection	0	0.01	0.093	0.026	0	Ordinary
Repairability	0	0	0.088	0.044	0	Ordinary
Mistake proofing	0.061	0.004	0	0	0	Excellent
Safety	0.081	0	0.022	0	0	Excellent
Accessibility	0	0	0.079	0.004	0.007	Ordinary
Human element engineering	0.032	0.032	0.078	0	0	Ordinary
Maintenance time	0	0	0.102	0.041	0	Ordinary
System	0.353	0.062	0.463	0.116	0.007	Ordinary

the specific evaluation level of each index in rule layer 1, the corresponding maintainability design content can be traced.

4 RESULT ANALYSIS

In order to illustrate the accuracy and applicability of the proposed method, the relative comparisons were carried out.

1. Comparison about weight

The result obtained by proposed method was compared with the results by other methods: Analytic hierarchy process (AHP) [7], Fuzzy analytic hierarchy process based on consistency matrix (FAHP) [10], Fuzzy analytic hierarchy process based on triangular fuzzy number (SFAHP) [11]. All the results are listed in Table 11.

As can be seen from the above table, the weights obtained by AHP and proposed method are very

Table 11. Weight comparison.

	W_A	W_B	W_C	W_D	W_E
AHP	0.1	0.095	0.125	0.136	0.066
Proposed method	0.1	0.095	0.129	0.132	0.065
FAHP	0.112	0.111	0.112	0.112	0.106
SFAHP	0.108	0.106	0.119	0.121	0.085
	W_F	W_G	W_H	W_I	
AHP	0.103	0.088	0.141	0.147	
Proposed method	0.103	0.09	0.141	0.143	
FAHP	0.112	0.111	0.113	0.112	
SFAHP	0.11	0.102	0.124	0.125	

Table 12. Evaluation results between the proposed method and cloud model method.

	Cloud model method	Proposed method
Simplification	Excellent	Excellent
Standardization and interchangeability	Excellent	Excellent
Detection	Good	Ordinary
Repairability	Good	Ordinary
Mistake proofing	Ordinary	Excellent
Safety	Excellent	Excellent
Accessibility	Good	Ordinary
Human element engineering	Good	Ordinary
Maintenance time	Good	Ordinary
System	Good	Ordinary

close, but the range of result obtained by proposed method is slightly smaller than that of AHP. The main reason is that, although the judgment matrix of 1~9 scale is inconsistent, the error is not very large due to its high precision. And the weight obtained by FAHP is close, which is result from the poor precision of 0.1~0.9 scale. The result obtained by SFAHP is not good and the procedure of SFAHP is more complicated than other methods.

2. Comparison of evaluation results

The evaluation results obtained by the proposed method is compared with the cloud model method [12] based on golden partition evaluator and normal similarity [13–14], all the results are listed in Table 12.

According to Table 12, the qualitative evaluation results obtained by the proposed method are more conservative than that of cloud model method. As the using of golden partition evaluator and normal similarity is complex, and the proposed method is more suitable for the engineering application of maintenance evaluation.

5 CONCLUSION

Maintainability initiative design is an important way to improve the maintainability of equipment. And the maintenance evaluation can effectively grasp the maintenance design content and the maintenance level. From the purpose of the engineering application of maintenance evaluation, the appropriate scale can be selected which can weaken the influence of subjective factors after the comparisons between different scales are conducted. According to the grade that experts give, the consistency matrix is constructed to solve the problem about the consistency of expert opinion, and the related method and procedure are optimized. By using the AHP method, the expert experience can be translated into weight values. In addition, the fuzzy comprehensive evaluation method also can translate the designer's understanding about maintainability design content into values. Combined both, the fuzzy evaluation about the system can be obtained. Finally, a reasonable and simple evaluation process is given to make practical application more convenient.

Simultaneously, the proposed method is applied on heavy vehicles to study the maintenance evaluation. To verify the accuracy of the proposed method, two comparisons are conducted. The first comparison is the weight values between different methods: the proposed method, AHP, the FAHP based on fuzzy consistency matrix using 0.1~0.9 scale, and the FAHP based on triangular fuzzy number. The result proves that the proposed method has higher accuracy. The second comparison is that the evaluation result of the proposed method is compared with that of cloud model algorithm. The result of second comparison verifies the applicability and accuracy of the proposed method applied on heavy vehicles. The efficiency of the evaluation also can be improved. The proposed method effectively combine expert experience and maintainability design content to provide a basis for the follow-up maintenance initiative design and improvement.

ACKNOWLEDGEMENT

This paper is supported by Pre-Research Project in the years 2016–2020 (41404060103) and Basic Research Program of MIIT in the years 2015–2017 (A0920132002). We are grateful to the editors and reviewers for the suggestions that improve the draft of this paper.

REFERENCES

[1] Ralf Gossinger, Hanna Helmke, and Michael Kaluzny, 2017. Condition-based release of main-

tenance jobs in a decentralised production-maintenance system – An analysis of alternative stochastic approaches, *International Journal of Production Economics*, 193: 528–537.

[2] John Andrews and Claudia Fecarotti, 2017. System design and maintenance modeling for safety in extended life operation, *Reliability Engineering and System Safety*, 163: 95–108.

[3] Tom Vaneker and Tijmen van Diepen, 2016. Design Support for Maintenance Tasks using TRIZ, *Procedia CIRP* 39: 67–72.

[4] Gao, Xu, at al, 2014. Application of the Model Based on Fuzzy Consistent Matrix and AHP in the Assessment of Fire Risk of Subway Tunnel, *Procedia Engineering*, 71: 591–596.

[5] Jacek Skorupski, 2016. The simulation-fuzzy method of assessing the risk of air traffic accidents using the fuzzy risk matrix, *Safety Science*, 88: 76–87.

[6] Xiong, Liang and Wang, 2005. Method Research on Selection and Valuation of Numeric Scale in Analytic Hierarchy Process, *Systems Engineering Theory & Practice*, 3: 72–79.

[7] Xu Zeshui, 2000. A simulation-based evaluation of several scales in the analytic hierarchy process, *Systems Engineering Theory & Practice*, 20(7): 58–62.

[8] Luo Zheng-qing, Yang Shan-lin, 2004. Comparative Study on Several Scales in AHP, *System Engineering Theory & Practice*, 9: 51–60.

[9] Zhang, Liu, et al., 2010. Application of fuzzy comprehensive assessment to evaluation for reliability and maintainability of military engineer machinery, *Mining machinery*, 38(8): 42–45.

[10] Liu and Li, 2008. A Method of Improving the Consistence of the Judgment Matrix Based on the Property of the Fuzzy Consistent Judgment Matrix, *Journal of University of Jinan (Sci. & Tech.)*, 2(22): 200–202.

[11] Lu Zhong and Sun You-chao, 2007. Research on Maintainability Evaluation Model Based on Fuzzy Theory, *Chinese Journal of Aeronautics*, 20: 402–407.

[12] Han, Wang, et al., 2017. Research on Complex Equipment Maintainability Index Evaluation Based on Cloud Theory. *Journal of Ordnance Equipment Engineering*, (3): 72–76.

[13] Li Hailin, Guo Chonghui, and Qiu Wangran, 2011. Similarity Measurement between Normal Cloud Models. *Acfa Elegtronica Sinica*, 11(39): 2561–2567.

[14] Wang Jian, et al., 2010. An Improved Effectiveness Evaluation Method based on Cloud Model. Fire Control & Command Control, 7(35): 139–141.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Towards a model based asset deterioration framework represented by probabilistic relational models

Haoyuan Zhang & D. William R. Marsh
Risk and Information Management Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, UK

ABSTRACT: Most asset deterioration tools are designed for a specific application, as a consequence, a small change of the specification may result in a complete change of the tool. Inspired by the model-based approach of separating problem specification from analysis technique, we propose a model-based asset deterioration assessment framework using probabilistic relational models. The probabilistic relational models express abstract probabilistic dependency covers a range of deterioration modelling assumptions. An expert in the domain of asset deterioration can then use his knowledge of the factors that affect deterioration to customise the abstract models to a specific application, without requiring a detailed understanding the underlying computational framework. We illustrate the use of the framework with multiple variants of deterioration models.

1 INTRODUCTION

Traditionally, inspection and maintenance of infrastructure has followed a fixed time interval. One idea to make inspection more cost-effective is to use a statistical model to predict the rate of asset deterioration and use the predictions to plan detailed inspections or maintenance. A range of deterioration models has been developed in different field, from railway track (Guler et al., 2011) to bridge (Sobanjo, 2011). Despite having common objectives such as condition prediction and using the prediction to leverage maintenance planning, the models differ in many ways. For example, the deterioration distribution and the grading system for asset condition may differs depending on the asset type or the standards set by inspection agencies. Our aim is to build a unified framework that is general enough to encode a wide variety of deterioration models.

The approach of providing unified tools, so called model-based system engineering, has been advocated in both industry and academia. It aims to provide descriptive modelling of systems, common to different system analysis techniques. Importantly, this approach aims to enable decision makers to use analyses without a detailed knowledge of the underlying mathematical models. We review previous work in Section 2 and consider how it applies to deterioration models.

In Section 3, we describe a framework for maintenance domain experts, which does not require a detailed understanding of the underlying deterioration models. Our framework extends standard hierarchical Bayesian models with relational schema, allowing model variants to be expressed using domain concepts. In Section 4, we illustrate the use of the framework with a variety of deterioration model.

2 MODEL-BASED APPROACH

The emerging field of model-based system engineering focuses on bridging the gap between problem specification and modelling (Estefan, 2007), with an agile modelling formalism to meet different modelling requirements without changing the entire tool (Prosvirnova et al., 2013). The model-based approach formalises the system development process using a unified language (e.g. SysML language) to provide a platform integrating different modelling approaches and system analysis methods. This formalism has been extended to the safety and reliability domain, so called model-based safety assessment (MBSA) (see Lisagor et al. (2011)). MBSA aims to unify classical safety and reliability modelling methods (e.g. fault tree and stochastic process), and to generate an integrated structure for a range of safety and reliability analysis (e.g. fault tree analysis and system diagnosis). For example, the AltaRica modelling language (Arnold et al., 1999), separates system specification from analysis with a range of reusable techniques.

The MBSA concept has been applied in deterioration assessment in recent years, with the modelling of stochastic process and Markov chain for complex system developed in project AltaRica 3.0 (Prosvirnova et al., 2013). However, to our knowledge, the current practice of MBSA does not yet encompass learning from data, which is a component of deterioration modelling when since we wish to learn deterioration rates from inspection data. Fortunately, advances in machine learning provide a promising perspective for tackling these problems.

Previous studies have shown the power of a hierarchical Bayesian network (BN) based approach to learn asset deterioration rates from data and how it can be adapted when there is insufficient data, both with expert knowledge (Frangopol et al., 2004, Zhang and Marsh, 2018), or by learning from similar groups (Memarzadeh et al., 2016, Zhang and Marsh, 2018). In the work of Zhang and Marsh (2018), six generic BN models for asset deterioration were developed, which both provides us the possibility of adopt different deterioration models, but also enables us to include alternative data and unused expert knowledge. These model variants cannot yet be presented to an asset deterioration domain expert in a unified framework: adapting the underlying concepts to a particular context requires a deep understanding of their implementation as BNs.

In model-based machine learning (MBML) (see Bishop (2013) and Ghahramani (2015)), models and problem specifications are defined in a compact language, while inference or machine learning algorithm codes are generated automatically. Bayesian networks are such a language, though they lack structure. More recently, probabilistic programming languages such as Figaro (Pfeffer, 2009), has been developed which could also be used in our framework.

Model-based approaches, both in MBML and MBSA, often use the object-oriented paradigm to provide a library of generalized models for reuse. This is not provided by traditional BNs, with a fixed set of variables and relationships. This issue has been widely researched for BNs, with proposals including idioms in Neil et al. (2000) and fragments in Laskey and Mahoney (2000). Probabilistic relational models (PRM), developed by Koller (1999) combines relational structure with probabilistic graphical models (i.e. BNs). A PRM combines probabilistic dependencies with a relational schema that describes the entities in the problem domain. This representation provides a separation between model library and structure relationships.

Therefore, we propose to develop a model-based framework for asset deterioration assessment in the spirit of MBSA. The framework separates reusable low-level models from modelling choices and asset descriptions. The framework is encoded with a PRM representation of a hierarchical Bayesian network, with a range of generalised models for asset deterioration each represented by its probabilistic dependencies, and the problem specification of the target domain is represented as the relational schema.

3 MODEL-BASED ASSET DETERIORATION ASSESSMENT FRAMEWORK

3.1 *Asset deterioration model using hierarchical BNs*

3.1.1 *A simple deterioration model*

For a system that is either working or failed, given historical data on the times that it remained in working condition, we can estimate the distribution of time for its transition to the failed state and so predict its likelihood of failing. A basic deterioration BN model, from Zhang and Marsh (2016), is shown in Figure 1. This is a hierarchical BN model that both learns from data and can be used for decision support.

However, this specific model can only be used to describe a type of asset with two-state and deterioration that follows a one-parameter distribution. This is not usually the case in asset deterioration, for example a 4 point grading system is used to describe bridge condition, and a two-parameter Weibull distribution is used to fit the bridge transition distribution in Sobanjo (2011). So instead, the model has to be adapted: the variables are similar but the number of them and links between them must change.

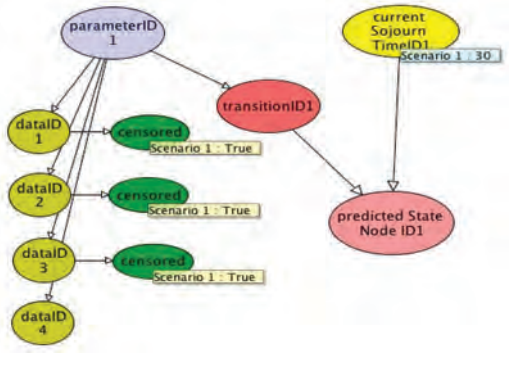


Figure 1. A simple deterioration model.

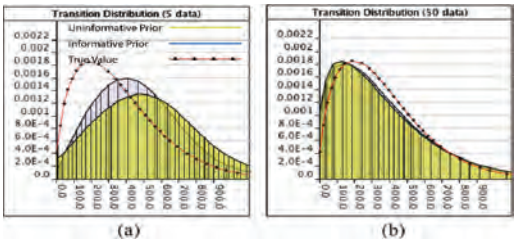


Figure 2. Effects of prior knowledge and data quantity in distribution training.

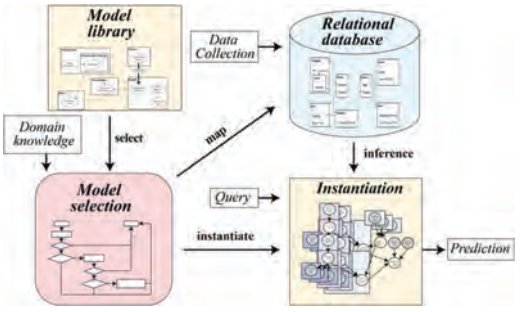


Figure 3. Stages of model-based asset deterioration assessment framework.

The structure of the model may also need to change when we have limited domain data. Although the parameters of the transition distribution are learnt from the historical transition data, some prior knowledge of the transition distribution is also required. In Figure 2, two different prior probability distributions have been used: i) an uninformative prior and ii) an informative prior, available when there is good knowledge representation of the deterioration. Figure 2(a) shows that with good prior, we can provide a good estimate of the true distribution with only a little data, while Figure 2(b) shows a larger dataset will give a correct estimate of the parameter even the prior is weak.

However, when failure data is scarce (which is the usual case in slow deteriorating asset, for example, bridges) or knowledge is poor for a particular asset class (which is also usual for new assets), we can combine data from asset types that, though not identical, are similar and so are believed to have the similar deterioration rate (Morcous, 2011). This kind of approximation is necessary, especially for assets types that are inspected infrequently so that the deterioration dataset is not large enough for each asset types. Two techniques are proposed in Zhang and Marsh (2016): one is to add another layer of parameter (hyper-parameter) to form a hierarchical BN that can group or pool data from different asset types sufficiently to overcome an uninformative prior; the other technique is to use influencing factors to adjust the transition distribution of a specific asset from distribution learned from a pool of similar assets (i.e. assets of the same type). Both methods may change the BN's structure depending on how assets are assigned to groups and what other factors influence the transition times. We refer to these (and related) issues as 'modelling assumptions'.

3.1.2 *A framework for expressing modelling assumptions*

To address the problem of many variants of the deterioration models, we provide a framework to help domain experts express modelling assumptions. The stages of model-based asset deterioration assessment framework can be illustrated in Figure 3:

- The model library encodes the possible dependencies of probabilistic models in the problem domain.
- Model selection uses modelling assumptions to determines what models, knowledge and data are included in the problem model.
- The relational database includes the configuration and failure data; its schema is derived from the modelling assumptions.
- Instantiation and inference, performed automatically, are used to evaluate queries on the model for domain decision support.

The following sections describe each aspects of the framework.

3.2 *Model library: Abstract PRM*

The generic models in the model library are represented as abstract probabilistic relational models (PRMs). Figure 4 shows an example developed from Zhang and Marsh (2018).

In Figure 4, a square (called a class) defines a group of identical objects that share the same set of variables or probabilistic models. An oval defines a variable, and directed edge defines the dependency of variables. Aggregation is defined by a bold arrow. *N* represents a fixed multiple relationship, and * represents a multiple relationship of uncertain degree. The classes are as follows:

Class	Purpose
Asset	We wish to predict the state of a specific asset, conditioned on its previous inspection and the deterioration data of similar assets. This prediction will inform decisions about maintenance and inspection.
Transition	Objects of this class represent transitions in a Markov chain, where the conditional probability of moving into future state S_t+_1 at time $t+1$ given the present state S_t at time t follows a distribution with parameters learnt from data.
Data	Data is gathered from inspections, each giving information about the current state of an asset. Different types of data are used: for example, it is common to have only censored data giving a time after which a transition occurred. This is modelled as constraints on the transition time.
Group	Assets of the same type form a group. A group is represented by the model by parameters of the distribution (for each transition) learnt from historical data for this type of assets. Since the parameters are learnt, their values are uncertain and the model include them as probabilistic variables.
Parameter	The population as a whole also has distribution parameters. The similarity of each group of assets to the population as a whole is judged and this establishes a way to learn a group's distribution parameters from data of other closely related groups.
Factor	The idea of assets of the 'same type' is defined in relation to properties of the asset that influence deterioration rate. However, if all relevant factors are used to distinguish groups then there is likely to be insufficient historical data to estimate the transition distribution parameters. Therefore, groups need to be defined by the factors that are most important (and are known for all assets). Other factors—for example, loading and environment condition—can be used to estimate a target asset's distribution.

We describe this model as 'abstract' as it represents number of different structures. In particular, the following issues need to be resolved to give a specific model:

- The distribution (Weibull or exponential) used for the transition and the number of parameters needed.
- The population priors.
- The number of state of deterioration and therefore the number of transitions between states.

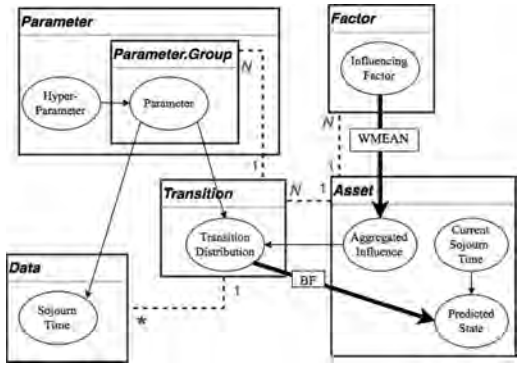


Figure 4. Probabilistic dependencies represented as an abstract PRM.

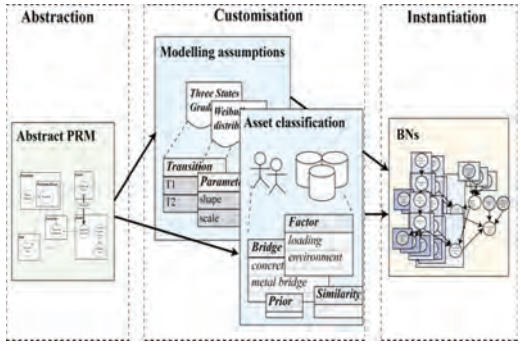


Figure 5. Process to customise the abstract PRM.

- The number of asset groups (or types) and the factors (e.g. material used in construction) used to define membership of a group.
- The similarity of each group to the population as a whole.
- The remaining factors that adjust the transition distribution, possibly varying by group, and the weighting used to aggregate the effect of these factors.
- The types of data available.

Although these factors are uncertain they are not part of the probabilistic reasoning. Instead, these are the decisions made by domain experts to apply the generic models to a specific situation. The next two subsections describe how this is done: the first covers 'modelling assumption' and the second asset classification. Together, as shown in Figure 5, these processes turn the abstract probabilistic model into a model that can be run.

3.3 *Customising the abstract PRM with modelling assumptions*

This aspect of the customisation covers four issues: a) the choice of transition distribution; b) the number of deterioration states; c) prior distributions and d) available inspection data.

3.3.1 *Transition distribution and parameters*

Different distributions can be used to estimate transition times, based on their goodness of fit. The goodness of fit of the distribution is usually done by visual observation and hypothesis test, such as coefficient of determination (R^2) and Anderson Darling (AD) test (Mendenhall et al., 2012). A range of study has been developed to find the best fit distribution of asset state sojourn times. For example, the exponential distribution has been used for railway track (Guler et al., 2011) and the Weibull distribution for bridges (Sobanjo, 2011). The number of parameter in the distribution's survival function fixes the number of instances of the Parameter class for each Transition. An example is showed in the left side of Figure 6: there are two instances of class of Parameter if the guideline shows a Weibull distribution is normally adopted in practice.

3.3.2 *Deterioration states used for grading*

Each asset is usually rated with a state representing its functionality. For example, a 4 point grading system is used in Sobanjo (2011). Grading systems are normally adopted from industry standards and are often used to identify and priorities maintenance actions. They vary for different infrastructure type, countries, and sometimes, inspection agencies. An *n*-states grading system results in *n–1* transitions represented in the instantiation of class Transition. An example is showed in Figure 6's right side: there are two instances of class Transition since it is rated by a three-state grading system.

3.3.3 *Asset deterioration characteristics: Prior*

Classical statistical methods, such as maximum likelihood estimators or least square method can be used to estimate priors if the data are sufficient. An alternative source is the expertise from experienced engineers (Welte and Eggen, 2008) from whom a prior range can be elicited. In addition, each group of assets is also characterised by its degree of similarity to the overall population. These group parameters are modelled used a truncated error distribution, with a mean inherited from the value of the learnt hyper-parameters, and the elicited degree of similarity defining the variance.

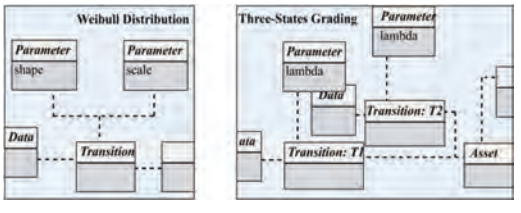


Figure 6. Example customisations of transition distribution and grading system.

3.3.4 *Inspection data: Inferring state sojourn time*

Continuous monitoring can provide exact transition times but periodic inspection is more common. In periodic inspection, the state of asset is only known at the inspection times. Therefore, the time an asset stayed in a state before deteriorating to another state is only constrained by the inspection result. Fortunately, this type of censored data can be modelled by different variants of the Data class. For example, periodic inspection requires three types: left, interval and right censored representing the state transition happening before, between and after the inspection respectively.

3.4 *Asset classification*

Asset deterioration rate may be influenced by many contributing factors, such as age, loading and environment (Fu and Devaraj, 2008, Wellalage et al., 2014). By classify assets into groups with the same factor levels, we can expect groups to have similar patterns of deterioration (Veshosky et al., 1994). Our framework provides two ways to adjust deterioration based on such factors:

1. **Grouping**: some factors are chosen to define groups of assets so that historical data can be pooled and used to learn assets deterioration parameters. The number of groups fixed the number of instances of the Group class in the abstract probabilistic model.
2. **Adjusting**: other factors are used to adjust the deterioration rates learnt from data. A range of studies has been developed to identify these impact factors, for example, closeness to the coast, galvanic response level and structure type are identified in Yianni et al. (2016) as the key contributing factors for railway bridge deterioration. The number of these factors fixes the number of instances of the class *Factor*. The importance of each factor in the aggregated effect is modelled by a weight and this is used to shift effect of the learnt parameters.

3.5 *Model instantiation*

The final stage of the process shown in Figure 5 is the instantiation of the BN. This set is provided by

the framework as the domain expert has expressed all the information needed in the steps described in sections 3.3 and 3.4.

As Figure 3 shows, the customisation process also defines the schema of the database that holds both configuration parameters and the records of assets and inspections results. The class of each asset is defined either directly or inferred from the values of factors held in the database.

Suppose that investigated asset is x, where x belongs to group g, which is similar to group h. We suppose that some time has elapsed since x was last inspected and we wish to estimate its current state. The following steps are involved to instantiate the BN:

1. Creating transition variables for each deterioration stage of asset classes g and h.
2. Creating hyper-parameters for each transition of group and at the population level. The group parameters approximate the population ones, using the similarity degree defined for the group.
3. Create the variables for the state of x and its adjustment by the factors (the ones that do not define the grouping). The values of the factor variables are extracted from the database and used as observations in the probabilistic calculation, but note that the model still operates if any of the values are missing (provided that priors have been provided).
4. Create variables for the all the inspection data—taken from the database—for both groups g and h. Starting from the current state, the inspection reports show either that one transition has occurred since the previous inspection or that no transition has occurred. By selecting the appropriate variant of data object both of these observations can constrain the transition time. It is even possible for more than one transition to have occurred between inspections.

For illustration, we assumed that it was known that group h needed to be include alongside data for assets in group g, the group to which x belongs. Two steps are involved in automating this. Firstly, we need to determine whether the number of observations in group g is sufficient to give a good estimate of the deterioration rate. We could evaluate this either from the variance of the learnt parameters or from a threshold value on the absolute number. The second step is to find the group that is closest in characteristics to g, perhaps by the proportion of the values of the factors that define the groups shared between the two groups.

3.6 *Inference and refinement*

Inference of the ground BN is performed automatically in the model-based machine learning framework. As suggested previously, probabilistic programming languages with an extensive list of inference algorithms can be adopted. For example, when dealing with hybrid Bayesian network that contains both discrete and continuous variables, Gibbs sampling can be used.

The query – to predict the unknown state of asset x – has to be expressed in domain terms and translated to a query on a BN variable. The result is a probability distribution over the possible states and further 'decision rules' or guidelines are needed to determine an action. For example, a small probability of the worst state of deterioration may determine that an urgent inspection is required.

Model evaluation can be made by comparison of different variants of the model, with the metrics such as predictive accuracy or computational speed. Further refinement of data sources (e.g. from other source domain), expert knowledge (e.g. different experts or different types of expertise), and variations in the models (e.g. different groupings of assets) are possible. This process repeats until a level of acceptable performance is accepted or it exhausts all the resource. As a future study, with the success of automatic inference software (Bishop, 2013), the refinement process can be made automatically with a defined threshold in a model-based machine learning framework.

4 REPRESENTING ASSET DETERIORATION MODEL WITH DIFFERENT STRUCTURE

Customised instantiation of this framework for practical uses have been developed, for example, general maintenance problems in Zhang and Marsh (2016), and rail bridge deterioration in Zhang and Marsh (2018). Since the focus of this paper is to show how the model-based asset deterioration framework to deal with different modelling assumptions that may happen in practice, we present a series of instantiation variants of a basic deterioration model, which is sufficient to convey the idea.

4.1 *The simple deterioration model*

The basic deterioration model is showed in Figure 1, and Table 1 shows its customisation setting. Notes that since the focus of this section is the model structure, we only show the setting that have influence on the shape of the model structure.

From the practice guideline, we decide there is only one transition (because it only has a binary state: working or failed), one group (representing the entire group has only one subgroup, the prior of

Table 1. The setting of the model in Figure 1.

	Transition distribution	Grading system	Prior	Sojourn time
Customisation	Exponential (λ)	Binary states	λ ~ Uniform (0, 0.05)	15 < sojourn < 35 12 < sojourn < 24 12 < sojourn < 36 sojourn = 24

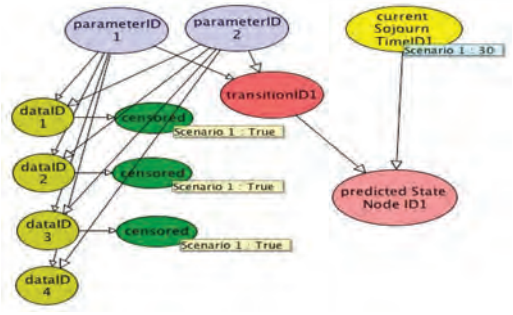


Figure 7. Deterioration follows a two-parameter distribution.

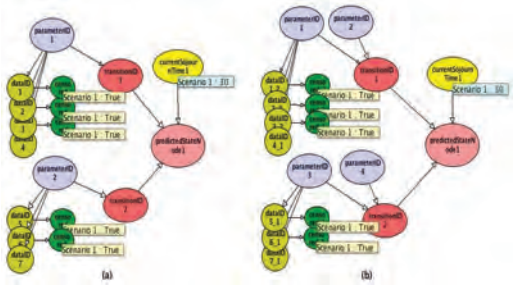


Figure 8. Deterioration under multi-states: (a) Markov chain based; (b) semi-Markov chain based.

the hyper-parameter becomes the prior of parameter), and one parameter (because its distribution is exponential). The expression of transition follows the distribution parameter in Parameter class, while the parameter's prior is given by experts, here is a uniform distribution between 0 and 0.05. Four sojourn time data are inferred from the inspection data, and three of them have censorships.

Based on this model, a range of variants can be extended to represent different assets, whose underlying modelling assumptions vary.

4.2 *Variant 1: Deterioration that follows a two-parameter distribution*

Extended from the basic model, Figure 7 presents a deterioration model of asset that follows a two-parameter distribution. Studies have found that transition probabilities between states of some assets are better fitted with distributions with two or even more parameters. For example, the two-parameter Weibull distributions for bridge deterioration are suggested in Ng and Moses (1998) and Sobanjo (2011).

This is achieved by the one-to-many relationship encoded in the relational database. Each instance of Transition class is linked to two identical instances of Parameter class.

4.3 *Variant 2: Deterioration under multi-states*

Asset may degrade with several stages representing the decrease of its functionality. For safety and reliability reasons, they can be rated from perfectly working to completely failure accompanied with several intermediate states. Extended from the last two subsections, Figure 8 shows two deterioration models for multiple states asset.

Degradation of asset with multi-states is modelled in the form of a Markov chain (Figure 8 (a)) by a sequence of states (represented by the transition nodes) representing the condition of an asset over time. Markov chain deterioration models are widely accepted in modelling most asset's life-cycle performance (Frangopol et al., 2004), but they also bear with the assumption that the transition probabilities between states follow the same stationary transition rate, which do not change over time. This property implies the sojourn time follows an exponential distribution regardless how long it has been in the current state (Ng and Moses, 1998). This is modelling is performed by the one-to-many relationship in the relational database: one instance of Asset class with two instances of transition class.

This restriction can be relaxed by semi-Markov model (Figure 8 (b)), which allows the modelling of transition probabilities to follow non-stationary distributions depending on current state and its next state. This extension enables the modelling of multi-state deterioration that follows multi-parameters based distributions. This modelling is performed by the many-to-many relationship: one instance of Asset class with two instances of transition class, and each instance of transition class links to two instances of parameter class.

4.4 *Variant 3: Learning from similar assets*

Assets classified into different groups may share similar deterioration rate, which gives a potential to learn from others. Two types of learnings from similar assets are presented:

1. Figure 9 shows an example of pooling all the available data to learn a universal distribution in the domain, and distinguish a specific asset by defining the influence of aggregated external factors on deterioration rates. A suggestion use of this form is in the situation when the entire population has little data. This is achieved by the instantiation of Factor class.
2. Figure 10 shows an example of pooling available data within its associated group to learn their own distribution but governed by their shared hyper-parameter introduced by hierarchical BN. This hyper-parameter helps the transfer learning of the weekly learnt group (typically target group with little data) from strongly learnt group (source groups with lots of data). A suggestion use of this form is in the situation when the groups are highly correlated. This is achieved by the introduction of the class hierarchy by adopting layer 3 data sources.

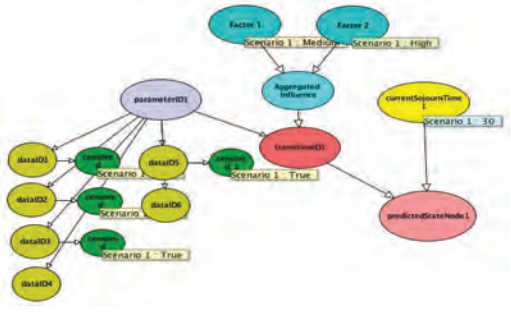


Figure 9. Distinguish an asset's deterioration from factors.

Figure 10. Learning from other groups hierarchically.

5 CONCLUSION AND FUTURE STUDY

We have argued the need to provide a generalised asset deterioration framework encoded by probabilistic relational models, which can be adapted to model assets with different modelling assumptions. The emerging field of model-based approach gives us a suitable formalism for separating specifications from analysis techniques, and we have applied this to asset deterioration.

We also used several variants of the deterioration model to demonstrate how it can be adapted to a variety of applications with different modelling assumptions. In the future, we hope to extend the amounts of models in the model library, and extend the applications to more safety and reliability related problems.

ACKNOWLEDGEMENT

This work is supported by European Research Council (Funding code: ERC-2013-AdG339182-BAYES_KNOWLEDGE), and Agena Ltd for software support. H.Z. is supported by China Scholarship Council (CSC)/Queen Mary Joint PhD scholarships.

REFERENCES

Arnold, A., Point, G., Griffault, A. & Rauzy, A. (1999) The AltaRica formalism for describing concurrent systems. *Fundamenta Informaticae,* 40, 109–124.

Bishop, C.M. (2013) Model-based machine learning. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371.

Estefan, J.A. (2007) Survey of model-based systems engineering (MBSE) methodologies. *Incose MBSE Focus Group,* 25.

Frangopol, D.M., Kallen, M.J. & Van Noortwijk, J.M. (2004) Probabilistic models for life-cycle performance of deteriorating structures: review and future directions. *Progress in Structural Engineering and Materials,* 6, 197–212.

Fu, G. & Devaraj, D. (2008) *Methodology of Homogeneous and Non-homogeneous Markov Chains for Modelling Bridge Element Deterioration*, Michigan Department of Transportation.

Ghahramani, Z. (2015) Probabilistic machine learning and artificial intelligence. *Nature,* 521, 452–459.

Guler, H., Jovanovic, S. & Evren, G. (2011) Modelling railway track geometry deterioration. *Proceedings of the Institution of Civil Engineers-Transport.* Thomas Telford Ltd.

Koller, D. (1999) Probabilistic relational models. *International Conference on Inductive Logic Programming.* Springer.

Laskey, K.B. & Mahoney, S.M. (2000) Network engineering for agile belief network models. *IEEE Transactions on knowledge and data engineering,* 12, 487–498.

Lisagor, O., Kelly, T. & Niu, R. (2011) Model-based safety assessment: Review of the discipline and its challenges. *Reliability, Maintainability and Safety (ICRMS), 2011 9th International Conference on.* IEEE.

Memarzadeh, M., Pozzi, M. & Kolter, J.Z. (2016) Hierarchical modeling of systems with similar components: A framework for adaptive monitoring and control. *Reliability Engineering & System Safety,* 153, 159–169.

Mendenhall, W., Beaver, R.J. & Beaver, B.M. (2012) *Introduction to probability and statistics*, Cengage Learning.

Morcous, G. (2011) Developing deterioration models for Nebraska bridges. *M302 Final Report 26-1122-0003-001.*

Neil, M., Fenton, N. & Nielson, L. (2000) Building large-scale Bayesian networks. *The Knowledge Engineering Review,* 15, 257–284.

Ng, S.-K. & Moses, F. (1998) Bridge deterioration modeling using semi-Markov theory. *A.A. Balkema Uitgevers B. V, Structural Safety and Reliability.,* 1, 113–120.

Pfeffer, A. (2009) Figaro: An object-oriented probabilistic programming language. *Charles River Analytics Technical Report,* 137, 96.

Prosvirnova, T., Batteux, M., Brameret, P.-A., Cherfi, A., Friedlhuber, T., Roussel, J.-M. & Rauzy, A. (2013) The altarica 3.0 project for model-based safety assessment. *IFAC Proceedings Volumes,* 46, 127–132.

Sobanjo, J.O. (2011) State transition probabilities in bridge deterioration based on Weibull sojourn times. *Structure and Infrastructure Engineering,* 7, 747–764.

Veshosky, D., Beidleman, C.R., Buetow, G.W. & Demir, M. (1994) Comparative analysis of bridge superstructure deterioration. *Journal of Structural Engineering,* 120, 2123–2136.

Wellalage, N.K.W., Zhang, T. & Dwight, R. (2014) Calibrating Markov Chain–Based Deterioration Models for Predicting Future Conditions of Railway Bridge Elements. *Journal of Bridge Engineering,* 20, 04014060.

Welte, T. & Eggen, A. (2008) Estimation of sojourn time distribution parameters based on expert opinion and condition monitoring data. *Probabilistic Methods Applied to Power Systems, 2008. PMAPS'08. Proceedings of the 10th International Conference on.* IEEE.

Yianni, P.C., Neves, L.C., Rama, D., Andrews, J.D. & Dean, R. (2016) Incorporating local environmental factors into railway bridge asset management. *Engineering Structures,* 128, 362–373.

Zhang, H. & Marsh, D.W.R. (2018) Generic Bayesian network models for making maintenance decisions from available data and expert knowledge. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, doi.dx.org/10.1177/1748006X17742765.

Zhang, H. & Marsh, W. (2016) Bayesian network models for making maintenance decisions from data and expert judgment. *European Safety and Reliability Conference 2016 (ESREL 2016).* Glasgow, CRC Press.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Industry 4.0 and real-time synchronization of operation and maintenance

J. Vatn
Department of Mechanical and Industrial Engineering, NTNU—Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: Industry 4.0 represents a trend in manufacturing which includes cyber-physical systems, the Internet of things, cloud computing and cognitive computing. Cyber-Physical Systems (CPS) refers to smart systems that include engineered interacting networks of physical and computational components. The term digital twin refers to a digital replica of physical assets, processes and systems that can be used in real time for control and decision purposes. The digital twin representation is seen as a prerequisite for effective synchronization of operation and maintenance within the manufacturing industry as well as in other industries. The relation between production plans and activities and actual production can to some extent be described by deterministic. The relation between maintenance plans and activities and the production system availability on the other and requires probabilistic representation. The term stochastic digital twin is therefore introduced. An ambition of Industry 4.0 is to support real-time processing whenever possible. This paper discusses elements of Industry 4.0. A case study is provided to demonstrate these terms and challenges to the mathematical modelling required for optimal synchronization of operation and maintenance.

1 INTRODUCTION

1.1 *Background*

Nowadays Industry 4.0 and digitalization are frequently used terms for the changes that are taking place in industry, civil engineering, transportation, public services and so on. The "4.0" refers to the forth industrial revolution and points to the opportunities communication over the internet gives with respect to real-time control of processes at almost every level. Industry 4.0 and related concepts as cyber-physical systems, internet of things, cloud computing and digital twins give new opportunities for both production and maintenance, but even more important the synchronization and coordination of the two.

1.2 *Objective*

The objective of this paper is to clarify basic terms and elaborate on basic elements of Industry 4.0 in relation to real-time synchronization of operation and maintenance. A case study from the railway sector is used to exemplify the concepts.

2 DEFINITIONS AND CONCEPTS

Industry 4.0 is a collective term particularly used in manufacturing to emphasize technologies and concepts of value chain organizations. Further the terms Cyber-Physical Systems, the Internet of Things, Cloud computing and the Digital Twin are often used in relation to Industry 4.0. Although the term originates from the manufacturing industry, the elements of Industry 4.0 are relevant for most businesses.

The current usage of the term Industry 4.0 has been criticized as essentially meaningless. The 4.0 points to the forth industrial revolution under a premise that digitalization is the really new thing. But why digitalization and not Nano technology? Further, the content of Industry 4.0 also seems to vary from industry to industry, and from author to author. From a scientific point of view it might therefore be better to avoid a precise definition but rather focus on Industry 4.0 *elements*.

The aim of this paper is to shed light on Industry 4.0 elements that are relevant for the interaction between production and maintenance. Here production has a very broad meaning, it could cover manufacturing, logistics, transportation systems, hospitals, power supply and so on.

The *Internet of Things* (IoT) is the network of physical devices, production facilities, cars, airplanes and in general items embedded with electronics, software, sensors, actuators, and network connectivity which enable these objects to connect and exchange data. Each "thing" is able to interoperate within the existing Internet infrastructure.

The IoT allows objects to be sensed or controlled remotely across existing network infrastructure, creating opportunities for more direct integration of the physical world into computer-based systems. When IoT is augmented with sensors and actuators, the technology becomes an instance of the more general class of *Cyber-Physical Systems* (CPS).

Cyber-physical systems (CPS) refers to smart systems that include engineered interacting networks of physical and computational components.

Cloud computing is an information technology paradigm that enables access to shared pools of configurable system resources. The companies can focus on their core businesses instead of expending resources on computer infrastructure and maintenance. Downsides of such a strategy could be unexpected operating expenses if administrators are not familiarized with cloud-pricing models and vulnerabilities and security issues. In some presentations the term Internet of Services (IoS) are used rather than cloud computing.

The term *digital twin* refers to a digital replica of physical assets, processes and systems that can be used in real-time for control and decision purposes. The digital twin representation is seen as a prerequisite for effective synchronization of operation and maintenance within the manufacturing industry as well as in other industries. The relation between production plans and activities and actual production can to some extent be described by deterministic models. The relations between maintenance plans and activities and the production system availability on the other and require probabilistic representations.

A *stochastic digital twin* is a computerized model of the stochastic behaviour of a system where the model is updated in real time based on sensor information and other information accessed via the internet and the use of cloud computing resources.

To be useful a digital twin needs "what-if" capabilities. This means that the decision makers, i.e., humans or computers, shall be able to "ask" the digital twin what will be the consequences of various decisions. For a stochastic digital twin this means that the "answer" is given as a set of probability statements.

A *real-time model* is a model where it is possible to obtain values of system performance and system states in real-time. With real-time we mean that data referring to a system is analysed and updated at the rate at which it is received. As for the digital twin a real-time model typically connects to the "real world" via the IoT, although other means of communication is also possible. A real-time model is also referred to as an on-line model.

A *test model* is a mathematical model describing relations between future and current values of the variables of interest, but where we are not able to monitor system performance and system sates in real-time. Such a model is often referred to as an off-line model. A test model is still valid in order to establish decision rules to be used in real-time.

Most methods and models used in production planning and optimization as well as in maintenance planning and optimization are off-line models. These models can be used for establishing optimal strategies, but they can not give real-time decision support. Areal-time model is often used to describe a limited part of a system, whereas a digital twin aims at giving a complete digitalized representation of the system and decision processes.

A *real-time decision support systems* is a system where relevant data is collected and processed into relevant information in real time. This means that the raw data stream is automatically collected and processed into information. Information is further interpreted in such a way that it gives meaningful decision support.

A *real-time execution system* is a system which implement algorithms to determine optimal decisions at time, and then execute these decisions. An example of a real-time execution system is an Automated Replenishment Program (ARP). The aim is to provide automated replenishment of products based on real-time demand information to the production, warehouses and distribution processes in the supply chain. This corresponds to real-time control in control theory. Similarly for maintenance a real-time execution system will automatically issue a work order with task descriptions and due date.

Predictive maintenance builds on the idea to utilize the condition of a component and the future expected loads in order to judge the correct time for "hard" maintenance such as overhaul, replacement of worn parts, calibration and so on. Sensor technology is usually used to capture the condition of components or a system, and the term 'condition monitoring' is often used to describe the collection and analysis of state data relevant for predictive maintenance. It should be noted that manual inspection and use of "human sensors" to capture noise, smell, vibration could also be treated as condition monitoring.

3 THE DIGITAL TWINS

This section presents principal elements of the digital twins for maintenance and production.

3.1 *Maintenance*

To a large extent the Computerized Maintenance Management System (CMMS) could be seen as a digital twin for maintenance. Principal content found in the CMMS are the asset register covering all components, the Preventive Maintenance (PM) program covering the type of maintenance and the

plan for maintenance. The CMMS will also contain required spare parts, resources and tools for conducting maintenance and so on. But there is relevant information not found in the CMMS which is essential for the stochastic digital twin to be developed. First of all the CMMS has no inherent mathematical models to be used for degradation development and time to failure. Further information regarding component condition is often not part of the CMMS, and needs to be obtained from stand alone systems operated in parallel to the CMMS. Further the CMMS is not connected to the supervisory control and data acquisition (SCADA) system and other systems giving information regarding process parameters and future loads from production and the environment.

It is beyond the scope of this presentation to write out the details regarding the content of a stochastic digital twin for maintenance. For illustrative purposes and for use in the case study presented later a very simple digital twin is presented in the following. Although we in many situations can do much better, the classical failure rate function is used as a basis. The situation relates to so-called delay time models (Christer 1987), often referred to as PF-interval models. The situation is as follows: A component is put into service at time $t = 0$. Then after a random time the component enters a degraded state. This state is often referred to as a potential failure. It is assumed that a condition monitoring activity can reveal such a potential failure with some detection probability, say $1 - q$. If no action is taken the component will fail after another random time T_{PF}. A Cox proportional hazard rate function (Cox 1972) is used as a basis for formulating the failure rate function, $z(t) = f(t)/R(t)$ for T_{PF} (t is running time after the potential failure has occurred). T_{PF} is often referred to as the PF-interval, and the corresponding failure rate function is:

$$z(t \mid \mathbf{y}, \overline{\mathbf{x}(t)}) = z_0(t) e^{\beta_1 \mathbf{y}} e^{\beta_2 \overline{\mathbf{x}(t)}} \quad (1)$$

where $z_0(t)$ is a baseline failure rate function, typically on the form $z_0(t) = \alpha\lambda^{\alpha} t^{\alpha-1}$ in the Weibull case. $\mathbf{y}$ is a vector of state variables at the point of time of the potential failure is observed, $\overline{\mathbf{x}(t)}$ and is the average load profile t time units ahead. β_x and β_y and are regression coefficient vectors established by for example statistical analysis of data.

The failure rate function in eq. (1) is a classical model and it could be questioned whether this model represent at digital twin. A prerequisite for being at least a part of a digital twin is that $\mathbf{y}$ could be accessed from sensor readings and communicated via the internet. Further $\overline{\mathbf{x}(t)}$ needs to be accessed in real time from enterprise resource planning (ERP) systems and other system for future production plans.

If eq. (1) is part of the stochastic digital twin we may now "ask" for the probability of failure if we wait for example t time units before the potential failure is fixed:

$$F(t \mid \mathbf{y}, \overline{\mathbf{x}(t)}) = 1 - e - \int_0^t z(u \mid \mathbf{y}, \overline{\mathbf{x}(u)}) du \quad (2)$$

Only a few aspect of the "maintenance twin" are elaborated here. For real applications it will be an enormous amount of work to structure the raw data, information, knowledge, models and so on to have a digitalized stochastic maintenance twin.

3.2 *Production*

There are so many aspects to deal with when it comes to production and logistic optimization that we will not even make an attempt to cover these in this presentation. However, with respect to maintenance there are some important aspects that we will emphasize when setting up the digital twin for production.

3.2.1 *Objective function*

Operations Research (OR) is the systematic approach to optimize production under various constraints (Phillips, Ravindran, & Solberg 1976). The objective function, Z, is typically the quantity to maximize or minimize with respect to some vector of decision variables, say $\mathbf{x}$, i.e., $Z = Z(\mathbf{x})$.

3.3 *Constraints and conditions*

Usually there are constraints to take into account in the optimization, for example a set of functions, say $g_j(\mathbf{x})$ should all be positive. In addition to these constraints we also have to optimize $Z = Z(\mathbf{x})$ subject to $\mathbf{S}$, where $\mathbf{S} = [s_1, s_2, \ldots, s_n]$ is the state vector of the components in the system. For example $s_i = 1$ could represent that component i is functioning, and $s_i = 1$ represents a fault state.

It should be emphasized that both the objective function and the constraints and conditions are changing all the time. It is therefore required to have real-time access via the internet to the "physical" plant, existing orders, inventory levels and so on.

3.4 *Maintenance interaction*

The digital twin for production will also be a *stochastic* twin due to the probabilistic nature of production optimization. From classical OR examples variability in supply and demand are the main sources for uncertainty. However, we will focus on the relation to maintenance. Important aspects that need to be structured as part of the digital twin for production are:

- Slots for maintenance, i.e., possible opportunities for doing maintenance

- Specifications of how utilizing possible slots will affect the objective function $Z = Z(\mathbf{x})$ and the constraints $g_j(\mathbf{x})$
- Specification of possible "relaxes" in production, for example avoid running a component with full load if a "potential failure" has been revealed
- Specifications of how such "relaxes" will affect the objective function $Z = Z(\mathbf{x})$ and the constraints $g_j(\mathbf{x})$.

Note that the objective function $Z = Z(\mathbf{x})$ in traditional OR does not include maintenance. Since the objective of this paper is to investigate synchronization and coordination of activities in the production and maintenance departments, the objective function should cover both departments.

4 CASE STUDY

4.1 *Introduction*

A railway example is used to demonstrate challenges in synchronization and coordination of activities in the production and maintenance departments. Only few aspects are dealt with, and issues related to really establish the stochastic digital twins and have them to play together is not addressed in this presentation. One aspect of "digitalization" within maintenance is related to increased used of predictive maintenance. Turnouts (switches) are important components in the railway infrastructure, and failure of a turnout will usually give large problems with the circulation, and delays are expected. Although various condition monitoring techniques exist for turnarounds, they have not been implemented on the Norwegian railway network due to high cost. In Norway Bane NOR is a state-owned company responsible for the Norwegian national railway infrastructure. In recent years Bane NOR has been running a test project on a simplified predictive maintenance strategy for turnouts based on measuring only power as function of time when the traction motor is activated to change the position of the turnout. The time required for changing the position of the turnout varies from one to up to 20 seconds. The idea is that the power as function of time for each individual turnout is a "signature" for that turnout, and deviation from that signature could be seen as a potential failure as discussed in Section 3.1. The main advantage of this system is that the information is available more or less "free of charge". The challenge is to use it in an efficient way.

The system has been piloted over a period of almost 3 years. As part of the pilot project data have been analysed for one of the turnouts. In a follow up project it is planned to conduct more comprehensive analyses. During the test period 11 failures were observed. For 3 of these failures a potential failure was not observed at all. Thus the reliability of the condition monitoring system is only some 70%. The analysis was conducted by visual analysis of the power/time curve for all movements of the turnout for a period of 10 days prior to the failure. More comprehensive analysis could obvious give a higher reliability.

4.2 *The PF-interval model*

The average PF-interval, i.e., the estimate of $E(T_{PF})$ were found to be 80 hours. However, 2 failures had an observed PF-interval of less than 3 hours. To estimate the parameters in the failure rate function in eq. (1) assuming a Weibull distribution and ignoring covariates $\mathbf{y}$ and $\mathbf{x}(t)$ we may use the following procedure:

1. Let x be such that $F_T(x) = p_x$ where both x and p_x are known. It can be shown that the following iterative scheme may be used to estimate α: $\alpha_{i+1} = \frac{\ln(-\ln(1-p_x))}{\ln(x\Gamma(1+1/\alpha_i)/E(T_{PF}))}$
2. For the location parameter we use set: $\lambda = \Gamma(1 + 1/\alpha)/E(T_{PF})$.

In the example we had $p_x = 2/(11 - 3)$ and $x = 3$ hours, and $E(T_{PF}) = 80$ hours. Applying the procedure this will give $\hat{\alpha} \approx 0.49$ and $\hat{\lambda} = 0.026$. It should be noted that $\alpha < 1$ means that the PF-interval is not very consistent. The reason for the low value of α is that we are mixing several failure mechanisms. There are three main failure mechanisms with quite different characteristics here, i.e., snow and ice with short PF-interval, lack of lubrication with a medium PF-interval, and mechanical failure with a rather long PF-interval. This means that we need to apply the procedure above for each separate failure mechanisms. From the failure statistics obtained from the pilot project we do not have sufficient number of observations to apply the procedure above. For the case study we will proceed with assuming that the failure mechanism is related to lack of lubrication and without any statistical support we set $\hat{\alpha} = 2$ and $\hat{\lambda} = 0.0246$ corresponding to $E(T_{PF}) = 36$ hours.

4.3 *The initial cost model*

The operational hindrance cost of executing a "hard maintenance" task, and the cost of a failure depends on the position of the turnout, the time of the day, the traffic and so on. Therefore an example situation is presented in the following.

The location of the turnout is assumed to be on a part of the line where access only can be made by means of a work train. We assume a single track line where access by the work train will disturb the circulation. Investigating the time table for today four opportunity windows have been identified. They are shown in column 1 in Table 1. The first

Table 1. Optimization results.

t (hours)	Delay (min)	c_{PM}	c_F	c_{Tot}
3	30	18 500	441	18 941
5	15	11 750	1 218	12 968
7	10	9 500	2 370	11 870
9	0	5 000	3 880	8 880

Table 2. Optimization results—with covariates.

t (hours)	Delay (min)	c_{PM}	c_F	c_{Tot}
3	30	18 500	822	19 322
5	15	11 750	3 422	15 172
7	10	9 500	9 812	19 312
9	0	5 000	22 562	27 562

three of these will, however, cause delays in circulation. The expected delay minutes for each window is shown in column 2 in Table 1. In average there are 150 passengers per train and a minute delay cost per passenger of 3 NOKs is used by Bane NOR. In addition to the delay cost there is a fixed cost of NOK 5 000 for ordering the work train and associated personnel cost. If the failure can not be "caught" in due time, the expected total delay is 3 hours. The cost equation to minimize is:

$$C(t) = c_{PM}(t) + c_U F(t) \tag{3}$$

where $c_U = 3 \cdot 150 \cdot 60 \cdot 3 = 81\,000$ NOKs, $F(t)$ is given by eq. (2). Table 1 shows that in this situation one should utilize the last maintenance window since the circulation is not affected, and the probability of failure is still rather low. It can be shown that if the failure mechanism is ice and snow, and assuming $E(T_{PF}) = 10$ hours and $\hat{\alpha} = 2$ the risk is much higher, and one should rather use the first opportunity. Note that the optimization here is seen from the maintenance department, i.e., the only "production" related cost is the increased c_{PM}-cost by rushing the maintenance.

4.4 *The refined cost model*

So far the covariates $\mathbf{y}$ and $\overline{\mathbf{x}(t)}$ have been ignored. We now introduce two variables, y which is a measure of degradation at the point of detection of the potential failure, and x as the number of times per hour the turnout will be operated. The proposed Cox proportional hazard model reads:

$$z(t \mid y, x(t)) = z_0(t) e^{\beta_y y + \beta_x x t} \tag{4}$$

For simplicity we have assumed that the number of train passages per hour is constant over the day. From the case study we do not have sufficient data to estimate β_y and β_x. We will therefore proceed with illustrative values for these parameters. That is, for the example we proceed with $\beta_y = \ln 2 \approx 0.69$ and $\beta_x = 0.1 \ln 2 = 0.069$.

Now, assume that at the time of the potential failure we assess $y = 0.15$ by analysing the power vs time curve from the condition monitoring system, and further from the time table we wind $x = 2$. Table 2 shows the result when the covariates are taken into account. Compared to the original situation we have to advance the point of time for doing hard maintenance, i.e., lubrication and required adjustments. The number of times per hour we operate the turnout, x, is a decision variable seen from operation. Since the failure rate function is increasing with increasing value of x, we should investigate whether it pays off to reduce x. Now, assume that wecan completely remove the need for operating the turnout by changing the station where trains are crossing. This corresponds to set $x = 0$ in the model. Rerunning the model shows that the optimal value of t is $t = 9$. The cost has been reduced from 15 172 to 12 249, i.e., a total saving of ≈ NOK 3 000. However, if this causes total delays of more than 7 minutes the delay cost will be higher than the savings. For the railway case it seams unrealistic that changing the crossings for the actual station for an entire day will not cause more than 7 minutes of total delay.

A first attempt to formalize such a "relax" strategy is to add an extra cost term in the objective function, $c_R(x)$:

$$C(t,x) = c_{PM}(t) + c_U F(t \mid y, x) + c_R(x) \tag{5}$$

The joint optimization of t and x is not pursued further in this presentation.

5 DISCUSSION

The objective of this paper has been to investigate "Industry 4.0 solutions" to facilitate synchronization and coordination of operation and maintenance. By a "paper exercise" it is rather easy to demonstrate how this can be done, and potential savings. This section discusses challenges when such ideas are to be implemented for real systems.

5.1 *Slots for maintenance and consequences for the production model*

In order to synchronize and coordinate production and maintenance it is essential that the digital twin on request can provide time slots for maintenance and evaluate the production consequences for each possible slots. In the example we assumed that "some" could establish the time slots at 3,5,7 and 9 hours. Here, "some" could be a train manager

at the Train Control Centre (TCC). But this is not really a part of the "digital twin" for production. To develop a digital twin all production plans, cost optimization functions etc. need to be implemented in a computerized system supported with algorithms to both find possible slots, and do calculations to evaluate the consequences. For the railway example we are far from realizing such systems. To the author's knowledge the situation is the same in most Norwegian industries.

The way forward is therefore to develop simplified production models. For example in Norway a simplified circulation model for use by the TCC-personnel upon traffic deviations to assist planning has been developed. The model acts like a "what-if" tool that can simulate the consequences if crossing is moved to station A rather than on the scheduled station B. It is hard to spot significant achievements here unless modelling competence within the companies is significantly increased. A vision behind "cloud computing" is that ready to use models could just be plugged in whenever needed. But still this is a vision.

5.2 *Specification of possible "relaxes" in production*

In the example, and in many real case situations a mitigating measure upon a component degradation is to reduce the load on that component to increase residual life. An even more realistic example than the railway example is maintenance and operations of wind farms. A wind farm can be difficult to access in periods of the year due to harsh weather conditions. Upon a potential failure of for example the main bearing of the turbine it may be better to close down the turbine in situations with high wind loads. Although this will reduce the power produced for some hours, it might prevent a failure which would have made the turbine unavailable for weeks and even months.

The digital twin for production therefore need to respond upon request on what are the possible "relaxes" that could be made in production that will have a positive impact on residual life of a component. In addition to respond on *what* can be done, the digital twin also needs to specify the consequences, for example by quantifying the reduced production.

5.3 *Maintenance models*

The literature in the field of maintenance optimisation produces every year a huge number of models. Very few of these models are used in practice. One reason for this is that it is hard to get access to statistical data for estimation of model parameters. In our example we need to estimate α, λ, β_Y and β_X. Further this have to be done for all failure mechanisms. We can easily imagine an enormous workload. Next, comes the question whether the Cox-proportional hazard model is the appropriate model to use. It is rather simple, but it does not really take into account the physical aspects of the phenomenon causing a failure.

The prospects for the maintenance twin is therefore also not that promising. Again, starting with a set of rather simplified models seems a natural first step.

5.4 *Machine-learning*

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed. Machine-learning is quite different from the model based approach advocated here. A strength of machine-learning is it's efficiency to produce huge amount of results without the explicit need to do all the "hard work". From a model based approach perspective most of us are reluctant to just "let the computer work out the answers". However, for sub-problems like establishing a failure model, looking into machine learning approaches are more acceptable.

5.5 *Real-time execution models*

An objective of Industry 4.0 solutions is to have automated decision processes. For simple situations such as replenishment in retail we see automated replenishment policies. However for mixed problems as discussed here it is a long way to go to get trust in real-time execution models.

6 CONCLUSIONS

This paper has discussed steps in synchronization of operation and maintenance. An example was provided to illustrate some of the challenges and opportunities this will give. With idealized examples and simplified assumptions we are able to carry out relevant modelling. Still, these models are test (off-line) models and integration into real-time (on-line) models require significant effort. To succeed it is recommanded to start with a relative small set of standardized models for critical processes in the value chain of the company.

REFERENCES

Christer, A. (1987). Delay-time model of reliability of equipment subject to inspection monitoring. *Journal of the Operational Research Society 38 (4)*, 329–334.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B. 34 (2)*, 187–220.

Phillips, D., A. Ravindran, & J. Solberg (1976). *Operations research: Principles and practice*. New York: John Wiley & Sons.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Bayesian update and aperiodic maintenance policy for deteriorating systems with unknown parameters

E. Mosayebi Omshi
School of Mathematics, Statistics and Computer Sciences, University of Tehran, Tehran, Iran

A. Grall
ICD/LM2S – UMR CNRS 6281, Université de Technologie de Troyes, Troyes, France

S. Shemehsavar
School of Mathematics, Statistics and Computer Sciences, University of Tehran, Tehran, Iran

ABSTRACT: Many papers are dedicated to maintenance strategies of deteriorating systems and almost all of them share a common assumption that the parameters of the degradation process are known. In this paper, we deal with a dynamic and aperiodic condition-based maintenance of single-unit systems with unknown parameters of deterioration. It has been considered that the deterioration is governed by an Inverse Gaussian (IG) process. The time interval between two successive inspections is scheduled based on the Remaining Useful Life (RUL) of the system. The Bayes method is employed to use the available information of degradation paths and update the information about parameters during the time. The proposed maintenance decision rule aims to avoid too frequent and costly inspections by implementing an aperiodic planning. The decision process is dynamically improved with the successive Bayesian update of the degradation parameters. The ability of the proposed modeling framework to drive the Bayesian update while controlling the number of inspection is analyzed through numerical experiments. The global maintenance cost is considered over a finite time horizon.

1 INTRODUCTION

Maintenance decision making is of prime importance to improve the global performance of industrial systems and structures during their useful life. Different strategies for maintenance can be considered, see Wang (2002). Their choice depends mainly on the system failure process and on the associated monitoring system. This paper is related to systems or structures which gradually deteriorates from initial "new" state to failure and whose degradation level can be perfectly observed through inspections. In this general context, predictive and Condition-Based Maintenance (CBM) strategies are understandably relevant candidates. Actually, they allow to adapt decisions about maintenance actions and the inspection scheduling to the current system state and possibly to remaining lifetime estimation including prospective usage. The global design of a predictive or condition-based maintenance policy requires mathematical modeling which brings together deterioration model, maintenance decision-rule and cost function for policy assessment and optimization.

Stochastic processes are natural choices to model deterioration over time. The Wiener process, the gamma process, and the Inverse Gaussian (IG) process are commonly used, Alaswad and Xiang (2017). The last two processes have the particularity to exhibit monotone evolutions of degradation indicator. The most popular stochastic process employed in maintenance literature is the gamma process; see van Noortwijk (2009) for a review in this subject. The IG process is an alternative stochastic process introduced by Wang and Xu (2010) to the reliability literature. Ye and Chen (2014) precisely investigated the IG process properties and mentioned its advantages. Having a clear physical interpretation, flexibility in incorporating random effects and covariates or even prior information, and the existence of explicit formulas for important related functions are some of such advantages. Although the IG process is employed progressively (see e.g. Pan et al. (2016), Peng et al. (2014), Peng (2015), and Ye et al. (2014)), the literature on CBM policies are scarce. Chen et al. (2015) investigated the optimal CBM policy with periodic inspections when the system degradation follows an IG process with random-drift model.

In this paper, we discuss the condition-based maintenance of a single unit system whose degradation follows an IG process. The considered system has degradation parameters that can be different from one unit to the other among a population. Hence it is considered that parameters of the model are unknown but a prior information like expert judgment is available. As an illustrative example, one can consider the degradation of roads subject to longitudinal cracking. For a road section, the degradation level corresponds to the ratio of the fissured length on the total length of the section. A maintenance action is a complete resurfacing of the road section. The dynamics of crack propagation depends on several parameters which may represent features of the road foundation and are unknown. As a consequence, there are some differences from a section to the others and the specific parameters for a given section are unreachable. An adaptive Bayesian method is employed to update the information about parameters after inspections as the degradation state of the system is measured. In order to reduce unnecessary inspections and to control maintenance actions, an aperiodic maintenance policy is considered. The time interval between two successive inspections depends on the system current degradation level and on degradation prediction. The knowledge about the degradation process can be improved on-line through the adaptive Bayesian method, simultaneously. The aim is to introduce this on-line improvement structure within the maintenance decision rule: the next inspection is planned based on the remaining useful life (RUL) of the system. At each inspection, the RUL is estimated according to the last update of degradation parameters.

The aim of the paper is twofold. First, the performance of the Bayesian update process has to be checked especially because of the inspection schedule which is driven by the aperiodic maintenance decision rule. Hence the times for updates are nonperiodic and scarce. Secondly, the effect of on-line updates of RUL prediction on the global maintenance cost is assessed and analyzed for different alternatives.

The remainder of the paper is organized as follows. The next section is devoted to the description of the stochastic process for degradation modeling and RUL prediction. Section 3 is related to the proposed maintenance policy including the procedure for on-line adaptation of the decision based on Bayesian update and the cost considered for assessment. Some numerical simulations are proposed in Section 4 to illustrate the behavior of the considered policy. Section 5 concludes and shows further possible extensions of this paper.

2 MODEL DESCRIPTION FOR RUL CALCULATION

Consider a single component system which is subjected to degradation. Let X_t denote the degradation state of the system at time t. In the absence of repair or replacement actions, the evolution of the system degradation is assumed to be strictly increasing. Then X_t can be modeled by an increasing stochastic process. Moreover, other assumptions are considered:

- The initial state X_0 is 0.
- The system is failed if its degradation crosses a critical threshold level L.
- The system failure is not self-announcing and if it fails, it remains failed until the next inspection. This down time imposes some extra cost.

2.1 *Stochastic degradation process*

The IG process is a stochastic process with independent, non-negative increments such that for each $t > s \geq 0$, the increment $Y = X_t - X_s \sim IG(\mu(t - s), \lambda(t - s)^2)$ with following probability density function (pdf):

$$f(y) = \sqrt{\frac{\lambda(t-s)^2}{2\pi y^3}} \exp\left\{\frac{-\lambda(y-\mu(t-s))^2}{2\mu^2 y}\right\}, \qquad (1)$$

where $\mu, \lambda > 0$. Then, the mean and the variance of X_t are μt and $\mu^3 t/\lambda$, respectively.

Known parameters are common assumptions in the maintenance literature. However, in practice, the model parameters are unknown and must be estimated. Here, we use the Bayesian approach to overcome this difficulty, also we consider that all information in hand like expert opinions are reflected in prior information. The assumption of the conjugate priors can be a good option to reduce the complexity of finding the posterior distribution. To this end, let λ have the gamma density function,

$$f(\lambda) = \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} \exp\{-\lambda/\beta\}, \qquad (2)$$

and let $\delta = 1/\mu$ have the conditional normal density function with mean ξ and variance σ^2/λ,

$$f(\delta \mid \lambda) = \sqrt{\frac{\lambda}{2\pi\sigma^2}} \exp\left\{-\frac{\lambda(\delta-\xi)^2}{2\sigma^2}\right\}. \qquad (3)$$

Then, the joint prior distribution of (δ, λ) is given by $f(\delta, \lambda) = f(\delta|\lambda)f(\lambda)$. Furthermore, to avoid the probability of getting non-feasible degradation slopes, it is supposed that $P(\delta \leq 0)$ is negligible.

In other words, the degradation model can be rewritten as:

$$\begin{aligned} &X_t \mid \mu,\lambda \sim IG(\mu t,\lambda t^2),\\ &\delta \mid \lambda \sim N(\xi,\sigma^2/\lambda),\\ &\lambda \sim \Gamma(\alpha,\beta). \end{aligned} \tag{4}$$

2.2 *Degradation model update*

Assuming the priors mentioned above, the posterior distribution of (δ, λ) can be obtained as soon as the new observations are available. Let $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ be new observations of the system's degradation state at times t_1, t_2,…,t_n. It can be shown that the joint posterior distribution of (δ, λ) is:

$f(\delta, \lambda|\text{Data}) = f(\delta|\lambda, \text{Data}) f(\lambda|\text{Data})$,

where

$$f(\delta \mid \lambda,\text{Data}) = \sqrt{\frac{\lambda}{2\pi\sigma^{2*}}}\exp\left\{-\frac{\lambda(\delta-\xi^*)^2}{2\sigma^{2*}}\right\},$$
$$f(\lambda \mid \text{Data}) = \frac{\lambda^{\alpha^*-1}}{\Gamma(\alpha^*)\beta^{*\alpha^*}}\exp\{-\lambda/\beta^*\}.$$

The updated hyperparameters are:

$$\alpha^* = \alpha + n/2, \qquad \beta^* = (1/\beta + 1/D)^{-1},$$
$$\xi^* = B/A \quad \text{and} \qquad \sigma^{2*} = A^{-1};$$

where

$$A = \sum_{i=1}^{n}\Delta x_i + \frac{1}{\sigma^2}, \qquad B = \sum_{i=1}^{n}\Delta t_i + \frac{\xi}{\sigma^2},$$
$$C = \sum_{i=1}^{n}\frac{(\Delta t_i)^2}{\Delta x_i} + \frac{\xi^2}{\sigma^2}, \quad D = \frac{1}{2}\left(C - \frac{B^2}{A}\right),$$

and $\Delta x_i = X_{t_i} - X_{t_{i-1}}$, $\Delta t_i = t_i - t_{i-1}$ when X_{t_0} is considered as the degradation level of the current time t_0. Then, new observations can be employed to update the information about δ and λ through the Bayesian method. It can happen at each inspection ($n = 1$) or at the end of each cycle (n would be the number of inspection in the cycle).

2.3 *Remaining useful life*

Herein, the objective is to determine the distribution of the remaining useful life (RUL) of the system. At a given time t and with knowing the current state, the RUL of the system can simply be computed as a first passage time when the degradation X_t crosses the threshold L. Hence, we define the RUL, R, of a system at time t as:

$$R = \inf\{r > 0 : X_{t+r} \geq L \mid X_t\},$$

where X_t is the observed degradation at time t. Assuming the parameters of the IG process are known, the cumulative distribution function (CDF) of R given the parameters can be expressed as:

$$\begin{aligned} F_{R|\delta,\lambda}(r \mid \delta,\lambda,X_t) = &\Phi\left(\frac{\sqrt{\lambda}\left(r-\delta\left(L-X_t\right)\right)}{\sqrt{L-X_t}}\right)\\ &-\exp(2r\lambda\delta)\Phi\left(-\frac{\sqrt{\lambda}\left(r+\delta\left(L-X_t\right)\right)}{\sqrt{L-X_t}}\right), \end{aligned} \tag{5}$$

where Φ is the CDF of the standard normal distribution. Then, in case of unknown degradation parameters and knowing the joint distribution of (δ, λ), the CDF of R can be obtained by:

$$F_R(r \mid X_t) = \int_0^{\infty}\int_{-\infty}^{\infty} F_{R|\delta,\lambda}(r \mid \delta,\lambda,X_t) f(\delta \mid \lambda) f(\lambda)\,\mathrm{d}\delta\,\mathrm{d}\lambda,$$

where $F_{R|\delta,\lambda}(r \mid \delta,\lambda,X(t))$ is given in (5). With some similar calculation given in Peng (2015), we have:

$$F_R(r \mid X_t) = \sqrt{\frac{\beta}{2\pi}}\frac{\Gamma\left(\alpha+1/2\right)r}{\Gamma\left(\alpha\right)}\int_{L-X_t}^{\infty} g\left(z\right)\mathrm{d}z, \tag{6}$$

where,

$$g\left(z\right) = z^{-\frac{3}{2}}(\sigma^2 z+1)^{-\frac{1}{2}}\left(1+\frac{\beta(\xi z-r)^2}{2z\left(\sigma^2 z+1\right)}\right)^{-\left(\alpha+\frac{1}{2}\right)}.$$

3 MAINTENANCE POLICY

3.1 *Maintenance policy structure*

Let $\{T_n\}_{n\in\mathbb{N}}$ be the aperiodic inspection times ($T_0 = 0$). At each inspection, one must decide about the required maintenance action. This decision is driven based on the knowledge of the system condition after the inspection. Here, we assume that the maintenance actions are performed in a negligible time and T_n^- refers to the time just before the maintenance date. To control the system failure occurrence, a preventive threshold $M < L$ is chosen. The possible scenarios which can arise are:

- If $X_{T_n^-} \geq L$, the system is failed and it is correctively replaced with the cost of C_c.
- If $M \leq X_{T_n^-} < L$, the system is not failed but it is too deteriorated and cannot function appropriately. Hence, a preventive replacement with the cost of C_p is performed.

- If $X_{T_n^-} < M$, the system is still properly functioning. Then, there is no need for replacement and the system is left as it is.

Both preventive and corrective replacements are perfect and reset the system to "as good as new" condition. Then, the system condition after the inspection X_{T_n} would be:

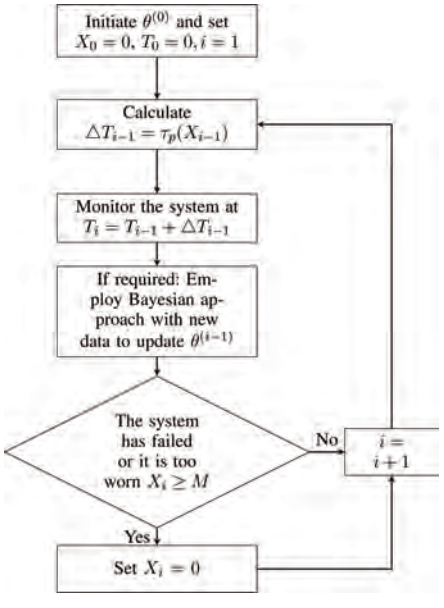
$$X_{T_n} = \begin{cases} 0 & \text{if } X_{T_n^-} \geq L \\ 0 & \text{if } M \leq X_{T_n^-} < L \\ X_{T_n^-} & \text{if } X_{T_n^-} < M \end{cases}$$

In all the above cases, the RUL based inspection is carried out. Therefore, the time for the next inspection T_{n+1} is determined from time T_n by:

$$T_{n+1} = T_n + \Delta T_n \text{ with } \Delta T_n = \tau_p\left(X_{T_n}\right),$$

where $\tau_p\left(X_{T_n}\right)$ is the p-quantile of the RUL distribution given in Equation (6).

The maintenance decision rule is dynamically updated through the Bayesian method. The time between two successive inspections is derived from the RUL assessment which depends on the stochastic degradation process hence on the hyperparameters. Different strategies can be considered for the update frequency. Typically it can be at each inspection time or at each replacement time. The general flowchart describing the evolution of the maintained system state submitted to the proposed aperiodic maintenance policy with the Bayesian update is given hereafter.



$\theta^{(i-1)}$ is the vector of hyperparameters at iteration i and $\tau_p\left(X_{i-1}\right)$ is the p-quantile of the RUL distribution obtained with $\theta^{(i-1)}$.

3.2 *Maintenance cost*

The inspections are planned discretely and each of them incurs a cost C_i. At each inspection, the decision of replacement is checked and a preventive or corrective action is performed, if needed, with costs C_p and C_c respectively. As the maintenance performed on a more deteriorated system is more complex and hence more costly, $C_c > C_p$. Moreover, since the failure can only be detected through inspections, there is a system downtime after failure and the additional cost is incurred from the failure time until the next replacement time at a cost rate C_d. The cumulative maintenance cost is:

$$C(t) = C_i N_i(t) + C_p N_p(t) + C_c N_c(t) + C_d d(t)$$

where $N_i(t)$, $N_p(t)$, and $N_c(t)$ are respectively the number of inspections, the number of preventive replacement, and the number of corrective replacement in $[0,t]$. Also, $d(t)$ is total time passed in a failed state in $[0,t]$.

Here, we use the expected cost function over a finite time horizon T_{end}, as the objective function to assess and optimize the maintenance policy.

4 SIMULATION STUDY

In order to illustrate the behavior of the aperiodic maintenance policy, let consider the case of a system which deteriorates according to an IG process with fixed parameters $\mu_{\text{real}} = 1$ and $\lambda_{\text{real}} = 1$. The prior information about the system is given by the values of hyperparameters such that $\alpha = 1.5$, $\beta = \frac{2}{3}$, $\xi = 1$ and $\sigma = \frac{1}{\sqrt{3}}$. The limit threshold is $L = 9$. The Bayesian update introduced before helps us to get better information about the model parameters (δ, λ) in comparison to the initial information which may be wrong or partly wrong in some cases. Figure 1 shows the mean and variance

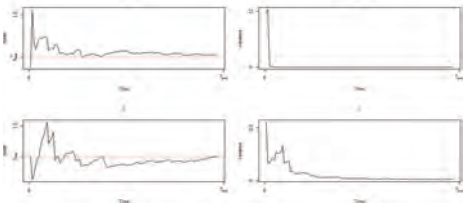


Figure 1. The mean (on the left) and the variance (on the right) of the distribution of δ (top) and λ (down) over the time.

of the evolving distribution of (δ, λ) over time with the update at each inspection. It is clear that after a while the means tend to their correct values, μ_{real}^{-1} and λ_{real}, while their variances vanish to zero.

The maintenance cost is the expected cumulative maintenance cost over a finite horizon with $T_{\text{end}} = 100$, where $C_i = 0.2$, $C_p = 4$, $C_c = 10$ and $C_d = 4$. It is estimated by a Monte Carlo simulation with 5000 samples over the considered finite horizon. Four different versions of the proposed maintenance policy are considered corresponding to four configurations of the model used for RUL prediction. The p-quantile of the RUL hence the time for the next inspection is successively obtained from Equation (6) or as the first passage time of the IG process used for degradation simulation. This last case is hereafter referred as "perfect" case because it assumes that the model used for prognosis is exactly the model used to describe the degradation of the system. Equation (6) describes the case of unknown degradation parameters. Three considered options depend on hyperparameters. In the first one, the values of hyperparameters are fixed to the ones given by experts. This case is called "no update" case. The second and third cases consider updates of α, β, ξ and σ respectively at each inspection time (case "inspection") and at each replacement time (case "cycle").

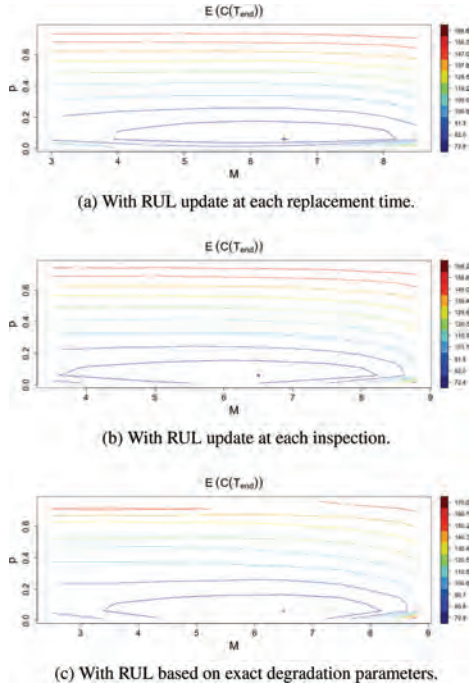
As an example, specific values of the decision parameters p and M are chosen, which are the optimal decision parameters obtained under the assumption of known degradation parameters (case "perfect"). The values obtained by Monte Carlo simulation are $p^*_{\text{known}} = 0.03$ and $M^*_{\text{known}} = 6.5$. Table 1 gives different results obtained with these values of decision parameters for the four different configurations of maintenance decision rule over the finite time horizon $[0; T_{\text{end}}]$. The given quantities are the cumulated expected maintenance cost, its variance, the mean numbers of maintenance actions (inspections, corrective and preventive replacements), the mean unavailability duration and the mean number of renewal cycles.

Table 1. Mean values for comparison of different maintenance decision rule options applied on degradation data generated from IG process with fixed parameters $\mu_{\text{real}} = 1$ and $\lambda_{\text{real}} = 1$.

Case	Perfect	Insp.	Cycle	No update
Cost mean	69.26	72.20	74.78	97.416
Cost Var.	11.41	11.86	12.25	7.62
Inspections	32.8	44.7	62.3	193.3
Prev. actions	11.5	11.4	11.7	12.8
Corr. actions	1.3	1.38	1.23	0.7
Unav. dur.	0.85	0.91	0.76	0.12
Nb of cycles	13.8	13.8	13.9	14.5

The variations of the cost for the considered policies are in line with logical thinking. The considered decision parameters correspond to the minimal cost for the highest possible level of information i.e. for known real deterioration parameters. The optimal cost is used as a reference. For the lowest level of information i.e. for RUL assessment based on initial hyperparameters values without any update, the cost increase with respect to the reference is about 40%. The introduction of the Bayesian update leads to maintenance costs which are close to the best case. The increase is close to 4.2% and 7.5% if the RUL prediction for inspections scheduling is respectively updated at each inspection time (i.e. around 44 times on $[0; T_{\text{end}}]$) or at each replacement (i.e. around 14 times). The main differences between the costs of the four considered options are due to the number of inspections. In the absence of updates, the probability law of the RUL is spread out and the p-quantile causes small inter-inspection times. With updates, the pdf becomes sharper and it leads to increasing periods of time between successive inspections.

Table 1 contains results for one specific value of $(p; M) = (p^*_{\text{known}}; M^*_{\text{known}})$. In order to illustrate the performance of the proposed maintenance policy for different settings, the expected cumulative maintenance cost as a function of the two decision parameters is depicted in Fig. 2. The two



(a) With RUL update at each replacement time.

(b) With RUL update at each inspection.

(c) With RUL based on exact degradation parameters.

Figure 2. Cost as a function of the maintenance decision parameters p and M for 3 different decision rules with $C_i = 0.2$, $C_p = 4$, $C_c = 10$ and $C_u = 4$ and $T_{\text{end}} = 100$.

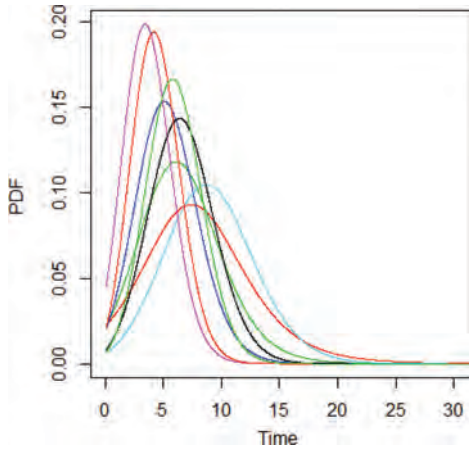


Figure 3. The pdf of the RUL updated at each inspection.

cases with the Bayesian update are investigated as well as the case with fixed and known parameters. It can be seen that the shapes of the different cost surfaces are close to each other. It means that the performance evolves in the same way when the decision parameters are modified. For example, one can see that the sensitivity to the parameter M is low and that too small values of p can produce a sudden cost increase due to the increase of the number of cycles. In order to show how the RUL pdf becomes peakier in each update, the RUL pdf are depicted with successive update at each inspection in Figure 3.

The obtained numerical results show that the efficiency of the Bayesian update procedure as a way to adapt the decision rule of the aperiodic maintenance policy. It allows implementing an auto-adaptive maintenance policy with a small number of decision parameters.

5 CONCLUSIONS

A predictive maintenance policy for deteriorating systems with unknown parameters is proposed in this paper. It includes a 2-parameter maintenance decision rule based on the Bayesian update which allows to jointly decide when to inspect the system and what to do about possible replacement at inspection time. The whole modeling framework based on IG process for degradation modeling and cumulative maintenance cost on a finite horizon is described. Some numerical results are given to illustrate the behavior of the maintenance policy. The inspection times are driven by the maintenance decision rule and used for hyperparameters updates. Two different versions are considered: with the update at each inspection time and with the update at each system replacement time. Whatever the version is, the shape of the cost function is shown to be close to the case of known deterioration parameters. It allows considering possible choices of decision parameters without unexpected behaviors on cost value. According to the Bayesian update procedure, the fast evolution of the means degradation parameters to their exact values make the proposed maintenance policy promising. The problem related to the optimal choice of the decision parameters will be considered in a future work.

REFERENCES

Alaswad, S. & Y. Xiang (2017, January). A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliability Engineering & System Safety 157*(Supplement C), 54–63.

Chen, N., Z.-S. Ye, Y. Xiang, & L. Zhang (2015, May). Condition-based maintenance using the inverse Gaussian degradation model. *European Journal of Operational Research 243*(1), 190–199.

Pan, D., J.-B. Liu, & J. Cao (2016, April). Remaining useful life estimation using an inverse Gaussian degradation model. *Neurocomputing 185*(Supplement C), 64–72.

Peng, C.-Y. (2015, January). Inverse Gaussian Processes with Random Effects and Explanatory Variables for Degradation Data. *Technometrics 57*(1), 100–111.

Peng, W., Y. Liu, Y.-F. Li, S.-P. Zhu, & H.-Z. Huang (2014, October). A Bayesian optimal design for degradation tests based on the inverse Gaussian process. *Journal of Mechanical Science and Technology 28*(10), 3937–3946.

van Noortwijk, J. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering and System Safety 94*, 2–21.

Wang, H. (2002). A Survey of Maintenance Policies of Deteriorating Systems. *European Journal of Operational Research 139*, 469–489.

Wang, X. & D. Xu (2010, May). An Inverse Gaussian Process Model for Degradation Data. *Technometrics 52* (2), 188–197.

Ye, Z.-S. & N. Chen (2014, July). The Inverse Gaussian Process as a Degradation Model. *Technometrics 56*(3), 302–311.

Ye, Z.S., L.P. Chen, L.C. Tang, & M. Xie (2014, September). Accelerated Degradation Test Planning Using the Inverse Gaussian Process. *IEEE Transactions on Reliability 63*(3), 750–763.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

An opportunistic maintenance policy for heterogeneous components

P.A. Scarf
University of Salford, Manchester, UK

C.A.V. Cavalcante & R.S. Lopes
Federal University of Pernambuco, Recife, Brazil

ABSTRACT: A hybrid maintenance policy that combines inspection in early life with opportunistic replacement in later life is developed for a one-component system with a heterogeneous lifetime. Inspections and replacements use opportunities that arise periodically, for example due to visits of a maintenance vessel to a turbine in an offshore windfarm. Components may be weak or strong, and the inspection phase of the policy operates like a burn-in period. Inspections are modelled using the delay time concept. The policy mimics reality whereby new systems are carefully maintained and older systems are replaced at events determined by operational constraints. We determine the cost-rate and reliability of the hybrid policy. Using a numerical example we show that the inspection phase offers benefits when the delay time is sufficiently large and/or early failure is a significant possibility. The policy would be relatively easy to implement in practice.

1 INTRODUCTION

Maintenance management of technical systems uses preventive and corrective replacement, inspection, repair and such like (De Almeida et al. 2015; Lee & Cha, 2016) to increase system reliability and availability and to reduce total cost of ownership of industrial assets (Berrade et al. 2013; Xia et al. 2015; Zheng, et al. 2016). For some systems, stoppages may provide opportunities for the execution of preventive maintenance with less disruption and at a lower cost than scheduled preventive maintenance. An example is the loss of the cold water supply to a soft-drinks production line caused by pump failure (Wang et al. 2000), whereby preventive maintenance of bottling and packing sub-systems may be carried out ahead of schedule. The models of opportunistic maintenance policies develop this idea in theory (e.g. Dekker & Smeitink, 1991, Zheng, 1995; Tan & Kramer, 1996; Mohamed-Salah et al. 1999; Budai et al. 2006; Laggoune et al. 2010; Xia et al, 2017b, c; Zhang & Zeng, 2017) and for practice (e.g Ding and Tian, 2011; Shafiee et al., 2015; Yildirim et al. 2017; Hu & Zhang, 2014; Nilsson et al. 2009; Cavalcante & Lopes, 2015; Xia et al. 2017a; Garambaki et al. 2016). Typically, opportunities arise from economic and structural interdependencies among components or parts (Dekker & Smeitink, 1991). Grouped maintenance policies (Wildeman et al. 1997) also exploit such dependencies to maintain groups of parts (Vu et al. 2015; Peng & Zhu, 2017), but opportunistic maintenance is different because it aims to maintain a part or parts when another part of the system causes a stop.

In this paper, we consider opportunities in different way. We suppose that the opportunities arise deterministically and periodically, as they might arise when shutdowns of a plant of which the system of interest is a part are seasonal or when maintenance resources are available only occasionally, as in the case of visits of a maintenance vessel to a turbine in an offshore windfarm.

We further suppose that the system is subject to a two-stage failure process according to the delay time model (Christer, 1999), whereby a defective but operational state precedes the failed state. Then, opportunities may be utilized for inspection, and for replacement if required. Additionally, we model component heterogeneity that may arise, for example, in the context of variable maintenance quality (Scarf & Cavalcante, 2012), supplier selection (Berrade et al. 2012), reliability (Castet and Saleh, 2010), and analysis of failure warranty data (Attardi et al., 2005; Lee et al., 2016). This heterogeneity means inspection in early life has an important role that is similar to operational "burn-in" (Zhang et al. 2014).

Few papers consider this connection between opportunistic maintenance and inspection (Wang & Christer, 2003; Berrade et al. 2017). Cavalcante et al. (2018) exploit this gap in the literature. In particular, they develop a model that generalizes hybrid inspection and replacement (Scarf et al. 2009, Scarf & Cavalcante, 2010) and opportunistic maintenance,

and in which a system is periodically inspected in the first phase of its life and replaced at opportunities during the second phase. In the policy they consider, opportunities arise at random, due to system stoppages that are caused by the failure of other structurally connected systems. In our paper here, we take a different approach and suppose that opportunities are deterministic and periodic. This makes the policy simpler to study (fewer decision variables) and easier to implement in practice (akin to block replacement rather than age replacement, Barlow & Proschan, 1966), but nonetheless applicable to maintenance of windfarms (e.g. Shafiee, 2015), transportation systems (e.g. Corman et al. 2017), and manufacturing systems (e.g. Zahedi et al. 2017).

The structure of the paper is as follows. We begin next with the description of the system and the policy. Then in section 3 we develop the cost rate. Section 4 briefly develops the system reliability. This is followed by a numerical example to illustrate the policy. We conclude with a discussion.

2 THE POLICY

2.1 *Description of the system*

We consider a single component which when in its socket performs an operational function (Ascher and Feingold, 1984). The component is in one of three states, good, defective or failed. The time in the good state, X, the time to defect arrival, has a mixture distribution $F_X(t) = pF_1(t) + (1-p)F_2(t)$, with mixing parameter p. Thus components arise from a mixed population of "weak" and "strong" sub-populations. F_1 and F_2 are in general increasing failure rate distributions, and in our particular example they are Weibull distributions with characteristic lives η_1, η_2 and shape parameters $\beta_1, \beta_2 > 1$. The corresponding density and reliability (survival) functions are denoted by f_X and $\bar{F}_X$ respectively.

The component in its socket constitutes a one-component system. This system is a non-critical system so that, on failure, the system stoppage is revealed only at a subsequent opportunity, whereat the component is replaced and the system returns to operation. Inspection determines whether the system is good or defective. When the system is defective it continues to operate albeit with a degraded function (e.g. a noisy bearing). Inspections are perfect in that an inspection reveals the true system state. At a positive inspection (system is defective), the component is replaced. Component replacement corresponds to system renewal.

Opportunities arise deterministically every S time units, as a result of, for example, seasonal shutdowns or periodic availability of maintenance resources.

The sojourn in the defective state, H, the delay time, has density $f_H(h)$, distribution function $f_H(h)$, and reliability (survival) function $\bar{F}_H(h)$. X and H and opportunities are mutually, statistically independent.

2.2 *Description of the policy*

Opportunities arise every S time units. Inspections are scheduled at each of the first K opportunities from when the system is new and replacement is scheduled at the Mth opportunity $(M \geq K)$. On failure, the system is replaced at the first opportunity that follows the failure. Maintenance interventions (inspections and replacements) cannot occur at instants other than opportunities. In this way, replacements are always synchronized with opportunities, and so renewal occurs only at an opportunity. The policy has two decision variables: K and M. The cost parameters are as follows:

- the cost of an inspection is C_I;
- the cost of a replacement of a defective component is C_O;
- the cost of a preventive replacement of a component at age MS is also C_O;
- the cost of a replacement of a failed component is C_F $(C_I < C_O < C_F)$.

The innovation of this policy is the early-life inspection phase [0, KS], whereby a component is replaced if it is found to be defective. In this way, great care is taken of the system during early life.

We might study a more general policy in which inspections are scheduled every N-th opportunity, that is, at times NS, $2NS$, …, KNS (from new). This policy may be appropriate when opportunities arise very frequently (e.g. at shutdowns or under reduced timetables of transportation systems that occur on a weekly basis). However, in this paper, we focus on the simpler two-variable policy.

We might also study a policy in which the cost of failure is not fixed, but instead related to the unavailability of the system (from the point of failure until the replacement at the subsequent opportunity). This is perhaps a more realistic policy for the case in which the unavailability cost is large relative to the cost of remedial work on a failed system. For a turbine in a windfarm, however, unavailability cost is not large (relatively), because the cost of lost power generation will be small relative to the cost of damage to the turbine as a result of failure. Furthermore, unavailability costs across a large installation are typically factored into the wind power capacity model, so that operators should focus on direct maintenance costs (Shafiee and Sorensen, 2017).

3 CALCULATION OF THE COST-RATE

The cost-rate is determined by calculating the probabilities of the three types of renewal scenario that arise: related to failure; related to preventive replacement; and related to replacement of a defective component at inspection.

A failure can arise as a result of a defect that itself arises either in early-life during the inspection phase or in later-life during the replacement phase. The costs are different in each case, so we develop the calculations separately.

In the first case, a defect arises in the interval $[(i-1)S, iS)$ and fails before time (age) iS. This occurs with probability

$$P_{1,\mathrm{F}_i} = \int_{(i-1)S}^{iS} F_H(iS-x) f_X(x)\,\mathrm{d}x,$$

for $i = 1,\ldots,K$. The cost associated with this event is $C_\mathrm{F} + (i-1)C_\mathrm{I}$ and the length of the renewal cycle is iS, noting that replacement has to wait until the next opportunity and that the cost of unavailability is ignored.

In the second case, a defect arises after KS and fails in the interval $((i-1)S, iS)$, for $i = K+1,\ldots,M$. For $i = K+2,\ldots,M\,(M > K+1)$, this occurs with probability

$$P_{2,\mathrm{F}_i} = \int_{(i-1)S}^{iS} F_H(iS-x) f_X(x)\,\mathrm{d}x + \sum_{j=K+1}^{i-1} \int_{(j-1)S}^{jS} \{F_H(iS-x) - F_H((i-1)S-x)\} f_X(x)\,\mathrm{d}x,$$

and for $i = K+1$ we have

$$P_{2,\mathrm{F}_{K+1}} = \int_{KS}^{(K+1)S} F_H((K+1)S-x) f_X(x)\,\mathrm{d}x.$$

The cost associated with this event is $C_\mathrm{F} + KC_\mathrm{I}$ and the length of the renewal cycle length is iS.

For renewal related to preventive replacement at MS, this can occur only if either no defect arises before MS or a defect arises before MS but after KS and it survives to MS. In either case the cost of K inspections and the cost of the preventive replacement are incurred. Thus the probability of preventive replacement at MS is

$$P_M = \int_{KS}^{MS} \bar{F}_H(MS-x) f_X(x)\,\mathrm{d}x + \bar{F}_X(MS),$$

and the associated cost of the renewal cycle is $C_\mathrm{O} + KC_\mathrm{I}$ and the length of the renewal cycle is MS.

The final renewal scenario is replacement of a defective component at inspection. This can only occur in the early-life phase, and the defect must arise in an interval between opportunities and survive (not fail) by the end of the interval. The probability of renewal at the i-th inspection, for $i = 1,\ldots,K$, is

$$P_{\mathrm{D}_i} = \int_{(i-1)S}^{iS} \bar{F}_H(iS-x) f_X(x)\,\mathrm{d}x,$$

and the associated cost of the renewal cycle is $C_\mathrm{O} + iC_\mathrm{I}$ and the length of the renewal cycle is iS.

Thus, associated with each type of renewal event is a cost, and the expected cost of a renewal cycle is the sum of the products of the cost of each renewal event and the respective probability of each renewal event. Therefore

$$E\{U(K,M)\} = (C_\mathrm{O} + KC_\mathrm{I})P_M + \sum_{i=1}^{K} \left\{(C_\mathrm{F} + (i-1)C_\mathrm{I})P_{1,\mathrm{F}_i} + (C_\mathrm{O} + iC_\mathrm{I})P_{\mathrm{D}_i}\right\} + (C_\mathrm{F} + KC_\mathrm{I}) \sum_{i=K+1}^{M} P_{2,\mathrm{F}_i}.$$

The expected cycle length is found similarly:

$$E\{V(K,M)\} = MSP_M + \sum_{i=1}^{K} iS(P_{1,\mathrm{F}_i} + P_{\mathrm{D}_i}) + iS \sum_{i=K+1}^{M} P_{2,\mathrm{F}_i}.$$

Then we use the renewal-reward theorem to define the cost-rate (the long run average cost per unit time) $C_\infty(K,M) = E(U)/E(V)$. This is the objective function we use to determine the optimum values of the decision variables K and M.

4 RELIABILITY

We can also quantify the reliability of the maintained system (Lewis, 1987; Scarf et al. 2005) using the "long-run time between failures":

$$\frac{\text{expected cycle length}}{\Pr(\text{renewal on failure})} = \frac{E\{V(K,M)\}}{\sum_{i=1}^{K} P_{1,\mathrm{F}_i} + \sum_{i=K+1}^{M} P_{2,\mathrm{F}_i}}.$$

We denote this by $\mu_\infty(K,M)$. We might also describe its inverse as the "failure rate", a term often used in reliability engineering but commonly misunderstood (Ascher & Feingold, 1984; Gaens, 2017). The system availability might also be calculated, but we omit this for brevity.

5 NUMERICAL EXAMPLE

We set the time unit is one year and an arbitrary cost unit, so that all costs are multiples of C_O. The values of parameters (defined in sections 2.1 and 2.2) are typical of wind turbines (e.g. Faulstich

et al. 2011). In the base case (grey fill in row 2 of Table 1), we set $\eta_1 = 2$, $\beta_1 = 2$, and $\eta_2 = 10$, $\beta_2 = 5$ so that the populations of weak and strong components are reasonably well separated. The value of the mixing parameter ($p = 0.1$) is not untypical of mechanical systems (e.g. Gales, 2015). The mean delay time and the time between opportunities are both chosen to be the same (one year), so that we might expect interesting effects to be observed. Some results are shown in Table 1 and Figure 1.

Figure 1 suggests that the cost-rate is sensitive to both M, which determines the age at replacement, and K, which determines the length of the inspection phase. We can further see that the cost-rate increases sharply as inspection reduces ($K = 2$ vs $K = 3$), so the benefit of the inspection phase is clear. This benefit is absent when there are no weak components ($p = 0$ in Table 1). Final point to make about this figure is that it is apparent that if one is unsure about how much inspection to carry out, more ($K > 3$) is better than less.

The effect of heterogeneity (through p) is large, and as p increases we see in Table 1 that the inspection phase extends. Furthermore, there is a three-fold increase in the failure rate between $p = 0$ and $p = 0.2$. Presumably the increase would be greater without inspection, and this is responsible for the high cost-rate of the pure inspection policy in row 3 of Table 2. This is confirmed by the very high failure rate for the pure inspection policy in row 3 of Table 3.

The mean delay time $1/\lambda$ also has a large influence on the policy. As the mean delay time decreases, the usual response in an inspection policy is to carry out more frequent inspection. However, in the policy here, this frequency is fixed and only the length of the inspection phase can be increased. Thus, the opportunity to prevent failures through

Table 1. Optimal policy for various η_1, β_1, p, λ, and S. Other parameters fixed at $\eta_2 = 10$, $\beta_2 = 5$, $C_I = 0.08$, $C_O = 1$, $C_F = 8$. Base case is row 2.

η_1	β_1	P	λ	S	K	M	C_∞	$1/\mu_\infty$
1	2	0.1	1	1	2	6	0.301	0.0132
2	2	0.1	1	1	3	6	0.313	0.0134
4	2	0.1	1	1	6	7	0.307	0.0123
2	1	0.1	1	1	2	6	0.313	0.0155
2	4	0.1	1	1	3	6	0.308	0.0126
2	2	0.0	1	1	0	6	0.213	0.0066
2	2	0.2	1	1	6	7	0.372	0.0192
2	2	0.1	0.5	1	3	7	0.276	0.0123
2	2	0.1	2	1	2	6	0.351	0.0208
2	2	0.1	∞	1	0	6	0.387	0.0298
2	2	0.1	1	0.5	6	12	0.339	0.0111
2	2	0.1	1	2	2	3	0.312	0.0155

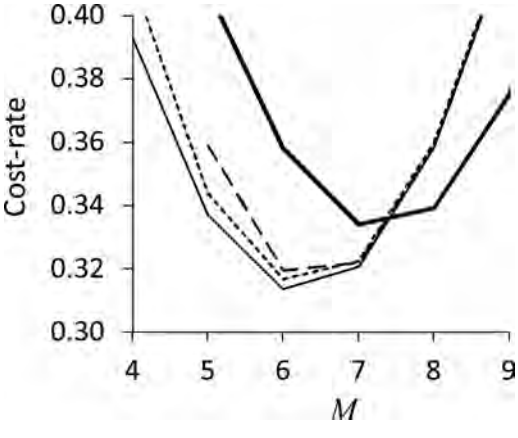


Figure 1. In the base case, cost-rate versus M for various values of K: $K = 2$ ( ), $K = 3$ (—), $K = 4$ (------); $K = 5$ (— — —).

Table 2. Policy cost-rate comparison. $C_I = 0.08$, $C_O = 1$, $C_F = 8$.

			K, M policy	Pure inspection ($M = \infty$)	Pure replacement ($K = 0$)	Whole life inspection, (K=M)
p	λ	S	Cost-rate, C_∞			
0	1	1	0.213	0.446	0.213	0.276
0.1	1	1	0.313	0.477	0.338	0.326
0.2	1	1	0.372	0.512	0.465	0.381
0.1	0.5	1	0.276	0.357	0.304	0.287
0.1	2	1	0.351	0.631	0.359	0.372
0.1	∞	1	0.387	0.965	0.388	0.466
0.1	1	0.5	0.339	0.445	0.339	0.369
0.1	1	2	0.312	0.562	0.335	0.324

Table 3. Policy failure-rate comparison. $C_I = 0.08$, $C_O = 1$, $C_F = 8$.

			K, M policy	Pure inspection ($M = \infty$)	Pure replacement ($K = 0$)	Whole life inspection, (K=M)
p	λ	S	Failure rate, $1/\mu_\infty$			
0	1	1	0.0066	0.0379	0.0066	0.0041
0.1	1	1	0.0134	0.0411	0.0233	0.0129
0.2	1	1	0.0192	0.0448	0.0434	0.0192
0.1	0.5	1	0.0123	0.0238	0.0219	0.0098
0.1	2	1	0.0208	0.0635	0.0259	0.0163
0.1	∞	1	0.0298	0.1117	0.0299	0.0299
0.1	1	0.5	0.0111	0.0245	0.0250	0.0088
0.1	1	2	0.0155	0.0601	0.0231	0.0155

inspection and consequently the replacement of defective components is limited by the frequency of opportunities ($1/S$). Nonetheless, inspection at these opportunities is still beneficial.

Although S is not a decision variable in our formulation of the model, we should study its effect. And its effect on the cost-rate and the failure rate is quite large. There is a 40% increase in the failure rate between $S = 0.5$ and $S = 2$. Notice however that the length of the inspection phase (KS) changes little and the age at replacement (MS) does not change at all. The increased cost-rate when $S = 0.5$ is explained by the increased cost of inspection. Thus, the statistical properties of the weak subpopulation determine broadly the length of the inspection phase and then the inspection regime follows from this and the frequency of opportunities. Obviously the frequency of opportunities limits how often inspections can be carried out.

The effect of η_1 is curious (when the life of the weak components is shortest, the cost rate is not the largest) and we have no explanation for this.

Focusing more broadly on Tables 2 and 3 and the policy comparisons (noting that each policy reported is the respective optimum policy in its class), we can see the use of an inappropriate policy has a more profound effect on cost and reliability than do the failure characteristics of the system. Pure replacement is always a high-cost and low-reliability. Whole life inspection gives the highest reliability and this is to be expected; in the circumstances, additional inspections only have a cost-disadvantage. Pure replacement and whole life inspection are the extremes of the K, M policy, and again Tables 2 and 3 suggest that if a maintainer is unsure about the appropriate frequency of inspection then more is better than less.

6 DISCUSSION

We study a hybrid inspection and replacement policy in which the opportunities to carry out inspections and replacement are periodic but limited and somewhat out of the control of the maintenance decision-maker. We consider in particular the effect of heterogeneity of component quality, so that early-life defects are possible. The model is related to the hybrid inspection and replacement policy proposed by Scarf et al. (2009) and to the general hybrid opportunistic policy of Cavalcante et al. (2017) in which opportunities are also limited but occurring at random. The policy proposed in the paper here is natural in the context of the maintenance of an offshore wind-farm.

The calculation of the long-run average cost per unit time (cost-rate) for the policy is presented. We also calculate a measure of the operational reliability of the system. We illustrate the policies with a numerical example. We compare, in terms of cost-rate, the proposed policy with a number of policies that are special cases including pure inspection and pure replacement.

Our results indicate a number of points. Component heterogeneity drives the length of the inspection phase, and then the frequency of inspect-ion during this phase is determined not by the decision-maker but by the frequency of opportunities. Similarly the age limit for replacement is driven by the system failure characteristics and the costs rather than the frequency of opportunities. When the quality of replacements is most poor then the inspection phase of the hybrid policy is most beneficial. Finally, if the maintainer is in doubt about the length of the inspection phase, then more inspection is better than less, provided that inspections do not induce defects (Scarf and Cavalcante, 2012).

The fixed frequency of opportunities for maintenance can simplify maintenance planning, and the policy we propose provides a simple framework in which to mitigate for the effects of variable quality in the execution of maintenance interventions. An interesting question relates to the effectiveness of continuous condition monitoring in the circumstances when the maintenance frequency is fixed. In a simple, single component system, alarms are not useful if the system cannot be accessed. Nonetheless, if failure can be prevented through remote shutdown, then we would expect continuous monitoring to be beneficial.

REFERENCES

Ascher H. & Feingold, F. 1984. *Repairable Systems Reliability.* Marcel Dekker.

Attardi, L., Guida, M. & Pulcini, G. 2005. A mixed-Weibull regression model for the analysis of automotive warranty data. *Reliability Engineering & System Safety* 87(2): 265–273.

Barlow, R.E. & Proschan, F. 1996. *Mathematical Theory of Reliability*. Society for Industrial and Applied Mathematics.

Berrade, M,D,. Scarf, P.A., Cavalcante, C.A.V. & Dwight, R.A. 2013. Imperfect inspection and replacement of a system with a defective state: a cost and reliability analysis. *Reliability Engineering and System Safety* 120(1): 80–87.

Berrade, M.D., Cavalcante, C.A.V. & Scarf, P.A. 2012. Maintenance scheduling of a protection system subject to imperfect inspection and replacement. *European Journal of Operational Research* 218(3): 716–725.

Berrade, M.D., Scarf, P.A. & Cavalcante, C.A.V. 2017. A study of postponed replacement in a delay time model. *Reliability Engineering & System Safety* 168: 70–79.

Budai, G., Dekker, R. & Nicolai, R.P. 2008. Maintenance and production: A review of planning models.

In Kobbacy K., Murthy D.N.P. (eds.) *Complex System Maintenance Handbook,* Springer, pp. 321–344.

Castet, J-F. & Saleh, J.H. 2010. Single versus mixture Weibull distributions for nonparametric satellite reliability. *Reliability Engineering & System Safety* 95(3): 295–300.

Cavalcante, C.A.V. & Lopes, R.S. (2015). Multi-criteria model to support the definition of opportunistic maintenance policy: A study in a cogeneration system. *Energy* 80(1): 32–40.

Cavalcante, C.A.V., Lopes, R.S. & Scarf, P.A. (2018). A general inspection and opportunistic replacement policy for one-component systems of variable quality. *European Journal of Operational Research* 266(3): 911–919.

Christer, A.H. 1999. Developments in delay time analysis for modeling plant maintenance. *Journal of the Operational Research Society* 50(11): 1120–1137.

Corman, F., Kraijema, S., Godjevac, M. & Lodewijks, G. 2017. Optimizing preventive maintenance policy: A data-driven application for a light rail braking system. *Journal of Risk and Reliability* 231(5), 534–545.

De Almeida, A.T., Cavalcante, C.A.V., Alencar, M.H., Ferreira, R.J.P., de Almeida-Filho, A.T. & Garcez, T.V. 2015. *Multicriteria and multiobjective models for risk, reliability and maintenance decision analysis*. Springer.

Dekker, R. & Smeitink, E. 1991 Opportunity-based block replacement. *European Journal of Operational Research*, 53(1), 46–63.

Ding, F. & Tian, Z. 2011. Opportunistic maintenance optimization for wind turbine systems considering imperfect maintenance actions. *International Journal of Reliability, Quality and Safety Engineering* 18(5): 463–481.

Faulstich, S., Hahn, B. & Tavner, P.J. 2011. Wind turbine downtime and its importance for offshore deployment. Wind Energy 14(3), 327–337.

Gaens, T. 2017. Tribute to Harry Ascher (1935–2014). https://www.asqrd.org/tribute-to-harry-ascher-1935-2014/ (accessed 25/10/2017).

Gales, T. 2015. Pump reliability in the food and beverage sectors. *Maintenance and Engineering,* September 2015, pp.6–10.

Garambaki, A.H.S., Thaduri, A., Seneviratne, A.M.N.D.B. & Kumar, U. 2016. Opportunistic inspection planning for railway emaintenance. *IFAC-Papers OnLine* 49(28): 197–202.

Hu, J. & Zhang, L. 2014. Risk based opportunistic maintenance model for complex mechanical systems. *Expert Systems with Applications* 41(6): 3105–3115.

Laggoune, R., Chateauneuf, A. & Aissani, D. 2010. Impact of few failure data on the opportunistic replacement policy for multi-component systems. *Reliability Engineering & System Safety* 95(2): 108–119.

Lee, H. & Cha, J.H. 2016. New stochastic models for preventive maintenance and maintenance optimization. *European Journal of Operational Research* 255(1): 80–90.

Lee, H., Cha, J.H., & Finkelstein, M. 2016. On information-based warranty policy for repairable products from heterogeneous population. *European Journal of Operational Research* 253(1): 204–215.

Lewis, E.E. 1987. *Introduction to Reliability Engineering.* Wiley.

Mohamed-Salah, A-K.D., & Ali, G. 1999. A simulation model for opportunistic maintenance strategies. In *Proceedings of International Congress of Energy Technologies and Factory Automation*, Vol.1, Barcelona, 703–708.

Nilsson, J., Wojciechowski, A., Strömberg, A-B., Patriksson, M. & Bertling, L. 2009. An opportunistic maintenance optimization model for shaft seals in feed-water pump systems in nuclear power plants. In *2009 IEEE Bucharest Power Tech.*

Peng, H. & Zhu, Q. 2017. Approximate evaluation of average downtime under an integrated approach of opportunistic maintenance for multi-component systems. *Computers & Industrial Engineering* 109: 335–346.

Scarf, P.A. & Cavalcante, C.A.V. 2010. Hybrid block replacement and inspection policies for a multi-component system with heterogeneous component lives. *European Journal of Operational Research* 206(2): 384–394.

Scarf, P.A. & Cavalcante, C.A.V. 2012. Modelling quality in replacement and inspection maintenance. *International Journal of Production Economics* 135(1): 372–381.

Scarf, P.A., Cavalcante, C.A.V., Dwight, R. & Gordon, P. 2009. An age based inspection and replacement policy for heterogeneous components. *IEEE Transactions on Reliability* 58(4): 641–648.

Scarf, P.A., Dwight, R. & Al-Musrati, A. 2005. On reliability criteria and the implied cost of failure for a maintained component. *Reliability Engineering and System Safety* 89(2), 199–207.

Shafiee, M. & Sorensen, J.D. 2017. Maintenance optimization and inspection planning of wind energy assets: Models, methods and strategies, *Reliability Engineering & System Safety* (in press), https://doi.org/10.1016/j.ress.2017.10.025.

Shafiee, M. 2015. Maintenance logistics organization for offshore wind energy: current progress and future perspectives, *Renewable Energy* 77, 182–193.

Shafiee, M., Finkelstein, M. & Bérenguer, C. 2015. An opportunistic condition-based maintenance policy for offshore wind turbine blades subjected to degradation and environmental shocks. *Reliability Engineering & System Safety* 142: 463–471.

Tan, J.S. & Kramer, M.A. 1997. A general framework for preventive maintenance optimization in chemical process operations. *Computers & Chemical Engineering* 21(12): 1451–1469.

Vu, H.C., Do, P., Barros, A. & Bérenguer, C. 2015. Maintenance planning and dynamic grouping for multi-component systems with positive and negative economic dependencies. *IMA Journal of Management Mathematics* 26(2): 145–170.

Wang, W. & Christer, A.H. 2003. Solution algorithms for a nonhomogeneous multi-component inspection model. *Computers & Operations Research* 30(1): 19–34.

Wang, W., Scarf, P.A. & Smith, M. (2000). On the application of a model of condition based maintenance. *Journal of the Operational Research Society* 51(11): 1218–1227.

Wildeman, R.E., Dekker, R. & Smit, A.C.J.M. 1997. A dynamic policy for grouping maintenance activities. *European Journal of Operational Research* 99(3): 530–551.

Xia, T. B., Tao, X. Y. & Xi, L. F. 2017a. Operation process rebuilding (OPR)-oriented maintenance policy for changeable system structures. *IEEE Transactions on Automation Science and Engineering* 14(1): 139–148.

Xia, T., Jin, X., Xi, L. & Ni, J. 2015. Production-driven opportunistic maintenance for batch production based on MAM–APB scheduling. *European Journal of Operational Research* 240(3): 781–790.

Xia, T., Xi, L., Pan, E. & Ni, J. 2017c. Reconfiguration-oriented opportunistic maintenance policy for re-configurable manufacturing systems. *Reliability Engineering & System Safety* 166: 87–98.

Xia, T., Xi, L., Pan, E., Fang, X. & Gebraeel, N. 2017b. Lease-oriented opportunistic maintenance for multi-unit leased systems under product-service paradigm. *Journal of Manufacturing Science and Engineering* 139(7): 071005.

Yildirim, M., Gebraeel, N. & Sun, X. 2017. Integrated predictive analytics & optimization for opportunistic maintenance and operations in wind farms. *IEEE Transactions on Power Systems* (in press).

Zahedi, F., Scarf, P.A. & Syntetos, A. 2017. Joint optimisation of inspection maintenance and spare parts provisioning: a comparative study of inventory policies using simulation and survey data. *Reliability Engineering and System Safety* 168: 306–316.

Zhang, M., Ye, Z. & Xie, M. 2014. A condition-based maintenance strategy for heterogeneous populations. *Computers & Industrial Engineering* 77: 103–114.

Zhang, X. & Zeng, J. 2017. Joint optimization of condition-based opportunistic maintenance and spare parts provisioning policy in multiunit systems. *European Journal of Operational Research* 262(2): 479–498.

Zheng, X. 1995. All opportunity-triggered replacement policy for multiple-unit systems. *IEEE Transactions on Reliability* 44(4): 648–652.

Zheng, Z., Zhou, W., Zheng, Y. & Wu, Y. 2016. Optimal maintenance policy for a system with preventive repair and two types of failures. *Computers & Industrial Engineering* 98: 102–112.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Influence of selected external factors on satellite navigation signal quality

K. Krzykowska, M. Siergiejczyk & A. Rosiński
Warsaw University of Technology, Warsaw, Poland

ABSTRACT: Signal monitoring is one of the basic tasks, which are included in the satellite system maintenance. Currently, the civil aviation, in terms of navigation, above all, develops solutions based on satellites, indicating them as future-orientated. This activity is coordinated by the ICAO (International Civil Aviation Organization), which oversees the operations of the Global Navigation Satellite System (GNSS). The analysis of satellite system errors is a major aspect limiting the operational functioning of such systems in air transport. From the point of view of this study, the tropospheric and ionospheric errors deserve special attention. It turns out that the time of year and even time of day can have a significant impact on the quality of the satellite signal and, therefore, on the operational safety of aircraft. Relationships occurring between selected external factors (temperature, pressure, cloudiness, precipitation, air humidity) and their very effect on the signal interferences—will be tested using fuzzy reasoning.

1 INTRODUCTION

The high safety level in aviation is placed on top of the pyramid of industrial challenges for modern operators and service providers'. The rationale for the selection of the research problem is the fact that the satellite systems are considered to be the future of navigation and surveillance in aviation. Failure to meet the requirements set out for satellite signals prevents their operational use (Siergiejczyk & Krzykowska 2014). The satellite systems play a significant role in programmes relating to the development of the aviation technology, including the SESAR programme (Single European Sky ATM Research), which is a technological component of the SES (Single European Sky) project implemented in the EU (Kierzkowski & Kisiel 2016). The conditions set out for the use of satellite systems in, for example, air traffic operations are, therefore, associated with four defined, main signal parameters: accuracy, availability, continuity and integrity.

2 SATELLITE SIGNAL PARAMETERS

The satellite signal used in aviation is subject to particularly stringent functional requirements (Siergiejczyk et al. 2015). Therefore, it is crucial to satisfy them. The aforementioned requirements are determined with navigational parameters (International Civil Aviation Organization 2006). They include:

- **accuracy** – defined by an error in determined position; in GNSS it is the difference between the determined and actual position; the probability for a determined position should be at least 95% – the measurement error is then within the specified accuracy;
- **integrity** – is characterized as a measure of confidence in the validity of information provided by a system; it covers the capability of a system to deliver appropriate warnings (alarms) to a user within a predetermined time, which include information on when not to use the system;
- **continuity** – is the ability of a system to utilize the assumed function without unplanned interruptions during an executed flight operation;
- **availability** – can be defined as a percentage of time, during which a satellite system can be used for navigation, and during which reliable information is passed on to the crew, a control system or other aircraft flight management systems.

The requirements in relation to accuracy indicate that in a large set of independent samples, at least 95% should meet specified conditions (stated in metres, per each satellite system type). Such accuracy must be satisfied in relation to the worst geometry of a satellite constellation, for which the system is to be available. It should be noted that position errors, in the case of, e.g. a GPS system, consist of satellite clock and ephemeris errors. They do not include ionospheric and tropospheric delays, multipath errors or self-receiver noise. The latter are in each case included in the standards regarding receivers (International Civil Aviation Organization 2006).

In the context of integrity, in order to determine whether a location error is acceptable—an alarm limit is specified, which allows to reflect the maximum, permissible position error that will not

undermined the executed flight operation. It should be noted that satellite system navigation, thus, a satellite signal, is simultaneously transmitted to many objects (aircraft) over a large area—often one or more continents. Therefore, the impact of losing integrity of a satellite system on an air traffic management system will be much more significant than in the case of conventional navigation methods. Hence, the stringent requirements regarding the parameters. An information about the loss of signal integrity (or exceeding the permissible values of other parameters) delivered sufficiently early, should result in an abandonment of using satellite navigation or discontinuation of the operation (in case of a take-off or landing). Furthermore, an individual, as well as a unique GNSS navigation feature is adapting the navigation capabilities over time, depending on the changing satellite constellation. The impact of changes in the space segment may be increased with an additional fault in the ground segment, e.g., damage to one of the components (International Civil Aviation Organization 2006).

In the case of en-route, approach and landing operations—the continuity of the service is associated with the capability of a navigation system to deliver output data with a specified integrity and accuracy over the course of the operation, assuming that the data were available at the beginning of the operation. Due to the fact that the length of individual operations is variable, the requirement regarding the continuity is defined as a range of signal discontinuity probability values per hour. The bottom range value is the minimum continuity value, at which a system may be used in areas with low traffic and a complex airspace structure (these are areas with a low number of navigation system failure per the number of aircraft). The top value enables the application in area with heavy traffic and a complex airspace structure (these are areas with a higher number of navigation system failures per the number of aircraft). It is worth noting that flight planning may not be approved if it is based solely on GNSS navigation, in which a signal is burdened with a high risk of a continuity loss at the time of planning the executed operation (International Civil Aviation Organization 2006).

Defining the requirements concerning GNSS availability should be considered in terms of the expected level of the provided service. Certain requirements will be set out for a system, which is to replace the existing navigation infrastructure, and different ones for a system supporting the current infrastructure (International Civil Aviation Organization 2006). Basically, the determination of a GNSS signal availability criterion for given operations or areas, should be based on:

a. traffic intensity and complexity;
b. the presence of back-up navigation aids;
c. covering an area with primary and secondary surveillance;
d. guidance procedures to another airport;
e. navigation system used at a back-up airport;
f. duration of interruptions in signal availability;
g. geographical range of interruptions.

In addition, according to the International Civil Aviation Organization, GNSS availability should be determined through engineering, analysing and modelling processes, and not only by measuring them. The signal availability model should take into account, among others, ionospheric and tropospheric errors, as well as receiver faults, which it utilizes to determine integrity via calculated HPL (Horizontal Protection Level) and VPL (Vertical Protection Level) indicators (Januszewski 2012) (Januszewski 2013).

3 LITERATURE STATE OF THE ART

The research problem in the presented paper is not only the analysis of satellite signal parameters, but also a search for external factors causing interference of that signal. Weather conditions will be certainly among those observed. The matter and its essence are already known in the domestic subject literature. R. Zieliński sets forth the issue of thermal noise and their presences in Earth—satellite and satellite—Earth links (Zieliński 2009). Ground receiving antenna receives noises through sky luminance temperature (sky radiation), whereas for a satellite antenna—the noises are the Earth's surface with a defined thermodynamic temperature. Attention was also paid to additional losses arising as a result of precipitation. Signal attenuation they cause, depends on the extent of the precipitation itself, most often expressed in mm/h. The European Broadcasting Union elaborated on measurement results, which present the phenomenon of signal attenuation, depending on the magnitude of precipitation, for the frequency of 11.5 GHz. This statistic provides attenuation distribution function values expressed in dB for 99% and 99.9% of the time for the worst month in Europe. The dependency of meteorological conditions and GNSS in the context of the position determination accuracy can be also found in elaboration of renowned scientific journals (Wilgan et al. 2015). However, in most cases, the impact of weather on satellite signal propagation has been described in the literature in light of tropospheric errors. The following factors are determined: dry-air density, pressure and temperature, and humid-air humidity caused by clouds, rain, fog.

It should be noted that the analysis of external factors, including meteorological, is of large significance also in the case of other fields of air traffic. The elaborations include studies on, inter alia, the impact of meteorological conditions on the

execution of aircraft landing operations or modelling external factors in the context of air traffic. These publications evidence the great importance of different factors present in air traffic. The fact that meteorological factors should not be underestimated in any of the air traffic fields seems crucial (Rychlicki & Miszkiewicz 2013).

In light of the received satellite signal, one should adopt a relevant correction, taking into account its passage through different layers of the atmosphere. For example, in order to determine a correction taking into account the passage of the signal through the ionosphere, a vertical component of electron density TEC [el/m^2] is adopted. TEC (Total Electron Content) is the total number of electrons concentrated between two points, along a column with a cross-section of 1 m^2. The studying of the TEC variable, its modelling and forecasting became a popular issue for contemporary science in the context of the satellite technology development (Rius et al 1994). Due to the demonstrated significant impact of the TEC value on the quality of the received satellite signal, a lot of attention was focused on TEC predictions, also via artificial neural networks (Paul & Sur 2013).

4 RESEARCH METHODOLOGY

The use of fuzzy sets for controlling and modelling processes has a long history. Originally, the creators of fuzzy logic saw their theory in fields such as economics or psychology, where human perception plays a crucial role, and the phenomena can be described in an unclear manner, thus, a fuzzy one. However, already in the 1970s, the possibility to control processes through this tool was observed. Nowadays, the greater part of its application concerns controlling, quite often, technical systems (Żurek & Grzesik 2015) (Robinson et al. 2005) (Skorupski & Uchroński 2016) (Stańczyk & Stelmach 2014) (Losurdo et al. 2017).

In principle, the literature sources define two approaches towards fuzzy control-descriptive and prescriptive (Zadeh 2008). The first one is based on the expertise of an operator who, based on his experience, knows how to control a process. It is a traditional approach, not based on a model. The second approach assumes the existence of a stochastic or deterministic model, and defines how, in an optimal manner, to control it. The traditional approach is closer to the studies in this paper. This is due to the fact that a model of a process defining the output (satellite signal interference) as an input function (weather condition) is not known. This means that the process is a, so-called, black box. However, knowledge on how to correctly control a process, i.e., which control to choose for a current output, is available (Kacprzyk 2001).

The construction of a model and conducting the aforementioned tests is subject to having a sufficient number of data (Siergiejczyk & Krzykowska 2017). The satellite signal analysis was based on EGNOS system data. A total of 181 measurements for a PRN120 satellite, conducted in the 1st half of 2014 at a station in Warsaw for an APV-1 vertical guidance approach operation shall be used for constructing the model. The selection of this period, area and satellite is justified by the completeness of data, which is compulsory in such studies—conditioning high reliability of the results. By agreement with the Institute of Meteorology and Water Management in Warsaw, a list of weather condition measurements (cloud cover, humidity, precipitation, pressure and temperature) was received, for the same period of time and area. The Space Research Centre of the Polish Academy of Sciences publicizes the data on solar activity, including the number of sunspots, number of observations, standard deviations. That data was also used in the research.

The article presents a model regarding the satellite signal accuracy. Therefore, input data was identified. It included:

- cloud cover;
- air humidity;
- precipitation;
- average air temperature;
- atmospheric pressure;
- solar activity expressed as the number of sunspots (daily sunspot number).

At the same time, it can be concluded that cloud cover, humidity and precipitation form a first group of input data, which determines the humid tropospheric part, associated with water vapour content. Two other factors-temperature and pressure constitute a fragment of the dry part of the troposphere and form a second data group. According to the literature (Wołoszyn 2009), the dry factor causes almost 90% of the tropospheric error (delay). The last, but not least, factor-solar activity, forms a third group of input data associated with the ionosphere and ionospheric delay (Siergiejczyk & Krzykowska 2017).

Nonetheless, the determination of an output variable remains important. It was an accuracy error, determined by the average value of HNSE and VNSE.

The next table shows a summary of linguistic values for all variables and their assigned numerical values, representing the total membership to a fuzzy set. This classification, for input variables, was based on known literature of the subject (Wołoszyn 2009) (Moszkowicz & Tuszyńska 2006). The definition of output variable ranges

Table 1. Linguistic values of input and output variables.

I/O data	Linguistic variable	Linguistic value				
I_1	Cloud cover	–	small	average	big	–
I_2	Humidity	v. low	low	average	high	v. high
I_3	Precipitation	–	small	average	big	–
I_4	Temperature	–	low	average	high	–
I_5	Pressure	–	low	average	high	–
I_6	Solar activity	v. low	low	moderate	high	v. high
O	Accuracy error	–	small	average	big	–

Table 2. The function of linguistic variable membership.

I/O data	Measure	Measurement value range	Numerical value
I_1	cloud cover (in octane scale)	0–8	0 4 8
I_2	%	42–100	42 57 72 86 100
I_3	mm water head	0–25	0 12.5 25
I_4	average air temperature in °C	(-15) – 24	-15 4.5 24
I_5	hPa	985–1024	985 1005 1024
I_6	sunspot number	40–222	40 87 110 150 222
O	average HNSE and VNSE value in metres	5–8	5 6.5 8 9.5

was solved otherwise. It applied the expertise and applicable requirements regarding the use of a satellite signal in civil aviation imposed by ICAO (International Civil Aviation Organization 2006).

The input and output variables were assigned the Gaussian membership function according to the ranges given in Table 2. 126 rules were set out.

5 RESEARCH RESULTS

The following visualizations present selected configurations of atmospheric factors with the most significant impact on the accuracy of a satellite signal.

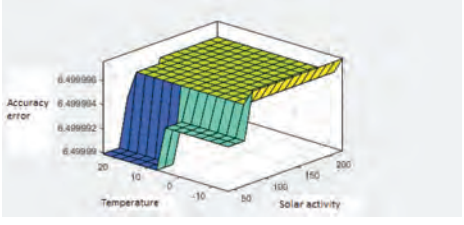


Figure 1. Graphic representation of the impact of temperature and solar activity on the accuracy error.

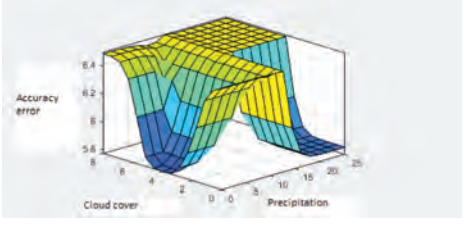


Figure 2. Graphic representation of the impact of cloud cover and precipitation on the accuracy error.

In the presented model, solar activity (I_6) was more important than temperature (I_4), nonetheless, the impact of low temperature on the increase of the accuracy error is noticeable. The configuration of precipitation (I_3) with cloud cover (I_1) and their impact on the satellite signal is interesting. Figure 2 shows an increase of the accuracy error in conditions of big and small cloud cover and rather low precipitation. This leads to a very important conclusion—I are able to indicate the day, during which it was not the weather factor, which impacted the increased signal accuracy error, because pressure changes (I_5) remained insignificant, whereas humidity (I_2) contributed to the error to a small extent. Such was the situation, for example, on 14 March 2014, when the applied principle (no. 78) had the following form:

If the cloud cover is average *and* humidity is very low *and* precipitation is low *and* temperature is average *and* pressure is average *and* solar activity is moderate *then* the accuracy error is high.

In most other cases, when the accuracy error reached a linguistic value of "high", solar activity would assume a linguistic value of "high" or "very high". The presented example, in some ways, confirms the effectiveness of the model and the tool.

6 SUMMARY

The described cases are undoubtedly a big advantage of the model. In many cases, the knowledge that the error source should not be sought among weather conditions is more important. Especially, if the factors are not accompanied by explicitness regarding the impact on signal interference. However, it can be concluded that the tool and the model implemented for it, constitute a clear base for the evaluation of the signal quality, depending on the tropospheric and ionospheric factors.

REFERENCES

International Civil Aviation Organization, Annex 10 to the Convention on international civil aviation. Air communication. Vol. I Radio-navigation Aids, ICAO, 2006.

Januszewski, J. 2012. How the ionosphere affects positioning solution using terrestrial and satellite navigation systems, w *Communication in Computer and Information Science. Telematics in the Transport Environments*, Katowice—Ustroń.

Januszewski, J. 2013. How the troposphere affects positioning solution using satellite navigation systems w *Communication in Computer and Information Science. Telematics in the Transport Environment*s, Activities of Transport Telematics.

Kacprzyk, J. 2001. *Wieloetapowe sterowanie rozmyte [Multistage fuzzy control]*, Warszawa: Wydawnictwa Naukowo—Techniczne.

Kierzkowski, A. & Kisiel, T. 2016. Simulation model of security control system functioning: A case study of the Wroclaw Airport terminal, Journal of Air Transport Management, http://dx.doi.org/10.1016/j.jairtraman.2016.09.008.

Losurdo, F. & Dileo, I. & Siergiejczyk, M. & Krzykowska, K. & Krzykowski, M. 2017. Innovation in the ICT infrastructure as a key factor in enhancing road safety: a multi-sectoral approach, Proceedings 25th International Conference on Systems Engineering ICSEng 2017, Las Vegas.

Moszkowicz, S. & Tuszyńska, I. 2006.*Meteorologia radarowa. Podręcznik użytkownika informacji radarowej [Radar meteorology. Radar information user manual]* IMGW, Warszawa: Instytut Meteorologii i Gospodarki Wodnej.

Paul, A. & Sur D. 2013. Comparison of standard TEC models with a Neural Network based TEC model using multistation GPS TEC around the northern crest of Equatorial Ionization Anomaly in the Indian longitude sector during the low and moderate solar activity levels of the 24th(...), *Advances in Space Research,* tomvol. 52, no. 5.

Rius, A. & Zarraoa, N. &Sardón, E.1994. Estimation of the transmitter and receiver differential biases and the ionospheric total electron content from Global Positioning System observations. *Radio Science*, vol. 29, no. 3.

Robinson, V.B., Cobb, M.A. & Petry, I.E. 2005. Fuzzy Modeling with Spacial Information for Geographic Problems. Springer.

Rychlicki, M. & Miszkiewicz, A. 2013. Ocena dokładności danych lokalizacyjnych odbiorników GPS. *[The assessment of GPS receiver localization data accuracy]* Prace Naukowe Politechniki Warszawskiej. Transport, no. 92.

Siergiejczyk, M., Krzykowska, K & Rosiński, A. Evaluation of the influence of atmospheric conditions on the quality of satellite signal, Proceedings of the Conference on Marine Nagivation and Safety of Sea Transportation (TransNav 2017) Gdynia, Poland, 21–23 June 2017, Editor: Adam Weintrit, CRC Press Taylor&Francis Group, London, UK 2017. ISBN: 978-1-138-29762-3.

Siergiejczyk, M. & Krzykowska, K. 2014. Some issues of data quality analysis of automatic surveillance at the airport, *Diagnostyka. Applied Structural Health, Usage and Condition Monitoring,* vol. 15, no. 1.

Siergiejczyk, M. & Krzykowska, K. 2017. Selected aspects of implementation of weather conditions monitoring service in context of the national traffic management system, (w:) D. Pyza, Prace Naukowe Politechniki Warszawskiej. Transport. Z. 118, Wydawca: Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.

Siergiejczyk, M., Rosiński, A. & Krzykowska, K. 2015. Parameters analysis of satellite support system in air navigation, w *23rd International Conference on Systems Engineering ICSEng 2014. Progress in Systems Engineering*, Las Vegas.

Skorupski, J. & Uchroński, P. 2016. A fuzzy system to support the configuration of baggage screening devices at an airport. *Expert Systems With Applications*, nr 44.

Stańczyk, P. & Stelmach, A. 2014. Modelowanie ruchu samolotu podczas operacji startu i lądowania z wykorzystaniem sztucznych sieci neuronowych. *[Aircraft movement modelling during take-off and landing operations with the use of artificial neural networks]* in *Współczesne problemy inżynierii ruchu lotniczego. Modele i metody.* Warszawa, Oficyna Wydawnicza Politechniki Warszawskiej.

Wilgan, K., Rohm, W. & Bosy, J. 2015. Multi-observation meteorological and GNSS data comparison with Numerical Weather Prediction model. *Atmospheric Research*, vol. 156, no. 1.

Wołoszyn, E. 2009. *Meteorologia i klimatologia w zarysie [Outline of meteorology and climatology].* Gdańsk: Wydawnictwo Politechniki Gdańskiej.

Zadeh, L.A. 2008. Is there a need for fuzzy logic? *Information Sciences*, nr 178 (2008).

Zieliński, R.J. 2009. *Satelitarne sieci teleinformatyczne [Satellite ICT networks]*, Warszawa: Wydawnictwa Naukowo—Techniczne Sp. z o.o.

Żurek, J. & Grzesik, N. 2015. Selected methods of aviation safety estimation, including use of fuzzy logic inference systems. International Journal Of Computer And Information Technology, vol. 4, no. 2.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Bayesian approaches to lifetime prediction

F. Marsili & J. Bödefeld
Federal Waterways Engineering and Research Institute, Karlsruhe, Germany

P. Croce & F. Landi
Department of Civil and Industrial Engineering, University of Pisa, Italy

ABSTRACT: In this paper a framework to predict the remaining lifetime for existing waterways infrastructures based on stochastic modeling of deterioration processes and Bayesian analysis is presented. The application of the Bayes' theorem is motivated by the availability of expert knowledge as well as the collection of both qualitative and quantitative data from the structure. An original method is proposed to derive the prior statistical parameters of the gamma distribution describing the stochastic deterioration process, based on the assumption that the lifetime distribution can be approximated by the Birnbaum-Saunders statistical model. An appropriate Bayesian Network is finally implemented to improve the classification of the structure with respect to its proneness to damage. The outcome of the research work is to assist the owners of large infrastructural network in planning and prioritizing maintenance interventions.

1 INTRODUCTION

In the last decade the management of assets of infrastructures is becoming an increasingly important engineering task. As it is also stated in ISO 55000 (2014), accurate prediction of lifetime distributions are required to develop strategic asset management plan, and the use of a life-cycle management approach is fundamental to realize value from the asset. The German Federal Waterways Engineering and Research Institute (BAW) has developed a management system based on fixed inspection intervals, allowing, in principle, to optimize the repair or strengthening interventions, which are planned and carried out according to the outcomes of the inspections themselves. Although plenty of data regarding structures' condition are available in BAW, further steps should be undertaken in order to extract information regarding asset's lifetime (Haider 2012). The transformation of data into useful information is not always straightforward, and several approaches could be found in literature: for example, Trappey (2012) evaluates the asset lifetime using logistic regression; Lim & Mba (2013) estimate the remaining useful life implementing Kalman filter, while Tse & Shen (2013) pursue the same scope using support vector machines.

Another way to reach this goal is through a more accurate lifetime prediction, which could be obtained by modelling degradation phenomena through stochastic processes (Riascos-Ochoa 2016); the parameters of the process can be also updated applying the Bayes' Theorem, given any new information about the state of damage (Ang & Tang 2007). However this approach, which has already been adopted by Bousquet (2014) and Haowei (2015), has the disadvantage that only quantitative data can be considered for the updating, and only the uncertainty directly affecting model parameters can be reduced.

This paper thus proposes a new procedure, which allows updating prior knowledge considering both quantitative data and qualitative information through the definition and the implementation of a Bayesian Network. Besides, using the Bayesian Network allows also identifying the stochastic process better describing the degradation phenomena affecting the structures; the Bayes' Theorem is thus applied in order to reduce the uncertainty affecting the parameters of the identified stochastic process.

The paper is subdivided in the following way: in Chapter 2 the current approach to the asset management implemented by the BAW is presented, focusing the attention on its advantages and drawbacks; in Chapter 3 the available information for lifetime prediction is examined; in Chapter 4 the new approach to lifetime evaluation is presented, also briefly reviewing the theoretical background to gamma processes, Bayesian Analysis and Bayesian Network; in Chapter 5 the new approach is applied to a real case study, paying particular attention to the elicitation of the prior distribution and the Bayesian Network; in Chapter 6 conclusion and outlook are drawn.

2 CURRENT APPROACH TO ASSET MANAGEMENT

The BAW has developed tools for the management of a huge number of waterways infrastructures. The entire portfolio comprises several types of construction such as locks, weirs, culverts, canal bridges and lighthouses, also having different ages. A huge part of the asset is represented by massive structures older than 100 years and designed according to empirical methods with limited and simplified static calculations. These structures are characterized by sections with a large thickness, in which a multi-axial stress state takes place. Therefore they exhibit significant "reliability reserves" and relevant plastic resources, as demonstrated by the low number of recorded collapses which were mainly characterized by ductile failure modes. Despite their satisfactory performance at the ultimate limit states, these structures often fail to meet serviceability requirements such as crack width limitations or deformations (BAW 2015). Furthermore, several other time-dependent factors may affect their service life, also depending on climate change (Orcesi 2016) and obsolescence (Langston 2011). But, above all, main degradation phenomena affecting such kind of structures are spalling and corrosion. In order to manage maintenance intervention, the BAW developed an *ad hoc* maintenance management system called EMS-WSV (Bödefeld & Kloè 2012), shortly summarized in the following.

A database software, called WSVPruf, is used to store data collected on each structure during execution, inspection and maintenance. All damages are rated on an increasing scale from level 1 to level 4 (1: good condition; 4: critical condition) and they are recorded in a standard format by the program. Another program called 'Zustandsprognose' forecasts future deterioration stages of the structure for the next 20–30 years. Here, different approaches have been considered:

- Survival functions are applied to describe the deterioration of components where no evident damage is detected at the actual inspection time;
- A method based on discrete-time Markov processes is implemented in order to forecast the deterioration of detected damages. The parameters of the transition matrix are also determined according to survival functions;
- In some cases, physical equations have been used in order to validate Markov Chains.

Once the survival functions and Markov Chains have been defined, the evolution of the damage scale in time can be determined in an almost deterministic way, as a unique process. The remaining lifetime is also expressed in a deterministic way, and it represents the time lapse interval required to the damage scale in order to reach the critical condition, corresponding to level 4.

Obviously, the above mentioned methods present some advantages and drawbacks. On one hand, physical equations are usually considered as deterministic; they model only some deterioration processes and they require a huge amount of data to determine the main parameters of the physical laws. On the other hand, survival functions and Markov chains are powerful and flexible methods and they can be easily adapted to different deterioration processes. But, in both cases, it is difficult to take into account the influence of different factors on future degradation stages. In case of Markov Chains, the state of the structure is described through a unique variable, and there are no consolidated methods according which the parameters of the transition matrix can be defined. It must be also underlined that degradation phenomena largely depend on other factors such as the environmental condition and the quality of the materials, which is mainly the quality of concrete. Anyhow, despite they are crucial to determine the proneness to damage of the structure and to predict the remaining lifetime, these factors are almost completely disregarded by the current approach.

3 AVAILABLE INFORMATION FOR LIFETIME PREDICTION

Whether survival functions of Markov Chains are implemented, the parameters of the models should be determined from real data obtained from inspections, surveillance or *ad hoc* investigations. Obviously, provided that enough data exist, the parameters can be obtained through some sort of regression analysis or statistical investigation; but, since inspections are usually carried out every six years, surveillance is executed not later than three years after each principal inspection, and *ad hoc* inspections are only required after accidental shocking events such as ship impacts or flooding, available data are often not sufficient.

As underlined also by Haider (2012), asset lifecycle management is an information intensive task that requires generating, processing and analyzing huge amount of information. Data usually consist in both qualitative and quantitative information, able to describe the static, constructional and hydromechanical condition of the structure; however quantitative data collected during the inspections, being mainly obtained through simple measuring instruments, are quite rough, and for this reason they are affected by great uncertainties.

Another source of information is represented by expert knowledge. In the present case, this information was previously collected through a Delphi Interview submitted to a total amount of 28 experts (BAW 2009). In the Delphi Interview it was asked to answer the following question: "When does a special damage appear for the first time in your opinion?", or in other words, it was asked to elicit the survival functions given a certain degradation process and a specified degradation level. Three degradation levels were especially considered, notionally corresponding to damage levels 2, 3 and 4, subdividing the asset into three categories: fragile, normal and robust constructions, and a choice sheet with several different predefined time intervals (decades) was provided to the experts to facilitate the comparison of the answers.

Other relevant information can be also extracted by unstructured data and secondary database of tests results by conducting data analysis with suitably developed algorithms and software, as shown by Gao & Koronios (2012) and Croce (2018). This information can be supplemented by the results of *ad hoc* investigation such as material or chemical tests carried out on the considered structure or on similar constructions, built in in the same time period in a given geographical area, adopting comparable construction techniques.

Finally, data need to be acquired about relevant climatic actions influencing some degradation processes, like temperature or moisture, as well as about effects of climate change on them.

Aleatoric and epistemic uncertainties are inevitably associated to the above mentioned information, the second one representing a lack of knowledge, which can be reduced as soon as further data become available.

4 A NEW APPROACH TO LIFETIME PREDICTION

4.1 *Description of the new procedure*

In order to obtain more realistic lifetime prediction, an innovative method is proposed in the paper, where the deterioration phenomenon is modelled by an appropriate stochastic process, a gamma process, while epistemic uncertainties are suitably reduced resorting to the Bayesian Theorem.

Although updating the parameters of a stochastic process through data collected during inspection was already suggested (Bousquet 2014, Haowei 2015), the issue here presented involves some complications that have not been yet considered. One complication is due to the fact that different levels of epistemic uncertainties affect the lifetime prediction. A first level is connected to the proneness of the structure to damage: in fact, even if three categories of constructions can be defined (fragile, normal and robust) according to Chapter 3, the "category" to which the structure belongs is not known *a priori*. A second level is represented by the uncertainty affecting the parameters of the gamma process describing the degradation phenomenon within each category.

Moreover, as information derived from the structure is both qualitative and quantitative, but the degradation phenomenon is mainly described in quantitative terms, the problem becomes more and more complex. Nevertheless, qualitative data are important in order to figure out the sensitivity of the structure to damage, and the variables better describing the degradation process. Thus the question becomes: *how could qualitative and quantitative data be considered in order to reduce the two levels of epistemic uncertainties previously identified*?

A possible answer to this question is to resort to a Bayesian Network (BN) and to use it as a Naïve Bayesian Classifier in order to identify the "Proneness to damage" category of the structure and to remove the first level of epistemic uncertainty. Once the category has been identified, the parameters of the gamma process describing the degradation phenomenon within that class could be further updated considering the data about the damage progression collected during the inspection on the structure.

In the following, the theoretical background to the proposed approach is briefly introduced.

4.2 *Gamma process*

Gamma process is always applicable to model positive and strictly increasing degradation data, as suggested by Nicolai (2007) and van Noortwijk (2009).

A non-stationary gamma process $Y(t)$ with shape function $\Lambda(t) > 0$ and scale parameter β has the following properties:

1. $Y(0) = 0$ with probability one;
2. $Y(\tau) - Y(t) \sim \Gamma\left(\Lambda(\tau) - \Lambda(t), \beta\right), t \in [0, \tau)$;
3. $Y(t)$ has independent increments;

where $\Gamma(.)$ is the gamma function. Conditions (1) and (2) provide the probability density function of a gamma process:

$$f\left(Y, \Lambda(t), \beta\right) = \frac{\beta^{-\Lambda(t)}}{\Gamma\left(\Lambda(t)\right)} Y^{\Lambda(t)-1} \exp\left(\frac{-Y}{\beta}\right). \quad (1)$$

Thus the expected deterioration at time t can be expressed by a power law:

$$E\left(Y(t)\right) = \frac{\Lambda(t)}{\beta} = \frac{\alpha t^c}{\beta}, \quad (2)$$

where $\alpha/\beta > 0$ and $c > 0$ is a parameter describing the shape of the expected deterioration. Values of the parameter c for some relevant deterioration processes are given by van Noortwijk (2009).

Suppose now that the lifetime ξ is defined as the time when $Y(t)$ reaches a suitable failure threshold D (first passage). Then the cumulative distribution function (CDF) of ξ can be obtained as:

$$F(t) = \frac{\Gamma\left(\Lambda(t), D_\beta\right)}{\Gamma\left(\Lambda(t)\right)} \tag{3}$$

where $D_\beta = D/\beta$. Park & Pedgett (2005) showed that the exact probability distribution function (*pdf*) of ξ for a gamma process can be extrapolated from Equation (3). However, since the resulting expression is too complex for practical application, they proposed to approximate the CDF of ξ with the Birnbaum-Saunders (BS) distribution (Birnbaum & Saunders 1969), so that the mean lifetime $\bar{\xi}_{BS}$ simply results:

$$\bar{\xi}_{BS} = \left[D\frac{1}{\alpha\beta} + \frac{1}{2\alpha} \right]^c. \tag{4}$$

4.3 *Bayesian updating*

As known, Bayes' theorem represents an actualization principle and it allows the calculation of conditional probabilities or conditional densities.

In the discrete case, it describes the updating of $p(A_i)$ to $p(A_i \mid B)$ once observed the event B.

In the continuous case, given two random variables X and Z, with conditional distribution of X given Z $f(x \mid z)$ and marginal distribution $g(z)$, it describes the conditional distribution of Z given X:

$$f(z \mid x) = \frac{f(x|z)\, g(z)}{\int f(x|z)\, g(z)\, dz} \tag{5}$$

Depending on the interpretation of probability, the meaning of Bayes' theorem differs significantly. If probability (or density) is interpreted in a frequentistic manner, the theorem expresses the proportion of an event given the occurrence of another event. Conversely, if probability reflects the relative plausibility or degree of belief attributed to a certain event, Bayes' theorem forms the mathematical basis for adjusting or updating the probability as soon as more evidence becomes available.

Also the uncertainty on the parameters θ of a model could be described by a probability distribution, which has to be interpreted in most of the cases as a degree of belief. If the model is represented by a density function, a two-levels hierarchical model can be obtained, in which the second level is represented by the *pdf* on the statistical parameters of the probability model at the first level. The statistical parameters of the second level distribution are usually called 'hyperparameters', and the *pdf* 'hyperdistribution'. Anyhow, the degree of belief can be updated when new data are available, so that, once defined the posterior distribution, statements about the parameter can be made.

The computation of the posterior distribution involves the calculation of several integrals and for this reason is not straightforward. Simplified approaches have been sought in order to facilitate the updating of the prior distribution. The most popular approach is to resort to the so called conjugate prior distributions, which have the appealing features that prior and posterior distributions have the same functional form, and the updated parameters can be computed in an analytical way. In case of the gamma distribution, the conjugate prior distribution is characterized by rate parameter $\theta = 1/\beta$ following a gamma distribution, denoted as $\theta' \sim \Gamma(a', \nu')$ where $\nu' = 1/b'$ is the rate parameter and b' is the scale. Let $\boldsymbol{x} = \left[x_0, x_1, \quad, x_n \right]$ the observed degradation data, $t = \left[t_0, t_1, \quad, t_n \right]$ the corresponding times; denoting with $\Delta x_i = x_i - x_{i-1}$ and $\Delta t_i = t_i - t_{i-1}$ the degradation and the time increments, respectively, the posterior shape and rate parameter for θ can be computed in the following way (Ang & Tang 2007):

$$a'' = a' + \alpha\left(t_n - t_0\right) \tag{6}$$

$$\nu'' = \nu' + x_n - x_0 \tag{7}$$

and the posterior mean of θ can be written as:

$$E(\theta|\Delta\boldsymbol{x}) = \frac{\varepsilon''}{\nu''}. \tag{8}$$

from which the scale parameter β is easily obtained.

4.4 *Bayesian Network*

A Bayesian Network (BN) is a flexible tool that allows a rigorous processing of both quantitative and qualitative information (Kjærulff & Madsen 2008). It is defined as a directed acyclic graph (DAG) which determines a factorization of a joint probability distribution, as the nodes of the DAG represent the variables and the directed links describe the factorization. For each direct link from a node X to a node Y (where X is here the 'parent' and Y the so-called 'child'), a conditional probability $p(Y \mid X) = z$ is attached to Y. The conditional probability expresses a rule that assumes the following form: if $X = x$ then $Y = y$ with probability z, where y and x denote the state of Y and

X, respectively. Usually Y represents the effect of X, typically not observable by itself, but whose state is inferable via prior *pdf* $p(X)$, conditional *pdf* $p(Y \mid X)$ and Bayes' theorem:

$$p(X = x|Y = y) = \frac{P(Y = y, X = x)}{\sum_X P(Y = y, X = x)} \quad (9)$$

Observation may have the form of hard evidence if zero probability is assigned to all but one state; otherwise it is said to provide soft evidence. So far it is implied that the nodes of the BN represent discrete variables, at which probability tables expressing conditional probabilities are attached. However it is also possible to have continuous variables: in this case it is necessary to specify a density function for each combination of states for the parent variables. Whether discrete or continuous variables, BN can be applied to solve a wide range of problems, and, *inter alia* for classification purposes. In this specific case, they are especially called Naïve Bayesian Classifier, because of the strong (naïve) independence assumptions assumed among the features that determine the class label drawn.

5 CASE STUDY

In order to clarify the proposed approach, a case study is here developed. The attention is mainly focused on the elicitation of the prior gamma distribution describing the stochastic process; this task is carried out exploiting the approximation of the lifetime distribution with the BS distribution. The case study aims to illustrate a practical application of the proposed method referring to a unique damage detected on an existing lock. Extensions to cases involving several damages and different deterioration processes will be discussed in future works.

5.1 *Elicit the prior distribution*

A prior distribution for the model parameters is elicited according to the expert knowledge collected through the Delphi interview described in Chapter 3. It would be easier if experts could have been asked to represent their opinion in statistical terms, for example: "*Given the degradation phenomena modelled through a gamma process, what would be the shape and the scale parameters of the gamma distribution?*". However it is unlikely that experts could give an accurate answer. Except in case of symmetric distribution, when mean value and standard deviation can be easily elicited, they usually think in terms of percentile. Moreover, expert knowledge actually concerns expected lifetimes rather than damage increments; for this reason, the prior information should be transformed in a form leading to an easy identification of the prior gamma distribution. One way to do this is by using Equation (4) which links the parameters of the gamma distribution with the expected lifetime. Given the expected lifetime required to reach certain degradation levels, and the speed of the process, represented in some way by the value of c, the parameters of the gamma distribution can be easily obtained by solving the system of equations. In this case, two equations are required to determine the two parameters: those related to damage levels 2 and 3 are considered, as expert knowledge about events that happened earlier should be more reliable compared to that related to future events. The uncertainty on the parameter is similarly defined, considering the uncertainty affecting the expected lifetimes and assuming a confidence interval. The procedure, which is very general, can be applied to elicit prior distribution for any degradation process.

In the specific case, in Table 1 the average of the expert predictions are summarized in terms of time lapses required by the crack width to reach the levels corresponding to different degradation class, while in Table 2 the confidence intervals of the predicted time lapses, corresponding to probability of exceedance of 75% and 25%, respectively are reported. Given the progress in time of the degradation process, it is also possible to conclude that the degradation speed can be assumed constant, at which $c = 1$ corresponds. For example, in case of fragile construction, the system of equations results:

$$\begin{cases} 10 = \dfrac{0.3}{\alpha\beta} + \dfrac{1}{2\alpha} \\ 20 = \dfrac{0.7}{\alpha\beta} + \dfrac{1}{2\alpha} \end{cases}. \quad (10)$$

The values of parameters α and β obtained by solving all the systems of equations are shown in

Table 1. Expert knowledge results—Delphi Interview – (DR: damage level, CW: crack width, $\bar{\xi}$: expected lifetime, F: fragile, N: normal, R: robust).

		$\bar{\xi}$ (years)		
DL	CW	F	N	R
2	0.3	10	25	50
3	0.7	20	50	100
4	1	30	75	150

Table 3, while the simulation of the gamma process for fragile, normal and robust structures with these values are shown in Figures 1, 2 and 3, respectively.

For analytical and mathematical tractability, it is then assumed that only β is random and it also follows a gamma distribution, so that the prior distribution is then conjugate through the likelihood function to the posterior, simplifying the updating calculation. The prior distribution can be defined by considering the uncertainty affecting the expected lifetime: each time lapse interval corresponds to an interval for the parameter β. Assuming the same probabilities of exceedance, it will be possible to elicit the shape a' and scale b′ (or the rate v') parameters of the prior distribution (Table 3).

5.2 *Elicit the Bayesian Network*

The Bayesian Network is elicited according to prior information and the previously defined gamma process. In the remainder of the paper, only the most fundamental variables will be considered; anyhow, as soon as that further analysis will be carried out, it will be possible not only to consider the relationships among a greater number of variables, but also to elicit the structure of the network and the conditional probabilities from data itself rather than from expert knowledge. The variables that will be considered are: the proneness to damage of the structure (also called damage category (DC), characterized by three possible states: Fragile (F), Normal (N) and Robust (R)), the concrete quality (CQ, characterized by three possible states: Bad (B), Normal (N), Good (G)), the damage quantity (DQ, characterized by three possible states: extended (E), limited (L), sporadic (S)) and

Table 2. Confidence intervals for the lifetime ξ predicted by the experts (F: fragile, N: normal, R: robust).

$\xi_{25\%} - \xi_{75\%}$ (years)		
F	N	R
8–12	20–30	40–60
16–24	40–60	80–120
24–36	60–90	120–180

Table 3. Elicited parameters of gamma process and statistical parameters of the gamma distribution describing the uncertainty over the scale β.

	α	$\beta_{50\%}$	$\beta_{25\%}$	$\beta_{750\%}$	a′	b′	v′
F	0.2	0.2					
N	0.08	0.2	0.15	0.25	7.27	0.75	1.33
R	0.04	0.2					

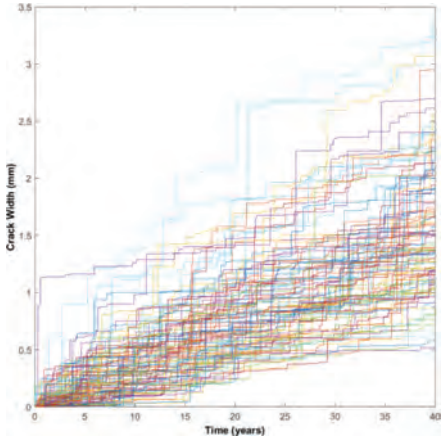


Figure 1. Simulation of the gamma process for fragile structures (100 sample paths).

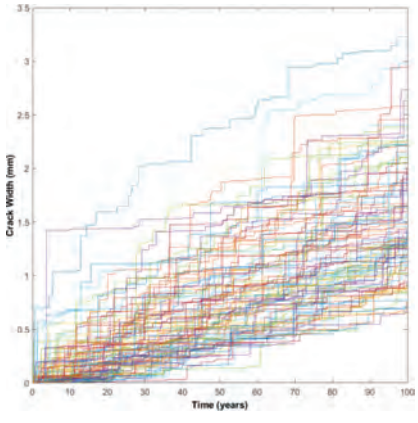


Figure 2. Simulation of the gamma process for normal structures (100 sample paths).

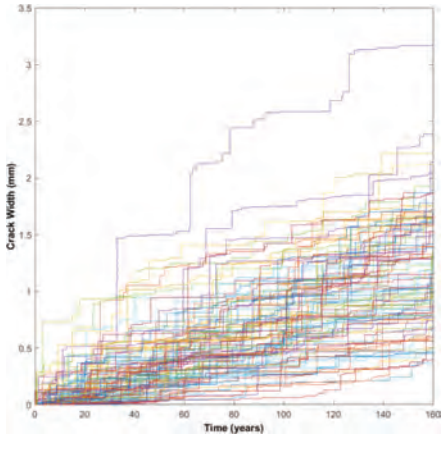


Figure 3. Simulation of the gamma process for robust structures (100 sample paths).

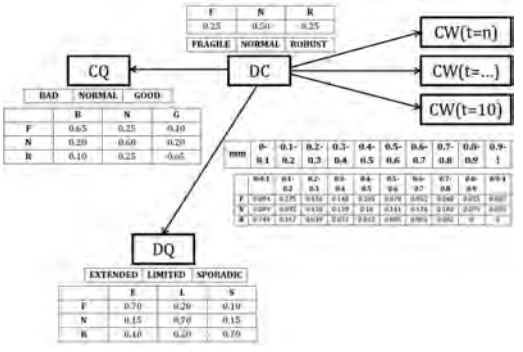


Figure 4. Bayesian Network implemented in order to classify the structure with respect to its proneness to damage.

the crack width referred to different time steps (CW), characterized by 10 possible states.

The states of each variable, the dependency relationships and the associated probability tables are shown in Figure 4. For the sake of simplicity only the probability table referred to t = 10 years is here showed.

The concrete quality, the damage quantity and the crack width can be observed, while the updating of the probability of the damage categories, that cannot be observed, is a key objective. It is assumed that the observable variables are independent, given the variable that it should be classified. Although this assumption facilitates calculation of posterior probabilities, it is actually inaccurate, as it is often the case with Naïve Bayesian Classifier. Nonetheless it has been demonstrated that the crude independence model could perform surprisingly better than more complicated alternatives (Hand 2001); furthermore precise estimation for the posterior class probabilities are here not required, since they will be used for comparison purposes, as better clarified in Chapter 5.3. By collecting evidence regarding the observable variables, it will be possible to infer the proneness to damage of the structure. It is important to underline that assuming a parametric deterioration process simplifies the relationships among the variables; in effect, the deterioration at instant t can be considered independent from the deterioration at the previous instant, preventing us to resorting to a much complicated dynamic Bayesian Network (Straub 2009, Ramìrez & Utne 2015). Furthermore, the probability table associated to the variable 'crack width' can be easily obtained simulating the gamma process. The variable is actually continuous, and the degradation at each time step t is represented by a gamma distribution with shape αt and scale β (see Chapter 4.2). However BNs that involve continuous random variables imply some complications, which are outside the purposes of the present paper and will be discussed later in further studies. Moreover, as it can be seen later from Table 4, data about crack widths are actually given in terms of intervals rather than single values, since, as already stressed, the measuring devices are rough and measures are inevitably affected by great uncertainty. Also for this reason, the continuous variables are then discretized so that each state corresponds to collectable measures.

5.3 *Implementation of the Bayesian Network and updating of the gamma process*

Data referring to three different structures (A, B and C) and related to a damage event having a single cause are considered in the case study (Table 4).

First of all the data are used in order to classify the structure by using the previously defined BN. If all the data point out that one category has very high updated probability (for example higher than 0.70), then it is possible to directly consider the Gamma distribution corresponding to that category and the related uncertainty of the statistical parameters for a further Bayesian updating. However, in several cases, the results could not point out clearly at a certain category, because two adjacent categories have similar updated probabilities; in this case, it is possible to conservatively consider only the lowest category, or to collect further data.

Once that the prior model will be chosen, it is possible to update it resorting to the analytical formula expressed by Equations (6) and (7).

Table 4. Inspection results on three structures, A, B, C, (CQ: concrete quality, DQ: damage quantity, CW: crack width).

			1st inspection			2° inspection		
Str.	Year	CQ	Year	DQ	CW	Year	DQ	CW
A	1995	1	2005	1	0.5	2017	1	0.8
B	1975	2	2005	2	0.5	2017	2	0.7
C	1955	3	2005	3	0.3	2017	3	0.4

Table 5. Updated probabilities for damage category (DC) derived from the Bayesian Network, updated statistical parameters for gamma process, prior and updated lifetime prediction $\bar{\xi}$ in years, obtained simulating the gamma process.

	F	N	R	ε''	v''	$\bar{\beta}''$	$\bar{\xi}'$	$\bar{\xi}''$
1	0.96	0.03	0.01	9.67	1.62	0.16	28	33
2	0.07	0.72	0.21	8.23	1.52	0.184	65	70
3	0.01	0.09	0.90	7.75	1.42	0.183	135	140

Finally β'' can be used to improve the prediction of the structure lifetime when the limit damage level will be reached; the prediction can be obtained by simulating the gamma process considering the updated parameter.

The results are listed in Table 5.

6 CONCLUSION

An innovative procedure to lifetime prediction exploiting the potentialities of the Bayes' Theorem has been proposed, as depicted in the flowchart shown in Figure 5, based on the following points:

- The degradation phenomenon is modelled as a stochastic process in which degradation increments are gamma-distributed.
- Considering background information and qualitative and quantitative data collected during inspections, a Bayesian Network is adopted to classify the proneness to damage of the structure.
- Uncertainties affecting the parameters of the stochastic process are updated considering quantitative data about the degradation level reached by the structure by applying the Bayes Theorem.

Advantages of the proposed method are:

- Qualitative and quantitative data can be used at the same time, to refine lifetime predictions.
- Adoption of gamma processes to model degradation phenomena simplifies the structure of the Bayesian Network and the computation of the conditional probabilities.

A case study has been finally developed in order to show a practical application of the proposed method. Attention has been largely devoted to the elicitation of the prior distribution for the gamma process, and real data collected during inspection on real structures have been considered. In authors' opinion, this approach is very promising although some aspects need further studies. In effect, the value of the constant c, that regulates the speed of degradation, should be determined more precisely, also considering different degradation phenomena. An improved Bayesian Network including continuous variables, representing the state of damage at different time steps and whose parameters are directly derived from measurements, should be suitably developed. Uncertainties in the collected data, as well as uncertainties in the degradation limits should be duly considered too. At the final stage, once the proposed procedure is improved and suitably validated, its efficiency should be also verified by comparing theoretical time lapse predictions with empirical data on a sufficiently large number of existing structures. The approach can be implemented in order to plan in a more precise and realistic way asset life-cycle, and to identify on which structures maintenance interventions should be firstly performed.

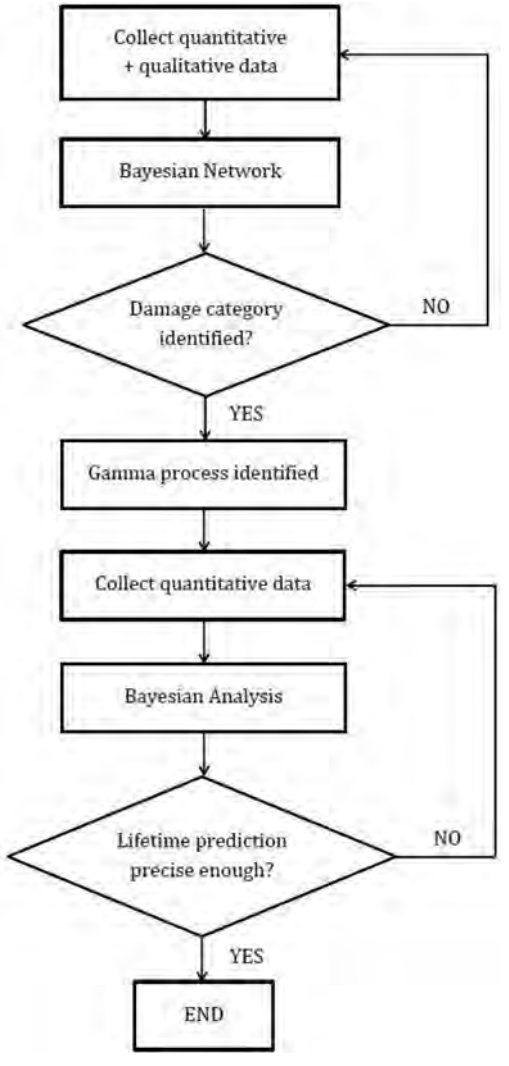


Figure 5. Flowchart summarizing the proposed procedure.

REFERENCES

Ang, H.S. & Tang, W.H. 2007. *Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering,* 2nd Ed. Hoboken, NJ, J. Wiley & Sons.

BAW 2009. Erhaltungsmanagementsystem für die WSV, Phase2 – Meilensteinbericht 2009 (internal report).

BAW 2015, Schadensklassifizierung an Verkehrswasserbauwerken (MSV), www.baw.de.
Birnbaum, Z.W. & Saunders, S.C. 1969. A new family of life distributions, J. of Appl. Prob., 6, 2, 319–327.
Bödefeld, J. & Kloè, K. 2012. Managementsystem for infrastructure of waterways, 3rd Int. Symp. on Lyfe Cycle Eng.
Bousquet, N. Fouladirad, M. Grall, A. Paroissin, C. 2015. Bayesian gamma processes for optimizing condition-based maintenance under uncertainty, Appl. Stoch. Mech. Bus. Ind., 31 360–379.
Croce, P. Marsili, F. Klawonn, F. Formichi, P. Landi, F. (2018), Evaluation of statistical parameters of concrete strength from secondary experimental test data, Construction and Building Materials, 163, 28, 343–359.
Gao, J. & Koronios, A. (2012), Unlock the Value of Unstructured Data in EAM. Proc. 7th W. Cong. Eng. Asset Man.
Haider A. (2012), Asset Lifecycle Data Governance Framework, Proc. 7th W. Cong. on Eng. Asset Management.
Hand, D.J. & Yu, K. (2001), Idiot's Bayes—not so stupid after all?, International Statistical Review. 69 (3): 385–399.
Haowei, W. Tingxue, X. Qiaoli, M. 2015. Lifetime prediction based on Gamma processes from accelerated degradation data, Chinese Journal of Areonautics, 28(1): 172–179.
ISO 55000, Asset management—Overview, principles and terminology, 2014.
Kjærulff, U.B. & Madsen, A.L. (2008). Bayesian Networks and Influence Diagrams, Springer.
Langston, C. (2001), Estimating the useful life of buildings, Conference Papers, Paper 30.
Lim, R. & Mba, D. (2013), Fault Detection and Remaining Useful Life Estimation Using Switching Kalman Filters, Eng. Asset Management—Systems, Professional Practices and Certification, Lecture Notes in Mech. Eng.
Nicolai, P. Dekker, R. van Noortwijk, J.M. 2007. A comparison of models for measurable deterioration: An application to coatings on steel structures, Rel. Eng. & Sys. Safety, 92, 1635–1650.
Orcesi, A.D. Chemineau, H. Lin, P.-H. van Gelder, P. van Erp, N. (2016), A Risk Analysis for Asset Management Considering Climate Change, Trans. Res. Proc., 14, 105–114
Park, C. & Padgett W.J. 2005. Accelerated Degradation Models for Failure Based on Geometric Brownian Motion and Gamma Processes, Lifetime Data Analysis, 11, 511–527.
Ramìrez, P.A. & Utne, I.B. 2015. Use of dynamic Bayesian networks for life extension assessment of ageing systems, Rel. Eng. & Sys. Safety 133, 119–136.
Riascos-Ochoa, J. Sànchez-Silva, M. Klutke, G.A. 2016. Modeling and reliability analysis of systems subject to multiple sources of degradation based on Lévy processes, Prob. Eng. Mech., 45, 164–176.
Straub D. 2009. Stochastic modeling of Deterioration Processes through Dynamic Bayesian Networks, J. of Eng. Mech. ASCE 135:1089–1099.
Trappey, A.J.C. Trappey, C.V. Tsao, W.-T. (2012), Engineering Asset Life Span Evaluation Using Logistic Regression, Proc. 7th W. Cong. Eng. Asset Man
Tse, P.W. & Shen, C. (2013), Remaining Useful Life Estimation of Slurry Pumps Using the Health Status Probability Estimation Provided by Support Vector Machine, Engineering Asset Management—Systems, Professional Practices and Certification, Lecture Notes in Mechanical Engineering.
Van Noortwijk, J.M. 2009. A survey of application of gamma processes in maintenance, Rel. Eng. & Sys. Safety 94, 2–21.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A methodology for selecting and defining maintenance tasks for critical equipment

M. Sousa
Department of Production and Systems, University of Minho, Guimarães, Portugal

I.S. Lopes
Department of Production and Systems, Algoritmi Research Centre, University of Minho, Guimarães, Portugal

ABSTRACT: This paper presents a methodology, developed through a case study, to support preventive maintenance tasks selection based on Reliability Centered Maintenance (RCM) methodology. The case study was carried out in an electronic goods company that follows the Total Productive Maintenance (TPM) philosophy. The methodology has six steps and it is intended to be applied in every mechanical and electromechanically equipment of the plant and it is applicable in equipment with some historical data of interventions. To validate the methodology, one system of the plant was selected and all the steps of the methodology were followed. With this application, a reduction of 33% in failures number and production losses was achieved, as well as a MTBF improvement of about 15% and therefore an increase in machine yield.

1 INTRODUCTION

1.1 *Background*

Companies have sought to improve their operational efficiency with the aim of offering high value added products (Ahuja & Khamba 2008; Eti et al. 2004; Konecny & Thun 2011). The growth of mechanization and automation of plants created a need to improve production equipment reliability, availability and productivity. This way, consumer needs can be pleased and at the same time competitiveness is improved (Bakri et al. 2012; Muchiri et al. 2011; Soni 2013).

The maintenance management area gives the means to deal with this challenge because it allows to improve systems performance (Kutucuoglu & Hamali 2001), instead of the traditional vision that maintenance actions are only applicable in repairing "broken items" (Ahuja & Khamba 2008).

Indeed, Moubray (1997) considers that modern management challenges are: to select the most appropriated techniques, to deal with all failure modes, to have cost effective processes and to satisfy customer expectations and of all the society.

1.2 *Objectives*

The main goal of this paper is to present a maintenance planning methodology. The methodology is based on Reliability Centered Maintenance (RCM) process and aims to be applicable in all mechanical and electromechanical equipment. By using failure data, it will define the most appropriated maintenance tasks, as well their periodicity. In addition, with the development of the methodology the following objectives are intended to be achieved:

- Balancing corrective and preventive maintenance tasks;
- Improving key performance indicators of the machines;
- Reducing total maintenance cost;
- Linking the maintenance planning methodology and Total Productive Maintenance (TPM) philosophy;
- Matching qualitative and quantitative maintenance management approaches.

1.3 *Paper outline*

This paper is structured in six sections. In the first one is presented the background, the goals and the structure of the paper. In the second one, a short literature review is produced to frame the subject of the paper. The third section is on the development of the critical equipment maintenance planning methodology and explains all its steps. In section four, one machine of the case study company´s plant is used to exemplify the methodology. The last section presents the paper conclusions.

2 LITERATURE REVIEW

Maintenance management has evolved in last decades and it can be divided in three main generations The first one is related to period before the

second world war (Moubray 1997). In this period, machines were simple and very reliable. Hence, maintenance was based on corrective actions to repair the equipment when some failure occurs (Ahuja & Khamba 2008; Garg & Deshmukh 2006; Soni 2013). After the second world war, equipment became more complex which triggered the first preventive maintenance tasks (Alsyouf, 2007; Moubray, 1997).

More recently, the research made in maintenance management field allowed to innovate these concepts and new techniques emerged, such as condition monitoring and failure mode and effects analysis (FMEA) (Moubray 1997). At the same time, TPM and RCM philosophies were developed (Garg & Deshmukh 2006; Soni 2013).

According to NP EN 13306 (2007) standard, maintenance is defined as a "combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function". The same standard also defines the maintenance interventions types or strategies. In a first level, the maintenance concept can be divided into preventive or corrective. Performing preventive maintenance tasks reduces the number of failures. As the cost of preventive maintenance increases with increasing frequency of interventions, the cost of corrective maintenance reduces. Thus, a replacement periodicity which minimizes maintenance costs should be found out (Figure 1).

Until the 60th decade, maintenance was based on the premise that every part has a defined life time at which it is necessary to do a general review, so it was applied the scheduled maintenance approach (Siddiqui & Ben-Daya 2009). However, in 1960, airline companies started to realize that this mindset was not cost-effective. The found solution was to study the failure pattern of every component.

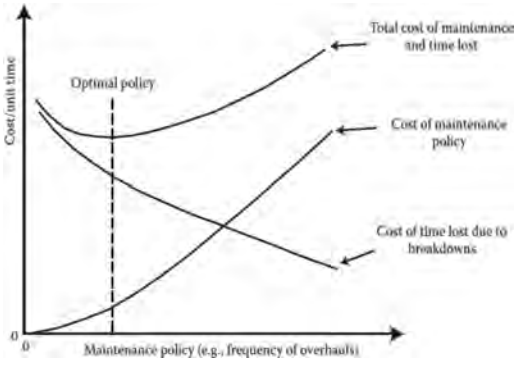


Figure 1. Maintenance total cost (source: Jardine & Tsang, 2013).

Hence, it was proposed the Maintenance Steering Group publications, that gives rise to RCM methodology named by Stanley Nowlan & Howard Heap (Siddiqui & Ben-Daya 2009). The Maintenance Steering Group is a coordination group that defines maintenance plans to new aeronautics projects. The main goal of RCM methodology is to keep system functions during its operational time. To achieve this intention, RCM uses some maintenance strategies such as time-based maintenance (TBM), condition-based maintenance (CBM), run-to-failure (RTF), hidden failures and design modification (Siddiqui & Ben-Daya 2009). This methodology is also supported by FMEA, Failure Tree Analysis and other decision making tools (Backlund & Akersten 2003; Eti et al. 2004; Siddiqui & Ben-Daya 2009; Waeyenbergh & Pintelon 2002). One definition of the RCM process given by Ahuja & Khamba (2008) is: "structured, logical process for developing or optimizing the maintenance requirements of a physical resource in its operating context".

RCM methodology is based on system functions, failure modes and effects study and its aim is to define the system maintenance requirements (Ahuja & Khamba 2008; Carretero et al. 2003; Mokashi et al. 2002; Siddiqui & Ben-Daya 2009). The selected maintenance tasks should be effective in reducing the possibility of occurrence of failure modes as well as applicable (Mokashi et al. 2002; Siddiqui & Ben-Daya 2009; Smith 1993). In this scope, "applicable" means that the task should avoid or detect the failure, while "effective" is related to costs (Siddiqui & Ben-Daya 2009).

In terms of application, Smith (1993) shows that RCM implementation should follow seven steps:

1. System selection and information collection;
2. System barriers definition;
3. System description and functional block diagram;
4. System functions and functional failures;
5. FMEA;
6. Decision tree analysis;
7. Task selection.

Eisinger & Rakowsky (2001) developed a new version of the decision tree because they believe that the questions should not be answered with total assurance. Thereby, the authors developed a tree in which the path is decided in terms of probabilities. A set of eight questions are answered according to the probability of the answer be "YES". After that, a value r_i is calculated using the traditional method of decision trees, where i is the index of each option. At the end, the option with bigger r_i is selected. This procedure was applied to a fire detection and extinction system.

Waeyenbergh & Pintelon (2002) developed a framework that combines TPM and RCM philosophies with an economic concern, once they claim that maintenance is now an economic concern. Hence, to each question of the decision tree, one is requested to answer first with a technical point of view and then with an economic criteria. Finally, the decision is the use of one the five possibilities of maintenance tasks or the search for additional data.

Rastegari & Mobin (2016) studied the case of a gearbox Swedish company. These authors settled up two flowcharts with two different points of view: technical and output-based. The types of questions are different to each flowcharts, however the result is the selection of one of the following tasks: Time-Based Maintenance (TBM), Breakdown Maintenance (BM), CBM or Improvement and Restoration.

Dehghanian, Fotuhi-Firuzabad, Aminifar, & Billinton (2013) recognized the RCM importance but also the barriers in the implementation of this methodology in energy industry. To improve the current situation, they developed a RCM version with a cost benefit analysis, as well as an output evaluation to help to improve maintenance plans. Following a set of steps similar to Smith (1993) approach, the authors have proposed a prioritization of maintenance tasks through Benefit-to-Cost Ratio (BCR) index.

The main benefits of RCM cited in the literature are:

- Maintenance plans quality improvement (Deshpande & Modak 2002);
- Increase of equipment reliability and availability (Backlund & Akersten 2003);
- Ensure safety (Backlund & Akersten 2003);
- Increase of equipment lifetime (Carretero et al. 2003);
- Maintenance costs reduction (Carretero et al. 2003; Waeyenbergh & Pintelon 2002);
- Unplanned downtime reduction (Backlund & Akersten 2003; Carretero et al. 2003).

On the other hand, the most cited disadvantage is the fact that RCM process is based in a qualitative base and focused only on system reliability and do not quantify the maintenance costs (Braglia et al. 2013; Waeyenbergh & Pintelon 2002).

3 METHODOLOGY

The proposed methodology aims to define a cost-benefit maintenance plan based on machine's failure modes. The process follows the RCM principles and uses some TPM concepts once the TPM philosophy is currently adopted by the case study company. Thus, the methodology has six steps and starts with the selection of the critical system, the team building and collection of some data about the system selected. The second step consists in system description and identification of analysis boundaries. Afterwards, the system functions and functional failures are defined. The fourth step uses a matrix to find the relationship between components and functional failures as base to apply the machinery FMEA technique. In the next step, a decision tree is proposed to help to find out which maintenance tasks can be used to mitigate the critical failure modes defined by FMEA. By the end, it is proposed a cost analysis step that allows choosing a cost-benefit plan for the machine.

3.1 *System selection, team building and data collection*

This step starts with a prioritization of equipment based on three criteria: Mean Time Between Failures (MTBF), spare parts cost (C) and preventive maintenance total time (T). Using Analytic Hierarchy Process (AHP) method, it was assigned weights to the mentioned criteria with the support of the case study company. This procedure was done with help of Saaty's relative importance scale (Table 1).

The result matrix is shown in Table 2 that was validated by a consistency test.

Table 1. Saaty's relative importance scale.

Intensity	Meaning	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Weak importance of one over another	Experience and judgment slightly favor one activity over another
5	Essential or strong importance	Experience and judgment strongly favor one activity over another
7	Demonstrated importance	An activity is strongly favored and its dominance demonstrated in practice
9	Absolute importance	The evidence favoring one activity over another is of the highest possible order of affirmation
2, 4, 6, 8	Intermediate values	When compromise is needed between two definitions

Table 2. Matrix of comparison of criteria.

	MTBF	C	T	Weight
MTBF	1	0,2	0,3333	0,1096
C	5	1	2	0,5813
T	3	0,5	1	0,3092
Sum	9	1,7	3,3333	1

Table 3. Criteria's scale.

Level	MTBF (min)	C (€)	T (min)
1	MTBF > 4897	C < 390	T < 333
2	4897 ≥ MTBF > 1825	390 ≤ C < 703	333 ≤ T < 442
3	1825 ≥ MTBF > 976	703 ≤ C < 1483	442 ≤ T < 664
4	976 ≥ MTBF > 766	1483 ≤ C < 5456	664 ≤ T < 976
5	MTBF ≤ 766	C ≥ 5456	T ≥ 976

At this point, it is necessary to solve another concern. Each criterion has its own measure, thus, a quantitative scale with five levels was defined, where 1 is assigned to the best scenario and 5 is assigned to the worst. Table 3 presents an example that was obtained with data of the case study company.

The levels shown in Table 3 were obtained by analyzing a random sample of the plant machines. For each type of machine the MTBF, spare parts cost and preventive maintenance time was collected from a period of 18 months. The levels were chosen using the percentiles 20, 40, 60 and 80 of the sample.

After the collection of the values, the machine ranking (R) is obtained by equation (1) considering the previously defined weights.

$$R = 0{,}1096*MTBF + 0{,}5813*C + 0{,}3092*T \quad (1)$$

The critical equipment is the one that has the lowest R value.

After the selection of the more critical equipment, the team is built according to the equipment type. It is also necessary to collect some information about the equipment, such as: supplier specifications manual, defined process rules for production and events data. The latter is composed by corrective, preventive and improvement actions.

3.2 *System description and limits*

The second step comprises the steps 2 and 3 of original RCM procedure. The procedure is simplified because this methodology will analyze machines with known operation behavior. Hence, this step consists in drawing the functional diagram of the equipment and lists the analysis limits, i.e. the system parts that will not be considered.

In this scope, a standard structure for all machines was considered, as shown in Figure 2.

Following the Figure 2 concepts, a "system" is a complex structure that has several functions. Each function is associated to a "subsystem" to facilitate the organization. In the next level appears the "parts" at which failure modes are associated. Lastly, the "causal elements", i.e., the reasons of failure modes (causes), are connected to the failure modes.

3.3 *Function analysis and functional failures*

In the third step, the subsystems functions are listed and each function should be described on basis of quantitative requirements, if possible. This way, it will be easier to understand the functions that make the machine useful.

Afterwards, the functional failures to each function, i.e., the conditions in which some function is not achieved, are listed. Only some of these functional failures will progress in the analysis. Hence, the team should choose the critical functional failures in terms of department objectives, costs, maintenance time required or breakdown frequency.

3.4 *Failure Mode and Effects Analysis (FMEA)*

In this step, the authors added a preparation phase before FMEA. Using a matrix with parts in line and functional failures in column, the team will identify what parts have relationship with what functional failures. This way, only critical parts will be analyzed in FMEA.

To each of these parts, the team will deliberate what failure modes can occur based on experience. To each failure mode is associated an effect at system level. This effect should be described in two ways: machine condition and error message. One example is "Machine stop with error 'NOK'", so after the failure mode has occurred, the machine has stopped and the message error was displayed on the monitor.

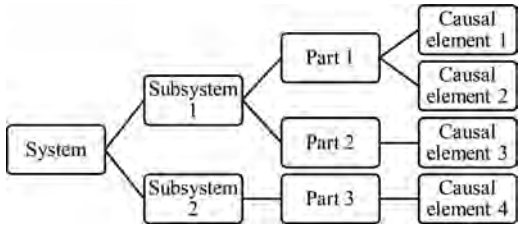


Figure 2. Standard structure of a system.

Afterwards, the team needs to find the causes. In this methodology three types of causes are considered: causal element, operation error or raw material issue. These three groups are adopted because the plant has the experience that an operation error or a raw material issue causes a breakdown of the equipment, so these problems should be considered to mitigate them since they affect machine and plant performance.

Detection and prevention actions should be filled in FMEA form. Examples of detection actions are: error messages, parameter changes and sensorial detection of an irregular behavior. It needs to be highlighted that not only detection before failure should be considered but also the detection modes after failure occurrence.

Regarding prevention actions, redundancies, tests or other methods that prevent the failure mode are considered. This field should not be confused with preventive maintenance actions.

Lastly, the Risk Priority Number (RPN) is calculated. In this methodology, a critical failure mode is characterized by a RPN > 100.

3.5 *Decision tree*

The proposed decision tree (Figure 3) is based on RCM original concepts but also in the autonomous maintenance and the planned maintenance TPM pillars. Thus, the alternative maintenance policies considered are: CBM, TBM, Design-Out Maintenance (DOM) and RTF.

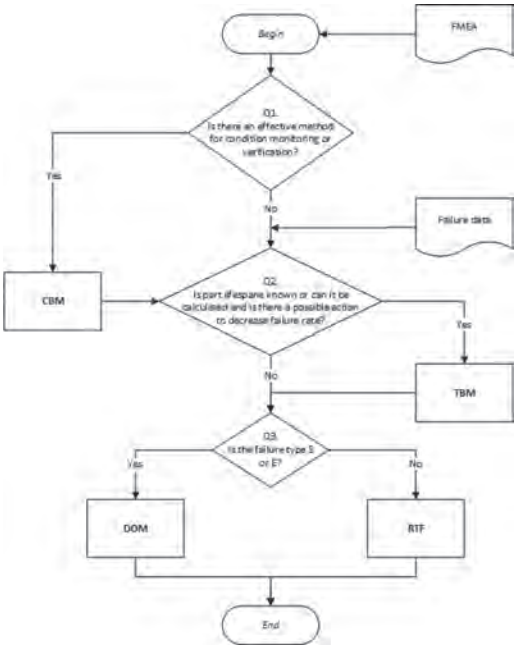


Figure 3. Decision tree.

The application of the decision tree is simple and intuitive. To each failure mode, the first question that is done is: "Is there an effective method of monitoring or checking the condition?" This question is related to CBM policy. If the answer is positive, the policy is taken as an option. However, the analysis continues once the goal is to find out all the applicable maintenance tasks. If the question is "No", the analyst goes directly to the second question.

Retaking the analysis, the second question is: "Is the component life span known or can it be calculated and is there an effective action that can decrease the failure rate?" In other words, to apply a time-based maintenance action it is needed to have some available historical failure data to calculate the life span of the part and an action that reduces the failure rate. Going down, the third question is: "Is the failure type S or E?" The failure modes can be classified in one of four types: safety (S), environment (E), operational (O) or none (N). According to RCM, a failure with impact on safety or environment should not occur, so for these types the option is a design modification (DOM). A RTF task can be considered to the other two types of failures.

For CBM, TBM and RTF policies there is the possibility to involve the operator in the maintenance action. So, after the selection of the applicable policies, it should be asked "Can the task be performed by the operator?" This shows the TPM influence in this methodology and it was considered because an autonomous maintenance task is more cost-effective, it makes maintenance planning easier and improves the operator skills and motivation.

3.6 *Economic evaluation and maintenance task selection*

The last step of this methodology has three main goals: to characterize the failure data distribution, to define the preventive intervention periodicity and to select the more cost-effective maintenance task.

The failure pattern of the component can be estimated by adjusting the failure data to a Weibull distribution. This distribution was chosen for the reason that it is adaptable because of its three parameters: form (β), scale (η) and position (γ) (Jardine & Tsang 2013).

In the end, it is possible to estimate average and standard deviation by equations (2) and (3).

$$\mu = A * \eta + \gamma \tag{2}$$

$$\sigma = B * \eta \tag{3}$$

The next goal is to calculate the optimal periodicity of preventive intervention, tp, if TBM was one of the selected policies. To obtain this value, maintenance costs should be balanced to minimize total cost (Jardine & Tsang 2013). Hence, the aged based policy is considered and involves two types of costs: failure cost (Cf) and preventive cost (Cp) (Jardine & Tsang 2013).

To obtain Cf spare part cost, workforce cost (MDO) and production loss (PL) cost are accounted for, as equation (4).

$$C_f = C_{spare\ parts} + C_{MDO} + C_{PL} \tag{4}$$

On the other hand, Cp is calculated by spare part cost and workforce cost – equation (5).

$$C_p = C_{spare\ parts} + C_{MDO} \tag{5}$$

This effort to obtain the cost values allows the use of the Glasser graph to find out the tp value (Fig. 4).

The Glasser graph allows reading z and ρ values. The variable z is used to obtain tp through equation (6) (Jardine & Tsang 2013).

$$tp = \mu + z * \sigma \tag{6}$$

The value ρ means the cost of optimal policy as a percentage of RTF policy cost (Jardine & Tsang 2013).

Finally, the annual cost of the different options is estimated. The authors have defined this period of time to make easier the cost comparison.

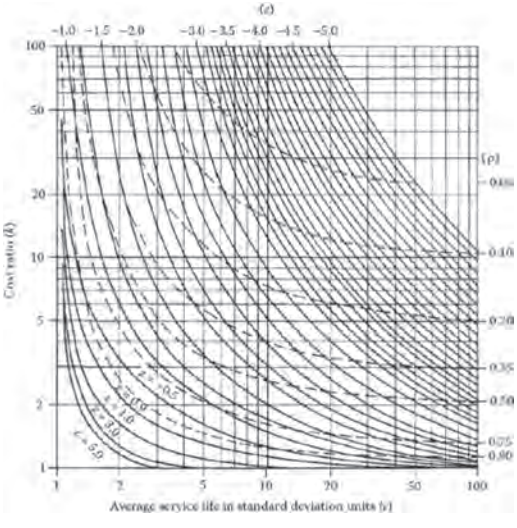


Figure 4. Glasser method for age-replacement policy.

The first one is CBM and its cost involves the preventive maintenance cost C_p, an inspection cost ($C_{inspection}$) and the investment value (Inv), if required.

$$C_{CBM} = f_1 * C_p + f_2 * C_{inspection} + (Inv/n) \tag{7}$$

Equation (7) assumes that failure is always predictable and f1 is the number of preventive intervention in a year and f_2 the number of inspection actions also in a year. The f_1 value can be calculated using a reliability study or estimated based in team experience. The f_2 value should be estimated the same way. The value of Inv is divided by n that indicates the expected number of years in which the condition monitoring equipment is expected to function well.

For TBM two distinct calculation formulas were considered.

$$C_{TBM} = \rho * f_3 * C_f \tag{8}$$

Equation (8) considers an optimal scenario where data is available and in which f_3 is the inverse of MTTF. About ρ, this variable allows to obtain TBM cost based on RTF cost. Hence, equation (8) represents the optimal TBM policy cost (with periodicity tp).

$$C_{TBM} = f_4 * C_p \tag{9}$$

On the other hand, equation (9) will be used when there is no available data to use Weibull distribution to estimate tp. Here, f_4 is an annual frequency defined by the team based on its experience.

For the RTF policy, the cost is obtained by equation (10).

$$C_{RTF} = f_3 * C_f \tag{10}$$

In case of equation (10), f_3 should be in an annual base.

After the calculation of cost through the previous equations, the maintenance task to apply the critical failure mode should be chosen. Between TBM, CBM and RTF policies, the cheaper one will be chosen. However, if DOM is an option, the team needs to analyze its benefits comparing to the cheapest task. In case of S or E effect of failures, it shall be analyzed if the probability of failure is reduced to a level considered adequate. Therefore, for each policy the associated probability of failure (or reliability) must be calculated and it should be made an analysis of costs together with an analysis of this probability.

4 APPLICATION

The described methodology was applied to electronic goods of the company plant.

The critical equipment was chosen between 10 types of machines. Table 4 shows the result of the prioritization (step 1).

In Table 4, it can be seen that GP is the machine with biggest rating value. This machine is a dispensing equipment whose process consists in dispensing a mixture with thermal conductivity properties.

After the selection of the critical equipment, a team of five members with multidisciplinary knowledge was formed. One of the members was the moderator.

Related to historical failure data, the interventions done during year 2016 were collected.

The first task of the team was to draw a functional diagram of the system GP. The result was a set of eight subsystems and it is shown in Figure 5.

The team has deliberated that subsystems 6 and 7 should be excluded from analysis because of their lower failure rate.

The third step is functions and functional failure study. In Table 5 it is possible to see an application example for dispensing subsystem.

In this case, the two functional failures go to next step. Following this thought, the team found out 22 functional failures and chose 13 ones, since some of them do not have great importance in terms of preventive or corrective interventions.

Through the relationship matrix the team crossed functional failure data with machine parts.

Table 4. Application of prioritization method.

Machine	MTBF	C	T	R
GP	5	5	4	4,6908
BUR	3	4	5	4,1996
LB	1	4	2	3,0529
ICT	3	3	3	3,0000
AOI	2	3	1	2,2721
RH	4	2	2	2,2192
AMF	5	1	3	2,0566
FCT	2	1	4	2,0370
MMC	1	1	1	1,0000
PR	1	1	1	1,0000

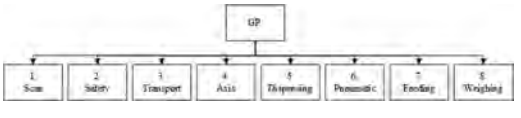


Figure 5. System structure.

The result was 12 parts with a strong relationship with system failures. Table 6 presents an example for 2 machine parts.

So, FMEA technique was applied to study the failure modes of these 12 parts and find out their causes. Using RPN rule, 25 critical failure modes was discovered. However, 3 of them are related to operation errors and raw material issues and, therefore, only 22 failure modes go to next step—the decision tree.

After running the decision tree 22 times, the team selected all the effective and possible tasks to each critical failure modes. Table 7 presents an excerpt of this step.

Calculating the costs of each task, Table 8 was obtained.

As example, for failure mode with code 1.1.1.1.1 TBM task was selected with a frequency of 7 hours and for 1.1.1.1.3 RTF was the only available option.

Table 5. Functions and functional failures analysis.

Subsystem	Function	Functional failures	Y/N
5. Dispensing	5.3 Mix materials A and B by 1:1	5.3.1 Do not mix materials A and B by 1:1	Y
		5.3.2 Do not mix materials A and B at all	Y

Table 6. Relationship between parts and functional failures.

	Functional failures				
Parts	1.1.1	1.1.2	1.1.3	2.1.1	2.1.2
Pump A	X	X	X		
Axis X				X	X

Table 7. Decision tree results.

Failure mode	Q1	Q2	Q3	Possible tasks
1.1.1.1.1	N	Y	N	TBM, RTF
1.1.1.1.3	N	N	N	RTF

Table 8. Economic evaluation of tasks.

Failure mode	Options	Annual cost	Frequency
1.1.1.1.1	TBM	8 526 €	7 hours
	RTF	11 392 €	
1.1.1.1.3	RTF	401 987 €	–

5 CONCLUSION

This paper presents a methodology for selecting and defining maintenance tasks for critical equipment. This procedure is based on RCM methodology and uses the AHP method to machine prioritization, the FMEA technique to study parts failure modes and an economical evaluation of maintenance tasks, the Weibull distribution and the Glasser Graph to define the most appropriate maintenance policy. In addition, some TPM concepts are taken into account by the methodology.

In short, the main differentiating factors regarding RCM are: the consideration of autonomous maintenance as a possible maintenance task to be selected and the economical evaluation of the tasks for this selection.

As results of application in the GP equipment of the case study company, the plant has achieved a 33% reduction in the number of failures, a 33% reduction in production losses, a MTBF improvement of 15% and an increase of machine yield of 1,8%. It is also expected to reduce total maintenance cost in mid-term.

Therefore, the application of the methodology makes the work of maintenance technicians more effective, since only appropriate and cost-effective preventive maintenance tasks are performed. It allows also the involvement of the operators in simple maintenance tasks as recommended by TPM. Since, the selection of maintenance tasks are oriented by costs, whenever adequate, the application of the methodology leads to a reduction in the overall maintenance cost.

ACKNOWLEDGMENT

This work has been supported by FCT—Fundação para a Ciência e Tecnologia in the scope of the PEst-UID/CEC/00319/2013.

REFERENCES

Ahuja, I.P.S. & Khamba, J.S., 2008. Total productive maintenance: literature review and directions. *International Journal of Quality & Reliability Management*, 25(7), pp. 709–756.

Backlund, F. & Akersten, P.A., 2003. RCM introduction—process and requirements management aspects. *Journal of Quality in Maintenance Engineering*, 9(3), pp. 250–264.

Bakri, A. et al., 2012. Boosting Lean Production via TPM. *Procedia—Social and Behavioral Sciences*, 65(65), pp. 485–491. Available at: www.sciencedirect.com.

Braglia, M., Castellano, D. & Frosolini, M., 2013. An integer linear programming approach to maintenance strategies selection. *International Journal of Quality & Reliability Management*, 30(9), pp. 991–1016.

Carretero, J. et al., 2003. Applying RCM in large scale systems : a case study with railway networks., 82, pp. 257–273.

Dehghanian, P. et al., 2013. A Comprehensive Scheme for Reliability Centered Maintenance in Power Distribution Systems—Part I: Methodology. *IEEE Transactions on Power Delivery*, 28(2), pp. 761–770.

Deshpande, V.S. & Modak, J.P., 2002. Application of RCM to a medium scale industry. *Reliability Engineering and System Safety*, 77(1), pp. 31–43.

Eisinger, S. & Rakowsky, U.K., 2001. Modeling of uncertainties in reliability centered maintenance—a probabilistic approach. *Reliability Engineering and System Safety*, 71, pp. 159–164.

Eti, M.C., Ogaji, S.O.T. & Probert, S.D., 2004. Implementing total productive maintenance in Nigerian manufacturing industries. *Applied Energy*, 79(4), pp. 385–401.

Garg, A. & Deshmukh, S.G., 2006. Maintenance management: literature review and directions. *Journal of Quality in Maintenance Engineering*, 12(3), pp. 205–238.

Jardine, A.K.S. & Tsang, A.H.C., 2013. *Maintenance, Replacement, and Reliability: Theory and Applications* 2nd ed., CRC Press.

Konecny, P.A. & Thun, J.H., 2011. Do it separately or simultaneously—An empirical analysis of a conjoint implementation of TQM and TPM on plant performance. *International Journal of Production Economics*, 133(2), pp. 496–507.

Kutucuoglu, K.Y. & Hamali, J., 2001. A framework for managing maintenance using performance measurement systems. *International Journal of Operations & Production Management*, 21(1), pp. 173–194.

Mokashi, A.J., Wang, J. & Vermar, A.K., 2002. A study of reliability-centred maintenance in maritime operations. *Marine Policy*, 26(5), pp. 325–335.

Moubray, J., 1997. *Reliability-centered Maintenance* 2nd ed., New York: Industrial Press, Inc.

Muchiri, P. et al., 2011. Development of maintenance function performance measurement framework and indicators. *International Journal of Production Economics*, 131(1), pp. 295–302. Available at: http://dx.doi.org/10.1016/j.ijpe.2010.04.039.

NP EN 13306, 2007. Terminologia da manutenção.

Rastegari, A. & Mobin, M., 2016. Maintenance Decision Making, Supported By Computerized Maintenance Management System. *IEEE*.

Siddiqui, A. & Ben-Daya, M., 2009. Reliability Centered Maintenance. In *Handbook of Maintenance Management and Engineering*. Springer, pp. 397–415. Available at: https://www.researchgate.net/publication/282253338.

Smith, A.M., 1993. *Reliability-Centered Maintenance*, USA: McGraw-Hill, Inc.

Soni, P.K., 2013. Total Productive Maintenance—An Implementation Experience., 2(5), pp. 263–267.

Waeyenbergh, G. & Pintelon, L., 2002. A framework for maintenance concept development. *International Journal of Production Economics*, 77(3), pp. 299–313.

Mathematical methods in reliability and safety

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A study of the relationship between sample size and the confidence level of MTTF for products with exponential failure distribution

Y. Wang, H. Cheng & D. Xu
NAA, Beijing, China

ABSTRACT: It is common in applications to ignore the influence of sample size and population size when evaluating confidence lower limit of Mean Time to Failure (MTTF) for products with exponential failure distribution. The conventional method assumes that failure times or lives of different samples are independent identically distributed random variables. This assumption holds reasonably when the population size is much larger than the sample size. However, it will induce substantial inaccuracies when the sample size is not negligible to the population. To this problem, the influence from the sample size to the confidence limit of MTTF is analyzed when the population is determined. First, the conventional method is reviewed. Then, the formulae based on hypergeometric distribution are derived to reflect the relationships among sample size, population size, confidence level and confidence limit. Finally, an example is presented to illustrate the suitability of the proposed method and to show the computation errors from the conventional method.

1 INTRODUCTION

Normally, the confidence lower limit of Mean Time to Failure (MTTF) for products with exponential failure distribution are verified by reliability verification test where some samples are picked up randomly (Jiang et al. 2012, Modarres et al. 1999) to conduct the test. Under most conditions, the number of samples is very small compared to the population. Then it is reasonable to assume that the failure times of different samples are independent and identically distributed (i.i.d.) random variables.

However, for complex system or product, the population size is not very large, usually. Hence, the sample size is not ignorable to the population in the reliability verification test plan (Li & Jiang 2006, Yan et al. 2014). The conventional method without considering the influence from sample size will result in computation error.

In this study, the conventional method ignoring the influence of sample size is reviewed. Then, a method considering the number of samples based on hypergeometric distribution (Ronald et al. 2009, Zhao & Mi 2015) is presented. Finally, the example is used to show the suitability of the proposed method. And the conventional method is used as well to show the computation error.

2 THE CONVENTIONAL METHOD

For a product that the failure time following exponential distribution, its failure rate (Tan et al. 2005) is λ. For mission time t, the reliability $R(t)$ is,

$$R(t) = \exp(-\lambda t) \tag{1}$$

where exp() denotes getting natural logarithm.

Meanwhile the expression for failure probability $F(t)$ is,

$$F(t) = 1 - \exp(-\lambda t) \tag{2}$$

Assume the number of population is infinity, and the sample size is n in the reliability verification test. For time t, the probability for r failures observed is,

$$P(X = r) = C_n^r F(t)^r R(t)^{n-r} \tag{3}$$

when n, t, r is specified in the reliability test plan, the formula to compute the confidence lower limit of $R(t)$ is,

$$1\text{-}\gamma = P(X \le r) = \sum_{i=0}^{r} C_n^i F(t)^i R(t)^{n-i} \tag{4}$$

where γ is confidence level.

The confidence lower limit of $R(t)$ can be derived from Eq. (4) through numerical method. Then we can use Eq. (1), where $\lambda = 1/\text{MTTF}$, to get the confidence lower limit of MTTF.

3 METHOD CONSIDERING POPULATION SIZE

Assume that the population size is N, the sample size for reliability test is n, the number of failures for acceptance is r in the test plan. Then, the probability for failure number r is,

$$\begin{aligned} P(X=r) &= \frac{C_M^r C_{N-M}^{n-r}}{C_N^n} \\ &= \frac{n!}{r!(n-r)!}\frac{M!}{(M-r)!}\frac{(N-M)!}{(N-M-n+r)!}\frac{(N-n)!}{N!} \\ &= C_n^r \frac{M(M-1)...(M-r+1)}{N^r} \times \\ &\quad \frac{(N-M)(N-M-1)...(N-M-n+r+1)}{N^{n-r}} \times \\ &\quad \frac{N^n}{N(N-1)...(N-n+1)} \end{aligned} \tag{5}$$

where M is the number of products that cannot pass the test in the population although it is unknown.

The probability for the failure number is equal to or less than r is,

$$P(X \le r) = \sum_{i=0}^{r} \frac{C_M^i C_{N-M}^{n-i}}{C_N^n} \tag{6}$$

By substituting Eq. (5) into Eq. (6) and denoting M/N with F, the confidence upper limit of failure probability, $F = 1 - R$ (R is the confidence lower limit of reliability), we can get,

$$\begin{aligned} 1-\gamma &= P(X \le r) \\ &= \sum_{i=0}^{r} C_n^r \left[F\left(F-\frac{1}{N}\right)...\left(F-\frac{r-1}{N}\right)\right] \times \\ &\quad \left[(1-F)\left(1-F-\frac{1}{N}\right)...\left(1-F-\frac{n-r-1}{N}\right)\right] \times \\ &\quad \frac{N^n}{N(N-1)...(N-n+1)} \end{aligned} \tag{7}$$

Base on the derivation of Eq. (7), the factor in the first bracket is senseless when r is zero. And a step further, when $n = 1$, $r = 0$, in the reliability test plan, Eq. (4) and Eq. (7) can be both rewritten as,

$$1-\gamma = P(X \le 0) = 1-F = R \tag{8}$$

4 EXAMPLE

For a product or system whose failure time follows exponential distribution, the population size N is 28. Assume that we need to perform reliability test to assess the confidence lower limit of MTTF. And the confidence level is 80%. To show the differences of the results of the conventional method and of the proposed method clearly, a series of test plans are presented and analyzed.

4.1 *Test plan 1*

In this reliability test, we set the parameters of the test plan as follows:

Population size $N = 28$.
Sample size of the test $n = 1$.
Acceptance number $r = 0$.
Accumulated test time $T = 4000$ h.
Confidence level $\gamma = 80\%$.

1. Conventional method

For conventional method, substitute the parameters above into Eq. (4), we could get the following equation,

$$\begin{aligned} &1-\gamma = P(X \le r) = \sum_{i=0}^{r} C_n^i F(t)^i R(t)^{n-i} \\ &\Rightarrow R(4000) = 1-0.8 = 0.2 \end{aligned} \tag{9}$$

That is,

$$R(4000) = \exp\left(-\frac{4000}{MTTF}\right) = 0.2 \tag{10}$$

From Eq. (10), we can derive the 80% confidence lower limit of MTTF,

$$\text{MTTF} = 2485.34 \text{ h} \tag{11}$$

2. Proposed method

For the proposed method considering the population size, substitute the parameters above into Eq. (7) or Eq. (8), we could get the following equation,

$$\begin{aligned} &1-\gamma = P(X \le 0) = 1\text{-}F = R \\ &\Rightarrow R(4000) = 0.2 \end{aligned} \tag{12}$$

As shown in Eq. (9) and in Eq. (12), for this test plan where sample size n is 1 and acceptance number r is zero, the proposed method is equivalent to the conventional method. This is because that there is statistical independence for the only one sample no matter what the population size is.

Hence, for the proposed method considering the population size, the assessed result of MTTF is 2485.34 as well.

4.2 *Test plan 2*

In this reliability test, we set the parameters of the test plan as follows:

Population size $N = 28$.
Sample size of the test $n = 2$.
Acceptance number $r = 0$.
Accumulated test time $T = 4000$ h (2000 h for each sample).
Confidence level $\gamma = 80\%$.

1. Conventional method
For conventional method, substitute the parameters above into Eq. (4), we could get the following equation,

$$\begin{aligned} &1-\gamma = P(X \le r) = \sum_{i=0}^{r} \mathrm{C}_n^i F(t)^i R(t)^{n-i} \\ &\Rightarrow R^2(2000) = 1-0.8 = 0.2 \end{aligned} \tag{13}$$

That is,

$$R^2(2000) = \left[\exp\left(-\frac{2000}{MTTF}\right)\right]^2 = 0.2 \tag{14}$$

From Eq. (14), we can derive the 80% confidence lower limit of MTTF,

$$\text{MTTF} = 2485.34 \text{ h} \tag{15}$$

The result is the same as the one from test plan 1.
2. Proposed method
For the proposed method considering the population size, substitute the parameters above into Eq. (7), we could get the following equation,

$$\begin{aligned} &1-\gamma = P(X \le r) \\ &= \left[(1-F)\left(1-F-\frac{1}{N}\right)\right]\left[\frac{N^2}{N(N-1)}\right] \\ &\Rightarrow \left[(1-F)\left(1-F-\frac{1}{28}\right)\right]\left[\frac{28^2}{28(28-1)}\right] = 0.2 \end{aligned} \tag{16}$$

That is,

$$\begin{aligned} &R(2000)\left[R(2000)-\frac{1}{28}\right]\frac{28}{27} = 0.2 \\ &\Rightarrow \exp\left(-\frac{2000}{MTTF}\right)\left[\exp\left(-\frac{2000}{MTTF}\right)-\frac{1}{28}\right]\frac{28}{27} = 0.2 \end{aligned} \tag{17}$$

From Eq. (17), we can derive the 80% confidence lower limit of MTTF,

$$\text{MTTF} = 2556.72 \text{ h} \tag{18}$$

The result is different from the one assessed from the conventional method. This is because that the proposed method has considered the interdependency of the two samples, but the conventional method has not.

And the computation relative error of the conventional method by take the result of the proposed method as standard is,

$$\varepsilon = \frac{2485.34-2556.72}{2556.72} = -2.8\% \tag{19}$$

4.3 *Test plan 3*

In this reliability test, we set the parameters of the test plan as follows:
Population size $N = 28$.
Sample size of the test $n = 10$.
Acceptance number $r = 0$.
Accumulated test time $T = 4000$ h (400 h for each sample).
Confidence level $\gamma = 80\%$.

1. Conventional method
For conventional method, substitute the parameters above into Eq. (4), we could get the following equation,

$$\begin{aligned} &1-\gamma = P(X \le r) = \sum_{i=0}^{r} \mathrm{C}_n^i F(t)^i R(t)^{n-i} \\ &\Rightarrow R^{10}(400) = 1-0.8 = 0.2 \end{aligned} \tag{20}$$

That is,

$$R^{10}(400) = \left[\exp\left(-\frac{400}{MTTF}\right)\right]^{10} = 0.2 \tag{21}$$

From Eq. (21), we can get the 80% confidence lower limit of MTTF,

$$\text{MTTF} = 2485.34 \text{ h} \tag{22}$$

The result is the same as the ones from test plan 1 and test plan 2.

2. Proposed method

For the proposed method considering the population size, substitute the parameters above into Eq. (7), we could get the following equation,

$$\begin{aligned} 1-\gamma &= P(X \le r) \\ &= \left[R\left(R-\frac{1}{N} \right) ... \left(R-\frac{9}{N} \right) \right] \\ &\left[\frac{N^{10}}{N(N-1)...(N-9)} \right] \\ &\Rightarrow \left[R\left(R-\frac{1}{28} \right) ... \left(R-\frac{9}{28} \right) \right] \times \\ &\frac{28^{10}}{28(28-1)...(28-9)} = 0.2 \end{aligned} \tag{23}$$

That is,

$$\begin{aligned} &R\left(R-\frac{1}{28} \right) ... \left(R-\frac{9}{28} \right) \times \\ &\frac{28^{10}}{28(28-1)...(28-9)} = 0.2 \\ &\Rightarrow \exp\left(-\frac{400}{MTTF} \right) \left[\exp\left(-\frac{400}{MTTF} \right) - \frac{1}{28} \right] ... \\ &\left[\exp\left(-\frac{400}{MTTF} \right) - \frac{9}{28} \right] \times \\ &\frac{28^{10}}{28(28-1)...(28-9)} = 0.2 \end{aligned} \tag{24}$$

From Eq. (24), we can derive the 80% confidence lower limit of MTTF,

$$\text{MTTF} = 3055.20 \text{ h} \tag{25}$$

And the computation relative error of the conventional method by take the result of the proposed method as standard is,

$$\varepsilon = \frac{2485.34 - 3055.20}{3055.20} = -18.7\% \tag{26}$$

From all the results of test plan 1, test plan 2 and test plan 3, we can see that for the conventional method the results always keep the same regardless of the sample size when the accumulated test time and acceptance failure number is determined. This is because that the conventional method has ignored the interdependency of different samples from the same population. And it could be verified easily that all the test plans in this example is equivalent to the conventional method, we will not present the verification process here.

5 CONCLUSIONS

In this study, we analyze the influence of population size to the computation accuracy of confidence lower limit of MTTF for products with exponential failure distribution in reliability verification test, and get the following conclusions:

1. A method considering the interdependency of different samples is proposed to compute the confidence lower limit of MTTF mainly based on hypergeometric distribution.
2. The conventional method will give out the same result regardless of how many samples are used in the reliability test when the accumulated test time and acceptance failure number are determined. This is because that the conventional method has ignored the interdependency of different samples from the one population.
3. According to the proposed method considering the interdependency of different samples and the results in the example, the confidence lower limit of MTTF will become greater with the increase of sample size when population size, accumulated test time, acceptance failure number are determined in the reliability test plan.

REFERENCES

Jiang, T. *et al*. 2012. *Reliability and Life Test*. National defense industry press, 196–200.

Li, G. & Jiang, T. 2006. Selection and Analysis of Test Scheme Parameters for Time Curtailed Reliability Qualification Test. *Acta Aeronautica ET Astronautica Sinica*, 27(2), 272–274.

Modarres, M., Kaminskiy, M. & Krivtsov, V. 1999. *Reliability Engineering and Risk Analysis*. Boca Raton: Taylor & Francis Group, 127–144.

Ronald, W. *et al.* 2009. *Probability & Statistics for Engineers & Scientists (8th edition)*. Beijing: China Machine Press, 113–115.

Tan, F., Jiang, Z. & Bai, T. 2005. A statistic Analysis of Early Failure Rates of Large Size Generating Units. *Journal of Shanghai Jiaotong University*, 39(12), 2093–2096.

Yan, S., Liu, X. & Li J. 2014. Reliability Test for Qualification Method of Shipborne Gun Fire Control System Based on Type-I. *Fire Control & Command Control*, 39(5), 172–175.

Zhao, Y. & Mi, X. 2015. Missile Batch Sampling Method Based on Reliability and Accuracy. *Journal of Sichuan Ordnance*, 36(8), 29–31.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Failure Mode Effects & Criticality Analysis (FMECA) using Bayesian Dirichlet-multinomial conjugate pair

W. Baun
Pratt & Whitney, East Hartford, CT, USA

ABSTRACT: Failure Mode, Effects & Criticality Analysis (FMECA) is a widely-used tool for system safety and reliability evaluations; one popular approach is outlined by MIL-STD-1629A. MIL-STD-1629A Criticality Analysis requires estimates of the Failure Mode Ratio (α_i) and the Failure Effect Probabilities (β_i) for each failure mode. To maximize impact on the design, FMECAs are initiated early in the Product Development Process (PDP), before reliability data is available for the new product. Typically, initial criticality estimates are derived via a combination of failure mode/effect data from similar legacy products and from engineering judgment. Later in the PDP, actual data on the new product becomes available. Such a situation is ideal for employing Bayesian methods. The data from which the Criticality parameters are estimated is Multinomial. The Dirichlet distribution is chosen to represent prior uncertainty in the Criticality parameters, thus taking advantage of the Bayesian conjugacy property for the Dirichlet-Multinomial pairing. This paper describes the application of Bayesian techniques to FMECA Criticality analysis, the structure of a Bayes-enabled FMECA, briefly outlines some expert elicitation approaches for developing prior distributions for α_i and β_i, and demonstrates the use of evidence in the form of field data to update those prior estimates.

1 CRITICALITY ANALYSIS INTRODUCTION

Failure Mode, Effects and Criticality Analysis (FMECA) is an inductive tool used to analyze the effects, and quantify the risks of failure modes within a larger system. FMECA is an extension of Failure Modes and Effects Analysis (FMEA), which adds a quantitative Criticality assessment to the basic FMEA process. The outputs of the criticality analysis enable risk mitigation activities to be undertaken in a systematic, prioritized fashion.

There are different approaches to conducting criticality analysis. This paper focuses on the FMECA approach outlined in MIL-STD-1629A (MIL 1980), which is also covered in EN 60812:2006. (EN 2006) It is worth noting that the methods outlined in this paper are not applicable to the SAE J1739 (SAE 2009) FMEA approach; SAE FMEA focuses on a very different aspect of design risk, and uses a qualitative risk ranking technique. Criticality analysis as performed in MIL-STD-1629 requires estimation of the following factors for every component-failure mode combination in the system being analyzed:

α: Failure mode ratio
β_i: Failure effects probability factors
λ: Component base failure rate

The failure mode ratio, α, is the probability that a particular component will fail by a certain failure mode. The sum of all failure mode ratios for any specific component must sum to 100%. Where a component has only one failure mode, the failure mode ratio would equal 100%. If a component has three equally-likely failure modes, then each failure mode would have a failure mode ratio equal to 33.3%. If one failure mode is more prevalent than others, then that failure mode would have the largest share of the 100% total. The number of failure modes is not limited, being a function of the various ways in which the particular component can fail.

The failure effects probability factors, β_i, are the probability that a particular component-failure mode will result in effects of a certain severity level. In the case of the MIL-STD-1629A FMECA approach as applied to the aerospace industry, the effects severity level is broken into four distinct categories (CAT):

In this categorization system, there would thus be four different β factors for each failure mode, β_1, β_2, β_3 and β_4, representing the probabilities that a failure by that component-failure mode combination results in CAT I, II, III, or IV effects, respectively. The sum of the β factors for a given component-failure mode combination must equal 100%. If a particular failure mode can only result in effects of one category, then the failure mode

Table 1. MIL-STD-1629A severity categorizations for aerospace application.

Category	Description	Details
I	Catastrophic	A failure which can cause death or loss of aircraft
II	Critical	A failure which can cause major system damage, mission abort
III	Marginal	A failure which can cause minor system damage, mission delay, loss of availability
IV	Negligible	A failure which will not cause system damage, mission delay or loss of availability

effect probability for that category would be 100%; all other categories would be 0%. Generally, a failure mode will have a most likely set of effects; this category would thus have the highest proportion of the 100% total for the four categories. If it is possible for other effect levels to occur, those categories should be given a non-zero rating which corresponds to the best estimate of their probability of occurrence.

While other effects severity rating systems having fewer or more categories are possible, the remainder of this paper will assume that four severity categories exist, as given above. The methods outlined in this paper are equally applicable to any number of severity effects categories.

The component base failure rate, λ, is the overall failure rate for a particular component, for all failure modes, and all outcomes (effects). Typically in FMECA (and other reliability analysis techniques), λ is assumed to be constant, with failure times exponentially-distributed. Under that assumption, the base failure rate would be calculated as the total number of failures of that component observed in the fleet (for all failure modes and effects) divided by the total operating hours for that component across the entire fleet. Failure rate in the aerospace application consistent with Table 1 has units of failures per engine flight hour.

In the MIL-STD-1629A approach to FMECA, Criticality is calculated as the product of the three factors above, and a fourth factor, time, t:

$$C_i = \lambda \alpha \beta_i t \quad (1)$$

For the remainder of this paper, the time factor will be omitted, under the assumption that the Criticality factors are calculated on a per-hour basis. Because there are four severity categories, there exist four criticality numbers for each component-failure mode, one for each category of effect severity.

The units of criticality are also in failures per hour, and represent the failure rate per hour for a given component-failure mode combination producing effects of a particular severity category.

Ideally, estimation of the factors used in criticality analysis will come from an extensive data set of actual field experience. The failure rate estimate for a given component comes from the total number of failures divided by total operating time, as discussed above. The failure mode ratio for a component-failure mode combination would be the number of times a component has failed by that particular failure mode divided by the total number of failures for that component for all failure modes. For a given failure mode, the failure effect probabilities would come from the number of times that failure mode has produced effects of that severity level divided by the total number of failures by that mode.

In practice, some or all of these factors will need to be estimated from other sources. FMECA is most effective as a risk prioritization and mitigation tool when applied early in the product design process. However, this is typically long before any actual field failure data from which to calculate the criticality factors would be available for the new system being designed. In this situation, the FMECA analyst must turn to similar legacy products as a source of such data. However, there may be components in the FMECA which have never experienced a particular failure mode in the legacy fleet. Furthermore, for high severity failure modes (CAT I), it is typically the case (fortunately) that they are rare events, thus there are few to no CAT I failures in the legacy data set from which to calculate a probability. Regardless, estimates must be made, in order to utilize the FMECA process to its full potential. In these situations, where limited / no field experience is available, estimates of criticality factors must be made by expert opinion.

The use of Expert opinion to estimate unknown quantities of interest is widely-used in reliability engineering applications. (Bedford et al. 2006, Hodge et al. 2001, Yadov et al. 2003) Bayes Rule provides a framework for starting with such subjective estimates of parameter(s) of interest, and then updating those estimates with actual data as it becomes available. The remainder of this paper will discuss a method for the application of Bayes Rule to the estimation of FMECA Criticality parameters α and β. Section 2 will provide a brief overview of Bayes Rule. Section 3 will outline the specific choice of distributions for modeling the unknown quantities of interest, α and β for the FMECA application. Section 4 will discuss the process of expert elicitation as it applies in this situation. Section 5 will discuss specific concerns related to prior distribution strength. Section 6 will provide

examples of the Bayesian FMECA updating process, showing how evidence can be combined with the prior estimates via Bayes Rule. Section 7 will discuss conclusions and future work.

2 OVERVIEW OF BAYES RULE

Bayes Rule is a formal mathematical method for updating one's state of belief about unknown parameter(s) of interest (POI) as new evidence is obtained. It enables initial, often subjective, estimates about those POI to be combined with actual data to produce those updated estimates. A classical formula for Bayes rule is:

$$P_f(A|E) = \frac{L(E|A) * P_0(A)}{\int_A L(E|A) * P_0(A) dA} \quad (2)$$

where A is the parameter of interest (POI), the unknown parameter or factor being modeled; $P_0(A)$ is the prior distribution of A (the "Prior"), which is the uncertain estimate of the value of A developed before any actual evidence is obtained, modeled as a distribution to represent the initial state of uncertainty regarding the value of A; E is the evidence; $P_f(A|E)$ is the posterior distribution of A (the "Posterior"), which is the updated estimate of A, after gaining new information in the form of evidence; $L(E|A)$ is the likelihood of the evidence, given parameter A. The denominator is a normalizing factor, sometimes referred to as the "total probability of the evidence", and is the summation of the likelihood of the evidence over all possible values of the parameter, A.

The prior distribution is the mathematical representation of the initial state of knowledge about the POI. Depending upon that state of knowledge, the prior can be diffuse, or concentrated. The prior distribution can be modeled using a wide variety of mathematical distributions to represent uncertainty. Where the POI can take on any value in a range, the initial state of knowledge is represented by a continuous Prior distribution, and the summation term in the denominator is an integral. Typical distributions employed to model continuous Priors include the uniform distribution, lognormal, gamma, beta, as well as many others. The integral is evaluated over the full set of unknown POI, A, which are the parameters of the distribution used to model the prior. If there is more than one parameter, then the integral becomes a double (or higher) integral. As such, the Bayes equation can quickly become difficult to solve. In many cases, a closed-form solution cannot be found, and numerical methods are required.

Fortunately, there are a special set of Prior Distributions and Likelihood Functions which make the calculation process relatively easy. These sets are called Conjugate Pairs. For a certain form of evidence expected, a specific likelihood function is required. Given a particular type of evidence (and thus likelihood function), the choice of a certain form of Prior distribution will make the calculation of the Posterior Distribution via Bayes Rule simple, and obviate the need to evaluate any integrals.

For example, if one is interested in estimating the failure rate, λ, from a constant failure rate process, then the evidence by which failure rate will be estimated will be of the form n failures in observation time T, also known as Poisson data. Given evidence of this form (Poisson) and the resultant likelihood function, the choice of a Gamma distribution with parameters α and β to model the Prior distribution of failure rate (the initial estimate of the unknown POI, λ) results in a Posterior distribution which is also Gamma, and which has parameters α+n, β+T. Other conjugate pairs exist for widely-modeled situations (Fink 2017), including the Beta-Binomial conjugate pair for modeling situations such as "Probability of Failure on Demand" where the value can range between 0 and 1, and the evidence will be in the form of k failures in n trials (Bernoulli process).

3 DIRICHLET-MULTINOMIAL CONJUGATE PAIR

Another Bayesian conjugate pair is the Dirichlet-Multinomial pair (Fink 2017). This pairing is used in situations where the evidence will be multinomial in nature; the prior distribution of the unknown POI will be modeled as a Dirichlet distribution. Multinomial data is applicable in situations where the data can fall into one of n categories (with binomial being the specific case where n = 2). It is easy to see that failure mode and failure effects data are both cases of multinomial datasets—failure modes can fall into one of n categories for a given component, where n = the number of failure modes for that component. Similarly, under the MIL-STD-1629A approach to FMECA, failure effects data can fall into one of four severity categories. Therefore, because the data for estimating FMECA Criticality factors will be of Multinomial form, the Dirichlet distribution will be employed as the choice of priors for the FMECA criticality analysis POI in order to take advantage of the mathematical benefits of the conjugate pairing.

As discussed previously, the criticality calculation for the MIL-STD-1629A FMECA has two parameter sets (POI) which must be estimated—the failure

mode ratio, α, and the failure effects probabilities, $\beta_1 - \beta_4$. For a given component, the failure mode ratio is the probability that the component fails by that failure mode (versus other possible failure modes) when that component fails. Thus, the evidence which will be used to update the prior estimates of failure mode ratios using Bayes rule will be multinomial in nature. Table 2 below shows an example of multinomial data for failure mode ratio for one component.

Similarly, the failure effects probabilities for a given component-failure mode combination are the probabilities that a particular failure mode will result in effects in each of the four different categories of severity from Table 1. Taking the data for failure mode C from Table 2, the failure effects probability data is also multinomial, as shown in Table 3.

The multinomial probability density function for the failure mode ratios, α_i, is as follows:

$$f\left(x_1, x_2, \ldots, x_k \middle| n, \alpha_1, \alpha_2, \ldots, \alpha_k\right) = \frac{n!}{x_1! \mathrm{K} x_k!} \alpha_1^{x_1} \ldots \alpha_k^{x_k} \tag{3}$$

where k is the number of failure modes for that component, x_i are the number of failures by each of the modes $x_1 - x_k$, n is the total number of failures observed for that component.

Similarly, the multinomial probability density function for the failure effect probabilities, $\beta_1 - \beta_4$ is:

$$f\left(x_1, x_2, \ldots, x_4 \middle| n, \beta_1, \beta_2, \beta_3, \beta_4\right) = \frac{n!}{x_1! \ldots x_4!} \beta_1^{x_1} \ldots \beta_4^{x_4} \tag{4}$$

Table 2. Example failure mode data and calculated failure mode ratios.

Failure mode	# Failures	Failure mode ratio
A	2	20%
B	1	10%
C	7	70%
All	10	100%

Table 3. Example failure effect data and calculated failure effect probability factors for failure mode C from Table 2.

Category	# Failures	Failure effect probability
I	1	14.3%
II	1	14.3%
III	2	28.6%
IV	3	42.9%
All	7	100%

where x_i are the number of failures for that mode that resulted in effects of severity level i, n is the total number of failures for that component-failure mode.

Because it is the conjugate prior for this form of evidence, the Dirichlet distribution is used to model the Prior state of knowledge of the criticality parameters α and β.

The Dirichlet distribution for the failure mode ratio parameters $\alpha_1 - \alpha_k$ for a given failure mode is written as follows:

$$f\left(\alpha_1, \alpha_2, \ldots, \alpha_k \middle| \rho_1, \rho_2, \mathrm{K}, \rho_k\right) = \left[\frac{\Gamma\left(\sum_{i=1}^{k} \rho_i\right)}{\prod_{i=1}^{k} \Gamma\left(\rho_i\right)}\right] \prod_{i=1}^{k} \alpha_i^{\rho_i - 1} \tag{5}$$

where $\alpha_1 - \alpha_k$ are the POI – the failure mode ratios for the k failure modes of that component, $\rho_1 - \rho_k$ are the parameters of the Dirichlet distribution.

The values $\alpha_1 - \alpha_k$ are not known exactly, but rather are uncertain quantities, with uncertainty represented by the Dirichlet distribution. The expected value of each of the failure mode ratios, α_i, for the prior distribution are simply the ratio of the parameter ρ_i for failure mode i to the sum of all parameters $\rho_1 - \rho_k$ (Agresti 2017):

$$E\left(\alpha_i\right) = \frac{\rho_i}{\sum_{i=1}^{k} \rho_i} \tag{6}$$

As discussed previously, evidence about the failure mode ratios will come in the form of the number of failures which were caused by a particular failure mode out of a total number of failures for all failure modes—Multinomial data. Because of the conjugacy property of the Multinomial data with the Dirichlet prior distribution, the Posterior distribution of the failure mode ratios $\alpha_1 - \alpha_k$ is also a Dirichlet distrbution, with the following characteristics:

$$\rho_{i_f} = \rho_i + x_i \tag{7}$$

where ρ_i are the initial Dirichlet parameters; x_i are the number of failures which occurred by that failure mode; and ρ_{if} are the final Dirichlet parameters.

The expected value of each of the failure mode ratios, α_{if}, for the posterior distribution are simply the ratio of the parameter ρ_{if} for failure mode i to the sum of all parameters $\rho_{1f} - \rho_{kf}$ (Agresti 2017):

$$E\left(\alpha_{i_f}\right) = \frac{\rho_{i_f}}{\sum_{i=1}^{k} \rho_{i_f}} \tag{8}$$

Unlike the failure mode ratio which can have any number of parameters, one for each failure mode, the failure effects probability can have only four in the MIL-STD-1629A approach to FMECA, since there are four categories of failure effect severity. The Dirichlet distribution for the failure effect probability parameters $\beta_1 - \beta_4$ for a given component-failure mode is written as follows:

$$f\left(\beta_1,\beta_2,\beta_3,\beta_4\middle|\delta_1,\delta_2,\delta_3,\delta_4\right)=\left[\frac{\Gamma\left(\sum_{i=1}^{4}\delta_i\right)}{\prod_{i=1}^{4}\Gamma\left(\delta_i\right)}\right]\prod_{i=1}^{4}\beta_i^{\delta_i-1} \tag{9}$$

where $\beta_1 - \beta_4$ are the POI – the failure effect probability factors for that failure mode; and $\delta_1 - \delta_4$ are the parameters of the Dirichlet distribution.

As with failure mode ratio parameters, the expected values of the prior failure effect probability factors are (Agresti 2017):

$$E\left(\beta_i\right)=\frac{\delta_i}{\sum_{i=1}^{4}\delta_i} \tag{10}$$

Similarly, the parameters of the Posterior Dirichlet distribution are the initial parameters, δ_i plus the number of failures which occurred by that failure mode resulting in effects having severity of that category, x_i.

$$\delta_{i_f}=\delta_i+x_i \tag{11}$$

The expected value of each of the failure effect probability factors, β_{if}, for the posterior distribution are simply the ratio of the parameter δ_{if} for failure effect i to the sum of all four parameters $\delta_{1f} - \delta_{4f}$ (Agresti 2017):

$$E\left(\beta_{i_f}\right)=\frac{\delta_{i_f}}{\sum_{i=1}^{4}\delta_{i_f}} \tag{12}$$

4 EXPERT ELICITATION OF PRIOR CRITICALITY PARAMETERS

For a Criticality analysis of a new product which is in the development process, where no testing (or failures) have yet occurred, the Dirichlet Prior distributions of α_i and β_i are the initial estimates of those parameters. By setting up the FMECA structure with a Bayesian framework, as multinomial data are gathered later in the design and verification/validation processes, the criticality estimates may be updated. This allows for updated FMECA-based risk assessments to be made as additional evidence is accrued over time.

Because of the stage of the process where the prior estimates will be needed, development of the Dirichlet prior distribution may require the elicitation of information from experts. A detailed review of the topic of expert elicitation is beyond the scope of this paper. Some basics as they apply to this situation will be briefly discussed in this section, but the remainder of the paper will assume any expert-elicited estimates utilized sound methods in the elicitation process.

In order to develop the Priors for α_i and β_i, experts must provide their estimates of each of the values, $\alpha_1 - \alpha_k$ and $\beta_I - \beta_{IV}$. Once values are obtained from experts, assuming more than one expert, their answers (which may differ) must be aggregated into a single set of estimates. There are a couple of different approaches which can be taken to this aggregation process, often categorized as either mathematical or behavioral approaches (Clemen et al. 1999). At a high level, the mathematical approach allows each expert to independently provide their assessments, and then the analyst combines them via some form of averaging. Weighted averages can be used to reflect differing degrees of expertise. The advantage of this approach can be that it requires less time in formal joint meetings with experts, whose time may be of limited availability. The alternative behavioral approach requires the experts to develop to a jointly-agreed upon estimate through discussion. This method has the advantage of allowing all of the experts to share their viewpoints and reasoning for their initial positions. As the experts share their unique knowledge and perspectives with each other, they may collectively arrive at an estimate which is more comprehensive and representative than if they had each provided independent assessments. The drawback to this method is the time it may take the experts to reach that consensus, and the difficulty in getting all experts together for the necessary consensus-reaching discussions.

One particular nuance in developing expert estimates of FMECA Criticality parameters is that the parameters being estimated are dependent. For example, when estimating the failure mode ratios for a component with k distinct failure modes, only k – 1 estimates can be provided; the last estimate is derived by subtracting the sum of the other estimates from 1, since the sum of the failure mode ratios must equal 100%. This dependency can produce some challenges to the elicitation process. Some techniques for eliciting estimates

for dependent parameters include overfitting as described in (Zapata-Vazquez et al. 2012).

5 PRIOR DISTRIBUTION STRENGTH

Once a set of criticality factor estimates have been developed, the experts must determine whether to apply those estimates as strong or weak data. The prior estimates of the Dirichlet parameters, ρ_i and δ_i, can be assigned any nominal value, as long as their relative values are consistent with the experts' opinions. For instance, assuming three failure modes, the following two sets of failure mode ratio prior distribution parameter estimates are all the same from a relative perspective. Both sets of prior distribution parameters imply the assumption that the expected value of Mode 1's failure mode ratio, α_1, is 10%, Mode 2's expected value is 65% and Mode 3's expected value is 25%. However, in the first case, the parameters are set as decimal values representing the percentage, whereas in the second case, the parameters are set as the integer values representing the percentage.

These two sets of prior parameters, both representing the same expected values for the failure mode ratios produce prior distributions with very different strengths. The sum of the ρ_i values in each case can be thought of as a quasi-number of failures previously observed. The stronger prior in this case is approximately the same as having observed 100 past failure events from which initial estimates of failure mode ratio have been made, whereas the weak prior is approximately equivalent to having observed only 1 failure. As actual field data is gathered and combined with these prior estimates to update predictions of failure mode ratios, the stronger prior will require significantly more data to overcome the initial estimates, if those initial estimates are proven to be incorrect by the field data.

For example, assume that actual field data is gathered on the failure modes in Table 4. Assume ten failures have occurred, and all 10 were due to Mode 1. This is not at all what was expected based on the prior estimates above (in both cases, the prior parameters would have predicted only 1 out of 10 failure would be via Mode 1). Table 5 below compares the Posterior estimates of the expected value for the failure mode ratios, using Bayes rule to combine the prior parameters with the field data. Note that the results obtained using the weak prior shifted significantly, and "feel" more correct in light of the field data. Conversely, the posterior estimates obtained using the strong prior have only marginally changed from their prior values, despite the significant disagreement between the prior estimates and the field results.

Table 4. Examples of two Dirichlet parameter sets with differing strengths for a given component.

Failure Mode	ρ_i – weak	ρ_i – strong
1	0.10	10
2	0.65	65
3	0.25	25

Table 5. Posterior estimates of failure mode ratio expected values as a function of prior strength.

Failure mode	α_i – weak	α_i – strong
1	92%	18%
2	6%	59%
3	2%	23%

The key takeaway here is to be cognizant of the fact that the choice of absolute values for the Dirichlet parameters, ρ_i, determines the strength of the prior distribution. The absolute values should be determined based upon the strength of the prior information from which those parameter estimates were developed. Pre-fitting of hypothetical field data can help the analysis team to see how such data would affect posterior estimates, and help the team to select a set of prior values which best represent the team's true state of prior knowledge.

Note that this same behavior is true for the prior parameters for the failure effect probability factors, β_1 – β_4, too—the prior distribution strength is positively correlated with the sum of the failure effect probability estimates for that failure mode.

6 APPLICATION OF EVIDENCE TO UPDATE PRIOR CRITICALITY FACTOR ESTIMATES

Data to enable Bayesian update of the prior FMECA criticality parameters will be in the form of a number of failure events categorized by failure modes and their resultant failure effects. The utility of the proposed FMECA with Bayesian structure will be shown via an example.

Assume experts have provided prior estimates of the Failure Mode Ratios and Failure Effect Probabilities for a particular component. Those prior estimates have been provided in the form of Dirichlet distribution parameters, ρ_i and δ_i, and their associated expected values, α_i and β_i, calculated from those parameters in Tables 6–8. Note that in this example, the expert-estimated parameters result in weak prior distributions for both Failure Mode Ratios and Failure Effects Probabilities, presumably reflecting their lack of certainty about those prior estimates. Such weak priors will be quickly influenced by any new evidence which is gathered.

Assume that later in the development process for the component/system in question, the following failure information is obtained:

Combining the prior Dirichlet distribution information with the new evidence following Bayes Rule as outlined in Section 3 results in the updated (Posterior) estimates of the Failure Mode Ratio and Failure Effect Probability Dirichlet parameters and expected values for this component in Tables 10–12.

Table 6. Example prior failure mode ratio estimates.

	Prior failure mode ratio	Estimates
Failure mode	Dirichlet parameters, ρ_i	Expected values, α_i
1	0.05	5%
2	0.30	30%
3	0.65	65%

Table 7. Example prior failure effect probability factor dirichlet parameter estimates.

	Prior failure effect probability factor			
	Dirichlet parameter estimates, δ_i			
Failure mode	CAT I	CAT II	CAT III	CAT IV
1	0.00	0.20	0.50	0.30
2	0.00	0.00	0.80	0.20
3	0.01	0.01	0.75	0.23

Table 8. Example prior failure effect probability factor expected value estimates.

	Prior failure effect probability factor			
	Expected value estimates, β_i			
Failure mode	CAT I	CAT II	CAT III	CAT IV
1	0%	20%	50%	30%
2	0%	0%	80%	20%
3	1%	1%	75%	23%

Table 9. Example failure mode and effects data.

		# Failures by failure mode/effect category			
Failure mode	#Failures by mode	I	II	III	IV
1	0	0	0	0	0
2	2	0	0	2	2
3	1	0	1	0	0

Table 10. Posterior failure mode ratio estimates.

	Posterior failure mode ratio estimates	
Failure mode	Dirichlet parameters, ρ_{if}	Expected values, α_{if}
1	0.05	1.3%
2	2.30	57.5%
3	1.65	41.3%

Table 11. Posterior failure effect probability factor dirichlet parameter estimates.

	Posterior failure effect probability factor			
	Dirichlet parameter estimates, δ_{if}			
Failure mode	CAT I	CAT II	CAT III	CAT IV
1	0.00	0.20	0.50	0.30
2	0.00	0.00	2.80	0.20
3	0.01	1.01	0.75	0.23

Table 12. Posterior failure effect probability factor expected value estimates.

	Posterior failure effect probability factor			
	Expected value estimates, β_{if}			
Failure mode	CAT I	CAT II	CAT III	CAT IV
1	0%	20%	50%	30%
2	0%	0%	93.3%	6.7%
3	0.5%	50.5%	37.5%	11.5%

The changes for the Failure Mode Ratio expected values are shown graphically in Figure 1. Based on the new evidence, one would conclude that the probability of occurrence of failure modes 1 and 3 are less than originally estimated (despite having observed one failure via failure mode 3), and the probability of occurrence of failure mode 2 has nearly doubled.

The changes for the Failure Effect Probability Factor expected values for Failure Mode 3 are shown graphically in Figure 2. Based on the new evidence, one would conclude that the probability of failure mode 3 producing CAT II severity effects is significantly higher than was originally estimated.

A key item to note in both examples above is that all categories' predictions change in response to new evidence, even if that new evidence didn't explicitly include events in that particular category.

This is one of the key benefits of the Bayesian approach to FMECA updating—new evidence showing that a failure occurred in one of the categories updated all of the other categories' predictions, too. For example, in the failure effects probability example, the single piece of new evidence (occurrence of a single event with CAT II failure effects) also informed the other three categories; it was a failure that did NOT result in effects from those other severity categories, and thus their predicted values shifted downwards by varying degrees.

A final benefit of approaching FMECAs in a Bayesian fashion is that it enables sensitivity and uncertainty analysis techniques to be applied to the results. Because the criticality factors are modeled in a Bayesian framework as uncertain parameters of interest, one can report FMECA criticality factors calculated based on the expected value of those parameters, or one can report criticality factors calculated based on worst-case values, such as upper confidence limit values. The confidence limits for any Dirichlet-distributed parameter are found by calculating the marginal Beta distribution for that parameter. (Farrow 2017) In the case of the uncertain estimate of the failure mode ratio for failure mode 3, α_3, from the previous examples,

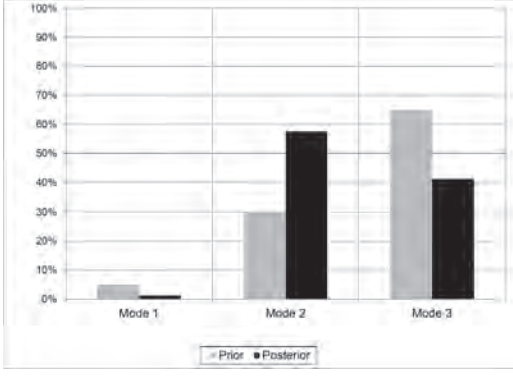


Figure 1. Comparison of prior and posterior failure mode ratio expected values.

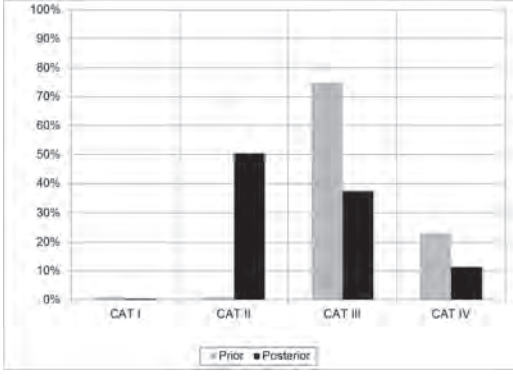


Figure 2. Comparison of prior and posterior failure effect probability factor estimates.

Table 13. Comparison of criticality estimates based on use of expected values versus 90% UCL values for failure mode ratio and failure effect probability estimates for failure mode 3, failure effect CAT II.

Parameter	Expected Value	90% UCL
Failure Mode Ratio, Mode 3, α3	41.3%	94.1%
Cat II Failure Effect Probability, β3	50.5%	97.7%
Component Base Failure Rate, λ	6.2E-8	6.2E-8
Cat II Criticality (failures per hour)	1.3E-8	5.7E-8

$$\alpha_3 \sim Beta\left(\rho_3, \sum_i \rho_i - \rho_3\right) \tag{13}$$

The same holds true for the failure effect probability Dirichlet parameters. Table 13 shows a comparison of the criticality values calculated for Failure Mode 3, Failure Effect Category II based on the data used in the examples above. Note that the table assumes no uncertainty in the base failure rate, λ, and thus the same failure rate is used for both the nominal criticality value calculation and the upper confidence limit criticality value calculation. (In practice, the base failure rate could also be treated as an uncertain POI, and modeled in a Bayesian framework). The upper confidence bound estimate of the criticality value for this failure mode effect combination is more than 4X higher than the nominal estimate. The Bayesian structure of the FMECA enables such uncertainty analysis, which can serve as a means of prioritizing future data-gathering efforts, or can allow for uncertainty-based risk evaluations. For failure modes with the highest severity effects, risk prioritization efforts can be based upon upper confidence limit values of the Criticality estimates.

7 CONCLUSIONS AND FUTURE WORK

FMECA is a widely-used tool in a broad range of industries for evaluating and quantifying the risks associated with a design. Because FMECA is most effective when applied early in the product design process, it is frequently the case that the criticality analysis must proceed before any actual data from

which to conduct that analysis is available on the system in question. The methods outlined herein provide a framework for allowing the synthesis of expert-elicited prior estimates of criticality factors with later-arriving direct evidence via Bayes Rule to produce updated estimates of those factors. This approach imparts a broad range of analytical benefits, including allowing uncertainty and sensitivity analyses to be conducted on the quantitative outputs of the FMECA.

Future work in this area will focus on the development of FMECA-specific techniques for eliciting expert estimates of criticality factors α and β, notably in dealing with the dependency of those factors, and in addressing the difficulties in estimating rare event probabilities.

REFERENCES

Agresti, A. & Hitchcock, D.B. Bayesian Inference for Categorical Data Analysis. Accessed 8 June 2017 www.stat.ufl.edu/bayes.pdf.

Bedford T., Quigley J., & Walls, L. 2006. Expert Elicitation for Reliable System Design. *Statistical Science Vol. 21, No. 4, pp. 428–450.*

Clemen, R.T., Winkler, R.L. 1999. Combining Probability Distributions from Experts in Risk Analysis. *Risk Analysis Vol. 19, No. 2, pp. 187–203.*

EN 60812 – Analysis Technique for System Reliability, Procedure for Failure Mode and Effects Analysis (FMEA). January 2006.

Farrow M. MAS3301 Bayesian Statistics. Newcastle University. Accessed 8 June 2017 www.mas.ncl.ac.uk/teaching/mas3301/week11.pdf.

Fink, D. A Compendium of Conjugate Priors. Environmental Statistics Group, Department of Biology, Montana State University, Bozeman, MT 59717. Accessed 8 June 2017. www.johndcook.com/CompendiumOfConjugatePriors.pdf.

Hodge, R., Evans, M., Marshall, J., Quigley, J., Walls, L. 2001. Eliciting Engineering Knowledge About Reliability During Design – Lessons Learnt from Implementation. *Quality and Reliability Engineering International 17, pp. 169–179.*

MIL-STD-1629A – Procedures for Performing a Failure Mode, Effects and Criticality Analysis, 24 November 1980.

SAE J1739 – (R) Potential Failure Mode and Effects Analysis in Design (Design FMEA), Potential Failure Mode and Effects Analysis in Manufacturing and Assembly Processes (Process FMEA), January 2009.

Yadav, O.P., Singh, N., Goel, P.S., Itabashi-Campbell, R. 2003. A Framework for Reliability Prediction During Product Development Process Incorporating Engineering Judgments, *Quality Engineering Vol. 15, No. 4, pp. 649–662.*

Zapata-Vazquez, R.E., O'Hagan, A., Bastos, L.S. 2012. Eliciting Expert Judgments About a Set of Proportions. *Journal of Applied Statistics Vol. 41, 2014, Issue 9, pp. 1919-1933.*

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Bringing formal methods on the rail: On automatic verifying railroad interlockings from railML models

Tim Gonschorek, Ludwig Bedau & Frank Ortmeier
Chair of Software Engineering, Faculty of Computer Science, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany

ABSTRACT: The theoretic foundations for formally verifying railway interlocking systems have already been studied extensively. There exist a lot of work covering the application of methodologies like model checking in this context. However, some design faults still remain undetected until final on-track evaluation of the system. This is strongly related to missing automation solutions for real-world models and standards as well as the high theoretical expertise required. There exist many well-developed tools each requiring different modeling formalisms and focusing on a different question/scenario. Without specific experience in formal system modeling, it is extremely complicated to model such complex systems. In this paper, we present a methodology for the automatic model generation and verification of railway interlockings in a tool-independent(!) way. Therefore, we define a generic template set of atomic track elements and safety properties in a formal modeling language applicable with precise semantics. This generic template enables us to verify the structure of any given track layout. The already existing tool support of VECS allows to automatically translate these specifications into various model checkers for verification. More important, we present a robust transformation of the upcoming data exchange format for railway interlocking systems railML into the presented specification template. As a consequence, this approach really may help to bridge the gap between formal methods and system design in railway interlockings. We evaluate this approach on a real-world case studies train station of Brain l'Alleud. We also show the tool-independent modeling by automatically translating the specification to different verification engines and compare their performance.

1 MODEL-BASED VERIFICATION AND INTERLOCKING SYSTEMS

Designing interlocking systems for large railway stations is a very complex task. Lots of different routes, not only for passengers but also for logistics and freight traffic, must be combined with a vast traffic network. Moreover, such railway components, e.g., the interlocking system structure and corresponding route network, are safety critical infrastructures. This means, errors in the network plan, as an unconnected or a dead-end track, an incorrect working switch or wrong scheduled and therefore crossing routes can lead to costly and dangerous hazards. Verifying such railroad interlocking systems by applying formal verification techniques (e.g., model checking (Clarke et al. 1999)) increases the quality a lot.

Even though there exists a vast amount of work on the formal verification of railway interlocking systems, techniques as model checking or deductive verification are not commonly used in practice. We think that one major problem is that there exists no off-the-shelve implementation that can be applied as a simple addon to a given interlocking design tool.

In particular, there exist several tools providing their domain specific language or an interface to a particular modeling tool that can do the verification task. However, either you need to find the one tool that applies to your specific modeling environment, or you must transform your model into the input language of your verification tool. The first approach can lead to the conclusion that there does not already exist such tool and the second is very erroneous and due to the time consumption and required skill level of the engineer.

To overcome this problem, we want to define a structural transformation from an accepted interlocking modeling standard into a formal representation of a railway interlocking system and its routing tables. Over the last ten years, a data scheme standard enabling the interchangeability of railroad design data has been developed by a consortium of leading companies from the interlocking and signaling domain. It is called Railway Markup Language (railML) (Nash et al. 2004) and defines a data scheme based on XML. By using this standard, it becomes possible to define a standardized verification approach for general interlocking system designs and logical routing tables. Since we think that this standard will be used through industrial application by the next years, a verification tool based on this can help to lower the hurdle of using formal verification techniques for

interlocking systems in practice. For example, it is imaginable that a an algorithm for the automatic processing of paper-written interlocking schemes generates railML output, as presented in Klockman et al. (2018), which is later on verified by the algorithm defined in this paper.

In this paper, we define a transformation from a given railML description into a formal model representation which can be used for model-based verification, i.e., for verifying safety properties and additional measures as probabilistic analysis or failure injection techniques. Therefore, we defined a set of template automata for essential interlocking elements and their instantiation with the information derived from the railML representation. Further, we present a realworld case study of the Belgian railway station Braine l'Alleud and a representation of the German train station of the city of Leipzig. This is, to provide an idea of the capabilities enabled by the automatic transformation and corresponding model checking tools.

We think that the fully automatic transformation from an accepted railway data scheme will lower the hurdle for the application of formal verification techniques within the development of interlocking systems and even increases the acceptance of the assessment. This is especially the case since we try to provide a 1:1 transformation be keeping as much information of the original data, which also ensure a full, and easy to understand, traceability from the design to the formal model.

Related Work Of course, the verification of railway interlocking systems has already been researched during the last years by several publications as Cimatti et al. (1998a) or Haxthausen et al. (2011), Cappart and Schaus (2016), Cimatti et al. (1998b), Bonacchi (2013), or Limbrée et al. (2016). These authors showed the overall applicability of formal verification, in particular, model checking, in the context of the analysis of railway interlocking systems. However, Banci et al. (2004) presented a first general representation of interlocking components using statemate and Harel state charts (Harel 1987). Further, the first transformation of a railway interlocking system from a specific Domain Specific Language into a formal modeling language, Event-B Abrial (2010), was presented by Iliasov and Romanovsky (2012) in the context of the SafeCap project (Iliasov et al. 2013) intending to improve the time-effectiveness of route plans without violating the safety properties of the systems. Therefore, the B verification engine ProB (Leuschel and Butler 2003) was used. However, the main focus of this project was not on the formal verification but the optimization of the system.

The remainder of the paper is the following: In section two we present preliminary background information about the implemented interlocking components and railML. Further, we show the generation algorithm for the formal model by presenting abstractions for the minimal set of required elements (switches, tracks, etc.) and the corresponding safety properties to be verified (derailment, collision, etc.). Section four presents the real-world case study and experimental results of the verification using different model checking tools. The conclusion and further work are given in section five.

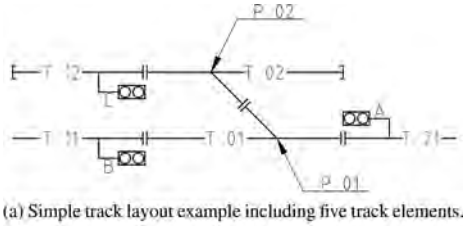
2 THE BASIC ELEMENTS OF AN INTERLOCKING SYSTEM AND RAILML

Before we introduce the mentioned approach, we need to clear the atomic elements of an interlocking system which must be implemented for a correct abstraction of an interlocking system.

Figure 1 presents a simple example of a railway interlocking system's track layout (Fig. 1a) as well as the corresponding interlocking table (Fig. 1b) representing the available routes.

The track layout is divided into connected track elements like T01, T02, T11, etc. Track T12, for example, is a dead-end track ending with a bumper. Openended track elements T11 and T21 can connect the system to other interlockings. Tracks T01 and T02 are connected by switches (the points P01 and P02) for branching between several track elements.

Points (P01 and P02) can be in normal position so that trains will drive straight on, or in reverse position so that trains will branch off the current track element. Further, the given interlocking table supports flank protection, i.e., Switch P02 is also



(a) Simple track layout example including five track elements.

Route	Direction	Tracks	P01	P02
A11	$A \rightarrow T11$	T01, T11	N	N^F
A12	$A \rightarrow T12$	T01, T02, T12	R	R
B21	$B \rightarrow T21$	T01, T21	N	N^F
C21	$C \rightarrow T21$	T02, T01, T21	R	R

(b) Interlocking table of the example track layout containing four simple routes.

Figure 1. A simple interlocking system example with track layout (a) and the route definitions given by an interlocking table with the corresponding direction, track elements, and position of the switches (P01, P02) as nominal (N) or reverse (R).

switched as a safety function. In normal position, it will disconnect track T12 from the main track T11↔T01↔T21 to prevent wagons from rolling back into main track. Of course, an industrial real-world interlocking system plan contains more elements than the presented ones. However, this is sufficient to verify the basic system topology and on top of that the logical route planning and scheduling of the system.

Railway Markup Language If we want to process the data of a given representation in different steps and tools, we are required to base our transformation on some standard scheme. As mentioned before, we, therefore, focus on the infrastructure definition of railML. The definition of railML in general contains schemes for *Timetable Rostering, Infrastructure* definitions, *Rollingstock,* and *Interlocking* tables and signal plans. It would have also been interesting to use the Interlocking scheme which is meant to support the definition of interlocking and route tables. Unfortunately, this scheme is under development at the moment and therefore not available. Instead, we use a simple csv representation of the route table as given in Fig. 1.

In the following, we present an excerpt of the infrastructure railML representation of the example in Fig. 1. In particular, these are elements covered by route *C21* (cf. Fig. 1b), i.e., *T12*, *T02*, *T01*, and correlated elements like switches and signals. In the following, we give a short explanation of the listing in Fig. 2, without providing a complete description of railML. The basic elements of the infrastructure are ① tracks with an *id*. For each *track* element we define its *trackBegin*, its *trackEnd*, and the connections to other track elements or a *bufferStop* if it is a blind end ②. Each *trackBegin* and *trackEnd* has, in addition, a position that is used to define the direction on the track. From the position with the lower value to the higher value the position is *up*wards, otherwise *down*wards. The connections are realized by references to other *trackBegin* or *trackEnd* elements via their unique *id*. Further, we find ③ *switch* elements within a *track*, which can be connected to corresponding switch elements on other tracks, also via *id* of the elements. Signals ④ can be modeled, too, by using *signal* elements with a specific position and a direction in which they work, e. g., *down* if they work in the downwards direction of the track.

Using this representation, we get an applicable standard for exchanging interlocking data between different modeling and verification tools. The original stand contains, of course, more than only these elements, but they are sufficient for proofing the applicability of the presented approach.

```
1  <railml>
2  <infrastruxcture id='example'>
3  <tracks>
4  ① <track id='T12'>
5   <trackTopology>
6    <trackBegin id='T12a' pos='0'>
7  ②  <connection id='T12ac' ref='T02bc'/>
8    </trackBegin>
9    <trackEnd id='T12b' pos='42'>
10 ②  <bufferStop id='T12bc'/>
11   </trackEnd>
12  </trackTopology>
13  <ocsElements>
14   <signals>
15 ③   <signal id='SigC' pos='10' dir='down' type='main'
         → />
16   </signals>
17  </ocsElements>
18 </track>
19 ① <track id='T02'>
20  <trackTopology>
21   <connections>
22 ④   <switch id='P02' pos='21'>
23     <connection id ='P02c' ref='P01c' orientation='
         → incomming'/>
24    </switch>
25   </connections>
26   <trackBegin id='T02a' pos='0'>
27 ②  <bufferstop id='T01ac'/>
28   </trackBegin>
29   <trackEnd id='T02b' pos='42'>...</trackEnd>
30  </trackTopology>
31 </track>
32 ① <track id='T01'>
33  <trackTopology>
34   <connections>
35 ④   <switch id='P01' pos='21'>
36     <connection id ='P01c' ref='P02c' orientation='
         → outgoing'/>
37    </switch>
38   </connections>
39   <trackBegin id='T01a' pos='0'>
40 ②  <connection id='T01ac' ref='T21bc'/>
41   </trackBegin>
42   <trackEnd id='T01b' pos='42'>...</trackEnd>
43  </trackTopology>
44 </track>
45 </tracks>
46 </infrastructure>
47 </railml>
```

Figure 2. Excerpt from the railML representation of the example track layout.

3 A TEMPLATE SYSTEM FOR INTERLOCKING SYSTEMS

The basic idea of our transformation system is to provide a set of accepted templates in a formal language, e.g., SAML in our prototype, for single track elements and to derive the instantiations and the links between elements from the railML representations.

The class diagram in Fig. 3 shows the templates and their relations. All infrastructure elements

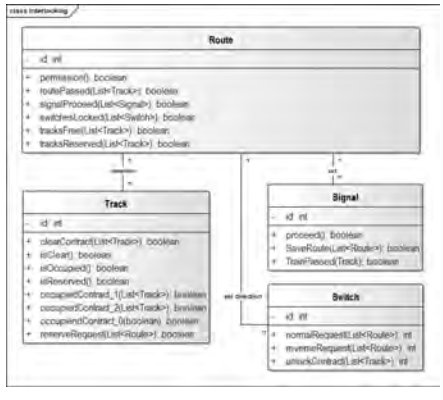


Figure 3. Class model of the templates.

(tracks, signals, switches) are controlled via route automata. Each route holds information about the track segments it contains and the corresponding signal states and switch positions. Therefore, all activities are correlated to the currently active routes, i.e., which tracks must be blocked, which signal must be set, or which switch must be set to a particular position.

Route Scheduling The route scheduling itself is implemented indeterministic, i.e., each nonactive route can request to get active at each time if all required track segments are not blocked by another route. If more than one route tries to get active, we solve this race condition by choosing the route with the lower id (e.g., route 1 before route 2). Overall, more than one route can be active at the same time, but only one additional route can get active at each time step.

Logical Trains At the moment, the train movement is modeled in an indeterministic, logical, way. This means a train can leave a specific track segment or not, without taking into account physical parameters like train velocity and length of the track segment. Each train consists of one to n adjacent track elements to model the movement from one to another track element and trains of different size. If required the formal model can easily extend with physical behavior. Further, we define two trains per route for modeling collisions, flank protection, and other safety properties.

In the following, we present the internal automata of the templates and their basic instantiation idea.

3.1 *Track*

In general, a track can be *clear, reserved*, or *occupied*. Since we are verifying the system with two trains, the track can be occupied by train No. 1 or train No. 2 (*occupiedBy1* or *occupiedBy2*). Model checking, in combination with non-deterministic behavior, covers all possible combination of train positions and routes and therefore two trains are sufficient for analyzing all possible train to train situations.

railML relation In the following, we need information on adjacent tracks. This information is derived from the railML element track (cf. Fig. 2 ①) and underlying connection information (cf. Fig. 2 ②). To identify, which tracks are before and behind a track element, we use the railML direction conventions, where the direction is from the lower *pos* value to the higher. From this, tracks that are referenced *track-Begin* are precedent tracks and those references at *trackEnd* are subsequent. The assignment of tracks to routes and their required direction (*normal* or *reverse*) is extracted from the route table.

Behavior A track can be requested by a route. This is done indirectly by checking the current state of all routes containing the track.

$$reserve\ Request\ := \bigvee_{r \in requireingRoutes} r.state = commanded$$

Tracks will be occupied in addiction to their neighbor tracks.

$$\begin{aligned} occupiedContract1 :&= trackLeft.occupiedBy1 \\ &\vee\ trackRight : occupiedBy1 \\ occupiedContract2 :&= trackLeft.occupiedBy2 \\ &\vee\ trackRight.occupiedBy2 \end{aligned}$$

The initial state of the model is with empty tracks, so there has to be a possibility to place some trains on the tracks. Therefore, all possible starting tracks have a non-contradictional formula *occupiedContract*0 with

$$occupiedContract0 := \bigvee_{r \in requireingRoutes} r.commanded$$

The transition from occupied to clear is enabled when the next track is occupied, and the previous track is clear. This assures that a train will not be split or removed.

Further a train is not forced to enter or leave a track in every step modeling trains with different length.

$$clearContract := (trackLeft.occupied \wedge trackRight.state = clear) \vee (trackLeft.state = clear \wedge trackRight.occupied)$$

3.2 *Signals*

Signals can be in one of two states – *stop* (red) or *proceed* (green). States which allow the trains to proceed with low speed are not modeled.

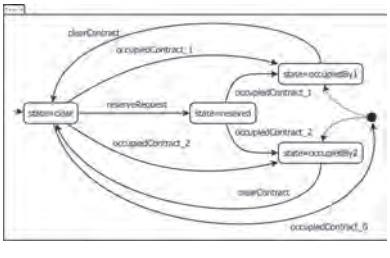


Figure 4. Automat model of tracks.

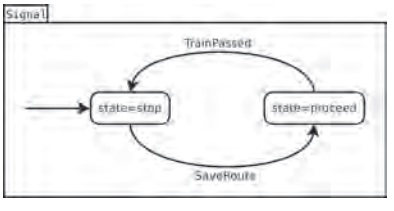


Figure 5. Internal automaton for signals.

railML relation For the signals, we need to know on which track the signal is placed and the direction of the signal, as well as the next track behind the signal to define when the train entered a route and set the signal respectively. These information can be derived from the signal element of the railML description (cf. Fig. 2 ③) and the parent track element.

Behavior The complete automaton of the signal template is given in Fig. 5. Its initial state is *stop*. It changes to *proceed* if the following route (behind this signal) is ready, i. e., the route is accessible. *Save Route* gets true one of the routes starting at the signal is accessible and the signal state changes to *proceed*.

$$saveRoute := \bigvee_{r \in requireingRoutes} r.proved$$

If a train passed the signal, it falls back to *stop*. Therefore, formula *TrainPassed* is connected to the track behind the signal to detect when a train passed it. This is detected by a subsequently occupied track.

$$trainPassed := trackBehind.isOccupied$$

3.3 *Switch*

A switch can be in two states according to its position: *normal* (straight on) and *reverse* (branching).

railML relation Besides the *id*, we must derive the information of the connection relations from the switch, i.e., which tracks are connected via *normal* and reverse position. This is done by utilizing the switch tag information (cf. Fig. 2 ④) and the references to the adjacent track or switch in combination with the parent track element.

Behavior A switch automaton has four states, two for each position *normal* or *reverse*, and two according to its possibility to change its position. This means a switch can be *locked* (no change possible), or *unlocked* (changing the position is possible) for preventing switches from changing while occupied by a train. The transitions *normalRequest* and *reverse Request* will cause the switch to change its position corresponding to the one required by the active route and lock the switch against any other requests.

$$normalRequest := \bigvee_{r \in requireingRoutesNormal} r.reserved$$

$$reverseRequest := \bigvee_{r \in requireingRoutesReserve} r.reserved$$

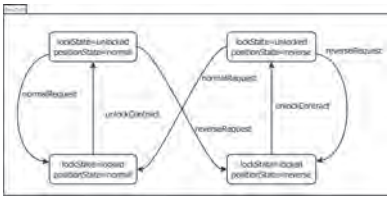


Figure 6. Internal automaton for switches.

The transition *unlockContract* will open the switch for new commands if both connected tracks are clear.

$$unlockContract := Track1.isClear \wedge Track2.isClear$$

3.4 *Route scheduling*

railML relation As mentioned before, the railML *Interlocking* scheme is not finished, i.e., not available, and therefore we have to process simple csv data. This has the same structure as given in the route table in Fig 1b.

Behavior A route controls and observes all track side elements necessary for a save train movement. It can be in one of three states (*idle*, *commanded*, or *occupied*). A route is initialized as *idle*. As previously mentioned, the routes are commanded nondeterministic.

If a route can get active, i.e., is in the state *commanded*, it causes the corresponding tracks to be reserved for this route and the switches to change their position as required. The route will check the state of the elements (*tracksReserved* and *switchesLocked*).

$$tracksReserved := \bigwedge_{t \in includedTracks} t.reserved$$

$$switchesLocked := \bigwedge_{s \in includedSignals} s.locked$$

If everything is correct the start signal is changed to *proceed*.

$$signalProceed := \bigwedge_{s \in includedSignals} s.proceed$$

After that, the route will be *occupied*. When the train passed the complete route, the route will change its state back to idle.

$$routePassed := \bigwedge_{t \in includedTracks} t.isClear$$

3.5 *Safety specifications*

In the following, we want to present the safety properties we provide for our method. Since the defined model is built after a given mechanism, we

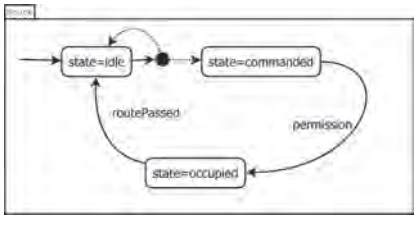


Figure 7. Internal automaton for routes.

think it can be easily extended with other safety specifications as the presented ones.

3.5.1 *Collision detection*

Two trains collide if they are at the same time at the same place. As described in 3.1 there are two or more track objects for one track element in the track layout to save additional information about the direction of the trains. Therefore, a collision occurs if a track is occupied by more than one train. With this definition, you can detect collisions on switches and head-to-head collisions.

To detect rear-end collisions, tracks have to recognize if a collision can occur in the next step. This is done by observing the reserved and occupied state of the track elements. If a track should get occupied in the next step, *occupiedContract_a* gets true. If it is also already occupied by another train (*track.state = occupied By_b*) we found a possible collision in the next step.

$$collision(Trackt) :=$$
$$(t.occupiedContract\_1 \,\&\, t.state = occupiedBy2)$$
$$|\,(t.occupiedContract\_2 \,\&\, t.state = occupiedBy1)$$

3.5.2 *Derailment detection*

In this model derailment due to wrong switch positions is checked. The two cases which can happen are i) derailment because a train is moving over a switch while it is changing its position and ii) derailment because a train is passing a switch which is in a wrong position.

The specification will ensure that the position of a switch correlates with the track object which is occupied. So it will check if the switch is in the correct position according to the direction of the train. In the transformation, we derive the correct position for a switch from the route table and encode this within the model. For readability, we refer here to *inRightPos* that evaluate to *true* if the switch has the required position. As a second aspect, the specification will check if the switch is locked every time a train is on a track linked to this switch.

$$isDerailed :=$$
$$\bigvee_{t \in Tracks} (t.isOccupied$$
$$\bigwedge_{s \in switch(t)} s.inRigthPos \wedge s.isLocked)$$

3.5.3 *Flank protection check*

Especially on sidings, some parked wagons can start rolling unadvisedly. To prevent accidents between those wagons and trains specific switches which are not on the route of the train can be commanded to a position so that they route the wagons to the other track than used by the train. Nearly the same problem is caused by trains which miss stopping at red signals.

(a) Train flank protected. (b) Train flank not protected. (c) Flank protection not available.

Figure 8. Verification of flank protection.

For checking this safety guideline every time a train is on a switch a backward search is performed. This backward search will start at the not used branch of the switch and will progress hand over hand along the tracks. If it reaches a switch which is in the position to guide wagons away from the train, it will stop (8a). If the switch is in the other position, the search will proceed (8b). If the search reaches a uniting switch, it will continue on both branches (8c). The specification ensures that the backward search will never reach a station or main track.

This formula is also derived automatically from the track layout for each route, i.e., for each route that is active, all adjacent tracks, which are not directly on a function *isFlankProtected(r)* is defined. It returns true if the switch in in the correct position.

$$isFlankProtected(Router) :=$$
$$\bigwedge_{s \in Switch} s.sFlankProtected(r)$$

4 CASE STUDY BRAINE L'ALLUDE

The validation of the presented modeling methods was executed on a real-world case study of the Belgian train station Braine l'Alleud, taken from (Cappart and Schaus 2016). Braine l'Alleud station consists of four platforms, twelve switches, twelve signals, and 18 track segments. From the given layout, 32 different routes are available.

From this layout, we generated a model with about 100 state machines. This contains representations of the track elements, switches, routes, etc. In the following, we present the verification times of different model checking engines for the defined specifications from sec. 3.5. For providing information about the computation times and capabilities of different verification methods, different tools implementing divers qualitative model checking techniques were chosen. Further, we checked these specifications on a correct model, i.e., the specifications hold for the system. On the other hand, we injected faulty behavior to provoke erroneous behavior. The failures can be categorized as follows:

- a route does not reserve a needed track
- a route requests a wrong position of a switch
- a route does not request any position of a needed switch

To check the verification time for the erroneous models, we also model checked the specifications and calculated the mean time for each specification, not separating for the found error (i.e., whether a switch or a track was not commanded correctly). We hope that the evaluation of different tools may help users, which are not familiar with model checking, in choosing a proper tool for their purpose.

4.1 *Experiments with the model checking tools*

To check the modeled interlocking system the model checking tools iimc, nuXmv, and aigbmc were used. The tests were performed on a computer with Intel i7 core (3.2 GHz). The target language of our prototype is the System Analysis and Modeling Language (SAML) (Güdemann and Ortmeier 2010). Using SAML, we can verify the imported qualitative model with several states of the art model checking tools (nuXmv, iimc, UUPAAL), for which the SAML IDE VECS (Verification Environment for Critical Systems) provides implemented connectors. We used this connectors for the experiments with the different model checking tools.

aigbmc This tool is a bounded model checker built on top of the AIGER distribution provided by Biere (2007). Being a bounded model-checker makes it necessarily incomplete (if the system diameter is unknown), but allows for very efficient counter-example search, as only the base case is checked and the induction step is not encoded.

iimc This tool, described by Hassan et al. (2012), is the evolution of the original IC3 method. It uses different proof engines (BMC, IC3, FAIR), depending on the type of property to verify. The tool is one of the most efficient model-checkers for sequential circuits and allows for multi-threaded verification.

nuXmv This tool is the evolution of NuSMV, described by Cavada et al. (2014). It supports infinite state spaces via real-valued variables with an analysis based on the SMT solver MathSAT5 and also PDR-style verification using SMT solvers (Cimatti and Griggio 2012).

4.2 *Evaluation of the results*

In the following, we present the results of the verification experiments. Table 1 presents the verification run time for the different tools and specifications. In this diagram, the BMC-based algorithms are missing since BMC is not capable of proving the correctness of a system rather than falsifying specifications. One most important fact is that all tools that terminated for a given specification also produced correct verification results for both, the correct and the incorrect model. For the verification of the correct model iimc performed best for all specifications. It took between 120 s for the collision specification and 608 s for the flank protection, which is a quite acceptable computation time. In comparison, the nuXmv model checker had quite more difficulties with about 2400s for the derailment specification (iimc: 120 s).

Table 1. Verification results of the evaluation. The star * marks the specifications of the defect model.

Spec	iimcIC3	iimcBMC	nuXmv	AIGBMC
collision	120*s*	—	1419*s*	—
collision*	45*s*	2*s*	52*s*	51*s*
derailment	120*s*	—	2402*s*	—
derailment*	38*s*	3*s*	33*s*	29*s*
flank protect	608*s*	—	—	—
flank protect*	223*s*	—	437*s*	—

For the verification of the incorrect model we also used the BMC-based tools. The results show that for medium complex formulae, as the derailment or the flank protection specification, BMC-based algorithms outperformed the inductive ones. However, they were not able to complete their analysis for the flank protection (2 s iimc BMC compared to 45 s iimc IC3 for the derailments specification). This behavior is not a surprise. The reason is that the BMC algorithm is very fast for simple specifications and short counterexample paths (10 to 11 steps for derailment and collision). However, the algorithm is not very applicable for complex specifications and, moreover, with high bounds, it was not able to compute the results for the flank protection specification within 24 h. This is connected to the algorithm itself since for each step towards the bound, new formulas are added increasing the complexity of the underlying SAT problem. In contrast to that, it is quite fast for the short counter examples since the computation of the invariants is quite complex, without a direct correlation to the number of steps.

Summarizing, BMC has shown up to be a suitable method for finding bugs in a system, especially if the error bound is small. Nevertheless an IC3 implementation is needed for proving the correctness of the system and, moreover, even for the discovery of errors in a deep bound, e.g., the presented flank protection faults. This means iimc would be the best tool out of our small collection for the verification of the interlocking systems since it contains both, a fast BMC and a fast IC3 engine.

5 CONCLUSION

In this paper, we presented an approach for reducing the complexity of the formal verification of

railroad interlocking systems. Therefore, we presented an approach for automatically generating a formal model from a given track layout and the corresponding route definitions. Moreover, the declaration rules for basic safety specifications, i.e., derailment, collision avoidance, and flank protection were presented.

To validate our method we generated a model of a real-world case study of the Belgian train station Braine l'Allude. For giving an insight view into available verification algorithms and tools, we verified the specifications with four state-of-the-art verification tools implementing the currently most effective algorithms, IC3 and Bounded Model Checking. In addition to the experiments where the specifications hold, we also examined the behavior of the verification for the error case and injected failures into the model.

The results show, in our point of view, the applicability of the transformation and the computability of the verification. This opens new perspectives for the analysis of interlocking systems since the presented target modeling language SAML supports both, qualitative as well as quantitative verification methods.

REFERENCES

Abrial, J.-R. (2010). *Modeling in Event-B: system and software engineering*. Cambridge University Press.

Banci, M., A. Fantechi, & S. Gnesi (2004). The role of formal methods in developing a distributed railway interlocking system. In *Proceedings of FORMS/FORMAT*.

Biere, A. (2007). The aiger and-inverter graph (aig) format. *Available at fmv. jku. at/aiger*.

Bonacchi, A. (2013). Formal safety proof: a real case study in a railway interlocking system. In *Proceedings of the ISSTA 2013*.

Cappart, Q. & P. Schaus (2016). A Dedicated Algorithm for Verification of Interlocking Systems. In *SAFECOMP*.

Cavada, R., A. Cimatti, M. Dorigatti, A. Griggio, A. Mariotti, A. Micheli, S. Mover, M. Roveri, & S. Tonetta (2014). The nuXmv Symbolic Model Checker. In *Proceedings of CAV*.

Cimatti, A. & A. Griggio (2012). Software Model Checking via IC3. In *Proceedings of CAV*.

Cimatti, A., F. Giunchiglia, G. Mongardi, D. Romano, F. Torielli, & P. Traverso (1998a). Formal verification of a railway interlocking system using model checking. *Formal Aspects of Computing*.

Cimatti, A., F. Giunchiglia, G. Mongardi, D. Romano, F. Torielli, & P. Traverso (1998b). Model checking safety critical software with spin: an application to a railway interlocking system. *Proceedings of SAFECOMP 1998*.

Clarke, E.M., O. Grumberg, & D. Peled (1999). *Model checking*. MIT press.

Güdemann, M. & F. Ortmeier (2010). A framework for qualitative and quantitative model-based safety analysis. In *Proceedings of HASE 2010*.

Harel, D. (1987). Statecharts: A visual formalism for complex systems. *Science of computer programming*.

Hassan, Z., A. Bradley, & F. Somenzi (2012). Incremental, Inductive CTL Model Checking. In *Proceedings of CAV*.

Haxthausen, A.E., J. Peleska, & S. Kinder (2011). A formal approach for the construction and verification of railway control systems. *Formal aspects of computing*.

Iliasov, A. & A. Romanovsky (2012). Safecap domain language for reasoning about safety and capacity. In *Proceedings of WDTS-RASD 2012*.

Iliasov, A., I. Lopatkin, & A. Romanovsky (2013). The safecap platform for modelling railway safety and capacity. In *Proceedings of SAFECOMP 2013*.

Klockman, M., M. Filax, F. Ortmeier, & M. Reiß (2018). On the similarities of fingerprints and railroad tracks: Using minutiae detection algorithms to digitize track plans. In *Proceedings of DAS 2018*.

Leuschel, M. & M. Butler (2003). Prob: A model checker for b. In *Proceedings of FME*.

Limbrée, C., Q. Cappart, C. Pecheur, & S. Tonetta (2016). Verification of railway interlocking-compositional approach with ocra. In *Proceedings of RSSRail 2016*.

Nash, A., D. Huerlimann, J. Schütte, & V.P. Krauss (2004). Railml—a standard data interface for railroad applications. *WIT Transactions on The Built Environment*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Mathematical modelling of critical infrastructure reliability

D. Vališ, K. Hasilová & Z. Vintr
University of Defence in Brno, Brno, Czech Republic

M. Forbelská
Mendel University in Brno, Brno, Czech Republic

ABSTRACT: Safety and dependability of technical systems are key aspects when it comes to a system quality. Critical infrastructure covers many parts such as mains, transportation system, networks and buildings. In our paper, we put an emphasis on a water distribution network where high availability, reliability and safety are very much demanded. Such network usually consists of more lines with different hierarchical importance and age. Moreover, various materials are usually used to manufacture water pipes. The in-field operation of a water distribution system is not very well recorded since it usually does not contain all details about failure occurrences. We have interesting water distribution system recordings covering a wide time span, but, unfortunately, they provide only the numbers of failures during respective months. In this paper, we apply a selected form of dynamic linear models—a modified Kalman filter with the implementation of a structural break point.

1 INTRODUCTION

Critical Infrastructure (CI) is a vital part of each country. It includes important elements such as energy supply networks and distribution systems of, e.g. gas, water, etc. Availability, safety and security of CI elements are key aspects for a country to function smoothly. Some CI elements and nodes are continuously monitored to watch their condition. Failures, which might occur time and again, are also recorded. Unfortunately, not all the records provide every detail about undesirable event occurrence such as failure.

Our paper includes an approach to analyse water distribution system field data related to failure occurrence. Our intention is to introduce the way of modelling some dependability measures despite having only limited records. During our research, we have noticed that there is a certain break point in the displayed data course. It might be explained by the gradual improvement of the water distribution network when old lines are being replaced by new ones. This results in improving some reliability measures. We would like to describe this process analytically and predict how it is going to be developed in the future. For this purpose, we use a dynamic linear model—the Kalman filter with a structural break point.

1.1 *State of the art*

CI is studied from several perspectives—from the conceptual ones to specific applications. A detailed review can be found, for example, in Wilt et al. (2016). Here, we mention only a few articles concerning failures of the critical infrastructure.

Zio (2007) initiated a change in a point of view to frameworks capturing properties emerging from a complex system of critical infrastructure such as water supply, transportation, information and other networks. Eusgeld et al. (2009) proposed a methodological framework for the analysis of critical infrastructures using topology-driven analysis of vulnerabilities. The authors applied the proposed methodology on the Swiss high-voltage grid. Jaskolka and Villasenor (2017) focused on interactions in a distribution system and analysed dependencies in the system. Their approach is based on concurrent Kleene algebra, which offers sequential and concurrent compositions. Korkali et al. (2017) pointed out that mechanism of failures in models provided so far differs from reality where interconnections can be present. They compared several models to understand the impact of network topology. They concluded that understanding both benefits and risks of interconnections is crucial to designing robust and resilient systems.

We would like to focus on water distribution systems. These systems are studied with respect to several factors such as plans for maintenance, rehabilitation and replacement (M/R/R); application to a specific water network; methods and models for water mains failures.

Failures in water networks were studied by Cemagref group. They presented methods to support maintenance of the network based on a survival

analysis model (Le Gat & Eisenbeis 2000). Reliability of the water distribution system was studied in (Jung et al. 2016). The authors investigated sixteen study networks to develop linear reliability models using linear regression. Failures caused by external influences are studied, e.g., in Otřísal et al. (2017). Classification of a water distribution system was given in (Hwang & Lansey 2017). The authors used a graph theory combined with classification to determine adequate parameters for describing the water distribution system.

Evaluation of the efficiency of urban water infrastructure and determination of the optimal time of rehabilitation is given in (Karamouz et al. 2017). The study is completed by examination of the model in a real water distribution network.

Failures are a common keyword of all above mentioned sources. Hazard function (or failure/intensity has been studied by many authors, beginning with Ascher (1970), unified approach for both repairable and non-repairable items by Hokstad (1997), and its modifications by Woch & Vališ (2017). For more complex (non-reconfigurable) flow networks, one can use a topological or algebraic approach (Todinov 2012, Hošková-Mayerová et al. 2013, Ameri et al. 2016).

However, having a limited though interesting data set, we turn our attention to statistical methods. Regarding repairable systems, the non-homogeneous Poisson process is taken into account and its properties were studied by MacLean (1974) or Krivtsov (2007). The data set is in the form of the time series; therefore, Kalman filter is a suitable tool to study the series. Kalman filter was introduced by Kalman (1960) and has been used since then, e.g. in the latest study by Arthur et al. (2018).

2 ANALYSED DATA DESCRIPTION

We have a data set which contains the failures of water lines in three different structures—a magistral/main line, a local line and a branch line to a house. However, the failure records cover the number of failures during single months only, and, unfortunately, there is no more information available. In practice it is rather common to keep records this way. The data set, however, is quite rich when we consider how long it took to record the information—it was a period of fifteen years.

For the purpose of our further analysis, we convert the failure records into an event occurrence rate during the period we are interested in. It is common to use the occurrence rate per one day or a month. However, if we wanted to assess how the seasons affected the process for example, the data would be sorted by quarters, but this is not our case since we are working with the number of failures per month.

Table 1. Example of water distribution system field failures records.

Sample nr.	Month/ year	Nr. of failures	Failure frequency
1	01/2000	6	0.193 548
2	02/2000	12	0.413 793
3	03/2000	27	0.870 968
4	04/2000	23	0.766 666
5	05/2000	18	0.580 645
6	06/2000	8	0.266 667
…	…	…	…

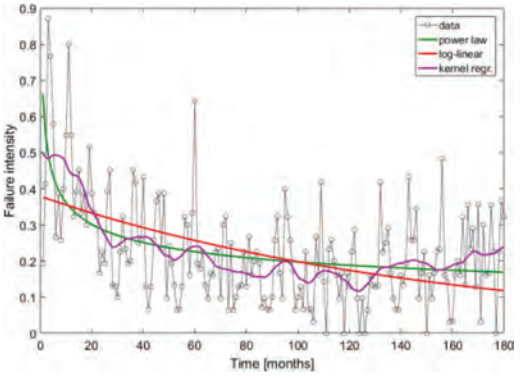


Figure 1. Water distribution system field failures—in the form of failure frequency time series with typical courses applied for non-homogeneous Poisson processes.

In Table 1 there is an example of the way to record the failures and convert them into event frequency which we develop further. The data converted into failure frequency are illustrated in Figure 1. Since this is the Non-Homogeneous Poisson Process (NHPP), a common way to assess this kind of data is modelling with the use of a power-law model and a log-normal model (Rausand & Høyland 2004, NIST/SEMATECH 2017) as we can see in Figure 1. These parametric courses are complemented by a non-parametric kernel regression estimate.

3 THEORETICAL FORM OF MATHEMATICAL MODEL

In order to analyse the data described above, we use in this case time series modelling. If we are to comply with some assumptions for the time series (e.g. an equidistant time increment), it will be necessary to work with a pseudo measurement, failure frequency, where this principle is observed. Another

assumption is a correlation rate in data which will be proved by performing further tests.

For data assessment, we apply generalized dynamic time series model based on linear Kalman filter (LKF) (Bhar 2010, Bain & Crisan 2009). Let the dynamic process X_t follow a transition equation

$$x_t = f(x_{t-1}, w_t) \tag{1}$$

and we also assume that we have a measurement Y_t such that

$$y_t = h(x_t, u_t). \tag{2}$$

In the above equations (1) and (2), w_t and u_t are two mutually uncorrelated sequences of temporally uncorrelated sequence of normal random variables with zero mean and covariance matrices Q_t and R_t, respectively. Additionally, w_t is uncorrelated with x_{t-1} and u_t is uncorrelated with x_t. The prior process estimate is defined as

$$x_{t|t-1} = E[x_t]. \tag{3}$$

which is the estimate of x_t at time t–1 just to making the measurement at time t. Similarly, we define the posterior estimate as

$$x_{t|t} = E[x_t \mid y_t]. \tag{4}$$

which is the estimate at time t after the measurement at t has taken place. We also have the corresponding estimation errors $e_{t|t-1} = x_t - x_{t|t-1}$ and $e_t = x_t - x_t|y_t$. These give us the estimate of the error covariates as

$$P_{t|t-1} = E[e_{t|t-1} \cdot e'_{t|t-1}], P_{t|t} = E[e_t \cdot e'_t]. \tag{5}$$

In order to compute the above mean and covariance, we need the corresponding conditional densities $p(x_t|y_{t-1})$ and $p(x_t|y_t)$. These are determined iteratively via transition and measurement updates. The basic idea is to define the probability density function corresponding to the hidden state x_t given all the measurements made up to that time i.e. $y_{1:t}$. The transition step is based on Chapman-Kolmogorov equation

$$\begin{aligned} p(x_t \mid y_{1:t-1}) &= \int p(x_t \mid x_{t-1}, y_{1:t-1}) p(x_{t-1} \mid y_{1:t-1}) dx_{t-1} \\ &= \int p(x_t \mid x_{t-1}) p(x_{t-1} \mid y_{1:t-1}) dx_{t-1} \end{aligned} \tag{6}$$

following the Markov property. The measurement update step is based on Bayes rule

$$p(x_t \mid y_{1:t-1}) = \frac{p(y_t \mid x_t) p(x_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})} \tag{7}$$

and $p(y_t|y_{1:t-1}) = \int p(y_t|x_t)p(x_t|y_{1:t-1})\mathrm{d}x_t$. At this point it is instructive to specialize the transition and the measurement equations (1) and (2) for a linear system and state the updating equation in a form amenable for easier implementation.

Let us focus on a linear state space system with transition equation of the form

$$x_t = T_t x_{t-1} + c_t + w_t \tag{8}$$

and the measurement equation

$$y_t = Z_t x_t + d_t + u_t \tag{9}$$

where c_t and d_t are possible time dependent vectors of compatible dimensions. Similarly, the matrices T_t and Z_t are of dimensions compatible with the length of the state vector x_t and the measurement vector y_t, respectively.

For our purposes, however, we apply two forms of dynamic local linear models.

3.1 *Dynamic linear local level model*

First form of model—called dynamic Local Level Model (LLM) has the following form:

$$\text{observed series: } y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$$

$$\text{latent level: } \mu_t = \mu_{t-1} + \eta_t \qquad \eta_t \sim NID(0, \sigma_\eta^2).$$

Looking at the data plot (see Figure 1), we can notice a negative spike around year 2002 (after approx. 24 months). Recall that the model we just fit the data to (local level plus noise) assume that the variance matrices σ_ε^2 and σ_η^2 are constant over time. Therefore, one way to improve the accuracy of this model and take the jump in failure level (around March 2002) into account is to assume that the variance did change in this year. The new model LLM therefore becomes:

$$\text{observed series: } y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$$

$$\text{latent level: } \mu_t = \mu_{t-1} + \eta_t \qquad \eta_t \sim NID(0, \sigma_\eta^2(t))$$

$$\text{where } \sigma_\eta^2(t) = \begin{cases} w & t \neq t_{break} \\ w* & t = t_{break} \end{cases}$$

3.2 *Dynamic linear local level model—seasonal*

The other type of the applied dynamic linear local model is the one which takes into account seasonality. The dynamic Local Level Model (LLM) described in the previous sub-paragraph is extended by seasonality here and has the following form:
observed series:

$$y_t = \mu_t + \gamma_{1,t} + \varepsilon_t \qquad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$$

latent level with jump:

$$\mu_t = \mu_{t-1} + \eta_t \qquad \eta_t \sim NID(0, \sigma_\eta^2(t))$$
$$where\ \sigma_\eta^2(t) = \begin{cases} w & t \neq t_{break} \\ w^* & t = t_{break} \end{cases}$$

latent seasonality:

$$\gamma_{1,t} = -\sum_{j=1}^{s-1} \gamma_{j,t-1} + \omega_t \qquad \omega_t \sim NID(0, \sigma_\omega^2)$$
$$\gamma_{2,t} = \gamma_{1,t-1}$$
$$\vdots$$
$$\gamma_{s-1,t} = \gamma_{s-2,t-1}$$

By applying this type of model, we intend to trace potential influence the seasons could have on the course of the observed measure of time series failure intensity.

Before we actually start the modelling of both types of defined dynamic models, we have to apply the Show test sequential algorithm (Bai et al., 1997a, 1997b, 1998, 2003) in order to find the structural break points.

3.3 *Optimal break-points estimation*

The foundation for estimating breaks in time series regression models was given by Bai (1994) and was extended to multiple breaks by Bai (1997a) and Bai & Perron (1998). We implement the algorithm described in (Bai & Perron, 2003) for simultaneous estimation of multiple breakpoints. The distribution function used for the confidence intervals for the breakpoints is given in (Bai 1997b). The ideas behind this implementation are described in Zeileis et al. (2003).

In order to determine the optimal number of *m* break-points we use sequential tests and compare the criteria based on Bayes Information Criterion (BIC) and Residual Sum of Squares (RSS). The principle is based on the RSS calculation and it shows that the bigger the number of regression coefficients which are supposed to be another break-point, the higher the penalization which affects the result of the BIC criterion. The results for the optimal numbers of *m* break-points are put in Table 2.

The selection of the number of break-points *m* might be shown the following way, see Figure 2. Based on the results, we then select one break-point in our further research.

Table 2. Reference criteria for break-points *m* choice.

m	0	1	2	3	4	5
RSS	3.964	2.284	2.719	2.658	2.622	2.636
BIC	–164.6	–216.3	–212.7	–206.4	–198.5	–187.1

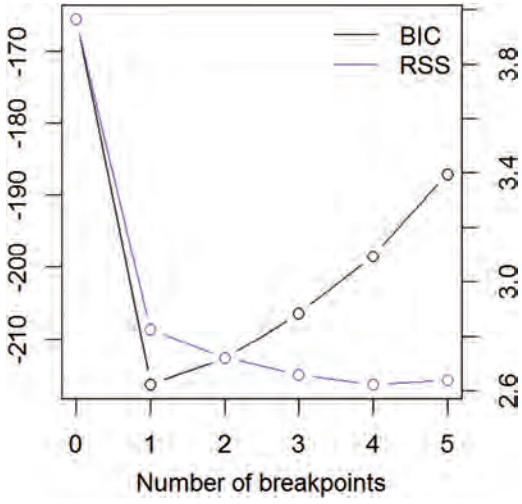


Figure 2. Graphical representation of number of break-points *m* selection criteria RSS and BIC.

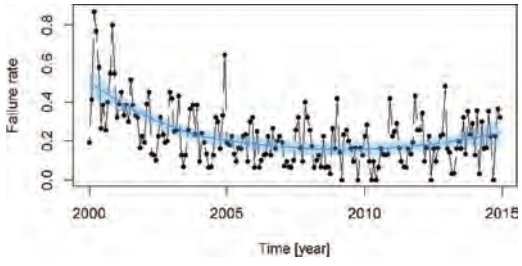


Figure 3. Water distribution system field failures – failure frequency/intensity time series with loess course and its confidence intervals.

4 RESULTS OF SELECTED MODELLING

In this part, we are gradually introducing the results achieved by using the mathematical tools mentioned in paragraphs 3.1 and 3.2. For modelling and simulating, we use the product R-Studio (R Core Team 2005).

First we introduce the approximation of the basic data course, failure intensity, using non-parametric kernel estimation—loess function, see Figure 3.

Before we start the actual modelling, we first construct a static model which is later used for determining initial parameters of the dynamic linear models. Although during the modelling and simulation the software system would be able to estimate the initial parameters of the dynamic models, our own input is by all means better and therefore the static model is made first. This static model is put in Figure 4.

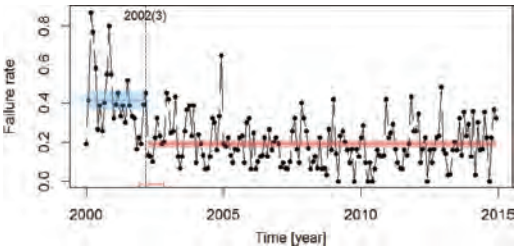


Figure 4. Construction of static linear model with confidence intervals for observed series—failure intensity.

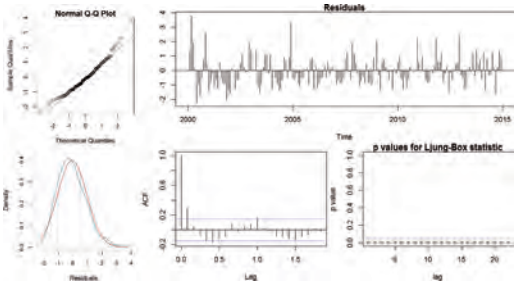


Figure 5. Analysis of residuals of LLM based on Q-Q plot, Auto Correlation Function (ACF) and *p*-value of Box-Ljung.

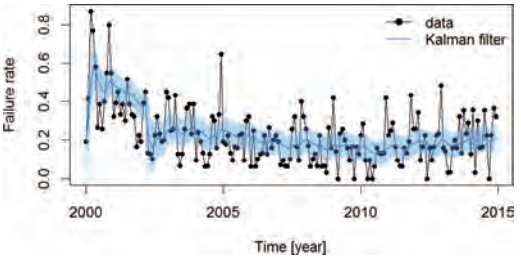


Figure 6. Kalman filter of LLM with confidence intervals for observed series—failure intensity.

4.1 *Results of pure LLM*

In the following steps, we deal with the results of pure LLM. First we analyse and test residuals to find required normality. It is useful to put the results in a graph, see Figure 5. They show clearly that in the normality test (here Q-Q plot) there is a decent compliance. However, the ACF and *p*-value Box-Ljung analyses show that there might be a certain correlation among the data (events-failures).

In our next step, we make the Kalman filter for the LLM model. The result is put in Figure 6 and it clearly shows a few potential break-points. The only break-point we have selected is illustrated in Figure 6 by a dashed vertical line.

For illustration, we also introduce the decomposition of single components of this applied LLM type Kalman filter and the observed series, see Figure 7.

Next, we construct the Kalman smoother for our analysed data and the LLM model. This is shown in Figure 8, where we can also clearly see the break-point illustrated by a vertical dashed line.

There are also the decompositions of the Kalman smoother for our LLM model and the observed series, see Figure 9.

After, we set the course for the Kalman predictor of LLM and the observed series—failure intensity. The predictor is predicted 10% steps ahead which is about 18 months. The result is shown in Figure 10.

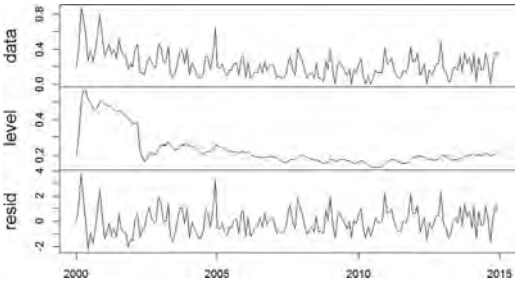


Figure 7. Kalman filter of LLM for observed series—decomposed.

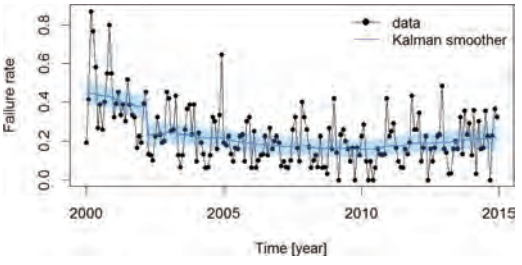


Figure 8. Kalman smoother with confidence intervals of LLM for observed series—failure intensity.

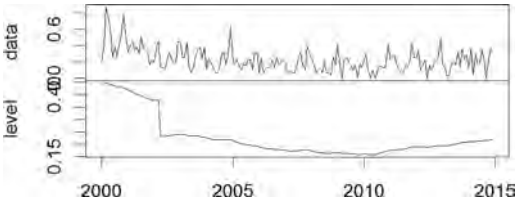


Figure 9. Kalman smoother of LLM and observed series—failure intensity—decomposed.

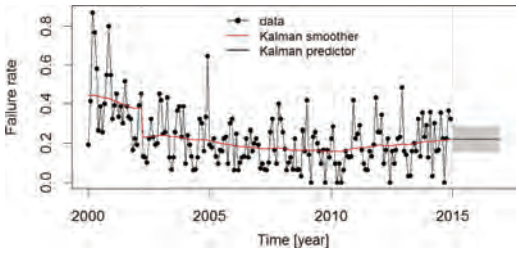


Figure 10. Kalman predictor of LLM with confidence intervals for observed series—failure intensity.

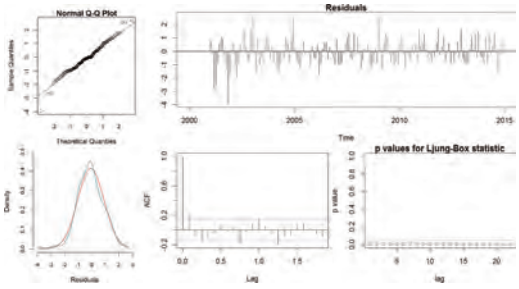


Figure 11. Analysis of residuals of LLM—seasonal based on Q-Q plot, Auto Correlation Function (ACF) and *p*-value of Box-Ljung.

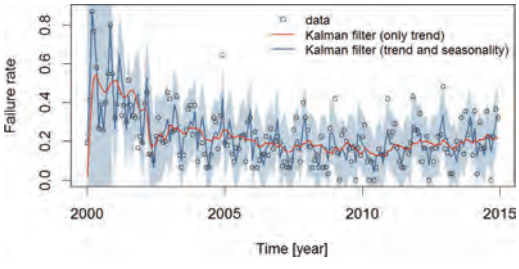


Figure 12. Kalman filter with confidence intervals of LLM—seasonal for observed series—failure intensity.

4.2 *Results of extended LLM—seasonal*

Next, we deal with the results of the LLM model extended of seasonality. First, we analyse test residuals to find required normality. Again it is useful to put the results in a graph, see Figure 11. Again they show clearly that in the normality test (here Q-Q plot) there is a decent compliance which might be even higher than in the previous model. However, the ACF and *p*-value Box-Ljung analyses show that there must be a certain correlation among data (events-failures).

After, we construct the Kalman filter for the introduced seasonal LLM model. The result is put in Figure 12 and it clearly shows a few potential break-points. The only break-point we have selected is illustrated in Figure 12 by a dashed vertical line.

The effect of seasonality is also nicely remarkable in the figure.

For illustration, we also decomposed single components of this applied Kalman filter of seasonal LLM for the observed series, see Figure 13.

In the subsequent stage, we construct the Kalman smoother for our analysed data and the seasonal LLM model. This is shown in Figure 14, where also a break-point in the form of a vertical dashed line along with the effect of seasonality are clearly seen.

For illustration purposes, we also introduce the decomposition of the Kalman smoother for our seasonal LLM model for the observed series, see Figure 15.

After that we set the course for the Kalman predictor of the seasonal LLM for the observed series, failure intensity. The predictor is predicted again 10% steps ahead which is approximately 18 months. The result is put in Figure 16.

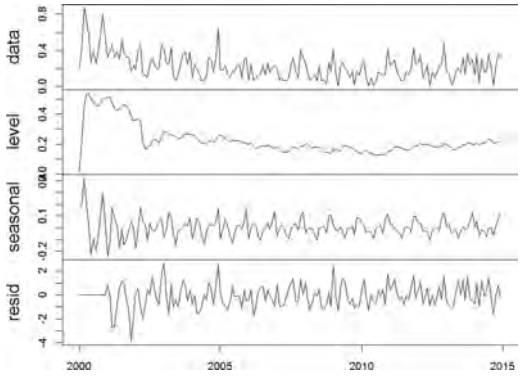


Figure 13. Kalman filter of LLM—seasonal for observed series—failure intensity—decomposed.

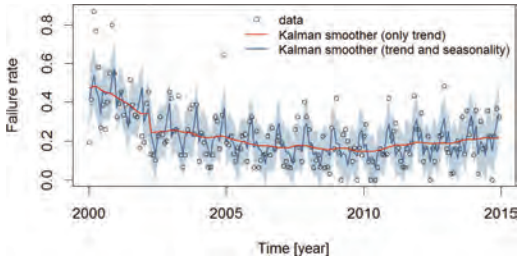


Figure 14. Kalman smoother of LLM—seasonal with confidence intervals for observed series—failure intensity.

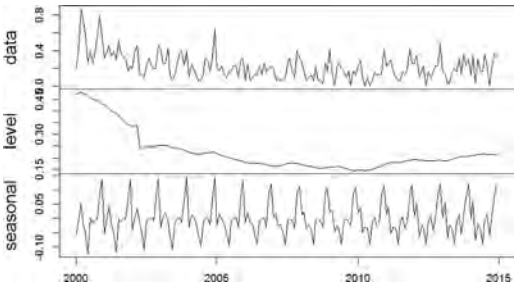


Figure 15. Kalman smoother of LLM—seasonal for observed series—failure intensity—decomposed.

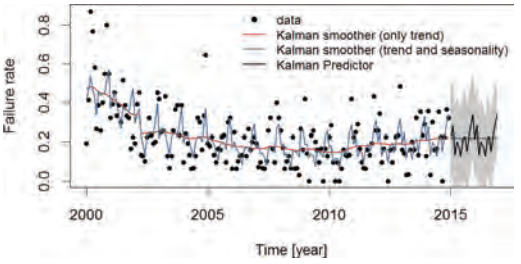


Figure 16. Kalman predictor of LLM—seasonal for observed series—failure intensity.

5 DISCUSSION

All the numerical values representing single mean value courses of the Kalman filter (KF), the Kalman smoother (KS) and the Kalman predictor (KP) from the applied normal and seasonal LLM models are put in Tables 3 and 4. For comparison purposes, we also included a basic data set for the failure intensity observed series. The results clearly show that each model has a specific estimation power for the observed time series; therefore, the course results slightly vary. The great advantage of the applied models is their ability to estimate the course and predict a few steps ahead. However, when predicting future development, a constant trend is seen for the LLM.

The above estimations show us that if we follow the Kalman predictor, we can spot about four to nine failures on the observed mains during the next few months. We could also provide an accurate calculation for each separate month. Since this information has not been available yet (NA), we compare our results after it is available.

5.1 *Conclusion*

The aim of our article is to show possible ways of appropriate modelling even if there are only short and insufficient records of device failures available. We wanted to introduce the modelling of a pseudo-measure and also show that we are able to estimate possible development of the assessed system behaviour, similarly as Woch & Zieliński (2015).

The observed and later transformed data record is interesting because it shows that the course of the observed failure intensity decreases in time. It might indicate a change in the system properties which are manifested by reliability change. In

Table 3. Numerical values of KF, KS and KP of LLM.

Time	Original series	LLM		
		KF	KS	KP (KF/KS)
01/00	0.1935	0.1935	0.4414	NA
02/00	0.4138	0.304	0.4429	NA
03/00	0.8709	0.4949	0.4446	NA
04/00	0.7667	0.5643	0.4437	NA
...				
09/14	0.0	0.1953	0.2149	NA
10/14	0.2258	0.1976	0.2165	NA
11/14	0.3667	0.2102	0.2180	NA
12/14	0.3226	0.2186	0.2186	NA
01/15	NA			**0.2186**
02/15	NA			**0.2186**
03/15	NA			**0.2186**
04/15	NA			**0.2186**
05/15	NA			**0.2186**
06/15	NA			**0.2186**

Table 4. Numerical values of KF, KS and KP of LLM seasonal.

Time	Original series	LLM—seasonal			
		KF	KFseas	KS	KSseas
		KP	KPseas	KP	KPseas
01/00	0.1935	0.0161	0.1935	0.4729	0.4124
02/00	0.4138	0.2230	0.4137	0.4765	0.4696
03/00	0.8709	0.4444	0.8709	0.4811	0.5346
04/00	0.7667	0.5289	0.7667	0.4800	0.4667
...					
09/14	0.0	0.2080	0.1466	0.2148	0.1521
10/14	0.2258	0.2079	0.2260	0.2157	0.2289
11/14	0.3667	0.2186	0.2964	0.2166	0.2955
12/14	0.3226	0.2163	0.3375	0.2164	0.3375
01/15	NA	**0.2163**	**0.2142**	**0.2163**	**0.2142**
02/15	NA	**0.2163**	**0.2806**	**0.2163**	**0.2806**
03/15	NA	**0.2163**	**0.1372**	**0.2163**	**0.1372**
04/15	NA	**0.2163**	**0.2000**	**0.2163**	**0.2000**
05/15	NA	**0.2163**	**0.1785**	**0.2163**	**0.1785**
06/15	NA	**0.2163**	**0.1278**	**0.2163**	**0.1278**

this case, it would be advisable to replace-revitalize gradually all water pipe, thereby restoring its original properties.

Moreover, the results of prediction estimations enable us to expect a certain amount of failures on the observed main pipeline section. This can help with planning an operation system and a technical maintenance system.

In our future work, we are going to apply different approaches, different forms of dynamic models based on the observation of the structural changes of the observed pipeline. Some inspiring results, however, which we are going to use for that, have been already introduced in Hasilová & Vališ (2018), Pietrucha-Urbanik et al. (2017 a, b), Pilch et al. (2014), Rojek & Studzinski (2014), Romaniuk (2016), Vališ et al. (2015, 2017 a, b), Vališ & Pokora (2015) and Woch et al. (2015).

ACKNOWLEDGEMENTS

This paper has been prepared with the support of the Ministry of Defence of the Czech Republic, Partial Project for Institutional Development "PASVR" (K-110, Department of Tactics) and "MOBAUT" (K-202, Department of Combat and Special Vehicles), University of Defence, Brno.

REFERENCES

Ameri, R., Amiri-Bideshki, M., Saeid, A.B. & Hoskova-Mayerova, S. 2016. Prime filters of hyperlattices. *Analele Stiintifice ale Universitatii Ovidius Constanta* 24(2): 15–26.

Arthur, J., Attarian, A., Hamilton, F. & Tran, H. 2018. Nonlinear Kalman filtering for censored observations. *Applied Mathematics and Computation* 316: 155–166.

Ascher, H. 1970. Hazard functions, renewal rates and peril rates. *SAE Technical Papers*, article number 700627.

Bai, J. 1997a. Estimating Multiple Breaks One at a Time. *Econometric Theory* 13: 315–352.

Bai, J. 1997b. Estimation of a Change Point in Multiple Regression Models. *Review of Economics and Statistics* 79: 551–563.

Bai, J. & Perron, P. 1998. Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* 66: 47–78.

Bai, J. & Perron, P. 2003. Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics* 18: 1–22.

Bain, A. & Crisan, D. 2009. *Fundamentals of Stochastic Filtering*. London: Springer.

Bertein, J.C. & Ceschi, R.. 2010. *Discrete Stochastic Processes and Optimal Filtering*. London: Wiley.

Bhar, R. 2010. *Stochastic Filtering with Applications in Finance*. Singapore: World Scientific.

Eusgeld, I., Kröger, W., Sansavini, G., Schläpfer, M. & Zio, E. 2009. The role of network theory and object-oriented modeling within a framework for the vulnerability analysis of critical infrastructures. *Reliability Engineering and System Safety* 94(5): 954–963.

Fomin, V. 1999. *Optimal Filtering—Volume I: Filtering of Stochastic Processes*. Berlin: Springer Science.

Hasilová, K. & Vališ, D. 2018. Non-parametric estimates of the first hitting time of Li-ion battery. *Measurement* 113: 82–91.

Hokstad, P. 1997. The failure intensity process and the formulation of reliability and maintenance models. *Reliability Engineering and System Safety* 58(1): 69–82.

Hošková-Mayerová, Š., Chvalina, J. & Nezhad, A.D. 2013. General actions of hyperstructures and some applications. *Analele Stiintifice ale Universitatii Ovidius Constanta* 21(1): 59–82.

Hwang, H. & Lansey, K. 2017. Water Distribution System Classification Using System Characteristics and Graph-Theory Metrics. *Journal of Water Resources Planning and Management* 143(12). DOI 10.1061/(ASCE) WR.1943–5452.0000850.

Jaskolka, J. & Villasenor, J. 2017. An Approach for Identifying and Analyzing Implicit Interactions in Distributed Systems. *IEEE Transactions on Reliability* 66(2): 529–546.

Jung, D., Yoo, D.G., Kang, D. & Kim, J.H. 2016. Linear model for estimating water distribution system reliability. *Journal of Water Resources Planning and Management* 142(8): article number 04016022.

Kalman, R.E. 1960. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering* 82(1): 35–45.

Karamouz, M., Yaseri, K. & Nazif, S. 2017. Reliability-based assessment of lifecycle cost of urban water distribution infrastructures. *Journal of Infrastructure Systems* 23(2): article number 04016030.

Korkali, M., Veneman, J.G., Tivnan, B.F., Bagrow, J.P. & Hines, P.D.H. 2017. Reducing Cascading Failure Risk by Increasing Infrastructure Network Interdependence. *Scientific Reports* 7: article number 44499.

Krivtsov, V.V. 2007. Practical extensions to NHPP application in repairable system reliability analysis. *Reliability Engineering and System Safety* 92(5): 560–562.

Le Gat, Y. & Eisenbeis, P. 2000. Using maintenance records to forecast failures in water networks. *Urban Water* 2: 173–181.

MacLean, C.J. 1974. Estimation and testing of an exponential polynomial intensity function within the nonstationary Poisson process. *Biometrika* 61(1): 81–85.

NIST/SEMATECH. *e-Handbook of Statistical Methods*, USA, 2012. URL http://www.itl.nist.gov/div898/handbook/ (accessed 24.10.2017).

Otřísal, P., Florus, S. & Karkalić, R. 2017. Resistance of barrier materials against toxic compounds permeation and its evaluation in accordance with new European norms. In: *Conference Proceedings 3 "Applied Technical Sciences and Advanced Military Technologies" of the 23rd International Conference "The Knowledge-Based Organization"*. Sibiu: Nicolae Balcescu Land Forces Academy, p. 224–233. ISBN 978-973-153-275-2.

Pietrucha-Urbanik, K., Vališ, D. & Vintr, Z. 2017a. Perspective renewal model for water distributions systems. In: *Risk, Reliability and Safety: Innovating Theory and Practice*. London: Taylor & Francis Group, p. 1050–1055. ISBN 978-1-138-02997-2.

Pietrucha-Urbanik, K. & Studzinski, A. 2017b. Case Study of Failure Simulation of Pipelines Conducted in Chosen Water Supply System. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 19(3): 317–323.

Pilch, R., Szybka, J. & Tuszynska, A. 2014. Application of Factoring and Time-Space Simulation Methods for Assessment of the Reliability of Water-Pipe Networks. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 16(2): 253–258.

Rausand, M. & Høyland, A. 2004. *System Reliability Theory: Models, Statistical Methods, and Applications*. John Wiley: New Jersey, USA. ISBN 9780471471332.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rojek, I. & Studzinski, J. 2014. Comparison of Different Types of Neuronal Nets for Failures Location within Water-Supply Networks. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 16(1): 42–47.

Romaniuk, M. 2016. On Simulation of Maintenance Costs for Water Distribution System with Fuzzy Parameters. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 18(4): 514–527.

Rozovskii, B.L. 1990. Stochastic Evolution Systems – Linear Theory and Applications to Non-linear Filtering. Berlin: Springer Science.

Shen, B., Wang, Z. & Shu H. 2013. *Nonlinear Stochastic Systems with Incomplete Information*. London: Springer.

Todinov, M.T. 2012. Algorithms for minimising the lost flow due to failed components in repairable flow networks with complex topology. *International Journal of Reliability and Safety* 6(4): 283–310.

Vališ, D., Milazzo, M.F., Ancione, G. & Brkic, V.S. 2017a. Investigation of crane operation safety by analysing main accident causes. In: *Risk, Reliability and Safety: Innovating Theory and Practice*. London: Taylor & Francis Group, p. 74–80. ISBN 978-1-138-02997-2.

Vališ, D., Pietrucha-Urbanik, K. & Vintr, Z. 2017b. Water network condition assessment using analytic hierarchy process. In: *Safety and Reliability – Theory and Applications*. London: Taylor & Francis Group, p. 371–376. ISBN 978-1-138-62937-0.

Vališ, D., Žák, L., Vintr, Z. & Hasilová, K. 2016. Mathematical Analysis of Soot Particles in Oil Used as System State Indicator. In: *IEEM 2016*, p. 486–490. ISBN 978-1-5090-3665-3.

Vališ, D. & Pokora, O. 2015. Application of selected diffusion processes on system state assessment. In: *Safety and Reliability: Methodology and Applications*. London: Taylor & Francis Group, London, p. 911–916. ISBN 978-1-138-02681-0.

Wilt, B., Long, S. & Shoberg, T. 2016. Defining resilience: A preliminary integrative literature review. In S. Long, C. Downing, E.H. Ng & B. Nepal (eds), *Proceedings of the American Society for Engineering Management 2016 (ASEM)*, Huntsville: American Society for Engineering Management, p. 151–160.

Woch, M.K., Kurdelski, M. & Matyjewski, M. 2015. Reliability at the checkpoints of an aircraft supporting structure. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 17(3): 457–462.

Woch, M.K. & Vališ, D. 2017. Comparison of different methods for calculation of aircraft structure failure probability per single flight. In: *Risk, Reliability and Safety: Innovating Theory and Practice*. London: Taylor & Francis Group, p. 1406–1410. ISBN 978-1-138-02997-2.

Woch, M.K. & Zieliński, W. 2015. Development and reliability testing of a new filtering algorithm for noisy data from flight data recorder. In: *Safety and Reliability of Complex Engineered Systems*. London: CRC Press, p. 2141–2146.

Zeileis A., Kleiber C., Krämer W. & Hornik K. 2003. Testing and Dating of Structural Changes in Practice. *Computational Statistics and Data Analysis* 44: 109–123. doi:10.1016/S0167-9473(03)00030–6.

Zio, E. 2007. From complexity science to reliability efficiency: a new way of looking at complex network systems and critical infrastructures. *International Journal of Critical Infrastructures* 3(3/4): 488–508.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Self-healing networks: A theoretical approach to smart grids' resilience

A. Scala
Istituto Sistemi Complessi CNR, Università "Sapienza" Roma, Rome, Italy

F. Morone & H. Makse
City College of New York, New York, USA

ABSTRACT: To ensure a high quality of service to the users in the next generation of smart grids, self-healing capabilities are a crucial feature to be introduced. We show how distributed communication protocols can enrich complex networks with self-healing capabilities; an obvious field of applications are infrastructural networks distributing a commodity via a flow, like gas, water or electric power. We consider the case where the presence of redundant links allows to recover the connectivity of the system. We then introduce a theoretical framework to calculate the fraction of nodes still served for increasing levels of network damages. Such framework allows to analyse the interplay between redundancies and topology, a key point in improving the resilience of networked infrastructures to multiple failures.

1 INTRODUCTION

The functioning of any advanced society relies on networks that distribute commodities like water, gas, energy. The increasing urbanization and the accelerated growth of the size and the numbers of mega-cities (Facchini et al. 2017, Kennedy et al. 2015) requires such networks to be not only robust—i.e. able to sustain natural hazards, random failures or intentional attacks—but also to be resilient, i.e. to quickly recover from such hazards. Thus, the key feature to be implemented in network systems is *resilience* (Francis and Bekera 2014, Ganin et al. 2016), i.e. the ability of recovering an acceptable level of service in the face of faults, failures, accidents and attacks. In this paper, we described a simplified model for networks who are able to increase their resilience through self-healing capabilities while considering the effects both of the topology and of the redundancy on the capability of network recovery. After introducing the concept of resilience in sec.2, we will introduce in sec.3 a percolative model of self-healing reconnection. In sec.4 we will describe the analytic approaches needed to solve the problem at the mean-field level; we will compare the results of the model prediction with numerical simulation in 5.

2 RESILIENCE

Resilience is a complex property and is best expressed via the function describing the recovery history of a system (see Fig. 1, upper panel). Loosely speaking, there are three main factors characterizing resilience: the initial after-event state s^*, the recovery time τ and the recovery level s_C. In the usual resilience cycle, an accident happens at time t^* and the quality of service s drops quickly to value s^* that depends on the robustness (Cohen and Havlin 2010) and on the reliability (Chaturvedi 2016) of the network; after a recovery time τ that will depend on the restoration plan, service would have been restored up to a level s_C that in some cases can be equal or even higher than the initial one. In this paper we will concentrate on the restored quality of service s_C and we will introduce a self-healing percolation model for characterizing such quantity.

Self-healing is a crucial feature for implementing resilience in the smart networks of the future (Quattrociocchi et al. 2014). An example of already existing self-healing infrastructures are telecommunication networks, which rely on self-healing procedures based on routing protocols to restore traffic using redundant links (Bhandari 1998). This is a typical example of a strategy based on the redundancy in the interconnectivity of its components to ensure the continuity of a system; for example, when a hole is punched in a leaf, the remaining vessels are capable to sustain the extra flow necessary to keep the tissues alive (Katifori, Szöllősi, & Magnasco 2010). On the other hand, self-healing in infrastructural networks should be instead though as a constrained mechanism in which only a limited amount of resources is available.

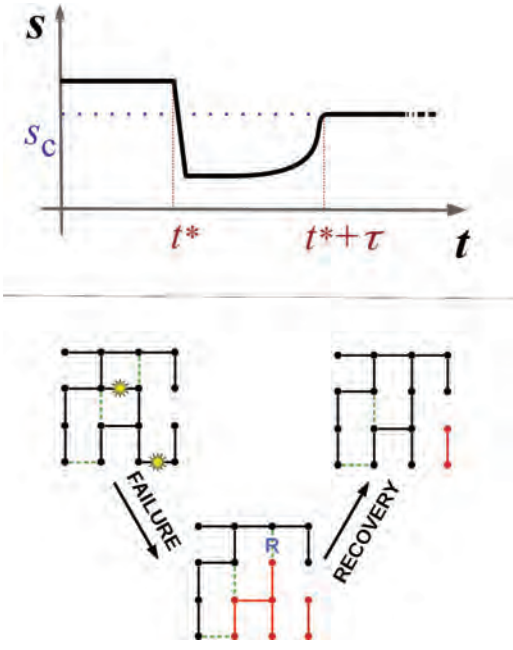


Figure 1. Relations among self healing and Resilience. Upper panel: Cartoon of a recovery function describing the resilience of a system; in this case we are plotting a metric s measuring the quality of service versus the time t. In the picture, an accident happens at time t^* and the s drops to a low value that depends on the robustness and on the reliability of the network; after a time τ, recovery plans will have restored the service s up to a level s_C that in some cases can be equal or even higher than the initial one. While the order of the restoration events determines the recovery time τ, any optimal restoration plan will achieve the same level of service s_C given the same resources. In this paper we will concentrate on the magnitude of s_C and not on the recovery time. Lower panel: Cartoon of a self-healing process in a distribution network.

Such a strategy is also common to material science where new polymeric compounds are capable of self healing due to the presence of small amounts of healing agents that gets released and activated upon cracking (White et al. 2001, Toohey et al. 2007).

While the security of large-scale infrastructural networks—like long range, high-voltage electric power networks—is based on redundancy, most of local (regional, rural and city level) networks due to economic constrains are tree-like objects with few redundant links (Quattrociocchi et al. 2014). To describe and characterize a core feature of the resilience of such local networks, we introduce a new percolation problem that models their capability of recovering connectivity upon random failures. Hence, our model for self-healing networks is inspired by local distribution networks like gas, water and medium/low voltage electric power networks. In real networks, cables and pipes—especially in urban areas—will likely follow the topology of the street networks. Also, tree-like networks allow operators to minimize costs (building physical links in the network requires huge investments) and to have an easy accountability of the consumptions. However, few redundant links must be present in order to recover the connectivity of the networks in case of accidents. Thus, after a link failure, some redundant links can be activated to recover (at least partially) the functionality of the network. In real infrastructures, such a procedure is often implemented manually, while in the smart networks of the future it should function automatically, possibly by embedding distributed algorithms automating the network recovery (Quattrociocchi et al. 2014). In the lower panel Fig. 1 we present a cartoon of the self-healing process in real networks.

3 MODEL

In our scenario we consider network systems distributing some utility; for sake of simplicity, we will consider a single node to be the source of the quantity to be distributed on the network. Examples of such network utilities are water, power, gas or oil pipelines or electric power distribution. At each instant of time, the topology of the network distributing the utility (the *active tree*) is assumed to be a tree; this assumption is partially verified in the above mentioned system; in particular, it is mostly verified in the case of electric power distribution (Pagani and Aiello 2011). In fact, such a structure meets the infrastructures' managers needs—i.e., to measure (for billing purposes) in an easy and precise way how much of a given quantity is served to any single node of the network. Finally, as a further simplification we will not take into account the magnitudes of flows—i.e., all links and sources are assumed to have infinite capacity—but we will focus on maximizing the connectedness of the system in order to serve as many nodes as possible. Notice that in the case of real flows, such assumption can be unrealistic since links and nodes in real networks have limits beyond which they become unoperational.

In order to implement our strategy and its self-healing capabilities, we consider the presence of *dormant* backup links—i.e., a set of links that can be switched on. Nodes are assumed to be able to communicate with their neighbors by means of a suitable distributed interaction protocol with a limited amount of knowledge: in particular, nodes are supposed to possess information only about the state of neighboring nodes connected either via active or via dormant links.

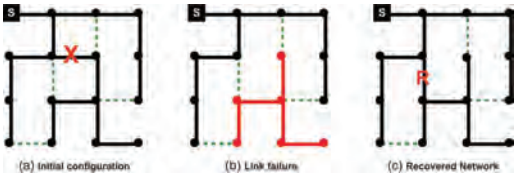


Figure 2. Example of the healing procedure: (Left Panel) In the initial state, the source node (filled square, upper left corner) is able to serve all 16 nodes through the links of the active tree. The 4 dashed lines (green online) represent dormant backup links that can be activated upon failure. The redundancy of the system is $p = 4/9$ as only 4 of the 9 possible backup links are present. The link marked with an X is the one that is going to fail. (Central panel) A single link failure disconnects all the nodes of a sub-tree; in the example, a sub-tree of 6 nodes (red online) is left isolated from the source—i.e., the system has a damage $\Delta = 6$). (Right Panel) By activating a single dormant backup link, the self-healing protocol has been able to recover connectivity for the whole system, in this case bringing back the number of served nodes at its maximum value 16. The link that has recovered the connectivity is marked with an R. Notice that in real networks, due to the physics of the flows and to the constraints on links and nodes capacities, not all the possible reconnection could be viable and on the contrary could lead to cascading behavior.

When either a node or a link failure occurs, all the nodes below the failure will disconnect from the active tree and become unserved. Such unserved nodes can now try to reconnect the active tree by waking up through the protocol some dormant backup links. Such a process will reconstruct a new active-tree that can restore totally or partially the flow, i.e. heal the system. Fig. 2 presents a graphical sketch of the healing procedure.

In the following, we will indicate with $T_s(G)$ a tree on the graph G routed in the node s; we indicate with $R \subseteq G - T$ the set of redundant links and with N the number of nodes. The set of redundant links will be also described by its adjacency matrix R_{ij} that takes the value 1 if a redundant link is present among nodes i and j, 0 otherwise. Moreover, we will describe the damages inflicted to the distribution tree by the matrix $q_{ij} = 0$ if link ij is removed, $q_{ij} = 1$ otherwise. The relevant metric for the effectiveness of a redundancy pattern R given an attack q is the fraction of served nodes FoS, i.e. the number of nodes connected to the root normalized by N.

4 METHODS

Let us first notice that a node i is connected to a node j if and only if a message from node i can reach node j. To solve such a problem, we use a cavity based approach for message passing (Mezard and Montanari 2009).

First of all let us define the set $S(i)$ as the set on nodes which are sons of node i on the original tree T (i.e. not considering the redundant edges): $S(i) \equiv \{k : k \text{ is a son of } i\}$. Moreover we will call $F(i)$ the unique father of node i. Finally, we define $R(i) = \{j : R_{ij} = 1\}$ the set of nodes connected to i via redundant links; obviously, we have $F(i) \notin R(i)$ and $S(i) \cap R(i) = \varnothing$. Then, let us consider a node i and a node $j \in S(i)$ which is one of the sons of node i on the original tree.

We introduce the following quantities:

- $d_{i\to j}$ is probability that node i is connected to the root s, when the son-node $j \in S(i)$ is absent from the tree.
- $u_{i\to j}$ is the probability that node i is connected to the root s, when the father-node $j = F(i)$ is absent from the tree.
- $r_{i\to j}$ is the probability that node i is connected to the root s when i and j connected by a redundant edge and j is absent from the tree.

We will derive first the recursive equation for $d_{i\to j}, j \in S(i)$. When j is absent from the graph, then i is connected to the root s if at least one among this possibilities is realized:

1. its father $F(i)$ is connected to the root. The probability that such event does NOT happen is

$$\pi_1 = \left(1 - q_{iF(i)} d_{F(i)\to i}\right)$$

2. one of its sons $S(i)$ – except j – is connected to the root. The probability that such event does NOT happen is

$$\pi_2 = \prod_{k \in S(i) \setminus j} (1 - q_{ki} u_{k\to i})$$

3. one of the neighbours connected to i via a redundant link is connected to the root when i is absent. The probability that such event does NOT happen is

$$\pi_3 = \prod_{k=1}^{N} (1 - R_{ik} r_{k\to i})$$

The total probability that i is connected to the root when its son j is absent is thus given by $1 - \pi_1 \times \pi_2 \times \pi_3$, i.e.

$$d_{i\to j} = 1 - \left(1 - q_{iF(i)} d_{F(i)\to i}\right) \times \prod_{k \in S(i) \setminus j} (1 - q_{ki} u_{k\to i}) \prod_{m=1}^{N} (1 - R_{im} r_{m\to i}) \quad (1)$$

Following the same procedure we can write the equation for $u_{i\to j}$:

$$u_{i\to j} = 1 - \prod_{k\in S(i)} (1 - q_{ki} u_{k\to i}) \times \prod_{m=1}^{N} (1 - R_{im} r_{m\to i}) \qquad (2)$$

Finally, the equation for $r_{i\to j}$ is:

$$r_{i\to j} = 1 - \left(1 - q_{iF(i)} d_{F(i)\to i}\right) \times \prod_{k\in S(i)} (1 - q_{ki} u_{k\to i}) \prod_{m=1, m\neq j}^{N} (1 - R_{im} r_{m\to i}) \qquad (3)$$

Equations (1–3) constitute the self consistent equations of the problem. They are valid for any given realization of the random tree and redundant links. Equations (1–3) can be considered as describing messages running on the edges of the tree and on redundant edges, and they can be solved by simple iteration.

Once a solution to Eqs. (1–3) has been found, we can compute the total probability p_i that node i is connected to the root

$$p_i = 1 - \left(1 - q_{iF(i)} d_{F(i)\to i}\right) \times \prod_{k\in S(i)} (1 - q_{ki} u_{k\to i}) \prod_{m=1}^{N} (1 - R_{im} r_{m\to i}) \qquad (4)$$

Disregarding correlations, we can estimate the average fraction of served nodes *FoS* as

$$FoS = \frac{\sum_{i=1}^{N} p_i}{N} \qquad (5)$$

5 RESULTS

We now compare the results of our eqs. (1–3) to the numerical simulation of our self-healing procedure. In particular, we are considering the case of networks generated by random trees at which a fraction α of random recovery links are added. After an initial fraction f of links in the tree is deleted at random, recovery links are activated whenever they reduce the number of connected components. Finally, the *FoS* is calculated by checking the fraction of nodes connected to the origin via the surviving links plus the redundant. Notice that when averaging eq. (1–3) to solve the mean-field equations, we are using $\langle q_{ij} \rangle = f$ and $\langle R_{i,j} \rangle = \alpha/(N-1)$.

First, we perform simulations on random trees on a complete graph of 1000 nodes. The random trees are generated according a flat-sampling procedure in the space of possible trees (Broder 1989, Aldous 1990, Wilson 1996). Edges on the tree are removed at random; the fraction of removed edges is indicated as $f = 1 - \sum_{i,j} q_{ij} / |T|$, where $|T| = N - 1$. We parametrize with α the population of redundant link existing among two nodes i and j where the link i,j does not belong to T; thus, R_{ij} are random variables on $\{0,1\}$ where $R_{ij} = 1$ with probability $\alpha/(N-1)$. In each simulation, a random tree is generated, a random vertex is assumed to be the source, a fraction α of redundant links are added to the tree and a fraction f of links is erased at random from the tree; then the fraction of *FoS* of sites connected to the root (via links either in $T - q$ or in R) is calculated. The results for the average *FoS* are presented on Fig. 3.

We then consider the average behavior of Eqs. (1–3) over the randomness in the model, i.e. over the possible realizations of the q_j and C_j. Performing the average we obtain the following self consistent equations:

$$d = 1 - \left[1 - (1-f)d\right] e^{-\alpha r} \times \sum_k \frac{k(k-1)}{\langle k^2 \rangle - \langle k \rangle} P(k) [1 - (1-f)u]^{k-2} \qquad (6)$$

$$u = 1 - e^{-\alpha r} \frac{k}{\langle k \rangle} P(k) [1 - (1-f)u]^{k-1}$$

$$r = 1 - \left[1 - (1-f)d\right] e^{-\alpha r}$$

$$\sum_k \frac{k}{\langle k \rangle} P(k) [1 - (1-f)u]^{k-1}$$

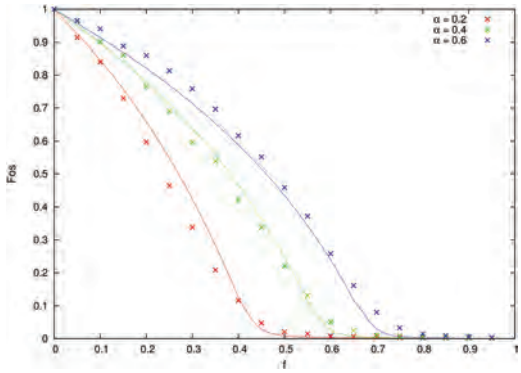


Figure 3. Comparison of simulations and analytical approximation. Depicted are the curves for the fraction of served nodes *FoS* (i.e. nodes connected at the origin after self-healing) versus the initial fraction of failed links f. Crosses correspond to average values of the *FoS* obtained by simulating random trees of 1000 nodes for different levels of redundancy $\alpha = 0.2$ (red X), $\alpha = 0.4$ (green X) and $\alpha = 0.6$ (blue X); the size of the symbols is of the order of thrice the error bars. As expected, curves are monotonically decreasing with f and monotonically increasing with α. Full lines correspond to the predictions of our analytical approximations eqs. (7).

where $P(k)$ is the degree distribution (i.e. the probability of having k neighbours) of nodes in the tree which are not leaves. In Fig. 3 we compare the theoretical predicted *FoS* obtained by averaging eq. (5) with the results of numerical simulations.

6 CONCLUSION

In this paper we have discusses the recovery of networks upon a minimal self-healing procedure that exploits the presence of redundant edges to recover the connectivity of the system. Our scenario is inspired by real-world distribution networks that are, often for economic reasons, tree-like and in the meantime are also often provided with alternative backup links that can be activated in case of malfunctioning; as an example, this is the case for low-voltage distribution networks (ENEL 2011).

Our model, albeit schematic, is realistic in the sense that it could be readily and easily implemented with the current technologies. In fact, routing protocols represent a vast available source of distributed algorithms able to maintain the connectivity of a system. Therefore, our scheme could be implemented by coupling an ICT network to current infrastructures. Our case is an example in which interdependencies enhance the resilience instead of introducing catastrophic breakdowns (Buldyrev et al. 2010). However, since in we are assuming that links capacities are infinite, our model applies to cases where the network is not stressed. For real flow networks, the physics of the flows and their constrains can forbid some of the possible reconnections that, in the worst cases, could even lead to cascading failures like he one observed in model power grids (Pahwa et al. 2014). Moreover, we are considering the case of a single source; in the case of multiple sources, leaving the system disconnected (islanding) could even improve its robustness (Mureddu et al. 2016). Notice that since our model predicts only the final level of service of a network, also timescales are an important element that should be introduced to allow for a characterization of the full resilience curve of the system.

In this paper we have introduced the cavity equation describing our model and compared an analytical approximation for the average values of the connectivity under random failures with numerical simulations. We find a promising accordance of such an approximation with numerical results opening the field for future investigations.

The first direction to be investigated is the study of our approximation when both the fraction of failures f and the fraction of redundant links α is small. In fact, in real systems the number of concurrent failures is small: this is the reason at the basis of the $N-1$ criterion in engineering. On the same pace, for economic reasons also the fraction α of redundant links is doomed to be small: hence, our approximation is valid for small f's and α's, where analytical expressions could be linearly expanded. On the other hand, if failures are not independent but happen in a correlated and perhaps catastrophic way like in cascading events (Pahwa et al. 2014), the approximation must be enhanced to hold for the entire f range; work in progress is done in this direction.

Most importantly, cavity equations (1–3) can be applied to single networks with a given topology and set of redundant links to calculate the *FoS* as an alternative method to numerical simulations. Having a set of closed equations allows then to easily analyze scenarios in which the set of redundant links is varied: coupling our approach with the introduction of a cost function for the links is of importance for the design of networks since it would allow to optimizing the redundancy. As an example, numerical simulations on planar topologies suggest that a very effective strategy to strengthen planar networks is to add long range links (Quattrociocchi et al. 2014): since such links are overly expensive in networks like electric distribution, the feasibility of such a strategy depends on cost-benefit analysis about their implementation of physical long-range links in *PNI*s. A further direction of study would be to consider the effects of more detailed structural characteristics on the dynamics of the system (D'Agostino et al. 2012). However, it is important to remember that in optimizing the system the cost of the links is as much important as the increase in resilience of the system. In fact, a simple constrain like keeping the number of redundant links fixed would lead to very unrealistic topologies in which the source is at the center of a star regardless of the length of the links (Quattrociocchi et al. 2014).

REFERENCES

Aldous, D.J. (1990, November). The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discret. Math. 3*(4), 450–465.

Bhandari, R. (1998). *Survivable Networks: Algorithms for Diverse Routing*. Norwell, MA, USA: Kluwer Academic Publishers.

Broder, A. (1989). Generating random spanning trees. In 30*th Annual Symposium on Foundations of Computer Science*, pp. 442–447.

Buldyrev, S.V., R. Parshani, G. Paul, H.E. Stanley, & S. Havlin (2010). Catastrophic cascade of failures in interdependent networks. *Nature 464*(7291), 1025–1028.

Chaturvedi, S.K. (2016). *Network reliability: measures and evaluation* (1 ed.). Performability engineering series. John Wiley.

Cohen, R. & S. Havlin (2010, August). *Complex Networks: Structure, Robustness and Function*. Cambridge University Press.

D'Agostino, G., A. Scala, V. Zlatic, & G. Caldarelli (2012). Robustness and assortativity for diffusion-like processes in scale-free networks. *EPL (Europhysics Letters) 97*(6), 68006.

ENEL (2011). Private communication.

Facchini, A., C. Kennedy, I. Stewart, & R. Mele (2017). The energy metabolism of megacities. *Applied Energy 186*, *Part 2*, 86–95. Energy and Urban Systems.

Francis, R. & B. Bekera (2014). A metric and frameworks for resilience analysis of engineered and infrastructure systems. *Reliability Engineering & System Safety 121*, 90–103.

Ganin, A.A., E. Massaro, A. Gutfraind, N. Steen, J.M. Keisler, A. Kott, R. Mangoubi, & I. Linkov (2016, January). Operational resilience: concepts, design and analysis. *Scientific Reports 6*, 19540–.

Katifori, E., G.J. Szöllősi, & M.O. Magnasco (2010, Jan). Damage and fluctuations induce loops in optimal transport networks. *Phys. Rev. Lett. 104*, 048704.

Kennedy, C.A., I. Stewart, A. Facchini, I. Cersosimo, R. Mele, B. Chen, M. Uda, A. Kansal, A. Chiu, K.-g. Kim, C. Dubeux, E. Lebre La Rovere, B. Cunha, S. Pincetl, J. Keirstead, S. Barles, S. Pusaka, J. Gunawan, M. Adegbile, M. Nazariha, S. Hoque, P.J. Marcotullio, F. Gonzalez Otharan, T. Genena, N. Ibrahim, R. Farooqui, G. Cervantes, & A.D. Sahin (2015). Energy and material flows of megacities. *Proceedings of the National Academy of Sciences 112*(19), 5985–5990.

Mezard, M. & A. Montanari (2009). *Information, Physics, and Computation*. New York, NY, USA: Oxford University Press, Inc.

Mureddu, M., G. Caldarelli, A. Damiano, A. Scala, & H. Meyer-Ortmanns (2016, October). Islanding the power grid on the transmission level: less connections for more security. *Scientific Reports 6*, 34797–.

Pagani, G.A. & M. Aiello (2011, Sept). Towards decentralization: A topological investigation of the medium and low voltage grids. *Smart Grid, IEEE Transactions on* 2(3), 538–547.

Pahwa, S., C. Scoglio, & A. Scala (2014, January). Abruptness of cascade failures in power grids. *Sci. Rep. 4*, –.

Quattrociocchi, W., G. Caldarelli, & A. Scala (2014, 02). Self-healing networks: Redundancy and structure. *PLoS ONE 9*(2), e87986.

Toohey, K.S., N.R. Sottos, J.A. Lewis, J.S. Moore, & S.R. White (2007, August). Self-healing materials with microvascular networks. *Nat Mater 6*(8), 581–585.

White, S.R., N.R. Sottos, P.H. Geubelle, J.S. Moore, M.R. Kessler, S.R. Sriram, E.N. Brown, & S. Viswanathan (2001, February). Autonomic healing of polymer composites. *Nature 409*(6822), 794–797.

Wilson, D.B. (1996). Generating random spanning trees more quickly than the cover time. In *Proceedings of the 28th annuan ACM Symposium on the Theory of Computing*, pp. 296–303. ACM.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A new hybrid Bayesian network approach for modeling reliability

F. Petiet, O. François & L. Bouillaut
Grettia, Ifsttar, University of Paris Est, France

ABSTRACT: In this paper, a hybrid discrete-continuous Graphical Duration Models is proposed. Since the interest of the Weibull density was demonstrated for reliability analysis, this paper focuses on the use of Weibull densities for modeling sojourn times in each state of the system. This extension of the standard Graphical Duration Model (GDM) requires a specific structure that we call Weibull-Hybrid Graphical Duration Models (W-HGDM). The main contribution of this study lays in the proposal of a specific inference algorithm for such hybrid networks. Finally, comparisons of reliability estimation will be proposed for both standard GDM and W-HGDM.

1 INTRODUCTION

Reliability analysis is an integral part of system design and operating, especially for systems performing critical applications.

A wide range of works about reliability analysis is available in the literature.

Most of the time, the system failure is caused by the failure of one or more components. In this case, it is possible to use statistical distributions as the exponential distribution to model the lifetime. The Weibull Distribution (Weibull 1951) allows to describe each phase of the bathtub curve. In (Bertholon 2001) a new modeling of aging is proposed.

These distributions do not allow to focus on the dynamics of degradation. They cannot be applied to a system that is repaired when it fails. They cannot be used to assess the expected number of failures during the warranty period, or maintain a minimum mission reliability, or determine when to replace or overhaul a system. These methods are often called "classics".

Dynamic models explicitly take into account the temporal aspect modeling the evolution of the degradation of the system over time by stochastic methods (Cocozza-Thivent 1997).

Rather than considering the different components of the system, it is also possible to consider the whole system. Several methods are commonly used to model the different states of a dynamic system, in order to analyze its reliability, such as Markov Chain, Petri net, or Bayesian Network (Demri 2009).

Recent works have shown the interest of using Bayesian Networks (BN) (Jensen 1996) in the field of reliability. For example (Boudali & Dugan 2005) shows how to model the reliability of a complex system using Bayesian networks.

Weber and Jouffe (2003) explains how to use dynamic Bayesian networks (DBN) (Murphy 2002) to study the reliability of a multi-state system that depends on a certain context.

However DBN suppose that the sojourn time in each state are exponentially distributed, whereas most of the industrial applications underline non Markovian behaviors. In these cases, a Markovian degradation process modeling can introduce non negligible biases.

So an original Bayesian Network structure was proposed, named Graphical Duration Models (GDM), in order to fit systems whose sojourn time in each state are not necessary exponentially distributed (Donat, Leray, Bouillaut, & Aknin 2010). A GDM is characterized by a duration variable, allowing the use of any kind of distribution for modeling the sojourn time in each state of the considered system.

However the complexity is directly related to the size of the discrete space of the sojourn time variable. It can induce some technical problems in terms of storage capacity and computation time. A solution could be to consider a continuous duration variable.

In the theory of Bayesian Networks, there are several approaches that contain both continuous and discrete variables.

In hybrid Bayesian networks, where both discrete and continuous variables appear simultaneously, it is possible to apply inference schemes similar to those for discrete variables. The first model that allowed exact inference in hybrid networks was based on the Conditional Gaussian (CG) distribution (Lauritzen 1992).

The restriction of discrete variables with continuous parents in CG may also be partially lifted using logit or probit function, generalized in the multinomial case by the softmax function (Murphy 1999).

Another way to lift the restriction caused by discrete variables with continuous parents is the using of a mixture of exponentials (Koller, Lerner, & Angelov 1999), but the inference is approximated.

The Mixture of Truncated Exponential model has been introduced in (Moral, Rumí, & Salmerón 2001). The advantage with respect to CG is that discrete nodes with continuous parents are allowed and inference can be exact (Cobb & Shenoy 2006).

Given that hybrid approaches thus exist in the Bayesian Network framework, such a perspective could be envisaged in a Graphical Duration Model.

According to expert feedback, sojourn-time variables follow a Weibull distribution in many systems (Weibull 1951).

Our goal is to integrate sojourn-time variables following a Weibull distribution in a graphical duration model by proposing a new approach.

First, this paper will briefly describe the formalisms of the Bayesian Networks of the Graphical Duration Models. We then introduce the our proposed formalism named Weibull-Hybrid Graphical Duration Models (W-HGDM) and the inference algorithm associated.

Then a toy system is introduced. Before some conclusions and prospects, a comparison of reliability analysis results, obtained from both Graphical Duration Models and Weibull-Hybrid Graphical Duration Models approaches, will be done.

2 INTRODUCTION OF THE FORMALISM

2.1 *Bayesian Network*

Bayesian Networks (Jensen 1996) are mathematical tools relying on the probability theory and the graph theory. They allow to qualitatively and quantitatively represent uncertain knowledge.

Bayesian Networks are Probabilistic Graphical Models that allow to intuitively represent the distribution of a set of random variables $\mathbf{X} = (X_1,...,X_N)$. Basically, a BN is defined as a pair $\mathcal{M} = (\mathcal{G},(p_n)_{1\le n\le N})$. $\mathcal{G} = (\mathbf{X},\mathcal{E})$ is a Directed Acyclic Graph in which each node i is associated to a random variable X_i, that takes its values in a finite and countable set $\mathcal{X}_i$, and in which each directed arc $(i,j)\in\mathcal{E}$ represents dependencies between random variables X_i and X_j. $(p_n)_{1\le n\le N}$ is a set of Conditional Probability Distributions (CPD) such that each p_n denote the conditional probability distribution associated to random variable X_n given its parents X_{pa_n}, pa_n referring to the sequence of parents indices of the random variable X_n in $\mathcal{G}$.

The conditional independence relationships introduced by the arcs of the graph enable to factor the joint probability distribution of the set of random variables $\mathbf{X}$ as follows:

$$P(\mathbf{X}) = P(X_1,\ldots,X_N) = \prod_{n=1}^{N} P(X_n \mid X_{\mathrm{pa}_n}) \qquad (1)$$

Besides, tools have been developed to automatically learn the structure and the parameters of the graph and those of the CPD from complete or incomplete data or if a priori knowledge is available (e.g. expert opinion) (Neapolitan 2003).

Using BN is particularly interesting because of the possibility to propagate knowledge through the network. Indeed, various inference algorithms can be used to compute marginal probabilities of the system variables. One of the most classical inference procedures relies on the use of a junction tree (Lauritzen & Spiegelhalter 1988). Nevertheless, in our experiments, we use the elimination algorithm (Dechter 1999). This choice is motivated by the simplicity and the efficiency of the method.

2.2 *Dynamic Bayesian Network*

Inspired by the formalism of classical BN, the Dynamic Bayesian Networks (DBN) framework (Murphy 2002) allowed to unify many approaches from modeling of time series such as hidden Markov models. A DBN aims to model the probability distribution of a random variables set $(\mathbf{X}_t)_{1\le t\le T} = (X_{1,t},\ldots,X_{N,t})_{1\le t\le T}$. It consists of a pair of Bayesian Networks $(\mathcal{M}_1,\mathcal{M}_{\rightarrow})$. $\mathcal{M}_1$ defines the prior distribution $P(X_{1,1},\ \ldots,X_{N,1})$ as in (1). $\mathcal{M}_{\rightarrow}$ defines the transition model which describes the dependencies between variables in slice $t-1$ and variables in slice t, i.e. the distribution of $\mathbf{X}_t \mid \mathbf{X}_{t-1}$.

$$\begin{aligned} P(\mathbf{X}_t \mid \mathbf{X}_{t-1}) &= P(X_{1,t},\ldots,X_{N,t} \mid X_{1,t-1},\ldots,X_{N,t-1}) \\ &= \prod_{n=1}^{N} P(X_{n,t} \mid X_{\mathrm{pa}_{n,t}}) \end{aligned} \qquad (2)$$

where $\mathrm{pa}_{n,t}$ refers to the sequence of parents indices of the random variable $X_{n,t}$ in the graph of $\mathcal{M}_{\rightarrow}$.

A Dynamic Bayesian Network is actually a static Bayesian Network, that is repeated several times. So DBN inherit the convenient properties of static BN, particularly with regard to learning and inference.

The state of the system at future time $t+1$, $\mathbf{X}_{t+1}$, is decided by the system state at the current time t, $\mathbf{X}_t$, and does not depend on the state at earlier time

instants 1,…,$t-1$; and the conditional probability distributions of $\mathcal{M}_{\rightarrow}$ do not depend on the slice t; so the distribution of time spent in each state is geometric. Whereas most of industrial applications have not this behavior. Such a modeling can introduce non negligible biases in the estimations.

2.3 *Graphical Duration Model*

A Graphical Duration Model (Donat, Leray, Bouillaut, & Aknin 2010) relies on the two following variables: the system state X_t and the duration variable S_t describing the time spent in any system state (remaining sojourn time) (Fig. 1).

Firstly, the Conditional Probability Distribution associated with the distribution of the initial system state is defined as follows, over the discrete and finite domain $\Omega_X = \{1,\ldots,N_X\}$:

$$P(X_1 = i) = P_{X_1}(i) \tag{3}$$

The initial sojourn-time CPD gives the distributions for each initial state. This CPD is defined over the discrete and finite domain $\Omega_S = \{1,\ldots,N_S\}$:

$$P(S_1 = k \mid X_1 = i) = P_{S_1}(i,k) \tag{4}$$

Then, it is necessary to define the system state and the Sojourn-time transition CPDs.

A transition occurs if and only if $S_{t-1} = 1$.

$$P(X_t = j \mid X_{t-1} = i, S_{t-1} = 1) = Q^{sys}(i,j) \tag{5}$$

where Q^{sys} is a $N_X.N_X$ matrix, called static system transition matrix.

A new sojourn-time is selected according to the following CPD:

$$P(S_t = k \mid X_t = i, S_{t-1} = 1) = F^{sys}(i,k) \tag{6}$$

where F^{sys} is a $N_X.N_S$ matrix.

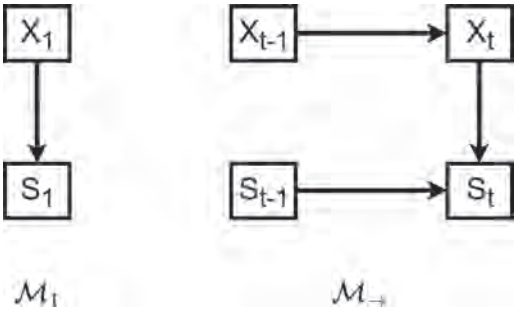


Figure 1. Discrete duration graphical model.

While there is no transition, the system deterministically remains in the previous state i:

$$P(X_t = j \mid X_{t-1} = i, S_{t-1} \geq 2) = I(i,j) \tag{7}$$

and the sojourn-time in the current state is decreased deterministically by one unit:

$$P(S_t = k \mid X_t = i, S_{t-1} = k' \geq 2) = \begin{cases} 1 & \text{if } k = k'-1 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

2.4 *Reliability computation*

Let assume that the set of system state Ω_X is partitioned into two sets $\mathcal{U}$ and $\mathcal{D}$ respectively for "up" states and for "down" states (i.e. OK and failure situations).

The discrete-time system reliability is defined as the function R: $\mathbf{N}^* \mapsto [0,1]$ where $R(t)$ represents the probability that the system has always stayed in an up state until moment t, i.e.

$$R(t) = P(X_1 \in \mathcal{U};\ldots;X_t \in \mathcal{U}).$$

In addition, it is possible to derive some interesting metrics such as the failure rate or the MTTF (Pham 2006) from the reliability definition.

Hence, this issue boils down to an inference problem, i.e. to the computation of $P(X_1 \in \mathcal{U};\ldots;X_t \in \mathcal{U})$.

The system cannot repair itself. The failure state is absorbing. The reliability is then equal to the availability, ie

$$R(t) = P(X_t \in \mathcal{U}) = 1 - P(X_t \in \mathcal{D}).$$

Then the reliability computation is reduced to the computation of the marginal distribution of X_t.

2.5 *Limitation of GDM*

The larger N_S is, the more accurate the representation of the sojourn-time distribution. On the other hand, choosing a too large N_S value will have immediate consequences on both the space needed to store the Conditional Probability Distributions, i.e. $N_X.N_S$ values, and the time complexity of inference.

That can induce some technical problems in terms of storage capacity and computation time. A solution could be to consider a continuous duration variable. With c the number of parameters needed by the chosen distribution, the CPD will be constituted of $N_X.c$ values, so the space needed to store the continuous CPD is smaller than if they are discrete.

Since the interest of the Weibull density was demonstrated for reliability analysis (Weibull

1951), this paper focuses on the use of Weibull densities for modeling sojourn times in each state of the system.

3 WEIBULL-HYBRID GRAPHICAL DURATION MODELS

In the following paragraphs, we propose a new formalism to represent the evolution of a dynamic system over time, in order to estimate its reliability. This particular model is called Weibull-Hybrid Graphical Duration Models (W-HGDM). In the following, we will describe its conditional probability distributions, and an inference algorithm in order to compute the reliability.

3.1 *Generalities*

The collection $(X_t)_{1\leq t\leq T}$ represents the system state over a sequence of length T. The collection $(S_t)_{1\leq t\leq T}$ represents the remaining time before a system state modification. More clearly, we refer to the random variable S_t as the remaining sojourn time in the current system state. These variables are called duration variables or sojourn-time variables. A Weibull-Hybrid Graphical Duration Model is illustrated in Figure 2.

Expressions (3), (4), (5), (6), (7) and (8) become expressions (9), (10), (12), (13), (14) and (15).

The Conditional Probability Distribution associated with the distribution of initial system state must be defined as follows, over the discrete and finite domain $\Omega_X = 1,\ldots,N_X$:

$$P(X_1 = i) = p_i^{ini} \tag{9}$$

The initial sojourn-time CPD gives the distribution for each initial state, it is defined over the continuous domain $\Omega_S =]0,+\infty[$:

$$f_{S_1|X_1=i} = f_{\mathcal{W}_{\alpha_i,\beta_i}} \tag{10}$$

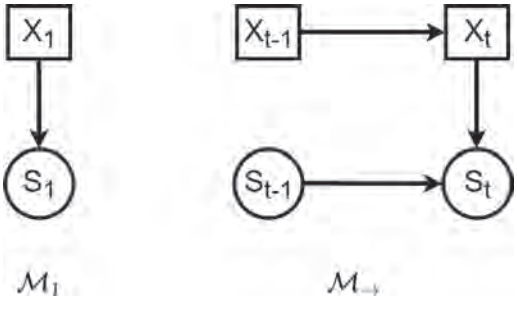


Figure 2. Continuous duration graphical model.

with

$$f_{\mathcal{W}_{\alpha,\beta}}(s) = \frac{\beta}{\alpha}\left(\frac{s}{\alpha}\right)^{\beta-1} e^{-\left(\frac{s}{\alpha}\right)^{\beta}} \tag{11}$$

where α is the scale parameter and β the shape parameter.

Then, it is necessary to define the system state and the sojourn-time transition CPD:

A transition occurs if and only if $S_{t-1} < 1$.

$$P(X_t = j \mid X_{t-1} = i, S_{t-1} < 1) = Q^{sys}(i,j) \tag{12}$$

where Q^{sys} is a $N_X.N_X$ matrix, called static system transition matrix.

A new sojourn-time is selected according to the following CPD:

$$f_{S_t=s|S_{t-1}<1,X_t=i} = f_{\mathcal{W}_{\alpha_i,\beta_i}}(s) \tag{13}$$

While there is no transition, the system deterministically remains in the previous state *i*:

$$P(X_t = j \mid X_{t-1} = i, S_{t-1} > 1) = I(i,j) \tag{14}$$

and the sojourn-time in the current state is decreased deterministically by one unit:

$$f(S_t = s \mid X_t = i, S_{t-1} = s' > 1) = \begin{cases} 1 & \text{if } s = s' - 1 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

3.2 *Inference*

As shown in 2.4, the reliability of a non-repairable system is equal to the availability of the system:

$$\begin{aligned} R(t) &= P(X_t \in \mathcal{U}) \\ &= 1 - P(X_t \in \mathcal{D}) \\ &= 1 - \sum_{\substack{X_1,\ldots,X_{t-1} \\ S_1,\ldots,S_{t-1}}} P(X_t \in D, X_{t-1},\ldots,X_1, S_{t-1},\ldots,X_1) \\ &= 1 - \sum_{\substack{X_1,\ldots,X_{t-1} \\ S_1,\ldots,S_{t-1}}} P(X_t \in D \mid \mathrm{pa}(X_t)) \end{aligned} \tag{16}$$

We have to compute successively the following distributions:

- $\Phi_{t,u} = P(X_t|X_u, S_{u-1})$ for $u \in [\![2;t-1]\!]$ and $\Phi_{t,1} = P(X_t|X_1)$.
- $\Lambda_{t,u} = P(X_t|X_u, S_{u-1})$ for $u \in [\![2;t]\!]$ and $\Lambda_{t,1} = P(X_t)$

The required distributions are expressed as follows:

For $u \in [\![2;t-1]\!]$

$$\Phi_{t,u} = \int_0^{\infty} P(X_t \mid X_u, S_u = s_u) f(S_u = s_u \mid X_u, S_{u-1}) \mathrm{d}s_u = \int_0^{\infty} \Lambda_{t,u+1} \times f(S_u = s_u \mid X_u, S_{u-1}) \mathrm{d}s_u \quad (17)$$

and

$$\Lambda_{t,u} = \sum_{x \in \Omega_{X_u}} P(X_t \mid X_u = x, S_{u-1}) P(X_u = x \mid X_{u-1}, S_{u-1}) = \sum_{x \in \Omega_{X_u}} \Phi_{t,u} \times P(X_u = x \mid X_{u-1}, S_{u-1}) \quad (18)$$

For $u = 1$

$$\Phi_{t,1} = \int_0^{\infty} P(X_t \mid X_1, S_1 = s_1) f(S_1 = s_1 \mid X_1) \mathrm{d}s_u = \int_0^{\infty} \Lambda_{t,2} f(S_1 = s_1 \mid X_1) \mathrm{d}s_u \quad (19)$$

and

$$\Lambda_{t,1} = \sum_{x \in \Omega_{X_1}} P(X_t \mid X_1 = x) P(X_1 = x) = \sum_{x \in \Omega_{X_1}} \Phi_{1,1} P(X_1 = x) \quad (20)$$

The elements required by the computation (17) are given by (18), (13) and (15); and those required by (19) are given by (18) and (10).

The elements required by the computation (18) are given by (17), (12) and (14); and those required by (20) are given by (17) and (9).

Equations 17, 18, 19, 20 define an iterative algorithm, that give at the end: $\Lambda_{t,1} = P(X_t)$, that allows to have the value of the reliability:

$$R(t) = 1 - P(X_t \in \mathcal{D}) = \sum_{x \in \mathcal{U}} P(X_t = x) \quad (21)$$

The computation in (17) and (18) can be developed as follows:

In (17):

$$P(X_t \mid X_{t-q}, S_{t-q-1} = s) = \quad (22)$$

$$\begin{cases} Q^{(q)} & \text{if} \quad s < 1 \\ Q^{sys} * Q^{(q-1)} & \text{if} \quad 1 < s < 2 \\ \quad \cdots & \\ Q^{sys} * Q^{(q-j)} & \text{if} \quad j < s < j+1 \\ \quad \cdots & \\ Q^{sys} & \text{if} \quad q < s < q+1 \\ I & \text{if} \quad q+1 < s \end{cases} \quad (23)$$

with $Q^{(q)}(x,y) = \sum_{k=1}^{q} \int_{k-1}^{k} f_{\mathcal{W}_{\alpha_x,\beta_x}}(s)\mathrm{d}s \times [\, Q^{sys} Q^{(q-k)}\,](x,y) + \int_q^{\infty} f_{\mathcal{W}_{\alpha_x,\beta_x}}(s)\mathrm{d}s \times I(x,y)$.

In (18):

$$P(X_t \mid X_{t-q-1}, S_{t-q-1} = s) = \quad (24)$$

$$\begin{cases} Q^{sys} * Q^{(q)} & \text{if} \quad s < 1 \\ Q^{sys} * Q^{(q-1)} & \text{if} \quad 1 < s < 2 \\ \quad \cdots & \\ Q^{sys} * Q^{(q-j)} & \text{if} \quad j < s < j+1 \\ \quad \cdots & \\ Q^{sys} & \text{if} \quad q < s < q+1 \\ I & \text{if} \quad q+1 < s \end{cases} \quad (25)$$

$Q^{sys} \cdot Q^{(q)}$ is actually the natural transition of the system between q slices.

4 ILLUSTRATIONS

The W-HGDM used has been implemented in MATLAB® environment, using the open source Bayes Net Toolbox (BNT). The objective of this study is the inference part, not the learning part. So toy systems are created. The parameters of Conditional Probability Distributions are not learned from data, they are determined.

4.1 *Test*

To validate this algorithm, we consider a two-state system, one of the states being "up", the other being "down". The reliability can be written as follows:

$$\begin{aligned} R_{sys}(t) &= P(X_1 \in \mathcal{U} \cap \ldots \cap X_t \in \mathcal{U}) \\ &= P(S_1 \geq t \mid X_1 \in \mathcal{U}) P(X_1 \in \mathcal{U}) \\ &= 1 - F_{\mathcal{W}_{\alpha_{\mathcal{U}},\beta_{\mathcal{U}}}}(t) \\ &= e^{-\left(\frac{t}{\alpha_{\mathcal{U}}}\right)^{\beta_{\mathcal{U}}}} \end{aligned} \quad (26)$$

The application of the inference algorithm presented in 3.2 to such a model is supposed to return a Weibull distribution.

The Figure 3 shows that the estimated reliability is exactly the reliability given by the Weibull distribution.

4.2 *Comparison with standard GDM*

Let's illustrate our approach using a GDM modeling the behavior of a 3-states production machine, i.e. $\Omega_X = \{1,2,3\}$, with $\mathcal{D} = \{3\}$. The transition rate between system states is given in Table 1. The param-

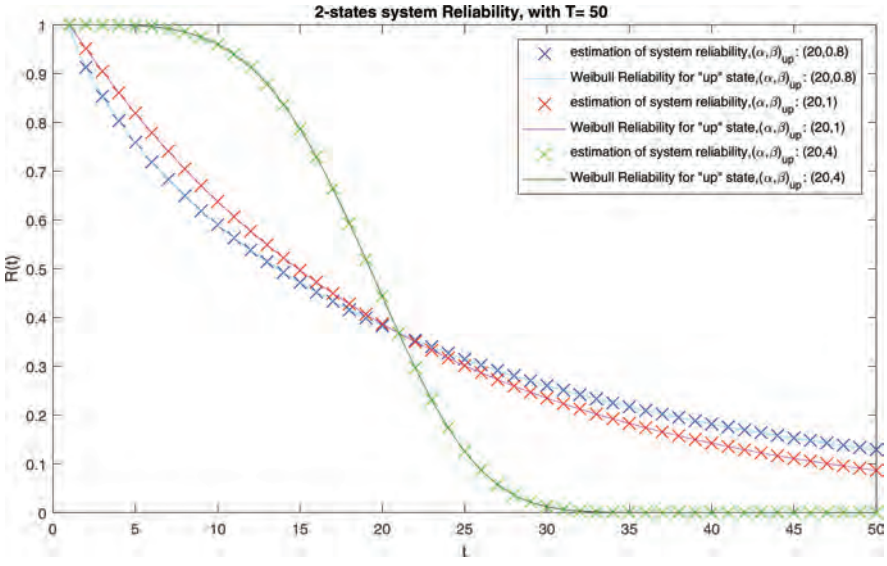


Figure 3. Reliability of a 2 states-sytem and of Weibull distributions.

Table 1. System transition conditional probability table.

	State 1	State 2	State 3
State 1	0	0.9	0.1
State 2	0	0	1
State 3	0	0	1

Table 2. Sojourn-time distribution for each state.

	α	β
State 1	30	1
State 2	20	1

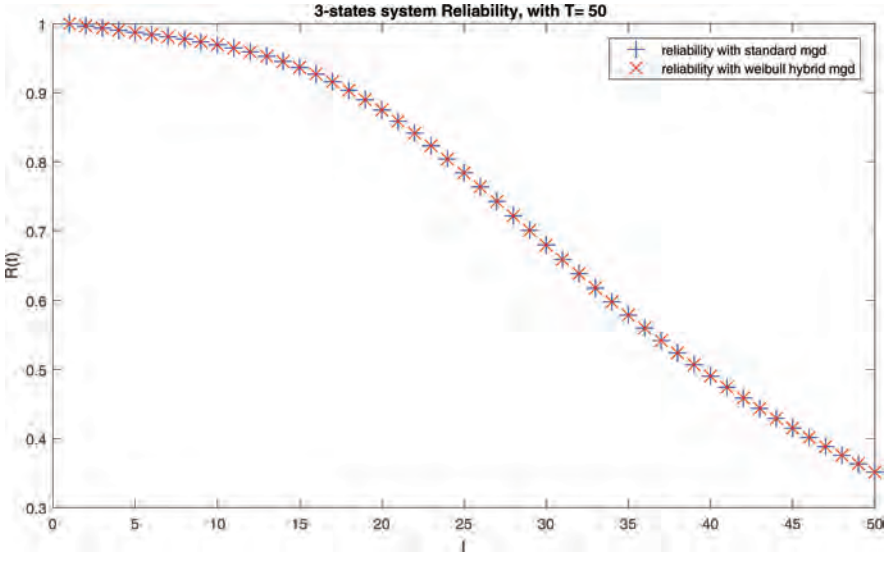


Figure 4. Comparison between discrete GDM and W-HGDM.

eter of the Weibull distribution that characterize the sojourn-time variable when a transition occurs are given in Table 2. The algorithm presented in subsection 3.2 was used to make the computations.

A standard GDM is then built. The state system variable of this standard GDM model has the same static system transition matrix as the W-HGDM (Table 1). The distribution of sojourn-time variable of this standard GDM model is constructed by discretizing the continuous Weibull distribution, with the same parameters as the W-HGDM (Table 2), and with $N_S = 150$ (Donat, Bouillaut, Aknin, Leray, & Bondeux 2008).

The method presented in 3.2 has been used to compute reliability estimations presented in Figure 4.

We obtain the same results with both methods.

5 CONCLUSIONS

In this paper, a new hybrid approach for the duration model in which sojourn-time follows a Weibull distribution, called Weibull-Hybrid Graphical Duration Model, has been proposed. An associated inference algorithm has been therefore developed and introduced allowing the estimation of the reliability of the system in such models.

Finally, first results introduced in this paper underline the feasibility of the proposed approach in terms of inference accuracy. This validates the interest we have had in proposing such a hybrid approach for the modeling of degradation dynamics. Indeed, the expected advantage of W-HGDM leads mainly in the non discretization of the sojourn variable in GDM that can become a drag to reliability computation for complexity reasons.

Now, many things can be done and investigated through this new approach. The first perspective is to go further in the algorithm validation, by comparing the complexities and the speeds between standard GDM and W-HGDM. We also have planned to add some nodes to the presented W-HGDM, such as context variables, and, of course, maintenance actions. The long-term goal of this study is to improve VirMaLab (Virtual Maintenance Laboratory), a generic decision support tool developed by the GRETTIA (Bouillaut, Aknin, Donat, & Bondeux 2011) in order to evaluate, optimize and compare maintenance strategies for discrete state space system, by allowing sojourn-time variables to follow a Weibull distribution.

REFERENCES

Bertholon, H. (2001). *Une modélisation du vieillissement*. Ph.D. thesis, Université Joseph Fourier, Grenoble.

Boudali, H. & J.B. Dugan (2005). A discrete-time Bayesian network reliability modeling and analysis framework. *Reliability Engineering and System Safety 87*, 337–349.

Bouillaut, L., P. Aknin, R. Donat, & S. Bondeux (2011). VirMaLab—A generic approach for optimizing maintenance policies of complex systems. In *WCRR 2011–9th World Congress on Railway Research*, Lille, France, pp. 11p. WCRR. WCRR 2011–9th World Congress on Railway Research, Lille, FRANCE, 22/05/2011–26/05/2011.

Cobb, B.R. & P.P. Shenoy (2006). Inference in hybrid Bayesian networks with mixtures of truncated exponentials. *International Journal of Approximate Reasoning 41*(3), 257–286.

Cocozza-Thivent, C. (1997). *Processus stochastiques et fiabilité des systèmes*. Mathématiques et Applications. Springer Berlin Heidelberg.

Dechter, R. (1999). Bucket elimination: A unifying framework for reasoning. *Artificial Intelligence 113*(1), 41–85.

Demri, A. (2009). *Reliability estimation of mechatronic systems using functional and dysfunctional analysis*. Ph.D. thesis, Université d'Angers.

Donat, R., L. Bouillaut, P. Aknin, P. Leray, & S. Bondeux (2008, June). Specific graphical models for analyzing the reliability. In *2008 16th Mediterranean Conference on Control and Automation*, pp. 621–626.

Donat, R., P. Leray, L. Bouillaut, & P. Aknin (2010, January). A dynamic Bayesian network to represent discrete duration models. *Neurocomputing 73*(4–6), 570–577.

Jensen, F.V. (1996). *Introduction to Bayesian Networks* (1st ed.). Secaucus, NJ, USA: Springer-Verlag, New York, Inc.

Koller, D., U. Lerner, & D. Angelov (1999). A general algorithm for approximate inference and its application to hybrid Bayes nets. pp. 324–333.

Lauritzen, S.L. (1992). Propagation of probabilities, means and variances in mixed graphical association models. *Journal of the American Statistical Association 87*, 1098–1108.

Lauritzen, S.L. & D.J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Methodological) 50*(2), 157–224.

Moral, S., R. Rumí, & A. Salmerón (2001). Mixtures of truncated exponentials in hybrid Bayesian networks. In *Proceedings of the 6th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, ECSQARU '01, London, UK, pp. 156–167. Springer-Verlag.

Murphy, K.P. (1999). A variational approximation for Bayesian networks with discrete and continuous latent variables. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, UAI'99, San Francisco, CA, USA, pp. 457–466. Morgan Kaufmann Publishers Inc.

Murphy, K.P. (2002). *Dynamic Bayesian Networks: Representation, Inference and Learning*. Ph.D. thesis, University of California, Berkeley.

Neapolitan, R.E. (2003). *Learning Bayesian Networks*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Pham, H. (2006). *System Reliability Concepts*, pp. 9–75. London: Springer London.

Weber, P. & L. Jouffe (2003, June). Reliability modelling with dynamic Bayesian networks. pp. 57–62. IFAC.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of applied mechanics 18*(3), 293–297.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A method of road network vulnerability identification taking into account travelers' heterogeneous risk attitudes

B. Lv & J. Zhang
Department of Railway Information Engineering, School of Information Science and Technology of Southwest Jiaotong University, Emeishan, China

Y.L. Liu
Department of Production and Quality Engineering, Norwegian University of Science and Technology, Trondheim, Norway

Y. Huang
Chengdu Metro Operation Co. Ltd., Chengdu, China

ABSTRACT: The analysis of road network vulnerability is a challenging and important research subject in the field of transportation reliability engineering because of the complex coupling relationships among travelers, vehicles, roads and environment. In view of the deficiency of existing researches on the description of travelers' risk averse and bounded rational behavior characteristics, both random utility theory and random regret theory are used to describe travelers' decision-making behaviors. Then a traffic assignment model expressed by variational inequality is constructed, in which travelers' risk aversion and bounded rationality as well as resultant heterogeneity of travelers' decision-making behaviors are explicitly taken into account. Subsequently, the vulnerability index of road network based on accessibility is defined, and a vulnerability identification model is built, then corresponding heuristic algorithm is also proposed. The example results show that the consequences of link closures could be misjudged and the vulnerability rankings could be misidentified, if ignoring the effects of the congestion levels of road network, travelers' perception errors and regret aversion degrees, as well as travelers' route choice decision criteria. Therefore, it is necessary to capture the travelers' behavior characteristics in the process of vulnerability analysis.

1 INTRODUCTION

1.1 *Background*

Robust road traffic networks have been regarded as one of the preconditions for a high quality of life. However, a traffic network can be vulnerable to various natural or man-made disasters. For example, adverse weather such as heavy snow and flooding could severely degrade the network performance. Although the occurrence probability of these major incidents is low, their consequences could be sufficiently large to generate a major problem that threats remedial actions. Therefore, it is vital to understand the potential vulnerability of traffic networks under such major incidents, so as to manage their risks.

A key issue in the vulnerability analysis is to identify the critical links of a network, where the failures of those links would have the most serious impacts on the whole network (Chen et al., 2012, Yin and Xu, 2010). After identifying the critical links, the network robustness can be enhanced through reinforcing these identified critical links or constructing new alternative parallel paths (Matisziw and Murray, 2009).

Modeling travelers' behavioral responses to link failures is another key issue involved in critical link identification (Chen et al., 2012). Subjected to link failures, the high degree of demand uncertainty and/or link capacity degradation will inevitably yield travel time variability, and consequently imposes additional disutility on travelers. Many empirical studies have revealed the significant influence of travel time uncertainty on travelers' route choice behavior (Liu et al., 2004, Shao et al., 2006, Wu and Nie, 2011). Travelers under travel time uncertainty tend to choose reliable shortest path, not only dependent on travel time saving, but also on reduction of travel time variability. This risk averse behaviors under travel time uncertainty have received considerable attention (Shao et al., 2006, Lo et al., 2006, Siu and Lo, 2008). However, most of the previous studies adopt Expected Utility Theory (EUT) and/or Random Utility Theory (RUT) to quantify travelers' perceptions of network uncertainty, and assume that the travelers are homogeneous. It is well known that both EUT and RUT are based on the assumption that travelers are absolutely rational

when making route choice decisions. In reality, however, part of travelers' behaviors may be bounded rational, and can be influenced by his or her personality, psychological state, and environmental elements, etc. It goes without saying that the usefulness of EUT or RUT as a descriptive model of choice behaviour has been fiercely debated both inside and outside the transportation domain. It is particularly interesting to note that Regret Theory (RT) (Loomes and Sugden, 1982) and its extended Random Regret Theory (RRT) (Chorus, 2014), being widely considered one of the most prominent competitors of both EUT and RUT in the behavioral decision sciences, has been virtually ignored in the route choice domain. RT or/and RRT postulate that when choosing, people anticipate and try to avoid the situation where a non-chosen alternative outperforms the chosen one (which would cause post-decision regret). It is a pity that, to our best knowledge, travelers' risk aversion and bounded rationality as well as resultant travelers' heterogeneous behavior responses due to link closures have not yet been considered simultaneously in the studies of critical link identification.

In view of the above, this study proposed a method of road network vulnerability identification taking into account travelers' risk aversion, bounded rationality and heterogeneity simultaneously under travel time variations subjected to link failures. We assume that the total travel demand comprises of two parts, completely rational travelers and bounded rational travelers. A new vulnerability index based on accessibility is introduced to evaluate the consequences of a link closure with consideration of their effects.

1.2 *Outline*

The remainder of this paper is organized as follows. In Section 2, a stochastic mixed user equilibrium model is built. It follows with the definition of vulnerability index based on accessibility for evaluating the consequences of a link closure with consideration of travelers' risk aversion, bounded rationality and heterogeneity. In Section 4, numerical examples by means of the Nguyen and Dupuis network is provided to demonstrate the proposed model. Finally, the conclusions are given in Section 5, together with future research directions.

2 A STOCHASTIC MIXED USER EQUILIBRIUM MODEL

2.1 *Distributions of travel times under link closures*

Consider a road network represented by a strongly connected graph G = (N, A), where N and A are the sets of nodes and links respectively. Let W denote the set of Origin-Destination (OD), R_w represent the set of paths between OD pair w, w ∈ W.

Because of link failures, the high degree of demand uncertainty and/or link capacity degradation will inevitably lead to travel time uncertainty. Assume that a link travel time is a random variable. Let T_a represent the travel time on link a. Furthermore, assume that T_a follows the normal distribution with mean value t_a and variance $\rho_a t_a$, where ρ_a represents the variation coefficient of T_a. The mean travel time t_a can be described by the following BPR (Bureau of Public Roads) function:

$$t_a = t_a^0\left(1+\beta\left(\frac{v_a}{c_a}\right)^n\right), a \in A \tag{1}$$

where t_a^0 is the free-flow travel time on link a, v_a and c_a are the flow and the capacity on link a respectively, β and n are the constant parameters of BPR function.

Let the travel time on path k between the OD pair w be represented as T_k^w, which can be calculated according to the relationship of link and path, as follows:

$$T_k^w = \sum_{a\in A} T_a \delta_{a,k}^w, k \in R_w, w \in W \tag{2}$$

where $\delta_{a,k}^w$ is the link-path incidence variable, $\delta_{a,k}^w$=1 if link a is on path k, otherwise $\delta_{a,k}^w = 0$.

In order to simplify the problem, it is assumed that link travel times are independent of each other. Because a path is composed of several independent links, according to the Central Limit Theorem, a path travel time should obey the normal distribution approximately. According to equations (1) and (2), the mean and standard deviation of T_k^w can then be expressed as below:

$$t_k^w = \sum_{a\in A} \delta_{a,k}^w t_a^0\left(1+\beta\left(\frac{v_a}{c_a}\right)^n\right), k \in R_w, w \in W \tag{3}$$

$$\sigma_{T_k^w} = \sqrt{\sum_{a\in A} \delta_{a,k}^w \rho_a t_a}, k \in R_w, w \in W \tag{4}$$

where t_k^w and $\sigma_{T_k^w}$ are respectively mean and standard deviation of T_k^w.

The empirical researches show that, the travelers not only want to save travel time, but also hope to avoid the risk caused by the travel time uncertainty. Therefore, Lo et al. (2006) put forward the concept of Travel Time Budget (TTB), which is used to describe the route choice behavior of travelers avoiding travel risk. Let $\xi_k^w(\omega)$ be the TTB of route k between OD pair w, which can be expressed as follows:

$$\xi_k^w(\omega) = t_k^w + \Phi^{-1}(\omega)\sigma_{T_k^w}, k \in R_w, w \in W \tag{5}$$

where $\Phi^{-1}(\cdot)$ is the inverse function of standard normal distribution, ω denotes the reliability parameter reflecting the probability that the actual trip time is within the TTB.

1.2 *Travel decision model for completely rational travelers based on RUT*

Considering the travelers' risk aversion in route choice decisions, it is assumed that the completely rational travelers use TTB as their route choice criterion. In addition, because of the travelers may not have perfect information on the travel time distributions, the travelers' perception errors on the travel times should be also taken into account. Therefore, according to RTU, the travel disutility of the completely rational travelers can be regarded as a random variable, which can be expressed as follows:

$$U_{w,k}^{\mathrm{CR}} = \xi_k^w(\omega) + \zeta_{w,k}^{\mathrm{CR}},\ k \in R_w, w \in W \tag{6}$$

where $U_{w,k}^{\mathrm{CR}}$ and $\zeta_{w,k}^{\mathrm{CR}}$ respectively represent the perceived travel disutility and perception error when the completely rational traveler' choose the route k between the OD pair w.

Furthermore, assuming that the perception error $\zeta_{w,k}^{\mathrm{CR}}$ are identically and independently Gumbel distributed random variables with mean zero, then the probability that a completely rational traveler chooses the route k between OD pair w can be described as follows:

$$p_{w,k}^{\mathrm{CR}} = \frac{\exp\left(-\theta^{\mathrm{CR}}\xi_k^w(\omega)\right)}{\sum_{r\in R_w}\exp\left(-\theta^{\mathrm{CR}}\xi_r^w(\omega)\right)},\ k \in R_w, w \in W \tag{7}$$

where $\theta^{\mathrm{CR}} > 0$ is the perception error parameter of the completely rational travelers which is used to measure the degrees of travelers' perception errors. It is noted that a higher θ^{CR} means smaller perception errors.

1.3 *Travel decision model for bounded rational travelers based on RRT*

In reality, due to the influences of information conditions, personality, preferences and other factors, not all travelers' route choice behaviors are completely rational, and some travelers show bounded rationality when choosing a route. In this paper, RRT is used to describe this phenomenon. In contrast with RUT, which postulates that a route's disutility is a function of its own performance only, RRT postulates that in addition, the performance difference with the competing route codetermines a route's disutility. In other words: RRT assumes that the traveler is regret averse in travel decision-making, that is, when the traveler make choice among all alternatives, the total value of regret that the current choice scheme compares with the other alternatives is considered, and the scheme with minimum value of regret is chosen.

Based on the above analysis, assume that the bounded rational travelers are risk averse and regret averse, who take perceived regret value as their route choice criterion. According to RRT, the travel disutility of the bounded rational travelers can be represented as below:

$$U_{w,k}^{\mathrm{BR}} = u_{w,k}^{\mathrm{BR}} + \zeta_{w,k}^{\mathrm{BR}},\ k \in R_w, w \in W \tag{8}$$

where $U_{w,k}^{\mathrm{BR}}, u_{w,k}^{\mathrm{BR}}$ and $\zeta_{w,k}^{\mathrm{BR}}$ respectively represent the perceived travel disutility (the perceived regret value), mean of the perceived travel disutility (mean of the perceived regret value) and perception error when the bounded rational traveler' choose the route k between the OD pair w.

Suppose that the bounded rationality travelers use the TTB as the absolute travel disutility, a specifically functional form of $u_{w,k}^{\mathrm{BR}}$ that satisfies RRT requirements can be stated as follows:

$$u_{w,k}^{\mathrm{BR}} = \sum_{m\neq k, m\in R_w} \exp\left(\gamma(\xi_k^w(\omega) - \xi_m^w(\omega))\right),\ k \in R_w, w \in W \tag{9}$$

where $\gamma > 0$ is a regret aversion parameter. When γ increases, regret becomes more and more important.

Similarly, assuming that the perception error $\zeta_{w,k}^{\mathrm{BR}}$ are identically and independently Gumbel distributed random variables with mean zero, then the probability that a bounded rational traveler chooses the route k between OD pair w can be described as follows:

$$p_{w,k}^{\mathrm{BR}} = \frac{\exp\left(-\theta^{\mathrm{BR}}u_{w,k}^{\mathrm{BR}}\right)}{\sum_{r\in R_w}\exp\left(-\theta^{\mathrm{BR}}u_{w,r}^{\mathrm{BR}}\right)},\ k \in R_w, w \in W \tag{10}$$

where $\theta^{\mathrm{BR}} > 0$ is the perception error parameter of the completely rational travelers which is used to measure the degrees of travelers' perception errors.

1.4 *Stochastic mixed user equilibrium model*

It is assumed that there are two kinds of travelers in the road network, completely rational travelers and bounded rational travelers, respectively. The completely rational travelers use the perceived travel time budget as their travel disutility and the bounded rational traveler use the perceived regret value as their travel disutility. In the process of routes selection, two kinds of travelers try to find

the routes with the minimum travel disutility. The network is called to achieve the stochastic mixed user equilibrium state when each type of travelers can not decrease their travel disutility by unilaterally changing the routes.

According to the principle of stochastic user equilibrium (Sheffi, 1985; Huang, 1994), the network equilibrium condition can be expressed as follows:

$$f_{w,k}^{\mathrm{CR}} = q_w^{\mathrm{CR}} p_{w,k}^{\mathrm{CR}}, k \in R_w, w \in W \tag{11}$$

$$f_{w,k}^{\mathrm{BR}} = q_w^{\mathrm{BR}} p_{w,k}^{\mathrm{BR}}, k \in R_w, w \in W \tag{12}$$

$p_{w,k}^{\mathrm{CR}}$ and $p_{w,k}^{\mathrm{BR}}$ in formulas (11) and (12) are determined by formulas (7) and (10) respectively, and the following conditions of flow conservation constraint are required:

$$q_w = q_w^{\mathrm{CR}} + q_w^{\mathrm{BR}} = \sum_{k \in R_w} f_{w,k}^{\mathrm{CR}} + \sum_{k \in R_w} f_{w,k}^{\mathrm{BR}}, w \in W \tag{13}$$

$$f_{w,k} = f_{w,k}^{\mathrm{CR}} + f_{w,k}^{\mathrm{BR}}, k \in R_w, w \in W \tag{14}$$

$$f_{w,k}^{\mathrm{CR}} \geq 0, k \in R_w, w \in W \tag{15}$$

$$f_{w,k}^{\mathrm{BR}} \geq 0, k \in R_w, w \in W \tag{16}$$

$$v_a^{\mathrm{CR}} = \sum_{w \in W} \sum_{k \in R_w} f_{w,k}^{\mathrm{CR}} \delta_{a,k}^w, a \in A \tag{17}$$

$$v_a^{\mathrm{BR}} = \sum_{w \in W} \sum_{k \in R_w} f_{w,k}^{\mathrm{BR}} \delta_{a,k}^w, a \in A \tag{18}$$

$$v_a = v_a^{\mathrm{CR}} + v_a^{\mathrm{BR}}, a \in A \tag{19}$$

where $q_w^{\mathrm{CR}}, q_w^{\mathrm{BR}}$ and q_w respectively denote the completely rational travelers demand, the bounded rational travelers demand and the total demand between OD pair w; $f_{w,k}^{\mathrm{CR}}, f_{w,k}^{\mathrm{BR}}$ and $f_{w,k}$ respectively represent the completely rational travelers flow, the bounded rational travelers flow and total flow on route k between OD pair w; $v_a^{\mathrm{CR}}, v_a^{\mathrm{BR}}$ and v_a respectively represent the completely rational traveler flow, the bounded rational travelers flow and total flow on the link a.

The equilibrium conditions of (11) and (12) can be described by the following equivalent variational inequality (VI) model.

Find $f_{w,k}^{\mathrm{CR}*}, f_{w,k}^{\mathrm{BR}*} \in \Omega$, making it satisfy the condition:

$$\sum_{w \in W} \sum_{k \in R_w} \left(\xi_k^w(\omega) + \frac{1}{\theta^{\mathrm{CR}}} \ln \frac{f_{w,k}^{\mathrm{CR}*}}{q_w^{\mathrm{CR}}} \right) \left(f_{w,k}^{\mathrm{CR}} - f_{w,k}^{\mathrm{CR}*} \right) + \sum_{w \in W} \sum_{k \in R_w} \left(u_{w,k}^{\mathrm{BR}} + \frac{1}{\theta^{\mathrm{BR}}} \ln \frac{f_{w,k}^{\mathrm{BR}*}}{q_w^{\mathrm{BR}}} \right) \left(f_{w,k}^{\mathrm{BR}} - f_{w,k}^{\mathrm{BR}*} \right) \geq 0, \quad \forall f_{w,k}^{\mathrm{CR}}, f_{w,k}^{\mathrm{BR}} \in \Omega \tag{20}$$

where the superscript "*" is used to designate the solution of the VI problem; Ω is the feasible set for route flows satisfying the constraint condition formulated as formulas (13)–(16).

Let $\mathbf{f}^{\mathrm{CR}}$ and $\mathbf{f}^{\mathrm{BR}}$ denote the column vectors composed of $\left\{ f_{w,k}^{\mathrm{CR}}, k \in R_w, w \in W \right\}$ and $\left\{ f_{w,k}^{\mathrm{BR}}, k \in R_w, w \in W \right\}$ respectively, $\mathbf{v}^{\mathrm{CR}}(\mathbf{f})$ and $\mathbf{v}^{\mathrm{BR}}(\mathbf{f})$ denote the column vectors composed of $\left\{ \xi_k^w(\omega) + \frac{1}{\theta^{\mathrm{CR}}} \ln \frac{f_{w,k}^{\mathrm{CR}}}{q_w^{\mathrm{CR}}}, k \in R_w, w \in W \right\}$ and $\left\{ u_{w,k}^{\mathrm{BR}} + \frac{1}{\theta^{\mathrm{BR}}} \ln \frac{f_{w,k}^{\mathrm{BR}}}{q_w^{\mathrm{BR}}}, k \in R_w, w \in W \right\}$ respectively. Because $\mathbf{v}^{\mathrm{CR}}(\mathbf{f})$ and $\mathbf{v}^{\mathrm{BR}}(\mathbf{f})$ are continuous with respect to $\mathbf{f}^{\mathrm{CR}}$ and $\mathbf{f}^{\mathrm{BR}}$ respectively, and the feasible set Ω is a bounded closed convex set, there exists at least one solution of VI problem expressed by formula (20) according to the variational inequality theorem (Facchinei and Pang, 2003).

2 VULNERABILITY IDENTIFICATION MODEL OF ROAD NETWORK

The identification of key links and their criticality are important question in the road network vulnerability evaluation. The key links refers to the links that the failure will result a significant impact on the network vulnerability. In the literature, various vulnerability indices have been proposed to evaluate the consequences of link closures. For example, Jenelius et al. (2006) used the increase of the generalized cost, weighted by the demand, as a vulnerability measure to a link closure. Chen et al. (2007) proposed the utility-based accessibility index to take account of travelers' behavioral responses to the link closure. In this paper, the road network vulnerability is evaluated by the change of road network accessibility, and then identify the key links.

Accessibility can be defined as the convenience for travelers to arrive at a destination from a origin to a destination by the certain way in the certain period of time (Taylor et al., 2006, Taylor, 2008). The accessibility index of a single OD pair can be defined as follows:

$$AC_w = \frac{\sum_{w \in W} \left(q_w^{\mathrm{CR}} + q_w^{\mathrm{BR}} \right)}{\sum_{w \in W} \sum_{k \in R_w} \left(f_{w,k}^{\mathrm{CR}} + f_{w,k}^{\mathrm{BR}} \right) \xi_k^w(\omega)}, w \in W \tag{21}$$

where AC_w expresses the accessibility between the OD pair w.

As shown in the formula (21), it can be seen that when the travel time of a unit OD travel demand is higher, the accessibility index of the OD pair is lower, which indicates that the traveling relative convenience is lower.

According to the formula (21), the accessibility index of a road network can be defined as follows:

$$AC(G)=\frac{\sum_{w\in W}\left(q_w^{\mathrm{CR}}+q_w^{\mathrm{BR}}\right)AC_w}{\sum_{w\in W}\left(q_w^{\mathrm{CR}}+q_w^{\mathrm{BR}}\right)} \tag{22}$$

where $AC(G)$ express the accessibility of road network G.

In this paper, the road network vulnerability is evaluated by the relative change of road network accessibility before and after link failures. The road network vulnerability index under the failure of link a is defined as follows:

$$VUL_a=\frac{AC_0(G)-AC_a(G)}{AC_0(G)},a\in A \tag{23}$$

where $AC_0(G)$ is the road network accessibility under normal condition; $AC_a(G)$ indicates the road network accessibility under the failure of road link a (removing road link a from road network). Obviously, it reflects the change of road network accessibility caused by the failure of link a.

The failure of single link is the lightest case of link failures in a road network. It is also the basis for studying the multi-link failures. For simplicity, this paper selects the situation of single link failure to identify the road network vulnerability.

Based on the above analysis, the specific vulnerability evaluation scheme is given below:

Step 1: Calculate road network accessibility $AC_0(G)$ under normal condition. Method of successive average (MSA) algorithm is used to calculate the equilibrium route flows and every route's TTB under normal conditions. Subsequently, the road network accessibility under normal conditions is calculated according to formula (21) and formula (22).

Step 2: Remove each link from the road network in turn, and the road network accessibility $AC_a(G)$ after the removal of link a is calculated according to the formulas (21) and (22).

Step 3: Calculate the road network vulnerability VUL_a. According to $AC_0(G)$ and $AC_a(G)$ obtained from Step 1 and Step 2, the road network vulnerability index VUL_a after the failure of link a can be calculated in turn by formula (23). If the removal of link a would result the road network is no longer connected directly, then the link a is immediately identified as the most critical section, making it $VUL_a=\infty$.

Step 4: Identify the critical links. Sort the VUL_a in descending order, and $rank_a$ express the order value of VUL_a (namely critical degree), and the key links are selected from the first N links with the minimum values of the $rank_a$, where N is the number of key links set in advance.

Where the steps of MSA algorithm are as following:

Step 1: Initialization. Set error parameter ε = 0.01 and iteration counter n = 1; initialize the route flows. The reasonable initial route flows of the completely rational travelers are set as $\mathbf{f}^{\mathrm{CR}(n)}=\left\{f_{w,k}^{\mathrm{CR}(n)},k\in R_w,w\in W\right\}$, and the initial route flows of the bounded rational travelers are set as $\mathbf{f}^{\mathrm{BR}(n)}=\left\{f_{w,k}^{\mathrm{BR}(n)},k\in R_w,w\in W\right\}$.

Step 2: Based on the current path flow $\mathbf{f}^{\mathrm{CR}(n)}$ and $\mathbf{f}^{\mathrm{BR}(n)}$, the vector of TTB $\xi^{(n)}=\left\{\xi_k^{w(n)},k\in R_w,w\in W\right\}$ and the vector of mean regret value $\mathbf{u}^{\mathrm{BR}(n)}=\left\{u_{w,k}^{\mathrm{BR}(n)},k\in R_w,w\in W\right\}$ are calculated respectively.

Step 3: Seek the iterative direction of route flows for the completely rational travelers $\mathbf{g}^{\mathrm{CR}(n)}$ and that of the bounded rational travelers $\mathbf{g}^{\mathrm{BR}(n)}$. Let $\mathbf{g}^{\mathrm{CR}(n)}=\mathbf{f}^{\mathrm{CR}(n)'}-\mathbf{f}^{\mathrm{CR}(n)}$, $\mathbf{g}^{\mathrm{BR}(n)}=\mathbf{f}^{\mathrm{BR}(n)'}-\mathbf{f}^{\mathrm{BR}(n)}$. where $\mathbf{f}^{\mathrm{CR}(n)'}=\left\{f_{w,k}^{\mathrm{CR}(n)'},k\in R_w,w\in W\right\}$, $f_{w,k}^{\mathrm{CR}(n)'}=q_w^{\mathrm{CR}}\exp\left(-\theta^{\mathrm{CR}}\xi_{w,k}^{(n)}(\omega)\right)\Big/\sum_{r\in R_w}\exp\left(-\theta^{\mathrm{CR}}\xi_{w,r}^{(n)}(\omega)\right),k\in R_w$, $w\in W$; $\mathbf{f}^{\mathrm{BR}(n)'}=\left\{f_{w,k}^{\mathrm{BR}(n)'},k\in R_w,w\in W\right\}$, $f_{w,k}^{\mathrm{BR}(n)'}=q_w^{\mathrm{BR}}\frac{\exp\left(-\theta^{\mathrm{BR}}u_{w,k}^{(n)}(\omega)\right)}{\sum_{r\in R_w}\exp\left(-\theta^{\mathrm{BR}}u_{w,r}^{(n)}(\omega)\right)},k\in R_w,w\in W$.

Step 4: Update flow. Set $\mathbf{f}^{\mathrm{CR}(n+1)}=\mathbf{f}^{\mathrm{CR}(n)}+\frac{1}{n}\mathbf{g}^{\mathrm{CR}(n)}$ and $\mathbf{f}^{\mathrm{BR}(n+1)}=\mathbf{f}^{\mathrm{BR}(n)}+\frac{1}{n}\mathbf{g}^{\mathrm{BR}(n)}$.

Step 5: Check the convergence. If $\sum_{w\in W}\sum_{k\in R_w}\left|\frac{f_{w,k}^{\mathrm{CR}(n)}-f_{w,k}^{\mathrm{CR}(n-1)}}{f_{w,k}^{\mathrm{CR}(n)}}\right|+\sum_{w\in W}\sum_{k\in R_w}\left|\frac{f_{w,k}^{\mathrm{BR}(n)}-f_{w,k}^{\mathrm{BR}(n-1)}}{f_{w,k}^{\mathrm{BR}(n)}}\right|<\varepsilon$, then stop iteration, otherwise, set $n=n+1$, turn to Step 2.

3 NUMERICAL STUDIES

In this section, the Nguyen and Dupuis network (Nguyen & Dupuis,1984) shown in Figure 1 is provided to demonstrate the proposed model, which consists of 13 nodes, 19 links, 25 routes, and 4 OD pairs. The free-flow travel time and design capacity for each link are shown in Table 1.

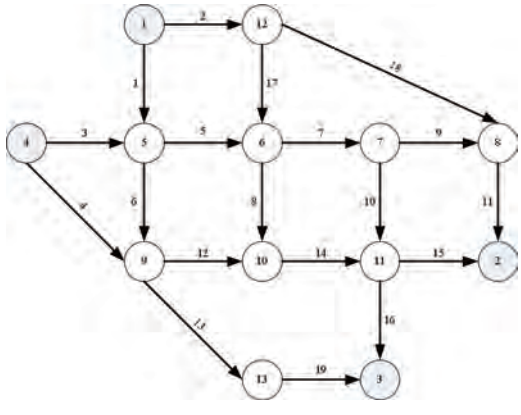


Figure 1. Nguyen and Dupuis network.

Table 1. Link characteristics.

Link	Free-flow travel time/min	Design capacity/pcu.h^{-1}
1	12	2500
2	12	2500
3	12	2500
4	24	2500
5	12	2500
6	12	2500
7	12	2500
8	12	2500
9	12	2500
10	12	2500
11	12	2500
12	12	2500
13	24	2500
14	12	2500
15	12	2500
16	12	2500
17	12	2500
18	36	1500
19	12	2500

For illustration purpose, the OD demand for each OD pair is described as follows:

$$q_w = \mu q_w^0 = \mu\left(q_w^{\mathrm{CR},0} + q_w^{\mathrm{BR},0}\right) = \mu\left(\eta q_w^0 + (1-\eta) q_w^0\right) \; w \in W \tag{24}$$

where q_w^0 represents the potentially maximum demand of OD pair w; μ is the multiplier of OD demand; η represents the proportion of completely rational travelers.

Other relevant parameters are set as: the parameters of BPR function in formula (1) are $\beta = 0.15$, $n = 4$; the variance coefficient $\rho_a = 0.5$, $a \in A$; the reliability parameter $\omega = 0.90$; the regret aversion parameter $\gamma = 0.15$; the perception error parameters $\theta^{\mathrm{CR}} = 1.0$ and $\theta^{\mathrm{BR}} = 0.2$ respectively; the maximum OD demands $q_{12}^0 = 5000$ (pcu·h^{-1}), $q_{13}^0 = 4000$ (pcu·h^{-1}), $q_{42}^0 = 4000$ (pcu·h^{-1}), and $q_{43}^0 = 3000$ (pcu·h^{-1}), respectively; the multiplier of OD demand $\mu = 0.5$; the proportion of completely rational travelers $\eta = 0.5$; the number of key links $N = 6$.

Table 2 shows the results of vulnerability evaluation for the 6 most critical links in the Nguyen and Dupuis network. It can be seen that the closures of these critical links decrease the network accessibility by 2% or more. The worst case is the closure of link 2, which leads to a 3.11% decrease in the network accessibility.

To test the effects of congestion level on network vulnerability, the consequences of link closures under various demand multipliers μ values are depicted in Table 3. From the table, it can be seen that as increases in demand level, the network vulnerability under link closures tends to go up. For example, when demand multiplier $\mu = 0.05$, the vulnerability index VUL_a is equal to 0.34% after the link 2 closure. However, when demand multiplier $\mu = 0.99$, the vulnerability index VUL_a rises to 5.45% under the link 2 closure. The results are not surprise because if the network is more congested, owing to closure, less spare capacity is available to absorb the rerouted traffic. In addition, From Table 3, it can be found that the vulnerability rankings may be varied owing to the congestion level. For instance, link 15 is ranked at $Rank_a = 5$ when $\mu = 0.05$. However, this link was ranked at $Rank_a = 6$ when $\mu = 0.49$, and $Rank_a = 7$ when $\mu = 0.99$.

In order to illustrate the effects of travelers' perception errors on network vulnerability in terms of different values of θ^{CR} and θ^{BR}, the results of link closures arising from various values of θ^{CR} and θ^{BR} are shown in Table 4. It should be pointed out

Table 2. The results of network vulnerability evaluation.

$Rank_a$	Link	VUL_a
1	2	3.11%
2	11	2.52%
3	14	2.44%
4	13	2.31%
5	19	2.31%
6	15	2.26%

Table 3. The effects of the congestion levels on network vulnerability.

	$\mu = 0.05$		$\mu = 0.49$		$\mu = 0.99$	
$Rank_a$	Link	VUL_a	Link	VUL_a	Link	VUL_a
1	2	0.34%	2	3.06%	2	5.45%
2	14	0.30%	11	2.48%	11	4.43%
3	7	0.29%	14	2.40%	14	4.26%
4	11	0.26%	13	2.27%	13	4.12%
5	15	0.26%	19	2.27%	19	4.12%
6	13	0.23%	15	2.22%	7	3.77%
7	19	0.23%	7	2.11%	15	3.76%
8	1	0.19%	1	1.76%	1	3.05%
9	3	0.17%	16	1.60%	16	2.83%
10	6	0.15%	3	1.57%	3	2.73%
11	4	0.15%	4	1.33%	4	2.43%
12	5	0.13%	5	1.26%	5	2.28%
13	16	0.12%	6	1.05%	6	1.83%
14	12	0.12%	18	0.92%	18	1.67%
15	17	0.11%	12	0.86%	12	1.50%
16	9	0.09%	17	0.80%	17	1.46%
17	18	0.09%	9	0.57%	9	1.02%
18	10	0.05%	10	0.43%	10	0.82%
19	8	–0.01%	8	0.27%	8	0.58%

that a higher θ^{CR} or θ^{BR} means smaller perception errors and vice versa. It can be observed from the table that travelers' perception errors can result in some impact on network vulnerability evaluation. Specifically, as change in values of θ^{CR} and θ^{BR}, the vulnerability rankings and the vulnerability index may be different accordingly. For example, link 13 is ranked at $Rank_a = 4$ when $\theta^{CR} = 5.0$ and $\theta^{BR} = 1.0$ whereas this link was ranked at $Rank_a = 7$ when $\theta^{CR} = 0.1$ and $\theta^{BR} = 0.02$.

Table 5 shows the impact of the regret aversion parameter on network vulnerability in terms of different values of γ. It can be found from Table 5 that the parameter γcan result in certain impact on network vulnerability evaluation. The vulnerability rankings may be varied due to the different γ values. For example, link 15 is ranked at $Rank_a = 4$ when $\gamma = 0.05$ whereas this link was ranked at $Rank_a = 6$ when $\gamma = 0.30$.

Table 6 depicts the impact of travelers' route choice decision criteria on network vulnerability according to different values of η. As shown in Table 6, with the variation on type structure of travelers, the vulnerability rankings and the vulnerability indices may change accordingly. For instance, when $\eta = 0.99$, which means that the completely rational travelers are dominant in the network, link 15 is ranked at $Rank_a = 6$ and the vulnerability index $VUL_a = 2.19\%$. However, this link was ranked at $Rank_a = 8$ and the vulnerability index $VUL_a = 2.06\%$ when $\eta = 0.01$.

The above analysis shows that ignoring the effects of the congestion levels of road network, travelers' perception errors and regret aversion

Table 4. The effects of the perception errors on network vulnerability.

	$\theta^{CR} = 5.0$, $\theta^{BR} = 1.0$		$\theta^{CR} = 1.0$, $\theta^{BR} = 0.2$		$\theta^{CR} = 0.1$, $\theta^{BR} = 0.02$	
$Rank_a$	Link	VUL_a	Link	VUL_a	Link	VUL_a
1	2	3.14%	2	3.11%	2	2.90%
2	11	2.55%	11	2.52%	14	2.55%
3	14	2.39%	14	2.44%	7	2.54%
4	13	2.33%	13	2.31%	11	2.28%
5	19	2.33%	19	2.31%	15	2.09%
6	15	2.19%	15	2.26%	19	2.02%
7	7	2.14%	7	2.15%	13	2.02%
8	1	1.74%	1	1.79%	1	1.63%
9	16	1.67%	16	1.63%	3	1.43%
10	3	1.56%	3	1.59%	6	1.33%
11	4	1.37%	4	1.35%	4	1.29%
12	5	1.26%	5	1.28%	5	1.10%
13	6	0.96%	6	1.07%	12	1.04%
14	18	0.94%	18	0.93%	16	1.03%
15	17	0.83%	12	0.88%	17	1.01%
16	12	0.81%	17	0.81%	9	0.82%
17	9	0.57%	9	0.58%	18	0.76%
18	10	0.46%	10	0.44%	10	0.40%
19	8	0.35%	8	0.28%	8	–0.08%

Table 5. The effects of the regret aversion parameter on network vulnerability.

	$\gamma = 0.05$		$\gamma = 0.15$		$\gamma = 0.30$	
$Rank_a$	Link	VUL_a	Link	VUL_a	Link	VUL_a
1	2	3.12%	2	3.11%	2	3.10%
2	11	2.52%	11	2.52%	11	2.51%
3	14	2.45%	14	2.44%	14	2.43%
4	15	2.31%	13	2.31%	13	2.34%
5	13	2.29%	19	2.31%	19	2.34%
6	19	2.29%	15	2.26%	15	2.20%
7	7	2.17%	7	2.15%	7	2.13%
8	1	1.83%	1	1.79%	1	1.74%
9	16	1.60%	16	1.63%	16	1.67%
10	3	1.58%	3	1.59%	3	1.62%
11	4	1.33%	4	1.35%	4	1.38%
12	5	1.30%	5	1.28%	5	1.27%
13	6	1.09%	6	1.07%	6	1.04%
14	18	0.94%	18	0.93%	18	0.93%
15	12	0.89%	12	0.88%	12	0.87%
16	17	0.82%	17	0.81%	17	0.80%
17	9	0.59%	9	0.58%	9	0.57%
18	10	0.45%	10	0.44%	10	0.44%
19	8	0.27%	8	0.28%	8	0.29%

Table 6. The effects of the travelers' route choice decision criteria on network vulnerability.

	$\eta = 0.01$		$\eta = 0.49$		$\eta = 0.99$	
$Rank_a$	Link	VUL_a	Link	VUL_a	Link	VUL_a
1	2	2.93%	2	3.11%	2	3.13%
2	14	2.72%	11	2.52%	11	2.53%
3	13	2.47%	14	2.45%	14	2.34%
4	19	2.47%	13	2.31%	19	2.32%
5	11	2.37%	19	2.31%	13	2.32%
6	7	2.27%	15	2.26%	15	2.19%
7	3	2.07%	7	2.15%	7	2.12%
8	15	2.06%	1	1.79%	1	1.72%
9	16	1.71%	16	1.63%	16	1.59%
10	1	1.63%	3	1.60%	3	1.45%
11	4	1.63%	4	1.35%	4	1.35%
12	6	1.45%	5	1.28%	5	1.23%
13	5	1.31%	6	1.08%	18	0.94%
14	12	1.13%	18	0.93%	6	0.90%
15	18	0.83%	12	0.88%	17	0.82%
16	17	0.79%	17	0.81%	12	0.79%
17	9	0.64%	9	0.58%	9	0.57%
18	10	0.43%	10	0.44%	10	0.44%
19	8	0.15%	8	0.28%	8	0.31%

degrees, as well as travelers' route choice decision criteria could misjudge the consequences of link closures and misidentify the most critical links.

4 CONCLUSIONS AND FUTURE RESEARCH

This study proposed a method of road network vulnerability identification taking into account travelers' risk aversion and bounded rationality simultaneously under travel time variations subjected to link failures. It is assumed that there are two kinds of travelers in the road network, completely rational travelers and bounded rational travelers, respectively. The completely rational travelers use the perceived travel time budget as their travel disutility while the bounded rational traveler use the perceived regret value as their travel disutility. According to the travelers' postulated route choice decision criteria, a mixed stochastic traffic assignment model formulated as variational inequality is constructed, a new vulnerability index of road network based on accessibility is defined, and a vulnerability identification model is built, and corresponding heuristic algorithm is also proposed. Numerical examples on the Nguyen and Dupuis network made apparent that the consequences of link closures could be misjudged and the vulnerability rankings could be misidentified, if ignoring the effects of the congestion levels of road network, travelers' perception errors and regret aversion degrees, as well as travelers' route choice decision criteria. Therefore, it is necessary to capture travelers' behavior characteristics for the vulnerability analysis.

The proposed vulnerability analysis only considers the scenarios of single link closure, and the consideration of multiple link closures is an important extension. Another valuable extension of this study is to take into account day-to-day adjustment processes for modeling travelers' behavioral responses to link closures.

REFERENCES

Chen, A., C. Yang, S. Kongsomsaksakul, & M. Lee (2007). Network-based accessibility measures for vulnerability analysis of degradable transportation networks. *Networks & Spatial Economics*. 7, 241–256.

Chen, B.Y., W. H. K Lam, & A. Sumalee, et al. (2012). Vulnerability analysis for large-scale and congested road networks with demand uncertainty. *Transportation Research Part A*. 3, 501–516.

Chorus C G (2014). A generalized random regret minimization model. *Transportation Research Part B*. 68, 224–238.

Facchinei F. & J.S. Pang (2003). *Finite-dimensional variational inequalities and complementarity problems*. Berlin Heidelberg, New York: Springer.

Huang, H.J. (1994). *Urban transportation network equilibrium analysis: theory and practice*. Beijing: China Communications Press.

Jenelius, E., T. Petersen, & L.G. Mattsson (2006). Importance and exposure in road network vulnerability analysis. *Transportation Research Part A*. 40, 537–560.

Liu, H.X., W. Recker, & A. Chen (2004). Uncovering the contribution of travel time realibility to dynamic route choice using real-time loop data. *Transportation Research Part A*. 38, 435–453.

Lo, H.K., X.W. Luo, & B.W.Y. Siu (2006). Degradable transport network: travel time budget of travellers with heterogeneous risk aversion. *Transportation Research Part B*. 40, 792–806.

Loomes, G. & R. Sugden (1982). Regret theory: an alternative theory of rational choice under uncertainty. *The Economic Journal. 368*, 805–824.

Matisziw, T.C., A.T. Murray (2009). Modeling *s–t* path availability to support disaster vulnerability assessment of network infrastructure. *Computers & Operations Research. 3*6, 16–26.

Nguyen, S., & C. Dupuis (1984). An efficient method for computing traffic equilibria in networks with asymmetric transportation costs. *Transportation Science*. 18, 185–202.

Sheffi,Y. (1985). Urban transportation networks: equilibrium analysis with mathematical programming methods. Englewood Cliff, New Jersey: Prentice-Hall.

Shao, H., W.H.K. Lam, Q. Meng, & M.L. Tam (2006). Demand-driven traffic assignment problem based on travel time reliability. *Transportation Research Record*. 1985, 220–230.

Siu, B.W.Y., H.K. Lo (2008). Doubly uncertain transportation network: degradable capacity and stochastic demand. *European Journal of Operational Research. 1*91, 166–181.

Taylor, M.A.P, S.V.C, Sekhar, & G.M. D'Este (2006). Application of accessibility based methods for vulnerability analysis of strategic road networks. *Networks and Spatial Economics*. 3, 267–291.

Taylor, M.A.P (2008). Critical transport infrastructure in urban areas: impacts of traffic incidents assessed using accessibility-based network vulnerability analysis. *Growth and Change*. 4, 593–616.

Wu, X., & Y. Nie (2011). Modeling heterogeneous risk-taking behavior in route choice: a stochastic dominance approach. *Transportation Research Part A* . 45, 896–915.

Yin, H.Y., & L.Q. Xu (2010). Vulnerability assessment of transportation road networks. *Journal of Transportation Systems Engineering and Information Technology*. 3, 7–13.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Safety analysis of autonomous driving using semi-Markov processes

Mattias Nyberg
Division of Mechatronics, Royal Institute of Technology (KTH), Stockholm, Sweden

ABSTRACT: The paper presents an approach of how safety of autonomous driving can be analysed by using semi-Markov processes. The approach can be used in development and assessment of vehicles implementing autonomous driving. Through a case study, it is indicated that a semi-Markov process model can capture relevant properties related to safety of autonomous driving. The case study particularly investigates if Level 3 autonomy, in which the driver is reponsible to take over when alerted by the system, can be made sufficiently safe. The paper also highlights how the current standard ISO26262 is insufficifient for autonomous driving where the system itself affects the exposure of operational situations. Therefore, as complement to ISO26262, it is shown how the proposed approach can be used to derive top level safety requirements.

1 INTRODUCTION

Autonomous driving is today one of the, if not *the*, major technological drivers in the area of on- and off-road vehicles. d new business opportunities. One of the main arguments for, and also main arguments against, autonomous driving is safety. It is argued that autonomous driving is safer that human driving, since a computer is always alert, always takes rational decisions etc. On the other hand, huge technological obstacles remain on how to realize autonomous driving in complex traffic situations, bad weather, and bad road conditions, where available commercial sensor technologies have yet to prove its sufficiency (Watzenig & Horn 2016).

Even though autonomous driving can be made safe, how can it be *proven* to be safe? In more technologically mature areas, there are applicable standards providing systematic methods to argue for safety, but this is still lacking in the area of autonomous driving. Current standards, such as ISO26262 (2011) cover only the case when a human driver is responsible for the driving.

To assess if autonomous driving is safe or not, systematic and formal analyses are needed. Therefore, as the main contribution, the present paper presents a method for formal analysis of safety of autonomous driving by using so called *semi-Markov processes*.

The use of Markov processes for analysis of dependability, including safety, is an established practice. The study of Markov processes has a long history but a mile-stone for its industrial usage for dependability analysis (Fuiqua 2003) was the introduction of the standard IEC61508 (2010). Since IEC61508 is a mother standard for many other standards, the usage of Markov processes has spread to many application areas. In particular, it is a large part of the ISO26262 standard, which uses continuous Markov-chains to asses reliability of hardware components. However, the literature is sparse on the usage of Markov processes for analysis of safety of autonomous driving. Markov process models of different kinds have indeed been used in the context of autonomous driving, but for the purpose of on-line perception or decision making, such as path planning, e.g. see (Hoekstra et al. 2013, Katrakazas et al. 2015, Wei et al. 2011). Also, there are studies on how to model *human* driving using Markov models, e.g. see (Mitrovic 2005, Oliver & Pentland 2000).

For reliability or safety analysis, the most common type of Markov processes used is discrete or continuous time Markov-chains (Fuiqua 2003). Markov chains are completely "memoryless" assuming that the probabilities of transitions are independent oftime, which implies that all transition times are assumed to be exponentially distributed. This is a limitation in modelling of real world systems, which do not always follow exponential distributions (Limnios 1997). Therefore, the present paper makes use of continuous-time so-called *semi-Markov processes*, introduced by Levy (1954) and Smith (1955). In semi-Markov processes, the probabilities of transitions from a state are allowed to depend on the time spent in the state. Thus, transition distributions in a semi-Markov process can be non-exponential.

The paper is focused around a case study, *Highway Pilot*, which is one typical autonomous driving function (Kirschbaum 2015). Autonomous driving

is classified into six levels (SAE 2016), and Highway Pilot is classified as Level 3 or 4. Level 3 means that the vehicle drives autonomously but requires the driver to be ready to take over whenever requested by the vehicle. Several manufacturers have argued that Level3 is not feasible since drivers can not be expected to be as alert as needed to manage the takeover reliably. However, other manufacturers argue that Level 3 is indeed possible as an intermediate step before Level 4, in which the driver has no responsibility. This is an example of a controversy in the area of autonomous driving. This is also an example where the semi-Markov analysis method proposed by the present paper can bring clarity and help answering questions such as: under what conditions can Level 3 be sufficiently safe, and, is it really easier to reach safety using Level 3 instead of Level 4?

The paper is organized as follows. In Sec. 2, selected parts of the theory of semi-Markov processes are shortly summarized. Then, in Sec. 3, a semi-Markov based approach for analysis of safety is presented. The approach utilizes the steady-state distribution in combination with an assessment of safety based upon the decision-theoretic concepts loss and risk. In Sec. 4 the case study highway pilot is presented, and also modeled by using a semi-Markov process. Then Sec. 5, based on the model, analyses the overall safety of highway pilot. Sec. 6 performs a sensitivity analysis on the parameters in the model and addresses the above mentioned questions related to Level 3 vs 4. Finally, Sec. 7 discusses how the proposed approach can be used in an engineering context.

2 SEMI-MARKOV PROCESSES

This section shortly introduces *continuous-time semi-Markov processes* in accordance with literature, e.g. (Limnios & Oprisan 2001, Limnios 1997). The purpose, and only focus, is to give the background theory needed for reading the rest of the paper.

Consider a system with a finite set of states $\{1,2,\quad,N\}$. Let X_n be a random variable representing the state after the n:th transition where $n \in \mathbb{Z}_{\geq 0}$. Let U_n be a random variable representing the time spent in state X_n, called the *sojourn* time. A semi-Markov process can then be described using a *transition function* in the form of a so called *semi-Markov kernel*, which is a matrix Q with elements

$$Q_{ij}(t) = \begin{cases} P(X_{n+1} = j, U_n \leq t \,|\, X_n = i) & i \neq j \\ 0 & i = j \end{cases} \tag{1}$$

Let M be the matrix of transition probabilities of the *embedded Markov chain*, i.e. $M_{ij} = P(X_{n+1} = j \,|\, X_n = i)$ which can be obtained from the semi-Markov kernel as $M_{ij} = \lim_{t\to\infty} Q_{ij}(t)$.

2.1 *Continuous-time Markov chains*

A special case of semi-Markov processes is so called *continuous-time Markov chains*, where the sojourn time U_n follows an exponential distribution, and the semi-Markov kernel takes the form

$$Q_{ij}(t) = \begin{cases} \frac{\lambda_{ij}}{\lambda_i}\left(1 - e^{-\lambda_i t}\right) & i \neq j \\ 0 & i = j \end{cases} \tag{2}$$

where $\lambda_i = \sum_{j\neq i} \lambda_{ij}$ and the parameters λ_{ij} are called *transition rates*. This further implies that the embedded Markov chain becomes $M_{ij} = 0$ for $i = j$ and

$$M_{ij} = \frac{\lambda_{ij}}{\lambda_i} \quad \text{for} \quad i \neq j \tag{3}$$

2.2 *Steady-state distribution*

Let $X(t)$ denote the random variable representing the state at time t. We now present how to compute the *steady-state distribution* of the semi-Markov process, defined by a vector π with elements $\pi_i = \lim_{t\to\infty} P\left(X(t) = i\right)$ for $i = 1, \quad, N$. To compute π, we will below use a formula containing two vectors $\mathbf{v}$ and $\mathbf{m}$.

Let $\mathbf{v}$ be the steady-state distribution of the embedded-Markov chain, which means that it is a row vector satisfying $\mathbf{v}M = \mathbf{v}$ and $\sum v_i = 1$. Note that $\mathbf{v}$ can be computed for example as $\mathbf{v} = \left[\,01\right]\Psi^T(\Psi\Psi^T)^{-1}$, where $\Psi = \left[\,(M - I)\,1\right]$, and 0 and 1 are vectors of 0's and 1's.

Let $\mathbf{m}$ be the column vector of expected sojourn times of the different states, i.e. $m_i = \mathrm{E}[U_n \,|\, X_n = i]$. Let $H_i(t)$ be the cumulative distribution function of the sojourn time of state i, i.e. $H_i(t) = P(U_n \leq t \,|\, X_n = i)$. Using one of the standard formulas for expectation, m_i can then be computed as

$$m_i = \int_0^\infty 1 - H_i(t)\,\mathrm{d}t = \int_0^\infty 1 - \sum_j Q_{ij}(t)\,\mathrm{d}t \tag{4}$$

The process is assumed to be *recurrent* meaning that it is possible, i.e. with a non-zero probability, to reach any state from any state, and the expected return time to any state is finite. Then the steady-state distribution of the semi-Markov process can be computed as

$$\pi = \frac{diag(\mathbf{v})\,\mathbf{m}}{\mathbf{v}\mathbf{m}} \tag{5}$$

where diag(**v**) is the diagonal matrix with the elements of **v** on its diagonal.

3 ANALYSING SAFETY BY USING SEMI-MARKOV PROCESSES

Inspired by the approach of *operational situations* in ISO26262 (2011), this section proposes a general loss- and risk-based framework for the analysis of safety. Although general, the intended target is here safety of autonomous driving.

3.1 *Loss and risk based measure of safety*

In accordance with so called decision theory (Berger 1985), each state *i* is associated with a *loss* denoted $L(i) \in \mathbb{R} \geq 0$ representing the level of "dangerousness" of being in state *i*. Then, expected loss, often referred to as *risk*, can be used as an overall measure of the dangerousness of the system. To obtain a measure independent of initial condition, we use the limiting expected loss:

$$\begin{aligned} &\lim_{t\to\infty} \mathrm{E}\left[L(X(t))\right] = \lim_{t\to\infty} \sum_i P(X(t)=i) L(i) = \\ &= \sum_i \lim_{t\to\infty} P(X(t)=i) L(i) = \sum_i \pi_i L(i) \end{aligned} \tag{6}$$

The actual value of the loss/dangerousness of being in state *i* can be chosen based on several principles. One approach is the one adopted in ISO26262 (2011), where metrics of *severity* and *controllability*, obtained from standardized tables, are combined into what here corresponds to loss. However, in the present paper, the dangerousness of being in state *i* will be assessed using the probability of fatality as a result of being in state *i*.

3.2 *Using probability of fatality as loss*

We consider two extra states, $\mathcal{F}$ and $\leftarrow\mathcal{F}$, residing in its own dimension, and where $\mathcal{F}$ represents "fatal accident" and is absorbing. Let the random variable $Y(t) \in \{\mathcal{F}, \neg\mathcal{F}\}$ represent if a fatal accident has occured or not at time *t*. We assume an exponential distribution of transition time from state $\leftarrow\mathcal{F}$ to state $\mathcal{F}$ but with a transition rate $\lambda_i^{\mathcal{F}}$ dependent on state *i*, i.e.

$$\begin{aligned} F_i(t) = P(Y(t) = \mathcal{F} \,|\, Y(0) = \neg\mathcal{F}, \forall \tau \in \\ [0,1]. X(\tau) = i) = = 1 - e^{-\lambda_i^{\mathcal{F}} t} \end{aligned}$$

While the transitions of *Y* have a dependency of *X*, we make the simplifying assumption that transitions of *X* do not depend on *Y*.

In accordance with ISO26262, we will evaluate safety based upon an assumption of 1 hour of driving. Therefore we use loss $L(i) = F_i(1\mathrm{h})$. It can then be realized that

$$\begin{aligned} &\lim_{t\to\infty} P(Y(t+1\mathrm{h}) = \mathcal{F} \,|\, Y(t) = \neg\mathcal{F}) \approx \\ &\approx \sum_i F_i(1\mathrm{h}) \lim_{t\to\infty} P(X(t) = i) = \sum_i L(i)\,\pi_i \end{aligned}$$

This means that the loss-based measure of safety (6) in fact corresponds to the limiting probability of fatality during 1h of operation.

4 MODELLING AUTONOMOUS DRIVING

This section presents the case study *Highway Pilot* (HP). HP is a general and well known function (Kirschbaum 2015), but the case study has been done in collaboration with Scania CV, a global manufacturer of heavy trucks. After giving an overview of HP, the approach for modelling and safety assessment presented in Sec. 2 and 3 is used to set up a semi-Markov process model of HP.

4.1 *Highway pilot*

When highway pilot is activated by the driver, the driving of the vehicle is completely taken over by the electronic system in the vehicle, and the driver does not need to be involved in the driving, not even monitoring. HP can only be activated when the vehicle is on a highway and the vehicle will continue to follow the highway as long as HP is on. HP is designed to function in a range of speeds, for example up to 120 km/h for a passenger car and 90 km/h for heavy trucks.

At the time of writing this paper, there is not yet a commercially available vehicle with HP. One reason is that there are a number of technical challenges present. One is the reliability of commercial radar and camera sensors. In more complicated road, weather, and traffic situations, the required performance is insufficient. But also under perfect conditions, these sensors have a relatively low reliability level causing them to miss or detect false objects at a non-negligible probability level. Independent of the reason, the situation when the sensors of the vehicle are fault free but still unable to obtain a correct interpretation of the surrounding environment, will in the following be referred to as *bad conditions*. Bad conditions may be detected by the system itself, but may also be undetected for a significant amount of time.

Another reason for the sensors to make an incorrect interpretation of the surrounding environment is that faults appear and become active. In contrast to bad conditions, which will always disappear after some time, faults do not disappear unless the system is repaired.

When HP is on and either bad conditions or faults are detected by the system, the system can not immediately deactivate HP. The reason is that the driver may not be ready to take over driving, e.g. he/she may be sleeping. Therefore, upon detection of badconditions or faults, the system will enter so called *degraded driving*. In degraded driving, faulty or unreliable sensors may be shut off. This typically causes perception accuracy or reliability to be lower. To compensate for this in the autonomous driving, a more safe driving style with more safety margins is typically adopted, including reducing the speed of the vehicle. During degraded driving, the system also tries to get attention from the driver so that he/she takes over the driving.

4.2 *States*

The states and transitions of the model are illustrated in Figure 1. Each of the states and their outgoing transitions are described below.

State S1 HP is off and the driver manually drives the vehicle. When the driver activates HP, there is a transition to S2.

State S2 HP is on and the electronic system of the vehicle drives the vehicle. The driver can at anytime deactivate HP which triggers a transition to S1. Alternatively, at any time, bad conditions or a fault can occur, which triggers transitions to S3 and S4 respectively.

State 3 HP is on and the electronic system of the vehicle drives the vehicle. However, undetected bad conditions are present so there is a greater risk for incorrect perception potentially leading to an incorrect and dangerous control action by the electronic system. The driver can at anytime deactivate HP which triggers a transition to S1. Alternatively, the system may detect the bad conditions and this triggers a transition to S5.

State 4 HP is on and the electronic system of the vehicle drives the vehicle. However, at least one undetected fault is present potentially leading to an incorrect control action by the electronic system. The driver can at anytime deactivate HP which triggers a transition to S1. Alternatively, the system may detect the fault[s] and this triggers a transition to S5.

State 5 HP is on but the system has detected bad condition or a fault so the driving is in a degraded mode. There is a warning indicated to the driver who is thereby urged to take over the driving by deactivating HP. If the driver deactivates HP, this triggers a transition to S1. If the driver does not deactivate HP within a predetermined time limit of 10 s, the system will initiate emergency stopping which means a transition to S6.

State 6 HP is on and the system is performing an emergency stopping. During emergency stopping, the system controls the brakes and steering of the vehicle in order to stop the vehicle as quickly as possible but in a controlled manner, e.g. making sure that the vehicle stops along the side, and not in the middle, of the road. The emergency stopping procedure does not involve the driver, and the driver is not able to terminate the procedure. When the vehicle reaches zero speed, there is a transition to S7.

State 7 HP is off and the vehicle is completely stopped, i.e. having a speed 0 km/h. When the driver again starts the vehicle, there is a transition to S1.

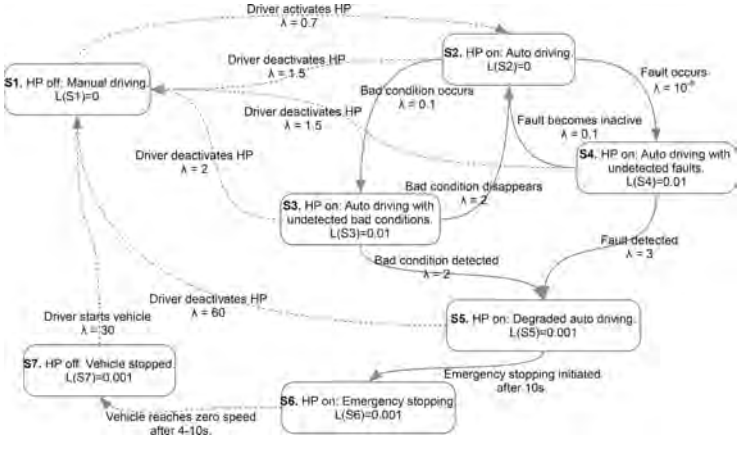


Figure 1. States and transition of highway pilot. Solid lines are used for transitions whose transition function is possible to affect by the design, and dashed lines for those not possible to affect by the design.

It can be noted that ISO26262 (2011) covers only safety with respect to faults, i.e. the right path (S2-S4-S5) of the model.

4.3 *Parameters of the model*

The model contains a number of parameters that need to be determined:

- The transition rate λ_{ij} for transitions respecting the Markov assumption, i.e. where the distribution of the sojourn time is exponential. These implicitly define the transition function $Q_{ij}(t)$ according to (2).
- Transition functions $Q_{ij}(t)$, for states with non-exponential distribution of sojourn times,
- The loss $L(i)$ associated with each state.

Now follows, for each state in the model, an explanation of the chosen parameter values in the model. All parameter values were derived by studying data and statistics from prototype systems and also from discussions with engineers. It should however be mentioned that the available data was limited or in many cases not applicable, due the prototype statusof the system, but also due to confidentiality reasons.

A general principle used is that as long as there is no clear argumentation for not choosing an exponential distribution, an exponential distribution is chosen. In all cases, transition rates are chosen by considering the resulting expected sojourn time, probability distribution of transitions to succeeding states, and for each succeeding state, the probability of transition within certain time periods.

State S1 The transition to S2 is considered to have constant rate $\lambda_{12} = 0.7$ and this corresponds to an expected soujourn time of 85 min and a probability of transition to S2 within 1 h to equal 0.50. The loss is $L(1) = 0$ since only loss (dangerousness) associated with usage of HP is considered.

State S2 The transition back to S1 is assumed to have a constant probability, with rate $\lambda = 1.5$. The transition to S3 is assumed to have constant rate $\lambda = 0.1$ and transition to S4 a constant rate $\lambda = 10^{-9}$. These three ratesresult in an expected sojourn time of 37 min. The probability of a transition back to S1 within 1 h becomes 0.75 while the corresponding probabilities of transitions to S2 and S3 respectively, become 0.05 and 10^{-9}. The loss is $L(2) = 0$ since S2 corresponds to when HP is working perfectly without any bad conditions or any active faults.

State S3 All transitions from S3 are assumed to occur with constant rates. The rate of the transition to S1 is assumed to be $\lambda_{31} = 2$, a bit higher than for the corresponding transition from S2 since when conditions are bad, the driver is likely to observe this and therefore deactivate HP. The rate of transition to S5 is assumed to be $\lambda_{53} = 2$ which corresponds to a probability of detecting the bad conditions within one hour to equal 0.49 and within 10 min to equal 0.24. The transition rate back to S2 were chosen to be also $\lambda = 2$. All this means that being in state S3, there in an equal probability of transitions to S1, S2, and S5. The expected sojourn time becomes 10 min. In contrast to the previous states, being in S3 isassociated with a non-zero loss due to possible incorrect perception caused by bad conditions. The loss is assumed to be $L(3) = 0.01$, but remember that this corresponds to the probability when being in state S3 for 1 hour and since the expected sojourn is only 10 min, the probability of fatality will be lower.

State S4 All transitions from S4 are assumed to occur with constant rates. The rate of the transition to S1 is assumed to be $\lambda_{41} = 1.5$, i.e. equal to the rate of corresponding transition from S2, since an undetected fault is likely notto change the behavior of the vehicle, and consequently not increase the probability of the driver to deactivate HP. The rate of the transition to S5 is assumed to be $\lambda_{45} = 3$, which is set a little bit higher than the rate of the corresponding transition from S3. The rate of transition back to S2 is set to $\lambda_{42} = 0.1$, which corresponds to an assumption that faults never disappears, but may become inactive. The loss of being in state S4 is estimated to be $L(4) = 0.01$, i.e. the same as for state S3.

State S5 During degraded driving the driver is alerted to take over driving so a transition to state S1 is set to a high and constant rate $\lambda_{51} = 60$. A transition to S6 will occur deterministically after 10 s, which means that this is not a constant rate and the sojourn time does not follow an exponential distribution. This implies that the probability of having a transition to S1 is $M_{51} = 1 - e^{-\lambda_{51} \cdot 10/3600} \approx 0.154$, and the probability of having a transition to S6 is $M_{56} = 1 - M_{51}$. This further implies

$$Q_{51}(t) = \begin{cases} 1 - e^{-\lambda_{51} t} & t < 10 \\ M_{51} & t \geq 10 \end{cases}$$

$$Q_{56}(t) = \begin{cases} 0 & t < 10 \\ M_{56} & t \geq 10 \end{cases}$$

The expected sojourn time can be computed using the formula (4):

$$m_5 = \int_0^\infty 1 - Q_{51}(t) - Q_{56}(t)\,\mathrm{d}t = \int_0^{10} 1 - \left(1 - e^{-\lambda_{51} t}\right)\mathrm{d}t = 9.2\mathrm{s}$$

In state S5, the loss is estimated to be $L(5) = 0.001$, i.e. significantly lower than in state S3 or S4. The reason is that the increased safety margin resulting from degraded driving.

State S6 From S6, the only possible transition is to S7 and it occurs at the time it takes to stop the vehicle. Since there are a number of uncertainties, e.g. road conditions and tire conditions, the time

is assumed to have a probability density function constant in the interval 4 to 10 s, and zero outside this interval. This means that the expected sojourn time is 7 s. The loss in S6 is estimated to be the same as in S5, i.e. $L(6) = 0.001$.

State S7 From S6, the only possible transition is to the starting S1. The transition is set to a fixed rate of $\lambda = 30$ which corresponds to an expected sojourn time of 2 min. The risk in S6 is estimated to be the same as in states S5 and S6, i.e. $L(7) = 0.001$. On one hand, the vehicle is standstill meaning that it can on its own not cause an accident. On the other hand, doing an emergency stop on a highway may result in stopping in a dangerous position on the road, thus increasing the probability of being hit by another vehicle.

5 COMPUTING THE RISK

To compute the risk, i.e. the limiting expected loss, according to (6), we first need to compute the steady-state distribution. For this, we will use (5) which requires **v**, the steady-state distribution of the embedded Markov-chain M, and **m**, the expected sojourn times.

For the states S1, S2, S3, S4, and S7, the corresponding rows in the matrix M, defining the embedded Markov-chain, are obtained by inserting the transition rates given in Section 4.3 into equation (3). For S6, there is only one possible transition so the row in M becomes by $M_{67} = 1$ and the other entries 0. For S5, both M_{51} and M_{56} have already been derived in Sec. 4.3, and the other entries are 0.

The expected sojourn times have all been given in Sec. 4.3. The resulting vector is

$$\mathbf{m} = [\,85\text{min}, 37\text{min}, 10\text{min}, 13\text{min}, 9.2\text{s}, 7\text{s}, 2\text{min}\,]$$

Now using (5) results in the steady-state distribution shown in Figure 2. Note that since we are only interested in the safety of HP, we condition on HP on, i.e. the bar chart shows the value $\lim_{t\to\infty} P(X(t) = i \,|\, \text{HP on})$ for each state i.

The risk, can now be computed by using (6), but use conditional expectation, conditioned by HP on:

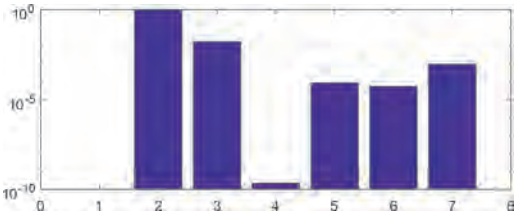


Figure 2. Steady-state distribution normalized for HP on.

$$\lim_{t\to\infty} \mathrm{E}[L(X(t)) \,|\, \text{HP on}] = 0.000165$$

This value is clearly too high. To investigate the explanation and to find means to reduce the value, the next section presents a sensitivity analysis with respect to the parameters in the model.

6 SENSITIVITY ANALYSIS

The model contains in total 14 transitions. Of these, 12 transitions are parameterized by a transition rate parameter λ_{ij}. The remaining two are the transitions from S5 to S6 and S6 to S7. The transition from S5 to S6 is therefore instead parameterized by the time of degraded autonomous driving, and the transition from S6 to S7 is parameterized directly by the expected sojourn time.

Each parameter is varied logarithmically from 10^{-6} to 10^{5} times the original value, and the resulting new risk is computed. For each parameter, the risk is plotted in Figure 3. In the figure, it is seen that some parameters have no effect on the total risk while other have a significant effect. For each of the 14 plots, a short interpretation of the observed behavior is now given.

S1 → S2 and S2 → S1 Since we measure the safety given that HP is on, solely activation and deactivation of HP does not affect the safety.

S2 → S3 If bad conditions rarely occurs (left part of the plot), then HP is very safe.

S2 → S4 If HP failure rate is changed, safety is not affected, since it is dominated by the effect of bad conditions.

S3 → S1 If the presence of undetected bad conditions are eliminated by the driver disabling HP, HP becomes very safe.

S3 → S5 If detection of bad condition is made more efficient, HP becomes safer since driver will faster disable HP. But to a limit, since driver reaction time from being alerted until action is not affected.

S3 → S2 If bad conditions quickly disappears, HP becomes very safe.

S4 → S1 The driverâ€™s rate of disabling HP, when a fault has occurred, does not affect HP safety. Since fault occurence is such a rare event.

S4 → S5 Increased efficiency in detecting faults does not affect safety since faults are anyway so rare.

S4 → S2 If faults quickly become inactive, HP becomes very safe.

S5 → S1 If the driver faster disables degraded HP, safety is not affected. Since degraded driving is comparably quite safe anyway.

S5 → S6 To more quickly, or slowly, activate emergency stopping, does not change safety. More time increases chance that driver disables HP, but also increase time of dangerous driving.

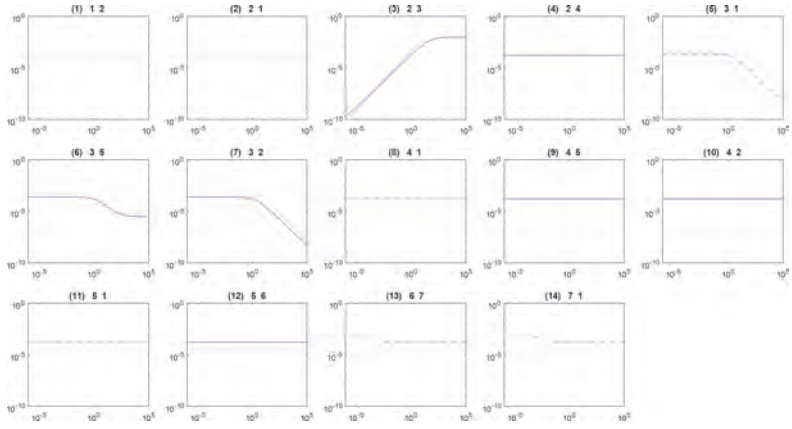


Figure 3. Sensitivity analysis where the probability of fatality is plotted w.r.t. variations in each of the 14 parameters in the model. Solid lines correspond to parameters possible to affect in the design of the HP, and dashed lines to parameters not affected by the design.

S5 → **S6** To more quickly start the emergency stopping after detection of fault, does not affect safety. Since fault detection is anyway so rare.

S6 → **S7** Stopping time can in fact not be affected, but if, then extreme long stopping time decreases safety, since time spent stopping is more dangerous than auto driving.

S7 → **S1** Time to restart vehicle after stopping can in fact not be affected, but if, then extreme long time before restart time decreases safety, since time spent stopping is more dangerous than auto driving.

6.1 *Conclusions of sensitivity analysis*

From the sensitivity analysis we can conclude that the model seems to be able to capture relevant properties of HP and their effects on safety. Another important conclusion is that the only way to reach sufficient safety of HP is to reduce time spent in state S3, i.e. to reduce transition rate from S2 to S3, i.e. to make sure that bad conditions do not occur, and according to the plot, the new value of the transition rate λ_{23} needs to be 10^{-6} times the original value 0.1, i.e. $\lambda_{23} = 10^{-7}$.

But then, it becomes Level 4 autonomous driving, i.e. the extra safety gained by demanding the driver to take over when bad conditions are detected is only marginal and not significant. However, to really know that we obtain Level 4 autonomous driving if $\lambda_{23} = 10^{-7}$, a second sensitivity analysis was performed with $\lambda_{23} = 10^{-7}$ and only the parameter λ_{51} varied. Although not shown in the paper due to space limitation, the result is that the risk is not dependent on λ_{51}. This implies that it truly becomes Level 4 autonomous driving.

7 USING THE MODEL IN ENGINEERING

This section discusses how a conceptual model of a technical function, like the one developed and analysed in the sections 4, 5, and 6, can be used generally by engineers to develop safe systems, in particular within automotive industry.

We will here view a transition function $Q_{ij}(t)$ for a transition dependent on the design (exemplified by the solid lines in Fig. 3) as a requirement on the system. This is a requirement to be taken as input to the succeeding engineering design activities. As a requirement it has to be formulated taking into account what is technically and commercially viable. In the case study, for example the transition function $Q_{56}(t)$ corresponds to the requirement

> "Emergency stopping shall start exactly 10 s after bad conditions or fault have been detected."

This requirement is in fact unrealistic in the sense that it is not so important that emergency stopping starts *exactly* after 10 s. And also, it is technically not feasible to build such a system since all components have tolerances. By following the principle of how safety requirements are formulated in industry with a Safety Integrity Level (SIL), such as ASIL in ISO26262, an example of a more realistic requirement is

P("Emergency ... ... detected.") ≥ 1 – 10^{-8}

This requirement corresponds in fact to a whole set of transition functions, namely all $Q_{56}(t)$ where there is a step somewhere between 8–12 s, and the integral of the corresponding density function is ≤10^{-8} excluding the step.

In conclusion, the transition function for each transition depending on the design must match the actual requirement and to align with principles of how requirements are written in industry, not one but a set of transition functions must be used. By using this insight, the following procedure summarizes how the proposed approach of the paper can assist the conceptual phase of engineering:

1. Set up a semi-Markov process with states and transitions modelling the system considered.
2. Identify the transition function $Q_{ij}(t)$ for the transitions not dependent on the design, e.g. transitions depending on the operator or other parts of the system (examplified by the dashed ones in Fig. 3).
3. Estimate the loss/dangerousness $L(i)$ of each state i, for example by estimating probability of fatality as result of being in the state a certain time.
4. For each transition depending on the design, specify a requirement and corresponding set of transition functions $Q_{ij}(t)$.
5. By using the formula (5), compute steady state distribution π, for each valid combination of transition functions picked from the sets in step 3, possible by using sampling if the set is too large.
6. For all steady-state distributions computed, compute the risk, i.e. limiting expected loss (6). If not acceptable, go back to step 4 and modify, possibly by first doing a sensitivity analysis, the requirements and corresponding transition functions $Q_{ij}(t)$.
7. Design the system according to the obtained requirements.

It can be noted that, in relation to the method of *hazard analysis and risk assessment* in ISO26262 (2011), the steady-state distribution, and the steps 1, 2, 4, and 5 leading up to it, replace the concept of *operational situations* in ISO26262. However, while the steady-state distribution is a result of the chosen requirements, the operational situations is a fixed input to the requirements elicitation and consequently, ISO26262 does not acknowledge themutual interplay between identification of requirements and the operational situations.

8 CONCLUSIONS

The paper has presented an approach of how safety of autonomous driving can be analysed using semi-Markov processes. The approach can be used both to complement the often informal discussion of whether autonomous driving is safe or not. But more importantly, the approach can be used also in development and assessment of vehicles implementing autonomous driving.

The case study has indicated that a semi-Markov process model can capture relevant safety related properties of autonomous driving. The case study also concludes that the extra safety obtained by Level 3 autonomous driving is only marginal and in fact, tomake Level 3 sufficiently safe, it needs to be Level 4.

The paper has highlighted how the current standard ISO26262 is insufficient for complex functions like autonomous driving where the system itself affect the exposure of operational situations. Instead of using the concept of operational situations from ISO26262, it has been shown how the proposed approach of semi-Markov processes can be used to derive top level requirements.

REFERENCES

Berger, J. (1985). *Statistical decision theory and Bayesian Analysis* (2nd ed.). New-York: Springer.

Fuiqua, N. (2003). The applicability of markov analysis methods to reliability, maintainability, and safety. *Selected Topics in Assurance Related Technologies 10*(2), 203–207.

Hoekstra, A. et al. (2013). Evolving markov chain models of driving conditions using onboard learning. In *IEEE Int. Con. on Cybernetics (CYBCONF)*, Switzerland.

IEC61508 (2010). Functional safety of electrical/electronic/programmable electronic safety-related systems. Standard IEC61508, International Electrotechnical Commision.

ISO26262 (2011). Road vehicles - functional safety. Standard ISO26262, International Organization for Standardization.

Katrakazas, C. et al. (2015). Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions. *Transportation Research Part C: Emerging Technologies 60*(Supplement C), 416–442.

Kirschbaum, M. (2015). Highly automated driving for commercial vehicles. In *6th International Munich Chassis Symposium*, Munich, Germany.

Levy, P. (1954). Processus semi-markoviens. In *Proc. Int. Gong. Math*., Amsterdam, pp. 416–426.

Limnios, N. (1997). Dependability analysis of semi-markov systems. *Reliability Engineering and System Safety 55*, 203–207.

Limnios, N. & G. Oprisan (2001). *Semi-Markov Processes and Reliability*. New-York: Springer.

Mitrovic, D. (2005, 07). Reliable method for driving events recognition. *Intelligent Transportation Systems, IEEE Transactions on 6*, 198–205.

Oliver, N. & A. Pentland (2000). Driver behavior recognition and prediction in a smartcar.

SAE (2016). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Standard J3016, Society of Automotive Engineers.

Smith, W. (1955). Regenerative stochastic processes. *Proceedings of the Royal Society A 232*(1188), 6–31.

Watzenig, D. & M. Horn (2016). *Automated Driving: Safer and More Efficient Future Driving*. New-York: Springer.

Wei, J. et al. (2011). A point-based mdp for robust single-lane autonomous driving behavior under uncertainties. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, Shanghai, China.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Research on bayesian reliability growth evaluation method for mechanical products

Jinyong Yao, Hao Wu & Tongmin Jiang
School of Reliability and Systems Engineering, Beihang University, Beijing, China

Yiliu Liu
Department of Mechanical and Industrial Engineering Faculty of Engineering, NTNU, Trondheim, Norway

ABSTRACT: The probability of failure of mechanical products often conforms to the Weibull distribution. If the Weibull distribution is used to construct the likelihood function, the Bayesian method is used to estimate the reliability, which involves a large number of integral operations and is not easy to select the appropriate prior distribution. In order to reduce the computational workload and improve the efficiency of the evaluation, based on the mature exponential distribution Bayesian method, this paper transformed the Weibull distribution into exponential distribution and chose the obverse-Gamma distribution as the prior distribution. Through a series of derivation and calculation, we obtain the Weibull distribution Life parameter estimation, and finally gives the corresponding examples to verify the feasibility of the method.

1 INTRODUCTION

Mechanical reliability growth technology is an important part in mechanical reliability theory. It exists in the whole lifespan of design, production and usage. Mechanical reliability growth technology aims to evaluate mechanical product reliability with the synthesized knowledge of different kinds, such as mechanical engineering, statistics and management. In the whole lifespan of the products, reliability growth test (RGT), data analysis and management are effective and economical methods to enhance the reliability stage.

Reliability growth test often faces the problem of too little data volume, especially for some mechanical products. The reliability test usually takes a long time and costly, and the test uses small samples or even very small samples, which makes the product reliable In the process of assessing small sample data. Bayes method can use the current test information together with the historical test data of prior information and similar products, which can be well applied to small sample data Statistical inference problems (Zhang 2000, Zhang 2003). The basic principle of Bayesian theory is to calculate the posterior distribution by using the pre-test distribution and sample information, so as to estimate the point estimate and confidence interval of the variable and further derive the estimated values of other related reliability features.

The specific method is generally based on a parameter as a random variable, through the construction of the likelihood function and pre-test distribution function, using the Bayesian formula to derive the form of a posterior distribution function, and then use the reliability test to get the failure data as a sample. Estimate the probability density function or distribution function of the relevant parameters, and finally get the reliability index. In this process. The determination of transcendental function and the construction of the likelihood function are the key points, and the difficulty is parameter estimation. Many scholars at home and abroad have done a lot of research on parameter estimation.

In engineering practice, the failure probability of mechanical products usually obeys Weibull distribution. The Bayesian parameter estimation of Weibull distribution is used to make mathematical derivation step by step by using the most direct method (Shi 1992). A large number of integral calculations are used, and the whole process is calculated Both the amount and computational complexity are high, increasing the workload for reliability assessment. The Weibull distribution is converted to the extreme distribution and then calculated using the Bayesian formula (Liu et al. 2005). Although the calculation is simplified, the workload is still very large. In response to this problem, this paper seeks a new method of evaluation and calculation to avoid a large number of integral calculations and reduce the computational workload. In the literature (Mao 1999), an exponential distribution evaluation method is introduced. The exponential distribution is a commonly used form of distribution, while the gamma distribution is the conjugate prior distribution of the exponential distribution. It is also suitable for describing the characteristic life of the mechanical and electrical products. Therefore, Bayes estimation of exponential

distribution, the choice of gamma distribution as a prior distribution, which will greatly simplify the calculation, compared to the direct use of Weibull Bayesian derivation, the efficiency is significant.

Based on this idea, this paper presents a method for Bayes reliability growth evaluation based on model transformation. We first convert the Weibull distribution obeying the product to an exponential distribution, and then return to the Bayes estimation of the parameter obeying the exponential distribution and finally return the parameters of the exponential distribution estimation to the parameters of the Weibull distribution, respectively, to obtain the estimate. At the end of this article, we use a practical case to verify the feasibility of the method.

2 MODEL DISTRIBUTION CONVERSION

Suppose the mechanical product life T obey Weibull distribution, the distribution function is:

$$F(t)=1-e^{-(t/\eta)^{\beta}},t>0,\beta,\eta>0 \tag{1}$$

Its probability density function is:

$$f(t)=\frac{\beta}{\eta}(t/\eta)^{\beta-1}\mathrm{e}^{-(t/\eta)^{\beta}} \tag{2}$$

where t is the time, β is the shape parameter, η is the characteristic life.

The cumulative probability distribution function is the integration of the probability density at time t_s:

$$F(t)=\int_0^{t_s}\frac{\beta}{\eta}(t/\eta)^{\beta-1}\mathrm{e}^{-(t/\eta)^{\beta}}dt \tag{3}$$

Assume $z = t^{\beta}$, which can be introduced as $dz=\beta t^{\beta-1}dt$, then the equation (3) can be expressed as:

$$F(t)=\frac{1}{\eta^{\beta}}\int_0^{t_s^{\beta}}\mathrm{e}^{-z/\eta^{\beta}}dz \tag{4}$$

Here the equation (4) can be further simplified as:

$$F(t)=\frac{1}{\theta}\int_0^{X}\mathrm{e}^{-z/\theta}dz \tag{5}$$

where $\theta=\eta^{\beta}$, $X=t_s^{\beta}$.

So, the equation (5) is the exponential distribution function, whose probability density function at time z is:

$$f(x)=\frac{1}{\theta}\mathrm{e}^{-x/\theta} \tag{6}$$

Among them, the exponential distribution function parameters are respectively: $\theta=\eta^{\beta}$, $x=t^{\beta}$.

That is to say, if the lifetime t obeys the Weibull distribution of the shape parameter β and the characteristic lifetime η, after exponential transformation, the experimental time can be expressed as an exponential distribution of x and the average life expectancy θ.

As can be seen from Equation (6), x is a random variable for the parameters t and β, and $1/\theta$ is an exponential distribution parameter which is related to the parameters η and β.

If β is known, then x can be treated as a random variable equivalent to the parameter t. The parameter $1/\theta$ depends only on the parameter η. The pre-test distribution of θ is equal to the pre-test distribution of η, and then the Bayes theory is used to infer it.

If β is unknown, θ depends on the parameters η and β simultaneously, which can be extrapolated using the multiple Bayes method, which is not discussed here.

In general, β as a shape parameter, for the same type of batch products, its value is often unchanged.

3 PRIOR DISTRIBUTION DETERMINATION

The basic principle of Bayes is to use the sample information to correct the prior information and estimate the approximate value closer to the truth. The formula is:

$$h(\theta|x)=\frac{p(x|\theta)\pi(\theta)}{m(x)}\propto p(x|\theta)\pi(\theta) \tag{7}$$

where $h(\theta|x)$ is the posterior distribution; $p(x|\theta)$ is the likelihood function; $\pi(\theta)$ is the prior distribution; $m(x)$ is the edge distribution.

Since $m(x)$ does not depend on the parameter θ, it plays a negligible role in calculating the posterior distribution of θ as a regularization factor. Therefore, the posterior distribution depends on the joint distribution density function and prior distribution.

When the β value has been determined based on prior information, then the prior distribution parameter θ depends on the feature life η.

According to the engineering experience, the prior distribution of the characteristic life η of mechanical products generally obeys the inverse gamma distribution, then the prior distribution of θ can be determined as:

$$\pi(\theta)\sim IGa(a,b) \tag{8}$$

Its distribution form is:

$$\pi(\theta)=\frac{b^{a}}{\Gamma(a)}\left(\frac{1}{\theta}\right)^{a+1}e^{-\frac{b}{\theta}} \tag{9}$$

And the mean and variance of the distribution respectively are:

$$E(\theta)=\frac{b}{a-1}$$
$$V(\theta)=\frac{b^2}{(a-1)^2(a-2)} \quad (10)$$

It can collect the corresponding fault information, calculate the average and variance statistics, and can calculate the value of hyper parameters a and b.

4 RELIABILITY ASSESSMENT

Assuming the end of a mechanical product reliability test using the censored method, in which the sample number n, the failure number is r, then the equation (6) of the likelihood function can be expressed as:

$$L(x,\theta)=f(x_1,x_2,...,x_r,\theta)=\frac{n!}{(n-r)!}\theta^{-r}\exp(-X_r/\theta)\propto\theta^{-r}\exp(-X_r/\theta) \quad (11)$$

where $X_r=\sum_{i=1}^{r}x_i+(n-r)x_r$

Then the posterior distribution of θ is:

$$h(\theta|x)\propto p(x|\theta)\pi(\theta)=\frac{b^a}{\Gamma(a)}\left(\frac{1}{\theta}\right)^{a+r+1}e^{-\frac{b+X_r}{\theta}}\propto\theta^{a-r-1}e^{-(b+X_r)/\theta} \quad (12)$$

At this point, the mean of θ is its point estimate, which is:

$$\hat{\theta}=\frac{b+X_r}{a+r-1} \quad (13)$$

It is known that the posterior distribution of θ obeys the inverse gamma distribution and parameter $1/\theta$ obeys the gamma distribution.

According to $\theta^{-1}\propto Ga(a+r,b+X_r)$, so:

$$2(b+X_r)\theta^{-1}\propto Ga\left(\frac{2(a+r)}{2},\frac{1}{2}\right)=\chi^2[2(a+r)] \quad (14)$$

According to the chi-square distribution, it can be concluded that the confidence interval is $1-\alpha$, the interval of θ is estimated as:

$$\left[\frac{2(b+X_r)}{\chi^2_{1-\alpha/2}(2(a+r))},\frac{2(b+X_r)}{\chi^2_{\alpha/2}(2(a+r))}\right] \quad (15)$$

The Bayes method can be used to estimate the corresponding parameters, and the corresponding reliability features can also be obtained. The mean time between failures (MTBF) of the product is:

$$MTBF=\eta\Gamma\left(1+\frac{1}{\beta}\right) \quad (16)$$

5 SIMULATION CASE CALCULATION

The data in this paper refer to the product failure time data in the paper (Zhang et al. 2005) as the historical data, which can be seen in Table 1. The product in this paper is a typical mechanical product. The distribution of failure interval obeys Weibull distribution. According to the data in the paper (Zhang et al. 2005), the shape parameter $\beta=1.3612$ of the Weibull distribution can be calculated by the least square method, and the value is taken as the shape parameter of the Weibull distribution in the failure interval of a certain mechanical product, and remain unchanged.

From the historical information data of the mechanical product in Table 1, when the parameter β is known, it can be seen from the calculation that the mean and variance of η_1^β, η_2^β,…, η_n^β are:

$$E(\eta_i^\beta)=11008.454,\quad V(\eta_i^\beta)=130857196.8$$

From equation (10), hyper parameters can be calculated as:

a = 3, b = 21203.

So, we can determine the prior distribution of the failure interval of mechanical products according to the inverted gamma ray distribution $IGa(3,21203)$.

According to known Weibull distribution shape parameters β, and the known parameter η^β, so the characteristic life parameter η can be calculated as 931.62h.

Using Monte Carlo simulation method, a set of 8 random number sequences obeying the Weibull distribution of two parameters (the shape parameter is 1.3612 and the size parameter is 931.62) are generated and arranged in order of size $t_1,t_2,\ldots,t_8$ as simulation test data. Assuming the experimental program for the sample size of 8 sets of the same batch of mechanical products, the number of failures for the censored is 5, the test data processing shown in Table 2, where n is the sample size, T is the censored time, $t_i^\beta(i=1,2,\ldots,5)$ is the β power of fault data, X_s is the sum of t_i^β and r is the truncation number.

Set the reliability of $1-\alpha$ is 90%, according to equation (13) ~ (15), get the following data:

The posterior point of θ is estimated as 9264.4.

θ interval is estimated as:

$$\left[\frac{129701.7}{\chi^2_{0.1}(16)},\frac{129701.7}{\chi^2_{0.9}(16)}\right]=[5509.8,13931.4]$$

Table 1. Fault interval time history data.

No.	1	2	3	4	5	6	7	8	9	10	11	12
η_i	79.45	104.11	143.56	196.16	248.77	250.41	276.71	367.12	395.07	397.81	408.22	446.03
No.	13	14	15	16	17	18	19	20	21	22	23	24
η_i	459.18	552.88	566.03	695.89	698.63	735.34	776.44	814.25	840.55	919.45	932.6	972.05
No.	25	26	27	28	29	30	31	32	33			
η_i	986.85	1261.37	1498.08	1590.14	1603.29	1708.49	1894.25	2326.58	2841.1			

Table 2. Monte Carlo simulation data.

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	n	T
47.6	301.2	432.5	482.6	735	936.9	1531.6	1644.8	8	735
t_1^β	t_2^β	t_3^β	t_4^β	t_5^β	t_6^β	t_7^β	t_8^β	X_s	r
188.8	2367.1	3873.5	4496.8	7972.4	7972.4	7972.4	7972.4	42815.8	5

Table 3. Data processing and analysis.

	β	η	MTBF (h)	Sample size (sets)	Test time (h)
Historical experience data	1.3612	922.14	852.06	33	2841.1
Simulation test data	1.3612	932.18	764.025	8	1644.8
Bayes method data	1.3612	820.4	758.05	8	735

Furthermore, the point estimate and interval estimation of η are 820.4 and [560.07,1107.07] respectively, and the estimated value of MTBF is 758.05 h.

Table 3 shows the comparison of empirical information and experimental data with several evaluation methods. The least squares method is used to estimate the historical data in the paper (Zhang et al. 2005), which represents the empirical information. The Bayes method combines the prior information and the experimental (Simulation) information, the calculation results and simulation data are similar, but the test time is about half.

6 CONCLUSION

In this paper, we presents a method for Bayes reliability growth evaluation based on model transformation, which can effectively solve several difficult issues in mechanical products reliability evaluation. It can be adopted in the following researches, but not limited to

1. this method avoids the complicated model of Weibull distribution, reduces the computational complexity of Bayes estimation, simplifies the calculation process and improves the efficiency of the algorithm;
2. this method also saves test time, and greatly reduces the number of samples.

REFERENCES

Chen, D.N. & Yao, C.Y. 2012. Reliability Analysis of Multi-state System Based on Fuzzy Bayesian network and Application in Hydraulic System. *Journal of Mechanical Engineering* 48(16): 175–183.

Liu, Z.X. & Zhou, Y.Q. 2005. Weibull distribution reliability assessment method. Journal of Quality and reliability.

Mao, S.S. 1999. Bayesian statistics. Beijing, China Statistics Press.

Shi, Y.M. 1992. Censored Life Testing Trials Weibull Distribution Bayesian Statistical Analysis. *Journal of Engineering Mathematics* 3(9):98–103.

Zhang, H.B., Jia, Y.Z. & Zhou, G.W. 2005. Research on Numerical Model of NC System Fault Time Distribution. *Journal of Harbin Institute of Technology* 37(2):198–200.

Zhang, S.F. 2000. Bayes Small Sample theory and its Application in Weapon System Evaluation. Changsha, National University of Defense Technology.

Zhang, X.P. 2003. Research on the Application of Small Sample Inference and Fusion Theory in Weapon System As sessment. Changsha, National University of Defense Technology.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Importance measure method for joint clearance of mechanism

Zhongchao Sun, Tianxiang Yu, Weimin Cui & Bifeng Song
School of Aeronautics, Northwestern Polytechnical University, Xi'an, Shaanxi, China

ABSTRACT: A method to measure the contribution of random joint clearances to mechanism output error, termed as Error Importance (EI), is presented in this paper. In this method, the 2-order original moment is used to characterize the deviation of mechanism output error and joint clearances from their ideal values, i.e. zero. Then the 2-order original moment of mechanism output error is decomposed into a series of fractions. These fractions are divided into two categories: individual effects and interaction effects. The total effect of one joint clearance is defined as the sum of the corresponding main effects and the interaction effects, and total importance index, i.e. EI index, is defined as the value of its total effect relative to 2-order original moment of mechanism output error. Similarly, the total importance index of a group of joint clearances are defined also. Then the mathematical and physical properties of the EI indices are discussed, and a Monte Carlo based evaluation method is offered. At last, the EI indices are applied to cabin door mechanism of aircraft landing gear. Simulation results revealed that the EI indices can reflect honestly the contribution of joint clearances to mechanism output error.

1 INTRODUCTION

Uncertainty joint clearances are inevitable in linkage mechanisms (Han et al. 2002, Zhang et al. 2015), and lead to mechanism output error (Li et al. 2015). The output error caused by joint clearances cannot be compensated with any kind of calibration (Chen et al. 2013). Furthermore, mechanism produce different error in different positions (Kumaraswamy et al. 2013), traditional calibration cannot ensure the motion accuracy across the whole workspace. With respect to these problems, joint clearances should be controlled under specified level, so as to ensure the mechanism output accuracy. In order to improve mechanism output accuracy maximally under limited resources, the study of *how different joint clearances contribute to mechanism output error* is very necessary.

Joint clearances are random variables of always positive. Thus the deviation of joint clearances from the ideal value, i.e. zero, is composed of variation and mean shift. Under the combined action of various joint clearances, the deviation of mechanism output from the ideal or design value is also composed of variation and mean shift, as shown in Figure 1. In engineering practice, not only the variation, we care about also the deviation of mechanism output from the ideal or design value. Thus both the contribution of variation and mean shift should be taken into account in EI of joint clearances.

With respect to the problem of identifying the most important joint clearances, Sensitivity Analysis (SA) of random variables is the most related

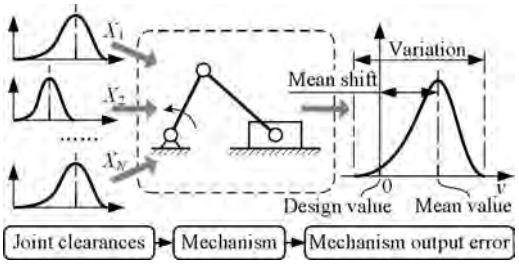


Figure 1. Mechanism output error under the joint effect of various joint clearances.

practice. SA is defined as "*the study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input*" (Saltelli 2002). There are mainly two kinds of SA methods, i.e. Local Sensitivity Analysis (LSA) methods and Global Sensitivity Analysis (GSA) methods (Sobol 2001, Borgonovo et al. 2016, Pianosi et al. 2016). GSA are also called uncertainty Importance Measure (IM) in some literatures (Aven et al. 2010, Borgonovo 2007).

SA methods have already been used by many researchers to study the relative importance of input errors. Caro et al. (2009) proposed partial derivatives based sensitivity indices for the geometric parameters and actuated variables of 3-RPR planar parallel manipulators. Hanzaki et al. (2009) use partial derivatives to study predict how the steering error is affected by manufacturing tolerances, assembly errors, and clearances. GSA methods, e.g. variance

based method, are also used to identify the most important factors that influence the output errors (Cheng et al. 2014, 2015).

However, the current SA methods are not suitable for the problem this paper addressed. From the definition of SA one can get that the present SA methods focus mainly on "uncertainty", rather than the deviation from the ideal or design values. Take variance based methods as an example, they use variance to characterize "uncertainty", implying that the deviation from the "mean values" is what we interested in. However, the mean value of joint clearance can never be zero. Thus variance based SA methods will produce unreasonable results if they are used to measure the contribution of joint clearances to mechanism output error. Similar conclusions can also be derived for other SA methods.

With respect to this problem, an importance measure method of joint clearances that can take into consideration both the contribution of variation and mean shift is presented in this paper. In addition, the presented method should satisfy the requirements of be "*global, quantitative, and model free*" that claimed by Saltelli (2002), and should be convenient to implement.

The following of the paper is organized as follows. In section 2, EI indices for one and a group of joint clearances are proposed based on the decomposition of the 2-order original moment of model output, and a Monte Carlo based evaluation method is offered. In section 3, the mathematical and physical properties of the presented EI indices are discussed. In section 4, an application case to cabin door mechanism of aircraft landing gear is offered. Section 5 provided a summary and some concluding remarks.

2 IMPORTANCE MEASURE METHOD FOR JOINT CLEARANCES

2.1 *2-order original moment decomposition of mechanism output error*

As the basement of IM method, a reasonable index to characterize the deviation of joint clearances and mechanism output from their design or ideal values is necessary. Taguchi et al. (1989) proposed the quadratic quality loss function to measure the deviation of parameters from design values. It took into consideration both the mean shift and variation, and has been adopted by many researches in error or tolerance allocation (Jin et al. 2015). The ideal values of both joint clearances and mechanism output errors are all zero. In this respect, we use the 2-order original moment to measure the deviation of input or output errors from their design values.

Assume a mechanism with N joint clearances, i.e. $\boldsymbol{X} = (X_1, X_2, \ldots\ldots, X_N)$, and one output, i.e. Y. The output error depends only on joint clearances. The PDFs of $\boldsymbol{X}$ are $f_{X1}(x_1), f_{X2}(x_2), \ldots, f_{XN}(x_N)$, and that of Y is $f_Y(y)$. The functional relationship between $\boldsymbol{X}$ and Y can be written as the following form:

$$y = g(\boldsymbol{x}) = g(x_1, x_2, \cdots, x_N) \tag{1}$$

If there are no joint clearances, the mechanism output error Y will obviously be zero, thus the following equation holds:

$$y = g(0) = 0 \tag{2}$$

Assume that the function $g(\boldsymbol{x})$ is $(m+1)$ times differentiable at the point of $\boldsymbol{x} = 0$. The m-th order Taylor expansion is performed to Equation 1:

$$\begin{aligned} y = g(\mathbf{x}) = g(0) &+ \sum_{i_1=1}^{N} \frac{\partial g(0)}{\partial x_{i_1}} \cdot x_{i_1} + \frac{1}{2!}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N} \frac{\partial g(0)}{\partial x_{i_1}\partial x_{i_2}} \cdot x_{i_1}x_{i_2} + \frac{1}{3!}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N} \frac{\partial g(0)}{\partial x_{i_1}\partial x_{i_2}\partial x_{i_3}} \cdot x_{i_1}x_{i_2}x_{i_3} \\ &+ \cdots + \frac{1}{m!}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\cdots\sum_{i_m=1}^{N} \frac{\partial g(0)}{\partial x_{i_1}\partial x_{i_2}\cdots\partial x_{i_m}} \cdot x_{i_1}x_{i_2}\cdots x_{i_m} + \frac{1}{(m+1)!}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\cdots\sum_{i_{m+1}=1}^{N} \frac{\partial g(\theta \boldsymbol{x})}{\partial x_{i_1}\partial x_{i_2}\cdots\partial x_{i_{m+1}}} \cdot x_{i_1}x_{i_2}\cdots x_{i_{m+}} \end{aligned} \tag{3}$$

where the last item is Lagrange reminder, with $0 \le \theta \le 1$. Because $g(0) = 0$, the first item in the right side of Equation 3 always be zero. Without the considering of the Lagrange reminder, Equation 3 can be simplified into the following form:

$$\begin{aligned} y \approx &\sum_{i_1=1}^{N}\left(a_{i_1} \cdot x_{i_1}\right) + \sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\left[a_{i_1,i_2} \cdot \left(x_{i_1}x_{i_2}\right)\right] \\ &+ \cdots \sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\cdots\sum_{i_m=1}^{N}\left[a_{i_1,i_2,\cdots i_m} \cdot \left(x_{i_1}x_{i_2}\cdots x_{i_m}\right)\right] \end{aligned} \tag{4}$$

where the polynomial coefficient, say, $a_{i1, i2, \ldots, is}$ can be computed as follows:

$$a_{i_1,i_2,\cdots i_s} = \frac{1}{s!} \cdot \frac{\partial^{(s)} g(0)}{\partial x_{i_1}\partial x_{i_2}\cdots\partial x_{i_s}} \quad (s = 1,2,\cdots,m) \tag{5}$$

Computing the 2-order original moment of Equation 4, we get:

$$A^{(2)}(Y) = E(Y^2) = E\left[\left(\sum_{i_1=1}^{N} a_{i_1} \cdot x_{i_1} + \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} a_{i_1,i_2} \cdot x_{i_1} x_{i_2} + \cdots + \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_m=1}^{N} a_{i_1,i_2,\cdots,i_m} \cdot x_{i_1} x_{i_2} \cdots x_{i_m} \right)^2 \right] \tag{6}$$

One can get that the right side of Equation 6 is also a polynomial of $M = m^2$ orders with the lowest order of 2. The equation can be expanded into the sum of a series of items. But in condition of high order Taylor expansion and a large number of variables, the expansion of Equation 6 will be very complicated. For convenience, we use $A_{i1, i2, \ldots, is}$ ($s = 2, 3, \ldots, M$) here to denote the polynomial coefficients, and Equation 6 can be written as the following form:

$$\begin{aligned} A^{(2)}(Y) &= E(Y^2) = E\left[\sum_{i_1=1}^{N} \sum_{i_2=1}^{N} A_{i_1,i_2} \cdot x_{i_1} x_{i_2} + \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{i_3=1}^{N} A_{i_1,i_2,i_3} \cdot x_{i_1} x_{i_2} x_{i_3} + \cdots \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_M=1}^{N} A_{i_1,i_2,\cdots,i_M} \cdot x_{i_1} x_{i_2} \cdots x_{i_M} \right] \\ &= \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} A_{i_1,i_2} E\left(x_{i_1} x_{i_2}\right) + \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{i_3=1}^{N} A_{i_1,i_2,i_3} E\left(x_{i_1} x_{i_2} x_{i_3}\right) + \cdots + \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_M=1}^{N} A_{i_1,i_2,\cdots,i_M} E\left(x_{i_1} x_{i_2} \cdots x_{i_M}\right) \end{aligned} \tag{7}$$

In fact, the polynomial coefficients, say, $A_{i1, i2, \ldots, is}$ need not to be evaluated. Items in Equation 7 can be divided into two categories:

1. The individual effects. The items that include only one joint clearance, e.g. ($A_{j,j}E(x_j^2)$, $A_{j,j,j}E(x_j^3)$, …, $A_{j,j\ldots,j}E(x_j^M)$, $j = 1, 2, \ldots, N$). They represents the individual contribution of one joint clearance to the 2-order original moment of mechanism output error.
2. The interaction effects. The items that include at least two different joint clearances, e.g. ($A_{i1,i2}$ $E(x_{i1}x_{i2}$, $A_{i1,i2,i3}$ $E(x_{i1}x_{i2}x_{i3}$, …., $A_{i1, i2,\ldots, iM}$ $E(x_ix_{i2}\cdots$ $x_{iM})$). They represent the interaction effects of the corresponding joint clearances.

The contribution that one joint clearance made to the mechanism output error should includes the corresponding individual effects and interaction effects. Thus we defined the "total effect" of X_j as the sum of its main effects and all the interaction effects that include X_j. In summary, the individual effect (termed as $A_M^{(2)}(X_j)$), the interaction effect (termed as $A_I^{(2)}(X_j)$), and the total effect (termed as $A_T^{(2)}(X_j)$) of X_j are illustrated as follows:

$$\begin{cases} A_M^{(2)}(X_j) = B_j^{(2)} E\left(x_j^2\right) + B_j^{(3)} E\left(x_j^3\right) + \cdots + B_j^{(M)} E\left(x_j^M\right) \\ A_I^{(2)}(X_j) = \sum\limits_{\substack{i_2=1 \\ i_2 \neq j}}^{N} B_{j,i_2} E\left(x_j x_{i_2}\right) + \left[\sum\limits_{i_2=1}^{N} \sum\limits_{i_3=1}^{N} B_{j,i_2,i_3} E\left(x_j x_{i_2} x_{i_3}\right) - B_j^{(3)} E\left(x_j^3\right) \right] \\ \qquad + \cdots + \left[\sum\limits_{i_2=1}^{N} \sum\limits_{i_3=1}^{N} \cdots \sum\limits_{i_M=1}^{N} B_{j,i_2,i_3,\cdots,i_M} E\left(x_j x_{i_2} x_{i_2} \cdots x_{i_M}\right) - B_j^{(M)} E\left(x_j^M\right) \right] \\ A_T^{(2)}(X_j) = A_M^{(2)}(X_j) + A_I^{(2)}(X_j) \end{cases} \tag{8}$$

where the polynomial coefficients, i.e. B, are only used to represent the model structure, and need not to be evaluated in the importance measures. In Equation 8, the lower the degree of nonlinearity between $\boldsymbol{X}$ and Y, the small of the polynomial coefficients in the higher order items, both in main effects and in interaction effects.

2.2 *Definition of the importance indices*

The value of one item divided by the 2-order original moment of Y, say, $A^{(2)}(Y)$ is used to denote the ratio of the item's contribution. With respect to X_j, the total importance index, denoted as EI_j, is defined as the value of its total effect relative to $A^{(2)}(Y)$:

$$EI_j = \frac{A_T^{(2)}\left(X_j\right)}{A^{(2)}\left(Y\right)} = \frac{A_M^{(2)}\left(X_j\right) + A_I^{(2)}\left(X_j\right)}{A^{(2)}\left(Y\right)} \tag{9}$$

With respect to a group of joint clearances, say, ($X_{j,1}$, $X_{j,2}$, ……, $X_{j,n}$), the total importance index, denoted as $EI_{j1, j2,\ldots, jn}$, can be defined in the same manner:

$$\begin{aligned} EI_{j1,j2,\ldots,jn} &= \frac{A_T^{(2)}\left(X_{j,1}, X_{j,2}, \cdots, X_{j,n}\right)}{A^{(2)}\left(Y\right)} \\ &= \frac{A_M^{(2)}\left(X_{j,1}, X_{j,2}, \cdots, X_{j,n}\right) + A_I^{(2)}\left(X_{j,1}, X_{j,2}, \cdots, X_{j,n}\right)}{A^{(2)}\left(Y\right)} \end{aligned} \tag{10}$$

where $A_M^{(2)}(X_{j,1}, X_{j,2}, \ldots\ldots, X_{j,n})$ denote the sum of the all the main effects of ($X_{j,1}$, $X_{j,2}$, ……, $X_{j,n}$) and all the interaction effects among ($X_{j,1}$, $X_{j,2}$, ……, $X_{j,n}$); $A_I^{(2)}(X_{j,1}, X_{j,2}, \ldots\ldots, X_{j,n})$ denote the sum of all the interaction effects between ($X_{j,1}$, $X_{j,2}$, ……, $X_{j,n}$) and the others; $A_I^{(2)}(X_{j,1}, X_{j,2}, \ldots\ldots, X_{j,n})$ denotes the

total effect, defined as the sum of the main effects and the interaction effects.

2.3 *Monte Carlo based evaluation method*

The derivation of EI indices is based upon the assumption that the model is m times differentiable. However, this assumption not always holds in engineering practice. Even this assumption holds, the computational burden induced by the evaluation of the high order partial derivatives is unaffordable, especially for complicated and implicit models. Thus an efficient method is desirable for the evaluation of the presented EI indices.

Looking at the expression of the main effects and the total effect of X_j, we noticed that each of their items can be seen as the multiplication of X_j with the others. In this respect, if X_i is assigned a value of zero, then not only the main effects, all the items that contain X_i will also be zero. This phenomena inspired an evaluation method for EI indices: given X_j equals zero, evaluate the conditional 2-order original moment of Y, i.e. $A^{(2)}(Y|X_j = 0)$. $A^{(2)}(Y|X_j = 0)$ represents the $A^{(2)}(Y)$ minus all the effects (including the main effects and the interaction effects) relevant to X_j. Then we minus $A^{(2)}(Y|X_j = 0)$ from $A^{(2)}(Y)$ and thus we get the total effect of X_j:

$$A_T^{(2)}\left(X_j\right) = A^{(2)}\left(Y\right) - A^{(2)}\left(Y \mid X_j = 0\right) \tag{11}$$

EI indices of X_j can be evaluated as follows:

$$EI_j = \frac{A^{(2)}\left(Y\right) - A^{(2)}\left(Y \mid X_j = 0\right)}{A^{(2)}\left(Y\right)} \tag{12}$$

Similarly, the evaluation method for the EI indices of a group of joint clearances can be defined as follows:

$$EI_{j1,j2,\ldots,jn} = \frac{A^{(2)}(Y) - A^{(2)}\left(Y \mid X_{j,1}, X_{j,2}, \cdots, X_{j,n} = 0\right)}{A^{(2)}(Y)} \tag{13}$$

The EI indices involve the evaluation of the conditional and unconditional 2-order original moment of Y. The most direct and simple method is the Mote Carlo method. The basic procedure is as follows:

1. Take L samples of $\boldsymbol{X}$ randomly based on their PDFs;
2. Run mechanism model to obtain the corresponding output error values y_i;
3. The 2-order original moment can be evaluated by the following equation:

$$A^{(2)}\left(Y\right) = \frac{1}{L}\sum_{i=1}^{L} y_i^2 \tag{14}$$

The larger the number of samples is, the more accurate the simulation results are.

4. In the same manner, the conditional 2-order original moment of Y, i.e. $A^{(2)}(Y|X_j = 0)$, can also be evaluated given $X_j = 0$.
5. EI indices of X_j can be evaluated based on Equation 12.

3 PROPERTIES DISCUSSION

Saltelli (2002) argued that SA should satisfy the requirements of "*global, quantitative and model free*". By global one means that the technique allows to take into consideration the entire input distribution. By model independent one means that no assumptions on the model functional relationship to its inputs is necessary in order for the SA method to produce accurate results.

With respect to X_j, the individual effects are averaged across the whole distribution range of X_j in the presented EI index. Thus all the possible values of X_j are considered in. Similarly, we average the interaction effects over the whole distribution range of the joint clearances they involved. Thus the presented EI index reflects the contribution of X_j over the whole distribution range of all the joint clearances. At the same time, the linear effect, the higher order effects, and the interaction effects are all considered in. Consequently, the presented EI index possess the property of "global". Same conclusion can also be derived for the EI index of a group of joint clearances.

Although the partial derivatives are used to deduce the importance indices, they need not to be computed in the evaluation of the importance indices. In addition, there are no extra assumptions about nonlinearity and monotonicity of the system model. Thus the indices possess the property of "model free". The presented EI indices reflect the contribution of system input errors to output error, clearly they have the property of "quantitative". Besides, the method presented here can handle the problem of a group of factors.

4 APPLICATION TO CABIN DOOR MECHANISM OF AIRCRAFT LANDING GEAR

4.1 *Problem statement*

The cabin door mechanism is used to cover the landing gear cabin when aircraft has taken off, and is an important subsystem of aircraft landing gear, as shown in Figure 2. Based on kinematic analysis one can get that the components in actuator and linkage mechanism are all under pulling or press forces,

thus they can be simplified into two-force rods. The cabin door is under bending moment, and can be simplified into a beam. In addition, the components can be seen to move in the same plane. Hence the cabin door mechanism can be simplified into a planar linkage mechanism, as shown in Figure 3.

4.2 *Synthesis of joint clearances*

With respect to two-force rods, the joint clearances can be synthesized into the rod length. Based on multi-body dynamic analysis we know that the rod R_{AB}, R_{BD}, R_{BE}, and R_{EF} are all under pulling force, thus their equivalent length can be evaluated:

$$L_{ij} = L_{ij}^0 + e_i + e_j \tag{15}$$

where $L_{i,j}$ denotes the equivalent length of the rod R_{ij}, L_{ij}^0 denotes its design length, and e_i denotes the clearance of the ith joint. Rod R_{CE} is under press force, thus its equivalent length is:

$$L_{CE} = L_{CE}^0 - e_C - e_E \tag{16}$$

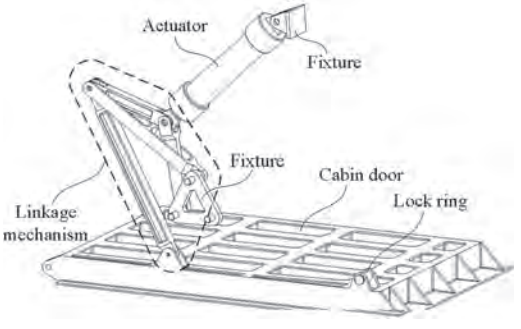


Figure 2. The cabin door mechanism.

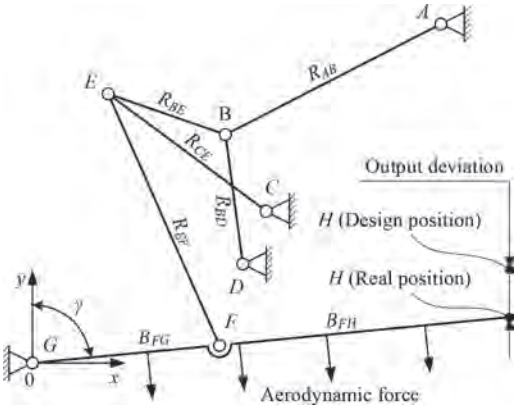


Figure 3. The simplified cabin door mechanism.

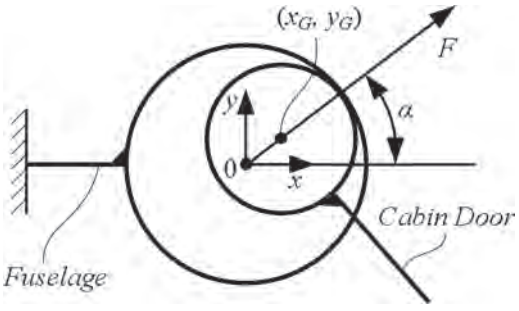


Figure 4. The joint center variation of joint G.

where L_{CE} denotes the equivalent length, L_{CE}^0 is the design length, and e_C and e_E are the clearance of joint C and joint E respectively.

Because the cabin door is under bending moment, the clearance of joint G should be integrated into the variation of the joint center:

$$\begin{cases} x_G = e_G \cdot \cos\alpha \\ y_G = e_G \cdot \sin\alpha \end{cases} \tag{17}$$

where x_G and y_G are the real coordinate of the joint G, e_G is the clearance of joint G, and the angle α is defined as the angle between x axis and the direction of force F, force F is the force applied on fuselage by cabin door, as illustrated in Figure 4.

4.3 *Results and discussion*

The position of joint A, C, D, G, and the design value of rods length are shown in Table 1. The joint clearances are all considered to follow truncated normal distribution, and their distribution parameters are shown in Table 2. In the tables, x_i^0 and y_i^0 denote the joint center position in the coordinate system illustrated in Figure 3, L_i^0 denotes the design value of the rod length, e_i denotes the value of the ith joint clearance.

The importance indices of joint clearances are evaluated by the variance based method, moment independent method, and the presented EI indices respectively. In addition, the partial derivatives (PD) of the output error with respect to joint clearances at zero points are also evaluated. Simulation results are shown in Figure 5. From Figure 5 and Table 1 one can get that e_C has the biggest PD value, mean value, and standard deviation. Thus it should contributes the most to the output error of latch hook. The ranking results of all the three methods consistent with this conclusion. In addition, the three methods produced the same ranking. In order to study the effects variation of distribution parameters on position error of latch hook, importance

Table 1. Design value of joints position and rods length.

Part name	Design value/mm	Part name	Design value/mm
x_A^0	800	L_{AB}^0	510
y_A^0	800	L_{BD}^0	270
x_C^0	510	L_{BE}^0	260
y_C^0	310	L_{CE}^0	420
x_D^0	450	L_{EF}^0	540
y_D^0	200	L_{FG}^0	408
x_G^0	0	L_{FH}^0	592
y_G^0	0		

Table 2. Distribution parameters of joint clearances.

Variable	Mean value/mm	Standard deviation/mm
e_A	0.14	0.06
$e_{B,AB}$	0.12	0.04
$e_{B,BE}$	0.12	0.04
e_C	0.16	0.08
e_D	0.16	0.08
$e_{E,BE}$	0.08	0.04
$e_{E,EF}$	0.08	0.04
e_F	0.08	0.04
e_G	0.06	0.03

indices are evaluated under different mean values and variations of of e_A. Simulation results are shown in Figure 6 and Figure 7 respectively.

From Figure 6 and Figure 7 one can get that: (1) By both the moment independent method and variance based method, importance indices of e_A increased accordingly along with the increase of σ_A, and the importance value of other joint clearances all have different degrees of decline; (2) By both the moment independent method and variance based method, importance indices of all the joint clearances kept roughly constant as μ_A increases; (3) By the presented EI indices, along with the increase of μ_A and σ_A, the importance value and order of e_A are all increased accordingly. Clearly the results of the presented EI indices consistent with with common sense, while that the other two methods are not. This is because the current SA methods, e.g. moment independent method and variance based method, focuses mainly on "uncertainty", rather than on mean shift from the ideal or design values.

In conclusion, the error importance measure is more appreciable when studying the contribution of joint clearances to mechanism output error. Furthermore, under the same modification level of μ_A and σ_A, a bigger variation of EI_A is caused by the shift of μ_A. It means that μ_A has a bigger influence on the variation of EI_A than σ_A. Based on this conclusion,

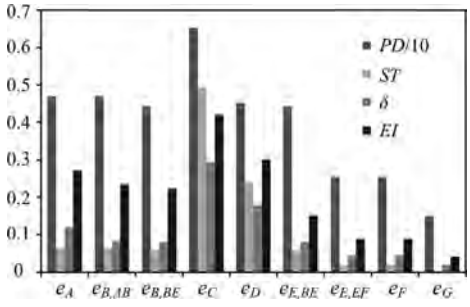


Figure 5. Importance values of joint clearances by different methods.

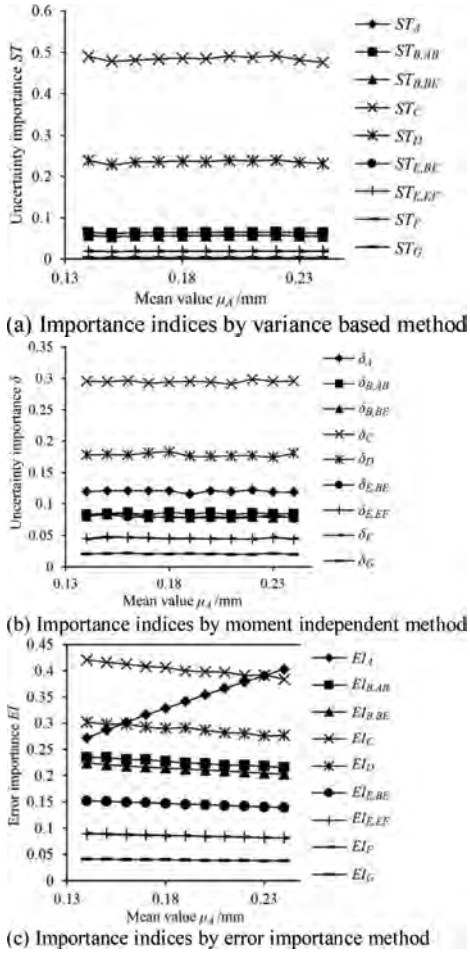


Figure 6. Variation of importance indices along with the shift of μ_A.

more attention should be paid on the decrease of μ_A so as to decrease the contribution of e_A to the position error of the latch hook.

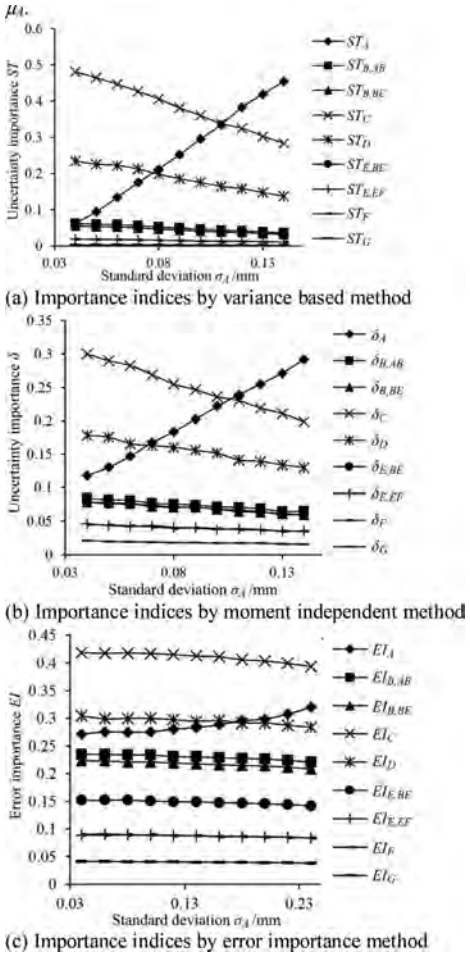


Figure 7. Variation of importance indices along with the shift of σ_A.

5 CONCLUSIONS

In this paper, EI indices to measure the contribution of joint clearances to mechanism output error is presented. The 2-order original moment is used to characterize the deviation of joint clearances or mechanism output error from their design values, i.e. zero. The 2-order original moment of output error is decomposed into a serial of fractions, and the total importance indices for one or a group of joint clearances are defined respectively. The presented importance indices possess the properties of quantitative, model free, and global. An evaluation method based on Monte Carlo simulation is offered, making it convenient to employ.

At last, the presented EI indices is applied to cabin door mechanism of aircraft landing gear. In the application case, the EI indices are used to measure the contribution of joint clearances to the position error of latch hook. For comparison, the variance based method and the moment independent method are also employed. Simulation results revealed that: The presented EI indices can reflect the effects of both variation and mean shift of joint clearances, and the variance based and moment independent methods reflect mainly the effects of variance. Consequently, the presented EI indices is more appreciable for the study of the contribution of joint clearances to mechanism output error.

ACKNOWLEDGEMENTS

This work is financially supported by National Natural Science Foundation of China (Grant No. 51675428).

REFERENCES

Aven, T. & Nøkland, T.E. 2010. On the use of uncertainty importance measures in reliability and risk analysis. *Reliability Engineering and System Safety* 95: 127–133.

Borgonovo, E. & Plischke, E. 2016. Sensitivity analysis: A review of recent advances. *European Journal of Operational Research* 248: 869–887.

Borgonovo, E. 2007. A new uncertainty importance measure. *Reliability Engineering and System Safety* 92: 771–784.

Caro, S., Binaud, N. & Wenger, P. 2009. Sensitivity Analysis of 3-RPR Planar Parallel Manipulators. *Journal of Mechanical Design* 131: 1–13.

Chen, G.L., Wang, H. & Lin, Z.Q. 2013. A unified approach to the accuracy analysis of planar parallel manipulators both with input uncertainties and joint clearance. *Mechanism and Machine Theory* 64: 1–17.

Cheng, Q., Zhang, Z. & Zhang, G. 2015. Geometric accuracy allocation for multi-axis CNC machine tools based on sensitivity analysis and reliability theory. *Proc IMechE, Part C: J Mechanical Engineering Science* 229(6): 1134–1149.

Cheng, Q., Zhao, H. & Zhang, G. 2014. An analytical approach for crucial geometric errors identification of multi-axis machine tool based on global sensitivity analysis. *Int J Adv Manuf Technol* 75: 107–121.

Han, C., Kim, J., Kim, J. & Park, F.C. 2002. Kinematic sensitivity analysis of the 3-UPU parallel mechanism. *Mechanism and Machine Theory* 37: 787–798.

Hanzaki, A.R., Rao, P.V.M. & Saha, S.K. 2009. Kinematic and sensitivity analysis and optimization of planar rack-and-pinion steering linkages. *Mechanism and Machine Theory* 44: 42–56.

Jin, Q., Liu, S.G. & Wang P. 2015. Optimal tolerance design for products with non-normal distribution based on asymmetric quadratic quality loss. *Int J Adv Manuf Technol* 78: 667–675.

Kumaraswamy, U., Shunmugam, M.S. & Sujatha, S. 2013. Unified framework for tolerance analysis of planar and spatial mechanisms using screw theory. *Mechanism and Machine Theory* 69: 168–184.

Li, X., Ding, X.L. & Gregory, S. 2015. Analysis of angular-error uncertainty in planar multiple-loop structures with joint clearances. *Mechanism and Machine Theory* 91: 69–85.

Pianosi, F., Beven, Freer, K., Hall, J.J.W. & Rougier, J. 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software* 79: 214–232.

Saltelli, A. 2002. Sensitivity analysis for importance assessment. *Risk Analysis* 22(3): 579–590.

Sobol, I.M. 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation* 55(1): 271–280.

Taguchi, G., Elsayed, E.A. & Hsiang, T.C. 1989. *Quality engineering in production system.* McGraw-Hill, New York.

Zhang, J.F. & Du, X.P. 2015. Time-dependent reliability analysis for function generation mechanisms with random joint clearances. *Mechanism and Machine Theory* 92: 184–199.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Multiaxial fatigue life prediction for turbine blades using finite element analysis

Jie Zhou, Hong-Zhong Huang, Yan-Feng Li, Junyu Guo & Xiang-Yu Li
Center for System Reliability and Safety, University of Electronic Science and Technology of China, Chengdu, Sichuan, P.R. China

ABSTRACT: The High-Pressure (HP) turbine blade of aero-engine is subjected to high temperature and high pressure, whose life is mainly governed by the fatigue. In order to model the working condition and obtain the critical area, the Finite Element Analysis (FEA) software ANSYS is typically used as a mathematical tool to solve the problems. Besides, the Chaboche model is employed as a constitutive model to describe the response behaviors of the materials under cyclic loadings in the software. The Fatemi-Socie (FS), Wang-Brown (WB) and Redefined Smith-Watson-Topper (Re-SWT) models are utilized to estimate the life of turbine blades on the basis of the critical plane criteria, which considers the effects of hardening due to non-proportional cyclic loading and mean stress on the multiaxial fatigue life of material. Furthermore, these models are applied to proportional and non-proportional loadings. The analysis results indicate that the critical region of HP turbine blade which should be enhanced the intensity or redesigned to reduce the stress concentration.

1 INTRODUCTION

It is known that the relationship between the fatigue life and characteristic of structure is very difficult to describe accurately in engineering applications because of variable working conditions. Fatigue life prediction plays a quite important role in the fatigue failure analysis. The blade in the engine is one of the highest risk components. According to the failure data of components in aircraft, 70% of those are ascribed to the blades (Tao et al., 2000). The blade should have good operating characteristics under different working conditions, and its life may be governed by a series of failure mechanisms, such as fatigue, creep, fracture, yielding, wear, corrosion, erosion, and so on. In order to decrease the incidence of turbine blades failure, all the aspects that degrade engine performances should be taken into account, such as material properties, loading spectrum, structure and working environment. To identify the mechanical behaviors of HP turbine blades in service, it is necessary to take advantage of the mechanical analysis software to model the blade and its working condition.

2 PREDICTION METHODS

The uniaxial fatigue life prediction models may be not suitable for turbine blade which suffers multiaxial loadings during flight missions. The methods of multiaxial fatigue life prediction can be divided into three categories based on different criteria: equivalent stress or equivalent strain criteria, the critical plane theory, and energy-based method. The critical plane theory is one of the most popular and potential theory in multiaxial fatigue life prediction, which is suitable for various loading condition, and it needs another effort to get the LCF (low cycle fatigue) data as preparatory work of multiaxial fatigue life prediction (Socie and Marquis, 1999).

A classical formulation (Kandil et al., 1982) was proposed to solve the multiaxial problems under biaxial loadings, in which the maximum shear strain range $\Delta\gamma_{\max}$ is taken as the main factor leading to failure, and the normal strain $\Delta\varepsilon_n$ on the same plane is regarded as the secondary damage parameter, given as

$$\begin{aligned} &\frac{\Delta\gamma_{\max}}{2}+s\Delta\varepsilon_n=A\frac{\sigma_f'}{E}\left(2N_f\right)^b+B\varepsilon_f'\left(2N_f\right)^c \\ &A=1+v_e+s\left(1-v_e\right) \\ &B=1+v_p+s\left(1-v_p\right) \end{aligned} \tag{1}$$

where v_e and v_p are the elastic and plastic Poisson's ratio; σ_f' is the fatigue strength coefficient; ε_f' fatigue ductility coefficient; b is fatigue strength exponent, c is fatigue ductility exponent, E is the Young modulus, N_f is the number of cycles to failure, s is a material parameter can be obtained from Eq. (2); τ_f' and b_1 are the shear fatigue strength coefficient and exponent, respectively; γ_f' and c_1

are shear fatigue strength ductility coefficient and exponent, respectively; G is the shear modulus, $G = E/2(1+v_e)$.

$$s = \frac{\frac{\tau'_f}{G}\left(2N_f\right)^{b_1} + \gamma'_f\left(2N_f\right)^{c_1} - \left(1+v_e\right)\frac{\sigma'_f}{E}\left(2N_f\right)^{b} - \left(1+v_p\right)\varepsilon'_f\left(2N_f\right)^{c}}{\left(1-v_e\right)\frac{\sigma'_f}{E}\left(2N_f\right)^{b} + \left(1-v_p\right)\varepsilon'_f\left(2N_f\right)^{c}} \quad (2)$$

Besides, the properties related to shear strain-life can be estimated by (Zhu et al., 2017)

$$\tau'_f \approx \frac{\sigma'_f}{\sqrt{3}}, \gamma'_f \approx \sqrt{3}\varepsilon'_f, b \approx b_1, c \approx c_1 \quad (3)$$

Considering the mean stress $\sigma_{n,mean}$ on the maximum shear strain plane, and may result in the loss of fatigue life. A modification of Eq. (1) based on the Morrow model is defined as (Wang & Brown, 1996)

$$\frac{\Delta\gamma_{\max}}{2} + S\Delta\varepsilon_n = A\frac{\sigma'_f - 2\sigma_{n,mean}}{E}\left(2N_f\right)^{b} + B\varepsilon'_f\left(2N_f\right)^{c} \quad (4)$$

Similarly, Fatemi-Socie used the normal tress to supersede the normal strain term in Eq. (4), then presented an improvement involving tension-torsion loadings and non-proportional hardening, shown as (Fatemi and Socie, 1988)

$$\frac{\Delta\gamma_{\max}}{2}\left(1 + k\frac{\sigma_{n,\max}}{\sigma_y}\right) = \frac{\tau'_f}{G}\left(2N_f\right)^{b_1} + \gamma'_f\left(2N_f\right) \quad (5)$$

where σ_y represents the yield strength; and k is the material constant which can be inferred by the uniaxial experimental data, shown in Eq. (6).

$$k = \left[\frac{\frac{\tau'_f}{G}\left(2N_f\right)^{b_1} + \gamma'_f\left(2N_f\right)^{c_1}}{\left(1+v_e\right)\frac{\sigma'_f}{E}\left(2N_f\right)^{b} + \left(1+v_p\right)\varepsilon'_f\left(2N_f\right)^{c}} - 1\right]\frac{2\sigma_y}{\sigma'_f\left(2N_f\right)^{b}} \quad (6)$$

In addition, Smith-Watson-Topper also established a formulation involving the mean stress effect under uniaxial cyclic loadings, and it also can be applied to the multiaxial loads under proportional or nonproportional conditions, and included the principal strain range $\Delta\varepsilon_1$ and the maximum stress $\sigma_{n,\max}$ on the principal strain range plane, given as (Smith et al., 1970)

$$\sigma_{n,\max}\frac{\Delta\varepsilon_1}{2} = \frac{\sigma'^2_f}{E}\left(2N_f\right)^{2b} + \sigma'_f\varepsilon'_f\left(2N_f\right)^{b+c} \quad (7)$$

Similarly, applying the SWT method on the on the maximum shear strain plane, the Eq. (7) can be rewritten as

$$\tau_{\gamma,\max}\frac{\Delta\gamma_{\max}}{2} = \frac{\tau'^2_f}{G}\left(2N_f\right)^{2b_1} + \tau'_f\gamma'_f\left(2N_f\right)^{b_1+c_1} \quad (8)$$

where $\tau_{\gamma,\max}$ is the maximum shear stress on the maximum shear strain plane

There is a simple assumption to decide the critical plane for SWT theory, if $\tau_{\gamma,\max} \geq \sigma_{n,\max}/\sqrt{3}$, Eq. (8) is applied to predict the fatigue life, or Eq. (7) is chosen. The Redefined-SWT (Re-SWT) model is shown in Eq. (9).

$$\begin{cases} \sigma_{n,\max}\dfrac{\Delta\varepsilon_1}{2} = \dfrac{\sigma'^2_f}{E}\left(2N_f\right)^{2b} + \sigma'_f\,\varepsilon'_f\left(2N_f\right)^{b+c}, \\ \qquad \tau_{\gamma,\max} < \sigma_{n,\max}/\sqrt{3} \\ \tau_{\gamma,\max}\dfrac{\Delta\gamma_{\max}}{2} = \dfrac{\tau'^2_f}{G}\left(2N_f\right)^{2b_1} + \tau'_f\gamma'_f\left(2N_f\right)^{b_1+c_1}, \\ \qquad \tau_{\gamma,\max} \geq \sigma_{n,\max}/\sqrt{3} \end{cases} \quad (9)$$

In general, the WB, FS and SWT models have already gained a certain acceptance in the life prediction field of multiaxial fatigue. Additional efforts of uniaxial experimental data (tensile or torsion test) are required to guarantee the accuracy of the models.

3 MODELLING OF HIGH-PRESSURE TURBINE BLADE

In order to ensure the accuracy of the results and reduce the difficulties of meshing, a simplified version of the three-dimensional model for the 1/n HP turbine disc (remain one fir-tree mortise) segment is created, illustrated in Figure 1. The blade-disc coupled system is modeled to gain the critical region of fir-tree mortise, which uses a nonlinear contact analysis technique.

The centrifugal forces and thermal stresses have a significant effect on the static analysis of turbine blade in LCF. The HP turbine blade is cast by high-temperature alloy GH4169, which has good performances of corrosion resistance, high temperature and high strength. The working temperature of blades is about 600–700°*C*. The material properties of GH4169 can be found in (Sun et al., 2010; Shi et al., 2001; Yu et al., 2017), listed in Table 1, where

K' and n' denote the cyclic strength coefficient and cyclic strain hardening exponent, respectively.

The Chaboche model is introduced to model the nonlinear kinematic hardening behavior of GH4169, which considers the effect of viscosity on elasticity (Chaboche, 2008), and it is exploited to model isotropic and kinematic hardening effect in the FEA. The Chaboche model with three evolution parts ($M = 3$) is given as

$$\frac{\Delta\sigma}{2} = \sigma_y + \sum_{i=1}^{M} \frac{C_i}{\gamma_i} \tanh(\gamma_i \frac{\Delta\varepsilon_p}{2}) \tag{10}$$

where the parameters C_i and γ_i $(i = 1,2,\cdots,M)$ can be obtained by the uniaxial test data.

In accordance with the flight mission and filed test for 800 hours of HP turbine blades, the loading spectrum can be specified by 3 different statements: S1 (0-max-0), S2 (idle-max-idle), and S3 (cruise-max-cruise).

After a series of prepared work, the stress-strain analysis of HP turbine blade is simulated to specify the stress concentration region, the fir-tree mortise, where begins with the cracks initiation, cracks propagate under continued cyclic loading, finally leads to the rupture or failure, shown in Figure 2.

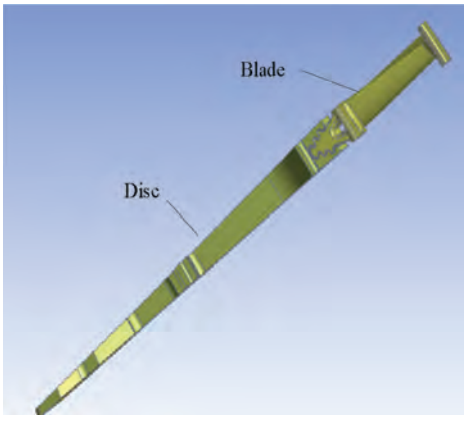


Figure 1. The geometry of blade and blade-disc coupled system.

Table 1. The material properties of GH4169.

T/°C	E/GPa	σ_f' (MPa)	ε_f'
650	182	1476	0.162
b	c	K′(MPa)	n'
–0.086	–0.58	1933	0.1483

Table 2. Loading spectrum of aircraft for HP blades.

Statement	Number of cycles n	Rotational speed ω (rpm)
S1	1306	0-18050-0
S2	2006	9520-18050-9520
S3	24326	16936-18050-16936

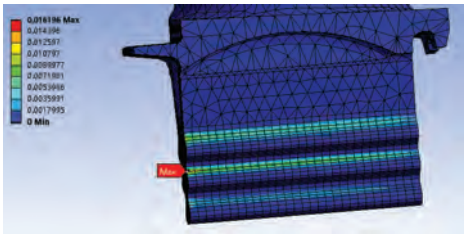


Figure 2. The equivalent plastic strain area of HP turbine blade.

4 EXPERIMENTAL VALIDATION

Referring to the estimation of the applicability and efficiency of the FS, WB and Re-SWT models,

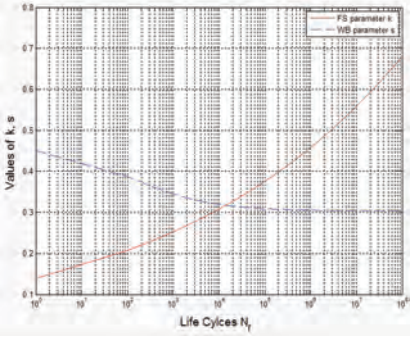


Figure 3. The FS and WB parameters vs. the fatigue life for GH4169.

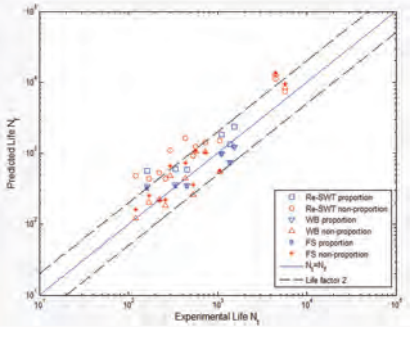


Figure 4. Predicted life N_f vs. the experiment life N_t for GH4169 at $T = 650$°C.

multiaxial fatigue experimental data were conducted on turbine blade alloy GH4169. It complied with the principle of proportional and non-proportional loading introduced in (Sun et al., 2010), which was intended for tension-torsion fatigue test at 650°*C*. Based on the material properties illustrated in Table 1, the material parameters *s* and *k* can be determined by Eqs. (2) and (6), as depicted in Figure 3. The predicted life cycles of Re-SWT, WB and FS models for multiaxial loading are compared with experimental data, is illustrated in Figure 4.

Figure 4 shows a good agreement for WB and FS models, which can be observed for most of the predicted results within the ±2 life factor. The Re-SWT model shows non-conservative results under non-proportional conditions compared with experimental lives.

5 APPLICATION

According to the loading spectrum and results of FEA, the predicted lives of Re-SWT, WB and FS models for HP turbine blade are summarized in Table 3. The life cycle of S3 trends to infinity, the damage is equal to 0. Here, the Miner rule (Miner, 1945), given in Eq. (11), is used to calculate the total damage of 800 hours. Furthermore, the predicted lives, formulated as Eq. (12), are calculated and illustrated in Table 4.

$$D_{800} = \frac{n_1}{N_{f1}} + \frac{n_2}{N_{f2}} + \frac{n_3}{N_{f3}} \tag{11}$$

$$T = 800 \cdot \frac{1}{D_{800}} \tag{12}$$

Table 3. Fatigue life prediction of HP turbine blade.

	S1	S2	S3
n_i	1306	2006	24326
N_{fi}(Re-SWT)	10294	28707	$>10^7$
N_{fi}(WB)	9833	41686	$>10^7$
N_{fi}(FS)	13624	101363	$>10^7$

Table 4. Working hours of HP turbine blade.

	Re-SWT	WB	FS
D_{800}	0.1967	0.1820	0.1164
T(h)	4067	4439	6872

6 CONCLUSIONS

In this paper, a set of experimental data for blade material GH4169 under proportional and non-proportional loadings are used to evaluate Re-SWT, WB and FS models, and the Re-SWT model is not suitable to predict the fatigue life for non-proportional loadings. Furthermore, these three models are also used to estimate the fatigue life based on the FEA of HP turbine blade under serious working conditions, and the prediction results of Re-SWT and WB models are more conservative than that of the FS model.

ACKNOWLEDGMENT

This research is supported by the National Natural Science Foundation of China under contract numbers 51775090 and the Fundamental Research Funds for the Central Universities under contract number ZYGX2014Z010.

REFERENCES

Chaboche, J.L., 2008. A review of some plasticity and viscoplasticity constitutive theories. International Journal of Plasticity, 24(10): 1642–1693.

Fatemi, A., & Socie, D.F., 1988. A critical plane approach to multiaxial fatigue damage including out-of-phase loading. Fatigue & Fracture of Engineering Materials & Structures, 11(3): 149–165.

Kandil, F.A., Brown, M.W., & Miller, K.J., 1982. Biaxial low-cycle fatigue failure of 316 stainless steel at elevated temperatures. In Mechanical Behaviour and Nuclear Applications of Stainless Steel at Elevated Temperatures.

Miner, M.A., 1945. Cumulative damage in fatigue. Journal of Applied Mechanics. 12(3): 159–164.

Shi, C.X., Yan, M.G. & Zhu, Z.Q., 2001. China aeronautical materials handbook. Standards Press of China, Beijing, China.

Smith, K.N., Watson, P. and Topper, T.H., 1970. A stress-strain function for the fatigue of metals. Journal of Materials, 5(4): 767–778.

Socie, D., & Marquis, G., 1999. Multiaxial fatigue. Warrendale, PA: Society of Automotive Engineers.

Sun, G.Q., Shang, D.G., & Bao, M., 2010. Multiaxial fatigue damage parameter and life prediction under low cycle loading for GH4169 alloy and other structural materials. International Journal of Fatigue, 32(7): 1108–1115.

Tao, C., Zhong, P., & Li, R.Z., 2000. Failure analysis and prevention for rotor in aero-engine. National Defence Industry Press, China.

Wang, C.H., & Brown, M.W., 1996. Multiaxial random load fatigue: life prediction techniques and experiments. In ICBMFF4.

Yu, Z.Y., Zhu, S.P., Liu, Q., & Liu, Y., 2017. A new energy-critical plane damage parameter for multiaxial fatigue life prediction of turbine blades. Materials, 10(5).

Zhu, S.P., Foletti, S., & Beretta, S., 2017. Probabilistic framework for multiaxial LCF assessment under material variability. International Journal of Fatigue.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Maintenance of a drone fleet

A. Segal & Y. Bot
BQR Reliability Engineering Ltd., Rishon LeZion, Israel

ABSTRACT: Maintenance and logistics optimization was applied to a fleet of drones, operating from three sites with a central stock. The optimization achieved a Life Cycle Cost reduction of 34% while the fleet availability increased. This paper presents the optimization process, methods and results. Similar methods can be applied to a variety of other fleets.

1 INTRODUCTION

1.1 *Motivation*

The drone industry is one of the fastest growing markets today (Forny & van der Meulen 2017). Drone failures pose both safety and financial risks, yet the drone failure rate is much higher than the failure rate of manned aircraft (Bone & Bolkcom 2003) Therefore, a great need exists for logistics and maintenance optimization of drone fleets.

In this paper we present an example for modelling and optimizing the maintenance policy of a fleet of drones.

1.2 *Case description*

A fleet of 11 surveillance drones, operating from 3 different sites is considered (4 drones in site 1, 4 drones in site 2, and 3 drones in site 3). A central stock services the three sites.

Site surveillance is considered as not operational when more than 1 drone is failed. During such downtime a penalty is paid by the drone operator to the site owner.

In order to optimize the fleet logistics and maintenance policy, the fleet operation had to be modeled. Following is a list of the parameters which were used to create a detailed model of the fleet behavior:

Reliability Data
- Component failure distribution
- Component failure modes
- Drone redundancies
- Operation profile

Maintenance Data
- Component repair/discard policy
- Repair time
- Corrective maintenance
- Preventive maintenance
- Inspections

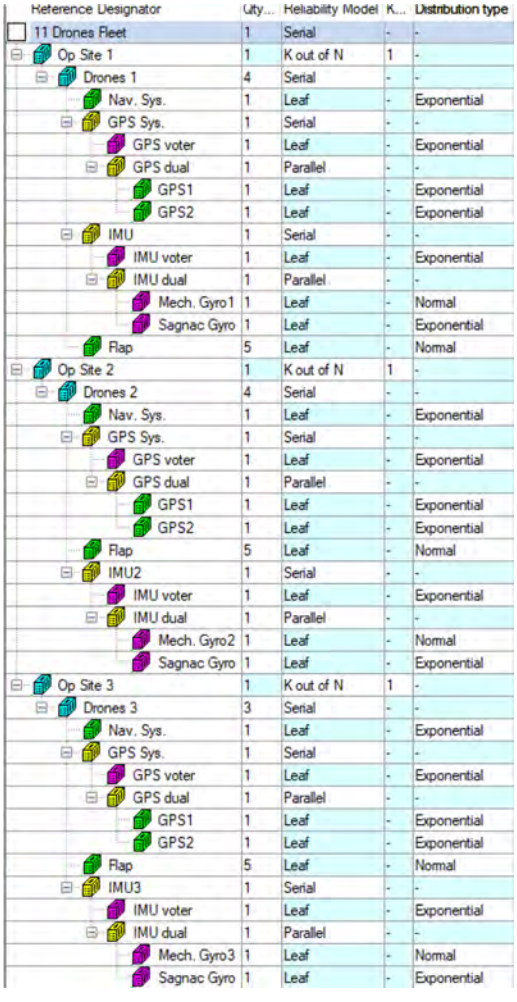
Logistic Data
- Spare parts
- Transportation times
- Procurement time

Financial Data
- Cost of spare parts
- Penalties due to operation agreement
- Corrective maintenance
- Preventive maintenance
- Inspections

Figure 1 (see next page) presents the fleet breakdown tree. The fleet tree includes three main branches (one for each operation site), and under each branch the drones and their components are described. This study focused on several drone sub-systems: Navigation, GPS, Inertial Measurement Unit (IMU), and the flaps.

The "Reliability Model" column in Fig. 1 describes the relevant model for each sub-system. For example: the GPS sub-system includes two redundant GPS units (parallel model).

The "Distribution Type" column in Fig. 1 presents the failure distribution type for each component. Electronic components were assigned an Exponential failure distribution whereas the mechanical gyros and flaps were given a Normal distribution that describes their ageing behavior.



Reference Designator	Qty...	Reliability Model	K...	Distribution type
11 Drones Fleet	1	Serial	-	-
Op Site 1	1	K out of N	1	-
Drones 1	4	Serial	-	-
Nav. Sys.	1	Leaf	-	Exponential
GPS Sys.	1	Serial	-	-
GPS voter	1	Leaf	-	Exponential
GPS dual	1	Parallel	-	-
GPS1	1	Leaf	-	Exponential
GPS2	1	Leaf	-	Exponential
IMU	1	Serial	-	-
IMU voter	1	Leaf	-	Exponential
IMU dual	1	Parallel	-	-
Mech. Gyro1	1	Leaf	-	Normal
Sagnac Gyro	1	Leaf	-	Exponential
Flap	5	Leaf	-	Normal
Op Site 2	1	K out of N	1	-
Drones 2	4	Serial	-	-
Nav. Sys.	1	Leaf	-	Exponential
GPS Sys.	1	Serial	-	-
GPS voter	1	Leaf	-	Exponential
GPS dual	1	Parallel	-	-
GPS1	1	Leaf	-	Exponential
GPS2	1	Leaf	-	Exponential
Flap	5	Leaf	-	Normal
IMU2	1	Serial	-	-
IMU voter	1	Leaf	-	Exponential
IMU dual	1	Parallel	-	-
Mech. Gyro2	1	Leaf	-	Normal
Sagnac Gyro	1	Leaf	-	Exponential
Op Site 3	1	K out of N	1	-
Drones 3	3	Serial	-	-
Nav. Sys.	1	Leaf	-	Exponential
GPS Sys.	1	Serial	-	-
GPS voter	1	Leaf	-	Exponential
GPS dual	1	Parallel	-	-
GPS1	1	Leaf	-	Exponential
GPS2	1	Leaf	-	Exponential
Flap	5	Leaf	-	Normal
IMU3	1	Serial	-	-
IMU voter	1	Leaf	-	Exponential
IMU dual	1	Parallel	-	-
Mech. Gyro3	1	Leaf	-	Normal
Sagnac Gyro	1	Leaf	-	Exponential

Figure 1. Fleet breakdown tree.

2 CALCULATION DETAILS

A commercial software (apmOptimizer) was used for the optimization. The software employs a combination of analytic methods (Birolini 1999) for calculating the fleet Life-Cycle Cost (LCC), and identifying cost and failure drivers. The analytic methods include:

- Markov chains for modelling spare parts supply, demand, and spare waiting times.
- Block mean failure rate calculations that account for component failure distributions, reliability models, scheduled maintenance, inspections and the mission profile.

While analytic calculation is not as flexible as Monte-Carlo simulations, the analytic method is much faster. The speed of evaluating each model allowed for a fast optimization of the maintenance and logistic policies using modified Dynamic Programming.

Dynamic Programming algorithms (Cormen et al. 2009) are ideal for bottom-up optimization of trees where the tree branches are independent.

However, in the fleet case the branches are not completely independent: A central stock services the three sites, therefore a failure in one site affects spare part availability in the other sites. A modified dynamic programming algorithm was used to account for the inter-site dependencies. For example: The Markov chain model that describes the GPS voter spare parts supply and demand accounts for all 11 operating units, serviced by a single central stock.

The optimization goal is to achieve high reliability and availability while minimizing the LCC. Optimization was achieved by using the following optimization modules:

- Optimal LOR: Level Of Repair Analysis—Optimization i.e. Repair/Discard policy. Repair is usually cheaper than buying a new component, however, long repair time (compared to procurement time) may require large and expensive safety stock. Therefore, the discard strategy is sometimes advantageous even when repair is cheaper than buying a new component.
- Optimal PM: Preventive Maintenance Optimization. Periodic maintenance is required for components that exhibit an ageing behavior (failure rate increases with time) and cannot be inspected for degradation. In our case scheduled maintenance is relevant to the mechanical gyros.
- Optimal PdM: Predictive Maintenance Optimization—inspections schedule. Periodic inspections are used to identify flap degradation. Beyond a degradation threshold, preventive maintenance is used to rejuvenate the flaps.
- Optimal I: Inventory Optimization. Optimal I finds the optimal combination of spare parts that minimizes the fleet LCC. The optimal spare part combination is as cheap as possible while ensuring a low probability of downtime due to stock out.

In order to emulate the case of under-maintained fleets, an initial maintenance policy was defined with few spare parts, inspections and scheduled maintenance events. In each optimization step some maintenance actions/spare parts were changed in order to find the optimal combination.

3 RESULTS

Fleet availability at each site as well as the fleet LCC were calculated at each step of the optimiza-

Table 1. Summary of fleet availabilities at the various sites and the total LCC at each optimization step.

Site	Initial A_0	Optimal-LOR	Optimal-PM	Optimal-PdM	Optimal-I
1	99.15%	99.15%	99.15%	99.4%	99.4%
2	99.15%	99.15%	99.15%	99.4%	99.4%
3	99.36%	99.36%	99.36%	99.5%	99.5%
LCC	$47.84M	$46.9M	$45.58M	$31.96M	$31.5M

tion process. Table 1 presents a summary of fleet availabilities at the various sites and the total LCC at each optimization step:

It can be seen from Table 1 that the optimizations resulted in increased fleet Availability and LCC reduction of 34%. Optimal-LOR, Optimal-PM, and Optimal-I decreased the LCC but had a small effect on site availabilities. Optimal-PdM had a strong effect on both availability and LCC. This is not surprising since increased availability means a lower downtime financial penalty.

4 CONCLUSIONS

Fleet operators can reduce operation costs without jeopardizing performance by using analytic tools such as the apmOptimizer.

The modelling and optimization methods which were used in the example are applicable to any fleet, and are therefore relevant to many industries: defense, rolling stock, aviation, and mining.

Furthermore, the method is also good for modelling MRI medical machines, industrial printers, and other sets of identical machines that are operated at different sites and are maintained by the OEM.

REFERENCES

Birolini, A. 1999. Reliability Engineering Theory and Practice, 3rd edition, Springer.

Bone, E. & Bolkcom, C. 2003. Unmanned Aerial Vehicles: Background and Issues for Congress. Report to Congress, Congressional Research Service, Library of Congress, pg. 2.

Cormen, T.H., Leiserson, C.E., Rivest, R.L., & Stein, C. 2009. Introduction to Algorithms, Third Edition. The MIT Press, Cambridge, Massachusetts, London, England. Pg. 359.

Forny, A.A. & van der Meulen 2017. R. Gartner. http://www.gartner.com/newsroom/id/3602317.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Statistical test planning using prior knowledge—advancing the approach of Beyer and Lauster

A. Grundler, M. Bartholdt & B. Bertsche
Institute of Machine Components, University of Stuttgart, Stuttgart, Germany

ABSTRACT: An approach is presented to include prior knowledge in the test planning of product reliability based on the approach of Beyer and Lauster (Beyer and Lauster, 1990) such that several and different sources of prior knowledge can be accounted for. Furthermore, this advanced approach is independent from Beyer/Lauster's prerequisite that the confidence C of the prior knowledge in the form of $R_0(t)$ is to be at $C = 63.2\%$. A nomogram is presented consistent with the pragmatic suggestions of Beyer/Lauster (Beyer and Lauster, 1990) yet extending the original version's applicability. This paper provides a mathematically sound extension to the original paper broadening its applicability to more than one source of prior knowledge while liberating the constraint of $C(R_0(t)) = 63.2\%$ while also allowing for a partial transfer of prior knowledge and accounting for accelerated lifetime tests.

1 INTRODUCTION

Physical product testing is an important tool to validate a product's maturity with respect to reliability. It is used to demonstrate the product's reliability. When planning such a validation test, budget, time and accuracy have restrictive effects. Budget may be limited and time to market may be constrained while accuracy of the derived conclusion is to be maximized. To meet the requirements of efficient and yet effective reliability validation, prior knowledge can be used, which could stem from former development stages, previous product generations or similar products such as benchmarks etc.

2 APPROACH OF BEYER AND LAUSTER

The approach of Beyer and Lauster is based on Bayes' theorem, which allows to link information on the product's reliability from current tests with information from prior knowledge, e.g. tests at a former development stage, previous product generations etc. Beyer and Lauster differentiate the approach in (Beyer and Lauster, 1990) into one part dealing with constant failure rates and consequently one part with non-constant failure rates ($\lambda_0 = f(t)$, i.e. $\beta_0 \neq 1$).

2.1 *Test planning with constant failure rate* λ_0

The failure probability of test samples with constant failure rates is based on the Poisson distribution (Kececioglu, 2002). The confidence level is then calculated by equation (1).

$$C = 1 - \sum_{i=0}^{x} \frac{\left(nt_t \lambda_{max}\right)^i}{i!} e^{-nt_t \lambda_{max}} \tag{1}$$

where n stands for the number of samples, t_t test time, λ_{max} the desired maximum failure rate and x the number of failures during the test.

Prior knowledge is considered through the corresponding failure rate λ_0 with t_{t0} and sample size n_0 ("0" connoting the prior knowledge). If the test to which the prior knowledge relates to is passed without failed samples, the failure rate's density function is then given by (2) with $n_0 t_{p0} = 1/\lambda_0$.

$$f(\lambda) = \frac{1}{\lambda_0} e^{-\frac{\lambda}{\lambda_0}} \tag{2}$$

If a Poisson distribution is supposed to describe the conditional probability of the current test results, the confidence as shown by (3) is calculated by means of Bayes' theorem.

$$C = 1 - \sum_{i=0}^{x} \frac{1}{i!} \left(\left(nt_t + \frac{1}{\lambda_0} \right) \lambda_{max} \right)^i e^{-\left(nt_t + \frac{1}{\lambda_0} \right) \lambda_{max}} \tag{3}$$

Note that the prior knowledge in form of the constant failure rate λ_0 needs to satisfy $C_0(\lambda_0) = 63.2\%$ which limits the approach's applicability.

2.2 *Test planning with non constant failure rates*

If failed samples are not replaced during a test—which is typically the case in practice—the Poisson

distribution fails to be able to describe the product's reliability. The density of reliability R is then described by means of a Beta distribution, cf. (4) for no failures.

$$f\left(R_0\right) = n_0 R_0^{n_0 - 1} \tag{4}$$

After combining the prior knowledge from (4) with the result of the current test encompassing n samples with x failures, a lifetime ratio L_V (test time t_l over the desired lifetime t_s) and shape parameter β of the underlying Weibull distribution, the confidence can be calculated by (5).

$$C = 1 - \sum_{i=0}^{x} \binom{n + \frac{n_0}{L_v^{\beta}}}{i} R\left(t_s\right)^{L_v^{\beta}(n-i)+n_0} \left(1 - R\left(t_s\right)^{L_v^{\beta}}\right)^{i} \tag{5}$$

The basic relation between the confidence level and the reliability in general, i.e. without failures and neglecting lifetime ratios etc. follows (6).

$$C_0 = 1 - R_0^{n_0} \tag{6}$$

With C_0 = 63.2%, an expression for n_0 results from (6) not containing a C_0 which is fixed.

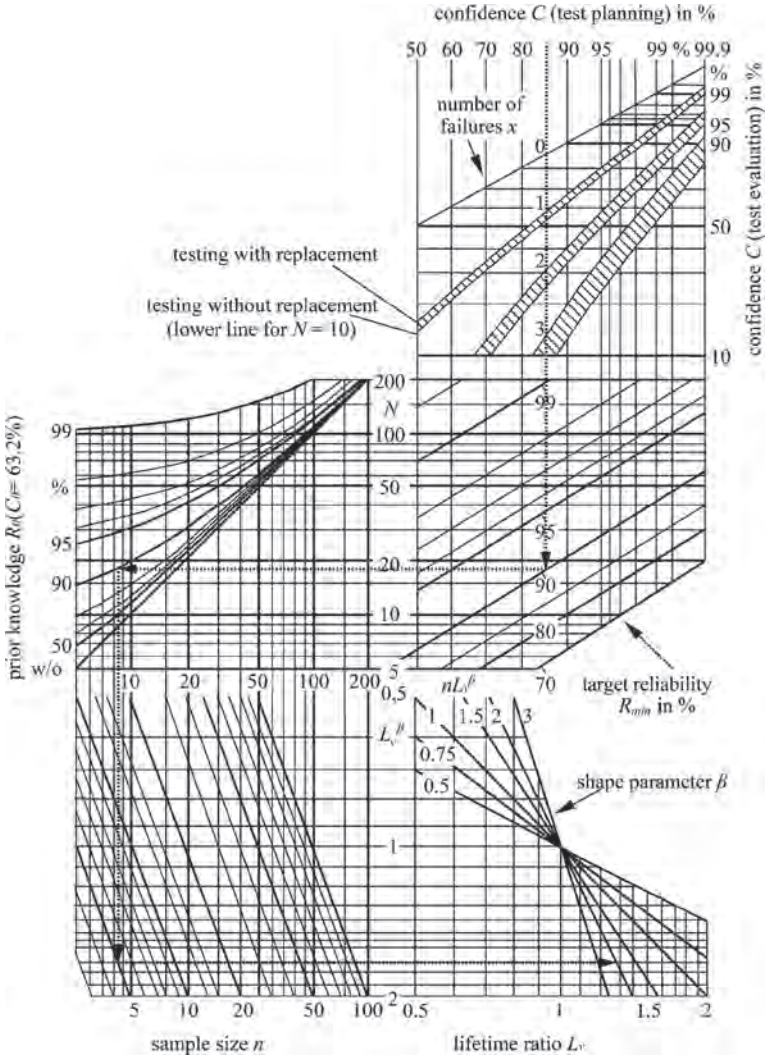


Figure 1. Nomogram for test planning with prior knowledge in the form of R_0 (C_0 = 63.2%) (Beyer and Lauster, 1990).

$$n_0 = \frac{1}{\ln\left(\frac{1}{R_0}\right)} \tag{7}$$

Equation (7) combined with (5) results in (8):

$$C = 1 - R(t_s)^{L_v^\beta n + 1/\ln\left(\frac{1}{R_0}\right)} \cdot \sum_{i=0}^{x} \binom{n + \frac{1}{L_v^\beta \ln\left(\frac{1}{R_0}\right)}}{i} \left(\frac{1 - R(t_s)^{L_v^\beta}}{R(t_s)^{L_v^\beta}} \right)^i \tag{8}$$

Please note that Equation (8) is only valid for $C_0 = 63.2\%$. For $R_0 \rightarrow 0$, i.e. for inexistent prior knowledge, (8) is reduced to (9):

$$C = 1 - R(t_s)^{L_v^\beta n} \sum_{i=0}^{x} \binom{n}{i} \left(\frac{1 - R(t_s)^{L_v^\beta}}{R(t_s)^{L_v^\beta}} \right)^i . \tag{9}$$

Equation (9) cannot be derived directly. Prior knowledge has to be transferred fully. In (Beyer and Lauster, 1990) a nomogram is offered for the practical application of equations (8) and (9) for test planning, cf. Figure 1. (The nomogram is also published and discussed in (Krolo, 2004, pp. 41–47)).

3 EXTENSION OF THE APPROACH OF BEYER AND LAUSTER

The extension of the original approach by Beyer and Lauster (cf. Chapter 2) is motivated by the following aspects:

- The consideration of prior knowledge in the form of R_0 at arbitrary confidence levels C_0.
- The consideration of several sources of prior knowledge.
- The consideration of accelerated lifetime tests by introducing an acceleration factor r in line with (Krolo *et al.*, 2002).
- The consideration of a nuanced incorporation of the existing prior knowledge by introducing a transformation factor Φ.

In order to allow for the fact that the prior knowledge might stem from previous product generations, former development stages or similar products, a transformation factor Φ is suggested in (Krolo, 2004). It ensures the ability to nuance the transfer of prior knowledge other than 100%, i.e. $0 \leq \Phi \leq 1$. Ways to quantify the transformation factor, i.e. the degree to which products or product generations are similar, are found in e.g. (Krolo, 2004; Hitziger and Bertsche, 2005; Schweizer *et al.*, 2015).

Ways to quantify the acceleration factor r are found in e.g. (Jakob *et al.*, 2017).

Both the acceleration factor r and the transformation factor Φ can be included in (8). Thus, (10) results, where the presumed confidence of R_0 is still $C_0(R_0) = 63.2\%$.

$$C = 1 - R(t_s)^{L_v^{\beta_t} r^{\beta_t} n + \Phi/\ln\left(\frac{1}{R_0}\right)} \cdot \sum_{i=0}^{x} \binom{n + \frac{\Phi}{L_v^{\beta_t} r^{\beta_t} \ln\left(\frac{1}{R_0}\right)}}{i} \left(\frac{1 - R(t_s)^{L_v^{\beta_t} r^{\beta_t}}}{R(t_s)^{L_v^{\beta_t} r^{\beta_t}}} \right)^i \tag{10}$$

Equation (10) contains a shape parameter β_t accounting for a possibly different shape of the Weibull distribution observed in the results of the accelerated test with the acceleration factor r as compared to the one observed in the field under regular conditions. It can be observed, that β_t tends to be slightly greater than β while it is commonly assumed that $\beta_t = \beta$ (Juskowiak and Bertsche, 2016).

From (6) results (11), similar to (7):

$$n_0 = \frac{ln(1 - C_0)}{\ln(R_0)} . \tag{11}$$

Prior knowledge with individually arbitrary confidence levels C_{0j} can be combined in the way (12) suggests. A total sample size n_{0sum} is thus calculated stemming from k different sources of prior knowledge. In doing so, individual transformation factors Φ_j account for different degrees of similarity between the product of interest and the source of prior knowledge.

$$n_{0sum} = \sum_{j=1}^{k} \Phi_j \frac{ln(1 - C_{0j})}{\ln(R_{0j})} \tag{12}$$

With (12), (10) is amended to (13):

$$C = 1 - \left(R(t_s)^{L_v^{\beta_t} r^{\beta_t} n + \sum_{j=1}^{k} \Phi_j \frac{\ln(1 - C_{0j})}{\ln(R_{0j})}} \right) \cdot \sum_{i=0}^{x} \binom{n + \sum_{j=1}^{k} \Phi_j \frac{\ln(1 - C_{0j})}{L_v^{\beta_t} r^{\beta_t} \ln(R_{0j})}}{i} \left(\frac{1 - R(t_s)^{L_v^{\beta_t} r^{\beta_t}}}{R(t_s)^{L_v^{\beta_t} r^{\beta_t}}} \right)^i . \tag{13}$$

4 A NOMOGRAM ENHANGING THE TEST PLANNING'S APPLICABILITY

4.1 *The Nomogram as proposed by (Beyer and Lauster, 1990)*

Beyer and Lauster offer in (Beyer and Lauster, 1990) a nomogram which aids decision makers at including prior knowledge into statistical test planning. Figure 2 shows the original version able to account for one source of prior knowledge in the form of $R_0(t)$ with confidence C_0 = 63.2%. This prior knowledge is assumed to be fully transferrable to the current product's version and the current tests need to be not accelerated.

4.2 *Example #1 – Beyer and Lauster's approach*

This example is extracted from (Beyer and Lauster, 1990). It is used in order to demonstrate consistency with the advanced approach for statistical test planning elaborated in this paper.

The target reliability level of a given product is set to be $R_{min}(t_s$ = 20000 h) = 90%, i.e. B_{10} = 20000 h. This reliability target is to be established with a confidence of C = 85%. From a previous and

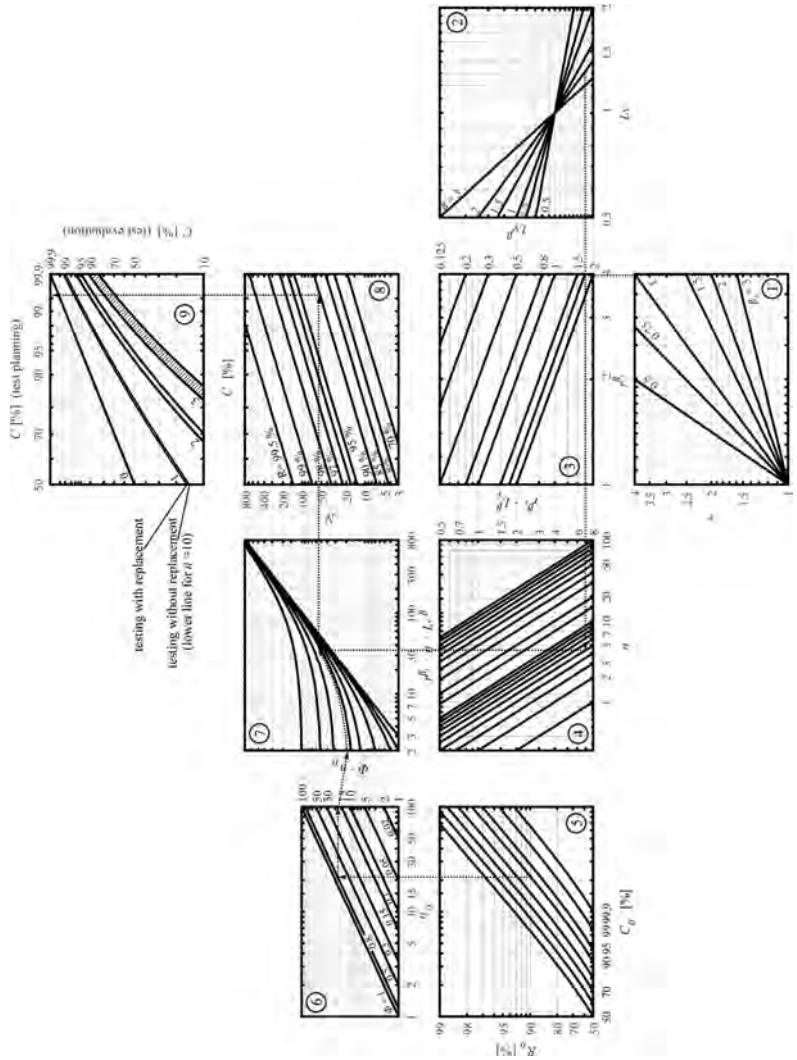


Figure 2. Nomogram for test planning with prior knowledge in the form of R_0 (C_0) according to the advanced approach.

very similar product generation it is known that $R_0 = 90\%$ and the then derived $R_0(t_s)$ follows a Weibull distribution with a shape parameter of $\beta_0 = 2$ and $\beta_0 = \beta$. From Figure 2 it follows, that

$$n \cdot L_v^{\beta} = 8.5 \tag{14}$$

If e.g. $n = 5$ samples are allowed for, then the lifetime ratio $L_v \approx 1.3$. The failure free test time per sample therefore amounts to about 26000 h, cf. (13). It becomes clear from (8) that if prior knowledge was neglected, 5 additional samples would be required to demonstrate the target reliability under the otherwise same conditions.

4.3 *Development of a nomogram for the advanced approach*

A nomogram is developed aiding at practical test planning in line with this paper's ambition as described above. It accounts for prior knowledge (R_{0j} with C_{0j}), which can be partially or fully transferred to the current product for which a certain reliability is to be demonstrated and whose demonstration tests may see accelerated test conditions.

Figure 2 illustrates the adapted version of the nomogram. Herein, example #2 is implemented. Its different fields are numbered.

Field 1: The acceleration factor r (here, $1 \le r \le 4$) of the current product's test (or the test of the product's current state) is taken to the power of β_t.

Field 2: The lifetime ratio L_v (test time t_t/t_s) of the current product's test is taken to the power β.

Field 3: The results from field 1 and field 2 get multiplied. The product is found by extending the resulting point in field 1 vertically and that of field 2 horizontally, followed by an extension parallel to the oblique lines in field 3 upon arrival at field 3.

Field 4: The result from field 3 gets multiplied with the sample size of the current product's test n. The sample size can be chosen in the form of one of the oblique lines. The point found in field 3 is horizontally extended to the given sample size line. The then found point in field 4 is extended vertically to serve as input to field 7.

Field 5: The prior knowledge is put into consideration here. Its reliability level R_0 as well as the corresponding confidence level C_0 are chosen. C_0 is represented by one of the oblique lines. Field 5 calculates what could be considered the sample size of the prior knowledge n_0:

$$n_0 = \frac{\ln\left(1 - C_{0j}\right)}{\ln\left(R_{0j}\right)}. \tag{15}$$

This version of the nomogram deals with one source of prior knowledge, albeit it can be seen from (12) and (13) that k sources can be dealt with. The prior knowledge's fictional sample size n_0 would then simply be the sum of the n_{0j} found through field 5.

Field 6: The result from field 5 (cf. (13)) gets multiplied with the transformation factor Φ which is represented by the oblique line adequate to the degree to which the prior knowledge is trusted to represent the current product (or the current product's state). The thus found point in field 6 is extended horizontally toward field 7.

Field 7: Here, the results from fields 4 and 6 are summed up to what could be considered the total sample size N containing the sample size n of the current product's test as well as the fictional sample size n_0 calculated previously. The corresponding point in field 7 is where the result from field 6, extended along the adequate plotted line in field 7 (i.e. the one entered at) intersects with the vertically extended line off of field's 4 resulting point.

Field 8: The desired reliability $R(t_s)$, represented by the corresponding oblique line, is taken to the power of N. C results if no failures were observed during the test of the n specimen, cf. (13).

Field 9: Field 9 accounts for the number of observed failures x during the test of the n specimen. Here, $0 \le x \le 3$. The resulting C can be concluded upon by extending the point of intersection of field 9 horizontally to the right. Each upper of the two equidistant lines represents the relationship between C and the number of failures with the supposed replacement of the failed units. In case that failed units are not replaced, Beyer and Lauster's suggestion is followed since (13) cannot be illustrated in the same way as (8) (Beyer and Lauster, 1990). A hatched area is introduced which is valid for large n. For large n, C from (13) is approximately equal to C from (8). Although it is suggested in (Beyer and Lauster, 1990) that this assumption holds true for $n \ge 10$, the authors recommend it for $n \ge 20$ (the deviation then amounting to 2,1%, given the conditions of example #1).

Note that the order of application through the described fields depends on the goal of the given lifetime test planning or evaluation. If the goal is to determine a confidence C which can be assured by the combination of a given test with prior knowledge, the order is one to nine. If the goal is to determine how many specimen n need to be tested under given conditions in order to assure a certain reliability R with a specific C_0, then the order is 9 through 4 and 1 through 4.

4.4 *Example #2*

This example adds an acceleration factor r, a confidence level C_0 and a transformation factor Φ to example #1. If $r = 2$, $C_0 = 90\%$ and $\Phi = 0.8$, then $C = 99.55\%$ results. As compared to the original Beyer and Lauster approach, where $C_0 = 63.2\%$, the resulting confidence was thus increased by $\approx 15\%$.

The implementation of this example is shown in Figure 2. The line laid over the structure described in chapter 4.3 constitutes the conditions of example #2 and their exemplary implementation in the developed nomogram.

5 SUMMARY

An extension to Beyer and Lauster's approach for including prior knowledge into statistical test planning was presented. For both, the original and the extended version of the approach, a nomogram was developed. It aids at the practical implementation of the definition of test plans and at analyzing given test settings in terms of the reliability's confidence $C(R)$.

The presented extension is capable of accounting for more than one source of prior knowledge with arbitrary confidences, results from accelerated tests of arbitrary length. Furthermore, it allows for a just partial transfer of prior knowledge to the product's current state based on the assessment of their similarity.

REFERENCES

Beyer, R. and Lauster, E. (1990), "Statistische Lebensdauerprüfpläne bei Berücksichtigung von Vorkenntnissen. Translated: Statistical Lifetime Test Plans Considering Prior Knowledge", QZ 35, *Qualität und Zuverlässigkeit*, No. 2, pp. 93–98.

Hitziger, T. and Bertsche, B. (2005), "An approach to determine uncertainties of prior information—the transformation factor", in Kołowrocki, K. (Ed.), *Advances in safety and reliability: Proceedings of the European Safety and Reliability Conference (ESREL 2005), Tri City (Gdynia—Sopot—Gdansk), Poland, 27–30 June, 2005*, vol. 1, Balkema, Leiden, pp. 843–849.

Jakob, F., Kimmelmann, M. and Bertsche, B. (2017), "Selection of Acceleration Models for Test Planning and Model Usage", *IEEE Transactions on Reliability*, Vol. 66 No. 2, pp. 298–308.

Juskowiak, J. and Bertsche, B. (2016), "Application and Simulation Study of Stress-Dependent Weibull Lifetime Models", *International Journal of Reliability, Quality and Safety Engineering*, Vol. 23 No. 02, 1650008-1-1650008-20.

Kececioglu, D. (2002), *Reliability engineering handbook: Volume 2*, DEStech Publications, Inc., Lancester, Pensylvania.

Krolo, A. (2004), "Planung von Zuverlässigkeitstests mit weitreichender Berücksichtigung von Vorkenntnissen", Dissertation, Institut für Maschinenelemente, Universität Stuttgart, Stuttgart, 2004.

Krolo, A., Rzepka, B. and Bertsche, B. (2002), "Application of Bayes statistics to reduce sample-size, considering a lifetime-ratio", in *Annual Reliability and Maintainability Symposium. 2002 Proceedings (Cat. No. 02CH37318), Seattle, WA, USA, 28–31 Jan. 2002*, IEEE, pp. 577–583.

Schweizer, V., Jakob, F., Bartholdt, M. and Bertsche, B. (2015), "Assessment and Visualization of Reliability Relevant Product Characteristics for the Application of Bayes Analysis", in Pham, H. (Ed.), *Proceedings/21st ISSAT International Conference Reliability & Quality in Design: August 6–8, 2015, Philadelphia, PA, U.S.A*, Internat. Soc. of Science and Applied Technologies, Piscataway, NJ, pp. 227–231.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Advances in component fault trees

Bernhard Kaiser
Assystem Germany GmbH, Berlin, Germany

Daniel Schneider & Rasmus Adler
Fraunhofer Institute for Experimental Software Engineering (IESE), Kaiserslautern, Germany

Dominik Domis & Felix Möhrle
Department of Software Engineering: Dependability, The University of Kaiserslautern (TUK), Kaiserslautern, Germany

Axel Berres
German Aerospace Center (DLR), Institute of Flight Systems, Braunschweig, Germany

Marc Zeller, Kai Höfig & Martin Rothfelder
Siemens AG, Corporate Technology, Munich, Germany

ABSTRACT: Component Fault Trees (CFTs) were invented in 2003 as a compositional extension to fault trees to better reflect the technical architecture of a system in its safety analysis model. Since then, a lot of research has been contributed regarding semantic extensions, evaluation techniques, and tighter linking between system and safety models. This paper addresses three main objectives. First, we summarize the most important contributions and shape a vision of better integrated system modeling and safety analysis. Second, we push forward standardization and sketch a new evaluation scheme for quantitative analysis using mdd. Lastly, an outlook on future improvement ideas is given to make CFTs a viable technique for loosely coupled systems and Cyber-Physical Systems.

1 INTRODUCTION

Fault Tree Analysis (FTA) is one of the most prominent safety and reliability analysis techniques. FTA has been applied for over 50 years and was standardized by several official or de facto industry standards like Vesely et al. (1981) and IEC 61025. Standardization in practice has also taken place through available tool solutions, such as Fault-Tree+, BlockSim, SafetyOffice, medini analyze and many more. Its application is recommended or even required by various safety standards such as IEC 61508 and ISO 26262.

FTA helps engineers identifying causes and influence factors for a supposed failure (called top event) by iteratively seeking backwards in the causal chain. Top events can be hazards (situations that may provoke accidents) in safety engineering, or system failures and unavailability in the domain of reliability engineering. Causes and influence factors are represented in a tree structure. The intermediate failures in the tree structure are joined by so-called gates, in particular the AND gate (indicating that all influencing factors together are necessary to cause the output failure) or the OR gate (indicating that at any of the influence factors causes the output failure). Originally, only Boolean logic connectives such as AND, OR, and N-out-of-M (sometimes called voter gate) were provided, as well as NOT (which cannot be handled by all evaluation algorithms). In later years, dynamic extensions like the Priority-AND gate or gates formulating constraints, such as the Sequence-Enforcing gate have been introduced. However, these gates require a closer look into their semantics, as they do no longer express Boolean proposition logic. The gates are usually represented by American style logic symbols, as used in circuit diagrams, or by European style rectangular symbols. At the leaves of the tree, the fundamental causes (basic events) for failures are represented. In the context of propositional logic, basic events represent failure conditions or failed states of technical components (e.g. "Relay 3 short circuit"). For quantitative analysis, failure probabilities (instantaneous values or functions over time) are assigned to the basic events. This is usually done by specifying a probability density function (e.g. exponential or Weibull) for the transition to the failed state along with its parameters, such as the

failure rate λ. Due to the tree structure, it is not possible to graphically indicate that a basic event may lead to different consequences in the tree (e.g. power supply failure influencing multiple parts of the system). This information can only be provided to evaluation tools by labeling the respective basic event as *repeated event*.

Fault Trees (FTs) can be analyzed in different ways. The most common qualitative analysis is the identification of "Prime Implicants" or "Minimal Cut Sets (MCSs)", i.e. combinations of basic events that cause the top-level failure or hazard. Cut Sets with only one element indicate single points of failures. The most important quantitative analysis is calculating the top-level failure probability (e.g. hazard probability or system unavailability). Other kinds of quantitative analysis are, for instance, importance analyses that unveil the relative importance of single failures w.r.t. the top event probability and give valuable indications where optimization makes sense and where not.

1.1 *Traditional ways of structuring fault trees*

FTs for real industry systems tend to be quite large, often comprising more than 1000 basic events. Therefore, it is not possible to present them on a single page, so some structuring measures are necessary. Traditionally, there are two means of structuring: (1) the transfer symbol and (2) splitting the trees into independent subtrees (modules). While the first method is purely graphical, the second has a semantic connotation, since it is possible to calculate the probability of failure of a subtree and include it in the parent tree as a basic event. This reduces computing time and enables IP protection by allowing component suppliers to deliver the aggregated failure rate instead of having to disclose the entire tree structure. Obviously, the second method is only feasible if a component's FT is a true subtree without any shared events. Multi-instancing of components is possible by allowing several instances of a module. Unfortunately, most available tools do not support this and the user has to copy and paste subtrees, which is prone to errors. Graphically, both methods appear the same, because subtrees are also separated by using the transfer symbol. Both methods are not formally related to the technical architecture of the system, so it is up to the user to divide the FT into meaningful parts that correspond to the components of the technical system.

1.2 *Introduction to CFTs*

To overcome these and other restrictions of traditional FTs, Component Fault Trees (CFTs) were proposed as a new component concept by Kaiser et al. (2003). The idea was to cut out parts of FTs that correspond to technical components and turn them into reusable units with input and output failure ports. At the same time, the tree structure was extended towards a Directed Acyclic Graph (DAG) structure. This avoids the artificial splitting of common cause errors into multiple "repeated events". Instead, it is possible for more than one path to start from the same basic event or subgraph. In addition, it is possible to have more than one top event, such as an accident when a primary failure coincides with the failure of a countermeasure, but only a system unavailability when the same primary failure occurs while the countermeasure is working. This saves analysts time, since large parts of the failure logic can be shared between the two top events.

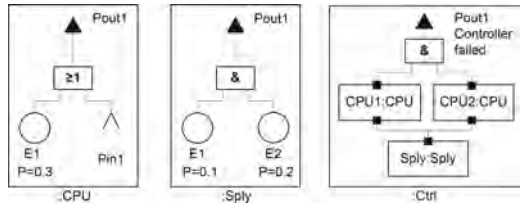


Figure 1. Example of a simple CFT.

Fig. 1 shows an exemplary controller system, including two redundant CPUs (two instances of the same component) and one power supply (which would be a repeated event in traditional FTA). The controller is unavailable if both CPUs are in the state "failed". The inner FT of the component type CPU is shown on a separate screen; as the CPUs are of identical type, they only have to be modeled once. The failure of a CPU can be caused by some inner basic event "E1" (the repetition of the ID "E1" in several components is not a problem, as each component constitutes its own name space). The failure of the CPU can also be caused by an external failure cause which is connected via an input port. As both causes result in a CPU failure, they are joined via a 2-input OR gate. The power supply is modeled as a separate component. Instead of a single large FT, the model consists of small, reusable and easy-to-review components.

1.3 *Quantitative evaluation using BDDs*

The most important quantitative analysis for CFTs, just as for traditional FTs, is the calculation of top event probabilities. The result is usually expressed as probability for a given time period or point in time, or by an equivalent failure rate. For traditional FTs, two approaches are commonly used: collecting all MCSs or prime implicants and to sum up their probabilities, or to transform the Boolean term represented by the FT into a Binary Decision Diagram (BDD). In the latter case, the

Figure 2. BDD fragments for the example CFT: main component (left) and subcomponent C1 (right).

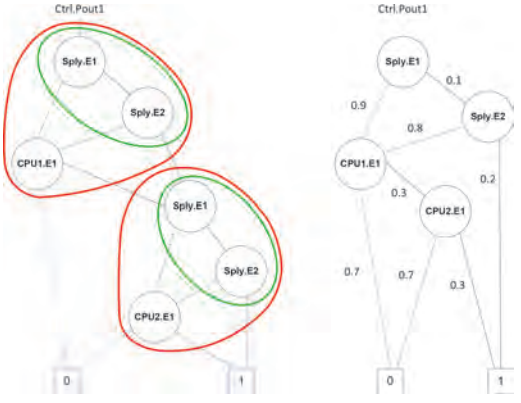


Figure 3. BDD integration, reduction and probability calculation.

event probabilities along each edge are multiplied and the final result calculated as the sum of probabilities of all paths. This BDD-based algorithm works also for CFTs. An example is given in Fig. 2 where the BDD for the main CFT in Fig. 1 is shown.

Fig. 3 shows the integrated BDD after inserting the fragments for the sub components into the main component at the position of the output port nodes. In this example, the probability is calculated as $P_{system\_fail} = 0.1 * 0.2 + 0.9 * 0.3 * 0.3 = 0.101$. Note that we use fixed probabilities to facilitate understanding, while in most cases in practice failure rates are applied.

When nodes representing CFT ports occur in the BDD, they have to be replaced by the sub-BDD that represents the corresponding FT segment with its root at that port. The algorithm for CFTs is explained in detail by Förster (2006).

Kaiser (2005) showed that CFTs not only help structuring large FTs in a more intuitive way, but that they also accelerate the quantitative analysis significantly. The reason is that the logical structure of the FT fragment of a component can be parsed and precompiled into a BDD, which makes up a significant part of the computation time. While optimizing a BDD by finding the lowest number of required nodes is NP-complete and thus not feasible in practice (see Bollig and Wegener 1996), the component concept leads to smaller BDD fragments that reduce the number of variables even if the ordering is not optimal. Furthermore, the nodes that belong to internal failure events of each component can be combined in one part, and the nodes that belong to input ports can be combined in the other part of the reduced structure of the BDD fragment. By a set of examples it could be shown empirically that this leads to an acceptably good variable ordering. The reduced BDD fragments for each port of each component can then be stored for later use. They can be instantiated many times if multiple instances of a component exist in a system (e.g. train braking system with multiple wheel brakes). Kaiser (2005) showed an algorithm to store precompiled and optimized BDD fragments and insert them into one virtual BDD node when instating.

The possibility of reducing/ordering and storing BDD fragments also helps suppliers to protect their intellectual property. In many cases, industrial companies refuse to hand over their safety analysis in full detail to the OEM, because a well-structured and well-documented FTA or FMEA allows reverse engineering the product to some extent. CFTs allow handing over just the signature of a component (name and types of failure input and output ports) plus a precompiled BDD. The BDD contains the full logical terms and the probability distribution of the inner failure events, but with the variables in a random order and without the names of any internal or immediate failure events. This is very similar to the principle of information hiding in software engineering, where compiled binaries and header files with signatures are handed out, but not the full program code.

2 SUBSEQUENT EXTENSIONS

2.1 *Integration with Markov Chains (MCs)*

An important extension was presented by Zocher (2005), which allows to consider not only stochastically independent events (which was the hypothesis of standard FTs) but also mutually exclusive events (e.g. failure modes "too high" and "too low"). Zocher managed to do so by replacing BDDs with Multi-Valued Decision Diagrams (MDDs) in the analysis. Instead of the binary logic TRUE/FALSE (which is interpreted as "working"/"failed" in FTA), now several mutually exclusive values are possible, e.g. "Working"/"in Failure Mode 1"/"in Failure Mode 2" etc. For tool vendors there are several software libraries available for handling MDDs (e.g. MEDDLY, Sylvan), but in the case that only a standard BDD engine is at work, Zocher also presents a canonical transformation of MDDs to BDDs. Zocher's main intention was to allow embedding MCs as subcomponents into

CFTs. However, his approach is useful in general, because mutually exclusive failure modes are often used and classic FTA cannot handle these correctly.

An example is given in Fig. 4. A Markov Chain subcomponent exports two failure states S1 and S2 for usage in the FTA, while the other internal states are hidden. For better understanding of the calculation, the fictitious probabilities of two failure states are given (instead of transition rates) and they have been chosen very high. On superordinate level, a CFT OR gate joins both failure modes to form the top event. Usually, the resulting probability according to the OR gate's formula would be calculated as $P_1 + P_2 - P_1 * P_2$.

Applying the MDD-based algorithm, the resulting MDD constellation would be identical to that described in Fig. 5. After integration and reduction, the probability is calculated correctly as $P_1 + P_2$ (there is no cut set of two mutually exclusive conditions), see Fig. 6. Putting an AND gate above the two mutually exclusive failure modes would correctly produce the output probability zero.

The same task of integrating MCs with more than one exported failure mode has been solved by

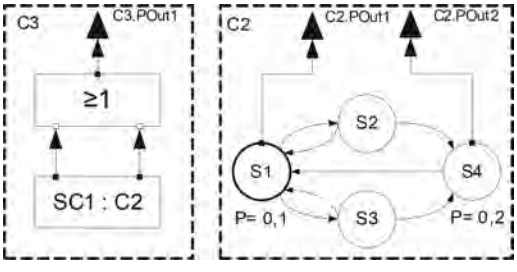


Figure 4. MC with several exported failure modes in a CFT.

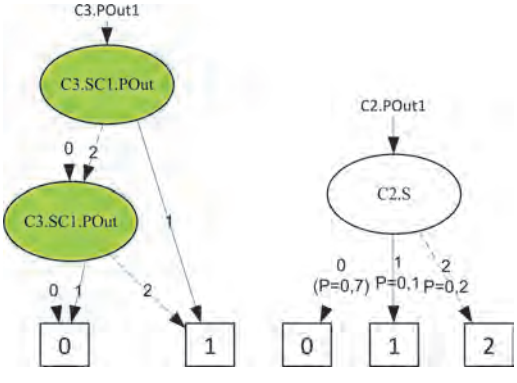


Figure 5. MDD fragments for the system and its subcomponent.

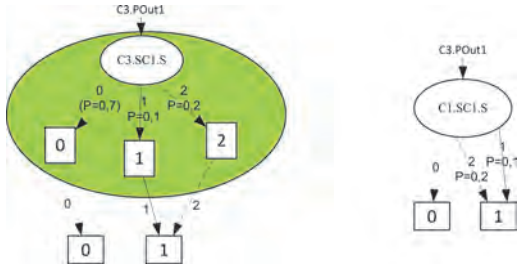


Figure 6. MDD after integration and reduction.

Adler et al. (2007) under the term "Hybrid CFTs". For more information, see Fig. 2 in the cited paper. Unlike Zocher, they do not use MDDs for analysis, but refer to another way of representing multi-state FTs, as proposed by Zang et al. (2003). However, both approaches ultimately produce the same quantitative result and the choice is simply a matter of taste.

2.2 *Component-integrated component fault trees*

Domis (2012) enhanced the CFT approach to enable traceability to system and architectural design models (C2FTs). The traces between CFT artifacts and architecture artifacts make it easier to maintain the fault models when changes are made to the design. Syntactically, the integration is realized by connecting (1) a CFT with a component and (2) every input or output failure port of the CFT with an architectural port of its connected component. While the syntactical enhancement has prevailed, we decided to revise the initial naming and stick to the term CFT.

2.3 *CFT generation from system models*

Many approaches have been proposed for modeling systems in UML, SysML and Simulink and expanding them for the generation of CFTs. Kaiser et al. (2015) show the highly-automated derivation of CFTs from SysML models. The Integration of CFTs with UML using a meta-model has been proposed by Adler et al. (2010). For that purpose, a UML profile was created that provides all necessary elements. As the UML addresses the modeling of software, using this profile makes it possible to model CFTs for the designed software.

The automated derivation of CFT frames from Simulink data flow models has been proposed and implemented by Ramich (2014) and has been validated by a small practical case study by Buono et al. (2015). A prototypical tool reads a hierarchy of Simulink models and transforms it into a hierarchy of CFT components. The integration is merely on topology level (each Simulink compo-

nent produces a CFT, each Simulink signal port a set of CFT ports, CFT edges reflect the Simulink connections). The internal failure logic must be inserted manually by a safety analyst, as in traditional FTA.

2.4 *Enhancement by failure taxonomies*

The introduction of failure type classifications has paved the way towards better semantic integration of system models and CFTs. Domis and Trapp (2008) and Domis (2012) systematically derive and integrate CFTs with a data flow based system architecture, providing a canonical procedure using interface-focused FMEA (IF-FMEA) and a hierarchical failure mode classification scheme. Möhrle et al. (2017a) automate the composition of CFTs on system level using semantic type annotations. For that purpose, architectural ports of components are annotated with flow types that classify the type of interaction (e.g. digital signals, material flows, or energy). For each flow type, safety engineers create a collection of failure types to classify related undesired behavior. These types are used to annotate interfacing model elements in CFTs to abstract from textual descriptions and provide a machine-readable vocabulary.

Fig. 7 shows a taxonomy including four layers. Layer L0 contains the root node *failure* as the most abstract failure type. In each subsequent layer, the undesired behavior is refined further into mutually exclusive sub-items. Taxonomies serve (1) as a template when creating CFTs and (2) as a means for automating the interconnection of ports when composing CFTs. For that purpose, a matching algorithm connects ports according to the relationship of their assigned types. Inconsistencies can be detected and located quickly by analysis tools. This allows a tight coupling between system model and safety analysis model since new CFTs can be generated with minimal effort when making changes to the architecture.

The idea of using failure taxonomies is also exploited by Kaiser et al. (2015). The system architecture is specified in SysML and annotated with semi-formal contracts (assumptions and guarantees) that specify expectations about the nominal externally observable behavior of the system

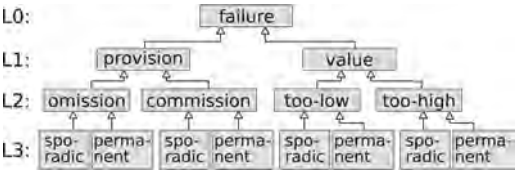


Figure 7. Excerpt of an example failure type taxonomy.

and each of its components. A contract can, for instance, contain the assertion (guarantee) that some value at an output port (e.g. voltage) shall always be in the range [0, 100], provided that all assumptions are met by the environment. A failure is consequently defined as contract violation. In the given example, a value lower than 0 would constitute the failure mode "too low" ("too high" analogously). With contract specification languages that also allow expressing temporal behavior, failure modes like "too late" can be expressed as well. Analyzing whether or not an output event is generated too late is not done inside the CFT analyzer, but must be performed in some discrete-state or hybrid analysis, model checking or simulation framework. Another proposed way to guess the failure propagation logic is to combine fault injection with simulation of the system models. For the CFT analyzer, the question whether or not an assertion is violated is still of Boolean nature. The failure mode classification scheme defined by Domis and Trapp (2008) has been used here as well, but can be refined as appropriate. In the same paper, a new graphical notation is used where the ports from the architecture model appear in the CFT as rectangular box around the triangular input and output failure ports (see Fig. 10a).

However, when switching to the technical viewpoint, potential technical failures like bit-flips, wire-breaks, failures of electronic components and the like will arise.

2.5 *Exploiting behavioral models to derive CFT failure models*

For traditional FTs, several proposals exist to also exploit behavioral models of components (e.g. State Diagrams) for deriving failure modes and failure propagation automatically. Of particular interest in the domain of CFTs is the approach by Berres and Schumann (2016), which can exploit behavioral models to generate CFTs by focusing on each component and its behavior individually.

2.6 *CFTs for layered architectures*

Höfig et al. (2015) describe a methodology called Architecture Layer FailuRE Dependency (ALFRED) which extends CFTs to better maintain the independence of model elements in different layers of a vertically decomposed system architecture. The approach uses so-called ALFRED connections to allow for CFTs on different layers of an architecture. These connections are used to generate a FT for the entire system and over all different architectural layers. This way, for every failure dependency relation, all basic events that are included in the CFT of the dependency

element are added to all output failure modes of the dependent component using OR gates.

ALFRED connections ease the modeling of common cause failures without explicitly modeling dependencies using information flow elements such as ports. This keeps layers in the model independent and supports compositional development. The approach can be used to reuse software on different hardware and ease the evaluation of different deployment variants in terms of safety. Using ALFRED connections, existing safety analysis models can be reused more effectively to support early safety assessments and head towards an automated safety qualification of future cps. The ALFRED approach is already being successfully applied for certification tasks in the railway domain.

2.7 *CFTs for product variants*

Möhrle et al. (2017b) extend CFTs further towards an automated exploration of the system design space in terms of safety. For that purpose, mutable parts of the architecture for which more than one design alternative exist are modeled as variants rather than concrete components. A variant specifies a functional interface by its type-annotated architectural ports (e.g. integer signals, energy flow). The failure logic to be connected to these ports is provided for each realizing entity (component or subsystem). For instance, multiple sensors may be applicable to measure the fill level in an oil tank. Hence, a variant is defined for the level measurement and one realization of the failure logic provided for each sensor. By including variants in the functional system model, multiple architectures can be derived for which CFTs are generated automatically. This way, FTA becomes a benchmark comparing various design alternatives in terms of safety.

3 PRACTICAL EXPERIENCE WITH CFTs

CFTs have been applied in several case studies and industrial projects. Below we present a case study of a situation display system, which was taken from a real world application. In the study, the modeling of the system was done by industrial engineers with some support from modeling experts.

The input to the system is sensor data of the world outside the system. The GPS receiver also obtains information from outside the system, but this information comes from a central authority and is not acquired by the situation display itself. The information is transmitted over two redundant channels and then collected and compared in the channel interface. Finally, it is transmitted to the processing component. During processing, sensor and GPS data are combined to obtain information of the situation outside the system. This information is redundant, since the outside world is evaluated by two independent sources. The system is shown in the center of Fig. 8.

Fig. 9 shows a classic FTA of the system for two top events. The top event *Loss of position data* implies that no information about the outside situation is available. This is the case if both signals are lost. The top event *Partial loss of position data* implies that either the GPS or the sensor data is not available. In this case, the analysts decided to make a worst case assumption by using an OR gate instead of the more precise XOR gate. The top event *Erroneous position data* is not depicted but it is shaped like the FT for the top event *Loss of position data* with only OR gates and erroneous basic events for each component. It models the situation where the data displayed is inaccurate. The analysts decided to create a pessimistic tree for this top event as well, where any failure in one of the components can contribute to the top event.

Fig. 8 shows the CFT analysis that was carried out for this system. The CFTs that are related to system components are depicted as breakout boxes. The CFT for the channel components is redundant

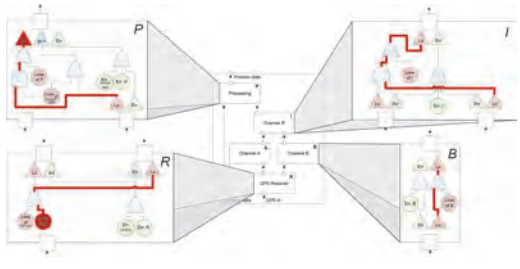


Figure 8. CFT analysis of the situation display system.

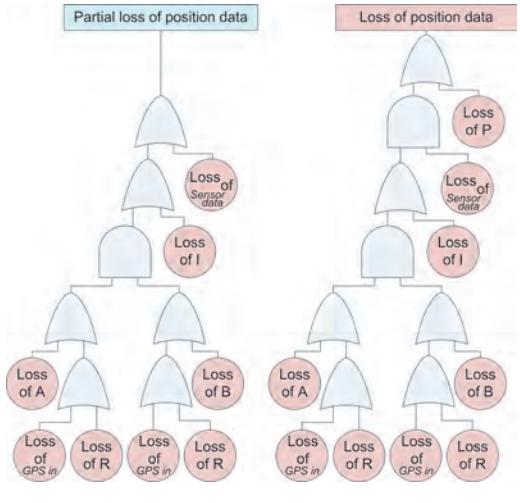


Figure 9. Classic FTA of the situation display system.

and therefore only depicted once. The triangles connected to the output port of the processing component represent the top events of the system, which are the same as for the classic FTA. *Lo* refers to *Loss of position data*, *pLo* refers to *Partial loss of position data* and *err* refers to *Erroneous position data* accordingly. CFTs facilitate the use of interfaces from the system model. For example, the triangles labeled *Lo* and *Err* connected to the right input port of the processing component model the failure modes *loss of* and *erroneous* that propagate over the interface from the component *channel interface* to component *processing*.

During the case study, some findings were observed. Some of them are related to benefits or drawbacks of the modeling strategies used, other findings can be used as hints for proper modeling.

Deep trees: It is a common practice during FTA to follow the chain of actions backwards through the system and document the intermediate failure modes. An example is provided in Fig. 9, where the FTs are deep instead of forming a flat structure as disjunction of all combinations that lead to the top event. Since CFTs follow the system structure, we can see that CFTs promote deep structures instead of flat disjunctions. CFTs therefore should start with the top event at the last processing component (actuator). The failure analysis is then done backwards or upstream to the participating components until all possible causes are identified, ending at the system inputs (sensors).

Localization: When changes need to be made, CFTs facilitate searching for affected parts of the failure analysis. Changes that can be narrowed down to a single component are constrained to a single CFTs. In the classic FT approach, restoring consistency after making changes lies in the hand of the analyst.

Redundancy: In classic FTs, homogeneous redundancy requires a precise distinction between repeated events and redundant, but not identical events. In the CFT approach, redundant components are expressed by simply duplicating the respective CFT, while common cause failures (which have to be identified by common cause analysis) stay outside the component and are connected via ports. In the situation display system, the channel components are redundant and the duplication of the CFT facilitates the modeling and the analysis of the system.

Organizational structures: As systems grow, classic FTs do not provide a proper divide-and-conquer strategy to cope with the increasing complexity. In CFTs, the entire failure behavior (internal) and failure propagation (interfaces) is documented at once. Large distributed development teams can take advantage of the clear structure and interact with one another using consistent interfaces for functional modeling and safety modeling. Hence, CFTs are a better choice for breaking down the complexity of a system in complex organizational structures.

Systemic faults: Uncovering systematic faults in critical applications is of great importance. If systems grow in complexity, failures resulting from flawed collaboration become more significant in comparison to local component failures. CFTs align with component interfaces in system models which often match different developer teams and therefore foster communication about overlooked failures.

Simplification and easier maintenance: The case study has several top events and therefore requires several FTs in classical FTA. This not only causes extra work, but increases the susceptibility to errors and impairs maintenance, as changes have to be applied to every affected tree. When changing details in a component, potential effects on all top events have to be checked. As CFTs filter only the relevant component failures for each top event, the analysis is facilitated for complex systems where many top events exist.

Prepared for later reuse: Reusing a component together with its CFT in later projects appears attractive and enables Off-the-shelf components ("safety element out of context"). However, caution is needed as safety issues may arise when reusing a previously verified component in a new environment. Yet, a reused CFT is a good starting point for new analyses. The combination with contract-based development, which logs assumptions and guarantees for each component, can mitigate the risk of blind reuse. If components are reused across many projects, they and their CFTs go through a maturing process and the chances of all potential failures being detected increase.

4 CONSOLIDATION TOWARDS "CFT 2.0"

After years of various new contributions since CFTs were initially proposed, it is time to collect the ideas and define "CFT 2.0". This hopefully encourages tool builders to support this method.

4.1 *Integration with architectural models*

A first proposal is discarding the term "Component-Integrated Component Fault Trees (C2FTs)" and refer to the overall concept as CFTs. CFTs can either be used as a safety analysis technique alone, or linked to functional or technical static architecture models, in particular SysML Internal Block Diagrams (IBDs). When used in combination with SysML IBDs, we suggest that it shall be mandatory that

1. for each component in the SysML architecture, there is exactly one corresponding CFT with the same name
2. the nesting hierarchy in the SysML IBD and the CFT is the same

3. cause-consequence edges in the CFT exist wherever and only where corresponding signal or service flows exist in the SysML IBD architecture
4. for each incoming flow port, at least one failure input port shall be assigned in the CFT (the same applies for outgoing flow ports and output ports)
5. for each required or provided service port (lollipop symbol), several failure input and output ports can be assigned, since failure consequences can be passed from service provider to requester and vice versa. Safety engineers are responsible for avoiding cycles in the causal chains.
6. all CFT failure ports are bound to one architectural port; graphically represented either by drawing the rectangular architectural port around its triangular failure ports, or by drawing edges between them.

Integration with other signal-flow-based modeling techniques should be standardized as well, in particular the integration with Simulink, e.g. based on the work of Ramich (2014) and Buono et al. (2015). Like for SysML IBD flow ports, the failure causality flow follows the signal flow. Attention must be paid because these models often contain cycles in the signal flows due to closed-loop control. These cycles must not lead to cycles in the causal edges of the FT, which are illegal. Avoiding cycles can be achieved by skillful modeling (e.g. modeling failure consequences on the open-loop signal chain sensors—controller—actuators), but there are also proposals how to resolve cycles automatically, see Vaurio (2007) and Domis (2012). However, these aspects should be handled externally and are not included in the CFT standard.

4.2 *Mutually exclusive failure modes*

We suggest BDD or MDD based evaluation of the resulting failure probabilities as the default algorithms for CFTs. Embedding MC components (which are not part of the CFT method) is a recommended practice for CFT tools. Using the MDD evaluation method, exporting more than one failure mode from a MC, as shown in Fig. 5, is possible and produces correct results. Even without embedding MCs, it should be possible to model mutually exclusive failure modes like "too low" and "too high" and notifying the tool about their special relationship. To allow this, we suggest as an extension to CFTs, that several basic events can be surrounded by a rectangular frame as shown in Fig. 10a), indicating that they form a mutually exclusive group. Similarly, failure ports encapsulated by the same rectangular architectural port shall be treated as mutually exclusive.

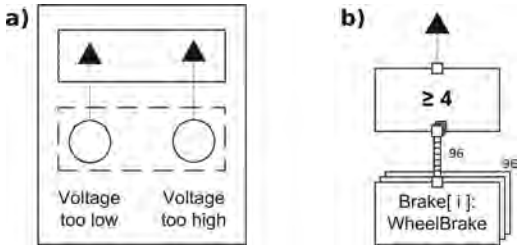


Figure 10. a) Two mutually exclusive failure modes b) Multiinstance component connected to a multi-input gate.

4.3 *Multiple instances and variants of components*

Another innovation we suggest for CFT2.0 is exploiting the multi-instance capabilities in a more convenient way for highly-redundant systems by introducing a multi-instance symbol and gates with multi-input ports. Imagine a train braking system with 96 wheel brakes of the same type, where a hazard is present whenever at least 4 wheel brakes have failed. Instead of manually creating 96 individual instances and connecting them to an n-out-of-m gate with 96 inputs, a tool can offer a multi-instance symbol like in Fig. 10b) with a parameter – 96 in this example—to indicate how many replicas are available. Note the input port symbol at the gate that indicates a multi-port and the different style of line indicating a multi-line. Of course, the parameter can easily be changed for later projects, so that only one parameter needs to be changed to adapt the model for a train with 72 wheel brakes. This enables efficient safety analysis for product lines. The semantics is straightforward because this is merely a matter of graphical representation.

To further support the product line approach, we suggest that also the variant approach from Möhrle et al. (2017b) be part of CFT2.0. Modelers can specify variants by modeling the signature (number and types of ports), and the concrete CFT model is attached later, selected from a model library and guided by a variants database. For example, if a hydraulic brake is replaced by an electric brake, only one more model needs to be added and the different solutions can be easily compared. However, this aspect is mainly a question of tool support and not of the CFT semantics themselves.

5 CONCLUSION AND FURTHER RESEARCH

By collecting and integrating scientific contributions from many researchers over more than a decade, we have provided an overview of the state-of-the-art in CFTs and made some proposals on

how to update their specification under the term CFT2.0. By referencing larger case studies, we aim to encourage tool developers and safety analysts to use CFTs in industrial projects and provide feedback for further method improvements. We are convinced that the CFTs have reached maturity for industrial use in safety-relevant projects. Nevertheless, we are inclined to work on further developments. We will further elaborate on a more formal integration of CFTs with SysML/UML by combining safety analysis with contract-based development. With regard to UML/SysML, we are working on integrating not only flow ports (dataflow oriented interaction) but also service ports (lollipop symbol), as this type of interaction (often asynchronous) is typical not only for software systems, but also for open networked systems and systems with user interaction. Here we will have to deal with the fact that failure propagation can occur in both directions over the architecture port. We will continue working on approaches to derive the internal failure propagation and mitigation logic of CFTs, in particular by combining failure injection and simulation. Also more traditional analytic approaches like FMEA or Hip-Hops to find inner failure modes and failure propagation should be better integrated with CFTs. We plan to extend the presented variability approach also towards runtime variance, which is important as spontaneously networking CPSs, so called System-of-Systems (SoSs) become ubiquitous. This would mean that for every legal configuration of an SoS (e.g. vehicle platoon or sensor network), parametric CFT templates will have to be re-combined and checks carried out whether or not the desired constellation is safe or reliable enough.

ACKNOWLEDGEMENTS

The work leading to this paper was partially funded by the German Federal Ministry of Education and Research under grant number 01IS16043Q (CrESt).

REFERENCES

Adler, R. et al. (2010). Integration of component fault trees into the UML. In *Workshops & Symposia at MODELS*.

Adler, R., M. Forster, & M. Trapp (2007). Determining configuration probabilities of safety-critical adaptive systems. In *21st Int. Conf. on Advanced Information Networking and Applications Workshops*, pp. 548–555.

Berres, A. & H. Schumann (2016). Automatic generation of fault trees: A survey on methods and approaches. In *ESREL 2016*.

Bollig, B. & I. Wegener (1996). Improving the variable ordering of obdds is np-complete. *IEEE Transactions on Computers 45*(9), 993–1002.

Buono, S., V. Ramich, B. Kaiser, & J. Zander (2015). Eine industrielle fallstudie zur teilautomatischen erzeugung von komponentenfehlerb äume aus simulink-modellen. In *Gemeinsamer Tagungsband der Workshops der Tagung Software Engineering*, pp. 41–50.

Domis, D. (2012). *Integrating fault tree analysis and component-oriented model-based design of embedded systems*. Verlag Dr. Hut.

Domis, D. & M. Trapp (2008). Integrating safety analyses and component-based design. In *Proc. of SafeCamp*.

Förster, M. (2006). Efficient quantitative evaluation of state/event fault trees. Master thesis, Hasso-Plattner-Institut für Softwaresystemtechnik (Uni Potsdam).

Höfig, K., M. Zeller, & R. Heilmann (2015). Alfred: A methodology to enable component fault trees for layered architectures. In *41st Euromicro SEAA*.

IEC 61025 (2006). Fault tree analysis (FTA).

IEC 61508 (1998). Functional safety of electrical/electronic/programmable electronic safety-related systems.

ISO 26262 (2011). Road vehicles—functional safety.

Kaiser, B. (2005). Extending the expressive power of fault trees. In *Proc. of RAMS*.

Kaiser, B. et al. (2015). Contract-based design of embedded systems integrating nominal behavior and safety. *Complex Systems Informatics and Modeling Quarterly (CSIMQ) (4)*, 66–91.

Kaiser, B., P. Liggesmeyer, & O. Mäckel (2003). A new component concept for fault trees. In *Proc. of the 8th Australian Workshop on Safety Critical Systems and Software*, pp. 37–46.

Möhrle, F. et al. (2017a). A formal approach for automating compositional safety analysis using flow type annotations in component fault trees. In *Proc. of the 27th European Safety and Reliability Conf.*

Möhrle, F. et al. (2017b). Towards automated design space exploration for safety-critical systems using typeannotated component fault trees. Int. Symposium on Model-Based Safety and Assessment (IMBSA).

Ramich, V. (2014). Teilautomatische erstellung von component-fault-trees aus simulink-modellen. Master's thesis, Master Thesis Universität Kassel.

Vaurio, J.K. (2007). A recursive method for breaking complex logic loops in boolean system models. *Reliability Engineering & System Safety 92*(10), 1473–1475.

Vesely, W.E., F. Goldberg, & D. Roberts, N. Haasl (1981). *Fault Tree Handbook (NUREG 0492)*. U.S. Nuclear Regulatory Commission.

Zang, X., D. Wang, H. Sun, & K.S. Trivedi (2003). A BDD-based algorithm for analysis of multistate systems with multistate components. *IEEE Transactions on Computers 52*(12), 1608–1618.

Zocher, A. (2005). Quantitative auswertung von multizustand-komponentenfehlerbäumen durch mehrwertige entscheidungsdiagramme. Master thesis, Hasso-Plattner-Institut für Softwaresystemtechnik (Uni Potsdam).

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Comparison of machine learning algorithms on data from the nuclear industry

E. Remy & E. Dautrême
EDF R&D, Chatou, France

C. Talon
Independent Researcher, Paris, France

Y. Dirat & C. Dinse Le Strat
EDF CEIDRE, Saint-Denis, France

ABSTRACT: With the increase in computing power and the rise of "big data", many machine learning algorithms have been developed or given a new lease of life. Their application is proved to be really efficient in various fields. However their implementation in the industrial sector, such as nuclear power generation, seems less widespread. This paper presents a comparison of several techniques on a case study from the nuclear industry. Their performance on the real dataset are compared and a discussion is proposed on their practical use, advantages and disadvantages, precaution for use and relevance in an industrial context.

1 INTRODUCTION

1.1 *Industrial context*

Shutdowns of nuclear power reactors are regularly planned for refueling and carrying out maintenance operations. The state of the reactor during its shutdown can be characterized by an indicator which is measured during the shutdown. If this indicator is higher than a given fixed threshold, it may impact the scheduled planning and lead to an extension of the shutdown. In addition to this indicator, around twenty parameters describing the operation conditions of the reactor just before the shutdown are available.

There currently exists neither physical-based model nor computer simulation code to characterize the indicator and predict if it will be higher than the given threshold. That is why, in order to anticipate the required logistical support for maintenance during the reactor shutdown, it seems relevant to "learn from the data" and try to take advantage of the information brought by the operation conditions before the shutdown to foresee the state of the reactor during its shutdown. The use of a "black-box" machine learning algorithm seems to be a promising solution in order to build the prediction function between the state before and the one during the shutdown.

Indeed, with the increase in computing power and the rise of "big data", many machine learning algorithms have been developed or given a new lease of life. Their application is proved to be really efficient for prediction purpose in various fields: banking, insurance, media, information technology, healthcare, transportation, sports… However their implementation in the industrial sector, such as nuclear power generation, seems less widespread and their relevance and effectiveness have to be confirmed in this area of application where data may be less voluminous.

1.2 *Objectives*

Different supervised machine learning techniques have been tested on the available data and the main objective of this paper is to present a comparison of the obtained results.

The remainder of this article is organized as follows. Section 2 describes the fundamental principles of the families of machine learning algorithms that have been used. Section 3 presents how the performance of the predictive models can be assessed and compared and it gives the main results obtained on the dataset. A general discussion is proposed in Section 4 on the practical use of the techniques, by exposing their advantages and disadvantages, precaution for use and relevance in an industrial "relatively small data" context.

2 PRESENTATION OF THE DIFFERENT FAMILIES OF MACHINE LEARNING ALGORITHMS

2.1 *Notations and preliminary assumptions*

Y will denote the indicator characterizing the state of the reactor during its shutdown. Y is a qualitative (or categorical) random variable taking two values: AT (for "Above Threshold") and BT (for "Below Threshold"). For more convenience, we will represent it by two equivalent numerical variables, Z and $\tilde{Z}$, using binary codes: $Z = 1$ if $Y = AT$ and $Z = 0$ if $Y = BT$, so that $Z = \mathbb{I}_{\{Y=AT\}}$ with $\mathbb{I}_{\{.\}}$ the indicator function, and $\tilde{Z} = 1$ if $Y = AT$ and $\tilde{Z} = -1$ if $Y = BT$. $\boldsymbol{X} = (X^{(1)},\ldots,X^{(p)})$ will denote the vector made of the p parameters describing the operation conditions of the reactor before its shutdown. In our case, $p = 25$. Each variable $X^{(j)}$, $1 \le j \le p$, can be either deterministic or random and either qualitative or quantitative (or continuous). If $X^{(j)}$ is qualitative with k_j levels, we will consider its numeric version using dummy coding of its levels and equivalently introduce $k_j - 1$ binary variables, only one of which is "on" at a time: $(\mathbb{I}_{\{X^{(j)}=h\}})_{1\le h\le k_j-1}$. If $X^{(j)}$ is quantitative, we will assume it is a standardized variable. n will denote the number of joined observations of $(Y_i, \boldsymbol{X}_i)_{1\le i\le n}$. In our dataset, $n = 89$. We will make the assumptions that $(Y_i, \boldsymbol{X}_i)_{1\le i\le n}$ are independent observations of (Y, X) and are not subject to measurement uncertainty.

Regarding the data $(Y_i, \boldsymbol{X}_i)_{1\le i\le n}$, we seek for a function, denoted by $f(\boldsymbol{X})$ (sometimes called a "classifier" since Y is a binary variable), for predicting Y given the predictor inputs $\boldsymbol{X}$. In the literature, this issue is called a "supervised" learning problem because of the presence of the two possible outcomes for Y (or equivalently for Z or $\tilde{Z}$) to guide the learning process and build classification function $f(.)$. Depending on the type of algorithm, $f(.)$ can predict either directly a value $\hat{Y}_i = f(\boldsymbol{X}_i) \in \{AT; BT\}$, $1 \le i \le n$, or an estimated probability that Y_i equals AT depending on $\boldsymbol{X}_i$. In that case, our decision rule will be to assign value AT if the estimated probability is higher than 0.35 (this conservative value has been tuned to obtain the best results in terms of performance indicators – see Sections 3.1 and 3.2).

2.2 *Logistic regression, stepwise, LASSO, Ridge, PLS and Sparse PLS*

From a general point of view, the logistic regression model can be used to estimate the probability of a binary response based on one or more predictor variables. A standard parametrization uses the logit function to link the target probability $\mathbb{P}(Y = AT \mid \boldsymbol{X}) = \mathbb{P}(Z = 1 \mid \boldsymbol{X})$ with predictors $\boldsymbol{X}$:

$$\begin{aligned} \log it\left(\mathbb{P}(Z=1\mid \boldsymbol{X})\right) &= \log\left(\frac{\mathbb{P}(Z=1\mid \boldsymbol{X})}{1-\mathbb{P}(Z=1\mid \boldsymbol{X})}\right) \\ &= \beta_0 + \sum_{j=1}^{p}\beta_j X^{(j)} \end{aligned} \quad (1)$$

$$\Leftrightarrow \left(\mathbb{P}(Z=1\mid \boldsymbol{X})\right) = \frac{\exp\left(\beta_0 + \sum_{j=1}^{p}\beta_j X^{(j)}\right)}{1+\exp\left(\beta_0 + \sum_{j=1}^{p}\beta_j X^{(j)}\right)} \quad (2)$$

with β_0 a constant called intercept and $(\beta_1, \ldots, \beta_p)$ a vector of regression coefficients to be estimated from the data. We will denote $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)$.

One interest of this parametric model is its interpretability, since $\exp(\beta_j)$, $1 \le j \le p$, quantifies the increase in $\mathbb{P}(Z = 1 \mid \boldsymbol{X})$ when (supposing) continuous variable $X^{(j)}$ increases by 1 unit.

Regarding $(Z_i, \boldsymbol{X}_i)_{1\le i\le n}$, $\boldsymbol{\beta}$ can be estimated by $\hat{\boldsymbol{\beta}}^{ML}$ using the Maximum Likelihood (ML) statistical inference method:

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{ML} &= \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\arg\min}(-2\ell(\boldsymbol{\beta})) \\ &= \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\arg\min}\Big(-2\sum_{i=1}^{n} Z_i \log\big(\mathbb{P}(Z_i=1\mid \boldsymbol{X}_i)\big) \\ &\quad + (1-Z_i)\log\big(\mathbb{P}(Z_i=0\mid \boldsymbol{X}_i)\big)\Big) \end{aligned} \quad (3)$$

which is equivalent to:

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{ML} = \underset{\beta_0,\ldots,\beta_p\in\mathbb{R}}{\operatorname{argmin}}\Big(&-2\sum_{i=1}^{n} Z_i\Big(\beta_0 + \sum_{j=1}^{p}\beta_j X^{(j)}\Big) \\ &- \log\Big(1+\exp\Big(\beta_0 + \sum_{j=1}^{p}\beta_j X^{(j)}\Big)\Big)\Big) \end{aligned} \quad (4)$$

where $\ell(\boldsymbol{\beta})$ denotes the log-likelihood function.

In order to improve the prediction accuracy and the interpretability of regression models, it may be interesting to alter the model fitting process to select only a subset of the provided predictors $\boldsymbol{X}$ for use in the final model rather than using all of the $X^{(j)}$, $1 \le j \le p$. A standard procedure for subset selection is the stepwise-selection strategy. The forward-stepwise selection starts with the intercept β_0, and then sequentially adds into the model the predictor that most improves the fit. The backward-stepwise selection starts with the full model including $\boldsymbol{X}$, and sequentially deletes the predictor that has the least impact on the fit. We used a hybrid stepwise-selection strategy that considers both forward and backward moves at each step, and selects the "best" of the two. The step function uses the Akaike Information Criterion (AIC) for weighing the choices, which takes proper account of the number of parameters to be fitted. At each step, an add or drop will be performed that minimizes the AIC score:

$$\mathrm{AIC} = 2(p^* + 1) - 2\ell(\hat{\boldsymbol{\beta}}^{ML}) \tag{5}$$

with $1 \le p^* \le p$ the number of predictors included in the model.

Because it is a discrete process—variables are either retained or discarded, stepwise procedure may not reduce the prediction error of the full model. Shrinkage methods are more continuous and do not suffer as much from high variability. Several shrinkage variants of the logistic regression are available. LASSO (for Least Absolute Shrinkage and Selection Operator) shrinks the regression coefficients $\boldsymbol{\beta}$ by imposing a $\mathbb{L}_1$ penalty on their size. It forces the sum of the absolute value of the regression coefficients to be less than a fixed value, which forces certain coefficients to be set to zero, effectively choosing a simpler model that does not include those coefficients:

$$\hat{\boldsymbol{\beta}}^{LASSO} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\operatorname{argmin}}(-2\ell(\boldsymbol{\beta})) \text{ subject to } \sum_{j=1}^{p} |\beta_j| \le t \tag{6}$$

The smaller the value of the tuning parameter t, the fewer the number of nonzero components in $\hat{\boldsymbol{\beta}}^{LASSO}$, thus leading to what is called "sparse" models. Thus LASSO does a kind of continuous subset selection. If t is chosen larger than $\sum_{j=1}^{p} \left|\hat{\beta}_j^{ML}\right|$, LASSO estimates are the same as the ML's. For $t = \frac{\sum_{j=1}^{p} \left|\hat{\beta}_j^{ML}\right|}{2}$ say, then the ML coefficients are shrunk by about 50% on average. Parameter t has to be carefully chosen in order to minimize an estimate of expected prediction error.

A "good" value of t can be obtained by cross-validation. Cross-validation is a general process which can be applied to many types of issues. It consists in dividing the dataset $(Z_i, \boldsymbol{X}_i)_{1\le i\le n}$ randomly into a certain number of equal parts, say 10. LASSO regression with a given value for t is fitted with the nine-tenths of the data (this dataset is called the "training dataset"), and the prediction error is computed on the remaining one-tenth (called the "test observations"). This process is done in turn for each one-tenth of the data, and the ten prediction error estimates are averaged. From this we can obtain an estimated prediction error curve as a function of t, enabling to identify a relevant value for t minimizing the estimated prediction error.

Ridge is another shrinkage method, but it shrinks the regression coefficients $\boldsymbol{\beta}$ by imposing a $\mathbb{L}_2$ penalty on their size:

$$\hat{\boldsymbol{\beta}}^{Ridge} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\operatorname{argmin}}(-2\ell(\boldsymbol{\beta})) \text{ subject to } \sum_{j=1}^{p} \beta_j^2 \le t \tag{7}$$

Contrary to LASSO and because of the $\mathbb{L}_2$ penalty, Ridge shrinks the size of the coefficients, but it does not set any of them to zero.

When there are many correlated predictors in a regression model (also called multicollinearity among $\boldsymbol{X}$), their coefficients $\boldsymbol{\beta}$ can become poorly determined. In this situation the coefficient estimates $\hat{\boldsymbol{\beta}}$ of the regression may change erratically in response to small changes in the model or the data and thus may suffer from unstability. By imposing a size constraint on the number of coefficients to be estimated, as in (6) or (7), this problem is alleviated. However another solution is to consider m new inputs, $1 \le m \le p$, denoted by $T^{(j)}$, $1 \le j \le m$, derived from the original predictors $\boldsymbol{X}$ and used in place of the $X^{(j)}$, $1 \le j \le p$, in the regression model:

$$\begin{aligned} \operatorname{logit}\left(\mathbb{P}\left(Z=1|\boldsymbol{T}\right)\right) &= \log\left(\frac{\mathbb{P}\left(Z=1|\boldsymbol{T}\right)}{1-\mathbb{P}\left(Z=1|\boldsymbol{T}\right)}\right) \\ &= \beta_0 + \sum_{j=1}^{m} \beta_j T^{(j)} \end{aligned} \tag{8}$$

Generally, m is chosen to be small and the $T^{(j)}$, $1 \le j \le m$, are orthogonal linear combinations of the $X^{(j)}$, $1 \le j \le p$, in order to avoid estimation unstability.

Partial Least Squares (PLS) constructs a set of linear combinations of the inputs for regression, by using Y in addition to $\boldsymbol{X}$ for this construction. It finds a linear regression model by projecting Y and $\boldsymbol{X}$ to a new space. It tries to find the multidimensional direction in the $\boldsymbol{X}$-space that explains the maximum multidimensional variance direction in the Y-space. A detailed mathematical description of how the $T^{(j)}$, $1 \le j \le m$, are defined can be found in (Wold et al. 1983).

Sparse PLS is a hybrid method, which can be viewed as a combination of LASSO and PLS. Details on this approach can be found in (Chun et al. 2010).

Stepwise, LASSO, Ridge, PLS and Sparse PLS are all variants of the logistic regression and they share the same advantage of being interpretable models, with fully explicit expressions for the estimated target probability depending on $\boldsymbol{X}$.

2.3 *Classification trees, bagging and boosting*

A classification tree is a directed acyclic graph consisting of nodes and directed edges. It has three types of nodes: a root node that has no incoming edges and zero or more outgoing edges; internal nodes, each of which has exactly one incoming edge and two or more outgoing edges; leaf or terminal nodes, each of which has exactly one incoming

edge and no outgoing edges. The full dataset sits at the root node at the top of the tree. Each leaf node is assigned a class label corresponding to one value of the target variable (Y in our case). The non-terminal nodes, which include the root and other internal nodes, contain attribute test conditions based on the predictors $\boldsymbol{X}$ to separate observations that have different characteristics, as shown in Figure 1. Stumps (or decision stumps) are trees with a single split (the root is directly connected to the leaves).

Algorithms for constructing classification trees usually work top-down and recursively, by choosing an input variable among the p and a split-value at each step that best split the set of observations into two data subsets. The process is continued until some stopping rule is applied.

A key advantage of the recursive binary tree is its interpretability, since it finally gives a set of fully understandable decision rules, as illustrated in Figure 1.

Different algorithms use different metrics for measuring "best". These generally measure the homogeneity of the target variable within the subsets. These metrics are applied to each candidate subset, and the resulting values are combined to provide a measure of the quality of the split.

How large should we grow the tree? Clearly a very large tree might overfit the data, while a small tree might not capture the important structure of the dataset $(Y_i, \boldsymbol{X}_i)_{1\le i\le n}$. Tree size is a tuning parameter governing the model's complexity, and a "good" tree size can be adaptively chosen from the data. A strategy is to grow the full initial tree, denoted by T_0, stopping the splitting process only when some minimum node size is reached (in our case 6). Then this large tree is pruned using a process called cost-complexity pruning. We define a subtree $T \subset T_0$ to be any tree that can be obtained by pruning T_0, that is collapsing any number of its internal (non-terminal) nodes. We index terminal nodes by

Figure 1. Illustration of a classification tree.

m, with node m representing data subset S_m. Let $|T|$ denote the number of terminal nodes in T, n_m, $1 \le n_m \le n$, the number of observations in subset S_m and $\hat{\mathbb{P}}_m$ the proportion of observations with value $Y = AT$ in node m: $\hat{\mathbb{P}}_m = \frac{1}{n_m}\sum_{\boldsymbol{X}_i \in S_m} \mathbb{I}_{\{Y_i = AT\}}$. We define the cost complexity criterion by:

$$C_\alpha(T) = \sum_{m=1}^{|T|} n_m Q_m(T) + \alpha|T| \tag{9}$$

where $Q_m(T)$ is an "impurity measure", in our case the Gini index defined by $Q_m(T) = 2\hat{\mathbb{P}}_m\left(1-\hat{\mathbb{P}}_m\right)$. We classify the observations in node m to class $Y = AT$ if $\hat{\mathbb{P}}_m > 0.5$ (that is to say if value $Y = AT$ is majority in node m), and to class $Y = BT$ otherwise. The idea is to find, for each $\alpha > 0$, the subtree $T_\alpha \subseteq T_0$ to minimize $C_\alpha(T)$. The tuning parameter α governs the tradeoff between tree size and goodness of fit to the data. Large values of α result in smaller trees T_α, and conversely for smaller values of α. With $\alpha = 0$, the solution is the full tree T_0. The determination of a "good" value for α can be achieved by cross-validation.

In order to avoid overfitting (see Section 3.2 for a detailed discussion on this issue), bagging (for "bootstrap aggregation") can be used. It consists in drawing randomly with replacement from the original dataset $(Y_i, \boldsymbol{X}_i)_{1\le i\le n}$ a large number $B > 1$ of new datasets of same size n. For each new dataset, a classification tree is built and the final bagged classifier is obtained by selecting the class between $Y = AT$ or $Y = BT$ with the most "votes" from the B trees. A deeper introduction to bagging can be found in (Breiman 1996).

For large problems, the performance of classification trees can be improved using a procedure called boosting. If we consider numerical target variable $\tilde{Z}$ introduced in Section 2.1, given the predictor variables $\boldsymbol{X}$, a classifier $f(\boldsymbol{X})$ produces a prediction taking one of the two values -1 ($Y = BT$) or 1 ($Y = AT$). The error (or misclassification) rate of $f(.)$ on the dataset is $err_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{Y_i \ne f(\boldsymbol{X}_i)\}}$. A weak classifier is one whose error rate is only slightly better than random guessing. The purpose of boosting is to sequentially apply a weak classification algorithm to repeatedly modified versions of the data, thereby producing a sequence of weak classifiers $f_m(X)$, $1 \le m \le M$. The predictions from all of them are then combined through a weighted majority vote to produce the final prediction:

$$f(\boldsymbol{X}) = \text{sign}\left(\sum_{m=1}^{M} \alpha_m f_m(\boldsymbol{X})\right) \tag{10}$$

where $\alpha_1, \ldots, \alpha_m$ weight the contribution of each respective $f_m(\boldsymbol{X})$. Their effect is to give higher influence to the more accurate classifiers in the sequence. The data modifications at each boosting step consist of applying weights $\omega_1, \ldots, \omega_n$ to

each of the observations $(Y_i, \boldsymbol{X}_i)_{1\le i\le n}$. Initially all of the weights are set to $\omega_i = \frac{1}{n}$, so that the first step simply trains the classifier on the data in the usual manner. For each successive iteration $m = 2,\ldots,M$, the observation weights are individually modified and the classification algorithm is reapplied to the weighted observations. At step m, those observations that were misclassified by the classifier $f_{m-1}(\boldsymbol{X})$ induced at the previous step have their weights increased, whereas the weights are decreased for those that were classified correctly. Thus as iterations proceed, observations that are difficult to classify correctly receive ever-increasing influence. Each successive classifier is thereby forced to concentrate on those training observations that are missed by previous ones in the sequence.

In our study, we used stumps as weak classifiers, $\alpha_m = \log\left(\frac{1-err_m}{err_m}\right)$ with $err_m = \frac{\sum_{i=1}^{n}\omega_i \mathbb{I}_{\{Y_i \neq f_m(X_i)\}}}{\sum_{i=1}^{n}\omega_i}$ and for each successive iteration $m = 2,\ldots,M$, the observation weights were modified as follows: $\omega_i \leftarrow \omega_i \exp\left(\alpha_m \mathbb{I}_{\{Y_i \neq f_m(X_i)\}}\right), 1 \le i \le n$.

Contrary to standard binary trees, bagging and boosting techniques lead to classification rules that are not easy to interpret. Indeed, since they aggregate results obtained from different trees, it is difficult to identify how the inputs concretely affect the output.

2.4 *Neural networks*

The name "neural networks" derives from the fact that they were first developed as models for the human brain. Each unit within the model represents a neuron and the connections (links in Figure 2) represent synapses. The term neural network encompasses a large class of models. Here we describe the most widely used neural network, often called the "single hidden layer back-propagation network" or the "single layer perceptron".

The central idea is to extract linear combinations of the inputs as derived features, and then model the target as a nonlinear function of these features. More formally, in our case, a neural network is a two-stage classification model. Derived features $\sigma(\sum ne_j\omega_j)$ and $\sigma(\sum ne_j\gamma_j)$ are created from linear combination of the inputs $\left(X^{(1)}, \quad , X^{(p)}\right) = \left(ne_1, \quad , ne_p\right)$ and then output Z is modeled as a function of linear combinations of these derived features: $g\left(\Theta\sigma\left(\sum ne_j\omega_j\right) + \Psi\sigma\left(\sum ne_j\gamma_j\right)\right)$.

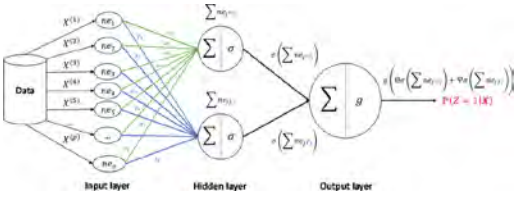


Figure 2. Illustration of a neural network.

$\sigma(.)$ is called the activation function and is usually chosen to be the sigmoid: $\sigma(v) = \frac{1}{1+\exp(-v)}$. $g(.)$ allows a final transformation and is usually chosen to be the logistic. $\boldsymbol{\beta} = \left(\left(\omega_j, \gamma_j\right)_{1\le j\le p}, \Theta, \Psi\right)$ are unknown parameters, often called weights, and we seek values for them that make the model fit the data "well", for instance by minimizing twice the opposite of the log-likelihood function $\ell(\boldsymbol{\beta})$ defined in Equation (3). The generic approach to this optimization problem is by gradient descent, called back-propagation in this setting. Because of the compositional form of the model, the gradient can be easily derived using the chain rule for differentiation. This can be computed by a forward and backward sweep over the network, keeping track only of quantities local to each unit. Details on the back-propagation algorithm can be found in (Rumelhart et al. 1986).

In our specific case, a neural network can be seen as a nonlinear generalization of the logistic regression model presented in Section 2.2. However, it lacks its interpretability and is often perceived as the "ultimate black-box model".

2.5 *Support vector machine*

In our classification context, a Support Vector Machine (SVM) consists of constructing a hyperplane that separates the data into two classes corresponding to the possible values for Y. Intuitively, if we suppose the two classes are linearly separable, a good separation is achieved by the hyperplane that has the largest distance to the nearest data point of any class, as in Figure 3.

More formally, a hyperplan is defined by:

$$\left\{\boldsymbol{X} \in \mathbb{R}^p : h(\boldsymbol{X}) = \beta_0 + \sum_{j=1}^{p}\beta_j X^{(j)} = \beta_0 + \boldsymbol{X}^T\boldsymbol{\beta} = 0\right\} \tag{11}$$

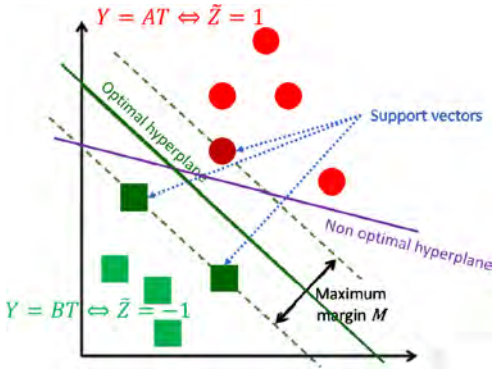


Figure 3. Illustration of an optimal separating hyperplane (in green) and a non optimal one (in purple).

with $\boldsymbol{\beta}$ a unit vector: $\boldsymbol{\beta} = 1$. If we consider target variable $\tilde{Z}$, the classification rule induced by $h(.)$ is $f(\boldsymbol{X}) = \text{sign}(\beta_0 + \boldsymbol{X}^T\boldsymbol{\beta})$.

If the two classes are separable, we want to find a function $h(\boldsymbol{X}) = \beta_0 + \boldsymbol{X}^T\boldsymbol{\beta}$ with $\tilde{Z}_i h(\boldsymbol{X}_i) > 0, \forall i = 1,\ldots,n$, that creates the biggest margin M between the data for class −1 and 1. The following optimization problem formalizes this idea:

$$\max_{\beta_0 \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^p} (M) \text{ subject to } \tilde{Z}_i(\beta_0 + \boldsymbol{X}_i^T\boldsymbol{\beta}) > M, 1 \le i \le n. \tag{12}$$

If the two classes overlap, one way to deal with this overlap is to still maximize margin M, but allow for some points to be on the wrong side of the margin. Define the slack variables $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$. One way to modify the constraint in Equation (12) is:

$$\tilde{Z}_i(\beta_0 + \boldsymbol{X}_i^T\boldsymbol{\beta}) > M(1 - \xi_i), \forall i = 1,\ldots,n \tag{13}$$

with $\xi_i \ge 0, 1 \le i \le n$, and $\sum_{i=1}^n \xi_i \le$ constant. ξ_i in constraint (13) is the proportional amount by which the prediction $h(\boldsymbol{X}_i)$ is on the wrong side of its margin. Hence by bounding the sum $\sum_{i=1}^n \xi_i$, we bound the total proportional amount by which predictions fall on the wrong side of their margin. Misclassifications occur when $\xi_i > 1$, so bounding $\sum_{i=1}^n \xi_i$ at a value C say, bounds the total number of misclassifications at C.

Whereas the original problem may be stated in the original space, it often happens that the two sets to discriminate are not linearly separable in that space. For this reason, it is proposed that the original finite dimensional space be mapped into a much higher dimensional space, presumably making the linear separation easier in that space, as illustrated in Figure 4. This transformation is called the "kernel trick".

To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function $\phi(\boldsymbol{X}, \boldsymbol{U})$ selected to suit the problem. The hyperplanes in the higher dimensional space (which may be nonlinear in the original input space) are defined as the set of points whose dot product with a vector in that space is constant. The vectors defining the hyperplanes can be chosen to be linear combinations with parameters α_i of images of feature vectors $\boldsymbol{X}_i, 1 \le i \le n$. With this choice of a hyperplane, the points $\boldsymbol{X}$ in the feature space that are mapped into the hyperplane are defined by the relation: $\sum_i \alpha_i \phi(\boldsymbol{X}_i, \boldsymbol{X}) = \text{constant}$. If $\phi(\boldsymbol{X}, \boldsymbol{U})$ becomes small as $\boldsymbol{U}$ grows further away from $\boldsymbol{X}$, each term in the sum measures the degree of closeness of the point $\boldsymbol{X}$ to the corresponding data base point $\boldsymbol{X}_i$. In this way, the sum of kernels above can be used to measure the relative nearness of each test point to the data points originating in one or the other of the sets to be discriminated. In our case, we tested four different kernel functions: linear $\phi(\boldsymbol{X}, \boldsymbol{U}) = \boldsymbol{X}^T\boldsymbol{U}$, Gaussian radial $\phi(\boldsymbol{X}, \boldsymbol{U}) = \exp(-\sigma\boldsymbol{X} - \boldsymbol{U}^2), \sigma \in \mathbb{R}$, inhomogeneous polynomial (of order $d \in \mathbb{N}$) $\phi(\boldsymbol{X}, \boldsymbol{U}) = (\gamma \boldsymbol{X}^T\boldsymbol{U} + r)^d, \gamma, r \in \mathbb{R}$, and hyperbolic tangent $\phi(\boldsymbol{X}, \boldsymbol{U}) = \tanh(\kappa \boldsymbol{X}^T\boldsymbol{U} + \theta), \kappa, \theta \in \mathbb{R}$. We chose $d = 2$ and "good" values for tuning parameters $\sigma, \gamma, r, \kappa$ and θ minimizing the estimated prediction error can be found by cross-validation as presented in Section 2.2. Further elements on SVM can be found in (Burges 1998).

As for neural networks, SVM provides a "black-box model" which does not allow any interpretable description of how the inputs affect the output.

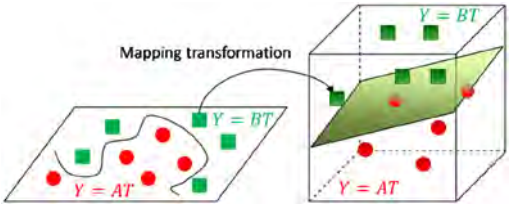


Figure 4. Illustration of the mapping transformation.

3 RESULTS

3.1 *Performance indicators*

Before showing the results obtained with the different families of machine learning algorithms presented in the previous sections, one must introduce indicators that will allow to assess and compare the performance of the techniques on the dataset.

These indicators are based on the "confusion matrix" (also called "error matrix") presented in Table 1. Once the prediction function $f(.)$ has been built using one of the available machine learning algorithms, it is easy to compare for each observation $(Y_i, \boldsymbol{X}_i)_{1 \le i \le n}$ if the predicted value $\hat{Y}_i = f(\boldsymbol{X}_i)$ is the same as the real Y_i observed in the dataset. Each row of the matrix represents the instances in a class predicted with the model while each column represents the instances in an actual class.
with $TP + FP + FN + TN = n$.

Error rate or misclassification rate on the dataset is defined as:

$$err_n = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{\{Y_i \ne f(\boldsymbol{X}_i)\}} = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{\{Y_i \ne \hat{Y}_i\}} = \frac{FP + FN}{TP + FP + FN + TN} \tag{14}$$

err_n lies between 0 and 100% and the lower err_n is, the better classifier $f(.)$ fits the data from a general point of view. Since the case when the

Table 1. Illustration of a confusion matrix.

		Actual class	
		$Y = AT$	$Y = BT$
Predicted class	$\hat{Y} = AT$	True Positive (TP)	False Positive (FP)
	$\hat{Y} = BT$	False Negative (FN)	True Negative (TN)

indicator characterizing the state of the reactor during its shutdown exceeds the fixed threshold is more critical than the other one, we prefer to correctly predict the class $Y = AT$ and that is why we also introduce the sensitivity defined as:

$$se_n = \frac{TP}{TP + FN} \tag{15}$$

se_n lies between 0 and 100% and the higher se_n is, the better classifier $f(.)$ is able to predict the observations where $Y_i = AT$, among the n of the dataset $\left(Y_i, \boldsymbol{X}_i\right)_{1 \le i \le n}$.

3.2 *Overfitting and bootstrap*

A prediction model $f(.)$ is built using some set of "training data", that is to say exemplary situations for which the desired output is known. Of course, the goal is that $f(.)$ has good performance on the training dataset but also performs well on predicting the output when fed "validation data" (or "test data") that was not encountered during its training. "Overfitting" (sometimes called "overtraining") is the production of a prediction model that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional new data or predict future observations reliably. Overfitting occurs when a model begins to "memorize" training data rather than "learning" to generalize from a trend. The potential for overfitting depends not only on the number of parameters and data but also the conformability of the model structure with the data shape, and the magnitude of model error compared to the expected level of noise or error in the data. To lessen the chance of, or amount of, overfitting, several techniques are available, among which cross-validation, as introduced in Section 2.2, or bootstrap. These procedures consist of testing the model's ability to generalize by evaluating its performance on a set of data not used for training, which is assumed to approximate the typical unseen data that a model will encounter.

The basic idea of a bootstrap iteration, indexed by $b, 1 \le b \le B$, is to randomly draw without replacement from the original data $\left(Y_i, \boldsymbol{X}_i\right)_{1 \le i \le n}$ two complementary subsets: one training subset, of size $1 < N < n$ (in our case $N = \frac{3}{4} n$), which will be used to fit prediction model $f(.)$, and one validation subset, of size $n - N$, on which $f(.)$ will be applied to assess its performance on data that were not used to train it. On this test subset, one can evaluate error rate err_b and sensitivity se_b using Formulas (14) and (15) applied to the $n - N$ validation data. This is done B times (in our case $B = 1000$), producing B bootstrap training and validation datasets. We refit the model to each of the B bootstrap learning datasets and systematically test it on the B bootstrap test datasets, leading to B values for error rate and sensitivity $\left(err_b, se_b\right)_{1 \le b \le B}$. From these B values, one can assess the bootstrap average error rate and the bootstrap average sensitivity:

$$\widehat{err}_{Boot} = \frac{1}{B} \sum_{b=1}^{B} err_b \quad \text{and} \quad \widehat{se}_{Boot} = \frac{1}{B} \sum_{b=1}^{B} se_b \tag{16}$$

These two quantities better estimate the "real" performance of classifier $f(.)$ than err_n and se_n do and easily allow to identify a potential overfitting issue if $\left(\widehat{err}_{Boot}, \widehat{se}_{Boot}\right)$ is far from $\left(err_n, se_n\right)$.

3.3 *Results*

To implement the different algorithms, we used R, an open source language and software environment for statistical computing (https://www.r-project.org/). More precisely, the following packages were used: 'stats' for stepwise logistic regression, 'glmnet' for Ridge and LASSO, 'plsRglm' for PLS, 'spls' for Sparse PLS, 'rpart' for classification trees, 'adabag' for bagging, 'ada' for boosting, 'nnet' for neural network and 'e1071' for the different SVM versions.

Table 2 gives the performance indicators $\left(\widehat{err}_{Boot}, \widehat{se}_{Boot}\right)$ for each of the machine learning algorithms that have been applied to our dataset.

Table 2. Summary of the results obtained with the different algorithms.

Algorithm	$\widehat{err}_{Boot}$ (%)	$\widehat{se}_{Boot}$ (%)
Stepwise	4.49	92.59
Ridge	13.48	66.67
LASSO	10.11	74.07
PLS	3.38	93.24
Sparse PLS	14.61	93.87
Classification tree	5.72	96.15
Bagging	5.61	95.71
Boosting	**3.37**	**96.32**
Neural network	4.69	96.17
Linear SVM	4.39	90.86
Gaussian radial SVM	3.46	91.03
Polynomial SVM	5.62	88.89
Hyperbolic tangent SVM	7.86	85.11

4 DISCUSSION

Boosting is the most efficient approach, with both the lowest $\widehat{err}_{Boot}$ and the highest $\widehat{se}_{Boot}$. PLS logistic regression and neural network come in second position. Classification tree and bagging have promising sensitivity but somewhat disappointing error rate. The different SVM kernels give quite close medium results. The two shrinkage variants of the logistic regression, Ridge and LASSO, give the worst models, quite far behind all the other techniques.

Based on this single study, it is of course impossible to draw any general conclusions about the potential superiority of one algorithm over the others. Many other numerical tests should be carried out on simulated (and not real industrial) data to achieve such an ambitious goal. Nevertheless it is interesting to highlight that the performance indicators are rather good, even excellent, compared with the size of the dataset. Indeed, in our case, we are far from what is called "big data", since we only have $n = 89$ observations and $p = 25$ variables to predict our binary target output. This finding runs counter the popular belief that machine learning algorithms necessarily require a huge amount of data. Nevertheless one must not be mistaken: it is unrealistic to imagine these black-boxes will solve any problems and always be efficient, even with little data.

From the perspective of the decision maker, even if boosting is the most efficient algorithm in our case study, he may prefer a more interpretable model, such as the PLS logistic regression. Indeed this approach has a quite honorable prediction performance and provides at the same time an interpretable description of how the inputs affect the output (see Equation (8)). If the interpretability and the explanation of the fundamental principles of the model are not necessary in various fields, they become significant arguments when trackability, auditability, transparency or physical justification are required, as in nuclear industry.

Otherwise it is often pointed out that machine learning algorithms are greedy ones requiring prohibitive computational time and/or memory to train the models or predict the outputs. With our small dataset, it was not an issue on a standard computer, even when carrying out bootstrap procedure. We should also mention that open source statistical software, such as R, make such machine learning algorithms financially accessible to any companies, even if their use requires relevant expertise in order to properly parametrize the algorithms and avoid the potential pitfalls (for instance overfitting). The transferability outside R&D divisions of such black-box predictive models is also a real issue, in particular to engineering divisions which often only have spreadsheet application software to make calculations.

Last but not least: before applying any machine learning algorithm, quality of the input data must first be ensured. Moreover, it can only be profitable that the data scientist, who designs and manipulates these black-boxes, discusses with the experts of the technical application field.

5 CONCLUSION AND PROSPECTS

Thirteen supervised machine learning techniques have been tested on a real dataset from the nuclear industry. The fundamental principles of these algorithms have been presented and their prediction performance has been assessed on the case study.

The most efficient methods give really promising results, especially compared to the small size of the available data. Nevertheless one would be well advised not to draw any general conclusions about the efficiency of these techniques.

Several prospects and extensions of this study can be envisaged. Other machine learning algorithms, such as discriminant analysis or random forests, could be tested on the same dataset to assess how they perform. An intensive study based on simulated data could also be carried out to try to identify if some algorithms are more efficient than others on datasets with characteristics close to those met in the nuclear industry.

REFERENCES

Breiman, L. 1996. Bagging predictors. *Machine Learning* 26: 123–140.

Burges, C. 1998. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining* 2(2): 121–167.

Chun, H. & Keles, S. 2010. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society – Series B* 72(1): 3–25.

Hastie, T., Tibshirani, R. & Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction – 2nd edition*. Verlag New York: Springer Series in Statistics.

Rumelhart, D., Hinton, G. & Williams, R. 1986. Learning internal representations by error propagation, in D. Rumelhart & J. McClelland (eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, Cambridge, MA., 318–362.

Wold, S., Martens, H. & Wold, H. 1983. The multivariate calibration problem in chemistry solved by the PLS method. In B. Kagstrom & A. Ruhe (eds), *Matrix Pencils. Lecture Notes in Mathematics*, vol. 973. Springer, Berlin, Heidelberg.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

New resilience performance indices based on the *k*-terminal reliability of the complete graph

C. Tanguy
Orange Labs, Orange/IMT/OLN/GDM/TRM, Châtillon, France

ABSTRACT: In network reliability, the best-known performance measure is the so-called all-terminal reliability $\mathrm{Rel}_A(G)$, i.e., the probability that all the nodes of the network (or its underlying graph G) are connected. A particular family of graphs, namely the complete graphs K_n, in which each node is connected to the $n-1$ others, have long been a key issue in this field. Brown, Cox and Ehrenborg have recently proposed new performance indices for networks based on the all-terminal reliability $\mathrm{Rel}_A(K_n)$: (i) the average reliability $\overline{\mathrm{Rel}_A}(K_n)$, (ii) the "average of the average" all-terminal reliability AvgAvg (G, n) of all the graphs with n vertices and at most one edge between two nodes. They showed that as n increases, these measures tend to 1. In this work, we generalize their idea to the *k*-terminal reliability—the probability that k nodes are connected—which can also be used to describe a network's resilience. The new measures can be derived from the *k*-terminal reliability of the complete graph. Since we are interested in the application of these indices to large systems (n >> 1), we have performed numerical investigations to assess their variations with n. We have identified the leading, analytical correction (in $1/n$) to unity. This corroborates a previous conjecture made on the asymptotic value of $\overline{\mathrm{Rel}_A}(K_n)$. These new results could be helpful as simple, ready-to-use evaluations/orders of magnitude of the resilience of a network, when its size is so large as to make exact or approximate computations either impossible or very cumbersome.

1 INTRODUCTION

Complete graphs are graphs in which each vertex is connected to all other vertices (see Figure 1 for the first examples of K_n, the complete graph with n vertices). They have attracted interest for a very long time. Among the pioneering papers on random graphs (Erdös and Rényi 1959, Gilbert 1959), one focused on the possible application to the probability P_N that N telephone central offices can call each other (Gilbert 1959), or that two offices can be connected (with probability R_N), assuming that the probability of connection between two nodes is p. The following simple expressions were given (Gilbert 1959, Frank and Gaul 1982)

$$P_N \sim 1 - N q^{N-1}, \tag{1}$$

$$R_N \sim 1 - 2q^{N-1}, \tag{2}$$

where $q = 1 - p$.

More recently, complete graphs K_n have been investigated in the context of the resilience of large networks (Sekine and Imai 1998, Imai et al. 1999, Tsitsiashvili 2011), and in particular for wireless networks (Park 2016). They also appear in performance measures recently defined by Brown and collaborators (Cox 2013, Brown et al. 2014).

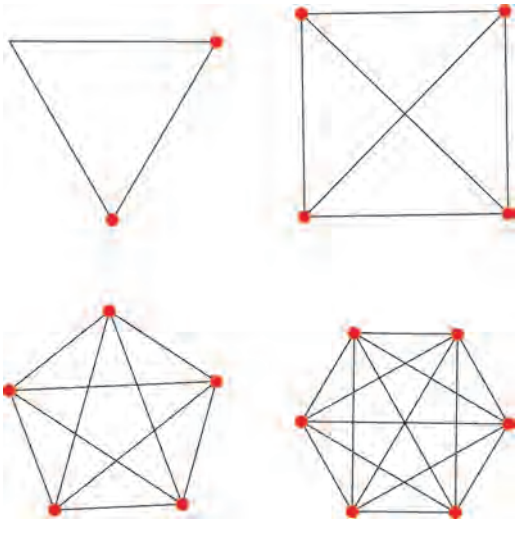


Figure 1. First complete graphs K_n ($3 \le n \le 6$).

The first one is simply the average all-terminal reliability of a graph G over the interval [0,1]:

$$\overline{\mathrm{Rel}_A}(G) = \int_0^1 \mathrm{Rel}_A(G;p)\,dp. \tag{3}$$

The second one is the "average of the average" reliability of simple graphs (a "simple" graph is such that there is at most one link/edge between two given vertices) G with n vertices, AvgAvg(G, n), which can be expressed (Cox 2013) as

$$\mathrm{AvgAvg}(G,n) = 2\int_0^{1/2} \mathrm{Rel}_A\left(K_n;p\right)dp. \quad (4)$$

Cox and collaborators (Cox 2013, Brown et al. 2014) showed that in the $n \to \infty$ limit, these two averages should go to one.

Multicast procedures in modern networks call for accurate assessments of the k-terminal reliabilities $\mathrm{Rel}_k(p)$ too. Let us recall that the k-terminal reliability is the probability that the k vertices of interest are connected. In this context, it is hardy surprising that the resilience of such procedures has stimulated a lot of work, from very mathematical approaches to more pragmatic ones. Its definition may also vary among authors: some of them (Colbourn 1993, Farley and Colbourn 2009) associate it with the number of nodes that are still connected, whereas others (Cox 2013, Brown et al. 2014, Heidtmann 2016) compute it as the average of all possible connections between k nodes of the system.

In a recent work (Tanguy 2017), we improved the asymptotic expansions of P_N and R_N, and generalized them to the k-terminal reliability for the complete graph K_n. We also provided an estimate of the asymptotic expansion of $\overline{Rel_A}\left(K_n\right)$, derived from numerical simulations.

In this work, we have addressed the generalization of equations (3) and (4) to the k-terminal reliability of k specific nodes of the system:

$$\overline{\mathrm{Rel}_k}(G) = \int_0^1 \mathrm{Rel}_k\left(G;p\right)dp. \quad (5)$$

The second generalization is the "average of the average" reliability of simple graphs G with n vertices, AvgAvg(G, n), which can be expressed (Cox 2013) as (see below for a justification)

$$\mathrm{AvgAvg}\left(\mathrm{Rel}_k;G,n\right) = 2\int_0^{1/2} \mathrm{Rel}_k\left(K_n;p\right)dp. \quad (6)$$

When all the link reliabilities are equal to p, all nodes are equivalent. The above expressions give therefore a direct expression for the resilience as viewed by Cox (2013), Brown et al. (2014) and (Heidtmann 2016).

This paper is organized as follows: we describe in Section 2 how to compute the k-terminal reliability for the complete graph K_n, and the associated averages. We then show in Section 3 a few figures showing how these averages approach 1 when n goes to infinity. We describe in Section 4 the determination of the first-order correction to unity for several values of k, from which we propose a closed-formed expression. We finally apply these results to the case of the all-terminal reliability of the complete graph K_n.

2 MATHEMATICAL PROCEDURES AND DEFINITIONS

2.1 *Asymptotic expansions of the reliabilities of the complete graph K_n*

The all-terminal reliability $\mathrm{Rel}_A\left(K_n;p\right) \equiv A_n$ can be obtained recursively by (Gilbert 1959).

$$A_n = 1 - \sum_{j=1}^{n} C_{n-1}^{j-1} A_j (1-p)^{j(n-j)}, \quad (7)$$

$$A_1 = 1, \quad (8)$$

where C_n^k is the binomial coefficient

$$C_n^k = \frac{n!}{(n-k)!k!}. \quad (9)$$

The k-terminal reliability $\mathrm{Rel}_k\left(K_n;p\right) \equiv B_n$ is *then* deduced from all the A_j's ($1 \le j \le n$) by

$$B_n = \sum_{j=k}^{n} C_{n-k}^{j-k} A_j (1-p)^{j(n-j)}, \quad (10)$$

2.2 *Asymptotic expansions of A_n and B_n*

In our previous work (Tanguy 2017), we extended Gilbert's results (Gilbert 1959) for the asymptotic expansions in the case of large n and fixed q's.

$$A_n \to 1 - n q^{n-1} + \frac{n(n-1)}{2}\left(-1+2q\right)q^{2n-4} + \cdots. \quad (11)$$

from which we derived

$$B_n \to 1 - k\,q^{n-1} + q^{2n-4}\left\{C_k^2 - k\left(n-1\right)\left(1-q\right) + C_{n-k}^{2-k}\left(1-q\right)\right\} + \cdots. \quad (12)$$

Equation (12) will be used in the following.

2.3 *Average k-terminal reliabilities*

We first start by defining, for k specific nodes,

$$\overline{\mathrm{Rel}_k}\left(K_n\right) = \int_0^1 \mathrm{Rel}_k\left(K_n;p\right)dp. \quad (13)$$

The second generalization is the "average of the average" reliability of simple graphs G

with n vertices, AvgAvg(G, n), which can be expressed (Cox 2013) as

$$\mathrm{AvgAvg}\left(\mathrm{Rel}_k;G,n\right)=\int_0^1 \mathrm{Rel}_k\left(K_n;p/2\right)dp. \qquad (14)$$

The origin of equation (14) is simple: all the k-terminal reliabilities are affine functions of each individual link reliability. Since we consider all possible graphs on n vertices, each link will be present (or not). The average is thus given by the corresponding reliability of the complete graph K_n, for which each link has a reliability equal to $(0+p)/2 = p/2$. We can thus write

$$\mathrm{AvgAvg}\left(\mathrm{Rel}_k;G,n\right)=2\int_0^{1/2} \mathrm{Rel}_k\left(K_n;t\right)dt. \qquad (15)$$

We deduce that

$$\mathrm{AvgAvg}\left(\mathrm{Rel}_k;G,n\right)=2\overline{\mathrm{Rel}_k}\left(K_n\right)-1 + 2\int_{1/2}^1\left(1-\mathrm{Rel}_k\left(K_n;t\right)\right)dt. \qquad (16)$$

Because of equation (12), the last integral in the above equation vanishes as $(2\,k)/(n\,2^n)$ when n goes to infinity. Consequently, the asymptotic expansions of the averages are linked by

$$1-\mathrm{AvgAvg}\left(\mathrm{Rel}_k;G,n\right)\sim 2\left(1-\overline{\mathrm{Rel}_k}\left(K_n\right)\right). \qquad (17)$$

By studying the behavior of $\overline{\mathrm{Rel}_k}\left(K_n\right)$ when $n \gg 1$, we can get the asymptotic expansion of AvgAvg(Rel_k; G, n) too.

3 EXAMPLES FOR $2 \le K \le 4$

In this section, we present the results of our numerical calculations of $\overline{\mathrm{Rel}_k}\left(K_n\right)$ and AvgAvg(Rel_k; G, n) as functions of n, for the first values of k.

We obtained the different $\mathrm{Rel}_k(K_n; p)$ from equation (10). Since they are polynomials in p with integral coefficients, the integrals in equations (13) and (14) were easily obtained.

3.1 $\overline{Rel_k}\left(K_n\right)$

We have represented in Figures 2–4 their variations with n. Obviously, these quantities go rapidly to 1, so much so that the curves cannot be distinguished from unity when $n \ge 200$.

3.2 *AvgAvg(Rel$_k$; G, n)*

We have displayed in Figures 5–7 the variation with n of AvgAvg(Rel_k; G, n) for $2 \le k \le 4$. Even though

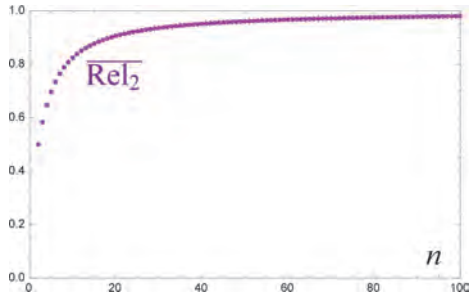


Figure 2. Variation of $\overline{\mathrm{Rel}_2}\left(K_n\right)$ with n.

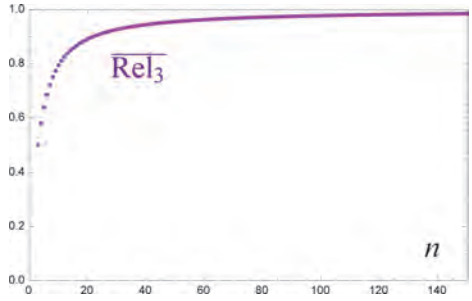


Figure 3. Variation of $\overline{\mathrm{Rel}_3}\left(K_n\right)$ with n.

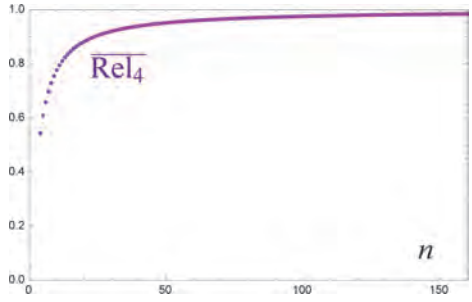


Figure 4. Variation of $\overline{\mathrm{Rel}_4}\left(K_n\right)$ with n.

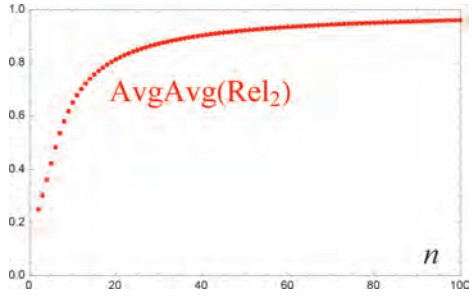


Figure 5. Variation of AvgAvg(Rel_2; G, n) with n.

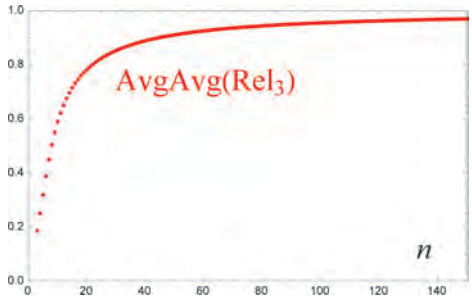


Figure 6. Variation of AvgAvg(Rel_3; G, n) with n.

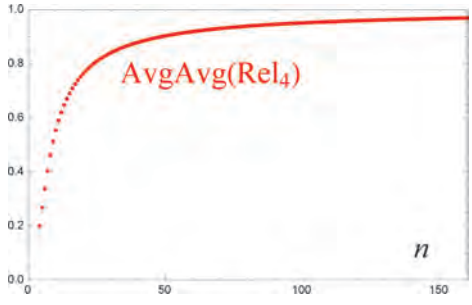


Figure 7. Variation of AvgAvg(Rel_4; G, n) with n.

the values go to 1 as n increases, they do so less rapidly than the average reliabilities of the preceding paragraph, in agreement with equation (17).

4 ASYMPTOTIC EXPANSIONS OF THE AVERAGES

From the numerical values obtained in Section 3, we were able to see that

$$\overline{\mathrm{Rel}_k}(K_n) = 1 - \frac{C_k}{n} + o\left(\frac{1}{n}\right). \tag{18}$$

The determination of C_k has been done by using "convergence acceleration methods". Even though the convergence to 1 is rather slow, it is possible to assess the limit C_k by using the Richardson extrapolation (Bender and Orszag 1999). Performing these computations, we found that $C_2 \approx 2.00000000$, $C_3 \approx 2.24999999$, and $C_4 \approx 2.4444444$, implying that C_2, C_3 and C_4 may well be 2, $\frac{9}{4}$, and $\frac{22}{9}$, respectively. This was a strong indication that all the C_k's are rational numbers. Using large values of n (up to 700), we were able to increase the accuracy of our evaluations of the first constants, the identifications of which are given in Table 1.

Table 1. Values of the first constants C_k ($2 \le k \le 10$).

k	C_k
2	2
3	$\frac{9}{4}$
4	$\frac{22}{9}$
5	$\frac{125}{48}$
6	$\frac{137}{50}$
7	$\frac{343}{120}$
8	$\frac{726}{245}$
9	$\frac{6849}{2240}$
10	$\frac{7129}{2268}$

The last part of the study was to infer the general formula for C_k. By trial and error, and taking the factorization of the denominators into account, we were able to find that

$$C_k = \frac{k}{k-1} H_{k-1}, \tag{19}$$

where $H_k = 1 + \frac{1}{2} + \cdots + \frac{1}{k}$ is the so-called harmonic number of order k.

As k increases, the precision of C_k gets greater. We have checked that equation (21) still gives satisfactory values; for instance, in the case $k = 150$, the difference between our estimate and C_{150} is less than 10^{-100}.

4.1 *Consequence for* $\overline{Rel_A}(K_n)$

In (Tanguy 2017), we had computed the average $\overline{\mathrm{Rel}_A}(K_n)$ as a function of n, and proposed that

$$\overline{\mathrm{Rel}_A}(K_n) \sim 1 - \frac{ln\, n + \mathbf{C}}{n}, \tag{20}$$

where $\mathbf{C}$ is the Euler gamma constant (using the notation of Gradshteyn and Ryzhik $\mathbf{C} \approx 0.5772156649$). This assumption was correct since for the all-terminal reliability, we have to replace k by n, and because

$$H_n \to \ln n + \mathbf{C} \tag{21}$$

as $n \to \infty$.

4.2 *Asymptotic expansion of* $AvgAvg(Rel_k; G, n)$

If we now turn to AvgAvg(Rel_k; G, n), we deduce from equation (17)

$$\mathrm{AvgAvg}\left(\mathrm{Rel}_k; G, n\right) \sim 1 - \frac{2\,k\,H_{k-1}}{k-1}\frac{1}{n}. \tag{22}$$

A by-product of this expression is of course

$$\mathrm{AvgAvg}\left(\mathrm{Rel}_A; G, n\right) \sim 1 - 2\frac{ln\,n + \boldsymbol{C}}{n}. \tag{23}$$

5 CONCLUSION AND OUTLOOK

We have generalized the asymptotic expansions of (Cox 2013, Brown et al. 2014) for averaged k-terminal reliabilities of the complete graph K_n, which are performance measures recently introduced for the resilience of large networks. The first correction to unity, as n goes to infinity, has a very simple form, and confirms the assumption made in a previous work (Tanguy 2017) on the averaged all-terminal reliability for the complete graph. Other performance measures for the complete graph are currently under investigation, and will be presented elsewhere.

ACKNOWLEDGMENTS

I wish to express my gratitude to my Orange colleagues Christian Bourliataud and Thomas Rivera for using their computing facility, and Éric Gourdin for discussions leading to this work.

REFERENCES

Bender, C.M. & S.A. Orszag (1999). *Advanced Mathematical Methods for Scientists and Engineers I Asymptotic Methods and Perturbation Theory*. Springer-Verlag, New York.

Brown, J.I., D. Cox, & R. Ehrenborg (2014). The average reliability of a graph. *Discrete Applied Mathematics 177*, 19–33.

Colbourn, C.J. (1993). Analysis and synthesis problems for network resilience. *Mathematical and Computer Modelling 17*(11), 43–48.

Cox, D.N. (2013). *On Network Reliability*. Ph.D. thesis, Dalhousie University, Canada.

Erdös, P. & A. Rényi (1959). On random graphs I. *Publicationes Mathematicae (Debrecen) 6*, 290–297.

Farley, T.R. & C.J. Colbourn (2009, October). Multiterminal measures for network reliability and resilience. In *7th International Workshop on Design of Reliable Communication Networks (DRCN 2009)*, pp. 107–114.

Frank, O. & W. Gaul (1982). On reliability in stochastic graphs. *Networks 12*(2), 119–126.

Gilbert, E.N. (1959). Random graphs. *Ann. Math. Statist. 30*(4), 1141–1144.

Heidtmann, K. (2016). ResiNet3: An immediately and easily usable tool for exact computation of network reliability and resilience. Technical report, Department of Informatics, University of Hamburg, Germany.

Imai, H., K. Sekine, & K. Imai (1999). Computational investigations of all-terminal network reliability via BDDs. *IEICE Transactions on Fundamentals E82-A*(5), 714–721.

Park, J.-H. (2016). All-terminal reliability analysis of wire-less networks of redundant radio modules. *IEEE Internet of Things Journal 3*(2), 219–230.

Sekine, K. & H. Imai (1998). Computation of the network reliability (extended abstract). Technical report, Department of Information Science, University of Tokyo.

Tanguy, C. (2017, June). Complete graph reliabilities: Asymptotic results. In *Mathematical Methods in Reliability (MMR 2017)*, pp. 1–6.

Tsitsiashvili, G. (2011). Cooperative effects in complete graph with low reliable arcs. *Reliabity Theory and Applications 2*(3), 58–62.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A mathematical programming approach to railway network asset management

C. Fecarotti & J. Andrews
Resilience Engineering Research Group, Department of Civil Engineering, The University of Nottingham, Nottingham, UK

ABSTRACT: A main challenge in railway asset management is selecting the maintenance strategies to apply to each asset on the network in order to effectively manage the railway infrastructure given that some performance and safety targets have to be met under budget constraints. Due to economic, functional and operational dependencies between different assets and different sections of the network, optimal solutions at network level not always include the best strategies available for each asset group. This paper presents a modelling approach to support decisions on how to effectively maintain a railway infrastructure system. For each railway asset, asset state models combining degradation and maintenance are used to assess the impact of any maintenance strategy on the future asset performance. The asset state models inform a network-level optimisation model aimed at selecting the best combination of maintenance strategies to manage each section of a given railway network in order to minimise the impact of the assets conditions on service, given budget constraints and performance targets. The optimisation problem is formulated as an integer-programming model. By varying the model parameters, scenario analysis can be performed so that the infrastructure manager is provided with a range of solutions for different combination of budget available and performance targets.

1 INTRODUCTION

The railway system is the result of the interaction of a number of different systems and infrastructure with the ultimate aim of transporting people and goods safely and on time. It consists of a diverse portfolio of assets, each bounds to deliver a specific function but all together contributing to ultimately provide a reliable and safe service. Each railway asset is subject to degradation and failure processes, and maintenance is performed in order to control the state of the assets and ensure that each asset's function is performed to the required standard. Maintenance policies are developed as a combination of periodic inspection, routine and emergency maintenance, enhancement and renewal activities, and these are specific to each railway asset. As maintenance resources and budget are limited, decisions have to be made on how to optimally allocate the available resources among all the asset on the network. Infrastructure asset management is the process of allocating maintenance resources among the assets comprising the system with the aim of minimising the whole-life costs while maximising the system performance. Optimal asset management involves decision making and selection of the best intervention strategy for each asset along the network in order to ensure that the required level of service reliability and safety risk is achieved within budget. Determining the best set of strategies for a given network does not simply consist in choosing the strategy which is optimal for each asset. When a network perspective is adopted dependencies among different assets and different sections of the network arise, due for example to resource availability. This implies that intervention strategies that are optimal when an asset is considered individually, might not be optimal when decisions are made at a network level.

1.1 *Modelling approaches to infrastructure asset management optimisation*

Optimisation models have been presented in the literature to support infrastructure asset management from different perspectives and to address different aspects of the problem. Two main approaches to the optimisation of infrastructure asset management can be identified: asset-level and system-level optimisation. Assetlevel optimisation aims at determining optimal maintenance policies for an individual asset, while systemlevel optimisation seeks the optimal combination of maintenance policies for all the assets comprising the system. The focus of this paper is on system-level optimisation; in the following, the modelling

approaches developed in the literature to determine the optimal set of maintenance policies for infrastructure systems composed of multiple assets are briefly discussed.

The authors in Yeo et al. (2012) address the problem of planning maintenance for a system of heterogeneous facilities undergoing stochastic deterioration over a finite time horizon. They develop a twostage bottom-up approach according to which optimal maintenance policies are first determined for each facility. The deterioration of each facility is modelled as a Markov process. The state of the facilities is known at the beginning of every year when inspection is performed and maintenance activities are selected year by year. The authors apply a dynamic programming approach to find the optimal activity as well as the alternative near optimal activities and associated costs for each facility. Then, a system-level optimisation is developed to obtain the combination of activities, one for each facility, that minimise the system expected cost-to-go while the agency cost (cost of the maintenance activities) is kept within a given budget. All facilities in the system are considered to be independent and the system-level optimisation problem is formulated as a constrained combinatorial problem.

A similar approach has been used in Furuya & Madanat (2013) with application to a hypothetical railway system, where facility-level and system-level optimisation are combined to obtain the best combination of activities for all facilities in a given network. The authors demonstrate their approach on a hypothetical dual redundant railway network. A number of facilities are associated to each link in the network, and a set of available maintenance activities is considered for each facility. As in Yeo et al. (2012), the degradation and maintenance of the railway assets is modelled as a Markov process, and the facility-level optimisation problem is formulated as a Markov decision process solved through dynamic programming. In the system-level optimisation problem, the budget constraint includes the cost reduction that can be achieved when adjacent facilities are maintained simultaneously. Constraints are also formulated on the minimum capacity to be guaranteed between an origin and a destination node and for each individual route. This enables to consider the loss of throughput due to maintaining adjacent facilities simultaneously. A numerical example is solved, which demonstrates how including both economic (opportunistic maintenance) and functional (capacity loss) dependencies arising between the assets when performing maintenance, has an impact on the optimal decision and associated lifecycle cost.

In Robelin & Madanat (2008) the authors address the optimisation of maintenance policies for a system of bridge decks with the objective of determining the optimal set of policies based on the current system conditions as well as the prediction of future conditions. The deterioration model of an individual deck is Markovian, where each state is defined in terms of the current condition of the deck, the last maintenance action performed and the time since the last intervention. The condition of a deck is given by its instantaneous probability of failure. A two-steps approach is suggested. First, a facility-level optimisation is solved to obtain the optimal cost of maintenance and replacement for each facility. The facility-level optimisation problem is solved for a discrete range of failure probabilities. Then, at system level, the cost of the system given by the combination of the cost for each facility, is minimised subject to budget constraint, and the optimal threshold of failure probability is obtained. This threshold is used backward within the facilitylevel optimisation to obtain the set of policies for each deck which are optimal at system level. Some of the assumptions the optimisation model in Robelin & Madanat (2008) is based on are too restrictive to be applied to the railway system. Many of the railway assets exhibit multiple failure modes, each with different probabilities and frequencies of occurrence. Different failure modes usually have different effects on system performance and must be therefore considered individually. Decisions on maintenance policies must account for the different failure modes so that different effects on service performance can be distinguished, and both safety and performance requirements can be addressed in a cost effective manner. Another simplifying assumption made in this paper is that at system level, the optimal threshold of probability of failure is the same for all the facilities. While this makes the optimisation problem easier to solve, it also produces a less realistic model. In real systems the location of the assets on the network may play an important role within the decision making process. The railway network includes lines and routes with different criticalities corresponding to different safety and service performance targets. It is often the case that in the trade-off between cost and performance, more expensive policies are likely to be implemented on assets located on lines with higher criticality, while lower performance is accepted on lower criticality lines.

The author in Durango-Cohen (2007) presents a method to simultaneously address the conditions and costs forecasting problem and the optimisation of maintenance action for transportation infrastructure facilities. Facilities deterioration is represented as an autoregressive moving average with exogenous input model (ARMAX). Decision variables can be investment levels or maintenance

rates and the optimisation problem is formulated as a dynamic program seeking the minimum expected discounted cost over the planning horizon. Decisions are made based on the information available at the beginning of the planning period. The use of the ARMAX model is based on the assumption that the effects of maintenance actions are linear and additive. This assumption however is too restrictive for many railway assets (e.g. track) as it completely disregards the complexity of the combined effects of different interventions on the future asset state and the consequent impact on costs.

The approach presented in the aforementioned papers is aimed at selecting the maintenance policies to be adopted year by year over a given time horizon. Inspection is not considered as part of the policies as it is assumed to be carried out at the beginning of every year. However, the frequency of inspection is an important aspect of every maintenance policy as it allows to reveal the conditions of an asset before failures occur or unacceptable degraded states are reached. Indeed it is the optimal combination of inspection frequency, threshold values for assets conditions triggering interventions and the time required to perform maintenance that make an effective maintenance strategy. Furthermore, most of the contributions use a Markov approach to model the degradation and maintenance processes of the assets. However, the Markov approach has a few limitations that prevent it from being an effective modelling tool for many of the railway assets. A significant limitation is the requirement of Markov models to restrict transitions between states on the model (generally representing degradation or repair) to occur at a constant rate. This means that the state residence times are exponentially distributed. The memoryless property of the Markov approach restricts the ability of the model to consider the maintenance history which is important in some of the railway asset components such as the track ballast. Furthermore, the size of a Markov model can experience a state-space explosion with the number of components considered, thus making difficult to model assets with many different components or formed linking several sections of track. One final significant limitation is its inability to represent a route or network perspective. If Markov models exist for two assets and it is required to account for their dependencies in constructing a route model, this can only be accomplished by the generation of a completely new model.

An alternative modelling technique that overcomes some of the limitations of the Markov approach in modelling railway asset degradation and maintenance is the Petri Net (PN) method. PNs are a formalism for modelling complex, dynamic systems characterised by concurrency and dependencies, synchronisation and resource sharing. PNs provide a valuable mathematical and graphical description of the system behaviour. PNs is a stochastic technique which allows far greater detail in comparison to the alternatives when modelling assets degradation and complex management strategies, whilst maintaining a manageable model size. PNs account for any distribution of degradation and failure times; thus increasing failure rate typical of components subject to wear-out can be considered. PNs also enable the modelling of complex maintenance processes including condition and risked based inspection and maintenance, replacement prior to failure based on either age, condition or use, reactive repair, refurbishment and renewal and all the rules for the implementation of such activities. The resulting PN models are usually smaller in size than the alternative Markov representation. An additional and very desirable feature of PN models is their modularity. Models of assets consisting of many interacting components can be built up in parts giving the model a modular structure which is easier to analyse. Monte Carlo simulation is the most common solution technique for PN models and produces distributions for the output variables of interest. The PN approach is suggested in this paper as a valid modelling technique to produce models that combine the degradation and maintenance processes involving the railway assets. Such models can be used as a tool to investigate the effectiveness of a variety of maintenance strategies for each railway asset, covering a range of performance and costs, so to provide the decision maker with a set of potential strategies among which the ones which are best from a system perspective can be selected.

2 THE METHODOLOGY

This paper presents a modelling approach to support decisions on how to effectively maintain a railway infrastructure system. First, for each railway asset, a modelling tool is required to assess the asset response to the implementation of a range of feasible maintenance strategies. Such modelling tools, called *asset state models* combine the degradation/failure processes affecting the asset with the intervention activities that can be performed in order to predict the future asset state. The asset state models developed for each asset inform a network-level optimisation model aimed at selecting the best combination of maintenance strategies to manage all the assets on a railway network under budget and performance constraints. The network-level optimisation model is formulated as an integer program with multiple constraints (Hillier and

Lieberman 2009). The model is bounded to select one option for each individual asset located in the considered railway network. Constraints are formulated on the overall available budget and on the availability required of each railway line. Different lines in the network may have a different criticality depending on the effect that failures have on service. This is strictly linked to the frequency of the service running on each line. Different lines criticality are accounted for by imposing different thresholds to the availability of each line. This modelling approach has the advantage to enable the evaluation of a variety of different scenarios by changing the model parameters such as the available budget or the threshold levels set for the lines availability.

2.1 *Network segmentation for strategic planning purposes*

The UK railway network is segmented for policy decisions. The whole network is divided into 19 Strategic Routes, each divided into a number of Strategic Route Sections (SRSs). An SRS is a section of the railway network characterised by broadly homogeneous infrastructure type and traffic levels. Therefore strategy decisions are taken at SRS level. It is assumed that the same maintenance strategy will be applied within the same SRS. Asset state models are developed for each asset type existing on each SRS and are used to assess the impact of a range of maintenance strategies on the assets' performance.

2.2 *Asset state models*

The PN method is adopted as the modelling approach to develop the asset state models. PNs are a formalism for modelling complex distributed systems characterised by concurrency and dependency, synchronization and resource sharing. Petri nets provide a valuable mathematical and graphical description of the system behaviour. A PN is a directed, weighted bi-partite graph where nodes are places and transitions connected by arcs (Murata 1989). A PN can be formally defined as follows.

Definition 1. A PN is a 5-tuple $PN = (P, T, A, W, M_0)$ where: $P = \{p_1, p_2, ..., p_m\}$ is the non-empty set of places, $T = \{t_1, t_2, ..., t_n\}$ is the non-empty set of transitions, $P \cap T = \varnothing$, $A \subseteq (P \times T) \cup (T \times P)$ is the set of arcs, $W : A \to \{1,2,\ldots\}$ is the multiplicity function, $M_0 : P \to \{0,1,2,\ldots\}$ is the initial marking.

Places may represent physical resources, conditions or the state of a component. Tokens are held in places and the number of tokens in each place defines the marking of the Petri net which represents the state of the system at a given time. The flow of tokens through the network is determined by transitions and represents the evolution of the system state over time. Transitions represent events that make the status of the system change. Arcs only connect places with transitions (input arcs) and vice versa (output arcs). Inhibitor arcs are defined as well, which can be used to stop the firing of a transition under certain circumstances. Arcs are characterised by a multiplicity. The marking and the multiplicity of the arcs determine the enabling conditions for each transition. Transitions can be deterministic or stochastic. The former have an associated constant firing time, while the latter sample their firing time from stochastic distributions. Firing of transitions is ruled as follow:

- If the number of tokens contained in the input places is at least equal to the multiplicity of the associated input arcs, and the number of tokens in the places connected by inhibitor arcs is lower than the arcs multiplicity, then the transition is enabled.
- Once the transition is enabled, it will fire after a time interval which is fixed for deterministic transitions. For stochastic transitions the firing time is sampled from a probabilistic distribution.
- When the firing time is reached and the transition fires, a number of tokens is removed from the input places and added to the output places according to the arcs multiplicity.

For the purpose of maintenance, a number of discrete states are usually considered, corresponding to levels of degradation that trigger maintenance interventions with different levels of urgency. The degradation process can therefore be represented as a chain of places and transitions as shown in Figure 1. Places $P_{deg,i}$ indicate different states which are relevant from a maintenance perspective, namely each state (except for the new state P_{new}) triggers maintenance with different level of urgency depending on the level of degradation. Transitions $T_{deg,i}$ represent the degradation from one state to the next (worse). These are stochastic transitions whose firing time is sampled from a stochastic distribution representing the distribution of times to degrade between two consecutive states. Asset conditions requiring a speed restriction or a line closure can be included as well, these being usually the last two levels of degradation. Inspection is performed periodically to reveal the current asset

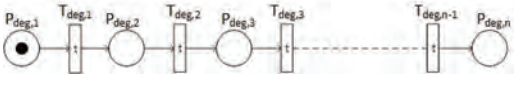


Figure 1. Degradation.

condition so that degraded states can be discovered and maintenance planned accordingly. In Figure 2) transitions $T_{rev,i}$ are timed deterministic and fire at a fixed frequency. Once a degraded condition is revealed, maintenance in planned depending on the level of urgency. Maintenance interventions are represented by transitions $T_{rep,i}$ (Figure 3). After maintenance, the asset is usually restored to a good condition (P_{good}) rather than to new, unless a renewal is carried out. If necessary, it is possible to account for the effectiveness of maintenance by adopting a probabilistic routing policy for transitions $T_{rep,i}$ so that the state after maintenance can be any of the degraded state with a given probability. It is also possible to keep track of the number of maintenance interventions performed. This is achieved by monitoring the marking of place *PI* which is marked every time an intervention is performed (and therefore any of transitions $T_{rep,i}$ fires). For some assets, the degradation might depend on the past maintenance history; an example is the ballast for which the rate of degradation increases with the number of tamping interventions performed. This can be accounted for if transitions $T_{deg,i}$ update their distributions of times to degrade according to number of interventions performed on the asset. This modelling approach enables the evaluation of a wide range of maintenance strategies, for each of which it is possible to specify the inspection frequency, the thresholds on the asset conditions that

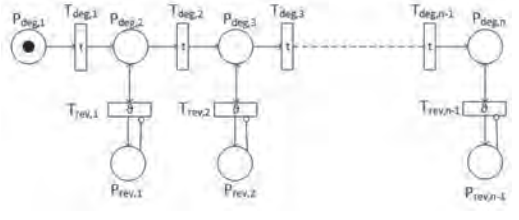


Figure 2. Degradation and inspection.

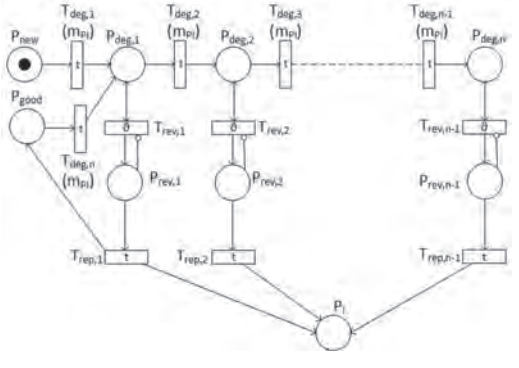


Figure 3. Degradation, inspection and repair.

trigger maintenance, the mean time to schedule and perform any maintenance activity. Furthermore, by keeping track of the marking during the simulation, it is possible to evaluate the probability of being in any of the considered states as well as the number of interventions performed. The probability of having a speed restriction and a line closure is of particular interest to evaluate the impact of a given strategy on service and safety risk.

This modelling structure can be used as a *modelling template* to describe a variety of railway asset exhibiting degradation during their lifetime. The number and features of places and transitions representing the degradation processes and the maintenance activities can be easily fitted to characteristics of the specific asset to be modelled. Example of degradation and maintenance models adopting a similar structure have been proposed in the literature for a number of railway assets such as track (Andrews 2012, Prescott & Andrews 2013, Andrews, Prescott, & De Rozieres 2014) and bridges (Le & Andrews 2016, Le, Andrews, & Fecarotti 2017).

2.3 *Network-level strategies optimisation*

The analysis conducted by means of the asset state models results in a set of potential asset management strategies covering a range of performance levels for each asset group. Given a set of potential strategies for each asset group, the infrastructure manager is faced with the task of selecting one strategy for each asset on the network given that a limited budget is available. Performance and safety targets are usually set for each route and line along the network, and these targets can be different depending on the route criticality. Decisions are therefore bounded by the available budget and are made with the aim of minimising the disruption caused to the railway service, while a certain level of availability is ensured for each line depending on the line criticality. Whatever asset fails, the impact on trains service is due to either a speed restriction, leading to delays, or a section closure leading to journeys cancellation. The extent of the disruption depends on both the duration of such control actions and the location of the section(s) involved. If a speed restriction or a section closure is imposed on a section belonging to a high frequency line, or to more than one line, then the number of journey affected by the disruption will be high. With regard to the impact of failures on service, for each section in the network is therefore fair to define two failure modes, each with a different effect on service: (i) section subject to speed restriction, and (ii) section subject to closure.

Let us define a Strategic Route as a set of SRSs $R = \{R_1, R_2, \ldots, R_i, \ldots, R_{n_R}\}$. Railway

services run along a set of railway lines $L=\{L_1,L_2,\ldots,L_l,\ldots,L_{n_L}\}$, each railway line consists of one or more SRSs. Therefore each railway line can be represented as a subset of set R, $L_l \subseteq R, \forall l=1,\ldots,n_L$. A railway line L_l will be unavailable if any of its SRS is unavailable. If a is the number of asset groups considered and b is the number of strategies available for each asset group, then the set of maintenance strategies for each SRS is given by all the possible combinations of the individual asset groups' strategies $n_S = a \cdot b$. The set $S=\{S_1,S_2,\ldots,S_j,\ldots,S_{n_S}\}$ is defined, containing n_S potential strategies available for each SRS, each corresponding to a given combination of the individual asset strategies. From now on the term strategy will be used to indicate a strategy for the individual SRS, among the available ones in set S. The index $j=1,2,\ldots,n_S$ will be used to refer to a generic strategy within set S while the index $j=1,2,\ldots,n_R$ will be used to refer to a generic SRS within set R. The vector of decisional variables X has components x_{ij} such that $x_{ij}=1$ if strategy j is applied to SRS i, 0 otherwise. The infrastructure manager is bounded to choose only one strategy per SRS. Following the implementation of a given strategy, each SRS will be subjected to a given probability, average number and duration of imposed speed restrictions and section closure during the considered planning period. Section closure contributes to define the availability of the SRS. In fact a section closure means that the section is not available for use and therefore all the journeys that use that section are cancelled or rerouted if possible. If a speed restriction is imposed, trains can still run but at a reduced speed; this implies delays and sometimes journey cancellations. Therefore we assume that the number and duration of imposed speed restrictions implicitly provide an indication of the impact on service delay. Similarly, we assume that the number and duration of imposed section closure implicitly provide indication of the deleted services due to section unavailability. The problem is formulated as follows:

$$\min Z(X)=\sum_{i=1}^{n_R}\sum_{j=1}^{n_S} n_{ij}^{(SR)} \cdot d_{ij}^{(SR)} \cdot f_i \cdot x_{ij} \quad s.t \tag{1}$$

$$\sum_{j=1}^{n_S} x_{ij}=1 \quad \forall\ i=1,2,\ldots n_R, \tag{2}$$

$$\sum_{i=1}^{n_R}\sum_{j=1}^{n_S} c_{ij} x_{ij} \le B, \tag{3}$$

$$Q_{L_l}(x_{ij}) \le Q_{L_l}^* \quad \forall L_l \in L, \tag{4}$$

$$x_{ij} \in \{0,1\} \quad \forall\ i=1,2,\ldots n_R; \quad j=1,2,\ldots n_S. \tag{5}$$

where the model parameters are:

- $n_{ij}^{(LC)}$ the average number of closures in SRS i following implementation of strategy j,
- $d_{ij}^{(LC)}$ the average duration of closures in SRS i following implementation of strategy j,
- $n_{ij}^{(SR)}$ the average number of speed restriction imposed on SRS i following implementation of strategy j,
- $d_{ij}^{(SR)}$ the average duration of speed restriction imposed on SRS i following implementation of strategy j,
- $q_{ij}^{(LC)}$ the probability of a closure in SRS i following implementation of strategy j,
- $Q_{L_l}^*$ the threshold on the unavailability of line L_l,
- c_{ij} the cost of strategy j implemented on SRS i,
- f_i the frequency of trains travelling on SRS i,
- B the available budget.

The objective function $Z(X)$ is representative of the impact that the selected combination of strategies has on service delay, which allows to compare different solutions. It represent the expected number of trains affected by a service disruption during the considered time horizon. Each term $n_{ij}^{(SR)} \cdot d_{ij}^{(SR)} \cdot f_i \cdot x_{ij}$ gives an indication of the contribution of each SRS to the overall service disruption. This contribution is proportional to the average number of speed restrictions imposed on the SRS and its average duration, and on the frequency of trains travelling through the SRS. The train frequency is used to weight each SRS proportionally to the normalised amount of flow travelling on the SRS. The frequency also implicitly weight each SRS based on the its centrality, namely its role in serving more than one line. The set of constraints 2 indicates that only one strategy can be selected for each link. Constraint 3 adds a bound on the overall costs according to the available budget. The set of constraints 4 put a threshold on the minimum value of unavailability of each line. A line is unavailable if any of its SRSs is closed. Therefore, the probability of line L_l being closed $Q_{L_l}(x_{ij})$ can be written as

$$Q_{L_l}(X)=1-\prod_{\forall i | R_i \in L_l}\left(1-\sum_{\forall j | S_j \in S} q_{ij}^{(LC)} \cdot x_{ij}\right) \tag{6}$$

The optimal solution X^* is given by the feasible combination of strategies that will provide the minimum impact on service as represented by the objective function $Z^*(X)$. The objective function in 1 $Z(X)$ is linear in X, as constraints (1) and (2), while constraints (3) are non-linear. Problem 1 is therefore a non-linear integer otpimisation problem.

2.3.1 *Solution method*

There are no general-purpose solution methods yielding the global optimum for non-linear (non-convex) constrained optimisation problems and

approximate solution algorithms are usually used. However, it possible to solve a linear approximation of the original problem if the non-linear functions (objective function and/or constraints) can be converted to an acceptable linear form.

Problem 1 is transformed into a linear integer programming model by replacing the left hand side of constraint 4 with its *rare event approximation* (Andrews & Moss 2002) as follows:

$$Q_{L_l}(X) = 1 - \prod_{\forall i | R_i \in L_l} \left(1 - \sum_{\forall j | S_j \in S} q_{ij}^{(LC)} \cdot x_{ij} \right) \leq \sum_{\forall i | R_i \in L_l} \sum_{\forall j | S_j \in S} q_{ij}^{(LC)} \cdot x_{ij}. \quad (7)$$

The rare event approximation is an upper bound to the top event exact probability and can be used when the probability of the basic events is low. This an acceptable approximation for the problem at hand as the probability of a link closure is usually small.

Integer programming is NP-hard, namely it can be solved in non-polynomial time. Therefore, depending on the problem size it can be difficult to solve in reasonable computational time. In such circumstances, the associated relaxed problem obtained through Continuous relaxation can be studied. The relaxed problem is a linear continuous programming model which can be solved by means of the simplex method. The optimal solution of the relaxed problem is a lower bound of the global optimum of the original problem.

3 NUMERICAL EXAMPLE

The optimisation approach presented in this paper has been applied to select the best combination of maintenance strategies for a set of SRSs comprising one of the UK Strategic Routes, the East Midlands (EM) Route. Details of the EM route and it SRSs can be found in (NetworkRail 2015). A schematic representation of part of the EM route showing seven of its eleven SRSs is given in Figure 4. The set of SRSs considered in this example are listed in Table 1 along with the train frequency.

Railway services running along the EM Route which have been considered here are listed in Table 2 along with the service type (Long distance high speed—LDHS, interurban and local), while Table 3 lists the SRSs included within each service.

For each railway service, different availability requirements are considered depending on the type of service. Three potential maintenance strategies are considered, $S = \{S_1, S_2, S_3\}$. The evaluation of

Figure 4. Map of part of the EM Route, including SRSs 11.01 to 11.07.

Table 1. SRSs and trains frequency.

SRS	Train per hour
01 London St. Pancras-Bedford	20
02 Bedford-Nottingham	8
03 Wichnor Jn/Long Eaton-Chesterfield	8
04 Chesterfield-Nottingham	4
05 Nottingham-Newark Castle	1
06 Matlock-Ambergate	1
07 Netherfield-Grantham	2

Table 2. Railway services.

Service name	Service type
London St. Pancras to Nottingham	LDHS
London St. Pancras to Sheffield (via Derby)	LDHS
Norwich to Liverpool	Interurban
Nottingham to Leeds	Interurban
Newark Castle-Nottingham-Derby-Matlock	local

Table 3. SRSs included within each railway service.

Service name	SRSs
London St. Pancras to Nottingham	{01, 02}
London St. Pancras to Sheffield (via Derby)	{01, 02, 03}
Norwich to Liverpool	{02, 04, 07}
Nottingham to Leeds	{02, 04}
Newark Castle-Nottingham-Derby-Matlock	{02, 03, 05, 06}

the maintenance strategies through the PN asset models yields the input parameters to the optimisation model. The values of the model parameters used to run this numerical example are detailed in Table 4 where c_i, q_i and n_i^{SR} indicate the cost, unavailability and number of speed restriction due to the implementation of the available strategies.

The optimisation model has been solved for eight different values of the available budget B_1 = 350, B_2 = 400, B_3 = 450, B_4 = 500, B_5 = 550, B_6 = 600, B_7 = 650, B_8 = 700, while the thresholds on the unavailability of each railway service remain unchanged and equal to $Q_{L_1}^*$ = 0.98, $Q_{L_2}^*$ = 0.98, $Q_{L_3}^*$ = 0.98, $Q_{L_4}^*$ = 0.95, $Q_{L_5}^*$ = 0.9. The results of the scenario analysis are summarised in Table 5 and Figure 5. Table 5 details the optimal maintenance strategies for each SRS, while Figure 5 shows the corresponding value of the objective function which is indicative of the expected number of trains affected by a speed restriction.

Table 4. Model parameters.

SRS	c_1	c_2	c_3	q_1	q_2	q_3	n_1^{SR}	n_2^{SR}	n_3^{SR}
01	50	70	85	0.9	0.95	0.99	4.7	3.8	2.5
02	50	70	85	0.9	0.95	0.99	4.7	3.8	2.5
03	60	80	95	0.9	0.95	0.99	4.7	3.8	2.5
04	60	80	95	0.9	0.95	0.99	4.7	3.8	2.5
05	60	80	95	0.9	0.95	0.99	4.7	3.8	2.5
06	50	70	85	0.9	0.95	0.99	4.7	3.8	2.5
07	60	80	95	0.9	0.95	0.99	4.7	3.8	2.5

Table 5. Maintenance strategies selected.

	SRS						
Scenario	01	02	03	04	05	06	07
1	–	–	–	–	–	–	–
2	–	–	–	–	–	–	–
3	–	–	–	–	–	–	–
4	–	–	–	–	–	–	–
5	S_3	S_3	S_3	S_3	S_1	S_1	S_2
6	S_3	S_3	S_3	S_3	S_1	S_3	S_3
7	S_3	S_3	S_3	S_3	S_3	S_3	S_3
8	S_3	S_3	S_3	S_3	S_3	S_3	S_3

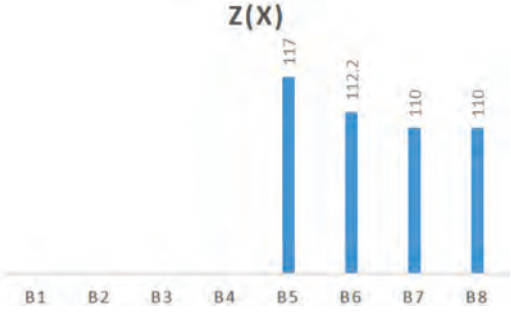


Figure 5. Expected number of disrupted trains for different available budgets.

Results show that no feasible solution can be found for budgets B_1 to B_4 as the strategies that would be achievable within the available budget do not ensure the required level of availability for each railway service. If the budget is increased solutions can be found. Budget B_5 is enough to find a feasible solution but does not allow the selection of the best strategy available for each SRSs. The algorithm selects less expensive strategies for SRSs 05, 06 and 07 as they belong to those railway services for which a less restrictive value of availability is required. Furthermore, the train frequency on those sections is lower than in the others. By further increasing the available budget, better strategies can be chosen and the impact on service decreases.

4 CONCLUSIONS

This paper presents a modelling approach to support decisions on how to effectively maintain a railway infrastructure system. First, for each railway asset, a modelling tool is required to assess the asset response to the implementation of a range of feasible maintenance strategies. Such modelling tools, called *asset state models* combine the degradation/failure processes affecting the asset with the intervention activities that can be performed in order to predict the future asset state. The modelling approach suggested to develop the asset state models is the PN method. A modelling template based on the PN method have been presented, which can be specified to represent a variety of railway assets undergoing degradation and ageing. The asset state models developed for each asset inform a network-level optimisation model aimed at selecting the best combination of maintenance strategies to manage all the assets on a railway network under budget and performance constraints. The network-level optimisation model is formulated as an integer program with multiple constraints. A numerical example has been presented to show the capabilities of the optimisation model. An advantage of mathematical programming formulation is that the model is not a black box. Furthermore, when the problem size is such that global solutions cannot be found in reasonable computational time, the mathematical programming formulation allows the use of tools to

estimate the goodness of approximate solutions. By varying the model parameters, scenario analysis can be performed so that the infrastructure manager is provided with a range of solutions for different combination of budget available and performance targets.

ACKNOWLEDGMENTS

John Andrews is the Royal Academy of Engineering and Network Rail Professor of Infrastructure Asset Management. He is also Director of The Lloyds Register Foundation Resilience Engineering Research Group at the University of Nottingham. Claudia Fecarotti is a Research Associate supported by Network Rail. They both gratefully acknowledge the support of these organisations.

REFERENCES

Andrews, J. (2012). A modelling approach to railway track asset management. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 227*(1), 56–73.

Andrews, J., D. Prescott, & F. De Rozieres (2014). A stochastic model for railway track asset management. *Reliability Enginering and System Safety 130*, 76–84.

Andrews, J.D. & T.R. Moss (2002). *REliability and Risk Assessment*. Professional Engineering Publishing.

Durango-Cohen, P. (2007). A time series analysis framework for transportation infrastructure management. *Transportation Research Part B 41*, 493–505.

Furuya, A. & S. Madanat (2013). Accounting for network for effects on railway asset management. *Journal of Transportation Engineering 139*, 92–100.

Hillier, F. & J. Lieberman (2009). *Introduction to Operations Research*. McGrow-Hill Higher Education.

Le, B. & J. Andrews (2016). Petri-net modelling of bridge asset management using maintenance-related conditions. *Structure and Infrastructure Engineering 12*(6), 730–751.

Le, B., J. Andrews, & C. Fecarotti (2017). A petri net model for railway bridge maintenance. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 231*(3), 306–323.

Murata, T. (1989). Petri nets: properties, analysis and applications. *Proceedings of the IEEE 77*(4), 541–580.

NetworkRail (2015). Route specifications: London north eastern and east midlands. Technical report.

Prescott, D. & J. Andrews (2013). A track ballast maintenance and inspection model for a rail network. *Proceeding of the Institution of Mechanical Engineers, Part O: J Risk and Reliability 227*(3), 251–266.

Robelin, C.A. & S.M. Madanat (2008). reliability-based system-level optimization of bridge maintenance and replacement decisions. *TRANSPORTATION SCIENCE*, 1–6.

Yeo, H., Y. Yoon, & S. Madanat (2012). Algorithms for bottom up maintenance optimization for heterogeneous infrastructure systems. *Structure and Infrastructure Engineering 1*, 1–12.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Operation and climate-weather change impact on maritime ferry safety

K. Kołowrocki & E. Kuligowska
Gdynia Maritime University, Poland

ABSTRACT: The paper is concerned with an application of the recently developed, a general safety analytical model of a critical infrastructure under the influence of an operation process related to climate-weather change process. The model is presented and applied to the prediction of the maritime ferry safety characteristics. As a result of this application, the ferry unconditional safety function and the risk function at changing in time operation and climate-weather conditions are determined. Moreover, the other significant safety indicators, i.e. the mean lifetime up to the exceeding a critical safety state, the moment when the risk function value exceeds the acceptable safety level, the intensities of ageing and the coefficients of the operation and climate-weather impact on the ferry intensities of ageing are presented.

1 INTRODUCTION

The paper presents the operation and climate-weather change influence on the safety of a critical infrastructure. The maritime ferry operation process is described in (Kołowrocki & Soszyńska-Budny 2011), whether the climate-weather change process for the ferry operating area is modeled in (Kuligowska 2017). The identification of the ferry operation process related to climate-weather change is performed in (Kołowrocki et al. 2017a). Having these processes identified, the safety prediction of the considered ferry under the operation process and climate-weather change influence is performed.

An analytical safety model of a complex technical system under the influence of the operation process related to climate-weather change process is proposed (Kołowrocki et al. 2017b). It is the integrated model of complex technical system safety, linking its multistate safety model and the model of its operation process related to climate-weather change process at its operating area, considering variable at the different climate-weather states impacted by them the system safety structures and its components safety parameters.

The maritime ferry safety characteristics, i.e. the unconditional safety function and the risk function at changing in time operation and climate-weather conditions (in February) are determined. Moreover, the safety and resilience indicators are presented: the mean lifetime up to the exceeding a critical safety state, the moment when the risk function value exceeds the acceptable safety level, the intensities of ageing and the coefficients of the operation and climate-weather impact on the maritime ferry safety.

2 CRITICAL INFRASTRUCTURE OPERATION PROCESS RELATED TO CLIMATE-WEATHER VARIABLE CONDITIONS

2.1 *Critical infrastructure operation process*

We assume that the critical infrastructure during its operation process is taking ν, $\nu \in N$, different operation states $z_1, z_2, \ldots, z_\nu$. Moreover, we assume that the critical infrastructure operation process $Z(t)$ is a semi-Markov process with the conditional sojourn times θ_{bl} at the operation states z_b when its next operation state is z_l, $b,l = 1,2, \ldots, \nu$, $b \neq l$.

Under these assumptions, the critical infrastructure operation process may be described by (Kołowrocki & Soszyńska-Budny 2011):

- the vector $[p_b(0)]_{1\times\nu}$ of the initial probabilities $p_b(0) = P(Z(0) = z_b)$, $b = 1,2, \ldots,\nu$, of the critical infrastructure operation process $Z(t)$ staying at particular operation states at the moment $t = 0$;
- the matrix $[p_{bl}]_{\nu\times\nu}$ of probabilities p_{bl}, b, $l = 1,2, \ldots,\nu$, $b \neq l$, of the critical infrastructure operation process $Z(t)$ transitions between the operation states z_b and z_l;
- the matrix $[H_{bl}(t)]_{\nu\times\nu}$ of conditional distribution functions $H_{bl}(t) = P(\theta_{bl} < t)$, $t \in <0,+\infty)$, $b,l = 1,2,\ldots,\nu$, $b \neq l$, of the critical infrastructure operation process $Z(t)$ conditional sojourn times θ_{bl} at the operation states.

The limit values of the critical infrastructure operation process $Z(t)$ transient probabilities $p_b(t) = P(Z(t) = z_b)$, $t \in <0,+\infty)$, $b = 1,2,\ldots,\nu$, at the particular operation states, can be found using the procedure given in (Kołowrocki & Soszyńska-Budny 2011).

2.2 *Climate-weather change process at the critical infrastructure operating area*

To model the climate-weather change process for the critical infrastructure operating area we assume that the process is taking w, $w \in N$, different climate-weather states c_1, c_2,…, c_w. Further, we define the climate-weather change process $C(t)$, $t \in <0,+\infty)$, with discrete operation states from the set $\{c_1, c_2,\ldots, c_w\}$. Assuming that the climate-weather change process $C(t)$ is a semi-Markov process it can be described by (Kołowrocki, Soszyńska-Budny & Torbicki 2017a,b):

- the vector $[q_b(0)]_{1\times w}$ of the initial probabilities $q_b(0) = P(C(0) = c_b)$, $b = 1,2,\ldots, w$, of the climate-weather change process $C(t)$ staying at particular climate-weather states c_b at the moment $t = 0$;
- the matrix $[q_{bl}]_{w\times w}$ of the probabilities of transitions q_{bl}, $b,l = 1,2,\ldots, w$, $b \neq l$, of the climate-weather change process $C(t)$ from the climate-weather states c_b to c_l;
- the matrix $[C_{bl}(t)]_{w\times w}$ of the conditional distribution functions $C_{bl}(t) = P(C_{bl} < t)$, $t \in <0,+\infty)$, $b,l = 1,2,\ldots, w$, of the conditional sojourn times C_{bl} at the climate-weather states c_b when its next climate-weather state is c_l, $b,l = 1,2,\ldots, w$, $b \neq l$.

The limit values of the climate-weather change process $C(t)$ transient probabilities $q_b(t) = P(C(t) = c_b)$, $t \in <0,+\infty)$, $b = 1,2,\ldots,w$, at the particular climate-weather states, can be found using the procedure given in (Kołowrocki, Soszyńska-Budny & Torbicki 2017a).

2.3 *Critical infrastructure operation process related to climate-weather change*

We assume as in (Kołowrocki et al. 2017c), that the critical infrastructure operation process related to climate-weather change $ZC(t)$, $t \in <0,+\infty)$, can take νw, ν, $w \in N$, different operation states $zc_{11}, zc_{12},\ldots, zc_{\nu w}$, described by:

- the vector $[pq_{bl}(0)]_{1\times \nu w}$ of initial probabilities of the critical infrastructure operation process related to climate-weather change $ZC(t)$ staying at the initial moment $t = 0$ at the operation and climate-weather states zc_{bl}, $b = 1,2,\ldots,\nu$, $l = 1,2,\ldots,w$;
- the matrix $[pq_{bl\,\tilde{b}\tilde{l}}]_{\nu w\times \nu w}$ of the probabilities of transitions of the critical infrastructure operation process related to climate-weather change $ZC(t)$ between the operation states zc_{bl} and $zc_{\tilde{b}\tilde{l}}$, $b = 1,2,\ldots,\nu$, $l = 1,2,\ldots,w$, $\tilde{b} = 1,2,\ldots,\nu$, $\tilde{l} = 1,2,\ldots,w$;
- the matrix $[HC_{bl\,\tilde{b}\tilde{l}}(t)]_{\nu w\times \nu w}$ of the conditional distribution functions of the critical infrastructure operation process related to climate-weather change $ZC(t)$ conditional sojourn times $\theta C_{bl\,\tilde{b}\tilde{l}}$, $b = 1,2,\ldots,\nu$, $l = 1,2,\ldots,w$, $\tilde{b} = 1,2,\ldots,\nu$, $\tilde{l} = 1,2,\ldots,w$, at the operation state zc_{bl}, $b = 1,2,\ldots,\nu$, $l = 1,2,\ldots,w$, when the next operation state is $zc_{\tilde{b}\tilde{l}}$, $\tilde{b} = 1,2,\ldots,\nu$, $\tilde{l} = 1,2,\ldots,w$.

Assuming that we have identified the unknown parameters of the critical infrastructure operation process related to climate-weather change $ZC(t)$, we can predict this process basic characteristics, e.g. the limit transient probabilities $pq_{bl}(t) = P(ZC(t) = zc_{bl})$, $t \in < 0, +\infty)$, $b = 1,2,\ldots,\nu$, $l = 1,2, \ldots, w$, at the particular states, according to the procedure given in (Kołowrocki et al. 2017c).

3 CRITICAL INFRASTRUCTURE SAFETY AT VARIABLE OPERATION CONDITIONS RELATED TO CLIMATE-WEATHER CHANGE

In the safety analysis of critical infrastructures at the variable operation conditions related to climate-weather change, to define the critical infrastructure with degrading components we assume that the changes of the process $ZC(t)$ states have an impact on the critical infrastructure's components and its structure (Kołowrocki 2014, Kołowrocki et al. 2017b). We denote the critical infrastructure asset A_i, $i = 1,2,\ldots,n$, conditional lifetime in the safety state subset $\{u, u + 1, \ldots, z\}$, while the operation process related to climate-weather change $ZC(t)$ is at the state zc_{bl}, $b = 1, 2, \ldots, \nu$, $l = 1,2, \ldots, w$, by $[T_i^4(u)]^{(bl)}$. Moreover, in this section we assume that the critical infrastructure assets at particular states have the exponential safety functions. According to (Kuligowska & Soszyńska-Budny 2017b), the conditional critical infrastructure safety function is defined by the vector

$$[S_i^4(t,\cdot)]^{(bl)} = [1,[S_i^4(t,1)]^{(bl)},\ldots,S_i^4(t,z)]^{(bl)}],t \in <0,+\infty),$$
$$b = 1,2,\ldots,\nu,\ l = 1,2,\ldots,w,\ i = 1,2,\ldots,n, \quad (1)$$

with the coordinates

$$[S_i^4(t,u)]^{(bl)} = P([T_i^4(u)]^{(bl)} > t \mid ZC(t) = zc_{bl}) = \exp[-[\lambda_i^4(u)]^{(bl)}t],\ t \in < 0, +\infty),$$
$$b = 1,2,\ldots,\nu,\ l = 1,2, \ldots, w,\ i = 1,2, \ldots, n, \quad (2)$$

where the intensities of ageing of the critical infrastructure assets related to operation and climate-weather impact, existing in (2), are given by

$$[\lambda_i^4(u)]^{(bl)} = [\rho_i^4(u)]^{(bl)} \cdot \lambda_i^0(u),\ u = 1, 2,\ldots, z,$$
$$b = 1,2,\ldots,\nu,\ l = 1,2, \ldots, w,\ i = 1,2,\ldots, n, \quad (3)$$

and $\lambda_i^0(u)$ are the intensities of ageing of the system components without operation and climate-weather impact and

$$[\rho_i^4(u)]^{(bl)},\ u = 1,2,\ldots,z,\ b = 1,2, \ldots, \nu,\ l = 1,2, \ldots,$$
$$w,\ i = 1,2,\ldots,n, \quad (4)$$

are the coefficients of operation and climate-weather change impact on the critical infrastructure assets' intensities of ageing without operation and climate-weather change impact.

Further, we denote the critical infrastructure unconditional lifetime in the safety state subset $\{u,u+1,\ldots,z\}$ by $T^4(u)$ and the system unconditional safety function by

$$\boldsymbol{S}^4(t,\cdot) = [1, \boldsymbol{S}^4(t,1), \ldots, \boldsymbol{S}^4(t,z)], \tag{5}$$

with the vector's coordinates defined by

$$\boldsymbol{S}^4(t,u) = P(T^4(u) > t),\ t \in <0,+\infty),\quad u = 1,2,\ldots,z. \tag{6}$$

In the case when the critical infrastructure operation time θC_{bl} is large enough, the coordinates of the unconditional safety function of the system defined by (5) are given by

$$\boldsymbol{S}^4(t,u) \cong \sum_{b=1}^{v}\sum_{l=1}^{w} pq_{bl}[\boldsymbol{S}^4(t,u)]^{(bl)},\ t \in <0,+\infty),\quad u = 1,2,\ldots,z, \tag{7}$$

where $[\boldsymbol{S}^4(t,u)]^{(bl)}$, $u = 1,2,\ldots,z$, $b = 1,2,\ldots,v$, $l = 1,2,\ldots,w$, are the coordinates of the system conditional safety function defined by (2)–(4) and pq_{bl}, $b = 1,2,\ldots,v$, $l = 1,2,\ldots,w$, are the system operation process limit transient probabilities (see section 2.3).

Further, we determine the mean values $\mu^4(u)$ and the standard deviations $\sigma^4(u)$ of the unconditional lifetimes of the critical infrastructure in the safety state subsets $\{u,u+1,\ldots,z\}$, $u = 1,2,\ldots,z$, the mean values $\overline{\mu}^4(u)$ of the unconditional lifetimes of the critical infrastructure in the particular safety states u, $u = 1,2,\ldots,z$, the risk function $\boldsymbol{r}^4(t)$ and the moment τ^4 when the critical infrastructure risk function exceeds a permitted level δ, after substituting for $\boldsymbol{S}^4(t,u)$, $u = 1,2,\ldots,z$, the coordinates of the unconditional safety functions given by (6).

4 MARITIME FERRY OPERATION PROCESS RELATED TO CLIMATE-WEATHER CHANGE

The maritime ferry technical system consists of a navigational subsystem S_1, a propulsion and controlling subsystem S_2, a loading and unloading subsystem S_3, a stability control subsystem S_4 and an anchoring and mooring subsystem S_5, which form a series structure. The detailed scheme and system description may be found in (Kołowrocki & Soszyńska-Budny 2011). The maritime ferry safety structure and the assets' safety depend on its changing in time operation and climate-weather states.

Taking into account the expert opinions and according to section 2.3 and (Kołowrocki et al. 2017a), the maritime ferry operation process related to climate-weather change process $ZC(t)$, $t \in <0,+\infty)$, can take

$$v \cdot w = 18 \cdot 6 = 108, \tag{8}$$

different operation states

$$\begin{aligned}&zc_{1\,1}, zc_{2\,1}, \ldots, zc_{18\,1};\\ &zc_{1\,2}, zc_{2\,2}, \ldots, zc_{18\,2};\\ &\ldots,\\ &zc_{1\,18}, zc_{2\,18}, \ldots, zc_{18\,18}.\end{aligned} \tag{9}$$

Considering the results of the identification of the unknown parameters of the maritime ferry operation process related to climate-weather change (Kołowrocki et al. 2017), it was possible to predict this process' basic characteristics. The limit values of the maritime ferry operation process related to climate-weather change process $ZC(t)$ transient probabilities pq_{bl}, $b = 1,2,\ldots,18$, $l = 1,2,\ldots,6$, at the particular states zc_{bl}, are given in the vector

$$\begin{aligned}[pq_{bl}]_{1\times 108} \cong [&0.015352, 0.000418, 0.017138, 0.00038, 0.004712, 0;\\ &0.000808, 0.000022, 0.000902, 0.00002, 0.000248, 0;\\ &0.02093, 0.004342, 0.000208, 0, 0.000182, 0.000338;\\ &0.02898, 0.006012, 0.000288, 0, 0.000252, 0.000468;\\ &0.287496, 0.067155, 0, 0, 0.005808, 0.002541;\\ &0.010244, 0.000416, 0.008086, 0.00078, 0.006474, 0;\\ &0.00197, 0.00008, 0.001555, 0.00015, 0.001245, 0;\\ &0.006304, 0.000256, 0.004976, 0.00048, 0.003984, 0;\\ &0.014578, 0.000592, 0.011507, 0.00111, 0.009213, 0;\\ &0.000788, 0.000032, 0.000622, 0.00006, 0.000498, 0;\\ &0.001182, 0.000048, 0.000933, 0.00009, 0.000747, 0;\\ &0.006304, 0.000256, 0.004976, 0.00048, 0.003984, 0;\\ &0.277992, 0.064935, 0, 0, 0.005616, 0.002457;\\ &0.02737, 0.005678, 0.000272, 0, 0.000238, 0.000442;\\ &0.01932, 0.004008, 0.000192, 0, 0.000168, 0.000312;\\ &0.001212, 0.000033, 0.001353, 0.00003, 0.000372, 0;\\ &0.00202, 0.000055, 0.002255, 0.00005, 0.00062, 0;\\ &0.005252, 0.000143, 0.005863, 0.00013, 0.001612, 0].\end{aligned} \tag{10}$$

5 MARITIME FERRY SAFETY PREDICTION INCLUDING OPERATION AND CLIMATE-WEATHER CHANGE IMPACT

5.1 *Maritime ferry safety parameters*

After discussion with experts, taking into account the safety of the operation of the ferry, we fix 5 ($z = 5$) safety states of the ferry technical system and we distinguish the following safety states:

- a safety state 4 – the ferry operation is fully safe;
- a safety state 3 – the ferry operation is less safe and more dangerous because of the possibility of environment pollution;

– a safety state 2 – the ferry operation is less safe and more dangerous because of the possibility of environment pollution and causing small accidents;
– a safety state 1 – the ferry operation is much less safe and much more dangerous because of the possibility of serious environment pollution and causing extensive accidents;
– a safety state 0 – the ferry technical system is destroyed.

Moreover, by the expert opinions, we assume that there are possible the transitions between the components' safety states only from better to worse ones.

Considering the assumptions and agreements from the previous sections, we assume that the components $E_{ij}^{(\upsilon)}$, $i = 1,2,\ldots,k$, $j = 1,2,\ldots,l_i$, of the subsystem S_υ, $\upsilon = 1,2,3,4,5$, at the system states z_{bl}, $b = 1,2,\ldots,18$, $l = 1,2,\ldots,6$, have the exponential safety functions, i.e. the coordinates of the vector

$$[S_{ij}^{4(\upsilon)}(t,\cdot)]^{(bl)} = [1,[S_{ij}^{4(\upsilon)}(t,1)]^{(bl)},\ldots,[S_{ij}^{4(\upsilon)}(t,4)]^{(bl)}],$$
$$t \in <0,+\infty), i = 1,2,\ldots,k, j = 1,2,\ldots,l_i, \upsilon = 1,2,3,4,5,$$
$$b = 1,2,\ldots,18, l = 1,2,\ldots,6, \tag{11}$$

are given by

$$[S_{ij}^{4(\upsilon)}(t,u)]^{(bl)} = P([T_{ij}^{4(\upsilon)}]^{(bl)}(u) > t \,|\, ZC(t) = zc_{bl}) = \exp[-[\lambda_{ij}^{4(\upsilon)}(u)]^{(bl)} t],$$
$$t \in <0,+\infty), u = 1,2,3,4, i = 1,2,\ldots,k, j = 1,2,\ldots,l_i,$$
$$\upsilon = 1,2,3,4,5, b = 1,2,\ldots,18, l = 1,2,\ldots,6. \tag{12}$$

Existing in the above formula the intensities of ageing of the components $E_{ij}^{(\upsilon)}$, $i = 1,2,\ldots,k$, $j = 1,2,\ldots,l_i$, of the subsystem S_υ, $\upsilon = 1,2,3,4,5$, at the system operation process states zc_{bl}, $b = 1,2,\ldots,18$, $l = 1,2,\ldots,6$, i.e. the coordinates of the vector of intensities

$$[\lambda_{ij}^{4(\upsilon)}(\cdot)]^{(bl)} = [0,[\lambda_{ij}^{4(\upsilon)}(1)]^{(bl)},\ldots,[\lambda_{ij}^{4(\upsilon)}(4)]^{(bl)}],$$
$$i = 1,2,\ldots,k, j = 1,2,\ldots,l_i, \upsilon = 1,2,3,4,5, b = 1,2,\ldots,18,$$
$$l = 1,2,\ldots,6, \tag{13}$$

are given by

$$[\lambda_{ij}^{4(\upsilon)}(u)]^{(bl)} = [\rho_{ij}^{4(\upsilon)}(u)]^{(bl)} \lambda_{ij}^{0(\upsilon)}(u), u = 1,2,3,4,$$
$$i = 1,2, \quad ,k, j = 1,2, \quad ,l_i, \upsilon = 1,2,3,4,5, \tag{14}$$
$$b = 1,2, \quad ,18, l = 1,2, \quad ,6,$$

where $\lambda_{ij}^{0(\upsilon)}(u)$, $u = 1,2,3,4$, $i = 1,2,\ldots,k, j = 1,2,\ldots,l_i$, are the intensities of ageing of the components $E_{ij}^{(\upsilon)}$, $i = 1,2,\ldots,k$, $j = 1,2,\ldots,l_i$, of the subsystems S_υ, $\upsilon = 1,2,3,4,5$, without of any impact, i.e. the coordinates of the vector

$$[\lambda_{ij}^{(\upsilon)}(\cdot)] = \left[0, \lambda_{ij}^{(\upsilon)}(1), \lambda_{ij}^{(\upsilon)}(2), \lambda_{ij}^{(\upsilon)}(3), \lambda_{ij}^{(\upsilon)}(4)\right],$$
$$i = 1,2,\ldots,k, j = 1,2,\ldots,l_i, \upsilon = 1,2,3,4,5, \tag{15}$$

and

$$[\rho_{ij}^{4(\upsilon)}(u)]^{(bl)}, u = 1,2,3,4, i = 1,2, \quad ,k, j = 1,2, \quad ,l_i,$$
$$\upsilon = 1,2,3,4,5, b = 1,2, \quad ,18, l = 1,2, \quad ,6, \tag{16}$$

are the coefficients of the operation and climate-weather change impact on the components $E_{ij}^{(\upsilon)}$, $i = 1,2,\ldots,k$, $j = 1,2,\ldots,l_i$, of the subsystems S_υ, $\upsilon = 1,2,3,4,5$, intensities of ageing at the system states zc_{bl}, $b = 1,2,\ldots,18$, $l = 1,2,\ldots,6$, i.e. the coordinates of the vector

$$[\rho_{ij}^{4(\upsilon)}(\cdot)]^{(bl)} = \left[\ 0, [\rho_{ij}^{4(\upsilon)}(1)]^{(bl)}, [\rho_{ij}^{4(\upsilon)}(2)]^{(bl)}\ \right]$$
$$i = 1,2,\ldots,k, j = 1,2,\ldots,l_i, \upsilon = 1,2,3,4,5, \tag{17}$$
$$b = 1,2,\ldots,18, l = 1,2,\ldots,6.$$

According to expert opinions, changing the maritime ferry operation process states have influence on changing the system safety structures and its selected components' safety parameters as well. For this system, the intensities of components' departure from the safety states subsets {1,2,3,4}, {2,3,4}, {3,4}, {4} without of any impact are given in (Kołowrocki, Soszyńska-Budny & Torbicki 2017d), whereas the intensities of departure related to the operation process influence on ferry safety are given in (Kołowrocki et al. 2017c). The intensities of departure related to the climate-weather influence on the maritime ferry safety are calculated as a multiplication of the coefficients given in Table 1 in the Appendix and the intensities without of any impact. Thus, considering the above results, the new intensities of departure related to the operation and climate-weather influence on the maritime ferry safety are calculated according to formula (14), where the coefficients of the operation and climate-weather change impact on the components' intensities of ageing at the particular states are calculated as follows:

$$[\rho_{ij}^{4(\upsilon)}(u)]^{(bl)} = [\rho_{ij}^{1(\upsilon)}(u)]^{(bl)} [\rho_{ij}^{3(\upsilon)}(u)]^{(bl)}, i = 1,2, \quad ,k,$$
$$j = 1,2, \quad ,l_i, \upsilon = 1,2,3,4,5, u = 1,2,3,4, b = 1,2, \quad ,18,$$
$$l = 1,2, \quad ,6,$$

where $[\rho_{ij}^{\xi(\upsilon)}(u)]^{(bl)}$, $\xi = 1,3$, are respectively the coefficients of operation impact and climate-weather change impact on the components' intensities of ageing at the particular states.

5.2 *Maritime ferry safety characteristics*

In (Kołowrocki & Soszyńska-Budny 2011), it is fixed that the maritime ferry technical system safety structure and its subsystems and components safety depend on its changing in time opera-

tion states. The influence of the system states changing on the changes of the system safety structure and its components safety functions is given in (Kołowrocki et al. 2017b). Thus, in the case when the operation time is large enough, according to (7) the maritime ferry technical system unconditional safety function is given by the vector

$$\boldsymbol{S}^4(t,\cdot) = [1, \boldsymbol{S}^4(t,1), \boldsymbol{S}^4(t,2), \boldsymbol{S}^4(t,3), \boldsymbol{S}^4(t,4)], \quad t \in \langle 0,+\infty), \tag{18}$$

where according to (7), the vector coordinates are given respectively for $t \in \langle 0,+\infty)$, $u = 1,2,3,4$, by

$$\boldsymbol{S}^4(t,u) \cong \sum_{b=1}^{18} \sum_{l=1}^{6} pq_{bl} [\boldsymbol{S}^4(t,u)]^{(bl)}, \tag{19}$$

where $[\boldsymbol{S}^4(t,u)]^{(bl)}$, $u = 1,2,3,4$, $b = 1,2,\ldots,18$, $l = 1,2,\ldots,6$, are the coordinates of the system conditional safety functions defined by (2)-(4) and pq_{bl}, $b = 1,2,\ldots,18$, $l = 1,2,\ldots,6$, are the system operation process limit transient probabilities given by (10).

The graph of the five-state maritime ferry technical system safety function is presented in Figure 1.

Considering (19), the expected values and standard deviations, given in years, of the maritime ferry technical system lifetimes in the safety states subsets {1,2,3,4}, {2,3,4}, {3,4}, {4}, respectively are

$$\mu^4(1) \cong 5.785139,\ \mu^4(2) \cong 3.161191,\ \mu^4(3) \cong 2.342224,\ \mu^4(4) \cong 1.879662; \tag{20}$$

$$\sigma^4(1) \cong 5.56368,\ \sigma^4(2) \cong 3.077813,\ \sigma^4(3) \cong 2.28222,\ \sigma^4(4) \cong 1.830362. \tag{21}$$

Consequently, the mean values of the maritime ferry technical system lifetimes in the particular safety states 1, 2, 3, 4, respectively are:

$$\overline{\mu}^4(1) = 2.623948,\quad \overline{\mu}^4(2) = 0.8189667,\quad \overline{\mu}^4(3) = 0.4625623,\quad \overline{\mu}^4(4) = 1.879662. \tag{22}$$

By (20) and (21), the mean and the standard deviation of the maritime ferry lifetime up to exceeding critical safety state $r = 2$ are

$\mu^4(2) = 3.161191$ years, $\sigma^4(2) = 3.077813$ years.

Since the critical safety state is $r = 2$, then the system risk function of the maritime ferry technical system, is given by

$$\boldsymbol{r}^4(t) = 1 - \boldsymbol{S}^4(t,2), \tag{23}$$

where $\boldsymbol{S}^4(t,2)$ is given by (19) and illustrated in Figure 1.

Hence, considering (23), the moment when the system risk function exceeds a permitted level, for instance $\delta = 0.05$, is given as follows

$$\tau^4 = \boldsymbol{r}^{4-1}(\delta) \cong 0.17 \text{ year}. \tag{24}$$

The maritime ferry intensities of ageing according to (Kołowrocki et al. 2017b) and considering (19) are:

$$\boldsymbol{\lambda}^4(t,u) = -\frac{d(\boldsymbol{S}^4(t,u))}{dt} \cdot \frac{1}{\boldsymbol{S}^4(t,u)}, \tag{25}$$

where particularly

$$\boldsymbol{\lambda}^4(t,1) \cong 0.216514,\ \boldsymbol{\lambda}^4(t,2) \cong 0.3860834,\ \boldsymbol{\lambda}^4(t,2) \cong 0.516497,\ \boldsymbol{\lambda}^4(t,3) \cong 0.6449959.$$

The graphs of the intensities of ageing for the maritime ferry are shown in Figure 3.

Considering (20) and applying (57) from (Kołowrocki et al. 2017d), the coefficients of the operation and climate-weather impact on the maritime ferry safety are

$$\boldsymbol{\rho}^4(t,1) \cong \frac{1/\mu^4(1)}{1/\mu^0(1)} \cong \frac{1/5.785139}{1/6.246} \cong 1.079663,$$
$$\boldsymbol{\rho}^4(t,2) \cong \frac{1/\mu^4(2)}{1/\mu^0(2)} \cong \frac{1/3.161191}{1/3.390} \cong 1.072381,$$
$$\boldsymbol{\rho}^4(t,3) \cong \frac{1/\mu^4(3)}{1/\mu^0(3)} \cong \frac{1/2.342224}{1/2.503} \cong 1.068642,$$
$$\boldsymbol{\rho}^4(t,4) \cong \frac{1/\mu^4(4)}{1/\mu^0(4)} \cong \frac{1/1.879662}{1/2.007} \cong 1.067745. \tag{26}$$

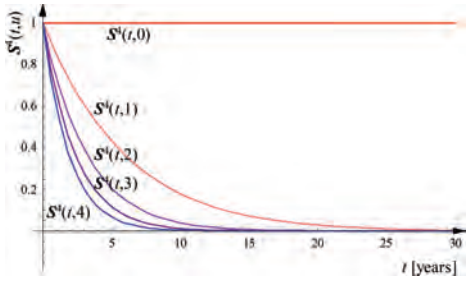


Figure 1. The graph of the maritime ferry safety function coordinates.

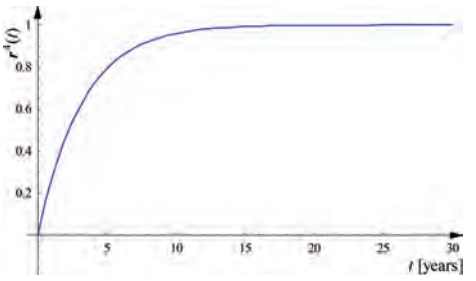


Figure 2. The graph (the fragility curve) of the maritime ferry risk function.

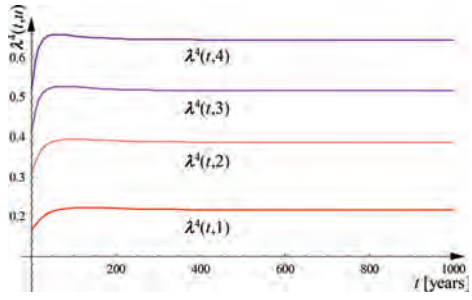


Figure 3. The graph of the intensities of ageing of the maritime ferry.

The resilience indicator, i.e. the coefficient of maritime ferry resilience to operation process and climate-weather change process impact is

$$\boldsymbol{RI}^4(t) = \frac{1}{\boldsymbol{\rho}^4(t,2)} \cong 0.9325047 \cong 93.25\%. \qquad (27)$$

6 CONCLUSIONS

The simplified impact model of critical infrastructure safety related to operation and climate-weather change impact was applied to the safety and risk evaluation for the maritime ferry operating at Baltic Sea waters. The predicted maritime ferry safety characteristics are different from those determined for this system operating at constant conditions without considering operation and climate-weather influence.

ACKNOWLEDGMENTS

The paper presents the results developed in the scope of the EU-CIRCLE project titled "A pan—European framework for strengthening Critical Infrastructure resilience to climate change" that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 653824, *http://www.eu-circle.eu/*.

REFERENCES

Kołowrocki, K., 2014. Reliability of Large and Complex Systems, Elsevier.

Kołowrocki, K., Kuligowska, E. and Soszyńska-Budny, J., 2015. Reliability of maritime ferry technical system, analytical assessment. Proc. European Safety and Reliability Conference - ESREL 2015, Zurich, Switzerland.

Kołowrocki, K., Kuligowska, E. and Soszyńska-Budny, J., 2017a. Identification and prediction of maritime ferry operation process related to climate-weather change. Journal of Polish Safety and Reliability Association, Summer Safety and Reliability Seminars, 8(2), 135–144.

Kołowrocki, K., Kuligowska, E. and Soszyńska-Budny, J., 2017b. Integrated Impact Model on Critical Infrastructure Safety Related to Its Operation Process Including Operating Environment Threats. Journal of Polish Safety and Reliability Association, Summer Safety and Reliability Seminars, 8(4), 11–20.

Kołowrocki, K., Kuligowska, E. and Soszyńska-Budny, J., 2017c. Safety of maritime ferry related to its operation process, Proceedings of European Safety and Reliability Conference - ESREL 2017, Portoroz, Slovenia.

Kołowrocki, K., Kuligowska, E., Soszyńska-Budny, J. and Torbicki, M., 2017. Critical Infrastructure Operation Process and Climate-Weather Change Process Data Processing – Identification and Prediction. Report for D6.4 Case Study 2 PL: Conduction.

Kołowrocki, K. and Soszyńska-Budny, J., 2011. Reliability and Safety of Complex Technical Systems and Processes: Modeling - Identification - Prediction - Optimization, Springer.

Kołowrocki, K., Soszyńska-Budny, J. and Torbicki, M., 2017a. Critical infrastructure operating area climate-weather change process including extreme weather hazards. Journal of Polish Safety and Reliability Association, Summer Safety and Reliability Seminars 8(2), 15–24.

Kołowrocki, K., Soszyńska-Budny, J. and Torbicki, M., 2017b. Identification methods and procedures of climate-weather change process including extreme weather hazards. Journal of Polish Safety and Reliability Association, Summer Safety and Reliability Seminars, (8)2, 85–106.

Kołowrocki, K., Soszyńska-Budny, J. and Torbicki, M., 2017c. Integrated Impact Model on Critical Infrastructure Safety Related to Its Operation Process and Climate-Weather Change Process. Journal of Polish Safety and Reliability Association, Summer Safety and Reliability Seminars, (8)4, 33–48.

Kołowrocki, K., Soszyńska-Budny, J. and Torbicki, M., 2017d. Safety of maritime ferry operating at Baltic sea open waters related to climate-weather change process including extreme weather hazards, Proceedings of European Safety and Reliability Conference - ESREL 2017, Portoroz, Slovenia.

Kuligowska, E., 2017. Identification and prediction of climate-weather change process for maritime ferry operating area. Journal of Polish Safety and Reliability Association, Summer Safety and Reliability Seminars, 8(2), 129–134.

Soszyńska-Budny, J., 2013. Modeling Safety of Multi-state Sys-tems with Application to Maritime Ferry Technical System, Reliability: Theory and Applications, 8(3): 24–39.

Xue, J. and Yang, K., 1995. Dynamic reliability analysis of coherent multi-state systems, IEEE Trans on Reliab. 4(44): 683–688.

APPENDIX

The coefficients $[\rho_{ij}^{4(\upsilon)}(u)]^{(bl)}$, $u = 1,2,3,4$, $i = 1,2,\ldots,k$, $j = 1,2,\ldots,l_i$, $\upsilon = 1,2,3,4,5$, $b = 1,2,\ldots,18$, $l = 1,2,\ldots,6$, of the operation process related to the climate-weather change process impact on the maritime ferry intensities of degradation are given in the table below.

Table 1. Coefficients of operation and climate-weather change impact on the maritime ferry intensities of degradation.

States	Oper. area	CW-CP	b	l	S_1	S_2							S_3					S_4		S_5		
					$E_{1,1}$	$E_{1,1\text{-}4}$	$E_{2,1\text{-}2}$	$E_{3,1}$	$E_{4,1}$	$E_{5,1}$	$E_{6,1}$	$E_{7,1}$	$E_{3,1}$	$E_{3,2}$	$E_{3,3}$	$E_{4,1}$	$E_{4,2}$	$E_{5,1}$	$E_{5,2}$	$E_{5,3}$
1	GDY	$C^{(1)}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	GDY	$C^{(1)}$	1	2	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
3	GDY	$C^{(1)}$	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	GDY	$C^{(1)}$	1	4	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
5	GDY	$C^{(1)}$	1	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	GDY	$C^{(1)}$	1	6	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
7	GDY	$C^{(1)}$	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	GDY	$C^{(1)}$	2	2	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
9	GDY	$C^{(1)}$	2	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	GDY	$C^{(1)}$	2	4	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
11	GDY	$C^{(1)}$	2	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	GDY	$C^{(1)}$	2	6	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
13	res	$C^{(2)}$	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	res	$C^{(2)}$	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	res	$C^{(2)}$	3	3	1.05	1.1	1	1	1.15	1.15	1.1	1.1	1.15	1.15	1.15	1	1.1	1	1	1
16	res	$C^{(2)}$	3	4	1.1	1.05	1	1	1.05	1.05	1.02	1.02	1.05	1.05	1.05	1	1.1	1	1	1
17	res	$C^{(2)}$	3	5	1.1	1.15	1	1	1.15	1.15	1.05	1.05	1.1	1.1	1.1	1	1.15	1	1	1
18	res	$C^{(2)}$	3	6	1.1	1.2	1	1	1.2	1.2	1.1	1.1	1.15	1.15	1.15	1	1.15	1	1	1
19	res	$C^{(2)}$	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20	res	$C^{(2)}$	4	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
21	res	$C^{(2)}$	4	3	1.05	1.1	1	1	1.15	1.15	1.1	1.1	1.15	1.15	1.15	1	1.1	1	1	1
22	res	$C^{(2)}$	4	4	1.1	1.05	1	1	1.05	1.05	1.02	1.02	1.05	1.05	1.05	1	1.1	1	1	1
23	res	$C^{(2)}$	4	5	1.1	1.15	1	1	1.15	1.15	1.05	1.05	1.1	1.1	1.1	1	1.15	1	1	1
24	res	$C^{(2)}$	4	6	1.1	1.2	1	1	1.2	1.2	1.1	1.1	1.15	1.15	1.15	1	1.15	1	1	1
25	open	$C^{(3)}$	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
26	open	$C^{(3)}$	5	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
27	open	$C^{(3)}$	5	3	1.05	1.4	1	1	1.4	1.4	1.4	1.4	1.3	1.3	1.3	1	1.25	1	1	1
28	open	$C^{(3)}$	5	4	1.15	1.1	1	1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1	1.1	1	1	1
29	open	$C^{(3)}$	5	5	1.15	1.3	1	1	1.3	1.3	1.3	1.3	1.2	1.2	1.2	1	1.2	1	1	1
30	open	$C^{(3)}$	5	6	1.15	1.5	1	1	1.5	1.5	1.5	1.5	1.3	1.3	1.3	1	1.3	1	1	1
31	KAR	$C^{(4)}$	6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
32	KAR	$C^{(4)}$	6	2	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
33	KAR	$C^{(4)}$	6	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
34	KAR	$C^{(4)}$	6	4	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
35	KAR	$C^{(4)}$	6	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
36	KAR	$C^{(4)}$	6	6	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
37	KAR	$C^{(4)}$	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
38	KAR	$C^{(4)}$	7	2	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
39	KAR	$C^{(4)}$	7	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
40	KAR	$C^{(4)}$	7	4	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
41	KAR	$C^{(4)}$	7	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
42	KAR	$C^{(4)}$	7	6	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
43	KAR	$C^{(4)}$	8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
44	KAR	$C^{(4)}$	8	2	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
45	KAR	$C^{(4)}$	8	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
46	KAR	$C^{(4)}$	8	4	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
47	KAR	$C^{(4)}$	8	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
48	KAR	$C^{(4)}$	8	6	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05

(*Continued*)

Table 1. (*Continued*).

49	KAR	$C^{(4)}$	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
50	KAR	$C^{(4)}$	9	2	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
51	KAR	$C^{(4)}$	9	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
52	KAR	$C^{(4)}$	9	4	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
53	KAR	$C^{(4)}$	9	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
54	KAR	$C^{(4)}$	9	6	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
55	KAR	$C^{(4)}$	10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
56	KAR	$C^{(4)}$	10	2	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
57	KAR	$C^{(4)}$	10	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
58	KAR	$C^{(4)}$	10	4	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
59	KAR	$C^{(4)}$	10	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
60	KAR	$C^{(4)}$	10	6	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
61	KAR	$C^{(4)}$	11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
62	KAR	$C^{(4)}$	11	2	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
63	KAR	$C^{(4)}$	11	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
64	KAR	$C^{(4)}$	11	4	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
65	KAR	$C^{(4)}$	11	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
66	KAR	$C^{(4)}$	11	6	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
67	KAR	$C^{(4)}$	12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
68	KAR	$C^{(4)}$	12	2	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
69	KAR	$C^{(4)}$	12	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
70	KAR	$C^{(4)}$	12	4	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
71	KAR	$C^{(4)}$	12	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
72	KAR	$C^{(4)}$	12	6	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
73	open	$C^{(3)}$	13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
74	open	$C^{(3)}$	13	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
75	open	$C^{(3)}$	13	3	1.05	1.4	1	1	1.4	1.4	1.4	1.4	1.3	1.3	1.3	1	1.25	1	1	1
76	open	$C^{(3)}$	13	4	1.15	1.1	1	1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1	1.1	1	1	1
77	open	$C^{(3)}$	13	5	1.15	1.3	1	1	1.3	1.3	1.3	1.3	1.2	1.2	1.2	1	1.2	1	1	1
78	open	$C^{(3)}$	13	6	1.15	1.5	1	1	1.5	1.5	1.5	1.5	1.3	1.3	1.3	1	1.3	1	1	1
79	res	$C^{(2)}$	14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
80	res	$C^{(2)}$	14	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
81	res	$C^{(2)}$	14	3	1.05	1.1	1	1	1.15	1.15	1.1	1.1	1.15	1.15	1.15	1	1.1	1	1	1
82	res	$C^{(2)}$	14	4	1.1	1.05	1	1	1.05	1.05	1.02	1.02	1.05	1.05	1.05	1	1.1	1	1	1
83	res	$C^{(2)}$	14	5	1.1	1.15	1	1	1.15	1.15	1.05	1.05	1.1	1.1	1.1	1	1.15	1	1	1
84	res	$C^{(2)}$	14	6	1.1	1.2	1	1	1.2	1.2	1.1	1.1	1.15	1.15	1.15	1	1.15	1	1	1
85	res	$C^{(2)}$	15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
86	res	$C^{(2)}$	15	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
87	res	$C^{(2)}$	15	3	1.05	1.1	1	1	1.15	1.15	1.1	1.1	1.15	1.15	1.15	1	1.1	1	1	1
88	res	$C^{(2)}$	15	4	1.1	1.05	1	1	1.05	1.05	1.02	1.02	1.05	1.05	1.05	1	1.1	1	1	1
89	res	$C^{(2)}$	15	5	1.1	1.15	1	1	1.15	1.15	1.05	1.05	1.1	1.1	1.1	1	1.15	1	1	1
90	res	$C^{(2)}$	15	6	1.1	1.2	1	1	1.2	1.2	1.1	1.1	1.15	1.15	1.15	1	1.15	1	1	1
91	GDY	$C^{(1)}$	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
92	GDY	$C^{(1)}$	16	2	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
93	GDY	$C^{(1)}$	16	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
94	GDY	$C^{(1)}$	16	4	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
95	GDY	$C^{(1)}$	16	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
96	GDY	$C^{(1)}$	16	6	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2

(*Continued*)

Table 1. (*Continued*).

97	GDY	$C^{(1)}$	17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
98	GDY	$C^{(1)}$	17	2	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
99	GDY	$C^{(1)}$	17	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
100	GDY	$C^{(1)}$	17	4	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
101	GDY	$C^{(1)}$	17	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
102	GDY	$C^{(1)}$	17	6	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
103	GDY	$C^{(1)}$	18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
104	GDY	$C^{(1)}$	18	2	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2
105	GDY	$C^{(1)}$	18	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
106	GDY	$C^{(1)}$	18	4	1.02	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
107	GDY	$C^{(1)}$	18	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
108	GDY	$C^{(1)}$	18	6	1.05	1.2	1.3	1.3	1.25	1.25	1.2	1.2	1.1	1.1	1.1	1.15	1.15	1.2	1.2	1.2

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Operating environment threats and climate-weather hazards impact on maritime ferry safety

K. Kołowrocki & E. Kuligowska
Gdynia Maritime University, Poland

ABSTRACT: In this paper, a general safety analytical model of a critical infrastructure under the influence of an operation process including operating environment threats related to climate-weather change is applied to safety prediction of maritime ferry. The main safety characteristics of the ferry in its corresponding operating area are evaluated. Namely, there are determined the unconditional safety function, the risk function, the mean lifetime up to the exceeding a critical safety state, the moment when the risk function value exceeds the acceptable safety level, the intensities of ageing and the coefficients of the operation and climate-weather impact on the ferry intensities of ageing.

1 INTRODUCTION

The paper is concerned with the operating environment threats and climate-weather hazards influence on the safety of a critical infrastructure. The maritime ferry operation process related to operating environment threats is described in (Kołowrocki et al. 2017a), whether the climate-weather change process for the ferry operating area is modeled in (Kuligowska 2017). The identification of the ferry operation process related to operating environment threats and climate-weather hazards is performed in (Kołowrocki et al. 2017b). Having these processes identified, the safety prediction of the considered ferry under the operation process including operating environment threats and climate-weather change influence is performed. The maritime ferry safety characteristics, i.e. the unconditional safety function and the risk function at changing in time operation including threats and climate-weather conditions (in February) are determined. Moreover, the safety and resilience indicators are presented: the mean lifetime up to the exceeding a critical safety state, the moment when the risk function value exceeds the acceptable safety level, the intensities of ageing and the coefficients of the operation and climate-weather impact on the maritime ferry safety.

2 CRITICAL INFRASTRUCTURE OPERATION PROCESS RELATED TO OPERATING ENVIRONMENT THREATS AND EXTREME WEATHER HAZARDS

2.1 *Critical infrastructure operation process related to operating environment threats*

We assume as in (Kołowrocki et al. 2017c) that the system during its operation process is taking ν', $\nu \in N$, different operation states $z'_1, z'_2, ..., z'_{\nu'}$. Further, we define the critical infrastructure operation process $Z'(t)$, $t \in < 0, +\infty)$ related to the critical infrastructure operating environment threats with discrete operation states from the set $\{z'_1, z'_2, ..., z'_{\nu'}\}$. Moreover, we assume that the critical infrastructure operation process $Z'(t)$ related to its operating environment threats is a semi-Markov process with the conditional sojourn times θ'_{bl} at the operation states z'_b when its next operation state is z'_l, $b,l = 1, 2, ..., \nu'$, $b \neq l$.

Under these assumptions, the critical infrastructure operation process may be described by (Kołowrocki, Kuligowska & Soszyńska-Budny 2017c):

– the vector $[p'_b(0)]_{1\times\nu'}$ of the initial probabilities $p'_b(0) = P(Z'(0) = z'_b)$, $b = 1, 2, \ldots, \nu'$, of the

process $Z'(t)$ staying at particular states at the moment $t = 0$;
- the matrix $[p'_{bl}]_{\nu' \times \nu'}$ of probabilities p'_{bl}, $b,l = 1, 2, ..., \nu'$, $b \neq l$, of the process $Z'(t)$ transitions between the operation states z'_b and z'_l;
- the matrix $[H'_{bl}(t)]_{\nu' \times \nu'}$ of conditional distribution functions $H'_{bl}(t) = P(\theta'_{bl} < t)$, $t \in <0,+\infty)$, $b,l = 1, 2, \ldots, \nu'$, $b \neq l$, of the system operation process $Z'(t)$ conditional sojourn times θ'_{bl} at the operation states.

The limit values of the process $Z'(t)$ transient probabilities $p'_b(t) = P(Z'(t) = z'_b)$, $t \in <0, +\infty)$, $b = 1, 2, ..., \nu'$, at the particular states can be found using the procedure given in (Kołowrocki, Kuligowska & Soszyńska-Budny 2017c).

2.2 *Climate-weather change process at the critical infrastructure operating area*

The climate-weather change process $C(t)$, $t \in <0, +\infty)$, for the critical infrastructure operating area is defined and modelled in (Kołowrocki, Soszyńska-Budny & Torbicki 2017a,b). We assume that the process is taking w, $w \in N$, different climate-weather states $c_1, c_2, \ldots, c_w$ and that this is a semi-Markov process. The process' parameters are described in (Kołowrocki & Kuligowska 2018), whereas the process' characteristics, e.g. the limit values of the climate-weather change process $C(t)$ transient probabilities $q_b(t) = P(C(t) = c_b)$, $t \in <0, +\infty)$, $b = 1, 2, ..., w$, at the particular climate-weather states, can be found using the procedure given in (Kołowrocki, Soszyńska-Budny & Torbicki 2017a).

2.3 *Critical infrastructure operation process related to operating environment threats and climate-weather hazards*

We assume as in (Kołowrocki et al. 2017a), that the critical infrastructure operation process $Z'C(t)$, $t \in < 0, + \infty)$, related to operating environment threats and climate-weather hazards can take $\nu' w$, ν', $w \in N$, different operation states $z'c_{11}, z'c_{12}, ..., z'c_{\nu w}$, described by:

- the vector $[p'q_{bl}(0)]_{1 \times \nu' w}$ of initial probabilities of the critical infrastructure process $Z'C(t)$ staying at the initial moment $t = 0$ at the particular states $z'c_{bl}$, $b = 1, 2, ..., \nu$, $l = 1, 2, ..., w$;
- the matrix $[p'q_{bl\tilde{b}\tilde{l}}]$ $\nu' w \times \nu' w$ of the probabilities of transitions of the critical infrastructure process $Z'C(t)$ between the states $z'c_{bl}$ and $z'c_{\tilde{b}\tilde{l}}$, $b = 1, 2, \ldots, \nu'$, $l = 1, 2, ..., w$, $\tilde{b} = 1, 2, ..., \nu'$, $\tilde{l} = 1, 2, ..., w$;
- the matrix $[H'C_{bl\,\tilde{b}\tilde{l}}(t)]_{\nu' w \times \nu' w}$ of the conditional distribution functions of the critical infrastructure process $Z'C(t)$ conditional sojourn times $\theta' C_{bl\,\tilde{b}\tilde{l}}$, $b = 1, 2, \ldots, \nu'$, $l = 1, 2, ..., w$, $\tilde{b} = 1, 2, ..., \nu'$, $\tilde{l} = 1, 2, ..., w$, at the state $z'c_{bl}$, $b = 1, 2, ..., \nu'$, $l = 1, 2, ..., w$, when the next state is $z'c_{\tilde{b}\tilde{l}}$, $\tilde{b} = 1, 2, ..., \nu'$, $\tilde{l} = 1, 2, ..., w$.

Assuming that we have identified the unknown parameters of the critical infrastructure operation process $Z'C(t)$ related to operating environment threats and climate-weather hazards, we can predict this process basic characteristics, e.g. the limit transient probabilities $p'q_{bl}(t) = P(Z'C(t) = z'c_{bl})$, $t \in <0, +\infty)$, $b = 1, 2, ..., \nu'$, $l = 1, 2, ..., w$, at the particular states, according to the procedure given in (Kołowrocki, Kuligowska & Soszyńska-Budny 2017a).

3 CRITICAL INFRASTRUCTURE SAFETY RELATED TO OPERATING ENVIRONMENT THREATS AND CLIMATE-WEATHER CHANGE PROCESS

In the safety analysis of critical infrastructures at the variable operation conditions under the influence of operating environment threats and related to climate-weather change, to define the critical infrastructure with degrading components we assume that the changes of the process $Z'C(t)$ states have an impact on the critical infrastructure's components and its structure (Kołowrocki 2014, Kołowrocki et al. 2017b). We denote the critical infrastructure asset A_i, $i = 1, 2, ..., n$, conditional lifetime in the safety state subset $\{u, u + 1, ..., z\}$, while the process $Z'C(t)$ is at the state $z'c_{bl}$, $b = 1, 2, ..., \nu'$, $l = 1, 2, ..., w$, by $[T_i^5(u)]^{(bl)}$. Moreover, in this section we assume that the critical infrastructure assets at particular states have the exponential safety functions. According to (Kołowrocki et al. 2017b), the conditional critical infrastructure safety function is defined by the vector

$$[S_i^5(t,\cdot)]^{(bl)} = [1, [S_i^5(t,1)]^{(bl)}, ..., S_i^5(t,z)]^{(bl)}], \; t \in <0, +\infty), \; b = 1, 2, ..., \nu', \; l = 1, 2, ..., w, \; i = 1, 2, ..., n, \tag{1}$$

with the coordinates

$$[S_i^5(t,u)]^{(bl)} = P([T_i^5(u)]^{(bl)} > t \mid Z'C(t) = z'c_{bl}) = \exp[-[\lambda_i^5(u)]^{(bl)}t], \; t \in <0, +\infty), \; b = 1, 2, ..., \nu', \; l = 1, 2, ..., w, \; i = 1, 2, \ldots, n, \tag{2}$$

where the intensities of ageing of the critical infrastructure assets related to operating environment threats and climate-weather impact, existing in (2), are given by

$$[\lambda_i^5(u)]^{(bl)} = [\rho_i^5(u)]^{(bl)} \cdot \lambda_i^0(u), \; u = 1, 2, ..., z, \; b = 1, 2, ..., \nu', \; l = 1, 2, ..., w, \; i = 1, 2, ..., n, \tag{3}$$

and $\lambda_i^0(u)$ are the intensities of ageing of the system components without any impact and

$$[\rho_i^5(u)]^{(bl)},\ u = 1, 2, ..., z,\ b = 1, 2, ..., \nu',\ l = 1, 2, ..., w,\ i = 1, 2, ..., n, \quad (4)$$

are the coefficients of operation and climate-weather change impact on the critical infrastructure assets' intensities of ageing without operation and climate-weather change impact.

Further, we denote the critical infrastructure unconditional lifetime in the safety state subset $\{u, u + 1, ..., z\}$ by $T^5(u)$ and the system unconditional safety function by

$$\boldsymbol{S}^5(t,\cdot) = [1, \boldsymbol{S}^5(t, 1), ..., \boldsymbol{S}^5(t,z)], \quad (5)$$

with the vector's coordinates defined by

$$\boldsymbol{S}^5(t, u) = P(T^5(u) > t),\ t \in <0, +\infty),\ u = 1, 2, ..., z. \quad (6)$$

In the case when the critical infrastructure operation time $\theta' C_{bl}$ is large enough, the coordinates of the unconditional safety function of the system defined by (5) are given by

$$\boldsymbol{S}^5(t,u) \cong \sum_{b=1}^{\nu'} \sum_{l=1}^{w} p'q_{bl}[\boldsymbol{S}^5(t,u)]^{(bl)},\ t \in < 0, +\infty),\ u = 1, 2, ..., z, \quad (7)$$

where $[\boldsymbol{S}^5(t,u)]^{(bl)}$, $u = 1, 2, ..., z$, $b = 1, 2, ..., \nu'$, $l = 1, 2, ..., w$, are the coordinates of the critical infrastructure conditional safety function defined by (2)-(4) and pq_{bl}, $b = 1, 2, ..., \nu$, $l = 1, 2, ..., w$, are the system operation process limit transient probabilities (see section 2.3).

Further, we determine the mean values $\mu^5(u)$ and the standard deviations $\sigma^5(u)$ of the unconditional lifetimes of the critical infrastructure in the safety state subsets $\{u, u + 1, ..., z\}$, $u = 1, 2, ..., z$, the mean values $\bar{\mu}^4(u)$ of the unconditional lifetimes of the critical infrastructure in the particular safety states u, $u = 1, 2, ..., z$, the risk function $r^5(t)$ and the moment τ^5 when the critical infrastructure risk function exceeds a permitted level δ, after substituting for $\boldsymbol{S}^5(t, u)$, $u = 1, 2, ..., z$, the coordinates of the unconditional safety functions given by (6).

4 MARITIME FERRY OPERATION PROCESS RELATED TO OPERATING ENVIRONMENT THREATS AND CLIMATE-WEATHER HAZARDS

Taking into account the expert opinions concerned with the operation process of the considered ferry technical system (Kołowrocki et al. 2017d), we assume that the maritime ferry operation process and safety may depend on operating environment threats and we distinguish the following 3 unnatural threats:

- ut_1 – a human error,
- ut_2 – a terrorist attack,
- ut_3 – a heavy sea traffic.

After assuming that the maritime ferry operation process including operating environment threats and the climate-weather change process at its operating area are independent, it is possible to evaluate the unknown basic parameter of the maritime ferry process $Z'C(t)$. Considering these identification results (Kołowrocki et al. 2017a), it was possible to predict this process' basic characteristics. The limit values of the process $Z'C(t)$ transient probabilities $p'q_{bl}$, $b = 1, 2, \ldots, 72$, $l = 1, 2, \ldots, 6$, at the particular states $z'c_{bl}$, are given in the vector

$$[p'q_{ij}]_{1\times 432} \cong [0.014948, 0.000407, 0.016687, 0.00037, 0.004588, 0;$$
$$0.0002424, 0.0000066, 0.0002706, 0.000006, 0.0000744, 0; 0, 0, 0, 0, 0, 0;$$
$$0.0001616, 0.0000044, 0.0001804, 0.000004, 0.0000496, 0;$$
$$0.000404, 0.000011, 0.000451, 0.00001, 0.000124, 0;$$
$$0.0002424, 0.0000066, 0.0002706, 0.000006, 0.0000744, 0; 0, 0, 0, 0, 0, 0;$$
$$0.0001616, 0.0000044, 0.0001804, 0.000004, 0.0000496, 0;$$
$$0.020125, 0.004175, 0.0002, 0, 0.000175, 0.000325;$$
$$0.000483, 0.0001002, 0.0000048, 0, 0.0000042, 0.0000078; 0, 0, 0, 0, 0, 0;$$
$$0.000322, 0.0000668, 0.0000032, 0, 0.0000028, 0.0000052;$$
$$0.028175, 0.005845, 0.00028, 0, 0.000245, 0.000455;$$
$$0.000483, 0.0001002, 0.0000048, 0, 0.0000042, 0.0000078; 0, 0, 0, 0, 0, 0;$$
$$0.000322, 0.0000668, 0.0000032, 0, 0.0000028, 0.0000052;$$
$$0.286704, 0.06697, 0, 0, 0.005792, 0.002534;$$
$$0.0004752, 0.000111, 0, 0, 0.0000096, 0.0000042; 0, 0, 0, 0, 0, 0;$$
$$0.0003168, 0.000074, 0, 0, 0.0000064, 0.0000028;$$
$$0.00985, 0.0004, 0.007775, 0.00075, 0.006225, 0;$$

0.0002364, 0.0000096, 0.0001866, 0.000018, 0.0001494, 0; 0, 0, 0, 0, 0, 0;
0.0001576, 0.0000064, 0.0001244, 0.000012, 0.0000996, 0;
0.001576, 0.000064, 0.001244, 0.00012, 0.000996, 0;
0.0002364, 0.0000096, 0.0001866, 0.000018, 0.0001494, 0; 0, 0, 0, 0, 0, 0;
0.0001576, 0.0000064, 0.0001244, 0.000012, 0.0000996, 0;
0.00591, 0.00024, 0.004665, 0.00045, 0.003735, 0;
0.0002364, 0.0000096, 0.0001866, 0.000018, 0.0001494, 0; 0, 0, 0, 0, 0, 0;
0.0001576, 0.0000064, 0.0001244, 0.000012, 0.0000996, 0;
0.014184, 0.000576, 0.011196, 0.00108, 0.008964, 0;
0.0002364, 0.0000096, 0.0001866, 0.000018, 0.0001494, 0; 0, 0, 0, 0, 0, 0;
0.0001576, 0.0000064, 0.0001244, 0.000012, 0.0000996, 0;
0.000394, 0.000016, 0.000311, 0.00003, 0.000249, 0;
0.0002364, 0.0000096, 0.0001866, 0.000018, 0.0001494, 0; 0, 0, 0, 0, 0, 0;
0.0001576, 0.0000064, 0.0001244, 0.000012, 0.0000996, 0;
0.000788, 0.000032, 0.000622, 0.00006, 0.000498, 0;
0.0002364, 0.0000096, 0.0001866, 0.000018, 0.0001494, 0; 0, 0, 0, 0, 0, 0;
0.0001576, 0.0000064, 0.0001244, 0.000012, 0.0000996, 0;
0.00591, 0.00024, 0.004665, 0.00045, 0.003735, 0;
0.0002364, 0.0000096, 0.0001866, 0.000018, 0.0001494, 0; 0, 0, 0, 0, 0, 0;
0.0001576, 0.0000064, 0.0001244, 0.000012, 0.0000996, 0;
0.2772, 0.06475, 0, 0, 0.0056, 0.00245;
0.0004752, 0.000111, 0, 0, 0.0000096, 0.0000042; 0, 0, 0, 0, 0, 0;
0.0003168, 0.000074, 0, 0, 0.0000064, 0.0000028;
0.026565, 0.005511, 0.000264, 0, 0.000231, 0.000429;
0.000483, 0.0001002, 0.0000048, 0, 0.0000042, 0.0000078; 0, 0, 0, 0, 0, 0;
0.000322, 0.0000668, 0.0000032, 0, 0.0000028, 0.0000052;
0.018515, 0.003841, 0.000184, 0, 0.000161, 0.000299;
0.000483, 0.0001002, 0.0000048, 0, 0.0000042, 0.0000078; 0, 0, 0, 0, 0, 0;
0.000322, 0.0000668, 0.0000032, 0, 0.0000028, 0.0000052;
0.000808, 0.000022, 0.000902, 0.00002, 0.000248, 0;
0.0002424, 0.0000066, 0.0002706, 0.000006, 0.0000744, 0; 0, 0, 0, 0, 0, 0;
0.0001616, 0.0000044, 0.0001804, 0.000004, 0.0000496, 0;
0.001616, 0.000044, 0.001804, 0.00004, 0.000496, 0;
0.0002424, 0.0000066, 0.0002706, 0.000006, 0.0000744, 0; 0, 0, 0, 0, 0, 0;
0.0001616, 0.0000044, 0.0001804, 0.000004, 0.0000496, 0;
0.004848, 0.000132, 0.005412, 0.00012, 0.001488, 0;
0.0002424, 0.0000066, 0.0002706, 0.000006, 0.0000744, 0; 0, 0, 0, 0, 0, 0;
0.0001616, 0.0000044, 0.0001804, 0.000004, 0.0000496, 0]. (8)

5 MARITIME FERRY SAFETY PREDICTION RELATED TO OPERATING ENVIRONMENT THREATS AND CLIMATE-WEATHER HAZARDS

5.1 *Maritime ferry safety parameters*

There are distinguished z = 5 safety states for the considered maritime ferry described in (Kołowrocki & Kuligowska 2018). Moreover, by the expert opinions, we assume that there are possible the transitions between the components' safety states only from better to worse ones.

Considering the assumptions and agreements from the previous sections, we assume that the components $E_{ij}^{(\upsilon)}$, $i = 1, 2, ..., k$, $j = 1, 2, ..., l_i$, of the subsystem S_υ, $\upsilon = 1, 2, 3, 4, 5$, at the particular states $z'c_{bl}$, $b = 1, 2, ..., 72$, $l = 1, 2, ..., 6$, have the exponential safety functions, i.e. the coordinates of the vector

$$[S_{ij}^{5(\upsilon)}(t,\cdot)]^{(bl)} = [1,[S_{ij}^{5(\upsilon)}(t,1)]^{(bl)}, ..., [S_{ij}^{5(\upsilon)}(t,4)]^{(bl)}],$$
$$t \in <0,+\infty), i = 1, 2, ..., k, \ j = 1, 2, ..., l_i,$$
$$\upsilon = 1, 2, 3, 4, 5, b = 1, 2, ..., 72, l = 1, 2, ..., 6, \quad (9)$$

are given by

$$[S_{ij}^{5(\upsilon)}(t,u)]^{(bl)} = P([T_{ij}^{5(\upsilon)}]^{(bl)}(u) > t \,|\, Z'C(t) = z'c_{bl}) = \exp[-[\lambda_{ij}^{5(\upsilon)}(u)]^{(bl)}t],$$
$$t \in <0,+\infty), \ u = 1, 2, 3, 4, i = 1, 2, ..., k, \ j = 1, 2, ..., li,$$
$$v = 1, 2, 3, 4, 5, b = 1, 2, ..., 72, l = 1, 2, ..., 6. \quad (10)$$

Existing in the above formula the intensities of ageing of the components $E_{ij}^{(\upsilon)}$, $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, l_i$, of the subsystem S_υ, $\upsilon = 1, 2, 3, 4, 5$, at the system states $z'c_{bl}$, $b = 1, 2, \ldots, 72$, $l = 1, 2, \ldots, 6$, i.e. the coordinates of the vector of intensities

$$[\lambda_{ij}^{5(\upsilon)}(\cdot)]^{(bl)} = [0,[\lambda_{ij}^{5(\upsilon)}(1)]^{(bl)},\ldots,[\lambda_{ij}^{5(\upsilon)}(4)]^{(bl)}], \quad i = 1,2,\ldots,k,\ j = 1,2,\ldots,li, \quad \upsilon = 1,2,3,4,5, b = 1,2,\ldots,72, l = 1,2,\ldots,6, \tag{11}$$

are given by

$$[\lambda_{ij}^{5(\upsilon)}(u)]^{(bl)} = [\rho_{ij}^{5(\upsilon)}(u)]^{(bl)}\lambda_{ij}^{0(\upsilon)}(u), \quad u = 1, 2, 3, 4, i = 1, 2, \ldots, k, j = 1, 2, \ldots, l_i, \quad \nu = 1, 2, 3, 4, 5, b = 1, 2, \ldots, 72, l = 1, 2, \ldots, 6, \tag{12}$$

where $\lambda_{ij}^{0(\upsilon)}(u)$, $u = 1, 2, 3, 4$, $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, l_i$, are the intensities of ageing of the components $E_{ij}^{(\upsilon)}$, $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, l_i$, of the subsystems S_υ, $\upsilon = 1, 2, 3, 4, 5$, without of any impact and $[\rho_{ij}^{5(\upsilon)}(u)]^{(bl)}$, $u = 1, 2, \ldots, k$, $j = 1, 2, \ldots, l_i$, $\upsilon = 1, 2, 3, 4,5$, $b = 1, 2, \ldots, 72$, $l = 1, 2, \ldots, 6$, are the coefficients of the operation impact including threats and climate-weather change impact on the components $E_{ij}^{(\upsilon)}$, $i = 1, 2, \ldots, k, j = 1, 2, \ldots, l_i$, of the subsystems S_υ, $\upsilon = 1, 2, 3, 4,5$, intensities of ageing at the states $z'c_{bl}$, $b = 1, 2, \ldots, 72$, $l = 1, 2, \ldots, 6$.

According to expert opinions, changing the maritime ferry operation process states have influence on changing the system safety structures and its selected components' safety parameters as well. For this system, the intensities of components' departure from the safety states subsets {1,2,3,4}, {2,3,4}, {3,4}, {4} without of any impact are given in (Kołowrocki et al. 2017d), the intensities of departure related to the climate-weather influence on the maritime ferry safety are given in (Kołowrocki & Kuligowska 2018) and the intensities of departure related to the operating environment threats influence on ferry safety are given in (Kołowrocki et al. 2017d). Thus, the intensities of departure related to the operating environment threats and climate-weather hazards influence on ferry safety are calculated according to formula (12), where the coefficients of the operation, operating environment threats and climate-weather change impact on the components' intensities of ageing at the particular states are the multiplication of the above mentioned coefficients of impact on the components' intensities of ageing at the particular states.

5.2 *Maritime ferry safety characteristics*

Assuming that the maritime ferry technical system safety structure and its subsystems and components safety depend on its changing in time operation states, the influence of the system states changing on the changes of the system safety structure and its components safety functions is given in (Kołowrocki et al. 2017b). Thus, in the case when the operation time is large enough, according to (7) the maritime ferry unconditional safety function is given by the vector

$$\boldsymbol{S}^5(t,\cdot) = [1, \boldsymbol{S}^5(t,1), \boldsymbol{S}^5(t,2), \boldsymbol{S}^5(t,3), \boldsymbol{S}^5(t,4)], \quad t \in <0, +\infty), \tag{13}$$

where according to (7) the vector coordinates are given respectively for $t \in <0, +\infty)$, $u = 1, 2, 3, 4$, by

$$\boldsymbol{S}^5(t,u) \cong \sum_{b=1}^{72}\sum_{l=1}^{6} p'q_{bl}[\boldsymbol{S}^5(t,u)]^{(bl)}, \tag{14}$$

where $[\boldsymbol{S}^5(t,u)]^{(bl)}$, $u = 1, 2, 3, 4$, $b = 1, 2, \ldots, 72$, $l = 1, 2, \ldots, 6$, are the coordinates of the system conditional safety functions defined by (2)–(4) and $p'q_{bl}$, $b = 1, 2, \ldots, 72$, $l = 1, 2, \ldots, 6$, are the limit transient probabilities given by (8).

The graph of the five-state maritime ferry technical system safety function is presented in Figure 1.

Considering (14), the expected values and standard deviations, given in years, of the maritime ferry technical system lifetimes in the safety states subsets {1,2,3,4}, {2,3,4}, {3,4}, {4}, respectively are

$$\mu^5(1) \cong 5.778643,\ \mu^5(2) \cong 3.157926,\ \mu^5(3) \cong 2.339679,\ \mu^5(4) \cong 1.877708; \tag{15}$$

$$\sigma^5(1) \cong 5.557708,\ \sigma^5(2) \cong 3.07524,\ \sigma^5(3) \cong 2.280044,\ \sigma^5(4) \cong 1.828565. \tag{16}$$

Consequently, the mean values of the maritime ferry technical system lifetimes in the particular safety states 1, 2, 3, 4, respectively are:

$$\overline{\mu}^5(1) = 2.620717,\ \overline{\mu}^5(2) = 0.8182476,\ \overline{\mu}^5(3) = 0.4619703,\ \overline{\mu}^5(4) = 1.877708. \tag{17}$$

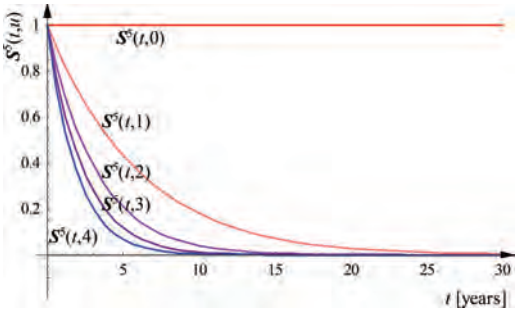


Figure 1. The graph of the maritime ferry safety function coordinates.

By (15) and (16), the mean and the standard deviation of the maritime ferry lifetime up to exceeding critical safety state $r = 2$ are

$$\mu^5(2) = 3.157926 \text{ years}, \quad \sigma^5(2) = 3.07524 \text{ years}.$$

The system risk function of the maritime ferry technical system, is given by

$$\boldsymbol{r}^5(t) = 1 - \boldsymbol{S}^5(t,2), \tag{18}$$

where $\boldsymbol{S}^5(t,2)$ is given by (14) and illustrated in Figure 2.

Hence, considering (18), the moment when the system risk function exceeds a permitted level, for instance $\delta = 0.05$, is given as follows

$$\tau^5 = \boldsymbol{r}^{5-1}(\delta) \cong 0.17 \text{ year}. \tag{19}$$

The maritime ferry intensities of ageing according to (Kołowrocki et al. 2017b) and considering (14) are:

$$\boldsymbol{\lambda}^5(t,u) = -\frac{d(\boldsymbol{S}^5(t,u))}{dt} \cdot \frac{1}{\boldsymbol{S}^5(t,u)}, \tag{20}$$

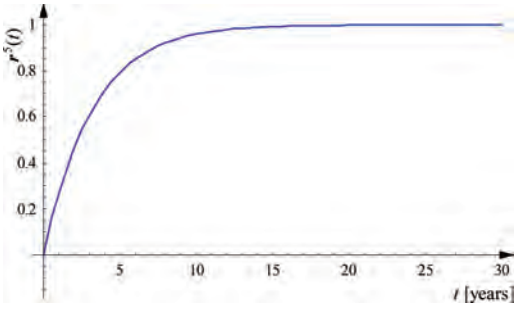


Figure 2. The graph (the fragility curve) of the maritime ferry risk function.

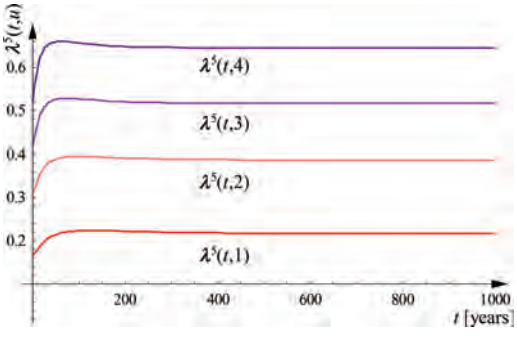


Figure 3. The graph of the intensities of ageing of the maritime ferry.

where particularly

$$\boldsymbol{\lambda}^5(t,1) \cong 0.2165163, \quad \boldsymbol{\lambda}^5(t,2) \cong 0.3860834,$$
$$\boldsymbol{\lambda}^5(t,2) \cong 0.5164971, \quad \boldsymbol{\lambda}^5(t,3) \cong 0.644996.$$

The graphs of the intensities of ageing for the maritime ferry are shown in Figure 3.

Considering (15) and applying (57) from (Kołowrocki, Soszyńska-Budny & Torbicki 2017d), the coefficients of operating environment threats and climate-weather change impact on the ferry safety are

$$\boldsymbol{\rho}^5(t,1) \cong \frac{1/\mu^5(1)}{1/\mu^0(1)} \cong \frac{1/5.778643}{1/6.246} \cong 1.080877,$$
$$\boldsymbol{\rho}^5(t,2) \cong \frac{1/\mu^5(2)}{1/\mu^0(2)} \cong \frac{1/3.157926}{1/3.390} \cong 1.073489,$$
$$\boldsymbol{\rho}^5(t,3) \cong \frac{1/\mu^5(3)}{1/\mu^0(3)} \cong \frac{1/2.339679}{1/2.503} \cong 1.069805,$$
$$\boldsymbol{\rho}^5(t,4) \cong \frac{1/\mu^5(4)}{1/\mu^0(4)} \cong \frac{1/1.877708}{1/2.007} \cong 1.068856. \tag{21}$$

The resilience indicator, i.e. the coefficient of maritime ferry resilience to operation process including threats and climate-weather change process impact is

$$\boldsymbol{RI}^5(t) = \frac{1}{\rho^5(t,2)} \cong 0.9315416 \cong 93.15\%. \tag{22}$$

6 CONCLUSIONS

The simplified impact model of critical infrastructure safety related to operating environment threats and climate-weather change impact was applied to the safety and risk evaluation for the maritime ferry operating at Baltic Sea waters. The predicted maritime ferry safety characteristics are different from those determined for this system operating at constant conditions without considering any impact. This approach makes the systems safety prediction much more precise.

ACKNOWLEDGMENTS

The paper presents the results developed in the scope of the EU-CIRCLE project titled "A pan—European framework for strengthening Critical Infrastructure resilience to climate change" that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 653824, *http://www.eu-circle.eu/*.

REFERENCES

Kołowrocki, K. & Kuligowska, E., 2018. Operation and Climate-Weather Change Impact on Maritime Ferry Safety. Proc. *European Safety and Reliability Conference—ESREL* 2018, Trondheim, Norway.

Kołowrocki, K. and Soszyńska-Budny, J., 2011. *Reliability and Safety of Complex Technical Systems and Processes: Modeling—Identification—Prediction—Optimization*, Springer.

Kołowrocki, K., 2014. *Reliability of Large and Complex Systems*, Elsevier.

Kołowrocki, K., Kuligowska, E. and Soszyńska-Budny, J., 2015. Reliability of maritime ferry technical system, analytical assessment. *Proc. European Safety and Reliability Conference—ESREL* 2015, Zurich, Switzerland.

Kołowrocki, K., Kuligowska, E. and Soszyńska-Budny, J., 2017a. Critical infrastructure operation process related to operating environment threats and extreme weather hazards. *Journal of Polish Safety and Reliability Association,* Summer Safety and Reliability Seminars 8(2), 41–58.

Kołowrocki, K., Kuligowska, E. and Soszyńska-Budny, J., 2017b. Identification and prediction of maritime ferry operation process including operating environment threats and extreme weather hazards. *Journal of Polish Safety and Reliability Association,* Summer Safety and Reliability Seminars, 8(2), 145–154.

Kołowrocki, K., Kuligowska, E. and Soszyńska-Budny, J., 2017c. Integrated Impact Model on Critical Infrastructure Safety Related to Its Operation Process Including Operating Environment Threats. *Journal of Polish Safety and Reliability Association,* Summer Safety and Reliability Seminars, 8(4), 11–20.

Kołowrocki, K., Kuligowska, E. and Soszyńska-Budny, J., 2017d. Integrated model of maritime ferry safety related to its operation process including operating environment threats. *Proceedings of European Safety and Reliability Conference—ESREL* 2017, Portoroz, Slovenia.

Kołowrocki, K., Kuligowska, E. and Soszyńska-Budny, J., 2017e. Safety of maritime ferry related to its operation process. *Proceedings of European Safety and Reliability Conference—ESREL* 2017, Portoroz, Slovenia.

Kołowrocki, K., Kuligowska, E., Soszyńska-Budny, J. and Torbicki, M., 2017a. Critical Infrastructure Operation Process and Climate-Weather Change Process Data Processing—Identification and Prediction. Report for D6.4 Case Study 2 PL: Conduction.

Kołowrocki, K., Kuligowska, E., Soszyńska-Budny, J. and Torbicki, M., 2017b. Integrated Impact Model on Critical Infrastructure Safety Related to Its Operation Process Including Operating Environment Threat and Climate-Weather Change Process Including Extreme Weather Hazards. *Journal of Polish Safety and Reliability Association,* Summer Safety and Reliability Seminars, 8(4), 49–66.

Kołowrocki, K., Soszyńska-Budny, J. and Torbicki, M., 2017a. Critical infrastructure operating area climate-weather change process including extreme weather hazards. *Journal of Polish Safety and Reliability Association,* Summer Safety and Reliability Seminars 8(2), 15–24.

Kołowrocki, K., Soszyńska-Budny, J. and Torbicki, M., 2017b. Identification methods and procedures of climate-weather change process including extreme weather hazards. *Journal of Polish Safety and Reliability Association,* Summer Safety and Reliability Seminars, (8)2, 85–106.

Kołowrocki, K., Soszyńska-Budny, J. and Torbicki, M., 2017c. Integrated Impact Model on Critical Infrastructure Safety Related to Its Operation Process and Climate-Weather Change Process. *Journal of Polish Safety and Reliability Association,* Summer Safety and Reliability Seminars, (8)4, 33–48.

Kołowrocki, K., Soszyńska-Budny, J. and Torbicki, M., 2017d. Safety of maritime ferry operating at Baltic sea open waters related to climate-weather change process including extreme weather hazards. *Proceedings of European Safety and Reliability Conference—ESREL* 2017, Portoroz, Slovenia.

Kuligowska, E., 2017. Identification and prediction of climate-weather change process for maritime ferry operating area. *Journal of Polish Safety and Reliability Association*, Summer Safety and Reliability Seminars, 8(2), 129–134.

Soszyńska-Budny, J., 2013. Modeling Safety of Multistate Systems with Application to Maritime Ferry Technical System, *Reliability: Theory and Applications,* 8(3): 24–39.

Xue, J. and Yang, K., 1995. Dynamic reliability analysis of coherent multi-state systems, *IEEE Trans on Reliab*. 4(44): 683–688.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Discussion on probabilistic and interval approaches applied to the Eurocode 7

Sónia H. Marques
Doctor in Geotechnics and Senior Professional Civil Engineer of the Portuguese Institution of Engineers, Portugal

ABSTRACT: The capabilities of nontraditional models for engineering computation under uncertainty have been under continuous review so that several nonprobabilistic approaches have been developed. The first applications of nonprobabilistic interval analysis in geotechnical engineering have been recently explored, considered a research area formally motivated by input information characterised by imprecision. Thus, the conventional probabilistic approach to uncertainty may be extended to include imprecise information in the form of intervals. For demonstration, results are provided on the analysis of a strip spread foundation designed by the Eurocode 7 methodology. A limit state imprecise interval analysis for bearing capacity is presented in the format of a sensitivity analysis. The limit state charts to safety assessment are separately sketched for the cases cohesion and friction angle interval scenario wherein the random variables are bounded on different levels of probability. The corresponding optimisation-based probability box structures are then sketched. The extension for high dimensional cases is as well considered through a limit state three-dimensional joint view to safety assessment, considered simultaneously the interval variables cohesion and friction angle. At last, the Eurocode 7 partial factor design is discussed on the basis of distinct levels of credibility.

1 INTRODUCTION

In recent years, several nonprobabilistic approaches have been developed to include uncertainties described by scarce information. If reliability is seen as a probability related to the satisfactory performance of a system under given circumstances, a nonprobabilistic concept of reliability holds on the acceptable range of performance fluctuations. According to the discussion of Haim & Elishakoff (1995), when considered only input interval variables, the nonprobabilistic concept of reliability is approached only by bounds as the safety margin is as well expressed in the interval format. Among nonprobabilistic approaches, the ordinary interval analysis involves the mapping of interval input to interval output quantities. The main advantage is the evaluation of the analytical enclosure of the true solution. Thereby, a considerable number of papers devoted to interval versus stochastic analysis have been published. Intervals represent an appropriate model to describe uncertainty in cases when a possible range between bounds is known and no other information concerning frequencies is available. Apart from, interval quantities may as well be included in computation based on other uncertainty models. Interval probabilities emerge from the consideration of a set of plausible probabilistic models in order to find the lower and upper probability bounds. This procedure is particularly convenient for sensitivity analysis, a very efficient tool to identify important design parameters and to address the performance level of real engineering systems in the context of optimisation procedures. Thus, whenever some probabilistic information is available, a mixed approach gathering imprecise information in the form of intervals may be explored. Vicig & Seidenfeld (2012) discuss the idea of replacing one exact probability value by introducing an indecision interval with two different exact one-sided values as endpoints. As the solution of practical problems in geotechnical engineering often requires judgement based on limited information, the assumption of interval-valued parameters in a mixed approach that admits as well probabilistic information verily complies with the imprecision of the input scenario. For demonstration, results are provided for a strip spread foundation designed by the Eurocode 7 methodology (EN 1997-1 2004), wherein the shear strength parameters of the foundation soil are implemented as intervals and then combined with other uncertain parameters in the form of random variables bounded on different levels of probability. A limit state imprecise interval analysis for bearing capacity is then presented in the format of a sensitiv-

ity analysis wherein the safety margin is as well expressed in the interval format.

2 LIMIT STATE IMPRECISE INTERVAL ANALYSIS

On the limit state imprecise interval analysis some selected parameters are separately or simultaneously implemented as imprecise intervals and then combined with other uncertain parameters in the form of random variables bounded under dependence; thereby, the random variables are bounded by using imprecise interval quantities in order to express different levels of probability. A brief explanation of the methodology for minimisation and maximisation which considers dependence is summarised next. The step by step optimisation algorithm is supported by transformations derived in compensative probability to express the performance function in the standard normal space of uncorrelated random and interval variables:

[Step 1] Express the distributions and the statistics of basic input variables;
[Step 2] Express the equivalent standard normal correlation matrix;
[Step 3] Express the outcome matrix from the Cholesky decomposition of the equivalent standard normal correlation matrix;
[Step 4] Express the set of random and interval variables vector in the standard normal space of uncorrelated random and interval variables;
[Step 5] Express the performance function and the limit state in the standard normal space of uncorrelated random and interval variables;
[Step 6] Express the set of constraints and whenever required the set of guess values for initialisation of the suboptimisation algorithm;
[Step 7] Select the suboptimisation algorithm properly and run the iterative process in a multistart approach;
[Step 8] Estimate the pair minimum-maximum and the corresponding coordinates in the standard normal space of uncorrelated as well as in the original space of correlated random and interval variables.

Thereby, the performance function in the standard normal space of uncorrelated variables is expressed by the following equivalent transformations $xi = \mathrm{f}(yi)$ and $\mathrm{f}(xi) = yi$ straightforwardly derived by the next Equation (1) in compensative probability:

$$xi = \mathrm{F_{Xi}}^{-1}\left[\Phi(yi)\right] \leftrightarrow \Phi^{-1}\left[\mathrm{F_{Xi}}(xi)\right] = yi \tag{1}$$

if xi = variable in the original space; yi = variable in the standard normal space; $\mathrm{F_{Xi}}$ = cumulative nonnormal distribution function; $\mathrm{F_{Xi}}^{-1}$ = inverse cumulative nonnormal distribution function; Φ = cumulative standard normal distribution function; and Φ^{-1} = inverse cumulative standard normal distribution function.

The correlation assignment in Equation (2) is also added to this standard normal space script:

$$\left[cyi^{*}\right] = \left[\mathrm{Ch}\right]\left[uyi^{*}\right] \tag{2}$$

if cyi^{*} = set of variables vector in the standard normal space of correlated variables; uyi^{*} = set of variables vector in the standard normal space of uncorrelated variables; and Ch = outcome matrix from the Cholesky decomposition of the equivalent standard normal correlation matrix.

Considered normal and lognormal distributions, the equivalent transformations for representation of variables xi in the original space as a function of variables yi in the standard normal space are further detailed as follows in the next Table 1 as explicit expressions for the equivalent transformations $xi = \mathrm{f}(yi)$ and $\mathrm{f}(xi) = yi$ as previously derived by Equation (1) in compensative probability.

A simplified approach may be considered on the computation of the equivalent standard normal correlation matrix wherein the values of the correlation coefficients between the variables are transformed by empirical relationship. Thus, the transformation F expressed in the next Table 2 is derived by exact relationship and used for the computation of the resultant equivalent standard normal correlation coefficient, when the random variables are normal and lognormal.

A test function as the Rosenbrock which is characterised by one challenging minimum point is very convenient for the purpose of testing, see the following Figure 1.

Table 1. Transformations for representation of variables xi in the original space as a function of variables yi in the standard normal space.

Distributions	$xi = \mathrm{f}(yi)$
Normal	$xi = \mu + \sigma \cdot yi$
Lognormal	$xi = \mathrm{e}^{(\mu l + \sigma l \cdot yi)}$
Distributions	$\mathrm{f}(xi) = yi$
Normal	$\dfrac{xi - \mu}{\sigma} = yi$
Lognormal	$\dfrac{\ln(xi) - \mu l}{\sigma l} = yi$

μ - mean value; σ - standard deviation; μl - log mean value; σl - log standard deviation.

Table 2. Transformation F by exact relationship for representation of the coefficient of correlation between two standard normal variables xi′ and xj′ as $\rho'_{xi'xj'} = F \cdot \rho_{xixj}$.

Variable xi	Variable xj	Transformation F
Normal	Lognormal	$\frac{cv_{xj}}{\sqrt{\left(\ln\left(1+cv_{xj}^2\right)\right)}}$

cv - coefficient of variation.

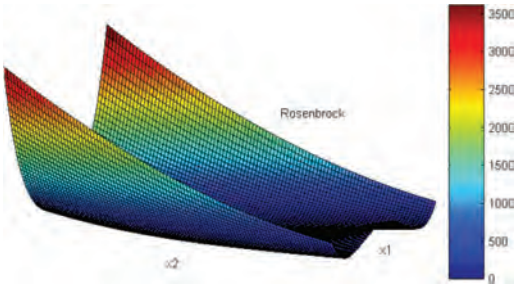


Figure 1. Rosenbrock test function overview if $x_1 \in [-3,+3]$ and $x_2 \in [-3,+3]$.

In fact, a considerable number of suboptimisation algorithms may be considered for the purpose but both local and global engines may fail under certain circumstances. In particular, conjugate gradient and quasi newton optimisation algorithms in a multistart approach are successful on the search for the minimum point of the Rosenbrock test function.

3 DESIGN EXAMPLE

The design example is referred to the strip spread foundation on a relatively homogeneous soil shown in Figure 2, wherein groundwater level is away. Considered the vertical noneccentric loading problem and the calculation model for bearing capacity, the performance function may be described by the simplified Equation (3):

$$M = f\left(B, D, \gamma_s, c_f, \varphi_f, \gamma_f, P, Q\right) \tag{3}$$

if B is the foundation width; D is the soil height above the foundation base; γ_s is the unit weight of the soil above the foundation base; c_f is the cohesion of the foundation soil; φ_f is the friction angle of the foundation soil; γ_f is the unit weight of the foundation soil; P is the dead load; and Q is the live load.

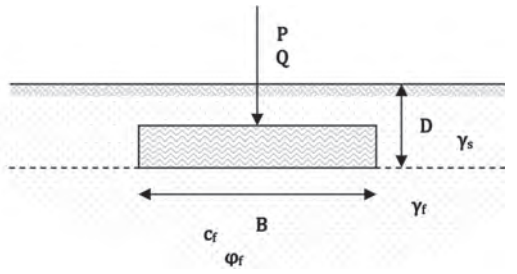


Figure 2. Strip spread foundation.

Table 3. Summary description of basic input variables.

Basic input variables	Distributions	Mean value	Coefficient of variation
B (m)	Deterministic	1.30	0.00
D (m)	Deterministic	1.00	0.00
γ_s (kN/m^3)	Normal	16.80	0.05
c_f (kN/m^2)	Lognormal	14.00	0.40
	Interval*	…	…
φ_f (°)	Lognormal	32.00	0.10
	Interval*	…	…
γ_f (kN/m^3)	Normal	17.80	0.05
P (kN/m)	Normal	370.00	0.10
Q (kN/m)	Normal	70.00	0.25

*Cases cohesion interval scenario [0.00,35.00] and friction angle interval scenario [25.00,35.00].

Table 4. Correlation coefficients between the basic input variables.

Correlation matrix

ρ_{x1x1}	ρ_{x1x2}	ρ_{x1x3}	ρ_{x1x4}	ρ_{x1x5}	ρ_{x1x6}
ρ_{x2x1}	ρ_{x2x2}	ρ_{x2x3}	ρ_{x2x4}	ρ_{x2x5}	ρ_{x2x6}
ρ_{x3x1}	ρ_{x3x2}	ρ_{x3x3}	ρ_{x3x4}	ρ_{x3x5}	ρ_{x3x6}
ρ_{x4x1}	ρ_{x4x2}	ρ_{x4x3}	ρ_{x4x4}	ρ_{x4x5}	ρ_{x4x6}
ρ_{x5x1}	ρ_{x5x2}	ρ_{x5x3}	ρ_{x5x4}	ρ_{x5x5}	ρ_{x5x6}
ρ_{x6x1}	ρ_{x6x2}	ρ_{x6x3}	ρ_{x6x4}	ρ_{x6x5}	ρ_{x6x6}

1.0	0.0	0.5	0.9	0.0	0.0
0.0	1.0	0.0	0.0	0.0	0.0
0.5	0.0	1.0	0.5	0.0	0.0
0.9	0.0	0.5	1.0	0.0	0.0
0.0	0.0	0.0	0.0	1.0	0.0
0.0	0.0	0.0	0.0	0.0	1.0

x_1-γ_s; x_2-c_f; x_3-φ_f; x_4-γ_f; x_5-P; x_6-Q; ρ-coefficient of correlation.

Table 3 summarises the description of basic input variables, with different types of distributions. The considered correlation coefficients between the

basic input variables are either presented in Table 4. The strip spread foundation is designed by the Eurocode 7 methodology, Design Approach DA.2*.

4 RESULTS AND DISCUSSION

A cautious estimate of the 95% reliable mean value for a known coefficient of variation is considered for each geotechnical characteristic value. The summary description of reliability estimates in interval scenario is further provided in Table 5 for the cases cohesion and friction angle interval scenario, wherein results are obtained from 5e6 simulations or sampling points, Monte Carlo simulation (MCS) considered. The generalised extreme value distribution is selected after a study with Kolmogorov Smirnov and Anderson Darling and Chi Squared goodness of fit tests. The shear strength parameters of the foundation soil are separately implemented as intervals and then combined with other uncertain parameters in the form of random variables under dependence. The optimisation-based probability box structures for the cases cohesion and friction angle interval scenario are drawn in the next Figure 3 and Figure 4 by using conjugate gradient and quasi newton optimisation algorithms in a multistart approach.

The 3.8 target reliability index is considered on the determination of the cohesion and friction angle values which satisfy the safety on every inter-

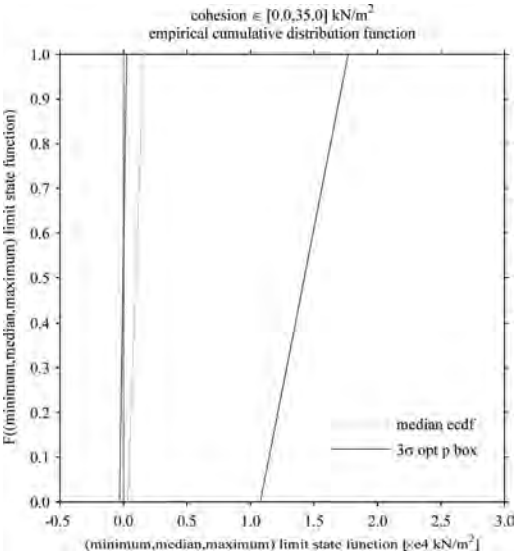


Figure 3. Optimisation-based probability box structure for the case cohesion interval scenario.

Table 5. Summary description of reliability estimates in interval scenario.

Case	Value	MCS failure probability	Reliability index β_{MCS}	FIT failure probability	Reliability index β_{FIT}
Cohesion interval scenario	0.00	4.4845e-2	1.6970	4.4800e-2	1.6974
	5.00	4.5420e-3	2.6089	5.0000e-3	2.5727
	10.00	2.0200e-4	3.5375	2.8887e-4	3.4419
	11.00	8.9800e-5	3.7461	1.4933e-4	3.6165
	12.00	4.1800e-5	3.9338	7.6973e-5	3.7846
	13.00	1.9800e-5	4.1098	3.8507e-5	3.9535
	14.00	7.2000e-6	4.3377	1.8511e-5	4.1253
	15.00	3.2000e-6	4.5127	9.1553e-6	4.2846
	20.00	≈0	≈∞	1.8173e-7	5.0872
	25.00	≈0	≈∞	2.5786e-9	5.8420
	30.00	≈0	≈∞	2.9561e-11	6.5459
	35.00	≈0	≈∞	3.0374e-13	7.1988
Friction angle interval scenario	25.00	2.2054e-3	2.8472	1.8000e-3	2.9189
	25.50	5.7560e-4	3.2507	3.8867e-4	3.3607
	26.00	1.3200e-4	3.6483	6.0447e-5	3.8443
	26.50	1.7600e-5	4.1369	5.9806e-6	4.3783
	27.00	1.8000e-6	4.6332	3.3652e-7	4.9690
	27.50	2.0000e-7	5.0690	8.9437e-9	5.6313
	30.00	≈0	≈∞	≈0	≈∞
	35.00	≈0	≈∞	≈0	≈∞

MCS results from 5e6 simulations in interval scenario.
FIT results from 5e6 sampling points for the generalised extreme value distribution.

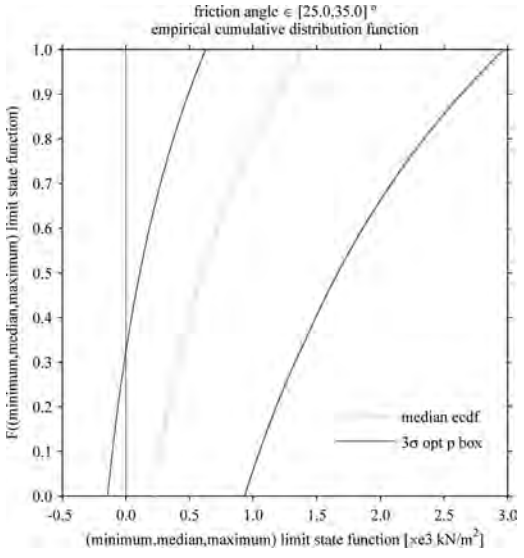


Figure 4. Optimisation-based probability box structure for the case friction angle interval scenario.

val scenario by simulation and distribution fitting. A boundless immeasurable reliability is further noticed on both cases. A comparative analysis of results from both methodologies shows appreciable differences whenever the cohesion values are approaching 15.0 kN/m^2 and the friction angle values are approaching 27.5º, noted that for the correspondent levels of reliability the Monte Carlo simulation failure probability error is estimated higher than 10%. Moreover, it is clearly shown that small variations in the friction angle input are very influential in that the median and the imprecise lower bound of probability correspond to a boundless immeasurable reliability, considered that satisfactory levels of reliability estimates are separately attained for a cohesion of about 12.0 kN/m^2 and for a friction angle of about 26.0º.

Regarding the graphs on Figure 3 and Figure 4, appropriate representation of the minimum and maximum branches comprehends respectively a number of 351 or 201 discretisation points uniformly distributed on the interval [0.0,35.0] kN/m^2 or [25.0,35.0]º. The stairstep median trend is comparatively drawn from a moderate number of discretisation points uniformly distributed on the intervals. The optimisation-based probability box structures are constructed comparatively to the median empirical cumulative distribution function, noted that the empirical cumulative distribution functions corresponding to the collection of minimum and maximum values are sketched at a central credibility level 3σ. From observation of Figure 3 and Figure 4, it is noted that the linear and nonlinear trends evinced by the graphs express the influence of the parameters cohesion and friction angle on the calculation model for bearing capacity. The position of the median trend, which separates the fifty percent chance cases, is further sketched on the safe side. Regarding the limit state, the probability of no failure across the cohesion and friction angle interval scenario at a central credibility level 3σ is respectively circa 50% or 70%, see the probability level for a positive outcome on the minimum branch of the graphs.

The limit state charts to safety assessment are represented on Figure 5 and Figure 6, considered respectively and separately the cases cohesion and friction angle interval scenario, in a sensitivity analysis wherein the random variables are bounded on different levels of probability. Distinct levels of credibility are used to sketch the lines which express the limit state bounds. It is possible to search for a credibility level which ensures no failure regardless of the parameter value on the horizontal axis, see safe level for arrow in Figure 6 against unsafe level for arrow in Figure 5. Conversely, it is possible to find the threshold parameter which ensures no failure for a given credibility level and then to proceed with proper ground investigation and testing or improvement, see in both charts the circles crossing the zero limit state boundary and the 0.9900 credibility level line. In decision making, this valuable approach may be extended by numerical analysis for high dimensional cases with several indecision variables in simultaneous, see Figure 7 and Figure 8.

According to the results on Table 5 and considered the cases cohesion and friction angle interval scenario, the threshold values which satisfy the 3.8 target reliability index are about 12.0 kN/m^2 and 26.0º, respectively. Following the vagueness of the indecision interval, this imprecise approach is interpreted altogether with a complementary limit state imprecise interval analysis for sensitivity, validation and decision. Thereby, it is concluded that satisfactory levels of reliability estimates are attained by the probability level of 0.9900 sketched on the limit state charts to safety assessment. On this framework, Figure 7 and Figure 8 display respectively the limit state three-dimensional joint view to safety assessment for a 0.9900 and a 0.9990 probability level. A three-dimensional representation of the zero limit state reference surface is crossed in one corner by the limit state lower bound surface constructed from a joint assumption of the values of the interval variables cohesion and friction angle. Considered the Figure 7, unsafe coordinates are limited to a cohesion parameter under a value between 7.0 kN/m^2 and 10.5 kN/m^2 or to a friction angle parameter under a value between 29.0º and 30.0º, considered the two worst combination

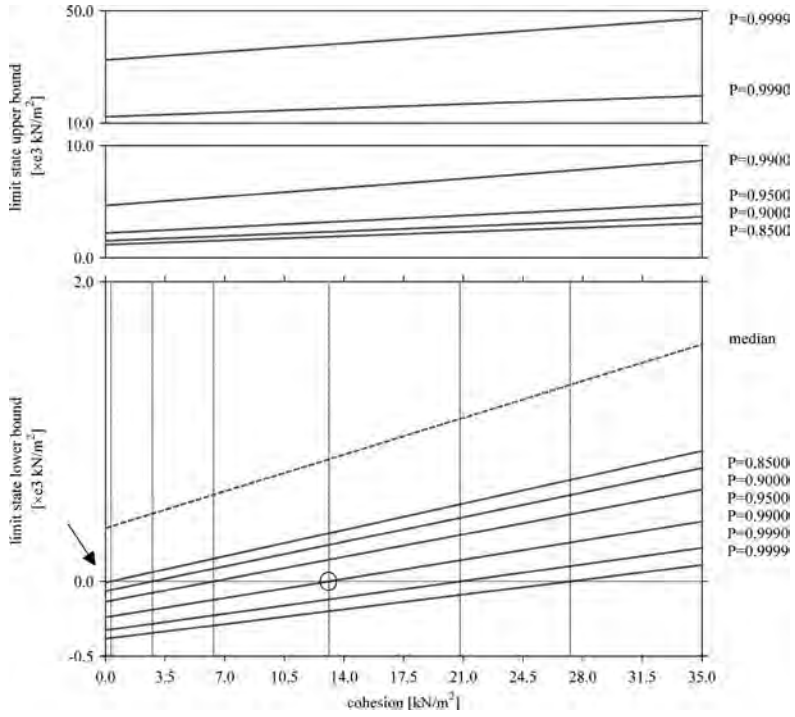


Figure 5. Limit state chart to safety assessment for the case cohesion interval scenario.

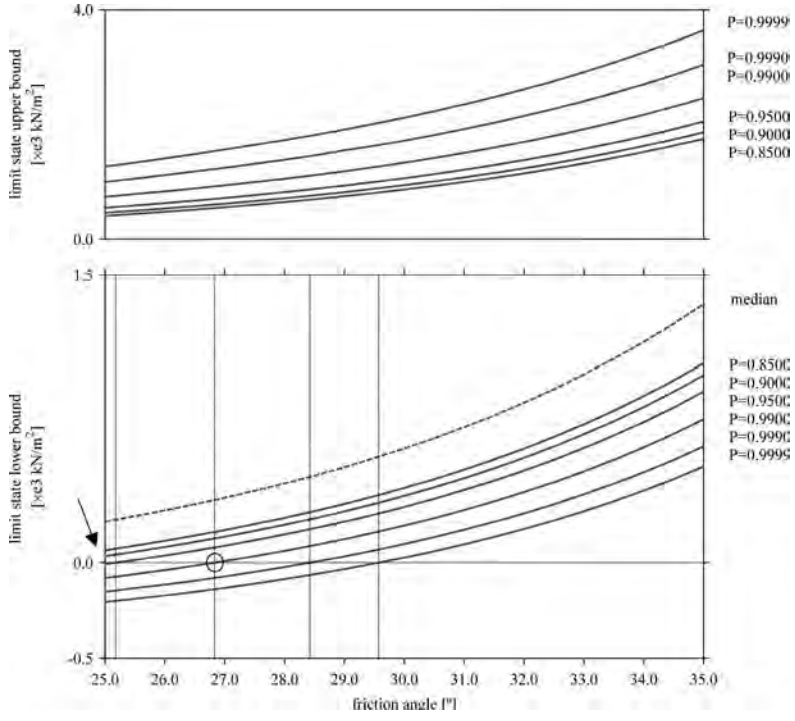


Figure 6. Limit state chart to safety assessment for the case friction angle interval scenario.

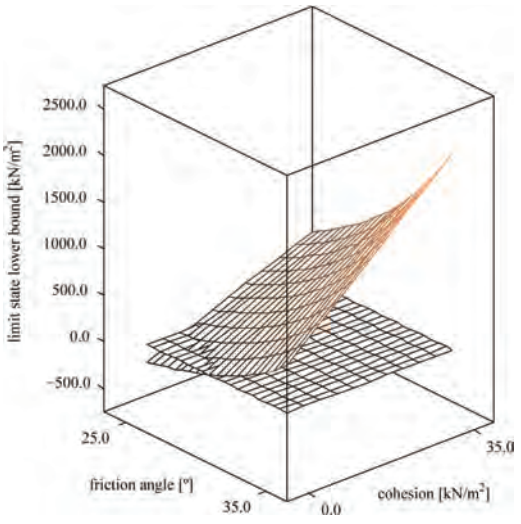


Figure 7. Limit state three-dimensional joint view to safety assessment for a 0.9900 probability level.

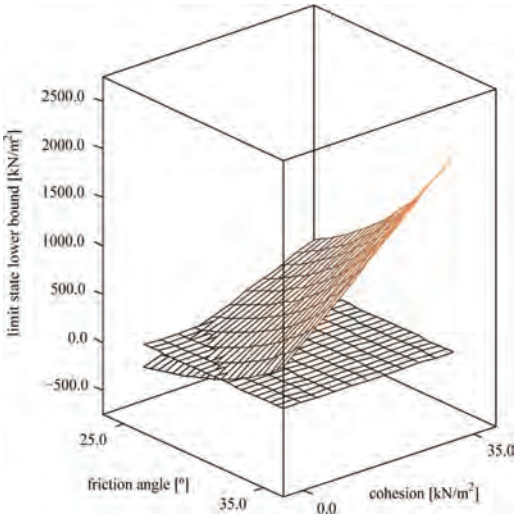


Figure 8. Limit state three-dimensional joint view to safety assessment for a 0.9990 probability level.

cases corresponding to a friction angle of 25.0º or to a cohesion of 0.0 kN/m^2, respectively. Considered the Figure 8, unsafe coordinates are limited to a cohesion parameter under a value between 10.5 kN/m^2 and 14.0 kN/m^2 or to a friction angle parameter under a value between 30.0º and 31.0º, considered the two worst combination cases corresponding to a friction angle of 25.0º or to a cohesion of 0.0 kN/m^2, respectively.

Unrealistic model assumptions are evinced by the limit state charts to safety assessment when higher levels of probability are considered, see the case cohesion interval scenario wherein the limit state upper bounds are obtained from unrealistic coordinates. Therefore, a critical evaluation from the interval scenario may influence the safety-based decision in ground investigation and testing or improvement. In addition, the interval uncertainty comprehends a probabilistic meaning for every combinatory possibility, see the case cohesion interval scenario wherein nonsatisfactory levels of reliability estimates are attained in a significant part of the cohesion interval. Another important issue is the position of the median trend which separates the fifty percent chance cases, here sketched on the safe side, noted that the linear and nonlinear trends evinced by the charts express the influence of the shear strength parameters on the calculation model for bearing capacity.

In the multivariate case the considered probability level prescribes a credible region in the hyperspace, then the imprecise interval analysis complies with a mixed set of probabilistic and nonprobabilistic interval models wherein different bounding measures may be applied in order to find the limit state lower and upper bounds in different scenarios. The proposed step by step optimisation algorithm for minimisation and maximisation further considers any important dependence relationships and whenever interval variables are involved proper constraints are required, see the case friction angle interval scenario. In fact, dependence is an important feature on the geotechnical engineering practice wherein design parameters encompass a physical meaning.

The summary description of reliability estimates in probabilistic scenario are finally provided in the next Table 6, wherein three methods are compared: the first order reliability method (FORM) and the second order reliability method (SORM), or the Monte Carlo simulation (MCS), the latter from counting among 5e6 trials. Regardless of the type of function and within acceptable margin of error, the global behaviour is approachable in every case.

Table 6. Summary description of model results for the reliability index and respective relative errors.

Reliability index β_{FORM}	Reliability index β_{SORM}	Reliability index β_{MCS}	β_{FORM} and β_{MCS} relative error (%)	β_{SORM} and β_{MCS} relative error (%)
3.3716	3.4392	3.4388	–1.9542	0.0102

MCS results from 1e6 simulations.

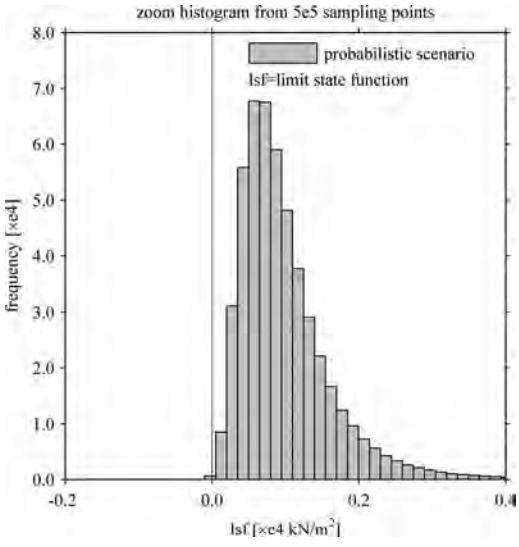


Figure 9. Limit state function histogram in probabilistic scenario.

A limit state function histogram in probabilistic scenario is further presented in the following Figure 9. The behaviour of the nonlinear bearing capacity model is then illustrated noted that the limit state function shows a nonnormal distribution under the probabilistic scenario characterised by the uncertain parameters and dependencies formerly described. Thorough observation reveals the presence of a small amount of data on the left of the zero limit state boundary, unsafe side. Thereby, failure may be considered a quite rare event among the considered 5e5 trials on nonsymmetric right skewed arrangement and with no single center.

5 CONCLUSION

The capabilities of nontraditional models for engineering computation under uncertainty have been under continuous review. Practical experience suggests that important dependence relationships should be considered on the proposed step by step optimisation algorithm for minimisation and maximisation. For demonstration, results are provided on the analysis of a strip spread foundation designed by the Eurocode 7 methodology, and precise versus imprecise probabilistic approaches are discussed. The limit state imprecise probabilistic analysis is interpreted altogether with a limit state imprecise interval analysis for bearing capacity. This sensitivity analysis may reveal unrealistic model assumptions and may provide meaningful results for safety-based decision in ground investigation and testing or improvement. Thus, the limit state charts to safety assessment are separately sketched for the cases cohesion and friction angle interval scenario wherein the random variables are bounded on different levels of probability.

The extension for high dimensional cases is as well considered through a limit state three-dimensional joint view to safety assessment, considered simultaneously the interval variables cohesion and friction angle. In the multivariate case the considered probability level prescribes a credible region in the hyperspace, then the imprecise interval analysis complies with a mixed set of probabilistic and nonprobabilistic interval models wherein different bounding measures may be applied in order to find the limit state lower and upper bounds in different scenarios. Thereby, the Eurocode 7 partial factor design is discussed on the basis of distinct levels of credibility. On this particular case study it is clearly shown that small variations in the friction angle input are very influential in that the median and the imprecise lower bound of probability correspond to a boundless immeasurable reliability. This conclusion is noticeable as well from the limit state chart to safety assessment for the case friction angle interval scenario.

REFERENCES

EN 1997-1 2004. Eurocode 7: geotechnical design-part 1: general rules. CEN.

Haim, Y.B. & Elishakoff, I. 1995. Discussion on: A non-probabilistic concept of reliability. *Structural Safety* 17(3): 195–199.

Vicig, P. & Seidenfeld, T. 2012. Bruno de Finetti and imprecision: imprecise probability does not exist! *International Journal of Approximate Reasoning* 53(8): 1115–1123.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Discussion on evaluation of probability bounds applied to the Eurocode 7

Sónia H. Marques
Doctor in Geotechnics and Senior Professional Civil Engineer of the Portuguese Institution of Engineers, Portugal

ABSTRACT: The underlying idea on the imprecise probability theory consists in modelling an imprecise probability distribution by a set of candidate probability distributions which are derived from the available data. The family is represented through a probability bounding approach applied to specify the lower and upper bounds of the imprecise probability distribution. A number of set-based uncertainty models derived from the probability bounding approach have been considered, namely the probability box structure. Designed from different approaches, probability boxes may differ meaningly from each other. A parametric approach may involve distributions with interval parameters or an envelope of competing probabilistic models. From search amid the number of candidate cumulative distribution functions the envelope of competing probabilistic models is expressed by a probability box function. In this way, a procedure for construction of a probability box structure by optimisation is advanced. Different dependencies may lead to quantitatively varied results so that a single scalar measure of a correlation coefficient may not be able to capture the complexity of the dependence model. Thus, the effects of correlation on the probability box structure may be comparatively considered. The technology is demonstrated on a synthetic exercise and on a design example referred to a strip spread foundation designed by the Eurocode 7.

1 INTRODUCTION

A probability bounding approach is hereafter applied to specify the lower and upper bounds of one imprecise probability distribution, see the review of Vicig & Seidenfeld (2012). By principle, variational correlation measures are insufficient to explore the possible nonlinear dependencies between variables in the form of a general relationship. In fact, the standard correlation concept is associated to a directional variation of paired random variables in that the word correlation is often reserved to be used only with random variables. In this way, a procedure for construction of a probability box structure by optimisation is advanced. In particular, the proposed procedure is aimed to consider how interval variables contained in a set between two endpoints may be related to other random variables characterised by a probability distribution. Regarding a sensitivity analysis, the effects of correlation on the probability box structure are comparatively demonstrated on a synthetic exercise, see the challenge problems in Oberkampf *et al.* (2004), and on a design example referred to a strip spread foundation designed by the Eurocode 7 (EN 1997-1 2004).

Considered that the parameter a follows a uniform distribution on the interval [0.1,1.0] and the parameter b follows a uniform distribution on the interval [0.0,1.0], Figure 1 represents comparatively the Kendall correlation pattern at 0.7 correlation coefficient for Archimedean parameterised copula families as Clayton or Frank or Gumbel, defined directly rather than being defined constructively from multivariate distributions. Considered the Pearson correlation pattern at 0.9 correlation coefficient, the equivalent Kendall rank correlation coefficient of 0.7 is determined at first to find the correspondent copula parameter α to launch the experiments accordingly by using a statistical toolbox. Afterwards, the Kendall correlation pattern at 0.7 correlation coefficient may be compared to the Pearson correlation pattern at 0.9 correlation coefficient, for the purpose see Figure 2. These approaches represent only a measure of the strength of the linear relationship between the two variables. It is further noted that the search for minimum and maximum values may reveal differences when compared the correlatedness and uncorrelatedness scenarios, particularly in the case of a higher correlation coefficient wherein the designed correlation pattern may be displayed as sharp at the endpoints of the intervals.

2 OPTIMISATION-BASED SYNTHETIC EXERCISE

The step by step optimisation algorithm is supported by transformations derived in compensative probability to express the performance function in the standard normal space of uncorrelated random and interval variables. Afterwards, the pair

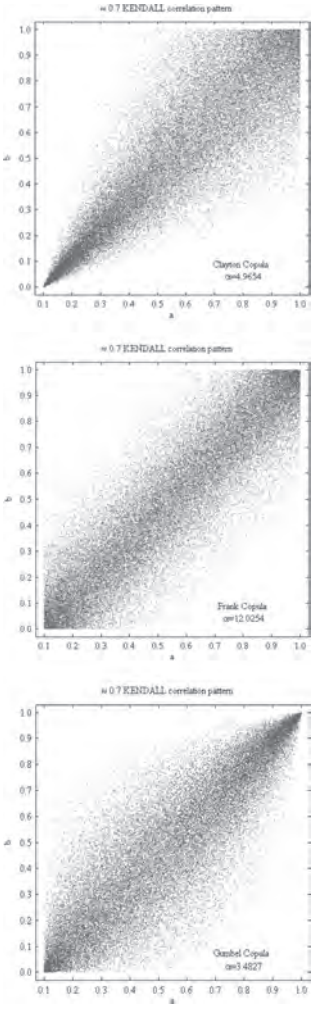


Figure 1. Kendall correlation pattern at 0.7 correlation coefficient, Clayton Copula and Frank Copula and Gumbel Copula.

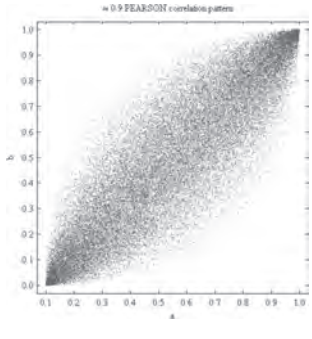


Figure 2. Pearson correlation pattern at 0.9 correlation coefficient.

minimum-maximum is estimated at a given central credibility level for every parametric combination across the indecision interval on the construction of the optimisation-based probability box structure:

[Step 1] Express the distributions and the statistics of basic input variables;
[Step 2] Express the equivalent standard normal correlation matrix;
[Step 3] Express the outcome matrix from the Cholesky decomposition of the equivalent standard normal correlation matrix;
[Step 4] Express the set of random and interval variables vector in the standard normal space of uncorrelated random and interval variables;
[Step 5] Express the performance function and the limit state in the standard normal space of uncorrelated random and interval variables;
[Step 6] Express the set of constraints and whenever required the set of guess values for initialisation of the suboptimisation algorithm;
[Step 7] Select the suboptimisation algorithm properly and run the iterative process in a multistart approach;
[Step 8] Estimate the pair minimum-maximum and the corresponding coordinates in the standard normal space of uncorrelated as well as in the original space of correlated random and interval variables.

The performance function in the standard normal space of uncorrelated random and interval variables is expressed by the following equivalent transformations $xi = f(yi)$ and $f(xi) = yi$ derived by Equation (1) in compensative probability:

$$xi = F_{Xi}^{-1}\left[\Phi(yi)\right] \leftrightarrow \Phi^{-1}\left[F_{Xi}(xi)\right] = yi \tag{1}$$

if xi = variable in the original space; yi = variable in the standard normal space; F_{Xi} = cumulative nonnormal distribution function; F_{Xi}^{-1} = inverse cumulative nonnormal distribution function; Φ = cumulative standard normal distribution function; and Φ^{-1} = inverse cumulative standard normal distribution function.

The correlation assignment in Equation (2) is also added to this standard normal space script:

$$[cyi^*] = [\mathrm{Ch}][uyi^*] \tag{2}$$

if cyi^* = set of variables vector in the standard normal space of correlated variables; uyi^* = set of variables vector in the standard normal space of uncorrelated variables; and ch = outcome matrix

from the Cholesky decomposition of the equivalent standard normal correlation matrix.

A considerable number of suboptimisation algorithms may be considered for the purpose but both local and global engines may fail under certain circumstances. A test function as the Peaks which is characterised by six minimum-maximum points is very convenient for the purpose of testing. In particular, conjugate gradient and quasi newton optimisation algorithms in a multistart approach are successful on the search for the six minimum-maximum points of the Peaks test function.

For demonstration, the proposed procedure is then applied on the synthetic exercise expressed by Equation (3) by using conjugate gradient and quasi newton optimisation algorithms in a multistart approach:

$$y = (a + b)^a \tag{3}$$

where the parameter a belongs to the interval [0.1,1.0] and the parameter b follows a uniform distribution on the interval [0.0,1.0].

Different degrees of correlation are considered to evince the effect of the parameter b on the interval for the parameter a. A sequence of transformations is exemplified next for the optimisation procedure by the group of Equations (5) and (7), respectively developed for the case minimisation to a 0.1 and 0.9 correlation coefficient at central credibility level 5σ and a = 1, wherein the group of Equations (4) and (6) presents the corresponding equivalent standard normal correlation matrix R and the outcome matrix from the Cholesky decomposition of the equivalent standard normal correlation matrix Ch:

Given

$$R = \begin{vmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{vmatrix} \quad \text{and} \quad Ch = \begin{vmatrix} 1.0000 & 0.0000 \\ 0.1000 & 0.9950 \end{vmatrix} \tag{4}$$

$$L(a,b) = (a + b)^a$$

$$L(ya,yb) = \left(F_U^{-1}\left[\Phi(1.0000 \cdot ya + 0.0000 \cdot yb)\right] + F_U^{-1}\left[\Phi(0.1000 \cdot ya + 0.9950 \cdot yb)\right]\right)^{F_U^{-1}\left[\Phi(1.0000 \cdot ya + 0.0000 \cdot yb)\right]}$$

$$-5 \le ya \le 5$$
$$-5 \le yb \le 5$$
$$a = F_U^{-1}\left[\Phi(1.0000 \cdot ya + 0.0000 \cdot yb)\right] = 1$$

$$M = \text{Minimise}(L, ya, yb)$$

$$ya = 5$$
$$yb = -5$$
$$a = F_U^{-1}\left[\Phi(1.0000 \cdot ya + 0.0000 \cdot yb)\right] = 1$$
$$b = F_U^{-1}\left[\Phi(0.1000 \cdot ya + 0.9950 \cdot yb)\right] = 3.8206e-6$$

$$L(ya,yb) = L(a,b) = 1.0000 \tag{5}$$

Given

$$R = \begin{vmatrix} 1.0 & 0.9 \\ 0.9 & 1.0 \end{vmatrix} \quad \text{and} \quad Ch = \begin{vmatrix} 1.0000 & 0.0000 \\ 0.9000 & 0.4359 \end{vmatrix} \tag{6}$$

$$L(a,b) = (a + b)^a$$

$$L(ya,yb) = \left(F_U^{-1}\left[\Phi(1.0000 \cdot ya + 0.0000 \cdot yb)\right] + F_U^{-1}\left[\Phi(0.9000 \cdot ya + 0.4359 \cdot yb)\right]\right)^{F_U^{-1}\left[\Phi(1.0000 \cdot ya + 0.0000 \cdot yb)\right]}$$

$$-5 \le ya \le 5$$
$$-5 \le yb \le 5$$
$$a = F_U^{-1}\left[\Phi(1.0000 \cdot ya + 0.0000 \cdot yb)\right] = 1$$

$$M = \text{Minimise}(L, ya, yb)$$

$$ya = 5$$
$$yb = -5$$
$$a = F_U^{-1}\left[\Phi(1.0000 \cdot ya + 0.0000 \cdot yb)\right] = 1$$
$$b = F_U^{-1}\left[\Phi(0.9000 \cdot ya + 0.4359 \cdot yb)\right] = 0.9898$$

$$L(ya,yb) = L(a,b) = 1.9898 \tag{7}$$

if ya = corresponding parameter a coordinates in the standard normal space; yb = corresponding parameter b coordinates in the standard normal space; F_U^{-1} = inverser cumulative uniform distribution function; and Φ = cumulative standard normal distribution function; noted σ as a unitary standard deviation in the standard normal space.

The following graphs are then designed at 0.1 correlation coefficient and at 0.9 correlation coefficient. Figure 3 and Figure 5 present the minimum and mean and maximum trends across the interval [0.1,1.0] for parameter a. Figure 4 and Figure 6 present the corresponding optimisation-based probability box structure. For appropriate representation of every sketched line, it is selected a number of 181 discretisation points uniformly distributed on the interval [0.1,1.0] for parameter a.

The optimisation-based probability box structure is constructed comparatively to the mean empirical cumulative distribution function, noted that the adjacent empirical functions corresponding to the collection of minimum and maximum values are sketched at a central credibility level 5σ. From observation of Figure 3 and Figure 5 it is noted that the position of the lowest minimum occurs on the inner part of the interval [0.1,1.0] for parameter a, with the position of the highest maximum on the coordinate 1.0. From observation of Figure 4 and Figure 6 it is noted that the gradual

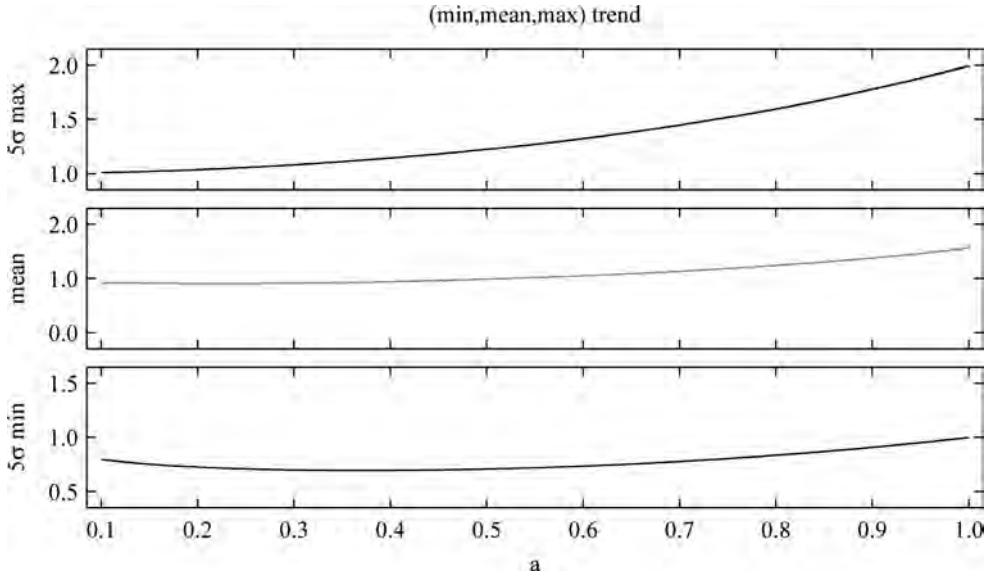


Figure 3. Trends for the synthetic exercise at 0.1 correlation coefficient.

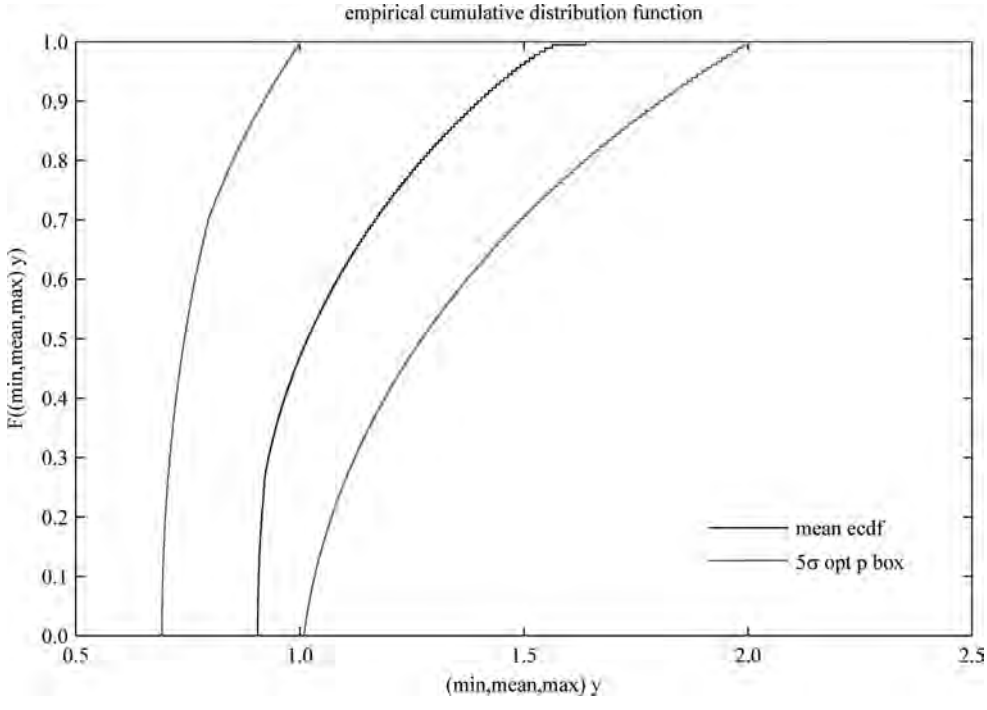


Figure 4. Optimisation-based probability box structure for the synthetic exercise at 0.1 correlation coefficient.

convergence of the major probability lines occurs whenever the correlation coefficient increases, with exception to the zone near the zero probability level on the minimum branch of the optimisation-based probability box structure at 0.9 correlation coefficient. This particular feature is observable because the lowest minimum occurs in the inner part of the interval [0.1,1.0] for parameter a.

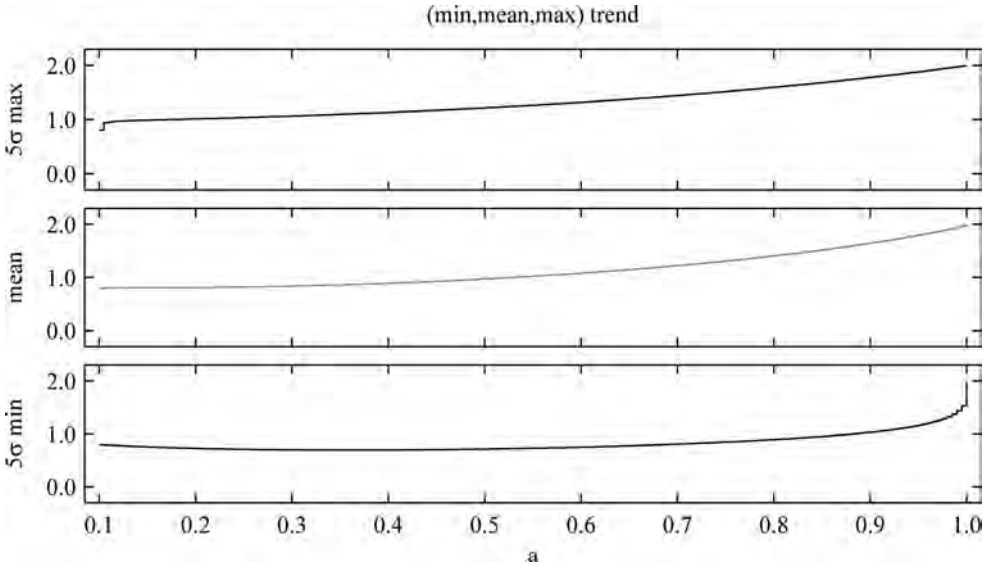


Figure 5. Trends for the synthetic exercise at 0.9 correlation coefficient.

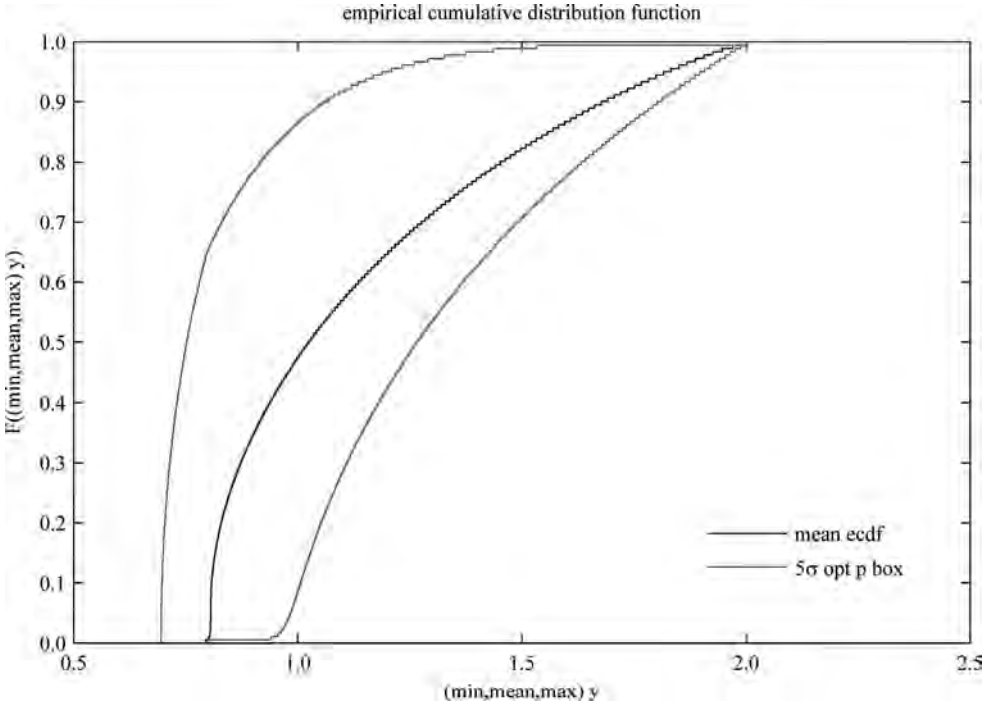


Figure 6. Optimisation-based probability box structure for the synthetic exercise at 0.9 correlation coefficient.

3 DESIGN EXAMPLE

The design example is referred to the strip spread foundation on a relatively homogeneous soil shown in Figure 7, wherein groundwater level is away. Considered the vertical noneccentric loading problem and the calculation model for bearing capacity, the performance function

may be described by the simplified Equation (8):

$$M = f\left(B, D, \gamma_s, c_f, \varphi_f, \gamma_f, P, Q\right) \quad (8)$$

if B is the foundation width; D is the soil height above the foundation base; γ_s is the unit weight of the soil above the foundation base; c_f is the cohesion of the foundation soil; φ_f is the friction angle of the foundation soil; γ_f is the unit weight of the foundation soil; P is the dead load; and Q is the live load.

Table 1 summarises the description of basic input variables, with different types of distributions. The considered correlation coefficients between the basic input variables are either presented in Table 2.

The strip spread foundation is designed by the Eurocode 7 methodology, Design Approach DA.2*, wherein partial factors are coupled with characteristic values.

A cautious estimate of the 95% reliable mean value for a known coefficient of variation is considered for each geotechnical parameter regarding the occurrence of a limit state. Scenarios wherein the parameters cohesion and friction angle of the foundation soil are separately implemented as intervals are further considered. The proposed procedures are then applied on the design example. The optimisation-based probability box structure is drawn by using conjugate gradient and quasi newton optimisation algorithms in a multistart approach for the case cohesion interval scenario and for the case friction angle interval scenario, represented respectively in Figure 8 and Figure 9. Considered the case cohesion interval scenario, for appropriate representation of the minimum and maximum branches, it is selected a number of 351 discretisation points uniformly distributed on the interval [0.0,35.0] kN/m^2. Considered the case friction angle interval scenario, for appropriate representation of the minimum and maximum branches, it is selected a number of 201 discretisation points uniformly distributed on the interval [25.0,35.0] °.

The optimisation-based probability box structure is constructed comparatively to the median empirical cumulative distribution function, noted that the adjacent empirical functions corresponding to the collection of minimum and maximum values are sketched at a central credibility level 3σ. The median trend is comparatively drawn from a moderate number of discretisation points uniformly distributed on the interval. From observation of Figure 8 and Figure 9 it is noted that the linear and nonlinear trends evinced by the graphs express the influence of the parameters cohesion and friction angle of the foundation soil on the calculation model for bearing capacity. The position of the median trend, which separates the fifty percent chance cases, is further sketched on the safe

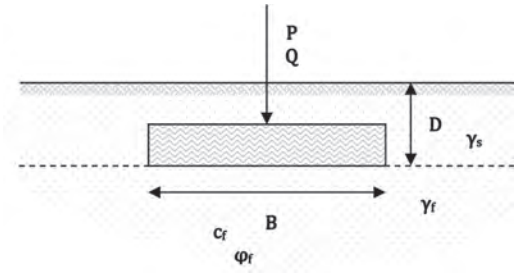


Figure 7. Strip spread foundation.

Table 1. Summary description of basic input variables.

Basic input variables	Distributions	Mean value	Coefficient of variation
B (m)	Deterministic	1.30	0.00
D (m)	Deterministic	1.00	0.00
γ_s (kN/m^3)	Normal	16.80	0.05
c_f (kN/m^2)	Lognormal	14.00	0.40
	Interval*	…	…
φ_f (°)	Lognormal	32.00	0.10
	Interval*	…	…
γ_f (kN/m^3)	Normal	17.80	0.05
P (kN/m)	Normal	370.00	0.10
Q (kN/m)	Normal	70.00	0.25

*Cases cohesion interval scenario [0.00,35.00] and friction angle interval scenario [25.00,35.00].

Table 2. Correlation coefficients between the basic input variables.

Correlation matrix					
ρ_{x1x1}	ρ_{x1x2}	ρ_{x1x3}	ρ_{x1x4}	ρ_{x1x5}	ρ_{x1x6}
ρ_{x2x1}	ρ_{x2x2}	ρ_{x2x3}	ρ_{x2x4}	ρ_{x2x5}	ρ_{x2x6}
ρ_{x3x1}	ρ_{x3x2}	ρ_{x3x3}	ρ_{x3x4}	ρ_{x3x5}	ρ_{x3x6}
ρ_{x4x1}	ρ_{x4x2}	ρ_{x4x3}	ρ_{x4x4}	ρ_{x4x5}	ρ_{x4x6}
ρ_{x5x1}	ρ_{x5x2}	ρ_{x5x3}	ρ_{x5x4}	ρ_{x5x5}	ρ_{x5x6}
ρ_{x6x1}	ρ_{x6x2}	ρ_{x6x3}	ρ_{x6x4}	ρ_{x6x5}	ρ_{x6x6}
1.0	0.0	0.5	0.9	0.0	0.0
0.0	1.0	0.0	0.0	0.0	0.0
0.5	0.0	1.0	0.5	0.0	0.0
0.9	0.0	0.5	1.0	0.0	0.0
0.0	0.0	0.0	0.0	1.0	0.0
0.0	0.0	0.0	0.0	0.0	1.0

x_1-γ_s; x_2-c_f; x_3-φ_f; x_4-γ_f; x_5-P; x_6-Q; ρ-coefficient of correlation.

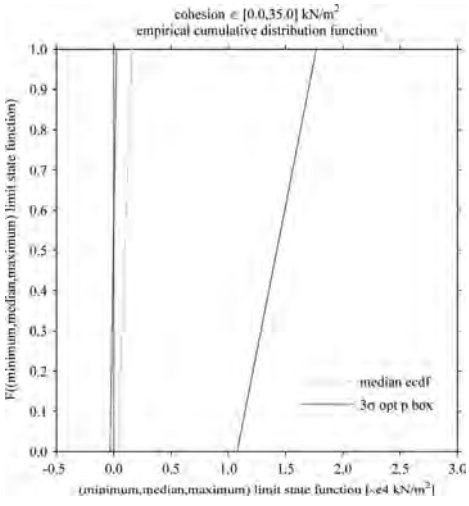


Figure 8. Optimisation-based probability box structure for the case cohesion interval scenario.

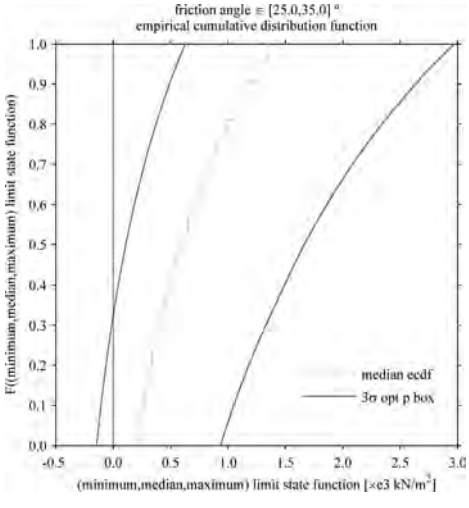


Figure 9. Optimisation-based probability box structure for the case friction angle interval scenario.

side. Regarding the limit state, the probability of no failure across the cohesion and friction angle interval scenario at a central credibility level 3σ may be determined, see the probability level for a positive outcome on the minimum branch.

4 CONCLUSION

A procedure for construction of a probability box structure by optimisation is advanced. Different dependencies may lead to quantitatively varied results and as the degree of correlation may be unknown, a single scalar measure of a correlation coefficient may not be able to capture the complexity of the dependence model. Thus, the effects of correlation on the probability box structure are comparatively demonstrated on a synthetic exercise and on a design example referred to a strip spread foundation designed by the Eurocode 7. The optimisation-based probability box structure opens a new path in the framework of engineering limit state design under dependence in order to consider the failure analysis on a number of different central credibility levels.

REFERENCES

EN 1997-1 2004. *Eurocode 7: geotechnical design-part 1: general rules.* CEN.

Oberkampf, W.L., Helton, J.C., Joslyn, C.A., Wojtkiewicz, S.F. & Ferson, S. 2004. Challenge problems: uncertainty in system response given uncertain parameters. *Reliability Engineering & System Safety* 85(1–3): 11–19.

Vicig, P. & Seidenfeld, T. 2012. Bruno de Finetti and imprecision: imprecise probability does not exist! *International Journal of Approximate Reasoning* 53(8): 1115–1123.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Reliability assessment model of technical object in aspect of catastrophic damage in the form of jamming—an outline

M. Zieja
Air Force Institute of Technology, Warsaw, Poland

M. Jasztal, S. Stępień & M. Ważny
Military University of Technology, Warsaw, Poland

ABSTRACT: This elaboration is an attempt to present a reliability assessment model of a technical object in the aspect of a catastrophic damage. As the technical object a mechanical device was adopted, in which, between operating elements a jamming may occur and as a diagnostic parameter—the operating time of device t. Determining the distribution of device operating time until jamming appeared was based on the Yule's process modified by Gercbach and Kordoński. Due to the adopted simplifications in recording of the above-mentioned model, the process of defining final equations was identified. This process, in a discrete system, depends on the system of equations typical for the accrual process of the diagnostic parameter. By performing a transformation of these equations the formula was determined for reliability of the technical object and density function of the object operating time until the catastrophic damage is found in the form of jamming.

1 INTRODUCTION

A characteristic feature of mechanical devices is the occurrence of movable mating elements creating a system of kinematic pairs. Depending on the intensity of using the device and the conditions in which work of the technical object is being performed, it might lead to wear and, in consequence, e.g. to jamming.

Due to device operation, the clearances between mating elements or resultants in kinematic sequences are changed and deviations are created from normal values, which harmfully affect the automatics operation. The values of deviations of clearances from normal values trigger a change in operating conditions of the technical object, what has a substantial impact on their operation reliability. Sometimes, interferences in automatics of operation of the technical object due to the change in operating conditions along with the increased value of clearances contribute to the formation of jamming, that is a kind of the catastrophic damage (Baranowski & Małachowski 2015, Idziaszek & Grzesik 2014, Tomaszek et al. 2016)

By making an analysis of the device technical condition an essential issue is the choice of a parameter, which will be leading (forecasting) in the assessment under consideration. It, therefore, appears to be appropriate to adopt that the parameter pursuant to which the possibility of occurrence of the catastrophic damage (jamming) is evaluated, will be a number of device completed working cycles. With the increase in the number of completed cycles, the device experiences wearing processes (destructive), which exert a considerable influence on the quality of the device's usage conditions (Dhillon 1999, Tomaszek et al 2013, Tomaszek et al. 2011, Ważny 2009).

2 DETERMINING THE DISTRIBUTION OF THE NUMBER OF THE DEVICE COMPLETED WORKING CYCLES UNTIL DAMAGE APPEARS (JAMMING)

For the diagnostic parameter (forecasting) according to which we will determine the change in technical condition of a device (wearing) the number of cycles completed by the device was assumed. This parameter is a speed function of completing the working cycles by the device and will be defined by equation (1):

$$z = v \cdot t \tag{1}$$

where v = speed of working cycles completed by the device; and t = device operating time.

In the function of the number of cycles completed by the device, the effects of destructive processes continue to increase in the form of wearing and clearances in kinematic chains. As there are more effects related to wearing of components, the chance that the device will jam also increase. Therefore, it might be concluded that with the increase of working cycles completed by the device, there is a change of its technical condition and increases a chance of the occurrence of a damage in the form of jamming. This suggests the possibility of using certain models to describe the risk of the occurrence of a damage in the form of jamming.

To determine the distribution of the number of completed cycles until jamming is found the Yule's process was applied by modifying it. The outline of this modification is provided in work (Gercbach & Kordoński 1968) without a detailed way of acquiring final results. Utilizing the results mentioned in this work (Gercbach & Kordoński 1968) needs supplementing the way they are presented with indirect operations enabling to derive final equations.

To define regularities of this model it is convenient to use discrete values of diagnostic parameter. Way of discretization is shown in Fig. 1 where E_k = discrete values of a diagnostic parameter defined as parameter value states; $\lambda\Delta t$ = probability of transition from E_k to E_{k+1} in a time interval with the length of Δt; λ = intensity of parameter's state changing described as

$$\lambda = \frac{P}{\Delta t}, \tag{2}$$

where P = probability of completion of one working cycle in a time interval with the length of Δt; $q_k(t)$ = probability of interrupting the development of the process of increasing a diagnostic parameter, depending on the process state; μ = increment of intensity of jamming occurrence along with the increase of device working cycle; μ_o = intensity of jamming for initial state (for t = 0); h = average value of the increment of a diagnostic parameter in time Δt; and t = device operating time.

Providing that $P_k(t)$ signifies the probability that for device operating time equaling t a diagnostic parameter takes the value $E_k (k = 0,1,2,\ldots)$. For the adopted arrangements, using the postulates of the Poisson process, a system of equations characterizing the accrual process of a diagnostic parameter may be obtained (DeLurgio 1998, Pham 2006, Werbińska & Zając 2015, Zio 2009).

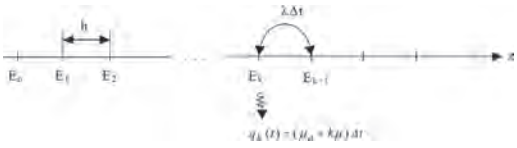


Figure 1. Way of discretization of a diagnostic parameter.

For k = 0:

$$P_0(t+\Delta t) = (1-\lambda\Delta t)(1-\mu_0\Delta t)P_0(t)$$
$$P_0(t+\Delta t) = (1-\mu_0\Delta t - \lambda\Delta t + \lambda\mu_0\Delta t^2)P_0(t).$$

After omitting higher order negligibles the formula takes the following form:

$$P_0(t+\Delta t) = \left(1-(\lambda+\mu_0)\Delta t\right)P_0(t) \tag{3}$$

For k = 1, 2, …

$$P_k(t+\Delta t) = (1-\lambda\Delta t)\left(1-(\mu_0+k\mu)\Delta t\right)P_k(t) + \\ + \lambda\Delta t\left(1-(\mu_0+k\mu)\Delta t\right)P_{k-1}(t);$$
$$P_k(t+\Delta t) = \left(\begin{array}{l} 1-(\mu_0+k\mu)\Delta t - \\ \lambda\Delta t + \lambda\Delta t(\mu_0+k\mu)\Delta t \end{array}\right)P_k(t) + \\ +\left(\lambda\Delta t - \lambda\Delta t(\mu_0+k\mu)\Delta t\right)P_{k-1}(t).$$

Again by omitting higher order negligibles the above-mentioned formula takes the following form:

$$P_k(t+\Delta t) = \left(1-(\lambda+\mu_0+k\mu)\Delta t\right)P_k(t) + \lambda\Delta t\ P_{k-1}(t). \tag{4}$$

Thus, we obtain the following system of equations:

$$\left.\begin{array}{l} P_0(t+\Delta t) = \left(1-(\mu_0+\lambda)\Delta t\right)P_0(t) + O(\Delta t) \\ P_k(t+\Delta t) = \left(1-(\mu_0+k\mu+\lambda)\Delta t\right)P_k(t) + \\ +\lambda\Delta t\ P_{k-1}(t) + O(\Delta t), \quad \text{for k} = 1, 2, \ldots \end{array}\right\}. \tag{5}$$

From system of equations (5) after transformation and movement to the limit from $\Delta t \to 0$ the following system of equations is received:

$$\left.\begin{array}{l} P_0'(t) = -(\mu_0+\lambda)P_0(t) \\ P_k'(t) = -(\mu_0+k\mu+\lambda)P_k(t) + \lambda\ P_{k-1}(t), \\ \qquad \text{for k} = 1, 2, \ldots \end{array}\right\}. \tag{6}$$

The initial condition for all of these equations may be written in the following form:

$$P_i(0) = \begin{cases} 1 & for \quad i = 0 \\ 0 & for \quad i \neq 0 \end{cases}. \tag{7}$$

System of equations (6) is solved by the use of a recursive method.

Solution for $k = 0$:

$$P_0'(t) = -(\mu_0 + \lambda) P_0(t),$$
$$\int_0^t \frac{P_0'(t)}{P_0(t)} dt = -\int_0^t (\mu_0 + \lambda) dt.$$

Therefore,

$$P_0(t) = -C_0 e^{-(\mu_0+\lambda)t}. \quad (8)$$

For $t = 0$, $P_0(0) = 1$ thus, $C_0 = 1$.

Solution for k = 1, 2, 3… can be determined in the following way. A differential equation takes the form:

$$P_k'(t) = -(\mu_0 + k\mu + \lambda) P_k(t) + \lambda P_{k-1}(t) \quad (9)$$

In this case, we predict the solution in the form of:

$$P_k(t) = C_k(t)\ e^{-(\mu_0+\lambda)t}. \quad (10)$$

A derivative of relationship (10) assumes the following form:

$$P_k'(t) = C_k'(t) e^{-(\mu_o+\lambda)t} + C_k(t)\left(-(\lambda + \mu_o)\right) e^{-(\mu_o+\lambda)t}. \quad (11)$$

Substituting the above equation for equation (9) the following formula was obtained:

$$C_k'(t)\ e^{-(\mu_o+\lambda)t} - (\lambda + \mu_o) C_k(t)\, e^{-(\mu_o+\lambda)t} = -(\mu_o + k\mu + \lambda) \underbrace{P_k(t)}_{C_k(t)\ e^{-(\mu_o+\lambda)t}} + \lambda \underbrace{P_{k-1}(t)}_{C_{k-1}(t)\ e^{-(\mu_o+\lambda)t}}.$$

Thus, we obtain an equation:

$$C_k'(t) + k\mu C_k(t) = \lambda C_{k-1}(t) \quad (12)$$

Equation (12) for *k=1* will be as follows:

$$C_1'(t) + \mu C_1(t) = \lambda. \quad (13)$$

Differential equation (13) takes the following general form:

$$y' + P(x)y = Q(x).$$

The solution of which is the following equation:

$$y = e^{-\int_0^t P dx} \left(\int_0^t Q\, e^{\int_0^t P dx} dx \right). \quad (14)$$

Using formula (14) the solution of formula (13) may be written in the following form:

$$C_1(t) = e^{-\int_0^t \mu dt} \left(\int_0^t \lambda\, e^{\int_0^t \mu dt} dt \right) = e^{-\mu t} \left(\int_0^t \lambda\, e^{\mu t}\, dt \right)$$
$$= e^{-\mu t}\ \lambda \frac{1}{\mu} e^{\mu t} \Big|_0^t = e^{-\mu t} \frac{\lambda}{\mu} \left(e^{\mu t} - 1 \right) = \frac{\lambda}{\mu} - \frac{\lambda}{\mu} e^{-\mu t}. \quad (15)$$

For *k = 2* equation (13) assumes the form:

$$C_2'(t) + 2\mu C_2(t) = \lambda \left(\frac{\lambda}{\mu} - \frac{\lambda}{\mu}\ e^{-\mu t} \right). \quad (16)$$

Solution of equation (16):

$$\begin{aligned} C_2(t) &= e^{-\int_0^t 2\mu t} \int_0^t \left(\frac{\lambda^2}{\mu} - \frac{\lambda^2}{\mu}\ e^{-\mu t} \right) e^{\int_0^t 2\mu dt} dt \\ &= e^{-2\mu t} \left(\frac{\lambda^2}{\mu} \frac{1}{2\mu} e^{2\mu t} - \frac{\lambda^2}{\mu} \frac{1}{\mu} e^{\mu t} \right) \Big|_0^t \\ &= e^{-2\mu t} \left(\frac{\lambda^2}{2\mu^2} e^{2\mu t} - \frac{\lambda^2}{2\mu^2} - \frac{\lambda^2}{\mu^2} e^{\mu t} + \frac{\lambda^2}{\mu^2} \right) \\ &= e^{-2\mu t} \left(\frac{\lambda^2}{2\mu^2} e^{2\mu t} + \frac{2\lambda^2 - \lambda^2}{2\mu^2} - \frac{\lambda^2}{\mu^2} e^{\mu t} \right) \\ &= \frac{\lambda^2}{2\mu^2} + \frac{\lambda^2}{2\mu^2} e^{-2\mu t} - \frac{\lambda^2}{\mu^2} e^{-\mu t} \\ &= \frac{\lambda^2}{2\mu^2} \left(1 + e^{-2\mu t} \right) - \frac{\lambda^2}{\mu^2} e^{-\mu t}. \end{aligned} \quad (17)$$

Equation describing function $C_2(t)$ was converted to the form, which suggests the general form of this function:

$$C_2(t) = \frac{\lambda^2}{2\mu^2} + \frac{\lambda^2}{2\mu^2} e^{-2\mu t} - \frac{\lambda^2}{\mu^2} e^{-\mu t} \Big| \cdot 2$$
$$2C_2(t) = \frac{\lambda^2}{\mu^2} - \frac{2\lambda^2}{\mu^2} e^{-\mu t} + \frac{\lambda^2}{\mu^2} e^{-2\mu t}.$$

Thus:

$$C_2(t) = \left(\frac{\lambda}{\mu} - \frac{\lambda}{\mu}\ e^{-\mu t} \right)^2 \cdot \frac{1}{2}. \quad (18)$$

Form of equation (18) allows to predict the general form of this function. This equation assumes the following form:

$$C_k(t) = \frac{1}{k!} \left(\frac{\lambda}{\mu} - \frac{\lambda}{\mu}\ e^{-\mu t} \right)^k. \quad (19)$$

Applying (19) the solution of equation (9) may be written. This solution adopts the following form:

$$P_k(t) = \frac{1}{k!}\left(\frac{\lambda}{\mu} - \frac{\lambda}{\mu}\ e^{-\mu t}\right)^k e^{-(\mu_0+\lambda)t}, \tag{20}$$

for $k = 1, 2, 3 \ldots$

Using equation (8) and (20) it is possible to determine reliability of the device. Therefore:

$$\begin{aligned} R(t) &= \sum_{k=0}^{\infty} P_k(t) \\ R(t) &= \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{\lambda}{\mu} - \frac{\lambda}{\mu}\ e^{-\mu t}\right)^k e^{-(\mu_0+\lambda)t}. \end{aligned} \tag{21}$$

It should be noted that the following equation is true:

$$\sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{\lambda}{\mu} - \frac{\lambda}{\mu}\ e^{-\mu t}\right)^k = e^{\frac{\lambda}{\mu} - \frac{\lambda}{\mu} e^{-\mu t}}. \tag{22}$$

Using equation (22) the formula for reliability takes the following form:

$$R(t) = e^{\frac{\lambda}{\mu} - \frac{\lambda}{\mu} e^{-\mu t}} \cdot e^{-(\mu_0+\lambda)t}.$$

Thus,

$$R(t) = e^{\frac{\lambda}{\mu}\left(1 - e^{-\mu t}\right) - (\mu_0+\lambda)t}. \tag{23}$$

Based on the above-mentioned equation the probability of jamming of the device for the number of completed cycles in time t will equal:

$$Q(t) = 1 - e^{\frac{\lambda}{\mu}\left(1 - e^{-\mu t}\right) - (\mu_0+\lambda)t}. \tag{24}$$

Distribution of the number of working cycles completed by the device until its jamming will be as follows:

$$f(t) = \frac{d}{dt} Q(t).$$

Thus,

$$f(t) = \left(\mu_0 + \lambda\left(1 - e^{-\mu t}\right)\right) e^{\frac{\lambda}{\mu}\left(1 - e^{-\mu t}\right) - (\mu_0+\lambda)t}. \tag{25}$$

The form of relation (25) presents the density function of working cycles completed by the device until catastrophic damage is found in the form of jamming. This distribution can be used to establish the risk of the occurrence of a catastrophic failure in the form of jamming.

3 CALCULATION EXAMPLE

The remarks as regards establishing distribution parameters λ, μ, μ_o shall also be given. The parameter λ will be evaluated by using the number of working cycles completed by the device. Calculation formula of parameter λ will be as follows (Ważny 2015):

$$\lambda^* = \frac{1}{\Delta t^*} \tag{26}$$

where Δt^* = the duration of one device working cycle.

For determining μ_0^* *i* μ^* data are used from observation of the device operating process due to which we obtain for N of devices the list of the number of cycles completed by the device until jamming appears. Hence, we acquire t_1, t_2,…, t_N. These data are used to build a histogram. The value μ_0^* is the ordinate of the histogram in point $t = 0$. Estimation value shall be established from the following expression:

$$\frac{1}{\lambda^*}\left(\frac{n(t)}{N} - 1\right) + t = \frac{1}{\mu^*}\left(1 - e^{-\mu^* t}\right) \tag{27}$$

where $n(t)$ = the number of efficient devices in time $t < t_N$; and t = device operating time (agreed value).

The left side of the equation for an agreed value t is changeable. A value μ^* is the value, which guarantees that the right side of the equation is equal to the left side of the equation in equation (27). For the purpose of estimating a jamming intensity of the device $\gamma(t)$ the following equation shall be used:

$$\gamma(t) = \frac{f(t)}{R(t)}. \tag{28}$$

Replacing equation (25) and (23) with the formula (28), equation (29) was obtained:

$$\gamma(t) = \left(\mu_0 + \lambda\left(1 - e^{-\mu t}\right)\right). \tag{29}$$

As it follows from equation (29) the probability of device's jamming increases with the increase of the number of working cycles completed by the device.

By attempting to perform the number verification of the presented model one adopted technical data characterizing a hypothetical technical object.

Table 1. Calculation results.

z	Δt^*	λ^*	$n(t)$	N	μ_0^*	μ^*
850	0,071	14,197	36	36	0	$1*10^{-9,5}$

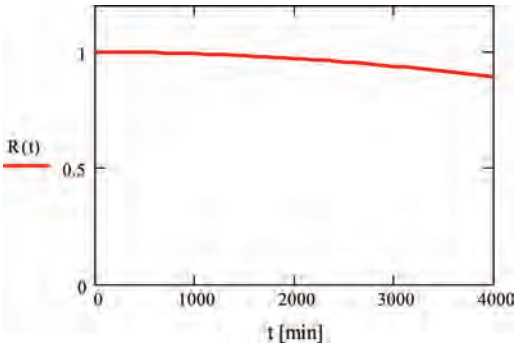


Figure 2. Graphic form of reliability function $R(t)$.

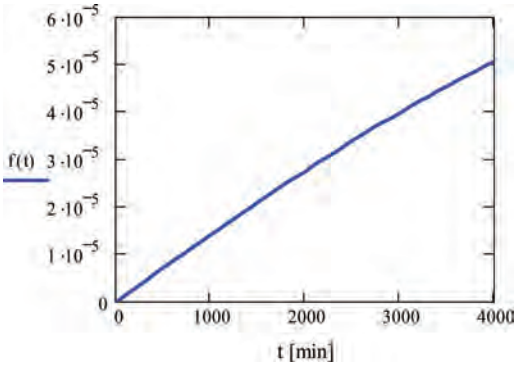


Figure 3. Graphic form of density function $f(t)$.

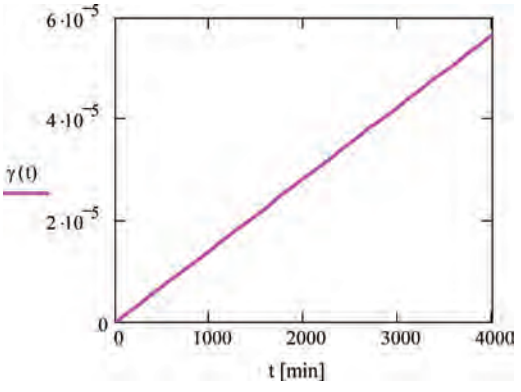


Figure 4. Graphic form of jamming intensity function $\gamma(t)$.

Input data used for calculations and obtained results were summarized in Table 1 where z is the number of working cycles completed by the device.

For the above data the characteristics $R(t)$, $f(t)$ and $\gamma(t)$ were established.

4 FINAL REMARKS

In a diagnostic parameter function there is an accrual process of the impacts of destructive processes such as surface wear, corrosion and other factors, what in consequence contributes to accruing of negative effects leading to the occurrence of the catastrophic damage, among others in the form of jamming, crack and other similar events of the assemblies of technical objects (devices).

The presented method of the assessment of device operation reliability in the aspect of the formation of the catastrophic damage in the form of jamming appears to be justified and correct. The presented calculation example enabled to perform number verification of a developed model and showed the application nature of the established method. Applying parameters determined on the basis of the described method it is possible to present the characteristics of technical properties of the object under consideration. In addition, the developed model can also be applied to estimate the reliability of other technical objects in the aspect of occurring sudden damage originating from relaxation stimuli.

REFERENCES

Baranowski, P. & Małachowski, J. 2015. Numerical study of selected military vehicle chassis subjected to blast loading in terms of tire strength improving. *Bulletin of The Polish Academy of Sciences: Technical Sciences*, 63 (4): 867–878.

Casciati, F. & Roberts, B. 1996. *Mathematical Models for Structural Reliability Analysis*. Boca Raton/New York/London/Tokyo: CRC Press.

DeLurgio, S.A. 1998. *Forecasting principles and applications*. University of Missouri-Kansas City: Irwin/McGraw-Hill.

Dhillon, B.S. 1999. *Design Reliability. Fundamentals and Applications*. Ottawa: Boca Raton/New York/London/Washington: CRC Press.

Gercbach, I.B. & Kordoński C.B. 1968, *Reliability models of technical objects*, Warsaw: Science and Technology Publishers.

Idziaszek, Z. & Grzesik, N. 2014. Object characteristics deterioration effect on task realizability – outline method of estimation and prognosis. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 3: 433–440.

Jacyna- Gołda, I. & Izdebski, M. & Podviezko, A. 2017. Assessment of efficiency of assignment of vehicles to

tasks in supply chains: A case study of municipal company. *Transport* 32(3): 243–251.
Jacyna-Gołda, I. & Lewczuk, K. 2017. The method of estimating dependability of supply chain elements on the base of technical and organizational redundancy of process. *Eksploatacja i Niezawodnosc-Maintenance and Reliability* 19(3): 382–392.
Kececioglu, D.B. 2002. *Reliability Engineering Handbook*. Lancaster: DEStech Publications.
Kowalski, M. & Magott, J. & Nowakowski, T. & Werbińska-Wojciechowska S. 2014. Exact and approximation methods for dependability assessment of tram systems with time window. *European Journal of Operational Research* 235 (3): 671–686.
Pham, H. 2006. *Handbook of Engineering Statistics*. London: Springer-Verlag.
Tomaszek H., Zieja M., Ważny M., 2016. A method for reliability assessment of structural components of aircraft and sea-going ships with taking into account a given failure generation model. *Polish Maritime Research* 23(2): 83–90.
Tomaszek, H. & Jasztal, M. & Zieja, M. 2011. A simplified method to assess fatigue life of selected structural components of an aircraft for a variable load spectrum. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 4: 29–34.
Tomaszek, H. & Jasztal, M. & Zieja, M. 2013. Application of the Paris formula with m = 2 and the variable load spectrum to a simplified method for evaluation of reliability and fatigue life demonstrated by aircraft components. *Eksploatacja i Niezawodnosc – Maintenance and Reliability 15(4): 297–304.*
Ważny M. 2015. The method of estimating lifetimes of aircraft devices operating under ageing-attributable wearing conditions. *Journal of Theoretical and Applied Mechanics* No. 4 Vol. 53. 981–990.
Ważny, M. 2009. The method for assessing residual durability of selected of devices in avionics system. *Maintenance and Reliability* 3.
Ważny, M. 2008. The method of determining the time concerning the operation of a chosen navigation and aiming device in the operation system. *Maintenance and Reliability* 2.
Werbińska-Wojciechowska, S. & Zając, P. 2015. Use of delay-time concept in modelling process of technical and logistics systems maintenance performance. Case study. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 17 (2): 174–185.
Zieja, M. 2015. A method of predicting reliability and lifetime of aeronautical hardware with characteristic function ap-plied. Transport Means – Proceedings of the International Conference, Kaunas, 22–23 October 2015. Kaunas Univ. Technol.
Zieja M., Ważny M., Stępień S., 2016. Distribution determination of time of exceeding permissible condition as used to determine lifetimes of selected aeronautical devices/systems. *Eksploatacja i Niezawodnosc-Maintenance and Reliability* 18(1): 57–64.
Zio, E. 2009. *Computational Methods For Reliability and Risk Analysis*. Singapore: World Scientific Publishing.
Żurek, J. & Smalko, Z. & Zieja, M. 2010. Methods applied to identify causes of air events. *Reliability, Risk and Safety: Theory and Applications*. CRC Press-Taylor and Francis Group: 1817–1822.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Extensions of the I&AB method for the reliability assessment of the spent fuel pool of EPR

M. Bouissou
EDF Lab Saclay, Palaiseau, France

ABSTRACT: The I&AB (Initiator and All Barriers) method was first introduced at ESREL 2016, as an efficient means to calculate, thanks to closed form formulae, the reliability of a very large repairable system with dependencies among components. The mathematical support of I&AB is continuous time Markov chains, and therefore it cannot be used for modeling the spent fuel pool of a nuclear power plant, because for this system, there are two kinds of *deterministic delays* that must be taken into account: grace times (for example, after the total loss of cooling of the pool, it takes exactly 14 hours for the water to start boiling), and deterministic failures due to the limited capacity of water tanks. In the present paper, we extend the I&AB method to account for deterministic delays. We explain how we could apply this method in the case of the fuel pool of the EPR (European Pressurized Reactor) starting from a model in the form of a BDMP (Boolean Logic Driven Markov Process), and how results and computation times compare to a Monte Carlo simulation of the same BDMP.

1 INTRODUCTION

The standard PSA method (based on fault tree linking) is not well suited for the reliability assessment of the spent fuel pool of a nuclear power plant, for several reasons. Firstly, the dynamics of the phenomena to be modeled are relatively slow because of the large amount of water available in the pool itself and in the safety systems. It is thus not sufficient to look at what can happen in only 24 hours after an initiator. Secondly, the fact that components are repairable, and the existence of multiple standby redundancies cannot be ignored.

EDF has developed several tools for creating and quantifying dynamic models, better suited for this kind of system study. In particular, BDMPs (Boolean logic Driven Markov Processes) are a powerful modeling tool for the dependability analysis of dynamic systems (Bouissou & Bon 2003). For more than ten years, they have been used for assessing the reliability, availability, and safety of complex reconfigurable systems. BDMPs have a graphical representation close to fault trees, yet they specify (potentially very large) CTMCs (Continuous Time Markov Chains). A BDMP model with the same detail level as fault trees of a standard PSA would not be quantifiable by analytical methods, even with classical approximations. On the other hand, it would require too large computation times with Monte Carlo simulation, because the probability of reaching a too low level in the spent fuel pool is very small.

For these reasons, we have developed a new approximate method for the quantification of very large BDMPs, and more generally any model able to generate minimal products containing one initiating event and the failures of the barriers activated after it in order to avoid the undesirable event. This is why the main foreseen application domain is nuclear PSA, all the more so as existing PSA models will be very easy to adapt to I&AB, merely by adding repair rates to component data. In a PSA context, I&AB can be used to take repairs into account instead of postulating that 24 hours after an initiating event, either the undesirable event is unavoidable, or the system is in a safe state. The I&AB (Initiator and all barriers) main principles were published in a paper at ESREL 2016 (Bouissou & Hernu 2016). In the present paper, we give all analytical formulae of I&AB and of its extension in the case of grace times and deterministic failures. We also give some numerical application examples, comparing the I&AB approximation to "exact" calculations performed on a dynamic model via Monte Carlo simulation.

2 THE INITIAL I&AB METHOD (2016)

2.1 *Hypothesis on the system and definitions*

Suppose we want to calculate the reliability of a repairable system with standby redundancies; it may be a good approximation to take into account only one level of dependences between

the components. In other words, one is capable to distinguish failures of "normal" components (they are called "initiating events") and failures of components in standby (that function only in case of failures of normal components). But one cannot discriminate between a component of "primary standby" (that assures the functioning of the system after a failure of the corresponding normal component) and a component of "secondary standby" (that operates only after a failure of the primary standby component).

The I&AB method relies on the two following approximations:

A0: When an initiating event occurs, all standby components are supposed to start functioning (or maybe refuse to start) immediately after the initiating event; then, they may fail and be repaired independently from each other until the initiating event is repaired.

A1: Once an initiating event is repaired, the system cannot anymore fail, whatever happens.

We suppose that the real system is described by a CTMC where the initial, "perfect" state is the state into which the system always returns, until it is absorbed by a failure state. Then its unreliability can be estimated from the following formula (Bouissou & Bon 1992):

$$\bar{R}(t) \le 1 - \exp(-\Lambda p t) \tag{1}$$

where Λ is the frequency of initiating events (sum of rates of all transitions exiting the initial state) and p is the probability that the initiating events lead to an accident before the system goes back to the perfect state.

In I&AB, in order to estimate p we use the "Minimal Content of (failure) Sequences" (MCS) of the Markov chain, as it was defined in (Bouissou 2006). The formal definition of a MCS is given ibid, but it can be defined informally as the result of a Boolean reduction of the following fault tree: a single OR gate with one son per failure sequence, each son being presented as an AND gate over the events appearing in the sequence. Initiating events must be distinguished from other events, so that for example, the MCS of a system made of 2 components Y and Z in active redundancy is {Y_init, Z} {Z_init, Y} and not simply {Y, Z}.

For real, complex systems, the MCS can be obtained in (at least) two ways: by building a PSA type model made of event trees and fault trees, and calculating its minimal cut sets, or by building a BDMP and applying the steps described in (Bouissou & Hernu 2016) to transform it into a standard fault tree whose minimal cut sets are the MCS of the Markov chain specified by the BDMP. In the remainder of the paper, we will therefore suppose that we have the MCS of the studied system at hand, and we will call its elements "minimal products".

2.2 *I&AB general formulae*

Let us suppose that there are n initiating events that can lead out the system from its perfect state. Then, according to (1), the system unreliability at time t can be found from

$$\bar{R}(t) \le 1 - \exp\left(-t \sum_{ie=1}^{n} \lambda_{ie} p_{ie}\right) \tag{2}$$

where λ_{ie} is the failure rate of initiating event *ie* and p_{ie} includes probabilities for all k minimal products corresponding to it.

In calculations we distinguish two time intervals. The first interval is the mission time figuring in (2). The second time interval is *infinite* and starts once an initiating event takes place. The probability that all components in a minimal product c fail within time interval $[0,\infty]$ is simply the unreliability of a parallel system made of these components $\overline{R_c}(\infty)$; then we can use the following upper bound for p_{ie}, that will be a good approximation when all failure probabilities are small:

$$p_{ie} \le \sum_{c=1}^{k} \overline{R_c}(\infty) \tag{3}$$

Using the Murchland approximation, we obtain:

$$p_{ie} \le \sum_{c=1}^{k} \mathrm{E}\left(N_c(\infty)\right) \tag{4}$$

where $N_c(\infty)$ is the number of failures of the minimal product c on an infinite horizon. What keeps p_{ie} small is the fact that in the initial state considered for c, the initiating event is realized with probability 1, but once repaired, it never fails again, contrary to other elements of the product. (this is the approximation *A1*).

In order to calculate $\mathrm{E}\left(N_c(t)\right)$ we need to give first a few definitions. We will utilize the following reliability characteristics:

- Unavailability $Q(t)$ – the probability that a component is in a failure state at time t;
- Unconditional failure intensity $W(t)$: $W(t)\Delta t$ is the mean number of failures of a component between t and $t + \Delta t$.

For markovian basic events, depending on their type, these quantities are given by the following expressions:

- Initiating event (the repair is definitive)

$$Q(t) = \exp(-\mu t)$$

$$W(t) = 0$$

– Failure in operation (it can fail several times)

$$Q(t) = \frac{\lambda}{\lambda+\mu}\left[1-\exp\left(-(\lambda+\mu)t\right)\right]$$
$$W(t) = \lambda\left(1-Q(t)\right)$$

– Failure on demand (the repair is definitive)

$$Q(t) = \gamma\exp(-\mu t)$$
$$W(t) = 0$$

Because of the lack of space, we will not recall here the demonstration given in (Bouissou & Hernu 2016) that leads to the following formula, written for a minimal product c containing l failures on demand and m failures in function.

$$E\left(N_c(t)\right) = \prod_{i=1}^{l}\gamma_{c,i} \times \int_0^t \exp\left(-\mu_{c,ie}x\right) f(x)\,dx \qquad (5)$$

with

$$f(x) = \exp\left(-x\sum_{j=1}^{l}\mu_{c,j}\right)\sum_{i=1}^{m} W_{c,i}(x)\prod_{\substack{j=1\\ j\neq i}}^{m} Q_{c,j}(x).$$

Equation (5) assumes that a minimal product contains at least one basic event with a failure in operation. However, sometimes minimal products are only composed of basic events corresponding to failures on demand (plus one initiating event, as usually). In such a case, we suppose that these events happen at $t = 0$ and the unreliability for minimal product c is given by:

$$\overline{R_c}(\infty) = \Pr\left(\text{top} = \text{true at } t=0\right) = \prod_{i=1}^{l}\gamma_{c,i}. \qquad (6)$$

2.3 *I&AB formulae in the markovian case*

The general equation (5) yields a closed form formula in the purely markovian case, where all components have constant failure and repair rates.

In order to simplify notations, we will omit the index c in the remainder of this section: we will implicitly give formulas for a single minimal product.

Taking an infinite time horizon and replacing $W_i(x)$ by its expression given in section 2.2 for a failure in operation, equation (5) becomes:

$$E\left(N(\infty)\right) = \prod_{i=1}^{l}\gamma_i \times \int_0^{\infty} \exp\left(-\mu_{ie}x\right) f(x)\,dx \qquad (7)$$

with

$$\begin{aligned} f(x) &= \exp\left(-x\sum_{j=1}^{l}\mu_j\right)\sum_{i=1}^{m}\lambda_i\left(1-Q_i(x)\right) \\ &\quad\times\prod_{\substack{j=1\\ j\neq i}}^{m} Q_j(x) \\ &= \exp\left(-x\sum_{j=1}^{l}\mu_j\right) \\ &\quad\sum_{i=1}^{m}\lambda_i\left[\prod_{\substack{j=1\\ j\neq i}}^{m} Q_j(x) - \prod_{j=1}^{m} Q_j(x)\right]. \end{aligned}$$

Here we need to introduce new notations in order to simplify upcoming formulas. Let:

$$\mu = \mu_{ie} + \sum_{j=1}^{l}\mu_j$$
$$r_i = \lambda_i + \mu_i$$

Hence, replacing the functions Q_j by their definitions and using these new notations, we obtain:

$$\begin{aligned} E\left(N(\infty)\right) &= \prod_{i=1}^{l}\gamma_i \sum_{i=1}^{m}\lambda_i \\ &\times\left(\prod_{\substack{j=1\\ j\neq i}}^{m}\frac{\lambda_j}{r_j}\int_0^{\infty} e^{-\mu x}\prod_{\substack{j=1\\ j\neq i}}^{m}\left(1-e^{-r_j x}\right)dx\right. \\ &\left.-\prod_{j=1}^{m}\frac{\lambda_j}{r_j}\int_0^{\infty} e^{-\mu x}\prod_{j=1}^{m}\left(1-e^{-r_j x}\right)dx\right). \end{aligned} \qquad (8)$$

Each integrand includes a product of functions, which can be represented in the following way:

$$\begin{aligned} \prod_{j=1}^{m}\left(1-e^{-r_j x}\right) &= 1-\sum_{i=1}^{m}e^{-r_i x}+\sum_{i=1}^{m}e^{-r_i x}\sum_{j>i}^{m}e^{-r_j x} \\ &-\sum_{i=1}^{m}e^{-r_i x}\sum_{j>i}^{m}e^{-r_j x}\sum_{k>j}^{m}e^{-r_k x}+\ldots+(-1)^m \\ &\times\exp\left(-\sum_{i=1}^{m}r_i x\right). \end{aligned} \qquad (9)$$

Hence, after the integration from 0 to infinity, we obtain an alternating series, every term of which, in its turn, is a sum of fractions. For instance, the second integral results in:

$$\begin{aligned} &\int_0^{\infty} e^{-\mu x}\prod_{j=1}^{m}\left(1-e^{-r_j x}\right)dx = \\ &\frac{1}{\mu}-\sum_{i=1}^{m}\frac{1}{\mu+r_i}+\sum_{i=1}^{m}\sum_{j>i}^{m}\frac{1}{\mu+r_i+r_j} \\ &-\sum_{i=1}^{m}\sum_{j>i}^{m}\sum_{k>j}^{m}\frac{1}{\mu+r_i+r_j+r_k}+\ldots\ldots \\ &+(-1)^m\left(\mu+\sum_{i=1}^{m}r_i\right)^{-1}. \end{aligned} \qquad (10)$$

The first integral is calculated in a similar way, the only difference is that one should exclude current element i from the product.

These analytical formulae (8) and (10) seem very cumbersome; however, they permit to considerably reduce the processing time (in comparison with a

numerical integration) while ensuring an excellent accuracy.

3 I&AB EXTENSIONS

3.1 *Taking grace times into account*

In this section, the focus is on systems such that, after the loss of all components subject to random failures in a minimal product, the undesirable event is delayed by some physical process that guarantees a deterministic grace time. The spent fuel pool is a good example: after the complete loss of the cooling system, the water will heat until it boils, but this process is deterministic and it would give an excessively conservative evaluation to replace the grace time by a random delay, exponentially distributed in order to stay in the markovian framework.

We first suppose that we need to quantify minimal products containing failures of components (with the same hypotheses as in § 2.3) and a single deterministic grace time. Let X_c be the failure time of the set A_c of markovian elements of the minimal product c, Y_c the time needed to repair at least one of the markovian components, starting from the state where they are all failed, and T_c the grace time. For sake of simplicity, we suppose that after a given occurrence of the initiator, the basic event corresponding to the grace time behaves like a Heaviside function: it becomes true at $X_c + T_c$ and stays true forever (it is "not repairable"). The probability p_{ie} to go from the state where the system is just after the initiator *ie* to the failure state can be estimated as:

$$p_{ie} \approx \sum_{c=1}^{k} E\left(N_c(\infty)\right) \cdot \Pr\left(Y_c > T_c\right). \tag{11}$$

The total repair rate when all markovian components are failed is the sum of their repair rates. Hence

$$\Pr\left(Y_c > T_c\right) = \exp\left(-T_c \sum_{i \in A_c} \mu_i\right). \tag{12}$$

As for $E\left(N_c(\infty)\right)$, it can be computed using the formulae of §2.3.

To conclude this section, let us mention that the grace delay may depend on the minimal product, and that a minimal product can contain two or more grace delays: in this case, only the *last one* must be taken into account (cf. §4.1.2 for more details about this choice).

3.2 *Taking deterministic failures into account*

If, after a non-recovered loss of cooling, the water starts to boil in the fuel pool, there is a possibility to add water coming from tanks. However, the capacity of tanks is limited and after a given time the water flow is interrupted: this is what we call a deterministic failure. After a given initiator, such failures can be considered as non-repairable: it is impossible to replenish the tanks in a short amount of time (the same applies to batteries). However, in a dynamic model, they can be associated to a repair (with a small repair rate, see discussion on that topic in §4.1.2) in order to allow the model to return to its initial state. In order to be consistent with general assumptions of I&AB, we will suppose that the "timers" associated to deterministic failures start just after the initiating event; this is obviously conservative, as in fact they start after some failures. This assumption has an immediate consequence: if there are two or more deterministic failures in a minimal product, the one associated to the greatest delay suffices to prevent the minimal product from becoming true until it happens. So, without loss of generality, we will consider in this section that we want to quantify a minimal product containing l failures on demand, m failures in function, and one deterministic failure.

We define the unavailability Q and unconditional failure intensity W, needed in equation (5), for this type of basic event. Q is a Heaviside function and W a Dirac distribution:

$$Q(t) = \bar{R}(t) = \mathrm{H}(t_0) = \begin{cases} 0, & t < t_0 \\ 1, & t \geq t_0 \end{cases}$$

$$W(t) = \delta(t - t_0).$$

With these notations, equation (5) can be written as follows (with an infinite time horizon, and omitting the minimal product index c):

$$E\left(N(\infty)\right) = \prod_{i=1}^{l} \gamma_i \times \int_0^{\infty} \exp\left(-\mu_{ie} x - \sum_{i=1}^{l} \mu_i x\right) \times \sum_{i=1}^{m+1} W_i(x) \prod_{\substack{j=1 \\ j \neq i}}^{m+1} Q_j(x)\, dx \tag{13}$$

Taking, as in §2.3,

$\mu = \mu_{ie} + \sum_{j=1}^{l} \mu_j$ and

$r_i = \lambda_i + \mu_i$

$$E\left(N(\infty)\right) = \prod_{i=1}^{l} \gamma_i \times \int_0^{\infty} \exp(-\mu x) \sum_{i=1}^{m} W_i(x) \times \left(\prod_{\substack{j=1 \\ j \neq i}}^{m} Q_j(x) \times \mathrm{H}(t_0) \right) dx + \int_0^{\infty} \exp(-\mu x) \times \delta(x - t_0) \prod_{j=1}^{m} Q_j(x)\, dx$$

Finally,

$$E\left(N(\infty)\right) = \prod_{i=1}^{l} \gamma_i \times \int_{t_0}^{\infty} \exp(-\mu x) \sum_{i=1}^{m} W_i(x) \times \prod_{j\neq i}^{m} Q_j(x)\,dx + \exp(-\mu t_0) \prod_{j=1}^{m} Q_j(t_0) \quad (14)$$

The second term (the integral) of equation (13) can be written, using the same notations as in §2.3:

$$\sum_{i=1}^{m} \lambda_i \left(\prod_{j\neq i}^{m} \frac{\lambda_j}{r_j} \int_{t_0}^{\infty} e^{-\mu x} \prod_{j\neq i}^{m} \left(1 - e^{-r_j x}\right) dx - \prod_{j=1}^{m} \frac{\lambda_j}{r_j} \int_{t_0}^{\infty} e^{-\mu x} \prod_{j=1}^{m} \left(1 - e^{-r_j x}\right) dx \right). \quad (15)$$

After integration from t_0 to infinity, we obtain for the second integral the following alternate sum:

$$\int_{t_0}^{\infty} e^{-\mu x} \prod_{j=1}^{m} \left(1 - e^{-r_j x}\right) dx = \frac{\exp(-\mu t_0)}{\mu} - \sum_{i=1}^{m} \frac{\exp\left(-(\mu + r_i) t_0\right)}{\mu + r_i} + \sum_{i=1}^{m} \sum_{j>i}^{m} \frac{\exp\left(-(\mu + r_i + r_j) t_0\right)}{\mu + r_i + r_j} - \sum_{i=1}^{m} \sum_{j>i}^{m} \sum_{k>j}^{m} \frac{\exp\left(-(\mu + r_i + r_j + r_k) t_0\right)}{\mu + r_i + r_j + r_k} + \cdots\cdots + (-1)^m \left(\mu + \sum_{i=1}^{m} r_i\right)^{-1} \exp\left(-\left(\mu + \sum_{i=1}^{m} r_i\right) t_0\right).$$

Of course, taking $t_0 = 0$, we obtain again the formula (10) given in §2.3.

All these formulae are so complicated that it is necessary to carefully validate their implementation in a program. The next section has two purposes: give what we believe is the result of I&AB (we cannot guarantee that our Python implementation is totally bug free) and see how I&AB approximations compare to more precise calculations made by Monte Carlo simulation (the only possible method because of deterministic times) on a truly dynamic model.

4 ACCURACY TESTS OF I&AB EXTENSIONS

The small examples of this section were designed just to make comparisons between I&AB and "exact" calculations performed with Monte Carlo simulation. In practice the models were input graphically as BDMPs in KB3 (see Figure 1 for an example), then processed both by I&AB and by the Monte Carlo simulator YAMS. An overview of EDF tools including KB3 and YAMS is given in (Bouissou 2005).

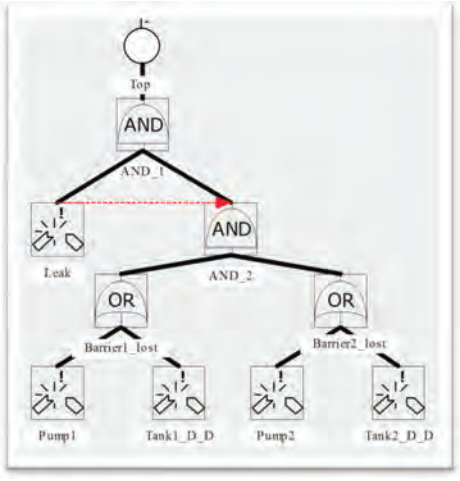


Figure 1. BDMP modeling a system with bounded capacities.

Table 1. I&AB accuracy on various simple test cases. Columns 2 and 3 are the estimations of the unreliability at 10000 hours computed by I&AB and Monte Carlo simulation (the last column is the width half of the 90% confidence interval of the YAMS result).

Test case	I&AB	YAMS	Conf. interval
4.1.1 a	$5.25\ 10^{-3}$	$5.05\ 10^{-3}$	$5.21\ 10^{-5}$
4.1.1 b	$1.17\ 10^{-3}$	$1.12\ 10^{-3}$	$5.50\ 10^{-5}$
4.1.1 c	$5.85\ 10^{-5}$	$5.72\ 10^{-5}$	$3.93\ 10^{-6}$
4.1.2 a	$7.08\ 10^{-3}$	$2.60\ 10^{-3}$	$2.64\ 10^{-4}$
4.1.2 b	$1.73\ 10^{-2}$	$3.90\ 10^{-3}$	$3.24\ 10^{-4}$
4.1.2 c	$2.89\ 10^{-3}$	$6.95\ 10^{-4}$	$4.33\ 10^{-5}$
4.1.2 d	$9.55\ 10^{-3}$	$1.03\ 10^{-3}$	$5.27\ 10^{-5}$
4.1.2 e	$3.54\ 10^{-4}$	$3.80\ 10^{-5}$	$1.01\ 10^{-5}$
4.1.2 f	$3.89\ 10^{-3}$	$1.00\ 10^{-4}$	$1.64\ 10^{-5}$
4.2.1 a	$1.87\ 10^{-1}$	$1.67\ 10^{-1}$	$8.67\ 10^{-4}$
4.2.1 b	$6.21\ 10^{-3}$	$5.44\ 10^{-3}$	$1.71\ 10^{-4}$
4.2.1 c	$1.65\ 10^{-3}$	$1.67\ 10^{-3}$	$9.50\ 10^{-5}$
4.2.2 a	$1.87\ 10^{-1}$	$1.42\ 10^{-2}$	$2.75\ 10^{-4}$
4.2.2 b	$1.87\ 10^{-1}$	$1.56\ 10^{-2}$	$2.88\ 10^{-4}$
4.2.2 c	$6.21\ 10^{-3}$	$9.86\ 10^{-4}$	$7.30\ 10^{-5}$
4.2.2 d	$6.21\ 10^{-3}$	$1.14\ 10^{-3}$	$7.86\ 10^{-5}$
4.2.2 e	$1.65\ 10^{-3}$	$8.28\ 10^{-4}$	$4.73\ 10^{-5}$
4.2.2 f	$1.65\ 10^{-3}$	$8.62\ 10^{-4}$	$4.83\ 10^{-5}$

In all calculations, failures are associated to a failure rate of 10^{-3}/h and repair rate of $2\ 10^{-2}$/h. The grace times and delays of deterministic failures are indicated in § 4.1 and 4.2. Table 1 gives a synthesis of all comparisons. The numbers in the first column correspond to the numbers of sections below that explain the test cases. All calculations with I&AB require a negligible time, whereas some of

the Monte Carlo simulations require a few minutes for sufficient precision.

Below are the descriptions of the test cases and comments on the results.

4.1 *Grace times*

4.1.1 *Single grace time*

The minimal product to quantify is {Initiator, A, B, grace_time}. In the dynamic model, there are only two sequences: Initiator, A, B, grace_time and Initiator, B, A, grace_time (A and B are in active redundancy). The grace time is successively taken equal to 25 h (line 4.1.1.a of Table 1), 50 h (line b), 100 h (line c).

In this case, I&AB works quite well, and it is not surprising, given the fact that the dynamic model corresponds exactly to the simplifying assumptions made in § 2.1.

4.1.2 *Two grace times*

The minimal product to quantify is {Initiator, A, grace_time_1, B, grace_time_2}. In the dynamic model, there is only one sequence: Initiator, A, grace_time_1, B, grace_time_2. The grace time is fractioned, and the component B can fail only after the failure of A and the end of grace_time_1. The two grace times (in hours) are successively taken equal to (5, 20) (line 4.1.2a of Table 1), (20, 5) (line b), (15, 35) (line c), (35, 15) (line d), (30, 70) (line e), (70, 30) (line f). Note that in the dynamic model, the basic event grace_time_1 is considered as repairable (with repair rate equal to 1/250 h), so that after an initiating event, the system can return to its initial state provided it does not reach the undesirable event; the value chosen for the repair rate is not sensitive as long as the mean time to repair components is much smaller than the mean time to repair the grace time: in such a case, after a given occurrence of the initiator, the grace time can be considered as "not repairable" just like in I&AB. In the simulation model, the order of the two grace times makes a difference if they are not equal. In I&AB, it is also the case because only the last grace delay is taken into account. We have also tested the idea of taking the sum of the two grace delays like a single one in I&AB: it yields results much closer to those of the dynamic model, but this approximation can produce optimistic results in some cases (for case f the result is $5.85\ 10^{-5}$). Intuitively, this is due to the fact that in the dynamic model only the last grace delay is competing with *all* repairs of the markovian elements of the cut set. Cf. also §5.

4.2 *Deterministic failures*

4.2.1 *Barriers in active redundancy*

Let us consider a little hydraulic system modeled by the BDMP below:

When the initiator Leak occurs, the two barriers (each one composed of a pump and a tank) are activated. The undesirable event occurs if, before the repair of the leak, the two pumping systems are lost, either because of a random failure of the pump, or because the tank is empty. The failure and repair rates for random events are as described at the beginning of §4, except that the repair rate of the Leak is 0.1/h in order to get small enough probabilities.

The results given in Table 1 correspond to the following values for the times after which Tank1 and Tank2 are empty: (40, 30) (line 4.2.1a), (80, 60) (line b), (150, 100) (line c). Note that the order of the two numbers is not important here, because of the symmetry of the two barriers. I&AB performs quite well on this example, where the minimal product containing the two deterministic failures is dominant. In the dynamic model as well as in I&AB, the amount of water in the biggest tank is the most influent parameter.

4.2.2 *Barrier 2 activated on failure of barrier 1*

In this case, in the dynamic model, the functioning times of the two tanks add up, unless a failure of pump1 forces to start barrier2 before depletion of tank1. It is therefore not surprising that I&AB is more conservative in this case than when the two barriers are in active redundancy. The BDMP corresponding to this case is not shown, because it is the BDMP of Figure 1 with just *one* additional trigger (red dotted line), going from gate Barrier1_lost to gate Barrier2_lost. There is no need to re-run the calculations with I&AB, since for this method, this case gives the same results as the previous one (active redundancy of barriers). But here, the capacities of the two tanks are not exchangeable in the dynamic model, this is why we ran YAMS with the following couples of values for deterministic delays: (40, 30) (line 4.2.2a), (30, 40) (line b), (80, 60) (line c), (60, 80) (line d), (150, 100) (line e), (100, 150) (line f). The unreliability increases a bit when the greatest delay is the last one. Going from line a to f, the results of I&AB range from extremely conservative (by a factor 10) to acceptably conservative (by a factor 2). On the other hand, using the sum of the delays instead of the greatest cannot be recommended because it could lead to optimistic results.

5 DIFFERENCES BETWEEN GRACE TIMES AND DETERMINISTIC DELAYS

In a dynamic model like a BDMP, both grace times and deterministic delays are represented as leaves

associated to a deterministic time to failure. So the difference between those two concepts is not obvious. In essence, the difference between a grace delay and a deterministic failure is that:

- Once the grace delay has started, whatever happens on the system can only postpone (case of a repair) the undesirable event, or leave it unchanged (case of a failure);
- In the case of a deterministic failure, whatever happens on the system can only make the undesirable event happen sooner (case of a failure), or leave it unchanged (case of a repair).

The I&AB theory makes a very clear distinction between the two concepts, because it considers that the grace time starts when all other components of the cut set have failed, whereas the timer of a deterministic failure starts just after the initiator. An intermediate grace time such as in the example of §4.1.2 corresponds to none of these cases, this is why in I&AB it should be simply ignored. It is the user's responsibility to mark leaves as grace delays, deterministic failures or "to be ignored" in the BDMP before transforming it into input data for I&AB.

On a large model, it is probable that the few minimal products with a too conservative quantification will be "hidden in the crowd" and that the global result will not be much affected.

6 APPLICATION TO THE SPENT FUEL POOL

To perform all our tests so far, we have used the implementation of I&AB that we described in (Bouissou & Hernu 2016). It is not the most efficient because it separates the search for minimal products from their quantification, therefore preventing the use of a probability threshold to discard at an early stage in the calculations most minimal products, as it is done by the MOCUS algorithm (Fussell & Vesely 1972). In spite of this limitation, we have been able to demonstrate impressive performances of I&AB in the spent fuel pool application.

We have built a model relative to the spent fuel pool of the European Pressurized Reactor and its support systems. Although less detailed than a classical PSA model, the BDMP we have built takes into account all dependances due to standby redundancies, common cause failures, sharing of electrical supplies… The model takes into account both the grace time of 14 hours before boiling of the water and deterministic failures of tanks used to replace evaporated water.

This BDMP (326 leaves, 77191 minimal products of order up to 6) could be processed by I&AB in a few minutes on a laptop. This model happened to be also quantifiable by YAMS: the Monte Carlo simulation gave a failure probability smaller than the result of I&AB by a factor around 2, but the calculation took 25 minutes to reach a 10% precision with 95% of confidence on the same machine.

Besides, with Monte Carlo simulation, it is very hard to get qualitative results: for that particular model, there is only one dominant sequence and all other sequences are much less probable: it would require many hours of simulation to get results comparable to the, say, 10 most probable minimal products that are easily identified by the I&AB method.

8 CONCLUSION

I&AB is an analytical method for the reliability calculation of large repairable systems with dependences between components. Two kinds of models can serve as input for this method: BDMPs or standard nuclear PSA models complying with the fault tree linking method. Both of them can be transformed into a set of minimal products that are the basis of the calculation. I&AB as it was described in (Bouissou & Hernu 2016) cannot readily be used for the fuel pool case, because for this system, there are two kinds of deterministic delays that must be taken into account: grace times, and deterministic failures due to the limited capacity of water tanks.

In the present paper, we have given two theoretical contributions: the analytical formulae of the I&AB method (so far, they were only available in the French patent file FR3044787) and their extension in the case of deterministic delays. In addition, we have shown on several examples that the extended method can yield reasonably conservative results, in times incomparably shorter than Monte Carlo simulation.

Thanks to a partnership between EDF and Lloyd's Register, I&AB will soon be available for the large community of users of the RiskSpectrum PSA tool. This could revolutionize PSA praxis in upcoming years.

ACKNOWLEDGEMENT

This paper is based on work done in collaboration with Olga Hernu, the co-author of the I&AB method and patent file (Bouissou & Hernu 2016).

REFERENCES

Bouissou, M. & Bon, J.-L. 1992. Fiabilité des grands systèmes séquentiels: résultats théoriques et applications dans le cadre du logiciel GSI. *Rev. Stat. Appl.* 40 (2).

Bouissou, M. & Bon, J.-L. 2003. A new formalism that combines advantages of fault-trees and Markov models: Boolean logic driven Markov processes. *Reliab. Eng. Syst. Saf.* 82: 149–163.

Bouissou, M. 2005. Automated dependability analysis of complex systems with the KB3 workbench: the experience of EDF R&D. *Proc. CIEM 2005, Bucharest, 2005.*

Bouissou, M. 2006. Détermination efficace de scenarii minimaux de défaillance pour des systèmes séquentiels. *Proc. 15ème Colloque de fiabilité et maintenabilité, Lille, 2006.*

Bouissou M. & Hernu, O. 2016. Boolean approximation for calculating the reliability of a very large repairable system with dependencies among components. *Proc. ESREL 2016, Glasgow, 2016.*

Fussell, J.B. & Vesely, W.E. 1972. A new methodology for obtaining cut sets for fault trees. *Trans. Amer. Nucl. Soc.* 15: 262–263.

Krcál, J. & Krcál, P. 2015. Scalable analysis of fault trees with dynamic features. *Proc. DSN 2015, Rio de Janeiro, 2015.*

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Structure function in analysis of multi-state system availability

M. Kvassay, V. Levashenko, J. Rabcan, P. Rusnak & E. Zaitseva
University of Zilina, Zilina, Slovakia

ABSTRACT: Mathematical representation of investigated system is important step of reliability analysis. There are two principal representations of considered system. The first one is known as Binary-State System (BSS), and it permits considering only two states in system/components performance—functioning and failed. Very often, this is not sufficient to adequate mathematical description of system behavior. Therefore, other types of mathematical representation are used. Multi-State System (MSS) is one of most prospective mathematical representations. This representation allows defining more than two states in describing behavior of the system and its components. A MSS can be described by structure function, which defines unambiguous correlation between all possible combinations of states of the system components and the system state. Typical disadvantages of MSS are large dimension and impossibility to form structure function based on uncertain data. We propose to consider application of mathematical methods of Multiple-Valued Logic (MVL) to form the MSS structure function. MVL is natural extension of Boolean algebra in MSS reliability assessment. This logic has been used for analysis of logic functions with more than two values. New method for structure function forming based on uncertain data and application of MVL methods is proposed. Multi-Valued Decision Diagrams (MDDs) are used for structure function representation, in particular. A MDD is useful data structure for representation of MVL function of large dimension. The method presented in this paper takes into account two disadvantages of MSS representation by structure function.

1 INTRODUCTION

System availability analysis can be implemented based on a mathematical representation of investigated system. The construction of the mathematical representation of the system has some factors that cause the final form of this representation. The first of them is definition of the number of availability levels (states) (Natvig 2011; Zaitseva & Levashenko 2017). The second factor is related to the mathematical method that is used for qualitative or quantitative analysis (Lisnianski & Levitin 2003; Zaitseva & Levashenko 2017). The third factor is caused by initial data uncertainty: the different mathematical representations are used for completely specified initial data and for uncertain initial data (Aven 2010).

According to these factors, mathematical representations for evaluation of system availability are divided as *Binary-State System* (BSS) and *Multi-State System* (MSS). A BSS allows representing the initial system as mathematical model with two possible states that are complete failure and perfect working. A MSS permits considering more than only two states in system behavior. In this paper, the MSS is considered for mathematical representation of the investigated system because it allows us to describe system behavior in more details. Both BSS and MSS can be used in analysis of systems for which the initial data is completely specified or uncertain. Four groups of methods are used for MSS analysis (Lisnianski & Levitin, 2003): stochastic processes, universal generating function, Monte Carlo simulation, and extensions of Boolean methods. The methods based on extensions of Boolean methods was historically the first in the MSS reliability evaluation (Murchland 1975), and they are based on the representation of investigated system in a form of the structure function.

The structure function defines dependency of system state on states of system components. There are many mathematical approaches in the reliability engineering for the structure function analysis. The most frequently used are fault trees (Kabir 2017), reliability block diagrams (Distefano & Puliafito 2009), and minimal cut/paths sets (Kvassay et al 2015a). Very often methods of Boolean algebra as *Binary Decision Diagram* (BDD) (Zaitseva et al 2015), Boolean function minimization (Di Maio et al 2016) and Logical Differential Calculus (Schneeweiss 2009) are used in the structure function analysis of BSSs. The similar methods have been developed for MSSs based on mathematical background of Boolean Logic (Xing & Amari 2015) and *Multiple-Valued Logic* (MVL) (Zaitseva & Levashenko 2017). For example, Multi-Valued Decision diagrams (MDDs) as generalization of BDDs are considered in (Xing & Dai 2008,

Zaitseva & Levashenko 2008) for analysis of MSSs. The methods for finding minimal cut/path sets of a MSS are proposed in (Kvassay et al. 2015b). Logical Differential Calculus for MSS importance analysis is considered in (Kvassay et al. 2017).

These investigations show the efficiency of MVL mathematical method use in MSS reliability analysis. However, methods based on structure function have some restrictions in analysis of MSSs, and they do not allow investigating systems for which initial data is incompletely specified. The structure function is constructed based on completely specified data as a rule. Therefore, MVL methods are combined with other methods to construct and analyze structure function of a MSS. One of possible methods has been proposed in (Zaitseva & Levahsenko 2016), and it allows us to construct the MSS structure function using Fuzzy Decision Tree (FDT). The FDT is inducted based on incompletely specified and uncertain data. The FDT permits creating a decision table that agrees with the structure function of the MSS. However, the table is not an optimal representation of MSS structure function because this function has large dimension as a rule. More suitable representation of the structure function of a MSS is MDD.

In this paper, we consider the MSS structure function mathematical representation based on completely specified data (section 2) and incompletely specified data (section 3). The MDD allows us to represent the completely specified structure function of a MSS. The complexity of the MDD depends on the ordering of variables of the structure function in this form. The FDT can be used in case of incompletely specified and uncertain data. The algorithm for induction of FDT for representation of the MSS structure function has been proposed in (Zaitseva & Levashenko 2016). However, this algorithm does not take into account specifics for ordering of variables in a MDD. Therefore, we propose use of other types of FDT that is ordered FDT (Levashenko et al 2007). A new algorithm for transformation of the ordered FDT into MDD is considered in section 4. The evaluation of accuracy of the FDT use in construction of a MDD representing the MSS structure function and efficiency of the ordering of variables in MDD are shown in section 5.

2 MATHEMATICAL REPRESENTATION OF SYSTEM BASED ON COMPLETELY SPECIFIED DATA

2.1 *Structure function*

A system of n components in the stationary state is considered in this paper. This system has M performance levels and is interpreted as a MSS. We suppose that this system is repairable and repairs end in the state as-good-as-new (for the individual components). This assumption allows defining MSS structure function as a time-independent function.

The structure function of MSS defines correlation of the MSS performance levels and its components states (Zaitseva & Levashenko 2017):

$$\phi(\boldsymbol{x}) = \phi(x_1,\ldots, x_n): \{0, \ldots, m_1 - 1\} \times \ldots \times \{0,\ldots, m_n - 1\} \rightarrow \{0,\ldots, M-1\}, \quad (1)$$

where $\phi(\boldsymbol{x})$ defines system performance levels from complete failure ($\phi(\boldsymbol{x}) = 0$) to perfect functioning ($\phi(\boldsymbol{x}) = M$ -1); $\boldsymbol{x} = (x_1,\ldots, x_n)$ is a state vector; x_i is the i-th component state that changes from failure ($x_i = 0$) to perfect functioning ($x_i = m_i$ -1).

Every system component is characterized by the probabilities of its states:

$$p_{i,s} = \Pr\{x_i = s\},\ s = 0,\ \ldots,\ m_i - 1. \quad (2)$$

The probabilities of the MSS performance levels, availabilities and unavailability of the system based on the structure function (1) are defined as (Kvassay et al. 2015b; Lisnianski & Levitin 2003):

$$P_j = \Pr\{\phi(\boldsymbol{x}) = j\},\ j = 0,\ 1,\ \ldots,\ M - 1, \quad (3)$$
$$A_h = \Pr\{\phi(\boldsymbol{x}) \geq h\},\ h = 1,\ \ldots,\ M - 1, \quad (4)$$
$$U = \Pr\{\phi(\boldsymbol{x}) = 0\}. \quad (5)$$

The MSS structure function (1) according to (Zaitseva & Levashenko 2017) is interpreted as MVL function that allows using MVL mathematical methods for MSS analysis. There are some descriptions of MVL function that can be applied for the MSS structure function representation. Two of most well-known are truth table (Zaitseva & Levashenko 2017) and *Multi-Valued Decision Diagram* (MDD) (Xing & Dai 2008).

For example, let us to consider the truth table of the simple service system analysed in (Kvassay et al 2015b) that consists of 3 components ($n = 3$): 2 service points (components 1 and 2) and the service system infrastructure (component 3). This system has 3 performance levels ($M = 3$): 0 – non-operational, 1 – partially operational, 2 – fully operational. Components 1 and 2 have 2 states ($m_1 = m_2 = 2$): functional (state 1) and dysfunctional (state 0). The third component has 3 states ($m_3 = 3$): failure (it is 0), partly working (it is 1) and perfect working (it is 2). The structure function of this system is shown in Table 1.

The basic measures (3)–(5) of this system are calculated by the truth table (Table 1) as:

$$P_2 = p_{1,1} \cdot p_{2,1} \cdot (p_{3,1} + p_{3,2}) + p_{1,1} \cdot p_{2,0} \cdot p_{3,2}, \quad (6)$$
$$P_1 = (p_{1,0} \cdot p_{2,1} + p_{1,1} \cdot p_{2,0}) \cdot p_{3,1} + p_{1,0} \cdot (p_{2,1} + p_{2,0}) \cdot p_{3,2}, \quad (7)$$

Table 1. The structure function of the simple service system.

Components states		x_3		
x_1	x_2	0	1	2
0	0	0	0	1
0	1	0	1	1
1	0	0	1	2
1	1	0	2	2

$$A_2 = p_{1,1} \cdot p_{2,1} \cdot (p_{3,1} + p_{3,2}) + p_{1,1} \cdot p_{2,0} \cdot p_{3,2}, \quad (8)$$

$$A_1 = (p_{1,0} \cdot p_{2,1} + p_{1,1} \cdot p_{2,0} + p_{1,1} \cdot p_{2,1}) \cdot p_{3,1} + p_{3,2}, \quad (9)$$

$$U = p_{3,0} + p_{1,0} \cdot p_{2,0} \cdot p_{3,1}. \quad (10)$$

The truth table of the MSS structure function has dimension $\prod_{i=1}^{n} m_i$. This dimension extremely grows with the increasing number of the function variables, i.e. with the number of system components. In MVL, another representation is used for function of large dimension. It is a MDD (Miller & Drechsler 2002).

2.2 *Multi-valued decision diagram*

A MDD is a graphical presentation of MVL function (Miller & Drechsler 2002). This form for the MSS structure function representation has been proposed in (Xing & Dai 2008, Zaitseva & Levashenko 2008). A MDD has been modified into Multi-state Multi-valued Decision Diagram (MMDD) in (Xing & Dai 2008). A MMDD has been proposed for the MSS in which the number of system performance levels is not equal to the number of component states, and each component in the system may have different number of states. The structure function of this MSS is defined according to (1). A MMDD analyzes the system states with success rather than all system states. In (Zaitseva & Levashenko 2017) the theoretical aspects of MVL mathematical approaches application for a MSS with the structure function (1) have been considered. These results allow us to use MDD to represent a MSS in which the system and each component may have different number of states.

A MDD is specified same as a BDD, except the nodes become more complex. This is due to the Shannon expansion for MVL-functions and this expansion is defined as (Miller & Drechsler, 2002):

$$\phi(\boldsymbol{x}) = 0 \cdot \phi(0_i, \boldsymbol{x}) + 1 \cdot \phi(1_i, \boldsymbol{x}) + \ldots + (m-1) \cdot \phi((m-1)_i, \boldsymbol{x}). \quad (11)$$

The Shannon expansion (11) is the basis for using MDD and a definition of manipulations over them. The format for MDD manipulation

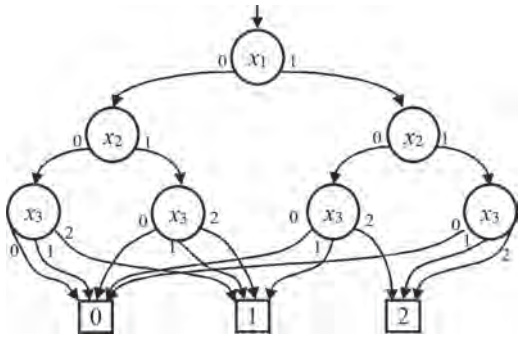


Figure 1. The MDD for the structure function of the simple service system (Table 1).

is defined as expression *case*(a, b_0, b_1, ..., b_{m-1}) = b_a (Miller & Drechsler 2002).

A MDD is a directed acyclic graph representing MVL function based on the Shannon expansion (11). This graph has M terminal nodes, labelled from 0 to (M – 1), representing M states of MSS. Each non-terminal node is labelled with a structure function variable x_i and has m_i outgoing edges. The non-terminal node outgoing edges are interpreted as component states.

For example, the MDD for the representation of the simple service system defined by Table 1 is shown in Figure 1. This MDD is uniquely described by procedures *case* and *if-then-else* as:

If $x_1 = 0$ *then*
    If $x_2 = 0$ *then case* (x_3, 0, 0, 1)
        else case (x_3, 0, 1, 1)
else
    if $x_2 = 0$ *then case* (x_3, 0, 1, 2)
        else case (x_3, 0, 2, 2)

The probabilistic interpretation of the MDD and rules for MSS measure calculation are based on the canonical Shannon expansion (11). The probabilistic interpretation of the MSS assumes that every edge from the i-th node labeled as s_j agrees the i-th component state probability $p_{i,s}$. Rules for the MSS evaluation according to (3) – (5) based on MDD are presented in details in (Zaitseva & Levashenko 2008). A calculation of these measures based on the MDD consists of analysis of paths from the root node to the terminal node "j", $j = 0, \ldots, M-1$. For example, the system availability A_2 based on the MDD in Figure 1 is defined by 3 paths from the root node labeled x_1 to terminal node 2:

$$A_2 = p_{1,1} \cdot p_{2,0} \cdot p_{3,2} + p_{1,1} \cdot p_{2,1} \cdot p_{3,1} + p_{1,1} \cdot p_{2,1} \cdot p_{3,2} = p_{1,1} \cdot p_{2,1} \cdot (p_{3,1} + p_{3,2}) + p_{1,1} \cdot p_{2,0} \cdot p_{3,2}.$$

This agrees with the availability A_2 (8) calculated by the truth table.

The truth table and the MDD are typically used for representation of completely specified function (Stankovic et al 2012). There are investigations to develop methods of the MDD construction for incompletely specified function. Most of these investigations have been provided in logic design (Popel & Drechsler 2003). The unspecified values of function in logic design can be interpreted arbitrarily (Stankovic et al 2012). The unspecified values of the structure function in reliability engineering appear if they cannot be measured or observed, but they must exist. Therefore, these values cannot be ignored (Aven 2010, Zaitseva et al 2017). These specifics require development of new methods and algorithms for representation of incompletely specified structure function because methods for MDD construction proposed in logic design cannot be used. One of possible approach has been proposed in (Zaitseva & Levashenko 2016, Zaitseva et al 2017). It is based on use of decision trees.

3 MATHEMATICAL REPRESENTATION OF SYSTEM BASED ON INCOMPLETELY SPECIFIED DATA

3.1 *Initial data*

According to the approach for the MSS structure function construction based on incompletely specified data in (Zaitseva & Levashenko 2016, Zaitseva et al 2017) this function is interpreted as classification procedure: all system state vectors $(x_1,\ldots, x_n)$ are divided into M classes. This classification is illustrated by MDD too. For example, the MDD of a simple service system in Figure 1 divides all state vectors into 3 classes.

The interpretation of the MSS structure function as classifier allows us to use methods for induction of classifiers based on incompletely specified data. Such methods are well known in Data Mining. One of them is induction of a decision tree. For example, the decision tree can be inducted by algorithm ID3/C4.5 developed by Quinlan (1987). This algorithm can be used for incompletely specified data that has crisp values. But initial data for reliability analysis is collected as expert data often. This data is latent or uncertain. There are different factors that cause uncertainty of data collected for reliability analysis (Aven 2010, Ley 2011). In this paper two of them are considered. The first is incompleteness of data. The second is ambiguity and vagueness of initial data.

The incomplete data is collected if it is impossible to indicate some values of the system components states or system performance levels. For example, it can be very expensive, or it needs unacceptable long time, or it is dangerous for the environment.

The ambiguity and vagueness of initial data are caused by methods for the data collection: measurement and/or expert's knowledge. The measurement can be inaccurate and with an error, that depends on accuracy of measuring device. Therefore, these data can be defined and used with some likelihood. In case of data collected based on expert's knowledge, data cannot be indicated exact because experts can have different opinions on one situation. The fuzzy logic makes it possible to define the structure function in a more flexible form.

The mathematical representation of a MSS in this case is interpreted as a classification problem for uncertain and incompletely specified data that can be decided with the application of FDT (Zaitseva & Levashenko 2016, Zaitseva et al 2017).

3.2 *Fuzzy Decision Tree*

A FDT is one of possible types of decision trees. A FDT allows taking into account not only unspecified values of system states but also uncertain values of components states. It is a formalism for expressing mapping of input attributes on output attribute/attributes, consisting of an analysis of attribute nodes linked to two or more sub-trees and leaves or decision nodes labeled with a class (in our case it is the system performance level) (Olaru & Whenkel 2003). A FDT defines the correlation between n input attributes $\{\mathrm{A}_1,\ldots, \mathrm{A}_n\}$ and an output attribute B. The important step in the use of FDT for the MSS mathematical representation is definition of the correlation between terminologies (formalisms) of FDT induction and MSS structure function.

In (Zaitseva & Levashenko 2016) the state vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ is interpreted as the set of input attributes $\mathbf{A} = \{\mathrm{A}_1, \ldots, \mathrm{A}_n\}$, and value of the structure function $\phi(\boldsymbol{x})$ as the output attribute B for FDT induction. Each input attribute (component state) A_i $(1 \le i \le n)$ is measured by a group of values ranging from 0 to $m_i - 1$, which agree with the values of states of the i-th component: $\{\mathrm{A}_{i,0}, \ldots, \mathrm{A}_{i,mi-1}\}$. Every value $A_{i,j}$ of fuzzy attribute A_i can be considered as a fuzzy set. These sets create the domain of a fuzzy attribute and their count is considered as a domain size. Each value of each instance is described by membership function $\mu_{A_{i,j}}(e) \in [0,1]$, where $\sum_{j=1}^{m_i} \mu_{A_{i,j}}(e) = 1$ and e represents an instance. These restrictions on representation of initial data are caused by the strategy for FDT induction for mathematical representation of the MSS. A FDT assumes that the input set $\mathbf{A} = \{\mathrm{A}_1,\ldots, \mathrm{A}_n\}$ is classified as one of the values of output attribute B. Value B_w of output attribute B agrees with one of the system performance levels and is defined as M values ranging from 0 to $M - 1$ $(w = 0,\ldots, M - 1)$.

For example, let us suppose that we have incompletely specified data with some likelihood about the simple service system (section 2.1). The data is represented in form of Table 2 (Zaitseva et al 2017). This data can be used for the system structure function construction based on likelihood of values (Table 3). But this structure function is incompletely specified (3 values are absent). Therefore, special methods or algorithms have to be used to "reconstruct" this MSS structure function. The method for structure function construction based on uncertain and incomplete data by FDT has been proposed in (Zaitseva & Levashenko 2016).

This method for structure function construction based on uncertain and incomplete data by FDT includes three steps (Zaitseva & Levashenko 2016):

- collection of data into the repository according to requests of FDT induction;
- representation of the system model in the form of a FDT that classifies components states according to the system performance levels;
- construction of the structure function as a decision table that is created by inducted FDT.

According to this method, the structure function is formed as a truth table that indicated the system performance level for each possible combination of components states. In this paper, we propose to transform inducted FDT into MDD for MSS mathematical representation, which is more acceptable for description of a MSS of large dimension. In particular, we propose to use the ordered FDT introduced in (Levashenko et al 2007) for the MDD creating.

Table 2. Uncertain data for the simple service system.

Components states							Performance levels		
x_1		x_2		x_3					
0	1	0	1	0	1	2	0	2	2
0.9	0.1	1	0	0.8	0.2	0	1	0	0
1	0	0.9	0.1	0.1	0.9	0	1	0	0
0.9	0.1	0.1	0.9	0.1	0.8	0.1	0	1	0
1	0	0	1	0	0.1	0.9	0	1	0
0.1	0.9	1	0	0.9	0.1	0	1	0	0
0	1	1	0	0.1	0.8	0.1	0	1	0
0.1	0.9	0	1	1	0	0	1	0	0
0.1	0.9	0	1	0.1	0.9	0	0	0	1
0.2	0.8	0.1	0.9	0	0.1	0.9	0	0	1

Table 3. The incompletely specified structure function of the simple service system.

Components states		x_3		
x_1	x_2	0	1	2
0	0	0	0	*
0	1	*	1	1
1	0	0	1	2
1	1	*	2	2

4 MULTI-VALUED DECISION DIAGRAM CONSTRUCTION BY FUZZY DECISION TREE

A FDT can have different properties. One of possible FDTs types is ordered FDT. This FDT has the same attributes at one level. It permits to use the predefined order of attributes in analysis of new samples. In this paper, the ordered FDT is used for the construction of the MDD based on uncertain data. The proposed algorithm includes three steps and differs from the method proposed in (Zaitseva & Levashenko 2016) only by the last step:

- collection of data into the repository according to requests of FDT induction;
- representation of the system model in the form of the ordered FDT that classifies components states according to the system performance levels;
- construction of the MDD by inducted FDT.

4.1 *Data collection and Fuzzy Decision Tree induction*

Collection of data in the form of a repository is provided by the monitoring of values of system component states and system performance level. This repository can be presented in the form of a table where the columns agree with the input and output attributes. This table is composed of $n + 1$ columns associated with n input attributes and 1 output attribute. The i-th column, for $i = 1,\ldots, n + 1$, is divided into m_i sub-columns. The j_i-th sub-column, for $j_i = 1, \ldots, m_i$, agrees with the j-th value of the attribute represented by the i-th column. Each row of the repository corresponds to one instance of collected data.

For example, the data about the service system in Table 2 satisfies the given conditions and can be interpreted as the repository.

There are different algorithms for FDT induction. The ordered FDT has been introduced by Levashenko et al (2007). In that paper, the algorithm for induction of ordered FDT based on cumulative information estimation of attributes for the selection of next node in tree's induction has been proposed.

The ordered FDT is inducted based on data in the repository by application of the cumulative

information estimates $I(B;A_{i_1},\ldots,A_{i_{q-1}},A_{i_q})$ where $A_{i_1},\ldots,A_{i_{q-1}}$ are input attributes that were used at previous levels of FDT. FDT induction is based on determination of expanded attribute where A_{i_q} are input attributes which have been used at previous levels of FDT. The selection criterion of expanded attributes A_{i_q} is defined as (Levashenko et al 2007):

$$i_q = argmax \cdot \left(\frac{\mathbf{I}\left(B;A_{i_1},\ldots,A_{i_{q-1}},A_{i_q}\right)}{Cost\left(A_{i_q}\right)\times \mathbf{H}\left(A_{i_q}\right)} \right), , \qquad (12)$$

where $Cost(A_{i_q})$ is an integrated measure that covers financial and temporal costs required to define the value of the A_{i_q} for an instance, and this value is defined a priori; $H(A_{i_q})$ is a cumulative entropy of input attribute A_{i_q}.

The cumulative mutual information in output attribute B about the attribute A_{i_q} and the sequence of attributes $A_{i_1},\ldots,A_{i_{q-1}}$ reflects the influence of attribute A_{i_q} on the output attribute B when sequence $A_{i_1},\ldots,A_{i_{q-1}}$ is known. Maximum value i_q in (12) facilitates the selection of expanded attribute A_{i_q}. This attribute will be associated with a node of the ordered FDT.

Two tuning thresholds α and β are in the algorithm for induction of the ordered FDT (Levashenko et al 2007). A tree branch stops to expand when either the frequency of the branch is below α or when more than β percent of instances left in the branch have the same class label. These values are key parameters needed to decide whether we have already arrived at a leaf node or whether the branch should be expanded further. Decreasing the parameter α and increasing the parameter β allow us to build large FDTs. On one hand, large FDTs describe datasets in more detail. On the other hand, these FDTs are very sensitive to noise in the dataset. We empirically select parameters near $\alpha = 0.10$ and $\beta = 0.80$.

The algorithm for the ordered FDT induction based on cumulative information estimates (12) is considered in (Levashenko et al 2007) in details. According to this algorithm we inducted the ordered FDT based on data in Table 2 for the mathematical representation of the simple service system (Figure 2).

The ordered FDT in Figure 2 allows us to form the decision table for all possible combinations of component states (all state vectors). For example, for state vector $x = (0\ 0\ 0)$ in Figure 2, the value of the output attribute is 0. This value is defined by the path from root node to leaf node $A_{3,0}$ with the confidence of 0.880 without analysis of other attributes. The values for other state vectors are computed in the similar way. All calculated values of the output attribute are equal to system states in Table 1.

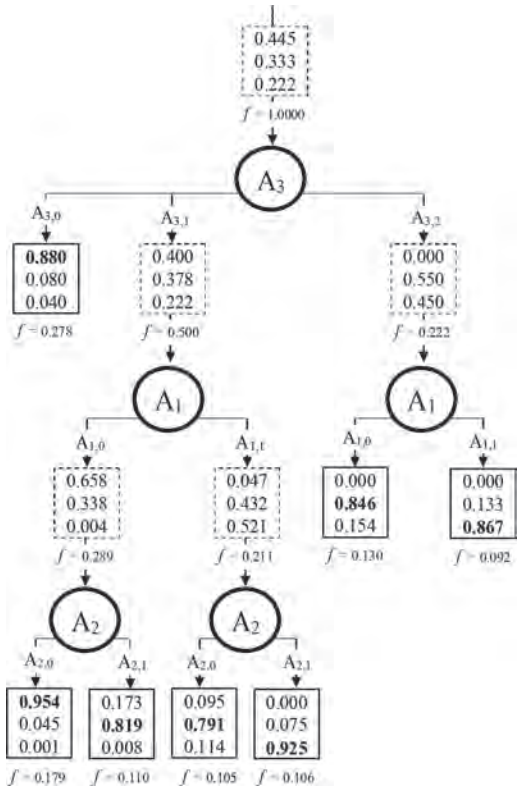


Figure 2. FDT of the simple service system (Table 2).

Therefore the ordered FDT allows represent the structure function uniquely. At the same time, there are many methods and algorithms for the MSS evaluation based on a MDD. The calculation of availability and other measures for the system reliability evaluation are considered in (Mo et al 2017, Xing &Dai 2008, Zaitseva & Levashenko 2008, Zaitseva et al 2013). So, let us consider the transformation of ordered FDT into MDD.

4.2 *Construction of multi-valued decision diagram from Fuzzy Decision Tree*

The important problem in MDD construction is the ordering of the variables: the complexity of MDD depends on the variables order in MDD (Miller & Drechsler 2002). This problem for MDD is being actively investigated and some methods have been proposed for approximate decisions for the optimal ordering of variables in MDD (Stankovic et al 2012, Popel & Drechsler 2003). At the same time the ordered FDT induction supposes the optimal (or quasi optimal) ordering of nodes. Therefore, the ordered FDT can be transformed into MDD with the optimal ordering of variables.

The algorithm for the transformation of the ordered FDT into the MDD includes three steps:

- the transformation of the ordered FDT into decision tree by the defuzzification of the ordered FDT nodes.
- the merger of the same leaf nodes;
- the tree's attributes substitution by appropriate variables.

Let us illustrate these steps by an example of the MDD construction for the service system (Table 2) based on the ordered FDT of this system (Figure 2). According to the proposed algorithm, the first step is defuzzification of the ordered FDT. It can be implemented by construction of decision tree based on maximal values of thresholds α and β. The decision tree for the investigated system is shown in Figure 3.

The second step is the merger of the same leaves that is illustrated in Figure 4. The result of the last step, which is substitution of tree's attributes by appropriate variables, is shown in Figure 5 as MDD of the structure function of the simple service system.

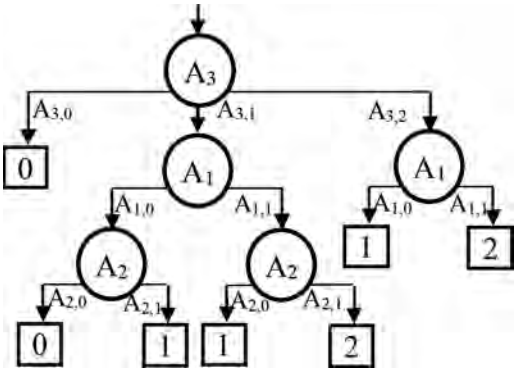


Figure 3. The decision tree of the simple service system formed from the ordered FDT in Fig. 2.

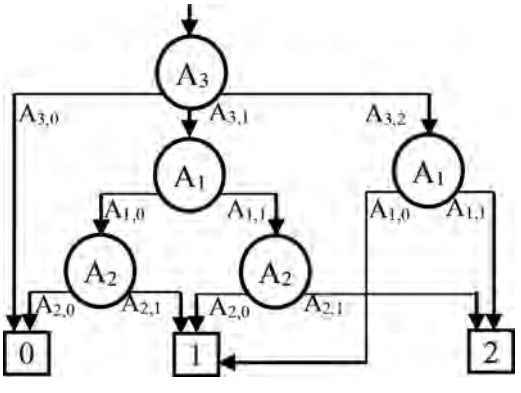


Figure 4. The merger of equal leaf nodes of decision tree in Fig. 3.

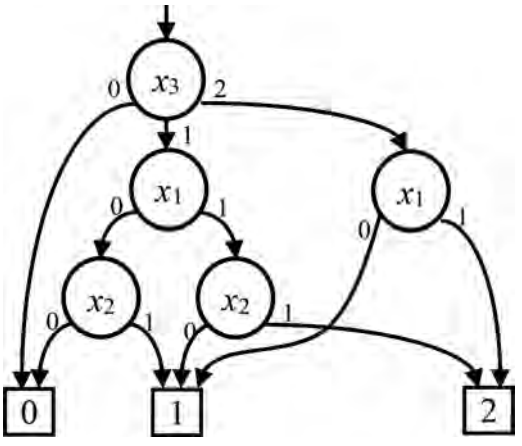


Figure 5. The MDD for the structure function of the simple service system.

The comparison of MDDs in Figure 1 and Figure 5 shows that the MDD constructed based on the ordered FDT has 5 non-terminal nodes and MDD constructed based on the truth table consists of 7 non-terminal nodes. This simple example illustrates that the MDD can be constructed for MSS by incompletely specified data and that the constructed MDD has better ordering of variables than the MDD constructed in section 2.2.

5 EVALUATION

In this section, the efficiency of the algorithm for the construction of the MDD of the MSS structure function based on uncertain data is considered. We use structure functions of three systems analyzed in (Zaitseva et al 2017): outline of an offshore electrical power generation system (Natvig 2011); army battle plan (Boedigheimer & Kapur 1994); a computer system with a memory subsystem subject to competing failure isolation and propagation effect (Xing & Levitin 2011). Basic characteristics of these systems are shown in Table 4. Two types of investigation were implemented.

The first of them is the influence of incompleteness of data on the accuracy of the MSS representation. For this investigation, all integer values representing components states and performance levels of the system were transformed to values with possibilities (Zaitseva et al 2017). We indicated known value with the possibility 1.00 and other values with possibilities 0.00 for components states and performance levels. To model incom-

Table 4. Characteristics of investigated system.

	Natvig (2011)	Xing & Levitin (2011)	Boedigheimer & Kapur (1994)
Numbers of state vectors	243	512	108
Numbers of components	5	5	4

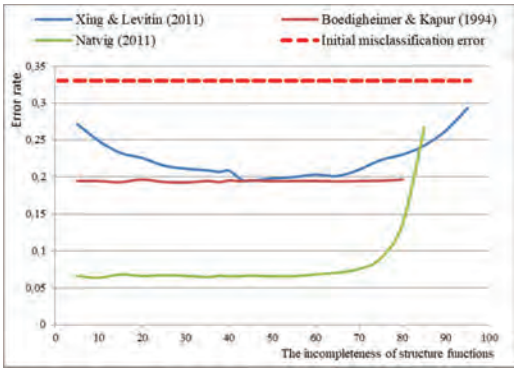


Figure 6. The error rate for the construction of the structure function for three systems.

pleteness of data, the structure functions of the MSSs were randomly transformed into incompletely specified by random deleting some values of the functions. The proportion of deleted values was changed from 5% to 90%. The transformed functions were interpreted as incompletely specified and restored by the ordered FDTs. A restored structure function and initial complete specified function were compared, and the error rate was calculated as a ratio of erroneous values of the structure function to the dimension of unspecified part of the function. The experiments were done for every structure function. The average values for all functions with specified range of deleted values were computed. Error rate for investigated system is less than initial misclassification error (maximal error rate) that is indicated by red line in Figure 6). The evaluation showed that the error increases significantly if the unspecified part is less than 10% or most than 85% for these systems (Figure 6).

The second type of the evaluation was analysis of MDD, in particular, the number of non-terminal nodes of MDDs that are constructed based on different approaches. In particular, we compared the MDD without special ordering of variables of structure function, the MDD formed based on FDT (this algorithm has been considered in (Zaitseva et al 2017) and MDD constructed by ordered FDT. This analysis was implemented for 16 structure functions of coherent systems (number of performance levels $M = 2$ and number

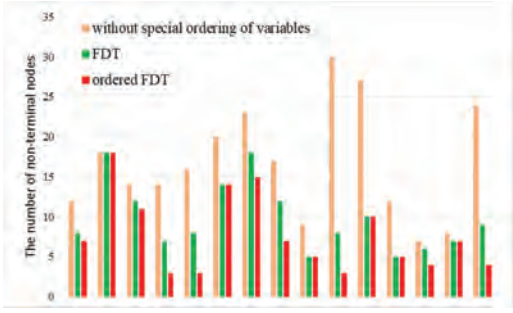


Figure 7. The efficiency of MDD constructed based on the truth table and FDT.

of variables $n = 5$). The evaluation of these MDDs showed that the MDDs created based on ordered FDT have less number of non-terminal nodes (Figure 7).

6 CONCLUSION

The main contribution of this paper is in development of the new algorithm for the construction of the structure function based on incompletely specified and ambiguous data in form of the MDD. It directly supports MSS reliability estimation and can cope with uncertain data for the analysis of system reliability/availability.

The analysis of the error rate of the proposed algorithm for the construction of the structure function based on ordered FDT showed that the method has good efficiency, which is similar to efficiency of the method for construction of structure function based on unordered FDT (Zaitseva et al 2017). This method is acceptable for incomplete data and the incompleteness of initial data can be indicated from 10% to 85%. The structure function constructed based on the proposed method has less error rate than maximal error rate in interval of the incompleteness.

ACKNOWLEDGMENT

This work was partly supported by the grants VEGA 1/0038/16 and 1/0354/17, APVV SK-FR-2017-0003.

REFERENCES

Akers, S.B. 1978. Binary Decision Diagrams. *IEEE Trans Computing* 27: 509–516.

Aven, T. 2010. Some reflections on uncertainty analysis and management, *Reliability Engineering and System Safety* 95(3): 195–201.

Boedigheimer R.A. & Kapur K.C. (1994). Customer-Driven Reliability Models for Multistate Coher-

ent System, IEEE Trans on Reliability, vol.43(1), pp.46–50.

Di Maio, F. et al. 2017. Determination of prime implicants by differential evolution for the dynamic reliability analysis of non-coherent nuclear systems, *Annals of Nuclear Energy* 102(4): 91–105.

Distefano, S. & Puliafito, A. 2009. Dependability Evaluation with Dynamic Reliability Block Diagrams and Dynamic Fault Trees, *IEEE Trans on Dependable and Secure Computing* 6 (1): 4–17.

Kabir, S. 2017, An overview of fault tree analysis and its application in model based dependability analysis, *Expert Systems with Applications* 77(1): 114–135.

Kvassay, M., et al. 2015a. Analysis of minimal cut and path sets based on direct partial Boolean derivatives, *Proc Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 230(2): 147–161.

Kvassay, M. et al. 2015b. Minimal cut sets and direct partial logic derivatives in reliability analysis, Safety and Reliability: Methodology and Applications—Proc of the European Safety and Reliability Conference, ESREL 2014, 241–248.

Kvassay, M. et al. 2017. Importance analysis of multistate systems based on tools of logical differential calculus, *Reliability Engineering and System Safety* 165: 302–316.

Levashenko, V. et al. 2007. Fuzzy Classified Based on Fuzzy Decision Tree, *Proc. of the IEEE Int. Conf. on Computer as a tool (EUROCON 2007),* 823–827.

Ley, D. 2011. Approximating process knowledge and process thinking: Acquiring workflow data by domain experts, *Proc. of IEEE Int. Conf. on Systems, Man, and Cybernetics*, 3274–3279.

Lisnianski, A. & Levitin, G. 2003. Multi-state System Reliability. Assessment, Optimization and Applications. *World Scientific, Singapore*, SG.

Miller, M. & Drechsler R. 2002. On the construction of multiple-valued decision diagrams, *Proc. of the IEEE 32nd Int. Symp. on Multiple-Valued Logic*, 264–269.

Mo, Y.C. et al. 2017. MDD-based performability analysis of multi-state linear consecutive-k-out-of-n: F systems. *Reliability Engineering & System Safety* 166: 124–131.

Murchland, J.D. 1975. Fundamental Concepts and Relations for Reliability Analysis of Multistate System. In: *Reliability and Fault Tree Analysis, Theoretical and Applied Aspects of System Reliability*, SIAM, 581–618.

Natvig, B. 2011. *Multistate Systems Reliability Theory with Applications,* Wiley, New York.

Olaru, C. & Whenkel, L. 2003. A Complete Fuzzy Decision Tree Technique, *Fuzzy Sets and Systems* 138(2): 221–254.

Popel, D.V. & Drechsler R. 2003. Efficient minimization of multiple-valued decision diagrams for incompletely specified functions, *Proc. of the 33rd Int. Symp. on Multiple-Valued Logic*, 241–246.

Quinlan, J.R. 1987. Simplifying decision trees, *Int. J. Man-Machine Studies* 27: 221–234.

Schneeweiss, W.G. 2009. A short Boolean derivation of mean failure frequency for any (also non-coherent) system, *Reliability Engineering & System Safety* 94(8): 1363–1367.

Stankovic, R.S. et al. 2012. *Representations of Multiple-Valued Logic Functions*, Morgan &Claypool Publishers.

Xing, L. & Amari, S.V. 2015. *Binary Decision Diagrams and Extensions for System Reliability Analysis,* Wiley.

Xing, L., & Dai, Y.S. 2008. A new decision-diagram-based method for efficient analysis on multistate systems, *IEEE Trans on Dependable and Secure Computing* 6(3): 161–174.

Xing L. & Levitin G. (2011). Reliability of Multi-State System Subject to Competing Failures, Proc. Annual Reliability and Maintainability Symposium (RAMS), pp.1–7.

Zaitseva, E. et al. 2015. Importance analysis based on logical differential calculus and Binary Decision Diagram, *Reliability Engineering and System Safety* 138: 135–144.

Zaitseva E. & Levashenko V. 2008. Decision diagrams for reliability analysis of multi-state system, *Proc of Int Conf on Dependability of Computer Systems (DepCoS-RELCOMEX)*, 55–62.

Zaitseva E. & Levashenko V. 2016. Construction of a Reliability Structure Function Based on Uncertain Data, IEEE Trans on Reliability 65(4): 1710–1723.

Zaitseva E. & Levashenko V. 2017. Reliability analysis of multi-state system with application of multiple-valued logic, *International Journal of Quality and Reliability Management* 34(6): 862–878.

Zaitseva E. et al. 2013. A Multi-Valued Decision Diagram for Estimation of Multi-State System, *Proc. of the IEEE Int. Conf. on Computer as a tool (EUROCON 2013)*, 645–650.

Zaitseva E. et al. 2017. Induction of structure function of Multi-State System based on uncertain data, *Safety and Reliability. Theory and Applications—Proc of the European Safety and Reliability Conference, ESREL 2017,* 241–248.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Advances in the simplification of Fault Trees automatically generated from AltaRica 3.0 models

M. Batteux
IRT SystemX, Palaiseau, France

T. Prosvirnova
LGI, CentraleSupelec, Gif-sur-Yvette, France

A. Rauzy
MTP, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: Safety and risk analyses rely on models. These models have several important characteristics. They are event-oriented. The system under study changes of state when events, such as failure, hazard, repair and so on, occur. They are probabilistic. The exact moment of the occurrence of a failure is in essence unpredictable. They are discrete. States are represented by means of variables that take their values into finite, usually very small, domains. The most widely used modeling formalisms such as Fault Trees, Block Diagrams and Event Trees rely on Boolean algebra. There are cases however where binary states are not sufficient. For instance, it is sometimes necessary to represent the level of degradation of a component, the quality of a signal, and so on. This kind of models can be easily represented with AltaRica 3.0 – a high level modeling language dedicated to safety analyses. AltaRica 3.0 is at the core of the OpenAltaRica project which aim is to develop a complete set of assessment tools for the language, including among others compilers to Fault Trees and Markov Chains, stochastic and stepwise simulators. In this article we study how the notion of prime implicants can be extended to finite domain calculus. We discuss the efficient implementation of finite domain calculus and show how these results can be applied to simplify Fault Trees, automatically generated from AltaRica 3.0 models. This simplification in its turn significantly improves the efficiency of the assessment of the automatically generated Fault Trees.

1 INTRODUCTION

Risk analysis relies on models. These models have several important characteristics:

- They are event-oriented. The system under study changes of state when events, such as failure, hazard, repair and so on, occur.
- They are probabilistic. The exact moment of the occurrence of a failure is in essence unpredictable.
- They are discrete. States are represented by means of variables that take their values into finite, usually very small, domains.

The last characteristic is pragmatic: given the difficulty to design models and computational complexity of the calculation of indicators, discrete abstractions are a necessary tradeoff. Hence the role of Boolean algebra in Reliability, Availability, Maintainability, Safety engineering. The most widely used modeling formalisms such as Fault Trees, Block Diagrams and Event Trees rely on Boolean algebra. There are cases however where binary states are not sufficient. For instance, it is sometimes necessary to represent the level of degradation of a component, the quality of a signal, and so on. This kind of models can be easily represented with AltaRica 3.0 – a high level modeling language dedicated to safety analyses (Prosvirnova, Batteux, Brameret, Cherfi, Friedlhuber, Roussel, & Rauzy 2013). AltaRica 3.0 is at the core of the OpenAltaRica project[1] which aim is to develop a complete set of assessment tools for the language, including among others compilers to Fault Trees (Prosvirnova & Rauzy 2015) and Markov Chains (Brameret, Rauzy, & Roussel 2015), stochastic and stepwise simulators (Aupetit, Batteux, Rauzy, & Roussel 2015).

In this article we study how the notion of prime implicants can be extended to finite domain cal-

[1]See https://www.openaltarica.fr.

culus and how to encode it efficiently. The contribution of this article is thus multiple. First, we present how the notion of prime implicants can be extended to finite domain calculus. Second, we discuss the efficient implementation of finite domain calculus. Finally we show how these results can be applied to simplify Fault Trees, automatically generated from AltaRica 3.0 models.

The remainder of this article is organized as follows. Section 2 describes a motivating example. Section 3 presents a theoretical work about finite domain calculus and discusses its implementation. Section 4 shows the application of the finite domain calculus to the simplification of Fault Trees automatically generated from AltaRica 3.0 models. Section 5 presents some experimental results using the motivating example. Section 6 concludes this article.

2 MOTIVATING EXAMPLE

Consider a parametric block diagram use case (see Figure 1) with three parameters:

- s the number of blocks in series;
- p the number of parallel blocks;
- q the level of recursivity (depth).

These relatively simple but large safety models can be easily represented in AltaRica 3.0 and handled simply and efficiently by means of the Fault Tree compilation tool chain.

Note that without loosing the efficiency of the assessment, in AltaRica 3.0, it is possible to represent multi-state blocks, e.g. consider the quality of data with the values `ok`, `lost` or `erroneous`, or the level of degradation with the values `ok`, `degraded` or `failed`.

This use case is both representative of a class of industrial models and parametric to show the scalability of the approach. We shall use it throughout the article to illustrate the advances in the simplification of Fault Trees.

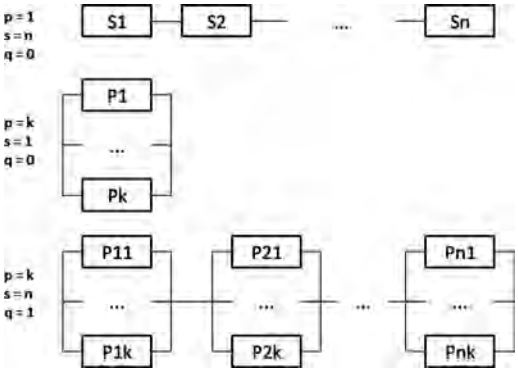


Figure 1. Parametric block diagram use case.

3 FINITE DOMAIN CALCULUS

3.1 *Definitions*

Let $\Xi = \{X_1, X_2, ..., X_n\}$ be a finite set of variables. Each Xi takes its values into a finite domain (a finite set of constants) denoted as $dom(X_i)$. The set of well formed formulas over Ξ is the smallest set such that:

- The two Boolean constants 0 (false) and 1 (true) are formulas.
- If X is a variable and c is a constant then $X = c$ is a formula. Such a formula $X = c$ is called a **literal** and makes only sense if $c \in dom(X)$.
- If f and g are formulas, then so are $f + g$ (disjunction), $f * g$ (conjunction), and $-f$ (negation).

We assume that the negation (–) has a higher priority than the conjunction (*), which has a higher priority than the disjunction (+).

A **product** is a set of literals interpreted as the conjunction of its elements. A product is said **fundamental** if it does not contain two literals built over the same variable. We shall consider only fundamental products. The empty product is denoted **1**.

A **minterm** is a product that contains a literal for each variable of Ξ. As we shall see, minterms play a fundamental role in the finite domain calculus for they are the atoms of the underlying Boolean algebra.

A **sum of products** is a set of products interpreted as the disjunction of its elements. The empty sum of products is denoted **0**.

A **variable assignment** of Ξ is a function $\sigma: \Xi \rightarrow dom(X_1) \times dom(X_2) \times ... \times dom(X_n)$, that associates to each variable X_i its value from $dom(X_i)$, $i = 1, ..., n$.

Let f and g be formulas and σ be a variable assignment over Ξ. The value of $\sigma(f)$ is calculated recursively as follows:

- $\sigma(1) = 1$, $\sigma(0) = 0$;
- $\sigma(X = c) = 1$ if $\sigma(X) = c$ and 0 otherwise;
- $\sigma(f + g) = max(\sigma(f), \sigma(g))$, $\sigma(f * g) = min(\sigma(f), \sigma(g))$, $\sigma(-f) = 1 - \sigma(f)$.

The variable assignment σ **satisfies** the formula f, if $\sigma(f) = 1$, otherwise is **falsifies** it.

There is a one to one correspondence between minterms and variable assignments: the minterm p corresponds to the variable assignment σ if for each variable $X \in \Xi$, $X = c \in p$ if and only if $\sigma(X) = c$.

3.2 *Implication, equivalence, properties*

Let $\Xi = \{X_1, ..., X_n\}$ be a finite set of finite domain variables. Let f and g be two formulas built over Ξ. f **implies** g, which is denoted as $f \Rightarrow g$, if any vari-

able assignment that satisfies f satisfies g as well. f is **equivalent** to g, which we denote as $f \Leftrightarrow g$, if both $f \Rightarrow g$ and $g \Rightarrow f$.

The usual properties of Boolean algebras hold for the finite domain calculus:

Neutral element: $f + 0 \Leftrightarrow 0 + f \Leftrightarrow f$ and $f * 1 \Leftrightarrow 1 * f \Leftrightarrow f$

Absorbing element: $f + 1 \Leftrightarrow 1 + f \Leftrightarrow 1$ and $f * 0 \Leftrightarrow 0 * f \Leftrightarrow 0$

Idempotence: $f + f \Leftrightarrow f$ and $f * f \Leftrightarrow f$

Commutativity: $f + g \Leftrightarrow g + f$ and $f * g \Leftrightarrow g * f$

Associativity: $f + (g + h) \Leftrightarrow (f + g) + h$ and $f * (g * h) \Leftrightarrow (f * g) * h$

Distributivity: $f + (g * h) \Leftrightarrow f * g + f * h$ and $f * (g + h) \Leftrightarrow (f + g) * (f + h)$

Double negation: $- - f \Leftrightarrow f$

de Morgan's law: $-(f + g) \Leftrightarrow -f * -g$ and $-(f * g) \Leftrightarrow -f * -g$

3.3 *Negation*

The real difference between the propositional and finite domain calculi stands in the negation.

Let Ξ be a finite set of variables, let X be a variable from Ξ, and finally let c be a constant of $dom(X)$. Then, $-(X = c) \Leftrightarrow \Sigma_{d \in dom(X),\, d \neq c}(X = d)$

Theorem 1 (Elimination of negations).: *For any formula of the finite domain calculus, there exists an equivalent formula involving no negation.*

Note that any formula is equivalent to the sum of minterms that satisfies it, which is a first way to demonstrate the theorem. A more syntactic proof consists in pushing negations down to literals, thanks to de Morgan's law, and then to transform negative literals as shown above.

3.4 *Subsumption, resolution*

Let p and q be two products built over Ξ. We say that p **subsumes** q if $q \Rightarrow p$, i.e. if and only if any literal of p is also a literal of q. If p subsumes q, then $p + q \Leftrightarrow p$.

Let X be a variable of Ξ, let $dom(X) = \{c_1, c_2, ..., c_k\}$ and let $p_1, ..., p_k$ be k products in which X does not show up. Let r be the product $p_1 * p_2 * ... * p_k$. Then the following implication holds: $(X = c_1) * p_1 + ... + (X = c_k) * p_k \Rightarrow r$.

The product r is called the **resolvent** of the products $(X = c_1) * p_1, ..., (X = c_k) * p_k$.

In the case, where there is a product p_j such that $p_j = r$, then the following equivalent holds:

$$(X = c_1) * p_1 + ... + (X = c_k) * p_k \Leftrightarrow (X = c_1) * p_1 + ... + (X = c_j) * p_j + ... + (X = c_k) * p_k + r$$

3.5 *Shannon normal form*

Let X be a variable of Ξ, let c be a constant of $dom(X)$ and finally let f be a formula built over Ξ. There exist two formulas f_1 and f_0 in which the atom $(X = c)$ does not show up such that:

$$f \Leftrightarrow (X = c) * f_1 + f_0$$

The above representation is called the pivotal **decomposition** of f with the respect to X and c.

Assume we are given an (arbitrary) order $<$ over the variables of Ξ and over the constants of the do-main of the variables of Ξ. The set of formulas in **Shannon Normal Form** is defined inductively as follows:

- The two constants 0 and 1 are in Shannon Normal Form.
- If f and g are two formulas in Shannon Normal Form, X is a variable and c is a constant of $dom(X)$, the formula $(X = c) * f + g$ is in Shannon Normal Form if
  - X does not show up in f, and
  - for all literal $(Y = d)$ showing up in g, either $X < Y$ or $X = Y$ and $c < d$.

3.6 *Representation theorem*

Let Ξ be a finite set of finite domain variables. Let X be a variable of Ξ, let c be a constant of $dom(X)$ and finally let $f = (X = c) * f_1 + f_0$ be a formula in Shannon Normal Form built over Ξ.

In the above representation we can assume without a loss of generality that:

- $f_1 \neq 0$ as $(X = c) * 0 + f_0 \Leftrightarrow f_0$
- $f_0 \neq 1$ as $(X = c) * f_1 + 1 \Leftrightarrow 1$

From now, we shall assume that these two **simplification rules** are systematically applied.

Theorem 2 (Representation).: *For any formula of the finite domain calculus, there exists at least one equivalent formula in Shannon Normal Form.*

In general, this equivalent formula is not unique. We shall see that two of the formulas that represent a given sum of products are of special interest: the first one can be interpreted as sum of disjoint products, the other one as the set of prime implicants. These two formulas are extremum in a sense we shall explain.

A formula in Shannon Normal Form can be interpreted as a sum of products. Namely,

- SumOfProducts [0] = 0;
- SumOfProducts[1] = 1;
- SumOfProducts[$(X = c) * f + g$] = $\{(X = c) * p; p \in SumOf\ Products[f]\} \cup SumOf\ Products[g]$

Theorem 3 (Sums-of-Products).: *Shannon Normal Formulas one-to-one correspond with Sums-of-Products (for a given order of variables and constants).*

3.7 *Factors and cofactors*

The **factor** and **cofactor** of a formula f with respect to a variable X, denoted respectively as $f|X$ and

$f \sim X$, are syntactic operations that select respectively the products of f that contain X and the products of f that do not contain X. The **factor** $f|X$ is defined recursively as follows:

- $0[X = 0$ and $1|X = 1$
- $[(X = c) * f + g]|X = (X = c) * f + [g|X]$
- $[(Y = c) * f + g]|X = 0$ if $X < Y$
- $[(Y = c) * f + g]|X = (Y = c) * [f|X] + [g|X]$ if $X > Y$

The **cofactor** $f \sim X$ is defined recursively as follows:

- $0 \sim X = 0$ and $1 \sim X = 1$
- $[(X = c) * f + g] \sim X = g \sim X$
- $[(Y = c) * f + g] \sim X = g$ if $X < Y$
- $[(Y = c) * f + g] \sim X = (Y = c) * [f \sim X] + [g \sim X]$ if $X > Y$

3.8 *Logical operations*

Let Ξ be a finite set of finite domain variables. Let X and Y be two variables of Ξ with $X < Y$. Let c, d and e be three constants such that $c, d \in dom(X)$ with $c < d$, and $e \in dom(Y)$. Finally let $f = (X = c) * f_1 + f_0$, $g = (X = c) * g_1 + g_0$, $h = (X = d) * h_1 + h_0$ and $I = (Y = e) * I_1 + I_0$ be four formulas built over Ξ in Shannon Normal Form. The following equivalences hold and they are used as recursive equations to perform logical operations on formulae in Shannon Normal Form:

- $f + g \Leftrightarrow (X = c) * [f_1 + g_1] + [f_0 + g_0]$
- $f + h \Leftrightarrow (X = c) * f_1 + [f_0 + h]$
- $f + I \Leftrightarrow (X = c) * f_1 + [f_0 + I]$
- $f * g \Leftrightarrow (X = c) * [f_1 * g_1 + f_1 * g_0 \sim X + f_0 \sim X * g_1] + [f_0 * g_0]$
- $f * h \Leftrightarrow (X = c) * [f_1 * h_1 + f_1 * g_0 \sim X] + [f_0 * g]$
- $f * I \Leftrightarrow (X = c) * f_1 + [f_0 * I]$
- $-f \Leftrightarrow [\sum_{d \in Dom(X), d \neq c} (X = d) * -g] + [-f * -g]$

3.9 *Subsumption*

As we shall see, it is of interest to remove from a formula f all the products that are subsumed by a product of a formula g. This operation, denoted $f \div g$, can be defined by means of the following recursive equations. Let Ξ be a finite set of finite domain variables. Let X and Y be two variables of Ξ ($X < Y$), let c, d and e be three constants such that $c, d \in dom(X)$, $c < d$, and $e \in dom(Y)$. Then:

- $f \div 0 = f$, $f \div 1 = 0$, $0 \div g = 0$ and $1 \div g = 1$
- $[(X = c) * f_1 + f_0] \div [(X = c) * g_1 + g_0] = (X = c) * [(f_1 \div g_1) \div g_0] + f_0 \div g_0$
- $[(X = c) * f_1 + f_0] \div [(X = d) * g_1 + g_0] = (X = c) * [f_1 \div g_0)] + f_0 \div g_0$
- $[(X = c) * f_1 + f_0] \div [(Y = e) * g_1 + g_0] = (X = c)[f_1 \div [(Y = e) * g_1 + g_0]] + f_0 \div [(Y = e) * g_1 + g_0]$
- $[(X = d) * f_1 + f_0] \div [(X = c) * g_1 + g_0] = (X = c)[f_1 \div g_0] + f_0 \div g_0$
- $[(Y = e) * f_1 + f_0] \div [(X = c) * g_1 + g_0] = [(Y = e) * f_1 + f_0] \div g_0$

3.10 *Prime implicants*

Let Ξ be a finite set of finite domain variables with an order over variables and constants. Let f and p be respectively a formula and a product built over Ξ.

- p is an **implicant** of f if $p \Rightarrow f$.
- p is a **prime implicant** of f if it is an implicant of f and no strict sub-product (subsuming product) of p is.

The set of prime implicants of f is denoted $PI[f]$.

Theorem 4 (Decomposition of Prime Implicants).: *Let f be a formula in Shannon Normal Form. Then $f = (X = c_1) * f_1 + ((X = c_2) * f_2 + ... + ((X = c_k) * f_k + f_0)) .. .)$ for some constants $c_1,..., c_k$ from $dom(X)$ and some formulas $f_1,..., f_k, f_0$ in Shannon Normal Form in which X does not occur.*

Let $h = (f_1 * f_2 * ... * f_k) + f_0$. Then, the set of prime implicants of f denoted by $PI[f]$ are calculated as follows:

$$PI[f] = \{(X = c_1) * p; p \in PI[f_i] \div PI[h]\}$$
$$...$$
$$\cup \{(X = c_k) * p; p \in PI[f_k] \div PI[h]\}$$
$$\cup \{PI[h]\}$$

The decomposition theorem gives an algorithm to calculate for any formula f in Shannon Normal Form an equivalent formula h such that:

$$g = SumOf\ Products[h] = PI[f]$$

Because all possible resolutions and subsumptions have been performed, g can be considered as the **most simplified form** of f.

At the opposite, we may want to transform f into an equivalent **sum of disjoint products** so to be able to calculate the exact probability of f. Disjoining products encoded by f is performed by the dual operation of calculating resolvents.

Let $f = (X = c_1) * f_1 + ((X = c_2) * f_2 + ... + ((X = c_k) * f_k + f_0)) .. .)$. Assume that $dom(X) = \{c_1,..., c_k\}$ (if some constant c_i of $dom(X)$ is missing we can always add the term $(X = c_i) * 0$). Then, f is equivalent to the following formula:

$$g = (X = c_1) * [f + f_0] + ((X = c_2) * [f_2 + f_0] + ... + ((X = c_k) * [f_k + f_0)] + 0) .. .).$$

By applying this transformation recursively, we get a sum of disjoint products, which is also **unique**, for a given order on variables and constants.

3.11 *Diagrammatic representation*

The idea is to represent sums of products in Shannon Normal Form by means of a variant of Bryant's **Binary Decision Diagrams** (Bryant 1992). The idea is therefore to represent formulas in Shannon Normal Form by means of Directed Acyclic Graphs with two types of nodes:

- Leaves, that are labeled with either 0 or 1.
- Internal nodes, that are labeled with a variable X and a constant c of $dom(X)$ and that have two out-edges called the 1-outedge and the 0-outedge. Such a node represents the formula $(X = c) * f + g$, where f and g are the formulas represented respectively by the node pointed by the 1-outedge and the node pointed by the 0- outedge.

The **Shannon Diagram** representing a formula is always built **bottom-up**. Nodes are maintained into a **unique** table (and accessed by means of a **hash**table). In this way, for any formula f, there is at most one node representing f in the table. Checking the equivalence of two formulas is thus performed in constant time once their Shannon Diagrams are built.

4 APPLICATION

One of the possible applications of the finite domain calculus presented above is the simplification of Fault Trees automatically generated from AltaRica 3.0 models. AltaRica 3.0 is an event-based high level modeling language dedicated to Safety Analyses (Prosvirnova, Batteux, Brameret, Cherfi, Friedlhuber, Roussel, & Rauzy 2013). Its semantics is based on Guarded Transitions Systems (Rauzy 2008).

4.1 *Guarded transitions systems*

A Guarded Transitions System (GTS) G is a quintuple $\langle V, E, T, A, \iota \rangle$, where:

- $V = S \cup F$ F is a set of variables, divided into two disjoint sets: a set S of state variables and a set F of flow variables.
- E is a set of events.
- T is a set of transitions. A transition is a triple $t = \langle e, G, P \rangle$, where e is an event from E, G is a Boolean expression built over variables from V and called the guard of the transition, and P is an instruction built over V and called the action or the post-condition of the transition.
- A is an assertion (i.e. an instruction built over V).
- ι is the initial (or default) assignment of variables of V.

A GTS $G = \langle V, E, T, A, \iota \rangle$ is an implicit representation of a labeled Kripke structure, i.e. a graph $\Gamma = (\Sigma, \Theta)$, where

- the set of nodes Σ represent the variable assignments (of V), and
- Θ is the set of edges labeled by the events from E.

Instructions of GTS are defined recursively as follows:

- "*skip*" is an empty instruction.
- If v is a variable and Exp an expression, then "v: = Exp" is an instruction (called "assignment").
- If C is a Boolean expression, I is an instruction, then " if C then I" is an instruction (called "conditional assignment").
- If I_1 and I_2 are two instructions, then so is "I_1; I_2" (called "parallel composition").

We shall consider two types of instructions. The "Actions" which are instructions in which left members of assignments are only state variables. The "Assertions" which are instructions in which left members of assignments are only flow variables.

Let denote by $\tau = Propagate(A, \iota, \sigma)$ a variable assignment obtained after applying the assertion A to the variable assignment σ., i.e. the calculation of flow variables value. $Propagate(A, \iota, \sigma)$ computes the values of flow variables using the instructions of the as sertion A and the values of state variables in σ. At the end if there are flow variables without any value, they are set to their initial values in ι and the assertion A is applyed to check that all the assignments are satisfied.

4.2 *Compilation to fault trees*

The compilation of AltaRica 3.0 models to Fault Trees works (Prosvirnova & Rauzy 2015) in 5 steps (see Figure 2):

1. The AltaRica 3.0 model is flattened into a GTS.

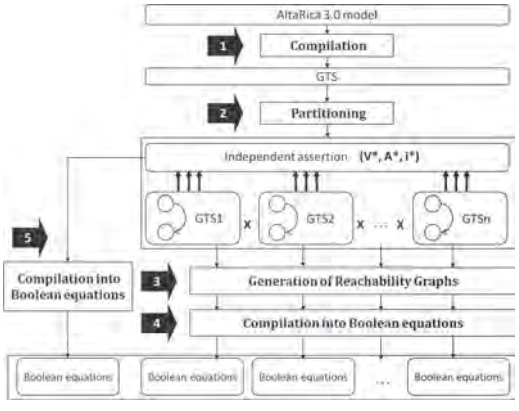


Figure 2. Compilation of AltaRica 3.0 models to Fault Trees.

2. The obtained GTS is partitioned into independent GTSs plus an independent assertion.
3. Reachability graphs of each independent GTS are calculated.
4. Each reachability graph is separately compiled into Boolean equations.
5. The independent assertion is compiled into Boolean equations.

The independent assertion $\langle V^*, A^*, \iota^* \rangle$ (the 5th step of the algorithm) is transformed into a set of Boolean formulas in the following way. For each pair (f, d), where $f \in V^*$ is a flow variable and $d \in dom(f)$ is its value, a formula $\phi_{(f,d)}$ is constructed according to the instructions in the assertion A^* and Boolean formulas $\{\phi_{(u,c)}, u \in U, c \in dom(u)\}$ obtained from the compilation of the independent GTSs.

In order to compile the assertion into Boolean formulas efficiently, one need to separate it into independent parts. The dependency relation between variables in the assertion A^* defines a dependency graph. This graph may contain cycles. The strongly connected components of this graph divide variables of A^* into sets and enable to decompose the assertion A^* into blocks of instructions A_i ($i = 1, ..., m$), where m is the number of strongly connected components: $A^* = A_1^*; A_2^*; \ldots; A_m^*$

Each block of instructions A_i^* is compiled into Boolean formulas recursively. Let denote by

- V_i^* – a set of variables labeling the vertices of the strongly connected component number i.
- A_i^* – an instruction that calculates the values of variables from V_i^*.
- ι_i^* – an initial assignment of variables from V_i^*.
- W_i^* – a set of variables such that variables from V_i^* depend on them in A_i^*.

For all variable v in V_{*_i} the formula $\phi(v, c)$ (where $c \in dom(v)$) is built as follows:

- Let $\Sigma = \underset{w \in W_i^*}{\times} dom(w)$ be the Cartesian product of the domains of variables from W_i^*.
- Let $\sigma \in \Sigma$ be an assignment of variables from W_i^*.
- Let ϕ_σ be a product built over W_i^* (as defined in Section 3.1) calculated as follows:

$\phi_\sigma = \prod_{w \in W_i^*} (w = \sigma(w))$

- Let τ be a partial variable assignment, $\tau: V_i^* \bigcup W_i^* \to \mathcal{C}$, such that:

$\forall w \in W_i^* \tau(w) = \sigma(w)$

- The partial variable assignment τ can be completed by propagating the assertion A_i^*:

$\tau = Propagate(A_i^*, \iota_i^*, \tau)$

- Then for each couple (v, c), with $v \in V_i^*$, such that $\tau(v) = c$, the formula associated with (v, c) is updated as follows

$\phi_{(v,\chi)} \leftarrow \phi_{(v,\chi)} + \phi_\sigma$

At the end of the algorithm, for all variables $v \in V^*$ and their values, we obtain a formula $\phi_{(v,c)}$ built over a finite set of finite domain variables W^*, such that v depends on them in the assertion A^*. We use the diagrammic representation as defined in Section 3.11 to represent these formulas.

As we have seen in Section 3.10, $\phi_{(v,c)} \Leftrightarrow PI[\phi_{(v,c)}]$ and it is the most simplified form of $\phi_{(v,c)}$.

For each variable $v \in V^*$ and its value $c \in dom(v)$, we compute $PI[\phi_{(v,c)}]$ and use this form, which greatly simplifies the generated Fault Tree.

4.3 *Example*

Consider the parametric block diagram use case presented in Section 2. Figure 3 illustrates how each basic block of these diagrams can be represented in AltaRica 3.0.

The variable *State* represents the internal state of a basic block and takes its value in the domain *BLOCKSTATE* = {*ok*, *ko*}. A domain is an enumeration having any finite number of values. The event *failure* represents the internal failure of a basic block. It is possible to associate different probability distributions to the events of basic blocks (e.g. exponential, constant, Weibull). The value of the parameters can also be changed. The behavior of a basic block is represented by a state machine given Figure 3. The variable *Out* is a flow variable, which represents the output of a basic block. The assertion of a basic block is an instruction, which calculates the value of this variable *Out* according to the value of the state variable *State*.

Figure 4 shows how two blocks in series can be modeled in AltaRica 3.0. The assertion of the whole model is

*B*1.*Out*: = *B*1.*State*;

*B*2.*Out*: = *B*2.State;

Out: = if (*B*1.*Out* = = *ok*) *and* (*B*2. *Out* = = *ok*) then *ok* else *ko*;

The compilation into Fault Trees is performed as follows.

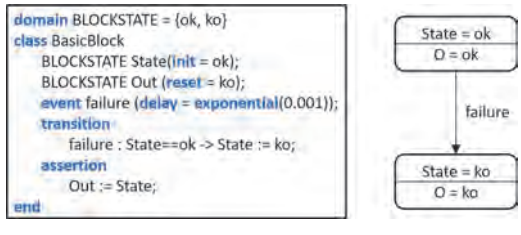


Figure 3. AltaRica 3.0 model of a basic block.

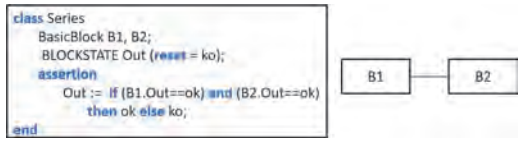


Figure 4. AltaRica 3.0 model of two blocks in series.

First, local reachability graphs are compiled:

$\phi(B1.State,ok) = true$

$\phi(B2.State,ok) = true$

$\phi(B1.State,ko) = B1.failure$

$\phi(B1.State,ko) = B2.failure$

Second, local assertions are compiled:

$\phi(B1.Out,ok) = (B1.State = ok)$

$\phi(B2.Out,ok) = (B2.State = ok)$

$\phi(B1.Out,ko) = (B1.State = ko)$

$\phi(B2.Out,ko) = (B2.State = ko)$

Third, the global assertion is compiled

$\phi(Out,ok) = (B1.Out = ok) * (B2.Out = ok)$

$\phi(Out,ko) = ((B1.Out = ok) * (B2.Out = ko)$
$+(B1.Out = ko) * (B2.Out = ok)$
$+(B1.Out = ko) * (B2.Out = ko))$

Figure 5 represents the last formula by means of a variant of Binary Decision Diagram (as presented in Section 3.11). It can be simpified using the algorithm presented in Section 3.10 as follows:

$dom(B1.Out) = dom(B2.Out) = \{ko, ok\}$

$f = (B1.Out = ko)$
$* [(B2.Out = ko) * 1 + [(B2.Out = ok) * 1 + 0]]$
$+[(B1.Out = ok) * [(B2.Out = ko) * 1 + 0] + 0]$

$f_1 = (B2.Out = ko) * 1 + [(B2.Out = ok) * 1 + 0]$

$f_2 = (B2.Out = ko) * 1 + 0$

$f_0 = 0$

$h = f_1 * f_2 + f_0 = (B2.Out = ko) * 1 + 0$

$PI[h] = (B2.Out = ko) * 1 + 0$

$PI[f_1] = 1$

$PI[f_2] = (B2.Out = ko) * 1 + 0$

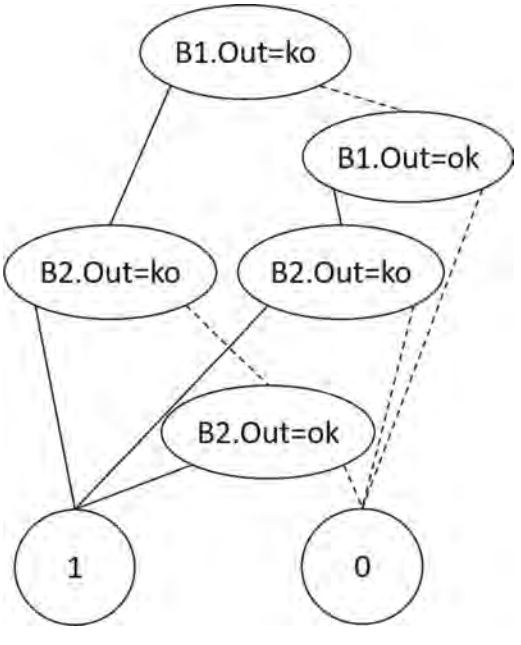


Figure 5. Diagrammatic representation of $((B1.Out = ok) * (B2.Out = ko) + (B1.Out = ko) * (B2.Out = ok) + (B1.Out = ko) * (B2.Out = ko))$.

$PI[f_1] \div PI[h] = 1$

$PI[f2] \div PI[h] = 0$

$PI[f] = (B1.Out = ko) * 1$
$+[(B2.Out = ko) * 1 + 0]$

which reads as $(B1.Out = ko) + (B2.Out = ko)$. It is the most simplified form of $\phi_{(Out,ko)}$.

5 EXPERIMENTS

The Fault Tree compiler of the the OpenAltaRica platform produces Fault Trees from AltaRica 3.0 models. More precisely, according to one or several Boolean observers, representing safety cases of the modeled system, the Fault Tree compiler generates Fault Trees in Open-PSA model exchange format (Hibti, Friedlhuber, & Rauzy 2012) with these Boolean observers as top events. The produced Fault Trees can be then assessed by XFTA (Rauzy 2012) to compute Minimal Cut Sets, probabilities of the top events, and so on.

We have implemented the algorithm presented in Section 4 and integrated it in the original version of the Fault Tree compiler. This algorithm greatly simplifies the generated Fault Trees, compared to those generated by the original version.

We have performed experiments with different values for the three parameters s, p and q of the motivating example pictured Figure 1.

In Table 1 we present the results obtained with the original version of the Fault Tree compiler and with the new one.

The first column is the number of the considered cases. The five next columns, from the second column to the sixth column, present the number of components contained in each sub-parts. We start with n_1 parallel blocks, in each block there are n_2 sub-blocks in series, into each sub-block there are n_3 parallel subblocks, and so on. For example, the first case means 3 parallel blocks, with 3 sub-blocks in series, each one containing 3 parallel sub-blocks; whereas the fifth case means 2 blocks in series, with 4 parallel subblocks, each block containing 4 sub-blocks in series, with 4 parallel sub-blocks into each one. The seventh column represents the total number of basic blocks in the AltaRica 3.0 model. Finally, the eighth and ninth columns represent the number of intermediate events in the Fault Trees generated by the original version of the Fault Tree compiler (ninth column) and the new one (eighth column).

The main observation is about the benefit of the number of generated gates with the new version of the Fault Tree compiler in comparison to the original one. In average, this benefit is of 56.9%. It means that in average with the new algorithm, the number of generated gates is less than 56.9% compared to the number of generated gates with

Table 1. Parametric block diagram use case—fault tree compilation.

Case	n_1	n_2	n_3	n_4	n_5	Number of blocks	Number of gates (new)	Number of gates (original)
1*st*	3	3	3	0	0	27	243	525
2*nd*	3	3	3	3	0	81	720	1158
3*rd*	4	4	4	0	0	64	537	1276
4*th*	4	4	4	4	0	256	2133	5114
5*th*	0	2	4	4	4	128	1135	3236
6*th*	0	3	3	3	2	54	558	1264
7*th*	0	3	3	3	0	27	243	596
8*th*	0	3	3	3	3	81	768	1914
9*th*	0	4	4	4	0	64	549	1562

the original one. The minimal value is 37.8% in the second case; and the maximum value is 64.9% in the fifth case.

The benefit obtained with the new version of the algorithm implemented in the Fault Tree compiler is important.

6 CONCLUSION

Boolean models are widely used for probabilistic safety analysis. There are cases however where binary states are not sufficient. For instance, it is sometimes of interest to represent the level of degradation of a component, the quality of signal, and so on. This kind of models can be easily represented with AltaRica 3.0, a high level modeling language dedicated to safety analyses. AltaRica 3.0 comes with several efficient assessment tools, amongst them a Fault Tree compiler.

In this article we presented how the notion of prime implicants can be extended to finite domain calculus. We discussed how the finite domain calculus can be efficiently encoded using a variant of Binary Decision Diagrams. We shown, using a parametric block diagram use case, how these results can be applied to simplify Fault Trees automatically generated from AltaRica 3.0 models. The number of generated intermediate events is on average divided by two, which greatly improves Fault Trees readability and the efficiency of their assessment.

REFERENCES

Aupetit, B., M. Batteux, A. Rauzy, & J.-M. Roussel (2015, September). Improving performances of the altarica 3.0 stochastic simulator. In L. Podofillini, B. Sudret, B. Stojadinovic, E. Zio, and W. Kröger (Eds.), *Safety and Reliability of Complex Engineered Systems: ESREL 2015*, Zu¨rich, Switzerland, pp. 1815–1824. CRC Press.

Brameret, P.-A., A. Rauzy, & J.-M. Roussel (2015). Automated generation of partial markov chain from high level descriptions. *Reliability Engineering & System Safety 139*, 179–187.

Bryant, R.E. (1992, sep). Symbolic boolean manipulation with ordered binary-decision diagrams. *ACM Comput. Surv. 24*(3), 293–318.

Hibti, M., T. Friedlhuber, & A. Rauzy (2012, June). Overview of the open psa platform. In R. Virolainen (Ed.), *Proceedings of International Joint Conference PSAM'11/ESREL'12*.

Prosvirnova, T., M. Batteux, P.-A. Brameret, A. Cherfi, T. Friedlhuber, J.-M. Roussel, & A. Rauzy (2013, September). The altarica 3.0 project for model-based safety assessment. In *Proceedings of 4th IFAC Workshop on Dependable Control of Discrete Systems, DCDS'2013*, York, Great Britain, pp. 127–132. International Federation of Automatic Control.

Prosvirnova, T. & A. Rauzy (2015). Automated generation of minimal cut sets from altarica 3.0 models. *International Journal of Critical Computer-Based Systems 6*(1), 50–80.

Rauzy, A. (2008). Guarded transition systems: a new states/events formalism for reliability studies. *Journal of Risk and Reliability 222*(4), 495–505.

Rauzy, A. (2012, June). Anatomy of an efficient fault tree assessment engine. In R. Virolainen (Ed.), *Proceedings of International Joint Conference PSAM'11/ESREL'12*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Enhancement of the AltaRica 3.0 stepwise simulator by introducing an abstract notion of time

M. Batteux
IRT SystemX, Palaiseau, France

T. Prosvirnova
LGI, CentraleSupélec, Gif-sur-Yvette, France

A. Rauzy
MTP, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: AltaRica 3.0 is an event-based, object-oriented modeling language dedicated to (probabilistic) safety analyses of complex systems. It makes it possible to design models at higher level than done with formalisms traditionally used for safety analyses (fault trees, Markov Chains, stochastic Petri nets, etc.), without increasing the complexity of calculations of risk indicators. Several assessment tools have been developed for AltaRica 3.0, including a stepwise simulator. This tool is of a great help for the design and the validation of AltaRica 3.0 models. It is the analog for modeling of debuggers for programming.

In this article, we show how the AltaRica 3.0 stepwise simulator has been greatly enhanced by the introduction of an abstract notion of time. The key mathematical property is that abstract and concrete simulation are bisimilar: any concrete (timed, stochastic) execution can be simulated by an abstract execution and reciprocally any abstract execution corresponds to at least one concrete execution. This important result paves the way to the design of efficient model-checking algorithms, e.g. generators of sequences of events leading to a failure state.

1 INTRODUCTION

AltaRica 3.0 is an event-based, object-oriented modeling language dedicated to (probabilistic) safety analyses of complex systems (Prosvirnova, Batteux, Brameret, Cherfi, Friedlhuber, Roussel, & Rauzy 2013). It makes it possible to design models at higher level than done with formalisms traditionally used for safety analyses (fault trees, Markov Chains, stochastic Petri nets, etc.), without increasing the complexity of calculations of risk indicators.

The semantics of AltaRica 3.0 is defined in terms of stochastic guarded transition systems (Rauzy 2008). AltaRica 3.0 executions are similar to those of other discrete event modeling formulisms: each time a transition gets fireable, it is scheduled and possibly fired after a certain real-valued delay, see e.g. (Cassandras & Lafortune 2008, Zimmermann 1976) for introductions to (stochastic) discrete event systems. Events labeling transitions may be either deterministic or stochastic. In the later case, AltaRica 3.0 provides both built-in distributions (exponential, Weibull, . . .) and empirical distributions.

Several assessment tools have been developed for AltaRica 3.0, see e.g. (Prosvirnova & Rauzy 2015, Brameret, Rauzy, & Roussel 2015, Aupetit, Batteux, Rauzy, & Roussel 2015), including a stepwise simulator. This tool is of a great help for the design and the validation of AltaRica 3.0 models. It makes it possible to perform interactive step by step simulations, i.e. to go forth and back in sequences of events, enabling in this way to track modeling errors, unexpected behaviors and so on. With that respect, stepwise simulators play a similar role for discrete event modeling as debuggers like GDB or DDD for programming, see e.g. (Matloff & Salzman 2008) for an introduction to the latter.

In this article, we introduce the notion of abstract executions of AltaRica 3.0 models. This notion is implemented into the new version of the AltaRica 3.0 stepwise simulator. The previous version of this tool did not consider the time at all. The reason was that it would have been much

too tedious for the analyst to enter by hand the delay associated with a stochastic transition each time this transition gets fireable. Moreover, infinitely many real-valued delays can be chosen, letting the analyst pondering which one is the most suitable for her or his purpose. However, ignoring delays had a major drawback: the stepwise simulator allowed the firing of sequences of events with no counterpart in stochastic simulation and more generally that did not obey the timed semantics of AltaRica 3.0.

The idea is therefore to abstract away the time in stepwise simulation: each transition is now associated with a time interval. Firing a transition may modify the time intervals associated with already scheduled transitions. This idea is by no means new: it enters into the general framework of Cousot's abstract interpretation (Cousot & Cousot 1977). The problem at stake was to make it work for the particular case of stochastic discrete event simulations. The key mathematical property here is that abstract and (concrete, in the sense of the semantics of AltaRica 3.0) simulations are bisimilar, see e.g. (Milner 1989) for an introduction to this important notion: any (concrete) execution can be simulated by an abstract execution and reciprocally any abstract execution corresponds to at least one (concrete) execution. In a word, abstract executions are in agreement with AltaRica 3.0 semantics.

This important result paves the way to the design of efficient model-checking algorithms, see e.g. (Clarke, Grumberg, & Peled 2000) for an introduction. In particular, it makes it possible the design of generators of sequences of events leading to a failure state.

The remainder of this article is organized as follows. Section 2 presents an illustrative example. Section 3 recalls fundamental notions about timed and stochastic guarded transition systems. Section 4 introduces the notion of abstract execution of guarded transition systems. Section 5 concludes this article and gives some perspectives.

2 ILLUSTRATIVE EXAMPLE

As an illustrative example, we shall consider a system made of two identical, periodically tested, components that evolve independently one another.

The behavior of such a component is represented by the state diagram pictured Figure 1.

The component alternates operation and test phases. It is initially working and starts with a first operation phase that lasts a constant time θ. All subsequent operation phases last a constant time π.

The component may fail when in operation, with a failure rate λ.

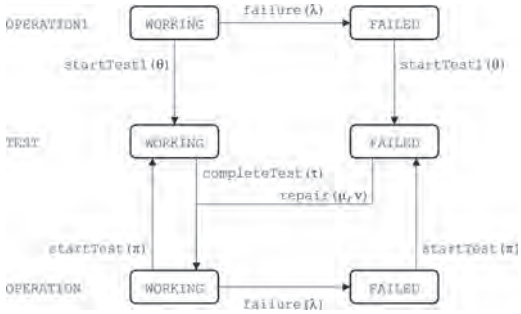


Figure 1. State diagram for a periodically tested component.

Table 1. A possible evolution of component for a periodically tested component.

Transition	Firing date
startTest1	$d_1 = \theta$
completeTest	$d_2 = d_1 + \tau$
failure	$d_3 = d_2 + \delta_1, 0 \ \delta_1 (\pi$
startTest	$d_4 = d_2 + \pi$
repair	$d_5 = d_4 + \delta_2, \mu (\delta_2 (\nu$
startTest	$d_6 = d_5 + \pi$
completeTest	$d_7 = d_6 + \tau$
⋮	⋮

If the component is working when it enters a maintenance phase, this maintenance phase lasts a constant time τ. If, on the contrary, it is failed when it enters a maintenance phase, the duration of its repair is uniformly distributed between two values μ and ν.

Finally, the component is as-good-as-new after a repair.

In the state diagram of Figure 1, transitions failure and repair are thus stochastic while transitions startTest1, startTest and completeTest are deterministic.

Table 1 shows a possible evolution of such a component.

We can assume that θ and π are relatively big compared to τ, μ and ν and that τ is smaller than μ (itself smaller than μ).

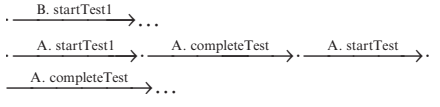
In order not to have the two components A and B out of service due to test or maintenance at the same time, it is reasonable to take different values of θ for A and for B (the values of the other parameters being identical for the two components). For instance, if an operation phase lasts normally six months ($\pi = 4380h$), the component A can be tested after three months (A.$\theta = 2190h$) while the component B is tested after six months (B.$\theta = 4380h$). It this way, tests/maintenances of A and B are shifted

Table 2. Typical values of the parameters.

parameter	A	B
θ	2190	4380
π	4380	4380
τ	0	0
μ	12	12
ν	24	24

by three months which improves the overall availability of the system. Table 2 gives some typical values of the parameters for components A and B that we shall use throughout the article.

With this values of parameters in mind, the reader sees immediately the problem of using a stepwise simulator that does not consider delays associated to transitions. Many executions that would be impossible with a timed semantics become possible with a non-timed semantics, e.g.



On other hand, asking the analyst to introduce interactively delays of stochastic transitions is not a practical solution. Not only it would be tedious, but it would let the analyst facing the choice of suitable delays, which gets quickly puzzling. Hence the need of an abstract notion of time which makes it possible to take into account delays of transitions without asking the analyst to enter them interactively.

To fulfill this need, the idea is to reason in terms of time intervals rather than in terms of dates. To explain how this idea works, we shall first recall the regular semantics of AltaRica in the next section. Then, we shall introduce its abstract semantics (in terms of time intervals) in Section 4.

3 TIMED/STOCHASTIC GUARDED TRANSITIONS SYSTEMS

The semantics of AltaRica 3.0 is defined in terms of stochastic guarded transitions systems (Rauzy 2008), (Batteux, Prosvirnova, & Rauzy 2017). We shall recall here only the notions that are important for the purpose of this article. The reader should refer to the cited articles for in depth presentations.

3.1 *Definition*

A guarded transitions system is a quintuple $S = \langle V, E, T, A, \imath \rangle$, where:

- V is a set of variables. V is the disjoint union of the set S of state variables and the set F of flow variables: $V = S \cup F$. . Each variable v of V takes its value into a finite or infinite set of constants called the domain of v and denoted as $dom(v)$. The global state of the system is thus a variable valuation, i.e. a member of the Cartesian product $\prod_{v \in V} dom(v)$.
- E is a set of events. Each event e of E is associated with a function $delay(e)$ that returns a non-negative real number. $delay(e)$ may be deterministic, in which case it returns always the same value, or stochastic, in which case it returns a value according to a certain cumulative probability distribution.
- T is a set of transitions, i.e. of triples $\langle e, G, P \rangle$, where e is an event of E, G is a Boolean expression built on variables of V (called the guard of the transition) and P is an instruction that modifies the value of state variables (called the action of the transition). For the sake of the clarity, we shall write a transition $\langle e, G, P \rangle$ as $G \xrightarrow{e} P$.
- A is an assertion, i.e. an instruction that modifies the values of flow variables.
- $\imath$ is a valuation of the variables of V, called the initial state.

Example The guarded transitions system encoding the periodically tested components described in the previous section is as follows.

The state of the component is represented by means of two state variables: state that takes its value in {WORKING, FAILED} and phase that takes its value in {OPERATION1, TEST, OPERATION}.

The events are startTest1, startTest, completeTest, failure and repair. They are associated with the delays described in the previous section.

The transitions are as follows.

$$\texttt{state} = \texttt{WORKING} \wedge \texttt{phase} \neq \texttt{TEST} \xrightarrow{\texttt{failure}} \texttt{state} \leftarrow \texttt{FAILED}$$

$$\texttt{phase} = \texttt{OPERATION1} \xrightarrow{\texttt{startTest1}} \texttt{phase} \leftarrow \texttt{TEST}$$

$$\texttt{phase} = \texttt{OPERATION} \xrightarrow{\texttt{startTest}} \texttt{phase} \leftarrow \texttt{TEST}$$

$$\texttt{state} = \texttt{WORKING} \wedge \texttt{phase} = \texttt{TEST} \xrightarrow{\texttt{completeTest}} \texttt{phase} \leftarrow \texttt{OPERATION}$$

$$\texttt{state} = \texttt{FAILED} \wedge \texttt{phase} = \texttt{TEST} \xrightarrow{\texttt{repair}} \texttt{state} \leftarrow \texttt{WORKING}, \texttt{phase} \leftarrow \texttt{OPERATION}$$

Finally, the initial state is defined by the variable valuation: state = WORKING, phase = OPERATION1.

3.2 *Composition*

One of the advantages of guarded transitions systems over some other similar formalisms is that they are highly compositional.

Formally, let $M_1 : \langle V_1, E_1, T_1, A_1, \iota_1 \rangle$ and $M_2 : \langle V_2, E_2, T_2, A_2, \iota_2 \rangle$ be two guarded transitions systems. Then the composition of M_1 and M_2, denoted as $M_1 \otimes M_2$, is simply the guarded transitions system $\langle V, E, T, A, \iota \rangle$ such that $V = V_1 \cup V_2$, $E = E_1 \cup E_2$, $T = T_1 \cup T_2$, $A = A_2 \circ A_1$ and $\iota = \iota_2 \circ \iota_1$.

The above principle extends to any number of guarded transitions systems.

Example To represent the system of the discussed example in the previous section, it suffices to create two copies of the above guarded transitions system and to compose them (which is done automatically by the AltaRica compiler).

In our example, it may be worth to introduce a flow Boolean variable failed to tell when the system is failed. The assertion defining this variable could be as follows.

failed ← A.state = FAILED ∧ B.state = FAILED

3.3 *Semantics*

The semantics of a guarded transitions system $S = \langle V, E, T, A, \iota \rangle$ is defined as the set of its possible executions (sometimes said "concrete" into this article, as opposite to abstract executions introduced into section 4).

To define formally the executions, we need to introduce the notion of schedule. A schedule of a guarded transitions system $S = \langle V, E, T, A, \iota \rangle$ is a function from T to $\mathbb{R} + \cup\{+\infty\}$.

A schedule Γ is compatible with a state σ of the guarded transitions system S and a date d if the following conditions hold for all transitions $t : G \xrightarrow{e} P$ of T.

- $d \le \Gamma(t) < +\infty$ if $G(\sigma) = true$.
- $\Gamma(t) = +\infty$ if $G(\sigma) = false$.

Intuitively, an execution of the guarded transitions system S is a sequence:

$$\langle \sigma_0, d_0, \Gamma_0 \rangle \xrightarrow{t_1} \langle \sigma_1, d_1, \Gamma_1 \rangle \xrightarrow{t_2} \ldots \xrightarrow{t_n} \langle \sigma_n, d_n, \Gamma_n \rangle$$

where $n \ge 0$, the σ_i's are states of S, the d_i's are dates, i.e. non negative real numbers verifying $0 = d_0 \le d_1 \le \ldots \le d_n$, each Γ_i is a schedule compatible with σ_i and d_i and finally the t_i's are transitions of S.

The set of valid executions of the guarded transitions system S is defined recursively as follows.

The empty execution $\langle \iota, 0, \Gamma_0 \rangle$ is a valid execution if the schedule Γ_0 is such that for all transitions $t : G \xrightarrow{e} P$ T:

- $\Gamma_0(t) = delay(e)$ if $G(\iota) = true$.
- $\Gamma_0(t) = +\infty$ if $G(\iota) = false$.

Now, if $\Lambda = \langle \sigma_0, d_0, \Gamma_0 \rangle \xrightarrow{t_1} \ldots \xrightarrow{t_n} \langle \sigma_n, d_n, \Gamma_n \rangle$, $n \ge 0$, is a valid execution, then so is the execution $\Lambda = \xrightarrow{t_{n+1}} \langle \sigma_{n+1}, d_{n+1}, \Gamma_{n+1} \rangle$ if the following conditions hold, assuming $t_{n+1} = G_{n+1} \xrightarrow{e_{n+1}} P_{n+1}$.

- $G_{n+1}(\sigma_n) = true$.
- $\sigma_{n+1} = A(P_{n+1}(\sigma_n))$, i.e. the firing of the transition t_{n+1} is performed in two steps: first, state variables are updated by means of the action P_{n+1} of the transition, then flow variables are updated by means of the assertion A.
- $d_{n+1} = \Gamma_n(t_{n+1})$ and there is no transition t of T such that $\Gamma_n(t) < \Gamma_n(t_{n+1})$.
- Γ_{n+1} is obtained from Γ_n by applying the following rules to all transitions $t : G \xrightarrow{e} P$ of T:
  - If $G(\sigma_{n+1}) = true$, then:
    - If $G(\sigma_n) = true$ and $t \ne t_{n+1}$, then $\Gamma_{n+1}(t) = \Gamma_n(t)$
    - Otherwise, $\Gamma_{n+1}(t) = d_{n+1} + delay(e)$
  - If $G(\sigma_{n+1}) = false$, then: $\Gamma_{n+1}(\tau) = +\infty$

Example Consider again our system of two components.

At time 0, 4 transitions are fireable:

Transition	Firing date
A. startTest1	2190
A. failure	5617
B. startTest1	4380
B. failure	4111

As A.startTest has the earliest firing date, it is fired (at 2190). After its firing, 3 transitions are fireable:

Transition	Firing date
A. completeTest	2190 + 0 = 2190
B. startTest1	4380
B. failure	4111

As A.completeTest has the earliest firing date, it is fired (at 2190). After its firing, 4 transitions are fireable:

Transition	Firing date
A. startTest	2190 + 4380 = 6570
A. failure	2190 + 6020 = 8210
B. startTest1	4380
B. failure	4111

As B.failure has the earliest firing date, it is fired (at 4111). After its firing, 3 transitions are fireable:

Transition	Firing date
A. startTest	6570
A. failure	8210
B. startTest1	4380

As B.startTest1 has the earliest firing date, it is fired (at 4380). After its firing, 3 transitions are fireable:

Transition	Firing date

A. startTest	6570
A. failure	8210
B. repair	4400

As B.repair has the earliest firing date, it is fired (at 4400). After its firing, 4 transitions are fireable:

Transition	Firing date
A. startTest	6570
A. failure	8210
B. startTest	4400 + 4380 = 8780
B. failure	4400 + 5201 = 9601

And so on. . .

This sequence shows how deterministic and stochastic transitions can be intricated. In particular, dates of tests are not decided once for all. They depend on times to failure and to repair of the component.

4 ABSTRACT SEMANTICS

4.1 *Principle*

The first idea to abstract the executions consists in associating an abstract delay *delay*∗ with each event of the model. *delay*∗(*e*) is simply the image of the function *delay*, i.e. an interval of non-negative real numbers.

We have to be a bit careful because some intervals that are the images of distributions are closed while some others are open (to the left and/or to the right) and that we have to consider infinite bounds. A solution consists in working only with closed intervals, but in a non-standard arithmetic built over the set $\overline{\mathbb{R}}^{+} = \mathbb{R}^{+} \cup \{\varepsilon, \infty\}$, , where ε and ∞ are respectively infinitely small and infinitely big numbers verifying: $\varepsilon + \varepsilon = \varepsilon$ and $\infty + x = \infty$ for all $x \in \overline{\mathbb{R}}^{+}$. In this way, the interval $]a, b[$, with $a, b, \in \mathbb{R}^{+}$, can be encoded as $[a + \varepsilon, b - \varepsilon]$. Table 3 gives the abstract delays associated with the most widely used distributions in AltaRica 3.0. Moreover, a transition whose guard is not satisfied in the current state is scheduled in the interval $[\infty, \infty]$.

The second idea is to consider not the date at which transitions are fired, but an interval of time within which they are fired.

Assume that we are building the sequence under study step by step and that the last transition we considered must be fired in the time interval $[l, h]$. Assume moreover that transitions $t_1, t_2, \ldots t_n$ are scheduled in time intervals $[l_1, h_1], [l_2, h_2], \ldots, [l_n, h_n]$. Then, we can make the following remarks.

1. We must have $l \le l_i$ for all $i = 1, \ldots, n$, because the next transition cannot be scheduled in the
2. We must have also $h \le h_i$ for all $i = 1, \ldots, n$, because if $h_i < h$ for some i, it means that the transition t_i must be fired before h_i, therefore the last transition must also be fired before h_i.
3. For the same reason, we can choose t_i as the next transition to be fired only if there is no other transition t_j such that $h_j < l_i$.
4. Again for the same reason, if the transition t_i is fired, it is necessarily fired in the interval $[l_i, h_{min}]$, where h_{min} is the smallest of the h_j's.
5. If the transition t_i is fired and the transition tj is such that $l_j < l_i$, then l_j must be changed to l_i so to obey our first remark.
6. Finally, if the transition t_i is fired and a transition t associated with the interval $[l_t, h_t]$ becomes fireable (t can be the transition t_i itself), then t must be scheduled in the interval $[l_t, h_{min}] + [l_t, h_t] = [l_i + l_t, h_{min} + h_t]$.

We are now able to define formally the abstract semantics of guarded transitions systems (and therefore for AltaRica 3.0).

4.2 *Formal definition*

The abstract semantics of a guarded transitions system $S = \langle V, E, T, A, \iota \rangle$ is defined as the set of its possible abstract executions.

To define formally the abstract executions, we need to introduce the notion of abstract schedule. An abstract schedule of a guarded transitions system $S = \langle V, E, T, \underline{A}, \iota \rangle$ is a function from T to closed intervals over $\overline{\mathbb{R}}^{+}$.

A schedule $\Gamma*$ is compatible with a state σ of the guarded transitions system S and the abstract date $[l, h]$ if the following conditions hold for all transitions $t : G \xrightarrow{e} P$ of T, with $\Gamma*(t) = [l_t, h_t]$.

- $l \le l_t < \infty$ and $h \le h_t$ if $G(\sigma) = true$.
- $l_t = h_t = \infty$ if $G(\sigma) = false$.

An abstract execution of the guarded transitions system S is a sequence:

$$\langle \sigma_0, d_0^*, \Gamma_0^* \rangle \xrightarrow{t_1} \langle \sigma_1, d_1^*, \Gamma_1^* \rangle \xrightarrow{t_2} \ldots \xrightarrow{t_n} \langle \sigma_n, d_n^*, \Gamma_n^* \rangle$$

where $n \ge 0$, the σ_i's are states of S, the d_i^*'s are abstract dates, i.e. time intervals, each Γ_i^* is an abstract schedule compatible with σ_i and d_i^* and finally the t_i's are transitions of S.

The set of valid abstract executions of the guarded transitions system S is defined recursively as follows.

Table 3. Intervals associated with delay functions.

Concrete delay	Abstract delay
Dirac (t)	$[t, t]$
Uniform Deviate (l, h)	$[l, h]$
Exponential (λ)	$[0 + \varepsilon, [$
Weibull (α, β)	$[0 + \varepsilon, [$
Empirical distribution	$[0 + \varepsilon, [$

The empty abstract execution $\langle \sigma_0, [0, 0], \Gamma_0^* \rangle$ is a valid abstract execution if the abstract schedule Γ_0^* is such that for all transitions $t : G \xrightarrow{e} P$ of T:

- $\Gamma_0^*(t) = delay^*(e)$ if $G(t) = true$. .
- $\Gamma_0^*(t) = [\infty, \infty]$ if $G(t) = false$.

If $\Lambda = \langle \sigma_0, [0,0], \Gamma_0^* \rangle \xrightarrow{t_1} \ldots \xrightarrow{t_n} \langle \sigma_n, [l_n, h_n], \Gamma_n^* \rangle$, $n \geq 0$, is a valid abstract execution, then so is the abstract execution, then so is the abstract execution, then so is the abstract execution $\Lambda \xrightarrow{t_{n+1}} \langle \sigma_{n+1}, [l_{n+1}, h_{n+1}], \Gamma_{n+1}^* \rangle$ if the following conditions hold, assuming

$t_{n+1} = G_{n+1} \xrightarrow{e_{n+1}} P_{n+1}$, $\Gamma_n^*(t_{n+1}) = [l^*, h^*]$ and $h_{min} = \min_{[l,h]=\Gamma_n^*(t), t \in T} h$.

- $G_{n+1}(\sigma_n) = true$.
- $\sigma_{n+1} = A(P_{n+1}(\sigma_n))$.
- There is no transition t of T such that $\Gamma_n^*(t) = [l, h]$ and $h < l*$.
- $[l_{n+1}, h_{n+1}] = [l*, h_{min}]$.
- $\Gamma_n^*(t)$ is obtained from Γ_n^* by applying the following rules to all transitions $t : G \xrightarrow{e} P$ of T and $\Gamma_n^*(t) = [l, h]$.
  - If $G(\sigma_{n+1}) = true$, then:
    - If $G(\sigma_n) = true$ and $t \neq t_{n+1}$, then

$$\Gamma_{n+1}^*(t) = [\max(l_{n+1}, l), h]$$

    - Otherwise,

$$\Gamma_{n+1}^*(t) = [l_{n+1}, h_{n+1}] + delay^*(e)$$

  - If $G(\sigma_{n+1}) = false$, then:

$$\Gamma_{n+1}^*(t) = [\infty, \infty]$$

Example We shall consider the abstract version of the execution given in the previous section.

At time 0, 4 transitions are fireable:

Transition	Abstract date
A.startTest1	[2190, 2190]
A.failure	[0 + ε, ∞]
B.startTest1	[4380, 4380]
B.failure	[0 + ε, ∞]

A.startTest1 is fired at the abstract date [2190, 2190]. After its firing, 3 transitions are fireable:

Transition	Abstract date
A.completeTest	[2190, 2190] + [0, 0] = [2190, 2190]
B.startTest1	[4380, 4380]
B.failure	2190, ∞]

A.completeTest is fired at the abstract date [2190; 2190]. After its firing, 4 transitions are fireable:

Transition	Abstract date
A.startTest	[2190, 2190]+[4380, 4380] = [6570, 6570]
A.failure	2190, 2190] + [0 + ε, ∞] = [2190 + ε, ∞]
B.startTest1	[4380, 4380]
B.failure	[2190, ∞]

B.failure is fired at the abstract date [2190 + ε, 4380]. After its firing, 3 transitions are fireable:

Transition	Abstract date
A.startTest	[6570, 6570]
A.failure	[2190 + ε, ∞]
B.startTest1	[4380, 4380]

B.startTest1 is fired at the abstract date [4380, 4380]. After its firing, 3 transitions are fireable:

Transition	Abstract date
A.startTest	6570, 6570]
A.failure	[4380, ∞]
B.repair	[4380, 4380] + [12, 24] = [4392, 4404]

B.repair is fired at the abstract date [4392, 4404]. After its firing, 4 transitions are fireable:

Transition	Abstract date
A.startTest	[6570, 6570]
A.failure	[4392 + ε, ∞]
B.startTest	[4392, 4404] + [4380, 4380] = [8872, 8884]
B.failure	[4392, 4404] + [0 + ε, ∞] = [4392 + ε, ∞]

And so on…

4.3 *Bisimulation*

The key mathematical property is that abstract and concrete executions are bisimilar: any concrete (timed, stochastic) execution can be simulated by an abstract execution and reciprocally any abstract execution corresponds to at least one concrete execution.

Theorem 1. *Let $S = \langle V, E, T, \iota, A \rangle$ be a guarded transitions system. Let $\Lambda = \langle \sigma_0, d_0, \Gamma_0 \rangle \xrightarrow{t_1} \ldots \xrightarrow{t_n} \langle \sigma_n, d_n, \Gamma_n \rangle$, be a (concrete) execution of S. Then it exists an abstract execution $\Lambda_a = \langle \sigma_0, d_0^*, \Gamma_0^* \rangle \xrightarrow{t_1} \langle \sigma_1, d_1^*, \Gamma_1^* \rangle \xrightarrow{t_2} \ldots \xrightarrow{t_n} \langle \sigma_n, d_n^*, \Gamma_n^* \rangle$, such that the following properties hold:*

- $\forall n \geq 0 d_n \in d_n^*$;
- $\forall n \geq 0 \forall t \in T\ \Gamma_n(t) \in \Gamma_n^*(t)$.

Theorem 2. *Let $S = \langle V, E, T, \iota, A \rangle$ be a guarded transitions system. Let $\Lambda_a = \langle \sigma_0, d_0^*, \Gamma_0^* \rangle \xrightarrow{t_1} \langle \sigma_1, d_1^*, \Gamma_1^* \rangle \xrightarrow{t_2} \ldots \xrightarrow{t_n} \langle \sigma_n, d_n^*, \Gamma_n^* \rangle$ be an abstract execution of S. Then Λ_a corresponds to at least one (concrete) execution $\Lambda = \langle \sigma_0, d_0, \Gamma_0 \rangle \xrightarrow{t_1} \ldots \xrightarrow{t_n} \langle \sigma_n, d_n, \Gamma_n \rangle$, such that the following properties hold:*

- $\forall n \geq 0 d_n \in d_n^*$;
- $\forall n \geq 0 \forall t \in T\ \Gamma_n(t) \in \Gamma_n^*(t)$.

Proofs of these two previous theorems are done recursively on the abstract executions or (concrete)

executions. They are out of the scope of this article and are not described.

5 CONCLUSION AND PERSPECTIVES

In this article, we introduce the notion of abstract executions of AltaRica 3.0 models. This notion implemented into the new version of the AltaRica 3.0 stepwise simulator. This notion of abstract executions enables to reconcile both stochastic and stepwise simulations of AltaRica 3.0 models.

We show that abstract and (concrete) simulations are bisimilar: any (concrete) execution can be simulated by an abstract execution and reciprocally any abstract execution corresponds to at least one (concrete) execution.

We illustrate our purpose using a motivating example that mix both stochastic and deterministic transitions.

The introduction of the notion of abstract executions to the stepwise simulator paves the way to the design of efficient model-checking algorithms, and in particular to the design of generators of sequences of events leading to a failure state.

The next step of our work is the application of the presented results for the development of an efficient sequence generator for AltaRica 3.0 models.

REFERENCES

Aupetit, B., M. Batteux, A. Rauzy, & J.-M. Roussel (2015, September). Improving performance of the AltaRica 3.0 stochastic simulator. In L. Podofillini, B. Sudret, B. Stojadinovic, E. Zio, and W. Kro¨ger (Eds.), *Proceedings of Safety and Reliability of Complex Engineered Systems: ESREL 2015*, pp. 1815–1824. CRC Press.

Batteux, M., T. Prosvirnova, & A. Rauzy (2017, September). Altarica 3.0 assertions: the why and the wherefore. *Journal of Risk and Reliability*.

Brameret, P.-A., A. Rauzy, & J.-M. Roussel (2015, July). Automated generation of partial markov chain from high level descriptions. *Reliability Engineering and System Safety 139*, 179–187.

Cassandras, C.G. & S. Lafortune (2008). *Introduction to Discrete Event Systems*. New-York, NY, USA: Springer. Clarke, E.M., O. Grumberg, & D.A. Peled (2000, February). *Model Checking*. Cambridge, MA, USA: MIT Press.

Cousot, P. & R. Cousot (1977). Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints. In *Conference Record of the Fourth Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, New York, NY, USA, pp. 238–252. ACM Press. Los Angeles, California.

Matloff, N. & P.J. Salzman (2008). *The Art of Debugging with GDB, DDD, and Eclipse*. San Fransisco, CA, USA: No Starch Press.

Milner, R. (1989). *Communication and Concurrency*. Prentice-Hall international series in computer science. Upper Saddle River, New Jersey, USA: Prentice Hall.

Prosvirnova, T., M. Batteux, P.-A. Brameret, A. Cherfi, T. Friedlhuber, J.-M. Roussel, & A. Rauzy (2013, September). The altarica 3.0 project for model-based safety assessment. In *Proceedings of 4th IFAC Workshop on Dependable Control of Discrete Systems, DCDS'2013*, York, Great Britain, pp. 127–132. International Federation of Automatic Control.

Prosvirnova, T. & A. Rauzy (2015). Automated generation of minimal cutsets from altarica 3.0 models. *International Journal of Critical Computer-Based Systems 6*(1), 50–79.

Rauzy, A. (2008). Guarded transition systems: a new states/events formalism for reliability studies. *Journal of Risk and Reliability 222*(4), 495–505.

Zimmermann, A. (1976). *Stochastic Discrete Event Systems*. Berlin, Heidelberg, Germany: Springer.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Reliability forecasting of components/systems in automobile applications by using two-dimensional stress functions

Abderrahim Krini
Robert Bosch GmbH Engineering of Remanufacturing and Quality, Schwäbisch Gmünd, Germany

Josef Börcsök
Department of Computer Architecture and System Programming, University of Kassel, Kassel, Germany

ABSTRACT: For remanufacturing automobile subsystems, the knowledge of future failure rates of systems produced in serial production is from great importance. Remanufacturing departments especially need information about the number of cores to remanufacture, the condition of those cores as well as the number of replacement system the market requires. On the basis of an already existing model that predicts failure rates using warranty data as input, a bivariate prognosis model is developed, which takes two variables of stress in field into account. By enhancing the existing model a more precise prognosis of future failure rates is expected as well as a better knowledge of the damage. This paper presents all necessary mathematical tools to perform a bivariate lifetime prognosis as well as the implementation in MATLAB.

1 INTRODUCTION

After serial production, automotive suppliers have the obligation to assure post series supply of replacement components and systems. The demanded number of replacement systems influence the production method. When the demand drops, from one point on, the remanufacturing of old systems represents the most economical method. In order to optimize the strategy of after series production, a suitable prognosis is needed. Requirement for this prognosis is a precise prediction of future failure rates in field in order to estimate the number of returning cores as well as the demand for replacement systems. Additionally, an accurate picture of the stress in field, the cores have experienced, is an advantage. It gives information about the costs that should be taken into account for remanufactureing the systems. On the basis of a univariate prognosis model by Pauli and Meyna [1], [2], [3] a bivariate model is developed. This model predicts future failures in field with respect to two different variables that represent the stress in field. Data basis for the model are warranty data. The knowledge of the distribution of stress in field for a car model allows a transformation to predict failure rates as a function of time. The prognosis is carried out by using bivariate stochastic distribution functions. Methods of fitting bivariate functions and comparing the goodness of fit between several distributions are presented. The discussed multivariate prognosis method has to be categorized between univariate prognosis models and prognosis models using neuronal networks.

2 THEORETICAL BACKGROUND

This chapter gives a brief overview of the basic tools the model is based on.

2.1 *Univariate prognosis model as basis for the new model*

As basis for reliability prediction models in general, data can be collected from different data origins: Laboratory testing, testing in field or real field data from the customers. The disadvantage of Laboratory testing and testing in field is the small sample size as well as the incomplete consideration of the variety of stress in field. Therefore, Pauli and Meyna developed an approach to predict the reliability of automobiles as well as automotive subsystems by using warranty data. These data must contain: Time in field until failure and mileage until failure for at least 50 failures. This method will be described briefly. For a better understanding, consider also reading references: [1], [2], [4] and [3].

In order to describe the stress in field, the driven mileage until failure is a suitable variable. Time in field is inappropriate due to varying user behavior [1].

Varying usage of automobiles can be described by the mileage distribution. Considering time in field and mileage (until failure), under the assumption of a linear driving behavior of each driver, one can develop a mileage distribution for a certain time period by using warranty data. Theoretical distribution functions can then be fitted to empirical data:

By cumulating the number of failures that occurred within a certain mileage interval, e.g. 3001 to 4000 km, a failure frequency is obtained. Based on the mileage distribution of the warranty period, it is known that only a share of all components of the sample has reached a certain mileage interval within warranty period. Until all components will have reached this interval, a certain number of additional failures are expected. The so called candidates can be estimated by correcting the observed failures by considering the mileage distribution for each interval.

In order to calculate a time dependent failure probability, it is necessary to fit a theoretical distribution function on the corrected failure frequency. By normalizing the resulting function with respect to the sample size, the mileage dependent failure probability can be obtained. The transformation to a time dependent failure probability is carried out by utilizing the mileage distribution for each date.

2.2 *Bivariate reliability parameters*

Considering the probability of failure as a bivariate function, the lifetime of a component is characterized by two positive continuous random variables X_1 and X_2. The system fails if both variables of stress, x_1 and x_2, exceed X_1 and X_2:

$$F(x_1,x_2) = P(X_1 \le x_1, X_2 \le x_2) \quad \text{for } x_1,x_1 \ge 0 \quad (1)$$

$F(x_1, x_2)$, the failure distribution function, indicates the probability for a component that experienced the stress x_1 and x_2 to be failed.

For $x_1, x_2 \to \infty$ the component fails:

$$\lim_{x_1,x_2\to\infty} F(x_1,x_2) = 1 \quad (2)$$

The derivation of $F(x_1, x_2)$ with respect to x_1 and x_2 gives information about the density of failures at a certain combination of stress:

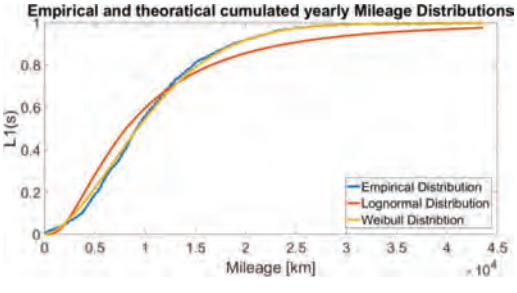


Figure 1. Empirical yearly mileage distribution, fitted lognormal distribution and fitted Weibull distribution.

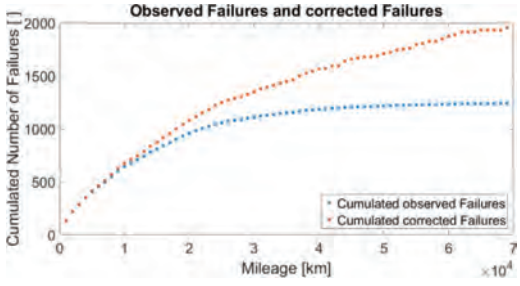


Figure 2. Observed failures and corrected failures.

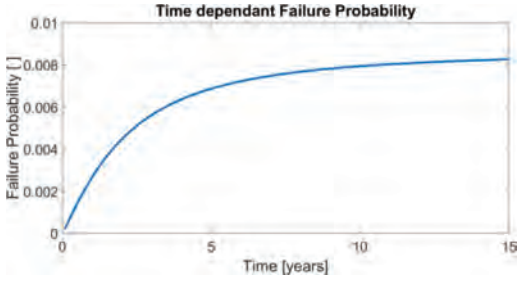


Figure 3. Time dependent failure probability.

$$f(x_1,x_2) = \frac{\partial^2 F}{\partial x_1 \partial x_2} \quad \text{for} \quad x_1,x_2 \ge 0 \quad (3)$$

The reliability R(t) is the complement of F(t). A component survives if one or both variables of stress, x_1 and x_2, remain below the random variables X_1 and X_2 [6][7].

$$\begin{aligned} R(x_1,x_2) &= P(X_1 > x_1, X_2 > x_2) \\ &= 1 - P(X_1 \le x_1, X_2 \le x_2) = 1 - F(x_1,x_2) \end{aligned} \quad (4)$$

By dividing the failure density through the reliability the hazard rate $h(x_1, x_2)$ can be determined. The hazard rate is an important indicator for the reliability of a component.

$$h(x_1,x_2) = \frac{f(x_1,x_2)}{R(x_1,x_2)} \quad (5)$$

Both, the failure distribution function and the reliability function, can be determined by empirical data. All other parameters can be derived. For a better distinction, the empirical functions are marked with a tilde. The sample size of all examined components (n_0), consists of the subsets of failed components (n_a) and surviving components (n_b):

$$n_0 = n_a(x_1,x_2) + n_b(x_1,x_2) \quad (6)$$

The empirical probability of failure is defined as:

$$\tilde{F}(x_1, x_2) = \frac{n_a(x_1, x_2)}{n_0} \tag{7}$$

The complement $\tilde{R}(t)$ is defined as:

$$\tilde{R}(x_1, x_2) = \frac{n_b(x_1, x_2)}{n_0} \tag{8}$$

In equations (7) and (8), the cumulated failures n_a and n_b must be used.

2.3 *Bivariate distribution functions*

For most of the common univariate distribution function, multivariate counterparts exist. Most important fields for application of such functions are financial (insurance) mathematics, meteorology and reliability statistics. In this chapter important bivariate distributions, in the context of reliability of automobile components, are presented.

In general, bivariate distribution functions are a useful tool for a dataset that fits the following criteria:

- The distribution of the events fits the same univariate distributions for each variable.
- The correlation between both variables is neither 0 nor ± 1.

In [5] a distribution function of the **bivariate normal distribution** is defined as follows:

$$f_{x_1x_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} * \exp\left\{ -\frac{1}{2(1-\sigma^2)} \left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x_1-\mu_1}{\sigma_1} * \frac{x_2-\mu_2}{\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 \right] \right\} \tag{9}$$

The parameters stand for:

For a given distribution function, the cumulative distributions function can be derived by double integrating the distribution function with respect to both variables [5].

Table 1. Parameters of the bivariate distribution function.

Parameter	Meaning	Domain of definition
μ_1, μ_2	First moment of X_1, X_2	$-\infty\ \mu_i; \mu_2\ \infty$
σ_1^2, σ_2^2	Variance of X_1, X_2	$\sigma_1^2; \sigma_2^2 > 0$
ρ	Correlation coefficient of X_1 and X_2	$-1 < \rho < 1$

$$F_{x_1x_2}(x_1, x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{v_1v_2}(v_1, v_2)\, dv_2\, dv_2 \tag{10}$$

Another important distribution for reliability models is the **bivariate lognormal distribution**. In [6] the distribution function is defined as:

$$f(x_1, x_2) = \frac{1}{2\pi x_1 x_2 \sigma_{y_1}\sigma_{y_2}\sqrt{1-\rho^2}} * \exp\left\{ -\frac{1}{2(1-\rho^2)} * \left[\left(\frac{\ln(x_1)-\mu_{y_1}}{\sigma_{y_1}}\right)^2 - 2\rho * \left(\frac{\ln(x_1)-\mu_{y_1}}{\sigma_{y_1}}\right) * \left(\frac{\ln(x_2)-\mu_{y_2}}{\sigma_{y_2}}\right) + \left(\frac{\ln(x_2)-\mu_{y_2}}{\sigma_{y_2}}\right)^2 \right] \right\} \tag{11}$$

One way to estimate the parameters of the distribution is the method of moments. On the basis of empirical failure data, an estimation for the parameters can be calculated using the equations given in (12):

$$\begin{aligned} &\sigma_{y_1}: \text{standard deviation} \\ &\sigma_{y_i} = \left[\log\left(1 + \frac{\sigma_{x_i}^2}{\mu_{x_i}^2} \right) \right]^{1/2} \\ &\mu_{y_1}: \text{first moment } \log(\mu_{x_i}) - \left(\frac{\sigma_{y_i}^2}{2}\right) \\ &\rho: \text{correlation coefficient} \\ &\rho = \frac{E[(Y_1-\mu_{y_1})(Y_2-\mu_{y_2})]}{\sigma_{y_1} * \sigma_{y_2}} \end{aligned} \tag{12}$$

The cumulative distribution function can be derived by applying equation (10).

Especially the reliability of electronic components can often be described by the exponential distribution. In [7] a **bivariate exponential distribution**, developed by Gumbel, is described.

$$F(x_1, x_2) = e^{-(\alpha x_1 + \beta x_2 + \theta\alpha\beta x_1 x_2)} \quad for \quad 0 < \theta < 1 \tag{13}$$

By derivating the cumulative distribution function with respect to both variables, the following distribution function can be obtained:

$$f(x_1, x_2) = \left\{ (1-\theta)\alpha\beta + \theta\alpha^2\beta x_1 + \theta\alpha\beta^2 x_2 + \theta^2\alpha^2\beta^2 x_1 x_2 \right\} F(x_1, x_2) \tag{14}$$

With:

$$E(X_1) = \frac{1}{\alpha}; E(X_2) = \frac{1}{\beta} \tag{15}$$

For the univariate Weibull distribution several bivariate counterparts were developed, for example by Hougaard [8]. Because of the difficult applicability of these functions, in the case of the bivariate Weibull distribution, the developed model works with copula functions. A copula function is a function that consists of marginal distributions as variables but contains additional parameters. For the developed reliability forecast model, three different copula functions fit the need. The goodness of fit with empirical failure data for all three functions is comparable [9].

Gumbel Copula:

$$F(x_1,x_2) = e^{-\left[\left(-\log(F_1(x_1))\right)^{\theta}+\left(-\log(F_2(x_2))\right)^{\theta}\right]^{\frac{1}{\theta}}} \quad \text{for} \quad \theta \geq 1 \tag{16}$$

Clayton Copula:

$$F(x_1,x_2) = \left(F_1(x_1)^{-\theta} + F_2(x_2)^{-\theta} - 1\right)^{-\frac{1}{\theta}} \quad \text{for} \quad \theta > 0 \tag{17}$$

Frank Copula:

$$F(x_1,x_2) = -\frac{1}{\theta} log\left[1 + \frac{\left(e^{-\theta F_1(x_1)} - 1\right)\left(e^{-\theta F_2(x_2)} - 1\right)}{e^{-\theta} - 1}\right] \quad \text{for} \quad \theta \neq 0 \tag{18}$$

$F_1(x_1)$ and $F_2(x_2)$ represent the cumulated marginal Weibull distributions of x_1 and x_2.

2.4 *Fitting of bivariate distribution functions*

The difficulty in dealing with multivariate distribution functions is the estimation of appropriate parameters. Most important methods will be explained briefly.

Method of moments
The method specified here makes use of the mathematical relationship between parameters of the function and characteristic values of the sample data. First step is the expression of the parameters of the function as a formula of moments of the distribution. The value of the moments is then calculated on the basis of the sample data. Important moments are for example mean and variance of a distribution. [5]

Maximum likelihood
A maximum likelihood estimate is the value of a parameter for which the sample data gain the highest possibility. For fitting univariate distributions, maximum likelihood is the predefined method MATLAB uses. [5]

Method of least squares
This method estimates parameters by minimalizing the deviation of the empirical and the theoretical distribution. The deviation is measured by the sum of the squared residuals. Those parameters for which this sum is minimal are chosen. The sum of squared residuals (SSE) serves as an index to compare the goodness of fit of several fitted theoretical distributions. [10] [11]

3 DISTRIBUTION OF STRESS IN FIELD

The above described univariate prognosis model uses the driven mileage as a variable to describe the stress in field. When dealing with ECU's, the finite number of storage processes the non-volatile storage can handle before wearing out, should be considered. Therefore switching cycles[1] are considered as an additional appropriate variable when modelling electronical components.

The sample data used in this paper contains failure data of ECU's in automobile application. The prognosis is carried out on basis of time in field, mileage and switching cycles at the time of failure for each sample. Nevertheless, the model is not limited on electronic components. It can be applied on all automobile systems and subsystems when two appropriate variables are determined to describe stress in field.

Exposure profiles of automobiles vary greatly from customer to costumer. To take this variety into account a distribution of stress in field is formed on the basis of all failures. A linear increase of mileage and switching cycles over time is assumed. An empirical distribution of stress in field, $lz_t(s, z)$, is obtained by considering all failure data, normalized to a certain time period. For example one year:

$$s_1 = s_F * \frac{t_1}{t_F}; \quad z_1 = z_F * \frac{t_1}{t_F} \tag{19}$$

S_1, z_1: mileage, switching cycle during one year
S_F, z_F: mileage, switching cycle until failure
t_1: one year
t_F: time in field until failure

[1]Switching cycle: Turning the ignition key in the off position forces the ECU to write data to the non-volatile memory.

By using all of the above described methods for fitting distributions, a bivariate normal-, lognormal-, exponential-distribution as well as a Weibull copula function can be fitted to the sample data. A brief description of the implemented fitting procedure for each distribution is given in the following:

Bivariate normal distribution
All parameters of the bivariate normal distribution are estimated by the method of moments. Values were calculated using the sample data (Table 1).

Bivariate lognormal distribution
On basis of mean, variance and covariance of the sample, the parameters were estimated, using equations (12).

Bivariate exponential distribution
Parameters α and β can be estimated by the first moments of the sample data, according to (15). Afterwards an estimator for θ is found by the method of least squares.

Bivariate Weibull distribution (Copula)
Both marginal distributions are estimated using maximum likelihood. In MATLAB this operation is carried out by "wblfit". Parameter θ for equations (16), (17) and (18) is then estimated by the method of least squares.

According to goodness of fit, carried out by SSE, the bivariate lognormal distribution and the bivariate Weibull distribution fit the sample data best:

The sum of squared residuals of the bivariate Weibull distribution, with a value of 0,00255, is smaller than the sum of squared residuals of the bivariate lognormal distribution with 0,00291. Therefor the stress in field for further calculations is considered to be a bivariate Weibull distribution with estimated parameters. The cumulative distribution function $LZ_t(s, z)$ is derived by double integrating $lz_t(s, z)$.

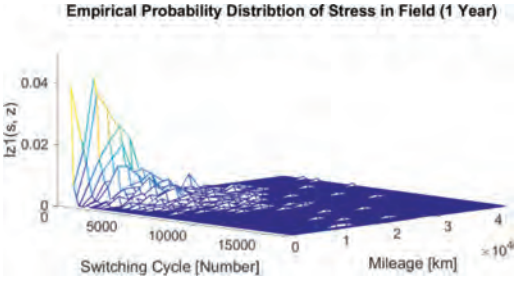


Figure 4. Empirical probability distribution of stress in field during a period of one year.

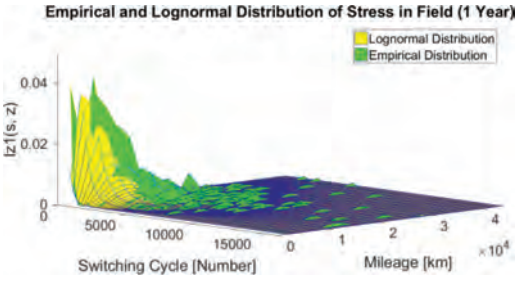


Figure 5. Empirical probability distribution of stress in field during a period of one year and fitted bivariate lognormal distribution.

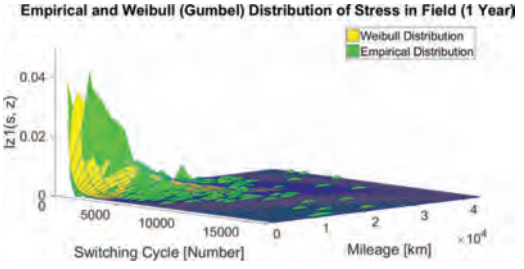


Figure 6. Empirical probability distribution of stress in field during a period of one year and fitted bivariate Weibull distribution (Gumbel copula).

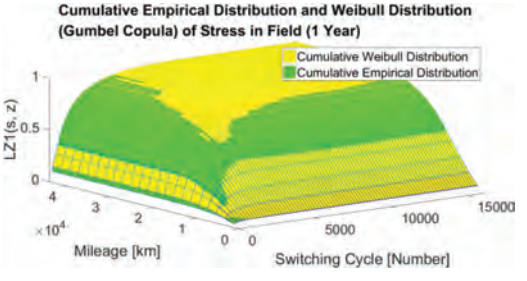


Figure 7. Cumulative empirical probability distribution of stress in field during a period of one year and fitted cumulative, bivariate Weibull distribution (Gumbel copula).

$$LZ_t(s,z) = \int_0^s \int_0^z lz_t(\sigma,\zeta)\, d\zeta\, d\sigma \qquad (20)$$

The cumulative distribution function can be transformed for arbitrary time periods, using the following equation:

$$LZ_t(s,z) = LZ_1\left(s\frac{t_1}{t}, z\frac{t_1}{t}\right) = LZ_G\left(s\frac{t_G}{t}, z\frac{t_G}{t}\right) \qquad (21)$$

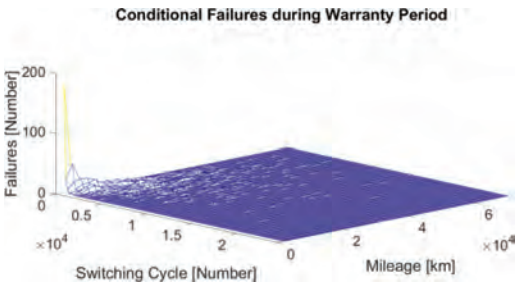


Figure 8. Distribution of the conditional failures during warranty period as a function of the class of stress in field.

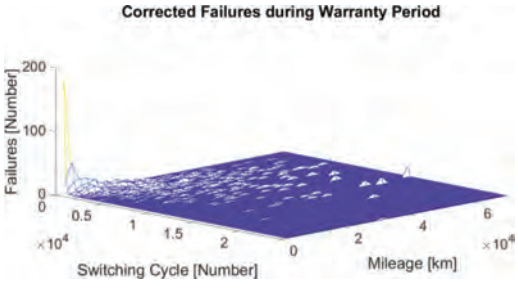


Figure 9. Distribution of the corrected failures during warranty period as a function of the class of stress in field.

LZ_1: cumulative distribution function for the period of one year

LZ_G: cumulative distribution function for the warranty period

t_G: warranty period

4 ESTIMATION OF CANDIDATES

Regarding warranty data, the sample is divided in two subsets: complained components and fully functional components still in field. For every complained component additional information exists:

- Time in field
- Mileage
- Switching cycles

For an efficient further processing in MATLAB, the data set of failed components is to be discretized. For the described application, a class size of 2000 km × 200 switching cycles was chosen. Each class contains $n_a(s, z)$ failures. These failures are called *conditional failures*, because until all components will have reached the considered class, additional failures are expected to occur: the *candidates*.

Considering the distribution of stress in field, the ratio of components that has already reached a certain class of stress during warranty period, can be estimated:

$$P(S \geq s, Z \geq z) = 1 - LZ_G(s,z) \tag{22}$$

The number of corrected failures n_k for each class composes of all conditional failures and all candidates.

$$n_k(s,z) = \frac{n_a(s,z)}{1 - LZ_G(s,z)} \tag{23}$$

5 STRESS DEPENDANT LIFETIME PREDICTION

Fitting an adequate theoretical distribution function to the distribution of corrected failures allows a forecast of the stress dependent lifetime of the regarded component. Procedure of fitting bivariate distributions is similar to the methods described in 3. The corrected failure density is calculated using the following equation:

$$\tilde{f}_k(s,z) = \frac{n_k(s,z)}{n_0} \tag{24}$$

A bivariate exponential, normal and lognormal distribution was fitted as well as a bivariate Weibull (copula) distribution. According to goodness of fit, carried out by SSE, the bivariate lognormal distribution and the bivariate Weibull distribution fit the corrected failure density $\tilde{f}_k(s,z)$ best:

For the calculation of a time dependent lifetime prediction goodness of fit of the theoretical distribution functions has to be compared. According to SSE, the Weibull distribution fits the data best.

SSE lognormal distribution: 33065

SSE Weibull distribution: 16945

For the present data set, the time dependent lifetime prediction is best carried out by a $\tilde{f}_k(s,z)$ represented by the Weibull distribution.

6 TIME DEPENDANT LIFETIME PREDICTION

To transfer the stress dependent lifetime distribution, determined in 5, in a time dependent distribution the distribution of stress in field must be considered:

$$F(t) = \int_0^\infty \int_0^\infty \tilde{f}_k(s,z) * \left(1 - LZ_1\left(\frac{s}{t}, \frac{z}{t}\right)\right) ds\, dz \tag{25}$$

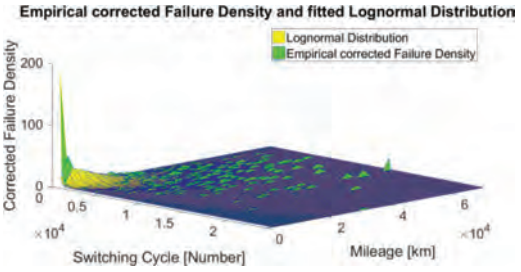


Figure 10. Empirical corrected failure density and the fitted lognormal distribution.

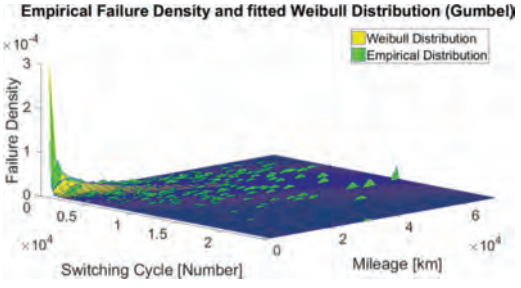


Figure 11. Empirical failure density and fitted Weibull distribution using Gumbel copula.

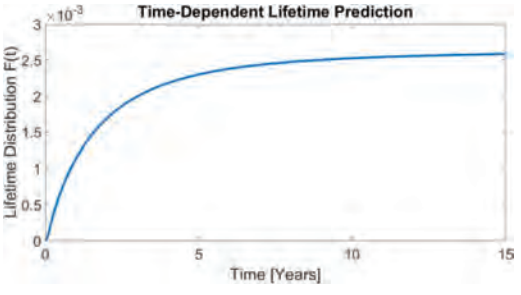


Figure 12. Time-dependent lifetime distribution as the result of a bivariate prognosis model.

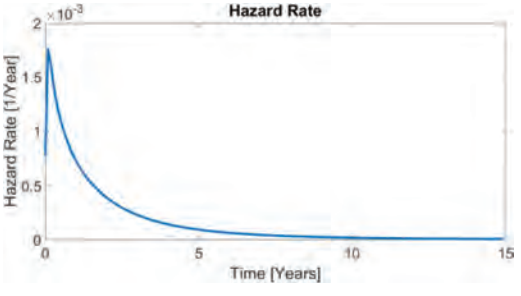


Figure 13. Time-dependent hazard rate.

The result of the bivariate prognosis model is the time-dependent lifetime prediction.

A time-dependent hazard rate can be determined using the following formula:

$$h(t) = \frac{f(t)}{R(t)} = \frac{dF(t)}{dt} * \frac{1}{1-F(t)} \tag{26}$$

The hazard rate shows a typical early failure characteristic.

7 ASSUMPTIONS/LIMITATIONS

Underlying the model a few assumptions concerning the data and the usage behavior were made:

- The warranty database is complete and all failures of the regarded period of time are registered.
- The warranty database contains an appropriate number of samples (Recommendation: Minimum 1 000 failed samples).
- A constant increase of mileage over time is considered for the complete lifetime of the component/the automobile.

The basic limitation of the described model is the ability to only describe one type of failure behavior (e.g. early failures, random failures or wear out failures). In order to describe the full life cycle of a product consider the superposition of three independent predictions.

8 CONCLUSIONS

The previous chapters prove the possibility to create a bivariate reliability forecast model as well as the successful implementation of said model in MATLAB. Comparing the calculated forecast with real field failure data, good accordance was determined. As a result it can be said that a model on basis of bivariate distribution functions with respect to two variables of stress in field can successfully be applied in quality management of automobile systems.

ACKNOWLEDGMENTS

Our thanks go to Christian Lachenmaier who put a great deal of data processing in the topic of bivariate prognosis models, associated with his master studies at Robert Bosch Automotive Steering GmbH.

REFERENCES

[1] A. Krini et al. (2015) "*Investigation of reliability for the remanufacturing of safety related automotive steering systems*," ICAT15, Turkey.

[2] B. Pauli (1999), "Eine neue Methode zur Bestimmung der kilometerabhängigen Lebensdauerverteilung von Kfz-Komponenten," *ATZ Automobiltechnische Zeitschrift (101)*.
[3] A. Krini et al. (2018), "*New approach to remanufacturing automotive steering systems with a reliability predict model*," ATINER15, Greece.
[4] A. Krini (2014), *A new reliability prevention model for reman systems in automotive*, Schwäbisch Gmünd: Bosch automotive steering.
[5] J.F. Lawless (1983): Statistical Methods in Reliability with Discussion. Technimetrics 25/4.
[6] B. Pauli (1999), "Eine neue Methode zur Bestimmung der kilometerabhängigen Lebensdauerverteilung von Kfz-Komponenten," *ATZ Automobiltechnische Zeitschrift 101*, pp. 256-261, 1999.
[7] B. Pauli & A. Meyna (2000), "Zuverlässigkeitsprognose für Kfz-Komponenten bei unvollständigen Daten," *ATZ Automobiltechnische Zeitschrift.*
[8] A. Krini et al. (2015) "*Investigation of a reliability prevention model for systems in automotive*," ICAT15, Bosnia & Herzegovina.
[9] W. Zucchini et al. (2017) "Universität Göttingen," 30 März 2017. [Online]. Available: http://www.statoek.wiso.uni-goettingen.de/veranstaltungen/statistik3alt/daten/.
[10] S. Yue (2000), "The bivariate lognormal distribution to model a multivariate flood episode," *HYDROLOGICAL PROCESSES.*
[11] S. Nadarajah et al. (2006) "Reliability for some bivariate exponential Distributions," *Mathematical Problems in Engineering*.
[12] C. K. Lee et al. (2009), "A Multivariate Weibull Distribution," Charlotte (North Carolina), Tainan (Taiwan).
[13] E.-J. Lee et al. (2011), "Life Expactancy Estimate With Bivariate Weibull Distribution Using Achrimedian Copula, "*International Journal of Biometrics and Bioinformatics.*"

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Newly enhanced computing algorithm to quantify unavailability of maintained multi-component systems

R. Briš & N.T.T. Tran
VŠB—Technical University of Ostrava, Czech Republic

ABSTRACT: In our previous work we developed an analytical algorithm which is able to carry out exact reliability quantification of highly reliable systems with maintenance (both preventive and corrective). A directed Acyclic Graph (AG) was used as a system representation. The unavailability of a node of an AG is in fact given by going over all possible combinations of probabilities of the input edges (such combinations that cause failure of the node). New improvement of the computing methodology will be presented in this paper, which efficiently reduces computing time and complexity. The improvement is based on applicable properties of internal non-terminal nodes. If an internal node has multiple-dimensionality, i.e. a lot of input edges, resulting in an excessive summarizing combinations, the number of combinations can be significantly reduced by multiple application of the pivotal decomposition. The effectiveness of the process will be demonstrated on selective systems.

1 INTRODUCTION

Estimating (un)availability of a highly-reliable multi-component and maintained system is a problem of great interest in different areas such as computer systems, telecommunications, mechanics, aircraft design, power utilities, and many other engineering fields. Increasing demand for system reliability cannot depend on the increasing reliability of components due to technological restrictions. Safety systems of nuclear power stations represent other example of highly reliable complex systems. They have to be reliable enough to comply with still increasing internationally agreed safety criteria and moreover they are mostly so called sleeping systems which start and operate only in the case of big accidents. Their hypothetical failures are not apparent (hidden or latent failures) and thus repairable only at optimally selected inspection times. Wide class of highly reliable fault-tolerant systems, which, through the use of redundancy, have the ability to operate properly in the presence of faults are investigated in (Villén-Altamirano 2014). Any system failure should have a small probability of occurring; that is, it should be a rare event. It is important to estimate such probabilities because when a rare event does occur, its consequences may be catastrophic. For example, network servers play an increasingly important role due to the rapid growth in demand for internet services, and a server breakdown event may cause significant financial losses. As a result, redundancy is usually built in to prevent services from breaking down.

A Monte Carlo simulation method (Marseguerra & Zio 2001) is used for the quantification of reliability when accurate analytic or numerical procedures do not lead to satisfactory computations of system reliability. Since highly reliable systems require many Monte Carlo trials to obtain reasonably precise estimators of the reliability, various variance-reducing techniques (Tanaka, Kumamoto & Inoue 1989), eventually techniques based on reduction of prior information (Baca A. 1993) have been developed. A direct simulation technique has been improved by the application of a parallel algorithm to such an extent that it can be used for real complex systems which can be then modelled and quantitatively estimated from the point of view of reliability without unreal simplified conditions which analytic methods usually expect (Briš 2008). However, if it is necessary to work and quantitatively estimate highly reliable systems, for which unreliability indicators (i.e. system non-functions) move to the order 10^{-5} and higher (i.e.10^{-6} etc.), the simulation technique, whatever improved, can meet the problems of prolonged and inaccurate computations.

In our previous work we developed a new analytical algorithm which is able to carry out exact reliability quantification of highly reliable systems with maintenance (both preventive and corrective). An exponential distribution for the time to a failure is supposed, possibly for the time to restoration. A generalization of the original methodology so as to be used for unavailability quantification of systems with ageing input components with optional lifetime distribution (i.e. where gener-

ally distributed failure and repair times were supposed) is developed in (Bris & Byczanski 2017a). A directed acyclic graph was used as a system representation. The algorithm allows take into account highly reliable and maintained input components. The unavailability of a node of an AG is in fact given by going over all possible combinations of probabilities of the input edges (such combinations that cause failure of the node). For example, having 20 input edges, we have regularly a million combinations to be summarized. This process has two disadvantages: first, computing errors can easily be committed, which were discussed and removed in (Bris & Byczanski 2013), and second, it is hard to be computed systems having big nodes with a lot of input edges.

New improvement of the computing methodology will be presented in this paper, which efficiently reduces CPU-time and computing complexity as well. The improvement is based on applicable properties of internal non-terminal nodes. If an internal node has multiple-dimensionality, i.e. a lot of input edges, resulting in a lot of summarizing combinations, the number of combinations can be significantly reduced by multiple application of the pivotal decomposition. The effectiveness of the process will be demonstrated on selective systems.

2 SYSTEM REPRESENTATION AND UNAVAILABILITY COMPUTATION

2.1 *The directed acyclic graph*

Example of a real system to be analyzed is shown in Figure 1. The system is demonstrated by means of a directed Acyclic Graph (AG), (Bris 2008). A graph is composed of nodes and edges. The highest node (here u1) represents functionality of the whole system (success, failure), internal and terminal nodes represent subsystems and components. All of the nodes are bounded by edges. Direction of the graph is not explicitly marked in Figure 1 it is given by itself—by projection to vertical direction. The graph is acyclic which means that it cannot contain feedback loops.

Terminal nodes, as for example T1 or T2, of the AG are marked by blue squares. They represent stochastic functionality of input system components given by a probability distribution of their time to failure and a maintenance model. From them we can compute a time course of the unavailability function of input components, using methodology of basic renewal theory (Bris 2007).

Internal nodes (non-terminal) are marked by blue triangles. They represent functionality of subsystems. A subsystem is well functioning in a given time point (success) just in the case when the number of well-functioning inferior edges reaches at least the number that is inside of the triangle, see Figure 1. Otherwise it is not-functioning (failure). For example internal node u3 is well-functioning when the number of well-functioning inferior edges is at least 1.

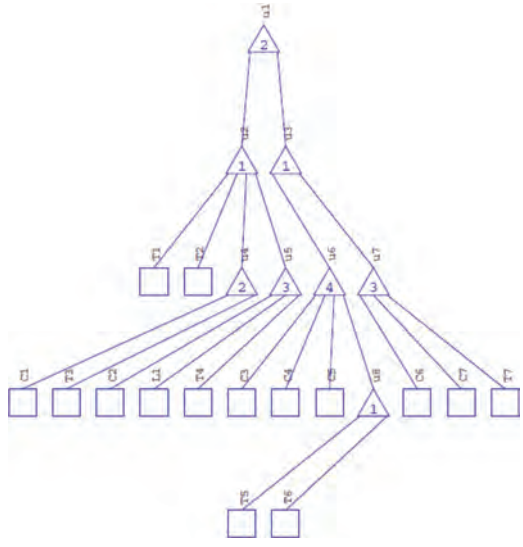


Figure 1. Graph structure of a real system.

The key problem is to estimate the point (or instantaneous) unavailability at any time t, which is the probability that the system is unavailable at time t due to a failure or due to an ongoing repair after the detection of a failure. System may be composed of highly reliable components. In (Bris 2010) we developed new procedure for exact reliability quantification of a highly reliable system. The procedure eliminates all errors made by a computing hardware system when calculations close to error limit are executed.

2.2 *Determination of system unavailability according to a graph structure*

The algorithm will be demonstrated on the system from Fig. 1, which is composed of different terminal nodes. The probability of a non-functional state of the system (represented by the u1 node) can be obtained upwards resulting from unavailability functions of independent terminal nodes.

For instance the unavailability of internal node u7 can be computed as follows:

- numerical expression of the unavailability of inferior terminal nodes, i.e. elements C6, C7 and T7 (having unavailability q_6, q_7 and q_{T7}).
- numerical expression of the unavailability q_{u7} of the internal node u7 which is given by the following sum:

$$q_{u7} = q_6.q_7.q_{T7} + (1-q_6).q_7.q_{T7} + q_6.(1-q_7).q_{T7} + q_6.q_7.(1-q_{T7}) + (1-q_6).(1-q_7).q_{T7} + (1-q_6).q_7.(1-q_{T7}) + q_6.(1-q_7).(1-q_{T7}) \quad (1)$$

In other words we go over all possibilities of non-function state of the node u7 in ascending order of functioning edges. The process is terminated just in the case when the number of well functioning inferior edges reaches 3 (i.e. the number that is inside of the triangle).

The unavailability of a node of an AG is in fact given by going over all possible combinations of probabilities of the input edges (such combinations that cause failure of the node). If a node has a lot of input edges, we can expect computational difficulties. For example, having 20 input edges we have regularly a million combinations to be summarized what results in an unaccepted excessive long computing time.

2.3 *Computational improvement*

In Figs 2–4 we give definitions of following nodes N1-N3, with input edges denoted as 1,2, …k:

We denote unavailability of j^{th} node as $U(Nj)$ and unavailability of i^{th} input edge as $U(i)$. Obviously we can write:

$$U(N1) = U(1).U(N2) + (1- U(1)).\ U(N3) = U(1).\ [U(N2) - U(N3)] + U(N3) \quad (2)$$

This operation we call as pivotal decomposition (applied to first input edge). Expression in square brackets can be simplified (a lot of elements are eliminated applying subtraction) to the following expression:

$$\sum_{\binom{k-1}{m-1}} \ldots \times U(i) \times \ldots \times \left(1 - U(j)\right) \times \ldots \quad (3)$$

indexes i are from {2,3,…k} ↓ number of factors in parentheses is "m–1".

And the last summand in (2) can be simplified by analogy (i.e. recurrently, by the same way as $U(N1)$). In other words the process of pivotal decomposition can be applied to second input edge, etc. By applying this process recurrently we obtain an alternative and time saving formula to compute $U(N1)$.

Example: Let us substitute in Figure 2 for m = 3 and k = 4. Then

$$U(N1) = U(1).\{(1- U(2)).(1- U(3)).U(4) + (1- U(2)).U(3).(1-U(4)) + U(2).(1-U(3)).(1-U(4))\} + U(2).\{(1-U(3)).U(4) + U(3).(1-U(4))\} + U(3).U(4) \quad (4)$$

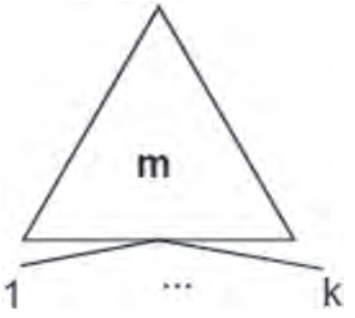


Figure 2. Node 1 (N1).

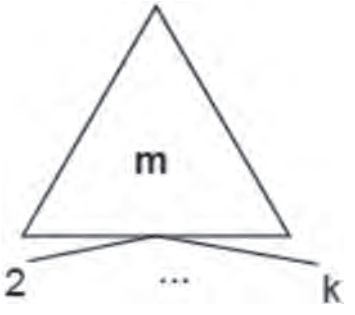


Figure 3. Node 2 (N2).

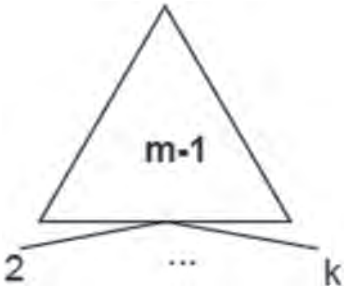


Figure 4. Node 3 (N3).

2.4 *Unavailability models of terminal nodes*

Most of component models (i.e. models of terminal nodes) including maintenance, both preventive and corrective, can be described by the following three models:

Model I with components that cannot be repaired. Final time dependent unavailability coefficient $U(t)$ for this simplest model is given by the distribution function of the time to failure of the component:

$$U(t) = F(t) = \int_0^t f(x)dx, \quad (5)$$

where $F(t)$, $f(t)$ is distribution function and probability density function (*pdf*) of the time to failure.

Model II with repairable components (CM – Corrective Maintenance) for apparent failures, i.e. a model when a possible failure is identified at its occurrence and immediately afterwards it starts a process leading to its restoration. In Model II, two random variables are immediately connected, i.e. the time to failure X, characterized by distribution function $F(t)$ and density $f(t)$, and the repair (or restoration) time Y, characterized by distribution function $G(t)$ and density $g(t)$. In this model we can apply well known relations from renewal theory and alternating renewal processes. In (Briš & Byczanski 2017b) we derived and proofed the following Theorem for time dependent unavailability $U(t)$:

$$U(t) = \int_0^t f(x)\cdot[1-G(t-x)]dx + \int_0^t (f*g)(x)\cdot U(t-x)dx \tag{6}$$

where * means convolution. This Theorem can be considered as a recurrent linear integral equation which helps us to compute $U(t)$.

Model III with repairable components with hidden failures, i.e. a model when a failure is identified only at special deterministically assigned times, appearing with a given period (moments of periodical inspections). In the case of its occurrence at these times an analogical restoration process starts, as in the previous case. Inspections are carried out periodically. Let us denote inspection time points as k_1, k_2, k_3…, then $k_{i+1} - k_i = T_p$ is period of inspections. In (Briš & Byczanski 2017b) we derived numerical formula to calculate the time dependent unavailability coefficient $U(t)$:

$$U(t) = \int_{k_n}^t f(x)dx + \sum_{i=1}^n S_i \cdot \left\{ \int_{k_n-k_i}^{t-k_i} (g*f)(x)dx + [1-G(t-k_i)] \right\} \tag{7}$$

where S_i is the probability that in the inspection time k_i a renew was realized. Numerical formulas to find probabilities S_i are as well derived in (Briš & Byczanski 2017b).

3 RESULTS WITH TESTED SYSTEM FROM REFERENCE

We consider a system model that is a generalization of the Highly Reliable Markovian System (HRMS) often used to represent the evolution of multi-component systems in reliability settings, and which has been studied in (Villén-Altamirano 2014, L'Ecuyer & Tuffin 2011), among others. In the HRMS model, the system has c types of component, with n_i identical components of type i. The system works if at least r_i components of each type i work. Each component is either in a failed state or in an operational state. Specifically we consider a system with 3 types of component, $c = 3$, with n components each. Although formulas (5)–(7) are numerically realized for any probability distributions $f(t)$ and $g(t)$, we assume exponential lifetime and repair times with failure rates $\lambda_1 = 0.01$, $\lambda_2 = 0.015$ and $\lambda_3 = 0.0002$, respectively, and repair rates $\mu = 1$ for all components to be compared with results in above mentioned reference. In other words, all components are of Model II. There are ample repairmen who work simultaneously on all the failed components. The system breaks down as soon as at least one component type had less than 2 operational units ($r = 2$). The redundancy is the same for all 3 types of component. Graph structure of the system for n = 8 is demonstrated in Figure 5. We realized comparison computations for n = 8, 12, 16 and 20.

The concern is to estimate transient measures, such as time dependent system unavailability, including steady state unavailability.

3.1 *Computed results*

The system is to such extent complex that it would be hardly computed by applying original methodology from (Bris 2010). Number of combinations of all input edges each of 3 internal nodes

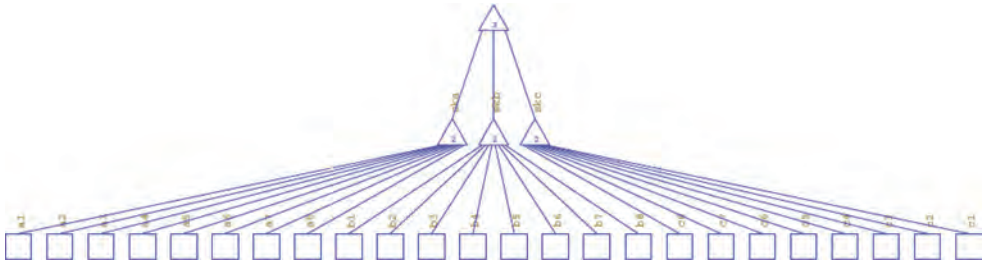


Figure 5. Graph structure of the referenced system for n=8.

is excessive. If computational improvements are realized, see formulas (2)–(4), all computations can be run in CPU-times less than 1 min. In Figs 6–8 we can see dependence of system unavailability on time for n = 8, 16 and 20, ending in steady state unavailability values.

3.2 *Comparisons with referenced results*

Advanced simulation methodology, so called RESTART estimators, to compute unavailability of this system was used in (Villén-Altamirano 2014). Table 1 brings comparison of our obtained numerical results of steady state unavailability with simulation results from the reference, including computing CPU-time.

As can be observed, very low probabilities were accurately estimated with reduced computational effort.

Our improved computational method as well as RESTART estimators can be extended to many other models of highly reliable components, such as non-Markovian models, without much additional effort. In first case computational experiments in (Bris & Byczanski 2017a) comparing exponential and Weibull distributions do not show relevant difference in computational time of both probability distributions. In second case author (Villén-Altamirano 2014) showed that the computational times in the two first system versions (for $n = 8$ and 12) were around 2.5–3 times greater when Raleigh-Erlang distributions were utilized, which was caused for two reasons (i) it is more time consuming to generate random numbers from these distributions than from exponential, and (ii) rescheduling with Erlang distribution is much more time-consuming than with exponential.

Table 1. Steady state unavailability of the referenced system with c = 3, r = 2 and changing n.

	Numerical results		Simulation results (with relative error = 0.1)	
n	U	CPU-time	$\hat{U}$	CPU-time
8	1.284×10^{-12}	< 1(min)	1.31×10^{-12}	0.67 (min)
12	8.74×10^{-20}	< 1(min)	8.76×10^{-20}	2.33 (min)
16	5.49×10^{-27}	< 1(min)	5.38×10^{-27}	3.60 (min)
20	3.258×10^{-34}	< 1(min)	3.19×10^{-34}	9.60 (min)

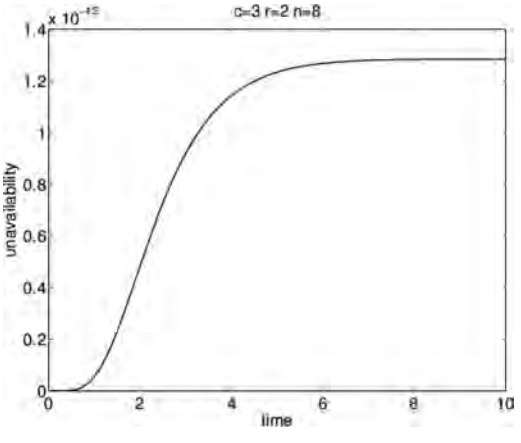


Figure 6. Dependence of system unavailability on time, n = 8.

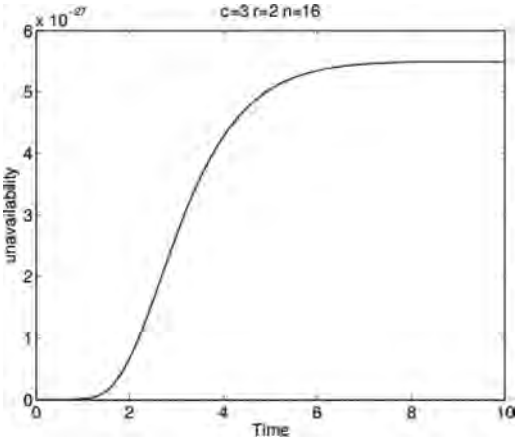


Figure 7. Dependence of system unavailability on time, n = 16.

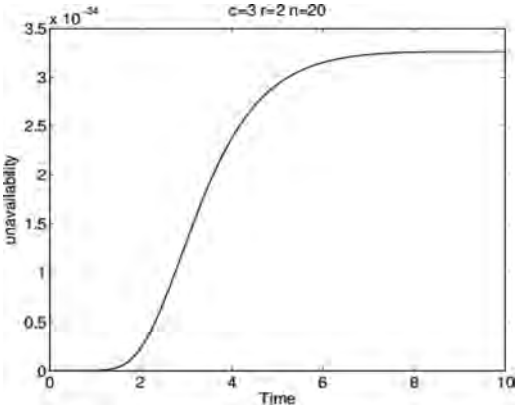


Figure 8. Dependence of system unavailability on time, n = 20.

4 CONCLUSIONS

We demonstrated that the innovative computational method for unavailability quantification is comparable with recent advanced simulation methodology. Even our numerical method gives better computational times particularly when $n \geq 12$.

The innovative methodology for high-performance computing enables exact unavailability quantification of a maintained and highly reliable system containing highly reliable components having optional probability distribution of the time to failure, as well as time to repair, i.e. a complex system with both ageing and non-ageing components can be analyzed. All frequently used component models with both preventive and corrective maintenance may be considered.

The most important advantage of the computing methodology is that it enables the analyst to calculate arbitrary small values of the unavailability function during a mission time exactly. Such small unavailability values as for example values of order 10^{-45} are hardly computed by other methods or software, including advanced simulation RESTART estimators, where computational time increases with increasing number of components.

Having system represented by AG, numerical expression of an unavailability value of one internal node of the AG has a combinatorial character. We have to go over all combinations of input edges leading to a non-functional state of the node. We showed that using pivotal decomposition this computational process can be efficiently improved.

The innovative computing methodology has been numerically realized within the high-performance language MATLAB. All computations above run on Intel (R) Core™ i7-3770 CPU @ 3.4 GHz 3.9 GHz, 8.00 GB RAM.

ACKNOWLEDGEMENTS

This work was supported partially by the European Regional Development Fund in the Research Centre of Advanced Mechatronic Systems project, project number CZ.02.1.01/0.0/0.0/16_019/0000867 within the Operational Programme Research, Development and Education, and partially by the VSB—Technical University of Ostrava in the SGS project number SP2018/68 – Applied Statistics.

REFERENCES

Baca A. 1993. Examples of Monte Carlo methods in reliability estimation based on reduction of prior information. *IEEE Trans Reliab*;42(4):645–9.

Bris R. & Byczanski P. 2013. Effective computing algorithm for maintenance optimization of highly reliable system. *Reliability Engineering & System Safety* 109:77–85.

Bris R. & Byczanski P. 2017a. On innovative stochastic renewal process models for exact unavailability quantification of highly reliable systems, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, Vol. 231(6) 617–627.

Briš R. & Byczanski P. 2017b. Advanced computing methodology for general highly reliable systems. In Walls, Revie & Bedford (eds), *Risk, Reliability and Safety: Innovating Theory and Practice*: p. 1466–1473, Taylor & Francis Group, London.

Bris R. 2007. Stochastic Ageing Models—Extensions of the Classic Renewal Theory. *Reliability: Theory & Applications*, ISSN 1932-2321, 2007;2(3-4):19–27.

Bris R. 2008. Parallel simulation algorithm for maintenance optimization based on directed Acyclic Graph. *Reliability Engineering & System Safety* 93: 852–62.

Bris R. 2010. Exact reliability quantification of highly reliable systems with maintenance. *Reliability Engineering and System Safety* 95: 1286–92.

L'Ecuyer P., Tuffin B. 2011. Approximating zero-variance importance sampling in a reliability setting. Ann. Oper. Res.; 189:277–97.

Marseguerra M. & Zio E. 2001. Principles of Monte Carlo simulation for application to reliability and availability analysis. In: Zio E, Demichela M, Piccinini N, editors. *Safety and reliability towards a safer world,* Torino, Italy, September 16–20. Tutorial notes. p. 37–62.

Tanaka T., Kumamoto H. & Inoue K. 1989. Evaluation of a dynamic reliability problem based on order of component failure. *IEEE Trans Reliab*; 38:573–6.

Villén-Altamirano J. 2014. Asymptotic optimality of RESTART estimators in highly dependable systems. *Reliability Engineering and System Safety* 130, 115–124.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

MLE versus MCMC estimators of the mixture of failure rate model

T.T. Thach & R. Briš
Department of Applied Mathematics, Faculty of Electrical Engineering and Computer Science, VSB–Technical University of Ostrava, Czech Republic

ABSTRACT: In this paper, the parameters and reliability characteristics of the failure distribution of the mixture of failure rates are estimated based on a complete sample using both Markov Chain Monte Carlo (MCMC) method and Maximum Likelihood Estimation (MLE). While MLE is the most frequently used method for parameter estimation, MCMC has recently emerged as a good alternative. The most popular MCMC method, called Metropolis-Hastings algorithm is used to provide complete analysis of the concerned posterior distribution. A simulation study is provided to compare MCMC with MLE, and differences between the estimates obtained by the two approaches are evaluated.

1 INTRODUCTION

Engineering systems, while in operation, are always subject to environmental stresses and shocks which may or may not alter the failure rate function of the system. Suppose p is the unknown probability that the system is able to bear these stresses and its failure model remains unaffected, and q is the probability of the complementary event. In such situations, a failure distribution is generally used to describe mathematically the failure rate on the system. To some extent, the solution to the proposed problem is attempted through the mixture of distributions (Mann et al. 1974, Sinha 1986, Lawless 2002). However, in this regard we are faced with two problems. Firstly, there are many physical causes that individually or collectively cause the failure of the system or device.

At present, it is not possible to differentiate between these physical causes and mathematically account for all of them, and, therefore, the choice of a failure distribution becomes difficult. Secondly, even if a goodness of fit technique is applied to actual observations of time to failure, we face a problem arising due to the non-symmetric nature of the life-time distributions whose behaviour is quite different at the tails where actual observations are sparse in view of the limited sample size (Mann et al. 1974). Obviously, the best one can do is to look out for a concept which is useful for differentiating between different life-time distributions. Failure rate is one such concept in the literature on reliability. After analyzing such physical considerations of the system, we can formulate a mixture of failure rate functions which, in turn, provide the failure time distributions. In view of the above, and due to continuous stresses and shocks on the system, let us suppose that the failure rate function of a system remain unaltered with probability p, and it undergoes a change with probability q. Let the failure rate function of the system in these two situations be in either of the following two states (Sharma et al. 1997):

1.1 *State 1*

Initially it experiences a constant failure rate model and this model may (or may not) change with probability $q(p = 1 - q)$.

1.2 *State 2*

If the stresses and shocks alter the failure rate model of the system with probability q, then it experiences a wear-out failure model.

In comparison with Sharma et al. (1997), this study brings distinctive generalization of the state by implementation of a new parameter, which enables to take into account also more general Weibull model.

In probability theory and statistics, the Weibull distribution is a continuous probability distribution, which is named after the Waloddi Weibull. This is the most commonly used distribution to model times until failure and provides a good description for many types of lifetimes (Rinne 2008, Lawless 2002). The Weibull distribution has two parameters, a shape parameter β and a scale parameter. Only shape parameters $\beta > 1$ correspond to an increasing failure rate, implying that ageing processes can be intensively studied. We will therefore not consider cases with $\beta \leq 1$. Recent studies that adopt a Weibull lifetime distribution include Xia et al. (2015), Zhou et al. (2015), and Xu & Cao (2015).

As a result of flexibility in time to-failure of a very widespread diversity to versatile mechanisms,

the two-parameter Weibull distribution has been recently used quite extensively in reliability and survival analysis particularly when the data are not censored. Much of the attractiveness of the Weibull distribution is due to the wide variety of shapes which can assume by altering its parameters.

Using such a failure rate pattern, the characterization of life-time distribution in the corresponding situation is given. Various inferential properties of this life-time distribution along with the estimation of parameters and reliability characteristics is the subject matter of the present study. Since the estimates based on the operational data can be updated by incorporating past environmental experiences on the random variations in the lifetime parameters (Martz and Waller 1982), therefore, the Bayesian analysis of the parameters and other reliability characteristics is also given.

The remainder of this article is organized as follows. Section 2 introduces the intended mixture of failure rates model including basic characteristics of the corresponding life-time distribution as well. Section 3 brings maximum likelihood estimators for two unknown parameters of the model on one hand and the Metropolis-Hastings algorithm based on MCMC method on the other hand. In addition, Bayesian methodology resulting from special likelihood function of the mixture model is demonstrated bringing formulas for alternative estimators in comparison with MLE. Section 4 shows short illustration how estimation procedures work. Section 5 reports simulation study on special selected situations where both MCMC and MLE are confronted.

2 CHARACTERISTICS OF THE LIFE-TIME DISTRIBUTION

Notations: Let

T: the random variable denoting life-time of the system.
$h(t)$: the failure rate function.
$f(t)$: the probability density function (p.d.f.) of T.
$F(t)$: the cumulative distribution function of T.
$R(t) = \mathbb{P}(T > t)$: the reliability/survival function.
$\mathbb{E}(T) = \int_0^\infty R(t)\,dt$: mean time to failure (MTTF).

2.1 *The mixture of failure rate model*

Let

p: the probability of the event A, that the system is able to bear the stresses and shocks and its failure pattern remains unaltered.

$q = 1 - p$: the probability of the complementary event A^c.

Further, let, the mixture of the failure rate function be

$$h(t) = p\lambda + (1-p)\lambda t^k, \quad \lambda, t > 0, 0 < p < 1 \tag{1}$$

for

1. $p = 1$; represents the failure rate of an exponential distribution.
2. $k = 1$ and $p = 0$; represents the failure rate of the Rayleigh distribution or Weibull distribution with shape parameter 2.
   Note: In our context $\beta = k + 1$
3. $k = 1$; represents the linear ageing process.
4. $0 < k < 1$; represents the concave ageing process.
5. $k > 1$; represents the convex ageing process.

In Weibull reliability analysis it is frequently the case that the value of the shape parameter is known (Martz & Waller 1982). For example, the Raleigh distribution is obtained when $k = 1$. The earliest references to Bayesian estimation of the unknown scale parameter are in Harris & Singpurwalla (1968). Since that time this case has been considered by numerous authors, see Sharma et al. (1997), Canavos (1974), Moore & Bilikam (1978), Tummala & Sathe (1978), Alexander et al. (2009) & Muhammad et al. (2014). This study is free continuation and generalization of the research originally introduced by Sharma et al. (1997).

2.2 *Characteristics of the life-time distribution*

Using the well-known relationship between p.d.f. and failure rate function

$$f(t) = h(t)\exp\left\{-\int_0^t h(x)\,dx\right\}, \quad t > 0 \tag{2}$$

and in view of (1), the p.d.f. of the life-time T is

$$f(t) = h(t)\exp\left\{-\left(p\lambda t + \frac{\lambda(1-p)}{k+1}t^{k+1}\right)\right\} \tag{3}$$

The reliability function is

$$R(t) = \exp\left\{-\left(p\lambda t + \frac{\lambda(1-p)}{k+1}t^{k+1}\right)\right\}, \quad t > 0. \tag{4}$$

The MTTF is given by

$$\begin{aligned} MTTF = \mathbb{E}(T) &= \int_0^\infty R(t)\,dt \\ &= \int_0^\infty \exp\left\{-\left(p\lambda t + \frac{\lambda(1-p)}{k+1}t^{k+1}\right)\right\}dt \end{aligned} \tag{5}$$

This integral can be obtained by using some suitable numerical methods.

3 ESTIMATION OF PARAMETERS AND RELIABILITY CHARACTERISTICS

Let $D: t_1,\ldots,t_n$ be the random failure times of n items under test whose failure time distribution is as given in (3). Then the likelihood function is

$$L(D \mid \lambda, p) = \lambda^n \left[\prod_{i=1}^{n}\left(p + (1-p)t_i^k\right)\right] \times \exp\left\{-\lambda \sum_{i=1}^{n}\left(pt_i + \frac{1-p}{k+1}t_i^{k+1}\right)\right\}. \quad (6)$$

3.1 *Maximum likelihood estimation*

The log-likelihood function can be written as

$$\log L(D \mid \lambda, p) = \sum_{i=1}^{n}\log\left(p + (1-p)t_i^k\right) + n\log\lambda - \lambda\sum_{i=1}^{n}\left(pt_i + \frac{1-p}{k+1}t_i^{k+1}\right). \quad (7)$$

From (7), we derive the likelihood equations for the two parameters p and λ, by taking the partial derivatives with regard to each of the parameters and equating them to zero

$$\frac{\partial log\, L}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n}\left(pt_i + \frac{1-p}{k+1}t_i^{k+1}\right) \quad (8)$$

$$\frac{\partial log\, L}{\partial p} = \sum_{i=1}^{n}\frac{1-t_i^k}{p+(1-p)t_i^k} - \lambda\sum_{i=1}^{n}\left(t_i - \frac{1}{k+1}t_i^{k+1}\right) \quad (9)$$

We get

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n}\left(pt_i + \frac{1-p}{k+1}t_i^{k+1}\right)} \quad (10)$$

and

$$\sum_{i=1}^{n}\left(\frac{1}{p + \frac{t_i^k}{1-t_i^k}}\right) - \frac{n\sum_{i=1}^{n}t_i\left(1+k-t_i^k\right)}{(k+1)\sum_{i=1}^{n}\left(pt_i + \frac{1-p}{k+1}t_i^{k+1}\right)} = 0. \quad (11)$$

Equation (11) may be solve for $\hat{p}$ by Newton-Raphson or other suitable iterative methods and this value is substituted in (10) to obtain $\hat{\lambda}$. By using the invariance property of MLE's,

1. The MLE for $R(t)$, say $\hat{R}(t)$, will be

$$\hat{R}(t) = \exp\left\{-\hat{\lambda}\left(\hat{p}t + \frac{1-\hat{p}}{k+1}t^{k+1}\right)\right\}. \quad (12)$$

2. The MLE for $h(t)$, say $\hat{h}(t)$, will be

$$\hat{h}(t) = \hat{\lambda}\left(\hat{p} + (1-\hat{p})t^k\right). \quad (13)$$

3. The MLE for MTTF will be

$$\hat{MTTF} = MTTF\left(\hat{p}, \hat{\lambda}\right), \quad (14)$$

which can be obtained by installing into (5) and integrating.

3.2 *The Metropolis-Hastings algorithm*

The Metropolis-Hastings algorithm is the most popular MCMC method (Hastings 1970, Metropolis et al. 1953). According to Navarro & Perfors, the basic problem is that MCMC provides a method for sampling from some generic distribution $p(x)$, say target distribution. The idea is that in many cases, we know how to write out the equation for the target distribution $p(x)$, but we do not know how to generate a random number from this target distribution. This is the situation where MCMC is very useful. In fact, for the Metropolis-Hastings algorithm we do not even need to know how to calculate $p(x)$ completely.

The basic idea behind MCMC is to define a Markov chain over possible x values, in such a way that the stationary distribution of Markov chain is in fact $p(x)$. That is, what we are going to do is to use a Markov chain to generate a sequence of x values, denoted ($x_0, x_1, x_2,$), in such a way that as $n \to \infty$, we can guarantee that $x_n \sim p(x)$. There are many different ways of setting up a Markov chain that has this property, one of which is the Metropolis-Hastings algorithm.

Here is how Metropolis-Hastings algorithm works. Suppose that the current state of the Markov chain is x_n, and we want to generate x_{n+1}. In the Metropolis-Hastings algorithm, the generation of x_{n+1} is a twostage process. The first stage is to generate a candidate, which we will denote x^*. The value of x^* is generated from the proposal distribution, denoted $q(x^* \mid x_n)$, which depends on the current state of the Markov chain, x_n. There is a few minor technical constraints on what we can use as a proposal distribution.

The second stage is the accept-reject step. Firstly, what we need to do is calculate the acceptance probability $A(x_n \to x^*)$, which is given by:

$$A(x_n \to x^*) = min\left(1, \frac{p(x^*)}{p(x_n)} \times \frac{q(x_n \mid x^*)}{q(x^* \mid x_n)}\right) \tag{15}$$

There are two things to pay attention to here. Firstly, notice that the ratio $\frac{p(x^*)}{p(x_n)}$ does not depend on the normalizing constant for the target distribution $p(x)$. The second thing to pay attention to is the behavior of the other term, $\frac{q(x_n|x^*)}{q(x^*|x_n)}$. What this term does is correct for any biases that the proposal distribution might induce. In this expression, the denominator $q(x^* \mid x_n)$ describes the probability of generating a x^* as the candidate given that the current state is x_n (i.e., what actually happened), whereas the numerator describes the probability that the "opposite" event would have occurred: that is, if the current state had actually been x^*, what is the probability that you would have generated x_n as the candidate value? If the proposal distribution is symmetric, then these two probabilities will turn out to be equal, $\frac{q(x_n|x^*)}{q(x^*|x_n)} = 1$. This special case of the Metropolis-Hastings algorithm is called the Metropolis algorithm.

Having proposed the candidate x^* and calculated the acceptance probability, $A(x_n \to x^*)$, we now either decide to "accept" the candidate (in which case we set $x_{n+1} = x^*$) or we decide to "reject" the candidate (in which case we set $x_{n+1} = x_n$). To make this decision, we generate a (uniformly distributed) random number between 0 and 1, denoted u. Then:

$$x_{n+1} = \begin{cases} x^* & \text{if } u \le A(x_n \to x^*) \\ x_n & \text{if } u > A(x_n \to x^*) \end{cases} \tag{16}$$

3.3 *Bayesian estimation*

For our mixture model, the Bayesian model is constructed by specifying a prior distribution for p and λ, and then multiplying with the likelihood function to obtain the posterior distribution. Given a set of data $D: t_1 \ldots, t_n$, the likelihood function is

$$\log L(D \mid \lambda, p) = \sum_{i=1}^{n} \log\left(p + (1-p)t_i^k\right) + n \log \lambda - \lambda \sum_{i=1}^{n} \left(pt_i + \frac{1-p}{k+1} t_i^{k+1}\right) \tag{17}$$

Denote the prior distribution of p and λ as $\pi(p,\lambda)$, the posterior distribution of p and λ given $D: t_1 \ldots, t_n$ is given by

$$\pi(p, \lambda \mid D) = \frac{L(D \mid p, \lambda)\pi(p, \lambda)}{\int_0^\infty \int_0^1 L(D \mid p, \lambda)\pi(p, \lambda)\, dp\, d\lambda} \tag{18}$$

Because the denominator in (18) is a normalizing constant, Bayes' theorem is often expressed as:

$$\pi(p, \lambda \mid D) \propto L(D \mid p, \lambda)\pi(\lambda, p) \tag{19}$$

Here the prior distribution is given beforehand, usually based on prior information of the parameters, such as that from historical data, previous experiences, expert suggestions, even wholly subjective suppositions, or simply from the point of mathematical conveniences.

The proposed priors for parameters p and λ may be taken as

$$\pi(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}, \quad a, b > 0. \tag{20}$$

$$\pi(\lambda) = \frac{\alpha^\beta}{\Gamma(\beta)} \lambda^{\beta-1} e^{-\alpha\lambda} \quad \alpha > 0, \beta > 0. \tag{21}$$

For these two parameters, we assume independent priors. Then the joint prior distribution for p and λ will be

$$\pi(p, \lambda) = \frac{\alpha^\beta}{B(a,b)\Gamma(\beta)} p^{a-1}(1-p)^{b-1} \lambda^{\beta-1} e^{-\alpha\lambda} \tag{22}$$

In view of the prior in (22), the posterior distribution of p and λ given $D: t_1 \ldots, t_n$ is given by

$$\pi(p, \lambda \mid D) \propto \lambda^{n+\beta-1} \left[\prod_{i=1}^{n} \left(p + (1-p)t_i^k\right)\right] \times e^{-\lambda\left(\alpha + \sum_{i=1}^{n}\left(pt_i + \frac{1-p}{k+1} t_i^{k+1}\right)\right)} p^{a-1}(1-p)^{b-1} \tag{23}$$

Then, under the square error loss function, the Bayes estimate of p, λ, failure rate function $h(t)$ and reliability function $R(t)$ are given by

$$p^* = \mathbb{E}(p \mid D) \tag{24}$$

$$\lambda^* = \mathbb{E}(\lambda \mid D) \tag{25}$$

$$h^*(t) = \mathbb{E}(h(t) \mid D) \tag{26}$$

$$R^*(t) = \mathbb{E}(R(t) \mid D) \tag{27}$$

In our study, we use adaptive Metropolis-Hastings sampling (Chivers 2012) to generate sample $\theta_i = (p_i, \lambda_i), i = 1, \ldots, n$ from the posterior distribution $\pi(p, \lambda \mid D)$. Then, Monte Carlo integration estimates p^*, λ^*, $h^*(t)$ and $R^*(t)$ by calculating the means:

$$p^* = \mathbb{E}(p \mid D) \approx \frac{1}{n} \sum_{i=1}^{n} p_i \tag{28}$$

$$\lambda^* = \mathbb{E}(\lambda \mid D) \approx \frac{1}{n} \sum_{i=1}^{n} \lambda_i \tag{29}$$

$$h^*(t) = \mathbb{E}(h(t) \mid D) \approx \frac{1}{n}\sum_{i=1}^{n} h(t; p_i, \lambda_i) \tag{30}$$

$$R^*(t) = \mathbb{E}(R(t) \mid D) \approx \frac{1}{n}\sum_{i=1}^{n} R(t; p_i, \lambda_i) \tag{31}$$

4 ILLUSTRATIVE EXAMPLE

In this section, we present an example to illustrate the estimation procedures discussed in this paper. We consider data given in Table 1, which were used as an illustrative example in our previous study (Bris & Thach 2016). The data set was generated in case $p = \frac{1}{3}$, $\lambda = 0.2$ and $n = 30$ and the parameter k is considered to be fixed to one. In this study, we use both MLE and MCMC method to estimate the parameters and reliability characteristics. In order to obtain MCMC estimators, prior parameters are arbitrarily taken as $a = b = 2$ and $\alpha = 0.1$, $\beta = 1$ and we ran the Metropolis-Hastings algorithm to construct Markov chain of length 50,000 with burn-in of 1000 and reduced the autocorrelation by retaining only every 5 iterations of the chain and obtain 9801 samples. Table 2 shows our MCMC point estimates and two-sided 90% of Bayes credible interval for p and λ, and Table 3 shows our MLE point

Table 1. Data from our previous study.

3.615	1.261	1.964	4.534	2.176	1.799
6.704	1.169	4.563	1.371	2.784	4.779
2.346	2.105	5.059	3.657	1.882	5.270
5.955	2.894	2.452	0.821	0.863	0.468
3.748	4.455	3.326	0.811	2.416	2.468

Table 2. Point estimates and two-sided 90% Bayes credible interval for p and λ.

	True value	MCMC	90% BCI
p	1/3	0.3242	[0.3089, 0.6169]
λ	0.2	0.2062	[0.2035, 0.2792]
MTTF	2.9844	2.9797	[2.4726, 3.4726]

Table 3. Point estimates and two-sided 90% bootstrap confident interval BCa (bias corrected and accelerated) for p and λ.

	True value	MLE	90% BCa
p	1/3	0.0065	[0.0000, 0.6276]
λ	0.2	0.1792	[0.1294, 0.2083]
MTTF	2.9844	2.9636	[2.6410, 3.3790]

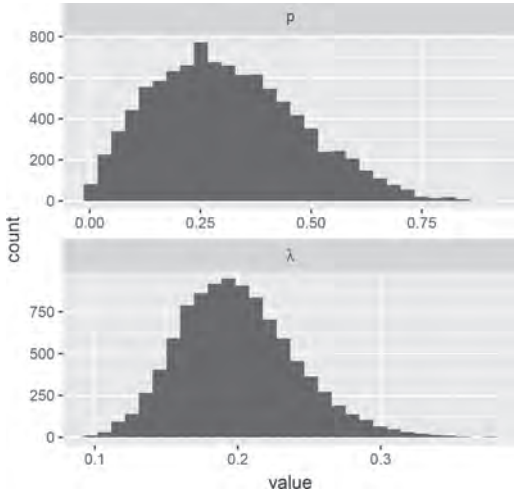


Figure 1. Histograms of each parameter of the Bayesian model.

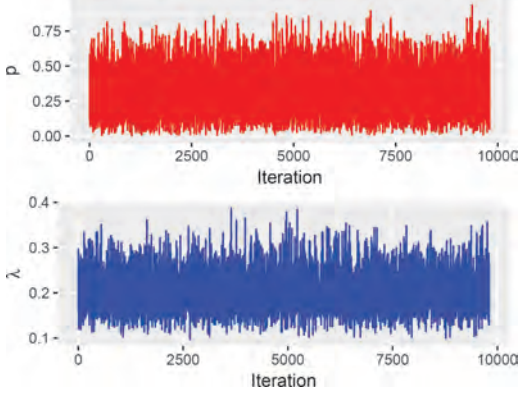


Figure 2. Traces of each parameter of the Bayesian model.

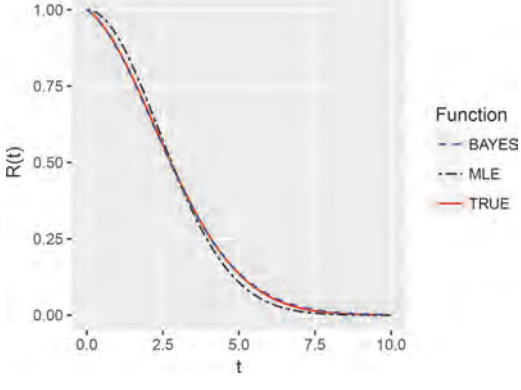


Figure 3. The time courses of reliability functions.

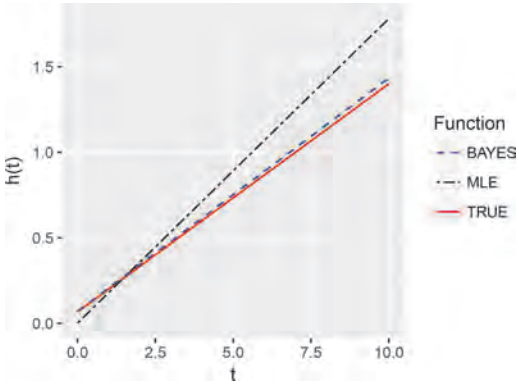


Figure 4. The time courses of failure rate functions.

estimates and two-sided 90% bootstrap confident interval BCa (bias corrected and accelerated) for p and λ. Figures 1–2 show posterior distributions and trace plots of each parameter of the Bayesian model obtained by Metropolis-Hastings algorithm, and Figures 3–4 show time courses of reliability function and failure rate function obtained by both MLE and MCMC method. On the basic of these results, we may conclude that MCMC is better than MLE.

5 SIMULATION STUDY

A Monte Carlo simulation study is conducted to compare the performance of MLE and MCMC estimators for the parameters of mixture failure rate. For each of the following choice of parameters, we simulate 1000 sets of data with sample size $n = 20$, 50, 100 and 200, respectively, and based on each set of data we computed the MLE and MCMC estimator for the model parameters. In order to obtain MCMC estimators, prior parameters are taken as in section 4, and we ran the Metropolis-Hastings algorithm to construct Markov chain of length 20,000 with burn-in of 1000 and reduced autocorrelation by retaining only every 5 iterations of the chain and obtain 3801 samples. Note that as discussed earlier, when $p = 1$, the hazard rate function $h(t)$ represents the failure rate of an exponential distribution, while when $p = 0$, it represents the failure rate of the Rayleigh distribution or Weibull distribution with shape parameter 2.

1. $p = 0.3$ and $\lambda = 0.2$
2. $p = 0.5$ and $\lambda = 0.2$
3. $p = 0.7$ and $\lambda = 0.2$

The Tables 4–6 list the results of the simulation study. Denote $\hat{\theta}$ as the MLE and θ^* as the MCMC

Table 4. Comparison of $\hat{\theta}$ and θ^* for $\theta = (0.3, 0.2)$.

n	Method	Bias p	MSE p	Bias λ	MSE λ
20	MLE	–0.0503	0.0657	0.0119	0.0036
	MCMC	0.0890	0.0152	0.0120	0.0019
50	MLE	–0.0431	0.0414	0.0016	0.0015
	MCMC	0.0500	0.0119	0.0070	0.0010
100	MLE	–0.0157	0.0239	0.0021	0.0007
	MCMC	0.0342	0.0098	0.0057	0.0005
200	MLE	–0.0131	0.0130	0.0006	0.0004
	MCMC	0.0128	0.0071	0.0028	0.0003

Table 5. Comparison of $\hat{\theta}$ and θ^* for $\theta = (0.5, 0.2)$.

n	Method	Bias p	MSE p	Bias λ	MSE λ
20	MLE	–0.0974	0.0864	0.0053	0.0038
	MCMC	–0.0530	0.0126	–0.0044	0.0018
50	MLE	–0.0505	0.0423	0.0026	0.0016
	MCMC	–0.0462	0.0140	–0.0018	0.0010
100	MLE	–0.0207	0.0221	0.0019	0.0009
	MCMC	–0.0317	0.0123	–0.0014	0.0006
200	MLE	–0.0139	0.0113	0.0010	0.0004
	MCMC	–0.0251	0.0090	–0.0011	0.0004

Table 6. Comparison of $\hat{\theta}$ and θ^* for $\theta = (0.7, 0.2)$.

n	Method	Bias p	MSE p	Bias λ	MSE λ
20	MLE	–0.1178	0.0867	0.0026	0.0048
	MCMC	–0.1750	0.0433	–0.0218	0.0024
50	MLE	–0.0559	0.0312	–0.0011	0.0018
	MCMC	–0.1218	0.0282	–0.0163	0.0013
100	MLE	–0.0308	0.0147	–0.0002	0.0009
	MCMC	–0.0779	0.0165	–0.0099	0.0008
200	MLE	–0.0115	0.0067	0.0001	0.0005
	MCMC	–0.0393	0.0076	–0.0055	0.0004

of θ. Bias is calculated as the mean of 1000 estimates minus the true value, and MSE is the mean square error, the mean of the squared differences between 1000 estimators and true value. And Figures 5–10 show the bias and MSE obtained in Tables 4–6.

From the comparison of estimates, we observe the following:

- For estimation of p and λ, although for some cases the biases of MLE are smaller than those of MCMC, MCMC has overwhelming advantage over MLE in the index of MSE. Therefore MCMC is more stable than MLE, in spite of the fact that when sample size is large (say, larger

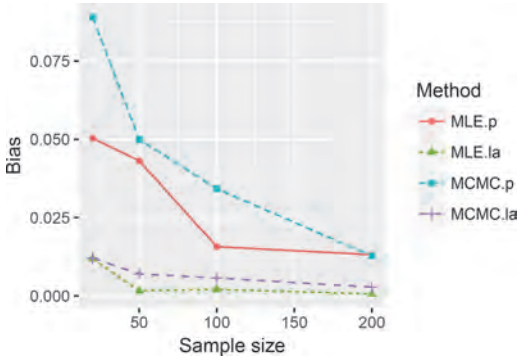


Figure 5. Comparison of bias of $\hat{\theta}$ and θ^* for $\theta = (0.3, 0.2)$.

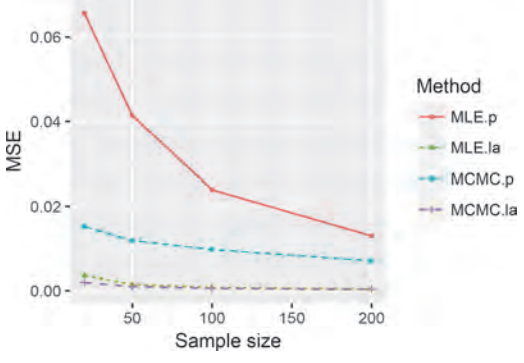


Figure 6. Comparison of MSE of $\hat{\theta}$ and θ^* for $\theta = (0.3, 0.2)$.

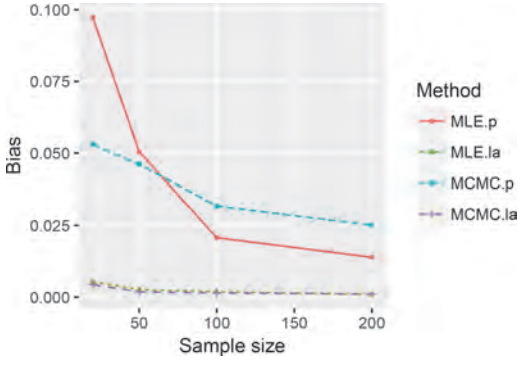


Figure 7. Comparison of bias of $\hat{\theta}$ and θ^* for $\theta = (0.5, 0.2)$.

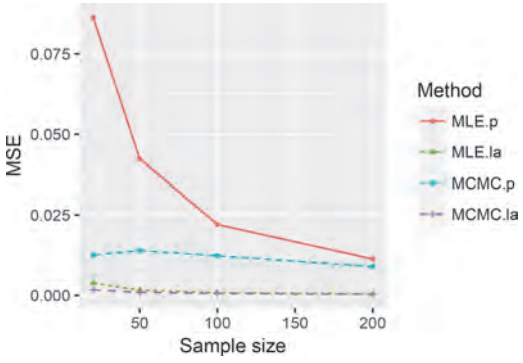


Figure 8. Comparison of MSE of $\hat{\theta}$ and θ^* for $\theta = (0.5, 0.2)$.

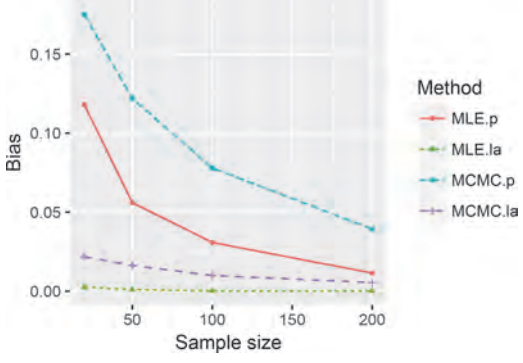


Figure 9. Comparison of bias of $\hat{\theta}$ and θ^* for $\theta = (0.7, 0.2)$.

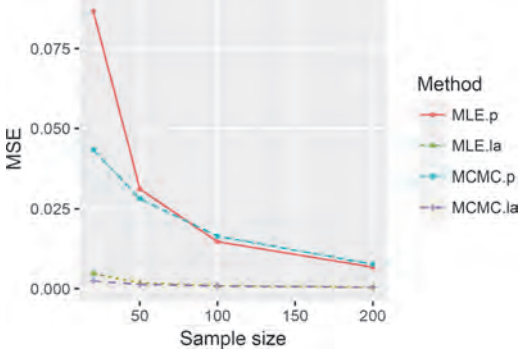


Figure 10. Comparison of MSE of $\hat{\theta}$ and θ^* for $\theta = (0.7, 0.2)$.

than 100), MLE has less bias than MCMC. In general MCMC is better than MLE.

- When the sample size is small (say, less than 100), MCMC behaves better than MLE with regard to the indexes, bias, and MSE. The advantage of MCMC over MLE is especially remarkable in the estimation of parameter λ.

6 CONCLUSION

Based on the simulation results, we may conclude that under square error loss function, the MCMC method can show better result than MLE method to estimate the parameters and reliability characteristics of the failure distribution of the mixture failure rate. We suggest the use of MCMC instead of MLE for point estimation when sample size is not very large. Even when sample size is large, MCMC is still more stable than MLE.

ACKNOWLEDGEMENT

This work was supported partially by the European Regional Development Fund in the Research Centre of Advanced Mechatronic Systems project, project number CZ.02.1.01/0.0/0.0/16˙019/00008 67 within the Operational Programme Research, Development and Education, and partially by the VSB – Technical University of Ostrava in the SGS project number SP2018/68 – Reliability and Risk Modeling.

REFERENCES

Alexander, A., H. Guo, A. Mettas, & D. Ogden (2009). Improving the 1-parameter weibull: A bayesian approach. *IEEE*.

Bris, R. & T.T. Thach (2016). Bayesian approach to estimate the mixture of failure rate model. *Proceedings of the 1st ICAMER. CRC Press*, 9–17.

Canavos, G. (1974). On the robustness of a bayes estimate. *Annual Reliability and Maintainability Symposium*, 432–435.

Chivers, C. (2012). *MHadaptive: General Markov Chain Monte Carlo for Bayesian Inference using adaptive Metropolis- Hastings sampling*. R package version 1.1-8.

Harris, C. & N. Singpurwalla (1968). Life distributions derived from stochastic hazard functions. *IEEE Trans, Reliab. 17*, 70–79.

Hastings, W.K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika Vol. 57*, 97–109.

Lawless, J.F. (2002). *Statistical Models and Methods for Lifetime Data (2 ed.)*. Wiley-Interscience.

Mann, N., E. Schaffer, & N. Singpurwalla (1974). *Methods for Statistical Analysis of Reliability and Life Data*. NY: Wiley.

Martz, H.F. & R.A. Waller (1982). Bayesian *Reliability Analysis*. New York: John Wiley and Sons.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, & A.H. Teller (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics Vol. 21*.

Moore, A.H. & J.E. Bilikam (1978). Bayesian estimation of parameters of life distributions and reliability from type ii censored samples. *IEEE Trans, Reliab. 27*, 64–67.

Muhammad, A., S.M.A. Kazmi, I. Ahmad, & S.H. Shah (2014). Bayesian estimation for parameters of the weibull distribution. *Sci.Int.(Lahore) 26(5)*, 1915–1920.

Navarro, D. & A. Perfors. The metropolis-hastings algorithm. *http://www.compcogscisydney.com/ccs-class.html*.

Pandey, A., A. Singh, & W.J. Zimmer (1993). Bayes estimation of the linear hazard-rate model. *IEEE Trans, Reliab. 42*.

Rinne, H. (2008). The weibull distribution: A handbook (1 ed.). *Chapmann and Hall/CRC*.

Sharma, K.K., H. Krishna, & B. Singh (1997). Bayes estimation of the mixture of hazard rate model. *Reliability Engineering and System Safety 55*, 9–13.

Sinha, S.K. (1986). *Reliability and Life Testing*. USA: Wiley Eastern Ltd.

Tummala, V.M.R. & P.T. Sathe (1978). Minimum expected loss estimators of reliability and parameters of certain lifetime distributions. *IEEE Trans, Reliab. 27*, 283–285.

Xia, T., X. Jin, L. Xi, & J. Ni (2015). Production-driven opportunistic maintenance for batch production based on mam-apb scheduling. *European Journal of Operational Research 240(3)*, 781790.

Xu, W. & L. Cao (2015). Optimal tool replacement with product quality deterioration and random tool failure. *International Journal of Production Research 53(6)*, 17361745.

Zhou, B., J. Yu, J. Shao, & D. Trentesaux (2015). Bottleneckbased opportunistic maintenance model for series production systems. *Journal of Quality in Maintenance Engineering 21(1)*, 7088.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Probabilistic safety assessment and state prediction of cranes based on fuzzy theory

G. Shen, X.J. Zhang, X.L. Tang & S.T. Wang
Beijing Materials Handling Research Institute, Beijing, China

G. Shen & D. Xiang
Tsinghua University, Beijing, China

ABSTRACT: Crane is the supporting equipment in the construction of engineering projects, thus ensuring its safe operation has become an important work. The critical load-bearing part, steel structure, in case of failure can affect reliable operation of the whole crane. Therefore, steel structure is chosen as the study object. First of all, we construct the testing system and data-integration platform to collect running data. Then, the article focuses on establishing the index system of safety assessment, determining the weight of each index based on the Analytic Hierarchy Process (AHP), and obtaining the membership matrix by utilizing the Fuzzy Theory. On this basis, we evaluate the safety level of its current state, and further express the possibility of each safety level occurring by probability form. Finally, we introduce the parameter, transition probability between different safety states, and build the model predicting future safety state of steel structure.

1 INTRODUCTION

1.1 *Research background*

Crane plays an indispensable role in the construction of engineering projects, and has been widely used in the pillar industry, such as machinery manufacturing industry, transportation and logistics industry, water conservancy and hydropower engineering, and nuclear power construction. In recent years, crane is developing rapidly in the direction of large scale and specialization, for adapting to the increasingly enlargement of basic industry and infrastructure.

Under actual operation conditions, the complex alternating load is frequently applied to cranes, which makes its safety to be the significant index in the process of design, manufacture and maintenance. Therefore, we know the crane's safety is of great importance. In case of failure, the economic loss produced tends to be extremely disastrous. Accordingly, carrying out the study on safety assessment and safety state prediction has become the critical breakthrough in solving the problem.

1.2 *Research reviews*

The latest research developments indicate that testing technology (Tian et al. 2009) and assessment method (Yang 2005) have been the focus in the field of crane's safety assessment.

In the aspect of testing technology, remote monitoring (Szpytko 1998, Li & Liu 2012) and spot inspection (Zhang 2017) are widely used in this field, and also become the research focus. Especially in the era of Internet of Things and Big Data, real-time monitoring of actual operation and timely capture of overload, emergency braking and other safety conditions can improve the safety and production efficiency of cranes. However, the current inspection of cranes is still based on the traditional way, which depends on spot experience and simple instrument inspection (Li & Yin 2011). Obviously, the current inspection level is relatively low, and the inspection content is narrow. It is just limited to the standard inspection and no inspection items are carried out according to the new market demand, such as safety level assessment and future state prediction. Moreover, many monitoring systems mainly focus on reflecting the real-time running parameters (Cen et al. 2015), and the in-depth analysis of these parameters is not enough. Thus, it is difficult to assess the current safety state and failure possibility of cranes (Wang et al. 2013, Makovskii 1994).

In terms of assessment method, the current research work mostly includes two aspects, the assessment of current safety state (Yang et al. 2009, Bucas et al. 2014) and the prediction of future safety state (Wu et al. 2010), and the latter is also called life prediction. At present, the assessment of

crane's safety state is mainly based on Grey Theory Method, Combination Weighting Method, Unascertained Measurement Theory, Fisher Discriminant Method, Support Vector Machine, Fuzzy Theory Method and Artificial Neural Network (Fan et al. 2011). In these methods, the application of Fuzzy Theory Method is the most extensive. On the other hand, scholars have conducted in-depth and systematic research by theoretical and experimental methods for life prediction (Zhou et al. 2012), and put forward several life prediction models for the whole crane or steel structure. However, due to the complexity and randomness of environment and loading conditions, these classical methods based on deterministic equations have being developing towards the direction of probability statistics.

On account of the insufficient research above, we accurately grasp the new market demand of safety assessment, and further carry out the research on index system of safety assessment, probabilistic safety assessment method, and future state prediction method for crane's steel structure. The research results will help to ensure the reliable operation of cranes for a long time and bring enormous economic benefits. Besides, these results can also promote the technological progress of hoisting machinery industry and enhance the international competitiveness of cranes, which have the long-term social benefits.

In this paper, the method, Analytic Hierarchy Process (AHP), has been used to assessing the safety level of cranes. AHP actually represents the decision process, namely at first resolving the total assessment target into several layers and then carrying out qualitative and quantitative analysis. The specific procedures mainly include: establishing the index system, defining the assessment level, constructing the judgment matrix, calculating the weight vector, checking the consistency, determining the membership and assessing the safety level, as shown in section 2 and section 3.

2 INDEX SYSTEM OF SAFETY ASSESSMENT

2.1 *Establishing the index system*

According to BS 7121-1:2006, GB 6067-1:2010 and GB/T 21920:2008, the influence factors have been sorted out and index system of safety assessment for cranes has been established, which includes three aspects, such as target layer, standard layer and index layer, as shown in Figure 1.

The target layer, namely the final result layer, refers to the safety assessment level of cranes, which is represented by the parameter *F*.

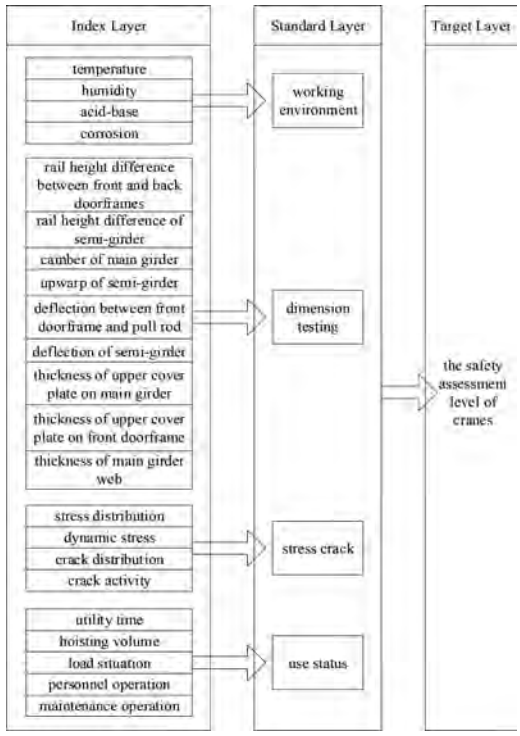


Figure 1. Index system of safety assessment for crane's steel structure.

The standard layer is sorted based on four aspects, including working environment, dimension testing, stress crack and use status. Namely, $V = [V_1\ V_2\ V_3\ V_4]$ = [working environment, dimension testing, stress crack, use status].

The index layer is the specific expansion of the factors in standard layer. For the working environment, $V_1 = [M_{11}\ M_{12}\ M_{13}\ M_{14}]$, where M_{11} represents the temperature; M_{12} represents the humidity; M_{13} represents the acid-base property; M_{14} represents the corrosion degree, respectively. For the dimension testing, $V_2 = [M_{21}\ M_{22}\ M_{23}\ M_{24}\ M_{25}\ M_{26}\ M_{27}\ M_{28}\ M_{29}]$, where M_{21} represents the rail height difference between front and back doorframes; M_{22} represents the rail height difference of semi-girder; M_{23} represents the camber of main girder; M_{24} represents the upwarp of semi-girder; M_{25} represents the deflection between front doorframe and pull rod; M_{26} represents the deflection of semi-girder; M_{27} represents the thickness of upper cover plate on main girder; M_{28} represents the thickness of upper cover plate on front doorframe; M_{29} represents the thickness of main girder web, respectively. For the stress crack, $V_3 = [M_{31}\ M_{32}\ M_{33}\ M_{34}]$, where M_{31} represents the stress distribution; M_{32} represents the dynamic stress; M_{33} represents the

crack distribution; M_{34} represents the crack activity, respectively. For the use status, $V_4 = [M_{41}\ M_{42}\ M_{43}\ M_{44}\ M_{45}]$, where M_{41} represents the utility time; M_{42} represents the hoisting volume; M_{43} represents the load situation; M_{44} represents the personnel operation; M_{45} represents the maintenance operation, respectively.

The index system above is established on the basis of AHP, which takes into account not only the technical elements such as dimension testing and stress crack, but also the non-technical elements such as working environment and use status. The method and index system are mainly used to assess cranes' comprehensive safety performance in the middle and later stages of service. Moreover, for the other mechanical products similar to cranes, the safety assessment can also be carried out by adjusting for the factors in the index system correspondingly.

2.2 *Constructing the testing system*

Under actual service conditions, the working environment of cranes exists uncertain factors, and heavy alternating load often acts on crane's steel structure at the same time. Therefore, collecting the real-time running data and testing data has become the basis of carrying out safety assessment research.

Based on the index system of safety assessment established in Figure 1, we construct the testing system with multiple modules, such as corrosion morphology testing module, dimension deformation testing module, stress distribution testing module, dynamic stress testing module, crack distribution testing module and crack activity testing module. Specifically, the advanced sensor technology has been applied on crane's steel structure to collect testing parameters, including corrosion parameter, deformation parameter, stress-strain parameter, crack parameter, crack growth parameter and so on. On this basis, the data-integration software platform is built to gather all the testing data, which are exactly the basic data source for assessing safety level of crane's steel structure.

3 SAFETY ASSESSMENT OF CURRENT STATE

3.1 *Definition of assessment level*

According to statistical data and previous experience, the assessment level of safety state has been classified in consideration of the severity of crane's failure accidents. There exists five kinds of safety level listed in descending order. It can be expressed as follows: $U = [U_1\ U_2\ U_3\ U_4\ U_5]$, where U_1 represents the excellent level; U_2 represents the good level; U_3 represents the available level; U_4 represents the require-repair level; U_5 represents the discard level, respectively.

3.2 *Weight of factors in standard layer*

The 1–9 scale method has been proposed to quantify the relative importance between working environment, dimension testing, stress crack, and use status, as shown in Table 1.

According to the scale method in Table 1, the judgment matrix can be constructed as follows.

$$N = \left(a_{ij}\right)_{n\cdot n} \tag{1}$$

where the element a_{ij} in matrix N indicates the relative importance between every two factors.

For the factor F in target layer, the matrix N reflecting the relative importance between working environment V_1, dimension testing V_2, stress crack V_3 and use status V_4 in standard layer, has been constructed and specifically expressed as follows.

$$N = \begin{bmatrix} 1 & 1/5 & 1/7 & 1/3 \\ 5 & 1 & 1/5 & 3 \\ 7 & 5 & 1 & 5 \\ 3 & 1/3 & 1/5 & 1 \end{bmatrix} \tag{2}$$

The maximum eigenvalue and eigenvector of matrix N have been calculated by MATLAB. The eigenvector exactly reflects the relative importance of factors in standard layer, which is also called the allocation of weight coefficient. The expression of eigenvector after normalization is given as follows.

$$A = (0.0521\ \ 0.2195\ \ 0.6194\ \ 0.1090) \tag{3}$$

In order to verify whether the allocation of weight coefficient is reasonable, it is necessary

Table 1. 1–9 scale method.

Scale value	Explanation
1	The two factors are of the equal importance
3	One is a little more important than the other
5	One is more important than the other
7	One is more important than the other intensely
9	One is more important than the other extremely
2,4,6,8	The intermediate values of adjacent scales above

to carry out the consistency check of judgment matrix N. The consistency ratio is defined as the parameter CR.

$$CR = \frac{CI}{RI} \tag{4}$$

$$CI = \frac{\lambda_{\max} - n}{n-1} \tag{5}$$

where $\lambda_{\max}$ represents the maximum eigenvalue; n represents the matrix dimension; RI represents the mean random consistency index.

When the consistency ratio satisfies the condition $CR < 0.1$, it is considered that the allocation of weight coefficient is reasonable. Otherwise, we should adjust the elements' value in judgment matrix N, and redistribute the weight coefficient.

Based on equations 4–5, the calculating result shows that $CR = 0.09 < 0.1$, which indicates the elements' value of matrix N and the weight coefficient distribution of eigenvector A are both reasonable and effective.

3.3 *Weight of factors in index layer*

For the factor V_1, the judgment matrix reflecting the relative importance between temperature M_{11}, humidity M_{12}, acid-base property M_{13} and corrosion degree M_{14} in index layer, has been constructed based on 1–9 scale method. The specific expression is shown as follows.

$$N_1 = \begin{bmatrix} 1 & 1/5 & 1/4 & 1/7 \\ 5 & 1 & 3 & 1/3 \\ 4 & 1/3 & 1 & 1/5 \\ 7 & 3 & 5 & 1 \end{bmatrix} \tag{6}$$

The maximum eigenvalue and eigenvector of matrix N_1 have been calculated by MATLAB. The expression of eigenvector after normalization is given as follows.

$$A_1 = (0.0516 \;\; 0.2605 \;\; 0.1276 \;\; 0.5604) \tag{7}$$

In aspect of the consistency check, the calculating result shows that $CR_1 = 0.067 < 0.1$, which indicates the elements' value of matrix N_1 and the weight coefficient distribution of eigenvector A_1 are both reasonable and effective.

For the factor V_2, the judgment matrix reflecting the relative importance between M_{21}, M_{22}, M_{23}, M_{24}, M_{25}, M_{26}, M_{27}, M_{28}, and M_{29} in index layer, has been constructed based on 1–9 scale method. The specific expression is shown as follows.

$$N_2 = \begin{bmatrix} 1 & 1/2 & 2 & 2 & 2 & 3 & 2 & 3 & 2 \\ 2 & 1 & 3 & 3 & 2 & 3 & 2 & 4 & 2 \\ 1/2 & 1/3 & 1 & 3 & 2 & 2 & 1/2 & 1 & 1/2 \\ 1/2 & 1/3 & 1/3 & 1 & 1/3 & 1/2 & 1/3 & 2 & 1/3 \\ 1/2 & 1/2 & 1/2 & 3 & 1 & 3 & 2 & 3 & 2 \\ 1/3 & 1/3 & 1/2 & 2 & 1/3 & 1 & 1/3 & 2 & 1/3 \\ 1/2 & 1/2 & 2 & 3 & 1/2 & 3 & 1 & 4 & 2 \\ 1/3 & 1/4 & 1 & 1/2 & 1/3 & 1/2 & 1/4 & 1 & 1/3 \\ 1/2 & 1/2 & 2 & 3 & 1/2 & 3 & 1/2 & 3 & 1 \end{bmatrix} \tag{8}$$

The maximum eigenvalue and eigenvector of matrix N_2 have been calculated by MATLAB. The expression of eigenvector after normalization is given as follows.

$$A_2 = (0.1660 \;\; 0.2136 \;\; 0.0969 \;\; 0.0488 \;\; 0.1338 \ldots \\ \ldots 0.0555 \;\; 0.1326 \;\; 0.0427 \;\; 0.1100) \tag{9}$$

In aspect of the consistency check, the calculating result shows that $CR_2 = 0.061 < 0.1$, which indicates the elements' value of matrix N_2 and the weight coefficient distribution of eigenvector A_2 are both reasonable and effective.

For the factor V_3, the judgment matrix reflecting the relative importance between M_{31}, M_{32}, M_{33}, and M_{34} in index layer, has been constructed based on 1–9 scale method. The specific expression is shown as follows.

$$N_3 = \begin{bmatrix} 1 & 1/3 & 1/6 & 1/7 \\ 3 & 1 & 1/4 & 1/5 \\ 6 & 4 & 1 & 1/3 \\ 7 & 5 & 3 & 1 \end{bmatrix} \tag{10}$$

The maximum eigenvalue and eigenvector of matrix N_3 have been calculated by MATLAB. The expression of eigenvector after normalization is given as follows.

$$A_3 = (0.0513 \;\; 0.1061 \;\; 0.2890 \;\; 0.5536) \tag{11}$$

In aspect of the consistency check, the calculating result shows that $CR_3 = 0.065 < 0.1$, which indicates the elements' value of matrix N_3 and the weight coefficient distribution of eigenvector A_3 are both reasonable and effective.

For the factor V_4, the judgment matrix reflecting the relative importance between M_{41}, M_{42}, M_{43}, M_{44} and M_{45} in index layer, has been constructed based on 1–9 scale method. The specific expression is shown as follows.

$$N_4 = \begin{bmatrix} 1 & 1/3 & 3 & 7 & 5 \\ 3 & 1 & 5 & 8 & 6 \\ 1/3 & 1/5 & 1 & 4 & 2 \\ 1/7 & 1/8 & 1/4 & 1 & 1/3 \\ 1/5 & 1/6 & 1/2 & 3 & 1 \end{bmatrix} \quad (12)$$

The maximum eigenvalue and eigenvector of matrix N_4 have been calculated by MATLAB. The expression of eigenvector after normalization is given as follows.

$$A_4 = (0.2705 \quad 0.5045 \quad 0.1153 \quad 0.0368 \quad 0.0729) \quad (13)$$

In aspect of the consistency check, the calculating result shows that $CR_4 = 0.045 < 0.1$, which indicates the elements' value of matrix N_4 and the weight coefficient distribution of eigenvector A_4 are both reasonable and effective.

3.4 *Obtaining the membership matrix*

According to the assessment level of safety state defined in section 3.1, the fuzzy subset for factors in index layer needs to be constructed. Namely, $D = [D_1\ D_2\ D_3\ D_4\ D_5]$ = [excellent, good, medium, poor, extremely poor], where D_1 represents the excellent state; D_2 represents the good state; D_3 represents the medium state; D_4 represents the poor state; D_5 represents the extremely poor state, respectively. Then, we analyze quantitatively the degree of each factor being subordinate to fuzzy subset D, and further obtain the membership matrix.

For working environment V_1, the membership matrix has been obtained as follows.

$$R_1 = \begin{bmatrix} 0.1 & 0.1 & 0.2 & 0.3 & 0.3 \\ 0.1 & 0.1 & 0.2 & 0.3 & 0.3 \\ 0.1 & 0.1 & 0.1 & 0.4 & 0.3 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (14)$$

The first row vector represents the membership degree of temperature factor M_{11} being subordinate to fuzzy subset D. As the temperature of cranes is changing during service, the membership degree corresponding to each state D_1-D_5 is the result of probability statistics. Through referring to the weather data, poor weather appears more frequently, and the membership vector of temperature M_{11} is [0.1 0.1 0.2 0.3 0.3]. In the same way, for the other two factors humidity M_{12} and acid-base property M_{13}, we can also obtain the membership vector based on previous data, as shown by the second row vector and the third row vector.

The fourth row vector represents the membership degree of corrosion factor M_{14} being subordinate to fuzzy subset D. It is remarkable that the membership vector is a definite value rather than the form of probability distribution because the corrosion degree is uniquely determined by the on-site testing results. According to the testing data, the degree of corrosion has reached state D_5, namely the extremely poor state. Therefore, the membership vector of M_{14} is [0 0 0 0 1].

For dimension testing V_2, the membership matrix has been obtained as follows.

$$R_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (15)$$

The first row vector represents the membership degree of factor M_{21} being subordinate to fuzzy subset D. We have collected six groups of testing data, which are respectively 7 mm, 10 mm, 14 mm, 6 mm, 3 mm and 7 mm. The results indicate that M_{21} has reached state D_4, namely the poor state, so the membership vector is [0 0 0 1 0].

The second row vector represents the membership degree of factor M_{22} being subordinate to fuzzy subset D. We have collected thirteen groups of testing data, which are respectively 32 mm, 9 mm, 18 mm, 13 mm, 6 mm, 4 mm, 12 mm, 1 mm, 3 mm, 25 mm, 1 mm, 23 mm and 25 mm. The results indicate that M_{22} has reached state D_3, namely the medium state, so the membership vector is [0 0 1 0 0].

The third row vector represents the membership degree of factor M_{23} being subordinate to fuzzy subset D. We have collected two groups of testing data, which are respectively 8 mm and 5 mm. The results indicate that M_{23} has reached state D_3, namely the medium state, so the membership vector is [0 0 1 0 0].

The fourth row vector represents the membership degree of factor M_{24} being subordinate to fuzzy subset D. We have collected two groups of testing data, which are respectively 76 mm and 133 mm. The results indicate that M_{24} has reached state D_5, namely the extremely poor state, so the membership vector is [0 0 0 0 1].

The fifth row vector represents the membership degree of factor M_{25} being subordinate to fuzzy subset D. The testing value of M_{25}, deflection

between front doorframe and pull rod, is 29 mm, which has reached state D_2, namely the good state, so the membership vector is [0 1 0 0 0].

The sixth row vector represents the membership degree of factor M_{26} being subordinate to fuzzy subset D. The testing value of M_{26}, deflection of semi-girder, is 100 mm, which has reached state D_4, namely the poor state, so the membership vector is [0 0 0 1 0].

The seventh row vector represents the membership degree of factor M_{27} being subordinate to fuzzy subset D. We have collected three groups of testing data, which are respectively 25.8 mm, 25.9 mm and 25.9 mm. The results indicate that M_{27} has reached state D_3, namely the medium state, so the membership vector is [0 0 1 0 0].

The eighth row vector represents the membership degree of factor M_{28} being subordinate to fuzzy subset D. We have collected three groups of testing data, which are respectively 14.5 mm, 14.6 mm and 14.5 mm. The results indicate that M_{28} has reached state D_3, namely the medium state, so the membership vector is [0 0 1 0 0].

The ninth row vector represents the membership degree of factor M_{29} being subordinate to fuzzy subset D. We have collected three groups of testing data, which are respectively 11.8 mm, 11.9 mm and 11.9 mm. The results indicate that M_{29} has reached state D_3, namely the medium state, so the membership vector is [0 0 1 0 0].

For stress crack V_3, the membership matrix has been obtained as follows.

$$R_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{16}$$

For use status V_4, the membership matrix has been obtained as follows.

$$R_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0.1 & 0.2 & 0.3 & 0.2 & 0.2 \\ 0.1 & 0.1 & 0.1 & 0.4 & 0.3 \\ 0 & 0.1 & 0.1 & 0.2 & 0.6 \end{bmatrix} \tag{17}$$

It should be noted that the membership degree of factors M_{31}-M_{34}, M_{41}-M_{45} being subordinate to fuzzy subset D can be determined in the same way by the on-site test and data analysis, so we won't repeat it.

3.5 *Assessing the safety level*

Based on equation 7 and equation 14 above, the assessing vector of working environment V_1 has been calculated. The expression of this vector B_1 after normalization is given as follows.

$$\begin{aligned} B_1 &= A_1 R_1 \\ &= (0.0440 \;\; 0.0440 \;\; 0.0752 \;\; 0.1447 \;\; 0.6922) \end{aligned} \tag{18}$$

Based on equation 9 and equation 15 above, the assessing vector of dimension testing V_2 has been calculated. The expression of this vector B_2 after normalization is given as follows.

$$B_2 = A_2 R_2 = (0 \;\; 0.1338 \;\; 0.5958 \;\; 0.2216 \;\; 0.0488) \tag{19}$$

Based on equation 11 and equation 16 above, the assessing vector of stress crack V_3 has been calculated. The expression of this vector B_3 after normalization is given as follows.

$$B_3 = A_3 R_3 = (0 \;\; 0 \;\; 0.4464 \;\; 0.5536 \;\; 0) \tag{20}$$

Based on equation 13 and equation 17 above, the assessing vector of use status V_4 has been calculated. The expression of this vector B_4 after normalization is given as follows.

$$\begin{aligned} B_4 &= A_4 R_4 \\ &= (0.0152 \;\; 0.0340 \;\; 0.0456 \;\; 0.0524 \;\; 0.8528) \end{aligned} \tag{21}$$

The following work is to integrate these assessing vectors, including B_1, B_2, B_3, and B_4, and then we can obtain the fuzzy relation matrix B of final index F in target layer, as shown in the following equation.

$$B = \begin{pmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix} \tag{22}$$

Based on equation 3 and equation 22 above, the assessing vector of final target F has been calculated. The expression of this vector C after normalization is given as follows.

$$C = AB = (0.0039 \;\; 0.0354 \;\; 0.4161 \;\; 0.4048 \;\; 0.1397) \tag{23}$$

Equation 23 indicates that each safety level U_1-U_5 has the possibility of occurrence based on the current testing data, and the vector C exactly represents the probability distribution as shown in Figure 2.

In Figure 2, the first column represents the probability of excellent level occurring, and the

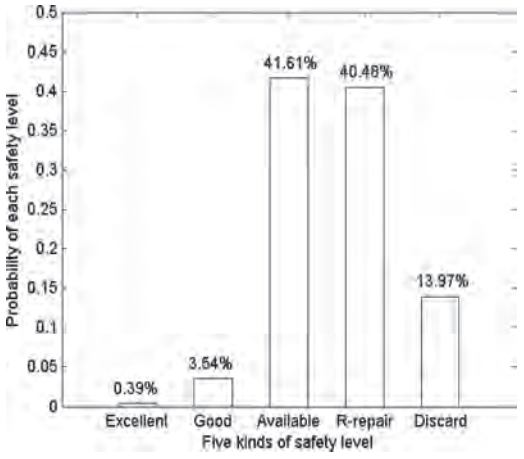


Figure 2. The probability distribution of current safety state.

value is 0.39%. The second column represents the probability of good level occurring, and the value is 3.54%. The third column represents the probability of available level occurring, and the value is 41.61%. The fourth column represents the probability of require-repair level occurring, and the value is 40.48%. The fifth column represents the probability of discard level occurring, and the value is 13.97%.

Next, the final score of current state are calculated considering the weighting coefficient, as shown in the following expression.

$$G = C \cdot S^T \tag{24}$$

where S represents the vector of weighting coefficient; G represents the final score of current state.

Based on equation 23 and equation 24, the safety score G is 60.99. Then, combining with the partition of safety state as shown in Table 2, the service state of crane's steel structure can be ultimately evaluated.

From Table 2, although the score 60.99 corresponds to the require-repair level, it is also near the discard level. Therefore, from the conservative perspective, the conclusion has been drawn that this crane should be discarded in the short term.

The above contents have described the process of analyzing cranes' safety state based on AHP and Fuzzy Theory. Then, we invite the experts in this field to discuss the assessment results obtained. Experts consider that the method is reasonable and the factors listed in index system are comprehensive, which correspond to the actual situation of cranes. To sum up, the assessment effect is very good. Similar to its application on cranes, the method can be further applied to assessing other critical mechanical products such as automobiles, wind turbines, nuclear power equipment, so as to provide important support for obtaining the current service state timely.

Table 2. Partition of safety state.

Safety score	Safety state
90–100	excellent level
80–89	good level
70–79	available level
60–69	require-repair level
< 60	discard level

4 SAFETY PREDICTION OF FUTURE STATE

4.1 *Transition probability between different safety states*

There exists five kinds of safety states for the factor F in target layer, namely $U = [U_1\ U_2\ U_3\ U_4\ U_5]$ = [excellent, good, available, require-repair, discard]. We introduce the parameter, transition probability p_{ij} (i,j = 1,2,3,4,5), to characterize the possibility of safety state i developing to safety state j, as shown in the following equation.

$$p_{ij} = p(U_i \to U_j) \tag{25}$$

It is necessary to point out that when the time interval of state developing is one year, p_{ij} is known as one-step transition probability. Therefore, parameter p_{ij} represents the transition probability of safety state developing from this year to the next year.

For crane's steel structure, the assessing result is a vector rather than a numerical value, which indicates that the assessing result includes several safety states and each state has the possibility of occurrence. In section 3.5, the assessing vector C of this year has been calculated and characterized by the form of probability distribution. On this basis, predicting the assessing vector C' of the next year has become the focus, and the transition probability matrix P can be expressed as follows.

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{bmatrix} \tag{26}$$

Based on equation 23 and equation 26 above, the assessing vector of future safety state can be obtained as follows.

$$C' = C \times P \tag{27}$$

$$\begin{aligned}&\left(c'_{U1} \quad c'_{U2} \quad c'_{U3} \quad c'_{U4} \quad c'_{U5}\right)\\&=\left(c_{U1} \quad c_{U2} \quad c_{U3} \quad c_{U4} \quad c_{U5}\right)\\&\times\begin{bmatrix}p_{11} & p_{12} & p_{13} & p_{14} & p_{15}\\p_{21} & p_{22} & p_{23} & p_{24} & p_{25}\\p_{31} & p_{32} & p_{33} & p_{34} & p_{35}\\p_{41} & p_{42} & p_{43} & p_{44} & p_{45}\\p_{51} & p_{52} & p_{53} & p_{54} & p_{55}\end{bmatrix}\end{aligned} \tag{28}$$

where c_{U1} represents the probability of excellent level occurring in this year; c_{U2} represents the probability of good level occurring in this year; c_{U3} represents the probability of available level occurring in this year; c_{U4} represents the probability of require-repair level occurring in this year; c_{U5} represents the probability of discard level occurring in this year; c'_{U1} represents the probability of excellent level occurring in the next year; c'_{U2} represents the probability of good level occurring in the next year; c'_{U3} represents the probability of available level occurring in the next year; c'_{U4} represents the probability of require-repair level occurring in the next year; c'_{U5} represents the probability of discard level occurring in the next year.

4.2 *Solving the transition probability matrix P*

Based on equation 27, the corresponding equations can be obtained.

$$C^{2011} = C^{2010} \cdot P \tag{29}$$

$$C^{2012} = C^{2011} \cdot P \tag{30}$$

$$C^{2013} = C^{2012} \cdot P \tag{31}$$

$$C^{2014} = C^{2013} \cdot P \tag{32}$$

$$C^{2015} = C^{2014} \cdot P \tag{33}$$

After expanding these equations above, we can solve the transition probability matrix P as follows.

$$P = K^{-1} \times L \tag{34}$$

$$K=\begin{pmatrix}C^{2010}\\C^{2011}\\C^{2012}\\C^{2013}\\C^{2014}\end{pmatrix},\quad L=\begin{pmatrix}C^{2011}\\C^{2012}\\C^{2013}\\C^{2014}\\C^{2015}\end{pmatrix} \tag{35}$$

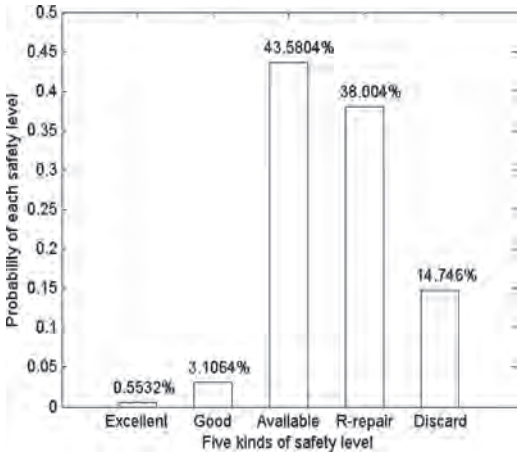


Figure 3. The predicting probability distribution of future safety state.

4.3 *Predicting the future safety state*

For the assessing vectors of the past years, C^{2010}, C^{2011}, C^{2012}, C^{2013}, C^{2014} and C^{2015} have been obtained based on the running data tested once a year. Accordingly, the matrix P can be solved and the assessing vector in 2016 can be predicted as follows.

$$\begin{aligned}C^{2016} &= (c_{U1}^{2016} \quad c_{U2}^{2016} \quad c_{U3}^{2016} \quad c_{U4}^{2016} \quad c_{U5}^{2016})\\&= (0.0055 \quad 0.0311 \quad 0.4358 \quad 0.3800 \quad 0.1475)\end{aligned} \tag{36}$$

The assessing vector in equation 36 exactly represents the prediction of probability distribution, as shown in Figure 3.

Contrasting and analyzing the assessing results between the testing result and the predicting result have been carried out. Equation 23 and Figure 2 represent the safety state in 2016 based on on-spot testing, while equation 36 and Figure 3 represent the safety state in 2016 based on predicting model. In general, the possible safety states in the later stage of service are the last three states. The comparison results show that the percent error of available level is 4.7354%, the percent error of require-repair level is 6.1166%, and the percent error of discard level is 5.5548%. All these three percent error are below 10%, so the prediction of probabilistic safety assessment has high accuracy.

5 CONCLUSIONS

The key load-bearing part, steel structure, in case of failure can affect reliable operation of the whole crane, therefore carrying out the study on safety assessment and safety state prediction has become

the critical breakthrough in solving the problem. Considering the complexity and diversity of influence factors, we adopt the method, Analytic Hierarchy Process (AHP), to assess the safety level of cranes. The final results have indicated that the application of AHP on safety assessment is effective. After our research work, the following conclusions have been drawn.

1. The index system of safety assessment including target layer, standard layer and index layer has been established, which almost covers the factors affecting crane's safety. On this basis, constructing the testing system with multiple modules, and collecting the real-time running data have been carried out.
2. Based on fuzzy theory, we obtain the weight vector of each factor and membership matrix of each factor, and then depict the probability distribution diagram of current safety state. Further more, the safety score 60.99 is calculated, which indicates that the current safety state has nearly reached the discard level, and therefore this crane should be discarded in the short term.
3. To predict the safety state of cranes, we introduce the parameter *p*, transition probability between different states, and further build the model that can obtain the future safety state based on testing data of previous years. Moreover, the prediction of probabilistic safety assessment has high accuracy through comparative analysis.

REFERENCES

Bucas S. & Rumelhart P. & Gayton N. & Chateauneuf A. 2014. A global procedure for the time-dependent reliability assessment of crane structural members. *Engineering Failure Analysis* 42(6): 143–156.

Cen Z.B. & Hong H. & Qiu F.J. 2015. Research on remote monitoring system of large crane based on network. *Mechanical Research and Application* 28(01): 149–152.

Fan X.N. & Xu G.N. & Wang A.H. 2011. Estimation of fatigue residual life for cranes based on artificial neural network. *Journal of Mechanical Engineering* 47(20):69–74.

Li D.B. & Yin C.B. 2011. Structural safety assessment of tower crane based on on-site testing. *Construction Machinery* 42(11): 25–31.

Li Y.M. & Liu C.L. 2012. Integrating field data and 3D simulation for tower crane activity monitoring and alarming. *Automation in Construction* 27: 111–119.

Makovskii A.M. 1994. A diagnostic system for lifting cranes: Technical Diagnostics and Nondestructive Testing. *NDT & E International* 27(4): 217.

Szpytko J. 1998. Advanced supervision laser based system of the overhead cranes. *IFAC Proceedings Volumes* 31(15): 355–359.

Tian J.J. & Chen Z.P. & Zhang J.Y. & Huang C.L. 2009. Review and prospect of bridge crane safety detection method. *Mechanical and Electrical Engineering* 26(03): 1–5.

Wang W.X. & Wang X.H. & Huang G.J. & Wang D.H. & Xie X.P. 2013. The role of state detection technology in the safety assessment of cranes. *China Special Equipment Safety* 29(05): 4–6.

Wu X. & Luo W. & Liu L. & Huang Y.Y. 2010. Prediction of metal structure fatigue life of bridge and gantry crane in service. *China Safety Science Journal* 20(02): 95–99.

Yang R.G. & Xu G. & Fan X.N. 2009. Reliability failure criteria and residual life assessment criteria for overhead traveling crane structure. *China Safety Science Journal* 19(10): 95–100.

Yang X.Y. 2005. *Research on fatigue life of overhead traveling crane girder*. Wuhan: Wuhan University of Science and Technology.

Zhang D.P. 2017. *Research on prediction methods of girder deflection for large tonnage bridge/gantry crane*. Chengdu: Southwest Jiaotong University.

Zhou K. & Ding S.B. & Lu J.F. 2012. Study on the safety and residual life assessment system of hoisting machinery. *Hoisting and Conveying Machinery* 05:19–21.

Prognostics and system health management

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Optimal prognostic maintenance policy for railway track systems using rolling contact fatigue data

F. Dinmohammadi
School of Engineering and Built Environment, Glasgow Caledonian University, Glasgow, UK
School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK

ABSTRACT: Railway tracks are one of the most important assets in rail transport that are subject to very high stresses and have to be made of very high-quality steel alloy. The dominant form of track degradation is Rolling Contact Fatigue (RCF) which occurs mainly due to cyclic loading. If RCF is not controlled by maintenance operations, it will result in intensive correction works, service disruption, and even train derailment. In order to reduce such negative consequences and improve the quality of transport services, the railway organizations are moving towards using prognostic analytics approaches for the optimisation of maintenance and repair actions. In this paper, we investigate a cost-optimal prognostic maintenance policy for a railway track system by incorporating RCF data. The mechanism of RCF is analyzed by means of Finite Element Analysis (FEA) and the fatigue propagation behavior is described by the Paris-Erdogan power law function. The length of the cracks over time is also modelled by a stochastic gamma process and its parameters are estimated using the Maximum Likelihood Estimation (MLE) method. An optimization model is proposed to determine fixed time interval and/or Million Gross Tons (MGT) achieving the best possible balance between non-destructive tests and emergency maintenance. The proposed model is applied to support the maintenance decision-making for a conventional rail track system 60E1 in steel grade R350HT. The results show that the use of the proposed prognostic maintenance allows a significant reduction of the costs compared to the strategy when maintenance is conducted on an as-needed basis.

1 INTRODUCTION

Britain's railway system is one the most reliable, comfortable and safest rail networks in the world. Nevertheless, the country's rail network is still confronted with serious problems arising from the failures of infrastructure assets that require costly and time-consuming maintenance work. The rail infrastructure includes those assets that are fixed such as tunnels, bridges, permanent way, tracks, stations, signaling equipment, etc. The key components of a rail infrastructure system are illustrated in Figure 1.

Conducting regular Maintenance and Renewal (M&R) is essential for railway infrastructure to ensure network availability and reliability, passenger safety and comfort, and also energy efficiency. Currently, the maintenance of the railway infrastructure assets is preventive in nature and includes repair or renewal of some certain components at regular time intervals or pre-determined Million Gross Tons (MGT). However, it has been reported in many case studies that a significant portion of maintenance resources (e.g., budget, time, manpower) is wasted due to insufficiency or inefficiency of current programs.

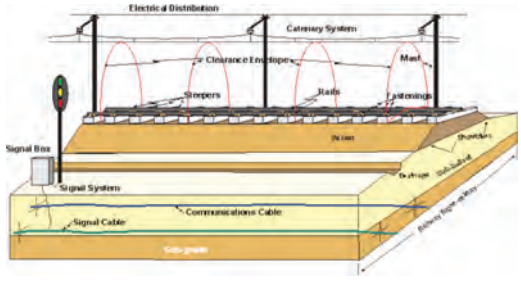


Figure 1. Railway infrastructure components.

To increase the cost-effectiveness of maintenance operations while achieving higher levels of reliability and service quality, several rail transport organisations across the Europe, such as Network Rail in the UK, Administrador de Infraestructuras Ferroviarias (ADIF) in Spain, ProRail in the Netherlands, Trafikverket in Sweden, and Österreichische Bundesbahnen (ÖBB) in Austria, etc. are moving towards using *prognostic analytics* approaches for the optimisation of maintenance programmes within the railway transport network (Dinmohammadi, 2018). However, a survey of the literature

shows that there are few studies examining the problem of optimised prognostic maintenance regimes for railway transport infrastructure assets.

Generally, railway infrastructure defects occur due to a number of specific causes that have been classified by many researchers. Olofsson and Nilsson (2002) divided the defects of steel tracks into two types of surface-initiated and subsurface-initiated cracks. Cannon *et al.* (2003) classified the steel track defects into three main groups: (i) defects originating from rail manufacture, (ii) defects originating from damage caused by inappropriate handling, installation and use, and (iii) defects caused by the exhaustion of the rail's inherent resistance to Rolling Contact Fatigue (RCF).

The RCF process in rail tracks is very complex as it depends on various factors such as age, traffic density, axle load, material properties, track geometry, curvature, speed, and accumulated MGT (Kumar, 2008). If RCF is not controlled by maintenance operations, it will result in intensive correction works, service disruption, and even train derailment. In order to analyse and control the rate of RCF growth, Non-Destructive Testing (NDT) is extensively used by railway operators. NDT includes an extensive range of inspection techniques such as ultrasonic, acoustic, eddy current, thermography, etc. that provide information about the presence of defect, its location, size and depth. The steel rail track has to undergo an emergency repair when the size or depth of crack exceeds a "warning" level. We assume that the costs for an NDT inspection and an emergency repair task are respectively C_I and C_r, where $C_r > C_I > 0$. The main problem encountered in this policy is to determine the optimal time interval (T), or MGT level (U), or number of fatigue cycles (N) for maintenance tasks such that the railroad availability is maximized and/or total inspection and repair cost is minimized.

In this paper, we formulate an optimal prognostic maintenance policy for steel rail tracks subjected to progressive RCF phenomenon. The mechanism of RCF is analyzed by means of Finite Element Analysis (FEA) and the fatigue propagation behavior is described by the Paris-Erdogan power law function. The length of the cracks over time is also modelled by a stochastic gamma process and its parameters are estimated using the Maximum Likelihood Estimation (MLE) method. The explicit expression of the long-run expected cost function per unit time is derived and the existence and uniqueness of the optimal solution are shown for the infinite-horizon case. The performance of the proposed prognostic policy in terms of cost is evaluated and compared to the case when only emergency repairs are considered.

The rest of this paper is organized as follows. The formulation of the optimization model and the properties of the optimal solution are discussed in Section 2. In Section 3, the model is applied to a real-life case study. Section 4 concludes this study.

2 MODEL FORMULATION

The aim of this model is to determine an optimal time interval or MGT level or number of fatigue cycles for the maintenance of steel railway tracks in case that they are subject to risk of progressive RCF. The RCF process typically involves three following phases: (i) initiation, (ii) propagation (or growth), and (iii) the failure. The step-by-step procedure on how to implement the model is presented below:

1. Suppose that the RCF processes initiate in the interval [0, t) following a Non-Homogeneous Poisson Process (NHPP), $\{N(t); t \geq 0\}$ with intensity function $m(t)$ and mean value function $M(t)$, i.e,

$$M(t) = \int_0^t m(x)dx, t \geq 0, \tag{1}$$

where t is the age of the track system and $M(t)$ is a non-decreasing function of t with $M(t) = 0$. Then, the probability that exactly j (= 0, 1, 2, …) RCF processes occur in the interval [0, t), $P_j(t)$ is given by

$$P_j(t) = P\{N(t) = j\} = e^{-M(t)} \times \frac{[M(t)]^j}{j!} \tag{2}$$

Let $T_j (j = 0, 1, 2, \ldots)$ denote the initiation time of the j^{th} RCF process in track system, where $T_0 = 0$. Then, the cumulative distribution function of the random variable T_j is given by

$$F_j(t) = P\{T_j \leq t\} = \sum_{i=j}^{\infty} p_i(t),\ j = 1,2,\ldots. \tag{3}$$

2. Propagation is the second phase of the RCF process which may be accelerated by adverse environmental conditions. Many models have been developed to study how RCF process in steel rail tracks propagate. For instance, Ringsberg (2001) proposed a crack growth model for railway tracks in which the crack propagation life is divided into three stages: (i) shear stress driven initiation at the surface; (ii) transient crack growth behavior; and (iii) subsequent tensile and/or shear driven crack growth (see Figure 2).

In this paper, the RCF propagation is modelled using a stochastic *gamma* process, which represents the evolution of RCF length/size/depth in time. The gamma process is a stochastic process with independent non-negative increments having a gamma

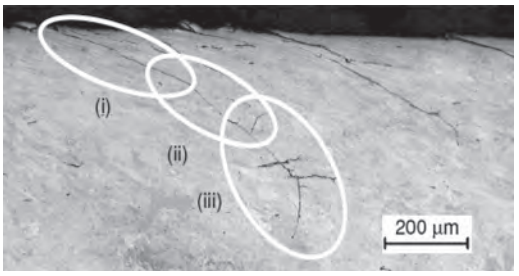


Figure 2. Crack propagation phenomenon in railway tracks (Ringsberg and Bergkvist, 2003).

distribution with identical scale parameter. The gamma process has been widely studied for different maintenance applications by several authors (see van Noortwijk (2009) for a thorough review on the use of gamma process in maintenance modeling). Also, it has been observed that the gamma process is satisfactorily fitted to data of different gradual degradation phenomena (such as wear and crack propagation) in railway industry (Meier-Hirmer *et al.*, 2005). Moreover, the existence of an explicit probability distribution function of gamma process permits feasible mathematical developments.

Let $X_j(t)$ be the length/size/depth of the j^{th} RCF process after t units of time from initiation. We assume that $X_j(t)$ has a homogeneous gamma process with shape and scale parameters given by αt and β respectively. Then, for $t > 0$, the density and the cumulative distribution function of the increment of the length/size/depth of the j^{th} RCF process is given by (Park and Padgett, 2008):

$$g_{\alpha t,\beta}(x) = \frac{\beta^{\alpha t}}{\Gamma(\alpha t)} x^{\alpha t-1} e^{-\beta x},\ x \geq 0, \tag{4}$$

and

$$G_{\alpha t,\beta}(x) = \frac{\gamma(\alpha t, \beta x)}{\Gamma(\alpha t)},\ x \geq 0;\ \alpha, \beta > 0, \tag{5}$$

where $\Gamma(\cdot)$ [$\gamma(.,.)$] denotes the gamma [incomplete gamma] function, i.e.,

$$\Gamma(\upsilon) = \int_0^\infty z^{\upsilon-1} e^{-z} dz;\ \gamma(\upsilon, u) = \int_u^\infty z^{\upsilon-1} e^{-z} dz,\ \upsilon, u > 0.$$

3. Steel railway track undergoes an emergency repair when the length/size/depth of RCF exceed a warning level D. In the event of a repair, the track segment returns to an "as-good-as-new" condition. Let U_j be the length of the interval between the initiation time of the j^{th} RCF process to the time that it attains the warning threshold D. Thus,

$$U_j = \inf\left\{t \geq 0 : X_j(t) \geq D\right\},\ j = 1,2,\ldots, \tag{6}$$

Then, from Eqs. (4) and (5), the density and the cumulative distribution function of U_j, respectively, are given by

$$g_{Uj}(t) \equiv g_U(t) = \frac{\beta^{\alpha t}}{\Gamma(\alpha t)} D^{\alpha t-1} e^{-\beta D},\ t \geq 0;\ \alpha, \beta > 0, \tag{7}$$

and

$$G_{Uj}(t) \equiv G_U(t) = \frac{\Gamma(\alpha t, \beta D)}{\Gamma(\alpha t)},\ t \geq 0;\ \alpha, \beta > 0. \tag{8}$$

We denote by S_j the time point that the length/size/depth of the j^{th} RCF process exceeds the warning level D. Then,

$$S_j = T_j + U_j,\ j = 1,2,\ldots. \tag{9}$$

Lemma. Let $\mathrm{I_A}(.)$ denote the indicator function that is defined as $\mathrm{I_A}(x) = 1$ for $x \in \mathrm{A}$, and 0 otherwise. Let $\{N_S(t);\ t \geq 0\}$ be the counting process associated with the random variables S_j ($j = 0, 1, 2, \ldots$), that is,

$$N_S(t) = \sum_{j=1}^{\infty} I_{[0,t]}(S_j), \tag{10}$$

Then, having in mind that the convolution of any functions $a(.)$ and $b(.)$ is given by

$$a(x) \bullet b(x) = \int_0^x a(x-t) db(t),$$

$\left\{N_S(t); t \geq 0\right\}$ is an NHPP with intensity function,

$$h(t) = m(t) \bullet g_U(t), \tag{11}$$

where $g_U(t)$ is given by Eq. (7) (Shafiee and Finkelstein, 2015).

Let X_r denote a maintenance cycle defined by the time interval between maintenance actions (either NDT inspection or emergency repair). Under the assumptions of the model, we have

$$X_r = \min(S_{[1]}, T), \tag{12}$$

where $S_{[1]}$ denotes the time that, for the first time, an RCF process exceeds the warning threshold D, i.e.,

$$S_{[1]} = \min\left\{S_j,\ j = 1,2\cdots\right\}, \tag{13}$$

Then, by using lemma, the survival function of $S_{[1]}$ is given by

$$\bar{F}_{S_{[1]}}(t) = P\{S_{[1]} > t\} = P\{N_s(t) = 0\} = \exp\left(-\int_0^t h(x)dx\right), \quad (14)$$

where $h(.)$ is the failure rate function of $S_{[1]}$, and is given by Eq. (11). Then, the expected length of a maintenance cycle $E[X_r]$, is given by

$$E[X_r] = \int_0^T \bar{F}_{S_{[1]}}(t)dt, T > 0. \quad (15)$$

Let $C(t)$ be the *s*-expected cost of operating the system for the time interval [0, t). From the *renewal reward theorem* (see Ross 1970, p. 52), the expected cost rate, denoted by $CR(t)$, is the expected operational cost incurred in a maintenance cycle divided by the expected cycle length, i.e.,

$$CR(T) = \lim_{t\to\infty} \frac{C(t)}{t} = \frac{C_r \times F_{S_{[1]}}(T) + C_I \times \bar{F}_{S_{[1]}}(T)}{\int_0^T \bar{F}_{S_{[1]}}(t)dt}, \quad (16)$$

where $F_{S_{[1]}}(\cdot)\left[\bar{F}_{S_{[1]}}(\cdot)\right]$ is the cumulative distribution [survival] function of $S_{[1]}$. The problem is to find the optimal value of T^* that minimizes the objective function $CR(T)$, given in Eq. (16). Therefore, the proposed optimization model can be formulated as follows:

$$minimise\ CR(T) = \frac{C_I + \int_0^T (C_r - C_I)h(t)\bar{F}_{S_{[1]}}(t)dt}{\int_0^T \bar{F}_{S_{[1]}}(t)dt}. \quad (17)$$

The following theorem solves this problem.

Theorem. If $h(T)$ is strictly increasing in t, and $\lim_{T\to\infty} h(T_{max}) > CR(T_{max})$, there exists an unique and finite minimum $T^* \in (0, T_{max})$ that verifies the following equation:

$$\bar{F}_{S_{[1]}}(T^*) + h(T^*) \times \int_0^{T^*} \bar{F}_{S_{[1]}}(t)dt = \frac{C_r}{C_r - C_I}, \quad (18)$$

whereas, if $h(T)$ is non-decreasing in t, and $(C_r - C_I)h(T_{max}) \le CR(T_{max})$, then $T^* = T_{max}$ (implying maximum preventive replacement interval).

3 SIMULATED CASE STUDY

In this Section, we present an application of the RCF-based maintenance decision-making model to a conventional rail track system 60E1 in steel grade R350HT. 60E1 is a common rail profile developed by the International Union of Railways (UIC) (http://uic.org/). It also has been widely used by Network Rail as the standard for all new high speed rail lines. The rail 60E1 profile was designed by 3D Abaqus simulation software and is shown in Figure 3.

UIC60 rails are made of high-quality steel alloys, resistant to Rolling Contact Fatigue (RCF), wear, plastic deformation and corrosion. Different sizes of UIC60 rails are available and used on railroads. In this study, a standard rail section as shown in Figure 4 was chosen for the analysis.

The material properties of the rail (e.g. mass, density and Young's modulus) were provided by the manufacturer, British Steel (britishsteel.co.uk). We assume that the arrival of RCF cracks on rail track follows a Poisson process with rate $\hat{m} = 0.144$/month (Shafiee *et al.*, 2016). The

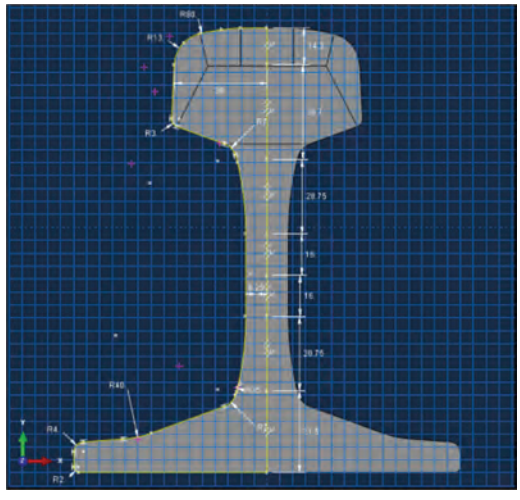


Figure 3. Rail 60E1 profile designed by 3D Abaqus simulation.

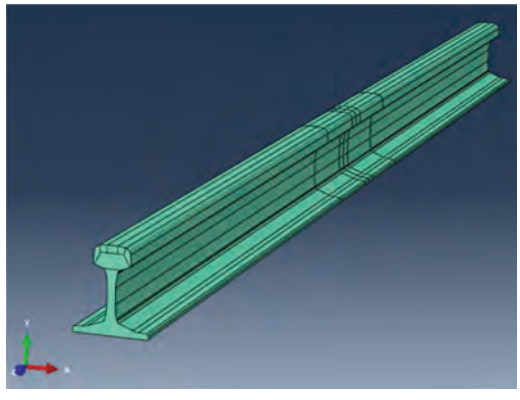


Figure 4. A rail section designed by 3D Abaqus simulation.

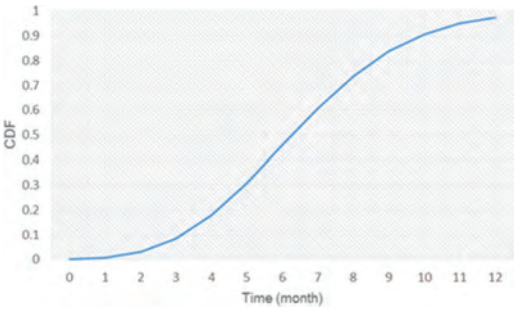


Figure 5. CDF of time-to-initiate a RCF crack.

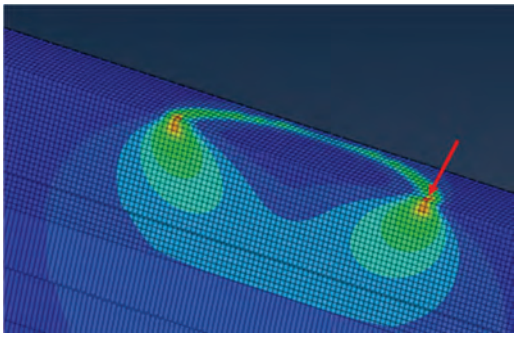


Figure 6. Cross section of contact patch localised on rail track.

Cumulative Distribution Function (CDF) of time-to-initiate a crack is illustrated in Figure 5. The mean-time-to-initiate a crack is estimated 6.94 months.

In order to estimate the rate of degradation propagation, an FEA simulation model was constructed by commercially available software package, Abaqus. The cross section of the contact patch localised on rail track is shown in Figure 6.

Many models have so far been developed *to describe the crack growth* process, e.g. Paris' law (also known as the Paris-Erdogan law). *The Paris' law is a power-function used to predict* crack evolution for structures subject to fatigue stresses. The Paris' law equation is given as follows (Paris and Erdogan, 1963):

$$\frac{da}{dN} = C(\Delta K)^m, \tag{19}$$

where a represents the crack length, N represents the number of load cycles, da/dN is the fatigue crack growth rate per cycle, and C and m are empirical constants (usually referred to as Paris' law parameters) which depend on material properties and operating environment. The range of the stress intensity factor, ΔK represents the difference between the stress intensity factor at maximum and minimum loads for a particular crack length and is calculated as:

$$\Delta K = K_{\max} - K_{\min} = \Delta\sigma Y\sqrt{\pi a}, \tag{20}$$

where $K_{\max}$ and $K_{\min}$ are, respectively, the maximum and minimum stress intensity factors, $\Delta\ \sigma$ is the range of cyclic stress amplitude, and Y is a dimensionless parameter that depends on the crack and loading geometries. The remaining cycles can be found by substituting Eq. (20) into Eq. (19):

$$\frac{da}{dN} = C(\Delta\sigma Y\sqrt{\pi a})^m, \tag{21}$$

For relatively short cracks, Y can be assumed as independent of a and the differential equation can be solved via separation of variables

$$\begin{aligned}\int_0^{N_f} dN &= \int_{a_i}^{a_c} \frac{da}{C(\Delta\sigma Y\sqrt{\pi a})^m} \\ &= \frac{1}{C(\Delta\sigma Y\sqrt{\pi a})} \int_{a_i}^{a_c} a^{-\frac{m}{2}},\end{aligned} \tag{22}$$

and subsequent integration

$$N_f = \frac{2\left(a_c^{1-\frac{m}{2}} - a_i^{1-\frac{m}{2}}\right)}{(2-m)C(\Delta\sigma Y\sqrt{\pi})^m}, \tag{23}$$

where N_f is the remaining number of cycles to fracture, a_c is the critical crack length at which instantaneous fracture will occur, and a_i is the initial crack length at which fatigue crack growth starts for the given stress range $\Delta\sigma$. If Y strongly depends on a, numerical methods might be required to find reasonable solutions. Basically, crack propagation can be divided into three stages: stage I (short cracks), stage II (long cracks) and stage III (final fracture). At stage I, once a fatigue crack is initiated, it propagates along high shear stress planes (45 degrees). When the stress intensity factor K increases as a consequence of crack growth or higher applied loads, slips start to develop in different planes close to the crack tip, initiating stage II. Finally, stage III is related to unstable crack growth as K_{max} approaches K_{IC}. At this stage, crack growth is controlled by static modes of failure and is very sensitive to the microstructure, load ratio, and stress state (plane stress or plane strain loading). The cross-section of a fatigue crack introduced to the model in 100 nm scale is shown in Figure 7.

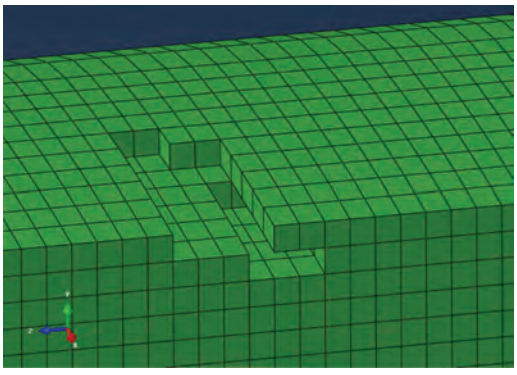


Figure 7. Cross-section of a crack that was introduced to model (scale bar: 100 nm).

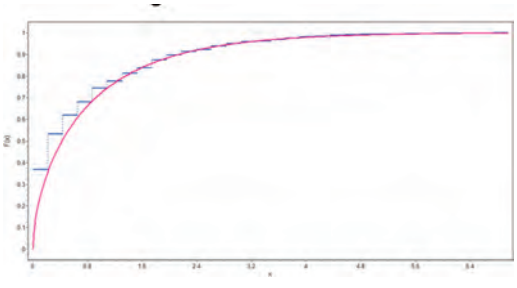


Figure 8. Cumulative distribution function for crack length.

The length of the fatigue cracks in time is modelled using the gamma distribution as presented in Equations (4) and (5). The parameters of the Gamma process are estimated by means of maximum likelihood estimators given in Kahle *et al.* (2016) and are reported as follows:

$\hat{\alpha} = 0.576$ and $\hat{\beta} = 1.50$ (α/β is 0.384 mm per month).

The best-fit models for the cumulative distribution function associated with the length of RCF processes are shown in Figure 8.

Average length of the rail repaired emergently is $L = 8$ meters (Patra, 2009). The cost of 60E1 railway track (including neutralization) per meter is 2,250 Monetary Unit (MU). Average labour cost per hour (including the track worker cost, track welder cost, and inspection personnel cost) is 625 MU. The hourly rate of hiring the welding equipment or service vessels for maintenance, replace or inspection of the railway track is 80 MU. The mean time required to perform a maintenance (either corrective or preventive) is 4 hours. However, the corrective type may cause traffic disruption that incurs an additional cost η to the route operator. The physical lifetime of 60E1 railway track is considered to be equal to six years (72 months). We wrote a MATLAB code for the minimization of the expected cost rate, as given in Eq. (17), to determine the optimal preventive NDT inspection interval T^*. The pictorial representation of the expected cost rate as a function of the inspection interval T for three different cases of (i) $\eta = C_r - C_I = 0$; $\eta = 5{,}000$, and $\eta = 10{,}000$ MU is shown in Figure 9. From Figure 9, it is found that the use of optimal prognostic inspection policy allows a significant reduction of the maintenance cost compared to the strategy when only emergency repair is considered (the corresponding cost is the asymptote of the path, when T tends to T_{max}). The optimal value of T^* and the corresponding expected cost rate, $CR(T^*)$ the expected cost rate for corrective maintenance policy, $CR(T_{max})$, and the percentage reduction of the maintenance cost, r are presented in Table 1.

It can be seen that as the cost parameter η increases, the optimal NDT inspection interval T^* becomes shorter, however the expected cost rate, $CR(T^*)$ increases. Also, when a large additional cost is likely to be incurred by the infrastructure

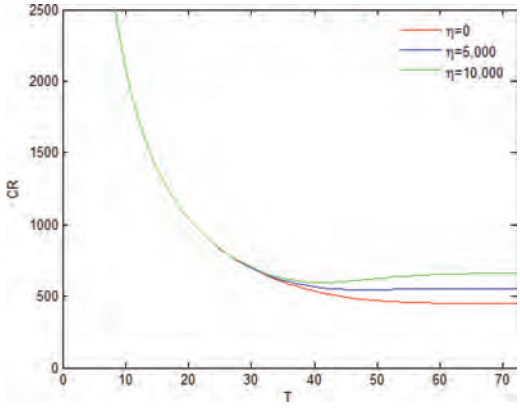


Figure 9. Expected cost rate for different values of η.

Table 1. Results of the optimization model for different values of η.

	Unit	$\eta = 0$	$\eta = 5{,}000$	$\eta = 10{,}000$
T^*	month	72	49	42
$CR(T^*)$	MU / month	448.38	546.40	594.81
$CR(T_{max})$	MU / month	448.38	554.94	661.49
Cost reduction	%	0	1.54	10.08

owner in emergency maintenance case, applying the prognostic inspection policy will be more efficient than corrective repair and has a huge potential to reduce the maintenance cost. For instance, when the cost parameter η is 10,000, the prognostic maintenance policy allows for approximately %10 reduction of the maintenance cost compared to the corrective repair policy.

4 CONCLUSIONS

In this paper, an optimal prognostic Non-Destructive Testing (NDT) and inspection policy is presented for railway track systems subject to progressive Rolling Contact Fatigue (RCF). The RCF phenomenon was modelled by means of Finite Element Analysis (FEA) and the fatigue propagation behavior was described by the Paris-Erdogan power law function. The length/size/depth of the RCF processes over time was formulated by a stochastic gamma process and a preventive inspection was conducted before the length/size/depth of cracks exceeds a "warning" level. This study can be extended in many directions to make it more practical in maintenance management of railway industry. The model can be formulated and analyzed for the case when RCF is only detected if its length/size/depth reaches a detection threshold c (> 0); and, a cost comparison can be made between the proposed prognostic inspection policy and other common strategies such as RCM.

REFERENCES

Aven, T. (1992). *Reliability and risk analysis*. Elsevier Applied Science, London.

Cannon, D.F., Edel, K.O., Grassie, S. L. and Sawley, K. (2003) Rail defects: an overview, *Fatigue & Fracture of Engineering Materials & Structures* 26(10), 865–886.

Dinmohammadi, F. (2018) *Data-driven risk-based modelling approaches to maintenance optimisation of railway transport assets.* Glasgow Caledonian University, UK.

Kahle, W., Mercier, S. and Paroissin, C. (2016) Gamma processes, In: Degradation processes in reliability, John Wiley & Sons, Hoboken, NJ, USA. DOI: 10.1002/9781119307488.ch2.

Kumar, S. (2008) *Reliability analysis and cost modeling of degrading systems*. Doctoral Thesis, Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden.

Olofsson, U. and Nilsson, R. (2002). Surface cracks and wear of rail: a full-scale test on a commuter train track, *In Proc. of the Institution of Mechanical Engineers,* 216(4), 249–264.

Meier-Hirmer, C., Sourget, F. and Roussignol, M. (2005) Optimising the strategy of track maintenance. In Proceedings of the European Safety and Reliability Conference (*ESREL*), June 27–30, Tri City, Poland, pp. 1385–91.

Park, C. and Padgett, W.J. (2008) Cumulative damage models based on gamma processes. *Encyclopedia of Statistics in Quality and Reliability.* DOI: 10.1002/9780470061572.eqr119.

Patra, A.P. (2009) *Maintenance decision support models for railway infrastructure using RAMS & LCC analyses*. Doctoral Thesis, Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden.

Ringsberg, J.W. (2001) Life prediction of rolling contact fatigue crack initiation. *International Journal of Fatigue* 23, 575–586.

Ringsberg, J.W. and Bergkvist, A. (2003) On propagation of short rolling contact fatigue cracks. *Fatigue & Fracture of Engineering Materials & Structures* 26(10), 969–983.

Ross, S.M., *Applied Probability Models with Optimization Applications,* Holden-Day, San Francisco, CA, 1970.

Shafiee, M., Patriksson, M., Chukova, S. (2016) An optimal age-usage maintenance strategy containing a failure penalty for application to railway tracks. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 230(2), 407–417.

Shafiee, M. and Finkelstein, M. (2015) An optimal age-based group maintenance policy for multi-unit degrading systems. *Reliability Engineering & System Safety* 134, 230–238.

van Noortwijk, J.M. (2009) A survey of the application of gamma processes in maintenance. *Reliability Engineering and System Safety* 94, 2–21.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

An evaluation method of methodology for integration of HALT, HASS and ADT

Tianji Zou & Peng Li
Technology and Engineering Centre for Space Utilization, Chinese Academy of Sciences, Beijing, China
University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing, China

Wei Dang, Kai Liu & Ge Zhang
Technology and Engineering Centre for Space Utilization, Chinese Academy of Sciences, Beijing, China

ABSTRACT: Accelerated Degradation Testing (ADT) is used to collect more performance degradation data under accelerated stress levels in a limited time. However, limited sample size or inadequate testing time may lead to the lack of degradation information, which causes inaccuracy of the evaluation of the lifetime. High Accelerated Life Test (HALT), High Accelerated Stress Screen (HASS) are common testing methods, whose information should be considered in evaluation of the lifetime and reliability. In this paper, Methodology for Integration of HALT, HASS and ADT (MIHHA) is proposed as the method with the multi-utilization of HALT, HASS and ADT. One difficulty of MIHHA is how to integrate the information of HALT, HASS and ADT to provide crucial information for product life prediction. We propose an evaluation method of MIHHA which integrated degradation information of HLAT, HASS into ADT evaluation based on Bayesian theory. This method would be a great help to the engineering application of MIHHA.

1 INTRODUCTION

During the past decades, Accelerated Degradation Testing (ADT) has drawn much more attention in both industry and academia, which can collect more performance degradation data under accelerated stress levels in a limited time (Ge et al. 2012). It is convenient to establish model of reliability and life assessment with degradation information through ADT technique. However, sometimes limited sample size or inadequate testing time may lead to the lack of degradation information, which causes inaccuracy of the evaluation of the lifetime and reliability. Usually, Highly Accelerated Life Testing (HALT) and highly Accelerated Stress Screening (HASS) are mainly applied in qualitative method to improve the reliability of products (Liu et al. 2016). But the information of HALT and HASS can make up for the lack of ADT information very well. Based on this, Methodology for Integration of HALT, HASS and ADT (MIHHA) was proposed to coordinate the design of HALT, HASS, ADT, and collect more information under the limited time and samples.

The different between ADT and MIHHA is MIHHA contains information of HALT, HASS. The key issue of MIHHA is how to integrate the information of HALT, HASS into ADT to predict reliability or lifetime of product. As quality and reliability of product improved, there was less failure in HALT and HASS. People pay more attention to the degradation information in HALT and HASS (Gray & Paschkewit, 2016). In this article, we consider to make full use of degradation information of HALT and HASS in evaluation method of MIHHA.

Bayesian inference is an important technique in statistical inference (Box & Tiao 1992). In Bayesian inference of ADT evaluation method, the degradation data of HALT and HASS can be regard as the prior information and degradation data of ADT can be considered as the sample information. The theory of Bayesian can successfully integrate degradation data of HALT, HASS into ADT evaluation method. To solve the problem that there are two sources of prior information come from HALT and HASS, a weighting fusion method is adopted to integrate them into one prior distribution based on correlation function. The technology roadmap of the evaluation method of MIHHA with integrated degradation information of HLAT, HASS is shown in Fig. 1.

In conclusion, this article proposes an evaluation method of MIHHA based on Bayesian inference, which integrated degradation information of HLAT, HASS and ADT. The evaluation method

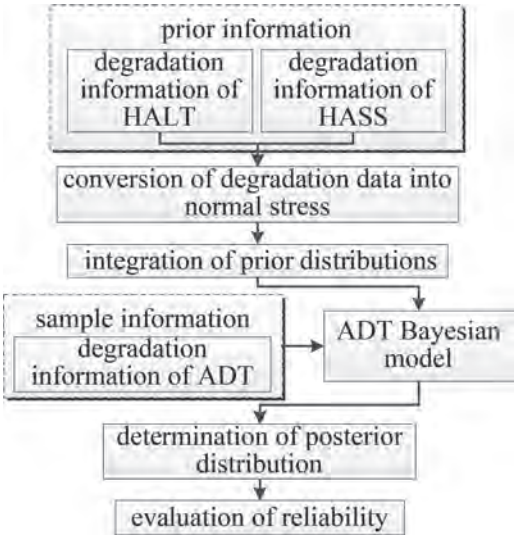


Figure 1. Technology roadmap of the evaluation method of MIHHA.

of MIHHA can make full use of various sources of information and improve the engineering application value of MIHHA.

2 BAYESIAN INFERENCE OF ADT MODEL

ADT Model and the relationship of parameters in Bayesian inference are essential parts of the evaluation method and integration method. In this paper, we describe the degradation paths through Wiener process and give the relationship of ADT model and parameters in Bayesian inference.

2.1 *ADT model based on Wiener process*

Wiener process is used for degradation data analysis widely (Whitmore & Schenkelberg 1997). The degradation model of Wiener process can be expressed as:

$$Y(t) = \sigma B(t) + d(s) \cdot t + y_0 \tag{1}$$

where $Y(t)$ is the performance degradation process of product, $B(t)$ means the standard Brownian motion, denoted as $B(t) \sim N(0,t)$, σ is a constant named diffusion coefficient, which is free from changes caused by stress or time. y_0 represents initial value of product performance. $d(s)$ means the drift coefficient, which represents the degradation rate of product under specific stress level s. Normally we use acceleration model to describe the relationship between $d(s)$ and s (Zou, T.J., et al., 2015). This article defines acceleration model as:

$$\ln d(s) = a + b\varphi(s) \tag{2}$$

where, $\varphi(s)$ is a function of stress s. Known from the property of the Wiener process, the degradation increment Δy during the unit time Δt is subject to a normal distribution with the mean of $d(s) \cdot \Delta t$ and the square deviation of $\sigma^2 \cdot \Delta t$, i.e., $\Delta y \sim N(d(s) \cdot \Delta t, \sigma^2 \cdot \Delta t)$. The probability density function of Δy is:

$$f(\Delta y) = \frac{1}{\sigma\sqrt{2\pi\Delta t}} \exp\left\{ -\frac{\left[\Delta y - d(s) \cdot \Delta t\right]^2}{2\sigma^2 \cdot \Delta t} \right\} \tag{3}$$

The literature (Whitmore & Schenkelberg 1997) suggests that the first passage time of Wiener process to a threshold follows the inverse Gaussian distribution. We assume y_f as the failure threshold of performance degradation. In other word, product failures when $Y(t) - y_f < 0$. Thus, the key of ADT evaluation method turn to the calculation of probability distribution of the first passage time t to the threshold y_f. The probability density function of the inverse Gaussian distribution with the mean $\mu = ((y_f - y_0)/d(s))$ and the shape parameter $\lambda = (y_f - y_0/\sigma)^2$ is:

$$f(t; y_0, y_f) = \frac{y_f - y_0}{\sigma\sqrt{2\pi t^3}} \exp\left\{ -\frac{\left[(y_f - y_0) - d(s) \cdot t \right]^2}{2\sigma^2 t} \right\} \tag{4}$$

The reliability function of degradation process (Lu, J., 1995) is given by:

$$R(t) = \Phi\left[\frac{y_f - y_0 - d(s)t}{\sigma\sqrt{t}} \right] - \exp\left(\frac{2d(s)(y_f - y_0)}{\sigma^2} \right) \cdot \Phi\left[-\frac{y_f - y_0 + d(s)t}{\sigma\sqrt{t}} \right] \tag{5}$$

From the reliability function above, the unknown parameters a,b,σ are the key parameters in ADT evaluation method, which are also known as prior parameters in Bayesian inference. We assume $\boldsymbol{\theta}$ as the prior parameter vector of model, i.e. $\boldsymbol{\theta} = (a,b,\sigma)$. The relationship of ADT model and prior parameters is shown in the Fig. 2.

If prior parameters are obtained, it is easy to accomplish the evaluation of lifetime and reliability (Wang et al. 2016). The crux of evaluation method of MIHHA is how to obtain reasonable valuations of prior parameters through degradation information of HLAT, HASS and ADT.

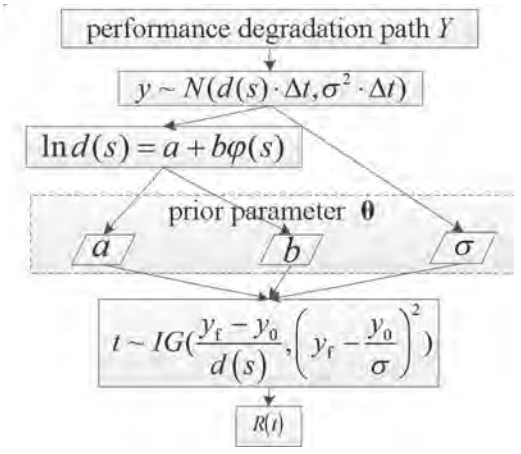


Figure 2. The relationship of ADT model and prior parameters.

2.2 *Bayesian inference of ADT model*

Bayesian inference is an important technique in statistical inference. Normally, we use Bayesian theorem to update the probability for a hypothesis as more evidence or information becomes available (Wang et al. 2016). It is suitable for integrating degradation information of HLAT, HASS and ADT in evaluation method. Bayesian inference is particularly important in the dynamic analysis of a sequence of data. Through Bayesian inference, this article can obtain the posterior distribution of the parameters to evaluate lifetime and reliability of product with degradation information of HALT, HASS and ADT.

In this paper, the observed data x (also known as sample information) can be obtained from degradation data of ADT and degradation data of HLAT, HASS can be regarded as prior information. The posterior distribution is the distribution of the parameters after taking into account the observed data, which combines the observed data and the prior information and forms the core of Bayesian inference. With the observed data x, the Bayesian theory suggests that the posterior distribution of parameters can be expressed as:

$$\pi(\theta \mid x) = \frac{\pi(\theta)\cdot p(x \mid \theta)}{\int_{\Theta} \pi(\theta)\cdot p(x \mid \theta) d\theta} \tag{6}$$

where, Θ is value range of θ. $\pi(\theta)$ is the prior distribution. $p(x|\theta)$ is the distribution of the observed data conditional on its parameters, which is termed the sampling distribution or likelihood function. In this paper, it is expressed as:

$$p(x \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} \exp\left\{ -\frac{\left[x - \exp(a + b\cdot\varphi(s))\cdot\Delta t \right]^2}{2\sigma^2\Delta t} \right\} \tag{7}$$

3 INTEGRATION OF PRIOR DISTRIBUTIONS

3.1 *Integrate degradation information from HALT and HASS*

Before the integration, we should make sure that the degradation data from HALT, HASS and ADT meet the following conditions: 1) the degradation data from HALT, HASS and ADT should belong to the same performance parameter. 2) the degradation mechanism of performance degradation data from HALT, HASS and ADT should be the same (Wang et al. 2013).

The degradation data of HLAT, HASS can be regarded as prior information. Thus, we assume $\pi_1(\theta)$ represents the prior distribution obtained from degradation information of HALT and $\pi_2(\theta)$ represents the prior distribution obtained from degradation information of HASS. For there are two prior distributions about parameters, a weighting fusion method is adopted to integrate into one prior distribution. Through using the weight coefficient α_1 and α_2, the prior distribution after integration is expressed as:

$$\pi(\theta) = \alpha_1\pi_1(\theta) + \alpha_2\pi_2(\theta) \tag{8}$$

where, the weight coefficient α_1 and α_2 are constrained by formula as follow:

$$\alpha_1 + \alpha_2 = 1 \tag{9}$$

According to the Bayesian theory, the posterior distribution after integration is expressed as:

$$\pi(\theta \mid x) = \frac{\alpha_1\pi_1(\theta)\, p(x \mid \theta) + \alpha_2\pi_2(\theta)\, p(x \mid \theta)}{\int_{\Theta} \alpha_1\pi_1(\theta)\, p(x \mid \theta) + \alpha_2\pi_2(\theta)\, p(x \mid \theta)\, d\theta} \tag{10}$$

It can be transformed into:

$$\pi(\theta \mid x) = \frac{\alpha_1 m_1(x)}{m(x)}\cdot\pi_1(\theta \mid x) + \frac{\alpha_2 m_2(x)}{m(x)}\cdot\pi_2(\theta \mid x) \tag{11}$$

where, $\pi_1(\boldsymbol{\theta}|x)$ and $\pi_2(\boldsymbol{\theta}|x)$ represent the posterior distribution of HALT and HASS independently.

$m(x)$ is known as the posterior predictive distribution, which is the distribution of a new data point, marginalized over the posterior:

$$m_i(x) = \int_{\Theta} \pi_i(\theta)\, p(x|\theta)\, d\theta \quad i = 1,2 \tag{12}$$

$$m(x) = \alpha_1 m_1(x) + \alpha_2 m_2(x) \tag{13}$$

Here, we assume:

$$\beta_i = \frac{\alpha_i m_i(x)}{m(x)} \tag{14}$$

So we can easily reach the conclusion:

$$\beta_1 + \beta_2 = 1 \tag{15}$$

And, the expression of posterior distribution after integration is transformed into formula (16), which is a weighted equation of posterior distributions with the weight coefficient β_1 and β_2.

$$\pi(\theta|x) = \beta_1 \pi_1(\theta|x) + \beta_2 \pi_2(\theta|x) \tag{16}$$

3.2 *Calculation based on correlation function*

Correlation function is a function that gives the statistical correlation between random variables. Sometimes correlation functions of different random variables are referred to as cross-correlation functions to emphasize that different variables are being considered. Through analyzing the correlation function, the correlativity between posterior distributions and prior distributions can be weighted.

In this paper, we set u_i as the mean of $\pi_i(\theta)$, u_i' as the mean of $\pi_i(\theta'|x)$. According to the formula(16), the mean of posterior distribution $\pi(x|\theta)$ can be descried as:

$$u' = \beta_1 u_1' + \beta_2 u_2' \tag{17}$$

Then we assume three equations as follow to describe the correlativity between prior and posterior distributions.

$$\begin{cases} u' = a_0 + a_1 u_1 + a_2 u_2 \\ u_1' = a_{01}' + a_{11}' u_1 + a_{21}' u_2 \\ u_2' = a_{02}' + a_{12}' u_1 + a_{22}' u_2 \end{cases} \tag{18}$$

There are three unknown coefficients in each equation in formula(18). Through simulation sampling, three groups of random data are obtained. So we can get three group of estimated value of u_i, u_i', u' to solve the equations.

We $v_i'^2$ as the square deviation of $\pi_i(\theta|x)$. The correlation function defined as (Wang et al. 2013):

$$r_i = \frac{d_{1i}' a_1 v_1'^2 + d_{2i}' a_2 v_2'^2}{\sqrt{d_{1i}'^2 v_1'^2 + d_{2i}'^2 v_2'^2} \cdot \sqrt{a_1^2 v_1'^2 + a_2^2 v_2'^2}} \quad i = 1,2 \tag{19}$$

From the formula(19) we can see $0 \le r_i \le 1$. A lager r_i means $\pi_i(\theta|x)$ is more closely associated with $\pi(\theta|x)$, which illustrates that $\pi_i(\theta)$ should have a stronger weight in $\pi(\theta)$. The weight coefficient α_1 and α_2 of prior distributions are obtained by formula (20):

$$\alpha_i = \frac{r_i}{r_1 + r_2} \quad i = 1,2 \tag{20}$$

With the values of α_i, we can easily obtain the posterior distribution $\pi(x|\theta)$ after integration. According to the reliability function of ADT, it is easy to achieve the evaluation of ADT.

4 A SIMULATION EXAMPLE

4.1 *Data simulation*

Accelerated test conditions are typically produced by testing units at higher levels of temperature, voltage, pressure, etc (Nelson 1971). In this paper, we assume the degradation process is mainly affected by high temperature. Arrhenius model is suitable to describe the relationship between the degradation rate and temperature (Alferink 2012), which means $\varphi(s) = 1/(s + 273.15)$. s is Celsius temperature.

In order to verify the effectiveness of the evaluation method of MIHHA, we simulated 240 degradation data points for each of the 6 samples under HALT and 30 degradation data points for each of the 6 samples under HASS. Performance monitor interval, ie.Δt, is one minute. Fig. 3 shows the degradation paths under different high-temperature level in HALT. Fig. 4 shows degradation paths of high temperature at 120°C involving vibration synchronously under different cycle in HASS. Both data of Fig. 3 and Fig. 4 are the prior information in Bayesian inference.

Then, we simulated a 2-day Step Stress ADT (SSADT) with 2880 degradation data points for each of the 6 samples under as observed data or sample information. Fig. 5 shows the degradation paths of SSADT under different stress.

According to the formula (2), we can convert all the degradation data under accelerated stress into the degradation data of 25°C by method of linear regression. The equivalent Δt_i at 25°C are shown in the Table 1. We can obtain the $d(s) \cdot \Delta t$ at 25°C of different samples under HALT, HASS and ADT through $\Delta y / \Delta t_i$.

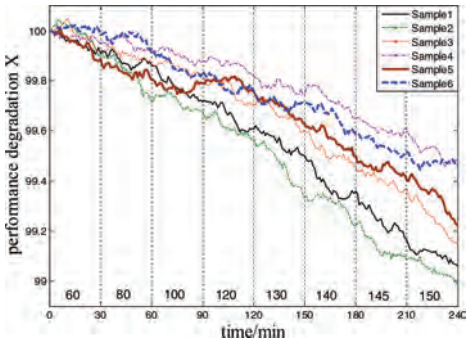


Figure 3. The degradation paths of high-temperature test in HALT.

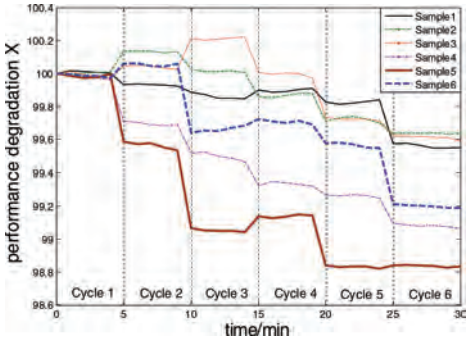


Figure 4. The degradation paths of high temperature vibration synthesis test in HASS.

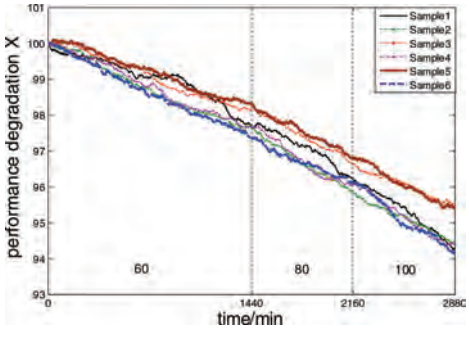


Figure 5. The degradation paths of SSADT.

4.2 *Integration of prior distributions*

As all the data have be converted into the data of 25°C, we can set $d(s) \cdot \Delta t$ as prior parameter replace a and b to reduce the calculation. The prior information is obtained from the distribution data of HALT and HASS as show in the Table 2.

With the help of Matlab, we simulate three groups of random samples from prior distribution

Table 1. Conversion factors of Δt_i

Accelerated stress	Δt/min	Equivalent Δt_i at 25°C/min
60°C	1	1.447711
80°C		1.730614
100°C		2.029593
120°C		2.341943
130°C		2.502343
140°C		2.665169
145°C		2.747404
150°C		2.830144

Table 2. Prior distributions.

Prior parameters	$d(s) \cdot \Delta t$	$\sigma^2 \cdot \Delta t$
HALT	N(0.0014, 0.00041)	N (2.34e-05, 2.61e-06)
HASS	N(0.0098, 0.0047)	N (1.21e-03, 9.19e-04)

$\pi_1(d(s) \cdot \Delta t)$, $\pi_1(\sigma^2 \cdot \Delta t)$, $\pi_2(d(s) \cdot \Delta t)$, $\pi_2(\sigma^2 \cdot \Delta t)$ and posterior distribution $\pi_1(d(s) \cdot \Delta t|x)$, $\pi_1(\sigma^2 \cdot \Delta t|x)$, $\pi_2(d(s) \cdot \Delta t|x)$, $\pi_2(\sigma^2 \cdot \Delta t|x)$ to calculate the value of three unknown coefficient in each equation in formula (18). Final, the posterior distributions after integration are obtained as the formulas shows:

$$\pi\left(d(s) \cdot \Delta t \mid x\right) = 0.9408\pi_1\left(d(s) \cdot \Delta t \mid x\right) + 0.0592\pi_2\left(d(s) \cdot \Delta t \mid x\right) \tag{21}$$

$$\pi\left(\sigma^2 \cdot \Delta t \mid x\right) = 0.9054\pi_1\left(\sigma^2 \cdot \Delta t \mid x\right) + 0.0946\pi_2\left(\sigma^2 \cdot \Delta t \mid x\right) \tag{22}$$

From the formula (21) and formula (22), we can see that there is a heavier weight coefficient of HALT than HASS, which illustrating the degradation information of HALT is more relevant to the observe data of ADT in the simulation example.

4.3 *Evaluation of reliability*

Through the posterior distribution after integration, we can obtain the assessed value of key parameters in reliability function:

$$\hat{\theta} = E\left(\theta \mid x\right) = \xi_1 E_1\left(\theta \mid x\right) + \xi_2 E_2\left(\theta \mid x\right) \tag{23}$$

where, E represents the mean of the posterior distributions.

We assume the initial value of product performance $y_0 = 100$ and the failure threshold of performance degradation $y_f = 50$. The evaluation

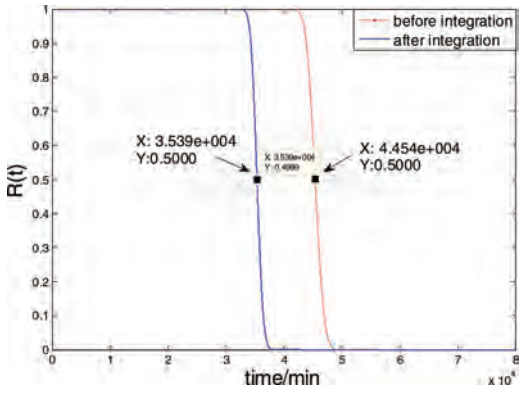


Figure 6. Evaluation of reliability before and after integration.

of reliability can be calculated by formula (5). To compare the different between evaluation results with information of HALT and HASS or not, the results of reliability before and after integration are showed in the Fig. 6.

The red line in Fig. 6 indicates the evaluation of reliability before integration, which is obtained from the information of ADT only. It represents the traditional ADT evaluation method. The blue line means the evaluation of reliability after integration, which integrated the information of HALT, HASS and ADT. It represents the evaluation method of MIHHA. From the Fig. 6 we can take the conclusion that the medium life is diminished after integration. We can hold the opinion that the evaluation result of MIHHA updates the evaluation result of ADT with information of HALT and HASS through Bayesian inference. Thus, the evaluation method of MIHHA contains more information than ADT; it can be a better indicator of reliability level of product.

5 CONCLUSION

Directing against the existing problem that ADT may lack degradation information for limited sample size or inadequate testing time, this article proposes an evaluation method of MIHHA. It solves the key issue of MIHHA that how to integrate HALT, HASS into ADT to predict reliability or lifetime of product. First, an ADT model based on Wiener process is proposed and the prior parameters are given. Then, based on the framework of Bayesian inference, the degradation information of HALT, HASS is regarded as prior information and the degradation information of ADT is considered as observed data. Through the prior information and observed data, we can obtain the posterior distribution which integrated the information of HALT, HASS and ADT. Third, the degradation information of HALT and HASS are integrated into one prior distribution based on weighting fusion method and the correlation function is adopted to calculate the weight coefficient of HALT and HASS, which solving the problem that the prior information comes from two sources. Finally, a simulation example is given to prove the feasibility of the evaluation method of MIHHA.

The evaluation method of MIHHA integrates degradation information of HALT, HASS and ADT. It provides a new path for evaluation method of product which makes full use of the information of HALT and HASS. This method would be a great help to the promotion and application of MIHHA.

ACKNOWLEDGEMENT

This study was supported by the National Natural Science Foundation of China (Grant No. 61703391) and Technology and Engineering Center for Space Utilization (Grant No. CSU-QZKT201714).

REFERENCES

Alferink, S. M. (2012). *Lifetime prediction and confidence bounds in accelerated degradation testing for lognormal response distributions with an arrhenius rate relationship.* Dissertations & Theses—Gradworks.

Box, G. E. P., & Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis.* Bayesian inference in statistical analysis /. Wiley.

Ge, Z.-Z., Jiang, T.-M., Han, S.-H. & Li. X.-Y. (2012). Design of accelerated degradation testing with multiple stresses based on d optimality. *Systems Engineering & Electronics,* 34(4), 846–853.

Gray, K. A., & Paschkewitz, J. J. (2016). *Next generation halt and hass.*

Liu, K. et al. (2016). Research on Integrated Accelerated testing and Reliability Assessment method for Space Electronic Products. *IEEE International Conference on Industrial Engineering and Engineering Management*, v 2016-December, p 1651–1654.

Lu, J. (1995). Degradation processes and related reliability models.

Nelson, W. (1971). Analysis of accelerated life test data—part i: the arrhenius model and graphical methods. *Electrical Insulation IEEE Transactions* on, EI-6(4), 165–181.

Wang, L., Pan, R., Li, X., & Jiang, T. (2013). A bayesian reliability evaluation method with integrated accelerated degradation testing and field information. *Reliability Engineering & System Safety,* 112, 38–47.

Wang, H. W., Xu, T. X., & Wang, W. Y. (2016). Remaining life prediction based on wiener processes with adt prior information. *Quality & Reliability Engineering International*, 32(3), 753–765.

Whitmore, G. A., & Schenkelberg, F. (1997). Modelling accelerated degradation data using wiener diffusion with a time scale transformation. *Lifetime Data Analysis,* 3(1), 27–45.

Zou, T. J., Li, X. Y., & Li Mei—Jun. (2015). Impact analysis of prior distributions on adt bayesian optimization design based on dic. 1–6.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Structural damage detection by integrating short time fourier transform, principal component analysis and logistic regression

Anand Kumar Agrawal & G. Chakraborty
Systems, Dynamics and Control Laboratory, Department of Mechanical Engineering, Indian Institute of Technology-Kharagpur, India

ABSTRACT: The aim of this work is to present a novel methodology for damage detection in a structure. In this method, new features based on Short Time Fourier Transform are used, dimensionality reduction was carried out by Principal Component Analysis and the classification was performed using Logistic regression. The efficacy of this method is demonstrated by detecting the presence of damage in a cantilever beam using the features based on transformation of displacement response.

Even though the difference between the displacement waveforms from damaged and the undamaged beam is not clearly perceptible, the difference is clearly visible in the principal components' space. Data belonging to damaged and undamaged classes are linearly separable up to a certain level of noise after which linear separability is lost. A linear decision boundary was obtained corresponding to a particular noise level after the mean normalization and feature scaling of the data. Feasibility of dimensionality reduction is ensured by checking the loss percent in the reduction process. And generalization ability of the classifier has been assessed on some test sets.

1 INTRODUCTION

Structural Health Monitoring (SHM) has been studied as a potential way to improve safety and reliability of various mechanical, aerospace and civil infrastructures. Current SHM techniques process dynamic responses of a structure in order to detect, localize or find out the extent of damage (Farrar & Worden 2013). There are two approaches for SHM i) The Model based and ii) The Data driven approach. In model based approach, the system is modeled based on its physics using a suitable modeling technique. Further the updating of this model is carried out using the field data. During this updating, initial model is tuned in a way that it produces outputs which match well with the real data. Eventually after updating stage a reasonably accurate model of the healthy system is obtained. Now the characteristics of the damage will be indicated in the form of change in parameter values on further updating using data from the monitoring stage. On the other hand in data driven approach instead of using a physics based model, the data obtained from the system is used to develop a statistical model for classification to achieve diagnosis. Major steps involved in the data driven approach are 1) Measurement of dynamic response 2) Feature extraction 3) Dimensionality reduction 4) Classification and 5) Prognosis. They are followed in sequence as shown in Figure 1.

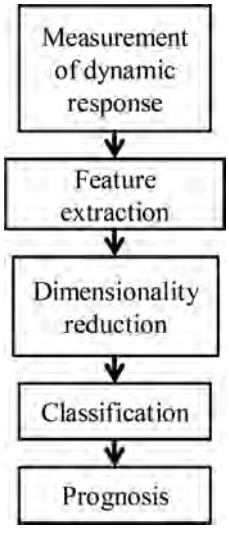


Figure 1. Major steps in SHM.

It is expected that accurate damage detection will require successful feature extraction, dimensionality reduction and signal classification. Appropriate selection of features and classification algorithm for an application depends on the nature of the dataset at hand. No single classification method can be preferred over others to

achieve good generalization. (Duda et al. 2001). Similarly by *Ugly duckling theorem*, no features can be called as best features irrespective of the application (Duda et al. 2001). Consequently many different classification methods and feature sets are explored in the SHM literature to suit a particular application. For example (Zhou et al. 2013) proposed a methodology for damage detection based on integrating data fusion and random forests. They used energy features extracted using wavelet packet transform for training. (He & Yan 2007) suggested a method where the classification was carried out using a wavelet based support vector machine and the feature extraction was done from wavelet energy spectrum constructed using wavelet packet transform. (Satpal et al 2016) have used support vector regression for localization of the damage using displacement values corresponding to first mode shape as the features. Although new methodologies are continually being proposed for damage detection still there are many to be explored. Damage detection accuracy can depend not only on the individual techniques used for feature extraction and classification but also on their various combinations.

The aim of this work is to present a novel methodology for damage detection in a structure. In this method, new features based on Short Time Fourier Transform (STFT) are used, dimensionality reduction is carried out by Principal Component Analysis and the classification is performed using Logistic regression. The efficacy of this method is demonstrated by detecting the presence of damage in a cantilever beam using the features based on transformation of displacement response obtained from a model of the beam.

Even though the difference between the displacement waveforms from damaged and the undamaged beam is not clearly perceptible, the difference is clearly visible in the principal components' space. Data belonging to damaged and undamaged classes are linearly separable up to a certain level of noise after which linear separability is lost. A linear decision boundary has been obtained for a particular noise level after the mean normalization and feature scaling of the data. Feasibility of dimensionality reduction is ensured by checking the loss percent in the reduction process. And generalization ability of the classifier has been assessed on some test sets.

2 METHODOLOGY

In this section the mathematical formulation of each step taken toward damage detection is provided.

2.1 *Data generation*

The purpose of this step is to acquire multiple signals (displacement time history) from a damaged and an undamaged beam. Here damaged and undamaged conditions are two different classes for which the classifier has to be designed. In a data driven approach the data are obtained either by carrying out experiments or with a reliable model. In this case we model the damaged beam based on Euler-Bernoulli hypothesis. The task of producing multiple signals has been accomplished as follows. Firstly the variational formulation of the damaged beam (shown in Figure 2) was done using Hamilton's principle and application of Ritz method to obtain the approximate response/signal. Subsequently the above signal was mixed with noise to obtain multiple signals corresponding to damaged beam with different damage locations. To get multiple signals for undamaged beam the damage size was substituted to be zero and the corresponding signal was mixed with noise. The mathematical formulation of this process is described below.

$$\text{By Hamilton's Principle: } \delta\int_{t_1}^{t_2}\left[K-U+W\right]dt=0 \qquad (1)$$

where, Kinetic energy, $K=\frac{\rho}{2}A\int_{0}^{L_1+\Delta L+L_2}\left(\frac{\partial w}{\partial t}\right)^2 dx-\frac{\rho}{2}b\delta\int_{L_1}^{L_1+\Delta L}\left(\frac{\partial w}{\partial t}\right)^2 dx$ Potential energy, $U=\frac{EI}{2}\int_{0}^{L_1+\Delta L+L_2}\left(\frac{\partial^2 w}{\partial x^2}\right)^2 dx-\frac{E\Delta I}{2}\int_{L_1}^{L_1+\Delta L}\left(\frac{\partial^2 w}{\partial x^2}\right)^2 dx$ Workdone by external forces, $W=f(t)w\left(x=L_1+\Delta L+L_2\right)$ where, ρ = desity, E = Young's modulus, I = moment of inertia

$f(t)$ = exitation force

w = deflection of the centreline of the beam

A = area of cross section

x is the direction along the length with its zero at the fixed end

To get the response using Ritz method we assume

$$w(x,t)=\{H\}^T\{p\}$$

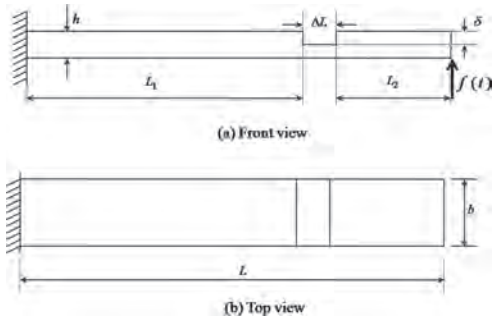


Figure 2. Beam geometry.

where, $\{H\}=\left[H_1(x),H_2(x)...H_N(x)\right]^T$ consisting of N linearly independent admissible functions. Such that $H_i(x)=\left(\frac{x}{L}\right)^2\left(1-\frac{x}{L}\right)^{i-1}, i=1,2,...,N$ & $\{p\}=\left[p_1(t),p_2(t),...p_N(t)\right]^T$ here $p_1(t),p_2(t),...p_N(t)$ are unknown coordinate functions.

The following equation is finally obtained

$$\left[\left[M_1\right]-\left[M_2\right]\right]\{\ddot{p}\}+\left[\left[K_1\right]-\left[K_2\right]\right]\{p\}=f(t)\{H\}_{x=L_1+\Delta L+L_2} \quad (2)$$

where,

$$\left[M_1\right]\triangleq\int_0^{L_1+\Delta L+L_2}\rho A\{H\}\{H\}^T dx,\left[M_2\right]\triangleq\int_{L_1}^{L_1+\Delta L}\rho b\delta\{H\}\{H\}^T dx$$

$$\left[K_1\right]\triangleq\int_0^{L_1+\Delta L+L_2}EI\{H\}''\{H\}''^T dx,\left[K_2\right]\triangleq\int_{L_1}^{L_1+\Delta L}E\Delta I\{H\}''\{H\}''^T dx$$

The above equations are subsequently solved.

The output from the sensor placed at $x = x_0$, is $w(x_0,n)=\{H\}^T_{x=x_0}\{p\}$, which is the required signal when the damage is at a location L_1 from the fixed end as shown in Fig. 2. By changing the location L_1 one can get signals corresponding to different damage locations. But, when these signals are acquired experimentally, the signal is bound to get distorted by noise. To simulate the real experiment the noise is added to the original signal giving multiple signals corresponding to each damage location. Similarly, one can get multiple signals from an undamaged beam as well. The process of mixing noise to the deterministic signal is expressed mathematically as

$$w_{\text{new}}(x_0,n)=\Psi(w(x_0,n),SNR)$$

where, Ψ is a function which adds noise to $w(x_0, n)$ with signal to noise ratio being *SNR*.

2.2 *Feature extraction*

Step 1: The set of time domain signals which one gets from data generation step are first transformed to a time-frequency domain signal by using the discrete Short Time Fourier transform (STFT) as shown in Equation 3.

$$S[m,\xi]=\sum_{n=0}^{N_s-1}w_{\text{new}}(x_0,n)\varphi(n-m)e^{\frac{-j2\pi\xi n}{N_s}} \quad (3)$$

where, $\varphi(n)$ is the window function given as:

$$\varphi(n)=\frac{I_0\left(\pi\alpha\sqrt{1-\left(\frac{2n}{\omega-1}-1\right)^2}\right)}{I_0(\pi\alpha)},\ 0\le n\le\omega$$

where, I_0 is the zeroth-order modified Bessel function of the first kind, ω is the width of the window and α is an arbitrary non negative real number that determines the shape of the window.

Step 2: In this step one forms a matrix [*P*] given in equation (4) using the outputs of the STFT given by Equation (3) in this matrix time increases across the columns of [*P*] and frequency increases down the rows, starting from zero.

First, two vectors namely the frequency and time vectors are defined as following:

$$\{F\}=\left[\xi_1,\xi_2,...\xi_{q_r}\right]^T,\ \left(\xi_i<\xi_{i+1}\right)$$

$$\{T\}=\left[m_1,m_2,...m_{q_c}\right],\ \left(m_i<m_{i+1}\right)$$

The entries of [*P*] denoted by $P(m_i, \xi_j)$ is defined as:

$$P\left(m_i,\xi_j\right)=\varsigma\left|S\left(m_i,\xi_j\right)\right|^2$$

where, $\varsigma=\frac{2}{f_s\sum_{n=1}^{\omega}|\varphi(n)|^2}$ when $\xi_i \neq 0$, Nyquist frequency $\varsigma=\frac{1}{f_s\sum_{n=1}^{\omega}|\varphi(n)|^2}$ otherwise

The time and frequency used in equation (3) are taken from $\{T\}$ and $\{F\}$ respectively. Finally the matrix [*P*] looks like

$$[P]=\begin{bmatrix} P(m_1,\xi_1) & P(m_2,\xi_1) & \cdots & P(m_{q_c},\xi_1)\\ P(m_1,\xi_2) & P(m_2,\xi_2) & \cdots & P(m_{q_c},\xi_2)\\ \vdots & \vdots & \cdots & \vdots\\ P(m_1,\xi_{q_r}) & P(m_2,\xi_r) & \cdots & P(m_{q_c},\xi_{q_r}) \end{bmatrix} \quad (4)$$

This matrix has been computed for each of the signals in the data generation step.

Step 3: Subsequently the eigenvalues of $[P]^T[P]$ are calculated and used as initial features. Therefore the data matrix [*X*] is given as:

$$[X]^T=\left[\{x\}_1\ \ \{x\}_2\ \ \cdots\ \ \{x\}_{n_s}\right] \quad (5)$$

where, $\{x\}_k = \left[f_1, f_2, ..., f_{q_c} \right]^T$ and *f, s* are the eigenvalues corresponding to the matrix $[P]^T [P]$ constructed from k^{th} signal and n_s is the number of signals used in training. In other words the rows of the data matrix [*X*] represent the feature vectors corresponding to each signal used in the training.

2.3 *Dimensionality reduction using principal component analysis*

Principal Component Analysis (PCA) attempts to project the data obtained in the feature extraction step into a lower dimensional vector space maintaining the maximum variance possible. The mathematical procedure to obtain such a transformation is as follows.

Step 1: Given the data matrix: [*X*] the variance-covariance matrix can be calculated as:

$$[S] = \begin{bmatrix} \mathrm{var}(F_1F_1) & \mathrm{cov}(F_1F_2) & \cdots & \mathrm{cov}(F_1F_{q_c}) \\ \mathrm{cov}(F_2F_1) & \mathrm{var}(F_2F_2) & \cdots & \mathrm{cov}(F_2F_{q_c}) \\ \vdots & \vdots & \cdots & \vdots \\ \mathrm{cov}(F_{q_c}F_1) & \mathrm{cov}(F_{q_c}F_2) & \cdots & \mathrm{cov}(F_{q_c}F_{q_c}) \end{bmatrix}$$

where, $F_i = [X_{1i}, X_{2i}, X_{3i}, ...X_{n_s i}]^T$ such that X_{ij} is an element of data matrix at i^{th} row and j^{th} column, $\mathrm{cov}(F_iF_j) = \frac{1}{n_s - 1} \sum_{k=1}^{n_s} (F_{ik} - \mu_i)(F_{jk} - \mu_j)$, and $\mathrm{var}(FF) = \frac{1}{n_s - 1} \sum_{k=1}^{n_s} |(F_{ik} - \mu)|^2$

Step 2: Now eigenvalues and eigenvectors of matrix [*S*] are calculated by solving the eigenvalue problem as given in Equation 6.

$$[S]\{v\}_i = \lambda_i \{v\}_i \tag{6}$$

Step 3: Subsequently λs are arranged in descending order as $\lambda_1 \rangle \lambda_2 \rangle \lambda_3 .... \lambda_{q_c}$

Step 4: For dimensionality reduction the first γ Eigen vectors corresponding to the Eigen values which are arranged in descending order were chosen. These Eigen vectors are stacked into a matrix [T] such that

$$[T] = \left[\{v\}_1 \{v\}_2 \{v\}_3 .... \{v\}_\gamma \right] \quad \& \quad \frac{\sum_{i=1}^{\gamma} \lambda_i}{\sum_{i=1}^{n_c} \lambda_i} \geq 0.99$$

The quantity $\left(1 - \frac{\sum_{i=1}^{\gamma} \lambda_i}{\sum_{i=1}^{n_c} \lambda_i} \right) * 100$ is referred as loss percentage.

Step 5: Finally the new data matrix corresponding to the new vector space can be obtained as:

$$[X]_{\text{reduced}} = [T]^T [X]^T \tag{7}$$

Each column of the new data matrix denotes the projection of a point in $\mathbb{R}^{q_c}$ on to the $\mathbb{R}^{\gamma}$. Hence, one gets all the q_c dimensional vectors converted to γ dimensional vectors where $\gamma < q_c$ The proof that the matrix $[T]^T$ indeed projects the data on to a subspace which maximizes the variance can be found in (Bishop 2006). For visualization of the data, γ must be ≤ 3.

2.4 *Classification using logistic regression*

Reduced dimensional feature vectors obtained after the PCA are first mean normalized and scaled and then augmented with unity to create the augmented feature vectors. These augmented feature vectors along with the corresponding class labels are used as training data. Logistic regression uses a hypothesis function given below (Ng 2012). The unknown parameter vector $\{\theta\}$ in the hypothesis function is estimated by optimizing the cost function given in equation (9) using a MATLAB function 'fminunc'. The value of the hypothesis function calculated for a feature vector can be treated as the probability of it belonging to the damaged beam.

Training set:

$$\left\{ \left(\{\chi\}^{(1)}, y^{(1)} \right), \left(\{\chi\}^{(2)}, y^{(2)} \right), \ldots \left(\{\chi\}^{(n_s)}, y^{(n_s)} \right) \right\}$$

Hypothesis function,

$$h_\theta(\{\chi\}) = \frac{1}{1 + e^{-\{\theta\}^T \{\chi\}}} \tag{8}$$

Cost function J($\{\theta\}$) is given as following:

$$J(\{\theta\}) = -\frac{1}{m} \left[\sum_{i=1}^{m} y^{(i)} \log h_\theta \left(\{\chi\}^{(i)} \right) + \left(1 - y^{(i)} \right) \log \left(1 - h_\theta \left(\{\chi\}^{(i)} \right) \right) \right] \tag{9}$$

where, $\{\chi\}^{(i)} = \begin{Bmatrix} 1 \\ c_i \end{Bmatrix}$

such that $\mathbf{c}_i = \begin{Bmatrix} \frac{PC_1^i - meanPC_1}{\sigma_{pc1}} \\ \frac{PC_2^i - meanPC_2}{\sigma_{pc2}} \end{Bmatrix}$

PC_1^i and PC_2^i are the values of 1st and 2nd principal components corresponding to column of matrix $[X]_{reduced}$, mean PC_1 and mean PC_2 are the means of 1st and 2nd principal components, while

σ_{pc1} and σ_{pc2} are the corresponding standard deviations and $y \in \{0,1\}$ n_s is the number of signals used in training, $\{\theta\}$ is the vector of parameters.

Assessment of the classifier was done by creating a test set following the same procedure as in the case of training data, except that during test set creation the noise added to the signals were altered. The accuracy obtained for various test set is tabulated in Table 2. The classifier thus made from the training data can be used for classifying the signal from the real structure.

3 RESULTS AND DISCUSSION

The method explained in the previous sections is now applied to a cantilever beam of the following specifications.

$$L = 45\ cm,\ b = 5.2\ cm,\ h = 0.5\ cm,\ \delta = 0.2\ cm$$

$$L_1 = (5\ cm,\ 10\ cm,\ 15\ cm,\ 20\ cm,\ 25\ \text{cm},\ 30\ cm)$$

$$\Delta L = 0.2\ cm,\ L_2 = L - \Delta L - L_1,\ E = 200\ GPa$$

$\rho = 7850\ kg/m^3$ where the geometric parameters are shown in Figure 2. For simulation of equation (2) the initial conditions were set to zero and the forcing function was a rectangular pulse of small width as shown in Figure 3. The displacement responses for different damaged and undamaged cases are shown in Figure 4. To verify the correctness of the above results the damage size was substituted to be zero in the mathematical model for the damaged beam and the first four natural frequencies obtained through Fast Fourier Transform (FFT) of the response with the ones obtained by analytical formula and ANSYS were compared. The comparison is shown in Table 1. It is clear from Table 1 that all natural frequencies calculated through FFT of the response match reasonably well with the natural frequencies obtained by the analytical formula. On the other hand when compared with the natural frequencies obtained by ANSYS the match is fairly good for the first two natural frequencies but there is a big difference between the third and fourth natural frequencies. The reason for above mismatch for higher frequencies can be attributed to the following reason. In ANSYS the field problem is solved over the entire domain without making any special assumptions as in Euler and Bernoulli's beam theory. It is known that at high frequencies more realistic assumptions like the inclusion of rotary inertia (Rayleigh beam theory) and shear deformation (Timoshenko beam theory) are to be incorporated (Shames 1985). The predicted results of Euler-Bernoulli beam overestimate the correct values. Looking at the displacement signals corresponding to various damage conditions it is almost impossible to figure out any difference between them. Time frequency plots given in Figure 5 do reveal some differences present between the signals. But after the feature extraction and dimensionality reduction by the STFT and PCA respectively a good separation can be seen in the final 2D feature space (principal components' space) as shown in Figure 6.The loss% during PCA is of the order of 10^{-6} which indicates that the data obtained after the feature extraction step has most variance in the first two principal directions. Another interesting thing can be noted in Figure 6 is that there are seven clusters which are

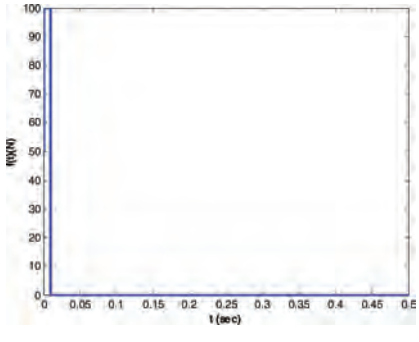


Figure 3. Excitation force.

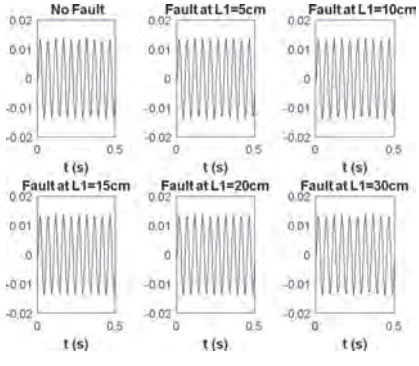


Figure 4. Displacement response for different damage locations.

Table 1. Comparison of natural frequencies.

Natural frequency number	From FFT of the MATLAB response	Analytical	(FEM) ANSYS
1*st*	20 Hz	20.1329 Hz	20.131 Hz
2*nd*	126 Hz	126.17 Hz	126.08 Hz
3*rd*	354 Hz	353.28 Hz	207.21 Hz
4*th*	694 Hz	692 Hz	328.4 Hz

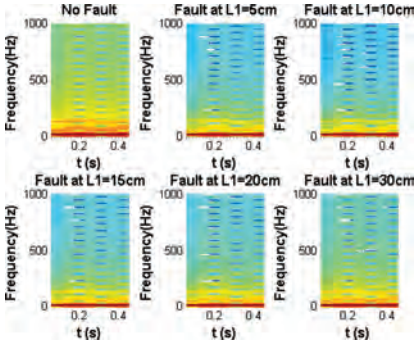


Figure 5. Spectrograms for different damage locations.

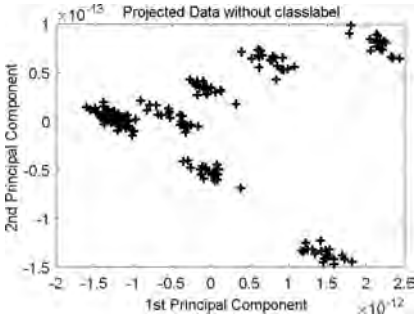


Figure 6. Data obtained after PCA.

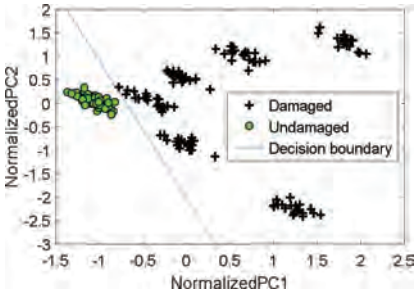


Figure 7. Decision boundary.

indicative of the nature of data set because the data set contains vectors corresponding to the six different damage locations and the undamaged system. Decision boundary obtained using logistic regression for training data obtained by mixing noise of $SNR = 91$ is shown in Figure 7. It can be noticed from Figure 7 that the accuracy obtained on the training set is 100%. Generalization assessment performed by calculating the accuracy achieved on different test sets in given in Table 2. It is clear that the accuracy achieved on test sets created with SNR ≥ 91 is 100% on the other hand the accuracy is 98.89% with the SNR = 90 in the test set. Hence the classifier is working quite well on the data outside the training set.

Table 2. Accuracy achieved for various test sets.

SNR	No. of test points	Accuracy
90	90	98.89%
91	90	100%
92	90	100%

4 CONCLUSION

The proposed structural damage detection methodology is based on integrating techniques used in the field of Signal processing and Machine learning in a novel way, but the combination of the methods and the features proposed are new. New features based on the Short Time Fourier Transform of the response signal have been successfully used to detect the presence of a damage in a beam. As these features are able to discriminate between the damaged and undamaged beams they can also be used with classifiers other than the one presented here. In addition, the method is also capable of capturing the characteristic of the data by creating different clusters in the principal components' space for different damage location indicating the possibility of damage localization as well. The assessment of the classifier showed the accuracy of 100% on the test sets with $SNR \geq 91$ and 98.89% for $SNR = 90$.

REFERENCES

Bishop, Christopher M. 2006. *Pattern_Recognition. Springer Science + Business Media, LLC.*

Duda, Richard O., Peter E. Hart, and David G. Stork. 2001. "Pattern Classification." *New York: John Wiley, Section*. doi:10.1007/BF01237942.

Farrar, Charles R., and Keith Worden. 2013. *Structural Health Monitoring: A Machine Learning Perspective*. doi:10.1177/1475921708090560.

He, Hao-Xiang, and Wei-ming Yan. 2007. "Structural Damage Detection with Wavelet Support Vector Machine: Introduction and Applications." *Structural Control and Health Monitoring* 14(1):162–76. doi:10.1002/stc.150.

Ng, Andrew. 2012. "1. Supervised Learning." *Machine Learning*, 1–30. doi:10.1111/j.1466-8238.2009.00506.x.

Satpal, S.B, A Guha, and S Banerjee. 2016. "Damage Identification in Aluminum Beams Using Support Vector Machine: Numerical and Experimental Studies." *Structural Control and Health Monitoring* 23 (3): 446–57. doi:10.1002/stc.1773.

Shames, Irving H. 1985. *Energy and Finite Element Methods in Structural Mechanics*. CRC Press.

Zhou, Q., Y. Ning, Q. Zhou, L. Luo, and J. Lei. 2013. "Structural Damage Detection Method Based on Random Forests and Data Fusion." *Structural Health Monitoring* 12 (1): 48–58. doi:10.1177/1475921712464572.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Fault diagnosis and remaining useful life prediction of multiple deteriorating components in hybrid dynamical system

Om Prakash, A.K. Samantaray & R. Bhattacharyya
Systems, Dynamics and Control Laboratory, Department of Mechanical Engineering, Indian Institute of Technology, Kharagpur, India

ABSTRACT: A Hybrid Dynamical System (HDS) includes a set of continuous dynamics in which a particular continuous dynamics is activated at a particular set of discrete events, termed as mode of the system. Thus, a different mode triggers a different continuous dynamics. Degradation evolutions of the components of HDS depend on the operating mode of the system. Thus, the existing approaches for continuous systems are not suited for Remaining Useful Life (RUL) prediction for HDS. In addition, a discrete mode fault may be possible besides the parametric faults (abrupt or progressive nature). This article presents an integrated approach to Fault Diagnosis (FD) and RUL prediction of multiple deteriorating components in an HDS. For improving FD scheme, dynamic fault signature matrices are utilized for parametric and discrete mode fault isolation, which minimize the possible suspected faults by using the possible deviation direction of the faults. If the detected fault is progressive in nature, then the FD scheme is further utilized to point out the severity change points of the degradation. Using the knowledge of each severity change points and the deviation direction of progressive fault, constrained parameter estimation method with dynamically updated parameter's bound is proposed for fast degradation states estimation. The estimates of the degradation at different time instances in a respective operating mode are further utilized for mode-dependent degradation model identification and RUL prediction. An online degradation model selection scheme is proposed for degradation model identification in different operating modes. The proposed method is able to identify the degradation model of multiple degrading components in a real time at different operating modes and can be adapted with new information of their degradation states estimated by the constrained parameter estimation during continuous monitoring. The proposed approach is demonstrated through numerical simulation of an example hybrid dynamical system.

1 INTRODUCTION

Nowadays, due to advances in technology and modern control systems, most of the process engineering systems, like drinking water systems, chemical process engineering systems, etc., can be best modeled as hybrid dynamical systems (HDS). Fault diagnosis (FD) and remaining useful life (RUL) prediction of HDS are complicated because the dynamical behaviour of such system is governed by a specific combination of discrete modes (autonomous mode or supervisory controller mode) (Narasimhan & Biswas, 2007; Wang et al. 2013; Borutzky, 2015). These systems show different continuous dynamics in different modes. Most of the existing FD and RUL prediction approaches (Medjaher & Zerhouni, 2009; Jha et al., 2016) are typically developed for continuous dynamical systems and assume the occurrence of single fault with constant rate of degradation throughout the components life cycle. However, many faults may be possible during the continuous operation of a system and RULs of all such faulty components must be provided to the plant engineers for the decision making and maintenance scheduling. Generally, component's degradation is a slow process where the degradation severity level changes in the time units of hours, days, weeks, months, or even years depending on the type of the system, its dynamics and the environmental conditions. These severity change time points for different components may be different, i.e., different components degrade at different rates. The existing approaches may not be easily applied to HDS since the degradation progress in the faulty components depend on system operating mode and the existing approaches may fail to diagnose the actual faulty components due to single fault assumption. Thus, the switching behavior of the system and possibility of occurrence of multiple faults make the FD and RUL prediction tasks very challenging for HDS, especially considering the

fact that some fault symptoms may be masked or compensated by pre-existing fault symptoms. Also, a discrete mode fault may be possible besides the parametric faults (abrupt or progressive nature). Examples of discrete mode faults are valve stuck on/off fault, mode transition failure, control command communication fault, etc. Thus, the existing approaches based on single fault assumption may provide unreliable results and lead to misdiagnosis. Therefore, there is a need for development of more effective and efficient techniques for FD and RUL prediction of HDS and it is also practical to unify both in a common methodology for health monitoring of the system. Also, degradation of a component is usually irreversible and the associated parameter value deviates monotonically in a certain direction (either increasing or decreasing parameter value). This information of parameter deviation direction can be used for improving parameter estimation algorithms through specification of appropriate constraints.

Recently, many works have been published using hybrid bond graph (HBG) technique with applications to modeling, control and FD of HDS (Narasimhan & Biswas, 2007; Wang et al. 2013; Borutzky, 2015; Low et al., 2010; Ghoshal et al., 2012; Levy et al., 2015). But, very limited works (Yu et al., 2015; Daigal et al., 2015) deal with RUL prediction along with FD of HDS. In (Yu et al., 2015; Daigal et al., 2015), HBG technique is utilized for system modelling and FD; however particle filtering/Monte Carlo technique is applied for identification of degradation rate. According to available literatures for HDS, no work is found related to RUL prediction of sequentially occurring multiple faults, where new fault's symptoms may be masked or compensated by the already existing degradations in other components.

In summary, paper proposes an integrated approach for real time RUL prediction of multiple degrading components in an HDS using HBG as a common framework. Constrained parameter estimation (CPE) technique with dynamically updated constraints supported by the information of parameter drift direction is proposed to speed up the degradation pattern identification and RUL prediction. The proposed method accommodates the influences of different operating modes and is adapted with new information of degradation states identified through continuous monitoring.

The remaining article is organized as follows. Section 2, presents the common terminology and methodology used for health monitoring of HDS using HBG approach. Section 3 presents the FD and RUL prediction of multiple deteriorating components. Section 4 shows the efficacy of the proposed method using simulation and Section 5 gives conclusions and perspectives.

2 HBG-BASED DIAGNOSIS METHODOLOGY

2.1 *Hybrid dynamical system*

A benchmark hybrid two-tank example system is shown in Figure 1, which is considered as the application example in this paper. This system includes tanks, valves, pipes, and PI-controlled pump flow (Q_P). Valve V_1 is on-off type and operates at different state (on or off state) according to the supervisory controller command (a_{v1}). Two drainage pipes L_1 and L_2 belong to two autonomous modes transitions (a_1 and a_2, respectively) of the system, which depend on the internal dynamics of the system, i.e. state of liquid level (H_1, H_2) of tanks T_1, T_2, respectively. When the level of liquid in T_1 exceeds the level H_{L1} then the autonomous mode a_1 is switched to its active mode and liquid starts flowing from T_1 to T_2 through pipe L_1. Likewise, when the level of liquid in T_2 exceeds the level H_{L2} then another autonomous mode a_2 is switched to its active mode and the liquid starts flowing from T_2 to atmosphere through pipe L_2. These changing modes complicate the FD and RUL prediction process. For example, performance of an on-off valve V_1 may degrade due fouling (buildup of sediment/lime-scale, etc.). But, the deterioration rate of V_1 is not always constant, as it depends on the liquid flow through this unit. During off command of V_1, there is no buildup of fouling as no liquid flows through this unit. Only in on command of V_1, fouling is possible and may increase at some unknown rate.

Also, two imaginary valves V_{Leak1} and V_{Leak2}, respectively, are modeled to simulate leakage faults in tanks T_1 and T_2. The pump saturation characteristic (Φ_P) and output law of PI-controller (Φ_{PI}) are, respectively, stated as

$$Q_P = \begin{cases} U_{PI}, & \text{if } 0 \le U_{PI} \le f_{max} \\ 0, & \text{if } U_{PI} \le 0 \\ f_{max}, & \text{if } U_{PI} > f_{max} \end{cases} = \Phi_P(U_{PI}) \qquad (1)$$

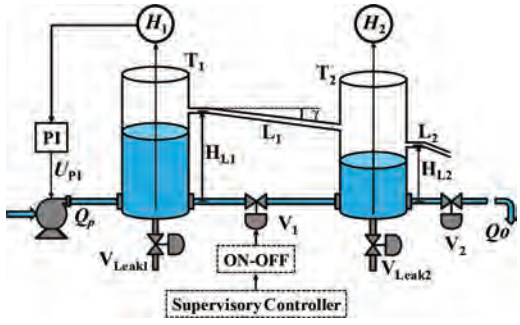


Figure 1. A schematic of hybrid two-tank system.

$$U_{\mathrm{PI}} = K_{\mathrm{P}}(S_{\mathrm{pt}} - \rho.g.H_1(t)) + K_{\mathrm{I}} \int (S_{\mathrm{pt}} - \rho.g.H_1(t))\,\mathrm{d}t = \Phi_{\mathrm{PI}}\left(H_1(t)\right) \tag{2}$$

where U_{PI} is PI-controller output, $f_{\max}$ is the maximum pump flow, S_{pt} is a controller set point, K_{P} is proportional gain and K_{I} is integral gain, g is acceleration due to gravity, ρ is liquid density.

2.2 *Diagnostic Hybrid Bond Graph (DHBG) model*

In bond graph, generalized elements, i.e., *Se*-effort source, *Sf* – flow source, *I* – inertia, *C* – compliance, *R*-resistor, 0 – equal effort junction, 1 – equal flow junction, *TF* – transformer, *GY* – gyrator, *De* – effort detector and *Df* – flow detector, are used to model the multi-physics system in a unified way (Samataray et al., 2008). The diagnostic BG technique is proposed for diagnosis of continuous system as in Samataray et al., 2006, Samataray et al., 2008. This technique is further extended for hybrid system, called DHBG technique (Low et al., 2010), which is well-suited for diagnosis of HDS by means of controlled/switched junctions and by feeding the measurements and mode information to the DHBG model. Also for the uncertain system, the nominal parameters are decoupled from their uncertain parts to account for the parameter uncertainties and modeled in linear fractional transformation (LFT) form by using feedback loops of internal variables, which is called LFT-DHBG model (Djeziri et al., 2007; Merzouki et al., 2012). For example, a LFT-DHBG model of hybrid two-tank system is presented in Figure 2, where the two level sensors (H_1, H_2) are fed to the model and the respective parameter uncertainty is decoupled from its nominal value to account for the uncertainty. In Figure 2, 1 – junctions with subscript a_{v1}, a_1 and a_2 are switched junctions related with discrete modes.

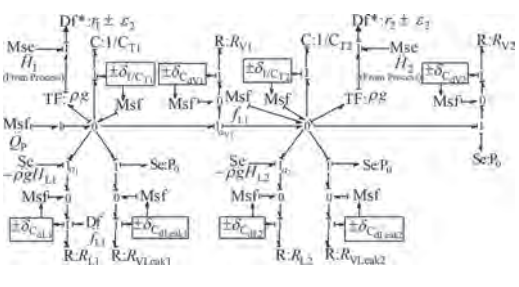


Figure 2. LFT-DHBG model of hybrid two-tank system.

2.3 *ARRs/GARRs and adaptive thresholds*

The analytical redundancy relations (ARRs) or global ARRs (GARRs) represent the conservation equations like continuity or mass balance equations, energy equations, momentum equations, etc. for the system. These relations are true at all times and at any working mode of the system. These relations do not depend on the past history of events. Evaluations of these relations provide the residuals which are ideally zero during fault free operation of system. But, due to measurements and parameters uncertainties, practically, residuals show small non-zero values and hence they are bounded with some time varying thresholds, called adaptive thresholds. The ARRs/GARRs and adaptive thresholds equations can be systematically derived from LFT-DHBG model, whose general form is

$$GARR_{ni}\left(Z,\theta,U,Y\right) \pm (\lambda_i + \lambda_{Si}) = 0 \tag{3}$$

where GARR_{ni}, λ_i and λ_{Si} represent the nominal residual (r_i) ($i = 1, 2, \ldots, n$; n is number of residuals), uncertain part due to parameter uncertainties and small static uncertain part needed to account for measurement errors, respectively. The uncertain parts of residual provide the adaptive threshold as $\varepsilon_i = \pm\mathrm{abs}(\lambda_i + \lambda_{Si})$. Also, $Z = [a_1, a_2, \ldots, a_k, \ldots, a_m]^T$ signifies the switched-junction ideal mode vector of m discrete components, $a_k \in (0, 1)$, $\theta = [\theta_1, \theta_2, \ldots, \theta_j, \ldots, \theta_p]^T$ signifies a known ideal parameter vector of p components, U and Y are the measured input and output variables of the system.

The two imaginary flow detectors (Df^*) in the LFT-DHBG model of hybrid two-tank system (Fig. 2) provide the two GARR_i ($i = 1, 2$) and uncertain parts (λ_i). These GARRs are nothing but represent continuity equations of the system in any mode.

Since absolute values of uncertain parts contribution is added in adaptive threshold, the small λ_{Si} part can be neglected, thus, the GARRs can be written as

$$\begin{aligned}\mathrm{GARR}_1 : {} & Q_{\mathrm{P}} - C_{\mathrm{T1}}\frac{\mathrm{d}}{\mathrm{d}t}\rho g H_1(t) - a_{\mathrm{v1}}C_{\mathrm{dv1}}\sqrt{|\rho g(H_1(t) - H_2(t))|} \\ & .\mathrm{sign}(H_1(t) - H_2(t)) - a_1 C_{\mathrm{dL1}}\rho g(H_1(t) - H_{\mathrm{L1}}) \\ & - C_{\mathrm{dLeak1}}\sqrt{|\rho g H_1(t)|} \pm \lambda_1 = 0,\end{aligned} \tag{4}$$

$$\begin{aligned}\mathrm{GARR}_2 : {} & a_{\mathrm{v1}}C_{\mathrm{dv1}}\sqrt{|\rho g(H_1(t) - H_2(t))|}\ .\mathrm{sign}(H_1(t) \\ & - H_2(t)) - C_{\mathrm{T2}}\frac{\mathrm{d}}{\mathrm{d}t}(\rho g H_2(t)) + a_1 C_{\mathrm{dL1}}\rho g(H_1(t) \\ & - H_{\mathrm{L1}}) - a_2 C_{\mathrm{dL2}}\rho g(H_2(t) - H_{\mathrm{L2}}) \\ & - C_{\mathrm{dv2}}\sqrt{|\rho g H_2(t)|} - C_{\mathrm{dLeak2}}\sqrt{|\rho g H_2(t)|} \pm \lambda_2 = 0\end{aligned} \tag{5}$$

where $C_{Ti} = A_i/g$, A_i is tank cross-section area, and

$$a_i = \begin{cases} 0, \text{ if } H_i(t) \le H_{\mathrm{L}i} \\ 1, \text{ if } H_i(t) > H_{\mathrm{L}i} \end{cases}, \; i = 1, \; 2.$$

Using Equations (1)–(2), two more ARRs, respectively, for actuator and controller (assumed to have no uncertainty) are obtained as

$$\mathrm{ARR}_3\colon Q_{\mathrm{P}} - \Phi_{\mathrm{P}}(U_{\mathrm{PI}}) = 0 \tag{6}$$

$$\mathrm{ARR}_4\colon U_{\mathrm{PI}} - \Phi_{\mathrm{PI}}(H_1(t)) = 0 \tag{7}$$

The uncertain parts λ_i (i = 1, 2) presented in Equations (4) and (5), respectively, are given as

$$\begin{aligned} \lambda_1 = {} & \left|\delta_{\mathrm{C1}} C_{\mathrm{T1}} \frac{\mathrm{d}}{\mathrm{d}t} \rho g H_1(t)\right| + \\ & \left|a_{\mathrm{v1}} \delta_{C_{\mathrm{dv1}}} C_{\mathrm{dv1}} \sqrt{\left|\rho g (H_1(t) - H_2(t))\right|}\right| \\ & + \left|a_1 \delta_{C_{\mathrm{dL1}}} C_{\mathrm{dL1}} \rho g (H_1(t) - H_{\mathrm{L1}})\right| \\ & + \left|\delta_{C_{\mathrm{dLeak1}}} C_{\mathrm{dLeak1}} \sqrt{|\rho g H_1(t)|}\right| \end{aligned} \tag{8}$$

$$\begin{aligned} \lambda_2 = {} & \left|a_{\mathrm{v1}} \delta_{C_{\mathrm{dv1}}} C_{\mathrm{dv1}} \sqrt{\left|\rho g (H_1(t) - H_2(t))\right|}\right| + \\ & \left|\delta_{\mathrm{C2}} C_{\mathrm{T2}} \frac{\mathrm{d}}{\mathrm{d}t} \rho g H_2(t)\right| + \left|a_1 \delta_{C_{\mathrm{dL1}}} C_{\mathrm{dL1}} \rho g (H_1(t) - H_{\mathrm{L1}})\right| + \\ & \left|a_2 \delta_{C_{\mathrm{dL2}}} C_{\mathrm{dL2}} \rho g (H_2(t) - H_{\mathrm{L2}})\right| + \left|\delta_{C_{\mathrm{dv2}}} C_{\mathrm{dv2}} \sqrt{|\rho g H_2(t)|}\right| + \\ & \left|\delta_{C_{\mathrm{dLeak2}}} C_{\mathrm{dLeak2}} \sqrt{\rho g H_2(t)}\right| \end{aligned} \tag{9}$$

2.4 *Fault signature matrix and coherence vector*

For different faults (parametric or mode faults), the GARRs are utilized to create the fault signature matrix (FSM), which depends upon the sensitivities of GARRs to the parameter variations (Low et al., 2010). In this study, global fault sensitivity signature matrix (GFSSM) and mode change sensitivity signature matrix (MCSSM) (Levy et al., 2015) are utilized, which have a power to discriminate between increasing ($\theta_j\uparrow$) and decreasing ($\theta_j\downarrow$) variations in the parameters. These dynamic matrices (see Equations (10) and (11)) are the extended forms of GFSM/MCSM matrices (Low et al., 2010) used for diagnosis of HDS.

$$GFSSM_{ji} = \begin{cases} -\operatorname{sign}(\partial r_i / \partial \theta_j), & \text{if } r_i \text{ is sensitive to } \theta_j\uparrow \\ +\operatorname{sign}(\partial r_i / \partial \theta_j), & \text{if } r_i \text{ is sensitive to } \theta_j\downarrow \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

$$MCSSM_{ki} = \begin{cases} -\operatorname{sign}(\partial r_i / \partial a_k), & \text{if } r_i \text{ is sensitive to } a_k\uparrow \\ +\operatorname{sign}(\partial r_i / \partial a_k), & \text{if } r_i \text{ is sensitive to } a_k\downarrow \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

For example, the signature for the leakage fault in tank T_1, i.e. $C_{\mathrm{dLeak1}}\uparrow$, is as

$$\begin{aligned} \mathrm{GFSSM}^{(C_{\mathrm{dLeak1}}\uparrow)} = {} & \left[-\operatorname{sign}\left(\frac{\partial GARR_1}{\partial C_{\mathrm{dLeak1}}} \right), \right. \\ & \left. -\operatorname{sign}\left(\frac{\partial GARR_2}{\partial C_{\mathrm{dLeak1}}} \right) \right] \\ = {} & \left[+\operatorname{sign}\left(\sqrt{|\rho g H_1(t)|} \right), \; 0 \right] \end{aligned} \tag{12}$$

Likewise, signatures for all faults are derived.

A coherence vector (***CV***) is used to generate the alarm state (0 or 1), whose standard form is given as $\boldsymbol{CV} = [cv_1(t), cv_2(t),..., cv_n(t)]$, where $cv_i(t)$, i = 1, 2,..., n, are generated from decision procedure, $\Theta\left(r_i(\mathrm{t})\right)$. For robust FD, each residual $r_i(t)$ is tested against an adaptive threshold $\varepsilon_i(t)$ as follows:

$$cv_i(t) = \Theta\left(r_i(\mathrm{t})\right) = \begin{cases} 0, \text{ if } -\varepsilon_i(t) \le r_i(t) \le \varepsilon_i(t) \\ +1, \text{ if } r_i(t) \ge \varepsilon_i(t) \\ -1, \text{ otherwise} \end{cases} \tag{13}$$

Here, it is assumed that $\varepsilon_i(t) \simeq \lambda_i(t)$. During monitoring, ***CV*** is derived at every sampled data for generating the alarm state. An alarm state shows 1 if any abnormality is found in the system with $\boldsymbol{CV} \neq [0, 0,\ldots, 0]$. After an alarm, the ***CV*** is matched with the GFSSM/MCSSM for the isolation of real degrading component. If a unique fault signature matches with the obtained ***CV*** the fault is isolated. A detectable and isolable component fault is represented by detectability index $D_{\mathrm{b}} = 1$ and isolatability index, $I_{\mathrm{b}} = 1$, respectively, in the fault signature matrix.

3 FD AND RUL PREDICTION OF MULTIPLE COMPONENTS DEGRADATIONS

A schematic flow diagram of integrated FD with RUL prediction of multiple degrading components occurring in a sequential way is presented in Fig. 3. The GARRs and adaptive thresholds are numerically evaluated at each and every time instance. If any threshold violation occurs for any unknown fault (abrupt/progressive parametric fault or discrete mode fault) then the obtained instantaneous ***CV*** is checked with both MCSSM/GFSSM and generates the initial set of suspected faults (SSF). Inclusion of discrete mode fault in SSF complicates the fault isolation process, since the discrete mode fault in a component shares the same fault signature as the partial parametric fault associated with the same component. First, it is assumed that the threshold violation occurs due to

a discrete mode fault, as discrete mode fault has more severe impact on the system dynamics and its stability. This paper presents a new method to discriminate discrete mode fault from the parametric fault based on magnitude of residual deviation after a fault, which is identified from Equation (14) as

$$D_{a_k}^{\eta} = \left|GARR_i(Z,\ \theta,\ U,\ Y)\right| - \left|\frac{\partial GARR_i}{\partial a_k}\right| \le \varepsilon_i, \quad (14)$$

where $D_{a_k}^{\eta}$ is the difference between absolute value of residual deviation after a fault and sensitivity of residual with respect to suspected discrete parameter a_k in the initial SSF. If any discrete mode fault is found, then the DHBG model is updated with faulty mode for subsequent detection of fault (if monitoring is continued). After updating the DHBG model, all residuals are forced to remain bounded within the updated thresholds even in presence of one or more faults. If it is found that the discrete modes are con-sistent, then it indicates that the threshold violation is due to parametric fault. Then, the discrete mode faults are removed from the initial SSF, and the algorithm is triggered for parametric fault identification only for the suspected parameters which remain in refined SSF. Using the knowledge of deviation directions of suspected parameters obtained from GFSSM, the bounds of the suspected parameters, $\theta_F \in [\theta_{FL},\ \theta_{FU}]$ are created. Bounds are created based on previous known nominal values of the parameters (θ_F) and their possible extreme variations after the fault, derived from the deep knowledge of the system, called technological specifications. For the time varying parameter, the parameter bound is also updated when the true magnitude of varying parameter is obtained after successive estimation of fault at different times in the respective mode. A CPE technique with dynamically updated parameter bounds is proposed by integrating the gradient-projection method with Gauss-Newton method. To further improve the parameter estimation, sensitivity BG technique (Samantaray & Ghoshal, 2007) is utilized which provides the gradient information of the cost function during optimization process. The gradient projection method is more efficient, particularly, when constraints include only bounds on the parameters (Nocedal & Wright, 2006). The objective function is formulated as:

$$\min_{\theta} J(\theta) = \frac{1}{2}\sum_{j=k-q}^{k} \mathbf{r}^{\mathrm{T}}(t_j).\mathbf{W}.\mathbf{r}(t_j)$$
$$\text{subject to}: \ \theta_L \le \theta \le \theta_U \quad (15)$$

where $\mathbf{r}(t_j)$ is the residual vector obtained by evaluating the *GARRs* at time instant t_j, k is the current sample time, $q \ge 0$ is the number of collection of past sampled data during monitoring, and $\mathbf{W} \in \mathbb{R}^{n\times n}$ is a positive semi-definite weighing function and can be considered as unit matrix, θ_L and θ_U are, respectively, the sets of lower and upper parameter bounds on the parameters.

Thus, CPE finally isolates (say $\theta_j\downarrow$ at mode $\boldsymbol{Z}=z^{(i)}$) the actual deteriorating component (θ_j) at k^{th} time instant and gives the first estimate $\theta_j^{\mathrm{E}}(k, z^{(i)})$ (fault magnitude at k^{th} instant). The parameter vector is updated as $\theta = [\theta_1, \theta_2, ..., \theta_j^{\mathrm{E}}(k, z^{(i)}), ..., \theta_P]^T$ and considered as a new nominal parameter vec-

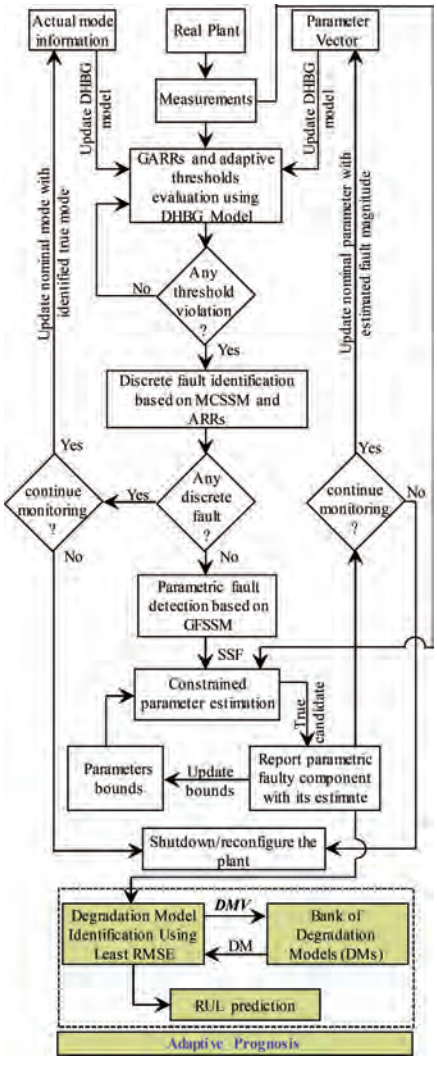


Figure 3. A schematic flow diagram of integrated approach.

tor at k^{th} instant. Also, initial bound $\theta_j \in [\theta_{jL}, \theta_{jU}]$ is updated with fault magnitude at k^{th} instant as $\theta_j \in [\theta_{jL}, \theta_j^f(k, z^{(i)})]$. The evaluation of residuals with updated parameters again forces the residuals to remain lie within the updated thresholds. If the identified degradation is progressive, the parameter $\theta_j(t, z^{(i)})$ varies slowly at mode $z^{(i)}$. Thus, the updated residuals again violate the same thresholds after some more time, and again CPE technique with the updated bound $\theta_j \in [\theta_{jL}, \theta_j^f(k, z^{(i)})]$ gives the second fault estimate $\theta_j^f(k+\Delta, z^{(i)})$ at $(k+\Delta)^{th}$ time instant. Parameter estimation method converges quickly as the search zone is reduced dynamically here. Likewise, more estimated data points are accumulated with dynamically updated bounds of the supervised system at a particular mode $z^{(i)}$. If the j^{th} mode switching occurs at $t^{(j)}$ time instant during sample data accumulation for CPE in the preceding mode $z^{(i)}$ then the fresh sampled data of a fixed time window at new mode $z^{(j)}$ are accumulated and deteriorating parameter, $\theta_j(t, z^{(j)})$ is estimated in this new mode $z^{(j)}$. This way, a set of estimated parameter values (degradation trend) is obtained at different time instants in different operating modes. Depending upon the number of such estimates, an interpolation curve is generated and that curve is extrapolated to obtain the RUL.

3.1 *Degradation model and RUL prediction*

Degradation estimates $D_{\theta_j}^{z^{(i)}} = \left(t_w, \theta_j^f\left(t_w, z^{(i)}\right)\right)$, $w = 1, 2, \ldots, k$, is further used for degradation model (DM) identification at mode $z^{(i)}$. Model fitting should contain sufficient number of estimated data points for precise degradation model identification. The proposed RUL prediction adapts to any new information of deteriorating state of the supervised system. Initially, for RUL prediction, linear DM is used which has good prediction accuracy with less information of data and as the monitoring is continued DM and RUL are adapted with the new facts of deteriorating state of the component or system. The initially predicted RUL provides some indication to the maintenance engineers for decision and maintenance planning. Thus, the selection of DM which depends on the information of estimated data points up to current time t_k at mode $\boldsymbol{Z} = z^{(i)}$ is represented as;

$$M_{\theta_j}^{z^{(i)}}: \begin{cases} \theta_j(t, z^{(i)}) = \mathrm{M}_{\mathrm{LIN}}, & \text{if } w < w_S \\ \theta_j(t, z^{(i)}) = \zeta_1 \mathrm{M}_{\mathrm{LIN}} + \zeta_2 \mathrm{M}_{\mathrm{PL2}} + \zeta_3 \mathrm{M}_{\mathrm{EXP}}, & \text{if } w \geq w_S \end{cases} \tag{16}$$

where w is number of estimated data points up to current time, w_S is sufficient number of estimated data points decided by the user for precise degradation model identification. $\mathrm{M}_{\mathrm{LIN}}$ is linear model, $\mathrm{M}_{\mathrm{PL2}}$ is second order polynomial model and $\mathrm{M}_{\mathrm{EXP}}$ is exponential model. Thus, if $w \geq w_S$, a particular degradation model is selected for data fitting according to the value of degradation model vector $\boldsymbol{DMV} = [\zeta_1, \zeta_2, \zeta_3]$. If $\boldsymbol{DMV} = [1, 0, 0]$ then $\mathrm{M}_{\mathrm{LIN}}$ is selected, if $\boldsymbol{DMV} = [0, 1, 0]$ then $\mathrm{M}_{\mathrm{PL2}}$ is selected and if $\boldsymbol{DMV} = [0, 0, 1]$ then $\mathrm{M}_{\mathrm{EXP}}$ is selected and the model which has least root mean square error (RMSE) to the data fit is selected as a best degradation model at mode $\boldsymbol{Z} = z^{(i)}$. Likewise, other equation models can be plugged into Equation (13).

However, this paper uses these three models only for RUL prediction since these models show the monotonic increasing or decreasing trend and can touch the set failure threshold after extrapolation of the models. It is also suggested that the polynomial models of higher order more than the second order model should be avoided in data fitting unless there is some known physical reason or any past experience of such type of degradation of the component. The RUL prediction with higher order polynomial models may give good interpolation result, but may give bad extrapolation result (Randall, 2011).

RUL is predicted for degrading parameter $\{\theta_j(t, Z), t > 0\}$ by extrapolating the data using the identified model $M_{\theta_j}^{z^{(i)}}$ with the future operating modes ($\boldsymbol{Z}$) of the system known up to the current time (obtained from past experience of the system). Thus, when the extrapolated trend of θ_j reaches a well set failure threshold θ_j^{fl} then the component is declared as end of life (EOL) component at time to failure (TTF) t_{fl}. Thus, the t_{fl} and RUL are defined as

$$t_{fl} = \inf\{t \in \mathbb{R} : \theta_j(t, Z) \geq \theta_j^{fl} \mid \theta_j(t_0, Z) < \theta_j^{fl}\} \tag{17}$$

$$\mathrm{RUL}\left(t, Z\right) = t_{fl} - t_0 \tag{18}$$

Also, the identified mode-dependent DM, $M_{\theta_j}^{z^{(i)}}$, are fed into DHBG model for detection of subsequent faults precisely. In case of multiple degrading components, RUL is predicted for every degrading component and the component with least predicted RUL requires more attention by plant technicians.

4 NUMERICAL SIMULATION

The proposed approach is demonstrated through simulation on two-tank system (Fig. 1). Two faults are injected, first in valve V_1 (mode, a_{v1}, dependent degradation, see Fig. 4) and second in tank T_1 at instances $t_{f1} = 225$ s and $t_{f2} = 1475$s, respectively, as

$$C_{\mathrm{dv1}}(t,Z)=\begin{cases}C_{\mathrm{dv1n}}(Z).a_{\mathrm{v1}}, & \text{if } t < t_{f1}\\ C_{\mathrm{dv1n}}(Z)\left(e^{-r(Z).t_{\mathrm{on}}}\right).a_{\mathrm{v1}}, & \text{if } t \geq t_{f1}\end{cases} \quad (19)$$

$$C_{\mathrm{dLeak1}}(t)=\begin{cases}C_{\mathrm{dLeak1n}}, & \text{if } t < t_{f2}\\ C_{\mathrm{dLeak1n}}+k_{\mathrm{L}}(t-t_{f2}), & \text{if } t \geq t_{f2}\end{cases} \quad (20)$$

where $C_{\mathrm{dv1n}}(\boldsymbol{Z})$ is the nominal discharge coefficient of V_1 at respective mode ($\boldsymbol{Z}$), $r(\boldsymbol{Z}) = 1.0 \times 10^{-4}$ s^{-1} at $\boldsymbol{Z} = \mathrm{z}^{(1)} = a_{\mathrm{v1}} = 1$, and $r(\boldsymbol{Z}) = 0$ s^{-1} at $\boldsymbol{Z} = \mathrm{z}^{(2)} = a_{\mathrm{v1}} = 0$, $t_{\mathrm{on}} = \int_{t_{f1}}^{t} a_{\mathrm{v1}}\mathrm{d}t$, C_{dLeak1n} is the nominal discharge coefficient of V_{Leak1} and $k_{\mathrm{L}} = 1.0 \times 10^{-6}$ kg$^{1/2}$ m$^{1/2}$ s^{-1}.

The system is simulated for a time span of 2400s, with a fixed step size of 0.02 s, by setting all state variables to zero at $t = 0$ s. The nominal parameters of system are $K_{\mathrm{P}} = 1$ ms, $K_{\mathrm{I}} = 5 \times 10^{-2}$ m, $S_{\mathrm{pt}} = 0.5$ m, $f_{\max} = 1$ kg/s, $A_i = 2.16 \times 10^{-2}$ m^2, $C_{\mathrm{dv}i} = 1.593 \times 10^{-2}$ kg$^{1/2}$ m$^{1/2}$, $C_{\mathrm{dL}i} = 1 \times 10^{-3}$ ms, $C_{\mathrm{dLeak}i} = 0$ kg$^{1/2}$ m$^{1/2}$, (i = 1, 2), $H_{\mathrm{L1}} = 0.58$ m, $H_{\mathrm{L2}} = 0.40$ m, $P_{\mathrm{atm}} = 0$ N/m^2, $g = 9.81$ m/s^2, $\rho = 1000$ kg/m^3.

4.1 *Implementation of proposed scheme*

The measurements (Q_{P}, H_1, H_2) and the modes (a_{v1}, a_1, a_2) information obtained from the simulation are used to evaluate the residuals and the adaptive thresholds presented in Equations (4)-(5) and Equations (8)-(9), respectively. Activation of mode (a_1) and no activation of mode a_2 are noticed according to respective prescribed condition. So, $a_2 = 0$ for all the duration of simulation. The response of residuals (r_1, r_2) and adaptive thresholds (ε_1, ε_2), using single fault assumption, is shown in Figure 5.

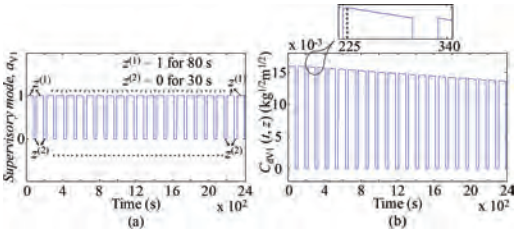


Figure 4. (a) mode, a_{v1} (b) mode-dependent degradation in V_1.

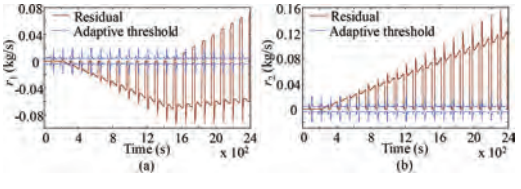


Figure 5. Residuals and thresholds with single fault assumption.

From the simulation results, $H_1 > H_2$ is found during the observation period 0 to 2400s, so the GFSSM and MCSSM are as presented in Table 1. Also, the obtained $\boldsymbol{CV}_{\boldsymbol{S}}$ (subscript $\boldsymbol{s}$ signifies single fault assumption) using previous existing approaches and obtained $\boldsymbol{CV}_{\boldsymbol{M}}$ (subscript $\boldsymbol{M}$ signifies multiple sequential faults assumption) using proposed method just after the both faults are also shown. Note that taking of absolute value of signatures in Table 1 without deviation direction provides the standard GFSM/MCSM. $\boldsymbol{CV}_{\boldsymbol{S}}$, just after 225 s, provides SSF = {a_{v1}, C_{dv1}} if mode $a_1 = 0$; otherwise SSF = {a_{v1}, C_{dv1}, C_{dL1}} if mode $a_1 = 1$ (see Table 1). In any mode, parametric fault (C_{dv1}) is not isolatable as a_{v1}, C_{dv1}, C_{dL1} share a common signature. Also, the inclusion of a_{v1} in SSF complicates the task of FD; so, tracking real mode ($a_{\mathrm{v1}} = 1$) just after the fault is also required. As dynamics of V_1 is dependent on mode a_{v1}, the evaluated residuals (r_1, r_2) at $a_{\mathrm{v1}} = 0$ forces the residuals to enter into the threshold bounds even after detecting this fault (see Fig. 5); and under such situation, FD is not possible unless the system enters into a different working mode ($a_{\mathrm{v1}} = 1$). Also, $\boldsymbol{CV}_{\boldsymbol{S}}$, just after 1475s, provides the same SSF as before. In such cases, the real fault (C_{dLeak1}) is not included in SSF which leads to misdiagnosis. Here, next component degradation (C_{dLeak1}) is concealed by the already known component degradation (C_{dv1}).

The response of residuals (r_1, r_2) and adaptive thresholds (ε_1, ε_2) using proposed modified method, in which LFT-DHBG model is dynamically updated after each fault estimate, is shown in Figure 6. It is observed that using proposed technique with locally updating the model, SSF includes the true faults

Table 1. GFSSM with MCSSM of the system.

Parameter	r_1	r_2	D_b	I_{b}
$C_{\mathrm{dv1}}\uparrow$	a_{v1}	$-a_{\mathrm{v1}}$	a_{v1}	0
$C_{\mathrm{dv1}}\downarrow$	$-a_{\mathrm{v1}}$	a_{v1}	a_{v1}	0
$C_{\mathrm{dv2}}\uparrow$	0	+1	1	0
$C_{\mathrm{dv2}}\downarrow$	0	−1	1	$1-a_2$
$C_{\mathrm{dL1}}\uparrow$	a_1	$-a_1$	a_1	0
$C_{\mathrm{dL1}}\downarrow$	$-a_1$	a_1	a_1	0
$C_{\mathrm{dL2}}\uparrow$	0	a_2	a_2	0
$C_{\mathrm{dL2}}\downarrow$	0	$-a_2$	a_2	0
$C_{\mathrm{dLeak1}}\uparrow$	+1	0	1	1
$C_{\mathrm{dLeak2}}\uparrow$	0	+1	1	0
$a_{\mathrm{v1}}\uparrow$	+1	−1	1	0
$a_{\mathrm{v1}}\downarrow$	−1	+1	1	0
$\boldsymbol{CV}_{\mathrm{S}}$ after 225 s	−1	+1	1	0
$\boldsymbol{CV}_{\mathrm{S}}$ after 1475s	−1	+1	1	0
$\boldsymbol{CV}_{\mathrm{M}}$ after 225 s	−1	+1	1	0
$\boldsymbol{CV}_{\mathrm{M}}$ after 1475s	+1	0	1	1

according to $\boldsymbol{CV_M}$ for the same faults scenario discussed beforehand. The suspected mode fault ($a_{v1}\downarrow$) is tested as per Equation (14) just after threshold violation and it is found that $a_{v1} = 1$ is consistent. Thus, using CPE technique only for the faults remain in the refined SSF, i.e., $C_{dv1}\downarrow$ and $C_{dLeak1}\uparrow$ are isolated as true faults just after 225 s and 1475s, respectively. Also, the predicted RULs for both faults using linear DMs, with less data points of deteriorating parameters ($C_{dv1}\downarrow$ and $C_{dLeak1}\uparrow$), are presented in Figures 7a and 8a, respectively. The horizontal lines in these figures indicate the degradation threshold or end of life of a component. Subsequently, updated RULs estimated with adapted models after getting the sufficient data points are presented in Figures 7b and 8b, respectively. The gradually evolving degradation pattern reconstructed from results of the simulation and parameter identification matches with the expected behavior defined in Equations (19)-(20).

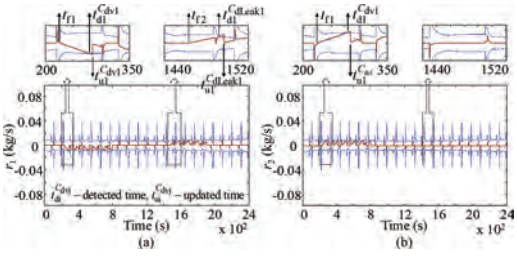


Figure 6. Residuals and thresholds using proposed method.

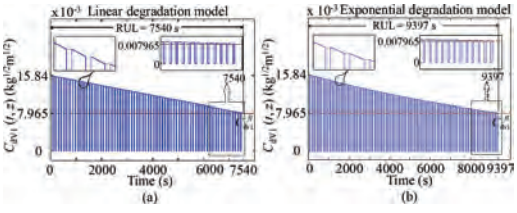


Figure 7. RUL of C_{dv1} using (a) initial linear model (b) identified true model, with a failure threshold $\left(C_{dv1}^{fl}\right)$.

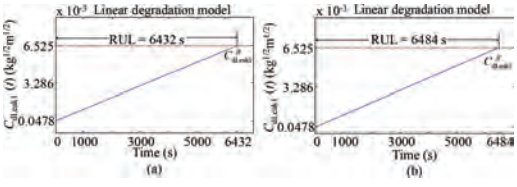


Figure 8. RUL of C_{dLeak1} using (a) initial linear model (b) identified true model, with a failure threshold $\left(C_{dLeak1}^{fl}\right)$.

5 CONCLUSION

The key features of the proposed technique are (1) the use of residual response to identify the minimal set of prognosis candidates for which parameter estimation is triggered, (2) detection of the time points when the parameter estimation needs to be triggered (3) the use of minimization of only inconsistent residuals for the degradation state estimation and the use of sensitivity of ARRs during estimation (4) successive ARRs updation with the degradation estimates in order to force the residuals to lie within the respective adaptive thresholds and thereby detect the further degradation and predict the degradation trends of multiple degrading components, and (5) RUL prediction is coupled with the operation mode of the system and adapts to any new state of health information on the system's components. The implementation of proposed CPE technique with SBG approach, dynamic model updation and use of GFSSM/MCSSM improve the FD and RUL prediction. The problem of misdiagnosis with multiple sequential faults is also discussed. However, the proposed work demonstrated through simulation and the predicted RULs are also deterministic in nature. In future, the proposed technique will be applied to some real experiment, and uncertainties and confidence limit will be taken into account for the prediction of the RULs. Also, in the current work, failure threshold for each component is selected arbitrarily just for demonstration. However, setting failure threshold for multi-component system based on system level performance limits is also a crucial design factor in prognosis, which can be considered as a future research problem.

REFERENCES

Borutzky, W. (2015) Bond Graph Model-based Fault Diagnosis of Hybrid Systems, Switzerland: Springer.

Daigle, M., Roychoudhury, I., Bregon, A. (2015) "Model-based prognostics of hybrid systems," In Annual Conference of the Prognostics and Health Management Society.

Djeziri, M.A., Merzouki, R., Ould Bouamama, B., Dauphin-Tanguy, G. (2007) "Robust fault diagnosis by using bond graph approach," IEEE/ASME Trans. Mechatronics, vol. 12, no.6, pp. 599–611.

Ghoshal, S.K., Samanta, S., Samantaray, A.K. (2012) "Robust fault detection and isolation of hybrid systems with uncertain parameters," Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering; vol 226, no. 8, pp. 1013–1028.

Jha, M.S., Dauphin-Tanguy, G., Ould Bouamama, B., "Particle filter based hybrid prognostics for health monitoring of uncertain systems in bond graph framework," Mechanical Systems and Signal Processing, vol 75, pp. 301–329, 2016.

Levy, R., Arogeti, S., Wang, D., Fivel, O. (2015) "Improved diagnosis of hybrid systems using instantaneous sensitivity matrices," Mechanism and Machine Theory, vol. 91, pp. 240–257.

Low, C.B., Wang, D., Arogeti, S., Luo, M. (2010) "Quantitative Hybrid Bond Graph-Based Fault Detection and Isolation," IEEE Transactions on Automation Science and Engineering, vol 7, no. 3, pp. 558–569.

Medjaher, K., Zerhouni, N. (2009) "Residual-based failure prognostic in dynamic systems," IFAC Proceedings, vol 42, no. 8, pp. 716–721.

Merzouki, R., Samantaray, A.K., Pathak, P.M. Ould Bouamama, B. (2012) Intelligent mechatronic systems: modelling, control and diagnosis, London: Springer.

Narasimhan, S., Biswas, G. (2007) "Model-based diagnosis of hybrid systems," IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, vol 37, no. 3, pp. 348–361.

Nocedal, J., Wright, S. (2006) Numerical optimization. New York: Springer.

Randall, R.B. (2011) Vibration-based condition monitoring: industrial, aerospace and automotive applications. John Wiley & Sons.

Samantaray, A.K., Ghoshal, S.K. (2007) "Sensitivity bond graph approach to multiple fault isolation through parameter estimation," Proc. Inst. Mech. Eng. Part-I: J. Syst. Control Eng., vol. 221, no. 4, pp. 577–587.

Samantaray, A.K., Medjaher, K., Ould Bouamama, B., Staroswiecki, M., Dauphin-Tanguy, G. (2006) "Diagnostic bond graphs for online fault detection and isolation," Simulation Modelling Practice and Theory, vol.14, pp. 237–262.

Samantaray, A.K., Ould Bouamama, B. (2008) Model-Based Process Supervision: A Bond Graph Approach, London:Springer.

Wang, D., Yu, M., Low, C.B., Arogeti, S. (2013) Model-based Health Monitoring of Hybrid Systems, New York: Springer.

Yu, M., Wang, D., Luo, M. (2015) "An integrated approach to prognosis of hybrid systems with unknown mode changes," IEEE Transactions on Industrial Electronics, vol. 62, no. 1, pp. 503–515.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Energy efficiency and predictive maintenance applications using smart energy measuring devices

S. Kotsilitis
System Reliability and Industrial Safety Laboratory, National Centre for Scientific Research "Demokritos", Athens, Greece
Department of Information and Communication Systems Engineering, University of the Aegean, Samos, Greece

E.C. Marcoulaki
System Reliability and Industrial Safety Laboratory, National Centre for Scientific Research "Demokritos", Athens, Greece

E. Kalligeros
Department of Information and Communication Systems Engineering, University of the Aegean, Samos, Greece

Y. Mousmoulas
Plegma Labs, Athens, Greece

ABSTRACT: This paper discusses novel technologies for energy efficiency and predictive maintenance using hardware accelerated energy disaggregation. The disaggregation process involves the use of custom designed smart sensors that collect and treat aggregated information on the current and voltage waveforms. The treated data are further on transmitted to the cloud where they are stored and processed to enable the extraction of advanced information on individual device consumption patterns and health status. This information can be extremely useful for the management of electric devices in residential or commercial sites as well as for predictive maintenance in industrial sites. The paper reviews the underlying methodologies, and presents preliminary work and results from data collection in the offices of a software company. The presented work involves the installation of measurement devices and the development of complementary hardware and software. This is part of the ongoing 4-year project PREDIVIS (PREdictive, Disaggregation Intelligent VIS (meaning "power" in Latin)).

1 INTRODUCTION

Nowadays, the ever-growing power demand of industries and households combined with the goals for carbon dioxide emission reduction, have led the communities to take action, by implementing conservation and energy efficiency programs. The first step in energy reduction actions is the rollout of smart meters to monitor energy consumption and the smart grid technologies to distribute the available energy more efficiently, combined with the wide adoption of renewable energy sources.

The energy consumption and carbon emissions are regulated by frameworks and directives, mainly focused on actions by industries in order to minimize their impact on climate change. In most cases these actions are costly and inefficient, and often industries are incapable of adapting new equipment and carbon dioxide emission reduction techniques which leads to increased taxes and fines when goals are not achieved.

Monitoring of energy consumption at appliance level is essential for predicting energy needs and monitoring appliance operation in a household, a building or an industrial system. Energy disaggregation refers to using data analytics and signal processing, to identify specific patterns and to break down electricity consumption to individual appliances. This is usually done in a non-intrusive manner by monitoring the utility connection meter, and has been a field of significant research work for over twenty years. Non-Intrusive Load Monitoring (NILM) is a process where the aggregated electricity consumption is metered at the Grid-consumer connection point, and by analyzing the changes in voltage and current wavelengths tries to identify which appliances are being used at a certain time. Still, NILM technology's main goal is to provide insights into energy consumption at appliance level, mainly to support energy efficiency actions with economic and environmental impact. There are novel techniques using various approaches of NILM for a great number of

applications, like safety on industrial environments, device health monitoring and predictive maintenance and demand response applications.

Equipment monitoring on industrial sites is a necessity and most of the time, a costly and complex process. Various industries, monitor their machinery and equipment to prevent malfunctions, minimize danger, service and repair costs, and to increase the overall operating time. NILM techniques could be a cheap alternative to equipment monitoring systems which are costly and require huge and complex installations. Monitoring equipment is vulnerable to failures due to its numerous sensors and measurement devices that are being deployed. NILM is not widely used in industrial and commercial environments because of the complexity of these environments: the great number of similar devices, power factor correction and load balancing equipment, as well as the huge number of harmonics in loads make this process really challenging.

2 TECHNICAL PROBLEM DESCRIPTION

Let a system of N devices. Devices can be of different types (let K denote the number of possible types, e.g., washing machine, PC, monitor, refrigerator, etc.). For each type $k = 1, 2, \ldots, K$ of device $n = 1, 2, \ldots, N$ there is a set S_k of possible operation states. The assumption here is that there is a mapping between the state of a certain device and its electrical footprint on an aggregated time series. Consequently, the sequence of operational states leaves a string of unique fingerprints on the time series of energy consumption measurements.

The usual case is that for each time segment, only the total energy consumption is measured and not the individual consumptions of each device. The device fingerprints are therefore mixed up. The disaggregation exercise consists of analyzing the system data in order to unravel the strands of each device, and enable further analysis of the device operation. The disaggregation process is accompanied by information on the operational pattern of each device type, for instance, continuous operation or interrupted, expected duration of each operational state etc.

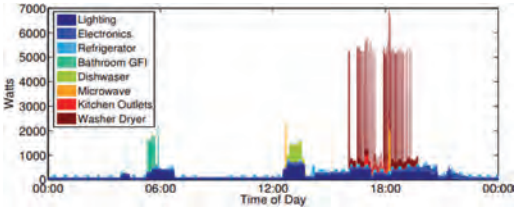


Figure 1. Example of a household total energy consumption from the REDD[1] dataset.

Looking more closely at the device types, it is also possible to extract and make use of more advanced information. Indeed, different device types usually generate slightly different harmonic distortions. The harmonic distortions can be identified if the resolution of the time series is sufficiently high.

3 STATE-OF-THE ART ON ENERGY DISAGGREGATION

Over the past 20 years, there have been many different approaches to addressing the problem of energy disaggregation and the subsequent monitoring of device health.

Since Hart [1] first introduced the concept of Non-Intrusive Appliance Load Monitoring (NIALM) in 1992, numerous techniques have been developed to address the problem. Initial approach was focused on examining a device as a state machine, trying to identify the states and disaggregate the device from the aggregated load. The method could perform well for large loads and devices with a finite number of states that are not always on, with discrete power changes between states.

Since then, other works [5, 6, 7] have proposed different techniques for electrical signature analysis to address the classification problem, with promising results. Such methods include, Support Vector Machines (SVM), Bayesian methods, k-Nearest Neighbors etc. The most common techniques used are Hidden Markov Models and Artificial Neural Networks using supervised, semi-supervised and unsupervised methods [8–12].

Regarding energy disaggregation, the non-intrusive approaches (NILM) promise adequate accuracy with lower installation costs and complexity compared to smart plug-based [2, 12] approaches. NILM methods using steady-state and transient load signatures are further classified according to data time series' measurement frequency. High-frequency [13] methods require custom hardware (high-frequency meters ~10^6 Hz/s) and employ an array of machine learning and pattern recognition methods. Low-frequency methods [14] (1 sec up to 1 hour) apply similar data processing techniques, but are not sufficiently tested to guarantee commercial-grade accuracy. NILM is so far mainly focused on households and small-scale buildings, so there is also the issue of its scalability to commercial buildings or even entire neighborhoods in order to extract useful information for demand response applications, as well as for grid and device health.

The models employed range from least square estimation to Hidden Markov models. Some approaches use Fast Fourier Transform and other transformations to reduce hardware, bandwidth

and storage cost. Research in this field focuses on finding the algorithm that increases the accuracy of energy disaggregation in each application case. More recently, new approaches with semi-/un-supervised algorithms are being studied [15]. An emerging research trend is to use IoT based architectures for data capturing and on-board [16] analytics on appliance level to provide energy efficiency solutions. By using IoT devices dedicated to energy monitoring and data manipulation, researchers aim to extract more information by analyzing the electrical characteristics of the appliances in the deployed sites.

At the moment, there is no universal Machine Learning (ML) algorithm that will fit multiple application cases. The specification of the ML algorithms varies with the constraints of each application case. A list of well performing algorithms in different application cases should include Artificial Neural Networks, SVM + kernels, Decision Tree, Random Forests etc. Marking a turning point in the history of Artificial Intelligence, Deep Neural Networks (DNN) are now widely used, e.g., for face recognition on smartphone cameras. Research in this field is ongoing in response to the evolving market interest for improved DNNs. It includes development of new hardware architectures implementing DNNs to improve on the current CPUs and GPUs [17]. Neuromorphic chips have reduced energy consumption and enhanced DNN capabilities in processing the vast volumes of information generated by the IoT [18].

Analyzing data from multiple sensors can provide critical information on each current state of the monitored devices and enable predictions of behavior in the future. The sensors can measure a variety of device/environmental attributes, such as the temperature [19]. Likewise, electrical consumption data, when added to other available device data, can provide significant input to predictive maintenance, and also minimize the data volume that needs to be processed and stored. Approaches to predictive maintenance through electrical consumption data has been made on specific cases.

4 PROPOSED APPROACH

4.1 *Description of solution*

Project PREDIVIS aims to develop novel tools for energy disaggregation and monitoring of device operation status, based on real-time pattern recognition/matchmaking of complex energy load data time-series, using hardware acceleration techniques. The proposed approach requires the design, development and implementation of complex algorithms on a reprogrammable Field Programmable Gate Array (FPGA), in order to create a network of distributed agents that performs the majority of the data analysis in real-time, and transmits events, instead of raw data, to a main server.

This project is trying to address the problem of energy Disaggregation on household and commercial/industrial environments. Due to the difference in complexity of the above mentioned two cases that this project is trying to address, we will need to utilize different approaches, algorithms and also fine-tune the sampling frequency needed per case to acquire sufficient data for the disaggregation process. A system like that depends on the specifications of each deployment site, e.g., different sites have different number of devices with different characteristics that might lead us in using completely different data acquisition rate. Typically, a NILM system design involves three main components: Data Acquisition and Storage, Analysis and Classification.

Most of the previous projects were using private generated/produced or open datasets to train models and validate the results and the algorithm efficiency. Previous works are using data collection methods at either High frequency (1 to 10^3 kHz) or Low frequency (10^{-6} to 10^{-3} kHz). The sampling rate may differ from one sample per 15 minutes or more, to a couple of millions per second. This project focuses on High frequency methods using a sampling rate between 8 kHz and 64 kHz to be able to extract more features from the available data to assist the classification process. Sampling rate determines the information that can be extracted from the sampled signals. Consider for example an electrical installation with fundamental power frequency of 5×10^{-2} kHz. Sampling the wavelength with higher sampling rate (8 kHz) fulfilling the Nyquist-Shannon theorem, enables our system to capture up to the 160-th harmonic. By analyzing the different harmonic distortions, we can identify and differentiate the device from the aggregated workload.

Using high frequency sampling rates requires large storage space to store the acquired data and huge bandwidth to transfer the data over the internet to a central powerful unit for further analysis. To minimize storage and bandwidth, some applications are using compression technics or on-site devices to analyze the data. These hardware devices require a great amount of power in order to perform the analysis and usually they are very expensive.

Project PREDIVIS will use a novel technique implementing on site disaggregation to limit bandwidth and storage needed (Figure 2). Using dedicated hardware implemented in FPGA devices will help in decreasing not only the overall bandwidth and storage but also the power consumption needed for the data analysis. Based on the installation,

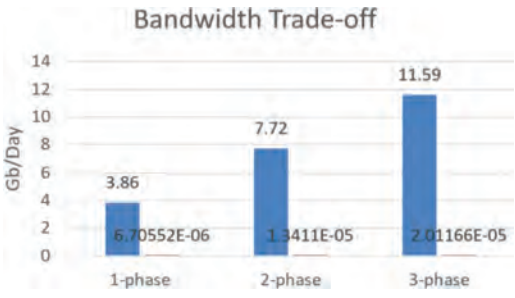


Figure 2. Bandwidth trade-off between raw data transfer and event reporting architecture.

which can be one-, two- or three- phase, the number of electric power transmission data to be used for the disaggregation, will increase proportionally. Note, however, that, the detected events will be more or less the same for each case, despite the number of phases.

For data storage, PREDIVIS will use a small local memory capable of storing the device signatures and a couple of hours of data stream. Aside from the local storage, data like on/off events, total energy consumption and total amount of operation time, anomalies on appliance electric characteristics etc., will be sent to a cloud infrastructure and will be saved in a No-SQL database. Each agent will have the ability to retrain when specific conditions are applied. The system will use adaptive learning techniques to adapt better on new and existing installations by sharing knowledge on already known devices between the employed agents through the central cloud system.

In terms of data analysis, various techniques will be used to extract information from the available data in order to identify:

- Event transitions, when a device is turned on or off. Event-based approaches detect only major changes and anomalies on the energy load time series.
- State transitions, when the device swifts from one state to another (e.g. from full operation to standby and vice versa). State-based approaches also detect the different states of device workload.

4.2 *Description of project*

Project PREDIVIS consists of three main components:

- Agents, which are custom hardware components for data collection, data analysis and load monitoring
- Cloud-based platform for data visualization, storage and NILM assisting mechanisms

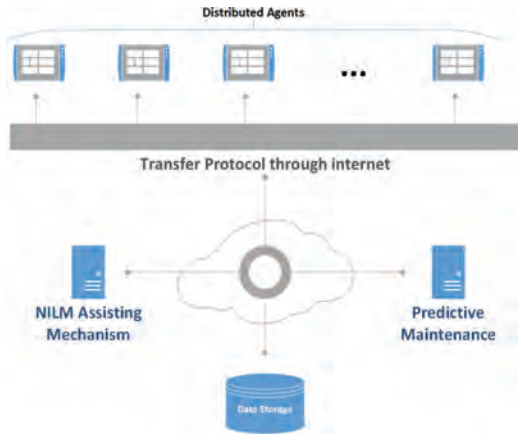


Figure 3. PREDIVIS architecture diagram.

- NILM and Predictive maintenance algorithm suite

The Agents mentioned above are custom hardware implementations using FPGA devices and different Intellectual Property (IP) blocks (e.g., ADC, DSP etc.) to address data collection and analysis on the deployment site. Hardware acceleration of NILM algorithms will help us take the computational load off the cloud infrastructure and minimize the bandwidth needed, as mentioned earlier. Each deployment site is unique, having a different number of N devices with different K states, making the case of using a universal algorithm/approach very challenging. Each device will be able to perform better on its deployment site through adaptive learning techniques, with the assistance of the Central platform and information gathered by other deployed agents with similar site characteristics or identical device types.

The Cloud-based platform will provide data visualization and display information through a friendly User Interface (UI) helping the user get insights on the energy consumption. Acting as a central point for reporting, the platform will complement distributed agents by collecting and analyzing information from each one of them about the deployment site and the site's devices, and will help distributing knowledge between them.

The Cloud platform will also host a NILM suite with several algorithms for analyzing real-time data to determine which algorithm/method is more suitable for that particular site's appliance mix, condition etc. Finally, the predictive maintenance suite will analyze the data and anomalies detected by the agent to help reduce hazardous machinery errors, downtimes and failures.

The consortium of this project comprises the following complementary partners:

- the System Reliability and Industrial Safety Laboratory, National Center for Scientific Research "Demokritos" as a research partner supporting NILM and predictive maintenance analysis.
- Plegma Labs S.A as Enterprise partner in IoT technologies supporting the cloud infrastructure, and the data storage and management.
- the Department of Information and Communication Systems Engineering, University of the Aegean, as a research partner supporting hardware development and data acquisition processes.

Figure 4 depicts the three main stages of this project. The project will run for 4 years and the work is currently at stage one.

4.3 *PREDIVIS technologies breakdown*

The project is a combination of the aforementioned techniques, ranging from hardware blocks to advanced software features. In a nutshell, this project will try to implement hardware designs for data collection using Analog to Digital conversion and Digital Signal Processing techniques, combined with embedded Artificial Intelligence functions and methods. The ability to reprogram over the air an FPGA device, can help each device adapt better to new or pre-installed environments. The software portion of this project includes:

a. the Cloud-based high-level software for data transport and storage,
b. the intelligent adaptive NILM algorithm suite to reprogram and calibrate DNN on agents, and
c. the Predictive maintenance analytics suite combined with statistical models and machine learning algorithms, to predict future failures and stimulate faults.

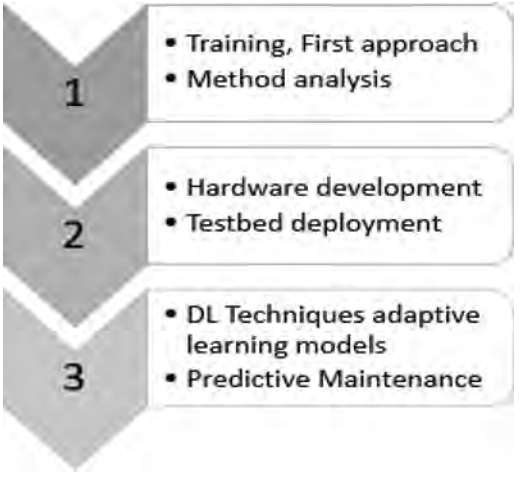


Figure 4. PREDIVIS project stages.

5 GENERATION OF DATA SETS

In order to develop and test different energy disaggregation methods and optimize the efficiency of PREDIVIS project, a variety of datasets will be used. These involve open High Frequency public datasets as well as privately generated data. Some of the public sets that we intend to use are the REDD [2] (2011), the Blued [3] (2012), and the UK-DALE [4] (2015). These datasets relate mainly to residential applications.

The private datasets will contain data from office sites and from industrial sites. Regarding the former, a set of measuring devices is currently installed at offices of a typical software SME. Figure 3 shows the layout of the company offices. The site is connected to the electricity grid through a three-phase power supply. Note that, three-phase data have entirely different characteristics compared to single or two-phase data and this will be considered during the analysis stages.

The installed measuring devices collect data logs from the main power circuitry connector, as well as from individual devices. The main power data involve the aggregated electric current and voltage waveforms, and these are measured at both low and high-frequency rates. The electric power of individual devices is monitored using one smart plug per device.

Seventeen (17) different entities are monitored, ranging from lighting to air-condition units. These include multiple devices of the same appliance type, for example 9 monitors and 3 laptops.

5.1 *Monitoring devices setup*

This section presents the employed technical equipment towards the generation of the dataset described above.

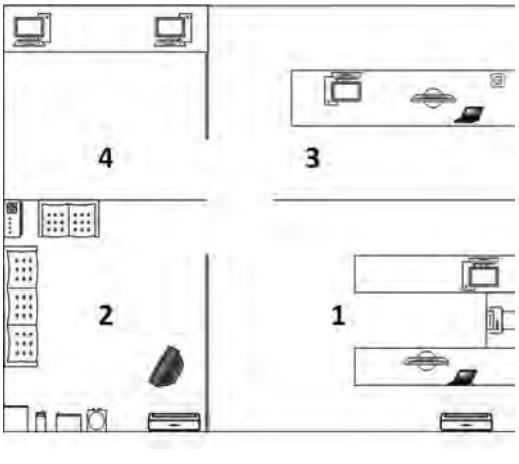


Figure 5. Plegma labs headquarters installation.

For the main power supply, the authors developed a custom implementation using (a) a set of voltage and electric current converters for the measurements, (b) an Analog to Digital Converter (ADC) and (c) an ARM based single board computer for data treatment. In particular:

a. The current sensors used for this project are current transformers with 1:1800 turn ratio with rated input of 100 Amperes to 50 milliamperes output. For the voltage sensing, in-house implemented voltage transformers have been used.
b. The employed ADS is the ADS 131E08 by Texas Instruments, which is capable of sampling simultaneously eight different channels. The number of available channels allows the measuring of electric current and voltage for each of the three phases, leaving two channels free for additional analog sensors (e.g. for temperature, humidity, luminosity etc.). The sampling frequency can range from 1 kHz up to 64 kHz. For the needs of the PREDIVIS project, data are collected at the maximum resolution. In the future, the analysis will indicate whether lower are sufficient (to reduce the sensor consumption) without losing on the quality of results.
c. The arm single board computer is a Raspberry Pi 3 by Raspberry Pi Foundation which receives data from the ADC through serial peripheral interface (SPI) communication.

In order to verify the data collected through the above custom implementation, a widely used industrial grade energy meter/analyzer is also connected to the system. The selected device is the BFM136 produced by SATEC ltd (see Table 1).

For the recording of individual devices' data, a Z-wave based system is implemented. The system uses two different types of smart-pugs, namely the Wall Plug by Fibaro and the Smart Switch 6 by Aeotec. Each plug is monitored continuously at the interval of 5 seconds or less, using a Z-wave USB adapter. The adapter is the Z-Stick S2 by Aeotec (see Table 1).

The two sets of measuring devices are accompanied by appropriate software components developed by the authors. These include:

Table 1. Devices used in the case study.

Metering device	Number
BFM136	1
100 A High Accuracy Current Sensors	3
ADS136E08 with RPi	1
100 A:50 mA Current Sensors	3
Voltage Sensors	3
Fibaro Wall Plug	4
Aeotec Smart Switch 6	13

a. a No-SQL database to store the data locally. The chosen format for the recorded data is in the form of time—voltage—current triplets. These follow a key-value format, with UNIX timestamp for the time.
b. a custom interface is herein implemented to collect and transmit the data and visualize the current and voltage waveforms. The data are sent over the internet to Plegma's cloud infrastructure.

All the devices reported in Table 2 are measured through the smart-plug installation. The two main monitoring devices are currently polled in 3 second (BFM136) intervals, as well as with frequency of 8 kHz (custom implementation). The main monitoring devices are measuring the three-phase installation as follows:

Table 2. Measuring devices and entities.

Entity	Measuring device	Room No
Main	BFM136 & ADS with RPi	Electricity board
Water cooler	Fibaro Wall plug	2
Microwave	Fibaro Wall plug	2
Coffee maker	Fibaro Wall plug	2
Refrigerator	Fibaro Wall plug	2
Work station 1 desktop with 2 monitors	Aeotec Smart Switch 6	1
Work station 2 laptop with 1 monitor	Aeotec Smart Switch 6	4
Work station 3 1 desktop with 2 monitors	Aeotec Smart Switch 6	3
Work station 4 1 high load desktop with 2 monitors	Aeotec Smart Switch 6	4
1 Monitor	Aeotec Smart Switch 6	3
1 Laptop	Aeotec Smart Switch 6	3
1 TV monitor	Aeotec Smart Switch 6	2
1 Router	Aeotec Smart Switch 6	2
1 Printer	Aeotec Smart Switch 6	1
Guest plug 1	Aeotec Smart Switch 6	1 or 3
Guest plug 2	Aeotec Smart Switch 6	1 or 3
2 Air-condition Units	Aeotec Smart Switch 6	1, 2

- Phase 1 lights for rooms 1, 2, 3 and 4,
- Phase 2 sockets of rooms 1 and 2,
- Phase 3 sockets of rooms 3 and 4.

5.2 *Future steps*

Due to the differentiation of the two cases this project is trying to address, it is necessary to be able to simulate different events and scenarios to test its efficiency. The two cases are divided in two major categories residential and industrial/commercial. In order to test the efficiency of the utilized algorithm on specific cases and validate our data different devices will be simulated on variable working states.

The collected data allow us to test different approaches of NILM, based on either high or low frequency data. This project is trying also to address the problem of device health monitoring, and predictive maintenance. In order to have sufficient data for the third phase of the project where a Predictive Maintenance suite will be implemented, we will tamper some specific days of the data from the devices, with different methods (e.g. leaving the fridge door open, turning devices on and off, modifying thermal loads etc.).

Additional data will be generated to see how the disaggregation mechanisms work, to test their ability to distinguish the anomalies produced and match them to the device or entity appropriately. More similar sites will follow so that the final dataset has adequate variety to allow the development of widely applicable algorithms and tools.

6 CONCLUSIONS

In this paper, we reviewed the fundamentals of NILM systems and the energy disaggregation problem. A novel technique is proposed to address this problem that will be applicable to the energy efficiency and predictive maintenance.

Recent works indicate that energy disaggregation is an active field, and there are a lot of different approaches to address it. Even, however, the most advanced methods have not achieved adequate results to be reliable for deployment on a large scale. The problem, therefore, is still open and the potential benefits of disaggregation, in terms of its ability to support end-users and utilities, cannot be fully exploited.

The project PREDIVIS presented here proposes a novel approach with custom hardware implementation of measuring devices. The combination of software and hardware modules can address the problem of data bandwidth and storage size. Adequate information on predictive maintenance can be obtained by combining electric consumption data with other monitoring data.

The availability of relevant and reliable data is crucial for the development of the disaggregation tools. For this reason, the project starts with the generation of datasets for residential, commercial and industrial energy use patterns. The collected residential and commercial data will be combined with publicly available datasets. For commercial and industrial environments, the lack of public datasets makes it a more challenging process.

Once the datasets are fixed, the main work involves the development of energy disaggregation algorithms, the design of data collection and smart on-site devices for hardware accelerated analysis. With continuous monitoring it is possible to produce useful information about the device and machinery health. Monitoring the full cycle of operation could support energy efficiency and predictive maintenance applications, by detecting abnormalities, predicting total operating time of components etc. The current approach could thus detect early-stage device malfunction in industrial sites, supported by energy consumption evidence as well as other sensor data (e.g. temperature). The project considers, at a later stage, the development of decision support systems for industrial, commercial and large residential sites. The success of non-intrusive electric load monitoring on predictive maintenance could provide a much cheaper alternative as compared to complex and expensive monitoring equipment.

ACKNOWLEDGEMENTS

Author SK acknowledges financial support through the Programme of Industrial Scholarships of the Stavros Niarchos Foundation.

REFERENCES

[1] Hart, G.W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12), 1870–1891.

[2] Kolter, J.Z., & Johnson, M.J. (2011, August). REDD: A public data set for energy disaggregation research. In Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA (Vol. 25, pp. 59–62).

[3] Filip, A. (2011). BLUED: A fully labeled public dataset for event-based non-intrusive load monitoring research. In 2nd Workshop on Data Mining Applications in Sustainability (SustKDD) (p. 2012).

[4] Kelly, J., & Knottenbelt, W. (2015). The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. Scientific data, 2, 150007.

[5] Kolter, J.Z., Batra, S., & Ng, A.Y. (2010). Energy disaggregation via discriminative sparse coding. In Advances in Neural Information Processing Systems (pp. 1153–1161).

[6] Barsim, K.S., Streubel, R., & Yang, B. (2014, June). An approach for unsupervised non-intrusive load monitoring of residential appliances. In Proceedings of the 2nd International Workshop on Non-Intrusive Load Monitoring.

[7] Bochao Zhao, Lina Stankovic, and Senior Member. On a Training-Less Solution for Non-Intrusive Appliance Load Monitoring Using Graph Signal Processing. IEEE Transactions on Smart Grid, 4, 2016.

[8] Bonfigli, R., Principi, E., Fagiani, M., Severini, M., Squartini, S., & Piazza, F. (2017). Non-intrusive load monitoring by using active and reactive power in additive Factorial Hidden Markov Models. *Applied Energy*, 208, 1590–1607.

[9] Beckel, C., Kleiminger, W., Cicchetti, R., Staake, T., & Santini, S. (2014, November). The ECO data set and the performance of non-intrusive load monitoring algorithms. In Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings (pp. 80–89). ACM.

[10] Li, G.A.O., Bo, Y.I.N., & ZHU, Z.C. (2017). Load Identification of Non-intrusive Load-monitoring System Based on Time-frequency Analysis and PSO-SVM. DEStech Transactions on Engineering and Technology Research, (EETA 2017).

[11] Shaw, S.R., Leeb, S.B., Norford, L.K., & Cox, R.W. (2008). Nonintrusive load monitoring and diagnostics in power systems. *IEEE Transactions on Instrumentation and Measurement*, 57(7), 1445–1454.

[12] Ridi, A., & Hennebert, J. (2014). Hidden Markov Models for ILM appliance identification. *Procedia Computer Science*, 32, 1010–1015.

[13] Bouhouras, A.S., Milioudis, A.N., & Labridis, D.P. (2014). Development of distinct load signatures for higher efficiency of NILM algorithms. Electric Power Systems Research, 117, 163–171.

[14] Liao, J., Elafoudi, G., Stankovic, L., & Stankovic, V. (2014, June). Power disaggregation for low-sampling rate data. In 2nd International Non-intrusive Appliance Load Monitoring Workshop, Austin, TX.

[15] Liu, B., Luan, W., & Yu, Y. (2016). A Fully Unsupervised Appliance Modelling Framework for NILM. In Proceedings of the 3rd international workshop on NILM.

[16] Villani, C., Balsamo, D., Brunelli, D., & Benini, L. (2015, May). Ultra-low power sensor for autonomous non-invasive voltage measurement in IoT solutions for energy efficiency. In SPIE Microtechnologies (pp. 95172I-95172I). International Society for Optics and Photonics.

[17] Knag, P., Kim, J.K., Chen, T., & Zhang, Z. (2015). A sparse coding neural network ASIC with on-chip learning for feature extraction and encoding. *IEEE Journal of Solid-State Circuits*, 50(4), 1070–1079.

[18] Sharma, H., Park, J., Amaro, E., Thwaites, B., Kotha, P., Gupta, A. & Esmaeilzadeh, H. (2016). Dnnweaver: From high-level deep network models to fpga acceleration. In the Workshop on Cognitive Architectures.

[19] Daily, J., & Peterson, J. (2017). Predictive Maintenance: How Big Data Analysis Can Improve Maintenance. In Supply Chain Integration Challenges in Commercial Aerospace (pp. 267–278). Springer International Publishing.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A diagnosis method for diesel engine wear fault based on grey rough set and SOM neural network

Silin Qian, Shenghan Zhou, Wenbing Chang & Yiyong Xiao
School of Reliability and Systems Engineering, Beihang University, Beijing, China

Fajie Wei
School of Economics and Management, Beihang University, Beijing, China

ABSTRACT: The paper aims to establish a model to identify wear fault of marine diesel engine based on grey rough set and Self-Organizing Map (SOM) network with oil monitoring data analysis. The empirical data indicates the wear fault takes great proportion in fault types of diesel engine. Through oil monitoring, the change of parameters of lubricating oil and the information of wear particle can be obtained to analyze status of components. Firstly, the paper constructs the two-dimensional fault decision table. Subsequently, the grey relational analysis and rough set theory are used to reduce the fault decision table horizontally and longitudinally. Next, the fault diagnosis model is established by SOM network. Finally, the proposed model is validated by empirical research. The result suggests that the proposed model is feasible in wear fault diagnosis problem. Moreover, compared with the traditional SOM neural network, the model has less error and better diagnosis effect.

1 INTRODUCTION

Marine diesel engine as the heart of the marine power plant, its safety and reliability is vital. However, due to the abominable work condition and the complexity of marine diesel engine structure, the fault of diesel engine is relatively frequent. In addition, there are many kinds of faults in diesel engines, and the proportion of faults caused by abnormal wear is the highest, which is about 37.5% (Jones et al. 2000). Therefore, monitoring of the running status in real time and timely recognition of the wear condition with marine diesel engine can effectively improve the reliability and economy of the operation of the ship (Xiang 2009).

During the operation of the marine diesel engine, the oil is circulated in various parts of the equipment. Through the oil monitoring, we can learn the changes of lubricating oil indexes and the wear particles of each friction pair. Further, the diesel engine wear state can be qualitatively and quantitatively identified. Spectral analysis and ferrography analysis, which are able to detect concentration of elements and identify the fundamental parameters of ferromagnetic wear particles in lubricating oil, is one of oil monitoring technologies and widely used for diagnosing wear-out fault of diesel engine (Gao et al. 2013).

Due to the complexity of the factors that affect the oil and the result of fault diagnosis is greatly influenced by the oil sampling period and oil change, the traditional "three-line value" method based on statistics is no longer applicable. Therefore, the best way is to excavate the nonlinear relationship behind the data as much as possible through data learning. The development of artificial intelligence algorithm provides more choices for fault diagnosis research (Li et al. 2017). Among them, Self-Organizing Mapping (SOM) network has good self-organization, self-adaptability and robustness. Due to the unsupervised learning method, the SOM network does not need to specify the category of the input vector. It is a kind of recognition network based on small sample training, which is different from the traditional neural network that requires a large number of training samples to ensure the accuracy of classification (Wen 2016). Therefore, SOM networks are widely used in pattern recognition and classification.

However, when inputting a large amount of complex data, the SOM network often suffers from slow convergence and low classification accuracy due to the complexity of the network. The rough set theory can identify and extract the hidden and valuable key data in input information, and remove redundant and invalid data. Yi et al. (2014) used rough set theory to simplify decision rules and removes redundant information, and proposed a fault diagnosis model for the lube oil system of gas turbine based on rough set and SOM neural network. Zhao et al. (2016) designed a neural network

fault diagnosis model based on rough set theory. Although these researches use the rough set theory to reduce the attribute of fault decision table, that is, the two-dimensional decision table is reduced by the longitudinal dimension. However, the other dimension of the two-dimensional decision table is ignored-the data of the horizontal dimension. Therefore, the traditional reduction of input data is not complete, but there are still some irrelevant and redundant data. These data will still be excessive learning in the SOM network, affecting the classification accuracy of the model.

Therefore, this paper makes a two-dimensional reduction of the fault decision table; first, the data reduction of the horizontal dimension is carried out by grey relational analysis, and then the rough set theory is used to reduce the redundant attributes of the longitudinal dimension in the fault decision table. Finally, a diesel engine wear fault diagnosis model based on Grey rough set and SOM neural network is established.

2 TWO DIMENSION REDUCTION OF FAULT DECISION TABLE

For the marine diesel engine, the fault diagnosis process caused by abnormal wear is actually a decision-making process. The basis of decision making can be a predetermined decision table which represents the experience knowledge of managers. It can also be a cumulative decision table which is abstracted and generated according to the summary of historical failure events (Gao et al. 2013). A common fault decision table is shown in Table 1.

Each column in the table represents an attribute, and each row represents an object. $a_1 \ldots a_n$ represent the conditional attributes, and D represent the decision attributes. In practice, there are many fault reporting events and oil monitoring information of diesel engine. Therefore, there are more condition attributes, and the objects in the horizontal dimension also have some irrelevant and redundant data.

Thus, in this paper, two-dimension reduction of two dimensional fault decision table is introduced. Indeed, the data reduction of the horizontal dimension in the fault decision table is carried out by grey relational analysis, and then the rough set theory is used to reduce the redundant attributes of the longitudinal dimension.

Table 1. Diagrammatic sketch of fault decision table.

U	a_1	…	a_k	…	a_n	D
X_1	1	…	2	…	1	0
…	…	…	…	…	…	…
X_i	2	…	1	…	2	1
…	…	…	…	…	…	…
X_m	1	…	1	…	1	0

2.1 *Grey relational analysis method*

Grey system theory is one of the important methods and techniques for studying uncertain systems. And grey relational analysis is a very active branch in the grey system theory, which basic idea is to divide the factors as sequence curve, and then through the similarity degree of geometric shapes to obtain the correlation degree of each factors (Gao et al. 2013). The closer the shape of the curve is, the greater the correlation of the corresponding sequence is determined.

The calculation steps of grey correlation degree are as follows (Liang & Zhang 2009):

Step 1: $X_0 = \{a_0(k) \mid k = 1,2,\cdots n\}$ was determined as a reference sequence, and $X_i = \{a_i(k) \mid k = 1,2,\cdots n\}$ were the comparative sequences. Among them, $i = 1, 2,\ldots, m$, and there are m sequences and n attributes.

Step 2: Calculate the correlation coefficient $\gamma(a_0(k), a_i(k))$ of each attribute relative to the reference sequence.

$$\gamma(a_0(k), a_i(k)) = \frac{\min\limits_{i\in m}\min\limits_{k\in n}\left|a_0(k)-a_i(k)\right| + \rho\max\limits_{i\in m}\max\limits_{k\in n}\left|a_0(k)-a_i(k)\right|}{\left|a_0(k)-a_i(k)\right| + \rho\max\limits_{i\in m}\max\limits_{k\in n}\left|a_0(k)-a_i(k)\right|} \tag{1}$$

where, ρ means resolution coefficient, and the smaller the value is, the greater the differentiation is. Generally, $\rho \in (0,1)$, and when $\rho \le 0.5463$, the discrimination performance is the best. Here ρ is taken as 0.5.

Step 3: Calculating grey relational grade.

$$\gamma(X_0, X_i) = \frac{1}{n}\sum_{k=1}^{n}\gamma(a_0, a_i) \tag{2}$$

In a diesel engine wear fault decision table, if each row is considered as a factor leading to diesel engine fault, then the grey correlation grade of these factors can be sorted from large to small, eliminating some invalid, low correlation and redundant data.

2.2 *Attribute reduction based on information entropy*

2.2.1 *Rough set theory*

Rough set theory is based on an information system IS = (U, A, F), where U = (X_1, X_2,…,X_i,…,X_m) is called the universe and A = (a_1, a_2,…,a_k,…,a_n)

denotes attribute sets . $F = \{f_k : U \to V_k (k \le n)\}$ is the relation sets between U and A, and V_k is the range of a_k (Jia et al. 2016).

Definition 1 Let IS = (U, A, F) be the information system of diesel engine, and V_d is the range of fault state D. The mapping from universe U to fault state is denoted as d: $U \to V_d$, then DIS = (U, A, F, d) is called fault diagnosis information system.

$$R_A = \{(X_i, X_j) \big| f_k(X_i) = f_k(X_j)(a_k \in A)\} \tag{3}$$

$$R_d = \{(X_i, X_j) \big| d(X_i) = d(X_j)\} \tag{4}$$

If $R_A \subset R_d$, DIS is called coordinated diagnosis information system. On the contrary, it is called inconsistent diagnostic information system.

For inconsistent diagnostic information systems, the processing is no longer a simple inclusion relation, but the inclusion degree between sets (Zheng et al. 2014). The concept of inclusion degree is introduced as follows.

$$\begin{aligned} &B \subseteq A, [X_i]_B = \{X_j \big| (X_i, X_j) \in R_B\} \\ &R_B = \{(X_i, X_j) \big| f_k(X_i) = f_k(X_j)(a_k \in B)\} \end{aligned} \tag{5}$$

$$U/R_B = \{[X_i]_B \big| X_i \in U\} = \{X_1, X_1, \cdots, X_s\} \tag{6}$$

$$U/R_d = \{D_1, D_2, \cdots, D_r\} \tag{7}$$

For $X_i \in U(j = 1,2,\cdots,r)$, the inclusion degree is:

$$D(\mathrm{D}_j/[\mathrm{X}_i]_B) = \mathrm{D}_j \cap [\mathrm{X}_i]_B / \big|[\mathrm{X}_i]_B\big| \tag{8}$$

In the formula, $|X|$ denotes the cardinality of the set X.

In order to enhance the anti-interference ability of the rough set model, Ziarko proposed a variable precision rough set model, introducing the classification accuracy $\beta(0.5 < \beta \le 1)$.

Definition 2 Let DIS be an inconsistent diagnostic information system, $B \subseteq A$, precision threshold $\beta \in (0.5,1]$. For $X \subseteq U$,

$$\begin{aligned} \underline{R}_B^{\beta}(X) &= \{\mathrm{X}_i \big| \mathrm{D}(X/[\mathrm{X}_i]_B) \ge \beta\} \\ &= \cup\{[\mathrm{X}_i]_B \big| \mathrm{D}(X/[\mathrm{X}_i]_B) \ge \beta\} \end{aligned} \tag{9}$$

$\underline{R}_B^{\beta}$ is called β lower approximation of X.

Definition 3 Let DIS be an inconsistent diagnostic information system, $B \subseteq A$, precision threshold $\beta \in (0.5,1]$. The β approximate dependence of the fault state D on the parameter set B is defined as:

$$\eta(B, D, \beta) = \left| \bigcup_{j=1}^{r} \underline{R}_B^{\beta}(D_j) \right| \Big/ |U| \tag{10}$$

For inconsistent diagnostic information systems, the classification accuracy is first selected, and the corresponding approximate dependence is calculated. If the evaluation index meets the requirements, the system reduction should be continued.

2.2.2 *Information entropy and conditional entropy*

For diesel engine fault diagnosis, the purpose of attribute reduction is to use fewer parameters to obtain the same diagnosis effect as many parameters, and improve the diagnosis efficiency. Information entropy can measure the uncertainty of knowledge, and reveals that the roughness of knowledge is essentially the description of the information contained in it (Li et al. 2012).

Definition 4 For DIS, the information entropy of parameter set B is:

$$H(B) = -\sum_{i=1}^{s} \frac{|X_i|}{|U|} \log_2 \frac{|X_i|}{|U|} \tag{11}$$

Definition 5 the conditional entropy of the fault state D is relative to the parameter set B:

$$H(D|B) = -\sum_{i=1}^{s} \sum_{j=1}^{r} \frac{|X_i \cap D_j|}{|U|} \log_2 \frac{|X_i \cap D_j|}{|X_i|} \tag{12}$$

2.2.3 *Heuristic attribute reduction algorithm*

Attribute reduction is one of the key problems in rough set theory. Searching for all reductions or optimal reductions is proved to be a NP (nondeterministic polynomial) complete problem. Therefore, heuristic algorithms are usually used to search for optimal reductions (Zhang et al. 2009).

Definition 6 the mutual information of fault status D and parameter set B is defined as:

$$I(D;B) = H(D) - H(D|B) \tag{13}$$

Mutual information is used to measure the amount of information obtained from the parameter set B of the fault state D.

Definition 7 the importance of any parameter $a \in A - B$ relative to the fault status D is defined as:

$$SGF(a, B, D) = H(D|B) - H(D|B \cup \{a\}) \tag{14}$$

The attribute importance measures the amount of information about the fault state D from the parameters $\{a\}$ under the condition of the known parameter set B.

In order to reduce the parameter redundancy in the longitudinal dimension of the fault decision table, a forward heuristic search algorithm is proposed for attribute reduction. From the rough set

theory, the relative kernel of any decision information system is unique. Therefore, the relative core can be used as the starting point of searching the optimal reduction (Tian et al. 2014). The algorithm steps are as follows:

Step 1: Calculate mutual information $I(D;A)$ between attribute set A and fault state D in DIS.

Step 2: Calculate the core of the attribute set A relative to the fault state D. If relative core is $C = \phi$, then $I(D;C) = 0$.

Step 3: Let B = C and repeat 1) to 3) for the parameter subset A–B.

1. If $I(D;B) = I(D;A)$, then turn to step 4.
2. For each parameter $a \in A - B$, calculate the importance $SGF(a,B,D)$ of the attributes.
3. Select the attribute with the most important attribute value, which is denoted as p. Let B: = B $\cup$ {p}.

Step 4: Output reduction parameter set B.

3 FAULT DIAGNOSIS MODEL OF DIESEL ENGINE WEAR BASED ON SOM NETWORK

3.1 *Self-organizing map network*

SOM network is a self-organizing, unsupervised learning, self-learning neural network composed of fully connected neuron arrays. The typical SOM network structure is shown in Figure 1, which consists of an input layer and a competition layer (also known as an output layer). The number of neurons in the input layer is n, and the competition layer is a two-dimensional plane array composed of $a \times b$ neurons. Each input neuron is connected to all the neurons in the two-dimensional plane array (Xu et al. 2014). The training process of SOM network is to adjust the weights of network nodes continuously, so that different input types correspond to different neurons in two-dimensional plane array.

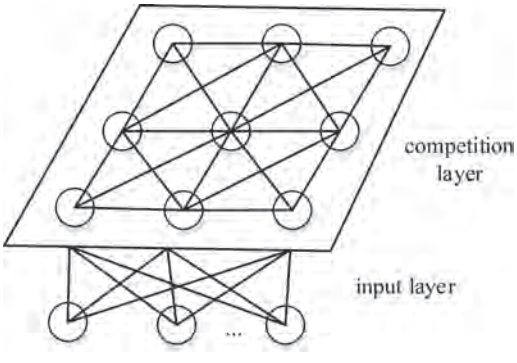


Figure 1. Network model of SOM.

The learning steps of the SOM network are as follows (Zhang et al. 2017):

Step 1: Network initialization. The random number is used as the initial value between the input neuron and the output neuron to connect weights.

Step 2: Accept input samples. The matrix $U = X1, X2, \cdots, Xm$ composed of the sample characteristic parameters is input to the SOM network.

Step 3: Calculate the Euclidean distance between the input vector and the connection weight vector. The neurons with the least Euclidean distance are determined, and they are denoted as the winning neuron O.

The Euclidean distance between the i-th input vector and the j-th neuron in the mapping layer is

$$d_j = \sqrt{\sum_{i=1}^{n} (X_i - \omega_{ij})^2} \tag{15}$$

where, ω_{ij} is the weight between the i-th neuron of the input layer and the j-th neuron of the mapping layer.

Step 4: Adjustment learning of the value. According to the formula (16), the weights between the winning neuron O and the neighboring neurons are corrected.

$$\Delta\omega_{ij} = \eta h(j, j^*)(X_i - \omega_{ij}) \tag{16}$$

where, η is the constant in the range of (0, 1).

$$h(j, j^*) = \exp\left(-\frac{|j - j^*|^2}{\sigma^2} \right) \tag{17}$$

In the formula, σ^2 decreases with learning and narrows the range of $h(j, j^*)$ from width to width. The weight adjustment is changed from coarse to fine. Ensure the accuracy of classification.

Step 5: stopping criterion. Determine whether the expected requirements, if achieved, then end. Otherwise, go back to step 2 and proceed to the next round of learning.

3.2 *The process of diesel engine wear fault diagnosis model*

By using grey relational analysis and rough set theory, the data of horizontal and vertical dimensions in a fault decision table are reduced. Therefore, the validity of the input data to SOM network is greatly improved, and the classification accuracy of the network is improved. The diesel engine wear fault diagnosis model based on Grey rough set and SOM network is shown in Figure 2.

The concrete steps are as follows:

Step 1: Initial fault decision table. The Ferrography and spectral analysis data of diesel engine oil monitoring information, as well as multi group sample data (Universe), form the initial fault decision table.

Step 2: The grey correlation grade of each row data is calculated and the data reduction of horizontal dimension is carried out.

Step 3: Discretization of continuous attributes and data reduction in longitudinal dimension.

Step 4: Before the analysis of the rough set method, the continuous attributes of the horizontal reduction are discretized by using the equal frequency binning. Then, attribute sets are reduced by using information entropy.

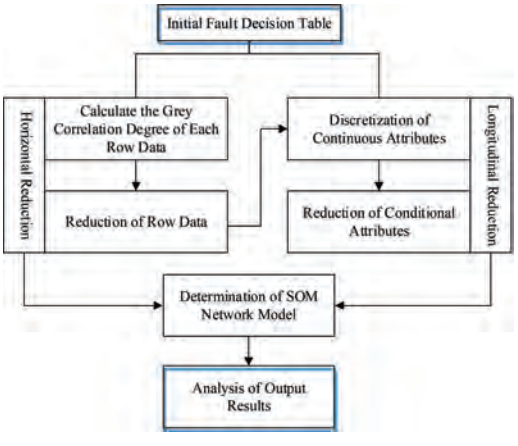


Figure 2. Diesel engine wear fault diagnosis model based on Grey rough set and SOM network.

Step 5: Determination of SOM network model. The number of neurons in the input layer is the number of attributes set B after reduction.

Step 6: Analysis of fault diagnosis results. The evaluation index of SOM network is defined as the classification accuracy, such as formula (18).

$$R(\%) = \frac{y_n}{y_m} \cdot \%$$ (18)

Where, y_n is the number of samples for correct classification and y_m is the total number of samples.

4 CASE EXPERIMENT

The oil spectrum and Ferrography data obtained in this paper are derived from a marine diesel engine. Rated power of marine diesel engine is 1760 KW and engine speed is 1800 r/min. There are 92 sets of oil monitoring data. Irregular sampling was carried out from 2013 to 2015, and the sampling position was the main oil duct of marine diesel engine. Oil is changed out regularly. There are 8 kinds of elements (Unit: ppm) in oil spectral analysis, including Cu, Fe, Cr, Ba, Zn, Si, Al and Pb. The oil Ferrography data is derived from Quantitative Ferrography Analysis, which contains the concentration value of large wear debris Dl (> 5 μm) and small wear debris Ds (1–2 μm) in the oil samples. In order to better characterize the wear conditions of diesel engines, we introduce the Total wear Dls = Dl + Ds; and Wear severity index IS = Dl^2-Ds^2. The data set (fault decision table) is shown in Table 2. Among them, the fault states include 3 types, namely, well condition (50 samples), degradation (21 samples),

Table 2. Oil monitoring data set.

U	Cu	Fe	Cr	Ba	Zn	Si	Al	Pb	Dl	Ds	Dls	IS	Wear-out status
X1	0.9	7.5	0.4	0.9	2.8	11.0	1.9	1.2	28.4	11.6	40	672	Well condition
X2	5.6	8	1.5	0.7	3.1	9.8	1.4	1.3	32.1	16.6	48.7	754.9	Well condition
X3	0.5	9.6	0.4	1.6	2.5	15.9	2	0.5	38.3	15.3	53.6	1171.1	Well condition
…	…	…	…	…	…	…	…	…	…	…	…	…	…
X50	0	3.9	0.5	0.6	0.3	21.1	1.8	0	33.5	20.9	54.4	685.44	Well condition
X51	13.2	298	3.3	0.5	0.4	40.6	23.9	5.8	96.9	88.8	185.7	1504.17	degradation
X52	24.8	296.7	2.4	0.7	2.9	30.4	1.4	8.2	143.5	106.5	250	9250	degradation
…	…	…	…	…	…	…	…	…	…	…	…	…	…
X71	0.8	489	55.4	0.4	3.5	40.6	1.6	2.9	168.3	142.7	311	7961.6	degradation
X72	1.9	221.5	0.3	0.9	2.5	34.3	17.1	1.8	98.5	132.7	231.2	–7907.04	fault
X73	1.3	186.3	0	1.0	0.4	40.6	15.8	1.3	69.2	96.1	165.3	-4446.57	fault
…	…	…	…	…	…	…	…	…	…	…	…	…	…
X91	76.2	100.6	2.1	0.5	1.5	39.1	17.8	24.3	96.9	88.8	185.7	1504.17	fault
X92	100.6	78.6	1.6	0.4	3.0	30.3	29.6	36.4	96.9	62.3	159.2	5508.32	fault

and fault (21 samples). The judgment for the condition of diesel engine is given by experts.

4.1 *Grey relational analysis*

Sample ID represents the cycle of marine diesel engine oil monitoring. Therefore, the oil sample with the first fault is selected as the reference sequence. It is considered that the sequence of grey relation less than 0.85 cannot reflect the wear fault state of diesel engine effectively. Therefore, these invalid redundant data samples are deleted. In this study, samples of ID 2, 14, 17, 29, 43, 55, 63, 67, 71, 76, 80, 82, 85 and 90 were excluded from the fault decision table.

4.2 *Attribute reduction*

Firstly, the method of equal frequency binning is used to discretize the data after horizontal reduction. Due to $R_A \not\subset R_d$, the fault decision table shown in Table 2 is inconsistent diagnostic information system. At the same time, when the classification accuracy β is different, the classification performance of the system is not the same. We select different values of β, and their classification performance is shown in Table 3.

Therefore, this paper chooses the precision threshold $\beta = 0.8$, and uses the heuristic algorithm of 2.2.3 section to reduce the attributes. Finally, the attribute reduction result is B = {Fe, Al, Cu, Pb, Dl, IS}.

4.3 *Establishment of SOM network model*

The parameters of the attribute set B are normalized and then used as input to the SOM network. The competition layer of SOM network is set to $6 \times 6 = 36$ neurons, and the number of training iterations is defined as 200. There are 78 groups of samples after horizontal data reduction. 45 groups were selected as training sets (including 25 well condition, 10 degradation and 10 fault), and the remaining 33 (including 20 well condition, 7 degradation and 6 fault) were the test sets.

The Matlab toolbox was used to establish SOM network structure and the results as shown in Figure 3, Figure 4 and Table 4. Hexagons with numbers represent the winning neurons in Figure 3; the number of training samples represented by the neuron is described numerically. The hexagons with numbers in Figure 4 represent neurons, and the straight line in the middle represents the connection of neurons. The distance between neurons was calculated by Euclidean distance formula, and the distance between the neurons was reflected by the hexagon color background of the connected neurons. The darker the color is, the farther the distance between neurons is. That is, the greater the difference between the two neurons.

Table 4 illustrates the neurons excited by each wear state. The distribution of neurons in the whole competition layer can be explained by Table 4 and Figure 4.

Table 3. Comparison table of classification performance for the samples.

β	1.0	0.9	0.8	0.7
$\eta(A, D, \beta)$	0.1563	0.5721	1.000	0.8109

4.4 *Results analysis*

The diagnosis results of diesel engine wear fault based on SOM network are shown in Table 5. As

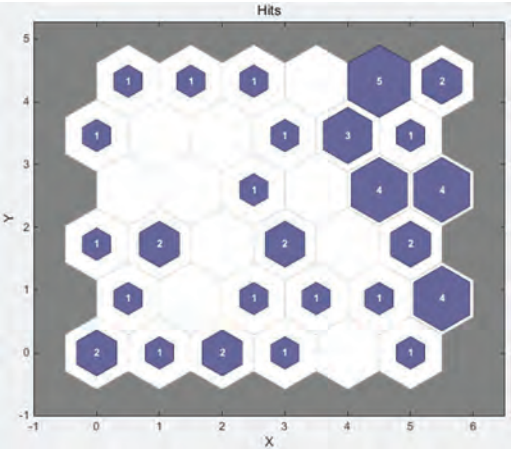


Figure 3. Winning neuron map.

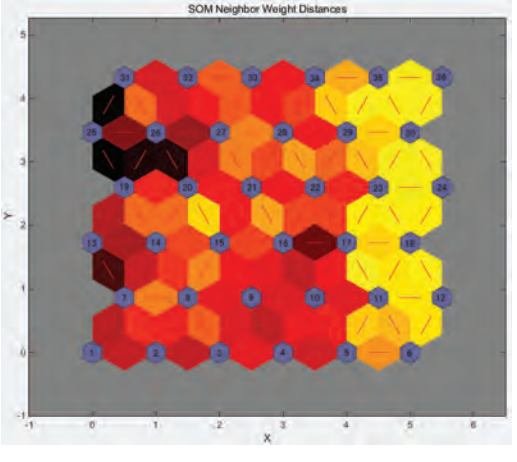


Figure 4. Distance between the neurons.

Table 4. Competitive layer neurons corresponding to wear states.

Wear-out status	Serial number	Index of excited neurons
Well condition	1	23, 30, 36, 24, 18, 12, 6, 11, 29, 35, 29
Degradation	2	3,13,2,1,14,7,31,25
Fault	3	4,9,10,32,33,28,21,3,16

Table 5. Diagnostic result statistics of SOM network model.

Sample type	Classification accuracy
Well condition	100%
Degradation	85.71%
Fault	100
Total	96.97%

Table 6. Classification accuracy of different models.

Model	Total classification accuracy
GRS-SOM	96.97%
RS-SOM	87.88%
PCA-SOM	75.76%
GRPCA-SOM	78.79%

shown in Table 5, the accuracy of the whole test set is 96.97%, and the diagnosis process is feasible and the results are satisfactory.

To illustrate the advantages of the model, we compare it with Rough set-SOM (RS-SOM) (without horizontal data reduction), Principal component analysis-SOM (PCA-SOM), and Grey relational principal component analysis-SOM (GRPCA-SOM) models. Using the same training set and test set, the results are shown in Table 6.

Compared with RS-SOM, PCA-SOM and GRPCA-SOM methods, GRS-SOM has higher classification accuracy. Therefore, it is feasible to use the SOM network model based on Grey rough set to diagnose the wear fault of diesel engine. In addition, it can solve the problem that the wear state of diesel engine is difficult to identify.

5 CONCLUSION

This paper presents a fault diagnosis method for marine diesel engine wear based on Grey rough set and SOM neural network. The horizontal and longitudinal dimensions of the two-dimensional fault decision table are reduced by using grey correlation analysis and rough set reduction theory respectively. The reduced data is used as the input of the SOM network, and the fault diagnosis model is established to identify the wear state of the marine diesel engine.

The grey relational analysis and rough set theory are used to reduce the input data, and the redundant data and attributes are completely removed. While simplifying the network mechanism, the accuracy of fault diagnosis is improved.

Compared with the traditional RS-SOM, PCA-SOM and GRPCA-SOM model, the fault diagnosis model based on Grey rough set and SOM neural network has smaller error and higher classification accuracy.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (Grant No.71501007 & 71672006). The study is also sponsored by the Technical Research Foundation and The Graduate Student Education & Development Foundation of Beihang University.

REFERENCES

Gao, S.Z., Wang, J.S. & Zhao, N. (2013) Fault Diagnosis Method of Polymerization Kettle Equipment Based on Rough Sets and BP Neural Network. *Mathematical Problems in Engineering*, 1–8.

Jia, X., Shang, L., Zhou, B. & Yao, Y. (2016) Generalized attribute reduct in rough set theory. *Knowledge-Based Systems,* 91, 204–218.

Jones, N.B. & LI, Y.H. (2000) A review of condition monitoring and fault diagnosis for diesel engines. *Lubrication Science,* 6, 267–291.

Li, L., Chang, W., Zhou, S. & Xiao, Y. (2017) An identification and prediction model of wear-out fault based on oil monitoring data using PSO-SVM method. *Reliability and Maintainability Symposium.*

Li, Z., Yan, X., Guo, Z., Zhang, Y., Yuan, C. & Peng, Z. (2012) Condition Monitoring and Fault Diagnosis for Marine Diesel Engines using Information Fusion Techniques. *Electronics & Electrical Engineering,* 123.

Liang, N. & Zhang, J.G. (2009) Combinational Forecasting Model Based on Neural Network and Gray Relational Analysis. *Natural Science Journal of Hainan University*.

Tian, J., Wang, Q., Bing, Y. & Dan, Y. (2014) A rough set algorithm for attribute reduction via mutual information and conditional entropy. *International Conference on Fuzzy Systems and Knowledge Discovery.*

Wen-Ying, L.I. (2016) Classification of coal gas permeability based on SOM artificial neural network. *Inner Mongolia Coal Economy*.

Xiang, Z. (2009) The Application of Oil Monitoring Techniques in the Failure Diagnosis of Ship Diesel Engine. *Lubrication Engineering*.

Xu, X., Yan, X., Zhao, J., Sheng, C., Yuan, C. & Ma, D. (2014) Remote fault diagnostic model for

tribological systems in marine diesel engine with two-level self-organizing map network. *Prognostics and System Health Management Conference.*

Yi, S., Zhao, N., Li, S. & Xu, Z. (2014) A study on fault diagnostic method for the lube oil system of gas turbine based on rough sets theory. *International Conference on Fuzzy Systems and Knowledge Discovery.*

Zhang-Yan, X.U., Hou, W., Song, W. & Yang, B.R. (2009) Efficient Heuristic Attribute Reduction Algorithm Based on Information Entropy. *Journal of Chinese Computer Systems,* 30, 1805–1810.

Zhang, Y., Tang, B., Han, Y. & Deng, L. (2017) Bearing performance degradation assessment based on time-frequency code features and SOM network. *Measurement Science & Technology,* 28.

zhao, R., Li, C. & Tian, X. (2016) A novel industrial multimedia: rough set based fault diagnosis system used in CNC grinding machine. *Multimedia Tools & Applications*, 1–14.

Zheng, K., Hu, J., Zhan, Z., Ma, J. & Qi, J. (2014) An enhancement for heuristic attribute reduction algorithm in rough set. *Expert Systems with Applications,* 41, 6748–6754.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Anomaly indicators for Kaplan turbine components based on patterns of normal behavior

M.A. Sanz-Bobi
Institute for Research in Technology, Comillas Pontifical University, Madrid, Spain

T. Welte
SINTEF Energy Research, Trondheim, Norway

L. Eilertsen
GLITRE Energy, Norway

ABSTRACT: This paper describes and proposes some indicators for continuous monitoring of anomalous conditions in the hydraulic system of a Kaplan turbine using SCADA data. The indicators are based on significant deviations between the estimated values for key variables describing the current working conditions of the components at the plant, and those actually observed. This monitoring strategy requires models describing the expected values for variables through the whole range of possible working conditions of the monitored components. These models are normal behavior models able to characterize the typical relationships between a set of variables used as inputs to the models and the corresponding output of a target variable whose expected value has to be predicted. The criteria to select the variables to use in the models are based on the physical working principles of the component. The paper is focused on models of normal behavior applied to a real case of condition monitoring of a Kaplan turbine regulating mechanism.

1 INTRODUCTION

Hydropower is the leading renewable global source for electricity generation supplying 71% of all renewable electricity and reaching 1,064 GW of installed capacity in 2016 (WEC, 2017). It generated 16.4% of the electricity produced in the world from all sources. Hydropower is the most flexible and consistent of all the renewable energy resources, capable of meeting base load electricity requirements, as well as with pumped storage technology, meeting peak and unexpected demand due to shortages or the use of intermittent power sources. Also, hydroelectricity is a source of electrical energy coming from water that is clean and safe.

A large number of data is logged in the SCADA system (Supervisory Control And Data Acquisition) in hydropower plants, but the current status in Norway and Sweden is that SCADA data—apart for their use to control the plant—is not much used for other purposes, such as condition monitoring and maintenance planning. Thus, there is a large potential for using SCADA data for these new purposes. This may contribute to increased availability and energy production due to prevention of failures and shut downs.

The identification of possible failure modes in a hydropower plant (Topliceanu, 2016) is one of the key points in order to identify how failures could be detected in an early state. The analysis of the causes and effects of these failure modes can suggest the variables that can be useful for the detection of abnormal behaviors or anomalies (Chandola, 2009). Several references can be found in scientific literature proposing different methods for anomaly detection, and, in general, fault detection in industrial processes (Garcia Matyos, 2013) based on values of some variables measured in real time. One area in hydropower plants with an important research activity is related with the vibrational analysis focused on some key components (Mohanta, 2017), also the health condition of the components observed through several types of measurements is the goal of other studies such as those referred to in (Jamil, 2013) and (Selak, 2014).

In this paper, the hydraulic system of a Kaplan turbine was identified as a target of analysis and in particular the detection of a possible oil leakage in the system. This analysis is part of the results obtained in the research project MonitorX – "Optimal utilization of hydropower asset lifetime by monitoring of technical condition

and risk". MonitorX is a joint industry project initiated and led by Energi Norge (Energy Norway—the Norwegian electricity industry association) in cooperation with Energiforsk (the Swedish Energy Research Centre), more than 20 Norwegian and Swedish power companies, a number of equipment manufacturers and service providers, and the research institutions Comillas Pontifical University, SINTEF Energy Research and the Norwegian University of Science and Technology as R&D partners. The project is financially supported by the Research Council of Norway.

The aim of the MonitorX project is to develop models and algorithms for condition monitoring and the detection of faults in hydropower equipment. The main focus in the project is on models based on machine learning and artificial intelligence. The project is case-driven, and several relevant cases have been identified in the beginning of the project, whereof the case related to monitoring of the Kaplan turbine regulating mechanism and corresponding hydraulic system was considered as relevant for further work. Since several components and parts of the system are difficult to inspect, models that can be used to monitor the system condition and detect failures are valuable. Furthermore, oil leakage from the hydraulic system may cause environmental damage, especially when oil leaks into the river.

Usually, no separate condition monitoring systems are installed in power stations to surveil the condition of the regulating mechanism. The data that normally is available is from the SCADA system of the plant that usually presents one hour average values. Thus, one of the aims of the presented case is to study if such type of data is useful for modelling the normal behavior of hydropower components and detecting with these models anomalies that are related to faults.

The paper is organized in the following sections. Section 2 describes the method and objectives used for the creation of normal behavior models and detection of anomalies. Section 3 presents a description of the hydraulic system of the hydraulic power plant analyzed. Section 4 includes the description and development of normal behavior models used as references for detection of anomalies. Section 5 presents several cases about how the normal behavior models can be used as reference patterns for the detection of anomalies. Finally, section 6 summarizes some conclusions of the analysis developed throughout the paper.

2 METHOD AND OBJECTIVES

This section describes the main steps of the process to build anomaly indicators for detection of abnormal behavior in some functional characteristics of components in a hydropower plant. These indicators are based on patterns previously obtained from observing the typical normal behavior of the monitored components. The following sequential steps are required in order to detect anomalies based on an estimation for these indicators:

a. Selection of a data training set for learning the typical normal behavior of the component. This includes data selection and filtering, removing of outliers and treatment of missing measurements.
b. Identification of failure modes that could be detected with the variables available in the SCADA system, and selection of variables. Information available in a Failure Modes and Effects Analysis (FMEA) may help to select relevant failure modes and variables. The variables will be used for the characterization of normal behavior patterns developed in the next step.
c. Building of normal behavior patterns of a component described through variables collected in real-time from the hydropower plant. The cases studied in this paper are based on data samples collected every hour. The patterns were built using multi-layer perceptrons (Bishop, 1995), (Bishop, 2006). This technique is supervised requiring previous knowledge of behavior considered as normal and covering all the typical working conditions of the plant. A good selection of this behavior, considered as normal, is crucial in this method because the normal behavior will be learnt by the models as a reference to watch when new information is coming from the power plant.
d. Estimation of anomaly indicators. Once the previous steps are completed, the indicators of anomalies can be estimated. Its objective is to warn about data collected from the hydropower plant that do not correspond to the expected behavior by the reference patterns. The evolution of the values of the anomaly indicators over time will suggest whether or not it is necessary to pay attention to the components monitored from the point of view of scheduled maintenance and operation.

Sections 4 and 5 will describe details about each of the previous steps with examples demonstrating their use.

3 SYSTEM ANALYSED

The cases analyzed in the paper are from Embretsfoss 4, which is a hydropower plant using a Kaplan turbine for the production of electric energy. The Kaplan turbine is a propeller type turbine

controlled by the operation of the turbine runner blades (turbine blades) and the wicket gates (guide vanes). See illustration in Figure 1. A Kaplan turbine is a typical run-of-river turbine, which can be operated at different flows and at varying head. For each head and flow, there is a given ideal combination of the wicket gate and runner blade position to ensure the best efficiency of the turbine.

A turbine regulator controls and operates the turbine. Based on information about head and flow it uses predefined combination curves for the runner and wicket gate. The regulator controls the turbine by adjusting the blade and wicket gate positions with a correlated movement between the two. The acting mechanism for the wicket gates and runner blades are based on high-pressure hydraulics where an HPU (high-pressure unit) and an accumulator bank provide high-pressure oil for actuation of hydraulics servomotors.

3.1 *The high-pressure hydraulic system*

The turbine regulator controls the wicket gates and the runner blades by the use of a high-pressure hydraulic system which consist of the following main components:

- Turbine governor oil sump tank with oil pumps (HPU – High Pressure Unit)
- Pressure accumulator banks. One bank for runner blades and one bank for the wicket gates
- Hydraulic oil cooling/heating system
- Wicket gate control system
- Runner blade control system

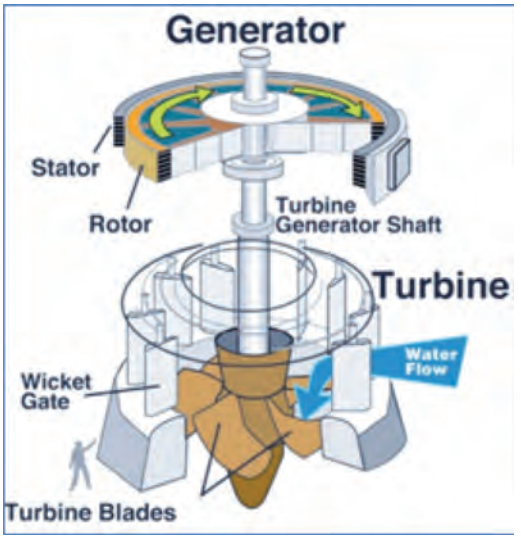


Figure 1. Illustration of the Kaplan turbine (Courtesy of Wikipedia).

- Quick stop/Emergency stop system
- Oil system for runner hub. The runner hub is the lowermost part of the runner. The cone part just below the runner blades. See Figure 1.

For a simplified view of the high-pressure hydraulics system, see Figure 2. The HPU is located at the turbine floor and it supplies the wicket gate and runner blade control system with high-pressure oil. The main components of the HPU are the oil reservoir, the oil pumps, valves, filters and coolers. In addition to supply oil to the control system, the HPU is "charging" in total five accumulator banks. The accumulator system is a safety system designed to handle a predefined number of safe shutdown cycles, in case of malfunction of the HPU system or a blackout of the station. The HPU have systems for monitoring the oil level, temperature and water-in-oil content. To prevent the pollution of the oil, each of the HPU pumps are equipped with a filter system.

For maintenance reasons, the oil reservoir is designed to be big enough for storage of all the oil in the system. However, during operation, the oil is in the different components of the hydraulic system hence only a limited amount of oil is contained in the reservoir. A minimum level is however required in the reservoir for avoiding dry running of the HPU pumps.

The hydraulics system has an oil cooling (and heating) system. The cooling system cools the oil during operation and the heating system heats the oil during standstill.

The wicket control system controls the wicket gates by the use of two hydraulic servos (cylinders). The servos actuate the control ring, which again provides the open/close movement on the wicket gates. When the control ring, seen from the top, turns clockwise, the wicket gates close.

The runner blade control system controls the position of the runner blades by the use of a servo actuator located in the runner hub. The actuator high-pressure oil supply/return is routed through the center of the turbine shaft via the oil supply head located at the top of the shaft.

The system is equipped with a system for safe emergency stopping of the turbine. This can be activated by a manual activation of the emergency stop or if the turbine is speeding and the overspeed trip valve is activated.

The turbine hub is filled with oil and has an oil pressure that is slightly higher than the surrounding water pressure. In the case of runner blade sealing degradation, this pressure prevents water from entering the hub. The oil pressure in the hub is a static pressure created by the elevated location of the hub oil tank (see Figure 2). The oil in the hub oil tank is pumped up from the HPU oil reservoir. The hub oil

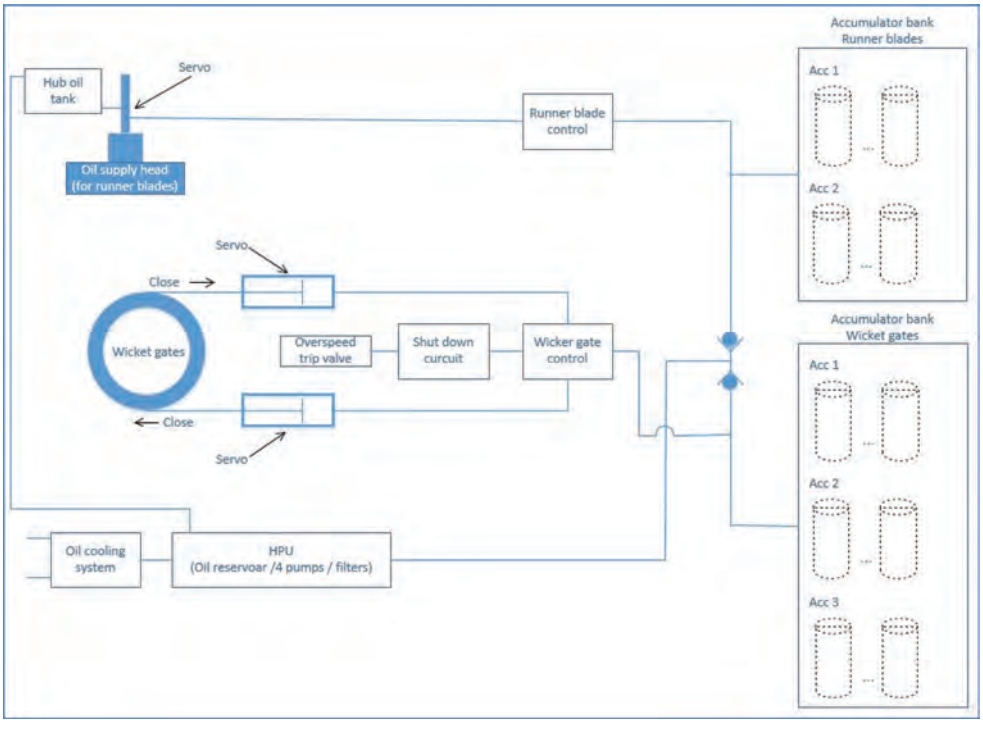


Figure 2. Simplified view of the hydraulic system.

tank and the hub oil are not a part of the high-pressure circuit, but a leakage in the runner blade servo will influence the oil level in the hub oil tank and will eventually sound an alarm or stop signal.

4 MODELS OF NORMAL BEHAVIOUR

An industrial component or system can be stressed due to normal operation, extraordinary operation and extreme environmental conditions or to a combination of all. Over time, these facts along with ageing factors can produce different ranges of typical values observed in measured variables even when the functional objectives of the component or system as expected have been reached (Sanz-Bobi, 2011). However, when a component has been stressed or overloaded over time, an increasing risk of occurrence of a failure is probable. For this reason, it is important to characterize the normal behavior expected for an industrial component or system when it is performing its function under several typical working conditions, because any deviation with respect to this behavior could alert about the presence of an incipient failure. The sooner this is detected, the sooner it is possible to mitigate the effect of a failure.

This section describes real examples of normal behavior models. These models are able to characterize the typical dynamical evolution of variables when the component is working under different operating conditions without symptoms of failure or stress.

In particular, the models developed and presented as an example in this paper, are based on information collected in real-time from a hydraulic power plant located in Norway. The models developed use neural networks based on multi-layer perceptrons (Bishop, 1995; Kruse, 2013) because this is a method able to approximate non-linear relationships among variables.

An basic model to characterize the normal behavior of the hydraulic power plant can be expressed by function *f* in Equation 1

$$P = f(GVP, WF, HW - TW) \tag{1}$$

where:

P: Power generated by the power plant in MW

GVP: Guide Vane Position in percentage

WF: Water flow through the turbine in m^3/s

HW-TW: Difference between headwater and tailwater levels in m.

Equation 1 tries to predict the power generated as function of the values of the main variables contributing to the power generation.

In order to build a normal behavior model characterizing the function *f* in Equation 1, a training set was selected covering different seasonal conditions from January 1 to August 20, 2015. The data set is based on hourly values for the variables considered. The model was developed with a multi-layer perceptron based on one hidden layer containing 20 neurons and using the Levenberg–Marquardt algorithm for learning. The model obtained is very good, as it can be observed in Figure 3, where the estimated values for the power generated and the real values observed are almost identical. The mean value of their difference (error of the trained model) is 0.0012 MW and the standard deviation is 0.067 MW. This error is distributed according to a normal distribution with narrow shape.

An interesting family of models will be presented in the following for the characterization of the normal relationships that exist between the tank oil level of the turbine regulator and variables observed in different components of the turbine regulator that uses this oil. It is important to monitor that the oil in the tank is at the expected level, because if this is not the case, a possible leakage could be present.

The first normal behavior model of the family that was tested is described in Equation 2 using the function *f*1.

$$OTL = f1(P, OTT, A1TR) \tag{2}$$

where:

OTL: Oil tank level in the HPU in percentage

P: Power generated by the power plant in MW

OTT: Oil tank temperature in ºC

A1TR: Oil level in accumulator 1 of the turbine runner.

Equation 2 tries to predict the oil tank level in the HPU of the turbine regulator knowing the working conditions of the plant, the level of one oil accumulator of the turbine runner and the temperature of the tank oil.

The model for *f*1 was obtained with a similar architecture as for *f* in Equation 1. Also, the same dates as in the previous case were used to obtain the samples of the training set. The model obtained is good, which can be observed in Figure 4 where the estimated values for the oil tank level and the real observed are very close. The oil tank level is measured in percentage (%). The mean value of their difference (error of the trained model) is 0.0007% and the standard deviation 0.0644%. This error is distributed according to a normal distribution shape.

The hydraulic power plant studied has another similar accumulator given the number 2 in the turbine runner. A normal behavior model was fitted and the results obtained were very similar to those obtained for accumulator 1 of the turbine runner.

Other important components in the turbine regulator of the hydraulic power are three oil accumulators for the guide vanes. These are very important for the correct regulation of the hydraulic turbine. Three models, one considering each of the oil accumulators, were developed such as in Equation 2. For simplicity, only one of them will be presented. Equation 3 describes it using function *f*2.

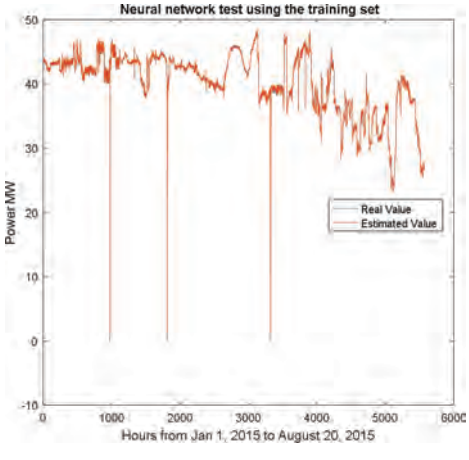


Figure 3. Estimated value for power generated predicted by the normal behavior model and the real value observed for the training set using the guide vane position, the flow through the turbine and the difference between headwater and tailwater level.

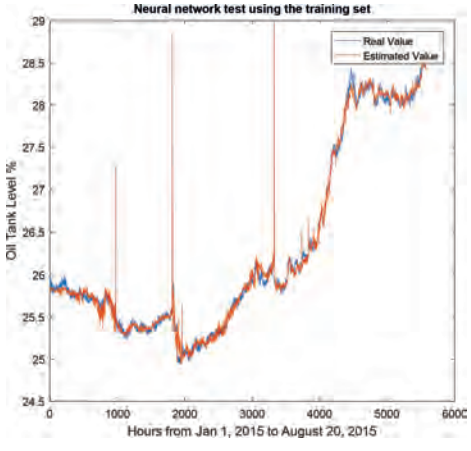


Figure 4. Estimated value for oil tank level in percentage predicted by the normal behavior model and the real value observed for the training set using as inputs the power generated, the oil tank temperature and the oil level in accumulator 1 of the turbine runner.

$$OTL = f2(P, OTT, A3GV) \quad (3)$$

where:

OTL: Oil tank level in percentage

P: Power generated by the power plant in MW

OTT: Oil tank temperature in ºC

A3GV: Oil level in the accumulator 3 for the guide vanes.

Equation 3 tries to predict the oil tank level in the turbine regulator knowing the working conditions of the plant, the level of the oil accumulator 3 for the guide vanes and the temperature of the tank oil.

The model for *f*2 was obtained following the same method as in the previous cases described. However, the main difference was that the data used in the training set covered the period from April 9, 2016 to October 13, 2016, because before that period some measurements of the oil accumulators of the guide vanes were not collected correctly. In any case, more than half of this period overlaps with the one used for obtaining *f* and *f*1. The model resulting for *f*3 obtained is good, as it can be observed in Figure 5 where the estimated values for the oil tank level (in percentage %) and the real observed values (in percentage too) are very close. The mean value of their difference (error of the trained model) is 0.0022% and the standard deviation 0.09%. This error is distributed according to a normal distribution shape.

Good results were also obtained for the two models that are similar to the one in Equation 3, where the variable oil level in accumulator 3 has been changed to the oil levels in the corresponding accumulators with numbers 1 and 2, respectively.

5 ANOMALY DETECTION BASED ON PATTERNS OF NORMAL BEHAVIOUR

Once a normal behavior model has been elaborated, it can be used in real time with real-time values from the required inputs. The output from the model can then be compared with the corresponding real measured output variable. The prediction will correspond to the expected value for normal behavior under the current working condition. Any incipient failure will produce a deviation between the expected value and the real value measured of the monitored variable. This section presents how the normal behavior models obtained in the previous section respond to new inputs of data collected after the training set dates. This will allow for the discovery of abnormal behavior different to the one expected.

Model *f* was used with data not contained in the training set, covering the period from November 25, 2015 to May 31, 2017. Figure 6 shows the results obtained by the model. The real behavior observed is very near to the predicted one and this confirms that the behavior observed in this new period of time is similar to the previous one in the training set. No abnormal behavior was detected in the power generation according to model *f*. The mean value of their difference (error) is –0.017 MW and the standard deviation is 0.7 MW. Both are higher than what was obtained for the training data set, but the prediction is still reasonable. Also, this error is distributed according to a normal distribution shape.

Furthermore, model *f*1 was used with data not contained in the training set, covering the period from November 25, 2015 to May 31, 2017. Figure 7 shows the results obtained from the model. The real behavior observed is near to the predicted

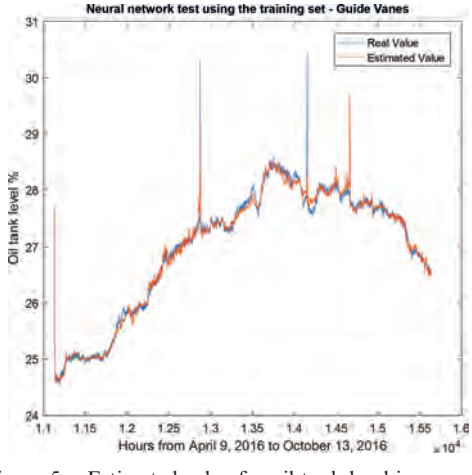


Figure 5. Estimated value for oil tank level in percentage predicted by the normal behavior model and the real value observed for the training set using the power generated, the oil tank temperature and the oil level in accumulator 3 for the guide vanes.

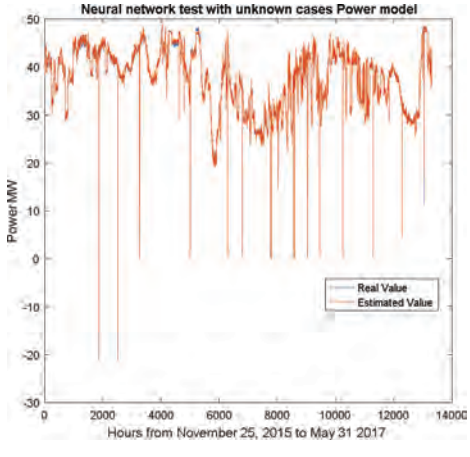


Figure 6. Estimated value for power generated predicted by the normal behavior model and the real value observed for the testing data set using the guide vane position, the flow through the turbine and the difference between the headwater and tailwater levels.

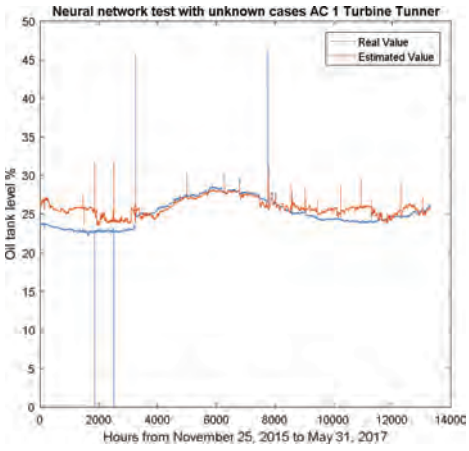


Figure 7. Estimated value for oil tank level in percentage predicted by the normal behavior model and the real value observed for the testing data set using as inputs the power generated, the oil tank temperature and the oil level in accumulator 1 of the turbine runner.

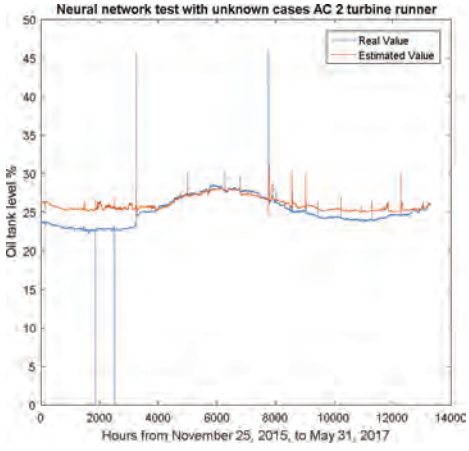


Figure 8. Estimated value for oil tank level in percentage predicted by the normal behavior model and the real value observed for the testing data set using as inputs the power generated, the oil tank temperature and the oil level in accumulator 2 of the turbine runner.

one in some cases in the central part of the figure and different in the rest of the period studied. This means that the behavior observed in the training data set is different from the one observed in the new test data set at some periods. An abnormal behavior was detected in the relationships between the output and input variables of this model for the test period. Once this was detected, it became necessary to investigate the cause.

The cause of abnormal behavior detected that is breaking the relationship modelled by *f*1 can be any of the variables used in this model. The variable power generated cannot be the cause due to the test carried out in model *f* and presented in Figure 6 which confirms that no abnormal generation of power exists. The rest of the variables could be candidates to be anomalous and they are related with the oil tank (level and temperature) and the accumulator 1 level of the turbine runner.

A model similar to the one presented in Equation 2 was developed replacing the variable A1TR (Oil level in accumulator 1 of the turbine runner) by another equivalent model, but measuring the oil level in accumulator 2 of the turbine runner. The model obtained was very good and similar to that presented in Figure 4. This model was checked with data not contained in the training set, covering the period from November 25, 2015 to May 31, 2017 as for accumulator 1 of the turbine runner.

The result is presented in Figure 8. The profile between predicted and real oil tank levels are almost the same in Figures 7 and 8. The same broken relationship is shown between the oil tank level and the oil level in accumulators 1 and 2 of the turbine runner. This induces the thought that it is not probable that the problem of the abnormal behavior observed is due to some anomaly in both turbine runner accumulators at the same time and it is therefore convenient to closely monitor the oil tank level.

In this way, model *f*2 was also tested with data covering the period from October 14, 2016 to May 31, 2017. This period includes data from sample 8000 till the end of the graphics in both Figures 7 and 8. Figure 9 presents the results of the application of model *f*2 to the data set mentioned. The discrepancy between predicted and real values for the oil tank level is clear. This is lower than expected for the working conditions of accumulator 3 of the guide vanes. In fact, it seems that the difference between the real and expected values for the oil tank level is increasing over time, except in the last part of the graphic in Figure 9 where the real and expected values are approaching.

Two similar models to *f*2 were built and tested during the same periods of time replacing the variable A3GV (Oil level in accumulator 3 for the guide vanes) by other equivalent elated respectively to accumulators 1 and 2 for the guide vanes. The results were similar.

According to the results obtained, all five models applied for anomaly detection in the oil tank level (three of them presented in Figures 7, 8 and 9) coincide in that they indicate a lower level of oil over time. This is an indicator of a possible leakage of oil in the oil tank level or surrounding locations. The accumulators are working as expected, but the total oil level in the tank of the HPU is decreasing. This was verified and a leakage was discovered from the oil side to the nitrogen side of the accumulators.

These examples demonstrate that the deviation values obtained from the comparison of the real value and predicted one by the patterns of normal

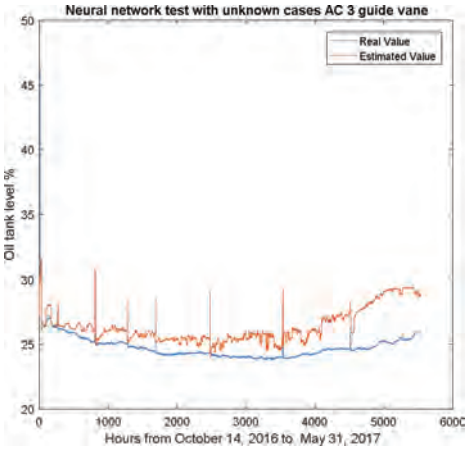


Figure 9. Estimated value for oil tank level in percentage predicted by the normal behavior model and the real value observed for the training set using the power generated, the oil tank temperature and the oil level in accumulator 3 for the guide vanes.

behavior can be good indicators for alerting when a typical relationship among variables could be broken.

In this case, the unexpected decreasing level in the oil tank must be monitored.

6 CONCLUSIONS

This paper describes a methodology for the early detection of anomalous behavior conditions of selected Kaplan turbine components. The method is based on discovering behavior patterns, also called normal behavior models, from the observation of the typical relationships existing between a set of variables used as inputs to the models and the corresponding output of a target variable whose expected value has to be predicted. The criteria to select the variables to use in the models are based on the physical working principles of the component in order to detect symptoms of abnormal behavior that can cause a possible failure mode.

The data set used for pattern discovering of normal behavior comes from the SCADA system of the plant. Abnormal behavior is any significant deviation or difference between the predicted output of the models and its corresponding real observation.

The paper presented some examples of normal behavior models for the cases of characterization of power generated by the hydropower plant and the oil tank level considered from different perspectives such as the oil level in the bank of accumulators of the turbine runner and the bank of accumulators of the guide vanes. Once the models were created, they were applied to new examples of operation. The predicted amount of generated power was always as expected, but the oil tank level was not. The analysis of deviations of normal behavior described in the paper shows that the oil levels in the accumulator banks were according to their working conditions, but the oil tank level was continuously decreasing during the time analyzed. This suggests a need for close monitoring of this level in order to search for the cause of this potential detected leakage.

In future works, an approach based on different algorithms working in parallel for anomaly discovering will be tested. This will improve even more the robustness of the anomaly detection method proposed.

ACKNOWLEDGEMENT

The research leading to these results has received funding from the MonitorX project from the Research Council of Norway (project no. 245317/E20) and the MonitorX industrial partners.

REFERENCES

Bishop, CM. 1995. *Neural networks for pattern recognition*. Clarendon Press.

Bishop, CM. 2006. *Pattern recognition and machine learning*. Springer.

Chandola, V., Banerjee, A. & Kumar, V. 2009. Anomaly detection: A survey. *ACM Computing Surveys,* 41(3).

García Matos, J.A., Sanz-Bobi, M.A. &Muñoz, A. 2013. Asset management overview focusing on fault detection in industrial processes – A state of the art. *International Journal of COMADEM*, 16(2): 43–54.

Jamil, I., Jamil, R., Jinquan, Z., Ming, L., Dong, W.Y. & Jamil, R. 2013. Condition-based maintenance decision-making support system (DSS) of hydropower plant. *International Journal of Innovation and Applied Studies* 4(3): 593–602.

Kruse, R, Borgelt, C, Klawonn, F, Moewes, C, Steinbrecher, M, Held, P, 2013. *Computational intelligence: a methodological introduction*. Springer.

Mohanta, R.K., Chelliah, T.R., Allamsetty, S., Akula, A. & Ghosh, R. 2017. Sources of vibration and their treatment in hydro power stations-A review, *Engineering Science and Technology, an International Journal*, 20 (2): 637–648.

Sanz-Bobi, M.A, Andrade Vieira RJ. 2011. A method for estimating stress of a failure mode in a component due to abnormal behaviour observed in a wind turbine, *24th International Congress on Condition Monitoring and Diagnostics Engineering Management,* Stavanger, Norway: 989–998.

Selak, L., Butala, P. & Sluga, A. 2014. Condition monitoring and fault diagnostics for hydropower plants. *Computers in Industry* 65(6): 924–936.

Topliceanu, L., Gabriel, P. & Furdu, I. 2016. Functional problems and maintenance operations of hydraulic turbines. *TEM Journal* 5(1): 32–37.

WEC, World Energy Council. 2017. Energy resources, Hydropower. https://www.worldenergy.org/data/resources/resource/hydropower/. Accessed in November, 2017.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Current status of the MFM suite for diagnostic and prognostic reasoning of industrial process plants

Harald P.-J. Thunem
Institute for Energy Technology, OECD Halden Reactor Project, Norway

ABSTRACT: This paper presents the status of a software system, the Multilevel Flow Modeling (MFM) Suite, dedicated to the design and analysis of MFM models related to diagnostic and prognostic analysis of physical processes. New and updated features of the system are described, as well as some examples of its practical use. The paper also briefly describes how the system facilitates the collaboration between control room and field operators via the Android-based MFM Viewer app.

1 INTRODUCTION

Multilevel Flow Modeling (MFM) (Lind, 2011) is a methodology for graphical modeling of industrial processes by representing the goals and functions of industrial plants. The purpose is to model the combined functions of any number of physical process components, which together provide the means to achieve one or more goals. The model may then be used for diagnostic and prognostic purposes to determine the possible causes and potential consequences of unwanted process events.

Since MFM models consist of graphical, interconnected elements arranged in specific structures, there is an apparent need for dedicated software to design them. The MFM models also need to be connected to the process, the functionality of which they represent.

For several years such a dedicated software system, the *MFM Suite*, has been under development at the Institute for Energy Technology (Thunem, 2013, Thunem and Zhang, 2015). The system will allow a user to graphically design and verify the semantic correctness of MFM models, in addition to creating graphical models of the industrial process. It will provide associations between process components and MFM functions. Using a dedicated MFM reasoning engine developed at the Technical University of Denmark (DTU), the system will perform diagnostic and prognostic analyses of anomalous events on online process data.

This paper provides an update of the system's features, a brief description of results from applying it in a practical experiment, and the functionality and use of the Android-based MFM Viewer app.

2 APPLICATIONS IN THE MFM SUITE

The MFM Suite primarily consists of three applications: the *MFM Editor* for graphical creation, editing and verification of MFM and process models, the *MFM Runtime* for real-time acquisition and analysis of on-line sensor data, and the *MFM Playback* for step-wise and simulated real-time playback and analysis of sensor data. Since the applications share a lot of functionality, they are all based on the *MFMApplication* Java class.

The MFM Suite also includes a simple launcher from which to start the applications.

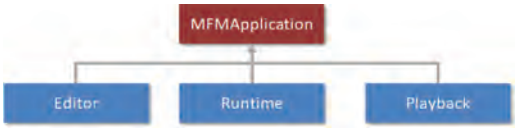


Figure 1. Inherited functionality in MFM Suite applications.

Figure 2. The launcher application.

3 THE EDITOR APPLICATION

The *MFM Editor* contains two essential graphical modelers, the MFM modeler and the process modeler. As both are based on the *ShapeShifter* framework (Thunem et al, 2011, Thunem, 2012), they share a lot of functionality.

3.1 *Meta-model display*

Most complex processes may be in more than one operating state (e.g. start-up, normal, shutdown), and the functionality of the process components and the process goals may depend on the operating state. Therefore, it may be necessary to create separate MFM models to describe the component functionalities and goals for each of the operating states.

The MFM Suite applications includes a meta-level display (Figure 3), which allows MFM models to be connected by arrows, indicating transitions from one operating state to another.

Note that only *one* MFM model may be considered *active* at any point, and any MFM reasoning analysis will be performed on the active model only. This furthermore requires that for each MFM model, the alarm limits for process sensors associated with MFM functions must be set individually, as their alarm limits may depend on the operating state.

3.2 *Click-and-drop model design*

Both the MFM and the process modelers provide graphical click-and-drop design of models in a manner similar to e.g. Powerpoint. For both modelers, a list of available components is shown to the left of the editing area (see Figure 4 and Figure 5). Clicking on one of them changes the cursor into a miniaturized version of the component. Clicking anywhere within the editing area will place the component at this point.

The MFM modeler provides another, faster method which also facilitates the creation of semantically correct models. When selecting an MFM function in the editing area, a list of allowed (according to MFM rules) connections (consisting of one MFM relation and one MFM function) will be displayed (see Figure 6). Clicking on one of them will create the connection from the selected function and place it at a suitable position. The modeler will automatically select the newly added function and update the list of available connections. In this way, complex MFM structures can be designed in relatively few steps.

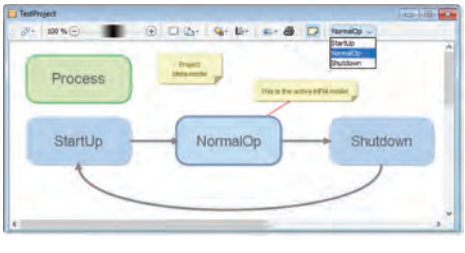


Figure 3. A project's meta-model.

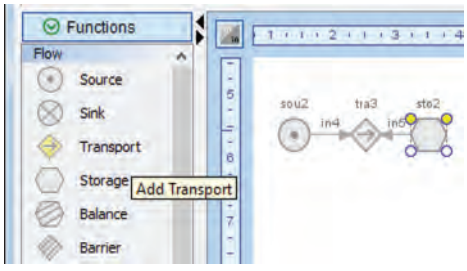


Figure 4. Designing an MFM model by click-and-drop.

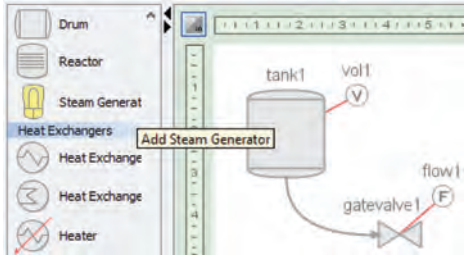


Figure 5. Designing a process model by click-and-drop.

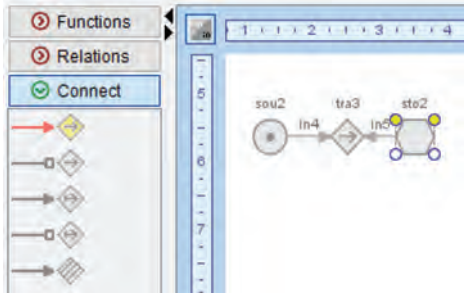


Figure 6. Assisted design of MFM structures.

4 THE RUNTIME AND PLAYBACK APPLICATIONS

The *MFM Runtime* and *MFM Playback* applications use the same *ShapeShifter*-based graphical components for displaying the MFM and process models as the MFM Editor, however in this case the editing functionality is disabled.

4.1 *Separate analysis threads*

The applications will perform diagnostic and prognostic analyses on a given MFM model using the reasoning rules that are explained in detail in (Zhang, 2015). The analysis processes will run in separate execution threads, which will set flags to avoid e.g. starting a new diagnosis before the previous has terminated (however, a diagnosis may run concurrently with a prognosis). This will prevent two problems; the analysis process will not lag the sensor sampling in cases where the sampling interval is shorter than the analysis time, and the user interface remains responsive since the main application thread is not blocked (Figure 7).

4.2 *Sensor display based on value and origin*

In addition to indicating the sensor range by changing the shape, color coding is used to indicate whether the sensor's value is obtained from the process, manually set by an operator, or inferred from an MFM model analysis.

An example is shown in Figure 8, where the shape (down-arrow) and color (red) of sensor PI444 indicates that it has a value below the low limit, and that the source is the actual process/simulator. This value, possibly along with other abnormal sensor values, has triggered an analysis in the associated MFM model.

The analysis has concluded that a *possible* cause of PI444's low value is a high value in sensor TI463, so this sensor is visualized with an up-arrow shape and a yellow color (indicating that the sensor state is inferred from an MFM analysis). Note that if the sensor is functioning correctly and shows a high value, it would be visualized using the red color. Since it is not, we can conclude that either this is not the cause of the abnormal value in PI444, or the sensor is malfunctioning.

The shape and color of sensor TI465 indicates that the sensor has a normal value/state (round shape), and that it has been set manually by an operator (green color), perhaps as a result of a visual inspection.

4.3 *Sensor trend display*

The process viewer component in the Runtime and Playback applications will display trend graphs for

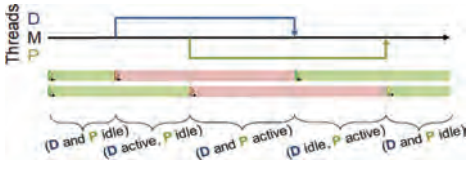


Figure 7. Concurrent main (M), diagnosis (D) and prognosis (P) threads.

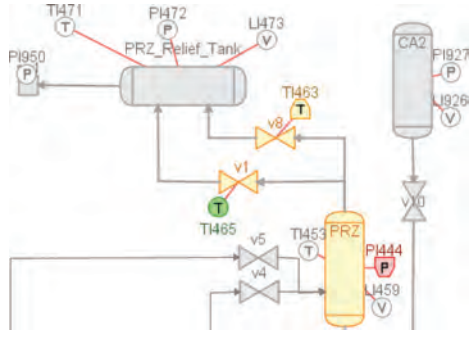


Figure 8. Sensor visualization.

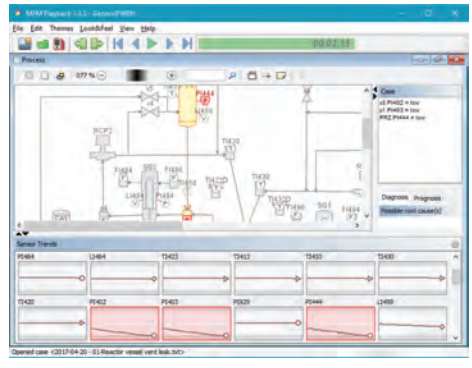


Figure 9. Process sensor trend panels.

selected process sensors (Figure 9). Through a dedicated settings menu the user may select which sensor trends to display (using a searchable selection panel), whether to show different markers depending on sensor type, and specify all colors used in the trends.

4.4 *Playback of existing data (Playback)*

The MFM Playback application will read stored sensor data from files to allow step-wise playback and analysis of previous experiments. The application also includes a slider to move back and forth freely. Furthermore, it can also replay a previous experiment in simulated real-time by activating a timer in a separate thread, and using the stored sampling intervals to update all sensors and re-run any analyses at correct times.

5 FEATURES COMMON TO ALL APPLICATIONS

5.1 *Server functionality*

The MFM Suite applications include functionality to allow external network clients to download the currently open MFM project (process and

MFM models) and to access updated MFM function states and sensor values. The MFM Suite also accepts commands from external clients for setting MFM function states and sensor value ranges (Figure 10). This allows e.g. a field operator to have a situation understanding similar to a control room operator and to prune away cause paths that have been determined irrelevant.

5.2 *Web export*

The MFM Suite applications will provide updated displays of models with dynamic data accessible through an ordinary web browser. When opening an MFM project, the applications will generate several files, depending on the number of models. An example of the files is shown in Figure 11.

Figure 10. Data server connectivity.

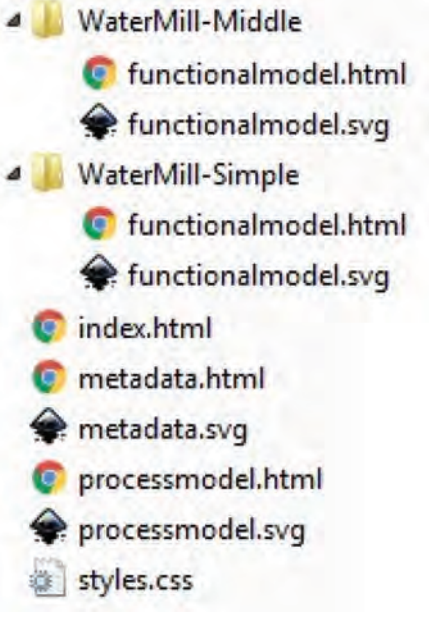


Figure 11. Web export files.

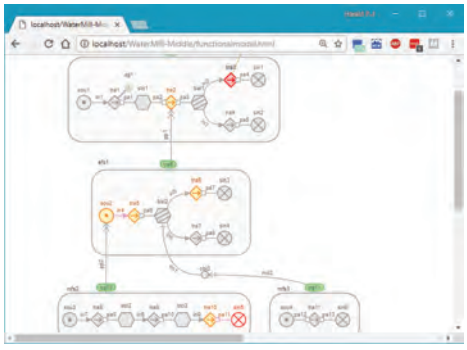


Figure 12. Web-page showing an MFM model.

The graphical models are exported as SVG (Scalable Vector Graphics) files, and since these graphics files are (as the name implies) scalable, they can be zoomed in the browser without the loss of any details (Figure 12).

Whenever the visual appearance of any element (MFM function, sensor etc.) in a model changes (e.g. due to an analysis affecting the elements), or whenever the user saves the project, the files are updated. Code is included in the HTML files to reload the pages at given frequencies (default every 5 seconds). This frequency is specified using the Settings dialog box of the applications, which also allows specifying an output folder, to which the generated files are exported. In order to make the web pages available to other devices in a network, this folder could be set to the active root folder of a web server application, such as e.g. Apache or IIS.

6 EXAMPLES OF USE

Several tests have been conducted using an MFM model of the primary side of two PWR simulators; the RIPS simulator which is based on the Ringhals power plant, and a generic PWR simulator (Zhang et al, 2014, Zhang et al, 2016, Thunem, 2017). Several scenarios were tested, including reactor trip cause by too low pressure in the steam generators, and a high pressure in the reactor coolant system.

Before running the experiment, the simulator was initialized to a normal running state. At this point, the values of all sensors were registered and used to set the individual alarm limits. The process model contained a total of 49 sensors, 21 of which were associated with MFM functions.

Before running each scenario, the simulator was reset to a normal running state. The control room operator would run the selected scenario by adjusting selected process parameters and letting the simulator run its course.

In most cases, the reasoning engine performed flawlessly. Figure 13 shows one scenario with multiple anomalous sensor readings (right side) and the successful diagnosis using MFM reasoning (left side).

It was, however, noticed that in some cases the selection of trigger function in the MFM model would cause a memory problem (stack overflow) in the reasoning engine. By manually selecting another trigger function, the problem would disappear.

The MFM reasoning engine uses the Java-based JESS inference engine (Friedmann-Hill, 2003), which employs recursion, a common cause of stack overflow problems. It is reasonable to assume that the selection of different trigger functions resulted in different recursion levels during the rule-based reasoning. One solution to this problem is to increase the Java call stack; however, the maximum size of the call stack may be platform dependent and difficult to ascertain.

Since the JESS system is no longer supported, the MFM reasoning engine has been re-implemented using the Drools inference engine, however, the experiments have not yet been re-run to verify any performance improvements.

7 MFM VIEWER

The *MFM Viewer* is a graphical model viewer designed for Android-based portable devices such as tablets and cell phones. The use of this application will facilitate a distributed team of e.g. control room and field operators to gain a common understanding of a process situation regardless of their location. For practical purposes a cell phone was used to create the images below, while in a work situation a tablet may be more suitable. The application will scale appropriately to all common display resolutions.

The main menu will allow the user to specify relevant settings, including the IP address and port number of the data server, i.e. any of the MFM Suite applications currently running. The main menu further provides a "Download project" menu item, which will instruct the MFM Suite applications to package the currently open project into a zip file and to transmit the zip file to the device. After receiving the zip file, the MFM Viewer will unzip the project files and display the project meta-model. Clicking on any of the model boxes (MFM or process) opens the model in a new window. In any of the models, the user may scroll and zoom using common finger gestures.

At the bottom of the MFM and process displays is a green button. Pressing this will cause the application to connect to the MFM Suite and retrieve and display the current MFM states and sensor values at a sampling frequency specified in the "Server Settings". The button will turn red, indicating the retrieval of data (Figure 14). Pressing the red button or returning to the meta-model will disable the connection. Note that for both models there is a search icon in the upper right corner. Clicking this will open a search field, which allows the user to find (by highlighting) any MFM function or process component whose outer label contains the given text string.

The MFM Viewer also allows the user to set the states of individual MFM functions or sensors. This is done by clicking on the desired element, which will open a dialog box, and clicking on the desired state (Figure 15). This will in turn trigger a re-analysis of the models in the MFM Suite application.

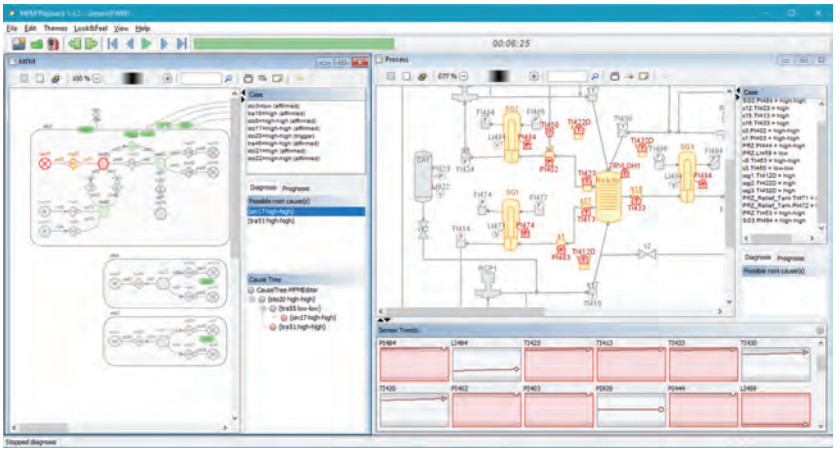


Figure 13. Scenario diagnostic.

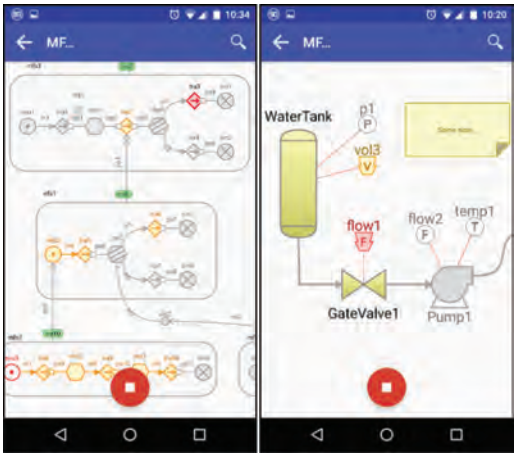


Figure 14. Displaying models and highlighting analysis results in the MFM Viewer.

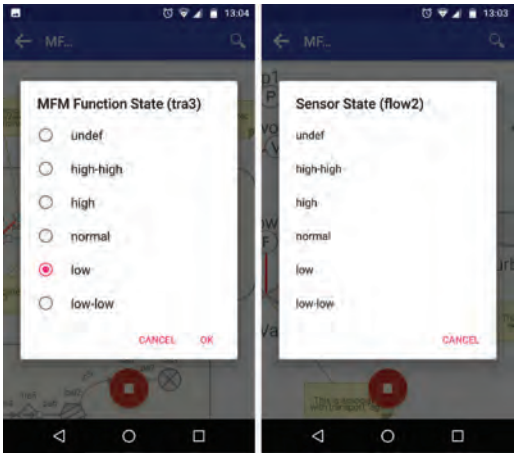


Figure 15. Setting MFM function and sensor states in the MFM Viewer.

8 FURTHER WORK

The primary focus on the MFM Suite development has until recently been on including necessary functionality to graphically design MFM and process models, create links between them, and perform online analysis with simulated process sensor data. When presenting the analysis results, focus has primarily been on how the MFM functions are affected, and which MFM functions have been identified as root causes. Interpreting the results has required some familiarity with the MFM methodology, which is unreasonable to expect from e.g. a control room operator.

To make the analysis results more accessible to someone without MFM training, the focus of the analysis presentation will therefore shift towards the process display, indicating which process components are the possible root causes and which components may be affected by the identified anomalies.

The activity will further evaluate various display modes to enhance the operators' comprehension of a given process situation in light of established analysis results, and to facilitate collaboration between control room and field operators via dedicated interaction mechanisms.

REFERENCES

Friedmann-Hill, Ernest, 2003. Jess in Action, *Manning Publishing Co* (ISBN 1930110898), USA, 2003 (http://www.manning.com/friedman-hill/).

Lind, Morten 2011. An Introduction to Multilevel Flow Modeling, *International Journal of Nuclear Safety and Simulation*, 2(1), pp. 22–32.

Thunem, Harald P-J et al. 2011. Using an Agent-oriented Framework for Supervision, Diagnosis and Prognosis Applications in Advanced Automation Environments, *Proceedings of the ESREL 2011 conference* (pp. 2368–2375, ISBN 978-0-415-68379-1), September 18–22, 2011, Troyes, France.

Thunem, Harald P-J 2012. Use-Case of an Agent-Oriented Framework for Supervision, Diagnosis and Prognosis Applications, *Proceedings of the combined ESREL2012/PSAM11 conference* (pp. 4910–4917, ISBN 978-1-62276-436-5), June 25–29, 2012, Helsinki, Finland.

Thunem, Harald P-J 2013. The development of the MFM Editor and its applicability for supervision, diagnosis and prognosis, *Proceedings of the ESREL2013 conference* (pp. 1807–1814, ISBN 978-1-138-00123-7), Sept 29 – Oct 2, 2013, Amsterdam, Holland.

Thunem, Harald P-J, 2017. Diagnostic Decision Support – Recent Development and Updates of the MFM Suite, *Halden Report HWR-1223*, Institute for Energy Technology, August 2017.

Thunem, Harald P-J & Zhang, Xinxin 2015. The continued development of the MFM Suite and its practical application on a PWR system, *Proceedings of the ESREL2015 conference* (pp. 2463–2471, ISBN 978-1-138-02879-1), Sept 7–10, 2015, Zürich, Switzerland.

Zhang, Xinxin 2015. Assessing Operational Situations, *PhD thesis*, Department of Electrical Engineering, Technical University of Denmark, June 2015.

Zhang, Xinxin et al. 2016. Applying MFM to the PWR Analysis: The experimental results and preliminary conclusions, *Halden Report HWR-1191*, Institute for Energy Technology, March 2016.

Zhang, Xinxin et al. 2014. Practical Application of the MFM Suite on a PWR System: Modelling and Reasoning on Causes and Consequences of Process Anomalies, *Halden Report HWR-1118*, Institute for Energy Technology, August 2014.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Prognostic and health management design for subsea applications

Xiaojing Gao, Octavian Niculita, Don McGlinchey & Babakalli Alkali
Department of Engineering and Built Environment, Glasgow Caledonian University, Glasgow, UK

ABSTRACT: The design of a subsea production system needs to ensure high reliability and safety figures since these assets will be deployed in harsh environments for extended periods of time. Maintenance costs associated with these systems represent a significant percentage of the total operational expenditure incurred by an Oil & Gas operator. Traditional reactive maintenance approaches applied on subsea equipment are starting to drop in their efficiency as degradation occurs on such systems resulting in prolonged downtime periods. Prognostics and Health Management is a relatively new topic and other industry sectors have demonstrated that it can provide a solution for reducing maintenance costs and improving systems' overall availability. This paper presents a prognostic and health management development process suitable for subsea production systems.

1 INTRODUCTION

When oil and gas exploration is not economically viable through tradition oil platforms, subsea production systems represent an alternative for the majority of operators. A Subsea Production System (SPS) is a collection of hydrocarbon extracting equipment located on the seabed and its main components consist of a seabed wellhead, subsea x-mas tree (XT), manifold, umbilical, riser, a network of pipelines, flowline as well as subsea power and control systems. The amount of shallow water oil and gas reserves is decreasing. This has led exploration and production into deep waters where SPS are the preferred solution to make development economically viable. The subsea industry is now facing a number of challenges on how to improve reliability and safety of critical assets in an economical way. Subsea lifecycle analysis demonstrated that Capital Expenditure and Reliability Availability Maintenance Expenditure (RAMEX) are the two major costs of a subsea asset, with downtime cost making up the majority of RAMEX. Prognostic and health management (PHM) has the capability to address these financial challenges and reduce the downtime cost by assessing and predicting the Remaining Useful Life (RUL) of a component/system while supporting the system's goal and compliance with high level requirements- safety, reliability, availability, maintainability etc.

At present, the main research focus area of PHM in the context of subsea applications is only targeting the development of such capability for isolated components. This paper introduces on an integrated approach to design the PHM for a SPS at the system level. This approach is mapped on a typical engineering design and operational process of a subsea system and it involves integration and concurrent analysis of multi-disciplinary sources of knowledge and interaction between several engineering functions.

Section 2 of this paper discusses the current prognostic approaches and the level of adoption of PHM for subsea equipment. Section 3 will cover four engineering disciplines and their potential use for the development of the PHM capability. Section 4 will present a novel PHM development process capable of integrating knowledge, information and data within concurrent engineering analysis. In section 5, an instantiation of the PHM development process for a XT will be presented.

2 STATE OF THE ART OF PHM

Four different types of prognostic approaches currently exist:

- Experience-based prognostic approaches are based on historical data and knowledge accumulated during the lifecycle of systems.
- Model-based approaches involve the construction of mathematical model which integrates the underlying physics of failure of the critical components of the system, their degradation and their failure modes.
- Data-driven prognostic approaches specify the behavior of a system through gathered operational data (CM data via sensors and/or event data). The data is processed and compared with key parameters/features to predict the probability of fault occurrence.

- Hybrid prognostics are a combination of the model-based and data-driven with parameters in the model being continuously updated when data from service becomes available with the ultimate purpose of improving the accuracy of the prediction (Vachtsevanos, Lewis, Roemer, Hess, & Wu, 2006) (Medjaher & Zerhouni, 2013), (da Silva & Radespiel, 2013).

The majority of industry sectors adopted the data-driven prognostic approaches as a first attempt for the development PHM capability, particularly for complex systems (Bykovsky, 2008) as this enables the development of the understanding of degradation without the construction of mathematical model capturing the physics of failure. Currently, in the subsea arena, there are very few Condition Based Maintenance (CBM) solutions deployed in field. The major Original Equipment Manufacturers (OEM) for SPS include companies like: Technip/FMC Technologies, Cameron, GE, Aker Solutions, OneSubsea etc. Based on information available in the public domain, Cameron provides a blowout preventer condition based monitoring system and riser annulus condition system. However, FMC has delivered a first attempt of a system level Condition and Performance Monitoring (CPM) capability for a subsea asset, currently being installed at the Gjøa field (Soosaipillai, Roald, Alfstad, Aas, Smith & Bressand, 2013).

3 ENGINEERING DISCIPLINES AND THE STAKEHOLDERS INVOLVED IN THE SUBSEA PHM DESIGN

3.1 *Overview*

To develop and implement any of the prognostic approaches mentioned in the previous section, different types of data and information is required (Vachtsevanos, Lewis, Roemer, Hess, & Wu, 2006). This information and data represents the output of several engineering disciplines being owned by various functional teams. The design stage for a subsea system is usually undertaken by the OEM and suppliers under the requirements established by an operator. During the operation period, maintenance teams will be hired by the operator to Inspect/Maintain/Repair (IMR) the subsea equipment, although, the maintenance of control systems will be the responsibility of the OEM. Very often, upgrades, overhauls and de-commissioning typically are done by different parties. During the life time of a subsea production system, multiple organizations are involved and a subsea system may include components and processes originating from all over the world. Throughout the entire lifecycle of a subsea field, interactions between different engineering disciplines belonging to different companies, which are operating under multiple languages and cultures, also take place. Hence, one of the challenges is the fact that the data/information required to develop a PHM solution is not unified and/or centralized. In the context of subsea applications, these engineering functions and the data/information associated to each of them is captured in Table 1.

3.2 *Subsea system design—initiation of the PHM design*

PHM Design must be underpinned by a level of understanding of the healthy state of a system. In the case of a subsea system, this characterization is developed during various stages of the design process (conceptual design, front end engineering design (FEED) and detailed design). Engineering modelling is carried out during the FEED. Modifications to the design (to include the PHM capability as an afterthought) might be extremely costly, therefore it is recommended to design-in the PHM function as the asset design progresses through various technical and business gates. It is instrumental at this stage to derive the PHM requirements from the subsea asset requirements, if such a capability is to be developed.

3.3 *Reliability and availability analysis—foundation of the PHM design*

In the offshore industry, Failure Modes and Effects Criticality Analysis (FMECA) tends to be one of most common approaches for reliability analysis and it has been increasingly implemented in the last ten years on subsea projects (DNV, 2013). FMECA also forms the foundation for good PHM design (Vachtsevanos, Lewis, Roemer, Hess, & Wu, 2006) although, in the context of subsea

Table 1. Engineering disciplines.

Engineering disciplines	Information
System design	Engineering models Schematics, Reports, Hierarchical levels Dependencies
Reliability & Availability	Failure concepts Criticality information (Re) Certification
Control and Instrumentation Condition and Performance Monitoring (CPM)	Operational conditions Condition, Performance Indicators Control data
Inspection, Maintenance and Repair (IMR)	Past operational conditions Maintenance records Failure history

equipment, it is mainly used to support qualification of new technologies or re-qualification of legacy systems. Reliability analysis must be based on the actual design of the system therefore a concurrent engineer design and reliability analysis is recommended to ensure the both disciplines are targeting the same point of truth. Input is required from IMR to ensure the efficiency and accuracy of this analysis. A FMECA study attempts a good understanding of system behavior under faulty conditions is instrumental to be able to design the diagnostic and prognostic capability. Historical data and knowledge can also be employed to adjust or re-design components/equipment/systems to improve reliability targets. This information usually resides with the operators and OEMs, but it is very often not available for the reliability team responsible with the qualification of the subsea equipment under investigation.

3.4 *Control & instrumentation/CPM—core of the PHM design*

A large number of parameters to assess the condition of an SPS is currently captured by CPM through sensors i.e. acoustic control systems, multiphase flow meters, accelerometers, pressure and temperature sensors, sand and leak detection systems, as well as detectors for dropped objects (ISO:13828, 2010). These types of sensors found their way into the subsea design either through regulations or recommendation and there is a consensus in the industry that they are not engineered for the purposes of ensuring higher availability figures. Production efficiency performance of subsea equipment has been particularly poor in recent years, reaching a low point of 60% in 2012 and averaging at 71% in 2015 although back in 2004 the production figures were above 80%. Degradation of production equipment is one of the contributor factors to this drop. Abnormal behavior of subsea equipment can be detected by condition monitoring solutions and information sensed by instrumentation is sent to the control module and interpreted by the operational teams (Markeset, Moreno-Trejo, & Kumar, 2013) to allow informed decisions guiding the operation and maintenance.

3.5 *Maintenance—exploitation of the PHM design*

Various types of maintenance strategies exist and can be implemented for subsea equipment in service. Corrective maintenance is typically applied after the failure has occurred. A scheduled maintenance regime aims at dealing with faults before they occur, but they are based on fixed time intervals. The conditional maintenance only supports non-dynamic estimation of the degradation and the most recent approaches take advantage of the PHM capability to estimate the RUL of a critical component and to plan the maintenance job according to these calculations (based on data-driven, model-based and hybrid prognostics algorithms). The PHM capability of a system is represented by a set of techniques and methods from different disciplines that combine knowledge and data to support predictive maintenance by detecting, diagnosing, predicting, advising and analyzing (postmortem) the failure information (Guillén, Crespo, Macchi, & Gómez, 2016). The information and data related to functional failures (failure modes) and physical failures (faults) of subsea equipment is scarce, so the traditional reactive maintenance approaches (corrective and scheduled) are still the preferred choices for oil and gas operators. The predictive maintenance regimes also present their own set of challenges and these must be considered during the design stage (the inability to accurately and reliably predict the RUL of a component/system; the inability of maintenance systems to document, learn and recommend that action should be taken; the lack of tools capable of demonstrating the effectiveness of a predictive maintenance program). SPS have been typically designed to operate over five years without failure, thus the operator will plan to carry out preventive maintenance every five years (Moreno-Trejo & Markeset, 2012). However, over the last decade, reliability data shows some components had to be maintained/replaced sooner to prevent failure. Hence, traditional maintenance can incur huge expense and consequential damage for an asset and the environment—such as pollution, loss of production, etc. (Markeset, Moreno-Trejo, & Kumar, 2013; Uyiomendo & Markeset, 2015). In recent years, traditional maintenance activities in the oil and gas industry are transforming through the adoption of CBM which are ensuring efficient maintenance, reducing lifecycle costs and improving the systems' overall availability. The application of CBM and PHM (as an extension of CBM) in the oil and gas industry has started to be exploited on for example, drilling systems, control systems and pumping systems.

4 SUBSEA PHM DEVELOPMENT PROCESS

To determine the Prognostic method, these four major disciplines need to share information and data with each other. However, PHM development for an entire subsea system is still a challenge from the view of big data management and information support, thus it is hard to build those disciplines into system.

In this paper, we propose an integrated approach to subsea PHM design process. This process is captured in Figure 1 and it highlights the data/information exchanged between the four main engineering disciplines, discussed in the previous

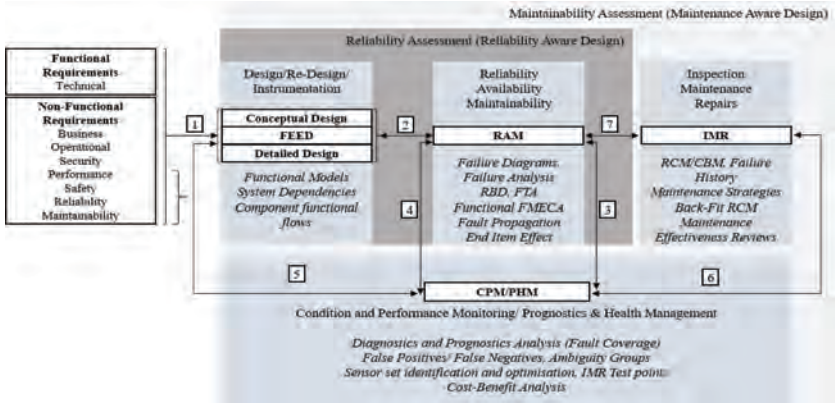


Figure 1. An integrated approach for exchanging engineering knowledge to support Subsea PHM Design.

section. The proposed subsea PHM development process includes feedback loops that are intended to enable enhanced data collection, exchange and analysis of knowledge related to the degradation of subsea components, addressed by different engineering disciplines as the subsea project moves through various technical and business reviews. Good communication during a generic design engineering process includes both historical and current information to be shared freely, problems to be reported, views to be exchanged, and positive interpersonal relationships to be retained, as the design progresses through various technical and business gates. However, the effective communication is also dependent on balancing the levels of information to the complexity of the task/topic, avoiding both over-complication and over-simplification. Hence, suitable levels of information must be delivered to the correct person at the correct time to communicate essentials without the receiver being overburdened with data (Parkes & Hodkiewicz, 2011).

Step 1 – The start of the PHM design process is represented by the requirements phase. High-Level (HL) requirements are typically divided into functional requirements (FR) and non-functional requirement (NFR) categories. Functional requirements define the technical details of a system (including the function of the system and the functions of each individual component) and non-functional requirements cover the attributes of the system (such as safety, reliability, maintainability, usability, performance, security, etc.). The Design/Re-design/ Instrumentation phase of the PHM development process coordinates various engineering analysis to ensure the final subsea design is meeting the requirements established at the start of the project. During the design phase, data associated with the environmental conditions, reservoir, well completion, process and operations, host facilities, safety and hazards should be considered when progressing through conceptual, FEED and detailed design. These engineering efforts are targeting the technical requirements (also known as functional requirements (FR)) of a subsea system since this design element is heavily regulated and supported through recommended practices (ISO 13628–1, 2010). Non-Functional Requirements are also considered by the current subsea design best practices and they cover safety, performance and security. However, there is no attempt to define and to target the reliability and maintainability requirements at the early stage of the design. This influences the PHM requirements definition as these are derived from the reliability and maintainability requirements. Only recently, the industry has generated recommended practices like the API-RP-17 N on topics related to reliability, technical risk and integrity management (API RP 17 N, 2009). However, they are not yet adopted due to the lack of tools, processes and meaningful reliability data to support their implementation. Reliability and maintainability requirements should be part of the non-functional requirement (NFR) of the system. Deriving PHM requirements should be done from system's high-level (HL) Non-Functional requirements. For example, a HL NF requirement for a XT can be affordability by reducing the down-time periods while keeping the same levels of safety. In this manner, cost and safety become main drivers for the development of a PHM solution. Derived PHM requirements can be represented: PHM Requirement 1 – the XT must have a feature that can reliably predict functional failures at least one week prior to the actual event and PHM Requirement 2: the XT must have the capability of offering mitigation/advisory generation in the context of current operational conditions. Having access to information provided by such features, the operator reasonable time to schedule a vessel, equipment, personnel to carry out an intervention on the faulty components.

Step 2 – To support the PHM within the subsea design phase, reliability analysis needs to be carried out concurrently with the subsea system design. Reliability analysis should be based on the actual design data to support the engineering analysis that can identify the means to detect and isolate the potential faults which may occur during the operation. One way to support the realization of this step is through the exploitation of functional models which can support reliability engineers to carry out analysis such as FMECA, Fault Tree analysis (FTA), and Reliability Block Diagrams (RBD) from the very early stages of the subsea design process. These methods and techniques can also underpin the Reliability Centered Maintenance (RCM) and Back-Fit RCM analysis at later stages of the lifecycle. RCM includes four elements that are critical to a maintenance program aimed at improving availability of a given asset. These elements are: preservation of the system function, identification of the failure modes that can lead to functional failures (and sequentially downtime), prioritization of failure mode candidates based on Occurrence (O), Severity (S) and Detectability (D) by highlighting the Risk Priority Number (RPN = O $\times$ S $\times$ D) of each of the candidates and finally selections of applicable and effective tasks to control the failure modes. Back-Fit RCM builds on the same RCM principles by incorporated operational reliability figures (by updating the O, S, D parameters) and evaluating the applicability and effectiveness of the control measures. We believe that the reliability, availability and maintainability analyses should be the foundation of the PHM design as it can highlight the risk associated with a brand new subsea design or a legacy system using information from service (provided to the reliability team during *Step 7* using the output of a bespoke maintenance analytics engines). Sequentially, using this information, the design will be assessed by the reliability authority during a technical review (*targeted reliability* assessment) and if the design fails to meet a specific target, the risk and the critical components must be addressed either through re-design, redundancy or addition of instrumentation. Informed trade-off studies between these three approaches must be in place to guarantee the final design meets all the requirements of the project. However, this topic is beyond the scope of this paper. Nevertheless, for far too long, reliability assessment on subsea equipment was carried out only to present a *measured reliability* figure to support re-certification of production equipment already in exploitation.

If the PHM is channeled as derived requirement from the RAM requirements, specific levels of targeted reliability can be achieved. Also, different PHM requirements and implementation strategies can be evaluated at this stage against specific sets of RAM requirements. Very often, in subsea applications, redundancy seems to be the option preferred by the system designers as this guarantees improved availability figures when the primary component/sub-system fails. The major drawback of the redundancy approach is that fact that it does not offer any indication of the RUL for the critical component, sub-system or system, therefore, our case for addition of instrumentation supporting the PHM capability. If instrumentation is required for diagnostics and prognostics purposes, this must be defined by the PHM analysts in collaboration with the subsea design and reliability teams. We believe that majority of the condition monitoring applications existent in a subsea environment were driven by vendors of sensors capable of targeting symptoms associated with specific failures. This approach captures failures in isolation and does not account for propagation of faults leading to functional failures of other components and sequentially to failure of the system. This limitation can be overcome using functional models of the system, functional relationships and failure/effects dependencies in a system for both functional and physical failures, defined using widely accepted, well defined failure taxonomies. Once the stakeholders of the asset have validated the propagation tables (component's reaction to functional failures of the system) generated against analysis capturing various end-item effects. The end-item effect is the consequence a failure mode has on the operation, functional output of the system at the highest indenture level (an item's position in the system hierarchy relative to the top-level item).

What we propose in step 2 is an implementation of the API 17 N recommended practices through a concurrent design-reliability analysis by evaluating the actual reliability of the design by highlighting the effects of failure modes leading to functional failures.

Step 3 – enables an informed dialog between RAM engineers and PHM analysts by allowing the use of system level propagation table (a collection of all the failure mode signatures—the effects (throughout the system, and not just at the point of occurrence) of a given failure mode universe. Currently, the job description of a PHM analyst falls somehow under the control team although we believe that its responsibility and involvement goes beyond the control systems. The PHM team should liaise very closely with the design team as it aims at the identification and optimization of sensor set solutions capable of detecting, isolating and making predictions of a given set of failure modes considered under the PHM analysis. The mean of realization of this dialog is represented by the reliability models populated with failure and criticality information of failure concepts. The automated PHM instrumentation analysis aims to identify and optimize, in a systematic manner, the sensor set configurations capable of supporting the detection, diagnosis, prediction functions of a subsea asset to further enable advisory

generation. It also aims to calculate, for each sensor set solution identified during this process, the fault detection and isolation characteristic representing the proportion of failure modes selected for PHM analysis that can be detected and identified by a given sensor set under consideration. The PHM instrumentation analysis must be able to allow modification of existing sensor arrangements based on user knowledge or trade-offs. Legacy sensors present on the system can be considered as part of this analysis, although qualification institutions and regulators will not easily accept interrogation of sensors used for fail-safe control purposes (functional safety sensors). During this step, criticality of failures affecting a subsea system must be considered during the PHM instrumentation analysis. For new subsea designs, no measured criticality information exists and this must be defined using input from various stake-holders (subsea designers, RAM team, Operators, IMR personnel) by taking into account qualification of new technology standards and recommended practices (DNV-RP-A203, 2011; DNV-DSS-401, 2012). For subsea legacy systems, the criticality should be considered given what failed in service and it can be characterized through occurrence, severity and detectability parameters by calculating a risk priority number for every single component of the subsea equipment. The PHM instrumentation analysis should be capable of running the identification and optimization algorithms for specific groups of components by focusing on specific targeted criticality. There is also a feedback loop between the PHM and RAM functions meant to ensure that the selected sensor set solution meets the reliability criteria of the system (as a sensor that will fail in service ahead of the component that it is monitoring for failure does not ensure higher levels of availability for the asset). This feedback loop is represented within the PHM development process by *Step* 4.

Step 5 of the PHM development process enables the dialogue between the PHM analysts and the subsea design team. Although represented as a separate link/step, this dialog should take place concurrently with the reliability analysis of the instrumented subsea system. Decisions regarding unfeasible sensor set solutions are taken, and trade-off studies related to cost, weight, coverage, location, reliability, probability of detection, probability of prediction, likelihood positives and negatives ratios, physical constraints, loads, environmental conditions should be carried out by a multi-disciplinary team led by the RAM function. During this step, the reliability and maintainability requirements are verified and validated by group of experts. As a mean of realization of this dialogue, we propose a model-based approach as this will allow rapid generation of sensor sets spanning multiple levels of hierarchy for a subsea system by considering technical and economical metrics.

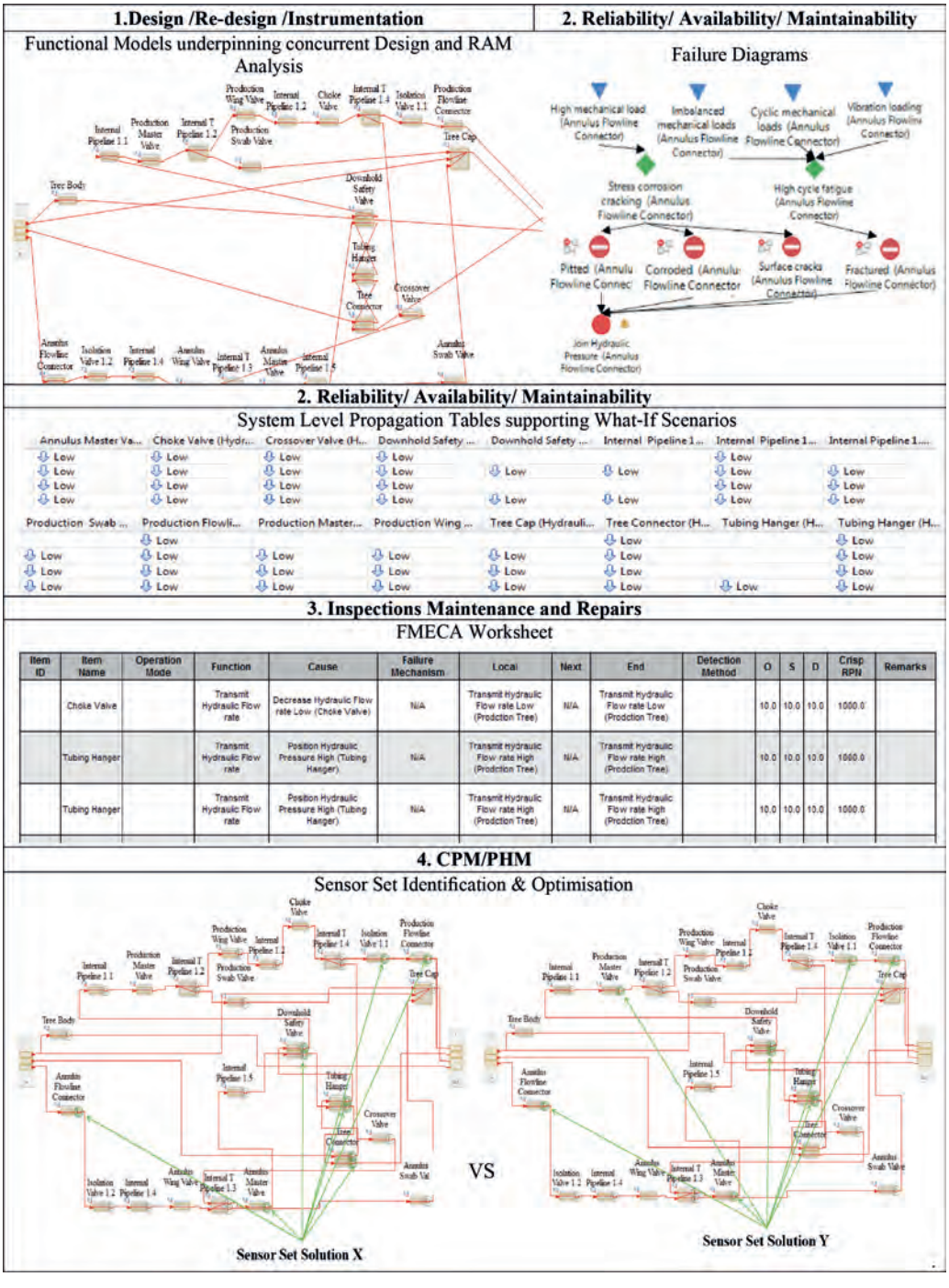
Step 6 facilitates the exploitation of the data provided by a PHM-enabled subsea system and the support offered to the IMR function. Data from sensors will be plugged to diagnostic and prognostic engines capable of supporting the IMR function on fault detection, fault isolation and ideally, prediction of the remaining useful life.

Step 7 of the PHM development process facilitates an implementation of an integrated analysis-drive sustainment activity. It provides traceability of the subsea maintenance activities when using PHM information. We recommend the use of function-based reliability models to gather, share and analyze maintenance data in order to enable automated failure and data reporting, analysis and corrective action system (FRACAS/DRACAS).

5 CASE STUDY

For the implementation of the integrated PHM development process for subsea equipment, a commercial-of-the-shelf software tool, namely Maintenance Aware Design environment (MADe™) developed by PHM Technology was employed. It was used to carry out an instantiation of the process on a XT. The selection of this software package was based on the previous success in using it in aerospace industry sector on fuel system and environmental control systems (Hess, Frith & Calvello, 2005). MADe™ is a 'model-based' engineering tool that can provide an integrated framework to manage, control and analyse the information and data throughout different disciplines including the design, safety, reliability, availability and maintainability for a high-value high-complex systems. However, a good inter-discipline communication requires expert document management and control, and in this instance this is achieved through a model gathering data and knowledge from various disciplined. The instantiation of the PHM design process was carried on a typical subsea XT. Several concurrent engineering analyses belonging to separate engineering disciplines, but derived from a single model of this XT (developed within MADe™) were carried out. The outcomes of some of these analyses are highlighted in Table 2 (a and b). These are briefly summarized further on. The functional model accommodates information describing the input and output flows of each component, the causal relationships between these flows that allow for systematic propagation of failures throughout the system and criticality data (retrieved from the 6th edition of the Offshore Reliability Data—Volume 2). Knowledge and data characterizing the failure of each of the component forming a subsea XT was added to the functional model using concepts defining the causes, mechanisms, faults, symptoms and the links to the functional failures previously

Table 2a. Outcomes of implementation of the integrated PHM development process.

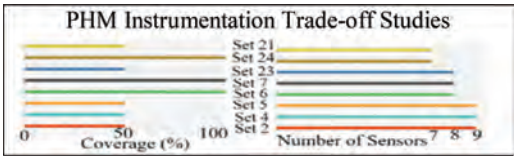


Item ID	Item Name	Operation Mode	Function	Cause	Failure Mechanism	Local	Next	End	Detection Method	O	S	D	Crisp RPN	Remarks
	Choke Valve		Transmit Hydraulic Flow rate	Decrease Hydraulic Flow rate Low (Choke Valve)	N/A	Transmit Hydraulic Flow rate Low (Prodction Tree)	N/A	Transmit Hydraulic Flow rate Low (Prodction Tree)		10.0	10.0	10.0	1000.0	
	Tubing Hanger		Transmit Hydraulic Flow rate	Position Hydraulic Pressure High (Tubing Hanger)	N/A	Transmit Hydraulic Flow rate High (Prodction Tree)	N/A	Transmit Hydraulic Flow rate High (Prodction Tree)		10.0	10.0	10.0	1000.0	
	Tubing Hanger		Transmit Hydraulic Flow rate	Position Hydraulic Pressure High (Tubing Hanger)	N/A	Transmit Hydraulic Flow rate High (Prodction Tree)	N/A	Transmit Hydraulic Flow rate High (Prodction Tree)		10.0	10.0	10.0	1000.0	

defined. Significant challenges were faced when trying to align the failure taxonomy used by the OREDA (based on the ISO: 14224 standard) and the failure taxonomy employed and defined by PHM Technology in MADe™.

A clear understanding of the causes, mechanisms, potential symptoms leading to a fault and the way this fault develops into a functional failure is instrumental in selecting the correct maintenance task or the PHM instrumentation, in an informed manner.

Table 2b. Outcomes of implementation of the integrated PHM development process.



Four components were selected during the criticality assessment as having a significant impact on the function of the XT, namely the choke valve, the down hold safety valve, the production master valve and the tubing hanger. For the scenario of these four functional failures, 100 sensor set solutions were automatically generated from the functional model, having between 7–9 sensors (measuring pressure and flow rate) offering between 50–100% fault coverage. Ambiguity groups were clearly highlighted during this process due the similarities in the fault signatures characterizing two of the faults. At this stage, the PHM development process allow investigations on the trade-off studies on the sensor set solutions very early on during the design process and various maintenance strategies can be benchmarked when using specific diagnostic and prognostic engines coupled to the instrumentation identified in the previous step.

6 CONCLUSIONS

The definition and articulation of cost-effective maintenance regimes is a challenging task for subsea assets. Very often, they are over-engineered since they are required to operate in harsh conditions for long period of times. Various stakeholders are involved with these assets throughout the entire lifecycle of these assets and knowledge related with the de gradation of this equipment is scattered throughout various organizations being owned and used by different engineering functions. In this paper, an integrated PHM development process was presented as a multi-disciplinary engineering analysis, complementing the current development of subsea equipment. The proposed development process aims to help subsea designers to integrate the development of the PHM capability of a subsea asset with the actual design of such systems. This is meant to happen at the early stages of the design process, but the process also enables the retrofit of such capabilities on legacy subsea fields to achieve higher availability and operational reliability figures. This is achieved by placing the reliability, availability and maintainability engineering analysis at the heart of the subsea asset and PHM system level design.

REFERENCES

American Petroleum Institute (API). 2009. *Recommended Practice for Subsea Production System Reliability and Technical Risk Management*, API RP 17 N, 1st Edition.

Bykovsky, V. K. 2008. *Data-driven modeling of complex systems*. Berlin: Springer.

DNV-RP-A203. 2011. *Qualification of New Technology*.

DNV-DSS–401. 2012. *Technology Qualification Management*.

DNV. 2013. *FMECA for Operational Planning of the Liwan 3-1 Subsea Development*. Singapore: DNV.

Goldsmith, R. & Ericson, R. 2003. *Lifetime Cost of Subsea Production Systems (JIP)*. Norway: DNV.

Guillén, J. A., Crespo, A., Macchi, M. & Gómez, J. 2016. On the role of Prognostics and Health Management in advanced maintenance systems. *Production Planning & Control*, 27(12), 991–1004.

Hess A., Frith P. & Calvello G. 2005. *Challenges, Issues, and Lessons Learned Chasing the Big 'P': Real Prognostics Part 1*, IEEE Aerospace Conference.

ISO 13374-4. 2015. *Condition monitoring and diagnostics of machine communication and presentation*.

ISO 13628-1. 2010. *Petroleum and natural gas industries – Design and operation of subsea production systems – Part1: General requirements and recommendations*.

ISO 14224. 2016. *Petroleum, petrochemical and natural gas industries – Collection and exchange of reliability and maintenance data for equipment*.

Ivanova, M. & Brkic, I. 2015. *Global Offshore Prospects and the Future of the North Sea*. London: Subsea Integrity and Efficiency Conference.

Markeset, T., Moreno-Trejo, J. & Kumar, R. 2013. Maintenance of subsea petroleum production systems: a case study. *Quality in Maintenance Engineering* 19(2), 128–143.

Medjaher, K. & Zerhouni, N., 2013. Hybrid prognostic method applied to mechatronics systems. *Advanced Manufacturing* 69(1–4): 823–834.

OREDA. 2015. *OREDA Offshore Reliability Data*. Høvik: OREDA Participants.

Parkes, K. & Hodkiewicz, M. 2011. The Role of Organizational Factors in Achieving Reliability in the Design and Manufacture of Subsea Equipment. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 22(6), 487–505.

da Silva, M.F. & Radespiel, E.S. 2013. *A Diagnostic and Prognostic Framework for Integrated Reservoir-Completion Management Using Intelligent Well Data*. Offshore Technology Conference.

Soosaipillai, C., Roald, P.K., Alfstad, D., Aas, T., Smith, G. & Bressand, J.Y. 2013. *Condition Performance monitoring for subsea: experience and value from Gjoa Filed, OTC-24182-MS*. Offshore Technology Conference.

Uyiomendo, E. E. & Markeset, T. 2010. Subsea Maintenance Service Delivery: Mapping Factors Influencing Scheduled Service Duration. *Automation and Computing* 7(2):167–172.

Uyiomendo, E. E. & Markeset, T. 2015. Subsea maintenance service delivery: A multi-variable analysis model for predicting potential delays in scheduled services. *Quality in Maintenance Engineering* 21(1): 34–54.

Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A., & Wu, B. 2006. *Intelligent Fault Diagnosis and Prognosis For Engineering systems*. New Jersey: John Wiley.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A particle filtering approach for temperature based prognostics

A. Bender & W. Sextro
Chair of Dynamics and Mechatronics, Paderborn University, Paderborn, Germany

ABSTRACT: Rubber-metal-elements are used in a wide range of applications for vibration and sound isolation. Nowadays it is state of the art to calculate the lifetimes of these elements under mechanical stress prior to their service life. To establish more reliable and safer rubber-metal-elements, continuous monitoring by different sensors can be used. Especially prognostics enable a rise in reliability, availability and safety. To establish these advantages, estimating the remaining useful lifetime of rubber-metal-elements should be realized during its service life based on current information on its condition. Therefore a suitable measure to monitor the condition of the element is necessary. This work focuses on temperature signals. This approach allows including the ambient temperature and thereby involving changing operating conditions. For estimating the RUL of rubber-metal-elements a model-based prognostics approach based on particle filtering is proposed. Its performance is analyzed regarding relevant parameters to enable the best performance for the applied data.

1 INTRODUCTION

1.1 *State of the art and motivation*

Rubber-metal-elements are used in a wide range of applications for sound isolation and in particular for isolating critical components from strong vibrations. Typical applications are trains, trucks and wind turbines. The bearing in focus is displayed in Figure 1. It consists of an inner steel ring, a rubber part and an outer steel ring which is slotted. This main part is within the outer hollow cylinder which is used in combination with the inner bolt for generating a prestress on the rubber part. Nowadays, it is state of the art to follow a preventive maintenance strategy handling these bearings. Thereby, the lifetime of the bearing needs to be estimated prior to its service life. Therefore this lifetime is estimated conservatively by the developer based on experience, lifecycle tests and the conditions of the planned application. This calculation is often based on linear damage accumulation theory (Spitz 2012). A preventive maintenance strategy shows some drawbacks regarding optimal utilization of the resource and costs. Moreover, today's industry develops growing expectations concerning efficiency of capabilities and availability. That is why condition monitoring gains more and more importance in the field of maintenance.

Figure 1. Rubber-metal-bearing.

1.2 *Maintenance strategies*

Maintenance can be divided in different strategies according to DIN EN 13306. The oldest strategy is the reactive maintenance. Technical systems were used until failure and needed to be repaired or replaced once they reached their end of lifetime. This procedure leads to a couple of problems concerning costs and time, for example possible high consequential costs due to unplanned downtime. Therefore the preventive maintenance strategy was developed. In that case mechanically lifetimes of technical systems are calculated based on experience, lifecycle tests and fatigue life calculations. However, this maintenance strategy does not enable exploiting the whole lifetime of a single product. The calculation is a generalized one that bases on assumptions regarding the expected loads over all bearings. Furthermore, safety factors are included in the calculations that ensure with a high degree

of certainty that every product is maintained or replaced previously to its end of lifetime. However, this strategy provides no information on the current state of individual bearings which experience individual loads during their lifetime and therefore degrade individually. That is why on the one side, possible early failures could occur and on the other side, bearings are replaced although their lifetimes are not yet exhausted. These disadvantages of the previous named strategies are the reason why the condition based maintenance strategy was developed. This strategy is mainly based on the condition of the product in focus. Using different kind of sensors, information about the condition of the product is acquired and progressed by condition monitoring methods. So, maintenance can be planned optimally based on the condition of the product and, in case of prognostics, additionally on the estimated remaining useful lifetime (RUL), which improves the reliability of the product and leads to an optimized efficiency.

1.3 *Structure of the following sections*

In this work the prognostics method which is used to estimate the RUL of these rubber-metal-bearings is analyzed regarding its performance on temperature data of these bearings. The used method is a particle filter. Due to the fact that different types of particle filters exist (Arulampalam et al. 2002, Jouin et al. 2016), in chapter 2 the used method is presented regarding type correlated differences. Aiming for realizing the best RUL prediction for rubber-metal-bearings, relevant parameters of the method for a performance analysis are identified. Chapter 3 focusses on necessary lifecycle tests and generated data for developing the condition monitoring system. Chapter 4 deals with the analysis of that temperature based prognostics. Two different measured values are implemented and particle filtering performance is analyzed based on varied parameters. In chapter 5 a conclusion and a short outlook are given.

2 PROGNOSTICS METHODS

2.1 *Types of particle filter*

Particle filters are Monte Carlo methods that base on Bayesian probability theory. These filters are model-based methods for state estimation that are appropriate for estimating non-linear behavior. Currently, it is a classical method for model-based predictions of RUL (Jouin et al. 2016). Moreover, particle filters create a probabilistic output which can be used to present uncertainty involved in RUL predictions. Additionally, this method was chosen because a multi model particle filter has been successfully applied to other signals of rubber-metal-elements (Bender et al. 2017b, Bender et al. 2017a, Bender et al. 2017c).

These filters can be divided in different types. The commonly known ones are Auxiliary particle filter, Unscented particle filter, Regularized particle filter, Sequential Importance Sampling (SIS) filter and Sampling Importance Resampling (SIR) filter (Arulampalam et al. 2002). In this work a SIR particle filter is used for estimating the RUL of rubber-metal-bearings due to the named advantages and the SIR related improvement of particle degeneracy. Particle degeneracy is a weakness of the classical SIS filter. The SIR particle filter is a further development of the SIS filter and prevents that degeneracy by resampling. All these Monte Carlo based filters use random samples that are called particles to estimate the state of the monitored product in the form of a distribution. Therefore, the samples' relevance is symbolized by weights. These weights are calculated based on a defined distribution and a comparison of the predicted and the measured values. In the case of degeneracy after little iteration most of the particle weights tend towards zero while only one particle has a bigger weight. That means that only one particle builds the base for the state estimation and the consecutive estimation of the RUL. Nevertheless all particles are still part of the estimation even if their influence on the result tends towards zero. This degeneracy problem can be solved by resampling. Thereby only relevant samples survive which means samples with a higher weight. Those samples build the base for the next prognostics step while the probably irrelevant particles are no longer considered. In that case a smaller variance of samples is used, but the result is more accurate. The RUL prediction is an iterative method. As long as measured data is available the weights can be updated and resampling can be proceeded (Arulampalam et al. 2002, Jouin et al. 2016).

The general structure of a particle filter is given in Figure 2. The models are developed based on data for training that is presented in chapter 3. Due to the complex, nonlinear behavior and multiple ways of degradation, no physical model of failure for rubber-metal-elements exists. Therefore empirical parameterized models are implemented. For every dataset these parameters are estimated by using Differential Evolution, a population based optimization algorithm (Elsayed et al. 2012). So, every model is related to one bearing. These models are used within the method for state estimation based on samples. Therefore a multi model version of a SIR particle filter is implemented. The general state equation to estimate the samples is given in Equation 1 (Vachtsevanos 2006, Arulampalam et al. 2002) and the particular state equation in Equation 2.

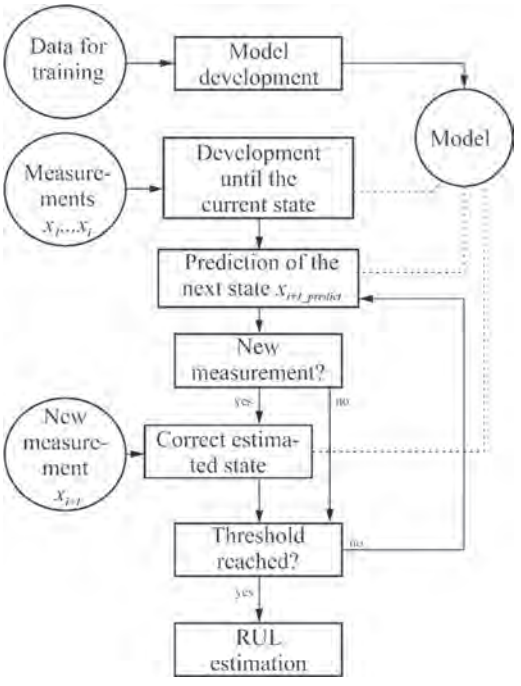


Figure 2. Structure of a particle filter.

$$x_i = f(x_{i-1}, mdl_{i-1}, v_i, t_i); \tag{1}$$

$$x_i = x_{i-1} \cdot \frac{\left[e^{p_{i-1,1} \cdot \left(\frac{t_i}{p_{i-1,2}} \right)^{p_{i-1,3}}} + e^{\frac{p_{i-1,4}}{p_{i-1,5} \cdot t_i + 1}} \right]}{2} + v_i; \tag{2}$$

where x_i is state vector at time ti, mdl_{i-1} is the model with parameters $p_{i-1,1-5}$ and vi is added noise. The model parameters are chosen based on the weights of the previous state vector. In this version the initial samples or initial states are in each case generated by one of the models and the first measurement. By an appropriate number of samples, model choice is equally distributed for the initial sample generation. Therefore different numbers of samples are evaluated in chapter 4. If new measurements are available, the estimated states can be corrected through resampling. With the aim of estimating the RUL, the prediction step is repeated until a given threshold is reached by the samples.

2.2 *Relevant parameters*

Variable parameters of a particle filter influence the accuracy of prognostics. The parameters to be analyzed are number of simulations, number of samples, the measured values and the resampling strategy.

Accuracy of particle filter strongly depends on the number of particles. That is because it is more likely that a big random sample of a defined distribution is able to show a good representation of that distribution than a smaller random sample. To show the influence of variable number of samples on predictions of RUL, three possible numbers of samples should be compared. In this context the number of simulations is analyzed as well.

The measured values in focus are temperatures acquired in or close to the bearing. It was observed that the temperature of rubber-metal-bearings changes over their lifetime, especially in the end of their lifetime. Due to the fact that bearing temperature is influenced by operating conditions, these conditions should be considered. In chapter 4 measurements based on similar conditions are implemented including similar exciter power, similar frequency and similar bearings. Nevertheless, there is one parameter that cannot be kept constant, the ambient temperature. That is why the ambient temperature is measured as well. The relative temperature ΔT involves both temperatures in the form of a subtraction, ΔT = T (bearing) – T (ambient). In the following chapter both measured values, absolute bearing temperature and relative temperature are presented.

To improve the degeneracy problem, resampling can be involved in the particle filter. Multiple resampling schemes exist (Arulampalam et al. 2002, Ignatious, Lincon 2013), in this work the SIR is implemented. One point of interested in this context is the question when to resample. Two possibilities are compared for the application of rubber-metal-bearings. The first continuous strategy enables resampling in every iteration step which is easy to implement but leads to high computational cost. The other strategy is based on a defined threshold for resampling. In this case resampling is only executed if the condition is fulfilled. The realized threshold is based on the effective sample size N_{eff} which is a measure for degeneracy. The effective sample size cannot be computed exactly, therefore an estimate $\hat{N}_{eff}$ of N_{eff} is used here, see Equation 3.

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N_S} \left(\omega_k^i \right)^2} \tag{3}$$

where ω_k^i is the normalized weight (Arulampalam et al. 2002). A threshold needs to be defined which

allows resampling when $\hat{N}_{eff}$ is smaller than that threshold. This resampling strategy needs less computational time because resampling is not realized in every iteration step.

3 LIFECYCLE TESTS

3.1 *Lifecycle tests*

Testing rubber-metal-elements is a complex task. Due to their nonlinear behavior and wide distributions concerning lifetime characteristics of rubber caused by manufacturing, lifetime estimation is not trivial (Steinweger 2006, Wallmichrath et al. 2009). That is why nowadays preventive maintenance based on prior calculated lifetimes, often using linear damage accumulation, is state of the art in applying rubber-metal-elements.

Due to a lack of real data, lifecycle tests are performed to generate data for prognostics. Here accelerated lifecycle tests with an increased excitation force are realized because of the long lifetime of these bearings. In the suspension system of trains they are used for up to 8 years (Bender et al. 2017b). These lifecycle tests are performed on a vibration analysis system using a hydraulic cylinder as exciter. It enables movements of the outer ring of the bearing, whereas the inner ring is fixed. The rubber between those rings allows a small movement. Under this mechanical stress the characteristics of rubber change over time due to degradation. Finding a suitable measure to monitor a rubber-metal-bearing condition is a challenging task due to non-linear rubber characteristics and many possible impacts on the lifetime of rubber. Moreover, the structure of the elements increases the difficulty of installing a sensor for a suitable and reliable measure. In this work the focus is on temperature, a measure that is used in other applications as well, for example ball bearings (Kimotho, Sextro 2015) or subsystems of wind turbines (Crabtree 2011). The correlated concept for temperature measurements in rubber-metal-bearings is introduced in (Bender et al. 2017c). Based on that work, a prototype of a rubber-metal-bearing was developed that enables temperature measurement inside the bearing. Integrating a sensor inside the rubber presents a weakness and could lead to a shorter useful lifetime. Moreover, (Molls 2013) showed that temperature inside the rubber part of rubber-metal-bearings have deviations of maximum 3°C compared to temperature measurements at its surface. Therefore, the used thermocouples are placed inside the outer ring of the bearing close to the surface of the rubber. Little pockets are shaped in the metal, in which the thermocouples are bonded. These pockets protect the sensible thermocouples from external influences. Additionally to the absolute temperature of the bearing, the ambient temperature is measured close to the lifecycle tests.

3.2 *Measurement data*

For temperature measurements sheath thermocouples of type K are inserted in the lifecycle tests that are able to monitor the temperature of the bearing. Moreover, they are robust to weather the conditions of the tests and real applications. Data is measured over the whole lifecycle test including data of the failure state. Prior to the prediction, the measured data is preprocessed for generating empirical models. As shown before these models are based on a combination of e-functions which describes the graph of the measurements. The characteristic graphs of the absolute temperature of three bearings are shown in Figure 3.

In the beginning the absolute temperature of bearings raises strongly, before it fluctuates during the main part of the life of a bearing. Bearing 2 shows a small fall of temperature during that time whereas the temperature of bearing 3 stays almost constant. In the last part all temperature curves rise until the end of lifetime is reached. Analyzing Figure 3, it becomes obvious that in addition to their common characteristics these curves differ in aspects such as starting and ending temperature, lifetimes of bearings and the corresponding graph. This has different reasons based on characteristics of the bearing and operating conditions, especially the ambient temperature. That is why the ambient temperature is involved in the second measured value, the relative temperature. The graph of the relative temperature for bearing 3 is depicted in Figure 4. In general the curve of the relative temperature shows similar characteristics like absolute temperature of bearings during its lifetime. The significant temperature rise in the beginning and in the end is related to the absolute temperature of

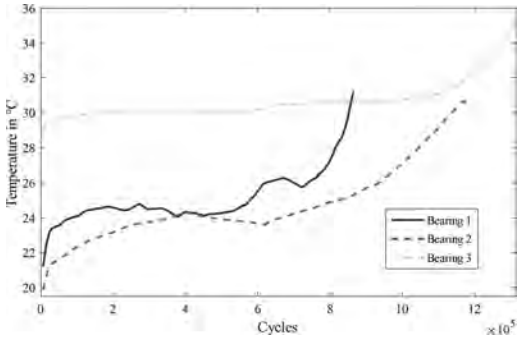


Figure 3. Absolute temperatures of bearings during lifecycle tests.

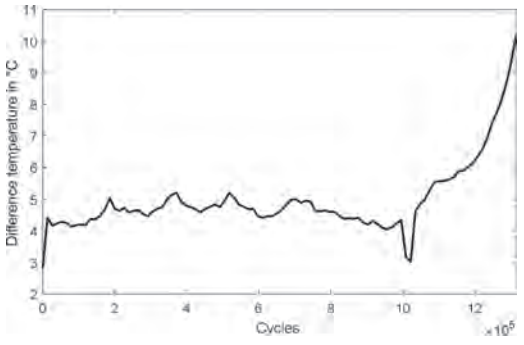


Figure 4. Relative temperature during lifetime of bearing 3.

the bearing. Due to the fact that the ambient temperature fluctuates more easily than the temperature inside the bearing, the relative temperature fluctuates during the main part of the lifecycle test. Moreover a stop of the test after about 10^6 cycles leads to a falling temperature because of a cooling. After starting the test again the temperature of the bearing raises quickly. Due to the similar graphs, all models of both measured values base on the same state equation, only the parameters differ. All in all, the relative temperatures have a more similar value range than the absolute temperatures of bearing. Therefore the models should fit better and might result in an improved prediction. Both measured values are implemented in the following analysis where the influence of the previously mentioned parameters comes under focus.

4 ANALYSIS OF PREDICTIONS

4.1 *Test structure*

In this chapter the presented measured values are used for estimating the RUL with the presented SIR particle filter. To find the best performance different parameters shall be implemented and compared. The following tests are evaluated on:

1. Measured values: absolute temperature of bearings and relative temperature
2. Number of simulations
3. Number or samples
4. Resampling strategy including different thresholds

For evaluating the performance a metric based on relative error is used. This metric is calculated analogue to Equation 4.

$$Error = \frac{RUL_{real} - RUL_{estimated}}{RUL_{real}} \cdot 100\% \qquad (4)$$

where RUL_{real} is the current RUL of the element and $RUL_{estimated}$ is the predicted RUL. In this work the RUL is estimated for different times from 15 to 95% of spent lifetime of the bearings. Thus the error is the mean error calculated as the mean of the RULs from different times. Thereby positive and negative errors compensate each other; therefore the number of negative errors is given in brackets to get an impression of the sign of single RULs. As an example Figure 5 depicts the RULs at the mentioned starting times for bearing 1. Illustrated are the real (grey circles) and the estimated RULs (black squares). The dashed lines symbolize an error band of 15%. Only one error is negative (1 N) which is good. Greater RULs present a too late prediction and thereby a possible unwanted breakdown of the system. The parameters used to generate this result are 100 particles, three simulations, resampling realized in every iteration step and the relative temperature as measured value.

4.2 *Results for bearing temperature*

In this paragraph prognostics base on absolute bearing temperature measurements and the associated models. The first parameter of interest is the number of simulations. Due to the fact that the SIR particle filter is based on probability it is necessary to reach a repeatable result within certain limits. To find a suitable number of simulations three different alternatives between three and 100 simulations are tested and the results are displayed for three test bearings in Table 1. The tests are numbered. An 'a' is added to the name as a symbol for measured value absolute temperature of bearings.

Table 1 shows that prediction of the RUL of these three test bearings do not lead to the same results regarding the best number of simulations. Each of the bearings shows the best performance for another number of simulations. However, the number of negative errors differs only slightly for

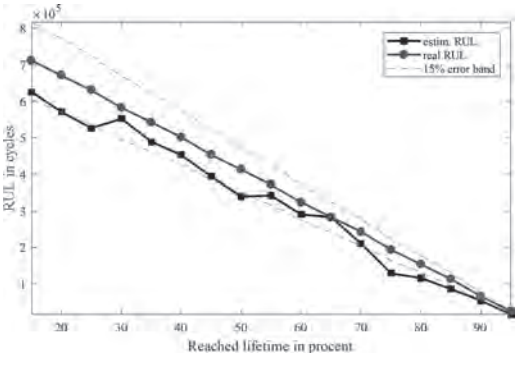


Figure 5. Estimated RUL at different times of bearing 1.

each bearing. SIR particle filter are sensitive to outliers (Arulampalam et al. 2002), that is why a high number of simulations is necessary to compensate outliers. To reach valuable results particle based methods need a suitable (minimum) number of samples or particles. Therefore in the previous simulations 100 particles were implemented.

For that reason 10 and 100 trails are analyzed again for varying number of samples from 100 to 1000 with the aim of improving the result. The performance metrics for the three test bearings are compared in Table 2. Again no consistent results over all bearings exist.

Bearing 3 performs best for 1000 samples, bearing 1 for more than 100 samples and bearing 2 for 100 samples. It indicates that especially bearing 2 leads to unexpected results. Furthermore, only 2/3 of the results exhibit a better performance for 100 simulations. This may be related to the small number of models. However, the difference between the results of variable trails decreases with a growing number of samples. To analyze the influence of the number of samples and resampling thresholds on the performance both parameters are varied in the next step for bearing 1. So far a continuous resampling was implemented; in Table 3 both resampling strategies are realized. The resampling threshold is in a first step based on the mean effective sample size measured during a continuous resampling strategy. In the following steps it is adapted to the performance metric. The chosen number of simulations is ten.

Table 3 underlines that a threshold based resampling strategy is able to improve the performance for both number of samples. While for 100 particles a threshold of around 50 leads to the best performance, for 1000 particles a threshold of 410 is the best. A threshold based resampling strategy can balance worse performance metrics based on a smaller number of simulations. As Table 3 shows, a parameter combination of 10 simulations, 1000 samples and a threshold of 410 (test 15a) leads

Table 1. Influence of number of simulations on prognostics of absolute temperature of bearings.

Test No.	Simulations	Error (bearing 1) in %	Error (bearing 2) in %	Error (bearing 3) in %
1a	3	12.6 (1 N)	–26.0 (17 N)	13.1 (3 N)
2a	10	15.7 (0 N)	–19.1 (17 N)	13.7 (0 N)
3a	100	14.0 (1 N)	–22.6 (17 N)	13.0 (1 N)

Table 2. Influence of simulations and number of samples for absolute temperature of bearings.

Test No.	Simulations	Number of samples	Error (bearing 1) in %	Error (bearing 2) in %	Error (bearing 3) in %
2a	10	100	15.7 (0 N)	–19.1 (17 N)	13.7 (0 N)
3a	100	100	14.0 (1 N)	–22.6 (17 N)	13.0 (1 N)
4a	10	500	13.3 (0 N)	–24.4 (17 N)	10.1 (1 N)
5a	100	500	13.5 (0 N)	–24.3 (17 N)	10.0 (1 N)
6a	10	1000	13.6 (0 N)	–24.5 (17 N)	9.6 (2 N)
7a	100	1000	13.4 (0 N)	–24.0 (17 N)	9.3 (1 N)

Table 3. Influence of number of samples and resampling strategy (bearing 1) for absolute temperature of bearings.

Test No.	Number of samples	Resampling threshold	Error in %
2a	100	–	15,7 (0 N)
8a	100	40	13.3 (1 N)
9a	100	42	13.1 (2 N)
10a	100	45	15.7 (0 N)
11a	100	50	12.0 (2 N)
6a	1000	–	13.6 (0 N)
12a	1000	390	13.8 (0 N)
13a	1000	400	14.5 (0 N)
14a	1000	405	14.1 (0 N)
15a	1000	410	13.4 (0 N)
16a	1000	412	15.5 (0 N)

to a similar result like a parameter combination of 100 simulations, 1000 samples and a continuous resampling strategy (test 7a). In the context of online prognostics this could be a great advantage, since less simulations and a threshold based resampling strategy need less computational cost. However, the error is smaller for 100 samples, but more sensitive to unwanted negative prediction errors.

4.2.1 *Results for a further position of measurement*

Molls suggested that the rubber temperature inside the rubber changes quite similar to the surface temperature (Molls 2013). The measurements in this work show similar temperature behavior between the bearing temperature and the temperature measured at the bolt of a bearing. In Figure 6 these two temperatures are visualized for bearing 3.

This leads to the possibility of testing the method and the models, based on those of absolute temperature of bearings, by new bearings whose bolt temperatures are known. The used parameters are 100 particles, 10 simulations and a resampling threshold of 50 effective samples. The resulting errors are 12.6% (0 N) for bearing I and 4.0% (8 N) for bearing II. Figure 7 depicts

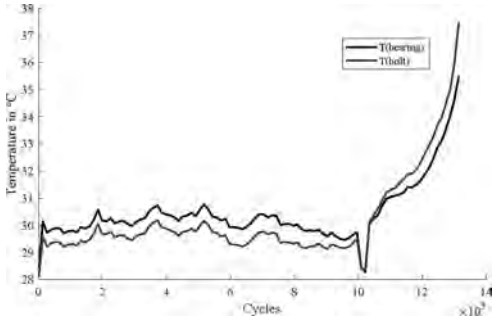


Figure 6. Absolute temperature of bearing and bolt temperature.

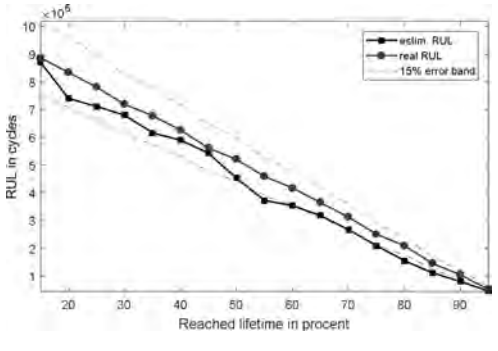


Figure 7. Estimated RULs for bolt temperature of new bearing.

the result of bearing I. The errors are within a 15% error band or just below it. The estimated RULs of bearing II fluctuate stronger as the large number of negative errors shows, while the mean error of 4.0% seems to be good.

These two tests show that temperature measurements close to the rubber-metal-element that have similar characteristics could be used for estimating RULs in the case of missing bearings' temperatures.

4.3 *Results for relative temperature*

In this paragraph predictions based on the relative temperature are evaluated. Because of the former results and the similarities between the two measured values, this evaluation is based only on bearing 1. The implemented models base on relative temperatures measurements. The structure of this analysis is similar to the one in chapter 4.2, the parameters simulations, number of samples and resampling strategy are varied and the performance metric is evaluated. In Table 4 the influence of the number of simulations on the performance of bearing 1 is shown for 100 particles. The names of the numbered tests for this measured value contain an added 'b'. The error falls with increasing number of simulations. The number of negative errors is small and nearly constant as before. The best result is based on 100 simulations that is why the predictions shown in Table 5 include 100 simulations.

Table 5 depicts an improved estimation of the RUL for an increasing number of samples. The best performance of 12.0% is predicted for 1000 particles. The former good result with less simulations and a threshold based resampling ($error_{15a}$) should be examined in the next step for the relative temperature. Table 6 shows the results for varying resampling strategies and thresholds based on

Table 4. Influence of simulations (bearing 1) for the relative temperature.

Test No.	Simulations	Error in %
1b	3	15.4 (1 N)
2b	10	12.7 (0 N)
3b	100	12.4 (0 N)

Table 5. Influence of number of samples (bearing 1) for the relative temperature.

Test No.	Number of samples	Error in %
3b	100	12.4 (0 N)
4b	500	12.2 (0 N)
5b	1000	12.0 (0 N)

Table 6. Influence of resampling strategy (bearing 1) for the relative temperature.

Test No.	Number of samples	Resampling threshold	Error in %
6b	1000	–	12.5 (0 N)
7b	1000	390	12.5 (0 N)
8b	1000	395	12.0 (0 N)
9b	1000	400	11.8 (0 N)
10b	1000	405	11.9 (0 N)
11b	100	45	12.3 (0 N)

10 simulations and 100 or 1000 samples. Similar to Table 3 Table 6 presents possible performance improvements based on suitable thresholds. Due to the small differences regarding the performance metric of different number of samples, the best performance for 100 and 1000 samples were evaluated. An error of 12.3% was achieved for a threshold of 45 using 100 samples. The best threshold of 400 enables an error of 11.8% using 1000 samples. Comparing the threshold based results of 1000 particles and 10 simulations (test 9b) to the one of continuous resampling with 100 simulations (test 5b), the threshold based resampling slightly improves the former results. It can be concluded that a threshold based resampling can lead to an improvement of the performance of the SIR particle filter. At least a saving of computational time is realized.

4.4 *Comparison of the results*

In this part of chapter 4 the results of the two measured values are compared. Regarding the number of simulations, 100 simulations are less outlier prone and therefore lead usually to the best results. Comparing the errors, the measured value relative temperature enables better performance regarding the number of simulations. While the smallest error related to absolute temperature of bearing 1 is 14.0%, an error of 12.4% is related to the relative temperature of bearing 1. The errors are in most cases reduced by an increased number of samples. Once again the relative temperature performs better than the absolute temperature of bearings ($error_{5b}$ = 12.0%, $error_{4a}$ (bearing 1) = 13.3%). In the end the analysis of different resampling strategies emphasis that a threshold based resampling with a suitable threshold leads to a similar good performance with a smaller number of simulations. Moreover, in the case of the relative temperature the performance is slightly improved ($error_{5b}$ = 12.0%, $error_{9b}$ = 11.8%).

All in all estimating RUL for rubber-metal-bearings is possible based on temperature measurements and SIR particle filter for almost constant conditions. In reality applied bearings experience variable changing conditions, e.g. changing excitation. If the operation conditions are changed to a great extent, the similarity between the measurements will not be given. Therefore, implementing or adapting models for different excitation forces seems to be necessary.

5 CONCULSIONS

In this paper a temperature based estimation of the RUL of rubber-metal-bearings is introduced. To evaluate and improve the performance of the SIR particle filter number of simulations, number of samples and resampling strategy are analyzed. The predictions base on two different measured values, absolute temperature of bearings and relative temperature that includes the ambient temperature and absolute temperature of bearings. Predictions based on relative temperature show a better performance than those based on absolute temperature. The reason lies in bigger differences between temperature curves that lead to more variance in the results compared to predictions based on relative temperature. Regarding the parameter, on average 10 to 100 trials and 1000 particles are a good choice for this application. Moreover, both predictions can be improved by a threshold based resampling strategy. It can be concluded that even if rubber-metal-bearings show nonlinear behavior and slightly changing characteristics a threshold based resampling in combination with a suitable threshold enables a relative good RUL estimation based on temperature measurements.

Open questions are related to the threshold of the previous predictions. In this work the end of lifetime is defined by the last measurement. Therefore a threshold needs to be estimated that marks the end of lifetime. Moreover, finding thresholds for effective resampling can be optimized.

ACKNOWLEDGEMENT

This work bases on cooperation with Jörn GmbH. The authors would like to thank especially Mr. Reinke and Mrs. Oberkirsch.

REFERENCES

Arulampalam, M.S.; Maskell, S.; Gordon, N.; Clapp, T. 2002: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* 50 (2), 174–188.

Bender, A.; Kaul, T.; Sextro, W. 2017a: Entwicklung eines Condition Monitoring Systems für Gummi-

Metall-Elemente. InEric Bodden, Falco Dressler, Roman Dumitrescu, Jürgen Gausemeier, Friedhelm Meyer auf der Heide, Christopf Scheytt, Ansgar Trächtler (Eds.): *Wissenschafts- und Industrieforum 2017: Intelligente Technische Systeme.* Paderborn, Germany, 347–358.

Bender, A.; Kimotho, J.K.; Kohl, S.; Sextro, W.; Reinke, K. 2017b: Modellbasierte Prognose der nutzbaren Restlebensdauer von Gummi-Metall-Elementen. In: *Tagungsband der 15. Internationalen Schienenfahrzeugtagung, 2017, vol.* 15. Dresden, Germany, 123–125.

Bender, A.; Sextro, W.; Reinke, K. 2017c: Neuartiges Konzept zur Lebensdauerprognose von Gummi-Metall-Elementen. In: *VDI-Berichte 2301:* VDI Wissensforum GmbH, 49–60.

Crabtree, C.J. 2011: Condition Monitoring Techniques for Wind Turbines. Dissertation. Durham University, Durham. School of Engineering and Computing Sciences.

Elsayed, S.M.; Sarker, R.A.; Ray, T. 2012: Parameter Adaption in Differential Evolution. *IEEE World Congress on Computational Intelligence*.

Ignatious, J.J.; Lincon, S.A. 2013: On the choice of importance of resampling schemes in particle filters. *American Journal of Engineering Research (AJER)* 02 (09), 228–233.

Jouin, M.; Gouriveau, R.; Hissel, D.; Péra, M.-C.; Zerhouni, N. 2016: Particle filter-based prognostics. Review, discussion and perspectives. *Mechanical Systems and Signal Processing* 72–73, 2–31.

Kimotho, J.K.; Sextro, W. 2015: Comparison and ensemble of temperature-based and vibration-based methods for machinery prognostics. In: *Proceedings of the Annual Conference of the Prognostics and Health Management Society 2015, vol.* 6.

Molls, M. 2013: Experimentelle und numerische Untersuchung ein- und mehrachsig belasteter Elastomerbuchsen unter besonderer Berücksichtigung des Reihenfolgeneinflusses. Dissertation. Universität Duisburg-Essen, Duisburg-Essen. Fakultät für Ingenieurwissenschaften, Abteilung Maschinenbau und Verfahrenstechnik.

Spitz, M. 2012: Modellbasierte Lebensdauerprognose für dynamisch beanspruchte Elastomerbauteile. Fakultät für Ingenieurwissenschaften, Abteilung Maschinenbau und Verfahrenstechnik der Universität Duisburg-Essen, Duisburg-Essen.

Steinweger, T. 2006: Lebensdauerberechnung und Lebensdauerprüfung von Elastomerbauteilen unter mehrachsiger dynamischer Belastung. Schlussbericht. With assistance of U. Weltin, M. Flamm, M. Kirstein, T. Steinweger. Technische Universität Hamburg-Harburg.

Vachtsevanos, G.J. 2006: Intelligent fault diagnosis and prognosis for engineering systems. Hoboken, N.J: Wiley.

Wallmichrath, M.; Lücker, E.; Jöckel, M. 2009: Elastomerbauteile – Charakterisierung und Prüfung. In: Deutscher Verband für Materialforschung und -prüfung e.V. (Ed.): *Elastomerbauteile.* DVM-Tag 2009, vol. 676. DVM-Tag. Berlin, 22.-24.4.2009: DVM (DVM-Bericht, 676), 171–180.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Prognostic and health management for safety barriers in infrastructures: Opportunities and challenges

A. Zhang, Y. Liu & A. Barros
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway

Y. Wang
School of Economics and Management, Tianjin Chengjian University, Tianjin, China
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: Different types of safety barriers are deployed in many infrastructures to reduce the occurrences of hazards, and protect people, environment and other assets in case the unexpected events have occurred and the capacity of these barriers against hazards can be weakened by degradations or the failures related to changes over time. It is natural to adapt the approaches of Prognostic and Health Management (PHM) to monitor the conditions and measurable parameters of safety barriers, and predict their future performance by assessing the extent of degradations. This study aims to identify the uniqueness and possible challenges when implementing PHM on safety barriers. Definitions and classifications of safety barriers will be discussed with considering their installation environment in infrastructures, in order to reveal what kind of characteristics of barriers can lead to higher demand on prognosis and heath monitoring. Another objective of this paper is to review the qualitative and quantitative measures for the capacity and performance of safety barriers, and to explore the possible methods and research gaps in the assessments for different PHM strategies, taking account their effects on safety barriers, and effects on the infrastructures being protected by the barriers.

1 INTRODUCTION

Maintenances can be defined as the activities to keep a system in a working order (Do et al. 2015). With the development of sensor technologies, the maintenances for many complex systems involve more and more condition-based and preventive activities to reduce maintenance costs on one hand, and improve their performance on the other hand (Sharma et al. 2017, Liu et al. 2017). Prognostics and Health Management (PHM), including fault detection, diagnostics, prognostics and health management, is a developing approach that enables real-time health assessment of a system and predicts of its future state based on up-to-date information. PHM has been conducted in many applications including manufacturing, aerospace systems, railway, energy, and military industry (Sun et al. 2012, Pecht and Rui 2010).

Safety barriers are installed in many critical systems and infrastructures to prevent hazardous events or mitigate their consequences, such as fire prevention systems and railway signaling systems. Technological safety barriers, such as shutdown valves in process, and airbags on cars, are also called as safety-critical system (Rausand 2014). But these safety barriers can also degrade and fail to accomplish their safety function under the evolving environment (Zio 2016). In case of failures of the barriers, serious accidents or disaster may occur. Many studies have been carried out on the operational and performance analysis of the safety barriers (Innal et al. 2015, Duijm and Goossens 2006, Innal et al. 2015, Rahimi et al. 2011, Cai et al. 2012), and most of them assume that the failures of components in the safety barriers follow the exponential distribution (Guo and Yang 2008, Jin and Rausand 2014, Catelani et al. 2011, Liu and Rausand 2011), meaning that their failure rate keep constant in any time.

According to IEC 61508 (2010) and IEC 61511 (2003), many technological safety barriers consist of three subsystems: sensor(s), logic solver(s) and actuating unit(s). The mechanical actuating units can degrade due to corrosion and wear-out etc, become more vulnerable along with time (Zio 2016), and so that the assumption of exponential distribution of failures is challenged. Based on this concern,

a growing attention is given to the predict degradations of safety barriers and offer suitable maintenances in advance to ensure the barrier adequacy. PHM can be a helpful approach in performance prediction and decision-making for maintenances.

The purpose of this paper is to review the techniques of PHM and designing and operational characteristics of safety barriers, so as to explore the research issues when the PHM approach is planned to be implemented for improving the integrity of safety barriers.

The remained of this paper is organized as follows: In section 2, the development and advantages of PHM are introduced; Section 3, includes the review of safety barriers in infrastructure and introduces technological barriers; Section 4 introduces several unmet problems and challenges related to using PHM on safety barriers. A conclusion is given in Section 5.

2 PROGNOSTICS AND HEALTH MANAGEMENT

2.1 *Development of PHM*

PHM is developed based on the concept of Condition-based maintenance (CBM). CBM is an approach to carry out maintenance actions based on the information collected through condition monitoring on systems in contrast to breakdown or time-based preventive maintenance. In order to make a timely decision on maintenance, prognostics is the key technology for CBM (Jardine et al. 2006, Shin and Jun 2015, Bousdekis et al. 2015). From this point, PHM is developed from the concept of CBM. A CBM program consists of three key steps (see Figure 1) (Lee 2004):

1. Data acquisition step;
2. Data processing step;
3. Maintenance decision-making step

Diagnostics and prognostics are two aspects in CBM. Diagnostics deals with fault detection, isolation and identification when it occurs (Jardine et al. 2006). Prognostics, in ISO-13381 (2015), is to estimate the time to failure and risk for one or more existing and future failure modes. The relative placement of detection, diagnostic and prognostic can be explained in Figure 2 (Gouriveau and Medjaher 2011).

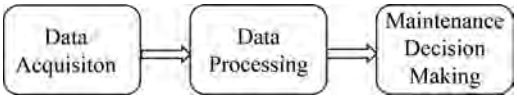


Figure 1. Three steps in CBM.

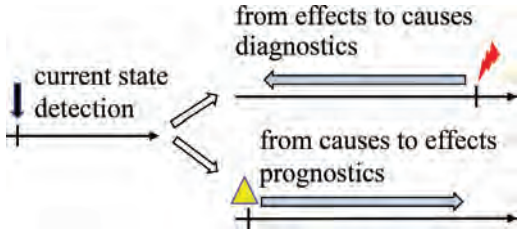


Figure 2. Complementarity of detection diagnostic and prognostic activities (Gouriveau, 2011).

In literature, prognostics is a process of health assessment and prediction, which includes incipient fault/failure detection, performance monitoring, life cracking and predicting residual useful lifetime (RUL) (Hess et al. 2005, Lee et al. 2014);

PHM is the extension of prognostics. According to CALCE (Center for Advanced Life Cycle Engineering) (2012), PHM is the means to predict and protect the integrity of equipment and complex systems, and avoid unanticipated operational problems leading to mission performance deficiencies, degradation, and adverse effects to mission safety.

Sun et al. (2010) regards PHM as a methodology to predict when and where failures will occur and to mitigate risks through evaluating the reliability of a system in its actual life cycle conditions. It is an enabling discipline of solving reliability problem in the process of design, manufacturing, operational and maintenance (Pecht and Jaai 2010). PHM is aiming to all information of an equipment in past, present and future while considering its environmental, operational and usage condition so as to detect its degradation, diagnose fault and predict and manage failures (Zio 2012).

Haddad (Haddad et al. 2012) regards PHM as a discipline that can used for: (i) evaluating the reliability of systems of their life cycle; (ii) determining the possible occurrence of failures and risk reduction; (iii) highlighting the Remanding Useful Lifetime (RUL) estimation. Actually, modern and comprehensive PHM systems take many issues into consideration, such as fault detection, fault isolation, useful life remaining, and performance degradation trending and then provides a broader set of maintenance benefits than any function by itself (Hess et al. 2005).

In this paper, we understand PHM as an approach to carry out dynamic management based on RUL which is predicted by status information collected through actual life cycle conditions, including environmental, operational and usage conditions.

2.2 *PHM architecture*

PHM means a complete process from capturing the data to decision-making (in maintenance,

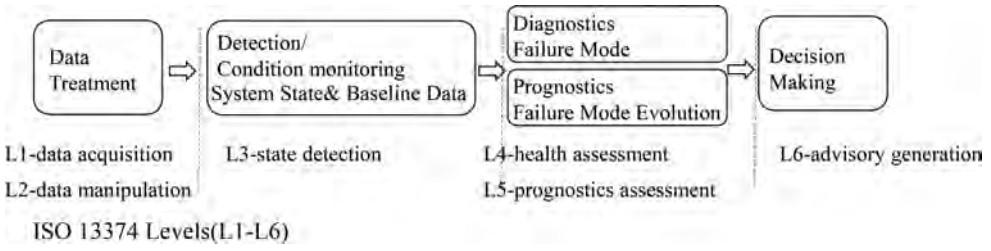


Figure 3. General process of PHM. Correlation with ISO 13374 (Guillén, 2016).

life time control, equipment design, etc.) (Guillén, Crespo, Macchi, & Gómez 2016), which is originally conceived by ISO 13374 and gradually becomes a standard in OSA-CBM (Open System Architecture for Condition Based Maintenance). As shown in Figure 3. The whole process of PHM is based on that of CBM, and can be divided into two parts. The first part (from Level 1/L1 to Level 5/L5) is related to health monitoring and prognostics, and the second part (Level 6/L6) is for health management.

In such a process, PHM attempts to answer several questions, e.g.:

- How is the status of system now? (Performance assessment).
- When will the system fail? (Remaining useful lifetime).
- What will the primary faults that cause system failure?
- Why does the incipient fault occur?

2.3 *PHM methodologies*

To answer the above questions, prognostics is currently carried out in different ways, namely with model-based, data-driven and hybrid prognostics (Brahimi et al. 2016).

The model-based approaches are based on a good knowledge of the physics of system and the available failure modes. Analysts can construct mathematical models with the above knowledge, and analyze those systems whose field operational and failure data is not enough (Lee et al. 2014, Luo et al. 2003). However, for many complex systems, one of limitations of the model-based approaches is the difficulty to create deliberate models representing the multiple physical processes (Pecht 2008). Moreover, it is very difficult to adopt the models built for some specific applications to the others, even though the systems are very similar.

The data-driven approaches are based on statistics and machine-learning techniques (Gu et al. 2007). In data-driven the remaining useful life would be predicted by fitting the monitoring data of developing fault to the degradation mechanism before it reaches the predetermined threshold level (e.g., see (Medjaher et al. 2012)). These methods are relatively simple to deploy due to the necessary of an analytical model of behavior and failure of the system.

The hybrid approaches are proposed in consideration of the pros and cons of the previous two groups (Lee 2004), in which prognostics results are claimed to be more reliable. The hybrid approaches have been used for the RUL prediction and maintenance of systems, such as (Kumar et al. 2008, García et al. 2010, Skima et al. 2015, Zhang et al. 2009).

PHM has been conducted in many areas, such as the infrastructures, aerospace industry, and in this paper we focus on the approach for safety barriers.

3 SAFETY BARRIERS

3.1 *Safety barriers and classification*

Safety barriers, or simply barriers are the equipment and features that are installed to protect people, the environment and other assets against harm should features or deviations occur in the most-designed system (Rausand 2013). Safety barriers are always related with a certain safety functions, which are defined by Sklet (2006) as the functions planned to prevent, control, or mitigate undesired events or accidents.

Figure 4 is a Bowtie diagram widely used in the field of risk analysis, where we can identify the two different roles of safety barriers. A hazardous event can occur due to some causes, so that some barriers can be located on the left side of the diagram (the causes side), to reduce the probability of the hazardous event. This kind of barriers are called as proactive barriers or prevention barriers, such as antilock braking system, electronic stability control system in automobiles. On the right side, some barriers are located on the right side (the consequences side), in case of the occurrence of a hazardous event, for reducing its effects or failure

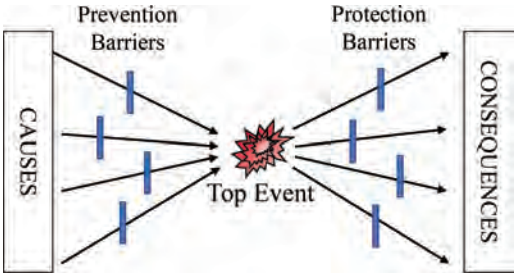


Figure 4. Bowtie diagram for a Top Event with prevention and protection barriers.

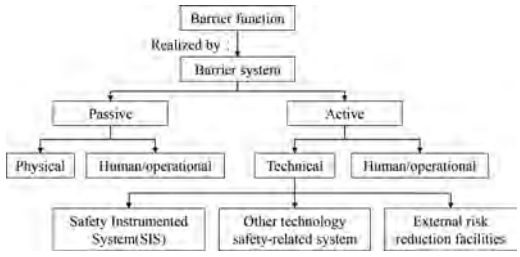


Figure 5. Classification of safety barriers (adopted from Sklet, 2006).

escalation, and they are regarded as reactive barriers, or protection barriers, e.g. seat belts, airbag systems (Hollnagel 2004, Rausand 2013, Groot 2016).

This classification is based on the objectives or functions of barriers. In addition, considering the operational modes of barriers, Rausand (2013) has distinguished safety barriers as passive and active barriers. An active barrier is dependent on some energy sources and a sequence of detection-diagnosis-action to perform its function, such as an airbag, Meanwhile, a passive system is not required to take an action and just by the presence of their elements to achieve its function (e.g. a seat belt).

Safety barriers also can be divided into on-line and off-line barriers. The on-line barriers operate continuously or so often, and on the contrary, the off-line ones are only used intermittently or infrequently. In practices, most protective barriers are off-line ones (Rausand & Arnljot 2004).

Sklet (2006), on the other hand, considers who are carrying out safety functions, and classifies barriers as the physical, technical, and human/operational barriers. Combining with the categorization based on the operational modes, we can obtain Figure 5. In the figure, technical barriers are always active. They are further divided into three groups: Safety Instrumented System (SIS), meaning that a technical barrier which involves the electric, electronic, and programmable electronic (E/E/PE) technologies, other technology safety-related systems and External risk reduction facilities. In the rest of this paper, we focus on technical barriers.

3.2 *Technological barriers*

A technological barrier, involving E/E/PE technologies and some mechanical items, generally consists of three subsystems: input element subsystem (e.g., sensors, transmitters), logic solver subsystem (e.g., programmable logic controllers [PLC]) and finial element subsystem (e.g., safety valves, circuit breakers). The main parts are illustrated in Figure 6.

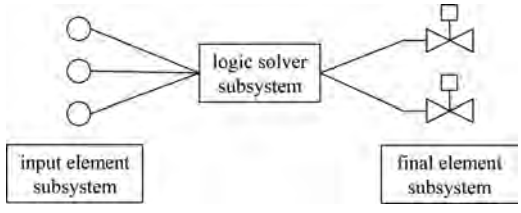


Figure 6. Main parts of a technological barrier.

The system protected by a technological barrier is called the Equipment Under Control (EUC). A safety-instrumented function (SIF) is a function that has been designed to protect the EUC against a specific demand. To enhance the reliability of a barrier, redundancy is often implemented in the system configuration.

4 DEVELOPMENT A PHM FOR SAFETY BARRIERS

We can introduce the PHM to safety barriers, with the purpose to assess the degree of deviation or degradation of barriers, and then plan maintenances in advances, so as to improve their availability and bring safety to EUC.

4.1 *Main functions of PHM on barriers*

Compared to the existing diagnostics of barriers, PHM is expected to predict failures from incipient failures or deviations in components. The main functions and potential benefits of the PHM on barriers can include:

- Advance warning of failures—Prognostics in PHM can evaluate the degradations of barriers, so as to detect incipient deviations. It is possible for maintenance staffs with prognostic results regarding the operational conditions to take actions on a barrier before a failure really occurs.

- Optimized maintenances—With prognostics, maintenance staffs also can estimate the remaining life of a component, especially a mechanical one, in a barrier, and then develop a maintenance, repair or replacement plan. Compared with scheduled maintenances, these condition-based and predictive maintenances eliminate unnecessary activities, and keep the barrier effective.
- Logistic support and cost reduction—Ideal prognostics tell the maintenance staffs when and where failures will occur, and thus they can identify and fix the failed components easily. PHM can reduce lead time and therefore increase the available time of safety barriers. Moreover, the "just-in-time" maintenances based on prognostics decrease the unnecessary costs of scheduled inspections and interruptions.

4.2 *Challenges of PHM on barriers*

Although PHM has been proved in many applications, we may meet challenges when we implement PHM on the technological safety barriers, due to the following design and operational characteristics of barriers:

4.2.1 *Operational modes of barriers*

Current PHM is always used for systems continuously running, while safety barriers have several operational modes in stead:

- Low-demand mode: where the safety function is only performed on demand, and where the frequency of demands is relatively low;
- High-demand mode: where the mechanism is same as low-demand, but the frequency of demands is relatively high;
- Continuous mode: where the safety function is a part of normal operation.

In the latest version of IEC 61508 (2010), the borderline between low-demand and high-demand is once per year in terms of demand frequency.

For those technologies barriers with demanding operational modes, they are usually in a dormant state and transit to an active state in case that demands come. The degradation mechanisms in different states are varied. Not many studies have been conducted so far on degradation prediction with state transitions. We need new approaches of parameters to predict the future performance of a barrier in response to demands during the durations of demands.

4.2.2 *Structures of barriers*

Redundancy structures are often used in barriers to improve availability and to enhance safety, e.g., two shutdown valves are installed in parallel to stop flow when the downstream pressure is too high. When one of them cannot activate, the process is still safe if the other works. Such kind of structures is called as 1-out-of-2 configuration. For a system with N channels, if at least K of the N channels need to be functional to ensure that the system is functional, the system has a K-out-of-N (Koo$N$) configuration.

Many barriers can be adaptive, meaning that they can change their configurations to perform safety functions when some expected occur. For example, a 2oo3 barrier can automatically transit to a 2oo2 configuration when one of the three channels fail. The challenge for PHM is to predict the effects of degradations in one channel on the entire barrier system with complex configuration and adaptivity, as well on the EUC.

4.2.3 *Failure modes and tests of barriers*

Failures of technological barriers can be classified as dangerous (D) failure and safe (S) failure. D failure refer to a failure that has the potential to put the barrier in a hazardous or fail-to-function state, while S failure does not leave the barrier in fail-to-function state (Rausand 2014), e.g. a valve shuts down unnecessarily.

The integrity of a technological barrier is highly related with tests, especially for those running in the low-demand mode. Regular proof tests are conducted on technological barriers (e.g. once per year), to reveal failures and then initiate maintenance activities if necessary. Many modern safety barriers have installed automatic self-testing modules, which has a diagnostic function and detects some failures. The D failures that can be found in diagnostic tests are called as dangerous detected (DD) failures, such as signal loss, signal out of range and final element in wrong position (Rausand 2014). The D failures that are not detected are called dangerous undetected (DU) failures. DU failures are only revealed in proof tests with regular intervals.

A research challenge of PHM is therefore to find suitable approaches to link the incipient failures or deviations with those D failures of interest in integrity of barriers. Most data-driven PHM approaches depend on the historical/training data to predict the trends of failure, but in those published data sources for technological barriers, such as Offshore Reliability Data (OREDA) and Process Equipment Reliability Data, we cannot find any clues. For model-based PHM approaches, no guidance is given to deal with those DU failures.

Another challenge is from the failure occurring in the redundancy structures. Common cause failures (CCFs) are the main contributor of the unavailability of redundant safety barriers (Hauge

et al. 2015). CCFs are the failures of multiple components simultaneously or with a short time interval due to a shared root cause or a common cause. It is valuable to identify those deviations that can lead to CCFs and predict their potential influences in PHM.

4.2.4 *Measures of technological barriers*

IEC 61508 (2010) suggests the average probability of failure on demand (PFD) as a measure for technological barrier of low-demand, and the probability of failure per hour (PFH) as the measure for technological barrier of high-demand. And then, for different results of PFD and PFH, safety barriers can be located at different integrity levels (SILs), from the loose SIL 1 to the strictest SIL 4. These measures are widely used, and they are calculated always on the basis of some basic assumptions (Jin and Rausand 2014, Wang and Rausand 2014, Rausand 2014), including: (1) each failure is assumed to occur at a constant rate (i.e. exponential distributed failures); and (2) the channels in a redundant structures are identical and independent.

We release these assumptions when implementing PHM, and so weaken the theoretical foundations of measure calculations, since we have realized that deteriorations in mechanical components of a technological barrier is unavoidable. However, to evaluate the effectiveness of a PHM program, we still need to utilize the widely accepted measures, and build a relationship between SILs and effects of PHM.

4.2.5 *Cost-benefit analysis of PHM*

Safety and availability are dominator in the assessment of safety barriers. But for PHM, the return-on-investment (ROI) needs to be considered (Saxena et al. 2008, Wang and Pecht 2011), especially for the fact where other test and diagnostics are also employed on safety barriers.

The main work for ROI analysis or cost-benefit analysis is to quantify the costs and benefits of PHM (Scanff et al. 2007). The costs of a PHM program can includes: the cost of acquisition and installation for data, such as sensors and microprocessors, the cost of re-design of host product, which can be a big investment (Sun et al. 2012). The benefit is more complex including the decrease of proof tests and maintenances. It is challenging on how we choose the indicators to calculate the ROI of a PHM program. Moreover, we also need to determine the best PHM program for a specific technological barrier.

5 CONCLUSIONS

In this paper, a short review of PHM is presented. PHM enables estimating the RUL of the in-service equipment which can provide timely decision for maintenance. Due to the vital role of technological barriers and the advantages of PHM, an idea for developing a PHM system for SIS is presented. Compared with mechanical systems, technological barriers have their own characteristics which propose new challenges.

Therefore, we propose several research topics to be addressed in future, specifically in a PhD project:

- New approaches for predicting degradations of a component with state transitions;
- Mechanism of incorporating redundancy structures and varied configurations in degradation modeling and analysis;
- Models to link the effectiveness of PHM with the measures for safety barriers;
- Methods to optimize PHM and other maintenance activities under the constraints of SIL requirements by safety barriers.

REFERENCES

Bousdekis, A., B. Magoutas, D. Apostolou, & G. Mentzas (2015). A proactive decision making framework for condition-based maintenance. *Industrial Management & Data Systems 115*(7), 1225–1250.

Brahimi, M., K. Medjaher, M. Leouatni, & N. Zerhouni (2016). Development of a prognostics and health management system for the railway infrastructure review and methodology. In *Prognostics and System Health Management Conference (PHM-Chengdu), 2016*, pp. 1–8. IEEE.

Cai, B., Y. Liu, Z. Liu, X. Tian, H. Li, & C. Ren (2012). Reliability analysis of subsea blowout preventer control systems subjected to multiple error shocks. *Journal of Loss Prevention in the Process Industries 25*(6), 1044–1054.

CALCE (2012, March). http://www.prognostics.umd.edu/PHM_March_Newsletter_Final.pdf.

Catelani, M., L. Ciani, & V. Luongo (2011). A simplified procedure for the analysis of safety instrumented systems in the process industry application. *Microelectronics Reliability 51*(9), 1503–1507.

Do, P., A. Voisin, E. Levrat, & B. Iung (2015). A proactive condition-based maintenance strategy with both perfect and imperfect maintenance actions. *Reliability Engineering & System Safety 133*, 22–32.

Duijm, N.J. & L. Goossens (2006). Quantifying the influence of safety management on the reliability of safety barriers. *Journal of Hazardous Materials 130*(3), 284–292.

García, C.M., T. Escobet, & J. Quevedo (2010). Phm techniques for condition-based maintenance based on hybrid system model representation. In *Annual Conference of the Prognostics and Health Management Society*.

Gouriveau, R. & K. Medjaher (2011). Chapter 2: Prognostics. Part: Industrial Prognostic—An Overview. In C.B.J. Andrews and L. Jackson (Eds.), *Maintenance Modelling and Applications*. ISBN: 978-82-515-0316-7, pp. 10–30. Det Norske Veritas (DNV).

Groot, A. (2016, 01). Advanced process safety barrier management by applying proactive incident investigation to failed or impaired barriers.

Gu, J., N. Vichare, T. Tracy, & M. Pocht (2007, Jan). Prognostics implementation methods for electronics. In 2007 *Annual Reliability and Maintainability Symposium*, pp. 101–106.

Guillén, A., A. Crespo, M. Macchi, & J. Gómez (2016). On the role of prognostics and health management in advanced maintenance systems. *Production Planning & Control 27*(12), 991–1004.

Guo, H. & X. Yang (2008). Automatic creation of markov models for reliability assessment of safety instrumented systems. *Reliability Engineering & System Safety 93*(6), 829–837.

Haddad, G., P.A. Sandborn, & M.G. Pecht (2012). An options approach for decision support of systems with prognostic capabilities. *IEEE Transactions on Reliability 61*(4), 872–883.

Hauge, S., A. Hoem, P. Hokstad, S. Habrekke, & M.A. Lundteigen (2015). Common cause failures in safety instrumented systems.

Hess, A., G. Calvello, & P. Frith (2005, March). Challenges, issues, and lessons learned chasing the "big p". Real predictive prognostics. part 1. In *2005 IEEE Aerospace Conference*, pp. 3610–3619.

Hollnagel, E. (2004). *Barriers and accident prevention*. Aldershot, Hampshire, England.

IEC 61508 (2010). Functional safety of electrical/electronic/programmable electronic safety-related systems. part 1–7.

IEC 61511 (2003). Functional safety—safety instrumented systems for the process industry sector.

Innal, F., Y. Dutuit, & M. Chebila (2015). Safety and operational integrity evaluation and design optimization of safety instrumented systems. *Reliability Engineering & System Safety 134*, 32–50.

ISO 13381-1:2015 (2015). Condition monitoring and diagnostics of machines—prognostics—part 1: General guidelines.

Jardine, A.K., D. Lin, & D. Banjevic (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing 20*(7), 1483–1510.

Jin, H. & M. Rausand (2014). Reliability of safetyinstrumented systems subject to partial testing and common-cause failures. *Reliability Engineering & System Safety 121*, 146–151.

Kumar, S., M. Torres, Y.C. Chan, & M. Pecht (2008, June). A hybrid prognostics methodology for electronic products. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 3479–3485.

Lee, J. (2004, 01). An integrated platform for diagnostics, prognostics and maintenance optimization.

Lee, J., F. Wu, W. Zhao, M. Ghaffari, L. Liao, & D. Siegel (2014). Prognostics and health management design for rotary machinery systems reviews, methodology and applications. *Mechanical systems and signal processing 42*(1), 314–334.

Liu, B., Z. Liang, A.K. Parlikad, M. Xie, & W. Kuo (2017). Condition-based maintenance for systems with aging and cumulative damage based on proportional hazards model. *Reliability Engineering & System Safety*.

Liu, Y. & M. Rausand (2011). Reliability assessment of safety instrumented systems subject to different demand modes. *Journal of Loss Prevention in the Process Industries 24*(1), 49–56.

Luo, J., M. Namburu, K. Pattipati, L. Qiao, M. Kawamoto, & S. Chigusa (2003, Sept). Model-based prognostic techniques [maintenance applications]. In *Proceedings AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference*. pp. 330–340.

Medjaher, K., D.A. Tobon-Mejia, & N. Zerhouni (2012). Remaining useful life estimation of critical components with application to bearings. *IEEE Transactions on Reliability 61*(2), 292–302.

Pecht, M. (2008). *Prognostics and health management of electronics*. Wiley Online Library.

Pecht, M. & R. Jaai (2010). A prognostics and health management roadmap for information and electronics-rich systems. *Microelectronics Reliability 50*(3), 317–323.

Pecht, M. & K. Rui (2010). Diagnostics, prognostics and system's health management. *PHM Centre, City Unversity of Hong Kong*, 7–23.

Rahimi, M., M. Rausand, & M. Lundteigen (2011). Management factors that influence common-cause failures of safety-instrumented systems in the operational phase. *Advances in Safety, Reliability, and Risk Management, ESREL 2011*, 2036–2044.

Rausand, M. (2013). *Risk assessment: theory, methods, and applications*, Volume 115. John Wiley & Sons.

Rausand, M. (2014). *Reliability of safety-critical systems: theory and applications*. John Wiley & Sons.

Rausand, M. & H. Arnljot (2004). *System reliability theory: models, statistical methods, and applications*, Volume 396. John Wiley & Sons.

Saxena, A., J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, & M. Schwabacher (2008). Metrics for evaluating performance of prognostic techniques. In *Prognostics and health management, 2008. phm 2008. International conference on*, pp. 1–17. IEEE.

Scanff, E., K. Feldman, S. Ghelam, P. Sandborn, M. Glade, & B. Foucher (2007). Life cycle cost impact of using prognostic health management (phm) for helicopter avionics. *Microelectronics Reliability 47*(12), 1857–1864.

Sharma, P., M.S. Kulkarni, & V. Yadav (2017). A simulation based optimization approach for spare parts forecasting and selective maintenance. *Reliability Engineering & System Safety*.

Shin, J.-H. & H.-B. Jun (2015). On condition based maintenance policy. *Journal of Computational Design and Engineering 2*(2), 119–127.

Skima, H., K. Medjaher, C. Varnier, E. Dedu, & J. Bourgeois (2015, March). Hybrid prognostic approach for micro-electro-mechanical systems. In *2015 IEEE Aerospace Conference*, pp. 1–8.

Sklet, S. (2006). Safety barriers: Definition, classification, and performance. *Journal of loss prevention in the process industries 19*(5), 494–506.

Sun, B., S. Zeng, R. Kang, & M. Pecht (2010). Benefits analysis of prognostics in systems. In *Prognostics and Health Management Conference, 2010. PHM'10*, pp. 1–8. IEEE.

Sun, B., S. Zeng, R. Kang, & M.G. Pecht (2012). Benefits and challenges of system prognostics. *IEEE Transactions on reliability 61*(2), 323–335.

Wang, W. & M. Pecht (2011). Economic analysis of canary-based prognostics and health management. *IEEE Transactions on Industrial Electronics 58*(7), 3077–3089.

Wang, Y. & M. Rausand (2014). Reliability analysis of safety-instrumented systems operated in high-demand mode. *Journal of Loss Prevention in the Process Industries 32*, 254–264.

Zhang, H., R. Kang, & M. Pecht (2009, Dec). A hybrid prognostics and health management approach for condition-based maintenance. In *2009 IEEE International Conference on Industrial Engineering and Engineering Management*, pp. 1165–1169.

Zio, E. (2012). Prognostics and health management of industrial equipment. *Diagnostics and prognostics of engineering systems: methods and techniques*, 333–356.

Zio, E. (2016). Some challenges and opportunities in reliability engineering. *IEEE Transactions on Reliability 65*(4), 1769–1782.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A deep variational auto-encoder based dimensionality reduction for fault diagnosis in ball bearings

G.A. San Martín & V. Meruane
Department of Mechanical Engineering, University of Chile, Santiago, Chile

E. López Droguett
Center for Risk and Reliability, University of Maryland, College Park, MD, USA
Department of Mechanical Engineering, University of Chile, Santiago, Chile

M.C. Moura
Department of Production Engineering, Federal University of Pernambuco, Recife, Brazil

ABSTRACT: One of the main challenges faced by the industry in the context of failure diagnosis is the high quantity and high dimensionality of the available data. Due to the increasing capability and availability of sensing technology, nowadays it is possible to acquire a large amount of (unlabeled) data on many operational and maintenance related variables from monitored machines. The problem lies on how to extract useful information from such data. A standard approach in fault diagnosis is to first apply a dimensionality reduction method. In this paper, we propose a method for dimensionality reduction based on Variational Auto-Encoders (VAEs). VAEs have shown good results in areas such as image processing, image generation and speech processing. In particular, in this paper, the VAE based method works on spectrograms generated from vibration signals measured during system's operation. This approach is applied to the fault diagnosis of ball-bearings.

1 INTRODUCTION

The early detection of faults in machinery is nowadays one of the main challenges faced by industrial sectors. In general, faults in machinery components yield to one of two possible results. Either the component is replaced according to the manufacturer instructions, which in the majority of the cases leads to a less than optimal useful life of the element, or the sudden failures occur that causes a performance drop in the machine. Nowadays, the development of new methods that can identify faults early and that make use of the entire useful life of the component is an active topic of research (Hasani, Wang, & Grosu, 2017; Liu, Li, & Ma, 2016).

One possible approach to this problem is to assume that data measured during the operation of a machine contains useful information about its health state and that such information can be extracted if a proper technique is used. Machine Learning has been used in the past to excerpt useful information from fault data (Blum & Langley, 1997; Cireşan, Meier, & Schmidhuber, 2012).

Nevertheless, the use of Machine Learning in conjunction with operational and maintenance data brings another problem to the table: the curse of dimensionality. Usually, data obtained during the operation of a system tends to be noisy and high dimensional. The use of a dimensionality reduction techniques is a popular strategy to improve diagnosis or prognosis performance; see for example (Shuang, Automation, ICMA, & 2007, n.d.) for Support Vector Machines and (Pan, Rust, Networks, IJCNN, & 2000, n.d.) for Artificial Neural Networks (ANN).

In fact, one type of model that has been developed with the objective of performing dimensionality reduction within the Machine Learning field are Auto-Encoders (AEs) (Hinton & Salakhutdinov, 2006). AEs are unsupervised models based on the use of neural networks. They are trained passing the data through an intermediate layer that is of lower dimensionality than the input, thus generating a latent representation and performing a dimensionality reduction. Then, the rest of the neural network tries to reconstruct the input from this reduced representation, forcing this latent space to be meaningful. In the area of fault diagnosis, AEs and its variants have been used recently with great success (Liu et al., 2016; Zhou, Gao, & Wen, n.d.).

Another model that also uses the idea of reducing the dimensionality of the data feeding it through a pipeline that generates a latent representation are Variational Auto-Encoder (VAE) (Kingma & Welling, 2013). VAEs nowadays are one the most auspicious unsupervised machine learning techniques, mainly because of their success in the processing of complex data such as images (Rezende, Mohamed, & Wierstra, 2014; Salimans, Kingma, & Welling, 2015) and speech (Hsu, Zhang, & Glass, 2017). VAEs use a combination of variational inference (Blei, Kucukelbir, & McAuliffe, 2017) and neural networks to solve the difficult problem of finding a good approximation of a posterior distribution in a Bayesian model. However, not attempt has been made to develop a dimensionality reduction method using VAE as the underlying model for fault diagnosis.

Therefore, this paper presents a VAE based dimensionality reduction approach for fault diagnosis in machinery using spectrogram images extracted from vibration signals, and evaluate it by comparing the classification results between the proposed approach and the case where no reduction in dimensionality is performed.

This paper is organized as follows. Section 2 introduces the VAE model. Section 3 discusses the proposed VAE's architecture for performing fault diagnosis. Then, Section 4 presents an example of application. Finally, Section 5 draws some concluding remarks.

2 VARIATIONAL AUTO-ENCODERS

Variational Auto-Encoders (VAEs) are generative models originally proposed by (Kingma & Welling, 2013) that are built on top of the concept of variational inference (VI) (Blei et al., 2017). VI is an approach used for solving the problem of approximating difficult to compute probability distributions. In a Bayesian framework, the observations x are assumed to be produced by a set of latent or hidden variables z. In general, the objective is to compute $p(z|x)$ because of two main reasons. First, for an observation x_i, the vector of latent variables that produce such observation can be of interest in itself, for example, when the latent variables represent physical magnitudes that are not easily measurable. Second, using $p(z|x)$ and the Bayes theorem, one would be able to compute $p(x)$ which represents the model that generates the data itself. This is of special interest for generative models where the objective is to produce new data that shares characteristics with the data that is already in the dataset.

The main difference between VAEs and others variational inference based approaches is the assumptions made over the distributions that controls the model. In the VAE model, such distributions are assumed to be parametrizable by a set of parameters (e.g., Normal or Bernoulli distributions). These parameters can be found by neural networks via a proper training. The combination of the assumptions made for the distributions and the use of neural networks is what made the VAE model so efficient in searching the approximation of the true posterior $p(z|x)$. For the proposed VAE based dimensionality reduction approach, in the following sections we discuss the parametric form for the prior over the latent variables $p(z)$ the data conditioned over the latent variables $p(x|z)$ and the approximate posterior distribution $q(z|x)$ along with the form that the objective function of the VAE.

2.1 *Prior over the latent variables $p(z)$*

The definition of the latent variables of a certain model is usually a complex problem. The nature of those variables or the relationship between them are very difficult to express beforehand without deep knowledge about the situation that we want to model. VAEs takes an easy approach to this, assuming that the latent variables are distributed according the following distribution:

$$p(z) \sim N(z \mid 0, I) \tag{1}$$

From Equation 1, the prior distribution for $p(z)$ assumes no prior information about the relationship between the latent variables or the nature of them (other than they are Normally distributed). As it is mention in (Doersch, 2016), this can be done because any distribution of k dimensions can be originated by a vector of k variables Normally distributed if they are processed with a function that is complex enough. This function corresponds to the neural network of the VAE's decoder.

2.2 *The data conditioned on the latent variables $p(x|z)$*

This distribution represents the probability of x being generated by the latent variables contained in z. Note that the family of $p(x|z)$ should be defined based on the nature of the data. For example, for black and white spectrograms, or gray scale ones that are real valued but restricted to the interval (0,1), $p(x|z)$ is chosen to be a Bernoulli Distribution, in the form of:

$$p(x|z) \sim Be(x \mid f(z, \theta)) \tag{2}$$

The function $f(z|\theta)$ is a neural network that takes as inputs both the latent variables z and a

vector of parameters θ that represent the weights and biases of the NN, and then outputs a vector of Bernoulli parameters, $\vec{\rho}$, to define the distribution. Here we can see that it does not matter if z is sampled from a simple distribution such as a standard Normal, because if we let f to have an acceptable level of complexity (i.e., number of neurons and layers), then f will be able to first find the mapping between those Normally distributed samples for z to the true but unknown distribution that controls the latent variables, and then find the relationship between that distribution and the vector of Bernoulli parameters. The Bernoulli distribution is chosen to represent the output of the decoder for cases where binary data is being reconstructed by the VAE, as it is the case in this paper where we use images as inputs for the model.

The structure formed by f and $p(x|z)$ is named the decoder of the VAE, since its job is to output a reconstruction of the original input from the latent variables z.

2.3 *The approximate posterior q(z|x)*

In variational inference, the true posterior $p(z|x)$ is approximated by a distribution $q(z|x)$ that is searched in a family of parametric distributions of probabilities Q. In the VAE model, Q is assumed to be the family of multivariate isotropic Normal distribution. So, the following is true for $q(z|x)$:

$$q(z|x) \sim N(z|\vec{\mu}(x,w_1),\vec{\sigma}(x,w_2)I) \tag{3}$$

In Equation 3, the distribution $q(z|x)$ have both of its vectors of parameters, $\vec{\mu}$ and $\vec{\sigma}$ parametrized by neural networks that take as inputs the input data and vectors of weights and biases denoted as w_1 for the means vector and w_2 for the variances vector. The structure formed by the distribution $q(z|x)$ in conjunction with the neural networks for $\vec{\mu}$ and $\vec{\sigma}$ is called the encoder of the VAE since its task is to transform the input data to the latent representation z. The choice of a Normal distribution to represent the approximate posterior in the VAE model serves a double purpose. First, as we shall see in the next section, this choice allows a very quick and efficient training of the VAE. Second, without adding excessive complexity, the use of a Normal distribution adds flexibility to the model, since every latent variable will be controlled by its own mean and variance.

Since VAEs uses variational inference to find the approximate distribution $q(z|x)$, the idea is to train the neural networks that parametrize the vectors $\vec{\rho},\vec{\mu}$ and $\vec{\sigma}$ optimizing the VAE's objective function. But before, we need to be precise about the form that the VAE's objective function will take with the assumptions made for $p(z)$, $p(x|z)$ and $q(z|x)$.

2.4 *Objective function of the VAE model*

As stated before, the VAE model uses variational inference to solve the problem of finding a good approximate to the true posterior distribution $p(z|x)$. Variational inference proposes the following objective function to optimize and find $q(z|x)$, named the Evidence Lower Bound (ELBO) function:

$$ELBO(q(z|x)) = E_{q(z|x)-Q}[\log p(x|z)] - KL(q(z|x)\|p(z)) \tag{4}$$

It can be proved (Blei et al., 2017) that by maximizing Equation 4, variational inference is able to find the parameters that determines the best possible approximate distribution $q(z|x)$ within the chosen family of distributions Q.

In the ELBO function, there are two terms that need to be computed efficiently in order to find $q(z|x)$. First, as both distribution $q(z|x)$ and $p(z)$ belong to the Normal family, the Kullback-Leibler divergence that appears in Equation 4 between them will have closed form. The KL divergence between an isotropic Normal and a standardized normal, as stated in (Doersch, 2016), is shown in Equation 5:

$$KL(N(\vec{\mu},\vec{\sigma}I)\|N(0,I)) = 2(tr(\vec{\sigma}I) + \vec{\mu}^T\vec{\mu} - k - \log det(\vec{\sigma}I)) \tag{5}$$

where k is the dimensionality of the distribution, i.e., the dimensionality of the latent space of the VAE model.

The first term that appears in the ELBO function is the expectancy of the logarithm of $p(x|z)$. Computing that expectancy by sampling from z would be extremely inefficient because that would require the evaluation of f (which is a neural network that may be very complex) many times in order to obtain a good estimation of the expectancy. The neural networks within the VAE model are trained using stochastic gradient descent (SGD), so we could also use SGD in the evaluation of this expectation by sampling one time from z, compute $logp(x|z)$ from that value of z and then use that results as an estimation of the expectancy above mention. This process will be repeated until the solution converges to a good result for the posterior approximation.

2.5 *The variational auto-encoder model*

In Figure 1 a graphical representation of the VAE that comprises all the elements discussed before model is portrayed.

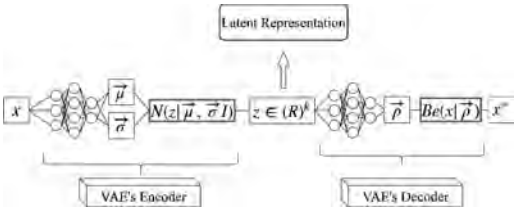


Figure 1. VAE model, with both neural networks, one for the encoder and the other for the decoder. Notice that it is the encoder that generates the latent representation z, and then the decoder, from that latent representation, reconstruct the input.

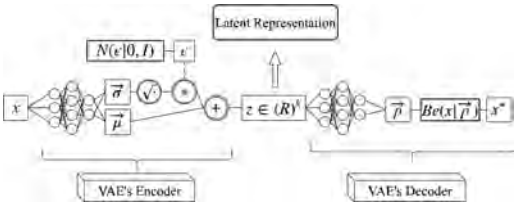


Figure 2. VAE model with the reparametrization trick applied.

Nevertheless, there is still one difficulty with the VAE model. It is known that neural networks use the back-propagation (BP) algorithm for performing a more efficient training. But BP does not work if there are stochastic units within the network, as it is the case when we sample from the encoder's distribution, since they have no gradient, therefore cannot propagate the error using the chain's rule. Kingma and Welling (Kingma & Welling, 2013) proposed a solution for this problem called "the reparametrization trick", in which all the stochasticity of the model is moved to an input, so the error can be propagated without problems. Figure 2 shows the VAE's model with the reparametrization trick applied.

3 PROPOSED VAE'S ARCHITECTURE FOR DIMENSIONALITY REDUCTION

The VAE model have as core components two neural networks, one for the encoder and other for the decoder. In what follows, the proposed architectures for such neural networks are described.

3.1 *Encoder's deep neural network architecture*

In Figure 3, a graphical representation of the encoder's neural network is portrayed.

The encoder's neural network is chosen to have three hidden layers with a number of units equal to 400, 200 and 100 for the first, second and third

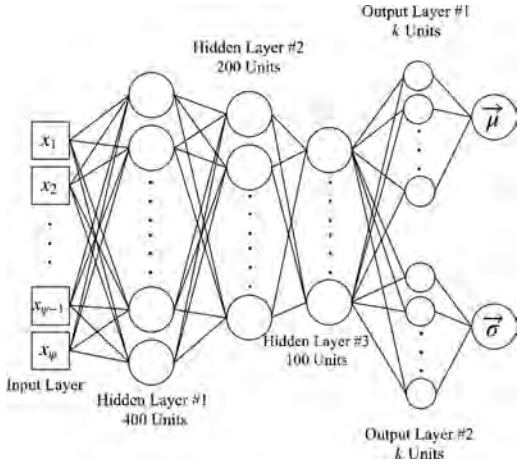


Figure 3. The encoder of the proposed VAE has three hidden layers, and two output layers, one for the means vector $\vec{\mu}$ and other for the variances vector $\vec{\sigma}$. The number of units in both output layers is k, the chosen dimension for the latent space of the VAE.

layer, respectively, and two different outputs layers, one for the vector of means, $\vec{\mu}$, and one vector of variances, $\vec{\sigma}$, to parametrize the approximate posterior distribution $q(z|x)$. In this case, both outputs layers have k units, where k corresponds to the target latent space dimension, i.e., the dimensionality of the data once the reduction is performed. This neural network uses the ReLU activation function in all the hidden layers.

3.2 *Decoder's deep neural networks architecture*

In Figure 4, the decoder of the proposed VAE model is portrayed.

As shown in Figure 4, the decoder of the proposed VAE has three hidden layers with 100, 200 and 400 units for the first, second and third hidden layers, respectively, and one output layers with ψ units, where ψ is the dimensionality of the data prior the reduction. Similarly to the encoder's case, this neural network has in all its hidden layers the ReLU activation function.

Both the VAE's encoder and decoder work together to first compress the data to its latent representation by computing the parameters that controls the distribution $q(z|x)$ to sample from it the latent variables and then decompressing the data from this latent representation to a reconstruction of its original form by first computing the vector of Bernoulli parameters $\vec{\rho}$ and then, depending on whether the data is binary or real valued restricted to the (0,1) interval, sample from the distribution or take $\vec{\rho}$ as the reconstruction itself.

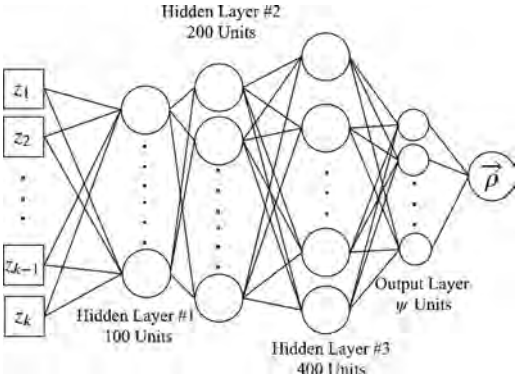


Figure 4. The decoder of the proposed VAE model has three hidden layers and one output layer to compute the vector of Bernoulli parameters $\vec{\rho}$. The number of units in the output layer is equal to ψ, the dimensionality of the input data.

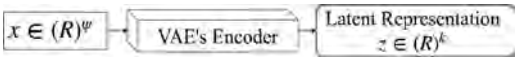


Figure 5. Dimensionality reduction method based on the Encoder of a VAE that has been trained successfully.

Once the VAE is trained, the neural network of the encoder can be used as a feature map to reduce the dimensionality of the input data to the lower dimensionality space by computing the means and variance vectors. Nevertheless, it is usually the case where the dimensionality reduction is preferable to be deterministic, i.e., it does not include any sampling in its process. This can be done if we note that the probability distribution of the encoder is a Normal distribution, so we could take the vector of means from the encoder's neural network as the latent variables directly in the form of $z = \vec{\mu}$. This provides a latent representation that does not vary each time a reduction over an specific data point is performed. A graphical representation of the dimensionality reduction process using the VAE's encoder can be seen in Figure 5.

4 EXAMPLE OF APPLICATION

4.1 *Dataset*

We use vibration data from the Society for Machinery Failure Prevention Technology (MFPT) open dataset (Bechhoefer, 2016) that was obtained during the operation of a NICE ball-bearing element. The dataset contains three main classes, each corresponding to one health condition: baseline (where no failure is present), outer race fault and inner race fault. The original signals for each class were first divided into portions of length $L = 1024$ points each. Besides, we use 50% of overlap between adjacent chunks. Then, by means of the Short Time Fourier Transform (STFT), a spectrogram of each section of data was computed, and then via a bilinear interpolation (Raveendran & Thomas, 2014), it was transformed into an image of 96 by 96 pixels. We chose to use spectrograms because they can portray both time and frequency information of the signal, instead of just time information as it would have been the case if we fed the original signal directly into the VAE model. Table 1 shows the total number of images per class in the dataset.

But, in order for the VAE model to work properly, as we can recall from Figure 3, the VAE's encoder has to receive as input a vector of CH131_181-E021.eps components, not a two dimensional array as it would be if we fed the spectrogram images directly to the model. Thus, we need to first reshape the images by stacking horizontally their rows to form vectors, as it is shown in Figure 6.

4.2 *Validation experiments*

We are interested in two situations. First, we want to know if the VAE's latent representation is stable across dimensions of the latent space. This is of important because the decision of the dimensionality k is not clear beforehand, and it usually relies on the person in charge of the analysis, so models that are less sensitive to variations of this parameter present an advantage over those who needs a fine tuning to be effective. For this purpose, we perform reductions using the VAE approach to a series of different values of k and then analyze if variations in the dimension of the latent space pro-

Table 1. Classes and number of samples per class for the MFPT dataset.

ID	Fault's location	# of samples
IR	Inner Race Ring	1981
OR	Outer Race Ring	5404
Baseline	No Failure	3423

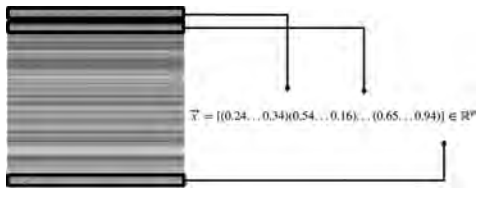


Figure 6. Diagram showing the process of reshaping the spectrogram images as vectors.

Table 2. Values for k and epsilon.

k	ε
2, 4, 8, 16, 64, 128	1%, 100%

duce variations in the accuracy of classification. The values of k tested can be seen in Table 2.

The second situation of interest is when the amount of labeled data available is low. This is usually the case in industrial applications, because it is usually cheap to acquire data from installed sensors, but the process of labeling such data for training machine learning models requires highly qualified personnel and often further analysis on the machine's degradation process. For this, we will test two extreme situations: where the size of the training dataset training for the fault diagnosis model is either the full training dataset or only 1% of it. This is represented by the parameter ε, the percentage of the training dataset available. This emulates situations where we might have a very reduced amount of labeled data versus when the availability of labeled data is not a concerning issue and from there, evaluate if the VAE based reduction generates a representation of the data that makes the differences between the different health states more explicit, so that the NN based classifier can still learn even if less data is available. Recall that as the VAE is an unsupervised model, it does not require labels for its training, so even when working with ε = 1%, the VAE can use the whole training dataset for its unsupervised training.

For the classification tasks, we use a neural network based classifier with a single hidden layer of 100 units and ReLU activation function. The regularization is performed via dropout with probability equals to 50%. For the output layer, softmax is used as the activation function. The model is trained using the cross entropy cost function, a learning rate of 10^{-4} and a maximum of 15000 epochs. The VAE's weights and biases are initialized with the Xavier initialization (Glorot & Bengio, n.d.) and with an array of zeros, respectively. For the NN based classifier, the weights and biases were randomly initialized with a Normal distribution of zero mean and unity variance.

The results shown in the next section were obtained following the steps below:

I. Choose a value for ε and k from Table 3. Then, divide the whole dataset into a training and testing set in the proportion 3:1.
II. Train the VAE in an unsupervised way using the full training dataset. This training is performed using a maximum of 500 epochs and learning rate equal to 10^{-4}.
III. Reduce the dimensionality of the training and testing datasets with the trained VAE's encoder as discussed in Section 3.
IV. If ε = 1%, extract that proportion from the training dataset, and train the NN based classifier. If ε = 100% use the whole dataset to perform the training.
V. Use the full testing dataset to evaluate the accuracy of the classification obtained.

We also compare the results obtained with the VAE model against results obtained when the spectrograms are fed directly to the NN-based classifier. To perform this type of experiment, the same procedure as before applies except for steps II and III.

The hardware and software used in the example of application are: Nvidia GPU Titan X, an Intel Processor i7 7700k, 32Gb of RAM DDR4 and Tensorflow 1.2.0 combined with CUDNN 8.0.

4.3 *Results and discussion*

In Figure 7 and Figure 8, results regarding the accuracy obtained in the classification of health states when using the VAE approach to reduce the dataset to different values of k are compared against the case where no reduction is performed to the dataset.

Table 3 shows the best accuracy results obtained with the VAE approach and the model that does not perform a reduction in dimensionality, for either ε = 1% or ε = 100%

As we can see from Figure 7 and Figure 8, fault diagnosis accuracies obtained with the VAE model are almost always better than the ones obtained when no reduction is performed. The only excep-

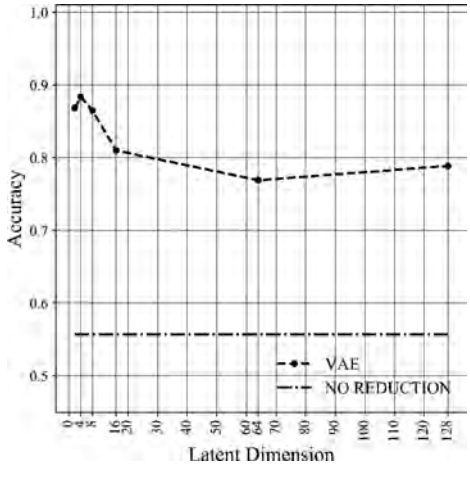


Figure 7. Accuracy versus latent space dimension for the case where only 1% of the training dataset is used to train the NN classifier.

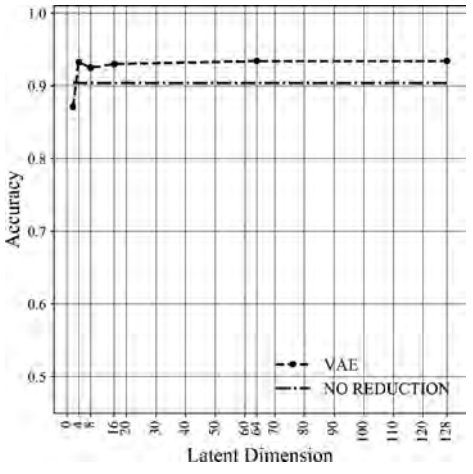


Figure 8. Accuracy versus latent space dimension for the case where the full training dataset is used to train the NN classifier.

Table 3. Best accuracy results for both models tested. The VAE model achieves its better results for values of k equal to 4 and 64 for ε = 1% or ε = 100% respectively.

VAE	No reduction
$\varepsilon = 100\%$	
93.43%	90.38%
$\varepsilon = 1\%$	
88.39%	55.76%

tion is when the full training dataset is used and the VAE model reduces the dimensionality to $k = 2$. This can be explained from the fact that the spectrogram images have an original dimensionality of $\psi = 9216$ data points (96 by 96 pixels), and to reduce from that dimension to $k = 2$ will induce some loss of information. Nevertheless, from Figure 8 and for every other choice of k (including $k = 4$) the VAE model surpasses in terms of fault diagnosis accuracy the model that do not perform reduction.

Also observe that the drop in accuracy caused by the reduction in the size of the training dataset is much less notorious for the VAE model. As we can see from Table 3, the drop in accuracy for the VAE model is approximately 5% for reducing the training dataset a 99% of its original size, but for the model that do not perform a reduction in dimensionality, the drop is about 45%. That clearly indicates the robustness of the VAE's latent space in making the different health states present in the dataset more differentiable for the NN based health state classifier.

In terms of the results obtained with the VAE by varying the latent space's dimensionality, the size of the training dataset does affect the stability of the model. For ε = 100%, except for $k = 2$, the model is very stable with differences in accuracies of maximum 2%. Instead, when ε = 1%, the model needs to be tuned prior its application in order to find the dimension that delivers the best performing latent space, since the differences in accuracies reach 12%.

5 CONCLUDING REMARKS

We have introduced a new method for performing dimensionality reduction in a dataset using deep Variational Auto-Encoders. The VAE model maps the data to a latent representation that can be chosen to have a lower dimensionality than the original data, thus producing a dimensionality reduction.

Results show that when using the STFT and spectrogram to preprocessed ball-bearings operational data provided by vibration sensors, the VAEs approach produces better accuracies in the classification tasks performed than the model that do not perform a reduction in dimensionality, especially when the available labeled data to train the fault classifier is limited.

REFERENCES

Bechhoefer, E. (2016). A quick introduction to bearing envelope analysis. Retrieved from http://www.mfpt.org/faultdata/MFPT Bearing Envelope Analysis.pdf.

Blei, D.M., Kucukelbir, A., & McAuliffe, J.D. (2017). Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773.

Blum, A.L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*. https://doi.org/10.1016/S0004-3702(97)00063-5.

Cireşan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column Deep Neural Networks for Image Classification. *Technical Report*.

Doersch, C. (2016). Tutorial on Variational Autoencoders.

Hasani, R.M., Wang, G., & Grosu, R. (2017). An Automated Auto-encoder Correlation-based Health-Monitoring and Prognostic Method for Machine Bearings.

Hinton, G.E., & Salakhutdinov, R.R. (2006). Reducing the dimensionality of data with neural networks (supporting online material). In *Reducing the dimensionality of data with neural networks*. https://doi.org/10.1126/science.1127647.

Hsu, W.-N., Zhang, Y., & Glass, J. (2017). Unsupervised Domain Adaptation for Robust Speech Recognition via Variational Autoencoder-Based Data Augmentation, (1). Retrieved from http://arxiv.org/abs/1707.06265.

Kingma, D.P., & Welling, M. (2013). Auto-Encoding Variational Bayes. https://doi.org/10.1051/0004-6361/201527329.

Liu, H., Li, L., & Ma, J. (2016). Rolling Bearing Fault Diagnosis Based on STFT-Deep Learning and Sound Signals. *Shock and Vibration*. https://doi.org/10.1155/2016/6127479.

Pan, Z., Rust, A., Networks, H.B.-N., IJCNN, 2000., & 2000, undefined. (n.d.). Image redundancy reduction for neural network classification using discrete cosine transforms. *Ieeexplore.ieee.org*. Retrieved from http://ieeexplore.ieee.org/abstract/document/861296/.

Raveendran, H., & Thomas, D. (2014). Image Fusion Using LEP Filtering and Bilinear Interpolation. https://doi.org/10.14445/22315381/IJETT-V12P282.

Rezende, D.J., Mohamed, S., & Wierstra, D. (2014). Stochastic Backpropagation and Approximate Inference in Deep Generative Models. https://doi.org/10.1051/0004-6361/201527329.

Salimans, T., Kingma, D., & Welling, M. (2015). Markov Chain Monte Carlo and Variational Inference: Bridging the Gap. In *Proceedings of the 32nd International Conference on Machine Learning*.

Shuang, L., Automation, L.M.-M. and, ICMA, 2007., & 2007, undefined. (n.d.). Bearing fault diagnosis based on PCA and SVM. *Ieeexplore.ieee.org*. Retrieved from http://ieeexplore.ieee.org/abstract/document/4304127/.

Zhou, F., Gao, Y., & Wen, C. (n.d.). A Novel Multimode Fault Classification Method Based on Deep Learning. https://doi.org/10.1155/2017/3583610.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Unsupervised deep generative adversarial based methodology for automatic fault detection

D.B. Verstraete & M. Modarres
University of Maryland, MD, USA

E. López Droguett
University of Chile, Santiago, Chile
University of Maryland, MD, USA

A.N. Ferrada & V. Meruane
University of Chile, Santiago, Chile

ABSTRACT: System health management is of upmost importance with today's sensor integrated systems where a constant stream of data is available to feed information about a system's health. Traditional methods to assess this health focus on supervised learning of these fault classes. This requires labeling sometimes millions of points of data and is often laborious to complete. Additionally, once the data is labeled, handcrafted feature extraction and selection methods are used to identify which are indicators of the fault signals. This process requires expert knowledge to complete. An unsupervised generative adversarial network based methodology is proposed to address this problem. The proposed methodology comprises of a deep convolutional Generative Adversarial Network (GAN) for automatic high-level feature learning as an input to clustering algorithms to predict a system's faulty and baseline states. This methodology was applied to a public data set of rolling element vibration data from a rotary equipment test rig. Wavelet transform representations of the raw vibration signal were used as an input to the deep unsupervised generative adversarial network based methodology for fault classification. The results show that the proposed methodology is robust enough to predict the presence of faults without any prior knowledge of their signals.

1 INTRODUCTION

Much of fault diagnostics involves the use of labeled data. This is challenging for new assets outfitted with sensor suites capable of generating massive amounts of data. Without knowledge of faults or their corresponding signals, engineers may not be able to diagnose faults effectively. Traditional methods include feature extraction and selection methods which attempt to use a specific feature of the signal to diagnose the faults. This method requires knowledge of which features are relevant for the task. Moreover, if an engineer has some knowledge of the fault, that knowledge could be biased or incomplete. Unsupervised fault diagnostics attempts to fill in that knowledge.

Deep learning algorithms can perform automatic feature learning to better understand the underlying data features that are most relevant. This automatic feature learning attempts to fill in the gaps of knowledge of relevant features to the fault signals. There are challenges with this automatic feature extraction and selection.

Unsupervised learning has been attempted for fault diagnostics previously. Indeed, Langone (2017) took pre-stressed concrete bridge natural frequency data and proposed an unsupervised adaptive kernel spectral clustering for damage events. Wang (2016) proposed unsupervised feature extraction via continuous Sparse Auto-Encoders (SAE). Once the SAEs extracted the features supervised learning was used on transformer faults. Lei (2016) proposed unsupervised sparse filtering feature learning. Faults were then diagnosed with supervised softmax regression. Jiang (2016) proposed unsupervised feature learning with SAEs for chemical sensor data. These features were fed into supervised softmax regression to diagnose faults. Sun (2016) took induction motor fault data and proposed the use of SAEs for unsupervised feature extraction. These features were again followed by supervised learning for classification by neural networks (NN). Of these approaches, only Langone et al. could be considered truly unsupervised. The rest are restricted to unsupervised feature learning followed by supervised fault diagnostics. Moreover,

apart from the use of SAEs, none of these methods would be considered deep.

In this paper, we propose a GANs based methodology application to unsupervised fault diagnostics on scalogram image representations. To validate the proposed methodology, the public Machinery Failure Prevention Technology (MFPT) Society bearing data set (Bechhoefer 2016) is used. The remainder of this paper is structured as follows. Section 2 gives an overview of GANs and the methodology. Section 3 presents results of the GANs based methodology applied to the MFPT data set. Section 4 provides conclusions.

2 GENERATIVE ADVERSARIAL NETWORKS

Generative adversarial networks (GANs) have at their core a minimax game which seeks to pit a forger, the generator network, against a detective, the discriminator network. The generator seeks to create fake data, or scalograms in this paper, to trick the discriminator who must discriminate between the real data and the fake data as shown in Figure 1. Back propagation is performed on the weights and biases and the process is repeated. The benefit to this training is while the generator seeks to develop an underlying distribution of the real data, the discriminator is feeding information back to the generator, not on the real data, on the weights and biases of the learned features. This helps to prevent overfitting of the data.

Within this minimax game, the objective function's goal is to maximize the value, V, to the point where the discriminator and generator no longer find it necessary to make changes to their weights and biases. While this is the goal of GAN training, there is functionally no mechanism with the training to control it. Therefore, there can be issues with convergence. More formally in Eq. 1 from Goodfellow (2014):

$$\min_G \max_D V(G,D) = \mathbb{E}_{x\sim P_{data}(x)}\left[\log(D(x))\right] + \mathbb{E}_{z\sim P_{noise}(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right]. \quad (1)$$

where, $P_{data(x)}$ is the data distribution, $P_{noise(x)}$ is the noise distribution, $D(x)$ is the Discriminator objective function, and $G(z)$ is the generator objective function.

The GANs based methodology used in this paper can be found in Figure 2. The methodology starts with developing a scalogram image representation of the raw data, and then proceeds to training of the deep convolutional generative adversarial network (DCGAN). Once the DCGAN training is completed and visual inspection of the generator output images is done, concatenation of the last activation layer of the discriminator is completed. Once the activations are concatenated, kmeans++ is used for clustering on the first two principal components. Visual inspection of the generator output is still needed within GANs training and is a crucial step within the training.

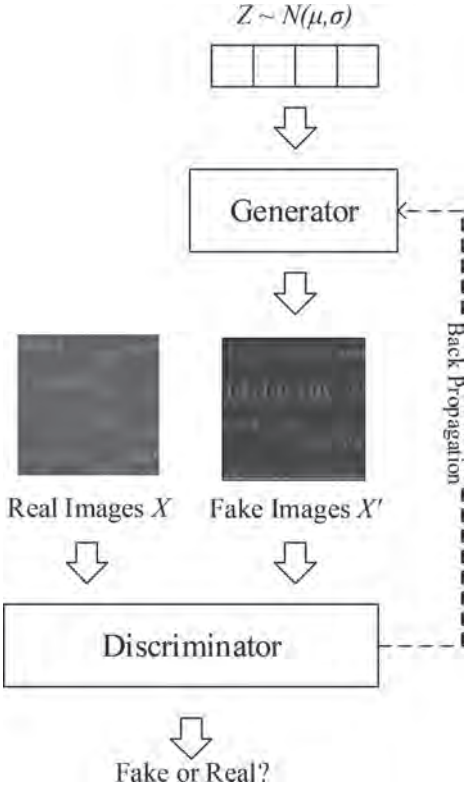


Figure 1. GAN training.

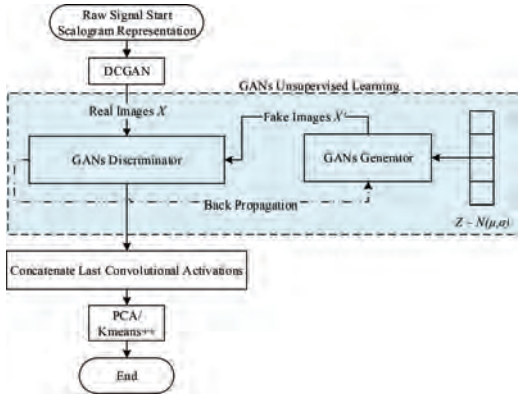


Figure 2. Proposed unsupervised GAN methodology.

There are two goals for the output of this methodology: 1) Separation of the baseline healthy data with the fault data, and 2) Separation of the individual faults. When a new sensor system comes online, the engineer needs to know when the system drifts from healthy signals to a signal with which to decide when to conduct planned maintenance. Once the engineer has familiarity with the system and signals can be identified as individual faults on the inner or outer raceway, then better predictions and a fully supervised methodology can be used (Verstraete 2017).

The GAN architecture used in this paper incorporates the guidelines proposed in Radford (2016); however, adjustments to that paper's architecture were made for handling the MFPT data set. Radford et al provides the following five GANs architecture guidelines: 1) generator and discriminator network pooling layer replacement with strided convolutions, 2) Batch normalization (BN) is required for both the discriminator and generator networks, 3) Fully connected hidden layers should be removed for deep architectures, 4) Rectified Linear Unit (ReLU) activation use in all layers of the generator except the output should use Tanh, and 5) Leaky ReLU activation use on all layers for the discriminator. DCGANs are used in this paper as a baseline to implement GANs. The combination of these five guidelines composes what is defined as deep convolutional generative adversarial networks (DCGANs).

2.1 *Strided convolutions*

The relationship between a convolutional operation's input shape, i_j, and the operation's output shape, o_j, of a convolutional layer along axis j are related to three factors: 1) kernel size (k_j), 2) stride (s_j), and 3) padding (p_j). Convolutional strides are generally set to $s_j = 1$ for most operations; however, for GANs strided convolutions of $s_j > 1$ are used in place of pooling layers. This is applied for the discriminator to learn its own downsampling, and for the generator to learn its own upsampling.

2.2 *Batch normalization*

Batch normalization (BN) is an important addition to the architecture between each convolutional layer. (Ioffe 2015) As the data moves through the convolutional layers the weight and bias values are adjusted. This has the potential to lead to the data increasing or decreasing to unrealistic values. Batch normalization prevents this from becoming an issue with the training by normalizing the data to a mean of zero and a variance of one for each data batch. Setting values of x over a mini-batch: $\beta = \{x_{1 \ldots m}\}$ to output the learned parameters γ and β, $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$. The mini-batch mean is $\mu_\beta \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i$, the mini-batch variance is $\sigma_\beta^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}\left(x_i - \mu_\beta\right)^2$, they are then normalized with $\hat{x}_i \leftarrow \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}}$, and scaled and shifted with, $y_i \leftarrow \gamma\hat{x}_i + \beta \equiv BN_{\gamma,\beta}(x_i)$.

2.3 *Activation layers*

The following activation functions are used throughout the architecture. For the generator, two activations functions are used: 1) Rectified Linear Unit (ReLU), $f(x) = \begin{cases} 0 \text{ for } x < 0 \\ x \text{ for } x \geq 0 \end{cases}$ and 2) Hyperbolic tangent (tanh), $\mathrm{f}(x) = \frac{2}{1+e^{-2x}} - 1$). Within the generator network ReLU is used between every layer except tanh activation is used after the last layer. For the discriminator, Leaky ReLU is used on every layer: Leaky ReLU $f(x) = \begin{cases} 0.01 \text{ for } x < 0 \\ x \text{ for } x \geq 0 \end{cases}$.

Leaky ReLU differs from ReLU in values less than 0.

2.4 *Neural network architectures*

The neural network architectures used in the proposed methodology incorporate the guidelines as proposed by Radford et al. Two networks were developed to account for the data set presented in Section 3. The generator network, as shown in Figure 3, takes the vector of noise and through deconvolution, BN, and activation functions creates an image. In this case the output of a 96×96 image of a scalogram of a signal. To do this, a 100×1 vector is projected and reshaped to deconvolve into a $6 \times 6 \times 512$ feature space. This space is then deconvolved to a $12 \times 12 \times 256$, then $24 \times 24 \times 128$, $48 \times 48 \times 64$, and finally a $96 \times 96 \times 3$ image.

The discriminator network, as shown in Figure 4, then takes that generated image and judges whether the image is real or fake. It does this by taking the real images and automatically learning the feature subspace. For the 96×96 images this results in a network of convolutional layers

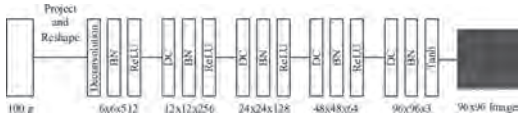


Figure 3. Generator network.

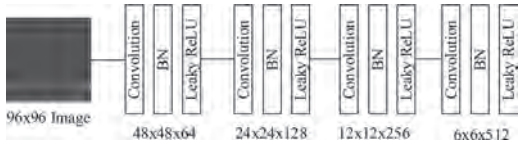


Figure 4. Discriminator network.

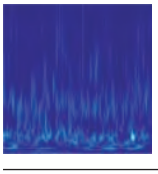
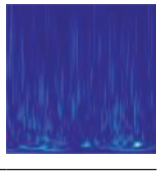
consisting of a $48 \times 48 \times 64$ layer, $24 \times 24 \times 128$, $12 \times 12 \times 256$, and $6 \times 6 \times 512$ layers. The output of this last activation holds a lot of information about the feature space and is useful for unsupervised fault diagnostics.

3 PROPOSED METHODOLOGY APPLICATION

The MFPT data set is a good test of any algorithm as the outer race fault and baseline conditions are difficult to separate. NICE bearings were used within an experimental test rig. Accelerometer data was gathered on three conditions. First, at a sampling rate of 97,656 Hz, a baseline condition at 270 lbs of load was captured. Second, a total of ten faults on the outer-raceway were gathered. At the same sampling rate and loading condition as the baseline, three outer race faults were tested, and the remaining seven outer race faults had the following load cases: 25, 50, 100, 150, 200, 250 and 300 lbs. These seven load cases had a sampling rate of 48,828 Hz. Third, with a sampling rate again of 48,848 Hz, seven inner race faults at a loading of 0, 50, 100, 150, 200, 250 and 300 lbs were gathered. From these raw signals, scalogram image representations were created with the following three classes as shown in Table 1: normal baseline (N), inner race fault (IR), and outer race fault (OR). In total 10,808 scalogram images were generated with 3,423 baseline, 1,981 inner race, and 5,404 outer race images. The training set used was fifty percent. Bilinear interpolation (Raveendran 2014) aided in reducing the original images to down to a trainable size for the GAN architecture.

The first step once the GANs training is completed is visual inspection of the generator image outputs. These can be seen in Figure 5. The different fault conditions can be identified within the images. This step is a key indicator for identification of mode col-lapse, vanishing gradients, non-convergence, or checkerboarding artifacts. With this completed, the last activation layer of the discriminator network is concatenated and clustering can be done.

Table 1. 96×96 pixel MFPT scalogram images.

Baseline	Inner race	Outer race
		

The first step once the GANs training is completed is visual inspection of the generator image outputs. These can be seen in Figure 5. The different fault conditions can be identified within the images. This step is a key indicator for identification of mode collapse, vanishing gradients, non-convergence, or checkerboarding artifacts. With this completed, the last activation layer of the discriminator network is concatenated and clustering can be done.

Kmeans++ is used for clustering within the paper to demonstrate how robust the GANs training can be towards a simple clustering algorithm. Figure 6 shows the resultant clustering predictions of the first two principal components of the last activation layer of the discriminator and colored by the predicted labels. There is overlap in the outer and inner race predictions, but the GANs training plus kmeans++ does an excellent job separating the baseline signals from the fault conditions.

Figure 7 shows the first two principal components of the last activation layer color coded by the real labels. It appears the GAN training with kmeans++ had the most difficulty with separating the fault conditions. A clustering algorithm more capable of handling the non-convex nature of the outer race fault could potentially increase the evaluation metrics but is beyond the scope of this paper.

Since the labels to the data are known, evaluation metrics like purity (Manning 2008), normalized mutual information (NMI) (Kuncheva 2004), and adjusted RAND index (ARI) (Hubert 1985) can be used to validate the architecture. Table 2 has the overview of these metrics for this methodology.

Overall these number could be improved; however, the first goal of this methodology is to

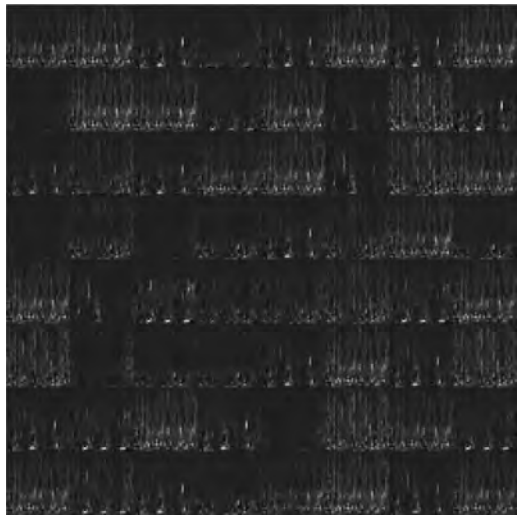


Figure 5. Output images of DCGAN generator training.

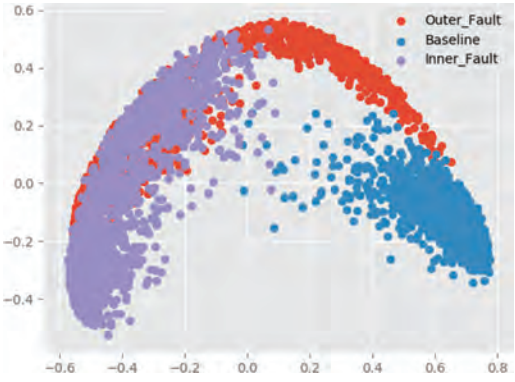


Figure 6. DCGAN PCA Kmeans ++ predicted.

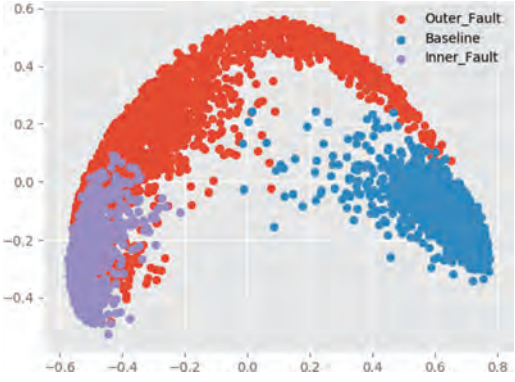


Figure 7. DCGAN PCA Kmeans ++ real.

Table 2. MFPT 96 × 96 generator output, DCGAN, Kmeans++.

ARI	Purity	NMI
0.50	0.79	0.62

separate the baseline healthy system state with that of the faults. This methodology proves it can handle that. More work can be done to improve these numbers and provide better information to the engineer regarding which individual fault case the signal is presenting itself as.

4 CONCLUSIONS

Generative adversarial networks and deep learning as a field stand to unlock numerous potential applications within the field of engineering research. This application is the first of its kind and shows great promise.

The proposed architecture demonstrates its abilities with automatic feature learning to a level with which a simple clustering algorithm can separate the healthy baseline signals with the fault data. An engineer can easily make an engineering decision on maintenance without the need for any knowledge of the individual signals.

The practical application of this paper has far reaching possibilities into many engineering fields and is not limited to rolling element bearings. Aerospace, automotive, oil & gas, and many other industries can utilize this unsupervised methodology.

ACKNOWLEDGMENTS

The authors acknowledge the partial financial support of the Chilean National Fund for Scientific and Technological Development (Fondecyt) under Grant No. 1160494.

REFERENCES

Bechhoefer, Eric. “A Quick Introduction to Bearing Envelope Analysis.” (2016).

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems 27*, 2672–2680.

Hubert, Lawrence, and Phipps Arabie. “Comparing partitions.” *Journal of classification* 2.1 (1985): 193–218.

Ioffe, Sergey, and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift.” *International Conference on Machine Learning*. 2015.

Jiang, Peng, et al. “Fault diagnosis based on chemical sensor data with an active deep neural network.” *Sensors* 16.10 (2016): 1695.

Kuncheva, Ludmila I. *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2004.

L. Wang, X. Zhao, J. Pei, and G. Tang, “Transformer fault diagnosis using continuous sparse autoencoder,” SpringerPlus, vol. 5, no. 1, p. 1, 2016.

Langone, R., Reynders, E., Mehrkanoon, S., & Suykens, J.A.K. (2016). Automated structural health monitoring based on adaptive kernel spectral clustering, *90*(June), 1–21. https://doi.org/10.1016/j.ymssp.2016.12.002.

Lei, Y., Jia, F., Lin, J., Xing, S., & Ding, S.X. (2016). An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data. *IEEE Transactions on Industrial Electronics*, *63*(5), 3137–3147. https://doi.org/10.1109/TIE.2016.2519325.

MFPT Data Set, http://www.mfpt.org/FaultData/FaultData.htm.

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv*, 1–15.

Raveendran, H., Deepa Thomas. "Image Fusion Using LEP Filtering and Bilinear Interpolation", International Journal of Engineering Trends and Technology (IJETT), V12(9),427–431 June 2014. ISSN:2231–5381.

Sun, Wenjun, et al. "A sparse auto-encoder-based deep neural network approach for induction motor faults classification." *Measurement* 89 (2016): 171–178.

Verstraete, D., Ferrada, A., Droguett, E.L., Meruane, V., & Modarres, M. (2017). Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings. *Shock and Vibration*, *2017*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Computer vision for structural damage quantification: A novel residual deep learning based approach

N. Astorga
Department of Mechanical Engineering, University of Chile, Santiago, Chile

E. López Droguett
Department of Mechanical Engineering, University of Chile, Santiago, Chile
Center for Risk and Reliability, University of Maryland, College Park, MD, USA

V. Meruane
Department of Mechanical Engineering, University of Chile, Santiago, Chile

ABSTRACT: Recent advances of deep neural networks have revolutionized the techniques of machine learning for practical applications involving computer vision tasks. The flexibility of these models has allowed difficult tasks such as image segmentation to be tackled by this type of algorithms with much improved results. In this work, we propose and explore the capabilities of a novel deep residual neural networks with atrous convolutions for pixel to pixel classification tasks to achieve localization and quantification of structural damage in noisy image datasets. The proposed model is applied to a dataset of images synthesized to resemble debonding damage in honeycomb structures.

1 INTRODUCTION

Critical infrastructures such as bridges or complex systems in the mining industry might present damage that severely reduces the reliability. Inspections are important to warrant that a dangerous and costly failure does not take place. In this context, Structural Health Monitoring is critical to ensuring cost-effective and safe operational efficiency.

Inspections of structures usually includes some form of visual tasks that heavily rely on individuals with domain knowledge. In scheduling these inspections, a preventative maintenance program must balance a system's safety and the cost of an expert analyst. Moreover, there are situations where these tasks are carried out in perilous of difficult circumstances.

To tackle these limitations, automated visual inspection systems might help to reduce costs, increase accessibility, and improve safety. For instance, unmanned aerial vehicles (UAVs) can autonomously produce videos and photographs of damaged areas. This can be further improved by methods based on computer vision to identify damage in real time.

For certain structures and equipment, the identification of the type or location of damage is not enough. It is also important to quantify the size of the damage as it can be can be used for the tracking of damage growth over time and therefore assess possible effects and their severity they may have on the structure or equipment health state. However, damage quantification is not widely used by current forecast algorithms (Douka et al, 2003) (Ohno et al, 2010) because it is a complex and complicated process as well as difficult to automate.

The existing methods for damage quantification are based on processing of images (Zhou et al, 2015) that in general require different parameters for each image and manually extracted features, a procedure that is labor intensive and costly as well as usually not adequate in real time damage quantification.

Thus, we propose a deep learning based model for damage quantification based on deep residual neural networks with atrous convolution layers. Applied in the context of semantic segmentation (L.-C. Chen, et al 2016) (J. Long, et al 2015), these convolution operations use contextual information to perform pixel-by-pixel image classification by identifying which pixels correspond to damaged areas in a structure or equipment. The proposed model is applied to a dataset of images synthesized to resemble debonding damage in honeycomb structures.

The remaining of the paper is organized as follows. Section 2 introduces the building blocks of residual networks and presents the proposed

model. Section 3 discusses the example of application and results based on the proposed model and compares it to fully connected neural network and k-means for image segmentation. Section 4 presents some concluding remarks.

2 PROPOSED DEEP RESIDUAL NEURAL NETWORK MODEL WITH ATROUS CONVOLUTIONS

In this section, we discuss the proposed model based on deep residual neural network with atrous convolutions and operating on images for damage quantification. We first introduce the several building blocks of the proposed model and then the proposed architecture is presented.

2.1 *Deep learning*

Deep learning is a set of techniques based on neural networks that in recent times have shown better results than traditional techniques. The great success of this framework is due to the increase of the computational power added to the increasing availability of larger datasets, thus being able to train more complex and robust models.

These techniques are based on the hierarchy of the layers of a neural network, which can vary from a few layers to hundreds (He et al, 2016). The success of deep learning lies in how information is distributed throughout the network, extracting the most abstract characteristics of the data in the first layers and locating the most specific characteristics in the latter layers.

Deep learning has established a change in how to solve machine learning problems by generating a variety of architectures that are not limited to supervised tasks. The applications range from unsupervised models such as auto-encoders (Vincent et al., 2010) to autoregressive models as recurrent networks (Gregor et al., 2014). In the case addressed in this paper, a supervised task is analyzed that encompasses a deep neural network architecture with multiple outputs, each corresponding to a pixel.

2.2 *Segmentation*

Semantic segmentation consists in the pixel to pixel classification of an image, allowing to predict whether a pixel is considered a part of the damage. This type of algorithm takes into consideration the pixel context in the image, establishing whether it corresponds to an isolated damage without importance to the problem at hand or belonging to a damage type of interest. This type of segmentation is not limited to binary classification, as they can also be extended to establish if the damage is considered of a certain type, that is, how it was generated since different damages result from different failure modes.

By obtaining the pixel to pixel prediction of the image it is possible to obtain the size of the damage (e.g., crack), essential information to diagnose the damage and predict the Remaining Useful Life (RUL) of a structure or machine.

One of the disadvantages of this type of algorithm lies in the construction of the dataset required for its training. In fact, since the prediction of the model is pixel by pixel, the classification task for the training images must also be pixel to pixel. An example of the data that should be available is shown in Figure 1 where the left image shows the training image while the right image is the corresponding ground truth. If the training set is properly constructed, a procedure can be automated that delivers a damage prediction that is flexible for a given application, which is determined by the ground truth.

2.3 *Convolution neural networks*

Convolutional neuronal networks are a type of neural networks that take advantage of the translational invariance that images have. In summary, the parameters that these networks learn are filters W that are convolved with a subarea of the features maps. As the convolution is done in only one sector of the feature maps, the number of parameters learned is smaller, and therefore the computational cost is reduced and more relevant characteristics are learned for data that fulfill the assumption of translational invariance. Mathematically, a convolutional layer can be written as shown in equation (1).

$$x_{i,j}^{l} = \sum_{m}^{f1} \sum_{n}^{f2} \sum_{c}^{C} w_{m,n,c}^{l} s_{i+m,\, j+n,c}^{l-1} + b^{l} \tag{1}$$

in which 1 is the number of the layer. The input of each layer is x, of dimensions $H \times W \times C$. These variables are the height, width and quantity of features maps, respectively, with x^0 the training image that is input to the network. The weights of the network are w. The variable s are the feature maps after the activation function; m and n iterate over the width

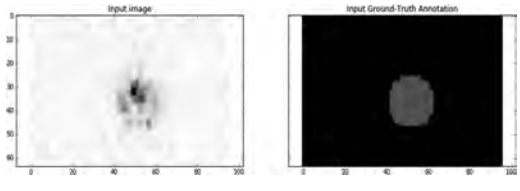


Figure 1. Left, Image of data training. Right, ground truth.

and height of the filter, f1 and f2. The variable c iterates over the number of channels C, i and j iterate over the width and height of the feature map, H y W. In addition, the convolution in equation (1) is carried out a number of times equals to the number of feature maps desired for the layer 1.

2.4 *Atrous convolutions*

In the context of semantic segmentation, it is relevant to establish the information of the context in which a pixel is found. To do this, previous algorithms based on deep networks (Krizhevsky et al, 2012) used a series of convolutions followed by pooling layers and with a subsequent layer known as deconvolution. This increases the size of the smaller feature map to establish a cost to what the pixel to pixel prediction should be. This type of model has a disadvantage: the resolution is lost with increasingly deeper network architectures, so it is not possible to use this type of transformation when the number of layers is large, as is the case of residual networks used in the proposed model. Atrous convolutions (Holschneider et al, 1989) try to solve this problem by creating filters that have zeros in between. For a traditional filter, this means increasing its size, placing zeros in between and making a standard convolution. Mathematically this shown in equation (2) for the case of one dimension (1D). If the rate r is equal to 1, we have the usual convolution; x is the input and K is the filter length w[k]. Figure 2 shows the variation of the filters used in the atrous convolution for different rates, with the green color representing 0 values in the filter. As these are parameters are not learned, performing atrous convolution does not increase the number of parameters. The blue color represents values of the filters that are learned by backpropagation.

$$y[i] = \sum_{i}^{K} x[i + r * k]\, w[k] \tag{2}$$

2.5 *Atrous pyramid pooling*

In the proposed architecture discussed in Section 2.7, the concept of pyramid pooling is used in which atrous convolution is performed in parallel using different resolutions and, therefore, capturing contextual information at different levels.

2.6 *Residual neural network*

In general, deep neural networks before 2015 did not exceed a number of layers greater than 20 as trying to increase this number resulted in performance deterioration. This is known as vanishing gradient, which consists in that the earliest layers of the neural network receive updates that are too small and, therefore, a null learning. This is due to the rule of the chain and how backpropagation is made.

Recently this difficulty was surpassed by a type of neural networks called residual networks (He et al, 2016) These networks allow the use of architectures with an even larger size reaching block numbers of up to 151, in which each block has several layers, thus overcoming the usual convolutional networks. For example, in ILSVRC & COCO 2015 Competition (ImageNet, 2015), residual networks obtained the first place in all categories (He et al, 2016). Some known residual networks are ResNet-50, ResNet-101 and ResNet-151 (He et al, 2016), where the name number represents the number of residual blocks it has.

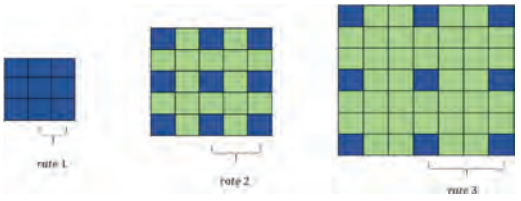


Figure 2. Atrous convolutions for different rates.

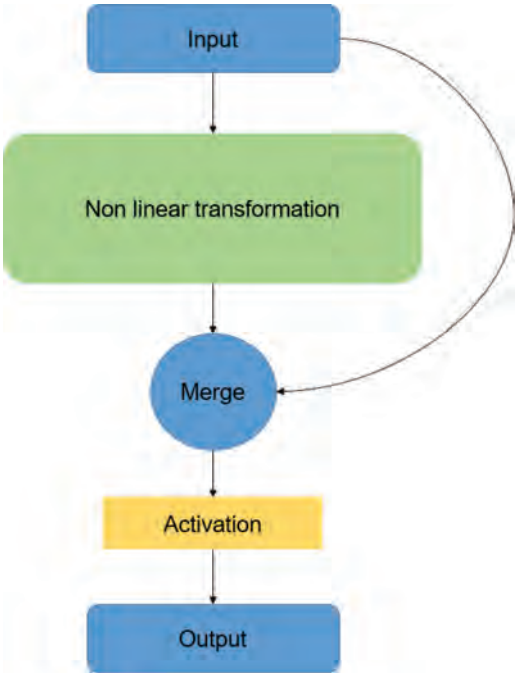


Figure 3. Residual block.

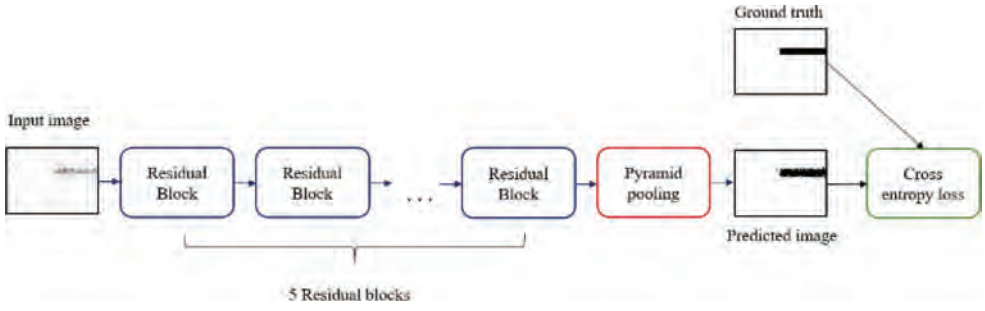


Figure 4. Proposed architecture.

The main idea underlying the residual networks is that if a network is working properly, it should not worsen its performance by increasing the number of layers. In this way, the residual networks are constructed in blocks as shown in Figure 3, where the information flows through the identity mapping if the network does not need to change the information that is reaching a certain layer. On the other hand, if it is required to make changes in the feature maps, mapping with different non-linear functions are applied.

In Figure 3, the classical components that have a residual block are shown, where the non-linear transformations correspond to the transformations that a traditional convolutional network could have. The "Merge" is the function designated to how the input of the block should be united with the output of the non-linear transformations. Following the "Merge", a non-linear activation is placed.

2.7 *Proposed architecture*

The proposed residual neural network architecture is comprised of five residual blocks followed by a pyramid pooling layer, as shown in Figure 4. Each residual block has the following structure (see Figure 5): (i) one convolution with stride of 1×1 and batch normalization; (ii) atrous convolution with filter of 3×3 and rate of 2 followed by batch normalization; (iii) another convolution with stride 1×1 and batch normalization; (iv) addition operation over all the resulting features maps and passed to a ReLU activation function. The resulting feature map is the input to the next residual block.

Note that the number of feature maps (FM) varies from block to block. For the first two blocks, the feature maps FM_1, FM_2 and FM_3 are equal to 256, 256 and 1024, respectively. For the remaining blocks, the number of feature maps FM_1, FM_2 and FM_3 are 512, 512 and 2056, respectively. After all the residual blocks, pyramid pooling is applied with 4 different rates, i.e., atrous convolutions with rates of 6, 12, 18 and 24. Moreover, the model uses cross entropy loss function and the parameters are updated via back-propagation with ADAM optimizer.

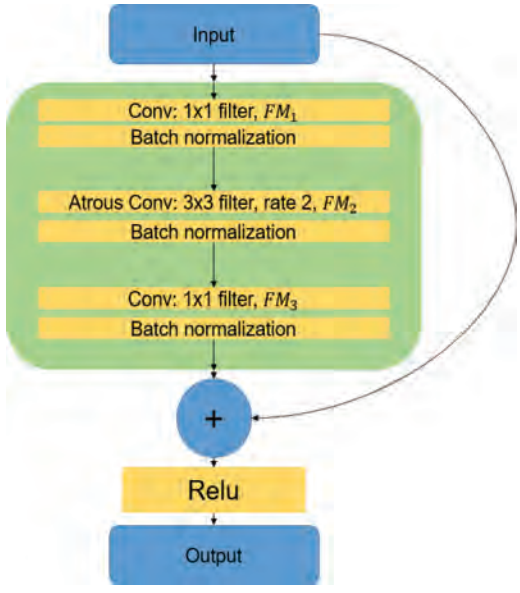


Figure 5. Residual block used.

3 EXAMPLE OF APPLICATION

3.1 *Data*

The dataset used in this example was obtained via finite element modeling of debonding damage in honeycomb structures and consists of 6000 noisy images of 98×69 pixels, of which 2000 images with circular damage, 2000 with rectangular damage and the remaining are undamaged. Note that the damage quantification is performed pixel by pixel, thus the classification task is separated into two: quantification of rectangular damage and

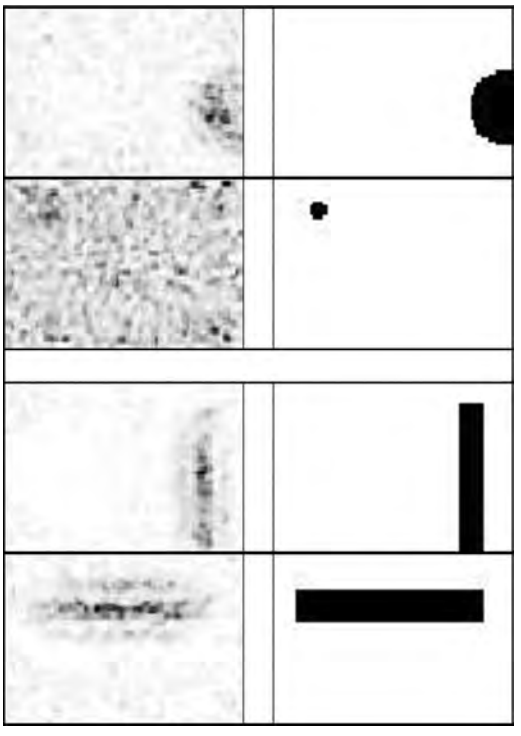


Figure 6. Examples of damage types and corresponding ground truth used for training: the two top images are for circular damage while the two bottom images are for rectangular damage.

quantification of circular damage. In both cases, the ground truth is the original image without noise. Figure 6 shows examples of rectangular and circular damage images.

The results presented in the next section were obtained by running the experiments in a computer with intel core i7 and 4.2 GHz processor capacity, 32 Gb RAM, Nvidia Titan X Pascal GPU with TensorFlow 1.3, Cuda 9 and Cudann 5.1 for compiling and neural networks optimizers.

3.2 *Metrics*

To analyze the damage quantification results, the following semantic segmentation metrics are considered: pixel accuracy (PA), mean accuracy (MA), mean intersection over union (MIoU) and frequency weighted intersection over union (FWIoU). Pixel accuracy in equation (3) corresponds to the overall accuracy. Mean accuracy, given in equation (4), is the average accuracy over all classes involved in the segmentation task. Mean intersection over union is shown in equation (5) and is the average of the correct pixel classification divided by the total number of pixels of that class, thus is an important performance metric for semantic segmentation as it encompasses true positives and false positives for the pixel by pixel classification. Frequency weighted intersection over union, as shown in equation (6), is similar to MIoU, but with the difference that FWIoU takes into account the number of data points in each class.

Therefore, MIoU and FMIoU are stricter metrics, while MIoU and MA are not sensible to unbalanced datasets, but FMIoU and PA could be inflated if that is the case.

$$\mathrm{PA} = \frac{\sum_i n_{ii}}{\sum_i t_i} \tag{3}$$

$$\mathrm{MA} = \frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{t_i} \tag{4}$$

$$\mathrm{MIoU} = \frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \tag{5}$$

$$\mathrm{FWIoU} = \left(\sum_k t_k \right)^{-1} \sum_i \frac{t_i n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \tag{6}$$

n_{cl}: Number of classes included in the ground truth segmentation.

n_{ij}: Number of pixels of class i predicted to belong to class j.

t_i: Total number of pixels of class i in the ground truth segmentation.

t_k: Total number of pixels in the ground truth segmentation.

3.3 *Results and discussion*

In this section, we present the results of the damage quantification obtained from the proposed model as well as comparison to two other approaches: K-means for segmentation with feature space in a similar fashion as presented in (Dhanachandra et al, 2015) and a deep fully convolution neural network (FCN) with the same architecture shown in (Long et al, 2015).

Indeed, Table 1 shows the results for damage quantification for the test dataset containing 200 images. For circular damage, both the proposed model and the FCN deliver quite similar performance metric results and significantly superior to the unsupervised K-means. This is mainly because the deep learning based models are supervised thus taking into consideration the source of the damage images. In terms of the MIoU metric, the proposed model performs better than the FCN because atrous convolutions do not lose detail information as occurs with the standard convolutions used in the FCN architecture.

Table 1. Segmentation results for exposed metrics. R: Rectangular damage, C: Circular damage.

	Proposed model %		FCN%		K-means clustering %	
	R	C	R	C	R	C
PA	97.8	98.5	97.5	98.6	96.5	97.2
MA	89.3	90.0	83.9	85.5	86.4	85.7
MIoU	81.2	82.6	76.7	81.6	76.5	76.9
FWIoU	96.1	97.2	95.5	97.4	94.2	95.3

The performance metrics for rectangular damage quantification for all models are worse than in the previous case as segmentation for this type of damage is a more complex task than the circular one because it requires obtaining greater contextual information and, at the same time, determining greater details (e.g., corners). Moreover, the proposed model outperforms the two other algorithms. In particular, the performance deterioration of the FCN model might be attributed to the information loss due to the use of pooling layers, whereas the proposed model uses atrous convolution that permits to capture contextual information without losing information.

Figure 7 shows examples of predicted damage quantifications delivered by the proposed and the FCN models for two test images: the top images are the noisy test data, the middle ones are the ground truths and the bottom images correspond to the predicted damages. It can be observed that the rectangular damage prediction from the proposed model is able to better capture the shape of the damage, in particular the corners, whereas the FCN model struggles in that task and thus leading to deterioration in its performance when quantifying this type of damage.

Table 2 and Table 3 show the no normalized and normalized versions of the confusion matrix for the proposed model for every pixel in test data and for both rectangular and circular damages. Note that the true negatives are 99.1% and 98.9% for circular and rectangular damages, respectively. These results can be explained by the unbalanced pixel count in the ground truth between no damage and damage. It is important to note also that the proposed model has higher scores for true positives with 85.9% and 79.7% for circular and rectangular damages, respectively, arguing in favor of the proposed model's robustness for the unbalanced data in this example of application.

The promising results obtained by the proposed model come at a computational cost that might be perceived as a disadvantage. In fact, the training time for the proposed model takes forty minutes for twenty epochs. However, it is significantly faster than the K-means in predicting the damage size of an unseen image: 0.07 second against 0.6 second for the latter. These results indicate that the proposed model might be a candidate for developing

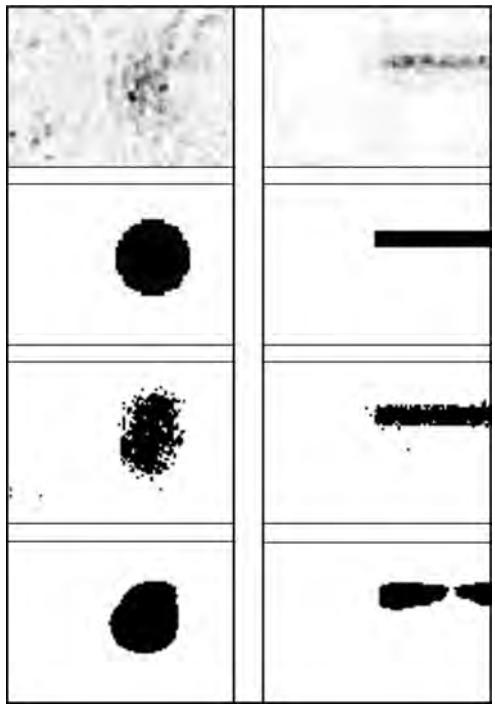


Figure 7. Top: test images; second row: ground truth; third row: damage prediction from proposed model; bottom: damage prediction from FCN.

Table 2. Confusion matrix of the proposed model. ND: No Damage, D: Damage.

	Circular damage		Rectangular damage	
	ND	D	ND	D
No Damage	1277298	11532	1262236	14340
Damage	8962	54608	15394	60403

Table 3. Normalized confusion matrix of the proposed model. ND: No Damage, D: Damage.

	Circular damage (%)		Rectangular damage (%)	
	ND	D	ND	D
No Damage	99.1	0.9	98.9	1.1
Damage	14.1	85.9	20.3	79.7

online monitoring systems where fast responses are required.

4 CONCLUDING REMARKS

This paper presented a novel residual atrous convolution neural network model for quantification of damage based on image processing. The proposed model was applied for damage quantification and segmentation of debonding damage in honeycomb structures.

The results show that the proposed model is a promising tool for damage quantification and segmentation with superior performance in noisier and more complex damage types and shapes than both deep fully convolutional networks and K-means algorithm.

REFERENCES

Arbelaez, P., B. Hariharan, C. Gu, S. Gupta, L. Bourdev, and J. Malik. Semantic segmentation using regions and parts. In CVPR, pages 3378–3385, june 2012.

Bishop, C.M. (2006). Pattern recognition. *Machine Learning*, 128.

Chen, L.-C., G. Papandreou, I. Kokkinos, K. Murphy, and A.L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv preprint arXiv:1606.00915, 2016.

Douka, E., S. Loutridis, A. Trochidis, Crack identification in beams using wavelet analysis, In International Journal of Solids and Structures, Volume 40, Issues 13–14, 2003, Pages 3557–3569, ISSN 0020-7683.

He, K., X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

Holschneider, M., R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in Wavelets: Time-Frequency Methods and Phase Space, 1989, pp. 289–297.

Karol Gregor, Ivo Danihelka, Andriy Mnih, Charles Blundell, and Daan Wierstra. Deep autoregressive networks. In Proceedings of the 31st International Conference on Machine Learning, 2014.

Kentaro Ohno, Masayasu Ohtsu, Crack classification in concrete based on acoustic emission, In Construction and Building Materials, Volume 24, Issue 12, 2010, Pages 2339–2346, ISSN 0950-0618.

Krizhevsky, A., I Sutskever, GE Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 1097–1105, 2012.

Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Lattanzi, D., & Miller, G.R. (2012). Robust automated concrete damage detection algorithms for field applications. *Journal of Computing in Civil Engineering*, *28*(2), 253–262.

Long, J., E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in CVPR, 2015.

Longpre, S., & Sohmshetty, A. (2016). Facial Keypoint Detection.

Nameirakpam Dhanachandra, Khumanthem Manglem, Yambem Jina Chanu, Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm, In Procedia Computer Science, Volume 54, 2015, Pages 764–771, ISSN 1877-0509.

Noh, H., S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, pages 1520–1528, 2015.

Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. The Journal of Machine Learning Research, 9999:3371–3408, 2010.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117.

Sutskever, I., Martens, J., Dahl, G.E., & Hinton,G.E. (2013). On the importance of initialization and momentum in deep learning. ICML (3), 28, 1139–1147.

Zhou, S.-B & Shen, A.-Q & Li, G.-F. (2015). Concrete Image Segmentation Based on Multiscale Mathematic Morphology Operators and Otsu Method. Advances in Materials Science and Engineering. 2015.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Return on investment on PHM systems

A. Segal & Y. Bot
BQR Reliability Engineering Ltd., Rishon LeZion, Israel

ABSTRACT: Return On Investment was calculated regarding a PHM system for a fleet of rolling stock. We show that the financial feasibility depends strongly on expected asset life as well as on the asset reliability model. The tool that was used for the analysis can be used for choosing the optimal PHM system for various assets.

1 INTRODUCTION

The goals of PHM systems are to provide advanced warning of system failures, enable Condition Based Maintenance (CBM), increase system availability, reduce Life-Cycle Cost (LCC), and reduce No Failure Found events (Pecht 2008).

While these goals are appealing, PHM integration projects include expensive installation of sensors, communication networks, big data storage and computing hardware or services (Feldman, Sandborn, and Jazouli 2008).

Therefore, it is not surprising that many asset owners are reluctant to venture into such projects. In order to assess the pros and cons of a suggested PHM project, Return On Investment (ROI) Analysis is required. The PHM system ROI is defined as:

$$ROI = \frac{C_1 + C_2 - C_3}{C_3} \quad (1)$$

where C_1 represents the financial cost reduction due to decreased downtime, C_2 is the financial savings due to reduced maintenance costs, and C_3 is the cost of PHM system procurement, installation and services.

While the cost of investment is relatively easy to quantify, the expected gain is hard to asses. A good PHM system is expected to reduce the occurrence of some failure modes. Therefore, Failure Mode, Effects, and Criticality Analysis (FMECA) is a good basis for PHM design (Meng and Zhang 2013; Banks, Reichard, Crow and Nickell 2009) and ROI analysis.

However, additional considerations are required. Maintenance policies such as inspections and scheduled maintenance affect the number of expected corrective and preventive maintenance events during the lifetime. Logistics policies affect the transportation and spare waiting times. These parameters heavily influence the asset/fleet maintenance cost as well as downtime penalties. Therefore, expected financial gain depends also on the facility/fleet operation profile, maintenance and logistics policies, and related LCC (Kacprzynski, M. J. Roemer and A.J. Hess 2002).

The general tradeoff of initial investment and maintenance cost for LCC optimization was discussed by W. Taylor (Taylor 1981). The question of PHM system ROI is a classic case for such tradeoff analysis.

Several PHM ROI evaluations were reported in the literature (see references in Feldman, Sandborn, and Jazouli 2008), mostly related to the defense industry. In recent years PHM and the Industrial Internet of Things (IIoT) penetrated other industries with focus on critical expensive equipment such as turbines in the Oil & Gas, wind farms, and aviation industries. These cases relate to new equipment with built in sensors.

In this paper we present an example regarding the ROI on a PHM system for an existing rolling stock asset. The LCC of several scenarios is considered and it is shown under which conditions PHM installation is financially advantageous.

2 CASE DESCRIPTION

A fleet of 17 trains located in two sites (10 trains in site A and 7 trains in site B) is considered. The trains operate 18 hours each day, and have 6 hours for scheduled maintenance and inspections during the night. A central stock services both sites.

The suggested PHM system will monitor the state of the rolling stock motors and pantographs. Adding the PHM system is expected to increase visibility and control of the rolling stock fleet, and to reduce failure events and downtime.

Following is a financial analysis of the ROI on the suggested rolling stock PHM system.

2.1 *Investment*

Adding the PHM system is expected to cost $20,000 per train for motor and pantograph sensors, and $160,000 for the central servers and operation center dashboard displays. An additional cost of $4,000 per month is paid for support and maintenance of the PHM system by the PHM provider.

This amounts to an investment of $1.22M for a 15 years period, or $1.94M for a 30 years period.

2.2 *Calculating expected gain*

In order to calculate the mean expected gain, the fleet behavior over the life period has to be predicted.

Most commercial system simulation software are based on the Monte-Carlo method. Indeed, Monte-Carlo simulations are flexible and easy to create. However, in order to achieve highly accurate results, many simulations have to be carried out. This is especially important in systems where rare events can have significant effect on the LCC.

We used the apmOptimizer software that includes a combination of analytic methods (Birolini 1999) for calculating the fleet Life-Cycle Cost (LCC), and identifying cost and failure drivers. The advantages of analytic calculations are speed and accuracy.

Calculations were carried out for various scenarios with and without the PHM system.

A detailed model of the existing fleet was constructed, accounting for:

Reliability Data

- Component failure distribution
- Component failure modes
- Rolling Stock redundancies
- Operation profile

Maintenance Data

- Component repair / discard policy
- Repair time
- Corrective maintenance
- Preventive maintenance
- Inspections

Logistic Data

- Spare parts
- Transportation times
- Procurement time

Financial Data

- Cost of spare parts
- Penalties due to service agreement
- Corrective maintenance
- Preventive maintenance
- Inspections

Using the input data, the apmOptimizer calculates the expected rolling stock availability, failures, maintenance, inspections, and LCC.

Figure 1 presents the breakdown tree of rolling stock in site A. The "Reliability Model" column in Fig. 1 describes the relevant model for each subsystem. The "Distribution Type" column in Fig. 1 presents the failure distribution type for each component. Electronic components were assigned an Exponential failure distribution whereas the mechanical components were given a Normal distribution that describes their ageing behavior.

The initial model described the fleet behavior without a PHM system. In order to calculate the behavior of the fleet with PHM, the original model was copied, and the following changes were implemented:

- Motor and several pantograph components have sensors; therefore, there is no need for scheduled maintenance, only Condition Based Maintenance (CBM).
- When CBM is conducted instead of scheduled maintenance, only the problematic components are treated. Therefore, each CBM event is cheaper than the corresponding scheduled maintenance event.

Four scenarios were considered:

- No PHM system, life-cycle of 15 years
- No PHM system, life-cycle of 30 years
- Added PHM system, life-cycle of 15 years
- Added PHM system, life-cycle of 30 years

Project	1	Serial	-	-
Rolling Stocks1	1	Parallel	-	-
RollingStock1	10	Serial	-	-
Sensors control console	1	Leaf	-	Exponential
Bogies	4	Serial	-	-
Sensors	1	Leaf	-	Exponential
Front/Back	2	Serial	-	-
Left/Right	2	Serial	-	-
Bearing1	2	Leaf	-	Normal
Case part	2	Leaf	-	Normal
Seal	1	Leaf	-	Normal
Shaft	1	Leaf	-	Normal
Wheel	1	Leaf	-	Normal
Brakes	1	K out of N	3	-
Brake units	16	Serial	-	-
Pneumatic unit	1	Leaf	-	Normal
Disc	1	Leaf	-	Normal
Motors	1	Serial	-	-
Gear	1	Leaf	-	Normal
Rotor	1	Leaf	-	Normal
Stator	1	Leaf	-	Normal
Windings	1	Leaf	-	Normal
Pantograph	1	Serial	-	-
Frame and ins.	1	Leaf	-	Normal
Valve plate	1	Leaf	-	Normal
Elavation sys.	1	Leaf	-	Normal
Arm	1	Leaf	-	Normal
Head	1	Leaf	-	Normal

Figure 1. Breakdown tree of rolling stock in site A.

For each scenario the optimal maintenance and logistics policies were found using apmOptimizer's enhanced dynamic programming algorithms. The optimization process was very fast due to the use of analytic calculations (as opposed to Monte Carlo). The optimizations goal is to minimize the LCC. LCC includes downtime penalty, therefore the optimization also ensures low downtime (high fleet availability). Furthermore, LCC includes the cost of spare parts, corrective and preventive maintenance and inspections.

The net gain of a PHM is equal to the calculated LCC of the fleet without PHM minus the LCC of the fleet with PHM, minus the PHM investment (this is $C_1 + C_2 - C_3$ from Eq. 1).

3 RESULTS

LCC was calculated for each optimized scenario. Table 1 presents a summary of the calculated results.

From Table 1 it is clear that a PHM system is not expected to yield financial savings when a lifecycle of 15 years is considered (ROI < 0). On the other hand, some savings are expected when a lifecycle of 30 years is considered (ROI = 0.278).

The analytic tool that was used produced additional useful information: it was found that the main contributor of rolling stock downtime is bearing failure. The suggested PHM program did not include bearing monitoring.

Table 1. Expected LCC for various scenarios.

Model	PHM investment	LCC	Total
Lifecycle 15 years			
No PHM	$0	$43.52M	$43.52M
With PHM	$1.22M	$42.35M	$43.57M
Lifecycle 30 years			
No PHM	$0	$100.9M	$100.9M
With PHM	$1.94M	$98.42M	$100.4M

Table 2. Expected LCC for various scenarios with enhanced PHM.

Model	PHM investment	LCC	Total
Lifecycle 15 years			
No PHM	$0	$43.52M	$43.52M
With PHM	$2.44M	$14.3M	$16.74M
Lifecycle 30 years			
No PHM	$0	$100.9M	$100.9M
With PHM	$3.88M	$29.38M	$33.26M

Next we consider the case where bearing sensors were also added. The bearing sensors, data analysis, storage and maintenance are expected to double the PHM cost. Table 2 presents a summary of expected LCC for the enhanced PHM system.

Data in Table 2 clearly demonstrates the financial advantage of PHM systems which are applied to the asset/fleet critical failure driver. The reduced LCC results from reduced maintenance costs as well as greatly reduced service penalties. ROI for 15 years is 10.97 while for 30 years the ROI is 17.43. These values are very high. One possible reason for the high ROI is the assumption that bearing failures are 100% prevented by the PHM system. The apmOptimizer can also account for non-ideal PHM systems.

4 CONCLUSIONS

A general conclusion is that PHM systems are most suited for asset intensive systems with long expected life-cycles (aircrafts, rolling stock, utilities, mining and O&G). The reason is as follows: While PHM reduces downtime penalties and maintenance cost, a high installation cost is incurred. For long life-cycles the accumulated savings overcome the initial PHM installation cost.

Another conclusion is that the right PHM system has to be selected in order to address the asset/fleet main drivers of failure and downtime. This requires preliminary field data collection and sensitivity analysis using modeling software such as the apmOptimizer.

REFERENCES

Banks, J., Reichard, K., Crow, E. and Nickell, K. 2009. *How Engineers Can Conduct Cost-Benefit Analysis for PHM Systems*, IEEE Aerospace and Electronic Systems Magazine (Volume: 24, Issue: 3).

Birolini, A. 1999. *Reliability Engineering Theory and Practice, 3rd edition*, Springer.

Feldman, K., Sandborn, P. and Jazouli, T. 2008. *The Analysis of Return on Investment for PHM Applied to Electronic Systems*, Proceedings of the International Conference on Prognostics and Health Management.

Kacprzynski, G.J., Roemer, M.J. and Hess, A.J. 2002. *Health management system design: Development, simulation and cost/benefit optimization*, IEEE Aerospace Conference Proceedings.

Meng, J. and Zhang, W. 2013. *Research on Missile PHM Design based on FMECA*, CHEMICAL ENGINEERING TRANSACTIONS, VOL. 33.

Pecht, M.G. 2008. *Prognostics and Health Management of Electronics*, CALCE, University of Maryland, Wiley Publication.

Taylor, W. 1981 *The use of Life Cycle Costing in Acquiring Physical Assets*. Long Range Planning, Vol. 14, No. 6, 32–43.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Reliability engineering based on operating data and monitoring systems within technical products: Challenges, requirements and approaches

S. Bracke & M. Hinz
Chair of Reliability Engineering and Risk Analytics, University of Wuppertal, Wuppertal, Germany

C. van Gulijk
Institute of Railway Research, University of Huddersfield, Huddersfield, England

F. Gronwald
Chair – Reliability of Technical Systems and Electrical Measurement, University of Siegen, Germany

M. Muenker
diondo GmbH, Hattingen, Germany

M. Inoue & S. Yamada
Department of Mechanical Engineering Informatics, Meiji University, Kanagawa, Japan

E. Patelli
Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK

B. Ulutas
Department of Industrial Engineering, Eskisehir Osmangazi University, Eskisehir, Turkey

M. Bonato
Valeo S.A., Paris, France

T. Yamada
Department of Informatics, University of Electro-Communications, Tokyo, Japan

ABSTRACT: The development process of complex technical products of the last years shows an increasing amount of sensors, electronic control units, data logging and monitoring systems within consumer goods (e.g. automobiles, washing machines) and industrial goods (e.g. machine tools, manufacturing systems). In many cases, the main goal of data logging is monitoring, controlling and optimisation of the product functionalities within the usage phase. A further aim is the fulfilment of the process capability of manufacturing processes. Therefore, the layout of the concept of operating data logging and monitoring systems—especially operating data (mainly type, volume, and format) as well as hardware (like sensors and storage)- is designed by the development engineer within the product concept development phase. This paper discusses challenges, requirements and approaches for future conceptual design of operating data logging concepts of technical products related to reliability engineering. Base of operations is the state of art. Based on that, the concept for operating data logging within a monitoring system is shown. The concept draft is subdivided in three parts which are divided as follows: part one deals with data analytics, part two contains data requirements, and part three focuses on hardware requirements. The presented research study was worked out on the international research platform "Computational Reliability Engineering in Product Development and Manufacturing (CRE) – 2017" and contains contributions of universities, institutes and original equipment manufacturers of industrial nations: Germany, United Kingdom, Japan, Turkey and France.

1 INTRODUCTION

The development process of complex technical products of the last years shows an increasing amount of sensors, electronic control units, data logging and monitoring systems within consumer goods (e.g. automobiles, washing machines) and industrial goods (e.g. machine tools, manufacturing systems). In many cases, the main goal of data logging is monitoring, controlling and optimisation of the product functionalities within the usage phase. A further aim is the fulfilment of the process capability of manufacturing processes.

Therefore, the layout of the concept of operating data logging and monitoring systems—especially operating data type, volume, format and storage—is up to this point of time in most cases designed by the development engineer within the product concept development phase. However, the design engineer is also responsible for the product functionality. Hence, the logged data is very often directly related to a technical discipline (e.g. automotive engineering: the rotation angle and cycle sensor is related to the antilock braking system which belongs to the division of chassis engineering). But this data can also serve as a foundation for the reliability analysis (e.g. automotive engineering: amount of steering turns, or frequency gathered from rotation angle sensor are live span variables, which can be used for statistical reliability models). Consequently, a comprehensive operating data logging (software) and monitoring system (hardware) for future technical complex product generations is needed with complementary functionality: Controlling product functionality and ensure product reliability of the actual and subsequently following generation.

2 GOAL OF RESEARCH ACTIVITIES

This paper discusses challenges, requirements and approaches for future conceptual design of operating data logging systems of technical products related to reliability engineering. In detail: (1) Requirements regarding operating data structure: e.g. data type, data volume, data format; (2) Requirements regarding data recording structure and hardware aspects: e.g. frequency, sensors and storage; (3) Aspects of data analytics based on gained operating data in the usage phase.

3 FUNDAMENTALS

3.1 *Design of the monitoring system within the product life cycle*

The product life cycle of technical products can be described in four main and eight subordinate phases, cf. (Bracke 2016):

1. Concept phase
   1a. Definition of the product characteristics
   1b. Development of the product concept
2. Development phase
   2a. Construction stages (different prototype levels and finalising the design)
   2b. Preparation of manufacturing
3. Production phase
   3a. Start of production (SOP)
   3b. Production
4. Sale/Usage phase
   4a. Sale of products to the markets
   4b. Usage phase and product observation

The concept of operating data logging and monitoring systems—especially operating data type, volume, format and storage – has to be designed by the reliability engineer of the Original Equipment Manufacturer (OEM) or Supplier within the product concept development phase (Phase 1b, cf. section 3.1).

3.2 *Design of operating data type*

In general, the operating data types can be subdivided in following different categories:

a. Secretly compiled data (OEM / Supplier):
   – Definition logging logic OEM,
   – Data encryption through OEM,
   – Storage strategy: “fleeting”, “semi-permanent”, “permanent”
b. Officially compiled data:
   Example automobile: eCall emergency system (since 31-03-2018), which gives an emergency call after an accident (“sleeping system”) and transfers basic automobile operating data.
c. Voluntarily compiled data:
   Example: Vehicle insurance, Policy with scoring option, logging function is always on.

The reliability engineer has to define the essential life span variables and operating data types in the concept development phase (cf. section 3.1). To ensure a long-term availability of the operating data regarding reliability analysis within the entire product life cycle, the storage strategy “permanent” (cf. numeration (a) above) is to be pursued. The strategy “fleeting” is only interesting for direct operating decisions, the strategy “semi-permanent” does not allow the data analysis after the end of product life. Officially compiled data is also interesting, if the storage strategy is “permanent”. Voluntarily compiled data is not in focus of this study.

4 OPERATING DATA AND MONITORING SYSTEM: DATA ANALYTICS, DATA AND HARDWARE REQUIREMENTS

Within this section, the data and hardware requirements, based on data analysis strategies, for an

operating data and monitoring system are shown. The concept draft is subdivided in three following parts: Part one deals with data analytics e.g. uncertainty, second life and lessons learned aspects (cf. section 4.1). Based on the goal of data analysis, part two and three contains data requirements (cf. section 4.2; e.g. structure and format) and hardware requirements (cf. section 4.3, e.g. sensors and storage availability within monitoring systems in products and facilities).

4.1 *Data analytics*

4.1.1 *Aspects of data uncertainty regarding a reliability model*

Operating data logging and monitoring systems are largely used to improve the knowledge of a specific system or component. However, data are always associated with some noise or measurement error, e.g. due to different environmental conditions. In turn, the model used to, e.g. predict the useful reaming life of a component, or schedule maintenance is also affected by uncertainty. If such uncertainties are neglected, some wrong and costly decision can be made (for instance, recall a product). Such uncertainty can also nullify the benefits of using machine learning frameworks for analysing the continuously increasing available data.

One of the current challenges in the capability is to discriminate when such machines and tools are providing reliable estimate or their prediction has been fooled by noise. One possible solution is to use past experience and predictions to determine precise levels of confidence for the new predictions (Shafer and Voyk 2008). Another popular approach is based on the Bayesian paradigm for inference. In such framework, Bayes' rule is used to update our believe on validity of the model prediction with information from empirical observations (data) taking into account the associate uncertainty in such observations. The reader is referred to (Aki Vehtari and Janne Ojanen 2012) for detailed description of Bayesian methods.

4.1.2 *Aspects of the use of product operating data for a second life cycle*

In order to avoid environmental issues, it is necessary to minimize the material and energy consumption during the whole product lifecycle (Yamada 2012). One of the potentials for material circulation environmentally and economically is to reuse the End-of-Life (EOL) assembly products by remanufacturing in the second life cycle. Remanufacturing is the process of bringing an assembly to like-new condition through replacing and rebuilding its components at least to current specification (Ilgin and Gupta, 2012). There are two essential processes in the remanufacturing: disassembly and re-assembly of the EOL products (Lambert and Gupta 2005).

To conduct the data analytics for the disassembling process in an environmental friendly and economical way, a parts selection method (Igarashi, et al., 2016; Kinoshita et al., 2016) including reuse (Hasegawa et al., 2017a, 2017b) shows which parts should be reused, recycled and disposed in terms of environmental impacts and costs. The operating data of a part in the usage stage of the first life cycle helps on the parts selection that can be disassembled. Here, one of the challenges is that the data affects the decision of the part selection itself by the recovery costs with different sales revenues for the parts.

4.1.3 *Transforming operating data in Lessons-Learned-Data-Structure for subsequently following product generations*

Application of data analytics to improve manufacturing operations and transferring critical information to following product generations are an integral part of data-driven decision making. When processes are better defined and more standardized, lessons can be combined into standards and guidelines. It is rather hard to share lessons in areas of complex or context-specific need, for topics that are rapidly changing, and where new problems are frequently being identified. Therefore, lessons should be written down and stored in a database in such a way that other related people can find and access the required knowledge. It is important to sort the individual lessons and store them under themes or topics in the lessons learned database. By this means, data can be filtered and previous actions of any problem can be considered during product's engineering design process. Updating the database (i.e., guidance documents, best practices and standards for the process) is also a crucial issue to sustain the garbage in, garbage out philosophy. The quality of lessons knowledge may change from extremely useful to completely unhelpful. Therefore, the ease and accuracy of transforming data for product generations depend on how data is collected, stored, and updated.

Commonly used data mining applications in manufacturing include failure evaluation, quality control, safety analysis, and capacity planning. Statistical Process Control (SPC) is one of the techniques suggested for real-time monitoring of operational performance of manufacturing systems. There is an ongoing research on better ways of collecting data and developing big data infrastructure technologies. Besides sensor technologies, Enterprise Resource Planning (ERP) and Manufacturing Execution Systems (MES) are also used for data collection technology the development of cloud service platforms.

4.1.4 *Aspects of reliability analytics: Probabilistic uncertainty based on input data/operating data*

The quality of the data can have a significant influence on the analysis results and their interpretation

which could cause a wrong conclusion of the product reliability. In fact, the amount, quality, format and processing of the data are only few of many other factors which have to be considered during the statistical analysis of operating data. It is still uncertain, whereupon and in which order it is necessary to pay attention by the examination of particular properties during the statistical analysis. A comprehensive list of factors which can influence the data analysis, or much more the results, is shown in (Hinz 2015). The proposed factors are divided into four groups:

- Data quality – demonstrates the requirements with respect to the compound of diverse inputs and properties of a data set (e.g. diverse load profiles or the unit on life span variables)
- Empiricism – knowledge based on experience regarding the application of the product fleet as well as market specific boundary conditions (e.g. product derivatives or user profiles)
- Aim of analysis – various purposes of the statistical analysis will result in the application of different methods which may cause further uncertainties (e.g. the kind of a damage case)
- Mathematical models – this group describes statistical models, equations and algorithms which can be used with regard to the reliability analysis (e.g. various methods for the estimation of the distribution parameters).

In plenty of cases, even the application of methods that can be expected to provide always reliable results may lead to high uncertainties. For example, the estimation of the shape parameters of Weibull distribution based on different estimators (here: Maximum Likelihood, Gumbel, Least Squares, Method of Moments, Nelson, and DIN 55303) and various sample sizes (varying between 10 and 1000) shall be considered. The results are shown in Figure 1. The mentioned methods are compared to the maximum likelihood (MLE) and the relative error is plotted as a function of the sample size.

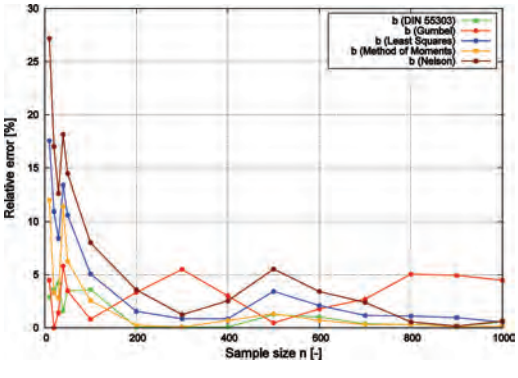


Figure 1. Relative error of the estimated shape parameters of Weibull distribution using various estimation models.

It can be easily observed, that especially for the small sample sizes, the application of different parameter estimators can cause big differences in the gathered results. For a sample consisting of 10 entries, the difference between MLE and Nelson equals to 28%.

In many cases combinations of multiple factors play a major role during the data analysis. This can cause a high uncertainty of the results which can differ exponentially from the reality.

4.2 *Data requirements: Operating data structure: e.g. data type, data volume, data format*

To some extent, the requirements for data are relatively flexible as long as that data supports the data-scheme for monitoring systems. The requirements therefore focus on data scheme and depend less on the actual data itself. The objective is to enable communication between the data-sender and the data-receiver. The data sender, say a logging system, has a data scheme to store relevant information in its own local database (which may be small if there are detectors only). The data may be system state messages, error messages or alarms that may contain a timestamp, serial numbers, identification codes, numeric values and meta-data. The date receiver, say a Matlab application for reliability engineering, has its own data scheme. Only part of the data from the data logger is useful for the Matlab application; some kind of data transformation is required. Such transformations can be made in many different ways; the key, however, is that the meta-data about the data-scheme is correct, informative and up-to-date. In many industries data standards have been developed to harmonize efforts of different industry partners; this tends to be efficient for many industries. For instance, the Oil and Gas industry uses ISO 15926 as an International Standard for the representation of process plant life-cycle information. This standard specifies a generic, conceptual data model that is suitable as the basis for implementation in a shared database or data warehouse. Many industries have developed similar standards; aligning with them in own field is well-worth the effort; IT solutions that do not follow the standards may not be accepted in the industry. Summarizing, data requirements focus on correct data scheme descriptions. For engineers, such descriptions are captured in technical reports. For the computer it is captured in the format of database.

4.3 *Hardware requirements*

4.3.1 *Aspects/requirements of hardware (sensor and storage technologies) within certain products*

The hardware components of data recording and monitoring systems typically represent a measurement chain that consists of four blocks for sensing, signal conditioning, signal processing, as well as data presentation and storage (Bentley 2005). Depending on the field of application, each hardware component has to fulfil certain requirements that are typically provided by established standards. An example of a general high-level standard is the RTCA-DO-160 standard for the environmental testing of hardware in an aerospace context (RTCA-DO160). The related tests are, for example, of a mechanical, chemical, or electromagnetic nature and, as a whole, rather involved and time-consuming. Depending on the field of application of a data recording and monitoring system, it might be neither feasible nor economical to perform the whole variety of desirable tests for each component. This leads to the challenge of determining, on a component level, meaningful hardware requirements that avoid excessive testing while keeping necessary standards. There is no general solution to this problem, which highly depends on the actual case. Existing standards, however, can provide valuable guidelines.

Moving from component to system level, the proper integration of a data recording and monitoring system into a complex technical product is required. Due to the general increase of complexity and electromagnetic sensitivity within technical systems, this task becomes increasingly challenging as well. The traditional approach for a proper integration involves the analytical and numerical modelling of possible unwanted electromagnetic couplings between different components (Tesche 1997). This is caused by the signal propagation between sensor and data unit along the aforementioned four blocks which is mainly of an electromagnetic nature. However, it has recently become apparent that increasing complexity requires, besides deterministic methods, also statistical methods that are adapted from reliability engineering to electrical engineering (Mao 2016). As a result and new development, both hardware requirements and aspects of uncertainty have to be considered as a whole during the system integration.

4.3.2 *Hardware requirements: Aspects of monitoring systems in complex technical facilities*

In an industrial environment a monitoring system will be developed, installed and operated only if a commercial benefit, either direct or on multi-level basis, can be expected.

The basic kind of monitoring is meant to prevent from system damage under use, offering upfront indications (e.g. life span variables like temperature, vibrations, etc.) for imperative service, usually combined with routine maintenance. It is applied for cheap and simple mass products. Second order monitoring is used to acquire data about the product quality during production (etc. weight, shape, homogeneity); it is recommended for mass products of some value and complexity. Ideally, this information is used to control the production process inline.

On the next level, sensor data combined with the operating parameter log can be used to continuously diagnose the present system status and so the product quality. It is combined with preferably non-destructive sample inspection of the product to verify the process' stability. Such kind of monitoring may also allow to adapt the maintenance frequency to the actual strain of the system. It is used for complex processes where reliability is the most important aspect (low-volume, costly or safety-related products).

Finally, these vast amount of information and data has to be merged with a system behaviour model to approve, recommend or tune warily operation modes and to venture a prediction for remaining lifetime, while the maintenance schedule is aligned with production requirements. On this level, the product is not necessarily a touchable item but could also mean energy (battery), information (data storage), or movement (aircraft engine).

Some of the economic effects of such monitoring efforts are obvious: increased lifetime, less downtime, higher throughput, less spares on stock and sufficient time to plan inevitable replacements. But there are also savings due to less failure in general, and less risk caused by unknown defective parts distributed into the market (product liability: documented monitoring is mandatory to defend claims). Wherever potential savings outbalance the costs an appropriate level of monitoring will be established.

4.3.3 *Conceptional aspects of standardisation of operating data recording within technical products*

The purposes of the operation data logging system regarding reliability engineering are detection of the cause of a failure or observation as well as prediction of failure and providence of maintenance action to user or producer. To formulate failure and its cause, monitoring system, and necessary and sufficient kind and number of sensors need to be allocated to product system. Hence, failure mode needs to be deployed into basic events by using FTA: Fault Tree Analysis (Lee et al.

1985) where designer selects appropriate sensor by reference to these events. On the other hand, many existing products such as automobiles and machine tools are already equipped with a monitoring system for attainment of its functionality. This monitoring system consists of sensors, ECU: Electronic Control Unit, and actuators and these components are modularized from the perspective of functionality usually by using DSM: Design Structure Matrix (Eppinger et al. 1994). Therefore, in the case of developing reliability monitoring system additionally, this system is desired not to change the structure of the existing functionality-it has monitoring structure which includes existing system modules. In addition, the occurrence of product failure depends on various elements such as the usage time, client usage, and other external factors. The sensors which are components of functionality-based system might not detect external factors such as temperature, humidity, and electromagnetic wave. Hence, for attainment of building failure-based monitoring system, designer needs to integrate undermentioned three steps: (1) deploying target failure mode to basic event by using FTA, (2) identifying necessary functionality-based monitoring systems and additional sensors for detecting external factors, and (3) modularizing these functionality-based sub-systems, additional sensors, and administration unit for transmitted signals. Figure 2 illustrates the concept structure of the failure-based monitoring system.

This failure-based monitoring system does not affect the structure of the existing functionality-based monitoring system. Therefore, this system has possibility to be optionally added to operating product system by upgrading without major design or structure changes.

4.3.4 *Hardware in use: Impacts on reliability of data recording safety within the product use phase*

The first impact of data recording systems within the product usage phase is related to scenarios in which the measurements are performed:

1. Field test: the goal is to measure the transient and steady state inputs of a vehicle as it operates over the real environment, in order to anticipate market region of use
2. Proving ground measurement: the goal is to replicate the most significant drive profiles from the field test, but in a more controlled environment (e.g. test track or climatic wind tunnel).

Field test are designed to capture all the environmental loadings that might affect the reliability of the vehicle or single components during its in-use phase. This type of measurements takes usually days or weeks. Data are recorded by the mean of a mobile data logger, which allows the simultaneous recording of a wide variety of sensor measurements. This type of device can be small and with integrated sensors, making them ideal for final customers' survey (Figure 3 left).

Proving ground measurements are based on the results from field test and focused on precise driving events during the development phase of a new vehicle. They are performed by an acquisition system, which guarantees more refined measurements, but it is less robust towards environmental stresses, and need to be interfaced to a laptop for data saving and storage.

Be either a road field test or a wind tunnel test, vehicle measurements are expensive. To perform such tests, one must first built a prototype (for Original Equipment Manufacturer - OEM) respectively buy or rent the selected car (for component suppliers). The required sensors need to be mounted, connected, and cabled. There must be enough room for all sensors, the cables, and a comfortable environment for both the driver and the acquisition system. Additional care must be taken when measuring the response of a component, which needs to be equipped with sensors (e.g. strain gages and thermocouples) by a reliable supplier, and then assembled in the vehicle.

Because of so many time and money consuming aspect, the key role of the test engineer is to make sure that measurement sessions are not jeopardize because of 1) improper sensor mounting 2) inadequate data acquisition/storage.

Once the suitable acquisition system has been considered for the measurements, the key aspect of a successful measurement lies on the type of sensors.

Sensors need to be tailored to the physical value of interest. It is therefore fundamental a prior

Figure 2. Structure of failure-based monitoring system.

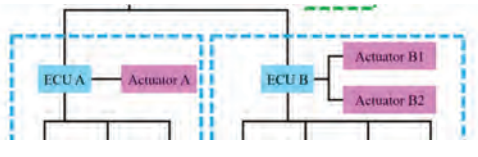


Figure 3. Small data logger with integrated sensors (left). PC interfaced Data acquisition system (right) (cf. (MSRDatenlogger)).

knowledge of the value range (maximum and minimum value, the frequency etc.).

Sensors must also be accurate, enough sensitive to properly measured small variations, but robust toward the inevitable environmental stress resulting from driving ground: shock, heat, humidity, and contamination associated to the potential adverse conditions of the road profile (dust, mud, water etc). In general, sensors must be operative under all ambient temperature conditions.

Similarly, the acquisition system must be tailored to the measurement type. There must be a trade-off between system performance and robustness and in some case its dimensions. As a typical example, integrated circuit piezoelectric (ICP) accelerometers are less intrusive (smaller and lighter) and more accurate than capacitive ones, but less resistant to high temperature and shock. ICP sensors would perfectly fit for vibration measurements on the chassis or cabin component, but be unreliable for engine vibrations, due to the high temperature reached during combustion.

High care should be taken when mounting the sensors: external factors that might interfere with the measurements are electrical leakage and shorts. Moreover, the electromagnetic compatibility of the acquisition systems and logger should be verified to avoid the presence of electromagnetic parasitic noise.

Some additional consideration and precaution should be used when planning long field measurements.

The settings of the data logger (usually mounted inside the cabin) require a trade-off between frequency of acquisition and storage memory. Particular care must be taken for acceleration measurements, since their high frequency and long acquisition time might be computationally demanding during data post-processing.

Solid State Drive (SSD) memory devices are preferred to Hard Disk Drives (HDD) because they are less affected by dust contamination and vibration loading. Moreover, SDDs do not require rotating component such as the platter or the fan system.

Planning of field measurements also need to consider the available memory and how it works, to avoid data loss. The most commonly used memory storage methods are i) erasable data storage systems (once the memory is full, after a certain time the system is erased) and ii) circular or buffer memory (oldest data gets overwritten when the memory is full).

During field or proving ground measurements, both data loggers and acquisition systems can be used to record data coming from the on Board Diagnostics (ODB) or the control area network (CAN) bus. It is obvious that when recording both data from sensor and form the vehicle on-board computer, the measurements needs to be synchronized.

A check-list to avoid potential problems encountered during field measurements could include the following topics:

– Sensors: must be robust, accurate and properly calibrated.
– Acquisition system: suitable to the type of measurements (portable PC interfaced vs. data logger).
– Memory storage: chosen with respect to the amount of expected data, limitation of the memory capability and post-processing computational effort.
– Post processing: properly labelling of measurements channels; data saved in an exploitable format.
– Privacy issue: field tests on final customers (e.g. users' fleet) must be compliant to privacy policy, which varies from one country to another.

5 SUMMARY

The development process of complex technical products of the last years shows an increasing amount of sensors, electronic control units, data logging and monitoring systems within consumer goods (e.g. automobiles, washing machines) and industrial goods (e.g. machine tools, manufacturing systems). This operating data can be a foundation for statistical analysis regarding the reliability of the product. The goals of data analytics (focus: data uncertainty, reliability analytics, second-life-cycle aspects, Lessons-Learned issues) are the base of operations for the data and hardware requirements.

Main requirements regarding the monitored data are as follows:

- Clear data Scheme regarding the local data storage system,
- Data content (storage): Messages, error, alarms, timestamp, serial number, identification codes, numeric values, meta data,
- Data receiver: Possibility of data transformation
- Consideration of industrial data standards, depending on product category,
- Possibility of technical report,

Main requirements regarding the monitoring system hardware are as follows:

- Considering standards for sensing, signal conditioning, signal processing, data presentation and storage,
- Considering possible electromagnetic sensitivity regarding signal propagation between sensor and data unit,
- Modularisation of monitoring system components,

- Considering upgrade possibility during product life cycle,
- Considering load profile regarding expected product life cycle within monitoring prototype testing (field test versus proving ground measurement).

The shown requirements can be used as a guideline for the reliability engineer for the design of operating data logging and monitoring systems within the product concept development phase of a new product generation.

REFERENCES

Bentley, J.P.: "Principles of Measurement Systems", 4th ed., (Pearson, Harlow, 2005).

Bracke, S., Hinz, M., Inoue. M., Patelli, E., Kutz, S., Gottschalk, H., Ulutas, B., Hartl, C., Mörs, P. and Bonnaud, P.: Reliability engineering in face of shorten product life cycles: Challenges, technique trends and method approaches to ensure product reliability. In: L. Walls, M. Revie, T. Bedford; Risk, Reliability and Safety: Innovating Theory and Practice; ESREL 2016, Glasgow, United Kingdom, 25th – 29th September 2016; European Safety and Reliability Association, ESRA (2016).

Eppinger, S.D., Whitney, D.E., Smith, R.P., & Gebala, D.A.: A Model-Based Method for Organizing Tasks in Product Development. *Research in Engineering Design* 6(1): 1–13. 1994.

Hasegawa, S., Kinoshita, Y., Yamada, T., Inoue, M., Bracke, S.: Disassembly Parts Selection for Recovery Rate and Cost Considering Reuse, The 24th International Conference on Production Research (ICPR-24), Poznan, Porland, July (2017a).

Hasegawa, S., Kinoshita, Y., Yamada, T., Inoue, M., Bracke, S.: Disassembly Parts Selection for Material-based CO2 Saving Rate and Cost Considering Reuse, The 3rd International Conference on Remanufacturing, (ICoR2017), Linkoping, Sweden, pp.175–188, Oct (2017b).

Hinz, M., Sochacki, S., Rosebrock, C. and Bracke, S.: Qualitative and quantitative analysis of uncertainties in the risk analysis of field data within the product usage phase. In: L. Podofillini, B. Sudret, B. Stojadinovic, E. Zio, W. Kröger; Safety and Reliability of complex Engineered Systems; ESREL 2015.

Igarashi, K., Yamada, T., Gupta, S.M., Inoue, M., Itsubo, N.: Disassembly System Modeling and Design with Parts Selection for Cost, Recycling, and CO2 Saving Rates using Multi Criteria Optimization, *Journal of Manufacturing Systems*, Vol.38, No.41, pp.151–164 (2016).

Ilgin, M.A., Gupta, S.M.: Remanufacturing Modeling and Analysis, Boca Raton, FL, USA: CRC Press; 2012.

Kinoshita, Y., Yamada, T., Gupta, S.M., Ishigak, A., Inoue, M.: Disassembly Parts Selection and Analysis for Recycling Rate and Cost by Goal Programming, *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, Vol.10, No.3, pp.1–15 (2016).

Lambert, A.J.D., Gupta, S.M.: Disassembly modeling for assembly, maintenance, reuse and recycling, Boca Raton, FL, USA: CRC Press; 2005

Lee, W.S., Grosh, D.L., Tilman, F.A. & Lie, C.H.: Fault Tree Analysis, Methods, and Applications: A Review. IEEE Transactions on *Reliability* R-34(3): 194–203. 1985.

Mao, C. and Canavero, F.: "System-Level Vulnerability Assessment for EME: From Fault Tree Analysis to Bayesian Networks—Part I: Methodology Framework", IEEE Transactions on *Electromagnetic Compatibility*, vol. 58, no. 1, (February 2016), pp. 180–187.

MSRDatenlogger – Source: MSR Electronics, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=18557569 and Müller-BBM GmbH, www.muellerbbm.com.

RTCA-DO160: "Environmental Conditions and Test Procedures for Airborne Equipment", Version G, (Radio Technical Commission for Aeronautics, 2010).

Shafer, G., Vovk, V.: A Tutorial on Conformal Prediction, Journal of Machine Learning Research 9 (2008) 371–421.

Tesche, F.M., Ianoz, M.V. and Karlsson, T.: "EMC Analysis Methods and Computational Methods", (John Wiley & Sons, New York, 1997).

Vehtari, A. and Ojanen, J.: A survey of Bayesian predictive methods for model assessment, selection and comparison Statistics Surveys, Vol. 6 (2012) 142–228.

Yamada, T.: Part 6, 6.2, design of closed-loop and low-carbon supply chains for sustainability. In: Soemon Takakuwa, Nruyen Hong Son, Nguyen Dang Minh (Editors), Manufacturing and environmental management. Hanoi, Vietnam: National Political Publishing House; 2012. pp. 211–221.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Enhanced hybrid prognostic approach applied to aircraft on-board electromechanical actuators affected by progressive faults

P.C. Berri, M.D.L. Dalla Vedova & P. Maggiore
Department of Mechanical and Aerospace Engineering (DIMEAS), Politecnico di Torino, Torino, Italy

ABSTRACT: In the last generation aircraft, the architecture of the powered flight control system adopted Electromechanical Actuators (EMAs). Being some on-board actuator safety critical, the practice of monitoring their behavior to determine their health condition is a task of growing importance. The choice of the best prognostic algorithm is driven primarily by their effectiveness in correctly identifying the health conditions of the system since each technique might be more or less useful in a given situation. In this contest, the authors propose a new GA-based fault detection tool, relying on a model-based approach, comparing the system output to that of a Monitor Model (MM), which is able to reproduce accurately the dynamic response of the actual EMA in terms of position, speed and equivalent current, even under the effects of different failure modes while keeping a reasonably low computational cost; this Fault Detection and Identification (FDI) algorithm have been extended to seven progressive failures. A numerical simulation test environment has been developed to simulate progressive faults and to evaluate the accuracy of this prognostic method. Results showed an adequate robustness and a suitable ability to early identify malfunctions with low risk of false alarms or missed failures. Moreover, the effect of a failures different from those considered was studied, to avoid safety concerns related to the missed identification of an incipient failure, hidden by another unknown failure mode.

1 INTRODUCTON

Electromechanical Actuators, or EMAs, are gradually replacing hydraulic systems in aeronautical applications, starting from the less safety critical uses (such as the actuation of trim tabs, cargo bay doors, or weapon and sensor systems of military aircraft) up to the most important ones, like primary and secondary flight controls. EMAs are composed of an electric motor driving the user through a mechanical transmission; then, the main advantages over a traditional electrohydraulic system are the absence of a centralized hydraulic power generation and distribution system, leading to an overall weight reduction, and the total absence of the hydraulic fluid itself, which is usually pollutant or flammable.

Given that the EMA technology is quite new and the reliability of these systems is not yet adequately known, risk reduction methods based on redundancy and scheduled maintenance and inspections shall be heavily employed, which somehow limits the diffusion of this new technology due the unavoidable cost increase. Moreover, those risk reduction methods can do nothing against failures caused by unexpected and extreme scenarios such as the exceeding of the flight envelope.

A new approach to the risk reduction, called Prognostics and Health Management (PHM), relies on the monitoring of functional parameters of the system to detect and identify the precursors of failures at an early stage (Vachtsevanos et al. 2006), in order to estimate the Remaining Useful Life (RUL) of the components. The monitored parameters shall be usually converted into electric signals, so a PHM approach is particularly convenient when applied to an electromechanical system, where most parameters are already in form of electric signals without the need for dedicated sensors and transducers, which would increase the overall costs and worsen the system basic reliability. In literature, many different Fault Detection and Identification (FDI) methods have been investigated: model-based techniques based on the direct comparison between the output of real and monitoring system (Raie & Rashtchi 2002, Byington el al. 2004, Alamyal et al. 2013), on the spectral analysis of well-defined system behaviors performed by Fast Fourier Transform (Mamis et al. 2013, Dalla Vedova et al. 2014), on combinations of these methods (Borello et al. 2009a, Dalla Vedova et al. 2015a,b) or on Artificial Neural Networks (Su & Chong 2007, Hamdani et al. 2011, Refaat et al. 2013, Dalla Vedova et al. 2016a).

This paper proposes an FDI algorithm relying on a model based approach and, in particular, on parametric estimation, for the prognostic analysis of a typical EMA according to the More Electric Aircraft and All Electric Aircraft paradigms (Quigley 1993, Howse 2003); the robustness of this algorithm is tested under different operating conditions and the effects of its use integrated with the traditional RAMS approach is investigated.

2 REFERENCE AND MONITOR MODELS

Two numerical models of the actuator were developed for this study. A very detailed reference model is used as a virtual test rig for the FDI algorithm, simulating the behavior of the faulty physical system. The computing time required by this model, however, is not compatible with the use in the FDI algorithm itself, which involves an iterative evaluation of the fitness function in Genetic Algoritm (GA). For this reason, a simplified monitor model was built to achieve a light computing cost and, at the same time, a high accuracy in reproducing the early effects of different incipient fault modes.

The Reference Model (RM), widely described by Dalla Vedova et al. (2016b), contains a detailed simulation of the physical phenomena acting in the EMA, in particular regarding the electromagnetic stator-rotor coupling (Haskew et al. 1999, Lee & Ehsani 2003, Halvaei et al. 2009, Çunkas & Aydoğdu 2010), end-of-travels, compliance and backlashes acting on the mechanical transmission (Borello et al. 2009b, Borello & Dalla Vedova 2014), dry friction acting on bearings, gears, hinges and screw actuators (Borello & Dalla Vedova 2012) and a precise model of the behavior of the power electronics, including the solid-state inverter and the PWM control of the three electrical phases.

The Monitor Model (MM) is a simplified representation of the system using, for example, an equivalent single-phase DC motor with a single feedback loop instead of the complex electromagnetic model of the BLDC. This requires the introduction of a shape function based model for the simulation of the electrical fault, which is not strictly related to the physics of the system, but allows to reproduce the effects of faults with goof accuracy, as shown by Berri, Dalla Vedova & Maggiore (2016).

3 FAULT MODES

Five different fault modes were considered for the study. Those were chosen among the most common for EMAs, as highlighted by (Kenjo & Nagamori 2003, Chesley 2011, Weiss, 2014); moreover, they are usually characterized by a progressive evolution, making possible an effective prognostic detection.

The considered faults are briefly listed below:

- Dry friction due to the wearing of mechanical components;
- Backlash of the reducer gearbox and/or rotary-to-linear conversion device;
- Partial short circuit of the BLDC stator coils;
- Rotor eccentricity due to the degradation of its support bearings;
- Control electronics fault resulting in the drift of the PID controller Proportional gain.

The implementation of the first four faults in both the RM and MM is described by Berri, Dalla Vedova & Maggiore (2017); the last one is modelled by varying the Proportional gain parameter in the Controller subsystem of both models. Despite the relatively straightforward implementation of this fault, non negligible difficulties were found due to the position of the affected subsystem in the feedback loop, in particular considering the interactions with other fault modes: in fact, its effects are hardly distinguishable from those of partial short circuit.

4 FITNESS FUNCTIONS

The objective function to be optimized by a GA is known as the fitness function. In the proposed FDI technique, it is the cumulative error in terms of equivalent single-phase current between the MM and RM. This results in a 8-variables function to be optimized: in fact, two of the considered fault modes have multiple degrees of freedom. The rotor eccentricity is characterized by a its magnitude ζ and its phase ϕ (i.e. the angular position of the minimum air gap measured from the reference rotor angular position); similarly, the partial short circuit can affect each of the three stator phases, which have to be treated separately in order to isolate this fault mode from the others.

The fault parameters are normalized as follows:

- $k(1)$ is the normalized friction: $k(1) = 0$ means nominal conditions, while $k(1) = 1$ means 300% of nominal condition;
- $k(2)$ is the normalized backlash: $k(2) = 0$ means nominal condition, $k(2) = 1$ means 100 times the nominal condition; although at a first glance this range may seem exaggerated, 100 times the nominal condition means about half a radian of mechanical play on the fast shaft, reduced by the gear ratio to $5.7 \cdot 10^{-2}$ degrees on the slow shaft or 20% of the already small chirp command amplitude.

– *k(3)*, *k(4)* and *k(5)* are respectively the normalized short circuit of phases A, B and C; for example, $k(3) = 0$ means a fully functional phase A, while $k(3) = 1$ means a complete short circuit for the same phase.
– *k(6)* is the rotor static eccentricity amplitude: $k(6) = 0$ means no rotor eccentricity, while $k(6) = 1$ means $\zeta = 1$; in fact, *k(6)* is equal to the eccentricity parameter ζ (Belmonte et al. 2015)
– *k(7)* is the phase of rotor eccentricity ϕ, i.e. the direction corresponding to the minimum air gap; $k(7) = 0$ means $\phi = -180°$, $k(7) = 1$ means $\phi = 180°$.
– *k(8)* is the normalized variation of the proportional gain: $k(8) = 0$ is a 50% reduction of the proportional gain, while $k(8) = 1$ means a 50% increase.

The fitness function is then computed with a modified total least squares method, which is tolerant to small phase lags cumulated between the two EMA models, even in presence of steep gradients and abrupt changes in the equivalent current (Markovsky & Van Huffel 2007, and Berri, Dalla Vedova & Maggiore 2016). The resultant error is therefore:

$$err = \int_0^{t_{sim}} \frac{\left(I_r - I_m\right)^2}{\frac{\left(dI_r / dt\right)^2}{k} + 1} dt \quad (1)$$

where I_r and I_m are the reference and monitor single phase currents, t_{sim} is the duration of the simulation and k is a constant used normalize the derivative of the reference current. Alternatively, computing the integral with a numerical scheme, considering the discrete nature of the simulations:

$$err = dt \cdot \sum_i \frac{\left(I_{r(i)} - I_{m(i)}\right)^2}{\frac{\left(dI_{r(i)} / dt\right)^2}{k} + 1} \quad (2)$$

It has to be noticed that in this formulation dt assumes the meaning of the finite time step of the numerical integration.

5 GA SETTINGS

The genetic algorithm used for this study is based on the *ga* function available in the MATLAB *Optimization Toolbox*. Employing an eight-variables fitness function requires a little calibration of the GA to achieve a good convergence. However, various parameters are left to their default value, so a further optimization of the algorithm settings is likely to allow a performance improvement, increasing the convergence speed. In particular, the *function tolerance* in the *stopping criteria* is tightened from the default $1 \cdot 10^{-6}$ to a value of $1 \cdot 10^{-9}$, to prevent the method to stop in a local minimum; moreover, the maximum iterations parameter is changed from 100 to 200, in order to avoid stopping the algorithm too early, before convergence is reached. Moreover, the *hybrid function* option is set to start a deterministic optimization with the *fmincon Interior Point* solver at the end of the genetic algorithm to refine results. The gradient based algorithm starts from the final point of the GA, offering a faster way to converge on the optimum solution, while the genetic algorithm provides a suitable start point to prevent convergence on local minima, ensuring robustness of the method.

6 RESULTS

In order to measure the accuracy of the fault identification in all the eight variables, a total error was defined as the quadratic mean of the errors on single variables:

$$e_{tot} = \sqrt{\frac{1}{8}\left(e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2 + e_6^2 + k_6^2 e_7^2 + e_8^2\right)} \quad (3)$$

The FDI algorithm was executed several times, with the fault parameters set to different values, in order to test the effectiveness and accuracy of the proposed method. For each combination of faults, the GA was executed ten times, because the method is inherently non-deterministic, involving the iterative random choice of points for evaluating the fitness function. This way it was possible to assess the repeatability of results, which were not strongly influenced by the aforementioned random choice.

Then, to assess the robustness of the algorithm, the variance σ^2 of results and total error is used:

$$\sigma^2 = \sum_{i=1}^{N} \left(k(j)_i - \overline{k(j)}\right)^2 \quad (4)$$

where N is the number of optimizations, $k(j)_i$ is the value of the j-th fault parameter resulting from the i-th optimization and $\overline{k(j)}$ is the average value of the j-th parameter over N optimizations.

The following table reports, as an example, the results obtained with three of the optimizations for the medium damage level combination of faults.

Tables 2 to 5 summarize the results of 40 optimizations, performed with multiple faults and different damage levels. For each damage level, the GA was executed ten times, and the arithmetic mean

Table 1. Optimizations for medium multiple damage.

Reference		Optimizations #1	#2	#3
k(1)	0.4	0.4129	0.3985	0.4090
k(2)	0.4	0.4620	0.4254	0.4303
k(3)	0.2	0.1756	0.1796	0.1798
k(4)	0.0	0.0060	0.0109	0.0000
k(5)	0.0	0.0067	0.0175	0.0005
k(6)	0.2	0.1912	0.2085	0.1961
k(7)	0.5	0.4986	0.5149	0.5037
k(8)	0.6	0.6131	0.6082	0.6024
Total Error e_{tot} [%]		2.48	1.43	1.33

Table 2. Results for a low damage combination.

Fault parameter reference value		Mean value (over 10 optimizations)	Variance
k(1)	0.1	0.0953	6.019E-05
k(2)	0.1	0.1109	3.424E-04
k(3)	0.1	0.0876	5.908E-04
k(4)	0.0	0.0069	8.215E-05
k(5)	0.0	0.0021	1.040E-05
k(6)	0.1	0.0944	3.706E-04
k(7)	0.5	0.5085	1.412E-04
k(8)	0.4	0.3940	7.776E-04
Total Error e_{tot}		1.51%	1.029E-04

Table 3. Results for a medium damage combination.

Fault parameter reference value		Mean value (over 10 optimizations)	Variance
k(1)	0.4	0.4017	1.774E-04
k(2)	0.4	0.4373	5.558E-04
k(3)	0.2	0.1809	1.780E-04
k(4)	0.0	0.0074	1.048E-04
k(5)	0.0	0.0088	1.414E-04
k(6)	0.2	0.1982	1.245E-04
k(7)	0.5	0.5039	2.010E-05
k(8)	0.6	0.6050	7.051E-04
Total Error e_{tot}		2.00%	8.916E-05

value of the Fitness Function input vector was computed, along with its variance. Moreover, average value and variance of the total error are provided. It can be noticed that, for damage levels low to medium (and therefore in the prognostic field of interest), both average total error and variance of the results are satisfactorily low, meaning that the proposed prognostic tool is both accurate and robust.

Table 4. Results for a high damage combination.

Fault parameter reference value		Mean value (over 10 optimizations)	Variance
k(1)	1.0	0.9727	2.719E-04
k(2)	1.0	0.9728	7.219E-05
k(3)	0.5	0.5623	4.757E-04
k(4)	0.0	0.0086	9.420E-05
k(5)	0.0	0.0071	7.816E-05
k(6)	0.5	0.6912	2.491E-04
k(7)	0.5	0.5019	1.554E-05
k(8)	1.0	0.9948	4.585E-05
Total Error e_{tot}		7.33%	4.190E-05

Table 5. Medium damage combination with noise.

Fault parameter reference value		Mean value (over 10 optimizations)	Variance
k(1)	0.4	0.3929	2.546E-04
k(2)	0.4	0.4072	1.131E-03
k(3)	0.2	0.1895	2.414E-04
k(4)	0.0	0.0028	9.730E-06
k(5)	0.0	0.0085	5.591E-05
k(6)	0.2	0.2135	5.635E-04
k(7)	0.5	0.5031	7.505E-06
k(8)	0.6	0.5996	1.064E-03
Total Error e_{tot}		1.97%	8.469E-05

The high level damage, with its error rising up to over 7%, starts to show the divergence between MM and RM; however, this case is considered only to assess the range of applicability of the model, but is not of practical interest. Eventually, the introduction of a white noise disturbance superimposed to the controller output signal in the RM seems not to affect the method accuracy nor its robustness (as shown in Table 5).

Figure 1 shows the probability distribution of the total error, with data gathered in all the performed optimizations. It can be noticed that, for most optimizations, the total error has a very low value, in the order of 1%. The smaller peak of the probability distribution, settled around 7%, is indeed caused by the simulations executed with a high damage level, which causes the behavior of the two models to diverge slightly. However, this case is not of practical interest for prognostic applications (in fact, such a high damage results in a jammed actuator response and therefore lies in the field of diagnostics), but rather is considered to assess the limits and the applicability of the MM damage implementation.

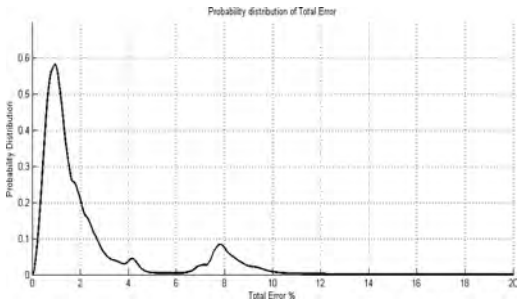


Figure 1. Probability distribution of the total error e_{tot}.

Table 6. Effect of unknown failure mode.

Initial conditions	Fitness function final value
k_f = 0.02 Nm/rad	0.0618
k_f = 0.05 Nm/rad	0.3609
k_f = 0.1 Nm/rad	1.8743
Baseline (low to medium damage, known faults only)	3.049E-04

In addition, in order to partially rule out the possibility of an unknown failure mode (i.e. one different from those considered in the models) not being detected by the GA, a fictitious failure consisting in the introduction of an elastic external force, with elastic constant k_f varying between 0.02 and 0.1 Nm/rad on the motor shaft, has been added to the RM only, to evaluate its effect on the FDI algorithm. The system health status in this case is partially misinterpreted, since the total error shows an increase in its value. However, the presence of an unknown failure is detected by the fitness function failing to converge to a near zero value.

Table 6 shows the optimized value of the fitness function for different entities of the added fictitious failure, compared to the mean baseline value referred to low to medium damage status with only known fault modes.

It can be seen that the introduction of an although small unknown failure results in an increase of the fitness function value after the optimization of 2 to 4 orders of magnitude. This increase can be easily related to the presence of a condition not identifiable by the algorithm, as an unknown failure mode or a too high damage level.

Each failure mode affecting the waveform of the monitored variables in a recognizable pattern, it appears unlikely that two failures can cancel each other resulting in a missed identification; however, this aspect shall be further investigated in future works.

7 RELATION BETWEEN RAMS AND PHM APPROACHES

Although the RAMS and PHM disciplines are seldom considered together in practical applications, they are closely related and their integration in a unified approach to the system life cycle management can lead to significant benefits in terms of safety and cost effectiveness (Dersin et al. 2017).

7.1 *Effect of PHM approach on RAMS*

The use of a prognostic tool to detect the early effects of incipient faults can be an effective and powerful method to improve the characteristics of Reliability, Availability, Maintainability and Safety of a system. A reliable PHM strategy can in fact enable the cost effective implementation of Predictive Maintenance and Condition Based Maintenance practices: with this approach, maintainability is affected by the improvement of the self-diagnostic capability of the system, which reduces the time to detect the fault to be fixed. In fact, with a PHM approach for the management of the system life cycle, most maintenance interventions are actually preventive scheduled maintenance. The corrective maintenance is then reduced to the replacing of those components that underwent failure modes impossible to detect in advance.

The overall Maintenance Man Hours/Flight Hour of a given aircraft system are then decreased, as well as the maintenance related operating costs.

The above mentioned effect on maintenance greatly improves the availability of the system. Being the corrective maintenance reduced, most of the necessary interventions can be scheduled in advance to be performed in the already slated ground time.

This allows a more cost effective management of the fleet, and in some cases also the reduction of the number of aircrafts necessary in the fleet for a given commercial airline or military service.

On the other hand, a prognostic fault detection and useful life estimation would virtually eliminate the risk of having faulty components flying on an aircraft in service. Then, the effective failure rate of the considered components is reduced, since worn components are replaced before their damage level starts affecting the performance of the system. Therefore, both the reliability and safety are improved, since they are both related to the failure rates of the system (the difference between them is only the potential effect of the considered failures).

7.2 *Use of RAMS tools for PHM purposes*

Various analytic tools commonly used in the RAMS disciplines can be employed to perform an

effective PHM activity, leading to safety improvement and cost reduction. Through the Failure Modes Effect & Criticality Analysis (FMECA) it is possible to identify the most significant failure modes of the system in terms of their impact on service reliability and on safety, based both on their effects at subsystem and aircraft levels and on their failure rate, estimated through statistical analysis of field data and return on experience. Therefore, focusing the PHM activity on the prediction of the occurrence of those failure modes can lead to a significant benefit in terms of improvement of safety and service reliability.

On the other hand, the Life-Cycle cost analysis is intended to identify the most cost affecting and availability affecting maintenance actions. Then, prognostic tools can be proficiently employed to reduce the number of required cost affecting maintenance actions and to plan the availability affecting ones into the already slated ground time windows, thus improving the availability of the aircraft and reducing its operating costs.

8 CONCLUSIONS

A satisfyingly effective FDI tool intended to be applied to an electromechanical actuator for flight control system has been implemented by the authors and tested in a simulated test bench in presence of different working conditions in terms of fault modes combinations of the monitored system.

The algorithm has shown an adequate robustness also in presence of a noisy input signal or the effects of failure modes not considered in the monitor model. In particular, the presence of an unknown failure mode leads to a partial misinterpretation of the system health status, but this condition is correctly recognized and reported coherently. On the other hand, the possibility of an unknown failure mode and a considered one cancelling each other effects, thus producing a missed fault identification, despite appearing as extremely unlikely, shall be further investigated. The future work on this FDI algorithm will include the extension of the models, taking into account a greater number of fault modes, and their validation on a physical test bench, also to ensure the required accuracy and resolution of the measured signals is matched by the currently available sensors. The implementation on a flying aircraft of an FDI algorithm similar to the one proposed (adapted to match the parameters of the particular actuator installed on-board), coupled with a statistical RUL prediction, could then reduce the maintenance related operating costs, reduce the downtime and increase both safety and availability of the system, integrating the traditional RAMS disciplines with the emerging PHM approach.

REFERENCES

Alamyal, M., Gadoue, S.M. & Zahawi, B. 2013. Detection of induction machine winding faults using genetic algorithm. *Diagnostics for Electric Machines, Power Electronics and Drives 9th IEEE Int.Symposium, Valencia, Spain*: 157–161.

Berri, P.C., Dalla Vedova, M.D.L. & Maggiore, P. 2016. A Smart Electromechanical Actuator Monitor for New Model-Based Prognostic Algorithms. *International Journal of Mechanics and Control (JoMaC)* 17(2): 59–66.

Berri, P.C., Dalla Vedova, M.D.L., Maggiore P. 2017. On-board electromechanical servomechanisms affected by progressive faults: proposal of a smart GA model-based prognostic approach. *Proc. of the 27th European Safety and Reliability Conference, Portoroz, Slovenia: 839–845*.

Belmonte, D., Dalla Vedova, M.D.L. & Maggiore, P. 2015. Electromechanical servomechanisms affected by motor static eccentricity: Proposal of fault evaluation algorithm based on spectral analysis techniques. *Safety and Reliability of Complex Engineered Systems – Proc. of the 25th European Safety and Reliability Conf. ESREL 2015*: 2365–2372.

Borello, L., Dalla Vedova, M.D.L., Jacazio, G. & Sorli, M. 2009a. A Prognostic Model for Electrohydraulic Servovalves. *Annual Conference of the Prognostics and Health Management Society, San Diego, CA*.

Borello, L., Villero, G. & Dalla Vedova, M.D.L. 2009b. New asymmetry monitoring techniques: effects on attitude control. *Aerospace Science and Technology* 13(8):475–487.

Borello, L. & Dalla Vedova, M.D.L. 2012. A dry friction model and robust computational algorithm for reversible or irreversible motion transmission. *International Journal of Mechanics and Control* 13(2): 37–48.

Borello, L., & Dalla Vedova, M.D.L. 2014. Flaps Failure and Aircraft Controllability: Developments in Asymmetry Monitoring Techniques. *Journal of Mechanical Science and Technology (JMST)* 28(11): 4593–4603.

Byington, C.S., Watson, W., Edwards, D. & Stoelting, P. 2004. A Model-Based Approach to Prognostics and Health Management for Flight Control Actuators. *IEEE Aerospace Conference Proceedings, USA.*

Çunkas, M., & Aydoğdu, O. 2010. Realization of Fuzzy Logic Controlled Brushless DC Motor Drives using Matlab /Simulink. *Mathematical and Computational Applications* 15(02): 218–229.

Dalla Vedova, M.D.L., Maggiore, P., & Pace, L. 2014. Proposal of Prognostic Parametric Method Applied to an Electrohydraulic Servomechanism Affected by Multiple Failures. *WSEAS Trans. on Environment and Development* 10: 478–490.

Dalla Vedova, M.D.L., Maggiore, P., & Pace, L. 2015a. A New Prognostic Method Based on Simulated Annealing Algorithm to Deal with the Effects of Dry Friction on Electromechanical Actuators. *International Journal of Mechanics* 9: 236–245.

Dalla Vedova, M.D.L., Maggiore, P., Pace, L. & Desando, A. 2015b. Evaluation of the correlation coefficient as a prognostic indicator for electromechanical servomechanism failures. *International Journal of Prognostics and Health Management* 6(1).

Dalla Vedova, M.D.L., De Fano, D. & Maggiore, P. 2016a. Neural Network Design for Incipient Failure Detection on Aircraft EM Actuator. *International Journal of Mechanics and Control (JoMaC)* 17(1): 77–83.

Dalla Vedova, M.D.L, Germanà, A. & Maggiore, P. 2016b. Proposal of a new simulated annealing model-based fault identification technique applied to flight control EM actuators. *Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016*: 313–321.

Dersin, P., Alessi, A., Lamoureux, B., Brahimi, M. & Fink, O. 2017. Prognostics and health management in railways. *Proc. of the 27th European Safety and Reliability Conference, Portoroz, Slovenia: 889–895*.

Halvaei Niasar, A., Moghbelli, H. & Vahedi, A. 2009. Modelling, Simulation and Implementation of Four-Switch Brushless DC Motor Drive Based On Switching Functions.*IEEE EUROCON 2009, 18–23 May, St.-Petersburg, Russia*.

Hamdani, S., Touhami, O., Ibtiouen, R. & Fadel, M. 2011. Neural network technique for induction motor rotor faults classification-dynamic eccentricity and broken bar faults. *IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics & Drives*: 626–631.

Haskew, T.A., Schinstock, D.E. & Waldrep, E. 1999. Two-Phase On' Drive Operation in a Permanent Magnet Synchronous Machine Electromechanical Actuator. *IEEE Transactions on Energy Conversion* 14.

Howse, M. 2003. All-electric aircraft. *Power Engineer*, 17(4).

Lee, B.K. & Ehsani, M. 2003. Advanced Simulation Model for Brushless DC Motor Drives. *Electric Power Components and Systems*, 31(9): 841–868.

Mamis, M.S., Arkan, M. & Keles, C. 2013. Transmission lines fault location using transient signal spectrum. *Int. J. Electr. Power Energy Syst.* 53: 714–718.

Markovsky, I. & Van Huffel, S. 2007. Overview of total least-squares methods. *Signal Processing* 87 (10): 2283–2302.

Quigley, R.E.J. 1993. More electric aircraft. Proc. of Eighth Annual IEEE Applied Power Electronics Conference - APEC '93, San Diego, CA: 906–911.

Raie, A., & Rashtchi, V. 2002. Using a genetic algorithm for detection and magnitude determination of turn faults in an induction motor. *Electrical Engineering* 84(5): 275–279.

Refaat, S.S., Abu-Rub, H., Saad, M.S., Aboul-Zahab, E.M. & Iqbal, A. 2013. ANN-based for detection, diagnosis the bearing fault for three phase induction motors using current signal. *IEEE International Conference on Industrial Technology (ICIT)*: 253–258.

Su, H. & Chong, K.T. 2007. Induction machine condition monitoring using neural network modelling. *IEEE Transactions on Industrial Electronics,* 54(1): 241–249.

Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A. & Wu, B. 2006. *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Wiley.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A method for wind speed generation

J. Ma, M. Fouladirad & A. Grall
Institut Charles Delaunay, LM2S, Université de Technologie de Troyes, CNRS, Troyes, France

ABSTRACT: This paper proposes a flexible continuous wind speed model based on first order Markov chain and Stochastic Differential Equations (SDE). This model permits to generate wind speed sequence with high frequency with statistical properties similar to an observed wind speed in a given geographical site. This model can be merged with the deterioration health indicators of wind turbine to make prognosis and plan maintenance actions.

1 INTRODUCTION

Since the Paris Agreement has been signed among the 195 members of the United Nations Framework Convention on Climate Change (UNFCCC), more and more countries have announced their plans to reduce the emission of CO_2 and to develop alternative clean energies like wind energy, solar energy, tidal energy, energy of natural gas, etc. Consequently, in the past decade, the world widely exploitation and technical development of wind energy have increased. In Europe, the capacity installed in 2016 reached 13489.9 MW and the estimated electricity production from wind power in the European Union (EU) in 2016 is 302.7 TWh (windeurope 2017a). By 2030, wind energy could cover 29.6% of EU's electricity demand (windeurope 2017b). However, because of the inbred randomness of wind speed, the wind power industry faces enormous challenges in practice and the development of wind power industry is very laborious.

To obtain a good prognostic of the wind turbine lifetime as well as to carry out efficient maintenance policies, it is necessary to study the impact of the wind speed on the wind turbine deterioration. It is important to rely on an appropriate wind speed model which permits to simulate the wind sequence related to a required geographical region with the same features as the real wind sequence and put in our disposable high frequency data of wind speed sequence.

The literatures on wind speed can be classified into three groups.

- Wind characteristics study on a specific site
- Wind speed generation
- Wind speed prognosis

The wind characteristics study is more quantitative and focuses on statistics properties of the wind speed, such as its average, maximum, minimum, standard deviation and its statistical distribution (usually consider the Weibull distribution function) ((Shu et al. 2016, Barthelmie et al. 2005, Celik 2004)). The daily patterns in wind energy production has been investigated (Scholz et al. (2014),Lennard (2014)), and studies the modelling for extreme events, refer to Baseer et al. (2017). Wind speed generation and wind speed prognosis are two different subjects. However, they have similarities regarding the modelling aspect. A great effort is deployed for wind speed modelling. To achieve the accuracy, models using environmental parameters such as temperature, pressure and humidity are proposed. Generally, these models require the computation of fluid dynamics in order to simulate the environmental conditions on different grids (Sezer-Uzol and Long 2006, Landberg et al. 2003). Since these latter are computationally burdensome they are not capable of generating a large data set and therefore not very suitable for applications requiring a good knowledge of the wind sequence in a short time.

Using probability distributions such as Weibull distribution to generate wind speed data which permits to reproduce wind speed data with the same statistical features as the wind data in our disposal. Regrettably, wind speed samples are measured at hourly time scale and data generation based on this type of data can only reproduce the macroscopic features of the wind behaviour.

Discrete models with memory applied in time series analysis such as Autoregressive Moving Average (ARMA) models (Poggi et al. 2003), Autoregressive Integrated Moving Average (ARIMA) models (Kavasseri and Seetharaman 2009) and Generalized autoregressive Conditional Heteroskedasticity (GARCH) models (Liu et al. 2011) are commonly used to reproduce wind speed data for a particular location depending on available

historical measured data. Although these models have efficient computational capacity for generating wind speed data, their linearity discard the possibility of nonlinearity in time series (Wanga et al. 2018). Since a simple model ignores the complexity of the wind behaviour and an elaborate model is not easy to deal with, some numerical methods based on computer capacity have been proposed to make predictions in wind behaviour only according to historical data.

Recently, wind speed prognosis based on machine learning and artificial neural networks (ANNs) has become very popular, see (Bilgili et al. 2007, Ji et al. 2007). Generally, the method is to generate wind speed by training the historical information about wind speed data. Easily trapped in local optimisation is a key problem related to ANNs. Moreover, without a model it is difficult to make instantaneous predictions and to control the behaviour of the engineering systems impacted by wind speed evolution.

Lately, in order to reduce the computational complexity and still propose a model for the wind evolution memoryless or short-memory stochastic models are getting more attention. In this framework, Markov processes, lévy processes or diffusion processes are proposed. For instance, based on Brownian motion and Langevin equations, the wind evolution is described by Wilson and Sawford (1996). More accessible, Markov chains have the capability to model wind speed data. Despite the simplicity of first-order Markov chain, it can generate wind speed time series with very similar statistical characters as the observed data (Nfaoui et al. (2004), Ettoumi et al. (2003), Yang et al. (2011)). However, Markov chains are not able to capture wind speed characteristics at high frequency. Research done by Brokish and Kirtley (2009) shows the limits of Markov chains for simulation data with time step smaller than fifteen minutes. To generalise, second or third order Markov chains and semi-Markov chains are used for wind speed modelling in order to improve the accuracy (D'Amico et al. 2013). Diffusion processes or stochastic differential equations are very flexible with short memory and as mentioned in (Zárate-Miñano et al. 2013), they seem to be suitable candidate wind speed modelling. Since SDE models are continuous, technically, they can be used to simulate wind speed at any time scales. Considering the diurnal and seasonal influences, the wind speed generation length of SDE models is limited to a few hours. SDE models are especially suitable for modeling wind speed turbulences.

An epitome could be summarized from the literatures is that the existent wind speed models focus on either short term modeling or long term modeling with large time scale. Hence, if these models are required for the analysis of wind turbine performances and operation related to wind speed their utility is limited. Short term models limit the analysis to a current short period; long term models can't provide sufficient and qualitatively useful data for analysis.

This problem is especially prominent in reliability analysis and failure prognosis and output power estimation of wind turbine. It has been highlighted that wind speed has a significant impact on degradation of wind turbine components. Particularly, the blade-pitch system has high failure rate among the components. Moreover, as it is indicated in Barthelmie et al. (2005), the wind turbulence intensity (variance divided by mean wind speed) has key influence on wind turbine lifetime. For instance, the fast variation of wind speed leads to frequent and sudden actions of the blade-pitch system. It has been demonstrated that turbulence intensity affects turbines power production ((Elliott and Cadogan 1990, Frandsen et al. 2000, Kaiser et al. 2007, Gottschall et al. 2006, Gottschall and Peinke 2007, Lubitz 2014)). Furthermore, from the point of view of prognosis and remaining useful lifetime (RUL) estimation, hourly average wind speed data is not satisfiable because it reflects the trend of wind speed by losing details, namely, the turbulence of wind speed. In addition, degradation occurring to wind turbine components is a long term process, and the RUL of wind turbine components are dynamically changing according to the working conditions of wind turbine.

Presently, wind speed at different time scales are applied to different problems, such as long term wind speed data are used to estimate wind energy for a future wind farm; diverse wind speed sequences are used to test a new control strategy; annual wind sequences with detailed information are indispensable to reliability research about the wind turbine, for example, the degradation of blade/yaw pitch system.

A method taking into account a continuous wind speed generation during a long period can fill the existing gap in the literature. This fact encourages us to propose a wind speed model that satisfies the following requirements.

- It can reflect the trend of wind in a long term period.
- It can contain turbulence information at small time scale.
- It can be accord with real wind speed data's probability distribution
- It can be generated quickly

In other words, this paper aims to propose a mathematical model which is able to generate satisfiable wind speed series for wind turbine components RUL study and accurate output power estimation.

The main contributions of this paper could be summarized as follows.

- Proposition of a new wind speed model considering Markov chains embedded with SED model;
- Consideration of different SDE models to be able to generate different wind profiles.

The remainder of this paper is organized as follows. After the model description a brief introduction on Markov chains and SDE models is given in section 2. In the same section, model parameters estimation methods are mentioned. An illustrative example is given in section 3. Conclusions in section 4.

2 MODEL DESCRIPTION

In wind industry, the Reynolds decomposition need to be considered for analyzing turbulence effects. Hence, the wind speed time series $U(t)$ is decomposed into its mean value $\bar{U}(t)$ and the fluctuation $u(t)$..

$$U(t) = \bar{U}(t) + u(t) \tag{1}$$

This gives us an idea to model mean wind speed and wind speed fluctuation separately. Refer to Calif (2012), $\bar{U}(t)$ can be considered as a low-pass filter corresponding to the hourly, daily, monthly, seasonal or annual effects; turbulence $u(t)$, which has a zero mean value and can be seen as a high-pass filter corresponding to turbulent effects.

Depending on research emphasis, wind speed can be modeled in different ways at different time scales. In wind energy industry, choosing continuously wind speed is more appropriate when we study the operation of wind turbine, as studies show that wind turbulence intensity has enormous effect on energy production and detected failures about wind turbine; hourly average wind speed is a good choice when we want to estimate the wind resource of a site; 10 minutes's average wind speed is noted down by SCADA system.

In this paper, we propose a Markov chain model embedded with SDE. Due to its properties, this model is very flexible. Markov chains is used to model macroscopical wind speed trend. It represents hourly average wind speed or different wind speed classes relating to wind turbulence intensity. The embedded SDEs are mainly used to model a continuous wind speed depending on the environmental states which are set by Markov chain. Different setting parameters and different applications of SDE give different continuous wind speed models. Other SDE models also can be considered for simulating different conditions such as gust or extreme weather. Hence this model is easily adapted according to the location of interest.

Here we would like to give two illustrative examples.

- One day's (24 hours) wind speed generation
  Each state in Markov chain represents hourly average wind speed and drives the choice of related SDE. In this case, 24 states are randomly generated to represent the average wind speed time series. In each state, continuous wind sequences are generated from specific SDE at small time scale like secondly data.
- Different wind condition generation
  Refer to IEC standards Commission et al. (2005), wind conditions contains normal, extreme and gust which could be generated by Markov chain. Inside Markov chain, SDEs are used to simulate wind speed associated with the certain wind condition.

2.1 *Markov chain*

A Markov chain process is uniquely defined by its state space, transition matrix and initial distribution. Markov chains wind speed modelling consists of two main steps: state classification and transition matrix estimation. The state classification is arbitrary depending on the purpose. In this paper, the range of wind speed $[U_{min}, U_{max}]$ is divided into serval intervals, each interval's index is identified by its central point, changing the event "the wind speed value is in the range [a, b]" to the event "the wind speed is $\frac{1}{2}(b + a)$". We introduce the notation for the wind speed states as $S = \{s_1, s_2, \cdots, s_N\}$; and make $\{X_t\}_{t\in\mathbb{N}}$ represent the wind speed time series. Hence, the event "$X_0 = s_5$" means at time t = 0, the wind speed belongs to state s_5. Table 1 shows an example for state assignment.

The initial distribution is estimated by dividing the dataset into bins according to the states, then normalising the vector of the occurrences in each bins. For example, to calculate $p_1^0 = \mathbb{P}\{X_0 = s_1\}$, which is the probability that the first element of the wind time series is in state s_1, we need to count the number of times that a value enters in state s_1 in the entire recorded time series and divide it by the total number of recorded values.

The empirical frequencies are the estimation of the transition probabilities

$$\begin{aligned}\mathbb{P}(X_{n+1} = s_j \mid X_n = s_i, X_{n-1} = s_{i_{n-1}}, \cdots, X_0 = s_{i_0}) \\ = \mathbb{P}(X_{n+1} = s_j \mid X_n = s_i) = p_{ij}\end{aligned} \tag{2}$$

where s_i, $s_j \in S$. The transition probability p_{ij} can be estimated by counting how many times a value from s_i goes to state s_j in the wind time series, normalised over the total number of occurrences of

Table 1. Example for Markov chain's state space.

State	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
Interval	[4, 8]	(8, 9.4]	(9.4, 10.2]	(10.2, 10.7]	(10.7, 11.2]	(11.2, 12]	(12, 12.8]	(12.8, 25)

values in state s_i. The transition matrix denoted π is filled with transition probabilities p_{ij}, $1 \le i, j \le n$.

2.2 *SDE*

After studying 1 million of wind speed fluctuation distributions, Calif (2012) concludes that it is more common to have three classes of wind speed fluctuation distribution. The distribution classes are as follows. The 90% wind speed turbulence follows a kind of symmetrical mono-modal PDF which is well fitted to a Gaussian PDF (class 1 shown in Figure 1). The 9% wind speed turbulence follows a kind of dissymmetrical mono-modal PDF which can be described by Gram-Charler series (class 2 shown in Figure 1) and the rest 1% follows a kind of bimodal PDF which is fitted by mixture of Gaussian PDF (class 3 shown in Figure 1). Figure 1 illustrates the mean PDF of wind speed turbulence for each class. In this paper, we consider more about the wind with speed fluctuation of class 1 and class 2.

2.2.1 *SDE model selection*

In order to randomly switch between the two classes, one considers an inner Markov chain inside each inner state s_i. To avoid confusions between the wind speed Markov chain and this SDE switching Markov chain, the latter one is designate as a SDE-Selection model (SSM) in this paper. A transition probability matrix of SSM is directly assigned as follow.

class	$Class_1$	$Class_2$
$Class_1$	0.9	0.1
$Class_2$	0.9	0.1

Since the majority of turbulence seems to follow a gaussian distribution and some can be fitted to mono-modal distributions, it seems natural to propose a stochastic process where increments are gaussian or functional of a gaussian distribution. Potential candidates are diffusion processes which are derived from stochastic differential equations. The general form of one-dimensional SDE equation is as follows:

$$dx(t) = a(x(t),t)dt + b(x(t),t)dW(t), t \in [0,T] \quad (3)$$

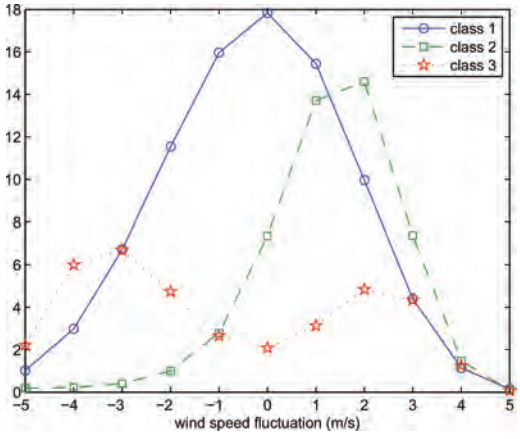


Figure 1. Wind speed fluctuation (Source: (Calif, R.2012)).

where $a(x(t), t)$ and $b(x(t), t)$ are the drift term and diffusion term, respectively. $W(t)$ is a standard Wiener process described as follows:

- $W(0) = 0$, with probability 1.
- For $0 \le t_i < t_{i+1} \le T$, the increments $\Delta W_i = W(t_{i+1}) - W(t_i)$ is a Gaussian distribution with zero mean and $\sigma = t_{i+1} - t_i$, namely, $\Delta W_i \sim \mathcal{N}(0, t_{i+1} - t_i)$
- for $0 \le t_i < t_{i+1} < t_{i+2} \le T$, the non-overlapping increments $\Delta W_i = W(t_{i+1}) - W(t_i)$ and $\Delta W_{i+1} = W(t_{i+2}) - W(t_{i+1})$ are independent.

Hence, a standard Wiener process describes a continuous Gaussian process whose increments are of unbounded variation. Since there is a large variety of functions $a(x(t), t)$ and $b(x(t), t)$, the diffusion process can model a large range of phenomenon. In this paper, we focus on a special case of SDE initially proposed by Vasicek (1977).

2.2.2 *Ornstein-Uhlenbeck process*

Figure 2 lists 3 histograms of wind speed turbulence. This group wind data is downloaded from the site *www.winddata.com*, recorded the wind speed during one hour at the site San Gorgonio, USA. It can be observed that the wind speed turbulence follow a kind of symmetrical mono-modal PDF and the mean value is almost around zero.

Combining with the research results of Calif (2012) (shown in Figure 1), Ornstein-Uhlenbeck

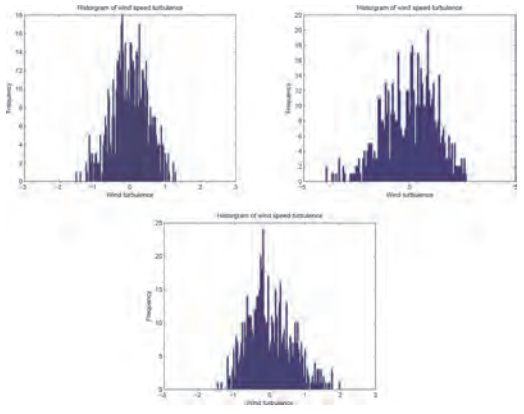


Figure 2. Histogram of wind speed turbulence (data downloaded from *www.winddata.com*).

processes appear to be suitable candidates for wind speed fluctuations modelling. In other words, the aim of SDE application here is to model a continuous wind speed during a short period, such as several seconds, 10 minutes or 1 hour.

Form A

Langevin's equation is considered to describe the first process that is chosen to depcit the increments of wind speed, namely,

$$\begin{aligned} dx(t) &= ax(t)dt + bdW(t), t \in [0,T] \\ x(0) &= x_0 \end{aligned} \tag{4}$$

where, $x(t)$ is the wind speed turbulence at time t, $W(t)$ is a standard Brownian motion, a and b are constants. $x(t)$ is a stationary autocorrelated Gaussian diffusion process.

Fouque et al. (2000) provides a maximum likelihood estimation method. The logarithm likelihood function is defined as

$$\begin{aligned} L_{a,b}(x_1,\ldots,x_n) &= -\frac{n}{2}\left(\ln(b^2) + \ln(v(a))\right) \\ &+ \frac{1}{2b^2 v(a)}\sum_{i=0}^{n-1}\left(x_{t_i+1} - \exp(a\Delta)x_{t_i}\right)^2 \end{aligned} \tag{5}$$

with $v(a) = \dfrac{\exp(2a\Delta - 1)}{2a}$. Therefore, we can get the estimated parameters.

$$\hat{a} = \frac{1}{\Delta}\log\left(\frac{\sum_{i=1}^{n} x - t_{i-1}x_{t_i}}{\sum_{i=1}^{n} a_{t_i-1}^2}\right) \tag{6}$$

$$\hat{b}^2 = \frac{1}{nv(\hat{a})}\sum_{i=1}^{n}(x_{t_i} - \exp(\hat{a}\Delta)x_{t_i-1})^2 \tag{7}$$

Form B

A classical mean-reverting Ornstein-Uhlenbeck process is chosen as the second one for wind speed simulation. It has the following form:

$$\begin{aligned} dx(t) &= -\left(x(t) - \mu\right)dt + \sigma dW(t) \\ x(0) &= x_0 \end{aligned} \tag{8}$$

where $x(t)$ is wind speed at time t, μ, σ are parameters and W_t is the standard Brownian motion. Make $\Delta t = t_i - t_{i-1}$, according to Deng et al. (2016) the transition density function is

$$\begin{aligned} \mathbb{P}(x_{t_i}, t_i \mid x_{t_{i-1},t_{i-1}}; \mu, \sigma) &= \frac{e^{\Delta t}}{\sqrt{\pi \cdot \sigma^2\left(e^{2\Delta t} - 1\right)}} \\ &\times \exp\left(-\frac{\left(x_{t_i} \cdot e^{\Delta t} + \mu\left(1 - e^{\Delta t}\right) - x_{t_{i-1}}\right)^2}{\sigma^2\left(e^{2\Delta t} - 1\right)}\right) \end{aligned} \tag{9}$$

Hence, the log-likelihood function is as follows:

$$\log L(\mu,\sigma) = \prod_{i=1}^{N}\mathbb{P}(x_{t_i}, t_i \mid x_{t_{i-1},t_{i-1}}; \mu, \sigma) \tag{10}$$

The parameters μ and σ are estimated by maximizing (10).

3 APPLICATION

3.1 *Experimental dataset*

Long term (as long as one year) hourly average wind speed data can be easily accessed at *www.ncdc.noaa.gov/cdo-web*, and *www.winddata.com* provides free time series of wind characteristics measured under different conditions during 10 minutes or 1 hour. The real data used in this paper is downloaded from the two websites.

3.2 *Wind speed generation*

It exists two difficulties when we use real wind speed data to estimate the transition probability matrix of Markov chain model:

- A Markov chain with a great deal of states could be constructed resulting in a huge transition matrix that might cause additional difficulties in estimation.
- The number of elements in certain states could be much smaller than others, resulting in a lot of small probabilities near 0 in the transition probability matrix.

In order to avoid the former phenomena, the number of states should be determined relatively smaller. The number of states could be determined

by the user or some criterions. Tang et al. (2015) recommend to determine the interval with the empirical quantiles. Suppose that the wind speed time series is stationary and ergodic. Hence, the empirical cumulative distribution function is a consistent estimator of the cumulative distribution of the invariant measure of this time series. The boundaries are taken to be $\hat{F}_N^{-1}(j/k)$, $j = 1, 2, \ldots, k$, where k is the number of state and $\hat{F}_N$ is the empirical cumulative distribution function.

The principal steps for wind speed data generation (shown in Figure 3) are as follows,

Step 1 Chose appropriate wind speed data for Markov chain and diffusion process model estimation separately

Step 2
- Determine the states of Markov chain
- Estimate transition probability matrix of Markov chain by using real wind speed data
- Estimate parameters of SDE with real wind speed data

Step 3 Generate hourly average wind speed

Step 4 Select SDE model and generate detail wind speed data inside each state

3.2.1 *Generation of wind speed in macroscopic scale using Markov chain model*

A sequence of hourly average wind speed during one year is used to estimate the transition probability matrix of the Markov chain model. Due to data acquisition issues, the available number of the data is 6180 instead of 8760 (24 hours/day × 365 days). The number of state is determined as 8, recall the states shown in Table 1, we can obtain the transition probability matrix of the Markov chain.

3.2.2 *Generation of wind speed at secondly time scale using SDE*

As Markov chain's states represent the hourly average wind speed, the aim of using SDE is to generate secondly wind speed data during a short period. Taking account the common data collection frequency of SCADA—10 minutes, a state contains 6 groups SDE models which represent the wind speed at secondly scale during 10 minutes, depending on the hourly average wind speed set by Markov chain. We class the real wind speed sequences according to their average values associated with the states determined in section 2.1.

An example of wind speed generation by SDE

To illustrate the performance of SDE wind speed model, the parameters with respect to the state s_3 are chosen to build the two Ornstein-Uhlenbeck process, with form A expressed by equation (11) and form B expressed by equation (12). For $x(0) = 0$

$$dx(t) = -0.0314x(t)dt + 0.2517dW(t), \quad (11)$$

$$dx(t) = -\left(x(t) - 10.0245\right)dt + 0.6459dW(t), \quad (12)$$

In Figure 4, the red line shows the probability density of a real wind speed sequence, the blue lines show the probability density of 500 samples simulated data generated by Equation (12).

SDE model has the ability to generate continuous wind speed.

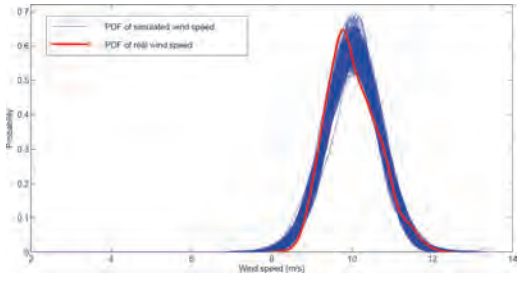


Figure 4. Probability density of real data and simulated data.

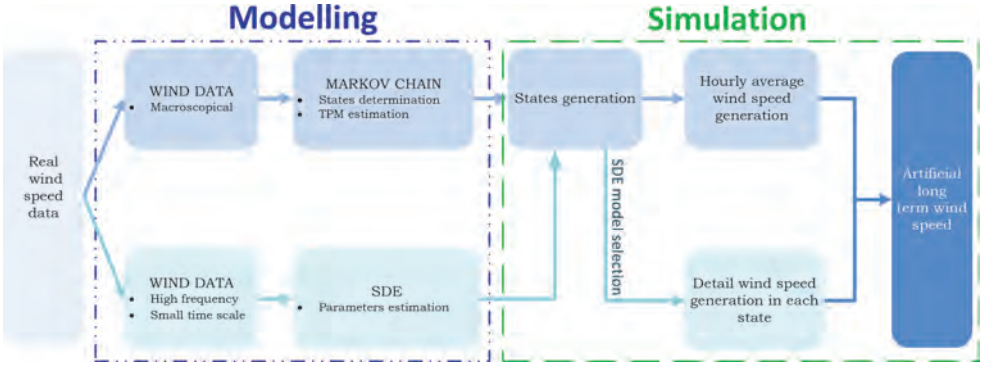


Figure 3. Steps for wind speed generation.

4 CONCLUSION

In this paper, a continuous wind speed generation model based on Markov chain and stochastic differential equation is proposed. This model is flexible enough to generate wind speed at different time scale without time length limitation. Having low computational requirement is another advantage of the model. Two forms of SDE are studied in this paper, other forms of SDE could be applied to this model according to the user's requirement. Particularly, this developed model is suitable to be merged with the deterioration model of wind turbine's key component, like blade-pitch system. Also, this model could be used in the analysis of wind turbine's power system such as dynamic behaviour of generator.

REFERENCES

Barthelmie, R., O.F. Hansen, K. Enevoldsen, J. Højstrup, S. Frandsen, S. Pryor, S. Larsen, M. Motta, & P. Sander-hoff (2005). Ten years of meteorological measurements for offshore wind farms. *Journal of Solar Energy Engineering 127*(2), 170–176.

Baseer, M., J. Meyer, S. Rehman, & M.M. Alam (2017). Wind power characteristics of seven data collection sites in jubail, saudi arabia using weibull parameters. *Renewable Energy 102*, 35–49.

Bilgili, M., B. Sahin, & A. Yasar (2007). Application of artificial neural networks for the wind speed prediction of target station using reference stations data. *Renewable Energy 32*(14), 2350–2360.

Brokish, K. & J. Kirtley (2009). Pitfalls of modeling wind power using markov chains. In *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*, pp. 1–6. IEEE.

Calif, R. (2012). Pdf models and synthetic model for the wind speed fluctuations based on the resolution of langevin equation. *Applied energy 99*, 173–182.

Celik, A.N. (2004). A statistical analysis of wind power density based on the weibull and rayleigh models at the southern region of turkey. *Renewable energy 29*(4), 593–604.

Commission, I.E. et al. (2005). Iec 61400-1: Wind turbines part 1: Design requirements. *International Electrotechnical Commission*.

D'Amico, G., F. Petroni, & F. Prattico (2013). First and second order semi-markov chains for wind speed modeling. *Physica A: Statistical Mechanics and its Applications 392*(5), 1194–1201.

Deng, Y., A. Barros, & A. Grall (2016). Degradation modeling based on a time-dependent ornstein-uhlenbeck process and residual useful lifetime estimation. *IEEE Transactions on Reliability 65*(1), 126–140.

Elliott, D.L. & J.B. Cadogan (1990). Effects of wind shear and turbulence on wind turbine power curves. Technical report, Pacific Northwest Lab., Richland, WA (USA).

Ettoumi, F.Y., H. Sauvageot, & A.-E.-H. Adane (2003). Statistical bivariate modelling of wind using first-order markov chain and weibull distribution. *Renewable energy 28*(11), 1787–1802.

Fouque, J.-P., G. Papanicolaou, & K.R. Sircar (2000). *Derivatives in financial markets with stochastic volatility*. Cambridge University Press.

Frandsen, S., I. Antoniou, J. Hansen, L. Kristensen, H.A. Madsen, B. Chaviaropoulos, D. Douvikas, J. Dahlberg, A. Derrick, P. Dunbabin, et al. (2000). Redefinition power curve formore accurate performance assessment of wind farms. *Wind Energy 3*(2), 81–111.

Gottschall, J., E. Anahua, S. Barth, & J. Peinke (2006). Stochastic modelling of wind speed power production correlations. *PAMM 6*(1), 665–666.

Gottschall, J. & J. Peinke (2007). Stochastic modelling of a wind turbine's power output with special respect to turbulent dynamics. In *Journal of Physics: Conference Series*, Volume 75, pp. 012045. IOP Publishing.

Ji, G.-R., P. Han, & Y.-J. Zhai (2007). Wind speed forecasting based on support vector machine with forecasting error estimation. In *Machine Learning and Cybernetics, 2007 International Conference on*, Volume 5, pp. 2735–2739. IEEE.

Kaiser, K., W. Langreder, H. Hohlen, & J. Højstrup (2007). Turbulence correction for power curves. *Wind Energy*, 159–162.

Kavasseri, R.G. & K. Seetharaman (2009). Day-ahead wind speed forecasting using f-arima models. *Renewable Energy 34*(5), 1388–1393.

Landberg, L., G. Giebel, H.A. Nielsen, T. Nielsen, & H. Madsen (2003). Short-term prediction—an overview. *Wind Energy 6*(3), 273–280.

Lennard, C. (2014). Simulating an extreme wind event in a topographically complex region. *Boundary-layer meteorology 153*(2), 237–250.

Liu, H., E. Erdem, & J. Shi (2011). Comprehensive evaluation of arma–garch (-m) approaches for modeling the mean and volatility of wind speed. *Applied Energy 88*(3), 724–732.

Lubitz, W.D. (2014). Impact of ambient turbulence on performance of a small wind turbine. *Renewable Energy 61*, 69–73.

Nfaoui, H., H. Essiarab, & A. Sayigh (2004). A stochastic markov chain model for simulating wind speed time series at tangiers, morocco. *Renewable Energy 29*(8), 1407–1418.

Poggi, P., M. Muselli, G. Notton, C. Cristofari, & A. Louche (2003). Forecasting and simulating wind speed in corsica by using an autoregressive model. *Energy conversion and management 44*(20), 3177–3196.

Scholz, T., V.V. Lopes, & A. Estanqueiro (2014). A cyclic time-dependent markov process to model daily patterns in wind turbine power production. *Energy 67*, 557–568.

Sezer-Uzol, N. & L.N. Long (2006). 3-d time-accurate cfd simulations of wind turbine rotor flow fields. *AIAA paper 394*, 2006.

Shu, Z., Q. Li, Y. He, & P. Chan (2016). Observations of offshore wind characteristics by doppler-lidar for wind energy applications. *Applied Energy 169*, 150–163.

Tang, J., A. Brouste, & K.L. Tsui (2015). Some improvements of wind speed markov chain modeling. *Renewable Energy 81*, 52–56.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of financial economics 5*(2), 177–188.

Wanga, J., T. Niua, H. Lub, Z. Guoc, W. Yanga, & P. Dua (2018). An analysis-forecast system for uncertainty modeling of wind speed: A case study of large-scale wind farms. *Applied Energy 211*(5), 492–512.

Wilson, J.D. & B.L. Sawford (1996). Review of lagrangian stochastic models for trajectories in the turbulent atmosphere. *Boundary-layer meteorology 78*(1), 191–210.

windeurope (2017a). Wind energy barometer 2017. Technical report, European Union. Accessed: 2017-11-21.

windeurope (2017b). Wind energy in europe and scenarios for 2030. Technical report, European Union. Accessed: 2017-11-21.

Yang, H., Y. Li, L. Lu, & R. Qi (2011). First order multivariate markov chain model for generating annual weather data for hong kong. *Energy and Buildings 43*(9), 2371–2377.

Zárate-Min˜ano, R., M. Anghel, & F. Milano (2013). Continuous wind speed models based on stochastic differential equations. *Applied Energy 104*, 42–49.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

The class of life time distributions with a mean residual life linear in time: Application to prognostics and health management

P. Dersin
RAM (Reliability-Availability-Maintainability) and PHM (Prognostics & Health Management), Alstom Digital Mobility, St-Ouen, France

ABSTRACT: Prognostics and Health Management (PHM) elicits increasing interest in view of its potential economic impact both through service-affecting failure avoidance and maintenance cost reduction. One of the key challenges in the PHM discipline is the estimation of Remaining Useful Life (RUL), i.e. the time remaining until failure in the absence of maintenance. The expectation of RUL, Mean Residual Life (MRL), has long been studied by reliability engineers, dating back to the nineteen sixties. There is a one-to-one relationship between MRL, reliability function and failure rate (also called hazard rate): each one of three functions uniquely determines the other two. Usually, MRL varies with time. We study here a special class of life-time distributions characterized by the fact that their MRL is a linear, non-increasing function of time, and highlight the central role of the time derivative of the MRL, called ageing rate, for PHM; and then generalize to piecewise-linear MRL.

1 INTRODUCTION

Since the pioneering work of (Barlow & Proschan 1965), the notion of MRL is well known, as well as its relation to reliability function and failure rate.

However, we have not seen in the literature any study of the class of life-time distributions which have a MRL that varies linearly with time, even though the literature on RUL is quite extensive; see e.g. (Si et al. 2011).

This class of distributions is important for two reasons:

- because a linear relationship is always a good approximation locally; more-general functions can be approximated by piecewise-linear ones.
- because this class contains important, well-known special cases: the exponential distribution, the uniform distribution, and the deterministic ('Dirac') distribution.

Explicit expressions for reliability function, MRL and failure rate are given here for this class. Cases of special interest are presented and the meaning of some relationships is discussed. A simple relation is derived between the coefficient of variation of the time to failure and the derivative of MRL with respect to time (which we call the rate of ageing). A confidence interval for the RUL is derived, in terms of this rate of ageing; the width of the confidence interval decreases as the rate of ageing grows from 0 (corresponding to the exponential distribution failure) to 1 (corresponding to the 'Dirac' deterministic case). Potential use of those results for PHM, and their generalization to piece-wise linear MRL, are discussed. In Section 2, a reminder of the key relationships between MRL, reliability function and failure rate is provided. In Section 3, expressions for those quantities are derived for the family of lifetime distribution with MRL linear with time, and special cases are identified. In Section 4, a confidence interval for the RUL is provided, in that family. In Section 5 an explicit relation is given between the rate of ageing and the coefficient of variation of the time to failure. Potential use in PHM and generalizations to the family of lifetime distributions with piecewise-linear MRL are discussed in Section 6. Finally, Section 7 presents the conclusions, followed by bibliographical references in Section 8.

2 REMINDER OF KEY RELATIONSHIPS

2.1 *Mean residual life, reliability function and failure rate*

If T denotes the time to failure of an item not subject to maintenance, the reliability function is defined by:

$$\begin{aligned} &R(t) = P[T > t], t > 0 \\ &R(0) = 1, \ \lim_{t \to \infty} R(t) = 0 \end{aligned} \tag{1}$$

Then the Mean Residual Life (MRL) at time t is the expectation of the remaining time to failure from time t under the condition that no failure has taken place until time t. The following relation

results directly from the definition. In what follows, the notation V(t) will be used for MRL(t).

$$V(t) = \frac{1}{R(t)} \int_t^\infty R(s)ds \tag{2}$$

From Equation 2 it is seen that, for t = 0, MRL(0) = MTTF. i.e., the a priori value for the mean residual life is the mean time to failure. The failure rate λ(t) (sometimes called hazard rate, especially in safety contexts), on the other hand, is given by:

$$\lambda(\mathrm{t}) = -\frac{R'(t)}{R(t)} \tag{3}$$

so that R(t) can be derived from it by integration (see e.g. Birolini 2017). Finally, under simple regularity conditions it can be shown (Swartz 1973) that the reliability function R(t) can be derived directly from the MRL function:

$$R(\mathrm{t}) = \frac{\mathrm{V}(0)}{V(t)} e^{-\int_0^t \frac{dx}{V(x)}} \tag{4}$$

Therefore, any one of those three functions determines the other two.

2.2 *Differential equation formulation*

There is also a simple relationship between the MRL and the failure rate which takes the form of a differential equation:

$$\frac{dV(t)}{dt} = \lambda V - 1 \tag{5}$$

Equation 5 can be derived immediately from Equation 2 and Equation 3 (Watson & Wells 1961). When Equation 5 is written in differential form:

$$dV(t) = -dt + \lambda V dt \tag{6}$$

its interpretation provides an interesting insight: usually

$$\frac{dV(t)}{dt} \le 0 \tag{7}$$

as the mean residual life does not increase with time.

Equivalently,

$$\lambda V \le 1 \tag{8}$$

In a time interval of length dt, the loss in mean residual life during the time span dt is given by :

$$dV(t) = -(1 - \lambda V)dt \tag{9}$$

Therefore it is usually less than the elapsed time dt.

The limiting case when: $\lambda V = 1$, corresponding to a distribution for which V(t) = MTTF is constant with time, which is the exponential distribution, also characterized by the property that its failure rate λ is constant in time.

Moreover, Equation 5 then yields:

$$\lambda(t) = \frac{1}{MRL(0)} = \frac{1}{MTTF} \tag{10}$$

a relation valid only in the exponential distribution case.

This property is the 'no aging' property of the exponential distribution in the reliability context.

More generally, in the next section, will be studied the class of time-to-failure distributions for which the MRL is a linear function of time, i.e.

$$\frac{dV(t)}{dt} = -k \tag{11}$$

where the constant k will be called the ageing rate. It is assumed that $0 \le k \le 1$ so as to ensure, according to Equation 11, that

$$-1 \le \frac{dV(t)}{dt} \le 0 \tag{12}$$

i.e. the mean residual life either is constant or decreases with time but no faster than time (in a time interval ΔT, the item does not lose more than an amount ΔT of its remaining life). Note that k is dimensionless.

The case k = 0 corresponds to the limiting case of the exponential distribution.

3 THE FAMILY OF LIFE DISTRIBUTIONS WITH MRL LINEAR IN TIME

3.1 *General case*

We study the family of time-to-failure distributions characterized by the property that their mean residual life MRL(t) is a linear non-increasing function of time or, equivalently, it satisfies Equation 11.

For convenience, from now on, the MTTF will be denoted by the symbol μ. Thus:

$$V(0) = MTTF = \mu \tag{13}$$

From Equation 11 and Equation 13, we obtain

$$V(t) = \mu - kt \tag{14}$$

The range of the time variable t is the interval $[0, \frac{\mu}{k}]$, so that

$$0 \le V(t) \le \mu \tag{15}$$

(In the limiting case of the exponential distribution, k is allowed to vanish asymptotically, and the range of the time variable becomes the nonnegative real half line).

From Equation 5 and Equation 14, there follows that the failure rate of such a distribution is given by:

$$\lambda(k,t) = \frac{1-k}{\mu - kt},\ for\ 0 \le t < \frac{\mu}{k} \tag{16}$$

From Equation 3 the reliability function, denoted from now on D(k,t), can be derived as

$$D(k,t) = e^{-\int_0^t \frac{1-k}{\mu - ks} ds} \tag{17}$$

which yields:

$$D(k,t) = \left(\frac{\mu}{\mu - kt}\right)^{1-\frac{1}{k}},\ for\ 0 \le t < \frac{\mu}{k} \tag{18}$$

And D(k,t) = 0 for $t > \mu/k$.

Thus Equations 14, 16, 18 characterize the one-parameter family of probability distributions, with parameter k ranging over the interval [0,1].

3.2 *Special cases*

3.2.1 *Exponential distribution*

If the parameter k is allowed to go to 0, the range of the variable t becomes (0, ∞).

The following limit is obtained from Equation 16:

$$\lim_{k \to 0} \lambda(k,t) = \frac{1}{\mu} \tag{19}$$

It can also be verified from Equation 18 that:

$$\lim_{k \to 0} D(k,t) = e^{-\frac{t}{\mu}} \tag{20}$$

Those properties characterize the exponential distribution.

As seen previously and confirmed from Equation 14, the mean residual life is then constant in time: $V(t) = \mu = MTTF$.

3.2.2 *Dirac distribution*

At the other extreme, if k = 1, the evolution of the mean residual life over time is characterized by:

$$V(t) = \mu - t,\ for\quad 0 \le t \le \mu \tag{21}$$

which means that, over any time interval Δt, the loss of mean residual life is precisely equal to Δt.

Equation 16 shows that the failure rate remains equal to zero as long as t<μ and jumps to infinity as t approaches the life limit μ. This is best seen by first taking k = 1–ε with ε close to 0 and letting t go to μ/k.

Likewise the reliability function has a discontinuity at t = μ, where it jumps from 1 to 0, which corresponds to a deterministic lifetime equal to μ. It can also be verified that the right-hand side of Equation 6 is always equal to –1, meaning that, over any time interval Δt, the loss of mean residual life is precisely equal to Δt.

3.2.3 *Uniform distribution*

The intermediate case $k = \frac{1}{2}$ defines the uniform distribution, over the range (0, 2μ).

Indeed, from Equation 18, the corresponding reliability function is given by:

$$D\left(\frac{1}{2}, t\right) = 1 - \frac{t}{2\mu},\quad 0 \le t \le 2\mu \tag{22}$$

The mean residual life (Eq. 14) is given by

$$V(t) = \mu - \frac{t}{2},\quad 0 \le t \le 2\mu \tag{23}$$

which states that, over a time interval Δt, the loss of mean residual life is equal to $\frac{\Delta t}{2}$.

The failure rate is given, as a function of time, by:

$$\lambda\left(\frac{1}{2}, t\right) = \frac{1}{2\mu - t} \tag{24}$$

Table 1. Family of life distributions with MRL linear in time.

Value of k	Range of t	D(k;t)	V(t)	λ(t)
$0 < k \le 1$	$\left[0, \frac{\mu}{k}\right]$	$\left(\frac{\mu}{\mu - kt}\right)^{1-\frac{1}{k}}$	$\mu - kt$	$\frac{1-k}{\mu - kt}$
0 (exp.)	[0, ∞]	$e^{-\frac{t}{\mu}}$	μ	$\frac{1}{\mu}$
$\frac{1}{2}$ (uniform)	[0, 2μ]	$1 - \frac{t}{2\mu}$	$\mu - \frac{t}{2}$	$\frac{1}{2\mu - t}$
1 (Dirac)	[0, μ]	1 for t < μ	μ–t	0 for t < μ

3.2.4 *Synopsis*

The results just derived are summarized in the table below.

4 CONFIDENCE INTERVAL FOR RUL

It is now possible to obtain the residual life-time distribution, i.e. the probability distribution of the Remaining Useful Life (RUL), or time to the next failure, under the condition that no failure has taken place up to time t, for the class of lifetime distributions under study. Note that the concept of 'failure' is understood in a broad sense and can also apply to pseudo-failures, such as the crossing of a threshold by a health indicator.

First let us rewrite Equation 18 slightly differently:

$$D(k,t)=\left(1-\frac{kt}{\mu}\right)^{\frac{1}{k}-1} \tag{25}$$

The probability distribution of the remaining useful life (RUL) at time t can be characterized by the conditional survival function. Let us denote by $D_t(s)$ the probability that the item survives at least for a duration s given that it has lived (without failure) up to time t. Then:

$$D_t(s)=\frac{D(k;t+s)}{D(k;t)} \tag{26}$$

Or, according to Equation 25),

$$D_t(s)=\left(\frac{1-\frac{k(t+s)}{\mu}}{1-\frac{kt}{\mu}}\right)^{\frac{1}{k}-1} \tag{27}$$

Or

$$D_t(s)=\left(1-\frac{ks}{\mu-kt}\right)^{\frac{1}{k}-1} \tag{28}$$

for $0<k\le 1$; $0\le t<\frac{\mu}{k}$

Let us now determine the confidence interval of the RUL at time t, with confidence level 1-α.

To that end, define s^+ and s^- by:

$$D_t(s^+)=\frac{\alpha}{2} \tag{29}$$

$$D_t(s^-)=1-\frac{\alpha}{2} \tag{30}$$

Then, if T denotes the random variable "survival time to failure after time t", the conditional probability of survival for a time between s^- and s^+ is:

$$P(t+s^-<T<t+s^+/T>t)=1-\alpha \tag{31}$$

i.e., the interval $(t+s^-,t+s^+)$ is the confidence interval with confidence level 1–α.

In other words, $t+s^-<\mathrm{RUL}<t+s^+$ with probability 1–α. Equations 28–30 determine the confidence bounds. Consequently, from Equation 28–29 one obtains:

$$s^+=\left(\frac{\mu}{k}-t\right)\left(1-\left(\frac{\alpha}{2}\right)^{\frac{k}{1-k}}\right) \tag{32}$$

Similarly, from (28) and (30), one obtains:

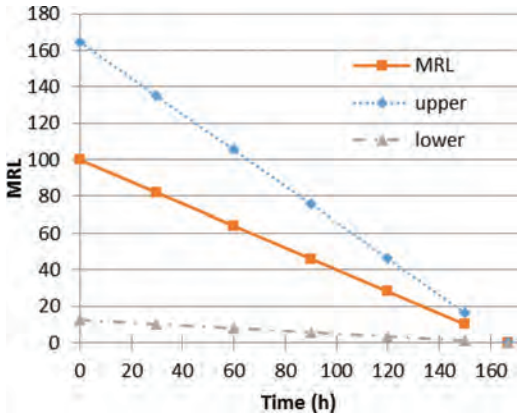


Figure 1a. Confidence interval for RUL for k = 0.6 (MTTF = 100 h).

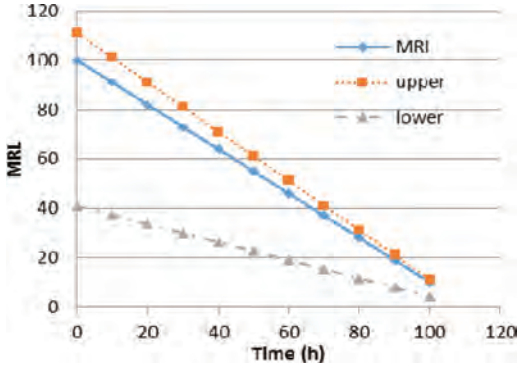


Figure 1b. Confidence interval for RUL for k = 0.9 (MTTF = 100 h).

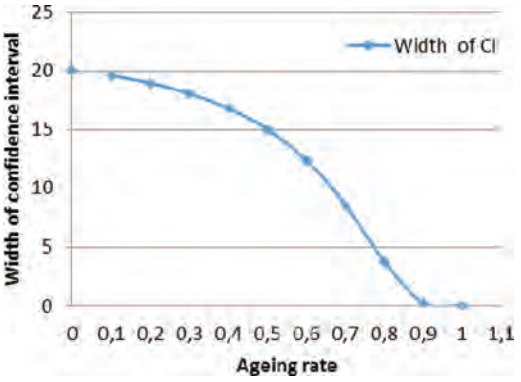


Figure 2. Width of confidence interval versus ageing rate k. (CI calculated at t = 50 h, for MTTF = 100 h).

$$s^- = \left(\frac{\mu}{k} - t\right)\left(1 - \left(1 - \frac{\alpha}{2}\right)^{\frac{k}{1-k}}\right) \tag{33}$$

Confidence intervals are illustrated in Figure 1: on (Fig. 1a) for k = 0.6 and on (Fig. 1b), for k = 0.9, where the confidence interval is narrower at any time.

For k = 1 (deterministic case), it is verified that the confidence interval reduces to a point (no uncertainty) and, for k = 0, the known result for the exponential distribution is found:

$$s^+ = -\ln\left(\frac{\alpha}{2}\right) \tag{34}$$

$$s^- = -\ln\left(1 - \frac{\alpha}{2}\right) \tag{35}$$

Indeed the width ($s^+ - s^-$) of the confidence interval decreases as k grows from 0 to 1, as illustrated in (Fig. 2).

Thus, the faster the asset ages, the narrower the confidence interval for the RUL is. This property is intuitive. For instance, in the limit of deterministic ageing (k = 1), the RUL is known with certainty and the failure rate jumps from 0 to infinity at t = RUL. In the other extreme (k = 0, exponential distribution), there is no ageing and the confidence interval is the widest and stays constant with time (Eq. 34–35).

5 RELATION BETWEEN AGING RATE K AND COEFFICIENT OF VARIATION OF TIME TO FAILURE

The probability density function f(k;t) for the time to failure is obtained from Equation 16 and Equation 18:

$$f(k;t) = \lambda(k;t).D(k;t) \tag{36}$$

$$f(k;t) = \frac{1-k}{\mu - kt}.\left(\frac{\mu}{\mu - kt}\right)^{1-\frac{1}{k}} \tag{37}$$

Or equivalently,

$$f(k;t) = \frac{(1-k).\mu^{1-\frac{1}{k}}}{(\mu - kt)^{2-1/k}} \tag{38}$$

It follows that the variance of the time to failure is given by:

$$\sigma^2 = \int_0^{\mu/k} \frac{(1-k).\mu^{1-\frac{1}{k}}}{(\mu - kt)^{2-1/k}}(t-\mu)^2\,dt \tag{39}$$

This integral can be calculated in closed form, which results in:

$$\left(\frac{\sigma}{\mu}\right)^2 = \frac{1-k}{1+k} \tag{40}$$

an expression for the coefficient of variation σ/μ as a function of the ageing rate k. Equivalently:

$$k = \frac{1 - \left(\frac{\sigma}{\mu}\right)^2}{1 + \left(\frac{\sigma}{\mu}\right)^2} \tag{41}$$

Equations 40 and 41 show that the coefficient of variation of the time to failure decrease from 1 (exponential distribution) to 0 (deterministic case) as the ageing rate increases from 0 to 1.

For instance, the uniform distribution (k = 1/2) corresponds to $\sigma/\mu = \sqrt{1/3}$, as is well known.

6 POTENTIAL USE FOR PHM AND GENERALIZATION

6.1 *Use for RUL estimation*

In PHM applications, one of the key concerns is to estimate the RUL of a monitored asset on the basis of past observations. If, for some reason, based for instance on physics, it is believed that MRL decreases linearly with time, then Equations 32–33 can be used to determine a confidence interval for the RUL at any time t. To that end, it is necessary to estimate the ageing rate k. If field data or simulation data are available that represent failure times or pseudo-failure times (a pseudo-failure may defined as the crossing of a threshold by a health

indicator, for instance), then those data can be used in order to obtain a MLE (Maximum Likelihood Estimator) of k, or a Bayesian estimator, on the basis of Equation 38. If on the other hand an estimate of the coefficient of variation of time-to-failure (or time-to-pseudo-failure) can be obtained from field data or simulations, then Equation 41 can be used in order to estimate k.

However, in general, the assumption of linear evolution of MRL with time may be unrealistic.

Below it is now shown that the results demonstrated so far can be generalized to the case of a MRL which is a piecewise linear function of time.

Therefore the generality of the method is considerably extended, as any continuous function can be approximated by a piecewise-linear function.

6.2 *Generalization*

Let us consider the case of a MRL consisting of two linear segments, as in Figure 3; it is then possible to generalize to any number of segments.

In that example, the MRL has a change of slope at $T_1 = 40$ h. It is defined as follows:

$$\text{For } 0 < t < T_1,\ V(t) = \mu - k_1 t \tag{42}$$

$$\text{For T1} < t < T_2, V(t) = \mu - k_1 T_1 - k_2 (t - T_1) \tag{43}$$

where $V(T_2) = 0$

Therefore the parameters are linked by the relation $\mu - k_1 . T_1 - k_2 (T_2 - T_1) = 0$.

In the example of Figure 3, ageing accelerates after $T_1 = 40$ h. The slopes are $k_1 = 0.5$ before T_1 and $k_2 = 0.9$ after T_1, and $T_2 = 133.33$ h

From Equation 42, the reliability function is obtained, using Equation 4.

In this example,

$$\text{For } 0 < t < T_1, D(k_1;t) = (1 - \frac{k_1 t}{\mu})^{\frac{1}{k_1}-1} \tag{44}$$

And, for $T_1 < t < T_2$,

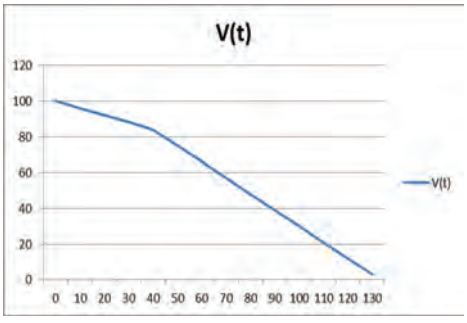


Figure 3. Piecewise linear MRL (2 segments).

$$D(k_1;t) = \left(1 - \frac{k_1 t}{\mu}\right)^{\frac{1}{k_1}-1} \left(\frac{\mu - (k_1 - k_2)T_1 - k_2 t}{\mu - k_1 T_1}\right)^{\left(\frac{1}{k_2}-1\right)} \tag{45}$$

From those expressions, it is possible to derive a confidence interval for the RUL in the same manner as was done in Section 4 for the linear case, i.e. from the conditional survival function.

Similarly, the probability density can be derived from Equation 44 and Equation 45 and therefore the MLE of parameters k_1 and k_2 can be obtained..

Increases in the ageing parameter k should alert the maintainer as they are reasonably correlated with reduction in the RUL.

Generalization to a piecewise linear function with any number of segments (instead of two) is relatively straightforward.

7 CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

Prognostics methods usual fall into the following categories: physics-of-failures analysis, data-driven methods, and fusion methods which combine both.

Unlike traditional reliability engineering, PHM focuses on individual assets rather than on a population of assets. However, the author contends that some methods of reliability engineering can be of use to PHM and can complement the other approaches, in particular for the elaboration of maintenance policies at fleet level. The recent work by (Huynh et al. 2017) points in that direction.

In this paper, a class of time-to-failure distributions characterized by a mean residual life evolving linearly in time has been studied, to derive confidence intervals for the RUL of assets whose time-to-failure follows such distributions.

Despite the fact that that class contains very important representatives such as the exponential distribution, the uniform distribution and the Dirac distribution, we have not found such a study in the literature.

The relationship between the ageing rate and the coefficient of variation of the time to failure is interesting to the extent that it exhibits an intimate link between the variability (measured by the coefficient of variation) and the rate of ageing: the higher the former, the lower the latter. It also allows to estimate one from the other.

The results have also been generalized to distributions with piecewise linear MRL, which can approach any distribution if enough segments are considered. The challenge then will consist of identifying the 'knees', i.e. the points of change of slope in the MRL. Methods such as trend filtering

(Tibshirani 2014), for instance, could be used for that purpose. It would be of interest to investigate how coefficient of variation of time to failure and ageing rate are linked in that general case; and to study in that respect the coefficient of variation of the RUL rather than that of the time to failure.

Recent contributions (Atamuradov et al. 2017) have highlighted the importance of changes in coefficient of variation of extracted features in detecting faults and in segmenting time series for health assessment and prognostics.

Since useful health indicators normally vary monotonically with RUL, changes in their coefficient of variation could be used as proxies for changes in that of the RUL.

In conclusion, the author believes that PHM stands to gain from cross-fertilization with traditional reliability engineering.

REFERENCES

Atamuradov, V., Medjaher, K., Lamoureux, B., Dersin, P., Zerhouni, N. 2017. Fault Detection by Segment Evaluation based on Inferential Statistics for Asset Monitoring, *Annual PHM Society Conference, St-Petersburg, FL, October 2017.*

Barlow, R.E. and Proschan, F., 1996. Mathematical theory of reliability; New York, Wiley 1965.

Birolini, A. (2017). *Reliability engineering: theory and practice. Springer*.

Dersin, P., Alessi, A., Lamoureux, B., Brahimi, M., Fink, O., 2017. Prognostics and Health Management in Railways, *ESREL 2017*.

Gouriveau, R., Medjaher, K., & Zerhouni, N. (2016). *From Prognostics and Health Systems Management to Predictive Maintenance 1: Monitoring and Prognostics*. John Wiley & Sons.

Huynh, K.T., Grall, A., & Bérenguer, C. (2017). Assessment of diagnostic and prognostic condition indices for efficient and robust maintenance decision-making of systems subject to stress corrosion cracking. *Reliability Engineering & System Safety, 159*, 237–254.

Si, X.S., Wang, W., Hu, C.H., & Zhou, D.H. (2011). Remaining useful life estimation–a review on the statistical data driven approaches. *European journal of operational research*, 213(1), 1–14

Swartz, G.B., 1973. The mean residual lifetime function. *IEEE Transactions on Reliability,* 22(2), pp.108–109.

Tibshirani, R.J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, 42(1), 285–323.

Watson, G.S. and Wells, W.T;,1961, On the possibility of improving the mean useful life of items by elimieliminating those with short lives, *Technometrics 3, 281–298*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Join optimization of detectors' fleet settings to maximize global detection power

P. Beauseroy & E. Grall-Maës
Institut Charles Delaunay/ROSAS Departement/ Systems Modelling and Dependability Team, CNRS, ICD/ROSAS/M2S, Université de Technologie de Troyes, Troyes, France

ABSTRACT: In many applications, detection methods are implemented to monitor systems. Detector may be built upon analytical model of the system behaviors, or based on recorded data. Lately lots of efforts have been put in developing methods that enable to transfer a detector from an assessed system to a new one when the 2 systems are similar. These so-called transfer learning approaches have shown efficiency when dealing with a fleet of systems composed of different similar systems. For the detector of each model of system a tradeoff between false alarm and non-detection must be chosen. One question that has not retained much attention up to now is the tuning of these detection systems. The common practice is to tune the detection system associated with each model individually. In this communication we tackle the problem of the global efficiency of this practice. How can we tune all those detection systems together so that the group of detectors satisfies some performance constraint? The formalization of the problem is introduced, discussed and commented based on detector ROC curves. Three toy examples are used to show the gain one could expect from join optimization of detectors and to illustrate optimization issues. Example results show that non detection probability can easily be reduced by up to 50%. In conclusion several extensions of this work are discussed.

1 INTRODUCTION

1.1 *Context*

In many applications, detection methods are implemented to monitor systems. The aim of these detectors is basically to setup an alarm when the monitored system behavior changes from its regular behavior. Detector may be built upon analytical model of the system behaviors (Tartakovsky et al. 2015), or based on recorded data (Pimentel et al. 2014). Lately lots of efforts have been put in developing methods that enable to transfer a detector from an assessed system to a new one when the 2 systems are similar (He et al. 2014), (West et al. 2007), (Evgeniou et al. 2005)]. These so-called transfer learning approaches have shown efficiency when dealing with a fleet of systems composed of different similar systems. The transfer learning idea is to adapt a decision system trained on a source system to a new target system using a limited amount of data coming from the target system. The adaptation process enables to keep most of the information coming from the source system, so to have good performances with few data. An informal way to put it would be to consider transfer learning as a calibration operation of a trained detector on a target system. To give an example of such a fleet, we can consider car engines in the case of K car models sharing the same engine. Since the models differ, the monitoring system may slightly differ from one model to another. It leads to define one detector D_i ($i = 1$ to K) with its specific tuning parameters for each car model.

For each detector, so for each model, a tradeoff between false alarm and non-detection must be chosen. One question that has not retained much attention up to now is the tuning of these detection systems D_i. A quite common practice is to tune the detection system associated with each model individually, so that each individual detector satisfies the global targetted performance constraint.

1.2 *Considered problem*

In this communication we tackle the problem of the global efficiency of this practice. Can we tune all the detection systems together so that the group of detectors satisfies some performance constraint? Does it make sense to tune all the detectors together? Is it feasible? And finally can we expect gain from such an approach?

2 FORMULATION OF THE PROBLEM

Consider K systems and the corresponding K detectors. Each detector makes decisions according

to measurements x between two hypotheses: H_0 the system is working fine and H_1 the system experiments a problem. Measurements x are considered as realization of a random variable X. Without loss of generality, we can assume that each detector D_i takes decisions depending on a function of the measurements $d_i(x)$. Each detector compares the output value of the function $d_i(x)$ with a tuned threshold t_i which is equivalent to decide according to:

$$D_i(x) = sign\left(d_i(x) - t_i\right). \tag{1}$$

H_1 is chosen and an alarm is set when $D_i(x) = 1$.

In this context the thresholds t_i must be tuned according to some criteria. Criteria are usually related to performance constraints. A very well-known performance constraint setting is the Neyman Pearson framework which is adopted here (DeGroot and Schervish 2002), (Neyman and Pearson 1933). So detectors are tuned according to a false alarm rate.

We assume that the owner of these systems considers that a global acceptable false alarm rate is α_G, so systems must be tuned according to this objective.

A usual mean to respect this objective is to tune each model so that its false alarm rate is α_G:

$$P_{fa_i} = P(D_i(X) = 1 \mid H_0) \tag{2}$$

$$= \int_{\{x \mid d_i(x) - t_i \rangle 0\}} f_i(x)\,dx \tag{3}$$

where $f_i(x)$ is the probability density fonction of X for system i.

A basic solution consists to choose $t_i^{(0)}$ so that $P_{fa_i} = \alpha_G$ for all systems i.

By doing so we ensure that the global performance, considering all models, named P_{fa_G} is α_G:

$$P_{fa_G} = \sum_{k=1}^{K} P_k P_{fa_k}. \tag{4}$$

where P_k is the a priori probability of using system k.

This solution is valid. But in such a case a very specific setting has been chosen among many possible ones. One could allow a larger false alarm rate on some models and a smaller one on others and still respects the constraint. As in Neyman Pearson problem, we consider that the best solution is the one that respects the constraint and that maximizes the detection power of the global system P_d:

$$P_d = P(D(X) = 1 \mid H_1) \tag{5}$$

$$= \sum_{k=1}^{K} P_k P(D_k(X) = 1 \mid H_1) \tag{6}$$

So the aim is to find $(t_1, t_2, \ldots, t_K)^*$ that maximize P_d under the constraint $P_{fa_G} \le \alpha_G$.

3 OPTIMIZATION

The optimization problem we want to solve is the following:

$$\begin{cases} \min\limits_{(t_1, t_2, \ldots, t_K)} P_d \\ \text{with} P_{fa_G} = \alpha_G \end{cases} \tag{7}$$

This problem could be optimized based on lagrangian formulation as in (Grall-Maes and Beauseroy 2009). But this would lead to a quite complex optimization problem for at leazt two reasons. The change of a threshold t_l has an indirect effect on both $P(D_l(X) = 1 \mid H_1, t_l)$ and $P(D_l(X) = 1 \mid H_0, t_l)$. To express the dual Lagrangian problem one would need to express the partial derivatives of these probabilities according to t_l for all l. Such expression cannot be computed in the case of trained detector, and in case of toy problem, as in this paper, this formulation would be rather complex even for simple laws on X. The second reason is that the function P_d may be non-convex leading to local minima.

3.1 *Proposed approach*

To tackle this optimization problem we propose a rather simple strategy that consists to first find a possible solution, next select 2 detectors l and m and change their settings so that their join probability of false alarm $P_{fa_{l,m}}$ does not change:

$$P_l P_{fa_l} + P_m P_{fa_m} = P_{fa_{l,m}} \tag{8}$$

and so that $P_{d_{l,m}}$ is maximized

$$P_{d_{l,m}} = P_l P_{d_l} + P_m P_{d_m} \tag{9}$$

This treatment is repeated with a new couple of detectors until the detection probability cannot be improved.

3.2 *Tuning two detectors*

To implement the proposed approach, for a given couple of detector (l, m) one need to change simultaneously their thresholds t_l and t_m so that equation 8 is satisfied.

So if t_l is changed to t'_l then P_{fa_l} is changed to P'_{fa_l} and thus the threshold t'_m that satisfies the relation:

$$P'_{fa_m} = \frac{P_{fa_{l,m}} - P_l P'_{fa_l}}{P_m}, \tag{10}$$

has to be found.

The impact of these changes can be obtained directly using the ROC curves of the two detectors since each ROC curve establishes a relation:

$$P_{d_k} = g_k\left(P_{fa_k}\right) \tag{11}$$

The ROC function $g_k()$ is a non decreasing function.

To maximize the detection probability for this couple, we search for a point that maximizes:

$$P_{d_{l,m}} = P_l g_l\left(P_{fa_l}\right) + P_m g_m\left(P_{fa_m}\right) \tag{12}$$

Combining with equation 10, we obtain

$$P_{d_{l,m}} = P_l g_l\left(P_{fa_l}\right) + P_m g_m\left(\frac{P_{fa_{l,m}} - P_l P_{fa_l}}{P_m}\right) \tag{13}$$

And deriving according to P_{fa_l} we obtain:

$$\begin{aligned}\frac{dP_{d_{l,m}}}{dP_{fa_l}}\left(P_{fa_l}\right) &= P_l \frac{dg_l}{dP_{fa_l}}\left(P_{fa_l}\right) + \\ &P_m\left(\frac{-P_l}{P_m}\right)\frac{dg_m}{dP_{fa_m}}\left(\frac{P_{fa_{l,m}} - P_l P_{fa_l}}{P_m}\right)\end{aligned} \tag{14}$$

which simplifies to:

$$\begin{aligned}\frac{dP_{d_{l,m}}}{dP_{fa_l}}\left(P_{fa_l}\right) = &\; P_l \frac{dg_l}{dP_{fa_l}}\left(P_{fa_l}\right) \\ &- P_l \frac{dg_m}{dP_{fa_m}}\left(\frac{P_{fa_{l,m}} - P_l P_{fa_l}}{P_m}\right)\end{aligned} \tag{15}$$

This formulation gives a mean to implement a gradient algorithm, increasing P_{fa_l} if the gradient is positive and decreasing it otherwise. P_{fa_m} is then set according to equation (10). The optimum point corresponds to an extreme point or to a point that cancels the derivative (equation 15). This formulation enables to optimize the local detection probability given by (9) under the constraint (8) according to the detectors false alarme rate instead of the thresholds (the threshold may be deduced afterward).

3.3 *Comments*

In order to speed up the convergence the idea is to choose the couple (l, m) so that the absolute value of the gradient is as large as possible at each iteration, meaning that the optimization focuses on the couple that locally brings the largest improvement.

It must be mentioned that the proposed alternate gradient approach can lead to a global optimum only if the starting point is located in a convex neighborhood of the optimal solution. Otherwise the algorithm converges to a local optimum.

To reach global optimum we propose to initiate the optimization process with different initial solutions when the number of detectors is large. When it is not too large a random optimization method can be applied without big computing time loss.

The experimental study that follows illustrates the optimization process for two detectors. The impact of prior on the solution are discussed. More complex cases are tested with four detectors, and the non convex case is illustrated with 2 examples.

4 EXPERIMENTAL RESULTS

We design the experiments to show the key issues related to the problem. The experimental study first illustrates the optimization process for two detectors. The impact of priors on the solution is discussed. More complex cases are tested next with four detectors, and the non convex case is illustrated with 2 examples.

4.1 *Two systems case*

Our first experiment corresponds to the simplest setting with 2 models. The ROC curves of the two detectors are shown on Figure 1. In that scenario

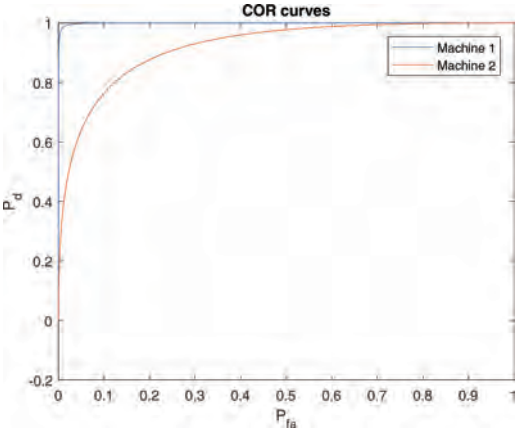


Figure 1. ROC curves of machine 1 and 2 detectors.

one detector is good while the second one is significantly less accurate.

The targeted global false alarme rate α_G is set to 0.1 and the systems are assumed equiprobable.

By tuning each detector to reach a false alarm rate of 0.1, the power of the global system is about 0.88 (Figure 2). By using the proposed optimization approach we improve the power of the global system to 0.925. To obtain this result the worst detector (system 2) is tuned to a higher false alarm rate (0.185) while the best one is reduced to less than 0.02 as shown by Figure 3.

It means that to optimize the global system we must be more demanding with the best detector: the one which detection rate decreases slowly when false alarm rate is reduced. On the contrary, the worst detector's false alarm rate is relaxed to gain a significant detection improvement. Logically, combining these two actions enables to respect the false alarm constraint and consequently improves the overall detection rate.

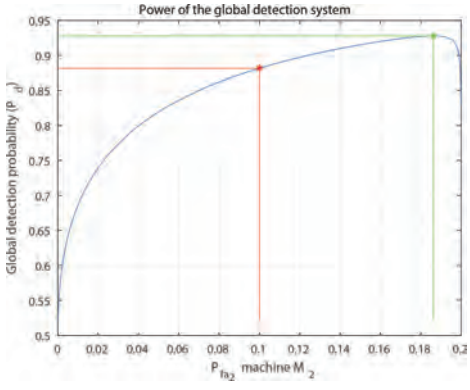


Figure 2. ROC curves of the complete system depending on systems 2 detector's false alarm with usual tuning (red) and proposed solution (green).

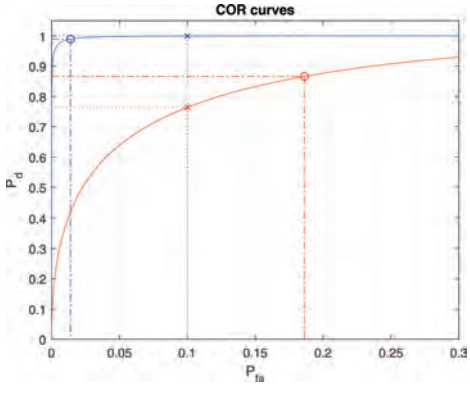


Figure 3. Tuning of each detector reported on its ROC curve—standard tuning (dotted lines) and proposed tuning (dash lines).

4.2 *Impact of models prior*

With a two system case the point which cancels the derivative (15) depends on the prior probabilities due to the argument $\frac{P_{faI,m} - P_I P_{faI}}{P_m}$ of its second term: $\frac{dg_m}{dP_{fa_m}}$. To study the impact of the prior probability the example of previous section is used again changing the prior P_1 from 0 to 1 and decreasing P_2 accordingly. The gain related to the global optimization is reported in Figure 4. It goes from 0 up to 0.047 which represents approximately a 30% reduction of the non detection rate (one minus the detection probability is plotted in Figure 5) which can be considered as significant.

The optimal false alarm tuning of machine 1 and 2 as a function of prior probability is illustrated by Figure 6. The false alarm rate of the best detector

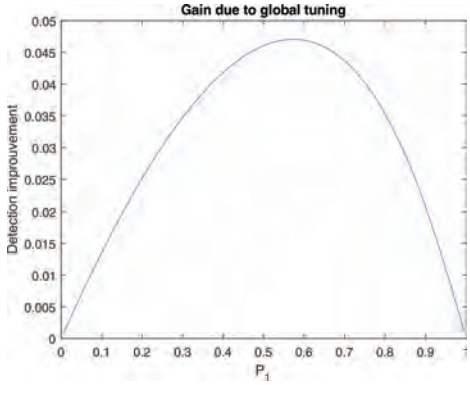


Figure 4. Performance gain as a function of the probability to use machine 1.

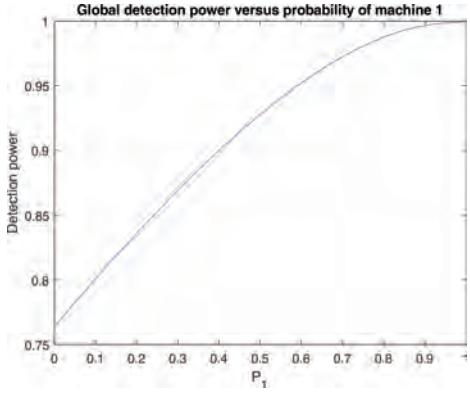


Figure 5. Power of the global system as a function of the probability to use machine 1.

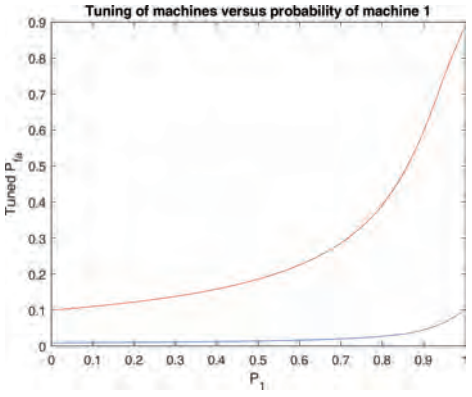


Figure 6. Tuned optimal false alarm probability for the detectors as a function of the probability to use machine 1 - machine 1 in blue, machine 2 in red.

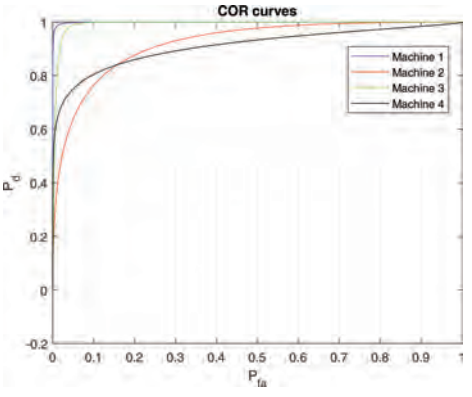


Figure 7. ROC curves of a 4 detector case.

(blue line) is always smaller that the targeted value $\alpha_G = 0.1$ but logically tends to α_G when P_1 tends to 1, while the false alarm rate of the worst detector (red line) is always larger than α_G.

4.3 *Simple four systems case*

To illustrate the proposed optimization process we designed a $K = 4$ systems case. The detector ROC curves corresponding to these systems are given by Figure 7. All priors are set to $\frac{1}{K}$ and $\alpha_G = 0.25$.

The optimal tuning is illustrated by Figure 8. The non detection rate corresponding to the reference setting $(\alpha_i = \alpha_G \ \forall i = 1..K)$ is 0.054 and decreases to 0.028 once optimized.

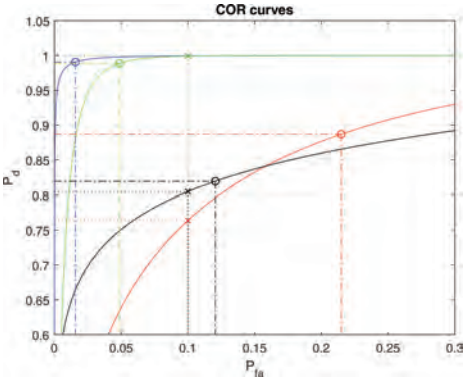


Figure 8. Tuning of each detector reported on its ROC curve—standard tuning (dotted lines) and proposed tuning (dash lines).

4.4 *Optimization limits*

To illustrate the limits of the proposed optimization process when the power function is not convex, a 3 system case has been designed and compared to a convex case.

Figure 9 shows the 3 detectors ROC curves for the convex case. The corresponding power function is displayed by Figure 10.

Figure 11 shows the 3 other detectors ROC curves. The corresponding power function is displayed by Figure 12.

In the second case the power function is not convex while the ROC curves are not so complex. It means that non convex case may not be an exception. In fact such case will appear each time the second derivative of a ROC curve becomes positive which is not so rare. It implies that the proposed optimization process may lead to a local optimum. So we have to find an optimization procedure that can cope with such non convex case.

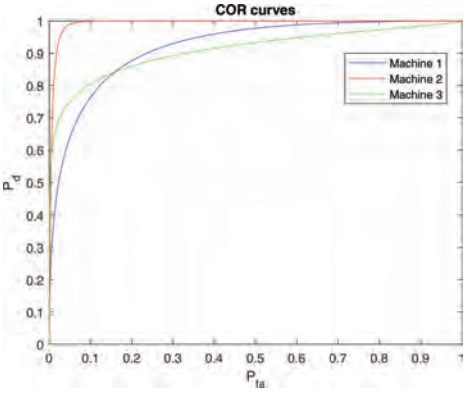


Figure 9. ROC curves of a simple 3 detector case.

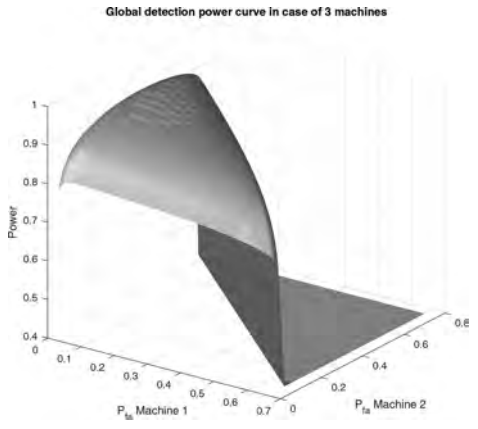


Figure 10. ROC surface of the global system as a function of detectors 1 and 2 tuning.

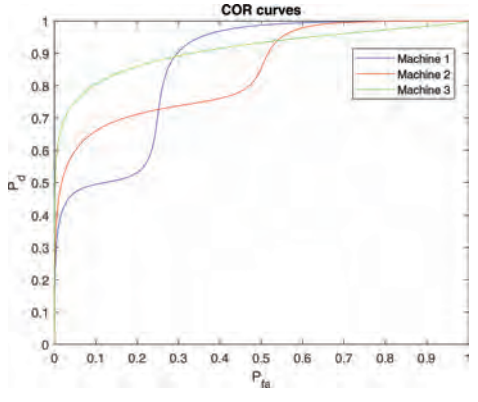


Figure 11. ROC curves of a more complex 3 detector case.

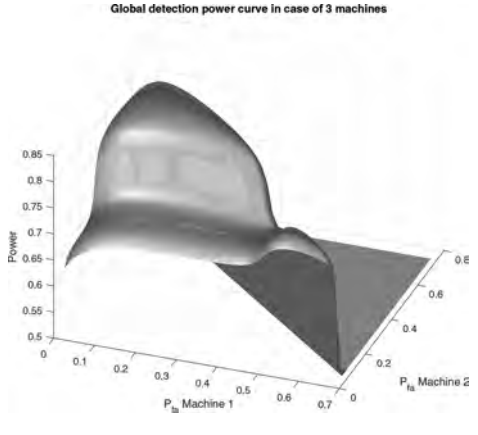


Figure 12. ROC surface of the global system as a function of detectors 1 and 2 tuning.

4.5 *Improved optimization process*

To improve the proposed optimization process, the idea is to initiate the optimization process from many different points. So we generate those initial solutions randomly. In our simulation we randomly chose initial false alarm rate for each machine and normalize those initial guess so that the global false alarm rate satisfies the constraint. Next, for each initial solution we apply the previous optimization procedure.

We apply this improved process on a 4 system example whose detector ROC curves are shown in Figure 13.

Figures 14 and 15 depict the results found with the initial optimization process and the results based on the improved one respectively.

Table 1 summarizes the tuning of each system and the overall performance in each case. The non detection rate decreases from more than 17% in that case to 10.3% in the case of a simple optimiza-

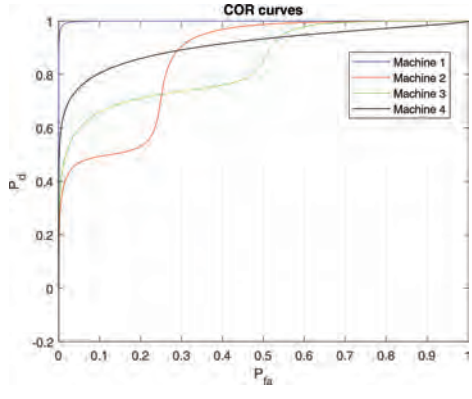


Figure 13. ROC curves of a complex 4 detector case.

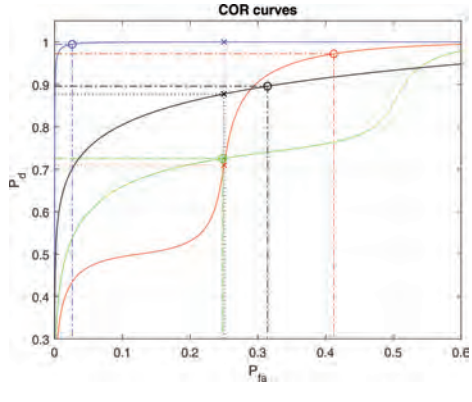


Figure 14. Tuning of each detector reported on its ROC curve—standard tuning (dotted lines) and proposed tuning (dash lines) obtained with the proposed optimization strategy.

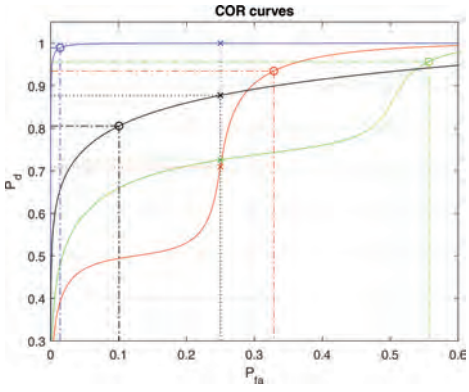


Figure 15. Tuning of each detector reported on its ROC curve—standard tuning (dotted lines) and proposed tuning (dash lines) obtained using random optimization.

Table 1. Tuning and performances for different settings.

Exp.	P_{fa_1}	P_{fa_2}	P_{fa_3}	P_{fa_4}	P_d
Ref.	0.25	0.25	0.25	0.25	0.828
Basic opt.	0.026	0.412	0.248	0.314	0.897
Imp. opt.	0.014	0.328	0.557	0.101	0.922

tion and is reduced to 7.8% based on the improved optimization proposed in this last section. These results illustrate the influence of detectors tuning on the overall performances and show that very significant gain can be achieved.

5 CONCLUSION

This communication proposes to tackle the problem of global detection optimization of a detector fleet. This can be encountered in distributed detection systems or in detection systems applied to a fleet of similar systems. In a distributed system the question that arises is to find an optimal fusion of the local detections that all relate to the same object (Blum and Kassam 1997). In the latter case the main concern is about the global performance of the system over decisions on different populations, typically a fleet of similar systems, with a specific detector for each kind of systems.

We propose a model for this problem that fits in the Neyman-Pearson's framework. We develop the analytical formulation and discuss the optimization issue of the global detection system in terms of detector ROC curves. An approach to optimize the global system is proposed and some simulations are designed to illustrate the main issues related to the problem and to the optimization process.

It is shown that the optimization can significantly improve the global performances compared to a reference naive approach. It tends to prove the utility of the proposed approach.

The next steps are the extension of this first study to real cases. This extension seems to be quite straightforward since the approach is based on ROC curves. The fact that the ROC curves correspond to maximum likelihood ratio tests as in the proposed examples or the fact that they are build upon trained detectors makes no real difference.

A more challenging extension is to develop the same approach in the case of production lines taking into account some real time issues. Typically, in recent production systems, the production of a line may evolve due to diverse causes such as production planning, availability of the line inputs, maintenance issues…The related questions are connected to our capacity to adapt, in real time, such detection or quality control systems to these changes and what gain can we expect from such a dynamic tuning system.

Other extensions can be studied such as the impact of the number of detectors. Can we use the same optimization approach or should we adopt more efficient ones? The case of detection with a rejection option should also be considered. The optimal criteria in the latter case must be redefined and the tuning strategy must be also reconsidered.

REFERENCES

Blum, R. S. & H.V. Kassam, S. A.and Poor (1997). Distributed detection with multiple sensors i. advanced topics. *Proceedings of the IEEE 85*(1), 64–79.

DeGroot, M. & M. Schervish (2002). Probability and Statistics. Addison-Wesley series in statistics. Addison-Wesley.

Evgeniou, T., C. Micchelli, & M. Pontil (2005). Learning multiple tasks with kernel methods. *Journal of Machine Learning Research 6*, 615–637.

Grall-Maes, E. & P. Beauseroy (2009). Optimal decision rule with class-selective rejection and performance constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence 31*, 2073–2082.

He, X., G. Mourot, D. Maquin, J. Ragot, P. Beauseroy, A. Smolarz, & E. Grall-Maes (2014). Multi-task learning with one-class svm. *Neurocomputing 133*, 416–426.

Neyman, J. & E. Pearson (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 231*, 289–337.

Pimentel, M., D. Clifton, L. Clifton, & L. Tarassenko (2014). A review of novelty detection. *Signal Processing 99*, 215–249.

Tartakovsky, A., I. Nikiforov, & M. Basseville (2015). *Sequential analysis: Hypothesis testing and changepoint detection*. Boca Raton, FL: Chapmann Book, Taylor & Franis.

West, J., D. Ventura, & S. Warnick (2007). *Spring Research Presentation: A Theoretical Foundation for Inductive Transfer*. Brigham Young University, College of Physical and Mathematical Sciences.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Assessment method of the deterioration degree of asphalt concrete airport pavements

M. Zieja, P. Barszc, K. Blacha & M. Wesołowski
Air Force Institute of Technology, Warsaw, Poland

ABSTRACT: An important factor, which affects the safety of carried out flight operations, constitutes proper management of airports on the basis of information on the pavement condition of their functional elements, obtained in a systemic manner. One of the elements of the technical condition estimation of airport pavements is the assessment of their deterioration degree on the basis of identified damage and carried out repairs. Such an approach allows to forecast essential resources necessary to carry out repairs and rationally plan overhauls. The multicriteria analysis presented in this article is a method of the weighted assessment supporting the deterioration degree estimation of airport pavements. While calculating the criteria, it is important to focus on parameters characterising the deterioration degree of airport pavements, both of civil and military facilities. Therefore, it is important to diagnose the functional elements' pavements of these airports within the framework of current five-year inspections, and to carry out inspections including the inventory of damage and made repairs within the annual intervals.

1 INTRODUCTION

In the literature and operational practice, many terms related to the life phases of airport facilities, such as, e.g. durability, service life, labour resource, technical service life resource, service life resources between repairs, are used (Shahin 2007, Zieja et al. 2016, Zieja 2015, Werbińska-Wojciechowska & Zając 2015, Barszcz & Blacha 2015 & 2016). The airport facilities' life phases are characterised by their operation strategies. Figure 1 graphically shows operation strategies of the airport functional elements' pavements, which currently function within the framework of operation of airport pavements.

In case of the functional elements' pavements of airports operated by air forces, mixed strategies of their operation are used; however, a series of works aimed at proceeding to the operation according to the technical condition is taken. The operation in accordance with the technical condition requires to obtain information, which would make it possible to monitor the technical condition of airport pavements. One of the parameters characterising the pavement's technical condition includes its deterioration degree. The deterioration degree is estimated on the basis of data obtained during the inventory. The inventory includes activities aimed at drawing up a detailed physical inventory listing of components, which specify the assessed feature for a given day. The inventory of damage and repairs to the functional elements' pavements of airports is carried out in accordance with the adopted assumptions with the use of the previously prepared bases. It involves the inspection of a basic element, which includes a slab (sample). The damage and performed repairs identified during the inspection are marked on the previously prepared bases with the use of adopted symbols with appropriate colours. Within the framework of the carried out inventories, it was assumed that the identified damage is marked with red, and repaired damage with black colour. The quantities which characterise a given type of damage and repairs are marked on the previously prepared bases (Żurek et al. 2014, Barszcz & Wesołowski 2015, Zio 2009). In accordance with the adopted nomenclature, the parameters characterising the deterioration degree of the airport pavement are marked, inter alia, as follows (Figure 2):

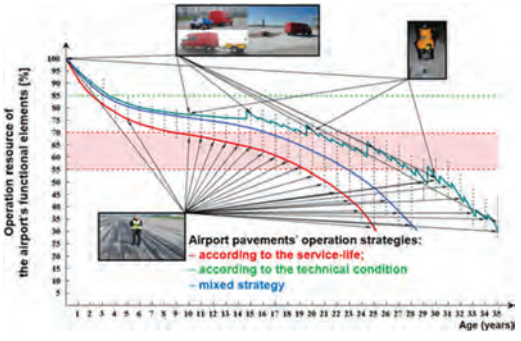


Figure 1. Operation strategies of the airport functional elements' pavements.

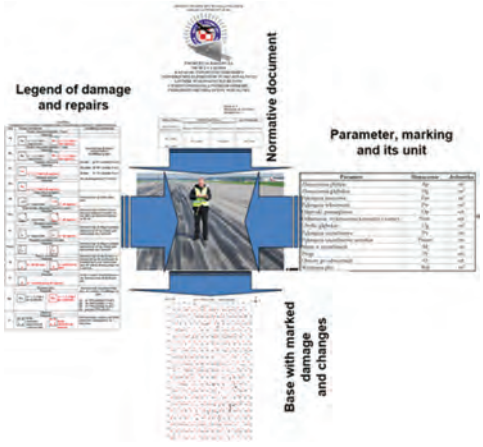


Figure 2. Inventory of damage and repairs performed by a prepared expert with the use of a visual method.

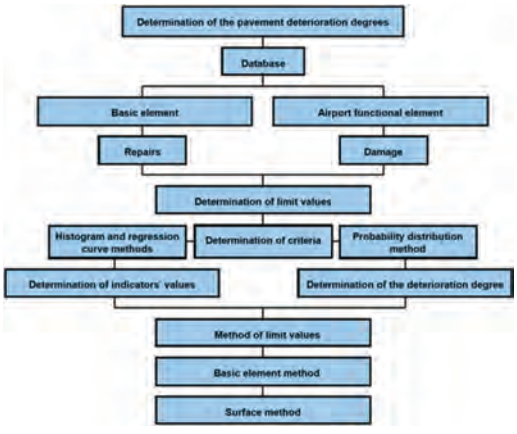


Figure 3. Algorithm of assessing the deterioration degree of airport pavements.

– alkali-silica reaction, alligator cracking;
– surface losses;
– blisters;
– chipping and loosening;
– cavity and wide cavity cracks;
– heaves or fractures;
– deep losses;
– ruts;
– boreholes.

One of the most important functional elements includes runways, and therefore, they are given a higher standard of maintenance than taxiways or airport aprons. Loose parts of the pavement can cause damage to the aircraft engines and propellers. The aircraft operational safety forces a way of conducting the runway inventory, where the slab is considered a primary element to be inspected. In case of the airports' functional elements, such as airport aprons or taxiways, the pavement tests can be carried out by analysing the number of selected samples on the tested section; however, at the same time, it is important to pay attention whether operated aircraft taxi with the use of their own drives or they are towed along the analysed test section (Smirnov & Mickiewicz 2002).

2 ALGORITHM OF ESTIMATING THE DETERIORATION DEGRE OF AIRPORT PAVEMENTS

The algorithm, presented in Figure 3, shows the way of assessment of the deterioration degree of the airport's functional element pavement.

In order to assess the pavement deterioration degree, a finite sequence of defined activities is performed, and it includes:

– determination of the parameter limit values of relative functional elements of airports on the basis of the analysis of:
  • histogram and regression curve;
  • probability distribution;
– determination of the parameter limit values characterising deterioration of a single slab on the basis of the analysis of:
  • histogram and regression curve;
  • probability distribution;
– determination of assessment criteria related to the deterioration degree of airport pavements;
– determination of the indicators' values characterising the pavement deterioration;
– determination of the pavement deterioration degree.

3 METHOD OF ESTIMATING THE DETERIORATION DEGRE OF AIRPORT PAVEMENTS

Deterioration is a slow and spread over time process. It mainly involves the reduction of construction properties by the impact of external factors, which as a result, generates changes in its structure. The deterioration degree of the airport pavement's functional element is affected by damage and performed repairs. This quantity is determined on the basis of 11 defined types of damage and repairs. In order to optimally select an indicator characterising the actual deterioration degree of the pavement surfaces, two variants of calculation are considered, where it is assumed that the performance of repairs affects the pavement deterioration in 20% or 50%. While analysing the indicators characterising the pavement deterioration, the impact of the types of damage and repairs on the aircraft operation safety

is taken into account by the introduction of properly selected weights. The indicator for assessing the pavement deterioration degree of the airports' functional elements counted with the use of a method of occupied surface including the impact of a specific parameter on the aircraft operation safety is calculated on the basis of data obtained during the inventory in accordance with the following formula:

$$\overline{D^{MF}_{BA\ w}} = w^{U}_{BA} \cdot W^{UF}_{BA} + w^{N}_{BA} \cdot W^{NF}_{BA}, \quad (1)$$

$$W^{wUF}_{BA} = \frac{\sum_{i=1}^{13} \frac{\left(w^{U}_{Ob}\right)_i \times \left(Ob^{U}_{BA}\right)_i \times \left(p^{U}_{BA}\right)_i}{F}}{\sum_{i=1}^{13} \left(w^{U}_{Ob}\right)_i} \times 100, \quad (2)$$

$$W^{wNF}_{BA} = \frac{\sum_{i=1}^{13} \frac{\left(w^{N}_{Ob}\right)_i \times \left(Ob^{N}_{BA}\right)_i \times \left(p^{N}_{BA}\right)_i}{F}}{\sum_{i=1}^{13} \left(w^{N}_{Ob}\right)_i} \times 100, \quad (3)$$

where D^{MF}_{BA} = pavement deterioration of the airport's functional element made of asphalt concrete; p_i = conversion rate of the parameter characterising damage and repairs on the surface including damaged and repaired areas; w^{i}_{BA} = statistical weight of the importance of damage and repairs in the assessment of deterioration of the airport's functional element pavement; w^{i}_{Obi} = statistical weight of the validity of specific damage and repairs in the assessment of deterioration of the airport's functional element pavement; Ob_i = measurement of damage and repairs of the airport's functional element pavement; F = total area of the tested pavement of the airport's functional element; U = damage to the airport's functional element pavement; and N – repairs of the airport's functional element pavement.

The indicator for assessing the pavement deterioration degree of the airports' functional elements calculated with the use of a method of limit values including the impact of a specific parameter on the aircraft operation safety is calculated on the basis of data obtained during the inventory in accordance with the following formula:

$$\overline{D^{MG}_{BA\ w}} = w^{U}_{BA} \cdot W^{UG}_{BA} + w^{N}_{BA} \cdot W^{NG}_{BA}, \quad (4)$$

$$W^{wUG}_{BA} = \frac{\sum_{i=1}^{13} \frac{\left(w^{U}_{Ob}\right)_i \times \left(Ob^{U}_{BA}\right)_i \times \left(p^{U}_{BA}\right)_i}{F \times \left(WG^{U}_{BA}\right)_i}}{\sum_{i=1}^{13} \left(w^{U}_{Ob}\right)_i} \times 100, \quad (5)$$

$$W^{wNG}_{BA} = \frac{\sum_{i=1}^{13} \frac{\left(w^{N}_{Ob}\right)_i \times \left(Ob^{N}_{BA}\right) \times \left(p^{N}_{BA}\right)}{F \times \left(WG^{U}_{BA}\right)_i}}{\sum_{i=1}^{13} \left(w^{N}_{Ob}\right)_i} \times 100, \quad (6)$$

where D^{MG}_{BA} = pavement deterioration of the airport's functional element made of asphalt concrete; and WG_i – limit value for specific types of damage and repairs.

The unloaded indicator for assessing the pavement deterioration degree of the airports' functional elements with the use of a method of occupied surface is calculated on the basis of data obtained during the inventory in accordance with the following formula:

$$\overline{D^{MF}_{BA}} = w^{U}_{BA} \cdot W^{UF}_{BA} + w^{N}_{BA} \cdot W^{NF}_{BA}, \quad (7)$$

$$W^{UF}_{BA} = \sum_{i=1}^{13} \frac{\left(Ob^{U}_{BA}\right)_i \times \left(p^{U}_{BA}\right)_i}{F} \times 100, \quad (8)$$

$$W^{NF}_{BA} = \sum_{i=1}^{13} \frac{\left(Ob^{N}_{BA}\right)_i \times \left(p^{N}_{BA}\right)_i}{F} \times 100. \quad (9)$$

The unloaded indicator for assessing the pavement deterioration degree of the airports' functional elements with the use of a method of limit values is calculated on the basis of data obtained during the inventory in accordance with the following formula:

$$\overline{D^{MG}_{BA\ w}} = w^{U}_{BA} \cdot W^{UG}_{BA} + w^{N}_{BA} \cdot W^{NG}_{BA}, \quad (10)$$

$$W^{UG}_{BA} = \sum_{i=1}^{13} \frac{\left(Ob^{U}_{BA}\right)_i}{F \times \left(WG^{U}_{BA}\right)_i} \times 100, \quad (11)$$

$$W^{NG}_{BA} = \sum_{i=1}^{13} \frac{\left(Ob^{N}_{BA}\right)_i}{F \times \left(WG^{N}_{BA}\right)_i} \times 100. \quad (12)$$

The indicator for assessing the pavement deterioration degree of the airports' functional elements is calculated in accordance with the following formula:

$$D = w^{U}_{BA} \cdot W^{U}_{BA} + w^{N}_{BA} \cdot W^{N}_{BA}, \quad (13)$$

$$W^{U}_{BA} = w^{wUG}_{BA} \cdot W^{wUG}_{BA} + w^{wUF}_{BA} \cdot W^{wUF}_{BA} + w^{UG}_{BA} \cdot W^{UG}_{BA} + w^{UF}_{BA} \cdot W^{UF}_{BA} \quad (14)$$

$$W^{U}_{BA} = w^{wUG}_{BA} \cdot W^{wUG}_{BA} + w^{wUF}_{BA} \cdot W^{wUF}_{BA} + w^{UG}_{BA} \cdot W^{UG}_{BA} + w^{UF}_{BA} \cdot W^{UF}_{BA} \quad (15)$$

4 ASSESSMENT OF THE DETERIORATION DEGREE OF OPERATED AIRPORT PAVEMENTS

While estimating the pavement deterioration degree, the test results, obtained during the inspections with a visual method of surfaces of the airport facilities' pavements, were taken into account. The analysed spectrum of tested facilities was divided into 7 groups, and the limits of ranges

characterising the assessment criteria of the pavement deterioration were determined. The indicators, on the basis of which the values of assessment criteria of the pavement deterioration were estimated, includes D^{UN} unloaded indicator and $D^{w(UN)}$ weighted index of deterioration of the airports' functional elements, which are defined on the basis of the indicator characterising W^U damage and W^N repairs of the airport's functional element. The values, which were achieved by indicators characterising the deterioration of airport pavements while taking into account appropriately selected weights, are shown in Figures 4 and 5. However, it should be noted that for a complete view of the pavement deterioration state, it is important to analyse indicators characterising damage and repairs, which was presented in Figures 6 and 7.

The Histogram, which has a bar graph form, was used in order to graphically display the variability of

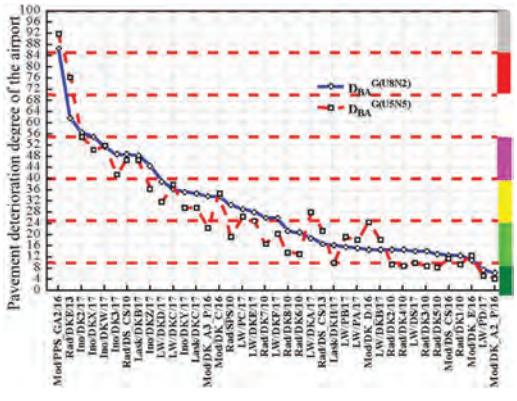


Figure 4. D^{G}_{BA} deterioration degree of the airport's functional element pavement.

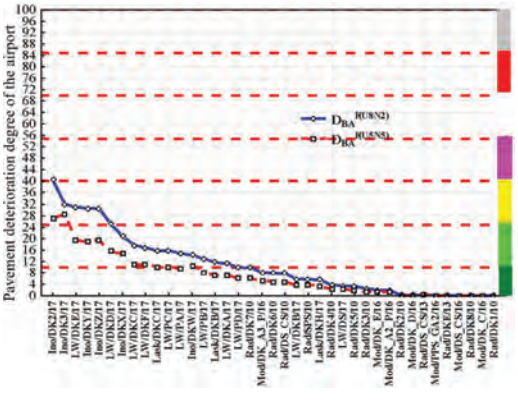


Figure 5. D^{F}_{BA} deterioration degree of the airport's functional element pavement.

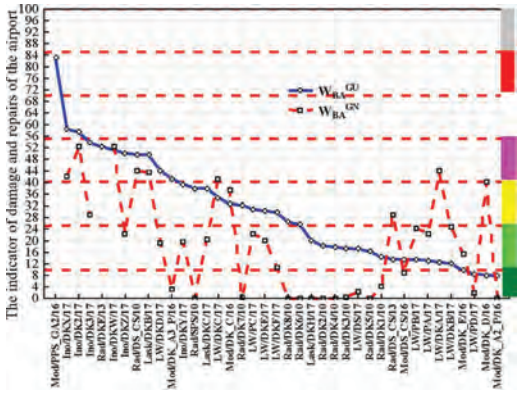


Figure 6. Indicator of the assessment of damage and repairs of the airport's functional element pavement.

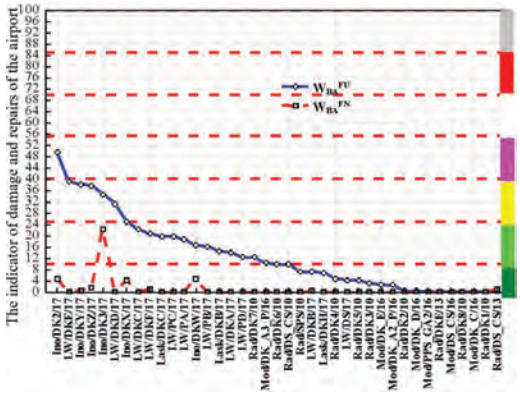


Figure 7. Indicator of the assessment damage and repairs of the airport's functional element pavement.

a particular set of data characterising the pavement deterioration. The organisation of a set of raw data is based on the division into ranges, the so-called classes. It allows to present the empirical distribution of characteristics for quantitative variables, and it determines the values at which the majority of results is located. Based on the probability distribution analysis, the corresponding distributions relevant to the nature of the probability density function were selected. The probability refers to the possibility of the occurrence of an event or several events. With data of the index development characterising the pavement deterioration degree, the probability of the occurrence of a specific event, which adopts a value in the range from 0 to 1, was determined. The probability scope of the occurrence of an event was divided into seven ranges and the possibility of the occurrence of a specific event with the specified probability was calculated.

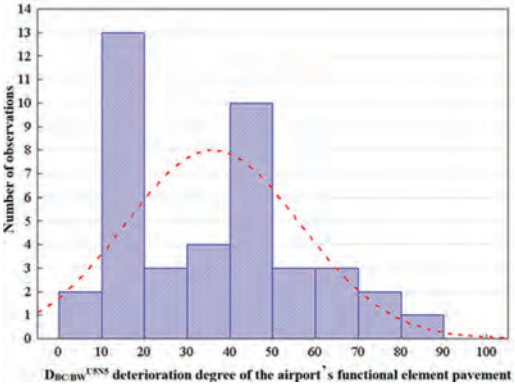


Figure 8. Standard distribution of the variability assessment indicator of the deterioration degree of the airport's functional element pavement.

The indicator offered by Air Force Institute of Technology, which characterises the pavement deterioration degree of the airport's functional elements, is within the range from 0 that means the perfect condition pavement to 100 that means the pavement unfit for further use. The calculation of D indicator is based on visual inspection results, during which different types of damage and repairs as well as their measurement are determined. The impact of a type of damage and repairs on the aircraft operation safety is included in the calculation by adopting weights estimated on the basis of the experts' method. The standard assessment scale of the pavement deterioration degree includes 7 levels, however, it is also possible to apply a simplified scale, where there are three decision-making levels of a description of the deterioration degree of the airport's functional element pavement. For each level, classes determining the pavement condition were assigned. The first one is a desired level, which includes new, renovated and operated pavements, with the assumption that these pavements will not require planned renovation works over the next five years. The indirect warning level identifies the pavement condition as the one in which it is reasonable to perform detailed tests in terms of conducting treatments in order to improve the pavement condition. The last one is a critical level, which determines the prompt performance of technical and operational research, in order to define activities aimed at the introduction of procedures to improve the pavement condition or taking the facility out of service. Figure 9 shows the relationship between decision-making levels and classes of the deterioration state of the airport's functional element pavement.

Interpretation of the pavement classes are shown in Table 1.

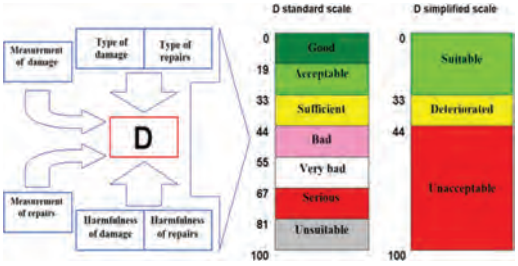


Figure 9. Assessment criteria of the deterioration degree of the airport's functional element pavement offered by Air Force Institute of Technology.

Table 1. Pavement classes.

State	$D^{F(UN)}$	$D^{G(UN)}$	$D^{FwUN)}$	$D^{Gw(UN)}$	$D^{F(UN)}$
Good	0–9	0–10	0–9	0–10	0–10
Acceptable	10–30	11–25	10–34	11–25	11–25
Sufficient	31–34	26–40	35–39	26–40	26–40
Bad	35–48	41–55	40–47	41–55	41–55
Very bad	49–51	56–70	48–51	56–70	56–70
Serious	52–54	71–85	52–69	71–85	71–85
Unsuitable	55–100	86–100	70–100	86–100	86–100

5 CONCLUSIONS

On the basis of verified parameters characterising the deterioration degree of airport pavements, it is possible to predict and estimate the period of safe operation of a particular functional element of the airport, which as a result, provides the proceeding to operation of pavements of the airport's functional element in accordance with the technical condition. The deterioration degree of airport pavements is estimated on the basis of accepted indicators with the use of selected weights verified according to the experts' method. In order to reliably predict the condition of airport pavements, it is necessary to use an objective, repetitive assessment system. Designing of the IT support system for managing pavements of the airport's functional element should be preceded by the analysis of processes, which operate within the organisational unit. Currently, while estimating the assessment criteria of the technical condition, a verified database, which was obtained within the framework of works carried out on the functional elements' pavements of tested civil airports, is used. While calculating the criteria, it is important to focus on parameters characterising the deterioration degree of airport pavements, both of civil and military facilities; therefore, it is important to diagnose these airport pavements within the framework of current five-year inspections, and to carry out inspections in order to inventory damage and repairs within the annual intervals.

REFERENCES

Barszcz, P. & Blacha, K. 2015: Estimation of assessment criteria of the deterioration degree of airport pavement functional elements made of cement concrete on the basis of a grouping method; *Proc. Scientific Conference "Contemporary problems of aviation logistics. Theory and practice". Dęblin.*

Barszcz, P. & Blacha, K. 2016 Multicriteria analysis in the assessment of the deterioration degree of airport pavement functional elements made of cement concrete. *Proc. 20th International Scientific Conference TRANSCOMP 2015. Zakopane.*

Barszcz, P. & Wesołowski, M. 2015 Estimation of assessment criteria of the deterioration degree of airport pavement functional elements made of cement concrete on the basis of data obtained from operated facilities; *Proc. 19th International Scientific Conference TRANSCOMP 2015. Zakopane.*

Casciati, F. & Roberts, B. 1996. *Mathematical Models for Structural Reliability Analysis*. Boca Raton/New York/London/Tokyo: CRC Press.

DeLurgio, S.A. 1998. *Forecasting principles and applications*. University of Missouri-Kansas City: Irwin/McGraw-Hill.

Dhillon, B.S. 1999. *Design Reliability. Fundamentals and Applications*. Ottawa: Boca Raton/New York/London/Washington: CRC Press.

Pham, H. 2006. *Handbook of Engineering Statistics*. London: Springer-Verlag.

Shahin, M. 2007 *Pavement Management For Airports, Road and Parking Lots*.

Smirnow, N. N. & Mickiewicz, A. A. 2002. Service and repair of the aircraft technology according to the state. *Translation of Air Force Institute of Technology. Warsaw.*

Tomaszek, H. & Zieja, M. & Ważny, M. 2016. A method for reliability assessment of structural components of aircraft and sea-going ships with taking into account a given failure generation model. *Polish Maritime Research* 23(2): 83–90.

Werbińska-Wojciechowska, S. & Zając, P. 2015. Use of delay-time concept in modelling process of technical and logistics systems maintenance performance. Case study. *Eksploatacja i Niezawodnosc – Maintenance and Reliability* 17 (2): 174–185.

Werbińska-Wojciechowska, S. 2007. The availability model of logistic support system with time redundancy. *Eksploatacja i Niezawodnosc-Maintenance and Reliability*.

Werbińska-Wojciechowska S. 2013. Time resource problem in Logistics systems dependability modeling, *Eksploatacja i Niezawodnosc–Maintenance and Reliability 15(4)*.

Zieja, M. & Ważny, M. & Stępień S. 2016. Distribution determination of time of exceeding permissible condition as used to determine lifetimes of selected aeronautical devices/systems. *Eksploatacja i Niezawodnosc-Maintenance and Reliability 18(1): 57–64.*

Zieja, M. 2015. A method of predicting reliability and lifetime of aeronautical hardware with characteristic function applied. *Transport Means – Proceedings of the International Conference, Kaunas, 22–23 October 2015.* Kaunas Univ. Technol.

Zio, E. 2009. *Computational Methods For Reliability and Risk Analysis*. Singapore: World Scientific Publishing.

Żurek, J. & Smalko, Z. & Zieja, M. 2010. Methods applied to identify causes of air events. *Reliability, Risk and Safety: Theory and Applications*. CRC Press-Taylor and Francis Group: 1817–1822.

Żurek J., Tomaszek H., Zieja M. 2014. Analysis of structural component's lifetime distribution considered from the aspect of the wearing with the characteristic function applied. *Safety, Reliability and Risk Analysis: Beyond the Horizon*: 2597–2602. Amsterdam: Balkema.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Applying Mahalanobis-Taguchi method to detect faults in rotating machinery

G.F.M. Souza, I.S. Melo & M.A.C. Michalski
Departamento de Engenharia Mecatrônica e Sistemas Mecânicos, Escola Politécnica da Universidade de São Paulo, São Paulo, Brazil

ABSTRACT: Monitoring systems for rotating machines has been largely used in industry even with the high reliability achieved in turbines and compressors. These systems can reduce the number of non-scheduled shutdowns and time for the scheduled ones. In this way, it remains a challenge for researchers and industry to develop better monitoring, diagnosis and prognostic techniques. These systems generally use multiple sensors to analyze the machine behavior, so the Mahalanobis-Taguchi strategy (MTS), which is recognized for its suitability for multivariate data analysis, was applied in order to integrate these sensors. Also, in order to compare and validate MTS results, a more classic approach, based in vibration analysis is also considered. Therefore, the objective of this paper is to present some initial results in the unbalance estimation in rotating machines using this method. A centrifugal compressor used in an off-shore unit is selected for application. Actual data and computational simulations were used to validate the method and to provide the unbalance estimation of the selected machine.

1 INTRODUCTION

Monitoring systems for rotating machines has been largely used in industry even with the high reliability achieved in turbines and compressors. This technology aims to reduce the non-scheduled shutdowns and the time for the scheduled ones minimizing the costs with downtime.

These problems are more severe in the off-shore facilities, because the logistics issues and weather condition play a role in reestablishing the production. In this way, it remains a challenge for researchers and industry to develop better monitoring, diagnosis and prognostic techniques.

Other peculiarity of these monitoring systems is that they have multiple sensors installed around the machine and integrate them in a troubleshooting analysis is not a simple task. So, the Mahalanobis-Taguchi strategy (MTS), which is recognized for its suitability for multivariate data analysis, was applied to improve this analysis (Cudney, 2015).

In order to verify the MTS' applicability to rotating machinery, the unbalance malfunction was chosen because, as described by Bently & Hatch (2002), this is the most common malfunction presented in this type of machines.

As it is not possible to insert an unbalance error in a real machine to validate the method, computational simulations were used to produce this data.

Also, in order to compare and validate MTS results, a more classic approach, based in vibration analysis is also considered.

Therefore, the objective of this paper is to present some initial results in the unbalance estimation in rotating machines using MTS. A centrifugal compressor used in an off-shore unit is selected for application.

2 MAHALANOBIS-TAGUCHI SYSTEM

As described by Cudney (2015), Soylemezoglu et al. (2011), Xin & Chow (2013) and John (2014), the Mahalanobis-Taguchi system (MTS) is a method that has been used in multivariable diagnostic applications. It creates two types of groups, the "normal" and "abnormal", using the Mahalanobis Distance (MD), and then optimize the number of variables used in the diagnostic by applying the Orthogonal Arrays (OA) and Signal-to-Noise ratios (SN). Finally, a prognostic can be done in order to avoid the application´s fault. The MTS can be generally summarized in four stages, as follows:

2.1 *Stage 1: Mahalanobis space construction*

The normal group is created in this step to be used as reference. The variables are selected and "healthy" samples of each one collected.

At first, calculate the mean of each variable in the "normal" condition per Equation 1:

$$\bar{x}_i = \frac{\sum_{j=1}^{n} X_{ij}}{n} \quad (1)$$

where n = number of samples of *i*th variable.

Then, calculate the standard deviation for each variable per Equation 2:

$$s_i = \sqrt{\frac{\sum_{j=1}^{n} (X_{ij} - \bar{x}_i)^2}{n-1}} \quad (2)$$

Normalize each variable per Equation 3:

$$Z_{ij} = \frac{(X_{ij} - \bar{x}_i)}{s_i} \quad (3)$$

Then, calculate the transpose of Z_{ij}, Z^T_{ij}.

Calculate the correlation matrix per Equation 4:

$$C_{ij} = \frac{\sum_{m=1}^{n} (Z_{im} Z_{jm})}{n-1} \quad (4)$$

Calculate the inverse of correlation matrix, C^{-1}_{ij}.

Calculate the MD per Equation 5:

$$MD_j = \frac{1}{k} Z^T_{ij} C^{-1} Z_{ij} \quad (5)$$

where k = number of variables.

2.2 *Stage 2: Validation of measurement scale*

In this stage is necessary to identify abnormal samples of each variable in order to calculate the abnormal Mahalanobis Distance (MD_A). The samples of abnormal variables are normalized using the mean and standard deviation of the respective normal variable. The normal correlation matrix is also used. Then, the MD_A is calculated. If the values of MD_A are greater than MD, it´s an indication that values used in the construction of MD should be right.

2.3 *Stage 3: Optimization*

To optimize the number of useful variables is necessary to construct a two-level OA. The variables will be placed in a row and will have two levels in each column. Level-1 indicates that the variable should be used in construction of Mahalanobis space and Level-2 means that it should not be used. Then, the MD_A will be recalculated following the OA.

The signal-to-noise ratio (η) is calculated per Equation 6.

$$\eta = -10 \log \left[\frac{1}{t} \sum_{j=1}^{t} \frac{1}{MD_j} \right] \quad (6)$$

where t = number of abnormal conditions; MD_j = the MD of the *i*th abnormal condition.

The gain in signal-to-noise ratio is calculated per Equation 7. If the value is positive, the variable is stored. If not, it is removed from the analysis.

$$Gain = \overline{S/Nratio}_{level-1} - \overline{S/Nratio}_{level-2} \quad (7)$$

2.4 *Stage 4: Diagnose*

As the optimization is done on stage 3, the Mahalanobis space will be reconstructed and the diagnosis process will be performed.

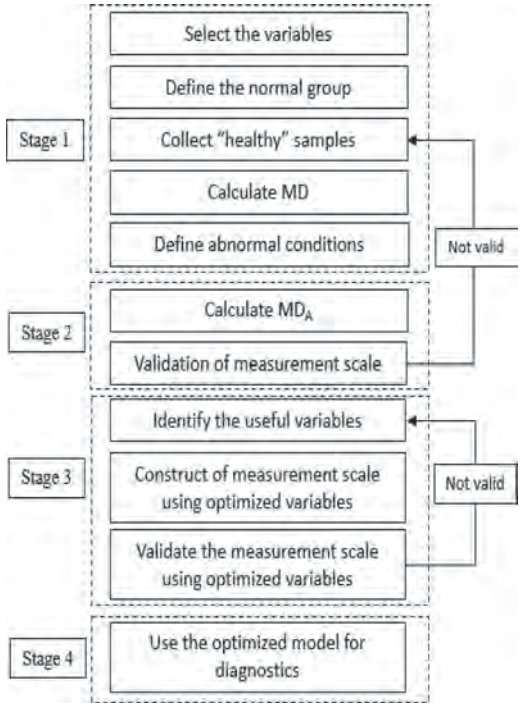


Figure 1. MTS process.

The Figure 1 summarize the four presented stages.

3 UNBALANCE

Unbalance is a centrifugal force produced by a difference between the mass center and geometrical center position of moving parts. This difference is related to manufacturing process that cannot produce a rotor with same geometrical and mass center. When the rotor is in operation, it tries to rotate around its geometrical center, but the presence of this not centered mass produces a centrifugal force resulting in unbalance vibration (Sun et al. 2017).

A strong radial vibration at the fundamental frequency is unbalance characteristic diagnostic symptom. As the response amplitude is related to the square of the rotational speed, unbalance is a dangerous condition in machinery that runs at high rotational speeds. At variable speed machines, the effects of unbalance will vary with the shaft rotational speed. At low speed machines, however, the high spot (location of maximum displacement of the shaft) will be at the same location as the unbalance. At increased speeds, the high spot will lag behind the unbalance location.

To analyze the vibration signals, and also the unbalance, a useful tool is the Fast Fourier Transform (FFT). The FFT is a mathematical algorithm that converts a periodic signal from the time domain to the frequency domain. So, when analyzing the spectrum, every periodic revolution will be presented by a peak and a respective frequency in the frequency domain (Al-Badour et al. 2011).

Every rotating machine has an unbalance degree, but when it is outside of limit, it becomes a problem (Walker et al. 2013). Unbalance is generally identified by a high synchronous vibration in frequency (1X) or time domain that, in excessive cases, may cause fatigue, internal rubs and damage bearings and seals (Bently & Hatch 2002).

Also, there are other malfunctions that produces a high synchronous vibration (for more details verify Bently & Hatch 2002), but when analyzing unbalance, it is important to remove the runout of the signal, because it can influence the 1X sensors readings and disturb the diagnosis (Bently & Hatch 2002).

Runout is a false vibration measurement which occurs when the rotor rotates its geometric center with no displacement of its center line. The runout can have two different sources, mechanical (defects on the shaft surfaces in the vibration probes area) and electrical (variance on electrical conductivity and permeability in the vibration probes area) (Bently & Hatch 2002).

Normally, there are four radial vibration probes that two are located on the drive (Fig. 3) and two on non-drive (Fig. 4) machine shaft ends mounted with 90° between each other. The recorded vibration data consists in the shaft radial displacement and phase angle measured by each probe. Also, a rotor orbit with 1X filtered signal is presented in Figure 5, the major axis is obtained by combination of the two probes displacement and phase angle.

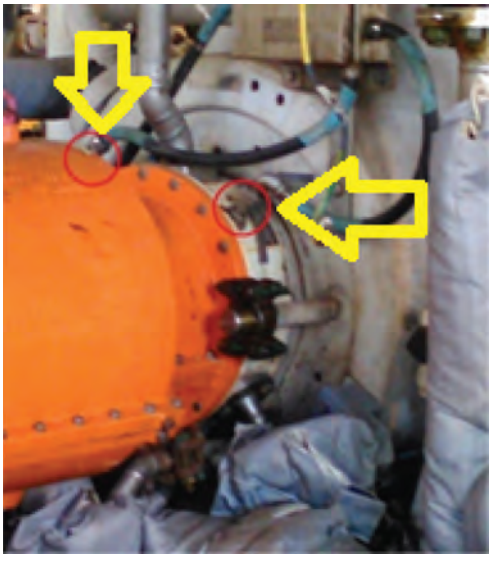


Figure 3. Drive end radial vibration probes.

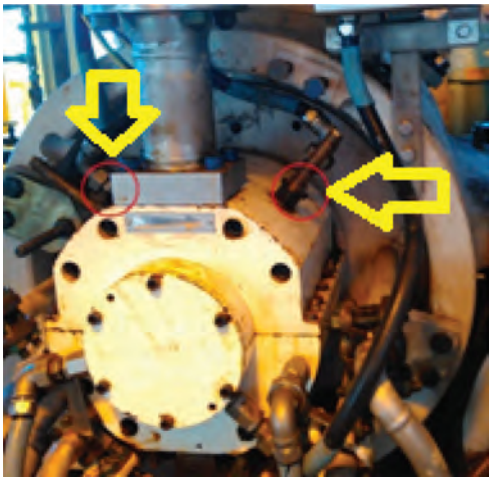


Figure 4. Non-drive end radial vibration probes.

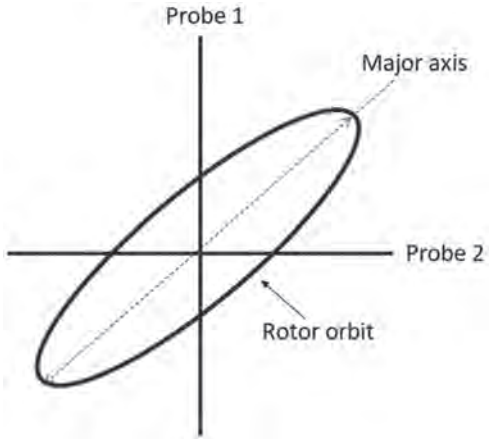


Figure 5. Rotor orbit – 1X filtered signal.

4 ROTORDYNAMICS MODELING

As described by Vance et al. (2010), rotodynamics modeling using beam elements models are generally adequate for modelling rotors. The main objective is to calculate the beam deflection to verify the internal clearances between the rotor and static parts and the rotor stability.

To construct this model, beam elements are defined with geometrical and material characteristics. Then, applying the theory of finite elements formulations, the stiffness, damping and inertia matrices are calculated.

Also, the machine´s vibrational performance is produced by the relationship between the bearings and rotor system dynamic properties. The modeling program used in this paper simulates the bearing stiffness and damping properties variation with the increase or decrease of rotor speed solving the Reynold´s equations. For more details, refer to He et al. (2005).

As described above, rotodynamic modeling is big deal, so the program called RotorLab developed by the Rotating Machine and Controls Laboratory (ROMAC) was chosen to simulate the rotor unbalance response. ROMAC is an industrial consortium at the University of Virginia with over 40 years of experience (Weaver et al. 2017).

Finally, the model was validated by the authors reproducing the API lateral analysis developed by the machine manufacturer.

5 UNBALANCE IDENTIFICATION USING VIBRATION ANALYSIS

As mentioned before, in order to compare and validate MTS results, a more classic approach, based in vibration analysis is considered. As widely reported in the literature, unbalance is strongly associated with 1X synchronous vibrations. Thus, considering the major axis displacement signal filtered at 1X the rotational speed in both bearings can give a clue about unbalance magnitude. Generally, it is expected that the greater the unbalance, the greater the vibration measured in 1x.

However, depending on the position and magnitude of the considered unbalance, the system modal response can assume different configurations and, therefore, the displacement amplitude in the bearings may not follow the logic described above.

In this case, an approach considering the virtual work (Lalanne & Ferraris 1998) of the forces acting on the shaft is applied. The bearings stiffness and damping terms are considered known and the shaft bending influence is neglect.

In fact, the bearings are considered symmetrical —since the unbalance response in both directions are almost identical—and without cross-coupled forces—since the machines has tilting pad bearings (He et al. 2005).

Considering just the forces acting on the shaft due the bearings stiffness, the virtual work can be presented per Equation 8.

$$\partial W_K = -k_b u \delta u - k_b w \delta w \tag{8}$$

where k_b is the bearing's stiffness and u and w are, respectively, the displacements in both directions.

From Equation 8 can be easily concluded that the maximum potential energy related to the shaft displacement, considering just the 1X component, occurs in the major axis direction.

The idea is to associate the energy level, considering the scalar value obtained for both bearings, with the rotor level of unbalance: the greater the unbalance, the greater the potential energy.

As the bearings' stiffness is a constant under the same speed conditions, the evaluated potential energy depends only on the major axis displacement ($d_{\max}$), as presented by Equation (9).

$$\frac{U_k}{k_b} = \frac{\left(d_{\max}\right)^2}{2} \tag{9}$$

6 CURRENT METHOD OVERVIEW

As described in Figure 6, a real tested machine will be considered as reference to construct the Mahalanobis space. Then, this machine will be modeled using the RotorLab software with two cases which produces the same vibration level as noticed in the tested machine (550 g.mm at the midspan and 275 g.mm in two points of the rotor and out of phase). This simulation will be done in order to

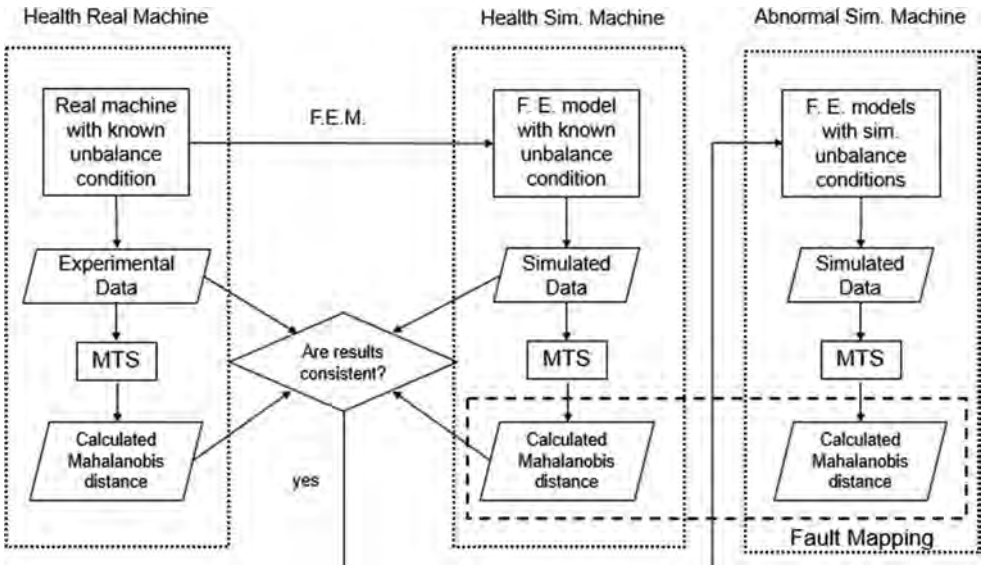


Figure 6. Method overview.

verify the similarity between the real machine and simulated data. Finally, more four unbalance cases will be simulated to calculate the abnormal Mahalanobis distance.

7 APLLYING MTS

As described on item 2, the MTS will be applied as follows.

7.1 *Stage 1: Mahalanobis space construction*

The first step is the Mahalanobis space construction which requires a "health" machine as standard. The activities developed in this step can be summarized as follows and will be detailed in the sequence.

- Collect the "health" vibration data with runout compensation of the four vibration probes during the mechanical running test (MRT);
- Filter the 1X signal (synchronous vibration) of each probe;
- Use the major axis displacement of drive and non-drive to construct the Mahalanobis space (there will be two variables to construct the Mahalanobis space);

The considered "health" machine was one that was tested and approved by API 617 7th edition MRT.

This test is conducted in a test bench with low pressure and consists in accelerating the compressor from zero to maximum continuous speed (MCOS) in increments of 10% of MCOS and run at MCOS until bearing temperatures and shaft vibrations have stabilized. Then, the speed is increased to the trip speed and kept at this level by fifteen minutes. Finally, the machine is decelerated to the MCOS and run in this condition during four hours, a schematic can be seen in Figure 7 (API 617 7th edition).

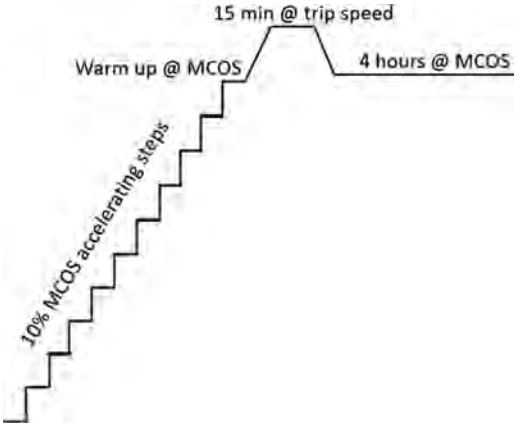


Figure 7. Mechanical running test schematic.

To construct the Mahalanobis space, vibration data were collected during the four hours operating at the maximum continuous speed with runout compensation (Fig. 8). Then, only the synchronous vibration signal (1X) of each probe was kept because unbalance is identified by high synchronous vibration.

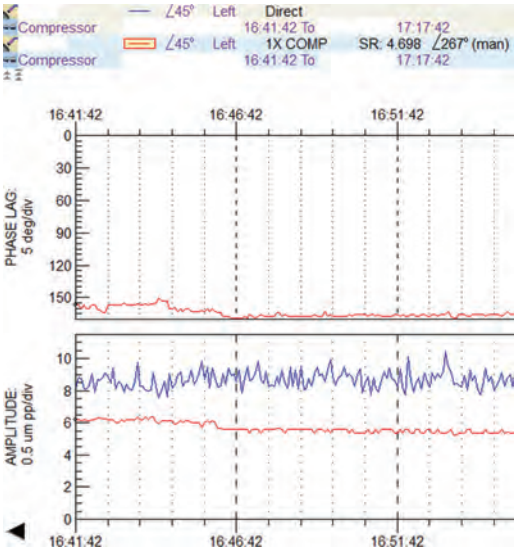


Figure 8. Trend of a radial vibration probe during the MCOS.

Table 1. Calculated Mahalanobis distance.

Minor value	−6,3
Major value	10,5
Standard deviation	0,7

Finally, the Mahalanobis space was constructed using the major axis displacement with the 1X filtered signal of drive and non-drive end, this choice eliminates the influence of phase angle.

Applying the Equations 1 to 5, the calculated MD is presented on Table 1.

7.2 *Stage 2: Validation of measurement scale*

At this point is necessary to identify abnormal behaviors to validate the measurement scale. The activities developed in this step can be summarized as follows and will be detailed in the sequence.

- Simulate four rotor unbalance cases;
- Verify the simulated vibration displacement and phase angle at the vibration probes;
- Analyze the distribution and variance of the major axis vibration data obtained during the MRT;
- Use the simulated unbalance rotor major axis displacement as the mean and same variance obtained in the previous step to create an abnormal vibration sample.

The six unbalance responses were simulated using the program RotorLab developed by ROMAC, two points following the API 617 7th edition paragraph 2.6.4 (550 g.mm at midspan to excite the first bending mode and two 275 g.mm 180° out of phase placed on the shaft maximum displacement to excite the second bending mode), one as per paragraph 2.6.4 (1100 g.mm at the coupling) and one with the same unbalance magnitude (1100 g.mm), but at the thrust disk. The other two are 1500 g.mm at the coupling and thrust disk.

At MCOS, the two first points presented the same vibration level as the "health" machine at the vibration probes. It shows that depending where the unbalance is placed and the vibration probes are located, this malfunction will not be identified. A similar case was identified in the company where one of the authors works, a centrifugal compressor was operating with a broken impeller without alarming the radial vibration probes, but presented a degraded performance and this root cause was identified during the overhaul.

The third point presented a major axis vibration of 26,1 microns pk-pk at drive end and 3,5 microns pk-pk at non-drive end. The fourth point presented a major axis vibration of 3,1 microns pk-pk at drive end and 25 microns pk-pk at non-drive end. The fifth point presented a major axis vibration of 31,4 microns pk-pk at drive end and 3,5 microns pk-pk at non-drive end. The sixth point presented a major axis vibration of 4,3 microns pk-pk at drive end and 34,3 microns pk-pk at non-drive end. A resume can be seen on Table 2.

Although the unbalance simulation produces a discrete result, the MD_A needs a sample of vibration measurements. So, it was decided to verify the major axis vibration statistic distribution on drive and no drive-end collected during the four hours mechanical running test. The drive end followed a Weibull distribution and the non-drive a mixed Weibull distribution as presented on Table 3 and Table 4, respectively.

Table 2. Unbalance response at MCOS.

	Unbalance	Position	Vibration
1	550 g.mm	Midspan	2,4 µm at drive end 2,3 µm at non-drive end
2	2 × 275 g.mm	180° out of phase	1,9 µm at drive end 8,4 µm at non-drive end
3	1100 g.mm	Coupling	26,1 µm at drive end 3,5 µm at non-drive end
4	1100 g.mm	Thrust disk	3,1 µm at drive end 25 µm at non-drive end
5	1500 g.mm	Coupling	31,4 µm at drive end 3,5 µm at non-drive end
6	1500 g.mm	Thrust disk	4,3 µm at drive end 34,3µm at non-drive end

Table 3. Weibull distribution on drive end.

F(x)	R^2
$1-\exp[-(x/11{,}3)^{25{,}7}]$	0,864

Table 4. Weibull distribution on non-drive end.

F(x)	R^2
$0{,}459\{1-\exp[-(x/5{,}15)^{1395{,}8}]\}$	0,798
$+0{,}355\{1-\exp[-(x/5{,}41)^{72{,}9}]\}$	0,765
$+0{,}186\{1-\exp[-(x/5{,}97)^{23{,}57}]\}$	0,848

As described by Lewis (1994), the Equations 10, 11 and 12 represent the Weibull distribution equations. According to Kececioglu & Wang (1998), the mixed Weibull distribution is described in Equation 13 and 14.

$$F(x)=1-\exp[-(x/\theta)^m] \tag{10}$$

where $F(x)$ = cumulative distribution function; θ = scale; and m = shape parameter.

$$\mu=\theta^*\Gamma(1+1/m) \tag{11}$$

where μ = mean.

$$\sigma^2=\theta^2[\Gamma(1+2/m)-\Gamma(1+1/m)^2] \tag{12}$$

where σ = variance.

$$f(x)=p^*f_1(x)+q^*f_2(x) \tag{13}$$

where $f(x)$ = probability density function; $f_1(x)$ and $f_2(x)$ = probability density function of two different subpopulation.

$$p+q=1 \tag{14}$$

where p and q = correspondent mixing weight of each subpopulation $f_1(x)$ and $f_2(x)$.

Then, per Equations 11 and 12, new parameters m and θ were calculated using the same variance (σ) of the measured sample with the mean (μ) as the simulated major axis vibration.

To create the new sample (the term x in Equation 10), aleatory numbers between 0 and 1 were generated for the term $F(x)$.

Finally, the MD_A is presented in Tables 5 and 7 for coupling unbalance and in Tables 6 and 8 for thrust disk unbalance. Both show higher values than the MD validating the measurement scale.

Table 5. Calculated MD_A for 1100 g.mm coupling unbalance.

Minor value	631,3
Major value	1.106,2
Standard deviation	60,8

Table 6. Calculated MD_A for 1100 g.mm thrust disk unbalance.

Minor value	4.254,8
Major value	4.841,9
Standard deviation	73,4

Table 7. Calculated MD_A for 1500 g.mm coupling unbalance.

Minor value	1.667,3
Major value	2.224,1
Standard deviation	77,9

Table 8. Calculated MD_A for 1500 g.mm thrust disk unbalance.

Minor value	10.886,1
Major value	12.175,0
Standard deviation	205,9

Table 9. MS – 550 g.mm at midspan.

Minor value	–16,4
Major value	22,5
Standard deviation	0,7

Table 10. MS – 2 × 275 g.mm 180° out of phase.

Minor value	–11,1
Major value	15,2
Standard deviation	0,7

As described previously, the cases 1 and 2 described in Table 2 presented a low vibration level. So, two health vibration samples were generated following the presented Weibull distribution and two new Mahalanobis spaces were calculated in order to compare the simulated data with the tested one. Comparing Table 1 with Tables 9 and 10, it can be seen that the simulated and tested date have similar results.

7.3 *Stage 3: Optimization*

As mentioned in item 2.3, the optimization aims to reduce the number of monitored variables. The Gain was calculated per Equation 7 per each variable (drive and non-drive end major axis vibration). The results of optimization considering the 1100 g. mm coupling unbalance and 1100 g.mm thrust disk unbalance are presented in Tables 11 and 12, respectively.

Following the optimization results in Table 11, only the drive end vibration shall be maintained, this result is consistent because when the unbalance is placed at the coupling for this machine, the high vibration signal is noticed only in the drive end probes. The same interpretation is noticed in the Table 12, at this point, only the non-drive end vibration shall be maintained, because an unbalance placed in the thrust disk, creates a high vibration signal only in the non-drive end probes.

So, in order to use the same Mahalanobis space, the optimization stage was removed from the analysis.

Table 11. Variables gain for coupling unbalance.

Vibration major axis	Gain	Result
Drive end	12,8	Keep variable
Non-drive end	–0,2	Not keep variable

Table 12. Variables gain for thrust disk unbalance.

Vibration major axis	Gain	Result
Drive end	–0,3	Not keep variable
Non-drive end	12,3	Keep variable

Table 13. MTS diagnoses.

	MD	Malfunction
1	–6,3 to 10,5	Normal operation
2	631,3 to 1.106,2	1100 g.mm Coupling unbal.
3	4.254,8 to 4.841,9	1100 g.mm Thrust disk unbal.
4	1.667,3 to 2.224,1	1500 g.mm Coupling unbal.
5	10.886,1 to 12.175,0	1500 g.mm Thrust disk unbal.

Table 14. Energy approach diagnoses.

	U_k/k_b	Malfunction
1	76,13	Normal operation
2	347,49	1100 g.mm Coupling unbal.
3	317,56	1100 g.mm Thrust disk unbal.
4	504,89	1500 g.mm Coupling unbal.
5	597,05	1500 g.mm Thrust disk unbal.

7.4 *Stage 4: Diagnose*

The diagnose is summarized in Table 13. There, it can be seen that the malfunctions have different MD from the normal operation showing the method´s effectiveness.

Considering the proposed method presented in section 5, in which the potential energy associated with the major axis displacement at 1X is related to the unbalance amount, the effectiveness of the MTS method can be confirmed. In Table 14 are presented the results considering the energy approach.

It can be seen, as in MTS method, that the malfunctions are also clearly associated with different energy levels.

8 CONCLUSION

Mahalanobis Taguchi strategy showed to be a useful tool when analyzing multivariate. This characteristic is a differential to produce a good malfunction diagnosis related to complex turbomachinery, which have many installed sensors.

In fact, considering MTS and the energy approach results, in both cases the malfunctions are clearly detected. Nevertheless, because the energy method is based on the vibration amplitude in the bearings, the results found for the analyzed cases demonstrate a much greater sensitivity to the unbalance value than to the unbalance position. In another hand, the MTS method presents different results for all cases, being sensitive to both the unbalance position and magnitude in the analyzed cases.

Considering these results, it can be said that MTS has the potential for a more complete analysis than methods based on more traditional approaches, such as vibration amplitude analysis, especially in cases with many monitored points.

However, some care shall be taken during the optimization step, because as seen in item 7.3, if the optimization were done, the different diagnosis would not be possible using the same Mahalanobis space.

Also, the first variables selection to construct the Mahalanobis space is not an easy task, some previous troubleshooting knowledge is necessary to do it, or the validation of measurement scale iterative procedure may take a long time.

This method has great potential to be applied in standard machines, because as they have similar behavior, more abnormal conditions may be identified for the same Mahalanobis space.

As well, the runout compensation is an important procedure to be done, mainly when construction the Mahalabonis space, because the “healthy”

machine presents vibration signals levels that runout can disturb them.

Finally, independently of the monitoring system, all malfunctions won´t be suitable to be identified, as described in item 7.2, two types of unbalance presented the same vibration level as the "healthy" machine on drive and non-drive end radial vibration probes.

ACKNOWLEDGEMENTS

The authors thank Petrobras for using the RotorLab software for academic purpose and CNPq for the financial support.

REFERENCES

Al-Badour, F.; Sunar, M. & Cheded, L. 2011. Vibration Analysis of rotating machinery using time-frequency analysis and wavelet techniques. *Mechanical System and Signal Processing* volume 25: 2083–2101.

API Standard 617 7th Edition. 2009. Axial and Centrifugal Compressors and Expander-compressors for Petroleum, Chemical and Gas Industry Services. Washington D.C.: American Petroleum Institute.

Bently, D.E. & Hatch, C.T. 2002. *Fundamentals of Rotating Machinery Diagnostics*. Minden: Bently Pressurized Bearing Press.

Cudney, E.G.A.A.E.A. 2015. Mahalanobys-Taguchi system: a review. *International Journal of Quality & Reliability Management* volume 32 (3): 291–307.

He, M.; Cloud, C.H.; Byrne, J.M. 2005. Fundamentals of Fluid Film Journal Bearing Operation and Modeling. *Proceedings of the Thirty-Fourth Turbomachinery Symposium*: 155–175.

He, Y.; Shi, L.; Shi, Z. & Sun, Z. 2017. Unbalance Compensation for HTR-10GT: A Frequency-Domain Approach Based on Iterative Learning Control. *Science and Technology of Nuclear Installations* (ID 3126738): 1–15.

Jin, X. & Chow, T.W.S. 2013. Anomaly detection of cooling fan and fault classification of induction motor using Mahalanobis–Taguchi system. *Expert Systems with Applications*. Volume 40 (issue 15): 5787–5795.

John, B. 2014. Application of Mahalanobis-Taguchi system and design of experiments to reduce the field failures of splined shafts. *International Journal of Quality & Reliability Management* volume 31 (issue 6): 681–687.

Kececioglu, D.B. & Wang, W. 1998. Parameter Estimation For Mixed- Weibull Distribution. *IEEE PROCEEDINGS Annual RELIABILITY and MAINTAINABILITY Symposium*: 247–252.

Lalanne, M. & Ferraris, G. 1998. *Rotordynamics Prediction in Engineering*. Second edition. Hoboken: John Wiley & Sons, Inc.

Lewis, E.E. 1994. *Introduction to Reliability Engineering*. Evanston: John Wiley & Sons, Inc.

Soylemezeglu, A.; Jagannathan, S. & Saygan, C. 2011. Mahalanobys-Taguchi System as a Multi-Sensor Based Decision Making Prognostics Tool for Centrifugal Pump Failures. *IEEE Transactions on Reliability* volume 60 (4): 864–878.

Vance, J.; Zeidan, F. & Murphy, B. 2010. *Machinery Vibration and Rotordynamics*. Hoboken: John Wiley & Sons, Inc.

Walker, R.; Perinpanayagam, S. & Jennions, I. Rotordynamic Faults: Recent Advances in Diagnosis and Prognosis. *International Journal of Rotating Machinery* volume 2013 (ID 856865): 1–12.

Weaver, B.; Tsukuda, T.; Rizvi, S.A.A.; Schwartz, B.; Nichols, B.; Griffin, D; Branagan, M.; Fittro, R.; Lin, Z. & Wood, H. 2017. Experimental Measurements of Turbomachinery Rotordynamics, Component Performance, and Dynamic Control at ROMAC – A Review. *Journal of the Gas Turbine Society of Japan*: 1–8.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Optimization of periodic inspection time of sis subject to a regular proof testing

H. Srivastav, A.V. Guilherme, A. Barros & M.A. Lundteigen
Department of Mechanical and Industrial Engineering, NTNU, Norway

F.B. Pedersen & A. Hafver
Group Technology and Research, DNV GL, Høvik, Norway

F.L. Oliveira
R&D Center, DNV GL, Rio de Janeiro, Brazil

ABSTRACT: Periodic testing is a method to ascertain the availability of Safety Instrumented Systems (SIS). These systems are generally passive and are activated only on demand. Testing is then required to diagnose their current state and to take the corresponding maintenance action. However, the testing procedure can provoke damage on some units of the SIS (especially the mechanical parts) and the system as a whole becomes more prone to failures. This situation is currently not well covered by standards under the so-called umbrella of imperfect testing. The decision maker must in practice come across to an optimization problem where the objective is to determine the optimal compromise between an accurate diagnostic of the current system state (high tests frequency) and the possible failures or degradation provoked by the testing procedure itself. The commonly used criteria to assess the performance of SIS are all related to the mean downtime of the SIS between two tests. The IEC 61508 provides subsequent analysis for multi-unit SIS when all the units are supposed to follow exponential lifetime distributions. It cannot be applied in this case as some parts of the system have a time varying failure rate which can increase after every test. We propose the use a of Markov process to model the degradation of the mechanical parts upon test and possible preventive maintenance after testing. Since the degradation due to tests is experienced at deterministic dates, we use the modelling framework of multiphase Markov processes to calculate the mean downtime. The paper is focused on explaining the optimization problem between the frequency of testing versus PFD_{avg} and find out the optimum frequency through simulations

1 INTRODUCTION

A Safety Instrumented System (SIS) is often used to detect hazardous events and to mitigate their consequences at facilities and plants that produce or handle hazardous substances, like e.g. hydrocarbon fluids and gases. Due to their criticality, they must obey to regulatory requirements and international standards on safety. IEC 61508 (1998) and related standards (such as IEC 61511 (2002) for the process industry sector) are key in framing the design and operation of SIS. One important requirement mandated by these standards is the need to verify, by quantitative analysis, that the safety performance is adequate in light of risk acceptance criteria. Most safety functions implemented by a SIS, the so-called Safety Instrumented Functions (SIFs), are seldom demanded as the normal operation is managed by a dedicated control system. According to the mentioned IEC standards, the SIFs are classified as operating in the low demand mode.

This means that the SIFs are passive most of the time and are supposed to act only when needed ("on demand"). The reliability of low demand SIFs is measured by the average probability of failure on demand (PFD_{avg}). PFD_{avg} is calculated over a time interval between two proof tests and corresponds to the mean downtime per unit of time between proof tests. The same measure is also used to express the reliability requirement for the each SIF„ but then the associated required value is derived on the basis of a risk analysis (Jin et al. 2012). IEC 61508 suggests four levels of safety integrity levels (SIL), each giving a specified range of PFDavg. For example, a SIF with a SIL 2 requirement must demonstrate that the PFD_{avg} is within $10^{-3}$and 10^{-2}.

The PFD_{avg} can be quantified using different reliability models. These models are based on assumptions and simplifications and in some situations they can lead to different results, depending on the dominating contributing factors. Lowdemand

SIS are periodically tested (proof tests) in order to confirm that they are able to act on demand. Length of intervals between such tests is an important contributor to PFD_{avg}. Normally, it is assumed that the proof tests are perfect, and that the equipment is restored to an to as-good-as-new condition (Shao-Ming et al. 1994). These assumptions imply that the proof tests are carried out in a manner and under conditions which are similar to a real demand, so that all dangerous failure modes,- i.e. failure modes that result in a failure to carry out the SIF, are revealed.The assumptions also imply that no degradation is experienced by the SIS due to the test itself (a non-destructive test). However, in reality, proof tests may not beperfect, and the equipment may degrade from exposures that are applied during the tests. The latter example is also identified by Brissaud et al. (2010). Rausand (2014a) gives one practical example on how the proof test can degrade a Downhole Safety Valve (DHSV) installed in to protect against releases from oil and gas wells. The DHSV is exposed to harsh conditions when operated (due to high pressures drop and in some cases high temperature). A perfect proof test, would imply that the DHSV is closed with full flow from the well (which would be the real demand situation). However, this type of exposure is known to degrade the performance of the DHSV, and the proof test is therefore carried out under non-perfect/imperfect test condtions by closing DHSV with downstream valves already closed. Still, it is interesting to understand better the impact of perfect versus non-perfect/imperfect test conditions. One approach has been suggested by Oliveira et al. (2016), where an additive test-step varying (ATSV) model was elaborated to reflect the increment of the failure rate after each proof test in a blowout preventer (BOP) system. Yet, it is still not clear how to implement the full effect of degradation for the quantification of PFD_{avg}. A review of the modelling framework was performed by Rouvroye & Brombacher (1999) and Bukowski (2005) and both promoted the use of Markov processes when other states than functioning and failed are to be included.

The objective of this paper is to demonstrate the implementation of the Markov process to model the combined effects of degradation due to equipment wear out (aging) and the exposure from the proof test. A simple homogeneous Markov process cannot be used,since the transition rates will change after each proof test. Instead, a multiphase Markov approach is suggested. This method was applied in Strand and Lundteigen (2015) to assess the BOP reliability and also in Innal et al. (2016) to establish new generalized formulas with repair time. Compared to simple Markov processes, multiphase Markov processes allows one to take into account changes of the transition rates at deterministic time points (Wu et al. 2018). The paper is organized as follows: Section 2 provides the problem statement and assumptions. The model is discussed in section 3, within a multiphase Markov framework. Section 4 describes the model implementation in terms of discrete event simulation and Monte Carlo simulations. The last section is devoted to numerical results and the consequent optimization problem.

2 MODELLING FRAMEWORK AND MODEL ASSUMPTIONS

PFD_{avg} is defined as Rausand (2014b):

"..If a demand of safety function of the item occur at a random time in future, the PFD_{avg} is the average probability that the item is not able to react and perform its safety function in response to demand.."

Theoretically, PFD_{avg} value stems from the risk analysis. For practical purposes, it is estimated on the basis of the reliability model of the SIF. In general, an estimator for PFD_{avg} ($\widehat{PFD_{avg}}$) can be interpreted as long run average value of unavailability, it can be defined as:

$$\widehat{PFD_{avg}} = \frac{1}{n}\sum_{k=1}^{n}\int_{(k-1)\tau}^{k\tau}\frac{U(t)}{\tau}dt$$

where:

- PFD_{avg} Probability of failure on demand on average
- n = Total number of inspection performed
- τ= Duration between two consecutive inspection
- $U(t)$ = Unavailability of the system at t

Inspection is an integral part of the proof test which reveals about the state of the system at the time of proof test. For all calculations, frequency of inspection is equal to frequency of proof test performed. In this situation $\widehat{PFD_{avg}}$ is proportion of time on average that the multiphase Markov process spends in the failed state. It is the dangerous failure rate that is considered in the calculation of $\widehat{PFD_{avg}}$, i.e. the failures that can prevent the SIF from functioning on demand.

The modelling framework to model this problem is described hereafter.

2.1 *Modelling framework*

There are basically two different mindsets for modelling degradation due to equipment wear out (aging) and degradation due to proof test. One mindset is more inherited from Reliability theory: the main idea is to model the degraded unit by a binary random variable moving from working

state to failed state and to consider that the transition rate between these two states will increase with time or with the number of tests experienced by the unit. In other words, the unit has a lifetime law with an increasing failure rate which is a function of the number of tests. Another mindset is more applied for people working in the framework of maintenance optimization. The unit is modelled by a random variable with more than two states. The state space can be a discrete finite space, an infinite discrete one or a continuous one. The main idea is that there exists intermediate states between the new one and the failed one. All the intermediate states can be considered as working states but with possibly degraded performances and they are taken as a health indicator of the system. They often correspond to degradation phenomena or symptoms that can be monitored, diagnosed and used as a decision indicator to trigger preventive maintenance actions. The advantage of such models is that

- We can make correspondence between degradation phenomena and the performance of the system (here 1-PFD).
- We can use the intermediate states to optimize and define preventive condition-based maintenance

However, if expert judgments can be relevant enough to define the number and the nature of intermediate states, the law of the sojourn time in every single state may be difficult to estimate. A model relying only on lifetime law and a binary random variable may be then more reasonable.

Most of the existing models that are described in the introduction are inherited from Reliability theory. The calculation of the PFD for SIS is mainly based on binary random variables. In this paper, we want to explore the use of intermediate states in a specific context when the tests have a negative impact on the system condition. We want to investigate such a framework because

- The literature, guidelines and practices related to negative impact of testing should be linked at some point to the identification of some degradation mechanism.
- This seems to be a good way to go ahead and prepare the future for condition-based maintenance and optimal use of condition monitoring.

As a preliminary study, we propose a model with two intermediate states. This number is arbitrarily chosen and we do not investigate any preventive maintenance. We only aim at showing that there is a trade off between the negative effect of tests (pushingthe system randomly into more degraded states) and the added value performing more tests to detect failures earlier.

Equipment wear out is modelled by a finite number of intermediate degraded states between the new state and the failed one. Degradation due to proof test is modelled by an increase of the transition rates between two states at inspection time. In addition, direct transitions are possible from any functioning state to the failed one: they model sudden failures that are not due to wear. Since the unit is passive, all the failures are undetectable without testing, whatever the failure mode is. At last, in order to develop further analytical formulations, we chose a Markovian framework. Because the transition rates are changing at inspection times, we refer it as a Multiphase Markov process. The current paper is only devoted to Monte Carlo simulations in order to demonstrate the relevance of the problem statement and the possible trade off that arises due to the negative effect of testings. Analytical formulations seems to be tractable but are left for further work.

2.2 *Assumptions*

Modelling degradation using Multiphase Markov process, we have used following assumptions:

- In general, we can consider that a SIF equipment is exposed to two types of failures:
  - Dangerous detected (DD) failures, i.e. the dangerous failures revealed by online diagnostics.
  - Dangerous undetected (DU) failures, i.e. the dangerous failures that are not DD and which are to be revealed by regular proof tests.
- For the sake of simplicity to begin with the modelling, we only consider the effect of DU failures in our analysis, since the equipment focused in our study (valves) have no or very limited facilities for diagnostics. However, effect of DD failures,for modeling purposes beyond equipment type in our study, will be considered in the future paper. From now, when we use the term “detected” or “detectable” , it is used to denote DU failures that are revealed by the proof test, in light of the real (non-perfect/imperfect) test conditions.
- DU—failures are of two types: they can be sudden or they can be due to a progressive degradation process named hereafter aging. Sudden failures are modelled by a failure rate λ_u, and aging is modelled by several intermediate states (degradation levels) between new state and failed one, with associated transition rates. Whatever the failure mode is, the system will stay in failed state until the next inspection, and then the system is repaired as per the chosen maintenance policy.

- There are 4 degradation levels: A, B, C, D. These are the states of a Markov process. (A: System working with no degradation, B: System working with degradation of system of level 1, C: System working with degradation of level 2, D: Failed)
- In our model the following instantaneous transitions are possible:
  - System can always jump to next higher state of degradation due to effect of aging.
  - System can always jump to failed state due to sudden DU failures.
  - System can not go to lower degraded state until the maintenance is performed.
- Instantaneous transitions rate for the multiphase Markov process are represented in the Figure 1.
- In the Figures above represents the effect of aging on the system, which changes every time when a proof test is performed on the system. We consider that the proof test has a negative effect on the system condition (shock leading to extra stress) and this negative effect increases the aging transition rates. The modelling of impact of negative effect of testing is done through the following model.

$$\lambda_a(t_0^+) = \begin{cases} 1.01 * \lambda_a(t_0^-) & \text{Current State A} \\ 1.03 * \lambda_a(t_0^-) & \text{Current State B} \\ 1.05 * \lambda_a(t_0^-) & \text{Current State C} \end{cases} \quad (1)$$

We assume here that a proof test is performed at $t = t_0$ and the current state is the state of the system at $t = t_0$.

  - The underlying idea behind this modelling is to show that the negative impact of the proof test increases with the degradation of the system

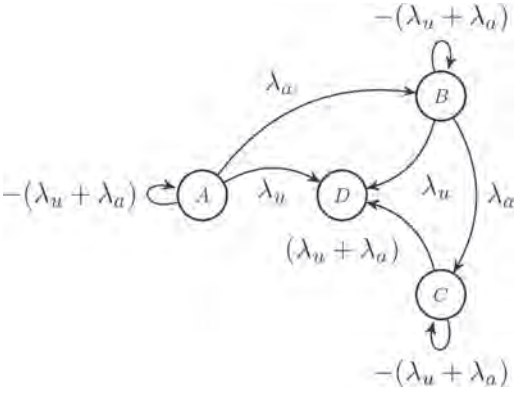


Figure 1. Instantaneous transition rates for the multiphase Markov process.

- Between two consecutive proof tests λ_a and λ_u remains constant.
- When a failure is detected after the proof test, we assume that the mean time to repair the system is negligible.

3 METHODOLOGY

The multiphase Markov process was analyzed using discrete event simulation and exponential distribution for the time spent in each state. System starts in state, degradation time (T_d) and failure time (T_f) are sampled from the exponential distribution of the respective parameters $\lambda_a(t)$ and $\lambda_u(t)$. Then based on the minimum of (T_d, T_f, τ) the next state of the process was chosen. Some specific decisions were made for the modeling:

- If system goes to a failed state (state *D*), the unavailability is calculated by measuring the time spent in the state *D* by the system. On inspection the maintenance action is taken and process is re-initiated.
- If the minimum is τ, then the system stays in the same state for the duration between two consecutive proof tests. Then the inspection is performed and we repeat the process with the increased $\lambda_a(t)$.
- If system goes to more degraded state, then the T_d',T_f', are again sampled from the corresponding exponential distributions. Now, the minimum is compared between (T_d',T_f', $\tau - T_d$). And the process repeats itself until system goes to failed state. Once the system fails, the unavailability is calculated,the maintenance action is taken and process is re-initiated.

The following maintenance policies were proposed when the system was found to be in the failed state on inspection:

- As-good-as-new (AGAN): System is reset to new state (*A*) and the failure rate of the system is reset to $\lambda_a[i + 1] = \lambda_a[1]$, ie we consider that system is as-good-as-new when we take away the effect of aging after maintenance of the system
- As-bad-as-old (ABAO): On maintenance, the new state of the system is set to *C* and the failure rate of the system is reset to $\lambda_a[i + 1] = \lambda_a[i]$

4 RESULTS AND DISCUSSION

Recall that the PFD_{avg} is the performance measure. Simulations were performed to estimate PFD_{avg} by calculating the average unavailability of the system. The proof test interval (τ) is varied from 3 days to 1 year, where represents the time

between two consecutive inspections/proof tests. We considered following values τ of for simulations: τ = (3 days, 6 days, 15 days, 21 days, 1 month, 2 month, 3 month, 4 month, 5 month, 6 month, 7 month, 8 month, 9 month, 10 month, 11 month, 12 month).

Values of parameters like λ_a, λ_u, and mission time are chosen based on industry guidelines on the performance measure. The mission time of the system for the purpose of simulation is chosen to be 5 years. Based on industrial guideline, the impact factor of the proof test is considered as per equation 1. For each value of τ, 500 random realizations were simulated to obtain average unavailability of the system.

Figure 2, shows the estimated value of PFDavg of the system for different values of τ. The borderlines of SIL 1 and SIL 2, showing that the .01 < PFD_{avg} < 0.1 for being within the range of SIL 1 and PFD_{avg} < 0.01 for being in the range of SIL 2. Left side plot in Figure 2 shows that when both λ_a and λ_u are of the order of 10^{-6} per hour, the PFD_{avg} remains within SIL 2 for both AGAN and ABAO maintenance policies for 15 days $\leq \tau \leq$ 1 year. Right side plot in Figure 2 shows that when λ_a and λ_u are increased to the order of 10^{-5} per hour, the PFD_{avg} increases for both maintenance policies. For AGAN maintenance policy, PFD_{avg} leaves the range of SIL 2 and enters SIL 1 at τ = 15 days and leaves the range of SIL 1 at τ = 6 months. For ABAO maintenance policy the PFD_{avg} leaves range of SIL 1 for $\tau \geq$ 4 months and $\tau \leq$ 15 days and stays within the range of SIL 1 for an optimal proof test interval (15 days < $\tau \leq$ 3 months).

In Figure 2, when the plots pertaining to AGAN maintenance policy are observed, it is found that the information gain through inspection is more significant over the negative effect of testing. This is because with AGAN maintenance policythe system did not carry the history of past tests experienced by the system.

The important conclusion that can be derived from Figure 2 is that when the maintenance policy ABAO is chosen, PFD_{avg} of the system shows a trade off between the negative effect of performing a proof test versus the gain of information by performing the proof test on the system. In other words, when the system undergoes through high frequency of proof tests, the unavailability represented by the PFD_{avg} increases instead of decreasing as it did for AGAN policy. At the same time, when the frequency of proof tests is reduced, the user does not get enough information about the state of the system. Therefore, there exists an optimum frequency of testing which minimizes the value of PFD_{avg} in the Figure 2.

Figure 3 shows the effect, of changing the values of λ_a while keeping the value of λ_u as constant 5×10^{-6} per hour, on the PFD_{avg}. Note that the tradeoff between multiplicative negative effect of testing by high frequency of testing versus loss of information by low frequency of testing, is an attribute of ABAO maintenance policy only. Hence, the maintenance policy considered in Figure 3 is ABAO. It is observed from the Figure 3 that the PFD_{avg} remains within the range SIL 2 when the value of $\lambda_a \leq 5 \times 10^{-6}$ per hour for $\tau \in$ [15 days, 5 months]. Plots show that for each value of λ_a, there exists an optimum value of τ for which PFD_{avg} attains a minimum value. It is also observed that the value of PFD_{avg} increases with increasing values of λ_a.

Figure 4 shows the effect, of changing the values of the failure rate λ_u, while keeping the value of λ_a constant 5×10^{-6} per hour, on the PFD_{avg}. ABAO maintenance policy is considered for obtaining these plots, using the same arguments as for plots in Figure 3. It is observed from the left side plot in Figure 4 that when λ_u is increased from 10^{-7} per

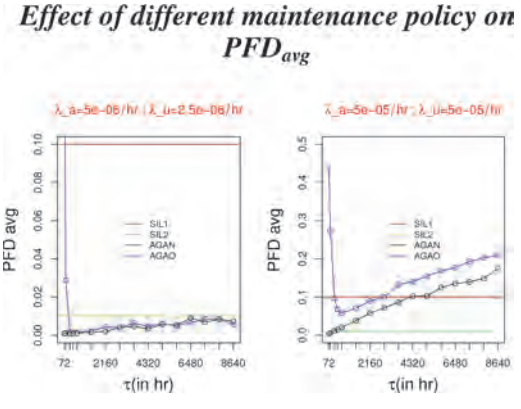


Figure 2. Effect of different maintenance policy on PFD_{avg}.

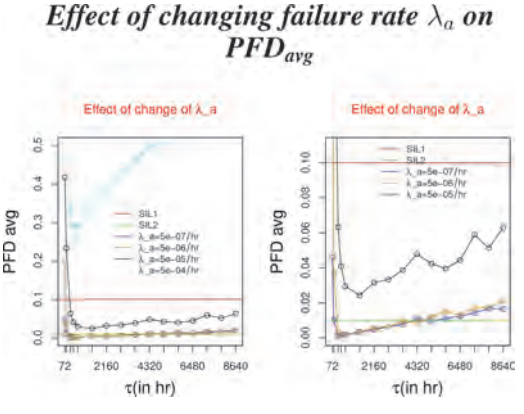


Figure 3. Effect of changing failure rate on PFD_{avg}.

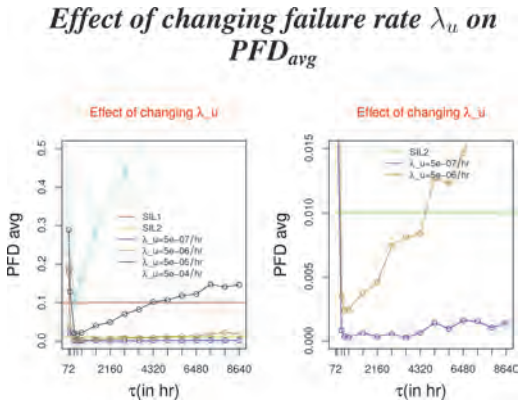


Figure 4. Effect of changing failure rate λ_u on PFD_{avg}.

hour to 10^{-4} per hour the shape of plot of PFD_{avg} changes from a flat convex to a steep convex indicating that PFD_{avg} increases with increase in λ_u.

The optimum time-interval (for performing proof test) which minimizes PFD_{avg} is significantly visible in Figure 4 for higher values of λ_u whereas for lower values of λ_u the curve needs to be zoomed up to observe the optimum time-interval (for performing proof test) as shown in the right side plot of Figure 4.

5 CONCLUSIONS AND IDEAS FOR FUTURE WORK

In AGAN maintenance policy, the technical state of the system is maintained to "as-good-as-new" after regular proof test meaning the system will not aggregate the negative effect of the regular proof test after maintenance. Hence, with AGAN we canmake PFD_{avg} as small as required by increasing the frequency of performing the proof test on the system. But using AGAN maintenance policy may not be economical in most of the practical situations, hence we focus on ABAO maintenance policy in this section.

5.1 *Conclusions*

Our case study showed that in case of ABAO maintenance policy, there are two competitive forces that can increase the PFD_{avg}. The first is the multiplicative negative effect of frequent proof tests, despite the maintenance that is carried out as part of the tests. This force becomes more dominant when the frequency of performing proof test is high. The second is the information obtained about the status of the system by carrying out the proof test. While the second force would like to increase the frequency of performing the proof test to lower PFD_{avg}. The first force would like to decrease the frequency of performing the proof test to obtain the same effect on the PFD_{avg}.

An optimum can be obtained for a regular proof test interval that can be verified against the constraints of the SIL requirement. It is therefore suggested that there exists an optimum frequency for performing the proof test that minimizes the PFD_{avg} of system whenever the following is true:

- The regular proof tests, that involves the inspection of technical state of the system, have some negative effect on the performance of the system due to test conditions and exposures.
- Some dangerous failure modes of the system can only be revealed by the regular proof tests, and not by other means (like e.g. diagnostic testing).
- System is maintained with the ABAO maintenance policy, meaning that the technical state is not "as-good-as-new" after a regular proof test. The ABAO maintenance policy will aggregate the negative effects of regular proof test.

5.2 *Ideas for future work*

The above studies were performed assuming no DD failures and mean time to repair as negligible. It would be an interesting proposition to see the effect of adding DD failures and mean time to repair to the above study. Analytical solutions need to be developed to find out the exact solution of the stochastic differential equation involved in the above situation. Two degraded states were chosen randomly in the above study, the connection between the physical phenomena of the degradation and quantification the degraded states needs to explored. Effect of the predictive maintenance and redundancies on the PFD_{avg} in this situation needs to be studied.

ACKNOWLEDGEMENTS

This paper has been written under the Norwegian Centre for Research based Innovation on Subsea Pro-duction and Processing (SUBPRO). The authors would like to thank the Research Council of Norway, as well to the industrial partners involved in this project.

REFERENCES

Bond, K. (2002). Iec 61511-functional safety: Safety instrumented systems for the process industry sector. In *Annual Symposium on Instrumentation for the Process Industries*, Volume 57, pp. 33–40. Instrument Society of America.

Brissaud, F., A. Barros, & C. B´erenguer (2010). Probability of failure of safety-critical systems subject to partial tests. In *Reliability and Maintainability Symposium (RAMS), 2010 Proceedings-Annual*, pp. 1–6. IEEE.

Bukowski, J.V. (2005). A comparison of techniques for computing pfd average. In Reliability and *Maintainability Symposium, 2005. Proceedings. Annual*, pp. 590–595. IEEE.

IEC, I. (1998). 61508 functional safety of electrical/- electronic/programmable electronic safety-related systems. *International electrotechnical commission*.

Innal, F., M.A. Lundteigen, Y. Liu, & A. Barros (2016). Pfdavg generalized formulas for sis subject to partial and full periodic tests based on multiphase markov models. *Reliability Engineering & System Safety 150*, 160–170.

Jin, H., M.A. Lundteigen, & M. Rausand (2012). Uncertainty assessment of reliability estimates for safety-instrumented systems. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of risk and reliability 226*(6), 646–655.

Oliveira, F., J. Domingues, A. Hafver, D.V. Lindberg, & F.B. Pedersen (2016). Evaluation of pfd of safety systems with time-dependent and test-step varying failure rates. *ESREL, Glasgow, UK* -, 413.

Rausand, M. (2014a). *Reliability of Safety-Critical Systems: Theory and Applications*, Volume -. Hoboken, Wiley.

Rausand, M. (2014b). *Reliability of safety-critical systems: theory and applications*. JohnWiley & Sons.

Rouvroye, J. & A. Brombacher (1999). New quantitative safety standards: different techniques, different results? *Reliability Engineering & System Safety 66*(2), 121–125.

Shao-Ming, W., H. Ren, & W. De-Jun (1994). Reliability analysis of a repairable system without being repaired "as good as new". *Microelectronics Reliability 34*(2), 357–360.

Strand, G.O. &M.A. Lundteigen (2015). Risk control in the well drilling phase: Bop system reliability assessment.

Wu, S., L. Zhang, A. Barros, W. Zheng, & Y. Liu (2018). Performance analysis for subsea blind shear ram preventers subject to testing strategies. *Reliability Engineering & System Safety 169*, 281–298.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Statistical comparison of three different measurement technologies

M. Hinz, A. Luecker & B. Bracke
Chair of Reliability and Risk Analytics, University of Wuppertal, Wuppertal, Germany

C. Klostermann
Klostermann Ingenieurbüro und Vertriebsgesellschaft mbH, Remscheid, Germany

ABSTRACT: Typical measurement technologies are coordinate measuring technology, structured-light 3D scanner and Computed Tomography (CT) scans. In this paper, an innovative measurement and analysis strategy for the statistical comparison of the mentioned technologies is developed. Furthermore, the measurement result can be proved with regard to the plausibility and sensitivity. This strategy can be used not only for the intern company purposes but also for the adjustment and harmonization of product and production improvement activities and corrective actions between original equipment manufacturers (OEMs) and suppliers. Plausibility checks of the gathered measurement results can be facilitated also in case of different applied technologies.

1 INTRODUCTION

To measure, to map and to analyse different parts or objects is interesting in many scopes of production respectively product development process. Especially in automotive, requirements on specification and product complexity is steadily increasing which requires high performance on measurement methods. Depending on product geometry, complexity and its critical specification, a proper measurement method has to be chosen.

Every measurement technology is basically suitable for the measurement of the product geometry but has also various advantages and disadvantages. Structured-light 3D scanner is well qualified for quick measurement of entire geometry but is sensitive for the creation of shades as well as product undercuts. An advantage of coordinate measuring technology is the high measuring accuracy of touched measure points which can be used for calculation and reconstruction of the actual product shape. In contrast, the measuring effort of explicit free-form surfaces is very high, which is a grate disadvantage of this method. Product cavities are measurable only after saw opening of a product in case of both, structured-light 3D scanner and coordinate measuring technology. A CT scan makes use of computer-processed combinations of many X-ray images taken from different angles and allows a survey of undercuts and cavities but cannot be used for all component materials (e.g. lead).

It is common that a complex geometry is measured with several, different measurement technologies in many practical applications. The problem is that the obtained results cannot be compared to each other due to different technology principles. It is uncertain whether all three technologies provide a sufficient precision of the product geometry or not.

Therefore, a study based on the statistical comparison of the three different measurement techniques regarding various dimensions of the measured objects is presented in this paper. Due to the overall understanding, first the measurement methods are discussed regarding the applicability as well as their advantages and disadvantages. In the subsequent part, the measured part, the material it is made of as well as the measured points are presented. The statistical methods used for the comparison, in this case the non-parametric statistical tests are discussed in detail. For the purpose of a sensitivity study, Monte Carlo simulated values of various dimensions are analysed first. In other words, it is necessary to understand the sensitivity of the tests before presenting and discussing the real measured values. Finally, the results are discussed in detail and the study is concluded.

2 MEASUREMENT METHODS

The following measurement methods have various advantages and disadvantages. In general there are three measurement methods, which were also used in this work:

- Coordinate measuring machine (CMM)
- Structured-light 3D scan method (3DSM)
- Computer tomography (CT)

Coordinate measuring machine is based on a tactile measuring method. Main components are the measuring head with a touch sensor, the measuring table and the positioning system with an incremental sensor technology. The measurement object and the measurement head move relative to each other; therefore CMM is suitable to measure space coordinates. In comparison to 3DSM and CT, CMM provides the best measurement result with highest measurement precision at the touched measurement points. Due to the relative movement of measurement object and head, measuring an object point by point takes a lot of time. Furthermore, it is difficult to measure objects with freeform surfaces and it is impossible to measure objects with undercuts in a non-destructive way.

The structured-light 3D scan method is an optical, non-contact method and structured by a projector and two cameras on a tripod. Measuring result is a point cloud of the surface of the measurement object and needs to be mapped by computer algorithms subsequently. It is also possible to create a CAD model out of these points. A quick measurement of the entire object is the big advantage of 3DSM and provides a good alternative to measure parts in a production line. Disadvantages are big scattering effects regarding measurement result and precision due to the creation of shades. Furthermore, it is not possible to measure objects with undercuts in a non-destructive manner.

Computer tomography is also called imaging method. It is constructed similar to an X-ray machine, whereby the operating principle is slightly different in comparison to an X-ray machine. The difference is that during a measurement with a CT a lot of images, out of various angles and directions, are made and recorded in a systematic way. Subsequently, a computer composes the recorded images with the help of complex algorithms and map them to a CAD model. CT method does not work very fast but it has the ability of measuring undercuts without destroying the measured object. Therefore, it is well suitable to measure complex objects and geometries. The measurement result and precision is strongly influenced by the material of the measurement object and lots of setup parameters of the machine itself. The thicker and the denser is the material, the more the measurement result and precision will be affected in a negative way due to the scattering effects.

3 MEASURED PART—DTM INTAKE SOCKET

3.1 *Geometry*

The part, that has been chosen for the study is an aluminium intake socket out of the DTM (German Touring Car Masters). Since the material it is made of is very important for the CT-measurement, a spectral analysis has been undertaken in order to define the proper material composition. The results are described in the following section.

The overall geometry was considered to consist of various measured dimensions as well as geometric tolerances. Furthermore, it was considered to have a complex, though still manageable complexity for the measurements. The chosen intake socket, presented in the Figure 1 was a good compromise of all these attributes.

Further requirements on the object were different specifications of the measurement units (in millimetre or degrees). The object was considered to consist of various surfaces, curvatures, as well as cavities and undercuts. The material had to be measureable in a CT. Is shall be additionally stated, that the found part was prepared and refurbished for the purpose of the study. The exact composition of the material was unknown at the beginning and had to be analysed before starting the measurements.

3.2 *Material*

As already stated in the previous section, the material the part has been made of, needed to be an appropriate one in terms of the applicability for the CT-system. Therefore, the chosen part had to be analysed first regarding the material composition. For this purpose, a small splinter has been removed in the inner part of the socket (in the upper flange) for the material analysis. Subsequently, the splinter has been prepared for the analysis which has been performed with the energy-dispersive X-ray spectroscopy (Lipták 2003). The result of the spectroscopy is shown in Figure 2.

Figure 1. Measured object—intake socket out of the automotive engineering.

The abscissa of Figure 2 shows the energy in [keV], and the ordinate sows the intensity in counts per minute [cpm] of a certain element in the composition. The main peak of the diagram indicates an aluminium alloy. Though, the proper estimation of the material composition cannot be performed without a material expert. Therefore, the overall estimation of the compound was discussed with a material scientist and provided the result (with a high probability) of an AlMgSi—wrought alloy.

The small amounts of other various elements are impurities. The analysed material is well suitable for the CT-measurement, since it is the aluminium wrought alloy.

3.3 *Measured points*

The chosen 19 points for the measurement are presented in Figure 3. The precise description of all point is out of scope for the purpose of this paper as well as not necessary for the understanding of the performed research, and will be therefore omitted. Nonetheless, the chosen points represent a variety of different lengths and angles with various geometric tolerances and tolerance specifications.

Though, the chosen points can be considered as representative for all parts of similar dimensions.

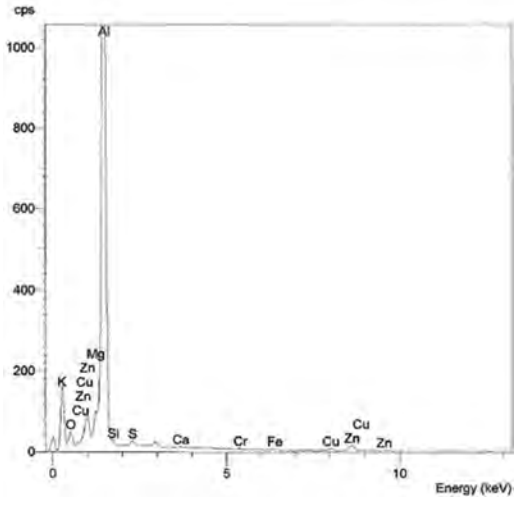


Figure 2. Spectral analysis of the material—intake socket out of the automotive engineering.

4 SAMPLE ANALYSIS BASED ON TEST STATISTICS

First step of the analysis of the measurements is the realisation of various hypothesis tests regarding the application of the proper analysis methods. Therefore, first the goodness of fit tests with the null (and alternative) hypothesis:

$$H_0 : F = F_0 \ \text{(and } H_1 : F \neq F_0\text{)} \tag{1}$$

that the sample data is distributed in a certain way (either normally or Weibull) has to be performed. For the normal distribution, Kolmogorov-Smirnov and Shapiro-Wilk (Hartung 1998) tests are common in use. Accordingly, Anderson-Darling, Kolmogorov—Smirnov and Chi-squared (Sachs 2009) tests can be applied for the Weibull distribution.

These testes are performed for two different tasks. First is the analysis of distributions regarding the measurement technologies itself, the second concentrates on the distributions of the measured points. The second analysis is mainly interesting for the different groups of geometric tolerances. In simple words, it shall be analysed, if e.g. the angles can be always fitted based on the same distribution function. The further purpose of the goodness of fit tests is the proper application of test statistics—parametric for the normal distributed samples and non-parametric in all other cases (Hinz 2014).

The goodness of fit tests applied for both cases provide heterogeneous results. It means, that no

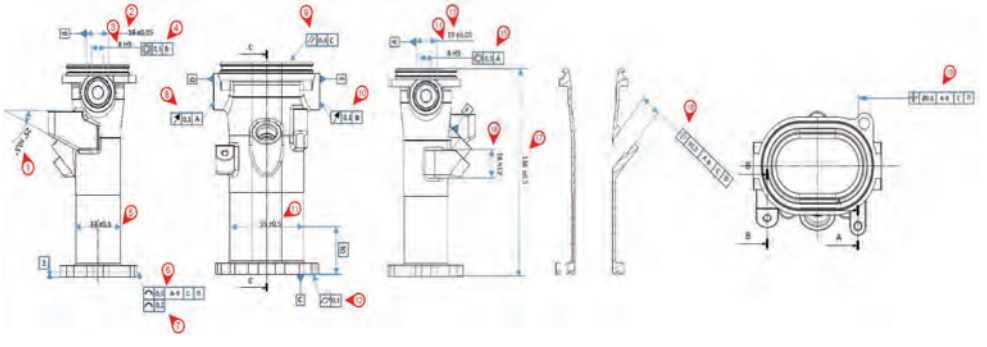


Figure 3. Measured points of the DTM socket.

specific distribution function can be fitted to all samples, either distinguished by the technology or by the measured points. Therefore, non-parametric statistics shall be applied for the further analyses.

For the comparison of samples, various significance tests according to different applications are available. Non-parametric significance tests do not require dependences on specific distributions. In comparison to similar application cases of parametric significance tests, further application cases and benefits such as simple calculation and detection of randomness of data are available.

For the purpose of this study, two nonparametric statistical tests were performed:

Mann-Whitney U test with the null (and alternative) hypothesis:

$$H_0 : F(z) = G(z) \quad (\text{and } H_1 : F(z) \neq G(z - \Theta)) \tag{2}$$

that the samples have the same focus, and

Levene's test with the null (and alternative) hypothesis:

$$H_0 : F(z) = G(z) \quad (\text{and } H_1 : F(z) \neq G(\Theta z)) \tag{3}$$

that the variances of the samples are equal.

A visual example of the statistical comparison based on the mentioned tests is shown in Figure 4.

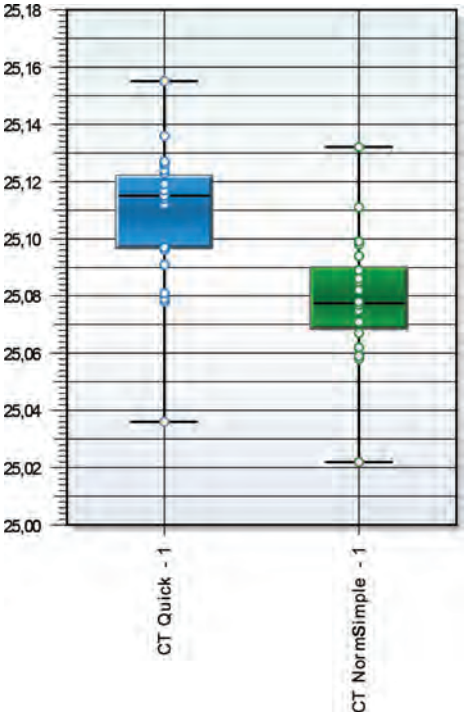


Figure 4. Boxplot—comparison of the samples.

Here, two different samples established with the CT-technology are statistically compared to each other.

5 SENSITIVITY STUDY

Preliminarily to the performed study based on real data, a sensitivity analysis was performed regarding the comparison of various samples based on simulated data. The main aim of the sensitivity analysis was the detection of critical boundaries, at which the test will distinguish the samples. In simple words, it shall be analysed when the samples can be defined as significantly different with regard to the measured value itself, the scattering of the value and the decimal digits with respect to the measurement technology.

For the purpose of the study, 27 various samples with sample size of 20 measurements were Monte-Carlo simulated according to the following conditions:

- All values are simulated with three decimal digits (due to the maximal resolution of the measured technologies)
- For every measured magnitude, three different values have been simulated: mean, upper specification limit, and lower specification limit
- All values were assumed to be normal distributed

Based on the given assumptions, different samples with different magnitudes has been simulated: 19 ± 0.05; 33 ± 0.5; and 136 ± 0.5, which provides all in all 27 samples. Note that all the values have no units, since the unit has no influence on the analysis itself and can represent any measurable value.

The simulated samples are analysed with both: Mann-Whitney U as well as Levene's test. Exemplary results are shown in Table 1. Here, the first column of the simulated value shows the simulated magnitude, the second shows the simulated scattering.

Table 1. Exemplary results of statistical analysis of simulated values.

Simulated value		U-Test	Levene's Test
$\mu = 19$	$s = 0{,}005$ vs. $s = 0{,}05$	$\mu_{005} = \mu_{05}$	$\sigma_{005} \neq \sigma_{05}$
	$s = 0{,}005$ vs. $s = 0{,}5$	$\mu_{005} = \mu_{5}$	$\sigma_{005} \neq \sigma_{5}$
	$s = 0{,}05$ vs. $s = 0{,}5$	$\mu_{05} = \mu_{5}$	$\sigma_{05} \neq \sigma_{5}$
$\mu = 33$	$s = 0{,}005$ vs. $s = 0{,}05$	$\mu_{005} = \mu_{05}$	$\sigma_{005} \neq \sigma_{05}$
	$s = 0{,}005$ vs. $s = 0{,}5$	$\mu_{005} = \mu_{5}$	$\sigma_{005} \neq \sigma_{5}$
	$s = 0{,}05$ vs. $s = 0{,}5$	$\mu_{05} = \mu_{5}$	$\sigma_{05} \neq \sigma_{5}$
$\mu = 136$	$s = 0{,}005$ vs. $s = 0{,}05$	$\mu_{005} \neq \mu_{05}$	$\sigma_{005} \neq \sigma_{05}$
	$s = 0{,}005$ vs. $s = 0{,}5$	$\mu_{005} = \mu_{5}$	$\sigma_{005} \neq \sigma_{5}$
	$s = 0{,}05$ vs. $s = 0{,}5$	$\mu_{05} = \mu_{5}$	$\sigma_{05} \neq \sigma_{5}$

The results of the statistical tests are shown in the columns U-Test (which is the short form of Mann-Whitney U Test) and Levene's test. It can be observed easily that only one hypothesis is rejected in case of the U-Test (seventh row) which means that the test statistics recognizes only one significant difference regarding the mean value. Though, it cannot be determined why this one particular test shows a significant difference. On the other hand, all test are rejected in case of the comparison of scattering of the samples. This means that independent on the magnitude of the simulated value, a factor 10 in the scattering will be always detected. This provides a very good detectability for the samples.

Since a detailed description of all results is out of scope for this paper, the most important results are summarized as follows: All 27 Levene's tests show a significant difference between the simulated scatterings. It means that a scattering with a difference of at least one magnitude can be always detected for such measured values. 26 of 27 U-tests (96.29%) show no significant difference between the samples regarding the focus of the sample. This means that a small scattering of the values has no influence on the detection of significance in the measured samples regarding the mean value. Furthermore, it means that once a significant difference will be detected in the measured samples, one can be sure that the technologies are significantly different.

Since no significant differences were determined in the scattering of the samples, a further analysis was performed. For this reason, further samples has been Monte-Carlo simulated based on the mentioned conditions with minor exception: The values are simulated with four decimal digits. Based on the assumptions, following samples has been simulated: 19 ± 0.002; and 19 ± 0.005.

All in all, 22 additional Mann-Whitney U and Levene's tests each were performed in the same manner. Once the scattering is much lower, the focus of the samples is much better differentiable. 18 out of 22 test show a significant difference regarding the focus of the samples. The variances are always differentiable based on the Levene's tests. This shows a very good applicability of the introduced hypothesis statistics.

6 ANALYSIS OF THE MEASUREMENTS

Based on the results and gained knowledge from the previous section, the measurements of the three described technologies shall be analysed. All measurements were performed 20 times for every technology each. Therefore, all discussed 19 measured points provide samples of 20 measurements. However, CT provides not the measured values itself, but.stl files which have to be additionally imported into a CAD program and processed. The.stl files are created based on the measurements and can be exported with very different qualities (the higher the number of elements in the.stl grid, the bigger the files). Therefore, different qualities of CT measurements were exported in order to analyse the quality of the mesh on the final results. The smallest.stl file was 120 MB and the biggest 3.5 GB big, which obviously determines also the export and calculation time.

The statistical test were performed exactly in the same manner as in case of the simulated values. All in all, over 250 statistical test were performed based on the measurements. The results provide the following conclusions:

- CMM has in 97.14% of all measurements a significant difference in the mean value
- CMM has in 98.85% of all measurements a significant difference in the scattering
- 3DSM has in 100% significant difference in both, mean and scattering, compared to CMM
- 3DSM has in 86.67% significant difference in mean value, compared to CT
- 3DSM has in 56.67% significant difference in the scattering, compared to CT
- CT has in 76.67% of all measurements a significant difference in the mean value

Basically, based on the gathered results, following statements are valid:

- CMM has a significant difference to the remaining technologies
- Based additionally on experience, it can be defined, that CMM is the most accurate technology
- The choice of the converting algorithm for the. stl files within the CT measurements has a significant influence on the quality of the results
- The difference between 3DSM and CT cannot be observed—both technologies has the same measuring accuracy

7 CONCLUSIONS

A very comprehensive study regarding the comparison of three different measuring technologies is presented in the present paper. For this purpose, an aluminium DTM socket has been measured on 19 different points. For the purpose of the comparisons, nonparametric test has been chosen and classified as suitable.

For the purpose of a sensitivity study, a number of samples has been Monte-Carlo simulated and analysed. It has been proven, that the chosen test

provide satisfactory results regarding the differentiability of the measured samples.

The analysis of the results show many significant differences, which means that the technologies itself are significant different, whereby the CMM is the most accurate one.

Based on the results of the analysed measurements as well as the advantages and disadvantages of the technologies, the decision about the application of the proper technology within a certain industry can be performed easily.

REFERENCES

Hartung, J. & Elpelt, B. 1998. *Lehr—und Handbuch der angewandten Statistik; mit zahlreichen, vollständig durchgerechneten Beispielen.* 11. Edition. München, Wien: Oldenbourg.

Hinz, M., Temminghoff, P. & Bracke S. 2014. *APTA approach: Analysis of accelerated prototype test data based on small data volumes within a car door system case study.* PSAM: The 12th Probabilistic Safety Assessment and Management, PSAM 12, Honolulu, Hawaii, USA, June 22th -27th, 2014.

Lipták, B.G. Instrument engineers' handbook. 4th. Ed. CRC Press, 2003.

Sachs, L. & Hedderich, J. 2009. *Angewandte Statistik Methodensammlung mit R.* (13). Berlin: Springer.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Machine learning modeling for massive industrial data: Railroad peak kips prediction

C. Contreras, M. López-Campos, P. Escalona, R. Stegmaier & T. Grubessich
Department of Industrial Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile

ABSTRACT: The exponential growth of industrial data being generated by sensors, modern equipment and devices is pushing the service sector to use more sophisticated analytics tools that can produce useful knowledge and predict certain events, especially for those which require reducing loss through preventive maintenance. This work presents the application of big data analytics for machine learning processing through a railway company problem approach, using one of the most powerful tools for large scale data management: the open-source Apache Spark platform. The practical implications of this, are in a reliable prediction of the condition of trains before being loaded and sent to a destination.

1 INTRODUCTION

The exponential growth of industrial data being generated by sensors, modern equipment and devices is pushing the service sector to use more sophisticated analytics tools that can produce useful knowledge and predict certain events, especially for those which require reducing loss through preventive maintenance.

However, processing large scale datasets and building predictive models with advanced algorithms, demands high efficiency in iterative computation tasks. In the last years, the open-source platform Apache Spark has experienced rapid growth due its outstanding performance and wide range of settings, which makes it well-suited for the development of machine learning (ML) applications using popular programming languages such as Java, Python, Scala and R.

This work shows how to build a predictive model in simple steps using the Spark environment for structured massive data manipulation. This model will consist of a decision tree for binary classification and a sensitivity analysis performance adjusting the main parameters. For academic purposes, the modeling will be set up on a pseudo-distributed single node cluster in Ubuntu 16.04, and the datasets loaded to Hadoop Distributed File System (HDFS), a reliable storage for large files that allows parallel processing, and Apache Spark 2.2.0 framework.

2 PROBLEM STATEMENT

In the railroad industry, the train wheels are critical components in terms of safety and one of the main priorities due their probabilities of failure, which may imply catastrophic consequences. One of the main concerns of the maintainers, is to be able to predict if according to several factors as the condition of the wheels, the size of the load to be transported, the distance to travel and the conditions of the road, the vehicle is suitable for go on a trip, ensuring its arrival on time and without problems to the destination. Otherwise, the necessary actions should be taken. The use of technology and especially counting on reliable databases, as well as mechanisms to provide intelligence and structure that data in predictive models, is relevant to support such decision-making.

Thanks to the development of new sophisticated sensors such as wheel impact load detectors (WILD) it is possible to monitor structural health trends and spot the critical wheels which need to be removed. However, setting out a car when it is loaded will cause a lot of loss to railroad companies, among other problems such delayed shipment and unnecessary disruptions in the network traffic. This situation urges the need of developing predictive models that can be used to project if a total vertical force imposed to the rails (denoted as peak kips) is above to certain values in the next loaded status. The WILD system scans millions of wheels per day throughout the international rail industry (Stratman, 2007), and it will provide us useful empirical datasets for training and testing for this endeavour.

3 PROPOSITION FOR METHODOLOGY

In real world applications, the data provided by different sources is not always structured. It needs to be preprocessed in order to generate the inputs to feed machine learning algorithms. In this particular case,

we propose the following framework for the creation and validation of a wheel peak kips predictive model.

3.1 *Machine learning data pipeline*

Data pipeline represents the flow of data for the machine learning process. This starts with a set of massive raw data which needs to be pre-processed and then fed to the machine learning algorithm. There are a large number of machine learning algorithms and different classifications, according to the type of problems they address (regression, classification, etc.), according to the data processing (linear, tree-based, neural network models) among others classifications. The ML algorithm learns, based on the fed data, to predict the future behavior of new supplied data. This is called "training" an algorithm. Subsequently, the data thrown by the ML algorithm need to be interpreted to finally obtain a diagnosis that is helpful in decision making. The entire process from the obtaining and pre-processing of raw data, to the interpretation and use of the predictions requested by the final users, is the called ML data pipeline.

The development and assembly of pipeline components need to support distributed computation and other requirements regarding data treatment, including fault tolerance, resource management, scalability and maintainability. This makes Apache Spark a very useful and reliable environment due its fundamental data structure for parallel processing (Resilient Distributed Datasets), and the Scala DataFrame API that allows to process from Kilobytes to Petabytes of data on a single node cluster and perform operations with only a few lines of code.

Figure 1 summarizes the steps followed by the ML data pipeline used in this work.

3.1.1 *Step 1: Data ingestion*

Data ingestion refers to the process of obtaining data to be used in the ML data pipeline. Data ingestion implies a non-trivial process of prioritizing data sources, validating the data obtained and sending them in an orderly manner to their next destination. All this must be done quickly and reliably so as not to lose information, especially when it is collected in real time. As previously mentioned, WILD system is the source of the two datasets used in this project: a training dataset and a testing dataset. These files contain the exhaustive information about a great number of variables monitored in relation to the performance of train wheels. It has very complete mechanical and physical information about the condition of each wheel, at each moment, corresponding to specific cars, trains and trips.

Data ingestion of the mentioned databases will be done using the Hadoop Distributed File System (HDFS), a reliable storage and highly fault tolerant for large files, and specially designed to be

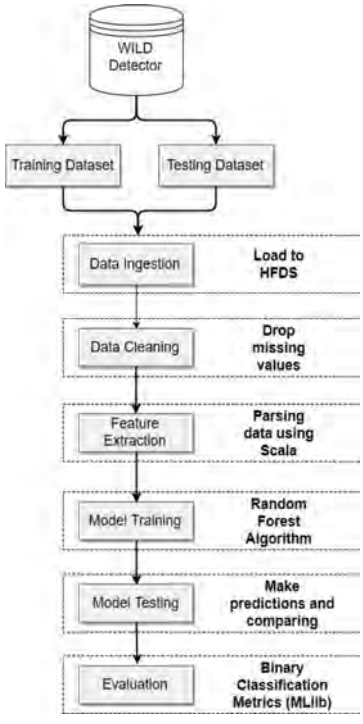


Figure 1. Machine learning data pipeline process flow diagram.

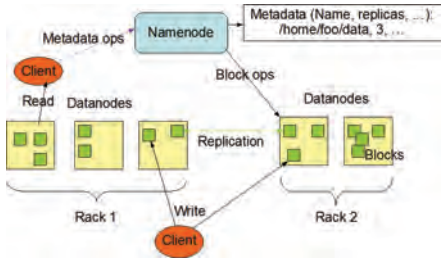


Figure 2. HDFS Archiquecture.

deployed on low-cost hardware. Its architecture is shown in Figure 2.

This distributed file system follows a master/slave architecture, the Namenode and Datanodes, respectively. For example, we have running five computers with 500 Gigabytes of storage each one with Hadoop environment installed. Accessing the storage from any of these five machines will work as a single large machine with total capacity of 2,5 Terabytes. This is when parallel processing can take advantage: if one single machine makes a task that takes 20 minutes with 500 Gigabytes of data, ten of them can complete such task in only 2 minutes.

In our particular case, the training and testing datasets consist in tabulated text files of 1.000

and 500 Gigabytes corresponding to 7 million and 2,5 million of rows, respectively. Both have headers with the name of the 22 features including the label, i.e. the values that we want to predict (peak kips). In our local machine terminal, the following commands are used to load these files:

```
>hadoop fs–put /localpath/training.tsv /training
>hadoop fs–put /localpath/testing.tsv /testing
```

3.1.2 *Step 2: Data cleaning*

Data cleaning is considered as a main challenge in the era of big data due to the increasing volume, velocity and variety of data in many applications (Tang, 2014). It is the process of detecting, fixing or removing incorrect and misleading records.

Having dirty source inputs is very likely and it can easily trigger runtime exceptions and therefore, terminate our whole process in Apache Spark.

There are several solutions and tools for data cleaning. In the Scala programming language, the main utility to achieve this task are the instances *Try*, *Success* and *Failure*.

```
> // Object to transform peak kips to binary
> def b(y: String): String =
> if (y.toDouble > = 90) "1" else "0"
>
> // Load from HDFS and clean data
> val file = sc.textFile("/training")
> val file2 = file
>      .map(_.split("\t")).
>      .flatMap(c = > Try{ b(c(1)+...)}.toOption)
```

3.1.3 *Step 3: Feature extraction*

Feature Extraction is a kind of dimensionality reduction that efficiently represents an initial set of data. The resulting reduced set is a non-redundant and good representation of the most relevant information of the entire data. Then, the following modeling training and testing processes can be executed over this reduced set of features (feature vector) instead over the complete initial database.

Since the incorporation of the Dataframe API to Spark in 2015, data processing and functional transformations have become much easier to code in general-purpose programming languages, and at the same time their performance have been improved. For the realization of this work, the selected features appear in Table 1.

3.1.4 *Steps 4 and 5: Model training and testing*

To train the data, we will use as ML algorithm, the Random Forest model for classification. One of the main advantages of this algorithm is that it does not require assumptions of normality of variables and it can deal with highly correlated and non-linear relationships between them.

Basically, the decision trees are built as it follows:

Table 1. Description of selected features.

Feature Name	Description
LOAD_EMPTY	Binary variable that indicates whether de equipment is loaded or empty
EQP_GRS_TONS	The tonnage of the equipment that the wheel is part of
EDR_EQP_SPD	The speed of the wheel at the time of the WILD measurement
TARE	Weight of the car when empty (in tons)
GRS_RAIL_WGT	The tonnage that the particular wheel type can sustain

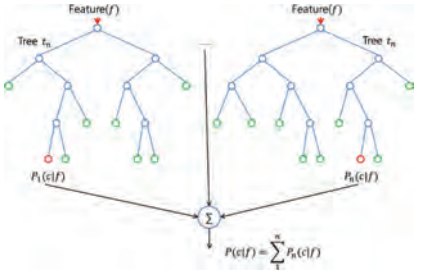


Figure 3. Graphical representation of a Random Forest decision tree.

- Randomly select *f* features from available features *F*
- Compute the best split point for tree *t* using the determined splitting metric (in our case of study, *Gini Impurity*), and split the current node into child nodes and reduce the number of features *f* from this node on
- Repeat steps 1 to 2 until either a maximum tree depth has been reached
- Repeat steps 1 to 3 in order to create a T number of trees

The result can be represented in Figure 3.

In the Spark platform, the generic code for building the model can be written as it follows:

```
> val model = Random-
Forest.trainClassifier (trainingData, numClasses,
categoricalFeaturesInfo, numTrees, featureSub-
setStrategy, impurity, maxDepth, maxBins)
```

3.1.5 *Step 6: Evaluation*

One of the most effective methods of evaluating the performance of a binary classifier system is the receiver operating characteristic (ROC) curve, which is defined as a plot of the true positive rate (y coordinate) against the false positive rate (x coordinate). Accuracy is measured by the area under the ROC curve (AUROC), and it's represented by the Equation (1)

$$AUROC = \int_0^1 \frac{TP}{P} d\left(\frac{FP}{N}\right) \quad (1)$$

This value fluctuates between 0,5 to 1, where 0,5 denotes a null prediction capacity, and 1 a perfect classifier.

To compute the raw scores on the test set, we can code the following lines:

```
> val predictionAndLabels = test.map {case
LabeledPoint(label,features)=>valprediction=
model. Predict(features)
 (prediction, label)}
> // Instantiate metrics object
> val metrics = new BinaryClassificationMet-
rics(predictionAndLabels)
> // Area under the ROC curve
> val auROC = metrics.areaUnderROC
> println("Area under ROC = " + auROC)
```

In addition, we can provide a simple sensitivity analysis by modifying the number of trees (numTrees) and maximum depth (maxDepth) to calculate the different values of AUROC (1).

4 RESULTS

As we can see in Table 2, the area under the receiver operating characteristic (AUROC) does not experience major fluctuations.

The most accurate model has 4 trees and maximum depth of 6. We can say that there is a 78% of probabilities that a defective train wheel will activate the alarm in the next loaded status predicted by the model (peak kips > 90) than a randomly chosen one. This represents the accuracy of prediction of the model.

5 DISCUSSION AND CONCLUSIONS

This work presents an application of big data analytics for machine learning processing with a railway company problem approach and using open-source tools. According to the obtained results, we can stand that the application of this methodology is adequate and valuable as a decision-making support, jointly used considering the experience of personnel, maintainers and technicians. At the same time, future analyzes can be tried to increase the predictive performance of the tool used in this project, for example, try to consider more features for the modeling, using a computer with more processing capacity.

We want to highlight the multipurpose Random Forest algorithm that can be used in many industrial sectors, in this case for preventive maintenance. Its outstanding performance is usable to get valuable predictive information that can be translated into new opportunities. However, due to limitations in hardware (settings on a single node cluster), training models with even larger input datasets may have a wide margin of improvement in terms of accuracy and processing speed.

Table 2. Area under ROC curve according to the parameters.

	Number of trees			
Max depth	3	4	5	6
3	0.7767	0.7708	0.7742	0.7702
4	0.7757	0.7828	0.7380	0.7812
5	0.7819	0.7632	0.7620	0.7622
6	0.7600	0.7864	0.7604	0.7622
7	0.7567	0.7660	0.7535	0.7622

REFERENCES

Aly, M., Yacout, S., & Shaban, Y. (2017, January). Analysis of massive industrial data using MapReduce framework for parallel processing. In Reliability and Maintainability Symposium (RAMS), 2017 Annual (pp. 1–6). IEEE.

Jamshidi, A., Faghih-Roohi, S., Hajizadeh, S., Núñez, A., Babuska, R., Dollevoet, R., ... & Schutter, B. (2017). A big data analysis approach for rail failure risk assessment. Risk analysis.

Ji, W., & Wang, L. (2017). Big data analytics based fault prediction for shop floor scheduling. *Journal of Manufacturing Systems*, 43, 187–194.

Liu, J., Dong, Y.F., Li, Y., Lei, S.Y., & He, S.Q. (2016). Composite Fault Diagnosis and Intelligent Maintenance Based on Data Driven. In Key Engineering Materials (Vol. 693, pp. 1357–1360). Trans Tech Publications.

Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., ... & Xin, D. (2016). Mllib: Machine learning in apache spark. *The Journal of Machine Learning Research,* 17(1), 1235–1241.

Shanahan, J.G., & Dai, L. (2015, August). Large scale distributed data science using apache spark. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 2323–2324). ACM.

Tang, N. (2014, September). Big data cleaning. In Asia-Pacific Web Conference (pp. 13–24). Springer International Publishing.

Xiaoshan, Y., Ligu, Z., Qicong, Z., & Dongyu, F. (2016, December). Research on Evaluation Method of Big Data Storage Utilization. In Applied Computing and Information Technology/3rd Intl Conf on Computational Science/Intelligence and Applied Informatics/1st Intl Conf on Big Data, Cloud Computing, Data Science & Engineering (ACIT-CSII-BCD), 2016 4th Intl Conf on (pp. 368–372). IEEE.

Zaharia, M., Xin, R.S., Wendell, P., Das, T., Armbrust, M., Dave, A., ... & Ghodsi, A. (2016). Apache Spark: A unified engine for big data processing. Communications of the ACM, 59(11), 56–65.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Adaptive meta-heuristic to predict dent depth damage in the fixed offshore structures

W. Punurai & M.S. Azad
Department of Civil and Environmental Engineering, Faculty of Engineering, Mahidol University, Thailand

N. Pholdee
Department of Mechanical Engineering, Faculty of Engineering, Khon Kaen University, Thailand

C. Sinsabvarodom
Department of Marine Technology, Norwegian University of Science and Technology, Norway

ABSTRACT: The jacket structures are often employed in the range of shallow-moderate water depth. The bracing systems and jacket legs typically use the circular section in order to compromise the hydrodynamic resistance and high torsional rigidity However, under lateral impact, these tabular bracing members are susceptible to local denting due to ship collisions or through impact of falling objects and that can weaken overall performance of the entire platform. It is a great significance for forecasting dent depth of these members accurately. This paper investigates the use of adaptive meta-heuristics algorithm to provide an automatic detection of denting damage in an offshore structure. A model is developed combining with the percentage of the dent depth of damaged member diameter and is used to assess the performance of the method. It is demonstrated that the small changes in stiffness of individual damaged bracing members are detectable from measurements of global structural motion.

1 INTRODUCTION

Offshore jackets play an important role in the oil and gas industry. The risk of platform failure is a higher risk issue when operating an ageing platform. As a result, the need for the development of techniques necessary to assess platform integrity is clearly established. Several standards and recommended practices list a detailed inspection methods and requirements for the underwater, splash zone and topsides structures (May et al. 2008). For the oldest of operational ageing platforms, the assessment is quite challenging due to the existence of deterioration or dent-damage of structural elements due to collision or impacts. The change of dent depth can dramatically reduce the axial and bending capacities of the individual jacket leg members and weaken overall performance of the entire structure (Bruin, 1995). Therefore, accurate prediction of dent depth and dent direction angle would provide valuable information for the operations and maintenance operations at offshore platforms.

Karamanos and Andreadakis (2006) studied on denting of tubular members subjected to lateral loads. The loading condition is quasi static and the tubes are internally pressurized. The relation between denting force and displacements were evaluated through experiments and finite element simulations. Also relation between normalized forces and normalized denting displacements was deviated following both procedures. Wedge shaped denting tool was used in the experiment as well. Outcome of this research deliberates that the internal pressure in the tubular member increases the resistance against denting and reduce the denting length. Travanca and Hao (2014) discussed on the response of jackets on response to ship impact. Finite element formulations and calculations are derived to incorporate the dynamic aspects as well. The jacket was considered as a cantilever beam and the Degree of Freedom (DOF) was reduced for easiness of study. The outcome of the reduction of DOF was significant and showed similar outcome as from the FEM of original structure so far. Comparison of response time histories, deformation of modes, normalized deformation at different stages and the comparison of response histories from the original model and equivalent reduced SDOF model was contemplated for the parametric study. Moreover a four-legged jacket, a tripod jacket and a jack-up model were considered

for case study. Furthermore the effect of top mass are included and it is found that it is significant while the natural period has large value and the ratio of top mass and the jacket mass is high. Two stage dynamic analysis is also suggested to achieve more accuracy. Cosham and Hopkins (2004) illustrated on the dent effect on oil and pipelines which are especially used for oil and gas transmission. Different types of defects along with different types of dents are studied in this study. Burst strength and fatigue life are investigated for different types of dents. It was found that plain and smooth dent doesn't affect significantly on the burst strength but has effect on the fatigue life. Moreover, smooth dent reduces the fatigue life to a great extent. Kinked dents may be are dangerous for longitudinal cyclic stresses and are risky for external pressures with the internals. Moreover, it is noted that the smooth dent along with gouge is very unsafe considering both burst strength and fatigue which reduces considerably. Khedmati and Nazari (2012) explored the behavior of tubular members on response to impact loads. Strength and deformation characteristics are inspected in this work. Moreover the axial shortening behavior is also included. The effect of preloading and quasi static lateral load are also delineated. Storheim and Amdahl (2014) deliberates the design process of offshore platforms susceptible to ship collision. The ship-platform interaction is also represented. Four different collision contexts are incorporated. Force-displacements curves for different scenarios are interpreted. The outcome of the study suggested to revise the design specification for collision as the vessel size and bow configurations are changed a bit in recent years. Cerik et al. (2016) inspected denting damages in tubular members for low mass impact. Both experimental and numerical investigation was incorporated. It is observed that the numerical outcome has satisfactory similarity with the experimental test models. Local denting is considered along with the load-indenter displacements are also delineated. The deformation characteristics of single member with clamps are vastly studied. Cho et al. (2010) investigated the denting damage consequences of tubular members in offshore structures. Experimental denting test along with bending test was directed and a relation between dent depth and denting force was established. Moreover the residual strength of dented tubes can be predicted through the established equation which was derived from the relation of residual strength and bending moment. Cho et al. (2015) deliberated the response characteristics of impact loading in the context of tubular members in maritime structures. Drop test along with statistical analysis was carried out to predict the behavior against dynamic impact loading. Consequences of local denting and global bending were discussed elaborately from the experimental and numerical outcome. Minor dented damages due to ship collision could be the cause of major reduction of ultimate capacity and it was illustrated by Pacheco and Durkin (1988). Moreover, a geometrically ideal dent can represent the actual damage in case of finite element analysis and it is also delineated in this research. Li et al. (2013) focused on the ship collision consequences on the tubular members of Jacket structures. Both elastic and plastic behaviors are elaborated. Different scenario of ship collisions were assessed for substantial conclusions. It is reported that in case of larger deformation, the standard guideline underestimate resistance-indentation relationship. The elastic response from the vessel-platform impacts should be noted more carefully to comprehend significant effect of it in terms of energy absorption during impact.

The damage detection can be considered as a problem of system identification or an optimization inverse problem (Farrar and Worden, 2012). The optimization techniques can be used to quantify the unknown parameters of the damage. Over the last few decades, the adaptive meta heuristic and predicting control methods have become more widely used in several scientific research especially in the damage detection under ambient vibration (Miguel et al, 2012) and in the optimization design of fixed jacket offshore platform under environmental loads (Nasseri et al, 2014). Although much have been reported, very little have been done for the dent examination of circular member in the jacket structure which is the most common problem type to investigate the structural damage after ship collision. Intelligent computational techniques such as metaheuristics can be served to highlight areas where the sensor technologies and structural integrity monitoring techniques might be useful.

In this work, an optimization problem for offshore dent detection is posed to find percentages of dent in element diameters and impact angles. Five well established self-adaptive metaheuristics include JADE (Zhang & Sanderson, 2009), CMAES (Hansen, Muller et al., 2003), SHADE (Tanabe and Fukunaga, 2013), L-SHADE (Tanabe and Fukunaga, 2014), and ASCDE (Bureerat and Pholdee, 2017) are used to perform detection within a simulated damage scenario of a finite element reduced model of an offshore jacket platform. The efficiency and accuracy of each optimizer is discussed. Finally, the main conclusions and recommendations of future work are summarized.

2 DESCRIPTION OF AN OFFSHORE JACKET PLATFORM FOR THE ADAPTIVE META-HEURISTIC ANALYSIS

2.1 *Fixed offshore platform model*

The offshore jacket platform is modeled with the 3D frame elements to conduct the adaptive meta-heuristic analysis. The physical configuration of the jacket platform in side view along with the front view is available in Figure 1.

The offshore jacket platform is simulated to operate at the water depth 65.53 m. The topside of the jacket platform emerges above the sea water level 17.68 m. The bracing system of the jacket platform consists of two types (single bracing and the K bracing). The bottom bracing systems of the jacket structure is the K-bracing so as to compromise the connecting angle between chord and brace members more than 30 degrees to avoid the welding problem. The remaining of bracing systems is the single bracing to optimize the structural weight.

The top side of the oil and gas platform is simplified as a lumped mass for finite element analysis. The total weight of the topside is 2500 tons, which is equally distributed over four legs where each leg, is conveying 625 tons at the top connection of the jacket structures platform.

The structural members of the offshore jacket platform are utilized according to the design specifications (DNV-RP-C203, 2001). This jacket structure consists of 11 groups of tubular cross section. The element properties are listed in Table 1. All the structural members have the same material density and elasticity modulus. The location of structural elements with different colors for each groups are illustrated in Figure 2. The annotations of the colors are clarified substantially in Table 1. The formations of the groups are based on the outer diameter and wall thickness of the tubular section.

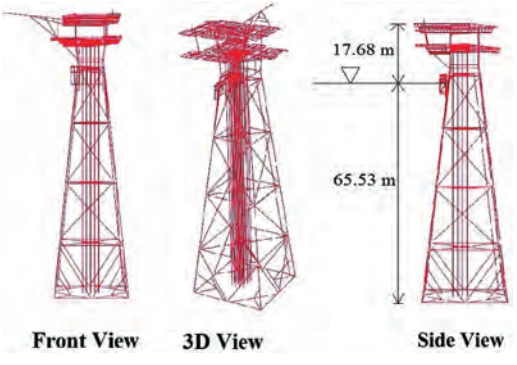


Figure 1. 3D and side views of offshore platform.

Table 1. Element properties.

Group name	Color indicator	Outside diameter	Thickness
G1		1.067	0.038
G2		0.457	0.010
G3		0.406	0.013
G4		0.356	0.010
G5		0.457	0.013
G6		0.356	0.013
G7		0.406	0.016
G8		0.324	0.010
G9		0.559	0.013
G10		0.559	0.019
G11		0.610	0.025

Modulus of Elasticity, E = 2.1 *10^{11} N/m^2;
Density of Steel = 7833 kg/m^3.

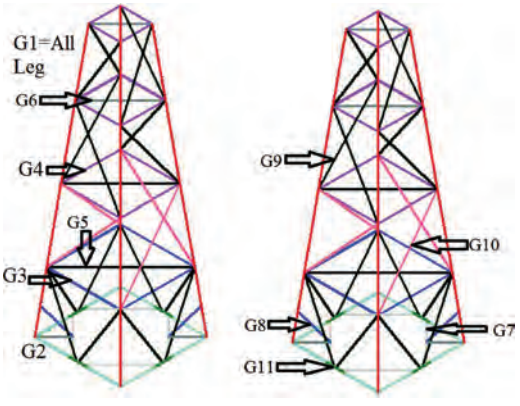


Figure 2. Member group specifications.

2.2 *Denting data and assumptions*

The assumption of denting members in the jacket platform comes from the consequences due to ship collision. There members from two different groups are assumed to be influenced due to ship collision. Member no 10 and 20 from G1 group and member no 88 from G9 group are the dented member. The vessel can impact to jacket platform in two directions either through bow or starboard direction. The location of dented members and impact direction are present in the Figure 3. Two members from leg and one bracing are impacted in according to the Figure 4. Denting is generally defined in terms of percentages of diameter reduc-

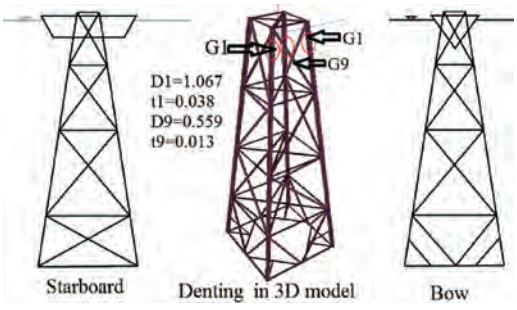


Figure 3. Ship collision and denting in members.

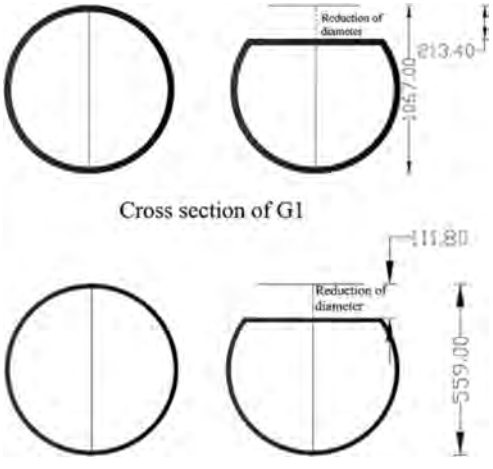


Figure 4. Denting sample in G1 and G9.

Table 2. Element groups of the jacket structure.

Group name	Element numbers	Outside diameter
G1	1–20	In element 10 & 20
G2	21–28	No denting
G3	29–32	No denting
G4	33–44	No denting
G5	45–52	No denting
G6	53–60, 117–120	No denting
G7	61–68	No denting
G8	69–72	No denting
G9	73–80, 85–92	In element 88
G10	81–84	No denting
G11	93–116	No denting

For members 41–44 and 57–60, are the members above the water level.

tion as per different code of practices (ISO 19901-3, 2014).

A sample of denting configuration for both type of members G1 and G9 are illustrated in Figure 4.

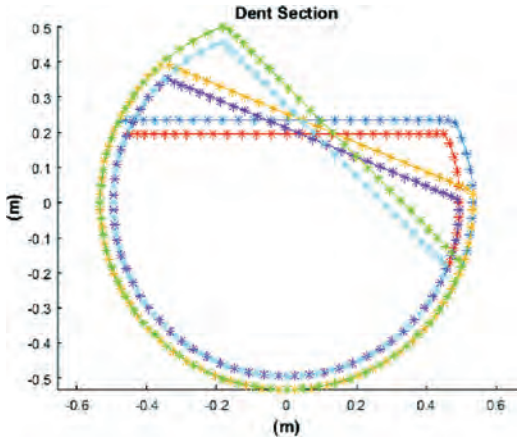


Figure 5. Denting of members for rotation of axis.

In this sample, 20% denting is contemplated. The diameters of the members as well as the dented length are identified in the following figure as well. Due to the denting effect, the stiffness of the member will be changed because of the changing of cross-sectional area and the moment of inertia but the mass will be unchanged.

It is assumed in this study that a single point of dented section will affect the whole member length. Moreover in reality, it is often not fixed whether the denting will occur along a certain axis or not.

As shown in Figure 5, the denting can occur in different angles if vessel impacts to the platform in the different direction of attack angle. The different local angle of denting member, as shown in Figure 5, can have different moments of inertia about two different axes. But the area will remain the same. For this reason, the angle for denting direction is also considered in this investigation and the range of the angle is between 0 and 360 degree. That is the ship can attack the offshore jacket platform from any direction and the denting can be occurred in any angles.

3 OPTIMIZATION PROBLEM OF OFFSHORE DENT DETECTION AND NUMERICAL EXPERIMENT SETUP

3.1 *Frequency and mode shape changes*

Dent detection of the offshore structure as detailed in the previous section is performed using vibration based damage detection. The main concept of vibration based damage detection is that updated mechanical properties of a mathematical model such as a finite element model and the modal data such as natural frequencies of the model agrees well the measured modal data. Damage of the

structure is identified by the detecting changes of the mechanical properties.

In this work, an optimization problem for offshore dent detection is posed to find percentages of dent in element diameters (P_d) and impact angles (θ) which consequently lead to the changes of mechanical properties (cross section area, second moment of area along y and z direction) and natural frequencies of the offshore structure. Since the natural frequencies can be accurately measured, an objective function in this case can be constructed in terms of changes in natural frequencies.

The percentage of dent damage in the tabular element can be found by solving an optimization problem to minimize the root mean square error (RMSE) between natural frequencies measured from the dented structure and natural frequencies computed by using the finite element model. The problem can be expressed as:

$$\text{Min: } f(\mathbf{x}) = \sqrt{\frac{\sum_{j=1}^{n_{mode}} \left(\omega_{j,damage} - \omega_{j,computed}\right)^2}{n_{mode}}} \quad (1)$$

where $\omega_{j,damage}$ and $\omega_{j,computed}$ are the structural natural frequency of mode j obtained from a dented structure and that from the finite element model respectively. A vector $\mathbf{x}$ is a set of design variables including percentages of dent in element diameters and impact angles ($\mathbf{x} = \{P_{d1}, P_{d2}, \ldots, P_{d,n}, \theta_1, \theta_2, \ldots, \theta_n\}^T$).

In this work, only element numbers 5, 10, 15, 20, 86, 88, 90 and 92 which are located at the sea level are set to have dent possibility. Therefore, the total number of design variables is set to be 16 (8 for percentages of dent in element diameters and other 8 for impact angles of the elements). The possibility of percentages of dent is set in rang of [0, 0.7] while impact angle is set in the set of {0, 22.5, 45}.

3.2 *Numerical experiment*

To investigate the search performance of optimization methods on solving the proposed problem of dent detection of the offshore structure, the percentage of dent in element diameter and impact angle are pre-defined while natural frequencies are simulated by means of finite element analysis instead of using real measuring data. The percentages of dent in element diameters and impact angles are set as 0.3 percent dent at elements 10, 20 and 88 with impact angles of 0, 22.5 and 45 degrees respectively. Natural frequencies for the first six modes of the dented and undented elements (as shown in Table 3) are used for the objective function calculation.

Table 3. Natural Frequency (Hz) up to six mode.

Modes	Undented	Dented
1	0.680	0.674
2	0.690	0.714
3	0.919	0.989
4	2.506	2.507
5	2.527	2.527
6	3.179	3.145

To minimize the objective function, five well-established self-adaptive meta-heuristics (MHs) are used. Details and notations of these methods are available in the literature in the corresponding references of the methods and will not be detailed here. The MHs used include:

- Adaptive Differential Evolution (JADE) (Zhang & Sanderson, 2009).
- Evolution Strategy with Covariance Matrix Adaptation (CMAES) (Hansen, Muller et al., 2003).
- Success-History Based Adaptive Differential Evolution (SHADE) (Tanabe & Fukunaga, 2013).
- SHADE with Linear Population Size Reduction (L-SHADE) (Tanabe & Fukunaga, 2014)
- Adaptive Sine Cosine algorithm with integrating Differential Evolution mutation (ASCDE) (Bureerat and Pholdee, 2017)

Each optimizer is used to solve the offshore structure dent detection test problem for 10 optimization runs. The population size is set to be 30 whereas the number of iterations is set to be 300. All methods will be terminated with two criteria: the maximum numbers of functions evaluation as 30×300, and the objective function value being less than or equal to 1×10^{-3}.

4 RESULTS AND DISCUSSIONS

After performing 10 optimization runs of all MHs on solving the offshore structure dent detection optimization test problem, the results obtained are given in Table 4. The mean of objective function are used to measure the algorithm rate of convergence in cases that the objective function threshold (1×10^{-3}) is not reached during searching. Otherwise, the mean number of FE runs is used as an indicator. The number of successful runs out of 10 runs is used to measure the search consistency. The algorithm that is terminated by the objective function threshold is obviously superior and any run being stopped with this criterion is considered a successful run (Bureerat and Pholdee, 2017).

Table 4. Comparison results among each optimizer.

Optimizers	Mean Obj.	Mean FEs	No. of successful run
CMAES	0.02428	8166	1
JADE	0.03626	8268	1
SHADE	0.03886	8367	1
LSHADE	0.02903	9000	0
ASCA	0.00095	2148	8

Table 5. The best results on finding the percentages of dent and impact angles obtained from all optimizers.

No	Simulated dent	CMAES	JADE	SHADE	LSHADE	ASCA
10	0.3	0.36	0.50	0.47	0.19	0.26
20	0.3	0.07	0.17	0.18	0.01	0.35
88	0.3	0.29	0.23	0.17	0.00	0.30
5	0	0.13	0.13	0.19	0.27	0.05
15	0	0.00	0.00	0.00	0.16	0.00
86	0	0.27	0.06	0.08	0.70	0.00
90	0	0.00	0.00	0.00	0.01	0.00
92	0	0.00	0.00	0.00	0.01	0.00
No	Simulated impact angle	CMAES	JADE	SHADE	LSHADE	ASCA
10	0	45	22.5	45	0	45
20	22.5	45	45	22.5	45	0
88	45	22.5	22.5	45	–	0
5	–	45	22.5	45	45	0
15	–	0	–	0	45	–
86	–	22.5	0	22.5	22.5	–
90	–	–	–	–	0	–
92	–	–	–	–	0	–
ω_1(hz)	0.674	0.674	0.675	0.673	0.672	0.673
ω_2(hz)	0.714	0.715	0.713	0.714	0.716	0.714
ω_3(hz)	0.989	0.989	0.988	0.989	0.988	0.989
ω_4(hz)	2.507	2.507	2.507	2.507	2.505	2.507
ω_5(hz)	2.527	2.527	2.527	2.527	2.527	2.527
ω_6(hz)	3.145	3.146	3.146	3.145	3.149	3.145

From Table 4, it can be seen that the best performer based on mean objective function values is ASCA while the second best and the third best algorithms are CMAES and LSHADE, respectively. When considering the number of successful runs, ASCA is said to be the most efficient optimizer which can detect the percentage of dent in element diameters and impact angles for 8 times out of totally 10 optimization runs with the average of 2,148 function evaluations.

Table 5 shows the best results on finding the percentage of dent and impact angles obtained from all optimizers. It was found that MHs will only detect impact angle values in the elements having percentage of dent higher than zeros. From Table 5, it can be observed that ASCA can correctly detect the percentages of dent in the offshore structure while the others failed to achieve such results. For the impact angles, the results of all optimizer are not accurate and need further improvement. One idea of such improvement is to introduce some nodal displacements that can be picked out of numerical mode shapes in the objective function, which will be explored and presented in future work.

5 CONCLUSIONS AND RECOMMENDATIONS

Five meta-heuristics optimizers were tested for the the dent damage of circular member in the jacket structure. The damage detection problems are based on vibration measurement and can be treated as an inverse optimization problem. The comparative results reveal that the ASCA is outstanding for predicting denting diameter but not denting angles. The results from ASCA could be use as the baseline for further improvement and investigation of dent damage examination using meta-heuristics.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation under the Marie Skłodowska-Curie grant agreement No. 730888.

REFERENCES

Bruin, W.M. 1995. Assessment of the residual strength and repair of dent-damaged offshore platform bracing. Lehigh University.

Bureerat, S., & Pholdee, N. 2017. Adaptive Sine Cosine Algorithm Integrated with Differential Evolution for Structural Damage Detection. In International Conference on Computational Science and Its Applications on 3–6 July, 2017. Trieste: Italy

Cerik, B.C., Shin, H.K., & Cho, S.R. 2016. A comparative study on damage assessment of tubular members subjected to mass impact. *Marine Structures*, 46: 1–29.

Cho, S.R., Kwon, J.S., & Kwak, D.I. 2010. Structural Characteristics of Damaged Offshore Tubular Members. *Journal of Ocean Engineering and Technology*, 24(4): 1–7.

Cho, S.R., Le, D.N.C., & Seo, B.S. 2015. Response of tubular structures under dynamic impact loadings. In *Proceedings of SNAK Autumn Conference, Society of Naval Architects of Korea, Geoje, Korea*.

Cosham, A., & Hopkins, P. 2004. The effect of dents in pipelines—guidance in the pipeline defect assessment manual. *International Journal of Pressure Vessels and Piping*, 81(2): 127–139.

DNV-RP-C203 2001. Recommended practice. Fatigue strength analysis of offshore structures. Høvik: Det Norske Veritas.

Farrar, C.R., & Worden, K. (First Edition) 2012. Structural health monitoring: a machine learning perspective: John Wiley & Sons.

Hansen, N., Müller, S.D., & Koumoutsakos, P. 2003. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary computation*, 11(1): 1–18.

ISO 19901–3 2014, Petroleum and natural gas industries, Specific requirements for offshore structures, Part 3: Topsides structures. Geneva, Switzerland: ISO

Karamanos, S.A., & Andreadakis, K.P. 2006. Denting of internally pressurized tubes under lateral loads. *International Journal of Mechanical Sciences*, 48(10): 1080–1094.

Khedmati, M.R., & Nazari, M. 2012. A numerical investigation into strength and deformation characteristics of preloaded tubular members under lateral impact loads. *Marine Structures*, *25*(1): 33–57.

Li, L., Hu, Z., & Jiang, Z. 2013. Plastic and elastic responses of a jacket platform subjected to ship impacts. *Mathematical Problems in Engineering*, 2013.

May, P., Sanderson, D., Sharp, J.V., & Stacey, A. 2008. Structural integrity monitoring: Review and appraisal of current technologies for offshore applications. In *ASME 2008 27th International Conference on Offshore Mechanics and Arctic Engineering* 15–20 June 2008. Estoril, Portugal.

Miguel, L.F.F., Miguel, L.F.F., Kaminski, J., & Riera, J.D. 2012. Damage detection under ambient vibration by harmony search algorithm. *Expert Systems with Applications*, *39*(10): 9704–9714.

Nasseri, T., Shabakhty, N., & Afshar, M.H. 2014. Study of Fixed Jacket Offshore Platform in the Optimization Design Process under Environmental Loads. *International Journal of Maritime Technology*, 2: 75–84.

Pacheco, L.A., & Durkin, S. 1988. Denting and collapse of tubular members—a numerical and experimental study. *International journal of mechanical sciences*, 30(5): 317–331.

Storheim, M., & Amdahl, J. 2014. Design of offshore structures against accidental ship collisions. *Marine Structures*, 37: 135–172.

Tanabe, R., & Fukunaga, A. 2013. Evaluating the performance of SHADE on CEC 2013 benchmark problems. In *Evolutionary Computation (CEC), 2013 IEEE Congress on* 20–23 June 2013. Cancun: Mexico.

Tanabe, R., & Fukunaga, A.S. 2014. Improving the search performance of SHADE using linear population size reduction. In *Evolutionary Computation (CEC), 2014 IEEE Congress on 6–11 July 2014*. Beijing: China.

Travanca, J., & Hao, H. 2014. Dynamics of steel offshore platforms under ship impact. *Applied Ocean Research*, 47: 352–372.

Zhang, J., & Sanderson, A.C. 2009. JADE: adaptive differential evolution with optional external archive. *IEEE Transactions on evolutionary computation*, 13(5): 945–958.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Strategic view of an assets health index for making long-term decisions in different industries

A. De la Fuente, A. Guillén, A. Crespo, A. Sola, J. Gómez & P. Moreu
Department of Industrial Management, School of Engineering, University of Seville, Seville, Spain

V. Gonzalez-Prida
Department of Industrial Management, School of Engineering, University of Seville, Seville, Spain
UNED, Madrid, Spain

ABSTRACT: An Asset Health Index (AHI) is a tool that processes data about asset's condition. That index is intended to explore if alterations can be generated in the health of the asset along its life cycle. These data can be obtained during the asset's operation, but they can also come from other information sources such as geographical information systems, supplier's reliability records, relevant external agent's records, etc. The tool (AHI) provides an objective point of view in order to justify, for instance, the extension of an asset useful life, or in order to identify which assets from a fleet are candidates for an early replacement as a consequence of a premature aging. This paper develops a model applicable to different classes of equipment and industrial sectors. A review of the main cases where the asset health index has been applied is included. Likewise, advantages and disadvantages in the application of this kind of tools are revealed, providing a guide for a research line related to the general application of this tool.

1 INTRODUCTION

Nowadays, network operators are facing many challenges in their assets management. There is an increasing trend for stakeholders (safety, reliability, environment, and financial impact) while assets are aging, increasing the risk of failure. The need to estimate the expected time to failure becomes more relevant every day, and planning for an optimal replacement or maintenance program to renew assets becomes even more essential. By having a large amount of assets, the maintenance manager's challenge is to decide which assets require more attention and what actions should be taken. The complexity of this decision increases because each asset class has different failure modes, and each failure has different consequences in the asset network (Vermeer et al. 2015).

The objective of this contribution is to highlight the most relevant recent studies related to the asset health index. The concept of asset health index (AHI) and its application appear throughout this document. The different models will be divided into three parts. The first part deals with data gathering and their treatment, the second part corresponds to the index composition, and the third part is the output of results and recommendations related to the index value. As a final part of the contribution, limitations of the models and the future scope of asset health index are discussed.

2 CONCEPT OF AN ASSETS HEALTH INDEX

An Asset Health Index (AHI) is an asset score, which is designed, in some way, to reflect or characterize the asset's condition and thus, its performance in terms of fulfilling the role established by the organization.

AHI represent a practical method to quantify the general health of a complex asset. Most of these assets are composed of multiple subsystems, and each subsystem can be characterized by multiple modes of degradation and failure. In some cases, it may be considered that an asset has reached the end of its useful life, when several subsystems have reached a state of deterioration that prevents the continuity of service required by the business (Hjartarson & Otal 2006). Therefore, the health index, based on the results of operational observations, field inspections and laboratory tests, produces a single objective and quantitative indicator. It may be used as a tool to manage assets, to identify capital investment needs and maintenance programs (Naderian et al. 2008). In addition to condition and operation factors, the health index requires also to contain static factors linked to its location. That means, when environmental conditions are changed independently of the asset itself or the lack of any other change over time (Scatiggio & Pompili 2013).

The critical objectives in the formulation of a complex Health Index are as follows (Hjartarson & Otal 2006):

- The index should be indicative of the asset suitability for a continued service and representative of the overall asset health.
- The index should contain objective and verifiable measures of asset condition, as opposed to subjective observations.
- The index should be understandable and readily interpreted.

3 MODELS

Next, four different models from the literature (proposed for the calculation of an asset health index) will be studied. In general terms, they all have inputs to the model that can be data related to condition, equipment operation, the availability of spare parts used in the maintenance and, in some cases, information from the geographic location.

For the algorithms used in the index calculation, it will be seen how in the different models, all the information sources are integrated by weighting factors and depending on the maturity level of the model implemented in each sector. Likewise, the output study for each model and their recommendations is shown.

The following scheme (Figure 1) represents the concept of an AHI. It tries to compile the different inputs to the model from the literature consulted, and the different outputs for making long-term decisions (Azmi et al. 2017). It is important to highlight that this paper is focused on making long-term decisions. In any case, there are also AHI models that are used as tools in the field of Prognostics and Health Management (P.H.M.) (Ludovic et al. 2011) (Abichou et al. 2012; Abichou et al. 2015).

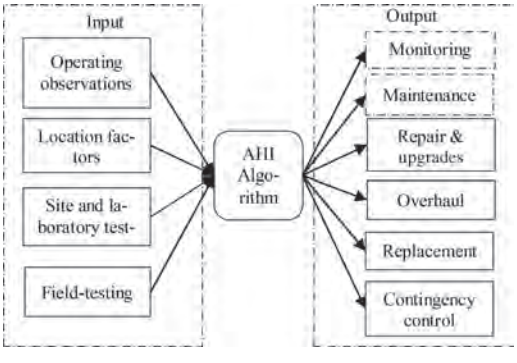


Figure 1. Concept of the health index within an asset management framework.

3.1 *Asset health index calculation model by Kinetrics*

The model developed by Kinetrics, Canada, proposes the overall assessment of transformers condition. The inputs to the model are data from different variables related to the operation and condition of the equipment throughout its useful life. The calculations for each variable, as well as detailed assessment are used to normalize the values, weighting them with corresponding weights and building a personalized asset health index.

3.1.1 *Inputs*

The inputs of the model, are the historical data of the variables of operation (load, number of operations, etc.), the results of oil samples labs tests (Oil quality, content in dissolved gases, acidity, etc.) and on-site tests carried out by technicians, such as insulation tests, thermography, corrosion status, etc.

3.1.2 *Index calculation methodology*

The methodology proposed for the index calculation is based on the normalization of each variable into a value between 0 and 4, together with a variable weight for the final composition in a single indicator.

As an example of standardization for the results obtained at a laboratory test, the following chart (Table 1) shows the link between concentration of dissolved gases in transformer oil and aging. In the table, concentrations of different dissolved gases are weighted, according to their relationship with the aging of the asset. Therefore, gases with greater weight are those that appear when the asset has reached a certain level of aging (Naderian et al. 2008).

Equation 1 below, calculated from the results of gases dissolved in oil, refers to the variable value which is one of the inputs to the AHI.

$$LTC\ Oil\ Quality = \frac{\sum_{i=1}^{4} S_i x W_i}{\sum_{I=1}^{4} Wi} \quad (1)$$

For the variables, the author proposes a weight between 1 and 10. Values close to 1 are assigned

Table 1. Concentration in ppm of gas dissolved in oil.

	Gas in oil concentration in tap changer				
Gas	ppm	ppm	ppm	ppm	W_i
CH_4	<50	50–150	150–250	≥250	3
C_2H_6	<30	30–50	50–100	≥100	3
C_2H_4	<100	100–200	200–500	≥500	3
C_2H_4	<10	10–20	20–25	≥25	3
Score (S_i)	1	2	3	4	

to variables whose relationship with the aging of the equipment is very small or null. On the other hand, for variables that take higher values (higher than 5 points), they are condition variables that more accurately reflect the aging of the equipment. The operating variables, such as the load factor of transformers and the power factor, are also related to the equipment aging speed, because they are good indicators showing when the equipment operates outside the design conditions.

3.1.3 *Model outputs*

For the model output, the author proposes a composition of all normalized variables in a single indicator ranging between 0 and 100. The value of 100 corresponds to a value of new equipment and the value of zero refers to a piece of equipment that has reached the end of its useful life, requiring to be replaced because it is already out of service. The following equation 2 is proposed by the author for calculating the health index (Naderian et al. 2008).

$$HI = 60\% \frac{\sum_{j=1}^{17} K_j HIF_j}{\sum_{j=1}^{17} 4K_j} + 40\% \frac{\sum_{j=18}^{20} K_j HIF_j}{\sum_{j=18}^{20} 4K_j} \quad (2)$$

The index value is related to a failure probability of the equipment, being divided into different ranges with their respective interpretations and recommendations. For the different index output ranges, some recommendations and measures are proposed in order to be taken into account for decision making in maintenance management. The following Figure 2 shows the relationship between the health index and the probability of failure (Naderian et al. 2009).

3.2 *Asset health index calculation model by DNV GL*

This model developed by DNV GL, Arnhem, the Netherlands, proposes a methodology for the calculation of a health index, calculated from the maximum admissible failure rates for the business, together with the asset criticality, in order to obtain an index to prioritize maintenance, overhaul and substitutions of parts. It uses as input variables the estimated useful life of the equipment, the current age and condition variables (load, on-site condition analysis, maintenance number, etc.). The output of this model is the remaining useful life of the equipment in years (Vermeer et al. 2015).

3.2.1 *Inputs*

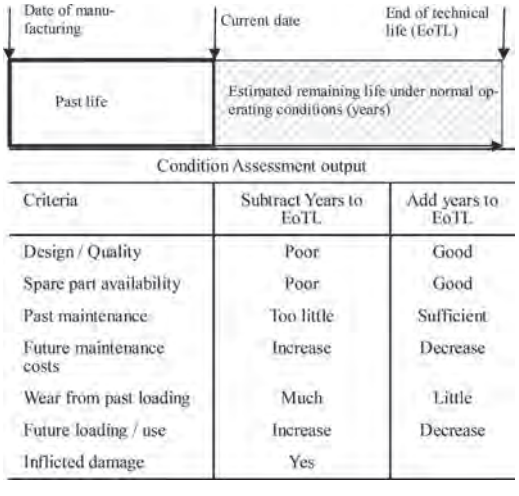
The model uses the useful life of the equipment as static data; this allows making a first estimation that will be corrected later with the information of the asset's condition and, at the same time, with the failure modes that appear throughout the asset life cycle.

3.2.2 *Index calculation methodology*

The methodology proposed by the author, is separated into three large blocks, depending on the type of data entry. The blocks are called as degradation function, static function and condition function. The model output is the relationship between the different functions for determining the health index, which is in this case the equipment remaining life.

The model application requires a previous estimation of the asset average age based on historical data. This average life becomes a technical life average that is later corrected with the specific condition data of the asset. In a simple way, the model increases or decreases the end of the asset technical life, depending on the real asset condition at the moment of its analysis, Figure 3, (Vermeer et al. 2015).

Figure 2. Asset health index ranges and the relationship with the probability of failure.



Condition Assessment output

Criteria	Subtract Years to EoTL	Add years to EoTL
Design / Quality	Poor	Good
Spare part availability	Poor	Good
Past maintenance	Too little	Sufficient
Future maintenance costs	Increase	Decrease
Wear from past loading	Much	Little
Future loading / use	Increase	Decrease
Inflicted damage	Yes	

Figure 3. Estimated remaining life corrected with the specific condition data of the asset.

3.2.3 *Model outputs*

Once calculated all proposed methodological functions, they are combined in order to provide the result of the asset remaining life. In order to make the calculation, first, the static function is calculated with condition function in series, while the degradation function is calculated in parallel with the others. Like the previous model, the index output provides an approximated value of the asset health status, which is in this case the equipment remaining life. In Figure 4, the combination between the different functions is observed for the calculation of the asset health index proposed by the author.

3.3 *Asset health index calculation model by TERNA*

This model developed by Terna Rete, Italy, proposes the calculation of the equipment health index based on static and dynamic parameters. Static parameters are associated with the location where the equipment is located, which are invariable in time and independent of the asset, for example, the recurrence of catastrophic phenomena, the probability of electrical storm, etc. Dynamic parameters are associated to the equipment and can be measured in situ by functional and visual tests, as well as in laboratory tests by analysis of oil samples, lubricants, etc.

The output of the model is an index between 0 and 0.5 which refer to the state as new and critical respectively. That is intended to justify technically and economically, making decisions for the investment of capital in replacement of equipment (Scatiggio et al. 2016; Scatiggio & Pompili 2013).

3.3.1 *Inputs*

The model author proposes static and dynamic variables for the model's inputs. Static variables do not depend on the asset itself but depend on the location (lightning frequency, catastrophic events, etc.). The dynamic variables proposed depend on the asset, and their value changes with the asset ages. Therefore, by capturing the change over time and comparing the maximum and minimum admissible for each kind of equipment, the condition status can be estimated at each moment of the asset life. The following condition parameters are those that are taken into account as inputs to the model (Pompili & Scatiggio 2015), each parameter is known as Health Index (HI).

- HI dielectric: parameters related to dielectric and thermal condition, as it may be obtained from dissolved gas analysis. These parameters are able to provide information on electrical (partial discharges, low energy discharges, arcing) and thermal problems (hot spots, overloads);
- HI thermal: parameters related to pure thermal condition of the insulating paper, as they may be obtained from the CO_2, CO and further periodical determinations;
- HI mechanical: parameters related to mechanical condition of the transformer, as they may be obtained from on-site electrical tests (inductance measurements, Sweep Frequency Response Analysis or SFRA, Frequency Domain Spectroscopy or PDC/FDS);
- HI oil: parameters related to insulating oil condition, as they may be obtained by water content, acidity, 50–60 Hz Breakdown Voltage (BDV) and Dielectric Dissipation Factor (DDF) determinations.

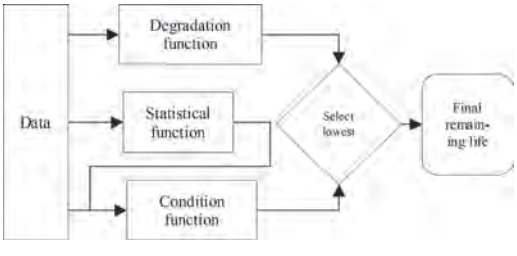


Figure 4. Schematic of how the functions are combined to give a health index.

3.3.2 *Index calculation methodology*

Due to the fact that different condition factors are very different from each other, they must first be standardized with their corresponding weights. In order to transform the value into a non-dimensional number, international guidelines and regulations (IEC, IEEE, CIGRE, etc.) are used.

Once the parameters have been standardized, from the following equation 3, the HI is calculated:

$$HI = \frac{HI_{dielectric} + HI_{thermal} + HI_{mechanical} + HI_{oil}}{HI_{MAX}} \quad (3)$$

where HI_{MAX} is a prefixed number and, as a consequence, the HI of each asset may be expressed per units (p.u.).

3.3.3 *Model outputs*

The model output is a HI value between 0 and 0.5. Higher and lower HI values are associated, respectively, to lower or higher levels of asset reliability.

In dependence on their HI, assets are classified in four classes. In Table 2, assets classified in "very good" and "good" condition may be managed following the common and standard maintenance practices, assets classified as "fair" or "doubtful" need an increase of analysis frequency or a deeper investigation (Scatiggio et al. 2016).

Table 2. Health Index (HI) evaluation.

Health Index (HI)	Condition
0–0,10	Very Good
0,10–0,20	Good
0,20–0,30	Fair
>0,30	Doubtful

The models that will be introduced in the paper are relevant, among other things, because:

- Asset managers need models to study options that maximise the value of an asset as it approaches the end of its useful life. Options may include (for example) changing the operating regime, partial asset replacement/refurbishment to extend useful life, or an indefinite ongoing 'patch-and-continue' programme, perhaps involving suppliers to provide necessary parts or services.
- Predicted performance supported by knowledge and asset information is available in many companies—normally based on good understanding of how assets degrade—but not incorporated in formal processes for capital investment. These models contribute to the decision process that seeks the optimal life cycle value.

4 CHALLENGES OF AHI APPLICATION

Currently, in order to respond the increasingly demanding requirements in terms of asset management, the application of AHI models offer the possibility to improve the process of decision making in maintenance situations. After a review of the literature, the best practices agree that using the asset health index offers the following advantages:

- Consolidate all information sources about the asset condition in a single integrated view of asset health.
- Provide an approaching indication of the asset at the end of its useful life.
- Condition assessment and asset performance.
- Report generation for maintenance attention.
- Needs identification at short and medium term for the replacement of individual equipment.
- Prediction of long-term needs replacement in large volumes of assets, identifying potential peaks with investment requirements.
- Identify problems, risks and opportunities for maintenance management.
- Provide information on asset deterioration trends that do not correspond to the rates of natural aging processes, which can be useful for planning appropriate maintenance strategies.
- Comparison between the assets condition by classes and locations, allowing taking actions in the operation and maintenance strategy of the organization.

On the other hand, any organization that decides to implement this tool in their strategic processes, with the purpose to improve its asset management, will have to take into account the below-mentioned considerations. Depending on the level of maturity of the organization, in some cases, they may be a challenge to overcome and, in others, an inconvenience to avoid or mitigate:

- The collection of data has a high cost. The capture of certain information requires a field technician in order to inspect and record the data.
- Uncertainty in evaluating asset conditions can create inconsistencies in the collection of data.
- Uncertainty about the return on investment, the valuation of costs and the financing of assets replacement or renewal, can make difficult to determine the information
- The lack of consistent and compatible methods to record, store and reference information can cause errors in the analytical phase.

Once these advantages and disadvantages have been seen, it's worth investigating the implementation of the AHI tool. The initial part of capture and processing of the data is critical; improving in this initial stage the assets management will ensure better results. For any organisation that decides to apply the AHI tools, its essential to incorporate in its asset management model, the condition study of the asset after replacement promoted by a decision based on the asset health. This will allow the learning and adjustment of the mathematical model based on their own experience, which will be benefited in better results in making long-term decisions.

5 CONCLUSIONS

Today, the use of tools for decision making about long-term renewal and replacement of equipment for organizations is quite extended. Thanks to Life-Cycle Cost analysis (LCC), it is possible to know from an economic point of view, the cost of an asset over its useful life and to estimate the time for replacement if needed. The disadvantage in many cases, is the large amount of variables that must be handled when estimating the real cost of an asset over its useful life, generating a scenario of high uncertainty (Durairaj et al. 2002). At that moment, it is where the AHI comes into play as a support tool, having a completely different calculation methodology, estimated from lab tests in order to know the asset condition, visual inspections,

operation and maintenance history and the age of the equipment and its components.

The roadmap for the definition of an AHI model is applicable to different kinds of equipment. It is currently under development and such development is generating the need to open new lines of research, in parallel to what is currently implemented in the field of electrical networks and more specifically in electrical transformers.

REFERENCES

Abichou, B., A. Voisin, and B. Iung. 2015. "Choquet Integral Capacity Calculus for Health Index Estimation of Multi-Level Industrial Systems." *IMA Journal of Management Mathematics* 26(2):205–24. Retrieved February 7, 2018 (https://academic.oup.com/imaman/article-abstract/doi/10.1093/imaman/dpu006/644089/Choquet-integral-capacity-calculus-for-health).

Abichou, B., A. Voisin, B. Iung, P. Do Van, and N. Kosayyer. 2012. "Choquet Integral Capacities-Based Data Fusion for System Health Monitoring." *IFAC Proceedings Volumes* 45(20):31–36. Retrieved February 7, 2018 (https://www.sciencedirect.com/science/article/pii/S1474667016347279).

Azmi, A., J. Jasni, N. Azis, and M.Z. A.Ab. Kadir. 2017. "Evolution of Transformer Health Index in the Form of Mathematical Equation." *Renewable and Sustainable Energy Reviews* 76(March):687–700. Retrieved (http://linkinghub.elsevier.com/retrieve/pii/S1364032117304306).

Durairaj, Senthil Kumaran, S.K. Ong, A.Y.C. Nee, and R.B.H. Tan. 2002. "Evaluation of Life Cycle Cost Analysis Methodologies." *Corporate Environmental Strategy* 9(1):30–39. Retrieved December 5, 2017 (https://www.sciencedirect.com/science/article/pii/S1066793801001415).

Hjartarson, Thor and Shawn Otal. 2006. "Predicting Future Asset Condition Based on Current Health Index and Maintenance Level." in *ESMO 2006–2006 IEEE 11th International Conference on Transmission & Distribution Construction, Operation and Live-Line Maintenance*.

Ludovic Rizzolo, Bouthaina Abichou, Alexandre Voisin, Naïm Kosayyer. 2011. "Aggregation of Health Assessment Indicators of Industrial Systems." in *The 7th conference of the European Society for Fuzzy Logic and Technology, EUSFLAT–2011*. Retrieved February 7, 2018 (https://hal.archives-ouvertes.fr/hal-00605121/).

Naderian, A., S. Cress, R. Piercy, F. Wang, and J. Service. 2008. "An Approach to Determine the Health Index of Power Transformers." *Conference Record of the 2008 IEEE International Symposium on Electrical Insulation* (July 2008):192–96. Retrieved (http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4570308).

Naderian, A., R. Piercy, S. Cress, J. Service, and W. Fan. 2009. "An Approach to Power Transformer Asset Management Using Health Index." *IEEE Electrical Insulation Magazine* Vol. 25(No. 2):2.

Pompili, M. and F. Scatiggio. 2015. "Classification in Iso-Attention Classes of Hv Transformer Fleets." *IEEE Transactions on Dielectrics and Electrical Insulation* 22(5):2676–83.

Scatiggio, F., M. Rebolini, Terna Rete Italia, and M. Pompili. 2016. "Health Index: The Last Frontier of TSO's Asset Management." 1–9.

Scatiggio, Fabio and Massimo Pompili. 2013. "Health Index : The TERNA's Practical Approach for Transformers Fleet Management." (June):178–82.

Vermeer, M., J. Wetzer, P. van der Wielen, E. de Haan, and E. de Meulemeester. 2015. "Asset-Management Decision-Support Modeling, Using a Health and Risk Model." *PowerTech, 2015 IEEE Eindhoven* 1–6.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Acoustic emission based fault diagnosis via a novel deep convolutional neural network method

D. González Toledo & V. Meruane
University of Chile, Santiago, Chile

E. López Droguett
University of Chile, Santiago, Chile
University of Maryland, College Park, USA

M. Modarres
University of Maryland, College Park, USA

ABSTRACT: Acoustic Emission (AE) has seen increased popularity in applications involving machine condition monitoring. AE applications usually involve higher sampling rate than vibration signals, not rare reaching 2 MHz. One of the main challenges involving AE based fault diagnosis is the need of preprocessing massive amounts of data generated by this technique, including engineering of appropriate features and dimensionality reduction so to be able to handle such massive datasets. In this paper, we propose a novel method based on Deep Convolutional Neural Networks (CNN) to handle raw AE signals for diagnosis of a system's health states. This method is flexible enough to not only handle the massive amount of AE data, but also to provide the means for automatic feature extraction by applying various filters to the raw AE signals, and thus identifying relevant frequencies related to different faults. The proposed CNN method is applied to fatigue crack detection on blades of an experimental rotor.

1 INTRODUCTION

Unscheduled maintenance of mechanical systems leads to loss of production and might as well affect safety. There are many ways to minimize this effect, some of them are: increase redundancy, programed maintenance to identify problems that could incur in extended downtimes and, condition monitoring (Rabiei et al., 2016).

A popular approach to condition monitoring is vibration analysis. However, vibration monitoring is usually less sensitive to detecting damages already developed, which pose a significant limitation in sensitive systems. On the other hand, Acoustic Emission techniques are gaining grounds because they can identify damage at early stages, with the tradeoff of introducing higher sample rates resulting in massive and higher data dimensionality.

Moreover, both AE and vibration monitoring require signal preprocessing and interpretation such as wavelets, fast Fourier transform and band filtering among others (Riaz et al., 2017), a labor intensive and expensive endeavor requiring specialized engineering expertise.

Machine learning techniques have become a popular choice for fault diagnosis and prognosis. Most of these shallow models heavily rely on manual feature identification and extraction (Ruiz-Gonzalez et al., 2014; Kane and Andhare, 2016; Li et al., 2016).

As discussed in (Verstraete et al., 2017), the performance of these methods is dependent on the quality of the hand-engineered features, which obviously requires significant understanding of the system's degradation processes.

To tackle these challenges, we propose a deep CNN-based method for fault diagnosis that operates on massive raw acoustic signals and allows for the automatic hierarchical "layer to layer" feature extraction to learn complex representations of the data.

The remainder of the paper is structured as follows. Section 2 introduces deep learning and CNNs and their architectural building blocks. Then, Section 3 discusses the proposed method, application and validation for fault diagnosis of an experimental rotor and compares its performance to a fully optimized shallow neural network. Section 4 presents some concluding remarks.

2 CONVOLUTIONAL NEURAL NETWORKS

2.1 *Artificial neural networks*

Artificial Neural Networks (ANNs) are comprised of simple units called neurons. Each of these neurons creates a linear combination for an input (x) between a weight (w) and a bias (b) parameters that are learned by the ANN.

Also, an activation function (f) adds the nonlinear behavior that allows to compute nontrivial responses. Then, the output (O) of a neuron is computed by Equation (1):

$$O(x) = f(wx + b) \tag{1}$$

2.2 *Deep learning*

Simply put, deep learning is a branch of the Machine Learning that uses many hidden layers to perform the learning. The deep learning based networks learn multiple features over the features learned by previous layers, integrating the concept of hierarchy between features implying different levels of abstraction (Deng and Yu, 2014). This is important to achieve high accuracy in tasks that have complex relationship among data such as image recognition and signal processing.

2.3 *Convolutional neural network overview*

Convolutional Neural Networks (CNNs) are a type of Neural Networks that are specialized for processing grid-topology data (Goodfellow, Bengio and Courville, 2017). CNNs have been shown to outperform shallow architectures in many image recognition tasks and have been applied to vibration based fault diagnosis (Verstraete et al., 2017).

The main characteristics of the CNNs are that the layers have sparse connectivity and parameter sharing. The first characteristic means that CNNs use filters that are considerably smaller than the input implying that the filters store less parameters than a shallow neural network and detect important features of the input. The second one implies that the filter weights in a convolutional layer are used multiple times across the input resulting in computationally efficient matrix multiplication.

2.3.1 *Convolutional layer*

As we are dealing with raw AE data, the convolution operation in the proposed model is also 1D. This means that the filters (w(t)) learned by the network are in the time domain and generate a filtered signal of the input (x(t)) highlighting features that represent the system's health state. The output signal of a 1D convolution, s(t), is computed as shown in Equation (2):

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)\, w(t-a) \tag{2}$$

2.3.2 *Batch normalization layer*

Batch Normalization (BN) is a method for accelerating the learning of deep networks by reducing internal covariate shift of the data (Ioffe and Szegedy, 2015). This is achieved by performing a normalization for each mini-batch with the learning of new normalization parameters: scale parameter gamma (γ) and shift parameter beta (β).

Given the p-dimension input to a BN layer $x = (x^{(1)},...,x^{(p)})$, the transformation is made with Equation (3) and Equation (4):

$$\hat{x}^{(i)} = \frac{x^{(i)} - E\left[x^{(i)}\right]}{\sqrt{Var\left[x^{(i)}\right]}} \tag{3}$$

$$y^{(i)} = \gamma^{(i)}\hat{x}^{(i)} + \beta^{(i)} \tag{4}$$

where Var[X] is the variance and E[X] is the expectation and are computed over the training set. Equation (3) is used to standardize features and accelerate convergence and Equation (4) restores the representation power of the network.

2.3.3 *Pooling layer*

Pooling layers are used to reduce the dimension of the input and achieve spatial invariance. This is usually accomplished by taking the maximum value of a pooling window and switching it for all values in that window. This lowers the resolution but taking only the most important feature.

2.3.4 *Activation function*

As discussed before, activation functions add nonlinear behavior to the network. In the proposed CNN architecture, we employ the Rectified Linear Units (ReLUs) as activation function for the convolutional layers as they provide increased sparsity compared with Tanh or Sigmoid activation functions, thus decreasing computation time (Maas et al., 2013). The commonly used ReLU is shown in Equation (5):

$$g(x) = \max(0, x) \tag{5}$$

For the fully connected layers (see Section 2.3.5), the softmax activation function is used to quantify the probability of a sample to correspond to a given system's health state. This function is displayed in Equation (6):

$$f_k(z) = \frac{e^{a_k}}{\sum_{k=1}^{K} e^{a_k}} \tag{6}$$

2.3.5 *Fully-connected layer*

Finally, a couple of fully connected layers with same dimension are responsible for the classification based on the feature maps from the last convolution.

2.3.6 *Network optimization*

For supervised learning, the network learns by optimizing (minimizing) a loss function on the difference between the predicted and true labels. The selected loss function is the cross-entropy between the estimated softmax output $q(x)$ and target class $p(x)$:

$$Loss = -\sum_{x} p(x)\log\left(q(x)\right) \tag{7}$$

The network is optimized via ADAM (Kingma and Ba, 2014), an algorithm inspired in Stochastic Gradient Descent (SGD). This optimizer works with adaptive estimates of lower-order moments and performs well with noisy and sparse gradients. This method combines two extensions of SGD: Adaptive Gradient Algorithm that improves performance on problems with sparse gradients by maintaining a learning rate per parameter. The other one is the use of Root Mean Square Propagation, which adapts the learning rate as a function of the magnitude of the gradients for the weights, improving the performance with noisy data.

2.3.7 *Regularization*

To tackle the CNNs tendency to overfit during training and to improve generalization performance, regularization is implemented by means of the following two approaches. First, we employ L2 regularization by adding a term to the loss function that penalizes high weights over the network (Peng *et al.*, 2015). Second, we use dropout, which consists of disconnecting some neurons during training to prevent co-adapting (Srivastava *et al.*, 2014), is implemented in the fully connected layer with 50% drop probability.

3 PROPOSED CNN METHOD

3.1 *Dataset*

The proposed CNN based method is conducted on a dataset generated from AE monitoring of an experimental rotor, as shown in Figure 1.

The setup is comprised of:

1. Mistras, Micro 30 Acoustic Emission sensors
2. Rotor and blades
3. DC Motor MY-1016, 24[V] 13.7[A]
4. MCP Q10-QS305 Power Source.

The rotor has 8 blades with one of them notched according to size and position in Table 1.

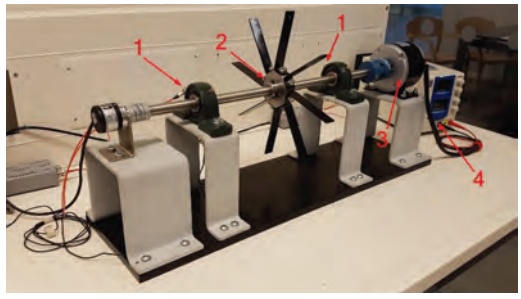


Figure 1. Experimental rotor setup.

Table 1. Size and position of notches.

Position [mm]	Size [mm]
5	3
20	6
	10

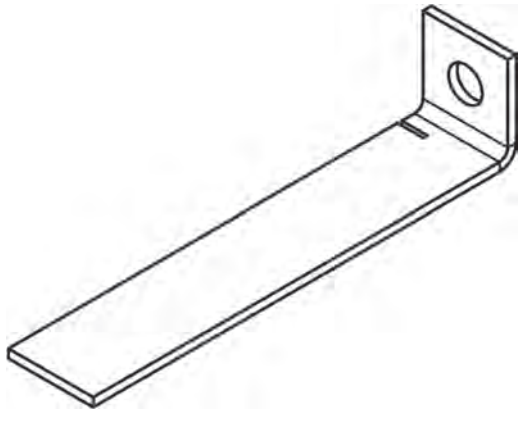


Figure 2. Sketch of a blade with a 6 mm notch at the 5 mm position.

A sketch of a blade is displayed in Figure 2.

The acquisition rate is 500 kHz and each combination between position and size of the notches are measured for 176.16 s divided in 168 files of 524,288 data points each.

In this dataset, there are seven health conditions: undamaged; 3 mm, 6 mm and 10 mm cracks at the 5 mm position; 3 mm, 6 mm and 10 mm cracks at the 20 mm position. For each health state, there are 53,477,376 data points and the CNN is fed with samples composed by slices of 49,152 points (1.77 rotor turn) of the raw signal with 50% overlap. Note that the proposed CNN method is trained for a 3-health state scenario corresponding to: (1) Undamaged, (2) Damage at 5 mm position obtained by combining the 3 mm, 6 mm and 10 mm cracks at that position; and (3) Damage

at 20 mm obtained by combining the crack sizes 3 mm, 6 mm and 10 mm at the 20 mm position.

Data augmentation is also implemented to enforce network generalization, thus improving the accuracy on unseen samples. The dataset has a total of 548,458,432 data points that are split into 80–20 proportion for training and testing, respectively.

Figure 3 shows examples of raw AE signals for each of the three health conditions. Notice that the signals have a significant amount of noise mainly because the rotor system is not perfectly balanced and has some degree of misalignment. In addition, vibrations from the bearings, coupling and motor add additional noise to the response.

However, the implementation of de-noising methods incurs in loss of information and encompasses pre-processing time that we want to avoid and handle with the proposed architecture.

Moreover, based on the raw signals and the amplitude spectrums shown in Figure 3, the health conditions are remarkably similar that, coupled with the signal noise levels, makes this dataset a significant challenging diagnosis task.

3.2 *Proposed deep CNN architecture*

The proposed CNN architecture, processing batches of 256 samples, consists of five convolutional layers as follows (see Figure 4): the first convolutional layer has 32 oversized filters of 128×1 designed to tackle background noise in the acoustic emission raw signal; this is followed by four convolutional layers with 32, 32, 64 and 128 filters of size 3×1, respectively, which are designed to automatically and hierarchically extract features from the AE data. The last convolutional layer's output is reshaped before being fed to the last two fully connected layers, each with 1024 neurons, that are responsible for processing the features obtained from the convolutional layers to perform fault diagnosis.

The proposed CNN method is trained for 15000 epochs, where one consists of all training samples. The CNN is regularized via dropout for the fully connected layers with 50% of keep probability and L2 regularization (see Section 2.3.7) as well as early stopping by saving the best epoch in terms of accuracy and generalization capability (train loss remaining low as test loss decreases).

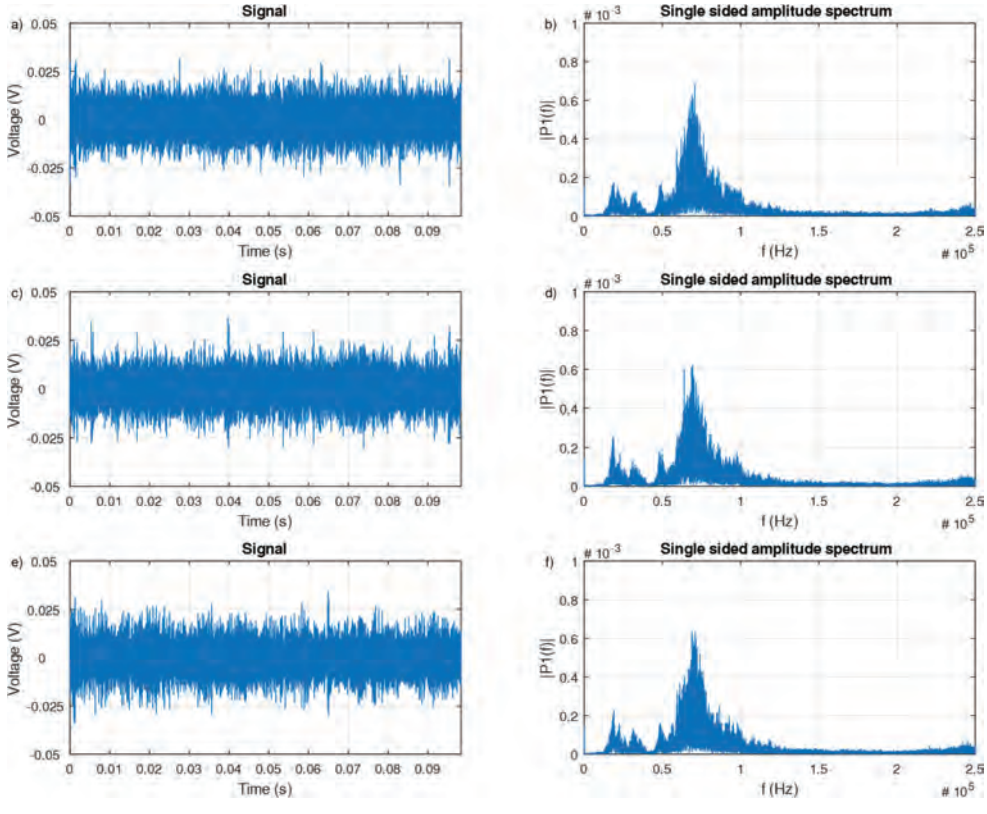


Figure 3. a) 5 mm sample signal, b) 5 mm amplitude spectrum, c) 20 mm sample signal, d) 20 mm amplitude spectrum, e) Undamaged sample signal and f) Undamaged amplitude spectrum.

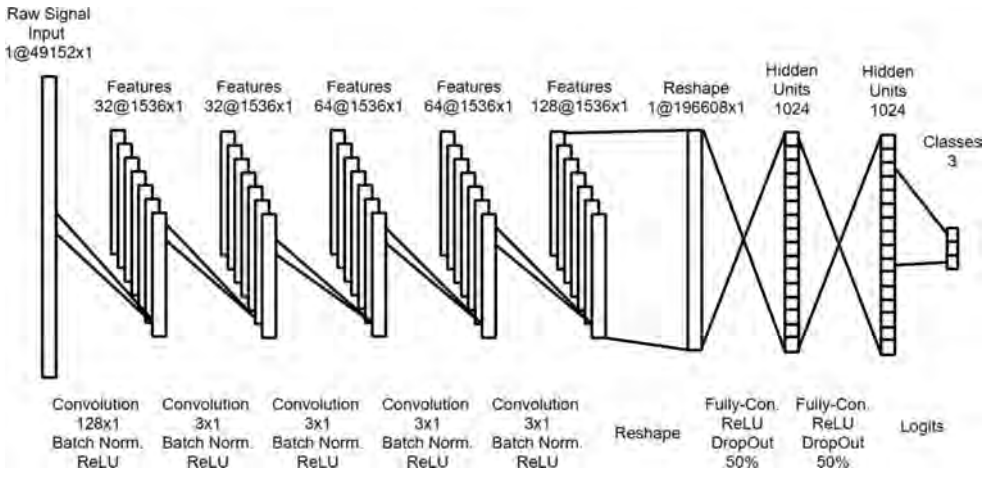


Figure 4. Architecture of the CNN.

Also of note is that the proposed CNN architecture does not have pooling layers. There are two underlying reasons: firstly, the proposed CNN method is not required to achieve spatial invariance as it deals with raw acoustic emission signals; secondly, the CNN marginally improved (in terms of accuracy and generalization) by the reduction in size of the feature maps resulting from the polling layers or, conversely, its performance deteriorated due to the loss of information incurred by the implementation of pooling layers.

In terms of activation functions, the fully connected layers use softmax, whereas ReLU is implemented in all five convolutional layers (see Section 2.3.4 for details). The CNN weights are initialized by Xavier Normal Initialization, a Normal Uniform distribution normalized by the size of the previous and next layer (Glorot and Bengio, 2010) and bias as 0.1 constant in all layers.

3.3 *CNN implementation*

All the results shown in the next section were obtained using the following hardware configuration at Smart Reliability and Maintenance Integration Laboratory (SRMILab) in the University of Chile: Intel® Core™ i7-6700 K CPU with 32 Gb RAM and a NVIDIA Titan XP GPU.

3.4 *Results and discussion*

The proposed CNN method is compared with a shallow ANN that has the same two fully connected layers, but lacks the convolutional layers. This allows us to assess the impact on the fault diagnosis performance of the convolutions as signal filtering and de-noising tool as wells as the quality and robustness of the extracted features. This ANN is fully optimized with ADAM adaptive gradient-based optimization algorithm and regularized via dropout (with 50% keep probability), weight regularization for both hidden layers and early stopping.

Table 2. Accuracy for health state of the system.

	Test accuracy (%)
ANN	33.7
CNN	93.0

Table 3. Performance measures in percentages [%] for the proposed CNN method.

	5 [mm]	20 [mm]	Undamaged
Sensitivity	89.77 ± 1.70	91.50 ± 1.48	97.65 ± 0.36
Specificity	94.69 ± 0.80	95.93 ± 0.62	99.04 ± 0.46
Precision	89.22 ± 1.62	91.53 ± 1.22	98.18 ± 0.92
F1 Score	89.48 ± 0.98	91.51 ± 0.88	97.91 ± 0.57
Accuracy	93.06 ± 0.80	94.50 ± 0.42	98.56 ± 0.39

Table 2 shows the overall test fault diagnosis accuracy. The proposed CNN method significantly outperforms the shallow ANN in terms of accuracy and generalization capacity, with the ANN barely learning from the complex AE dataset.

To corroborate these results, we collect the average values and corresponding standard deviations per health state from multiple runs for different performance metrics as shown in Table 3.

Table 4. Confusion matrix for the proposed CNN method.

	5 [mm]	20 [mm]	Undamaged
5 [mm]	363	35	6
20 [mm]	28	382	0
Undamaged	8	0	402

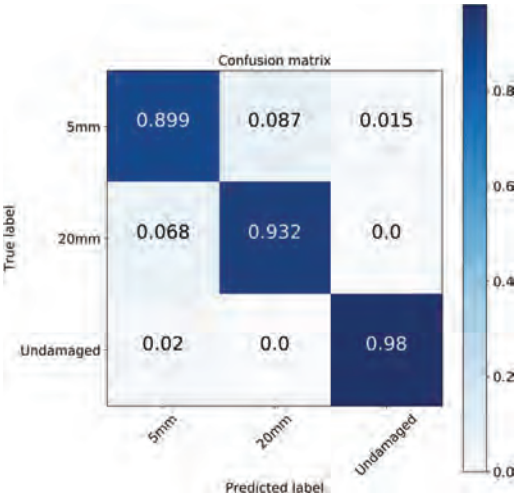


Figure 5. Normalized confusion matrix for the proposed CNN method.

Moreover, the unnormalized and normalized confusion matrices are shown in Table 4 and Figure 5, respectively.

Based on these results, the proposed CNN method outperforms the shallow ANN for the rotor's fault diagnosis based on acoustic emission monitoring. This is corroborated by observing Figure 6a) and c) that the CNN presents a monotonically descendent testing loss behavior that leads to improvement in the fault diagnosis accuracy. But, it should be observed that that the accuracy improvement to time ratio for the CNN is very low for the last epochs even though the network still learns. This could be driven by a very low learning rate for these epochs as ADAM adapts this hyperparameter.

However, as shown in Figure 6b) and d), the ANN barely learns from the raw AE data, which could be attributed to the complexity of the data as well as the meaningless features that the ANN extracts by treating the signals as independent points, problem that seems to be compensated by the convolutional filters in the CNN.

However, the superior performance achieved by the proposed CNN method comes at a much higher computational cost due to the significant number of learnable parameters and hyperparameters leads to extended training times.

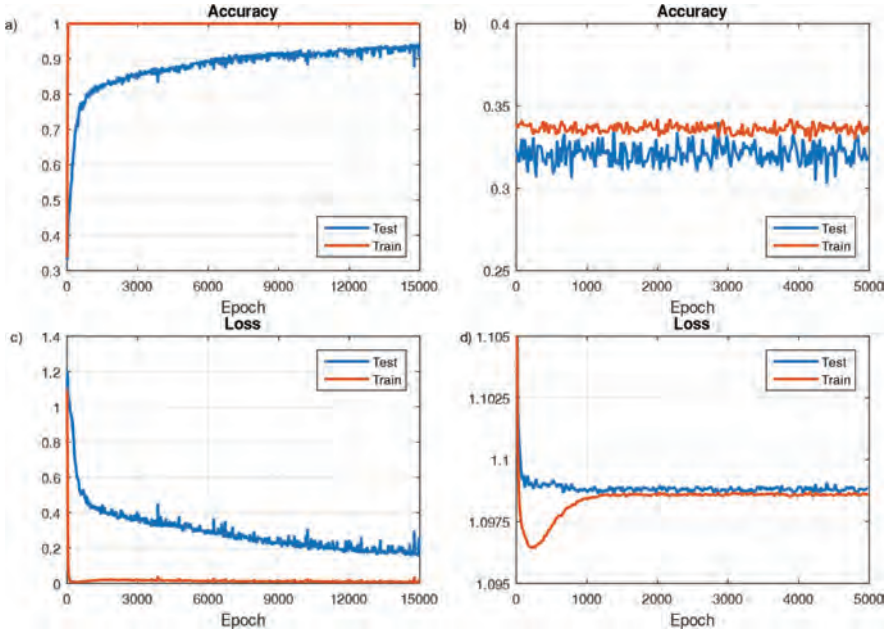


Figure 6. a) Accuracy behavior of the CNN, b) Accuracy behavior of the ANN, c) Loss behavior of CNN and d) Loss behavior of the ANN.

4 CONCLUSIONS

This paper has introduced a new deep CNN-based method for fault diagnosis using raw acoustic emission signals. The application of this method to an experimental rotor has shown that the proposed method delivers satisfactorily performance metrics for health state diagnosis. The CNN method was also compared to a fully optimized ANN, with the former significantly outperforming the shallow method.

These solid results in fault diagnosis are mainly due to the CNN's ability to automatically extract features from and efficiently handle the noisy acoustic emission signals. This also brings major advantages to the development of automated monitoring and fault diagnosis tools such as the possibility to bypass the intervention of the human element in the labor-intensive feature engineering process and reducing the need for preprocessing and de-noising of acoustic emission signals. Based on these preliminary results, the proposed CNN method is a promising tool for fault diagnosis.

ACKNOWLEDGMENTS

The authors acknowledge the partial financial support of the Chilean National Fund for Scientific and Technological Development (Fondecyt) under Grant No. 1160494.

REFERENCES

Deng, L. and Yu, D. (2014) 'Deep Learning: Methods and Applications', *Foundations and Trends® in Signal Processing*, 7(3–4), pp. 197–387. doi: 10.1561/2000000039.

Glorot, X. and Bengio, Y. (2010) 'Understanding the difficulty of training deep feedforward neural networks', 9, pp. 249–256.

Goodfellow, I., Bengio, Y. and Courville, A. (2017) *Deep Learning*. doi: 10.1007/s00287-016-1013-2.

Ioffe, S. and Szegedy, C. (2015) 'Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift'. doi: 10.1007/s13398-014-0173-7.2.

Kane, P. and Andhare, A. (2016) 'Application of psychoacoustics for gear fault diagnosis using artificial neural network', *Journal of Low Frequency Noise, Vibration and Active Control*, 35(3), pp. 207–220. doi: 10.1177/0263092316660915.

Kingma, D.P. and Ba, J. (2014) 'Adam: A Method for Stochastic Optimization', pp. 1–15. doi: http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503.

Li, C. *et al.* (2016) 'Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning', *Sensors (Switzerland)*, 16(6). doi: 10.3390/s16060895.

Maas, A.L., Hannun, A.Y. and Ng, A.Y. (2013) 'Rectifier Nonlinearities Improve Neural Network Acoustic Models', *Proceedings of the 30th International Conference on Machine Learning*, 28, p. 6. Available at: https://web.stanford.edu/~awni/papers/relu_hybrid_icml2013_final.pdf.

Peng, H. *et al.* (2015) 'A Comparative Study on Regularization Strategies for Embedding-based Neural Networks', (1). Available at: http://arxiv.org/abs/1508.03721.

Rabiei, E., Droguett, E.L. and Modarres, M. (2016) 'A prognostics approach based on the evolution of damage precursors using dynamic Bayesian networks', *Advances in Mechanical Engineering*, 8(9), p. 168781401666674. doi: 10.1177/1687814016666747.

Riaz, S. *et al.* (2017) 'Vibration Feature Extraction and Analysis for Fault Diagnosis of Rotating Machinery—A Literature Survey', *Asia Pacific Journal of Multidisciplinary Research*, 5(51), pp. 103–110. Available at: www.apjmr.com.

Ruiz-Gonzalez, R. *et al.* (2014) 'An SVM-Based classifier for estimating the state of various rotating components in Agro-Industrial machinery with a vibration signal acquired from a single point on the machine chassis', *Sensors (Switzerland)*, 14(11), pp. 20713–20735. doi: 10.3390/s141120713.

Srivastava, N. *et al.* (2014) 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting', *Journal of Machine Learning Research*, 15, pp. 1929–1958. doi: 10.1214/12-AOS1000.

Verstraete, D. *et al.* (2017) 'Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings', 2017, pp. 1–29.

Resilience engineering

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Best practices to improve public private people partnerships in the city resilience-building process

P. Marana, L. Labaka & J.M. Sarriegi
TECNUN, University of Navarra, Donostia-San Sebastian, Spain

ABSTRACT: Population is increasingly urban settled. Ensuring the welfare of society against upcoming crises derived from emerging challenges such as climate change, social dynamics and critical infrastructure dependencies within cities is seen as a priority for both scholar and practitioners. In this context, the concept of city resilience gains relevancy. Moreover, the task of ensuring the well-being of society increasing city resilience cannot completely rely on public entities; the contribution of private companies and citizens is also needed. There is a need to develop effective mechanisms such as Public Private People Partnerships (4Ps) to support the city resilience-building process. In order to develop effective and long lasting 4Ps that contribute to the city resilience building-process it is important to consider three dimensions; stakeholder relationship, information flow and conflict resolution. The aim of this paper is to present and describe a repository of best practices gathered from real city resilience-building processes that are currently taking place in different cities all over the world that contribute to the development of these three important 4P dimensions.

1 INTRODUCTION

According to the United Nations by 2030 more than 60% of people living in the world will be settled in urban areas (WHO 2017). Moreover, cities are currently at cross roads of challenges like climate change, social dynamics and the increasing dependence on the correct functioning of critical infrastructures that affect directly to the welfare of society (Gonzalez et al. 2017, Schauppenlehner-Kloyber & Penker 2016, Toubin et al. 2014, Elmqvist et al. 2013). Therefore, crises affecting cities derived from these challenges will potentially affect the welfare of citizens. This is why ensuring effective crisis management within cities will be increasingly important in the upcoming years in order to ensure the wellbeing of society (Toubin et al. 2014).

It is also important to bear in mind that the nature of the striking events that generate crises could be predictable or unpredictable. Moreover, predictable crises can also have unpredictable consequences due to potential cascading failures that may occur between complex interconnected systems (Pyrko et al. 2017). Therefore, a risk management approach that only consider predictable risks and consequences is not enough to deal with nowadays crises (Boin & McConell 2007). The resilience concept seems promising to address the need to deal with unexpected crises (FCOP 2011). Therefore, efforts are being made in promoting resilience in order to be able to face upcoming unpredictable crises that could potentially affect the welfare of society.

City resilience is an emerging concept that has been gaining popularity in the last few years. However, there is still a lack of consensus on its definition and has different approaches (Bång & Rankin 2016). Within this research, city resilience is defined as "the ability of a city or region to resist, absorb, adapt to and recover from acute shocks and chronic stresses to keep critical services functioning, and to monitor and learn from on-going processes through city and cross-regional collaboration, to increase adaptive abilities and strengthen preparedness by anticipating and appropriately responding to future challenges" (Hernantes et al. 2016).

Therefore, increasing city resilience will be a priority to ensure the welfare of society in the upcoming years (Toubin et al. 2014). In fact, ensuring the well-being of citizens in times of crisis is not a mission that can be delegated to public entities. In order to successfully fulfil this mission; the collaboration of additional city stakeholders, like private companies and citizens, is required (Kapucu 2012, Oxley 2013). While public entities should be the ones in charge of coordinating all the efforts being made to increase city resilience, it could also be beneficial to involve private companies and citizens (Gimenez et al. 2016). Private companies could contribute with technical expertise and additional resources in case the damage caused by an unpredictable event exceeds the capacities of public entities to deal with the crisis. Moreover, citizens could use their specific knowledge about the local community to better understand the needs of local

people regarding city resilience (O'Sullivan et al. 2015, Scolobig et al. 2015). Therefore, addressing unpredictable events affecting cities requires letting aside a silo-thinking mentality by involving different city stakeholders and coordinating their efforts to increase the city resilience level.

In light of this situation, developing meaningful public private people partnerships (4Ps) at the city level for strategic decision-making regarding city resilience could have a significant positive impact (Boyd & Juhola 2014, Ng et al. 2013). We define 4Ps as partnering arrangements, both formal and informal, that are developed between public entities, private companies and citizens with the aim of improving the city resilience-building process.

It is important to bear in mind that the objective of developing 4Ps is to foster meaningful collaboration among city stakeholders rather than to develop agreements per se. Rigid agreements have proven to be suitable for addressing expected crises, but trusting that rigid predetermined agreements can address unexpected events is not always effective (Stewart et al. 2009). Developing meaningful and long lasting 4Ps in which all the city stakeholders are represented enables to increase the adaptability and the improvisation capacity in times of crisis.

In order to develop effective 4Ps it is important to consider three dimensions; stakeholder relationship, information flow and conflict resolution (Marana et al. 2018a). The aim of this paper is to present and describe a repository of best practices (projects, strategies, activities, policies, methodologies and tools) gathered from real city resilience-building processes that are currently taking place in different cities all over the world that contribute to the development of 4P dimensions.

2 STATE OF THE ART

2.1 *Fragmented efforts between city stakeholders*

Emerging wide scope complex challenges like climate change, socio-political issues or critical infrastructure dependency, cannot be addressed by a single institution on its own.

The awareness regarding the importance of addressing the effects of these challenges is increasing among most city stakeholder groups (Gonzalez et al. 2017). For instance, the local government is developing and implementing climate change adapting plans; critical infrastructures are investing resources on understanding the existing interdependencies between them in order to prevent cascading failures that end up affecting several services, NGOs are focused on developing programs to reduce inequalities that affect vulnerable population and so on.

Each city stakeholder group can contribute to the city resilience-building process in different ways. Public companies (local, regional and national government, emergency services and so on) can contribute with their decision-making experience and capacity as well as with material resources. Private companies (Critical Infrastructure providers, businesses, insurance companies and so on) can contribute with technical and operational expertise as well as with additional material resources. People (citizens and NGOs) can contribute with knowledge of social, behavioral, economic and environmental issues (Scolobig et al. 2015).

All the approaches are equally valuable and require to create a holistic city resilience-building process. The key is not to consider them as isolated efforts but to integrate all of them in an effective way. Therefore, aligning the efforts of all the city stakeholders using mechanisms like 4Ps should be considered as a priority in order to improve city resilience.

2.2 *Dimensions of public private people partnerships*

Collaboration enables partners to share their knowledge, skills, resources and perspectives to use them in alternative and complementary ways (Gagnon et al. 2016, Jones & Barry 2011). Fostering 4Ps within the context of city resilience enables each stakeholder to contribute in the most appropriate way to the resilience-building process. In order to develop 4Ps in the most effective manner, the following dimensions should be considered (Marana 2018a) (Figure 1).

2.2.1 *Stakeholder relationship*

This dimension is related to the attributes and attitudes stakeholders must possess to work together successfully. We highlight the importance of promoting commitment of the city stakeholders to be

Figure 1. 4P dimensions.

active part of the city resilience-building process assuming that mutually beneficial goals could be achieved. Coordination and trust not only among city stakeholders but also with other systems and institutions with similar or complementary purposes are also considered within the scope of this dimension. It also embraces the need to increase adaptability of the partnership in the face of upcoming challenges. Finally, this dimension also considers the relevance of involving representatives of all the city stakeholder groups within the city (public institutions, private companies and citizens).

2.2.2 *Information flow*

This dimension is related to the communication channels and protocols that stakeholders must use to invest resources in the most effective manner. When we refer to this dimension we are highlighting the need to ensure the timeliness, accuracy and relevance of the shared information. Active participation of city stakeholders in planning, goal setting and execution of tasks are also considered within this dimension. This dimension also embraces how quickly information is available to relevant city stakeholder and the ease with which partners understand the information provided. Finally, it also considers to what extent critical and sensitive information is shared with other authorized partners.

2.2.3 *Conflict resolution*

This dimension is related to the techniques used to solve problems related to the correct functioning of the partnership. It highlights the importance of finding solutions to solve conflicts between city stakeholders in a constructive way in which the interests of all the involved partners are represented. This dimension also refers to the ability of the partnership to use lessons learnt in the past and to increase the effectiveness of future decisions. Finally, it also considers the importance of aligning the self-interests of each partner into a mutually beneficial goal.

3 METHODOLOGY

This research consisted of two different stages. The first stage consisted of an academic literature review and the second stage consisted on a revision of existing city resilience strategies.

3.1 *1st stage: Academic literature review*

A literature review was conducted in the Scopus database, in order to find best practices that contribute to the development of 4Ps in the city resilience-building process. This literature review enabled to find articles focused on the dimensions of multi-stakeholder collaboration in the context of city resilience.

The query used in the search in order to find relevant articles was the following: *"city resilience" OR "community resilience" OR "urban resilience" AND partnership OR collaboration.*

In order to ensure a more standard set quality only academic papers published in scientific journals were considered in this research. Although conference proceeding usually present interesting research projects, generally the main outcomes are published in scientific journals. Therefore, the type of publications was limited.

After conducting the search in the Scopus database a total amount of 96 articles were obtained. After reading the title and the abstracts a total amount of 52 research articles were analysed in full detail.

The aim of this academic literature review was to gather information about projects, strategies, activities, policies, methodologies and tools that have proven to be effective in the development of 4Ps in the city resilience-building process.

3.2 *2nd stage: Revision of city resilience strategies*

A revision of city resilience strategies was conducted in order to identify what cities are currently doing and planning to do in order to improve the effectiveness of 4Ps and consequently, increase city resilience.

City resilience strategies were obtained from the 100 City Resilience webpage (100 Resilient Cities, 2016c). 100 Resilient Cities is an initiative funded by the Rockefeller foundation whose aim is to help cities around the world to become more resilient to the physical, social and economic challenges that are a growing part of the 21st century.

The 36 city resilience strategies available in the 100 Resilient Cities website at the moment when this research was carried out were revised in order to find projects, strategies, activities, policies, methodologies and tools that contribute to improve each of the three dimensions of 4Ps in order to develop effective 4Ps in the city resilience-building process.

4 RESULTS

In the following section, the most important results obtained after reviewing scientific papers as well as city resilience strategies will be presented. Considering their final aim, the best practices gathered from the literature review have been classified into the three 4P dimensions.

4.1 *Stakeholder relationship*

Improving the interaction among stakeholders representing different city sectors is key for developing effective 4Ps that contribute to the city resilience-building process.

Representing the interests of all the city stakeholders when developing the basis of the city resilience-building process is key so that everyone accepts it and feels part of it.

The empowerment of citizens is key to foster engagement and a sense of belonging to the city. The city of Christchurch is developing alternative forms of public participation to promote awareness of issues and engage citizens in resilience related decision-making (100 Resilient Cities 2016d). The creation of community boards, advisory groups and working parties will enable that city stakeholders are more informed about the issues that community leaders have to make decisions.

The involvement of city stakeholders is not sufficient to ensure the effectiveness of resilience-building processes. There is a need to coordinate all the efforts being made by all the different groups. The city of Glasgow is conscious of this and has started to develop an integrated resilience plan for critical services in the face of long-term stresses in order to ensure the wellbeing of its citizens (100 Resilient Cities 2016b). The interdependencies among cross-sectoral critical services within Glasgow are identified with the objective to ensure that critical services remain functional and accessible regardless of the upcoming challenges.

Moreover, it is also important to realize that the resilience level of the city does not only rely on the city itself. Due to the increasing interconnection among cities throughout the world, learning from what others are doing is also key to address the challenge of developing resilience in cities. For instance, Bangkok is one of the cities that has addressed this challenge and is now interacting with different cities around the world with the support offered by the 100 Resilient Cities network (100 Resilient Cities 2017a). The city of Bangkok is currently working together with the city of Jakarta and Mexico City. The three of them are rapidly developing in mega cities where the challenge of efficient mobility exist. Therefore, they are working together to find solutions to this challenge.

4.2 *Information flow*

Improving the information flow among different stakeholder groups is key to improve the decision making process in the context of the city resilience-building process.

Developing communication channels to share information with city stakeholders is required to increase city resilience. The municipality of Dakar is aware of this fact and therefore is currently working on implementing tools and services to provide its city stakeholders with access to information on imminent crises in real time (100 Resilient Cities 2016a). These tools will enable that, a few months before a rainy season or time of high tides, the municipality could start communicating on disturbances observed to anticipate and reduce eventual damages caused by flooding. They realized that although public entities keep track of natural events preceding major disasters, such information is not actively used to the best advantage of the city. In light of this situation, they are working on creating a database of current data on the city's vulnerability state and they are establishing a network of community resilience champions.

This last initiative includes the integration of real-time information to the city's mobile application NAVIGEM in order to advise people when imminent risks are detected. However, communication with city stakeholders should be done not only using social media and apps. Usually, the most vulnerable groups in society like elderly people and children are not users of these communication channels. In order to reach to those groups alternative tools to proactive communicate upstream periods of vulnerability, like daily/weekly radio programs are being developed.

The timeliness, accuracy and accessibility of information is another key issue when talking about information in the context of city resilience.

Due to the technological updates conducted in cities in the last years, local governments are now increasingly considering to leverage the internet of things (IoT) to gather timely and accurate data that can enable to improve resilience related decision-making processes. Using the most current technology could help them to address disasters more efficiently and safely. However, this type of progress will require more than just employing the IoT to improve emergency preparedness and response; city stakeholders need to be ready to receive, interpret, and use the data in an effective manner. IoT sensors can be critical for urgent decisions like whether to evacuate an area at risk of earthquake, or how to guide residents to the safest exit routes ahead of an emergency. For instance, Santiago is working on applying sensors to improve the early warning systems as well as to use that information to design more effective evacuation protocols (100 Resilient Cities 2017b). In San Francisco, an early warning system called ShakeAlert has been started to be implemented in order to detect the first wave sent by an earthquake and to report is to citizens using this system (100 Resilient Cities 2016f).

Public entities must also know which communication channels work best to reach the affected city

stakeholders. For instance, if the at-risk population is predominantly Spanish-speaking, then the messages should be sent in Spanish. When dealing with an elderly population, the outreach can be done through television, newspapers, and radio rather than tech-driven channels like text alerts and apps. This targeted communication is a shift from the conventional "one size fits all" approach.

4.3 *Conflict resolution*

Improving conflict resolution to enable perspective alignment among stakeholders representing different interests will contribute to have a holistic view of city resilience.

Not only the involvement and contribution of all the city stakeholders is required to develop an effective city resilience-building process, an alignment on their perspective about resilience is also need. The city of New Orleans is aware of this need and is currently working on establishing a resilience center (100 Resilient Cities 2015). The aim of this center is to provide a space to build awareness and expertise of city stakeholders to develop projects and coalitions and to exchange ideas and practices both locally and globally. This space will enable to create synergies between different city stakeholders to work on initiatives that benefit all.

In order to increase the city's resilience level, the experience and lessons learnt by all the city stakeholders should be considered. The city of Melbourne is very aware of this need and is currently working on an initiative called (Monash University Disaster Resilience Initiative (MUDRI)), which explores how different stakeholders prepare to effectively response to crises and develops a Resilience Compendium that identifies leading practices that facilitate the sharing of best practices among different city stakeholders (100 Resilient Cities 2016e). This initiative enables to have a centralized resource for sharing and accessing information on resilience-building activities undertaken at the city. Consequently, this will prevent duplication of efforts and will promote a more efficient use of available resources.

In Rotterdam, the local government is also very aware of the need to empower all city stakeholder groups and has developed an action called the integration tours (100 Resilient Cities 2016g). In fact, talks and events aimed at encouraging cooperation and fostering dialogue between public entities, private companies and people are organized. These actions make citizens aware of their own roles in society and how they can better contribute to city resilience. Talks seek to break down the barriers created by self-interests to enhance effective dialogue among city stakeholders. These tours bring groups from different backgrounds and roles in society together to discuss different issues that are important for increasing the city's resilience level. Activities like this support knowledge sharing, strengthen mutual understanding, and enable the alignment of the different perspectives.

5 INTERCONNECTIONS BETWEEN 4P DIMENSIONS

This paper has presented the three different dimensions that should be considered when developing effective 4Ps that contribute to the city resilience-building process. However, it is important to bear in mind that all this dimensions are closely related among each other. The literature review has shown us that improving one 4P dimension has co-lateral effects on the other dimensions.

For instance, improving the relationship of different city stakeholders will have a potential impact on the amount and quality of the information they share between them (O'Sullivan et al. 2015). When a sense of belonging and trust among different entities exists, there is a bigger chance to improve the information flow among the partners. Improving the quality, accessibility and sharing of information also improves the coordination of city stakeholders and the sense of inclusiveness (Davenport et al. 2010).

Moreover, an improvement of the stakeholder relationship and of the information flow within the partnership has also an impact in the conflict resolution dimension (O'Sullivan et al. 2015). The better the relationship among city stakeholders and the more information is shared the easier to solve potential conflicts among city stakeholders will be.

Although all the best practices could have been classified within one 4P dimension, the effects of their implementation are usually transversal and affect not only to the improvement of their own dimension but also to the others. Therefore, further research should be conducted to better understand which the priority implementation order of these best practices should be in order to use the available resources in the most effective manner.

6 CONCLUSIONS

Although practitioners and academics are aware of the importance of improving collaboration between all the city stakeholders (public entities, private companies and citizens) to increase the resilience level of a city, there is not a concept to refer to this idea. This paper has presented the concept of public private people partnership (4P) as a new mechanism to foster collaboration through

formal or informal arrangements between city stakeholders in the city resilience building process.

This research has presented relevant best practices that are currently being implemented in certain cities all over the world. However, it is important to consider the limitations of the work presented. The best practices presented in this paper to illustrate how each dimension can be developed have been chosen in a pragmatic manner, without following a concrete methodology. The relevancy of these examples, will also decrease as available knowledge about city resilience building process increases and technology improves.

The efforts and resources dedicated to develop and implement an effective city resilience-building process are limited. Therefore, there is a need to set a priority order when implementing the resilience building best practices depending on the strengths and weaknesses of each particular city. Moreover, it is important to bear in mind that some cities could find easier to develop one dimension rather than other due to a cultural aspect. For instance, in some countries, like the ones in northern Europe, more attention is paid to the standardization of information sharing procedures. Therefore, these countries may find easier to improve the information flow dimension. However, other countries with a different cultural background, for instance the countries in the Mediterranean Sea, may find easier to establish informal relationships among stakeholders due to their feature of being more sociable. All these aspects should be addressed in future researches.

REFERENCES

100 Resilient Cities 2017a. Resilient Bangkok. p.46,60. Available at: http://www.100resilientcities.org/strategies/bangkok/ [Accessed 07 Dic. 2017].

100 Resilient Cities 2017b. Santiago Humano & Resiliente. p.131. Available at: http://www.100resilientcities.org/strategies/santiago-de-chile/ [Accessed 07 Dic. 2017].

100 Resilient Cities 2016a. Dakar Resilience Strategy, (December), p.65. Available at: http://www.100resilientcities.org/strategies/dakar/ [Accessed 07 Dic. 2017].

100 Resilient Cities 2016b. Our Resilient Glasgow A City Strategy. p. 41. Available at: http://www.100resilientcities.org/strategies/glasgow/ [Accessed 08 Dic. 2017].

100 Resilient Cities 2016c. Planning for Resilience. [online] Available at: http://www.100resilientcities.org/strategies/ [Accessed 20 Oct. 2017].

100 Resilient Cities 2016d. Resilient Greater Christchurch. p.61. Available at: http://www.100resilientcities.org/strategies/greater-christchurch/ [Accessed 07 Dic. 2017].

100 Resilient Cities 2016e. Resilient Melbourne. p.101. Available at: http://www.100resilientcities.org/strategies/melbourne/ [Accessed 06 Dic. 2017].

100 Resilient Cities 2016f. Resilient San Francisco. p.45 Available at: http://www.100resilientcities.org/strategies/san-francisco/ [Accessed 10 dic. 2017].

100 Resilient Cities 2016g. Rotterdam Resilience Strategy. p.60 Available at: http://www.100resilientcities.org/strategies/rotterdam/ [Accessed 08 Dic. 2017].

100 Resilient Cities 2015. Resilient New Orleans. p. 46. Available at: http://www.100resilientcities.org/strategies/new-orleans/ [Accessed 06 Dic. 2017].

Bång, Magnus and Rankin, Amy 2016. Survey report on worldwide approaches. Available at: http://smrproject.eu/fileadmin/user_upload/Documents/Resources/WP_1/D1.1.SMR_Final.pdf; [Accessed 06 Jan 2017].

Boin, A. & McConnell, A. 2007. Preparing for Critical Infrastructure Breakdowns: The Limits of Crisis Management and the Need for Resilience, *Journal of Contingencies and Crisis Management*, Vol. 15, No. 1, pp. 50–59.

Boyd, E., & Juhola, S. (2015). Adaptive climate change governance for urban resilience. *Urban studies*, *52*(7), 1234–1264.

Davenport, M.A., Bridges, C.A., Mangun, J.C., Carver, A.D., Williard, K.W., & Jones, E.O. 2010. Building local community commitment to wetlands restoration: A case study of the Cache River wetlands in southern Illinois, USA. *Environmental management*, 45(4), 711–722.

Elmqvist, T., Fragkias, M., Goodness, J., Güneralp, B., Marcotullio, P.J., McDonald, R.I., ... & Wilkinson, C. (Eds.). (2013). *Urbanization, biodiversity and ecosystem services: challenges and opportunities: a global assessment*. Springer.

Federal Office for Civil Protection (FOCP) 2011. Focal Report 7 : SKI Resilience and Risk Management in Critical Infrastructure Protection Policy: Exploring the Relationship and Comparing its Use, (December).

Gagnon, E., O'Sullivan, T., Lane, D.E., & Paré, N. (2016). Exploring Partnership Functioning within a Community-Based Participatory Intervention to Improve Disaster Resilience. *Journal of Higher Education Outreach and Engagement*, 20(2), 25–52.

Gimenez, R., Labaka, L., & Hernantes, J. 2017. A maturity model for the involvement of stakeholders in the city resilience building process. *Technological Forecasting and Social Change*, *121*, 7–16.

Gonzalez, J.J. et al. 2017. Stalking resilience Cities as vertebrae in society's resilience backbone. International Conference on Information Technology in Disaster Risk Reduction 2016 pp.31–45. Springer.

Hernantes, J. et al. 2016. Revised Resilience Maturity Model, Available at: http://smr-project.eu/about-the-smr-project. [Accessed 13 Dec 2017].

Jones, J., & Barry, M.M. 2011. Exploring the relationship between synergy and partnership functioning factors in health promotion partnerships. *Health Promotion International*, *26*(4), 408–420.

Kapucu, N. 2012. Disaster Resilience and Adaptive Capacity in Central Florida, US, and in Eastern Marmara Region, Turkey. *Journal of Comparative Policy Analysis: Research and Practice*, 14(3), pp.202–216. Available at: http://www.tandfonline.com/doi/abs/10.1080/13876988.2012.687620%5Cnhttp://www.tandfonline.com/doi/abs/10.1080/13876988.2012.687620#preview.

Marana, P., Labaka, L. & Sarriegi, J.M. 2018a. A framework for public-private-people partnership in the city resilience-building process. *Safety Science*. In press.

Ng, S.T., Wong, J.M., & Wong, K.K. 2013. A public private people partnerships (P4) process framework for infrastructure development in Hong Kong. *Cities*, *31*, 370–381.

O'Sullivan, T.L., Corneil, W., Kuziemsky, C.E., & Toal-Sullivan, D. 2015. Use of the structured interview matrix to enhance community resilience through collaboration and inclusive engagement. *Systems Research and Behavioral Science*, 32(6), 616–628.

Oxley, M.C. 2013. A "People-centred Principles-based" post-Hyogo framework to strengthen the resilience of nations and communities. *International Journal of Disaster Risk Reduction*, 4, pp.1–9.

Pyrko, I., Howick, S., & Eden, C. 2017. Risk systemicity and city resilience. Paper presented at *EURAM 2017*, paper no. 1446.

Schauppenlehner-Kloyber, E. & Penker, M. 2016. Between participation and collective action-from occasional liaisons towards long-term co-management for urban resilience. *Sustainability (Switzerland)*, 8(7).

Scolobig, A., Prior, T., Schröter, D., Jörin, J., & Patt, A. (2015). Towards people-centred approaches for effective disaster risk management: Balancing rhetoric with reality. *International journal of disaster risk reduction*, 12, 202–212.

Stewart, G.T., Kolluru, R. & Smith, M. 2009. Leveraging public-private partnerships to improve community resilience in times of disaster. *International Journal of Physical Distribution & Logistics Management*, 39(5), 343–364.

Toubin, M., Laganier, R., Diab, Y., & Serre, D. 2014. Improving the conditions for urban resilience through collaborative learning of Parisian urban services. *Journal of Urban Planning and Development*, 141(4), 05014021.

World Health Organisation (WHO) 2017. World Health Organisation: Global Health Observatory (GHO) – Urban population growth. Available at: http://www.who.int/gho/urbanhealth/situationtrends/urbanpopulation growth/en/ [Accessed 27 Sep 2017].

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Checklist for judgement of technical facility safety level and results obtained by its application in practice

D. Procházková & J. Prochazka
Faculty of Transportation Sciences, Czech Technical University in Prague, Czech Republic

ABSTRACT: Technical facilities safety is fundamental issue for human society and its development. It is reality that it is broken by many known and newly cognized risks that are related to forever grooving complexity of technical facilities and whole world. Today, we know that it is also necessary to consider the risks that are connected with interfaces among their subsystems and components. With regard to the world dynamic development, it is necessary to monitor the priority risks and to cope with them during the time. The measure of safety level is performed by help of logically arrangement of requirements of individual techniques used at work with risks in technical facilities, fragmented to 7 domains and Maximum Utility Theory principle. The paper also shows the results on safety levels for 5 technical facilities.

1 INTRODUCTION

On the basis of present level of knowledge that is e.g. represented by publications from the ESREL conferences (Ale et al. 2010, Bérenguer et al. 2011, Briš et al. 2009, Cepin & Bris 2017, Nowakowski et al. 2014, Podofillini et al. 2015, Steenbergen et al. 2013, Walls et al. 2016), which is summarized in (Procházková 2015, 2017), we perceive each technical facility as open complex system of systems, i.e. as several open systems that are mutually penetrated and are interfaced with vicinity.

The interfaces ensure the fulfilment of important operations and services, and simultaneously they cause the dependences that are the roots of specific vulnerabilities. Under specific conditions they originate highly unfavourable interfaces that lead to technical facility failure, which at certain circumstances distinctly also damage the technical facility vicinity. Therefore, at ensuring the technical facility safety it is necessary to consider that technical facility has various assets that are altered in dynamically variable world. The multiplicity and variability of assets cause that under certain conditions, the measures ensuring the safety of individual assets are conflicting, which means that methods using at risk management aimed to technical facility safety need to be multi criterial (Procházková 2017).

2 SAFETY AND RISKS OF TECHNICAL FACILITIES AND PRODUCTS

At present in advanced engineering disciplines described in works (Ale et al. 2010, Bérenguer et al. 2011, Briš et al. 2009, Cepin & Bris 2017, Novakowski et al. 2014, Podofillini et al. 2015, Steenbergen et al. 2013, Walls et al. 2016, Procházková 2015, 2017), the safety is understood as the attribute that emerges on the system level. The safety shows the quality of set of human measures and activities, which ensure that system is safe. Among the important quantities the following relations are valid: *dependable (reliable) system* is a system that performs required functions in a given place, a given time and a given quality during the whole life cycle; *secure system* is the dependable system that is protected against to internal and external disasters of all kinds; *safe system* is the secure system that does not endanger itself and its vicinity under all conditions; and *risk* je is understood as the probable size of losses, damages and harms on protected assets in real system that is calculated for unit of space and unit of time. It is dependent on the disaster size and on the local assets vulnerabilities. Safety and risk are in certain relation, but they are not complementary quantities. The risk reduction means the safety increase, but it is not always valid inversely (Procházková 2015). The complementary quantity to safety is the criticality; in some legislation, e.g. in the SEVESO directive, it is used the term recklessness instead of criticality. *Criticality* denotes the limit (boundary) from which the risk impacts are significant up to eliminative for followed system, which means that appurtenant risk needs to be always mastered.

3 DATA USED AT CHECKLIST FORMATION

On the basis of data given in publications (Ale et al. 2010, Bérenguer et al. 2011, Briš et al. 2009, Cepin

Figure 1. Items that influence the result of work with risks of technical product.

& Bris 2017, Novakowski et al. 2014, Podofillini et al. 2015, Steenbergen et al. 2013, Walls et al. 2016, Procházková 2015, 2017), and in Archives (ČVUT 2017), it is necessary to consider seven items (Figure 1) that influence the result of work with risks of technical facility, i.e. its safety, namely:

1. Context in which the risks, inherently connected with technical facility, are inserted.
2. List of considered sources of risks.
3. Type of risk form.
4. Ways of mastering the risks.
5. Process model of work with risks, application of the TQM and Coase theorem.
6. Technique of management and coping with risks of technical facility.
7. Way of management of risks in time.

Ad 1. It holds that the most general context to which the risks of technical facility are inserted has the assets: human life, health and security; property and public welfare; environment; and technologies and infrastructures. The process model ensuring the human security and development is in (Procházková 2015). On the basis of results in (Procházková 2015, 2017) and data obtained directly in practice (ČVUT 2017), in the technology sector it is often considered only the context of technical facility or context of enterprise that administrates the technical facility, and in many cases only the context of production facility or its part. It is understandable that the use of more limited context means the higher default of reality.

In practice, it means that the appurtenant solution does not consider some sources of risks and the impacts of risks´ realization on all public and enterprise assets. They are neglected: harmful phenomena from technical facility vicinity and phenomena induced by bad decisions of management of enterprise or administrative bodies; and the impacts of risks on humans, properties and environment in technical facility vicinity.

Ad 2. With regard to results in (Procházková 2011a, 2015, 2017) and data obtained directly in practice (ČVUT 2017), it holds that in practice there are used the following choices of sources of risks:

1. Sources of risks determined either by legislative, or by experiences of worker who solves the task.
2. Only technical sources of risks in a given technical facility. Usually, it goes on risks connected with: material (fulfilment of required parameters, supplier relations—alternative material etc.); construction and interfaces of components and facilities (free procedures, presence of unstable hazardous substances….); production procedures, e.g. at welding, specific works with millers, lathes etc.; and conditions that are necessary pro production of quality product, e.g. certain pressure, certain temperature or certain humidity of surrounding medium etc.
3. Technical sources of risks and human factor. To items given in point 2, they are added the risks connected with false operations of workers.
4. Technical sources of risks and human factor in the broadest interpretation. To items given in point 3, they are added risks connected with sources of organizational accidents (i.e. bad decision-making, using the false procedures etc.).
5. Technical sources of risks, sources of risks threatened the workers lives, health and safety, sources of organizational accidents and sources of risks in working environment.
6. The sources of risks given in point 5 plus external sources of risks.
7. The sources of risks given in point 6 plus sources of risks from interfaces of facilities, components and system that disturb the technical integrity, the originators of which are in automatization, education and good skill.
8. All Hazard Approach in the form described in (Procházková 2011a). This selection considers risks from the five basic domains (ca 77 sources).

The last set of risk sources is complete, but it is challenging on data, methods, knowledge, experience and time period. It requires the strategic, systemic and proactive approach and it has according the results of FOCUS project (Procházková 2015) a lot of deficits at use in present practice.

Ad 3. On the basis of results in (Procházková 2015, 2017) and data obtained directly in practice (ČVUT 2017), it holds that in technical practice, there are used the partial, integrated and integral

(systemic) risks. Partial risk is the risk connected with one asset. The partial risks are various, e.g. health risks, technological risks, risk of fire etc. For their determination, many legal rules and supporting software exist (Procházková 2011a). Integrated risk represents the sum or other aggregation of partial risks. It is used e.g. in protection of workers lives, health and safety (Procházková 2011a). Integral (systemic) risk is based on system concept of entity and it also includes the interfaces among the assets and components of technical facility (Procházková 2015, 2017). It is given by relation

$$R(H) = \left[\sum_{i=1}^{b} A_i(H) Z_i(H) + \sum_{i=1}^{n} \int_0^T \int_s F(H, A_i, Pi, 0, t) dS dt \right] \cdot \tau^{-1}$$

in which $\boldsymbol{H}$ is the hazard connected with given disaster in site of technical work; $\boldsymbol{A_i}$ are values of followed assets for i = 1,2,…, n; $\boldsymbol{Z_i}$ are vulnerabilities of assets for i = 1,2,…, n; $\boldsymbol{F}$ is the loos function; $\boldsymbol{P_i}$ are the occurrence probabilities of damage of assets for i = 1,2,…, it goes on conditioned probabilities; $\boldsymbol{O}$ is vulnerability of protective measures; $\boldsymbol{S}$ is the size of followed space; $\boldsymbol{t}$ is time measured from the disaster origin; $\boldsymbol{T}$ is the time period of losses origin; and $\boldsymbol{\tau}$ is the disaster return period.

It is evident that for long-term ensuring the safe technical facility, it is necessary to consider the integral risk. Because in above given formula, the loss function is not known, so in (Procházková 2015, 2017) there are given procedures used in practice for estimation of integral risk; they are based on the analysis of real and simulated disasters´ scenarios and expert judgement.

It is necessary to note that determination of individual types of risks also differ in exactingness on data and methods of their processing (Procházková 2011a, 2017); the lowest challenging is the determination of partial risks, and therefore, these are mostly used in practice, although their validity with regard to total technical facility safety is very limited.

Ad 4. On the basis of results of investigations given in (Procházková 2015, 2017) and data from practice (ČVUT 2017), three cases are found. In the first one, there are used the risks, which are determined and mastered only after the technical facility construction (Prochazkova 2015); this way is danger because some of important risks, which could be only mastered by specific technical measures in assignment of technical facility, can be only reduced by organizational measures that are lower effective than technical measures (ČVUT 2017, Procházková 2015, 2017). In the second one, the risks are considered from beginning the technical facility design up to its termination from operation. This way depends on requirements of legislation, knowledge and skill of designers, constructors and operators, which does not guarantee the consideration of all sources of risks. In the third one, risks are considered from beginning the technical facility design and at trade-off with them, it is used the Defence-In-Depth approach, which requires system thinking, multi sectoral and transdisciplinary knowledge and experiences (ČVUT 2017, Procházková 2015).

The ensuring the safety of technical facility and its vicinity depends on quality of work with risks and on accessible possibilities of both, the technical facility management and personnel, and the public administration (ČVUT 2017).

Ad 5. The risk mastering in given time and given site requires: knowledge; capabilities; competences, finance; material, technical and human sources. Therefore, in next we deal not only with alone work with risks, but also with practical procedures that are used at decision-making on the risk mastering. On the basis of results of investigations in (Procházková 2015) and data obtained directly in practice (ČVUT 2017), it holds that in technical practice it needs to use the process model shown in Figure 2.

It is evident that if we are not able to identify and analyse some risk, so we are not capable effectively to defend the followed entity against it. The error, which we do at risk analysis, is transferred to emergency, continuity and crises plans, and it reduces their value in relation to planned measures directed to protection of human lives and health, and also to operational capability of rescue units participating in performance of rescue operations.

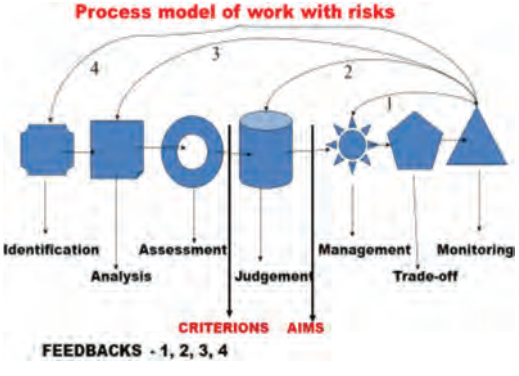


Figure 2. Process model of work with risks. Criterions = conditions that determined when the risk is acceptable, conditionally acceptable or unacceptable. Aims denote required states. Numbers 1,2,3,4 denote feedbacks that are used if the monitoring shows that followed requirements on safety are not fulfilled.

The aim of risk management is to find the optimum way, how to reduce the founded risks to required socially acceptable level, possible to keep up on this level. The risk engineering aim is then to find the way, how under available options, the risk management proposed measures and activities for risks mastering to realize and to ensure their reliability and function. The risk reduction is almost always connected with increase of expenses and claims on knowledge. The risk management is led by effort to find the boundary to which it is endurable the risk reduction, so the spend expenses would be socially acceptable.

In harmony with the public interest it is necessary so the risk acceptability might have the social dimension. Therefore, it is necessary to consider:

1. For whom the risk could be acceptable; for risk originators, for politicians or for public?
2. Who determines the risk acceptability; politicians adjudicate on that, which is legal, and so they could not adjudicate on that, which is acceptable.
3. If at risk determination there were discussed actually permitted risks, intolerant threshold values and attitudes of public to risks.

Risks are inherent factors of human system, i.e. they were, are and will be, and besides they will occur new ones. Therefore, the management of risks requires risk dimension and measurement of risk, which consider not only the physical damages, harms, victims and economic losses bulk, but also the social, organizational and institutional factors.

The outputs from risk management process for needs of good governance according to TQM are: *Risk assessment document* – records on all appurtenant risks; *Top risks list* – list of selected risks, the mastering with them has the highest claims on sources and time; *Retired risk list* – serving as the historical reference for future decision-making.

Technique of alone risk management from the point of provident handle with forces, sources and means, formally reviews before at each phase of work with risks the results of management and mastering the risks in the context of profits and expenses on outputs. The Coase theorem (Coase 1960) is used for determination of economic optimum in expenses on mastering the risks.

Ad 6. On the basis of results in (Procházková 2015, 2017) and data obtained directly in practice (ČVUT 2017), it holds that in technical practice it needs to understand that technical facility risk management and risk mastering are not the task of individual, not one organisation or one sector. It goes on the collective effort of all participants. It is evident that: professionals who have knowledge, data and capability to apply suitable methods can only determine the risk; and only persons who have appropriate competences can decide on handling with risk, i.e. legally determined representative of public administration or technical facility; and risk mitigating and control could be performed only by professionals who have appropriate knowledge, capabilities, skill, equipment, sources and means. The public is lawful participant at risk mastering because it goes on its security and quality of life. Because there are many risk sources, and countermeasures for their mastering are very often conflicting, it is necessary to use the risk management aimed to safety (Zairi 1991).

The negotiation with risks goes from present possibilities of human society and it lies in splitting the measures and activities for risk mastering into: prevention, mitigation, response and renovation.

So the executive body of organisation could effectively work with risks, it is necessary to determine the procedure for risk determination by legal rule, and simultaneously to determine the value scales by which the outputs of tools for determination of risks in organization are interpreted; i.e. it is necessary to determine which risk value is acceptable, which one is conditionally acceptable and which one is unacceptable. In tools for risk determination, it is necessary to distinguish the sophisticated tools for professional sphere and tools for administrative bodies for which the checklists are the most suitable.

Ad 7. On the basis of results in (Procházková 2015, 2017) and data obtained directly in practice (ČVUT 2017), it holds that in technical practice it needs to understand that from system viewpoint the ensuing the technical facility safety is the requirement on the complex system, not on its components.

Risks are inherent attribute of human system and each technical facility, and therefore, they need to be managed during the whole technical facility life time. The aim of risk management is to ensure the safe technical facility, i.e. also its competitiveness today and in future, i.e. it goes on determination the priority risks and their correct management. The risk management needs to ensure the technical facility safety at conditions normal, abnormal and critical.

On the basis of present knowledge given in (Procházková 2011a, 2015, 2017), the Safety Management System (SMS) of complex object is built on principles of process management and it includes the organization structure, responsibilities, practices, regulations, procedures and sources for determination and assertion of prevention of disasters or at least the mitigating their unacceptable impacts. Usually, it deals with many questions, apart from also the organisation, workers, identification and assessment of hazards and risk that follow from them, organization management, change

management in organization, emergency and crisis planning, safety monitoring, audits and review.

The process safety management is concentrated to six processes: concept and management; administrative procedures; technical matters; external co-operation; emergency preparedness; and documentation and investigation of accidents. These processed are further divided into sub processes that are in detail described in (Procházková 2015, 2017).

The processes coordination is aimed to ensuring the safe facility at conditions normal, abnormal and critical. The coordination in this context is understood as the controlled process, the aim of which is to create and to operate the technical facility in required quality; it follows the processes in spheres as: space and time, personnel, material, finance and documentation (Procházková 2015, 2017).

For support of safety management system, it is necessary to process the series of remedial tools as: security plans; on-site and off-site emergency plans; continuity plans; crisis plans; in practise the risk management plans for priority risks have been very came in useful (Procházková 2017). The most important is safety culture as it is stressed in fundamental work (Kongsvik, Almklov, & Fenstad 2010).

4 METHOD FOR CHECKLIST MAKE-UP AND METHOD OF ITS USE IN PRACTICE

Data and experiences from work with risks are given in (Procházková 2011a, 2015, 2017) and in works that are cited in them. They clearly show that the facility risks are pulled off, the better level of facility safety is reached. From this reason, we used at construction of tool for judgement of technical facility safety: the final judgement of used risk engineering methods quality (US EPA 2008), i.e. we compile the checklist; and the maximum utility theory principles (Keeny & Raiffa 1993). The checklist was proposed by procedure described in (Procházková 2011b) so the question answer "YES" in each aspect given in Chapter 3, belongs to the best way of aspect solution on the basis of present knowledge and experiences. The scale for judgement of total result is selected in agreement with recommendation in (Procházková 2017).

For real judgement of safety of technical facilities, we used the safety audit method (Procházková 2011a, b). At safety audit, the answer to each question is always separately formulated by 5 evaluators (technical director, security expert of technical facility, security expert of local public administration, security expert of regional public administration, authors) according to documentation of technical facility. The final evaluation of each question is made as median from partial evaluations. In case of significant doubts at certain real question judgement, the note was given in special column of check list; and the final results in these cases are finally obtained by panel discussion of experts.

5 CHECKLIST

Specific checklist compiled by procedure described in foregoing chapter is in Table 1. It contains 72 questions. The scale for its final evaluation (i.e. the determination of safety rate) is in Table 2.

Table 1. Check list for judgement of technical work safety according to judgement of work with risks. Answers: Y – YES, N – NO; R – Remark.

Question	Answer Y	N	R
Are in technical facility documentation distinguished the terms danger, hazard and risk?			
Is technical facility documentation based on context that considers only the technical work assets?			
Is technical facility documentation based on context that considers technical work assets and selected public assets (employee, contractors, and visitors, humans in work vicinity, working setting and environment)?			
Is technical work documentation based on context that considers technical facility assets and all public assets?			
Are only considered risk sources that are determined by expert experience?			
Are only considered only risk sources that are determined by legislative and expert experience?			
Are only considered risk sources that are connected with technical facility alone?			
Are considered risk sources that are connected with technical facility alone and human factor connected with badly performed working operation?			

(*Continued*)

Table 1. (*Continued*).

Question	Answer Y	N	R
Are considered risk sources that are connected with technical facility alone and human factor in the broadest concept?			
Are considered risk sources that are connected with technical facility alone, human factor in the broadest concept, workers health jeopardy and threatening the working environment?			
Are considered risk sources that are connected with technical facility alone, human factor in the broadest concept, workers health jeopardy, threatening the working environment and environment outside the technical facility?			
Are considered risk sources that are connected with technical facility alone, human factor in the broadest concept, workers health jeopardy and threatening the working environment in system context, i.e. also risk sources connected with linkages and flows in technical facility?			
Are considered risk sources according to All-Hazard-Approach?			
Are only considered partial risks?			
Are considered partial risk and integrated risk?			
Are considered partial risks, integrated risk and integral risk?			
Are risks in technical facility systematically followed?			
Are risks in technical facility systematically followed only after technical facility building?			
Are risks in technical facility systematically followed for its whole life cycle, i.e. from its design?			
Are risks in technical facility systematically followed for its whole life cycle, i.e. from its design and in its design and operation used the Defence-In-Depth approach?			
Is at work with risks in technical facility systematically used the process model of work with risks?			
Is at work with risks in technical facility systematically used the process model of work with risks that possesses clearly determined criterions for risks acceptance?			
Is at work with risks in technical facility systematically used the process model of work with risks that possesses clearly determined criterions for risks acceptance, which respect public interest (i.e. they have social dimension)?			
Is at work with risks in technical facility systematically used the process model of work with risks that possesses clearly determined criterions for risks acceptance and aims of risk management?			
Is at work with risks in technical facility systematically used the process model of work with risks that possesses clearly determined criterions for risks acceptance with regard to public interest?			
Is at work with risks in technical facility systematically used the process model of work with risks that possesses clearly determined criterions for risks acceptance with regard to public interest and corrected measures in monitoring for the case that risk will happen unacceptable?			
Is at work with risks in technical facility systematically determined and followed the set of priority risks?			
Does technical facility risk management technique ensure in each phase of work with risks the review of profits and costs connected with measures for risks mastering, so economical handling with forces, sources and means might be ensured in technical work?			
Does technical facility risk management technique ensure in each phase of work with risks the review of profits and costs connected with measures for risks mastering, so economical handling with forces, sources and means might be ensured in technical work and in public administration?			
Are in technical facility systematically performed the preventive measures for reduction or avert of some risks?			
Are in technical facility systematically performed the preventive measures for reduction or avert of all priority risks?			
Are in technical facility systematically performed the preventive measures for reduction or avert of all risks that have potential to cause important losses to technical facility?			
Are in technical facility systematically performed the preventive measures for reduction or avert of all risks that have potential to cause important losses to technical facility and unacceptable impacts on surrounding environment?			
Are in technical facility systematically performed preventive measures for reduction or avert of all risks and prepared the mitigating measures for reduction of some highest risk impacts?			
Are in technical facility systematically performed preventive measures for reduction or avert of all risks and prepared the mitigating measures for reduction of all priority risks impacts?			
Are in technical facility systematically performed preventive measures for reduction or avert of all risks and prepared the mitigating measures for reduction of all risks impacts that can cause the significant losses to technical facility?			

(*Continued*)

Table 1. (*Continued*).

Question	Answer Y	N	R
Are in technical facility systematically performed preventive measures for reduction or avert of all risks and prepared the mitigating measures for reduction of all risks impacts that can cause the significant losses to technical facility and unacceptable consequences for surrounding environment?			
Is technical facility insured against risks?			
Does technical facility possess the finance, material, technical, personal and organisational measures for response to important risk?			
Does technical facility possess the finance, material, technical, personal and organisational response for renovation after important risk realisation?			
Does technical facility possess the finance, material, technical, personal and organisational measures also for response and renovation after extreme unexpected realisation?			
Are at work with risks in technical facility only considered the results of preliminary risk analyses?			
Are at work with risks in technical facility preferred the results of standard, fast and low precise risk analyses before results of preliminary risk analyses?			
Are at work with risks in technical facility preferred the results of detailed risk analyses in synoptic concept before the results of preliminary risk analyses and standard, fast and low precise risk analyses?			
Are at work with risks in technical facility preferred the results of individual and specific risk analyses before the results of detailed risk analyses in synoptic concept, preliminary risk analyses and standard, fast and low precise risk analyses?			
Are at work with risks in technical facility determined the criterions for assessment?			
Are at work with risks in technical facility determined the criterions for assessment technical and economical items?			
Are at work with risks in technical facility determined the criterions for assessment technical and economical, external and internal items?			
Are at work with risks in technical facility determined the criterions for assessment technical and economical, external and internal and socially political items?			
Are at work with risks in technical facility determined the requirements for ensuring the safety?			
Are at work with risks in technical facility determined the requirements, standards and norms for ensuring the safety?			
Are at work with risks in technical facility determined the requirements, standards and norms for ensuring the safety and partial aims?			
Are at work with risks in technical facility determined the requirements, standards and norms for ensuring the safety, partial aims and methods and procedures?			
Are at work with risks in technical facility determined the requirements, standards and norms for ensuring the safety, partial aims, methods and procedures, and also limits and conditions?			
Are at work with risks in technical facility determined the requirements, standards and norms for ensuring the safety, partial aims, methods and procedures, limits and conditions and also the authorizations of persons or institutions?			
Does the technical facility administrator hold the safety management system that is compiled on the principles of process management and systemic work with risks?			
Does the technical facility administrator hold the safety management system (SMS) that contain the organizational structure, responsibilities, practices, rules, procedures and sources for determination and enforce of disaster prevention or at least for mitigating the unacceptable disasters impacts in technical facility and in its surrounding?			
Does the technical facility administrator hold the safety management system (SMS) that contain management of six processes: concept and management; administrative procedures; technical matters; off-site co-operation; emergency preparedness; and documentation and accident investigation?			
Does the technical facility administrator hold the SMS that contains the concept and management process with sub-processes for: overall concept; reaching the safety partial aims; safety governance; alone safety management system; personnel—human sources management, education and training, internal communication, working environment; audit and assessment of performance of safety aims?			
Does the SMS technical facility contain the administrative procedures process with sub-processes for: hazard identification from possible disasters and corresponding risk assessment; documentation of procedures (including the work permits); changes management; safety connecter with contractors; surveillance under products safety?			

(*Continued*)

Table 1. (*Continued*).

Question	Answer Y	N	R
Does the SMS technical facility contain the technical matters process with sub-processes for: research and development; design and montage; inherently safer processes; technical standards; storage of hazardous substances; and integrity maintenance and maintenance of equipment and buildings?			
Does the SMS technical facility contain the off-site co-operation process with sub-processes for: co-operation with public administration; co-operation with public and other involved (including the academic institutions); and co-operation with other enterprises?			
Does the SMS technical facility contain the emergency preparedness process with sub-processes for: on-site planning; facilitation of off-site planning (for which the public administration is responsible); and co-ordination of activities of resort organisations at ensuring the emergency preparing and the response?			
Does the SMS technical facility contain the documentation and accident investigation process with sub-processes for: processing the reports on disasters, accidents, near misses and other instructive experiences; investigation of damages, losses and harms and their causes; and response and consequential activities after disasters (including the application of lessons and information sharing)?			
Does the SMS technical facility contain the program for safety improvement in which there are given: roles of stakeholders; rules for safety culture improvement (golden rules); and relevant responsibilities?			
Does the SMS technical facility contain the program for safety improvement in which there is given: security plans (on strategic, tactical, functional and technical levels); on-site and off-site emergency plans; continuity plans; and crisis plans?			
Does the SMS technical facility contain the program for safety improvement in which there is given the risk management plan with clearly determined countermeasures and responsibilities?			
Does the SMS technical facility contain the program for safety improvement in which there is given the risk management plan with clearly determined countermeasures and responsibilities that only contains the technical risks?			
Does the SMS technical facility contain the program for safety improvement in which there is given the risk management plan with clearly determined countermeasures and responsibilities that only contains the technical and organisational risks?			
Does the SMS technical facility contain the program for safety improvement in which there is given the risk management plan with clearly determined countermeasures and responsibilities that contains the technical, organisational and external risks?			
Does the SMS technical facility contain the program for safety improvement in which there is given the risk management plan with clearly determined countermeasures and responsibilities that contains the technical, organisational, and external and cyber risks?			
Does the SMS technical facility contain the quality monitoring the both, the integral risk and all important partial risks and the corrective countermeasures for occurrence of unacceptable risks?			
Total			

Table 2. Value sale for safety level determination.

Safety level	Values v%	Number of answers "YES" in Table 1
Extreme high – 5	More than 95%	More than 68
Very high – 4	70–95%	51–68
High – 3	45–70%	33–50
Medium – 2	25–45%	19–32
Low – 1	5–25%	4–18
Negligible – 0	Lower than 5%	Lower than 4

6 RESULTS OF JUDGEMENT OF SAFETY LEVEL FOR SELECTED COMMON TECHNICAL FACILITIES AND DISCUSSION

For judgement of safety levels, it was selected 5 common technical facilities that do not belong to the critical objects in the Czech Republic; they belong to Small and Medium Enterprise (SME); specifically: chemical plant; machine plant; thermal power plant; airport; highway (ČVUT 2017). In all cases, the access to the facility documentation was regardless conditioned by agreement that real data on technical work will not be published. Therefore, it is only given the final result of investigation in the form:

1. Number of answers YES moves in interval 20–29 with mean value (median) 24.
2. The highest reached validities of work with risks:
   - there are only followed assets of technical facility,
   - there are only followed the risk sources that are in technical facility and human factor connected with wrongly performed operations,
   - there are considered partial risks and mostly integrated risk connected with workers´ health threatening,

– risks are followed only after technical facility building,
– at work with risks of technical facility it is used the process model of work with risks that has only clearly determined the acceptance criteria for risks inside, and sporadically for risks outside,
– there are performed the preventive measures only for reducing or averting the priority risks,
– it is ensured the insurance of technical facilities for case of realisation of famous risks,
– there are preferred the results of fast and less precise risk analyses,
– at work with risks there are only determined the criterions for technical and economical assessment,
– there are applied demands, standards and norms for ensuring the technical safety,
– the technical facility administrator has the safety management system based on principles of process management.

Comparison of number of answers YES with the scale in Table 2 shows that the safety level is medium in the followed SME. The judgement of level of reached validities of used techniques for work with risks shows that in practice, the system approach is missing and that in common technical facilities there are only respected the demands given in legislative and own experience with risks.

From the viewpoint of human system security, the obtained finding is not too comforting. Results of detail long-term research of technological facilities and infrastructure accidents, summarized in (Procházková 2017), fully agree with outcomes that are in many papers published in journal *Safety Science,* which are perfectly expressed in work (Konsvik, Almklov & Fenstad 2010). Structure of safety culture needs to start on top management level and to spread to lower management levels.

However, the safety culture is not all-powerful tool. Therefore, it is also necessary, so: the technological facility owner may not prefer the profit prior to human system security; and the tools used for safety formation need to correspond to present level of cognition. In these cases, the state government and legislation play important role. The government needs to ensure the corresponding education level, high qualified supervision and inspection under the technological facilities behaviour, namely starting from sitting, over building and operation up to decommission and decontamination of territory.

Subsidiary product of study of documentations of mentioned technical facilities and others (ČVUT 2017) is the detection that experts from different domains connected with technical facility do not co-operate; it is proved by records in documentations on conflicts that had not originated if expert communicate together.

7 CONCLUSION

An overview of areas that affect the selection of individual techniques work with the risks shows that where it comes on ensuring the safe technical facilities, it is a need to use the techniques for working with risks, which are based on the system concept and critical evaluation of all influences that can act on the technical facility, now and in the future.

The investigation of problems related to the work with risks of technical facilities showed that at common technical facilities, is the medium level of safety. It reflects low level of work with risks; i.e. low safety culture and weak power of government in formation of territory safety.

The judgement of validity of methods and procedures of work with risks, which are used in practice in the Czech Republic, shows that they still predominate the techniques that do not respect the system nature of technical facilities and the dynamics of development. From the study of followed technical facilities documentations, it is obvious that at formation of their safety, the experts from different fields work separately, which of course cannot guarantee optimal safety and optimal costs.

ACKNOWLEDGEMENT

Authors thanks for grant to EU and Czech Ministry for Education; project CZ.02.2.69/0.0/0.0/16_018/0002649.

REFERENCES

Ale, B., Papazoglou, I., Zio, E. 2010. *Reliability, Risk and Safety.* London: Taylor & Francis Group, 2448p.

Bérenguer, C., Grall, A., Guedes Soares, C. 2011. *Advances in Safety, Reliability and Risk Management*. London: Taylor & Francis Group, 3035p.

Briš, R., Guedes Soares, C. & Martorell, S. 2009. *Reliability, Risk and Safety. Theory and Applications.* London: CRC Press, 2362p.

Cepin, M., Bris, R. 2017. *Safety and Reliability—Theory and Applications.* London: Taylor & Francis Group, 3627p.

Coase, R.H. 1960. The Problem of Social Costs. *Journal of Law and Economics.* http://hdl.handle.net/10467/72582.

ČVUT 2017. Archives of Solved Tasks from Safety Management and Crisis Management. Praha: ČVUT.

IAPSAM 2012. Probabilistic Safety Assessment and Management Conference. Helsinki: IPSAM & ESRA, 6889p.

Keeney, R. L, Raiffa, H. 1993. *Decision with Multiple Objectives*. Cambridge: Cambridge University Press 1976, 569p.

Kongsvik, T., Almklov, P. & Fenstad, J. 2010. Organisational safety indicators: Some conceptual considerations and a supplementary qualitative approach. *Safety Science*, 48, pp. 1402–1411.

Nowakowski, T., Mlyńczak, M., Jodejko-Pietruczuk, A., Werbińska-Wojciechowska, S. 2014. Safety *and Reliability: Methodology and Application*. London: Taylor & Francis Group, 2453p.

Podofillini, L., Sudret, B., Stojadinovic, B., Zio, E., Kröger, W. 2015. *Safety and Reliability of Complex Engineered systems: ESREL 2015.* London: CRC press, 4560p.

Procházková, D. 2011a. *Analysis and Management of Risks*. ISBN: 978-80-01-04841-2. Praha: ČVUT, Praha, 405p.

Procházková, D. 2011b. *Methods, Tools and Techniques for Risk Engineering*. ISBN:978–80–01–04842–9. Praha: ČVUT, 369p.

Prochazkova, D. 2015. *Safety of Complex Technological Systems.* ISBN: 978-80-01-05771-1. Praha: ČVUT, 208p.

Procházková, D. 2017. *Principles of Management of Risks of Complex Technological Facilities.* ISBN: 978-80-01-06182-4. Praha: ČVUT 2017, 364p.

US EPA 2008. PHA Techniques in Chemical Emergency Prevention & Planning. *Newsletter*, No. 8, pp. 3–6.

Steenbergen, R., Van Gelder, P., Miraglia, S., Ton Vrouwenvelder, A. 2013. *Safety Reliability and Risk Analysis: Beyond the Horizon*. London: Taylor & Francis Group, 3387p.

Walls, l., Revie, M., Bedford, T. 2016. Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016. London: CRC Press, 2942p.

Zairi, M. 1991. *Total Quality Management for Engineers.* Cambridge: Woodhead Publishing Ltd.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

The Kursk submarine disaster in view of resilience assessment

A. Leksin & U. Barth
University of Wuppertal, Wuppertal, Germany

R. Mock
Zurich University of Applied Sciences, Winterthur, Switzerland

ABSTRACT: In August 12, 2000, the Russian Oscar-class submarine Kursk (K-141) sank during a navy manoeuvre in the Barents Sea killing all 118 personnel on board. The vessel was powered by two nuclear reactors and carry nuclear missiles which can be armed. The disaster is well documented and encompasses many socio-technical elements influencing the sequence of events finally leading to wreckage. For this, the disaster is considered as an archetypical event which might highlight the advantages as well as the limitations of resilience assessment approaches, e.g. in comparison with established risk assessment methodology. For this the paper starts with results of a literature survey with resilience metrics and areas of technical applications. The Kursk disaster is reviewed by available literature and research reports by Root Cause Analysis. The causing aspects (events, procedures, human factors, etc.) are then structured and classified according to their relevance and impact on vessel's resilience. In a next step, these aspects are contrasted to the risk assessment approach as defined, e.g. by ISO 31000. The methodological juxtaposition is intended to characterize the maturity level of resilience analysis in a real world framework as well as to elaborate major differences in validity of the underlying system analysis concepts. Finally, the pros and cons of the reviewing approach are discussed.

1 INTRODUCTION

In the context of risk analysis, the term resilience is often used nowadays. It is noticeable that both a generally accepted definition of this term and consequently a metric of resilience are missing. The differences between risk and resilience assessment often remain unclear, e. g. in connection with related terms such as availability, vulnerability, and Business Continuity Management (BCM). To a certain extent, this follows a tradition of dealing with indefinite terms such as, risk, which in turn is based on other terms that are not always clearly definable. For instance, there is a risk if several factors coincide: danger, exposure and vulnerability (cf. (Lenz 2009)).

The paper is an attempt to work out the differences and similarities between the two concepts of risk and resilience, where the approach follows the idea of "learning by doing" system assessments. An archetypical case was selected for this: The Kursk submarine disaster in 2000. On the one hand, a submarine is a self-contained socio-technical system, which simplifies considerations. The case itself, in turn, can be presented from a variety of sources. One of us (A. Leksin) can refer to less well known Russian literature as well as on feedback of one Russian accident investigator. The case was dealt with a root cause analysis (RCA), which was then used for the discussion on risk and resilience assessment.

The remaining paper is structured as follows: Chapter 2 compiles definitions of risk, resilience and the comparison of major system management terms. Chapter 3 describes the chronology of major events and causative aspects of the Kursk disaster and present a part of the resilience identification. Based on sequence of major events differences in risk and resilience assessment are elaborated in chapter 4. The results are discussed in chapter 5.

2 TERMINOLOGY

There is extensive literature research on the definition of the term resilience, e. g. (Husseini et. al. 2016, Francis et al. 2014). There is a consensus that resilience is concerned with socio-technical systems and their ability to respond to disturbances in order to maintain the specified performance. This paper follows the definition of (Lay et al. 2015) who defines resilience by a set of system abilities:

Resilience: System abilities to respond to disturbances, to monitor, to learn, and to anticipate developments.

Table 1. Comparison of major system management terms.

Term	Connotation	Intrinsic system property	Management	Focus
Risk	negative	no	external interference	(undesired) events
Resilience	positive	yes	Intrinsic	System performance
Vulnerability	negative	yes	external interference	(undesired) flaws
BCM	positive	no	external interference	(undesired) events
Availability	positive	yes	external interference	failures

Responsiveness considers all kind of disturbances into account, all deviations from specified performance levels, both positive and negative impacts. The term "disturbance" indicates that point of view is dominated by negative impacts.

Furthermore, responsiveness indicates systems immediate response to disturbances. Hence resilient systems are designed to react on disturbances in a self-managing way.

Looking at socio-technical system, humans are the carrier of its learning and anticipating abilities as covered by system management processes. Monitoring can be done both automatically/technically and by humans also depending on surveillance level.

The concept of risk is assumed to be known to the reader. The paper follows the well-established definition of risk of (Kaplan & Garrick 1981):

$\{Risk_i \mid s_i, f_i, c_i\}$,
where:
S_i: scenario identification or description
F_i: probability (or frequency) of that scenario
C_i: consequence or evaluation measure of that scenario, i.e., the measure of damage.

The frequency/consequence concept of risk is also along to common risk management standards, e.g. (ISO 31000 2009). Risk figures are usually computed by $f_i \cdot c_i$. Note, that scenario is not defined by (Kaplan & Garrick 1981) and (ISO 31000 2009). The authors will use it in terms of (imagined) sequence of events. Probability is a measurement of uncertainty of (future) events based on (statistical) data. As a consequence, risk becomes a concept of proactivity and finally preparatory by management.

Vulnerability is a well established term in IT security which is easily adaptable to any other engineered systems. According to (NIST 2012), the definition is:

Vulnerability: Weakness in an information system, system security procedures, internal controls or implementation that could be exploited by a threat source.

For this, the understanding of vulnerability follows the keyhole principle and is an intrinsic system property. Vulnerability management is then to reactively plug flaws.

However, the limits of the concepts are not always clear: According to (Lenz 2009:38–43.69), risk always coincides with danger, exposure and vulnerability. Additional components such as coping capacity and criticality (meaning and consequences upon entry) can be added to this assumption.

The concept of Business Continuity Management is a related system maintaining process to risk, resilience and vulnerability management, as defined by (SBA 2013):

Business Continuity Management (BCM) is a company-wide approach designed to ensure that critical business processes can be maintained in the event of major internal or external incidents.

The view is the management of single (major) undesired events in order to minimise their impacts. The management objective is to maintain the specified business performance level.

3 THE SUBMARINE KURSK (K-141) DISASTER

The Kursk submarine disaster took place in the Barents Sea on 12 August 2000, killing all 118 personnel on board. In this paper, the course of the disaster, if publicly known, serves as a test case for the methods of system analysis listed in chapter 1. The detailed description of the disaster is a significant part of the resilience identification step which is explained in chapter 4.2. The authors process the Kursk catastrophe on the basis of publicly available information and present a possible sequence of events. Further discussions will then be held on this basis. The case covers all elements that make an analysis interesting from very different perspectives, i.e. the interaction of people and technology in a stressful overall situation. However, the basic system performance remains simple: *ensuring the safety and health of the crew*. The question is to what extent the system analysis methods listed above would have been suitable for recognizing this accident in advance.

3.1 *Event of the Kursk disaster*

There are about 18 different disaster versions of the Russian Oscar-class submarine Kursk (K-141).

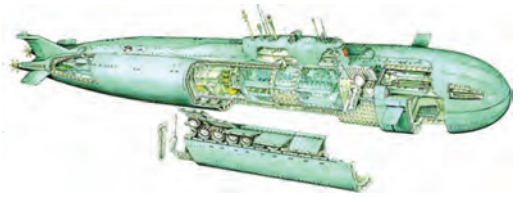


Figure 1. Example of the inner and outer hull construction with P-700 Granit "Shipwreck" cruise missiles on the bow side (Militaryarms 2017).

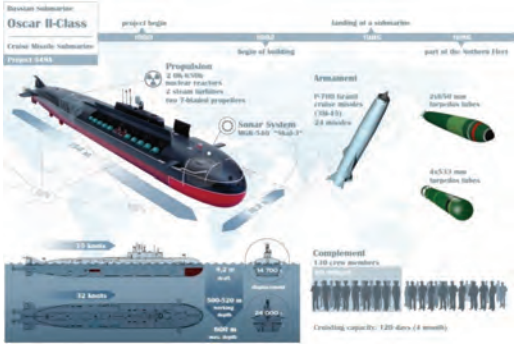


Figure 2. Characteristics of Oscar class submarine (Defending 2015).

This paper considers one of the official versions—an explosion of a torpedo but due to the influence of a second submarine. The chronology of major events and causative aspects (events, procedures, human factors, etc.) are structured and classified according to their relevance to give a better overview for the reader by describing only the important steps of the disaster. The RCA breaks down a complex scenario into individual steps (black boxes in the RCA diagram of Figure 3), which ultimately indicate a cause-effect chain. Secondary event chains can be added (grey boxes). For better understanding the boxes are numbered. The event numbers can also be found in the case description.

The description of the significant factors which had a strong influence on the worst case scenario can be traced back to 1999. Kursk was on the military mission in the Mediterranean Sea to monitor the United States Sixth Fleet responding to the Kosovo crisis. *(1)* After the successful mission the submarine returns into the stationing port of Vidyayevo. After a longer down time due to financial reason the commissioning of the submarine by the crew was under time pressure towards the end of May 2000 because of the Russian Navy large scale naval exercise planning for August 2000. Therefore the crew had a shortage of lack of planned training activities in the last approx. 9 months *(2)*. But due to the last successful mission in the Mediterranean Sea, it cannot be ruled out that part of the crew was self-confident. Either because of time pressure and/or the incorrect planning of the Marine areas by the Military-Maritime Fleet of the Russian Federation *(3)*, the way to the naval exercise area was over "underwater mountains" *(4)*. Such manoeuvre through areas of not deep-water sites of the sea can be dangerous for an Oscar-class submarine and other submarines because it is difficult to manoeuvre due to radar shadows of sonar and magnetic interference. The threat obviously increases with the condition that other countries submarines are always present in such naval exercises.

Table 2. Explosive characteristics of USET-80 and VA-111.

USET-80 (warshot torpedo)	Total weight – 2000 kg explosive weight – 200/300 kg
VA-111 (warshot torpedo)	Total weight – 2700 kg combat unit – 200 kg explosive weight 200 kg

On August 10th, 2000 the Kursk had begun the planned activities in the naval exercise near the Kola Bay. On August 12th, 2000 at 11:28 local time, two explosions were detected by various seismologists and hydroacoustics. The first explosion corresponded for ≈500 [kg] TNT equivalent and after 135 seconds the second explosion with ≈5000 [kg] TNT equivalent. Unfortunately the exact number of armed cruise missiles at the Kursk varied depending on references. Typical armament consist 24 of SS-N-19/P-700 Granit "Shipwreck" cruise missiles that were designed to defeat the best naval air defences. The missile containers are located on both sides of the deckhouse, outside the rugged boat hull. Based on the most references, photographic material and video footage the Kursk had during this naval exercise 24 of P-700 Granit "Shipwreck" cruise missiles on board. Due to the double hull construction of the Oscar-class submarine, the second explosion of the P-700 Granit "Shipwreck" cruise missiles did not initiated. Constructors considered such worst-case scenario and reinforced the inner hull with high content stainless steel about 45–68 [mm] thick. There is 200–350 [cm] gap to the 5–10 [mm] thick outer hull.

Therefore both detected explosions were in the 1st torpedo compartment. As before the exact number of dummy and warshot torpedos varies from 8 to 18 and even 24. Weapons included 18 of SS-N-16 "Stallion" (РПК-6 "Водопад"), hydrogen peroxide-fueled Type 65 torpedo (65–76A), USET-80 (УСЭТ-80) and their different types. Kursk was armed at that moment with dummy (65–76ПВ and USET-80) and warshot (65–76A, USET-80) as well as torpedo VA-111 Shkval.

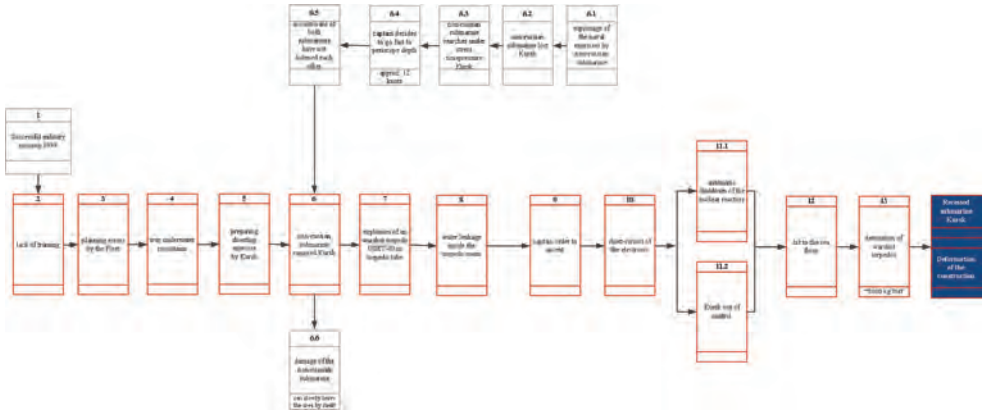


Figure 3. Sequence of events in form of RCA.

Although it was an exercise, Kursk loaded, as mentioned before, also with combat capable weapons. This means that some of the torpedo tubes are constantly in combat readiness with an armed warshot torpedo. Warshot torpedo which was used by Kursk in military mission is typically the torpedo USET-80. Table (2) shows short characteristics of the torpedoes USET-80 and VA-111 Shkval.

Based on these characteristics, the possibility of an USET-80 in the torpedo tube is very high and its TNT equivalent is near to 500 [kg] *(7)*. Also it was planned to launch the USET-80 torpedo as secondary in this naval exercise.

However, all versions of disaster reports agree on one—the first explosion was an explosion of a torpedo in a torpedo tube. As mentioned before, by all naval exercises other countries submarines are always present at the naval area as well as near main marine ports during the year. The history of underwater incidents between submarines is well known and documented in different languages and countries (Drew et al. 1998). "Among the specified accidents there are several tens of collisions of submarines, including 20 underwater collisions of Russian Navy submarines with foreign submarines. From these 20 examples 11 were in grounds of combat trainings (naval exercises) on the way to the main stationing sites of the Northern and Pacific fleet, including 8 in the north and 3 in the Pacific Ocean in a short time period from 1968 till 1993" (Aleksin 2001, Viperson 2001). Several accidents have also been registered since 1993 till nowadays. On August 12th, 2000 Kursk prepares for shooting practice in the predetermined and surveyed area radio and radio engineering investigation of surface forces of "opponent—Kirov-class battlecruiser Pyotr Velikiy" *(5)*. Due to force 3 at sea the speed of Kursk was approx. 8 knots. The Kursk had changes the depth level many times according to typical exercise. A second non-Russian submarine which monitors Kursk the last two days *(6.1)* has lost the contact *(6.2)* and couldn't find the Russian submarine *(6.3)*. They decided to emerge on periscopic depth *(6.4)* to explain this situation in order to prove if Kursk has also surfaced. On the way to periscopic depth the non-Russian submarine unexpected struck *(6.5)* with the lower cornice of a bow part from a high angle of attack to the top area of the right bow side of Kursk where were torpedo tube was charged with the warshot torpedo USET-80. Both submarines continue to move with a former speed (5.5 [m/s]), destroying each other's hulls *(6)*. Nuclear submarines of US and UK Navy are build only one 35–45 [mm] thick stainless steel hull. Thus Kursk damage was much higher. In a second after the struck with the torpedo tube located to the right board of Kursk it was crumpled on a half of the length which caused a detonation of the warshot torpedo USET-80 *(7)*. This detonation was on a line of least resistance to the hatch of the torpedo tube, destroying this and created a hole more as half a meter in diameter. Water flows inside the torpedo compartment and causes trim to the bow side *(8)*. The captain of Kursk order to ascent and increase the speed *(9)*. However short circuits of electrical networks happen because of water penetration *(10)* and due to this the emergency system block both nuclear reactors *(11.1)*. The Kursk was out of control *(11.2)* with a strong trim to the bow side and hits the seafloor *(12)*. The second explosion was initiating with the impact on the floor *(13)*. This explosion killed many crew members in the conning tower and control room (2 compartment), radioelectronics room (3 compartment), living room (4 compartment), room with diesel-generator, electrolysis installation for air regeneration, compressors of high pressure etc. (5 compartment). Although Kursk was designed

to withstand external pressure of depths of up to 1.000 [m], the second internal explosion destroyed the bulkheads between the compartments (probably till the compartment 5) which are calculated for only 10 atmospheres.

The inner hull is designed for 60 atmospheres, which prevented the explosion of the P-700 Granit "Shipwreck" cruise missiles as mentioned before in this paper.

Based on the RCA and literature statistics on similar incidents, the catastrophe must be considered by the submarine Kursk as an archetypal event. Next chapter discusses how such scenario could be implemented in the prospective of classical risk analysis as well as the possible approach of a resilience concept and its implementation problems.

4 RISK AND RESILIENCE ASSESSMENT

The established system assessment process can be summarized by three steps: *identification, analysis, evaluation*. These processes are well defined in risk assessment while there are methodological gaps in resilience assessment.

The following subsections outline these gaps and point to differences in risk and resilience assessment by exemplary application to the Kursk disaster.

4.1 *Risk assessment*

The established approaches and concepts of risk assessment are presumably known to the reader (i.e., how to perform Failure Mode and Effects Analysis (FMEA), Fault Tree Analysis (FTA), and others).

Also in risk assessment, a study starts with determination of system boundaries. When following the risk management process of, e.g., ISO 31000, then risk assessment includes the (technical) system, where the remaining risk managing processes are beyond. With regard to resilience assessment (cf. chapter 4.2), risk studies are based on a restricted system definition. As a consequence, some boxes of RCA in Figure 3 are excluded (the results of all selection criteria as compiled in this chapter are applied on RCA of the Kursk disaster and summarized in Table 3).

There are studies of navy available to support out established risk assessment approaches, e.g., (Holmboe et al. 1992) on likelihoods of threats, maturity of technologies, systems potential to develop a threat scenario.

The identification process starts with the specification of hazards and threats as well as vulnerabilities of the system and system components.

Hazard is commonly defined as a condition, circumstance or process what can cause damage. Furthermore, hazard is limited to accidental, undesired and sudden events.

Risk analysis needs the quantification of frequency and likelihood of an undesired event. Figure 5 shows the risk analysis model as applied by the authors.

The terms in Figure 5 are specified by:

- hazards are characterised by possibilities,
- results of threat factors. Scenario analysis used to anticipate how threats and opportunities might develop and are used for all types of risk with short and long term time frames,

Figure 4. Compartments of Kursk (Naked-science 2017).

Table 3. Risk analysis relevant actions according to RCA.

RCA steps	Action	*H*	*T*	*V*	*S&S*	*C*
2	negative	+	–	+	n.r.	–
3	negative	+	–	+	n.r.	–
4	negative	+	+	–	n.r.	+
5	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.
6	negative	+	+	+	+	+
7	negative	+	+	+	+	+
8	negative	+	+	+	+	+
9	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.
10	negative	+	–	+	+	+
11.1	positive	n.r.	n.r.	n.r.	n.r.	+
11.2	negative	+	–	–	–	+
12	negative	+	+	–	+	+
13	negative	+	+	+	+	+
End	negative	+	+	+	+	+

+: relevant impact regarding risk analysis; –: no-impact on defined system; n.r.: not relevant for risk analysis.

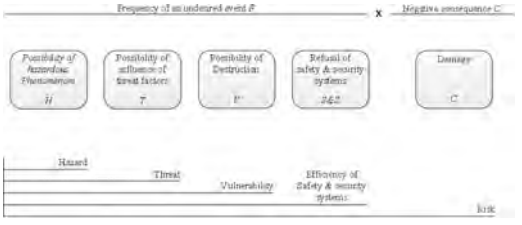


Figure 5. Risk analysis model.

- destruction of objects as a result of hazards are characterized by conditional probability,
- unavailability of safety and security systems because of combinations of non-reliability, human factors among others, are quantified by probabilities of scenario development from emergencies towards accident.

Depending on analysis goals (quantitative or qualitative) different approaches are in use. The Russian Army carries out FTA (personal communications with Saint Petersburg State Institute of Technology). However, FTA does not consider positive events and an analysis of top event *Recessed submarine* is then incomplete. The failure analysis based on qualitative approaches, as FMEA, could be added.

Within this defined framework for risk analysis Table 3 shows the relevant boxes of the RCA.

The selection process for identifying the RCA boxes relevant to risk analysis bases on the following rules:

Rule 1: The box event is within the defined system boundary.
Rule 2: The hazard is relevant within the defined system boundary.
Rule 3: If a threat (from outside) exists, it is taken into account.
Rule 4: Searching for vulnerabilities in the system.
Rule 5: Investigation of safety and security systems is a special task of risk analysis.
Rule 6: Negative consequences are always relevant for risk analysis independent from threats and vulnerabilities.

Thus, step *preparing shooting exercise (5)* is not relevant from the view of risk assessment. The *order by the captain (9)* is a positive measure and, thus not relevant for the defined scenario. The *automatic shutdown of the nuclear reactor (11.1)* is not part of the risk analysis because it is a planned safety process. Steps *(2)* and *(3)* are only considered in human reliability analysis. Finally risk is often evaluated by risk matrix. However, the risk evaluation process is not subject of this paper.

4.2 *Resilience assessment*

As mentioned in chapter 2, resilience considers extended socio-technical systems where (human as well as automated) actors are responsible for actions to positively or negatively affect system responsive-ness. Applied resilience assessment by case study brings further differences to risk assessment in understanding to light. The system assessment processes (i.e. aspect identification, analysis and evaluation) structures the following discussion.

System performance is a matter of documented system design specifications and other characteristics of embedding system entities. Then, resilience assessment of the Kursk disaster comprises all involved submarines and crews as well as the entire Northern Military-Maritime Fleet of the Russian Federation at the moment of the exercise and impacts from sea environment. The impact on the environment is not relevant. Hence, you can easily define actions, actors, and system boundaries in Kursk example in contrast to, e.g., infrastructures. Within this framework, the identification process starts with the specification of system performance P. For this, two approaches are common in resilience assessment (cf., e.g., Mock 2018): either the analysts decide to model time-depending performance $P(t)$ or they compile a set of n resilience impacting aspects $P = \{a_1; a_2; \ldots; a_n\}$. $P(t)$ can be easily defined (e.g. safe and secure transport of crew and cargo during mission time) but finding a corresponding measurement is not always as straightforward as, e.g. oxygen content of the breathing air during mission time. Note, that availability, as shown in Table 1, is a performance model $P(t)$ showing the probability course of operability of a system. Maintenance and repair are considered as activities to keep the system resilient, and are actions of responsiveness.

The identification process by compilation of a set of aspects influencing resilience appears plain, e.g. the number of redundancies of life supports systems, educational level of crew, repair, etc. However, time dependency and the representation of systemic relationships are lost then.

The resilience analysis process by $P(t)$ follows the common processes of formal mathematical/physical of system modelling and simulation and will not be discussed here. However, the RCA presentation of the Kursk disaster in Figure 3, which follows a timeline of succeeding events, is considered as a simple representor of $P(t)$ after revision towards resilience (see Table 4). $P(t)$ analysis needs the specification of normal operation bandwidths of total system performance. For instance, the oxygen content on a submarine can be above or below a lethal threshold. The life support system may be able to provide a breathable atmosphere again, but this can be too late for the crew. In terms of resilience analysis, the responsiveness of the entire submarine system is lost as safe transport has ended. These points to specific views in resilience analysis: Total loss of performance or functionality (worst case) is excluded from analysis ("If dead you are not resilient any longer"). The analysis of impacting aspects $\underline{P}$ needs the definition of a resilience metric which is still under discussion in academia. Table 4 summarises the findings in resilience identification and analysis by the Kursk example as represented in Figure 3.

Table 4. Resilience assessment for performance "safe and secure transport of Kursk crew and cargo during mission time".

RCA	Sub-system	*P*	Action	Actor(s)
1	fleet	+	success	Kursk crew, fleet
2	Kursk	–		Kursk crew
3	fleet	–	preparedness	Kursk crew, fleet
4	environment	–		
5	Kursk	+	exercise	Kursk crew
6	Kurs, other sub.	–	ram	Both crews
7	Kursk	–	explosion	
8	Kursk	–	leakage	
9	Kursk	+	order	Kursk's captain
10	Kursk	n.r	short-circuit, loss of control	
11.1	Kursk	n.r	shutdown reactor	
11.2	Kursk	n.r	loss of control	
12	Kursk, environment	n.r	grounding, loss of control	
13	Kursk	n.r	detonation, loss of control	
End	Kursk	n.r		

+/–: positive/negative impact on resilience; n.r.: not relevant for resilience assessment purposes.

As figured out in Table 2, resilience assessment ends with the loss of control of the Kursk ("If faint, then you are no longer resilient"). Step *(6)* can be similarly analysed by considering RCA steps *(6.1)* to *(6.6)* which introduces the second submarine into analysis. The marine environment (step *(4)*) has been identified as challenging for submarines which does not support safe transport.

Resilience evaluation is the process of assessment. Again, the analyst depends on how resilience analysis been performed. In case of modelling *P(t)* a characteristic value needs to be defined, e.g. the ratio of resilient operation mode to total mission time. This is equivalent, e.g. to reliability and availability analysis. The evaluation of the set of impacts *P* needs the definition of a resilience metric comparable to risk prioritisation value RPV in risk analysis and provided by FMEA. Evaluation criteria of acceptance/non-acceptance of resilience analysis results still needs to be defined (a resilience priority value is introduced in (Mock 2018)).

Table 5. Juxtaposition of risk and resilience assessment.

RCA events	Risk assessment	Resilience assessment
1	*N*	*Y*
2	*N*	*Y*
3	*N*	*Y*
4	*Y*	*Y*
5	*N*	*?*
6	*Y*	*Y*
6.1	*N*	*Y*
6.2	*N*	*Y*
6.3	*Y*	*Y*
6.4	*?*	*Y*
6.5	*?*	*Y*
7	*Y*	*Y*
8	*Y*	*Y*
9	*N*/*Y*	*Y*
10	*Y*	*N*
11.1	*Y*	*N*
11.2	*Y*	*N*
12	*Y*	*N*
13	*Y*	*N*
Summary	Y: 10 of 19	Y: 13 of 19

4.3 *Synopsis*

Based on RCA every single step is discussed from the side of resilience assessment in comparison to the risk assessment. Table 5 differs between the selections of *yes (Y)*, *no (N)*.

Avoidance of worst case scenario or disaster is the aim of a risk assessment. Therefore, positive or neutral (from the view of risk analysis) steps such *(1)*, *(3)*, *(5)*, *(6.1)*, *(6.2)* are not considered. But e.g. step *(9)* could be considered if the order of the captain is incorrect. Step *(2)* can be also considered only in case of human reliability analysis. Equivalent to a worst case scenario "meltdown of nuclear reactor", step *(13)* must be take into account by this disaster. Similar, step *(6)* could correspond to a "plane crash on nuclear power plant" scenario (external event). Step *(5)* is not a malfunction or optimization, but an important point in the overall process. Due to the defined performance indicator step *(6.2)* must be considered too (the second submarine is part of the whole system). As mentioned before, with defined performance indicator—safe implementation of the naval exercise for the crew—the resilience analysis ends with the step *(10)*. As mentioned in chapter 2, risk assessment often uses the basic frequency/consequence definition to get calculation values. However, the next step after description of the RCA from the side of resilience

analysis is complicated due to absence of any useful values and equations which could be support the calculations and as a result the evaluation of the defined system and performance indicator.

5 CONCLUSIONS

Resilience assessment should be different from risk assessment and other related concepts and approaches. For instance, risk assessment is basically restricted to undesired events and does not cover the extended view of technical systems. On the other hand, event identification highly depends on the definition of system performance indicating resilience as a measurement of system quality.

The issue of applied resilience assessment is shown by considering the archetypical case of Kursk submarine disaster. The detailed description of sequence of steps by Root Cause Analysis shows that a precious accident analysis is significant for identification of aspects which have impacts on resilience. The specification of system performance and the view of extended socio-technical systems increase resource requirements (time, expertise, etc.) of auditing. This way of thinking definitely uncovers additional elements of system disturbances.

However, resilience analysis is still in its beginnings and there is no commonly accepted methodology and metric. In summary, resilience assessment is different to risk assessment in some ways and shows promising aspects in extended system analysis. However, further steps towards operationalisation of the resilience concept are needed.

REFERENCES

Aleksin, W. (2001). Версия контр-адмирала Алексина: "Курск" уничтожила иностранная подлодка. https://i-korotchenko.livejournal.com/1136736.html?page = 1#comments.

Defending (2015). Подводные лодки проекта 949A «Антей» https://defendingrussia.ru/enc/apl_kr/podvodnyje_lodki_projekta_949a_antej-1949/.

Drew, C., S. Sontag & A.L. Drew (1998). Blind Man's Bluff: The Untold Story Of American Submarine Espionage. ISBN 1-891620-08-8.

Francis, R. & B. Bekera (2014). "A metric and frameworks for resilience analysis of engineered and infrastructure systems," *Reliability Engineering & System Safety*, vol. 121, pp. 90–103.

Holmboe E. & S. Seymour (1992). APL`S Submarine security program. *Johns Hopkins APL Technical Digest*, vol. 13, number 1.

Hosseini, S., K. Barker & J. E. Ramirez-Marquez (2016). A review of definitions and measures of system resilience, *Reliability Engineering & System Safety*, vol. 145, pp. 47–61.

ISO 31000 (2009). "Risk Management – Principles and Guidelines (Iso 31000:2009)." Geneva International Organization for Standardization (ISO).

Jean-Michel Carré (2004). *Video footage*: Kursk: A Submarine in Troubled Waters (French: Le Koursk, un sous-marin en eaux troubles). https://www.canal-u.tv/video/cerimes/koursk_un_sous_marin_en_eaux_troubles.13454.

Kaplan, S., & B. J. Garrick (1981). On the Quantitative Definition of Risk. *Risk Analysis 1, no. No. 1* (1981): 11–27.

Lay, E., M. Branlat & Z. Woods (2015). A practitioner's experiences operationalizing Resilience Engineering, *Reliability Engineering and System Safety*, pp. 63–73.

Lenz, S. (2009). Vulnerabilität Kritischer Infrastrukturen, Forschung im Bevölkerungsschutz, Bundesamt für Bevölkerungsschutz und Katastrophenhilfe, Bonn 2009.

Militaryarms 2017. Подводные лодки проекта 949A «Антей»: история создания, описание и характеристики. https://militaryarms.ru/voennaya-texnika/podvodnye lodki/949a-antej/. http://militaryrussia.ru/blog/topic-398.html.

Mock, R. (2018). A quantitative approach for applied resilience assessment audits. In Proc. of European Safety and Reliability Conference (ESREL 2018).

Mock, R. & C. Zipper (2017). Risiko: Das Ende Eines Konzeptes? Sicherheitsforum (2017): 3.

Mock, R. & C. Zipper (2017). Embedding resilience assessment into risk management. In Proc. of of European Safety and Reliability Conference (ESREL 2017).

Naked-science (2017). Как устроена атомная подлодка https://naked-science.ru/article/tech/kak-ustroena-atomnaya-podlodka.

NIST (2012). Guide for Conducting Risk Assessments—Information Security (SP 800–30 Rev.1), National Institute of Standards and Technology (NIST), Gaithersburg, Sept. 2012.

SBA (2013). Recommendations for Business Continuity Management (BCM), Swiss Bankers Association, Bale, Aug. 2013 http://shop.sba.ch/999925_e.pdf.

Viperson (2001). *Interview*: Гость программы – Контр-адмирал в запасе Валерий Алексин http://viperson.ru/articles/gost-programmy-kontr-admiral-v-zapase-valeriy-aleksin.

VK (2017). *Video footage*: Официальная версия в фильме о катастрофе АПЛ Курск. https://vk.com/videos543641?z=video61207156_456239177%2Fpl_543641_2.

Гибель "Курска" (2015). *Snipping from Video footage*: Гибель Курска. Следственный эксперимент (2015) https://www.youtube.com/watch?v=INJTLXL5GrQ&has_verified=1.

Рязанцев, В. 2017. В кильватерном строю за смертью. Почему погиб «Курск». ISBN: 978-5-906716-88-0.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A quantitative approach for applied resilience assessment audits

R. Mock
Institute of Sustainable Development INE, Zurich University of Applied Sciences, Winterthur, Switzerland

ABSTRACT: Today's infrastructural systems are expected to be safe and resilient. In this context, assessment of such systems faces two principal challenges: common approaches in risk assessment have reached their limits in methodology and feasibility in assessing complex and interconnected systems. On the other hand, resilience assessment is in its beginnings and lacks, e.g., a commonly accepted resilience metric. The paper starts to specify a practical definition of resilience and assigned metric: Resilience is characterised by influencing recovery properties of a socio-technical system. Actors and actions are carriers of these properties. This corresponds to the views of system representation by Use Case Diagrams (UCD). In order to quantify an UCD, actions are validated by assessing their compliance level L. Actors are associated with their abilities to respond, monitor, learning, and to anticipate developments. The result is given by the Resilience Priority Value REPV = $L \cdot I$ of actors and overall system. The resilience assessment process is exemplified by a case study of a car park guidance system.

1 INTRODUCTION

Current infrastructural systems show a high level of complexity and technical development will further strengthen this trend. As consequence, such systems will become increasingly difficult to handle for system operating organisations (private and non-private) and managers involved. It already looks as that methodological or practicable limits of, e.g., established risk assessment approaches have been reached. New terms reflecting newly desired system properties (e.g., resilience, smartness) are emerging too. However, the methodology of resilience assessment is in an early stage of development and not (yet) in the focus of most organisations. This is also due to the lack of a practicable, quantitative metric of resilience. In this context, the paper presents an approach to facilitate applied resilience assessment audits. Following the concept of system representation and resilience quantification, the remaining paper is structured as follows: Chapter 2 defines resilience and terms in use. The results of a literature survey on resilience definitions in specified engineering domains are given in Chapter 3. In Chapter 4, resilience assessment is utilised by using quantified Use Case Diagrams (UCD). The case study presented in Chapter 5 serves to proof the concept. The paper closes with discussion of pro and cons of approach and context.

2 TERMS

The view in applied research and development in resilience analysis covers the requirements of users in organisations and enterprises (mainly small to mid-sized enterprises SME). Hence, any resilience assessment approach needs to cover additional demands (cf. (ISO-31010 2009)), which might be unimportant to basic research. A major concern of organisation is method efficiency. Thus, the resilience assessment approach as introduced in this paper aims to finally reach practicability as known in basic risk assessment audits, fire and explosion inspections, annual tests of vehicle safety (Ministry of Transport (MOT) test), among others. According to the author's experience, such a system analysis must be typically performed from one person in about one day.

There are already exhaustive literature surveys on terminology of resilience, where the most common understanding of resilience is exemplified by Scholz et al. (2012): Resilience is the ability of the system to adjust its functioning […] following changes and disturbances, so that it can sustain required operations.

Hosseini et al. (2016) also consider system recovery abilities as crucial part of resilience, where recovery is the capability of a system to absorb and adapt to disruptive events.

Lay et al. (2015) labour characteristics and abilities of resilient systems in more detail, which is finally the definition as used in this paper:

DEFINITION 1 (RESILIENCE) *characterises the abilities of a system to respond to disturbances, to monitor, to learn and to anticipate developments.*

With this, resilience belongs to a set of related engineering terms characterising system capabilities by attributes or system performance function $P(t)$, e.g., availability $A(t)$:

Definition 2 (Availability) *is the probability of finding an unit in an operational condition at time t.*

This definition of availability follows, e.g., DINEN61703 (2002). Note that $A(t)$ encompasses maintenance, which is a system ability to respond to disturbances (failures, incidents, over-fulfilment, etc.), to monitor them (failure identification) and to learn (optimising maintenance processes) and anticipate trends (expected failures). The latter is covered by reliability management processes and preventive maintenance. So far, resilience looks like the generalisation of availability towards the analysis of extended socio-technological systems. Furthermore, management and associated processes are considered as an integral part of such a system in resilience assessment (cf. (Leksin et al. 2018)). By contrast, management tends to play the role of an external controller in risk assessment. Business continuity (BC) also follows the concept of system recovery but concentrating on business impacts:

Definition 3 (Business Continuity) *is a corporate capability. This capability exists whenever organisations can continue to deliver their products and services at accep*table *predefined levels after disruptive incidents have occurred (cf. ISO 22301: 2012).*

Resilience is in line with established approaches to manage deviations, e.g., risk management according to ISO-31000 (2009). Note, that any system assessment approaches cover the subprocesses of event identification, analysis and evaluation. The view of resilience, availability and business continuity is to describe system capabilities with associated performance functions where risk relates to (undesired) events.

3 STATE OF THE ART

The following results of a survey on resilience definitions concentrates around engineering domains which are then used to reason the way of utilisation of resilience assessment as proposed in Chapter 4. Hosseini et al. (2016) give an extended review of definitions. They state that the engineering domain "includes technical systems designed by engineers that interact with humans and technology, such as electric power networks". There, engineering resilience is defined in various points of views:

- Sum of the passive survival rate (reliability) and proactive survival rate (restoration) of a system.
- "Intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances, so that it can sustain required operations under both expected and unexpected conditions" (Hollnagel et al. 2010).
- "Infrastructure resilience is the ability to reduce the magnitude and/or duration of disruptive events. The effectiveness of a resilient infrastructure or enterprise depends upon its ability to anticipate, absorb, adapt to, and/or rapidly recover from a potentially disruptive event" (NIAC 2009).
- Factors, e.g., minimisation of failure, limitation of effects, administrative controls/procedures, flexibility, controllability, early detection.

Furthermore, Hosseini et al. (2016) state that:

- Many definition focus on the capability of system to *absorb* and to *adapt* to disruptive events, and *recovery* is considered as the critical part of resilience.
- For engineered systems, reliability is considered to be an important feature (e.g. nuclear power systems).
- Returning to steady state performance is needed for resilience; some definitions do not impose that the system returns to the pre-disaster state (e.g. infrastructure).
- Multidimensionality and threat-dependency of resilience definitions.

These lists and the findings of Chapter 2 substantiate: Resilient systems show abilities to preparedness and recovery in general. Then, preparedness is typically covered by a descriptive (i.e. qualitative) and case specific set of attributes $\max_{\bar{x}\in(0,\infty)} f(\bar{x}) - g(\bar{x}) - h(\bar{x})$. System performance $P(t)$ uses various modelling and simulation approaches to model system dynamics and performance $P(t)$ according to the resilience triangle concept. Definition of preparedness attributes follows methods to design and evaluate questionnaires, check lists, etc., and graphs represent relationships of system abilities. This notation is useful to characterise the common approaches of resilience system analysis:

- *Attributive*: Starting point is the compilation of system specific attributes $\mathbb{A} = \{a_1, a_2, \quad, a_n\}$, which characterises the presence of or impact on resilience properties, e.g., awareness, flexibility, risk management, competence, and redundancy. Then, analysis follows methods to evaluate questionnaires, check lists, etc. or uses any graphs to represent relationships.
- *Performance*: System performance modelling needs the specification of time dependent and system specific performance measurements, e.g., availability (i.e. function showing the alternation of operation and maintenance), returns (money), among others. Note that $P(t)$ already comprises the recovery properties. Modelling parameters of might base on $\mathbb{A}$. Hence, $P(t)$ can be modelled by graph theory, (e.g. Markovian

models, system dynamics, state diagrams), classical mechanics considering damped harmonic oscillation, (i.e., spring model) as well as control theory using proportional–integral–derivative controller (PID controller).

The definition of resilience performance $P(t)$ results in generic system statements:

- $max\left\{\left|\Delta P(t)\right|\right\} \leqq b$: There is a specified band width b (i.e. defined upper and lower performance levels) which defines system operability (Mock and Zipper 2017).
- $P(t) = 0$: Total system failure (worst case) which ends reparability. Recovery is not possible and system re-construction is equivalent to a different and thus new system. Reconstruction is only possible by supporting measures of the superior system (cf. definition of disaster). Hence, $P(t) = 0$ is associated with fully operable (new) system which is a common boundary condition in reliability analysis.
- $P(t) = k$ and $\dot{P}(t) = 0$: Nominal operation on constant performance level k.
- $\dot{P}(t) \neq 0$: System is in resilience mode.
- $\ddot{P}(t) \neq 0$: Acceleration of performance alteration is an indicator of resilience request.

In summary, there is no common understanding of resilience and how to model resilience. Many authors define key abilities of resilience by their own. The lowest common denominator is the ability of a resilient system to respond to disturbances and, hence, functional preserving capabilities (i.e. recovery). In order to verify these findings, Table 1 uses Def. 1 of resilience to allocate the specified attributes of resilience for infrastructure systems as named in references.

Table 1 also shows that "respond" to disturbance is the main property of a resilient system. The remaining properties are less frequently listed. From the author's point of view, this table exemplifies the uncertainty of how to deal with resilience key capabilities other than "respond", and that it is still necessary to utilise the resilience concept for concrete applications.

Table 1. Key abilities of resilience for infrastructure systems.

System	Respond	Monitor	Learn	Anticipate	Reference
System homeland security	robustness, consequence mitigation	threat and hazard assessments	adaptability, harmonisation od purposes, comprehensive of scope	risk-informed planning and investment	(Hosseini et al. 2016)
Telecommunication network	maintainability	reliability, safety, confidentiality, availability, integrity performance	–	–	(Hosseini et al. 2016)
Communication network	defend, remediate, recover	detect, diagnose	refine	–	(Hosseini et al. 2016)
Infrastructure system	absorb, recover	–	adabt	anticipate	(Lay et al. 2015)
Critical infrastructure	responsiveness, timely recovery, minimum level of service while undergoing changes, flexibility	–	–	coordinated planning	(Lay et al. 2015)
Infrastructure network	ability to regain a previous state	–	adopt the stress–strain model	–	(Bergström et al. 2015)
Infrastructure	recovery (bouncing back)	–	–	–	(Lundberg and Johansson 2015)
Critical infrastructures	robustness, (availability of redundancy, resourcefulness and efficiency of supporting measures)	–	–	–	(BABS 2013)

4 UTILISATION

Chapter 4 identifies the interrelationships among system elements by UCD and how to quantify system resilience by attributes as given in Def. 1.

4.1 *Use Case Diagram UCD*

The Unified Modeling Language (UML) is a quasi-standard of system representing diagrams offering conformance in syntax and semantics. UML 2.5 defines thirteen types of diagrams, divided into three major categories: Structure Diagrams, Behaviour Diagrams, and Interaction Diagrams (cf. www.uml.org) UML diagrams are standardised by (ISO-19501 2005), where UCD is the most simple structure diagram in UML. This Chapter gives a short introduction into the concept of UCD by referencing to the mentioned standard unless otherwise stated.

Definition 4 (Use case) *is a kind of classifier representing a coherent unit of functionality provided by a system, a subsystem, or a class as manifested by sequences of messages exchanged among the system (subsystem, class) and one or more outside interactors (called actors) together with actions performed by the system (subsystem, class).*

A use case is shown as an ellipse containing the name of the use case which characterises activities of actors.

Definition 5. *"An [actor] defines a coherent set of roles that users of an entity can play when interacting with the entity. An actor may be considered to play a separate role with regard to each use case with which it communicates".*

The standard stereotype icon for an actor is a "stick man" figure with the name of the actor.

There are three types of relationships among use cases (actions) and association

- *Association*: The participation of an actor in a use case. In Figure 1, associations are shown by solid lines.
- *Extend*: An extend relationship from use case A to use case B indicates that an instance of use case B may be augmented (subject to specific conditions specified in the extension) by the behaviour specified by A.
- *Include*: An include relationship from use case E to use case F indicates that an instance of the use case E will also contain the behaviour as specified by F.
- *Generalisation*: A generalisation from use case C to use case D indicates that C is a specialisation of D.

The author considers UCDs as especially useful for resilience assessment purposes in order to depict actors and associated actions on technical and organisational level (i.e., modelling socio-technical systems).

4.2 *Semi-quantified resilience assessment by UCD*

Establishing the resilience assessment audits at organisations needs an approach which is resource saving and follows established ways of system representation, e.g., by UML. In a first step, it is suggested to use the interrelationships among system elements by UCD and to assess system resilience by means of the resilience attributes as given in Def. 1. As mentioned above, the UCD differentiates between actors and actions. In engineering terms, actions can be evaluated by assessing their level of compliance with standards, best practices, etc. It is assumed that a high compliance level has a positive effect on the system resilience. Actors are the carriers of system resilience where their impact on recovery abilities is evaluated. For this, the Resilience Priority Value *REPV* of an actor is introduced, which uses the definition of resilience as given in Def. 1:

$$REPV = L \cdot I(d,m,l,a), \tag{1}$$

where

- *REPV*: Resilience Priority Value of an actor
- *L*: compliance fulfilment level of an use case (action)
- *I*: impact of recovery ability of an actor
- *d*, *m*, *l*, *a*: actor's abilities to respond disturbances, to monitor, to learn and to anticipate.

All assessments use ordinal scales of range [1, 2, …, 10], where 1 indicates best and 10 worst cases. The concept follows the familiar idea of estimating risk priority figures, even if resilience is understood as a positive system property.

So far, the assessment of *L* is the result of audits and expert judgement about the proven record of reached compliance levels of actions or associated technology, e.g., the operation of IT security management. In the best case, the rating of *L* bases on already available reports of compliance certifications, e.g., according to ISO/IEC-27002 (2005).

Actors are considered as the intrinsic carriers of resilience. As mentioned above and following Def. 1, the impact of recovery ability *I* depends on four attributes, which are rated by ordinal scales of range [1, 2, …, 10]. *I*(*d*, *m*, *l*, *a*) of an actor is assessed by the mean value of these abilities. The abilities of learning *l* and anticipation *a* are currently covered by humans. However, trends in smart manufacturing and artificial intelligence blur this classification.

The analysis of system resilience needs rules to make use of UCD. For a very first proof of concept, the following procedural steps are defined:

1. An estimated compliance fulfilment level $L_{i,e}$ of $i = 1,2, \quad ,k$ is assigned to each use case U_i.
2. Considering relationships for assessing L_i of U_i
   – Apply the mean value of all *incoming* extend associations
   – Apply the mean value of all values assigned to *outgoing* include associations
   – Compute the mean of both values
3. Rounded off to the next integer
4. Repeat the process until all use cases (actions) are assessed.

In summary, every use case (action) U_c is characterised by a number of extend and include relationships R (i.e., edges): $U_c\left(R_{c;ext};R_{c;inc}\right)$. For further resilience computation, only subsets of relationships are needed. For this, every U_c is assessed by the mean values $\bar{x}$ of compliance levels L of associated incoming extend relationships and outgoing include relationships, i.e. $U_{c;L}\left(\bar{x}_{c;L_{in-ex}};\bar{x}_{c;L_{out-incl}}\right)$. The mean of both values finally gives the looked for compliance level L_{U_c} of an action.

Next, every actor $A_j, j = 1,2, \quad ,k$ is evaluated by the following rules:

1. Assign values of $I_j\left(d_j,m_j,l_j,a_j\right)$
2. Compute the mean of assigned values in I_j which is the looked for impact value of recovery ability of an actor $I_{j,a}$.

Then every actor shows an impact value $I_{j,\,a}$ and is assigned with a use case value L_i (if there are more than two associations use the mean value of L_i's). With that, all values are given to compute $REPV_J$ as defined in Eq. 1. System resilience is estimated by the mean value of all actors' *REPV* and again rounded of to next integer.

4.3 *Proposed audit process*

The utilisation process of resilience assessment is finalised by auditing a system. The following steps roughly structure such an audit:

– Step 1 – Drafting use cases and actors of socio-technical system to be audited
– Step 2 – Transfer of use cases and actors into the UCD
– Step 3 – Quantification of UCD
– Step 4 – Evaluation of results and *REPV*.

Steps 1 and 2 follow the basic steps of creating any UCDs. In terms of risk and resilience assessment Step 1 covers the identification process and Step 3 the analysis process. The resulting REPVs might be evaluated by a matrix or threshold approaches as known in risk assessment. However, this step is not elaborated in this paper.

The suggested resilience audit process opens developments towards semi-automated processes to support auditors. The generation of UCDs is a well-known activity in software engineering and there are many tools available to do that (cf. Chapter 5). The computation process of UCDs follows ideas of using complexity metrics as common to characterise computer codes and associated UMLs (cf. (Mock et al. 2015)). Altogether, it is intended to develop the following audit supporting steps: The auditor has to identify actions and actors for UCD generation. Both aspects are plant or system specific. However, there are repetitive elements, e.g., associated with IT security, fire and explosion protection, and occupational safety. These elements are typically standardised and subject of compliance checks. Frequently occurring or, e.g., industry branch specific actions and actors can thus be deposited in a tool library. An auditor then selects the appropriate ones by a drop down menu.

In a next step, the auditor has to identify and create the relationships among actions and actors. This step is tool supported too.

Finally the auditor needs to input the estimated impact values *I* for every action with only one relationship and to assign $I_j\left(d_j,m_j,l_j,a_j\right)$ for every actor. The remaining computations will be done by the tool.

In the end, the auditor needs more knowledge in system relationships as, e.g., for filling check lists or to perform an FMEA. On the other hand, the usual actions and actors as well as associated $I_{j,\,a}$ and L_i ratings should be known by an experienced auditor as they are close to common checks and results of site-specific compliance checks.

5 CASE STUDY: CAR PARK GUIDANCE

The audited system in this case study is a car park guidance system as implemented in a Swiss city. The system is designed to manage and optimise car traffic flow between a parking lot outside town ("Castle") and a car park building in city centre. ("Town"). All parking spaces are equipped with sensors, networked and controlled by a Supervisory Control and Data Acquisition system (SCADA). Parking space allocation is visible for drivers by displays in "Town".

Step 1 – Drafting actors and use cases

The actors are defined by

– Driver (Family): The family is on a getaway. The Driver (Family) speaks German and strictly follows the parking guiding displays in

order to avoid looking for parking space. The DRIVER (FAMILY) first drives to the display at the car park in the city.
- DRIVER (TOURIST: The foreign DRIVER (TOURIST does not understand German and feels unconfident with display symbols. Hence, this driver ignores the parking guiding displays and makes ad-hoc decisions where to park.
- CAR PARK OPERATOR (TOWN): There are no specifications about sensors. CAR PARK OPERATOR (TOWN) is assumed to be responsible for car park and system operation. The operator might start parking place managing activities.
- PARKING LOTS OPERATOR (CASTLE): There are no specifications. PARKING LOTS OPERATOR (CASTLE) is assumed to be responsible for 84 parking lots and system operation. The operator might start parking place managing activities.

Actions (use cases) are defined as

- *Display:* The only car parking display is located at the car park "City" in town and shows the number of free parking spaces at "City" (max. 340) and "Castle" (two parking spaces small and big: 10 + 74 = 84) nearby the Castle. It is assumed that display hardware does not fail at any time within the observation period of 4.5 years of operation. The associated system software is remotely updated and patched via Internet.
- *Gateway*: Kerlink LoRa IoT Station (2 identical stations) "is an industrial solution suitable for people who want to mount the gateway outside and who have sufficient technical skills to connect, mount and maintain the device themselves. … somewhat older software, that is being used, [and] this device will do the job. A trained software engineer will be able to update the device using the [firm] software" (source: thethingsnetwork.org). The Gateways link the 84 *Sensors(Castle)* with the Internet by the Swisscom Mobile network.
- *Sensors (Castle)*: The "Fastpark Flush-Mounted Sensor" (in total 84 sensors) are part of Parking Management System (PMS). "The wireless system uses smart sensors installed in parking spaces and guides drivers to areas with vacancies via electronic panels …" (source: www.worldsensing.com). The *Sensors(Castle)* are linked with associated Gateways and Parking Management System PMS(Castle). Sensors might fail but are not maintained in observation time. The sensors are battery operated and uses the novel Low Power Wide Area (LPWA) technology for gateway communication.
- *PMS Operation*: PMS operation and associated data storage is done by a separated EU computing centre.
- *Sensors (Town):* There are no specifications about sensors of car boxes. It is only assumed that there are sensors which provide display data.
- *PMS (Castle)*: SCADA device in order to process and monitor data from *Sensors(Castle)*. The SCADA serves as Human Machine Interface (HMI). The operator is considered as an integral part of *PMS (Castle)* who then might startparking place managing activities.

Step 2 – Creating UCD

Information on actors and actions is used to build up the UCD of Figure 1. The software tool PlantUML (www.plantuml.com) creates UCDs from textual inputs. It is a plug-in, e.g., of Eclipse. The possibility of integrating PlantUML into various software development frameworks is considered as pre-condition for further resilience software tool development.

Step 3 – Quantification of UCD

Table 2 shows the quantification of UC as given in Figure 1.

Computation in Table 2 is exemplified by considering the Action U_8: The auditor estimates and

Figure 1. UCD of case study.

Table 2. Estimation of compliance fulfilment level L_i by use case (actions) U_i.

i	Action Ui	Li, e	L_i
1	Sensors Castle	8	$= L_{1,e}$
2	Gateways	–	8
3	Internet	9	$= L_{3,e}$
4	IT Security	8	$= L_{4,e}$
5	EU Comp. Centre	9	$= L_{5,e}$
6	Operates sensors in build.	9	$= L_{6,e}$
7	Low power WA	9	$= L_{7,e}$
8	Operates PMS	–	9
9	Display	–	7
10	reads display	–	9
11	Parks	10	$= L_{10,e}$

$L_{i,e}$: input by auditor; L_i: input by computation

Table 3. Estimation of impact value of recovery ability *Ij* of actors.

j	Actor A_j	d_j	m_j	l_j	a_j	I_j	Mean $\bar{x} = I_j$
1	Car par operator (town)	8	7	6	7	7	7
2	Parking lot operator (castle)	9	10	7	9	8	8
3	Driver (family)	9	7	6	5	6	6
4	Driver (tourist)	7	5	2	1	3	3

Table 4. Resilience priority value REPV of actors.

j	Actor A_j	L_i	I_j	$REPV_j$	Comment
1	Car par operator (town)	9	7	63	
2	Parking lot operator (castle)	9	8	72	
3	Driver (family)	10	6	60	$L_3 = \frac{10+9}{2}$
4	Driver (tourist)	10	3	30	

assigns a compliance fulfilment level of $L_{8,e} = 8$ to the action "operates PMS". This action points to three other actions by include relationships associated with (8+9+9). The mean value including $L_{8,e}$ gives 9. There is an input of an extend relationship $L_{4,e} = 8$ which then gives the final mean value of $L_8 = \frac{9+8}{2} = 9$ (rounded off to the next integer).

Every actor is assigned to an impact value of recovery ability using I_j d_j, mj, l_j, a_j.

Step 4 – Evaluation of results and REPV

As a result from Table 4 the actor Driver(Tourist) shows lowest resilience properties. The overall resilience value of the car park guidance system is the mean of all REPV's, i.e., $REPV_{syst} = 56$ indicating a system with medium resilience.

6 CONCLUSIONS

In view of extended socio-technical system analysis, developing a closed resilience assessment approach is subject of research (cf. (Mock and Zipper 2017). However, this research only makes sense if the understanding of resilience finally results in a different approach as already established by the concepts of, e.g., risk, BCM and availability. From the author's experience, discussion about resiliency often follows synonymous paths as already given by these established concepts (cf. (Leksin et al. 2018)).).

On the other hand, resilience assessment methodology is in its beginnings and still beyond entrepreneurial interests and has not fixed as state of technology yet. Thus, the paper is understood as a step toward utilisation of resilience assessments of complex systems. For this, a simple REPV is defined and the assessment process uses standardised system representation by UCD, which properly differentiates between actions and actors. This property covers well the inclusion of socio-technical aspects, where actors are carriers of major properties of resilience (e.g., learning). They are integral parts of the audited system, which is then becomes describable as a socio-technical system. By defining rules to quantify UCDs, the proposed resilience assessment approach opens paths for software tool development in order to support resilience assessment audits of, e.g., infrastructural systems. The case study serves as a proof of concept.

Discussions at ESREL conference in 2017 have given rise to fears that the inclusion and detailed understanding of the technical functioning of (infrastructural) systems could be neglected in resilience assessments. The use of UCD provide a practical way out of this situation, since UCDs are based on comprehensive descriptions of actions, actors and their relationships supporting a systemic analysis approach.

The proposed concept of system assessment supports auditors to check to what extend infrastructural systems are resilient. However, the approach still needs verification of quantification rules, which are presumably too simplistic. The approach also needs an extended review based on a broader application example. Further developments consider the inclusion of complexity measures in order to increase the meaningfulness of UCD quantification.

REFERENCES

BABS (2013, Apr. 29). Risikoausbildung BABS: Glossar der Risikobegriffe., Bundesamt für Bevölkerungsschutz. Bergström, J., R. van Winsen, & E. Henriqson (2015). On the rationale of resilience in the domain of safety: A literature review. *141*, 131–141.

DIN-EN61703 (2002). Mathematische Ausdrücke für Begriffe der Funktionsfähigkeit, Verfügbarkeit, Instand-haltbarkeit und Instandhaltungsbereitschaft. DIN EN 61703:2002–09, DIN Deutsches Institut für Normung.

Hollnagel, E., C.K. Tveiten, & E. Albrechtsen (2010, August). Resilience engineering and integrated operations in the petroleum industry. Technical Report SINTEF A16331.

Hosseini, S., K. Barker, & J. Ramirez-Marquez (2016). A review of definitions and measures of system resilience. In *Reliability Engineering & System Safety*, Volume 145, pp. 47–61.

ISO-19501 (2005). Unified modeling language specification (version 1.4.2). ISO/IEC 19501:2005(E), ISO.

ISO-31000 (2009). Risk management – principles and guidelines. ISO 31000:2009, ISO.

ISO-31010 (2009). Risk management – risk assessment techniques. ISO/IEC 31010:2009, ISO.

ISO/IEC-27002 (2005). Information technology code of practice for information security management. ISO/IEC 27002:2005(E), ISO/IEO.

Lay, E., M. Branlat, & Z. Woods (2015). A practitioner's experiences operationalizing resilience engineering. In *Reliability Engineering & System Safety*, pp. 63–73.

Leksin, A., U. Barth, & R. Mock (2018). The Kursk submarine disaster in view of resilience assessment (in print). In *Proc. of European Safety and Reliability Conference (ESREL 2018*), London. Taylor & Francis Group.

Lundberg, J. & B. J. Johansson (2015). Systemic resilience model. Volume 141, pp. 22–32.

Mock, R., B. Truninger, P. Brunner, G. Pociupa, & T. Hruz (2015). It risk audit tool to enhance IT risk assessments. In *Proc. of European Safety and Reliability Conference (ESREL 2015)*, pp. 4029–4036.

Mock, R. & C. Zipper (2017). Embedding resilience assessment into risk management. In *Proc. of European Safety and Reliability Conference (ESREL 2017)*, pp. 1009–1014.

NIAC (2009, Sept. 8). Critical infrastructure resilience: Final report and recommendations. Technical report, National Infrastructure Advisory Council (NIAC).

Scholz, R.W., Y.B. Blumer, & F.S. Brand (2012). Risk, vulnerability, robustness, and resilience from a decisiontheoretic perspective. In *J. of Risk Research*, Volume 15, pp. 313–330.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Enhancing metro system resilience after signaling perturbations by bus bridging service: The case of Beijing

Q. Wei, R. Niu, T. Tang & S. Su
State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Haidian District, Beijing, P.R. China

L. Yue
Communication and Signaling Branch Company, Affiliated with Beijing Mass Transit Railway Operation Corp. Ltd., Haidian District, Beijing, P.R. China

ABSTRACT: Signaling system is a typical safety critical system aiming to enhance the safety and efficiency of train operation. Any signaling failure or abnormity will force a fallback to the safe side (stop), causing the drop in driving efficiency. Therefore, a fast and efficient way to replace the metro system after signaling failure is of great significance for passengers. Considering the road traffic conditions, a method for generating emergency metro-bus bridging plan is presented to improve the metro system resilience by assessing the satisfied passenger travel demands. The plan is implemented by generating the bus bridging routes based on the constructed metro-bus network and allocating the limited bus resources to the generated routes optimally. The approach is applied to Beijing Metro Line 5 coping with high signaling failure probability and the simulation results show that the system resilience is significantly enhanced about 21%–43% with metro-bus bridging service.

1 INTRODUCTION

Nowadays, the operation of metro system relies more and more on signaling system with the increasing of the degree of automation. Any disruption in signaling system may cause the trains and passengers to be late, even more, resulting in the chaos of public traffic. In August 18, 2016, the Beijing Metro Line 1 was suspended for more than 2 hours during evening peak hours due to a 3 min interruption of signal transmission network. Coupled with the bad weather, thousands of people were trapped in the downtown area. Therefore, compared with the frequency of failures, metro operators are more concerned about the impact of signaling failures on train headways and passenger transport. However, RAM (Reliability, Availability, Maintainability), the commonly used performance indexes, are always assessed by the average data with a period time, and cannot clearly show the reduction and recovery of train passing efficiency. Therefore, the concept of resilience is introduced to measure the impact of failure to metro system.

The first systematic definition of resilience is represented by Holling (1973) in ecological system almost 40 years ago. He thinks "resilience determines the persistence of relationships within a system and is a measure of the ability of these systems to absorb changes of state variables, driving variables, and parameters, and still persist." Since that time, researchers covering numerous fields have gradually realized the advantages of resilience in system design and management. For social domain, US Department of Homeland Security (2006) tend to view social resilience in terms of the ability of system or asset to maintain its function or recover from attack or incident. When in economic system, Rose & Liao (2005) characterize dynamic resilience as the speed of which an entity or system recovers from a severe shock to achieve a desired state. In railway system, Adjetey-Bahun et al. (2016) state that the concept of resilience has been introduced to measure both the ability to absorb perturbations and rapidity recover from perturbations. Throughout all the definitions, all of them are emphasizing the ability of system to absorb the perturbation and adapt to it changes, as well as recover from it quickly. The most quoted concept of resilience is from Bureau et al. (2003), they state that the seismic resilience of a system can be achieved by reducing failure probabilities, reducing consequences from failures and reducing the time to recovery. Furthermore, it can be defined as four characteristics: robustness, redundancy, resourcefulness, and rapidity and a broad measurement of resilience capturing these features is proposed as a resilience triangle model.

Combining the characteristics of metro system, the resilience of metro system can be introduced as: the ability of system to reduce the probability of failure, adapt to the impact of failure and recovery

from the failure. In this point, Vugrin et al. (2010) define three capacities to quantify and design for a better resilience of system, these capacities are:

- Absorptive capacity:
  It mainly reflected before or the time perturbation intervention, when the system can automatically withstand the perturbation and minimize the consequence of failure, it's an endogenous feature of system. In metro system, it can be reflected as reducing the probability of system to failure through the reliability design. For example, the redundancy of the onboard wireless unit to make the system immediately switch to the rear one when the front unit failures, so that the train can normally receive the Movement Authority with no impact on the train operation.
- Adaptive capacity:
  During the disruption and recovery period, through ingenuity or extra efforts to deal with the impact of disruption, is a set of actions of self-organization to response to the perturbation. One example is during the signaling perturbations, the operation order of the metro system relies upon the command of the dispatcher in the degraded model.
- Restorative capacity:
  The ability of system to quickly recover back to its original or expectation state by maintenance management and the ability of system to be repaired easily. For example, training and testing of the maintenance plans of general failures to improve the human maintenance ability so that the system can be recovered quickly.

As quick and efficient substitution of metro service is necessary for accommodating metro passengers during signaling perturbations. Thus, an ad-hoc bus bridging service is set up after signaling perturbations to enhance the metro system resilience of adaptive capacity by transferring the passengers to the nearest or destination metro stations in the shortest time.

The reminder of the paper is organized as follows. Section 2 reviews the quantitative measurement methods of resilience as well as the strategies to enhance the resilience of system. Section 3 introduces the proposed metro-bus bridging emergency plan to enhance the system resilience. The plan is applied to Beijing Metro Line 5 in Section 4 and Section 5. And we end this paper with two conclusions in Section 6.

2 LITERATURE REVIEW

2.1 *Quantitative measurement of resilience*

The quantitative measurement of resilience proposed by Bureau et al. (2003) (shown in Figure 1.)

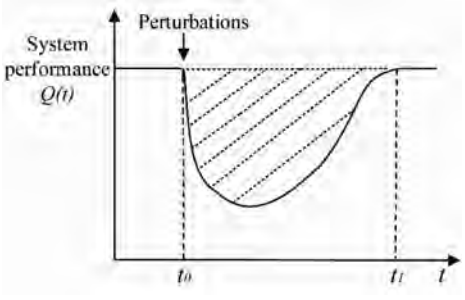


Figure 1. Resilience triangle model proposed by Bureau et al.

is extended and improved by many ways. Reed et al. (2009) propose the resilience of a system can be assessing by the ratio of the resilience triangle curve area to the time interval. Besides, Vugrin et al. (2013) proposed the system resilience can be assessed by resilience costs within systemic impact and total recovery effort. Where the systemic impact can be evaluated by the size of the resilience triangle curve, and the recovery effort is represented by the area of recovery effort curve.

Except for the size of resilience triangle, the important points of the triangle curve can also be used to measure the system resilience. Dorbritz (2011) puts forward the disaster resilience of transportation system can either be quantified by the resilience triangle area or three values: (1) the initial impact of the disaster; (2) the minimum value of system performance; (3) the time system get recovery. Nan et al. (2016) present an integrated metric for system resilience quantification, the metric combines three resilience capabilities in four phases based on resilience curve and is dimensionless and useful to compare different system resilience.

In addition to the traditional numerical methods used above, a stochastic method can also be used for assessing the resilience of system. Based on the resilience triangle, Chang & Shinozuka (2004) developed a probabilistic method for assessing system resilience. Where the resilience is defined as the probability of the system performance after the disruption lesser than the accepted performance. The proposed framework is applicable to both the infrastructure system and community, but the acceptable standard of the system is difficult to standard. Ouyang et al. (2012) introduce a time-dependent expected resilience metric as the mean ratio of the area between the real and target performance curve. The resilience metric is a stochastic one as the perturbation is modeled as Poisson distribution process and the resilience of system can incorporate one or multiple related hazards.

The above general resilience evaluation methods are concentrated on the resilience triangle model,

besides, the mathematical optimization models are also proved to be effective for assessing the resilience of system based on the network topologies structure of the system with one or more objective. As the scale-free graph nature of the system makes it easy to quantify the system resilience, especially for transportation system. Ip & Wang (2011) represent the transportation networks by an undirected graph with nodes as cities and edges as traffic roads, and the resilience of the network can be assessed by the feasible links between each node after disruption. Follow that, a multi-objective optimization model is proposed to evaluate the efficient edges and nodes to improve the network resilience. Beyond that, fuzzy logic is also used to quantify the resilience of system when involves several variables which are all important to the system.

2.2 *Strategies to enhance resilience in transportation system*

The definition and value of the system resilience can become useful and meaningful when used to devise effective resilience strategies for the system of interest. In transportation system, two ways are summarized to enhance the system resilience. On the one hand, it can be enhanced based on the definition of resilience expounded as the three capacities. As resilience is the embodiment of a variety of capacities, enhancing a specific capability contributes to enhance the resilience of system. Nan et al. (2016) test the strategies in an electric power supply system, where the resilience enhancing strategies can be interpreted as: (1) increasing the capacity of the battery to improve the absorptive capability during the disruption phase; (2) the improvement of human operators' ability to enhance the adaptive capability during the recovery phase; (3) the improvement of the efficiency of line operation to enhance the restorability capacity during the recovery phase.

In this point, organizational and management plans can be implemented to enable the system quick recovery and minimum the impact of disruptions. Adjetey-Bahun et al. (2016) propose a crisis management plan addressing the system capacity in order to assess the extent to which they increase the resilience of mass railway transportation system during perturbations, i.e. setting up temporary train services on part of the impacted line during perturbations can enhance the resilience of system effectively and decreasing the repair time of fault equipment contributes to the restorative capacity.

The same way also can be found in Chan et al. (2016) when assessing transportation system resilience with weather disruptions. Three common strategies are defined and offered for making transportation system more resilient: (1) hardening: building levees and floodwalls to prevent floodwaters; (2) redundancy: power distribution is a redundant system that provides a backup path between points; (3) elasticity: holding the aircraft on the ground to protect the fleet and passengers and taking off again quickly after the storm. Three aspects of resilience can work in combination to enhance the transit system resilience and support system management decisions.

On the other hand, pre- and post-disruptions actions are also impactful for system resilience improvement. After metro system disruption, an integrated local bus services and metro system to enhance metro network resilience is introduced by Jin et al. (2014), where a two-stage stochastic process model is presented to evaluate Singapore public transit network resilience and to optimize bus service routes that run in parallel with the metro lines. For more comprehensive, Faturechi & Miller-Hooks (2014) clearly address the pre-event mitigation and preparedness and post-event response in the disaster management life cycle are three decision process to the system resilience and a three-stage program is proposed to quantifying and optimizing the roadway network resilience.

3 METHODOLOGY

As the metropolises with large population and complex road conditions, the road congest conditions and the metro station exports are considered in the metro-bus bridging methodology to allow the passengers to reach their destinations in an available and quick way. In order to enhance the metro system resilience after signaling perturbations, the proposed bus bridging plan can be divided into three steps: (1) to the failure metro station as the endpoint, combined with the actual topology of the roadway to construct the metro-bus network; (2) generating the bus bridging routes between each O-D pair to avoid the congestion sections with minimum travel time; (3) allocating the limited bus resources to the bridging routes to increase the number of passengers demands that can be satisfied.

3.1 *Metro-bus network*

The metro-bus bridging network is modelled by a directed graph $G = (V, A)$, illustrated in Figure 2. In order to avoid passengers get stranded at the metro station during signaling perturbation, not only the disruption metro stations, the other metro stations which directly connected with the disruption metro stations are also considered to be bridged. Where set V is the set of nodes in the network, including metro station nodes and roadway nodes. The metro station nodes are the bus bridging stations and the roadway nodes are the end of roadways modelled according to the actual road network. Set A is the set of arcs in graph G, which is the connection of roadway

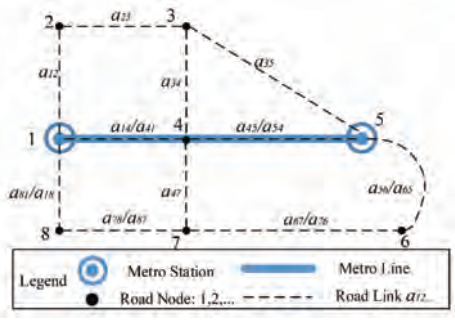


Figure 2. Metro-bus bridging topology network.

Table 1. Classification of roadway congest conditions in Beijing.

Congest conditions	h_{ij}
Smooth	0
Basically smooth	0.2–0.5
Mild congestion	0.5–0.8
Moderate congestion	0.8–1.1
Serious congestion	>1.1

node with metro node, or roadway node, labelled as $a_{ij}(i,j \in V)$. Based on the metro-bus network, the bus-bridging route can be modelled as the link of several arcs between origin nodes and destination nodes.

3.2 *Generation of bus bridging route*

Go through all the alternative routes, the shortest path may not be able to meet the constraints of actual demands, like route length constraints and road capacity limit. Thus, in the plan, Yen's (1971) *k*-shortest path algorithm is used to explore the feasible bus bridging routes under certain constraints.

Different from other research, the real-time traffic conditions $h_{ij}(i,j \in V)$ is considered in the plan, known as Traffic Performance Index (TPI) in Beijing, means the average travel time on a_{ij} is more than h_{ij} times as much as usual under congest conditions, for details shown in Table 1.

According to the real-time traffic conditions h_{ij}, the travel time $t_{ij}(i,j \in V)$ on each a_{ij} is calculated as:

$$t_{ij} = \frac{l_{ij}}{v} \cdot (1 + h_{ij}) \tag{1}$$

where v is the average travel speed of the bus, $l_{ij}(i,j \in V)$ is the length of a_{ij}.

The route constraints in the plan include two components: the minimum route travel time T_{min} and the maximum route travel time T_{max}. The final feasible generation route $b \in B$ between each O-D can be formulated as $f(w, k)$:

$$b: f(w,k) = min\{t_w x_w^k\} \tag{2}$$

subject to:

$$T_{min} \le t_w \le T_{max} (w \in W) \tag{3}$$

$$x_w^k = \{0,1\} \ (w \in W) \tag{4}$$

objective function (2) minimize the travel time of each O-D route, where route $w \in W$ is defined as all the available routes between each O-D and t_w is the travel time of route w. Constraint (3) is the travel time constraints for each route . In constraint (4), x^k_w is a binary decision, if $x^k_w = 1$, means the route w belongs the k-shortest path. Table 2. describes the process of generating the bus bridging route $b \in B$.

3.3 *Bus resource allocation*

According to Reed et al. (2009), the system resilience can be quantitative as (shown in Figure 3.):

$$R = \frac{\int_{t_0}^{t_f} Q(t)dt}{t_f - t_0} \tag{5}$$

Table 2. Steps of generating the bus bridging routes.

Input: $G = (V,A)$, T_{min}, T_{max}, k.
Output: $b \in B$.

Step 1	All the alternative routes are generated using k-shortest path algorithm between each O-D.
Step 2	Check all the generated routes for the minimum and maximum route constraints, if the route satisfied, then the route is accepted as a candidate route. Otherwise, it is removed.
Step 3	For all candidate routes, find the shortest one for each O-D, and kept in the set B as the final generated bus bridging routes.
Step 4	Output the set of B.

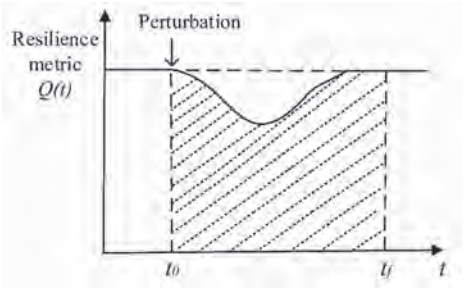


Figure 3. Resilience quantitative model proposed by Reed et al. (2009).

where R is the system resilience, can be calculated as an integral of the normalized resilience metric $Q(t)$ within an interval time, t_0 is the time perturbation intervention and t_f is the fixed time window. As the time interval $(t_f - t_0)$ can be defined flexibly, in the metro system, it can be varied with different types of signaling failure or defined as a fixed value. When resilience metric $Q(t)$ is a discrete value, the system resilience R can be defined as:

$$R = \frac{\sum_{i=1}^{i=N} Q(i \cdot \Delta T) \cdot \Delta T}{N \cdot \Delta T} \tag{6}$$

let N be a number such that:

$$\left(t_f - t_0\right) = N \cdot \Delta T \tag{7}$$

Mathematically, according to Chen & Miller-Hooks (2012), the metro system resilience metric $Q(t)$ can be expressed as the passenger travel demands between each Origin-Destination (O-D) pair that can be satisfied during the signaling perturbation, which known as Travel Demand Satisfaction Rate (*TDSR*) in the following.

Since the Genetic Algorithm (GA) provides a robust search as well as a near optimal solution in a reasonable time, the approach is employed to obtain an optimal system resilience accomplishing by allocating the limited bus resources to the generated bus bridging route $b \in B$ and the resilience metric (*TDSR*) of metro system is proposed as:

$$f(TDSR) = max\left\{\frac{\sum_{b \in B} n_b \times C}{\sum_{b \in B} d_b}\right\} \tag{8}$$

subject to:

$$0 \le n_b \le N (b \in B, n_b \text{ is integer}) \tag{9}$$

$$\sum_{b \in B} n_b = N \tag{10}$$

objective function (8) is to maximum the total number of passenger Travel Demands Satisfied Rate (*TDSR*), where d_b is the passenger demands for each O-D route $b \in B$ and obtained with fixed intervals Δt. An integer variable n_b indicates the number of buses allocated to each O-D route $b \in B$. C is the capacity of the bus and N is the total number of buses that available in the bus depots.

Hence the GA chromosome consists of integer gene values represents one solution of the bus allocation and is as follow:

$$[n_{b1}, n_{b2}, \ldots, n_{bm}] \tag{11}$$

where m is the total number of generated routes from set B. The objective of GA model presented here is to scientifically guide the bus resources allocation and select an optimum solution with the satisfied passenger travel demands being maximum. The implementation of the GA is shown in Table 3.

Table 3. The process of bus resources allocation.

Input:	$G = (V, A)$, d_b, B, N, C, m, Genetic Algorithm setting parameters.
Output:	$f(TDSR)$, $[n_{b1}, n_{b2,\ldots,} n_{bm}]$
Step 1	Input all the GA related parameters.
Step 2	Generate the GA population formulation for the current O-D pair set size m and initialize each chromosome randomly.
Step 3	Set generation = 1.
Step 4	According to each O-D pair passenger demands d_b, evaluate each chromosome by the objective function.
Step 5	Keep the current solution.
Step 6	Generate next generation. • Rank. • Selection. • Crossover. • Mutation.
Step 7	If the generation ≤ Max_Generation, increased generation by 1 and go to step 4; else update the current best solution if improved.
Step 8	Output the best bus-allocation solution from the best solution found, and its performance $f(TDSR)$.

4 CASE STUDY

According to the actual signaling system fault records, typical trackside signaling perturbations are summarized and the impact of different conditions to system resilience is discussed.

4.1 *Background of case study*

The model is applied to generate metro-bus bridging plan for Beijing Metro Line 5, whose average number of passengers on weekends can reach one million. The signaling failure data is provided by Communication and Signaling Branch Company Affiliated with Beijing Mass Transit Railway Operation Corp. Ltd. According to the statistics of the failure record in the past four years, the main trackside signaling equipment with high failure probability are summarized as: (1) WESTRACE TCOM system used to produce carrier frequency of track circuit; (2) track circuit transmitter or receiver; (3) the trackside APR used to send position information to the trains. Once the equipment fails, signal or multiple track sections will be affected and impact will last one or two hours.

Through the analysis of historical fault records, the failure of TCOM system has the most influence in the trackside equipment as the TCOM system failure will affect multiple sections for few hours. In this way, four typical areas with high failure rate of TCOM system are enumerated (shown in Figure 4.) and the bus bridging area is tabled in Table 4.

Due to the large passenger flow in Line 5, great disturbance to passenger travels will be caused when the signaling failure happens. Therefore, we assume the operators will call for bus-bridging if the perturbation lasts ten minutes. The parameters

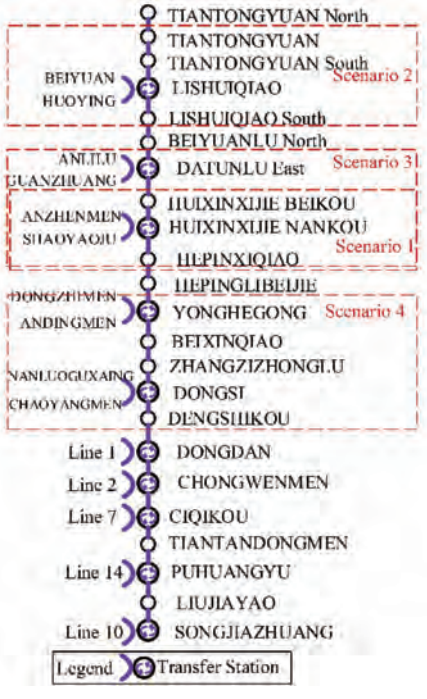


Figure 4. Typical areas with high failure rate of TCOM system in Beijing Metro Line 5.

Table 4. Four typical bus bridging scenarios.

Scenario NO.	Bus bridging areas
Scenario 1.	Huixinxijie Beikou-Huixinxijie Nankou-Hepingxiqiao-Anzhenmen
Scenario 2.	Tiantongyuan-Tiantongyuan South-Lishuiqiao-Lishuiqiao South-Huoying
Scenario 3.	Datunlu East-Huixinxijie Beikou-Huixinxijie Nankou-Hepingxiqiao-Anlilu-Anzhenmen
Scenario 4.	Yonghegong Lama Temple-Beixinqiao-Zhangzizhonglu-Dongsi-Dengshikou-Nanluoguxiang

related to the proposed bus-bridging plan are set as follow:

– The perturbation occurs at 7:00 am and lasts 2 hours.
– The average bus speed of bus: $v = 18$ km/h.
– Feasible bus resources: $N = 15$ during each time interval, $C = 80$ person/bus.
– Time constraints: 1 min–20 min.

4.2 *Modeling and bus-bridging strategy*

The path network models between metro stations are built based on google map. And then, shortest bridging route are selected with the improved k-shortest path algorithm. For example, in Scenario 1, set $k = 4$. Under the constraints of time conditions and passenger travel demands, there are 9 bus bridging routes are generated (shown in Figure 5.). And the bus bridging plan avoids the congestion section {*v7-v12-v15, v27-v28*} from Huixinxijie Beikou to Huixinxijie Nankou. But from Huixinxijie Nankou to Hepingxiqiao, the congestion section {*v15-v18-v27*} is still in the plan, as the other way cost too much travel time.

The travel demands for each O-D route are collected from the AFC (Automatic Fair Collection system) during 7 am to 9 am in the workday. The available bus resource in each time interval is 15. The *TDSR* of Scenario 1 with bus bridging plan is shown in Figure 6. And the resilience of the metro system is 0.80, improved by 29% than normal condition.

Figure 5. The bus bridging routes in Scenario 1.

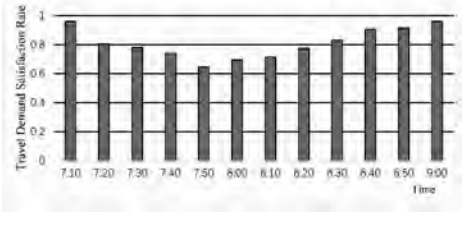


Figure 6. *TDSR* with bus bridging plan in Scenario 1.

5 DISCUSSION

5.1 *The influence of bus resources on bus bridging plan*

The system resilience metric (TDSR) can be calculated by the O-D demands that the bus bridging plan can satisfy in perturbation scenarios and shown in Figure 7. At the beginning of the implementation of the plan, as a large number of passengers stayed at the station, the TDSR presents a trend of decreasing due to the start of the failure. About half an hour later, the passengers are gradually evacuated with the bus bridging plan, and the resilience metric of system begins to recover. However, the time with the lowest TDSR is varied with the passenger travel demands delay at each metro station. In Scenario 2, the system is less resilience at the outset due to the greatest passenger demands at the time perturbation occurs. And in Scenario 3, the TDSR of the system has changed little because of the stable passenger demands during the perturbation.

According to Formula (2), the system resilience in the four scenarios with or without the bus bridging plan are shown in Table 5. It can be seen that with the proposed bus bridging plan the system resilience can be increased by 21%–43%.

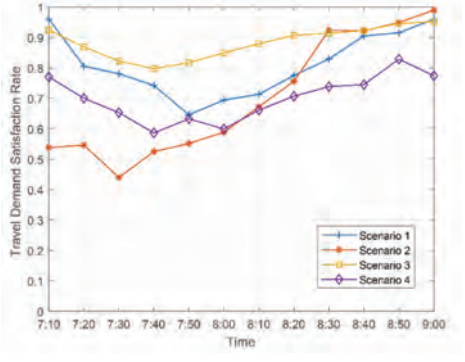


Figure 7. TDSR in different scenarios with bus bridging plan.

Table 5. The comparison of the metro resilience before and after the implementation of the bus bridge scheme.

Scenario NO.	System resilience (R1)	System resilience with bus bridging (R2)	Improvement (R2-R1/R1)
Scenario 1.	0.62	0.80	0.29
Scenario 2.	0.58	0.70	0.21
Scenario 3.	0.70	0.88	0.26
Scenario 4.	0.49	0.70	0.43

5.2 *The influence of bus resources*

In order to minimize the impact on normal bus transit system, the buses that can be used to the metro-bus bridging plan are limited. As the amount of buses will directly affect the improvement of system resilience, it is necessary to weight the bus transit system with the metro system resilience enhancement. Figure 8 shows the *TDSR* of system with different bus resources in Scenario 2. When the bus resources get 25 each interval time, the system resilience can get 0.91, up to 30% compared to M = 15. And from 7 am to 8 am, about 25 buses are needed at each time interval so that the *TDSR* of system can reach 0.8, and from 8 am to 8:30 am 20 buses are required, and from 8:30 am to 9 am, 10 or 15 buses are enough to make the *TDSR* to 0.8. It is provided a reference for the metro operator and the bus transit operator to design the bus bridging emergency plan adjusting the number of buses in different time periods to meet the needs of passenger demands without affecting the bus transit system.

5.3 *The influence of tidal flow*

For Beijing Metro Line 5, its northern end to Tiantongyuan, an important office worker residence, the flow of people in working days during morning and evening peak has obvious tidal trend (shown in Figure 9.). Therefore, in the bus bridging plan, the bus routes on the up and down lines will be reduced to one-side route according to the direction of passenger flow in order to support the specific passenger demands. In Scenario 2, before 8:30 am, the *TDSR* is about one third higher than that of the two-sides bus bridging and the *TDSR* of system are above 0.7 (shown in Figure 8.). There is an obvious improvement of system resilience as the bus bridging plan only considers one-side (uplink) passenger flow and it can be improved about 21%.

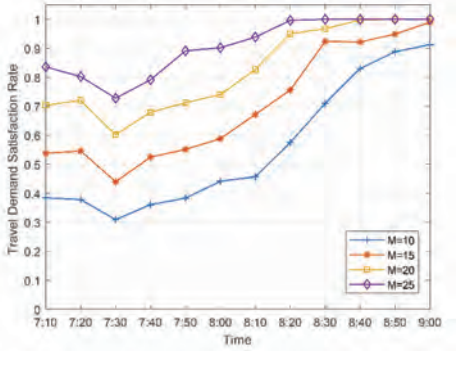


Figure 8. *TDSR* in different amount of buses over time in Scenario 2.

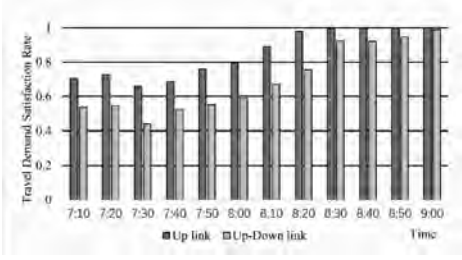


Figure 9. *TDSR* in uplink and up-down link bridging routes in Scenario 2.

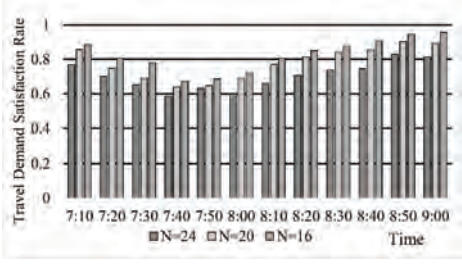


Figure 10. *TDSR* improvement from different bridging routes in Scenario 4.

Also, in Scenario 4, when removing part of the bridging routes that the system *TDSR* also get improvement (shown in Figure 10.). When the total bus bridging routes are 24, the system resilience is 0.7, when the total number of bridging routes streamlined to 16, the system resilience can be 0.83 improved by 19%. For general scenarios, with the limited bus resources, the metro operator can adjust the bus bridging routes appropriately according to the distribution of passenger flow in each time period and arrange the limited buses to the stations with most demands so as to meet the bridging routes with more demands and prevent the accumulation of stranded passengers in stations.

6 CONCLUSION

The resilience of metro system is mainly manifested as absorptive capacity, adaptive capacity and recovery capacity, and we focus on the adaptive capacity which can be enhanced with external bus bridging services after signaling failure. According to the above, the main conclusions are as follow:

1. In metropolis, urban rail transit system is the most important way for people's traveling, once the signaling system failures, a large number of passengers will be accumulated at the station in a short time during the peak period. With the bus bridging plan, the metro system resilience can be improvement about 21%–43% compared to the normal situations. And the bus bridging plan can be used to slow down the pressure caused by the passengers. The results show the passenger flow will return to normal state about an hour later without other actions taken by the metro operators. Therefore, we suggest that during the perturbations, except for the bus bridging plan, operators shall inform the passengers about the signaling failures with network or other media in time so as to avoid excessive passenger stranded at stations.
2. The trend of morning and evening tides of passenger flow is obvious in Line 5, the generated bus bridging routes can be properly adjusted to the change of passenger flows and the priority of the routes with more demands to enhance the metro system resilience. The results provide reference for the metro operator in daily disposal after signaling failure.

Further improvements of the metro-bus bridging plan, we are interested in the exits of the bridging metro stations, as the wide distribution of the metro exits will influence the bus bridging routes especially in transfer stations. In addition, further work can be carried out on the commands of the dispatcher during the perturbations, optimizing the strategy of dispatcher to restore the normal operation order of the train as soon as possible to enhance the resilience of the metro system.

ACKNOWLEDGEMENTS

This work was supported by the Fundamental Research Funds for the Central Universities under 2015 JBZ006, the Project U1434209 supported by National Natural Science Foundation of China.

REFERENCES

Adjetey-Bahun, K. et al. (2016). A model to quantify the resilience of mass railway transportation systems, *Reliability Engineering & System Safety*, 153, 1–14.

Bruneau, M. et al. (2003). A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthquake Spectra*, 19(4), 733–752.

Cen, N. & Sansavini, G. (2016). A quantitative method for assessing resilience of interdependent infrastructures. *Reliability Engineering & System Safety,* 157, 35–53.

Chan, R. & Schofer, J.L. (2016). Measuring transportation system resilience: response of rail transit to weather disruptions. *Natural Hazards Review,* 17(1), 05015004.

Chang, S.E. & Shinozuka, M. (2004). Measuring improvements in the disaster resilience of communities. *Earthquake Spectra,* 20(3), 739–755.

Chen, L. & Miller-Hooks, E. (2012). Resilience: an indicator of recovery capability in intermodal freight transport. *Transportation Science,* 46 (1), 109–123.

Dorbritz, R. (2011). Assessing the resilience of transportation systems in case of large-scale disastrous events.

Fan, W. & Machemehl, R.B. (2015). Optimal transit route network design problem with variable transit demand: genetic algorithm approach. *Journal of Transportation Engineering,* 132(1), 40–51.

Faturechi, R. & Miller-Hooks, E. (2014). Travel time resilience of roadway networks under disaster. *Transportation Research Part B Methodological,* 70(70), 47–64.

Holling, C.S. (1973). Resilience and Stability of Ecological Systems. *Annual Review of Ecology & Systematics,* 4(4):1–23.

Ip, W.H. & Wang, D. (2011). Resilience and friability of transportation networks: evaluation, analysis and optimization. *IEEE Systems Journal,* 5(2), 189–198.

Jin, J.G. et al. (2014). Enhancing metro network resilience via localized integration with bus services. *Transportation Research Part E Logistics & Transportation Review,* 63(2), 17–30.

Kepaptsoglou, K. & Karlaftis, M.G. (2009). The bus bridging problem in metro operations: conceptual framework, models and algorithms. *Public Transport,* 1(4), 275–297.

Ouyang, M. et al. (2012). A three-stage resilience analysis framework for urban infrastructure systems. *Structural Safety,* 36–37(2), 23–31.

Reed, D.A. et al. (2009). Methodology for Assessing the Resilience of Networked Infrastructure. *IEEE Systems Journal,* 3(2), 174–180.

Rose, A. & Liao, S. (2005). Modeling Regional Economic Resilience to Disasters: A Computable General Equilibrium Analysis of Water Service Disruptions. *Journal of Regional Science,* 45(1):75–112.

Vugrin, E.D. & Turgeon, J. (2013). Advancing cyber resilience analysis with performance-based metrics from infrastructure assessments. *International Journal of Secure Software Engineering,* 4(1), 75–96.

Vugrin, E.D. et al. (2010). A Framework for Assessing the Resilience of Infrastructure and Economic Systems. *Sustainable and Resilient Critical Infrastructure Systems.* Springer Berlin Heidelberg.

Yen, Jin Y. (1971). Finding the k shortest loopless paths in a network. *Management Science,* 17(11), 712–716.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

ISRA: IMPROVER societal resilience analysis for critical infrastructure

H. Rosenqvist
Danish Institute of Fire and Security Technology, Hvidovre, Denmark

N.K. Reitan
RISE Fire Research, Trondheim, Norway

L. Petersen
EMSC, Arpajon, France

D. Lange
RISE Research Institutes of Sweden, Borås, Sweden

ABSTRACT: Resilience of Critical Infrastructure (CI) has been a research focus for several years now, with efforts being made to develop methods for the analysis and assessment of CI resilience. However, these efforts are often carried out without consideration of enriching societal risk or resilience assessments with knowledge of the resilience of CI. Bearing in mind that the definition of CI according to the EU reflects the fact that it exists to deliver vital societal functions, the consideration of its resilience in isolation of the community it serves is only addressing part of the problem. The Horizon 2020 project IMPROVER has already developed methodologies for assessing and managing CI resilience. This paper proposes an evolution of the management framework for CI resilience which enriches societal resilience assessment with knowledge of the CI resilience. The framework and societal resilience analysis methodology are both described along with an application of the analysis method.

1 INTRODUCTION

In recent years, increasing resilience of Critical Infrastructures (CI) has been a major objective of the EU as well as worldwide. Disruptions of critical infrastructures, caused by natural or man-made events, are increasingly affecting today's society as the reliance on infrastructure systems to provide vital societal functions increases (Rinaldi et al. 2001). As one of the main purposes of CI is to deliver services to the society, resilience assessment of CI should be closely linked to the study of societal resilience. Societal resilience refers to "the ability of social groups or communities to cope with external stresses and disturbances as a result of social, political and environmental change" (e.g. Adger 2000, Folke 2006, Furedi 2007, Marshall 2010, Voss 2008). As being able to tolerate a reduction in the quality, quantity, or availability of a service provided by a CI demonstrates coping capacity, it can therefore be seen as a key component of societal resilience. Thus, CI resilience and societal resilience are closely linked, with each influencing the other. Indeed, many existing societal resilience analysis methodologies include indicators relating to critical infrastructure (e.g. Michel-Kerjan 2015, Boon et al. 2012, Cutter et al. 2010, Renschler et al. 2010). As such, increasing CI resilience also serves to increase societal resilience, creating a positive feedback loop between the two. The societal domain mainly becomes of interest when considering a higher level; regional or national, where several CI operate.

To that end, the EU Horizon 2020 project IMPROVER (Improved risk evaluation and implementation of resilience concepts to critical infrastructure) has developed the IMPROVER Societal REsilience Framework (IS-REF), which maps resilience management onto common risk management frameworks and includes aggregated CI resilience assessments as an input to the societal resilience analysis (Fig. 1). The IS-REF is an evolution of ICI-REF (IMPROVER Critical Infrastructure Resilience Framework), which is developed for managing CI resilience, and has a similar structure.

The initial step in IS-REF is to establish the context for societal resilience management. This includes the gathering of CI resilience assessments, which may have been conducted by using the ICI-REF framework with methodologies developed within IMPROVER, designed for CIs to manage

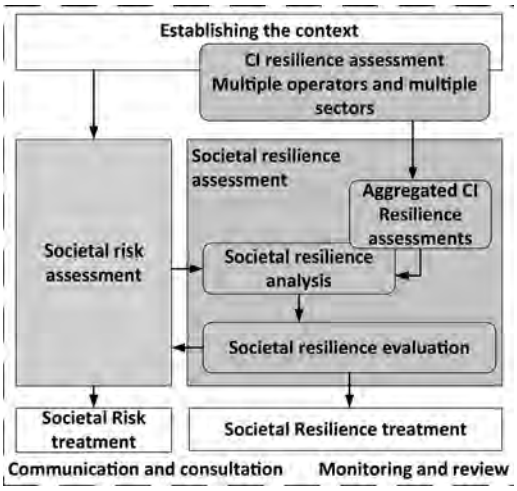


Figure 1. The structure and process of the IMPROVER Societal REsilience Framework (IS-REF).

Table 1. Success factors for IMPROVER's resilience management frameworks, with requirements to analysis methodologies.

System dimension	Success factor
Goal	Applicable to all types of CIs
Environment	Easy to use
	Effective and coherent crisis and disaster resilience management
	Provide relative resilience measurements
Structure	Supplements existing practice
	Taking into account public communication
Evolution	Arranged for being revised continuously
	Learning capabilities
	Willingness of utilisation

their technological and/or organisational resilience. In the next step of IS-REF, as a complement to societal risk assessment, a societal resilience assessment is undertaken. The results of such a resilience analysis shall be evaluated against previously decided upon criteria, and if necessary, a societal resilience treatment plan is devised. Throughout the entire IS-REF framework, communication and consultation as well as monitoring and review take place. For more details on the framework see IMPROVER's deliverable 5.1 (Lange et al. 2017).

Within both the ICI-REF and IS-REF, the resilience analysis can be performed using existing methodologies. However, IMPROVER has designed success factors for the performance of the resilience management frameworks and their content. Those of the success factors which generate requirements to resilience analysis methodologies are presented in Table 1. The success factors were designed to meet requirements from stakeholders and end-users, and have their basis in the Horizon2020 call text. With the background in systems theory (Roux-Rouquié & Moigne 2002), the success factors are categorised according to four fundamental system dimensions: goal, environment, structure and evolution, and are described in detail in IMPROVER's deliverable D6.1 (Reitan et al. 2017).

For the purpose of fulfilling the success factors, a well-defined methodology which is highly suitable for analysing societal resilience within the context of CI is currently being developed; the IMPROVER Societal Resilience Analysis (ISRA) methodology. This paper describes the background and structure of ISRA which is still in an early phase. Pilot tests, giving insight in how the methodology can be used, will aid the further development and optimisation of the methodology. A societal resilience analysis of one of the living labs in IMPROVER was conducted and is presented in this paper.

2 ANALYSING SOCIETAL RESILIENCE

Assessing and enhancing the resilience of critical infrastructures will not automatically result in a resilient society, since social and human dimensions have a strong influence in the achievement of a resilient society. Indeed, it is important to consider the link between physical and human systems to understand and enhance societal resilience (Chan et al. 2014). However, the concept of resilience is not yet fully operationalized, and many different approaches have been developed to achieve a measurement of resilience (e.g. Frankenberger et al. 2013, Boon et al. 2012, Cutter et al. 2010, Norris et al. 2008). While there are no generally agreed upon metrics, indicators can be an effective tool to help decision makers understand where their community stands in terms of resilience and as a base for developing plans and strategies to enhance resilience. Furthermore, for CI, a societal resilience analysis based on indicators provides a holistic picture of a community's strengths and weaknesses in times of disasters, offering an understanding about what kind of society the critical infrastructure operates in.

2.1 *Resilience capacities and dimensions*

In the field of societal resilience, the concept of coping capacity, adaptive capacity and transformative

capacity are common denominators that are used to categories capacities needed to achieve resilience (Keck & Sakdapolrak 2013). Coping capacity refers to the ability to respond, absorb and recover from a disruptive event and is generally related to a time frame close to the event. Adaptive capacity includes the ability to plan for and adjust to future challenges, which is related to a longer time-frame both before and after an event. Resilience is not only about quick recovery and adjusting to new circumstances; the aspect of transformation must also be taken into account. Transformative capacity refers to the ability to transform the stability landscape in order to create new, better, pathways for the system and is thus related to major changes in the long-term.

Resilience is by definition a complex and multi-dimensional topic, and a resilience assessment should ideally include all of these dimensions and their interdependencies (Sharifi & Yamagata 2016). However, to be able to operationalize the resilience concept, some trade-offs are necessary. From literature, six major societal resilience dimensions were identified as physical, social, human, natural, economic and institutional capital. Each of these dimensions has influence on the resilience capacities discussed above.

3 THE ISRA METHODOLOGY

3.1 *Overall structure*

The objective of the methodology is that the analysis should be able to inform a community on how to enhance coping, adaptive and transformative capacities. The starting position is to analyse a community's perceived capability to react, adapt and recover from a shock. Based on the resilience dimensions, a set of indicators has been identified to analyse societal resilience. Within ISRA, detailed individual assessments of CI resilience are used to provide an input to the physical dimension.

The indicators are further categorized by which capacity they mainly have influence on (Table 2). The indicators and their parent resilience dimensions can, in reality, influence more than one capacity, but in this first development phase, they are categorized according to the capacity they are assumed to have the most effect on. The proposed aim of using the methodology is to i) develop a common understanding of societal resilience, and ii) establish the current position (the result of the analysis).

However, it is important to keep in mind that ISRA is only a part in an overall resilience management strategy, the results of which need to be used in the resilience evaluation part of the framework to ensure sense making, before deciding how to go about improving resilience in the resilience treatment part of IS-REF. In this early development phase, the focus was to identify relevant indicators from literature and create a holistic picture of the societal resilience domain. Thus, at this stage, interdependencies or conflicting goals among the indicators were not taken into account.

3.2 *Aggregation of the indicators*

The assessment is performed by qualitatively scoring a set of indicators on a scale from 1 to 5. The indicators are categorized according to the six resilience dimensions at Level 2 of the ISRA structure (Fig. 2). The score for each resilience dimension is achieved by aggregating the indicators under each dimension and capacity to a single measurement, and thereafter aggregating the three capacity scores into one score for each resilience dimension. In this early development phase of ISRA, all indicators and subcategories are weighted equal, but one might want to weigh the importance of the capacities and dimensions in order to capture subjective resilience aspects. The indicators are aggregated by the weighted arithmetic mean (Eq. 1)

$$Level\ k\ indicator = \sum_{i=1}^{n} w_i x_i \quad (1)$$

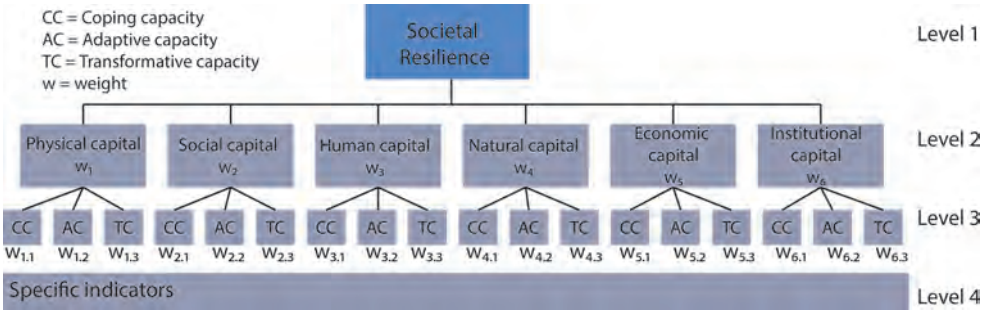


Figure 2. Structure of the ISRA methodology.

where n is the number of Level $k+1$ indicators, w_i the weighting coefficient for the individual Level $k+1$ indicator and x_i the scoring of the individual $k+1$ indicator. The aggregation can be done all the way up to Level 1 (see Fig. 2) resulting in a global score for societal resilience. However, a lot of information can be lost by doing that, so presenting the resilience dimensions in Level 2 provides a more detailed result.

4 PILOT DEMONSTRATION OF ISRA

4.1 *Case study*

The *Port of Oslo* has been one of the living labs in IMPROVER from the start of the project. The relevance of the Port of Oslo from a CI resilience perspective can be traced back to 2014, when the Norwegian Directorate for Civil Protection published an overall assessment of the management of the safety conditions in Sydhavna, and pinpointed improvement measures for the enterprises, the City of Oslo, the Port of Oslo and central government authorities.

In the wake of the report, the extensive exercise *HarbourEx15* was carried out at Sydhavna from 28 to 29 April 2015. The scenario for the exercise was an explosion and fire in containers with hazardous substances in the container area, fire in the fuel depot, evacuation of smoke-filled areas in parts of Oslo, and the grounding of a vessel and subsequent oil spill in the Inner Oslo Fjord.

The goal of the exercise was to improve emergency planning and rescue operations in the event of a major accident in Oslo. The exercise involved a total of ca. 40 participating organisations, including the rescue agencies, public authorities with responsibility for emergency planning, international rescue services and private actors at Sydhavna and in the City of Oslo.

In the evaluation of HarbourEx15, nine proposed initiatives were described, among these some of relevance to societal resilience, including population alert (acute alert) and providing information to the public (DSB 2016). Based on this work, a well-defined approach for societal resilience assessment is of interest to the area.

4.2 *The pilot test approach*

4.2.1 *The focus group*

In the development of ISRA, which is an operational methodology, users (authorities) were included at an early stage of the work to ensure that the methodology is fit-for-purpose. The focus group consisted of 5 relevant representatives from the HarbourEx15 evaluation. The participants were chosen to represent different parts of the community, as well as for their ability to provide insight and experience in areas connected to societal resilience. Their task was to evaluate the methodology from the perspective of the HarbourEx15 based on their expert knowledge about Sydhavna and the surroundings.

4.2.2 *Approach*

The exercise was in the form of a questionnaire. This was considered a sufficient method to combine a societal resilience analysis of the case study, and simultaneously receive the focus group's spontaneous evaluation of the indicators, as well as the quality of the questions that were designed to provide measures to the indicators. The questionnaires were designed as follows:

The focus group was asked to decide on a geographical area to perform the evaluation on. They were subsequent asked to evaluate the area according to the 56 societal resilience indicators. The indicators were evaluated according to a symmetric Likert scale from 1–5. To facilitate the self-evaluation, each indicator was described by a statement to which the respondents can specify their level of agreement from Strongly Disagree (corresponding to 1 on the scale) to Strongly Agree (corresponding to 5 on the scale). To support their evaluation, they were asked to provide evidence and support for their answers, along with a summary of the discussions that led to the response. Finally, they were asked to give their opinion about each indicator's relevancy and understandability. The feedback was then analysed using thematic analysis (Braun & Clarke 2006).

4.3 *Results*

In this section, general comments from the focus group and evaluations of the specific indicators are described.

4.3.1 *Comments and input*

Three main themes were identified from the comments and input regarding the chosen indicators and statements: lack of clarity, more than one question, and overlap.

For several of the indicator statements, focus group participants expressed the need for several clarifications. For example, one comment stated, "Unclear what it is you are looking for here." Another way to express lack of clarity was by asking directly for clearer definitions for the following words: Cooperation, Coordination, Community, Crowdsourcing, Community based, Actor, Stakeholder, Public, People in the community and Trust. Other times, there was uncertainty about which actor the methodology was asking about (e.g. "unclear if referring to decision making structures

Table 2. Societal resilience indicators according to their resilience dimension and supporting capacity, and scores from the pilot self-evaluation exercise. CC = Coping capacity, AC = Adaptive capacity, TC = Transformative capacity.

Resilience dimension	Indicator	CC	AC	TC	Reference	Self-eval. score
Physical capital	1.1 Preparedness	x			Pursiainen et al. (2016)	N/A
	1.2 Prevention	x			Pursiainen et al. (2016)	N/A
	1.3 Warning	x			Pursiainen et al. (2016)	N/A
	1.4 Response		x		Pursiainen et al. (2016)	N/A
	1.5 Risk assessment		x		Pursiainen et al. (2016)	N/A
	1.6 Recovery			x	Pursiainen et al. (2016)	N/A
	1.7 Learning			x	Pursiainen et al. (2016)	N/A
Social capital	2.1 Social welfare and family support	x			Sharma & Srivastava (2016)	4
	2.2 Isolation/decline in place attachment	x			Born (2014)	3
	2.3 Trust between citizens	x			Mayunga (2007)	4
	2.4 Attitudes towards sharing of resources	x			Coles & Buckle (2004)	4
	2.5 Perception of risk		x		Cabinet Office (2011)	2
	2.6 Participation in community organisations/projects	x			Burns et al. (2004)	3
	2.7 Minority groups in decision making structures		x		Magis (2010)	4
	2.8 Social cohesion			x	Poortinga (2012)	4
	2.9 Shared community values			x	Flora & Flora (2004)	3
	2.10 Perceptions of inclusiveness in decision making		x		Ahmed et al. (2004)	4
	2.11 Sense of identity in community		x		Paton & Johnston (2006)	4
	2.12 Engaging the public by using social technologies			x	Pursiainen et al. (2016)	4
	2.13 Information to public about their responsibilities in case of emergency/ disaster	x			Höppner et al. (2012)	4
	2.14 Exposure to media			x	Serafinelli et al. (2017)	5
	2.15 Interoperable communication among stakeholders		x		IMPROVER D2.2	5
	2.16 Information to public about hazards and risks			x	Höppner et al. (2012)	4
	2.17 Partnership between agencies, community groups and private enterprises		x		Evart & McLean (2017)	3
	2.18 Social network/good social infrastructure		x		Paton & Johnston (2006)	3
	2.19 Knowledge sharing by different stakeholder groups			x	ISO/DIS 22316:2017	5
Human capital	3.1 Diversity in resources and skills		x		Magis (2010)	4
	3.2 Health inequality	x			Chandra et al. (2011)	5
	3.3 Immunization coverage	x			WHO/EHA (1998)	5
	3.4 Water quality	x			Wilson (2012)	5
	3.6 Exercises and drills for disaster response		x		Collis et al. (2004)	3
	3.7 Attitudes towards change		x		Wilson (2012)	3
	3.8 Attitudes towards value of education			x	Cutter et al. (2010)	4
	3.9 School completion			x	UNDP (2014)	5
	3.10 Adoption of new technologies		x		Magis (2010)	5
	3.11 Free education			x	UNDP (2014)	5
	3.12 Experiences of disasters/emergencies			x	Wilson (2012)	3
	3.13 Knowledge about formal institutions, laws, legal frameworks and actors involved			x	Kuhlicke et al. (2011)	2

(*Continued*)

Table 2. (*Continued*).

Resilience dimension	Indicator	CC	AC	TC	Reference	Self-eval. score
Natural capital	4.1 Existence of green spaces			x	FEMA (2014)	5
	4.2 Communal resource management structures			x	Wilson (2012)	4
	4.3 Carbon footprint	x			Wilson (2012)	2
Economic capital	5.1 Economic inequality			x	Magis (2010)	3
	5.2 Economic resources available	x			Paton & Johnston (2006)	5
	5.3 Community support for maintenance of services		x		Magis (2010)	5
	5.4 Population covered by hazard insurance		x		Tierney (2007)	5
Insti-Tutional Capital	6.1 Disasters/emergency response plans		x		Chandra et al. (2011)	3
	6.2 Zoning ordinances for high hazard areas		x		Frankenberger et al. (2013)	5
	6.3 Land use and growth management plans		x		Frankenberger et al. (2013)	5
	6.4 Hazard mitigation and vulnerability assessments		x		Frankenberger et al. (2013)	5
	6.5 Disaster recovery plans			x	Frankenberger et al. (2013)	4
	6.6 Crowdsourcing platforms as decision support			x	Serafinelli et al. (2017)	2
	6.7 Usage of ICT for public awareness of disasters			x	Serafinelli et al. (2017)	2
	6.8 Public involvement in decision making and planning		x		Priest et al. (2016)	4
	6.9 Satisfaction with emergency managers			x	Frankenberger et al. (2013)	5
	6.10 Trust in local government and authorities	x			Paton & Johnston (2006)	3
	6.11 Satisfaction with local government			x	Frankenberger et al. (2013)	4
	6.12 Dialogue-oriented communication with the public			x	Kuhlicke et al. (2011)	4
	6.13 Emergency management procedures	x			Frankenberger et al. (2013)	4
	6.14 Early warning and contingency planning	x			Frankenberger et al. (2013)	4
	6.15 Building standards, codes and enforcement		x		Frankenberger et al. (2013)	5
	6.16 Transparency and accountability		x		Frankenberger et al. (2013)	5
	6.17 Regulatory mechanisms for use of pasture, water, agricultural lands and forest resources		x		Frankenberger et al. (2013)	5

in administrative decision-making systems or in civil society") or the scope of the question (e.g. "should we refer to the South Harbour and industrial duty, or should we refer to the community at large"). This theme appeared 18 times in the comments.

Another theme was that several indicator statements are actually comprised of multiple statements, showing up 11 times. This comment illustrates well the participants' views: "The statement contains two different elements/issues that can have different answers." There were also a few (five) comments with regards to certain indicators overlapping with one another. One such comment, "there is an overlap here with several of the other indicators," demonstrates well this theme.

Of special importance, though only mentioned once, was a comment related to the entire methodology instead of a single indicator or indicator statement. The participants asked, "How close should we attach the response to events in Oslo harbor?" This comment also shows a lack of clarity in how to use the method.

While the participants of the focus group had helpful critiques of the methodology, they also stated that while they found the exercise challenging, they also perceived it to be interesting and relevant to the assessment of societal resilience.

4.3.2 *Self-evaluation*

The preliminary scores from the pilot self-evaluation are shown in Table 2. Since there is no analysis made within IMPROVER project on the critical infrastructure in Sydhavna, there are no scores for the indicators under physical capital.

The focus group were also asked to provide evidence and discussion in regards to the self-evaluation score given for each indicator. However, evidence and discussion were provided for only 18 out of 56 indicators. Evidence and discussion provided ranged from very relevant and specific, for example for indicator 2.12 (see Table 2), "Emergency services and the municipality use social media, and have a large number of followers (c.f. HarbourEx15)", to relevant but lacking in empirical support, for example for indicator 2.3 Trust between citizens, "Based on the generally high level of trust in Norway, we answer that we agree," to suppositions, as for example indicator 6.7 Usage of ICT for public awareness of disasters, "we don't think so".

4.3.3 *Presentation of societal resilience*

The societal resilience of Sydhavna and its surrounding neighbourhoods is preliminary illustrated by the radar chart in Figure 3. All dimensions except physical capital are represented by aggregated scores from the self-evaluation, and the physical capital is purely fictively scored by the authors of this paper. Note that the radar chart does not represent absolute measurements of the societal resilience in Sydhavna, but is rather a way to show how the output from ISRA may be presented.

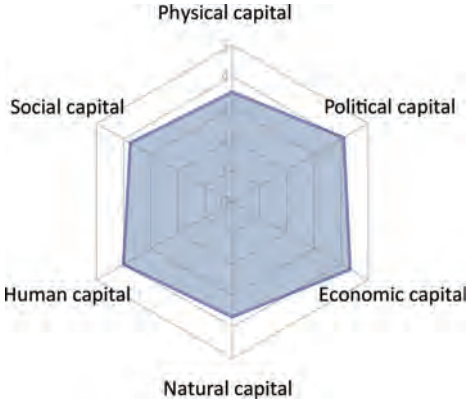


Figure 3. Presentation of the result from the pilot self-evaluation of societal resilience using ISRA.

4.3.4 *Sensitivity analysis of indicator weights*

At this early development phase, the indicators were weighted equally due to lack of information about relative importance of the indicators and dimensions. A simple sensitivity analysis was performed to study if, and how, different weightings would affect the results. The sensitivity analysis was performed by assigning the Level 3 indicators a weight of 0.5 one at a time, and assigning the remaining two Level 3 indicators a weight of 0.25. The resulting scores were compared to those obtained with equal weighting. The result from the sensitivity analysis is shown in Table 3.

5 DISCUSSION

5.1 *The outcome of the pilot test*

The methodology was perceived as valuable in terms of creating an overview of aspects that can affect coping, adapting and transforming capacity. However, the self-evaluation scores given to the indicators need to be supported by evidence or references. In the pilot test, only 18 out of 56 indicators were supported with evidence. This means that the output of the analysis should not be considered a realistic result in terms of maturity of societal resilience. In general, the focus group scored themselves high on the indicators and the resulting aggregated scores for each resilience dimension was around 4 on a scale from 1–5. This could mean that the studied area is mature regarding societal resilience, but since the evidence provided in the analysis were weak, strong conclusions should not be drawn from these results.

5.2 *Limitations of the study*

At this stage, the indicator hierarchy that ISRA is constructed of does not consider interdependencies

Table 3. Level 2 scores aggregated with equal weights, and w = 0.5 for each Level 3 indicator (capacity) by letting the rest of the weights remain equal (w = 0.25).

	Equal weights	Coping w = 0.5	Adaptive w = 0.5	Transform. w = 0.5
Physical capital	3.500	3.375	3.625	3.500
Social capital	3.802	3.768	3.744	3.893
Human capital	3.961	4.054	3.908	3.921
Natural capital	3.667	3.250	4.000	3.750
Economic capital	4.333	4.500	4.500	4.000
Institutional capital	3.944	3.875	4.125	3.833

and correlations between different resilience dimensions. However, as the socio-ecological system is a very complex system, there is a need to consider how the different parts interact and affect each other. Moreover, there is no consideration of conflicting goals among the indicators or if there are synergies that could result in emerging properties.

These limitations need to be explored, and the effects need to be evaluated, to further improve ISRA and produce valid results from the analysis.

5.3 *Going forward: Ways to improve the method*

The sensitivity analysis showed that weights can affect the end result, although neither a statistical sensitivity analysis was made, nor an analysis of weighting on each indicator. To further develop ISRA there is a need to investigate if and how to assign weights to the indicators.

One limitation of the indicator hierarchy that ISRA is constructed of is that it does not consider interdependencies and correlations between different resilience dimensions. As the socio-ecological system is a very complex system, there is a need to consider how the different parts interact and affect each other. This would increase the validity of the analysis.

The thematic analysis of the participants' comments identified three main areas for improving ISRA indicators. These are the need to clarify not only the definitions of words but also the scope and actors associated with the statements, ensure that each statement is only presenting the respondent with one question, and that the statements do not overlap with other indicator statements. As such, for future iterations of the methodology, we propose to add in a definitions section that clearly defines commonly used terms, as well as define certain terms that are used only once within the Statement provided. When operationalising the indicator into a statement, care needs to be taken to ensure that each indicator statement is indeed composed of only one question, thus a review of the statements to include only the most applicable question to the indicator will be done. Lastly, overlapping indicators will be evaluated to see if the overlap is necessary. If so, the statements will need to be rephrased in order to put the accent on the differences between the indicators and ensure they do not overlap. The focus group also demonstrated the importance of including ISRA in the IS-REF framework, and better explaining how these two work together. Indeed, the comment relating to the scope of the methodology demonstrates that the fact that the context was meant to be identified before ISRA was used was unclear to the participants.

Furthermore, the level of detail required for the evidence and discussion part of the methodology should be clearly defined, and more emphasis should be put on the importance of providing evidence to support ones self-evaluation. In this first pilot study, the scoring was, in general, not underpinned by strong evidence, and thus the results in terms of societal resilience maturity are not reliable. Indeed, the evidence allows future users of the evaluation to understand why the indicators were scored as they were, and also provides good input for the resilience treatment part of IS-REF.

While improvements need to be made, the overall feeling shared by the participants was that ISRA provides a novel approach on disaster risk management and may be useful for evaluating societal resilience, thus demonstrating the success of the method.

6 CONCLUSIONS

This paper has presented the Improver Societal Resilience Analysis (ISRA) and demonstrated the methodology. The methodology is being developed to enrich societal resilience with the knowledge of critical infrastructure resilience on a regional or national level. To ensure operability of the methodology, a focus group consisting of potential future users were involved from the start of the development process. Their feedback provides a basis for further development of the methodology, including increasing clarity, removing overlap and multiple statements, better explaining how ISRA fits into IS-REF, and putting more emphasis on the collection of evidence. These improvements will ensure that ISRA meets the needs of the intended user and is providing a useful process as well as valid results.

The ISRA methodology draws on existing indicators and frameworks for societal resilience analysis; however it introduces the results of critical infrastructure resilience evaluations to the societal resilience analysis as an overall high level indicator of the physical capital of a community. In combination with the IMPROVER IS-REF, as proposed in the introduction to this paper, ISRA is a tool which could be used to enrich societal risk assessments by providing an overview of the capacity of a society to cope with, adapt to and transform in the aftermath of a disaster or emergency. This information, while not necessarily adding value for a risk assessment addressing the frequency of an incident and the immediate consequences of an incident would add significant value and help to evaluate risks by providing an overview of the medium and long term ability of a community to cope with the incident.

ACKNOWLEDGMENT

We would like to thank the focus group participants for their time and valuable input. The IMPROVER

project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 653390.

REFERENCES

Adger, W. (2000). Social and ecological resilience: are they related? *Progress in Human Geography 24*(3), 347–364.

Ahmed, R., M. Seedat, A. Niekerk, & S. Bulbulia (2004). Discerning community resilience in disadvantaged communities in the context of violence and injury prevention. *S. Afr. J. Psychol. 34*(3), 386–408.

Boon, H., A. Cottrell, D. King, R. Stevenson, & J. Millar (2012). Bronfenbrenner's bioecological theory for modelling community resilience to natural disasters. *Natural Hazards 60(*2), 381–408.

Born, P. (2014). *Deepening community: Finding joy together in chaotic times*. San Francisco: BerrettKoehler.

Braun, V. & V. Clarke (2006). Using thematic analysis in psychology. *Qualitative research in psychology 3*(2), 77–101.

Burns, D., F. Heywood, M. Taylor, P. Wilde, & M. Wilson (2004). Making community. A handbook for development and assessment. The Policy Press, Bristol, UK.

Cabinet Office (2011). Strategic national framework on community resilience. Cabinet Office, London, UK.

Chan, S., W. Wey, & P. Chang (2014). Establishing Disaster Resilience Indicators for Tan-sui River Basin in Taiwan. *Social Indicators Research 115*(1), 387–418.

Chandra, A., J. Acosta, S. Howard, L. Uscher-Pines, M. Williams, D. Yeung, J. Garnett, & L. Meredith (2011). Building community resilience to disasters: A way forward to enhance national health security. RAND Corporation, Santa Monica, CA.

Coles, E. & P. Buckle (2004). Developing community resilience as a foundation for effective disaster recovery. *The Australian Journal of Emergency Management 19*(4), 6–15.

Collis, L., F. Schmid, & A. Tobias (2004). Managing incidents in a complex system; a railway case study. *Cogn Tech Work 16*, 171–185.

Cutter, S., C. Burton, & C. Emrich (2010). Disaster resilience indicators for benchmarking baseline conditions. *Journal of Homeland Security and Emergency Management 7*(1).

DSB (2016). Evaluation Report HarbourEx15. Norwegian Directorate for Civil Protection (DSB).

Evart, J. & H. McLean (2017). Swimming against the tide: How disaster agencies build political resilience. *International Journal of Public Administration 40*(7), 539–547.

FEMA (2014). How parks and open spaces can strengthen resilience. https://www.fema.gov/disaster/4085/updates/howparks-and-open-spaces-can-strengthen-resilience. Online. Accessed: 2017-09-30.

Flora, C. & J. Flora (2004). *Rural communities: Legacy and change* (2nd ed.). Boulder, CO: Westview Press.

Folke, C. (2006). Resilience: The emergence of a perspective for social-ecological systems analyses. *Global Environmental Change 16*(3), 253–267.

Frankenberger, T., M. Mueller, T. Spangler, & S. Alexander (2013). Community Resilience: Conceptual Framework and Measurement. Westat, Rockville, MD.

Furedi, F. (2007). The changing meaning of disaster. *Area 39*(4), 482–489.

Höppner, C., R. Whittle, M. Bründl, & M. Buchecker (2012). Linking social capacities and risk communication in Europe: A gap between theory and practice? *Natural Hazards 64*(2), 1753–1778.

ISO/DIS 22316:2017 (2017). Security and resilience—guidelines for ogranizational resilience. Standard, International Organization for Standardization, Geneva, CH.

Keck, M. & P. Sakdapolrak (2013). What is social resilience? lessons learned and ways forward. *Erdkunde 67*(1), 5–19.

Kuhlicke, C., A. Steinführer, C. Begg, C. Bianchizza, M. Bründl, M. Buchecker, B. De Marchi, M. Di Masso Tarditti, C. Höppner, B. Komac, L. Lemkow, J. Luther, S. McCarthy, L. Pellizzoni, O. Renn, A. Scolobig, M. Supramaniam, S. Tapsell, G. Wachinger, G. Walker, R. Whittle, M. Zorn, & H. Faulkner (2011). Perspectives on social capacity building for natural hazards: outlining an emerging field of research and practice in Europe. *Environmental Science & Policy 14*(7), 804–814.

Lange, D., D. Honfi, J. Sjöström, M. Theocharidou, G. Giannopoulos, N. Reitan, K. Storesund, L. Melkunaite, H. Rosenqvist, L. Petersen, R. Almeida, B. Rød, C. Bouffier, E. Serafinelli, & M. Lin (2017). Framework for implementation of resilience concepts to critical infrastructure. IMPROVER Project Deliverable 5.1 [To be published].

Magis, K. (2010). Community resilience: An indicators of social sustainability. *Society & Natural Resources 23*(5), 401–416.

Marshall, N. (2010). Understanding social resilience to climate variability in primary enterprises and industries. *Global Environmental Change 20*(1), 36–43.

Mayunga, J.S. (2007). Understanding and Applying the Concept of Community Disaster Resilience: A capital-based approach. Draft working paper prepared for the summer academy for social vulnerability and resilience building.

Michel-Kerjan, E. (2015). We must build resilience into our communities. *Nature 524*, 389.

Norris, F., S. Stevens, B. Pfefferbaum, K. Wyche, & R. Pfefferbaum (2008). Community resilience as a methaphor, theory, set of capacities, and strategy for disaster readiness. *American Journal of Community Psychology 41*, 127–150.

Paton, D. & D. Johnston (2006). *Disaster resilience: An integrated approach*. Springfield, Illinois: Charles C. Thomas.

Poortinga, W. (2012). Community resilience and health: The role of bonding, bridging, and linking aspects of social capital. *Health and Place 18*, 286–295.

Priest, S.J., C. Suykens, H.F. van Rijswick, T. Schellenberger, S. Goytia, Z.W. Kundzewicz, W.J. van Doorn-Hoekveld, J.C. Beyers, & S. Homewood (2016). The European Union approach to flood risk management and improving societal resilience: Lessons from the implementation of the Floods Directive in six European countries. *Ecology and Society 21*(4), 50.

Pursiainen, C., B. Rød, M. Alheib, G. Baker, C. Bouffier, S. Bram, G. Cadete, E. Carreira, P. Gattinesi, F.

Guay, D. Honfi, K. Eriksson, D. Lange, E. Lundin, A. Malm, L. Melkunaite, M. Merad, M. Mira da Silva, L. Petersen, J. Rodrigues, R. Salmon, M. Theocharidou, & A. Willot (2016). Report of criteria for evaluating resilience. Improver project deliverable 2.2.

Reitan, N., K. Storesund, C. Sesseng, R. Almeida, D. Lange, J. Sjöström, Ö. Durgun, L. Petersen, B. Rød, C. Bouffier, L. Vigh, & H. Rosenqvist (2017). Plan for the pilot implementation of a resilience management framework for critical infrastructure. IMPROVER Project Deliverable 6.1 [To be published].

Renschler, C., A. Frazier, L. Arendt, G. Cimarello, A. Reinhorn, & M. Bruneau (2010). Developing the 'peoples' resilience framework for defining and measuring disaster resilience at the community scale. In *Proceedings of the 9th US National and 10th Canadian Conference on Earthquake Engineering*, Toronto, Canada.

Rinaldi, S., J. Peerenbom, & T. Kelly (2001). Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems 21*(6), 11–25.

Roux-Rouquié, M. & J.-L. Moigne (2002). The systemic paradigm and its relevance to the modelling of biological functions. *C. R. Biologies 325*(4), 419–430.

Serafinelli, E., R. Stevenson, P. Reilly, L. Petersen, L. Fallou, & E. Carreira (2017). A Communication Strategy to build Critical Infrastructure Resilience. Improver project deliverable 4.2.

Sharifi, A. & Y. Yamagata (2016). On the suitability of assessment tools for guiding communities towards disaster resilience. *International Journal of Disaster Risk Reduction 18*, 115–124.

Sharma, M. G. & S. K. Srivastava (2016). Leveraging the social welfare chain to provide resilience during disaster. *International Journal of Logistics Research and Applications 19*(6), 509–519.

Tierney, K. (2007). Business and disasters: Vulnerability, impacts and recovery. In H. Rodriguez, E. Quarantelli, and R. Dynes (Eds.), *Handbook of Disaster Research*, pp. 275–296. New York, NY: Springer.

UNDP (2014). Understanding Community Resilience: Findings from Community-Based Resilience Analysis (CoBRA) Assessments. United Nations Development Programme, Drylands Development Centre, Nairobi, Kenya.

Voss, M. (2008). The vulnerable can't speak. An integrative vulnerability approach to disaster and climate change research. *Behemoth. A Journal on Civilization 1*(3), 39–56.

WHO/EHA (1998). Risk Assessments for Emergency. Management Emergency Health Training Programme for Africa. Panafrican Emergency Training Centre, Addis Ababa.

Wilson, G. (2012). *Community Resilience and Environmental Transitions*. Oxon: Routledge.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Novel methodologies for analysing critical infrastructure resilience

K. Storesund & N.K. Reitan
RISE Fire Research, Trondheim, Norway

J. Sjöström
RISE Research Institutes of Sweden, Borås, Sweden

B. Rød
UiT The Arctic University of Norway, Tromsø, Norway

F. Guay
INOV INESC Inovação, Lisbon, Portugal

R. Almeida
Danish Institute of Fire and Security Technology, Hvidovre, Denmark

M. Theocharidou
European Commission, Joint Research Centre, Ispra, Italy

ABSTRACT: In the field of Critical Infrastructures (CI), both policy and research focus has shifted from protection to resilience. The IMPROVER project has developed a CI resilience management framework (ICI-REF), applicable to all types of CI and resilience domains (technological, organisational and societal) allowing operators to understand and improve their resilience. IMPROVER has also developed methodologies to be used within the framework, accompanied with resilience indicators for operators to assess their technological and organisational resilience. The framework allows CI operators to incorporate resilience management as part of their risk management processes. The ICI-REF, the resilience analysis methodologies and indicators have been optimised, applied and demonstrated in a pilot implementation, focusing on the potable water supply in Barreiro, Portugal. Conclusions from the operators so far are that the indicators, well-defined and unambiguously described, are crucial for monitoring resilience activities, to ensure objective, consistent, repeatable and representative results from the assessed processes.

1 INTRODUCTION

Increasing Critical Infrastructure (CI) resilience is one is one of the main objectives for the European strategy towards a more secure Europe (COM, 2010). Through the Program for Critical Infrastructure Protection (EPCIP), issues and approaches to focus on are defined, where measures to facilitate implementation of resilience concepts to CI are identified (SWD, 2013). The concept of resilience has evolved from ecological resilience, via psychology, engineering to the disaster risk reduction field. There is thus a range of definitions of the concept of resilience and for this context we use that of UNISDR *"The ability of a system, community or society exposed to hazards to resist, absorb, accommodate to and recover from the effects of a hazard in a timely and efficient manner, including through the preservation and restoration of its essential basic structures and functions"* (UNISDR, 2009). In EU, CI is defined as: "*an asset, system or part thereof located in Member States which is essential for maintenance of vital societal functions, health, safety, security, economic or social well-being of people, and the disruption or destruction of which would have a significant impact in a Member State as a result of the failure to maintain those functions*" (Council, 2008).

An overall goal of the EU-funded Horizon 2020 project IMPROVER is to improve European CI resilience to crisis and disasters, through the implementation of technological, organisational and societal resilience concepts. To that end, the IMPROVER Critical Infrastructure REsilience Framework (ICI-REF) was developed. The framework is supported by resilience analysis methodologies and indicators, also developed in IMPROVER. It is inspired by existing stand-

ards and frameworks e.g. ISO 31000, ISO 22301, ISO 22316, Org. Resilience HealthCheck (Austr. Government, 2017), Benchmark Resilience Tool (Resilient Organisations, 2014) and Resilience Measurement Index (Petit et al, 2013).

To ensure that the developed ICI-REF framework, with supporting methodologies and indicators, is fit-for-purpose, it is optimised in pilot implementations, by application to relevant scenarios in semi-real environments at several living labs. One pilot implementation has recently been conducted, focusing on potable water supply in Barreiro, Portugal.

This paper describes structures and processes of the ICI-REF and resilience analysis methodologies, including preliminary results of the pilot implementation at the Barreiro living lab, Portugal.

2 IMPROVER CRITICAL INFRASTRUCTURE RESILIENCE FRAMEWORK (ICI-REF)

2.1 *The ICI-REF structure and process*

ICI-REF is a general and well-defined framework for managing the technological, organisational and societal resilience of CI (Lange et al., 2017a; 2017b). It includes the flexibility to account for the unique features of the various types of CI, giving CI operators an understanding of, and a capability to improve, their resilience. The framework extends standard risk procedures (ISO 31000) and considers resilience assessment as complementary to risk assessment. The framework is constructed such that it is easily incorporated within existing risk management processes by CI operators. Initial feedback by CI operators (Theocharidou et al., 2016) indicated this approach as the most feasible one, as it can improve their current practices and allow for risk and resilience management decisions to be taken based on the results of both assessments. ICI-REF allows operators to perform self-assessment or focused analysis of technological/organisational aspects in order to either monitor resilience over time, or compare to similar CI within the same sector. The ICI-REF structure is depicted in Fig. 1.

The ICI-REF process starts with *establishing the context*, implying the gathering of information, defining the resilience domain(s), etc. The defined context, risk identification and risk analysis are then fed into, and complemented by, the resilience assessment process. *Resilience assessment* comprises of *resilience analysis* and *evaluation* against pre-defined criteria. Three different resilience analysis methodologies have been developed (described in 2.3). The results from risk and

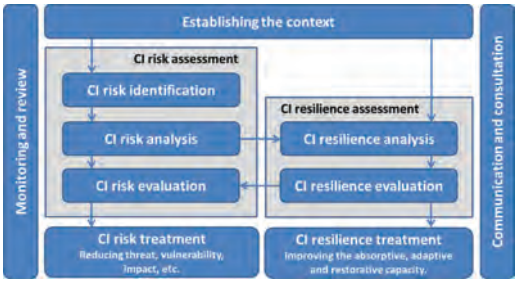


Figure 1. Structure of the IMPROVER Critical Infrastructure REsilience Framework (ICI-REF).

resilience assessments constitute the basis for designing treatment plans, describing how to both mitigate risk and improve resilience. This parallel process allows decision makers to select risk and resilience measures in a cost-effective way, especially when a measure can be implemented to address both risk and resilience objectives. Throughout the ICI-REF process, *Monitoring and review* as well as *Communication and consultation* are continuous background processes (see Lange et al., 2017).

CI resilience analysis can be performed by implementing existing methodologies, or by methodologies developed in IMPROVER. This paper focuses on technological and organisational resilience analysis, for which resilience assessments are performed at the CI level. Assessment and management of societal resilience shall instead be conducted on regional or national levels, using CI resilience assessments as input. A modified version of ICI-REF is developed for this purpose (Rosenqvist et al., 2018).

2.2 *Resilience indicators*

In the context of IMPROVER, the term "resilience indicators" is related to variables that can be used, either alone or in combination, as a representation of resilience. Qualitative, semi-quantitative or quantitative indicators are analysed and, when sufficient, aggregated to a measure of resilience.

The resilience indicators should be clearly defined, in order to ensure objectivity and a proper balance between generality and specificity. To monitor resilience over time or comparing to similar CI, the indicators must also provide reproducibility and repeatability. Measurement scales for the indicators and their possible weight factors should ideally be benchmarked at a sectoral level.

Based on literature and defined requirements from CI operators associated with IMPROVER, the resilience indicators to be included in the resilience analysis step of ICI-REF are developed and optimised. They relate to the various resilience

analysis methodologies used for different resilience domains.

2.3 *Resilience assessment*

As a first step, the CI operator may want to conduct an initial self-assessment to indicate strengths and weaknesses in its resilience; i.e. in which areas or domains a more in-depth assessment is required. For this purpose, the operator may find a resilience analysis methodology with high flexibility useful, such as the Critical Infrastructure Resilience Index (CIRI) developed in IMPROVER (Pursiainen et al., 2017).

This process may be sufficient, but if required, operators can perform re-assessment by using analysis methodologies which goes more into details. For this purpose, two different methodologies have been developed: the IMPROVER Technological Resilience Analysis (ITRA) and IMPROVER Organisational Resilience Analysis (IORA) for analysing technological or organisational resilience, respectively (Bram et al., 2017; Mindykowski et al., 2016). CIRI, ITRA and IORA methodologies are briefly described below.

2.3.1 *Critical Infrastructure Resilience Index (CIRI)*

Critical Infrastructure Resilience Index (CIRI) is a holistic and easy-to-use self-assessment methodology. It is applicable to all types of infrastructures, and built on a four level hierarchy of indicators, focusing mainly on the technological and organizational domain. The backbone for CIRI is the crisis management cycle (OECD, 2011; Pursiainen, 2017). The different phases in the cycle corresponds to the seven Level 1 indicators: Risk assessment, Prevention, Preparedness, Warning, Response, Recovery, and Learning, Fig. 2. Under each Level 1 indicator there is a subset of given generic indicators (Level 2).

Further, for each Level 2 indicator there is a new subset of mainly given, measurable indicators. However, as sectors use different metrics and measures (quantitative/qualitative) the exact measurement depends on the sector, referred to as Level 4 indicators, the bottom of the hierarchy.

For a common viewpoint, Level 4 indicators are transformed to qualitative maturity scale, scaling, from 0 to 5. At Level 3 and 4, the operator has the possibility to assign weight to the indicators according to their importance. After assessing the Level 4 indicators, results are aggregated up the hierarchy, and each Level 1–3 indicator get a score from 0 to 5. The result is presented in a radar chart with all the seven Level 1 indicators.

In addition, to present a more detailed analysis, it is possible to construct charts for all Level 1 over their respective Level 2 indicators, see Fig. 3.

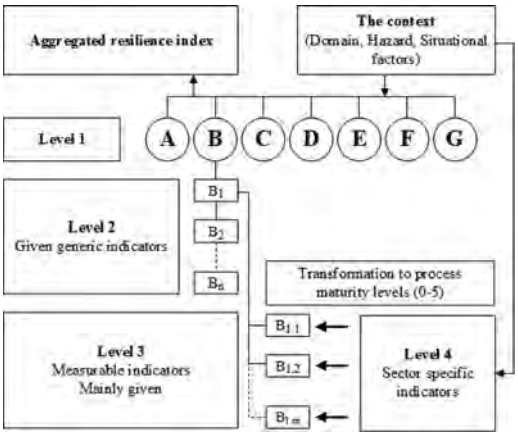


Figure 2. The hierarchical structure of Critical Infrastructure Resilience Index (CIRI). The Level 1 indicators, representing different phases in the risk management cycle, are here denoted (A) Risk assessment, (B) Prevention, (C) Preparedness, (D) Warning, (E) Response, (F) Recovery and (G) Learning.

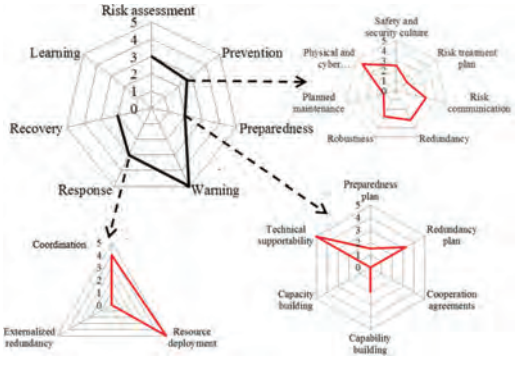


Figure 3. Analysis results for most Level 1 indicators, and Level 2 indicators under Prevention, Preparedness, and Response.

It should be noted here that this is a self-assessment methodology and thus not fully objective. However, the result is indicative of the CI's resilience level and highlights strengths and weaknesses of the infrastructure, both from the technological and organisational perspective. It can be used as the basis for further detailed analysis, using methodologies like IORA and ITRA.

2.3.2 *IMPROVER Technological Resilience Analysis (ITRA)*

Technological resilience is often visualised using the performance loss and recovery function or the area between the function and an uninterrupted capacity/performance. From the risk identification,

Fig. 1, a prioritised list of possible hazards which could impact the CI is used as input to the technological resilience analysis, which aims at quantifying the performance loss and recovery of the CI service. Thus, technological resilience is conditional on the occurrence of a specific hazard, following the procedure of the risk management of ISO 31000.

Estimating the functionality needs therefore suitable *intensity measure* of the hazard to which the vulnerability of the system's subparts can be evaluated through their *fragility.* Combining this information gives a measure of the *damage* to the system which should be transformed into one or several *performance measures* in order to focus on the core aspect of resilience: functionality of the system.

Once the performance measures loss and recovery functions are estimated they should be evaluated against other CI, historical performance or the needs and expectations of the infrastructure's end-users. It is therefore of vital importance to choose the performance criteria keeping in mind that: (i) they should be possible to translate from estimated damages, with sufficient accuracy and (ii) they should be constructed such that they can be compared to other CIs, historical performance or (preferably) the needs and tolerances from the end-user (Petersen, 2018).

2.3.3 *IMPROVER Organisational Resilience Analysis (IORA)*

IORA follows a similar structure to other organisational analysis methods. The purpose of the analysis is promoting resilient performance. Subsequent levels are functions, forms and processes which contribute to this purpose. The functions required to achieve this are: design of tasks and roles; design of the framework and its content, goals, rules, processes and procedures; strengthening collaboration; learning and redesign; underlying values and interpretations, Fig. 4.

Organisational resilience analysis process requires collection and processing of information about how the organisation's processes contribute to this. For the Barreiro implementation this is done via in-depth interviews based on a narrative of a historical event (saline intrusion in a fresh water well). Functions, forms and processes during this event form the basis for the analysis and the subsequent evaluation.

3 PILOT IMPLEMENTATION ON POTABLE WATER SUPPLY NETWORK

3.1 *Test object*

The object to be tested in the pilot implementation, comprises of the ICI-REF, its supporting methodologies for resilience analysis (CIRI, IORA and ITRA) and the developed resilience indicators. The test object will be denoted as ICI-REF in the remainder of the document for simplicity.

3.2 *Living lab: The potable water supply system of Barreiro*

Barreiro's municipality, with an area of 36.41 km^2, has, according to the Census 2011, a population of 78,764 people. It has 17 km river front to Tagus and Coina rivers and an important road-rail-river terminal. It is located about 40 km from Lisbon to which it is linked by two bridges, and about 35 km from Setúbal, the district capital. Barreiro's potable water supply system consists of 11 licensed ground-water intakes from a semi-confined aquifer, 7 reservoirs for treated water storage with the total capacity of 12.750 m^3, 7 treatment installations, for disinfection with the addition of sodium hypochlorite, 3 pumping stations, 5 blowers, 16.1 km of main ducts, and 308 km of meshed distribution pipes.

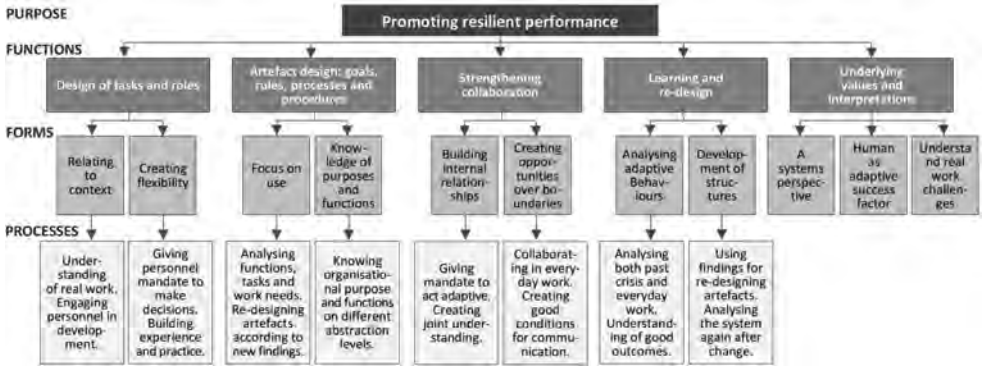


Figure 4. Indicators on different abstraction levels in the IMPROVER Organisational Resilience Analysis methodology (IORA).

The municipality has a remote management system that allows real time monitoring of pressure and flows in the water supply (and waste water) systems. The pilot implementation focuses on three pressure zones in the north, which combined account for 60% of the total water supply in the municipality.

Fig. 5 shows the area subject to the assessment.

3.3 *Systematic approach for testing and evaluating the performance of ICI-REF*

To make the pilot implementation robust, a triangular approach was used for testing and evaluating the performance of ICI-REF. Triangulation is the combination of two or more data sources, investigators, methodologic approaches, theoretical perspectives or analytical methods within the same study (Denzin, 1970; Kimchi et al., 1991). Using multiple methods decreases the "deficiencies and biases that stem from any single method" (Mitchell, 1986) creating "the potential for counterbalancing flaws or the weaknesses of one method with the strengths of another." Therefore; focus group, documentation, field studies and surveys were used to collect data for the critical evaluation of the performance of ICI-REF. The IMPROVER project embraces all these approaches in several steps and iterations for optimising ICI-REF.

3.3.1 *Collection of data*

A *focus group,* consisting of representatives from the operator at the Barreiro living lab, was selected based on their insight into current processes and methodologies for risk assessment at the Barreiro living lab. There has been close cooperation between the focus group and the project team throughout the project via continuous communication, and workshops. These were invaluable in addressing strengths and weaknesses of ICI-REF before the final pilot implementation. The focus group, as a qualitative, exploratory research method, has aided the understanding about not only the operators' opinions, but also how and why they think the way they do.

Field studies were performed for testing the application of ICI-REF in a semi-real environment. Field studies require detailed observation and evaluation, allowing conclusion of understanding and comparisons of the information generated from each site (Burgess, 1984; Denzin & Lincoln, 2011; Rossman & Rallis, 2011). An advantage of field studies is that they give better external validity than in laboratory experiments because a field experiment takes place in typically occurring social settings.

The field study relied on application of ICI-REF to a relevant hazard scenario. A scenario with high disaster risk was prioritised by structured expert judgement elicitation by the stakeholders. Fig. 6 shows a hazard map for the Barreiro living lab.

The hazard chosen to assess the resilience of the water supply system was an earthquake with liquefaction, which is considered the highest disaster risk for the water network combining consequence and probability. The assets susceptible to the hazard are:

- The reservoir, pipe system, pumps and the critical users being the hospital and the health centre.
- All technical equipment used to repair and to distribute redundant functionality.

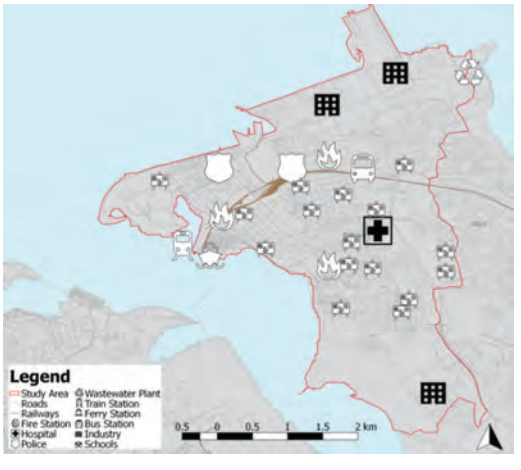


Figure 5. The northern part of the Barreiro municipality subject to the pilot implementation.

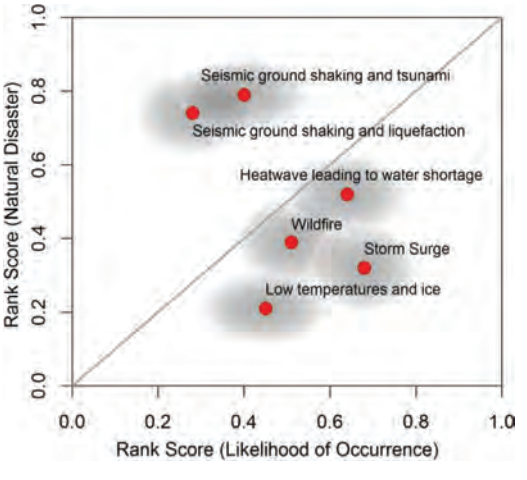


Figure 6. Plot of rank scores of six natural hazard scenarios based on their likelihood to cause disaster and to occur at Barreiro's water network in the next 5 years. (Pursiainen et al., 2015).

All staff and the entire organisation and the processes used in the preparatory, functional and administrative work.

Documentation was collected in order to analyse vital data from the CI. For example, the safety plans, and organisation chambers. These documents were used to assess the as-is situation of the CI. Typically, the analysis aims at visualising the current state process to clarify how the CI process works today, and what can be done to improve the current situation.

Different forms of *surveys*, aimed at the operator, project team members and other stakeholders were used in advance, during and after the pilot implementation. By using surveys, a broad range of data has been collected, e.g. tolerance levels; attitudes; opinions; beliefs; values; behaviour and factual. The surveys were used as basis both for defining performance criteria for resilience assessment and for the critical evaluation of the performance of ICI-REF.

3.3.2 *Critical evaluation*

Eighteen success factors were developed for the critical evaluation of the performance of ICI-REF. These ensure that ICI-REF meets stakeholders and end-users needs and are designed based on continuous input from the living labs during the project. The design science research methodology (Hevner et al., 2004) is used for the critical evaluation process in which the success factors are evaluated based on demonstration results and applications of ICI-REF.

The defined success factors of the project are primarily designed for critical evaluation of the overall ICI-REF framework, but they also implicitly set requirements to the relevance and quality of the tested analysis methodologies with indicators. Examples of success factors related to indicators are shown in Table 1.

Table 1. Examples of success factors for critical evaluation of the relevance and quality of resilience indicators.

Success factor	Defined by
The framework shall be applicable to all types of CI	The balance and definitions of indicators
	Clearly described and categorised indicators
The framework shall be easy to use	Guidance on how framework indicators can be interpreted in relation to resilient performance
The framework shall provide effective and coherent crisis and disaster resilience management	Resilience indicator follow-up should promote a shared view within the organisation on real work challenges.
The framework is arranged for being revised continuously	Existence of a system for recurring analysis, criticism and revision of the indicator framework and implementation

Table 2. Scale for analysing perception of indicators by the Barreiro operator.

Rating	Definition
A	The indicator was perceived, and there is evidence of the indicator
B	The indicator was perceived, but there is no evidence of the indicator
C	The indicator was not perceived
D	Not applicable

3.4 *Results from initial demonstrations*

CIRI, IORA and ITRA were all applied in the initial demonstrations. Evaluation from ITRA showed that the system most probably will meet the expectations of end-users for reasonable scenarios of damage. Also, despite being highly dependent on key personnel resources the flat organisation helps in fast recovery in times of crises, as shown in IORA evaluation.

A set of resilience indicators was tested within the CIRI methodology and assessed, using a software tool developed in IMPROVER (accessed at: http://improver-inov.herokuapp.com/). The indicators were discussed and evaluated by the operator according to the indicators' relevance and comprehensibility as means of assessing their resilience.

The operator was asked to assign resilience measurement scores to the indicators, and to rate them on the perception scale, according to how well they were understood by the operator. The scale for the perception ratings is presented in Table 2.

The structure and processes of resilience analysis methodologies proved functional in the demonstration. Based on the feedback from the operator, only minor modifications of the methodologies were required to optimise the relevance of the analysis results towards the main pilot implementation. The operator expressed the need for user-friendly, clear and not too complex assessments. They further concluded that the structure is not the main point of interest to the living lab, but the functionality of the assessment process, and the questions and goals related to the indicators. An issue pointed out by the living lab is which resources are required to perform the assessment; i.e. whether they need to employ external resources or can train internal resources.

The operator emphasised the crucial role of resilience indicators in the monitoring of resilience activities. However, to ensure objective, consistent, repeatable and representative results, the indicators and their designed questions must be defined using unambiguous terms. As long as the indicators are well described and leave little room for subjectivity, the high number of indicators is not a problem. The need for guidelines was also expressed.

Challenges related to the definitions of measurement scales and assignments of weights for qualitative or semi-quantitative indicators were pinpointed. E.g. the measurement scale used to assess the indicators should be well-defined since it is mandatory to understand the differences between the different measurement scales to perform benchmarking. It was also discussed how flexible the indicator structure should be; e.g. if CI operators shall be allowed to define their own scales and weights, and how this will affect the assessments and limit their relevance.

The perception of the operator that some of the indicators were too vague, needed to be better explained and that some were difficult to point out evidence for, led to adjustments and development of the overall set of indicators and how they are presented.

To address the need for proper descriptions and definitions of sector-specific indicators, "indicator cards" were developed for the complete developed set of technological and organisational resilience indicators at the lower CIRI level. Each individual resilience indicator card provides a detailed description of the sector-specific indicator subject to assessment as exemplified in Fig. 7.

The indicator cards consist of the following information:

- The assessed indicator and its parent indicators are listed.
- Detailed information about the context is given. The resilience domain (technological or organisational), hazard types (natural, non-malicious man-made, malicious man-made and multi-hazards) and situational factors (e.g. temporal, geographical or conceptual considerations for taking such an indicator into account) are indicated. Finally, the applicable sector (in this case potable water supply) is pointed out and if the indicator is generic or scenario specific.

A description of the indicator and guidance for assessing the maturity level is provided through a rationale of why this indicator is justified. Moreover, a question is provided, which can be asked to the operator for measuring the indicators in a clear and explicit manner with the 6 different maturity levels described (scale 0–5) and a reference for describing the indicator.

Resilience indicator card		
Level 1	Response	
Level 2	Communication	
Level 3	Interoperable information and communication technology	
Level 4	**Availability (SIRESP)**	
Context		
Domain	Technological/Organisational	
Hazard Type	General	
Situational factors	N/A	
Applicable sectors	Water Supply	
Generic or scenario specific	Generic	
Indicator description and assessment		
Rationale	SIRESP is the emergency communication system used in Barreiro, and it is crucial for effective coordination and exchange of information during emergency periods.	
Question	What is the availability level of the SIRESP system?	
Answer modalities	0	0-1 %
	1	1-10 %
	2	10-50 %
	3	50-90 %
	4	90-99 %
	5	>99 %
References	SIRESP manual	

Figure 7. Indicator card for a sector-specific indicator at a lower CIRI level (here level 4), showing its parent indicators, context, description and measurement scale.

3.5 *Results from pilot implementation*

Based on the feedback from the initial demonstrations, ICI-REF was optimised, and the pilot implementation was conducted.

The resilience assessments resulted in suggestions for resilience treatment. Raising public awareness as well as training were pinpointed from the three tested methodologies. Application of CIRI resulted in recommendations to prevent silos, i.e. to have quick and easy cooperation between management and people in the field and to have structures in place to ensure this. Application of IORA actually identified that such a characteristic existed in the Barreiro organisation however more as an unofficial way of working. This demonstrates the ability of the different methodologies to bring up different levels of details and different perspectives.

The performance of ICI-REF in the pilot implementation is currently being critically evaluated with regards to the success factors. Generally, indicators were perceived by the operator as clearly described and easy to interpret, hence the adjustments made after the initial demonstrations had

improved ICI-REF significantly. When weighting the indicators, information from the operator could be valuable with regards to the importance of an indicator, but at the same time, it is important that the indicators are not biased towards the operator. The operator was of the opinion that ICI-REF can be valuable both as an internal audit tool and also in everyday work, and that it was useful in promoting reflection around resilience of the organisation and resilience treatment. The ability to compare results with other operators in the same sector outside of Portugal would also be useful for benchmarking purposes.

In order to provide an overall resilience score, all relevant indicators must be assessed. Although the pilot implementation assessed only a sample set of all the defined indicators, the operator found the results valuable for prioritising future work and development within their organisation.

4 DISCUSSION AND WAY FORWARD

After the initial demonstration of ICI-REF at the Barreiro living lab, the operator was of the opinion that indicators are crucial for monitoring resilience activities. However, to ensure that the assessment results are objective, consistent, repeatable, and representative of the assessed processes, the indicators should be defined using unambiguous terms. It was strongly suggested that clear questions should be asked for the operator to better understand what the indicator is assessing. The main potential for improvements of ICI-REF therefore lies in the design of sector-specific resilience indicators.

The indicators must not only be comprehensible and clear, but also at the same time leave some room for site-specific information. The degree of indicator specificity has been discussed with several living labs through the project, and the need for a balance between generality and specificity has been emphasised. If an indicator is too general, this may reduce the ability to detect details or new areas of resilience improvements. On the other hand, information about the specific CI can also be lost if the indicator is too detailed, which can make a further comparison with similar CI less relevant.

Regarding the measurement scales for the sector-specific indicators, it is not only challenging to define the scales, but it may also be challenging to assign quantitative value to a qualitative indicator without introducing subjectivity. The operator should therefore provide evidence and comments to support their assigned values for each indicator.

Despite the challenges in defining the indicator scales, weights and degree of specificity, the need for including sector-specific indicators are unquestionable. It should be described in terms of guidelines or references, at which level the indicators' scales and weights should be defined to ensure legitimacy. Benchmarked indicators exist within certain CI sectors. Although it may not be a requirement for a CI to compare to similar CI's, the living labs have expressed the wish to perform such a comparison at a regional level. However, for indicators that are not benchmarked, the comparison between similar CI will not be applicable if the operators, themselves, define scales and weights.

The indicator cards for the Barreiro living lab were successfully tested in the pilot implementation, as the indicators were considered well described and easy to assess and respond to. Indicator cards are now being developed for application in the next pilot implementation at another of the living labs in IMPROVER; the M1 highway in Budapest, Hungary.

5 CONCLUSION

The ICI-REF, technological and organisational resilience analysis methodologies and indicators have been applied and demonstrated in a pilot implementation, focusing on the potable water supply in Barreiro, Portugal. These have been developed with the aim to smoothly extend current risk management practices into a resilience management framework. A set of technological and organisational resilience indicators has been designed and described in "indicator cards". Efforts are made to improve the clarity of definitions and descriptions of resilience indicators, since unambiguous description of indicators is crucial for monitoring resilience activities. Based on the feedback from the Barreiro living lab during the project, initial demonstrations and a pilot implementation, the ICI-REF and the developed resilience analysis methodologies proved functional based on preliminary results.

They are now ready to be fine-tuned towards the next pilot implementation in IMPROVER. This focuses on the M1 highway in Budapest, Hungary, covers several scenarios and will be finalised in 2018. Combining results from the two pilot implementations allows evaluating the performance of the ICI-REF framework, methodologies and indicators to different CI sectors and contexts. Based on this, European guidelines for the resilience management to CI will be developed, addressed both to CI operators and policy makers.

ACKNOWLEDGEMENTS

Barreiro Municipality are acknowledged for valuable feedback and active participation in the

pilot implementation. This work is funded by the EU's Horizon 2020 research and innovation programme, grant agreement No 653390. It does not reflect the official opinion of the EU. Information and views expressed therein is the responsibility of the authors.

REFERENCES

Austr. Government (2017). Organisational Resilience Healthcheck; https://www.organisationalresilience.gov.au

Bram, S. et al. (2017) *Organisational resilience concepts applied to critical infrastructure*, IMPROVER Project: Deliverable 4.3.

Burgess, R.G. (1984) *In the Field: An Introduction to Field Research*, London: Allen and Unwin.

COM (2010) 673 final. *The EU Internal Security Strategy in Action: five steps towards a more secure Europe.* Brussels, 22.11.2010.

Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and assessment of the need to improve their protection.

Denzin, N.K. (1970). *The research act: A theoretical introduction to sociological methods.* Chicago: Aldine.

Denzin, N.K. and Lincoln, Y.S. (2011). *The Sage handbook of qualitative research*. Sage.

Hevner, A.R. et al. (2004). *Design Science in Information Systems Research*, MIS Quarterly, vol. 28, pp. 75–105.

ISO 22301:2012, *Societal security—Business continuity management systems—Requirements.*

ISO 22316:2017, *Security and resilience—Organizational resilience—Principles and attributes.*

ISO 31000:2009, *Risk management—Principles and guidelines.*

Kimchi, J. et al. (1991). *Triangulation: Operational definitions*. Nursing Research, 40(6), 364–366.

Lange, D. et al. (2017a). *Framework for implementation of resilience concepts to Critical Infrastructure, IMPROVER Project: Deliverable 5.1.*

Lange, D. et al. (2017b). *Incorporation of resilience assessment in Critical Infrastructure risk assessment frameworks*, In: Safety and Reliability—Theory and Applications, ISBN 978-1-138-62937-0, p. 1031–1038.

Mindykowski, P. et al. (2016). *Physical exposure identification and mapping methodologies*, IMPROVER Project: Deliverable 3.1.

Mitchell, E.S. (1986). *Multiple triangulation: A methodology for nursing science.* Advances in Nursing Science, 8(3), 18–26.

OECD (2011). *Future Global Shocks, Improving Risk Governance*. OECD Reviews of Risk Management Policies. ISBN 978-94-09520-5. 139 p.

Petersen, L. et al. (2018). *Creating comparable public tolerance and technical performance indicators for CI resilience.* ESREL 2018 (to appear).

Petit, F.D. et al. (2013). *Resilience Measurement Index: An Indicator of Critical Infrastructure Resilience.* ANL/DIS-13-01, Argonne National Laboratory, USA.

Pursiainen, C. et al. (2015). *Report of criteria for evaluating resilience*. IMPROVER Project: Deliverable 2.2.

Pursiainen, C.H. (2017) The Crisis Management Cycle. Routledge ISBN 9781138643871.

Resilient Organisations (2014). Resilience Benchmark Tool, New Zealand; available at http://brt.resorgs.org.nz

Rosenqvist, H. et al. (2018). *ISRA: A societal resilience analysis methodology*. ESREL 2018 (to appear).

Rossman, G.B. and Rallis, S.F. (2011). *Learning in the field: An introduction to qualitative research*. Sage.

SWD (2013) *318 final, Commission Staff Working Document a new approach to the European Programme for Critical Infrastructure Protection Making European Critical Infrastructures more secure.* Brussels, 28.8.2013.

Theocharidou, M. et al. (2016). *Report of Operator Workshop 1*, IMPROVER Project: Deliverable 1.4.

UNISDR, United Nations Office for Disaster Risk Reduction, 2009 UNISDR terminology on disaster risk reduction; available at http://unisdr.org/.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Creating comparable public tolerance and technical performance measures for critical infrastructure resilience evaluation

L. Petersen
Euro-Mediterranean Seismological Centre, Bruyères-le-Châtel, France

E. Lundin, J. Sjöström & D. Lange
RISE Research Institutes of Sweden, Borås, Sweden

R. Teixeira
Divisão de Águas e Saneamento, Barreiro Municipality, Portugal

ABSTRACT: No consensus currently exists on how to measure and evaluate Critical Infrastructure (CI) resilience. Attempting to use the public's declared coping capacity as a target for CI resilience, this paper explores how to develop relevant resilience performance measurements that enable comparison to the tolerance levels of the general public. To do so, one must first establish the normal performance of the system and the applicable performance measures. Then, a survey is used to convert public perception into these measures as to enable comparison with the technical resilience performance. The CI resilience will be presented through a family of so-called resilience triangles which will illustrate the evolution of the performance, before, during and after a crisis event. A case study of the Municipal Water Network of Barreiro, Portugal, is used. The overall performance is preferably described with the categories quality, quantity and delivery. In quantifying the performance the importance of what is being assessed, to what hazard and for which end-user became evident.

1 INTRODUCTION

Critical infrastructure (CI) resilience is often defined as the ability to maintain a minimum acceptable level of service and the ability to rapidly restore full service in relation to a crisis event on the CI system. However, no consensus currently exists on how to measure these elements. Furthermore, most existing methodologies do not take into account human factors arising from the society which the CI serves. Since the general public, end users of CI services, appear to have reasonable expectations of CI operators in crisis times (Petersen et al., 2017), we suggest using their declared coping capacity as a reference for CI resilience measurements. In order to do so, one must convert public perception into measurable indicators, which can be comparable to the technical performance of the service, and examine how these indicators can be measured, forecasted or assessed over the course of an event. Thus, this paper explores how to identify relevant comparable performance measurements for the CI case of a water distribution network. The methodology will present the resilience measures through a family of so-called resilience triangles, which will illustrate the evolution of the performance, before, during and after a crisis event. A resilience triangle (Bruneau et al., 2003) is shown schematically in Figure 1.

The performance, Q, quantifies the performance of the system. For many CI systems, it is normal that the system performance decays slowly over time as a result of aging, reflected in by the difference between Q_0 and Q_1. A sudden drop in the performance represents the effect of a sudden shock to the system. How deep the drop goes and the steepness of the drop depends on how well the sys-

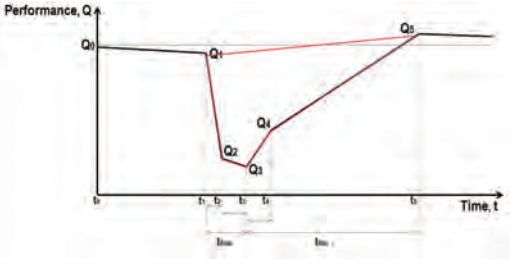


Figure 1. A schematically drawn "resilience triangle" illustrating the performance of a system over the course of a crisis event.

tem is able to absorb the initial shock (minimizing the impact) (Q_1 to Q_2) and respond to it (Q_2 to Q_3). The size of the triangle (marked in red) further depends on the system's capacity to recover (Q_3 to Q_5). The total performance loss of the system over the duration of the initial incident and recovery is however, partially an indirect result of how well the system has earlier targeted the measures of risk assessment, prevention and preparedness. The gain in performance over time after the performance loss reflects how well the system recovers from the shock and eventually learns from the shock. The learning phase enables and could likely lead to increased system functionality compared to before the shock, due to the renewal in restoration of the system, but also from the learning experience itself.

A case study of the Municipal Water Network of Barreiro, Portugal, is used to illustrate the development of the performance measuers needed to create the resilience triangles and is subject to a pilot implementation of a CI resilience management framework developed within the IMPROVER (Improved risk evaluation and implementation of resilience concepts to critical infrastructure) project, funded under the Horizon 2020 program. The implementation explicitly investigates public tolerance levels through a survey. The overall performance is well described and categorized in terms of quality (suitability for drinking and cooking), quantity (the amount of water accessible to the public) and delivery (the amount of water delivered to the tap) of water. In order to quantify these, the importance of understanding what is being assessed, to what hazard and for which end-user became evident. After describing the Municipal Water Network of Barreiro, Portugal the paper describes the development of performance measures, followed by a first look at the preliminary results of the resilience assessment through the resilience triangles.

2 BACKGROUND ON THE BARREIRO LIVING LAB CASE STUDY

The city of Barreiro is part of the Lisbon metropolitan area, located on the south bank of the Tagus River estuary, about 40 km from the city of Lisbon. It has a population of almost 80 000 people with an area of 36.41 km^2. Ferries and two bridges connect Barreiro to Lisbon.

2.1 *The municipal water network*

The Barreiro Municipal Water Network delivers potable water to the municipality of Barreiro and serves all its inhabitants and industries. It has an annual water flow of 6,200,000 meters cubed. The drinking water comes exclusively from an underground aquifer (Barreiro Municipality, 2009). The water supply system in Barreiro is constructed by 11 licensed ground-water intakes from a semi-confined aquifer, 7 reservoirs for treated water storage, 3 pumping stations, 5 blowers, 16.1 km of main, 263 km of ordinary pipes and a considerable amount of service connector pipes. The pipe system is made out of mostly fibre cement (FC) and some parts of PVC (polyvinyl chloride) and PE (Polyethylene).

The Barreiro municipality is divided into 7 different pressure zones. This study is limited to look at the resilience of 3 of these water supply zones that together serve about 60% of the population. Within the three pressure zones considered for this study there are three water storage units, including one high storage tank: Alto da Paiva high tank (SNZM Reservoir), and 2 semi-buried tanks: Alto da Paivo (SNZB1 Reservoir) and Sete Portais (SNZB2 Reservoir). The reservoirs have a reserve capacity of 12 750 m^3, which supplies the population for about 24 hours based on volume.

Several hazards may influence the water network in Barreiro including earthquakes (leading to sever ground shaking or liquefaction), droughts and heatwaves. A historical example can be found in 1969, when Barreiro's water network endured moderate damage due to a 6.8 Magnitude earthquake event which led to the unavailability of potable water for 24 hours. More recently, in 2012, rice and cereal agriculture in Setúbal were affected by a water shortage (Ioannou et al., 2016).

3 DEVELOPMENT OF THE METHODLOGY

3.1 *Interview with living lab*

The first step in developing a methodology to measure resilience was to interview the living lab. Semi-structured interviews were held with employees from the Municipal Water Network in order to further the understanding of the system function as well as how they would act in the case of a crisis. Specifically, questions were posed in regards to the technical details of the system, recovery time estimates, and emergency/contingency plans.

3.1.1 *Results*

The Barreiro Municipal Water Network operators provided models of their different pressure zones where all the details of pipes, valves, reservoirs and tanks are stored. The model can solve the hydraulic flow through the whole system using the EPANET freeware (Rossman, 2000), provided by the United States Environmental Protection Agency. The operators of Barreiro especially pointed out two

assets that are crucial for the overall supply of water in the system: a supported, semi-buried reservoir and one critical pipe, the General Distributor Conduit (DN350). The supported reservoir of Alto da Paiva supplies the zones, Zona Baixa 1 (SNZB1) and Zona Media (SNZM), while the General Distributor Conduit DN 350 mm in fiber cement supplies the Zona Baixa 2 (SNZB2). Figure 2 shows these assets including the water network in the study site.

When discussing recovery times and actions, the operators informed us that if there is a water outage longer than 24 hours, regardless of the cause, they intend to use two water tanks of 80 m^3 capacity provided by the Barreiro Firefighters to provide water to the public. However, the water will need to be boiled. They estimate that within 30 minutes they would be able to make the tanks available to the public if using their own water sources. However, if the crisis was larger and water would have to come from neighboring municipalities or elsewhere, it could take a few hours, depending on road accessibility. They also have the possibility to request assistance to the district civil protection command and to be able to have a collaboration between the firefighting corporations of the district, a total of 25 teams with about 1500 to 2000 m^3 (estimated values) of total water supply capacity. However, these resources are pending other duties for the firefighting corporations. If an earthquake also affects the neighboring municipalities, the assistance could come from either the north bank of the Tagus River, Lisbon, or even further south, using the same means. They are currently in the process of developing their Water Safety Plan that will include both emergency and contingency plans.

Based on the hazard assessment and interest of the operators, a scenario of liquefaction following a severe earthquake is considered for this study.

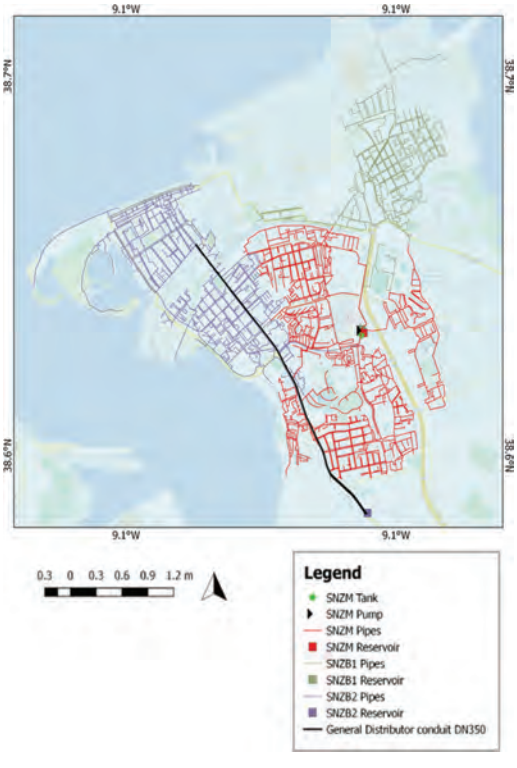


Figure 2. Barreiro Study Site Pipe network and Reservoir.

3.2 *Performance measures*

In order to define the performance measures, the first step was to define the normal performance. For this case, normal performance was considered as the normal domestic water use, covering consumption (drinking and cooking), hygiene (both personal and domestic cleanliness) and amenity use (i.e. watering plants, washing bikes) as listed by WHO (Howard & Bartram, 2003). In Barreiro the average water consumption to cover this is about 200 L/person/day. For survival and to avoid an outbreak due to lack of sanitation, the lowest acceptable quantity of water for survival is approximately 20 L/person/day (Cousins, 2013; Kameda, 2000; Mowll, 2012). In this study we have assumed that this is the minimum requirement for the first 3 days and that the requirement thereafter is increased to 50 litres per person per day, on account of the WHO recommendations.

A quantitative assessment of the performance of the water system was in this study measured as the percentage of the population that receives the services with the same performance as of a "normal day" before the earthquake for each performance measure. A similar study on the Los Angeles Water Service Restoration Following the 1994 Northridge Earthquake (Davis et al., 2012) defines five performance measures of the system: quality, quantity, delivery, firefighting and functionality. While the first three are measures directly impacting the public, functionality describes how the system performs its function in terms of efficiency, durability, sustainability and economics. This measure is not dealt with in this study. The availability of extinguishing water for firefighting is assumed covered by the abundant access to water around the municipality by the Targus River and is not covered here. Thus, the three performance measures in this study are:

1. Water delivery: Percentage of the population served by the pipe system through water on tap

(the water delivered may not meet the quality or quantity requirements).
2. Water quality: Percentage of population that have access to water at drinkable standards (not needing to boil the water).
3. Water quantity: Litres of water available per person per day.

3.2.1 *Estimating the performance measures*

First, an estimated time for service restoration is needed. To do so, the repair time was divided into two groups: emergency response time and recovery time. The emergency response time covers the period in which search and rescue is highly prioritized. This also includes work for road accessibility etc. Recovery time covers the period from when the repair starts until the entire network is repaired. These two times overlap and this is dealt with according to prioritizations inspired by the Oregon resilience plan (OSSPAC, 2013) and from literature on experiences and lessons learned from previous earthquakes (Bragado, 2016; EERI, 2007; Eidinger & Davis, 2012; Pedroso et al., 2013; Mowll, 2012). Based on this, a simple spreadsheet model is built which can estimate the emergency response time. The recovery time is estimated from simulations of perturbed EPANET models, as well as historical data for repair times for different pipe materials. These combined methods allow us to measure the water delivery and water quantity. The water quantity is defined as the total volume of water delivered on tap or to community service points for the public to carry home. This water is delivered either through tanks or, at later stages, also through pipes connected to these service points. At even later stages, the quantity also includes water being delivered to homes through the water distribution pipe system. Water delivery is calculated through the total amount of water that can be extracted from the network through nodes connected to households or other buildings as the system is repaired. Nodes that are somehow connected to a water source can often deliver the capacity prior to the incident. We assume that the "normal" delivery at each node is proportional to the number of people served at this point. Thus the relative delivery at all nodes to the delivery prior to incident is assumed equal to the fraction of people being served by the connected nodes. Based on the interview with the operators, the water quality indicator will need to remain at 0% of normal performance (i.e. water needs boiling before consumption but is suitable for washing) until the system has been repaired and thoroughly flushed.

3.3 *Public expectations*

Previous research on expectations/satisfaction of water service disruptions have shown that attitudes are not very strongly held on this subject matter (Vloerbergh et al., 2007). Most previous research does not deal with expectations/tolerances/satisfaction during times of crisis, but instead with "normal times" or planned works (Speers et al., 2002). As mentioned above, previous work within the IMPROVER project has found that the general public, end users of CI services, appear to have reasonable expectations of CI operators in crisis times (Petersen et al., 2017). As such, we suggest using their declared coping capacity as a target for CI resilience. In order to do so the public perception must be discovered in a way that is comparable to the technical performance measures of the service. To our knowledge, no studies comparing expectations to performance measures currently exist.

A common way to evaluate public expectations/user satisfaction for water operators is using a "willingness to pay" model. However in a disaster situation, this is not an appropriate measurement method as people need to have access to water, being a basic human need, regardless of the cost. As such, we propose to use a questionnaire in order to determine the public's coping capacity. Indeed, people themselves have been found to be good judges of their ability to deal with disturbance and change (Nguyen et al., 2013) and the idea of people being able to accurately judge themselves has been used in other domains, ranging from psychology to well-being and climate change adaptation (Jones & Tanner, 2015). Furthermore, previous research into water customer preferences has also used questionnaires to establish coping capacity in times of water shortage (Vloerbergh et al, 2007; Speers et al., 2002).

3.3.1 *Development of the survey*

The survey was developed with a view to real world performance capabilities of operators that were established in the interview. The performance measures were not asked about directly, meaning we did not use the terms delivery, quality or quantity. Instead, situational, laymen's terms were used to increase understanding of the survey (for more on the importance of the understandability of questionnaires, see OECD, 2013). The following paragraphs describe the questionnaire.

The respondent is presented with the following scenario: "*Imagine that a high magnitude earthquake occurs, where a large part of the population is left without access to potable water on tap without any previous warning*". Next, the respondent is reminded of the various needs for water following an earthquake (drinking, hygiene (showering, flushing toilet), cooking and cleaning). Then, the same measures as used for technical performance are used to create questions to find public expectations/tolerance levels for reduced service. This took

the form of three questions. The first question deals with the tolerance for recovery times of delivery by asking how long the respondent is willing to tolerate water being delivered via tanks (*Water would have to be delivered in tanks. How long would you tolerate these conditions?*). The next question addresses the recovery time tolerance level for quality by asking how long the respondent would be willing to boil water before drinking (*How long would you tolerate having to boil the water before drinking it?*). Lastly, to address the quantity indicator, respondents are asked how long they will tolerate having only X amount of water (going from 10 L to 100 L) per person per day (*How long will you tolerate having only the following amount of water per person per day?*). Respondents were also presented with some examples of water consumption (washing dishes by hand uses about 25 L, taking a 5-minute shower uses about 35 L). The time frame proposed to the respondent comes from not only historic examples and satisfaction surveys, but also realistic time frames for operators given in the interview. The proposed times are <12 h; 12–24 h; 1–2 days; 3–4 days; 5–6 days; 1 week; 2 weeks; 2 weeks – 1 month; More than 1 month. The questionnaire also asked about respondent's demographics and satisfaction levels with the current water service.

3.3.2 *Dissemination plan*

The questionnaire was designed as a telephone questionnaire. A representative sample of 1,005 (with a confidence level of 95.5% and an error value of +/–3%) based on age and gender of Barreiro residents was interviewed. The questionnaire was carried out by Pitagórica – Investigação e Estudos de Mercado SA. Respondents were interviewed in Portuguese. A few face-to-face interviews were also held, as there were issues finding enough young adults to answer the questionnaire via the telephone. Data collection was from 11 October 2017 to 5 November 2017.

4 APPLICATION OF THE METHODOLOGY

4.1 *Technical resilience analysis: performance measures resilience triangles*

The performance is briefly described below. The details of the models and assumptions involved will be further described in future publications of the same authors. Here, we show the result of one event only, a magnitude 7, peak ground acceleration 0.21–0.36 g shaking, which has a probability of 4–10% in 50 years for the specific location. It should be noted that the results shown here are preliminary and the final results will be subject to review throughout the complete pilot implementation. They serve, nevertheless, as a good example of how to compare estimated performance and public expectations.

The quantity of water during the response phase, in which resources are focused on rescue, clearing roads and electricity, is left to what is in tanks and what is stored in people's homes and stores. It is conservatively set to 5 L/day/person, assuming that all water in the storage tanks can be placed at community service points to reach the full population. After 24 hours water will be transported from adjacent regions via eight tank trucks of 40 m^3 to service points (~40 L/person). Since it is difficult to carry more than 20 L of water long distances, also considering that people have to carry for others (children, elderly etc.), a reasonable estimation of water availability when it's not delivered on tap, based on previous experience, is 20 L per person for the first days and twice that as needs become more pressing. When the network is being repaired the performance of the system is modelled using modifications and repeatedly running EPANET. By using fragilities of ductile and brittle pipes based on historical seismic events (Eidinger et al, 2001; Shih & Chang, 2006) we assigned a fragility of each pipe to liquefaction following earthquakes, describing the number of breaks in a pipe of certain material and size per unit length. The probability of a pipe breaking due to an event is given as, where is the breaks per unit length of the pipe and its length.

The pipes are then removed from the EPANET model using random numbers in comparison to the probability of failure. The flow in nodes connected to removed pipes are set to zero. Large negative pressures at nodes as a result of the perturbed network also results in the removal of nodes and any connected pipes. Running the model for the pressure and flow in the new system yields e.g. the demand (water extracted) at each node, which is summed up over all nodes. This is interpreted as a measure of the delivery of water. The pipe model just takes into account mains and not service lines from the street to the tap, which usually are the responsibility of the property owner.

Starting with the pipes of high priority (serving the hospital and community service points) and then according to their distance to reservoirs and tanks, the pipes are then replaced in the model, running it repeatedly to investigate the performance of the network until the network is back at original shape (note that other strategies of repair are being considered in the ongoing full study). The time to replace a pipe is treated as a function of its material and size based on empirical data (Porter, 2016) as well as the interviews with the operators. All in all several hundred of models are run controlled from Matlab. The result presented here in Figure 3 is very preliminary.

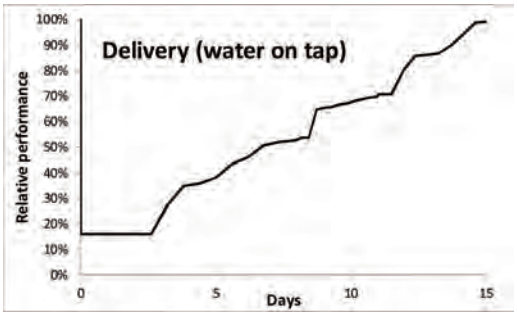


Figure 3. A simulation of the delivery to tap after the event using analytical models and EPANET modelling using fragilities of each pipe.

For the magnitude of earthquake considered here the response phase takes 2 days after which the repairs on the network start. After 48 hours some parts of the network distributed water will add to the water delivered at service points. As the network is repaired and additional nodes become connected, water is not uniformly distributed throughout the municipality. Instead water is delivered with close to normal service to a limited amount of people with access to it. There is likely a higher quantity capacity in the system than can be delivered.

It will be recommended that all water collected from service points or extracted from tanks should be boiled before consumption and, as such, the quality will not meet the drinkable water standards until the complete network is properly flushed.

4.2 *Evaluation criteria: tolerance triangles*

Preliminary results from the survey have been used to create tolerance triangles. Quantity is described as an average of the amount of water accepted weighted by the lowest fraction of population tolerating this (people tolerating 20 L also tolerate 40 L), Figure 4. Quality and delivery are defined as percentage of the population tolerating the action (boiling and collecting from service point, respectively). The expectations resilience triangles are thereafter created by the accumulated percentages of respondents willing to tolerate a given time frame.

4.3 *Technical resilience evaluation: performance vs tolerances*

To determine if the operator currently meets public expectations for service restoration time the tolerance and performance triangles are compared. The comparison of delivery tolerance against performance is shown in Figure 5.

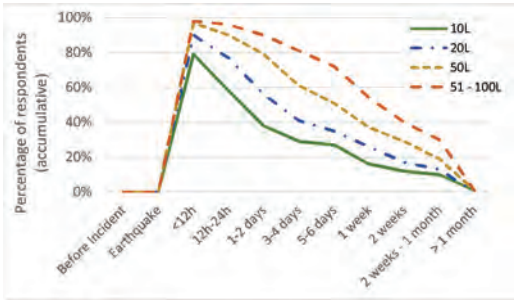


Figure 4. Percentage of respondent that tolerate a certain quantity of water per day and person after an earthquake.

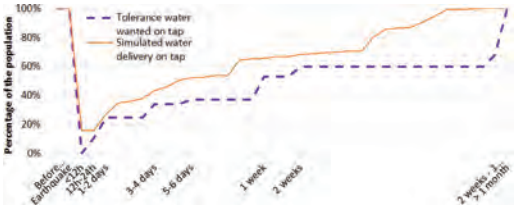


Figure 5. Resilience evaluation for delivery.

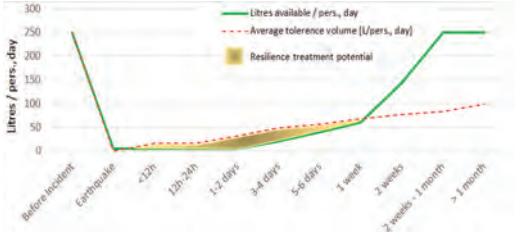


Figure 6. Resilience evaluation of quality.

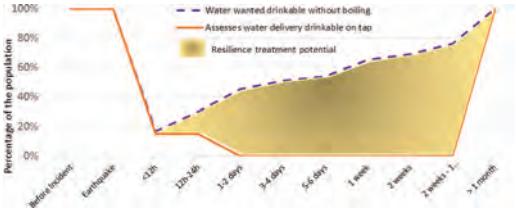


Figure 7. Resilience evaluation for quality.

The weighted average of the tolerance quantity at each moment in time is compared to the total quantities from our performance models, Figure 6. Advice to boil the water for everyone is considered the zero level for quality. The advice will be told until the whole network is restored. Clearly that does not meet the tolerances of the public, Figure 7.

5 DISCUSSION

5.1 *Meaning of the resilience evaluation*

The expectations of the public seem to be nearly met by the technical performance in terms of quantity, Figure 6. At the beginning the tolerance is somewhat below the performance, however, it seems that the endurance of the citizens is longer than the time needed to restore full capacity in terms of quantity. A month after incident the tolerances are for much lower quantities than expected performance. Also the simulated performance level during the first days is very conservative. This is because it is very difficult to assess road accessibility and to estimate the stored water capacity. We chose to ignore larger water storages in stores or homes. These can be significant and the resilience of the society might be much higher than the resilience of the specific infrastructure itself. Nevertheless, facing the mismatching identified, one way to close the resilience treatment potential could be to lower the expectations of the public during the initial stage and to increase the capacity through communication campaigns highlighting the need to store water for emergency situations.

Furthermore for delivery the performance is within a good range of the tolerances; the performance seems better than the tolerance for the entire time period studied. This is also in line with the previous findings in this project (Petersen et al., 2017).

One aspect where the public tolerances do not meet the performance is the quality. Apparently, boiling water for more than two days is too long of a time as it appears to constitute a major impact on many people's lives. The treatment to close this gap could be to use antiseptic tanks, which there is access to, but structures and practices of the usage are lacking. Also, quality of tap delivered water could possibly be guaranteed earlier by flushing separate areas of the system once they are repaired. Public campaigns could again be useful for changing the public's perception of and the need for boiling water.

5.2 *Performance measures chosen based on comparability of results*

In developing these performance measures the importance of what is being assessed, for which end-user and to which hazard became evident. The need of water could e.g. be larger in a heat wave and dry spell compared to an earthquake. The perception of the service provided to the public is not always measurable in technical evaluation of the system. From a personal perspective, it can be difficult to know how many liters of water one can tolerate to get by with. It is even harder to relate this to pressure or head in the system, parameters that are actually measured. Indeed, general performance indicators do not take into account the service that is provided to the public by a given infrastructure. Thus, by focusing on the normal performance of the service, the tolerance levels of the public to the change of service are able to be taken into account. Uncertainties in the modelling are found in the estimates of accessible resources and time efficiency of restorative capacity for the damage itself is done based on likelihood of exceeding a break probability. A natural way to solve this is to run the model in a probabilistic methodology to identify the spread in the result based on the random numbers. This is currently being done as this article is being written. One aspect is that the performance is here measured as a scalar number. It might not be the people of least tolerance who gets access to water first. Instead, even if the tolerance and performance curves matched perfectly a portion of the population would consider the system to underperform whereas another portion would consider the opposite. In addition, the tolerances are naturally a function of the magnitude of the event, something which is very difficult to identify in a survey. Previous findings show, however, that communication and information of the situation and the reasons for performance drops is vital to keep the public content during crisis events. Lastly, while the performance measures suggested could be used for almost any hazard, a scenario is necessary to be able to measure them effectively (for both technical performance and the survey).

5.3 *Survey limitations*

When responding to questionnaire surveys, people often respond by providing snap judgments based on available information and may be influenced by emotional or contextual factors (Schwarz & Stack, 1999). Also, question wording could affect stated tolerance levels, as research has demonstrated that when asked if they care about a given issue, people state concern for issues that do not exist (Herrmann et al., 1994). Further, respondents may choose to answer in their own self-interest, claiming to tolerate less so as to not give the CI operators an excuse to perform any lower than absolutely necessary. The opposite may be true, reporting that they are willing to tolerate more than they actually could handle in order to appear heroic. This is furthered by the fact that research has also shown that disaster victims rarely passively wait around for someone else to take care of their needs (Quarantelli, 1998), and having high expectations towards CI operators to act in a disaster may indicate a gap between expectations and the ability of citizens in responding to crisis situations. Lastly, expectations have been

found to be influenced by demographic factors, previous disaster experience and information provision (Petersen et al., 2016). However, with purposeful survey design and adequate sampling methods such as the one used here, many of these limitations are reduced and even overcome (Jones & Tanner, 2015).

5.4 *Success of the method*

These preliminary results show that public tolerance and technical performance of a critical infrastructure can be evaluated and compared to each other in the case of a drinking water distribution network. It seems from this case study that reasonable comparable performance measures have been found. The operators of Barreiro have expressed positive feedback to the methodology and think that the results provide relevant knowledge. However, a more descriptive resilience treatment where strategies of closing the gaps between expected performance and public tolerance are formed is currently work in progress during the writing of this article. The success of the method, pending on the usefulness to the operators, is not yet entirely apparent. However, this work shows that tolerance and performance can be compared if the survey asks the right questions and the right modelling work is conducted.

6 CONCLUSIONS

This paper describes the preliminary results of evaluating the expected performance of a critical infrastructure, in the form of a water distribution network, to the public tolerances of the service that the infrastructure provides. It is suggested that this comparison is a valuable measure reference of evaluating the performance compared to an otherwise arbitrary scale of performance. It is shown that comparisons can be done and that the chosen performance measures must be clearly defined prior to evaluating both performance and tolerances. When doing this, the different aspects of the service provided must be considered as the perception of the service usually is multifaceted and not necessarily directly linked to the scalar which might be the most straightforward to assess technically. The results of the resilience analysis that are presented here are based upon an initial study of the Barreiro municipality's potable water network. As such there are a number of simplifying assumptions and changes which are made which mean that the performance shown is not the true performance of the system (such as it being based on only one earthquake scenario). Nevertheless the results are useful as an illustration of the resilience evaluation methodology shown.

ACKNOWLEDGEMENTS

The IMPROVER project is funded from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 653390. Özüm Durgun and Moufid Salaimeh are acknowledged for help in developing the models.

REFERENCES

Barreiro Municipality. (2009). Gota a gota: Agua no barreiro: um tesouro Escondido. A água e o saneamento no concelho do Barreiro.

Bragado, A.D.D. (2016). Downtime Estimation Of Lifelines After An Earthquake (Masters Theses). https://upcommons.upc.edu/bitstream/handle/2117/88180/ADiaz-Delgado_Research.pdf

Bruneau, M. et al. (2003). A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities. *Earthquake Spectra: November 2003, Vol. 19, No. 4, pp. 733–752*. doi: 10.1193/1.1623497

Cousins, WJ. (2013). Wellington without water: impacts of large earthquakes. *GNS Science Report 2012/30*.

Davis, C. A. et al. (2012). Case study: Los Angeles water services restoration following the 1994 Northridge earthquake. *Proceedings of 15th world conference earthquake engineering, Lisbon*.

EERI. 2007. Preliminary Observations on the Niigata-Chuetsu Oki, Japan, Earthquake of July 16, 2007. *Learning From Earthquakes Program, Earthquake Engineering Research Institute EERI*.

Eidinger, J. et al. (2001). Seismic fragility formulations for water systems, Part 1: Guideline. *American Lifelines Alliance*.

Eidinger, J. & Davis, C. A. (2012). Recent Earthquakes: Implications for U.S. Water Utilities. *Water Research Foundation*.

Herrmann R. et al. (1994). Words matter. *California Agriculture*, *Vol. 58, Number 2*.

Howard, G. & Bartram, J. (2003). Domestic water quantity, service level and health. *World Health Organization*.

Ioannou, I. et al. (2016). Methodology for identifying hazard scenarios to assess the resilience of critical infrastructure. *IMPROVER Deliverable 2.1*.

Jones, L. & Tanner, T. (2015). Measuring 'subjective resilience' using people's perceptions to quantify household resilience. Overseas Development Institute.

Kameda, H. (2000). Engineering Management of Lifeline Systems Under Earthquake Risk. *12th World Conference on Eartquake Engineering, Auckland, New Zealand*. New Zealand Society for Earthquake Engineering.

Mowll, R. (2012). Lifeline utilities restoration times for metropolitan Wellington following a Wellington Fault earthquake. *Report to the Wellington CDEM Group Joint Committee from the Wellington Life-lines Group*. November 2012. 63p.

Nguyen, K. V., & James, H. J. (2013). Measuring household resilience to floods: A case study in the Vietnamese Mekong river del-ta. *Ecology and Society* 18(3): 13. 2013.

OECD. (2013). OECD Guidelines on Measuring Subjective Well-being. OECD Publishing.

Oregon Seismic Safety Policy Advisory Commission (OSSPAC). (2013). The Oregon Resilience Plan: Reducing Risk and Improving Recovery for the Next Cascadia Earthquake and Tsunami. *Report to the 77th Legistlative Assembly. Salem, OR*. http://www.oregon.gov/oem/Documents/Oregon_Resilience_Plan_Final.pdf.

Pedroso, F. et al. (2013). Input Paper: Post-Disaster Challenges and Opportunities: Lessons From the 2011 Christchurch Earthquake and Great Eastern Japan Earthquake and Tsunami. *Brazil, World Bank, Japan, Kyoto University, New Zealand: Resilient Organisations*.

Petersen, L. et al. (2016). Social resilience criteria for critical infrastructures during crises. IMPROVER project, Deliverable 4.1, European Commission H2020.

Peteresen, L. et al. (2017). Public tolerance levels of transportation resilience: a focus on the Oresund region within the IMPROVER Project. *CRITIS 2017*.

Porter, K. (2016). Damage and Restoration of Water Supply Systems in an Earthquake Sequence. Structural Engineering and Structural Mechanics Report Series 16–02, University of Colorado Boulder, 116 p., http://www.colorado.edu/ceae/node/1092/attachment

Quarantelli E.L. (1998). Major criteria for judging disaster planning and managing and their applicability in developing societies. *International Seminar on the Quality of Life and Environmental Risks* held in Rio di Janeiro, Brazil.

Rossman, L. A. (2000). EPANET 2 *Users manual. United States Environmental Protection Agency. EPA/600/R-00/05*.

Schwarz, N. & Strack, F., (1999). Reports of subjective well-being: Judgemental processes and their methodological implications. in: Jones, L. and Tanner, T. 2015. Measuring 'subjective resilience' using people's perceptions to quantify house-hold resilience. Working paper 423. *Overseas Development Institute, London.*

Shih, B.J. & Chang, C.H. (2006). Damage Survey of Water Supply Systems and Fragility Curve of PVC Water Pipelines in the Chi–Chi Taiwan Earthquake. *Nat Hazards 37: 71*.

Speers, A. et al. (2002). Determining Customer Service Levels - Development of a Methodology Overarching Report. *CSIRO*.

Vloerbergh, I. et al. (2007). Assessing consumer preferences for drinking water services - Methods for Water Utilities. *Techneau D6.2.2*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Lessons from the application of a resilience engineering based assessment method to evaluate the resilience of a train departure and arrival management system

E. Rigaud
MINES ParisTech, PSL—Research University, CRC, Sophia—Antipolis, France

C. Neveu & S.D. Langa
SNCF, Paris, France

ABSTRACT: Resilience Engineering is an original approach on safety management considering the development of agent's adaptive capacities to the diversity of situations that can occur, as the main target of safety management practices. The Resilience Analysis Grid aims supporting the assessment of key capacities of a resilient organisation. A specific instance is developed for supporting railway safety management. It is composed of a set of key indicators and a methodology to collect and analyse information. Application of the prototype to study trains station resilience capacities demonstrate its potential to understand resilience capacities with deducting resilience and fragility factors.

1 INTRODUCTION

Resilience Engineering is considered by Borys, Else and Leggett (2009) as the fifth age of safety following an age of integration age (Glendon et al. 2006) where safety management aims integrating technical, human, managerial and cultural factors in risk management practices such as risk analysis, barriers management and accident analysis (Hale and Hovden 1998) (Hudson 2007). Resilience Engineering perspective on safety management considers that the dynamic of evolution of safety practices doesn't provide to workers capacities to cope with the complexity of their environment and that theoretical and methodological innovations are necessary endowing systems the requisite imagination to respond and overcome to the diversity of situation that can possibly occur (Adamski and Westrum 2003, Woods and Hollnagel 2006). Safety evolves reactively after each event questioning the relevance of models structuring safety theories, methods and tools. The new factor or element founded that support the explanation of the event and the failure of safety management system is theorized and integrated in safety management practices (procedures, indicators, etc.). Then safety management system tries to constraint system dynamic in order to minimize the occurrence of such factors. The development of "human error", "organisational failures", "safety culture" theories, method, tools and operational practices illustrates this dynamic. The complexity of a system is associated among other properties, to non-linearity and to difference between reality and artifice (Chandlers 2014). Consequently, the Resilience Engineering perspective on safety aims to change the main focus of safety management from risk prevention to workers adaptive capacity to respond and overcome unwanted situations. This perspective considers that the absence of unwanted consequences is not caused by the efficiency of risk barriers but by the capacity of the system to be in control despite the variability and the complexity of situations and the lack of time, knowledge, competence or resources (Hollnagel and Woods 2006). The Resilience Engineering perspective on safety management considers resilience of a system as it's "ability to recognize and adapt to handle unanticipated perturbations that call into question the model of competence, and demand a shift of processes, strategies and coordination" (Woods 2006) and "the intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances, so that it can sustain required operations under both expected and unexpected situations" (Hollnagel 2011).

Aims of this paper are to firstly present indicators defined to assess the resilience of an organisation, then associated method for assessing resilience performance and finally results of its application to a train station.

2 PERFORMANCE INDICATORS

According to the resilience engineering perspective on safety management, organisation can be con-

sidered as resilient if they are able: 1) to respond and overcome to the diversity of situations that may arise; 2) to monitor that which changes, or may change in the near term that it will require a response; 3) to learn from both positive and negative experience of the past; 4) to anticipate development, threats, and opportunities further into the future (Hollnagel 2011). A first set of indicators, the Resilience Analysis Grid, has been proposed to support the assessment of these four dimensions (Hollnagel 2011). Starting from resilience engineering concepts and models and the Resilience Analysis Grid, a process of contextualisation to the railway domain has been conducted. Results from this process are nine indicators to be used to assess the resilience of railway systems. For each indicator, four rules are defined for supporting the evaluation. A system is evaluated with five stars if the four rules are true, four stars if the first three rules are true, etc.

2.1 *Capacity to respond and overcome to the diversity of situations that may arise*

Two indicators are related to the capacity to respond and overcome to the diversity of situations that may arise. The first one is related to routine situation and the second situations that are considered as abnormal. The definition of the indicators considers tasks related to the adaptive process (detect, recognize, decide to change behaviour, define behaviour, mobilise resources), the existence or not of procedures or good practices associated to the situation, the difference between the context of action (competence, knowledge, resources and time) available and the context required to respond to the situation, and the different dimensions of performance of the system (quality, reliability, safety, security, sustainability, etc.).

1. The first indicator is related to the capacity of operational agents to adjust their procedural and/or methodological framework or to be creative in order to carry out their normal activity in spite of the variability of their environment while respecting the temporal, economic and activity specific performance criteria. The four rules associated to the indicator are:
2. Agents know their work and associated performance criteria
3. They have the skills or know the procedures to follow and have the resources, time and information to carry out their work in accordance with the different performance criteria.
4. If they lack skills, resources, time or information they are able to be creative in carrying out their work according to performance criteria.
5. If the situation changes and the procedural framework is no longer applicable, they are able to be creative enough to carry out their work in accordance with performance criteria and have the necessary margins of maneuver.

The second indicator is related to the capacity of operational agents to adjust their normative and/or methodological framework or to be creative in order to face and overcome the occurrence of an urgent and/or unexpected situation, anticipated or not, while respecting the temporal, economic and activity specific performance criteria. The four rules associated to the indicator are

1. Agents are aware of the abnormal situations, the behavior to adopt when they occur, or what document to consult to know.
2. They have the skills, resources, time and information to respond to the situation in accordance with the different dimensions of performance.
3. If they lack skills, resources, time or information they are able to be creative in responding to the situation in accordance with the different dimensions of performance.
4. If the situation changes and the procedural framework is no longer applicable or there is no procedural framework, they are able to be creative in responding to the situation.

2.2 *Capacity to monitor that which changes, or may change in the near term that it will require a response*

Three indicators are related to the capacity to monitor that which changes, or may change in the near term that it will require a response. They are related to the capacity to evaluate and exploit retrospective safety performance, actual safety performance and prospective safety performance. The four rules associated to the three indicators are:

1. The system has indicators for measuring retrospective/actual/prospective performance
2. Retrospective/actual/prospective safety performance indicators are consistent with the system and are properly and regularly reviewed.
3. The nature, period, frequency of the measurement (qualitative or quantitative) of the indicators and the time between measurement and exploitation are correct
4. There is no conflict between safety performance indicators production performance indicators

2.3 *Capacity to learn from both positive and negative experience of the past*

Two indicators are related to the capacity learn from both positive and negative experience of the past. They are related to the capacity to acquire, disseminate and use experience in the occurrence of unwanted situations (incident, accident, etc.)

for the first one and of gained during the observation of daily operations for the second.

The four rules associated to the first indicator are:

1. The occurrence of an abnormal situation (non-compliance, incident, accident, disaster, etc.) is detected, listed, investigated and the results disseminated within the organization.
2. Criteria for identifying a situation to be investigated are clearly identified, shared and appropriate by the system. The direct and indirect causes sought and the reaction of the system are investigated.
3. Necessary skills for conducting the investigation are available, necessary information can be accessed and the study carried out independently of the stakeholders.
4. Lessons learned information are processed, capitalized and proactively transmitted to the other entities of the system.

The second indicator is evaluated with these four rules:

1. Audits are carried out to understand the real functioning of the system
2. Competences necessary to realise audits are sufficient
3. Audits results are used and capitalized by the station
4. Results are proactively transmitted to others stations

2.4 *Capacity to anticipate development, threats, and opportunities further into the future*

Two indicators are related to the capacity to learn from both positive and negative experience of the past. They are related to the capacity to identify, use and disseminate information about the consequences of a change of a component of the system for the first indicator and of a change in the environment of the system on safety performance for the second indicator.

The four rules associated to the first indicator are:

1. Anticipating the consequences of internal changes on safety is a dimension of the culture of the organisation.
2. The methodological approach for conducting the study of consequences of change is clearly formulated and based on adequate expertise
3. The anticipation of consequences of the change is carried out with sufficient time so that the consequences identified can be taken into consideration.
4. Outcomes relating to potential sources of threats or opportunities are shared within the organisation.

The second indicator is evaluated with these four rules:

1. Anticipating the consequences of external changes and of trends is a dimension of the culture of the organisation.
2. The methodological approach for conducting the study of consequences of external changes and trends is clearly formulated and based on adequate expertise
3. The anticipation of consequences of the external changes and trends is carried out with sufficient time so that the consequences identified can be taken into consideration.
4. Outcomes relating to potential sources of threats or opportunities are shared within the organisation.

3 METHODOLOGY

In order to assess and enhance the resilience of socio technical system resilience, a four phases method is proposed:

- Phase 1. Definition of the context of the diagnostic. System representative defines scope, schedule, working team and stakeholders of the diagnostic process.
- Phase 2. Performance assessment. Working team collects data necessary to evaluate performance indicators and write assessment report validated by stakeholders.
- Phase 3. Actions plan definition. Working team collects data necessary to define with stakeholder's actions plans aiming to improve gaps and preserve good practices.
- Phase 4. Conclusion of the study. Working team writes and presents final reports to system representatives.

In the perspective to experiment it, the method has been applied to the study the management of departures and arrivals of train processes of a train station. Accordingly, the following four paragraphs are outlining results of the application of these four phases in detail.

The study conducted aims studying both the resilience of the train management activities of the station and the relevance of the method. The scope will be operational activities dedicated to the management of departures and arrival of train in the station. Operational trains departure and arrival functions are relevant to study resilience because they involved technical agents, proximity managers, safety and production services, they are shape by schedule, procedures and time constraints, injuries may occur and they are sensitive to events occurring in the station and in the network.

The second phase of the process consists of the organisation of data collection. A set of workshops has been scheduled. First workshop was dedicated to a global presentation of the organisation of departure and arrival operational functions from schedules design to effective realisation of the tasks, observation and informal interview techniques were used. Results of this first workshop support the definition of a questionnaire. This questionnaire has been applied to interview twelve representative agents of the system (operation, technical manager, proximity manager, safety manager, head of the safety service). Data collected with interviews support the assessment of the different indicators. These results were discussed in a group workshop.

Next section is dedicated to the presentation of the results.

4 RESULTS

Data collected during interviews support the assessment of the different indicators that constitute the resilience performance. For each indicator rules, an evaluation is proposed and a set of fragility and resilience factors are proposed. Results for the indicator related to the capacity to respond and overcome to the diversity of situations that may arise are discussed in this section.

4.1 *The capacity to respond and overcome to the diversity of routine situations that may arise*

Results of the analysis of data collected related to the first indicator support the assessment of adaptive capacity of operational agents and of margin of manoeuver provided by the system to the variability of routines situations.

Agents know their work and associated performance criteria.

The assessment of the rule is based on the comparison of the description of the different tasks to be performed between both operational and managerial agents, procedures and observations.

According to the results of the analysis, the part of the rule dedicated to the knowledge of the activities is judged satisfactory. Agents seems to have a quite good perception of the reality of their tasks.

According to the results of the analysis, the part of the rule dedicated to the knowledge of the performance associated to the activities is judged satisfactory. Agents seems to know the different performance associated to their activities and aims satisfying safety and punctuality issues. Proximity managers aims finding the good trade-off between the satisfaction of objectives associated to their position and the management of the human dimension of their team.

They have the skills or know the procedures to follow and have the resources, time and information to carry out their work in accordance with the different performance criteria.

The assessment of the rule is based on the analysis of agent's testimonies about their work environment and their relation with competences, resources, procedures, time and information.

Operational work environment is judged maladjusted to work conditions such as winter and summer temperature, workplaces are not ameliorated and cleanness is judged insufficient. Moreover, they have the feeling that working in difficult condition is considered as normal and that budget is use for other issues than improving work conditions.

A period of empowerment is necessary for operational agents before they can overcome fear feelings to hazards associated to the tasks, during this one-two month period they can make mistakes. This time is longer for the traffic manager not caused by hazards buy by the complexity of the management tasks. Moreover, it is very difficult allocating training period to agent due to workforce issues and unplanned absence. These difficulties are compensated by initial training, the accompaniment of experimented agent during their first interventions, the vigilance of the hierarchy and mutual assistance between agents.

Procedures are learned in initial training by agents, monitoring achieved by proximity managers and their hierarchy helps check that they are fully known and understood. In case of doubt, agents ask their managers who are able to answer them. Agents used to mentally remind the procedure to be followed before performing critical and hazardous tasks. Agents consider that some tasks planed by procedure don't impact safety and consequently might don't apply them if they have constraints. This fragility is compensated by the monitoring by the hierarchy of the correct application of all the tasks of procedures. When a procedure is modified, each agent concerned have to attest he has considered the change. Due to the multiplication of changes, sometime minor, agent known that a change occur but don't take the time to study the change. Proximity managers compensate this issue with teaching and monitoring the correct understanding of changes that affect the work of the operational agents.

There is a difference between technical and human resources necessary to achieve goals in term of number of daily trains negotiated with stakeholders and resources available inducing recurring problems. There is a feeling that solutions provided to solve them by operational agents are not considered by the hierarchy. When tools are not available, exchanges are organized between agents. Agents absenteeism is important it is compensated by the

capacity of proximity managers and the hierarchy to perform operational tasks.

Work schedule is planned so that time is not a constraint for operational agents achieving tasks and time margins are planned between two work activities. The hierarchy considers that safety is the priority and don't put pressure on agents. A point of fragility is that agents enchain short period of activity and period of inactivity causing issues related to activation and concentration.

Information management is a major for all agents involve in the management of trains departures and arrivals. The use of information systems and informal communication networks helps manage constrains during both design and production phases. Information systems and communication protocol aims considering the constraints of each agents. Operational and proximity managers are continuously looking for information about the state of activities and of the network in order anticipating potential issues. For that they try to find information system and create a personal communication network. In case of lack of information operational agent don't hesitate to ask their managers who are responsible to find solution. Information systems may be saturated by the multiplication of message. Managers have to be careful delivering the good message to the right person at the right time in order to not induce perturbation in the system. Conflicts may arise caused by absence of answer or because of the tone or the type of message exchanged.

Related to all these positives and negatives factors the rules has been evaluated as medium.

If the situation changes and the procedural framework is no longer applicable, they are able to be creative enough to carry out their work in accordance with performance criteria and have the necessary margins of manoeuvre.

The assessment of the rule is based on the analysis of agent's testimonies about complicated and complex situations that can occur in routine situations.

Many situations related to incidents, delays, malfunctions require an adaptive response from agents and from the system. A routine situation may cause blockages due to its failure to be taken into account by procedures or by the definition of agent's roles. Some agents demonstrate initiative during such situations in order to insuring the train to be able to start on time and safely. They have margin of maneuver from hierarchy to take the time necessary to perform work safely for insuring safety even if it creates some delays for the train.

Based on all the fragility and resilience factors identified, the evaluation of the indicator is Acceptable for agents capacity to adapt to the variability of routines situation and the contribution of the organization as medium.

4.2 *The capacity to respond and overcome to the diversity of abnormal situations that may arise*

Results of the analysis of data collected related to the second indicator support the assessment of adaptive capacity of operational agents and of margin of manoeuver provided by the system to abnormal situations. A typology of abnormal situation has been firstly deduced and for each, factors of fragility and resilience has been identified.

Four abnormal situations have been studied:

- Situations caused by an increase of agents workload of agents resulting from the absence of two to three operational agents.
- Situations resulting from a safety incident occurring on one of the place of the station.
- Situations resulting from an incident occurring at the station or on the network which disrupts the production, where the management is under the responsibility of the train station.
- Situations resulting from an incident occurring at the station or on the network which disrupts the production, where a crisis management room is open under the responsibility of an authority external of the station.

Situations caused by an increase of agents workload of agents resulting from the absence of two to three operational agents.

When such a situation occurs, agents must perform the same tasks as in routine situations but in a different context. They may be required to perform more during the same service, to carry out them on numerous trains consecutively, to carry out them in a crisis atmosphere with many passengers seeking information to see in a hostile situation with aggressive and violent passengers. Agents adaptation is promoted with their acceptance that some days they must performed more activities than originally planned, with the ability of management to perform operational tasks and agent's knowledge on know how to prioritize their work with managing their production tasks as a priority and trying to find a solution for the passengers.

Situations resulting from a safety incident occurring on one of the place of the station.

When such a situation occurs, potential consequences of the psychological impact of the event on agent's activities in the short and medium term has to be considered as impact of the consequences of the absence of one or more agents for injuries, rest or suspension on the production capacity of the station. Ability of management supporting agents to overcome the event, the possible consequences (suspension, etc.) and to make the incident a source of awareness of risks not taken into account and of learning performing activities in compliance and safely.

Situations resulting from an incident occurring at the station or on the network which disrupts the production, where the management is under the responsibility of the train station.

When such a situation occurs, delays may affect the following services that need to be refined, Supervisors may have to leave the site and oversee operational activities to manage the crisis situation as a strain, the, some tasks that have to be carried out quickly in order to solve the situation may not be possible due to unavailability of technical resources or difficulties in setting up the stopover (access to storage areas, etc.). Culture of mutual assistance between services and agents in disturbed situations, the "Pride of the railway man", facilitate adaptation need to overcome such situations.

Situations resulting from an incident occurring at the station or on the network which disrupts the production, where a crisis management room is open under the responsibility of an authority external of the station.

When such a situation occurs, agents and managers have the feelings to not to be listened by the crisis management unit when they propose solutions appearing effective to them and that they undergo the decisions and their negative consequences on production and on customers. Moreover, there is a risk of chilliness caused by the fear of the perception and reaction of general headquarters. Culture of mutual assistance between services and agents in disturbed situations, the "Pride of the railway man", facilitate adaptation need to overcome such situations.

Based on all the fragility and resilience factors identified, the evaluation of the indicator is medium for agents capacity to adapt to the variability of abnormal situation and the contribution of the organization as medium.

5 DISCUSSION

Resilience can be perceived as a process, a set of properties, results of a dynamic of development, results of the response of a system to a situation, a combination of all the above. Resilience assessment is a complex process requiring to consider the system studied, a set of situations of adversity and a set of values.

As other equivalent work (Patriarca et al. 2017), a phase of translation of the initial RAG grid (Hollnagel 2011) is necessary to address concrete topics of the system studied.

The overall process is complicated to apply. Firstly, because topics studied are related to the system as he is and not as it should be, consequently conditions have to be favorable in order to have people accepting to describe their working conditions. Secondly, resilience performance is related to situation that don't occur frequently and for some situation they never occur. Consequently, it's difficult to collect relevant data allowing to predict how the system will respond to such a situation.

Results of the application are mainly qualitative because they aimed to promote dialog between actors of the organization and to support the definition of plan of actions for improving the system.

6 CONCLUSION

Application of the method to the train station demonstrates the potential of indicators and method to collect and analyse data on the resilience performance of an organisation. Fragility and resilience factors has been identified. If some topics where familiar with agents, others where more difficult to study, such as learning from normal situation, actual and prospective safety performance indicators or change management with operational agents.

Resilience factors have been identified, nevertheless the success of such factors are dependant of the availability and motivation of operational and managerial agents performing tasks that are not relevant to their duties, good cohesion of team and of vigilance on maintaining margins of manoeuver available.

Lessons learned from this experiment will structure the refinement of indicators and methodology in order to produce an operational method. The new method will be applied to other systems to continue validating its performance.

REFERENCES

Adamski, A. & Westrum, R. (2003). The Fine Art of Anticipating What Might Go Wrong. In: Erik Hollnagel (Ed.) Handbook of Cognitive Task Design, Lawrence Erlbaum Associates.

Woods, Nancy Leveson.

Woods D.D. & Hollnagel E. (2006). Prologue: Resilience Engineering Concepts. *In Resilience Engineering: Concepts and Precepts.* Ashgate. Erik Hollnagel, David D. Woods, Nancy Leveson.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Simulating the world described with the functional resonance analysis method

P. Smoczyński, A. Kadziński & A. Gill
Poznan University of Technology, Poznan, Poland

ABSTRACT: In the recent years there is a constant growth of interest in Resilience Engineering. According to its approach, accidents in complex socio-technical systems happen due to 'functional resonance' of many underdefined system functions. The resonance is often analysed and presented with help of Functional Resonance Analysis Method (FRAM). However, in many reported research, the analysis is limited to a separated group of functions and describes a process rather than a system as a whole. Contrary to this common practice, we assume that the processes are 'looped' and the system functioning—infinite. In the paper, we introduce a concept of a novel computer software which can demonstrate such systems described with FRAM models. The prepared software has been used for simulating a moving tram. Several simulations with different parameters have been compared to show the possible use of this software in further investigating of systems described with help of FRAM.

1 INTRODUCTION

The issues of safety have probably accompanied humankind since the very beginning of its existence. However, the need for a more formalised approach to this subject matter did not appear until the industrial revolution in the 18th century (Hollnagel 2014). Scientific research came even later—the essential book by Heinrich (Heinrich 1931) dedicated to work safety is worth mentioning here (Lundberg et al. 2009). In the period after World War II, reliability engineering developed, thanks to which many tools still in use today, such as the FMEA or FTA methods, were introduced.

Along with the noticeable improvement in the reliability characteristics of technical objects, undesirable events increasingly resulted from the errors of machine operators. In consequence, the models in use were supplemented with the so-called human factor. However, it turned out relatively quickly that an attempt at attributing responsibility solely to operators of technical objects does not bring the expected effect. It therefore became necessary to also include the organisations for which these operators worked in the analyses, which gave rise to safety management systems (Hollnagel 2014).

The results of scientific research translated into legal regulations concerning safety, in which departure from the focus on the technical details towards ways of making decisions and management could be observed since the 1970s (Hale et al. 1997). This trend was additionally reinforced by the results of studies on disasters from the 1970s and 1980s, such as the Piper Alpha oil platform disaster of 1988 (Paté-Cornell 1993), as well as the governments' will to withdraw from direct responsibility for the level of safety. Since the announcement of the results of the investigation of the causes of the Chernobyl disaster of 1986, also the need to enhance the culture of safety has been the topic of discussion (Wang & Liu 2012).

Experience in implementing safety management systems revealed a number of problems, however. Already in the 1990s, it was observed that people's behaviour changes under the influence of the ubiquitous procedures. Power (Power 1997) called this phenomenon the formation of an 'audit society', in which more emphasis is put on obtaining another certificate than on the actual effects of work. Forcing compliance with procedures does have certain positive effects, e.g. makes cooperation more predictable (Jeffcott et al. 2006), but it also leads to the marginalisation of the significance of the employees' knowledge and experience (Almklov et al. 2014).

The problems mentioned above are the reason that changes and additions to the approach to safety management are currently proposed. They concern changes in the manner of formulating safety procedures (Hale & Borys 2013) and moving from searching for the causes of undesirable events to searching for the causes of correct execution of system tasks (Hollnagel 2014). As a result, system resilience to unexpected changes in the manner of their operation is expected to increase.

One of the methods used for modelling of complex socio-technical systems in order to find the

resilience mechanisms is Functional Resonance Analysis Method (FRAM), proposed by Hollnagel (Hollnagel 2012). It has been applied in a variety of organisations and industries. The most recent papers describe e.g. sinter plant (Patriarca, Di Gravio, Costantino, et al. 2017), air traffic management system (Patriarca, Di Gravio, & Costantino 2017, Yang et al. 2017) or application in medical care (Pickup et al. 2017).

Most of the authors, however, tend to use the FRAM to analyse separated processes in the complex systems, from the beginning to the end of such a process. In our opinion, the FRAM can also be used to describe systems as a whole, where the processes are 'looped' and the system functioning—constant and infinite. The article aims to present an early version of a simulation software which can be used for this purpose, basing on a simple example of a moving tram.

In Section 2 we have presented the way how the approach to safety has changed over years to better and described the foundations of the FRAM. In Section 3 we have shown how these foundations are transformed into the proposed software. In Section 4, we show results of a tram ride simulation. The paper ends with conclusions in Section 5.

2 FUNCTIONAL RESONANCE ANALYSIS METHOD

Recently, a considerable increase in the complexity of systems, caused by the possibilities offered by new technologies and pressure to introduce changes as soon as possible and for as low a price as possible, could be observed. The consequence of this situation is often the lack of complete understanding of the processes occurring within the systems, which leads to their sudden failures. Since 2000, a rapid increase in the number of publications on this topic, dedicated to the issues of resilience engineering from the point of view of e.g. safety, system complexity, organisation or ecology, has been observed. The term 'resilience' is understood in four ways in these works (Woods 2015):

- As rebound – how a system rebounds from disrupting or traumatic events and returns to previous or normal activities.
- As robustness – expanding the set of disturbances the system is prepared for.
- As graceful extensibility – how a system extends performance or uses extra adaptive capacity in case of unpredictable events.
- As sustained adaptability – a policy making it possible to continue proper operation in the long term as external conditions change.

It is sometimes believed (Woods 2015) that only the latter two approaches are justified. Research of the sole process of returning to the initial condition is considered to be incomplete, as its course is determined by activities undertaken before the disturbance (which corresponds to increasing resilience as defined in no. 3 and 4). Preparing the system for the previously predicted events in turn corresponds to the phase of reacting to risk in the classic form of the hazard risk management process.

A different attempt at taking a complete look at the issue of resilience was made by Lundberg and Johansson, whose work (Lundberg & Johansson 2015) presents a resilience model including six functions: anticipation, monitoring, response, recovery, learning, and self-monitoring. An in-depth analysis of various definitions of resilience can also be found in (Adjetey-Bahun et al. 2016). In principle, resilience engineering is supposed to be as universal as possible (Haavik et al. 2016), and research on the topic has been carried out in many different areas of human activity, including medicine, aviation, and nuclear power stations (Le Coze 2016).

The Functional Resonance Analysis Method (FRAM), has been proposed by Hollnagel (Hollnagel 2012) for modelling how complex socio-technical system work. The method is based on four principles (Patriarca, Di Gravio, & Costantino 2017):

- Equivalence of failures and successes. Equivalence and successes come from the same origin, i.e. everyday work variability. This latter allows both things go right, working as they should and things go wrong.
- Principle of approximate adjustments. People as individuals or as a group and organizations adjust their everyday performance to match the partly intractable and underspecified working conditions of the large-scale socio-technical systems.
- Principle of emergence. It is not possible to identify the causes of any specific safety event. Many events appear to be emergent rather than resultant from a specific combination of fixed conditions. Some events emerge due to particular combination of time and space conditions, which could be transient, not leaving any traces.
- Functional resonance. The function resonance represents the detectable signal emerging from the unintended interaction of the everyday variability of multiple signals. This resonance is not completely stochastic, because the signals variability is not completely random but it is subject to certain regularities, i.e. recognizable short-cuts.

The principles constitute a new approach for understanding safety, called 'Safety-II' (Hollnagel 2014). In brief, the steps for safety system analysis using FRAM are (Patriarca, Di Gravio, & Costantino 2017):

– Identification and description of system's functions
– Identification of performance variability
– Aggregation of variability
– Management of variability.

The distinctive feature of the FRAM is the way how the functions are represented. The graphical form of this representation is shown in Figure 1.

As shown in Figure 1, each function is characterised through six aspects (Yang et al. 2017):

– Input (I): what the function transforms or processes or what starts the function
– Output (O): the result of the function, either an entity or a state change
– Preconditions (P): conditions that must exist before a function can be executed
– Resources (R): what the function needs when it is carried out (Execution Condition) or consumes to produce the Output
– Time (T): temporal constraints influencing the function (with regard to starting time, finishing time or duration)
– Control (C): how the function is controlled.

The definition of aspects is intuitive, but often too general for creating a consistent procedure of choosing them in particular circumstances. Anvarifar et al. (Anvarifar et al. 2017) notice that further work is required to test the applicability of the FRAM for detailed risk analysis in more complicated and data demanding case studies. Patriarca et al. (Patriarca, Bergström, et al. 2017) add that a comprehensive FRAM analysis might generate a representation, which is impressive in terms of its sheer number of functions and couplings, but hard to make interpretive sense for further analytical purposes. The solutions used to overcome this problem consists of decomposition schemes and various original computer software based on Monte Carlo simulation.

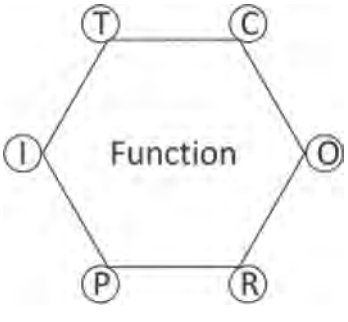


Figure 1. A hexagon characterising function in FRAM (Hollnagel 2016).

3 THE SIMULATION SOFTWARE

With our software presented in this paper we would like to simulate 'looped' systems described with FRAM models. It has been prepared using Microsoft Visual Studio Community 2017, which is available for free i.a. for academic purposes, enriched by a demo version of NMath library from CenterSpace. The program has been written with Visual Basic.NET programming language as a Console application and is running in the text mode only. It allows to focus more on the algorithms than on the visual part of the software. Following assumptions have been made for assuring consistency of FRAM-based models used as simulation basis:

– All the functions use information of generic type provided through their Input aspects to produce results of generic type, which are made available through the Output aspects
– The Control aspect is an input of generic type for information used by the function in order to minimise its own variability
– The Timing aspect is an input of date/time type that provides the earliest time when the function can be activated
– The Preconditions and Resources aspects are inputs of 'true' or 'false' type; the first aspect is checked whether it equals to 'true' at the activation of the function and the second aspect – constantly throughout its realisation.

The assumptions have been shown graphically in Figure 2.

With the assumptions (Fig. 2), the aspects have been effectively divided into two groups:

– Aspects responsible for determination if and when the function is performed: Timing, Preconditions and Resources
– Aspects responsible for how the function is performed: Input, Output and Control.

For the first group of aspects, it was possible to determine the type of variables used for communication between functions. The types correspond

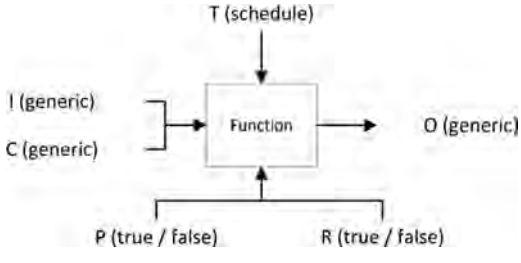


Figure 2. Interpretation of FRAM's aspects used in the proposed simulation software.

to the description in Fig. 2, i.e. Date for the Timing aspect and Boolean for the Precondition and Resources aspects. The second group of aspects is of generic type, as it depends on what the function actually does. The implementation of the 'inside' of the function has to be written manually directly into the program's code. However, in the future it is possible to introduce an interface similar to the LabView from National Instruments. It would allow to 'construct' the functions graphically from a set of predefined instructions (loops, conditions, etc.).

All the functions identified during the FRAM and present in the model are put by the software at the beginning of the simulation simultaneously in a 'stand-by' mode. Their actual activation depends on the respective Timing, Preconditions and Resources aspects an can be repeated throughout the simulation time. This approach makes the simulation never ends, just as the modelled systems never 'stop' their existence.

4 PERFORMANCE VARIABILITY OF A MOVING TRAM

For the testing purposes of the simulation software, we have decided to simulate a simple system of a tram moving through a city. The respective FRAM model has been presented in Figure 3.

The schemes as in Figure 3 can be created with help of specialised software, called FRAM Model Visualiser, which is available free of charge (Hollnagel 2016). Following two foreground functions have been considered in the model:

– *Exchange of passengers*, a function encapsulating the boarding activities between opening and closing doors
– *Move tram*, a function of keeping the vehicle running.

In addition, following background functions have been considered:

– *Prepare timetable*, giving the earliest time of moving away from a tram stop
– *Passenger flow*, a function that represents the passengers waiting for the tram on the tram stop
– *Traffic lights*, a technical function that periodically prevents the driver from leaving the tram stop
– *Produce electricity*, a technical function that supplies electricity which is inevitable for performing the function *Move tram*.

In the model, the Gauss distribution has been used for generating random numbers that describe the timing of traffic lights, producing energy, and the travel time between stops. The number of passengers increases with time and, when more boarding is needed than the schedule allows—the Exchange of passengers function uses warning signal to speed up the process. The warning signal device works with average efficacy of 1/*e*, i.e. in average one out of *e* signals lowers the remaining time of boarding by some value, that depends on the device type.

The sample simulation record has been shown in Figure 4. The simulation software allows to save the simulation record in form of a text file, which can be further elaborated with e.g. Microsoft Excel.

In the case study we have assumed that the departure time from stop $i + 1$ is 40 seconds later than the departure time from stop*i*. For the Gauss distributions determining the timing of background functions Produce electricity and Traffic lights (Fig. 3), we have used the following parameters:

– For Power = true: $\mu = 35$, $\sigma = 5$ [s],
– For Power = false: $\mu = 10$, $\sigma = 3$ [s],
– For Green light = true: $\mu = 4$, $\sigma = 4$ [s],
– For Green light = false: $\mu = 8$, $\sigma = 2$ [s].

In case that the time were negative, it is changed into zero. Additionally, the travel time is drawn

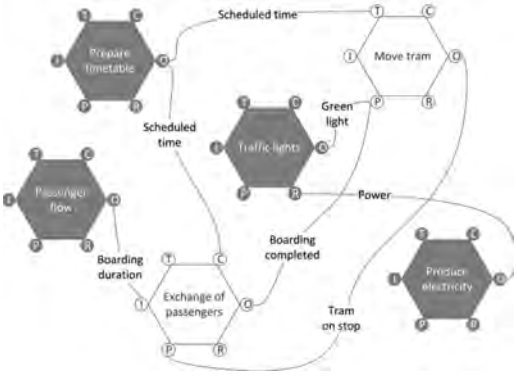


Figure 3. FRAM model of a tram moving through the city.

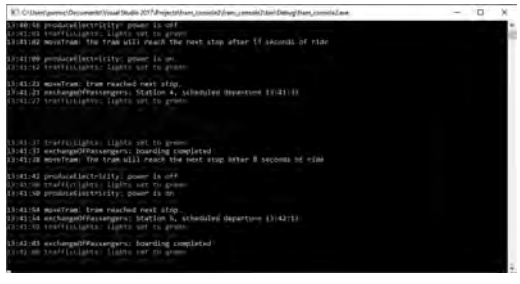


Figure 4. Sample record of a simulation performed in the software.

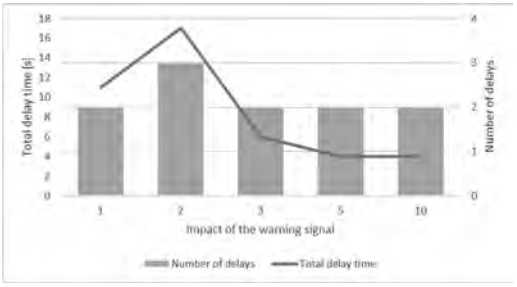


Figure 5. Total delay time and number of delays in respect to the impact of the warning signal.

according to the Gauss distribution of $\mu = 10$, $\sigma = 3$ [s]. The *Passenger flow* output is the boarding time that equals to the time between closing the doors at stop i and opening it at stop $i + 1$. Boarding time at the beginning of the simulation equals to 10 s.

The aim of the case study is to determine which one of five possible warning signal devices should be used in the tram after its renovation. The devices have the same efficacy $e = 0.4$, but differ in terms of their effectiveness in speeding up the boarding. Each time the signal is taken into consideration by the passengers (what happens with average probability 0.4), the signal lowers the remaining boarding time by 1, 2, 3, 5 or 10 seconds. This timespan will be called 'impact' of the device. For each of the warning signal device, a 10-minute simulation has been made. The results are summarised in Figure 5.

The number of delays during the simulation amounts to 2 in four cases and 3 in one case. Due to the different number of delays, the total delay time increases for impact 2 and then decreases and remains at the same level for impact 5 and 10. The results, although should only be considered as estimates, allow to opt for the device with impact 3 or 5 and suggest that investing in the device with impact 10 is not justified.

5 CONCLUSIONS

Models prepared with the FRAM can be used not only for description of processes, but also for simulating systems throughout their lifetimes. Models used as a basis for this kind of simulation will often be complex and, therefore, a strict and consistent understanding of the functions' aspects is needed. A proposal for such understanding has been presented in this paper together with an early version of a dedicated simulation software. Its applicability has been shown on an example of warning signal devices installed in a tram. Further work is needed to make the software fully intuitive and to integrate it with the official FRAM editor (Hollnagel 2016).

ACKNOWLEDGEMENTS

The research work financed with the means of statutory activities of Poznan University of Technology, No. 05/52/DSPB/0280.

REFERENCES

Adjetey-Bahun, K. Birregah, B. Châtelet, E. & Planchet, J.-L. 2016. A model to quantify the resilience of mass railway transportation systems. *Reliability Engineering & System Safety* 153: 1–14.

Almklov, P.G. Rosness, R. & Størkersen, K. 2014. When safety science meets the practitioners: Does safety science contribute to marginalization of practical knowledge? *Safety Science* 67: 25–36.

Anvarifar, F. Voorendt, M.Z. Zevenbergen, C. & Thissen, W. 2017. An application of the Functional Resonance Analysis Method (FRAM) to risk analysis of multifunctional flood defences in the Netherlands. *Reliability Engineering and System Safety* 158(October 2016): 130–41.

Le Coze, J.C. 2016. Vive la diversité! High Reliability Organisation (HRO) and Resilience Engineering (RE). *Safety Science*.

Haavik, T.K. Antonsen, S. Rosness, R. & Hale, A. 2016. HRO and RE: A pragmatic perspective. *Safety Science*.

Hale, A. & Borys, D. 2013. Working to rule, or working safely? Part 1: A state of the art review. *Safety Science* 55: 207–21.

Hale, A.R. Heming, B.H.J. Carthey, J. & Kirwan, B. 1997. Modelling of safety management systems. *Safety Science* 26(1–2): 121–40.

Heinrich, H.W. 1931. *Industrial Accident Prevention: A Scientific Approach*.

Hollnagel, E. 2012. *FRAM, the functional resonance analysis method: modelling complex socio-technical systems*. Ashgate.

Hollnagel, E. 2014. *Safety-I and safety-II/: the past and future of safety management*. CRC Press.

Hollnagel, E. 2016. The FRAM Model Visualiser [online].

Jeffcott, S. Pidgeon, N. Weyman, A. & Walls, J. 2006. Risk, trust, and safety culture in U.K. train operating companies. *Risk Analysis* 26(5): 1105–21.

Lundberg, J. & Johansson, B.J. 2015. Systemic resilience model. *Reliability Engineering & System Safety* 141: 22–32.

Lundberg, J. Rollenhagen, C. & Hollnagel, E. 2009. What-You-Look-For-Is-What-You-Find—The consequences of underlying accident models in eight accident investigation manuals. *Safety Science* 47(10): 1297–311.

Paté-Cornell, M.E. 1993. Learning from the Piper Alpha Accident: A Postmortem Analysis of Technical and Organizational Factors. *Risk Analysis* 13(2): 215–32.

Patriarca, R. Bergström, J. & Di Gravio, G. 2017. Defining the functional resonance analysis space: Combining Abstraction Hierarchy and FRAM. *Reliability Engineering and System Safety* 165(July 2016): 34–46.
Patriarca, R. Di Gravio, G. & Costantino, F. 2017. A Monte Carlo evolution of the Functional Resonance Analysis Method (FRAM) to assess performance variability in complex systems. *Safety Science* 91: 49–60.
Patriarca, R. Di Gravio, G. Costantino, F. & Tronci, M. 2017. The Functional Resonance Analysis Method for a systemic risk based environmental auditing in a sinter plant: A semi-quantitative approach. *Environmental Impact Assessment Review* 63(March): 72–86.
Pickup, L. Atkinson, S. Hollnagel, E. Bowie, P. Gray, S. Rawlinson, S. & Forrester, K. 2017. Blood sampling—Two sides to the story. *Applied Ergonomics* 59: 234–42.
Power, M. 1997. *The Audit Society. Rituals of Verification*. Oxford: Oxford University Press.
Wang, C.-H. & Liu, Y.-J. 2012. Omnidirectional safety culture analysis and discussion for railway industry. *Safety Science* 50(5): 1196–204.
Woods, D.D. 2015. Four concepts for resilience and the implications for the future of resilience engineering. *Reliability Engineering and System Safety* 141: 5–9.
Yang, Q. Tian, J. & Zhao, T. 2017. Safety is an emergent property: Illustrating functional resonance in Air Traffic Management with formal verification. *Safety Science* 93: 162–77.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Technical safety and reliability methods for resilience engineering

I. Häring & P. Gelhausen
Fraunhofer Ernst-Mach-Institut, EMI, Efringen-Kirchen, Germany

ABSTRACT: Resilience of technical and socio-technical systems can be defined as their capability to behave in an acceptable way along the timeline pre, during and post potentially dangerous or disruptive events, i.e. in all phases of the resilience cycle and overall. Hence technical safety and reliability methods and processes for technical safety and reliability are strong candidate approaches to achieve the objective of engineering resilience for such systems. This is also expected when restricting the set of methods to classical safety and reliability assessment methods, e.g. classical Hazard Analysis (HA) methods, inductive Failure Mode and Effects Analysis (FMEA), deductive Fault Tree Analysis (FTA), Reliability Block Diagrams (RBDs), Event Tree Analysis (ETA) and reliability prediction. Such methods have the advantage that they are typically already used in industrial research and development. However, improving the resilience of systems is usually not their explicit aim. The paper covers how to allocate such methods to different resilience assessment, response, development and resilience management tasks when engineering resilience from a technical perspective. In particular, the resilience dimensions of risk management, resilience objectives, resilience cycle time phases, technical resilience capabilities and system layers are used explicitly to explore their range of applicability. Also typical system graphical modelling, hardware and software development methods are assessed to document the usability of technical reliability and safety methods for resilience analytics and technically engineering resilience.

1 INTRODUCTION

As the number of applications of the concept of resilience to (socio) technical systems in mainly academic research and development rises, the question of how to successfully implement these approaches in the private sector, industry and small and medium enterprises is getting more and more prominent. The present paper addresses this challenge by surveying the suitability of classical, mainly analytical system analysis methods for assessing and improving resilience of (socio) technical systems.

The wording resilience analytics has been used recently quite general in the sense of resilience assessment in the societal-technical domain, see e.g. (López-Cuevas et al. 2017; Linkov, Florin 2016; Thorisson et al. 2017). However, the present use of the term analytical is to relate to classical system analysis methods, such as hazard analyses (HAs) including Hazard Lists (HLs), failure mode and effect analyses (FMEAs), fault tree analysis (FTA), event tree analysis (ETA), reliability block diagrams (RBD) and double failure matrix (DFM).

The application of established system modelling, analysis and simulation methods for resilience analysis covers Bayesian networks (e.g. Yodo et al. 2017) and Markov Models (e.g. Zhao et al. 2017). Fault propagation in a hazard and operability analysis (HAZOP) context for resilience assessment is described in (Cai et al. 2015). Also first approaches have been reported to use FMEA for resilience assessment with the aim of applying the functional resonance analysis methodology (FRAM) to a smart building (Mock et al. 2016). However, for instance fault tree analysis has not yet been used for resilience analysis.

First attempts of an evaluation of the suitability of system analysis, simulation and development methods including some selected classical system analysis methods for resilience assessment have been conducted in (Häring et al. 2016c; Häring et al. 2016b). In contrast to (Häring et al. 2017), the present approach does not provide generic considerations of the suitability of methods for technically driven resilience assessment and development process steps.

The present approach focuses on determining which contributions to resilience assessment can be expected from mainly analytical system analysis methods and their extensions. To this end, it resorts to already often used resilience dimensions such as resilience or catastrophe response phases, system management domains or resilience capabilities.

By resorting to 5 such resilience dimensions as detailed and motivated below, the expected relevancy of mainly analytical system analysis methods is assessed from different and complementary perspectives. Using the resilience dimensions, the suitability of method assessment is resolving the expected benefit in respective phases or resilience

aspects rather than aiming at an overall applicability scoring.

This is conducted based on expert judgement (Meyer, Booker 2001) and consensus feedback of scientists related to the research field of technical safety and risk analysis. Also groups of almost finished master students in security and safety engineering of an applied science university contributed, mainly trained in tabular system analysis methods.

Key motivations for focusing on classical system analysis methods include:

- Analytical system analysis methods are established and accepted by practitioners in industry;
- Expectation that resilience analysis can in parts be delivered with extensions of classical methods;
- Expected efficiency of semi-quantitative methods compared to quantitative approaches;
- Identification of implicit resilience activities within current existing risk analysis and management practice;
- Clarification of resilience concepts by specifying methods supporting their fulfillment;
- Identification of critical resilience aspects that need to be analyzed with more effort, i.e. going beyond classical system analysis methods.

The paper is structured as follows. In section 2, the approach is described how to assess the suitability of mainly classical analytical system analysis methods for resilience analysis by employing resilience dimensions suitable for technical resilience understanding. Section 2 also details the methodology. It illustrates the need of going beyond classical risk assessment with the help of resilience event propagation through logic and assessment layers.

In the following sections 3 to 7, for each of the listed resilience concepts, possible contributions from the methods are discussed. For each resilience dimension a matrix is filled with assessments of the suitability of the method for contributing to each of the resilience dimension attribute. Also, recommendations for the extension of the methods are given.

In section 8, the overall suitability of each method is summarized and conclusions regarding adaptations and further developments are drawn.

2 APPROACH TO ASSESS THE SUITABILITY OF METHODS

Before detailing the approach of suitability assessment of classical system analysis methods, the paper gives some general considerations on the necessary extension of methods for resilience assessment when compared to classical risk assessment.

Conditional probability expressions based and extending classical notions of risk have recently been used to quantify key objectives of resilient response (Aven 2017). This shows that resilience analysis may benefit from the application of traditional and more modern risk concepts.

The idea used as starting point in (Aven 2017) is that resilience behavior can be defined to occur post disruption events. Thus resilience event B, e.g. "system stabilizes post disruption", "system recovers", "system bounces back better", "recovery time shorter than critical time" or "sufficient system performance level reached within t" are always conditional previous events,

$$P(B \mid A), \tag{1}$$

where A is a "disruptive event" or equals a chain of events,

$$A = A_1, A_2, \cdots, A_n. \tag{2}$$

This approach relates with the often used definition of conditional vulnerability in risk expressions, see e.g. (Daniel M. Gerstein et al. 2016),

$$R = P(E)\, P(C \mid E)\, C, \tag{3}$$

where E is a threat event and C the consequence.

However, the classical vulnerability approach of (3) focuses on the quantification of the conditional consequence probability, whereas (1) refers to resilience behavior post disruption events.

As the vulnerability including risk definition of (3) is already an extension of the classical definition of risk

$$R = P(C)\, C, \tag{4}$$

and typical resilience expressions are further extending the definition of (1) and (3), it is expected that classical system reliability and safety approaches are challenged when used for assessing resilience. In particular (very) simple tabular approaches resort to risk concepts as described by (4) when applied in a traditional way, i.e. they focus on avoidance of events and system robustness in case of events only.

Generalizing (1), resilience expressions of interest typically are of the form (Häring et al. 2016a)

$$P(B \mid A) = \sum_{i=1}^{N} P(B \mid D_i) P(D_i \mid A), \tag{5}$$

where $D_i, i = 1,2,...,n,$ form a complete set of expansion events. Equation (5) uses the law of total probability and can be understood as an insertion of unity of all possible intermediate states

$$\sum_{i=1}^{n} P(\bullet \mid D_i) P(D_i \mid \bullet) \tag{6}$$

between any two known states. Equation (5) can also expresses the idea of possibly unknown transition states or disruptions which are included in the set D_i. In this case, A is just a system initial sate.

Of course, (5) can be generalized to consider multiple resilience layers or response and recovery phases, see (Häring et al. 2016a). Along the lines of interpretation given for (1), an interpretation of (5) reads for instance

$$\begin{aligned} &A = \text{“Disruption event”},\\ &\{D_i\}_{i=1,\ldots,n}, \text{Set of possible response and recovery events/Set of transition states},\\ &B = \text{“Final state of interest”}. \end{aligned} \tag{7}$$

When comparing risk and vulnerability expressions of the form (3) and (4) with resilience expressions of the form (1) and (5), it becomes obvious that it is not straightforward to expect that classical analytical system analysis methods can deliver assessment results regarding resilience. This motivates the question how such methods can contribute to resilience assessment.

For focusing the research question, the following system modelling, classical system analysis and system development methods are considered regarding their suitability for resilience assessment:

- SysML, UML;
- HL, PHA, HA, O&SHA, HAZOP;
- FMEA, FMECA, FMEDA;
- RBD;
- ETA;
- DFM;
- FTA, time-dependent FTA (TDFTA);
- Reliability prediction with standards;
- Methods for HW and SW development;
- Bit error correction methods.

To assess the suitability of methods for resilience engineering, the following resilience dimensions are used:

- 5-step risk management process (AS/NZS ISO 31000:2009), for review: (Purdy 2010), (Luko 2013), for critical discussion mainly regarding the coverage of uncertainty (Aven 2011);
- Resilience time-phase cycle, based on (Thoma 2014);
- Technical resilience capabilities, based on (Häring et al. 2016a);
- System layers, based on (Häring 2016a);
- Resilience criteria (Bruneau et al. 2003) (Pant et al. 2014);
- Resilience analysis and management process (Häring et al. 2017);

For the first 5 resilience dimensions, each combination of system analysis method and resilience dimension attribute is assessed using the three equivalent semi-quantitative scales

$$\begin{aligned} &\{1,2,3,4,5\},\\ &\{--,-,o,+,++\},\\ &\{\text{not suited (adaptation useless)},\\ &\ \text{rather not suited (or only with major modifications)},\\ &\ \text{potentially suited (after adaptation)},\\ &\ \text{suited (with minor modifications)},\\ &\ \text{very well suited (straightforward/no adaptions)}\}. \end{aligned} \tag{8}$$

Typical examples read as follows: (i) The identification of potential disruption events of systems can be supported by using the classical system analysis methods hazard list (HL) and preliminary hazard analysis (PHA). Hazard lists are very well suited for identifying hitherto unknown events when used as checklists of potential disruptions and asking the question “what if?”. Regarding the identification of possible disruptions for a system under consideration, the overall rating of HL could be “++” or “+” for PHA. This example shows that rather than assessing the generic suitability of a method, its use within a certain resilience assessment process or conceptual structuring is addressed.

(ii) Fault tree analysis (FTA) allows to consider combinations of events by using the AND gate. When only a known sequence of events is possible, the sequencing AND gate can be used, which enforces an order of occurrence of events. Such a sequence might be first “detection of threat”, second “decision to start counter-measure” and third “activation of counter-measure”. This order is then used for assessing the probability of success of a technical prevention measure. Sequential events can be analyzed with time-dependent Boolean differences to analyze sequential structure functions rather than classical combinatorial Boolean structure functions (Moret, Thomason 1984). Hence, FTA and even more TDFTA can be expected to cover after modifications also the response and recovery phase, resulting in a “+” assessment, respectively.

3 SUITABILITY ASSESSMENT WITH FIVE-STEP RISK MANAGEMENT SCHEME

The 5-step risk management scheme is only a very generic framework for identifying risks on resilience objectives. As discussed in the introduction, objectives in the case of resilience analysis are more second order (e.g. “fast recovery in case of disruption”) when compared to classical risk analysis and management (e.g. “avoid disruption”).

Table 1 assess the suitability of analytical system analysis and some development methods for resilience analysis sorted along the 5-step risk management scheme using the scale of (8).

Understanding the system sufficiently for resilience risk analysis is supported with graphical/semi-formal Unified/Systems modelling languages (UML/SysML) modelling, see the first two lines of Table 1.

The initial hazard analysis methods HL and PHL support the identification of possible disruptions. They are considered as a starting point. Refined analyses can be supported with SSHA, HA, O&SHA, and HAZOP, the differences of which are typically small and depend on the application; for small systems they can be summarized in one analysis.

Approaches that need substantial system knowledge include RBD, the inductive approaches ETA, FME(D/C)A and deductive approaches (TD) FTA, which are often summarized in a bow tie analysis. The success of FMEA variations is expected to be more efficient when depending on system functions (or services) as inductive starting points rather than system components or subsystems.

In the case of (TD) FTA, the success of application will strongly depend on the definitions of the top events, which should cover main resilience objectives.

Table 1. Suitability of analytical system analysis and HW/SW development methods for resilience analysis sorted along the 5-step risk management scheme.

Method\ 5-step risk management process steps	(1) Establish context	(2) Identify risk/ hazards	(3) Analyze/ compute risks	(4) Evaluate risks	(5) Mitigate risks
SysML	+	++	++	o	++
UML	+	+	+	o	+
HL	+	++	+	o	-
PHA	+	+	++	+	+
SSHA, HA	o	+	++	++	++
O&SHA, HAZOP	o	+	++	++	++
FMEA, FMECA	–	o	++	+	++
FMEDA	–	–	++	+	++
RBD	o	+	++	+	++
ETA	+	+	++	+	++
DFM	–	o	++	+	++
FTA, TDFTA	–	o	++	+	++
Reliability prediction with standards	–	–	++	+	++
Methods for HW and SW development	–	–	–	–	++
Bit error correction methods	–	–	–	–	++

4 METHOD USABILITY ASSESSMENT USING RESILIENCE RESPONSE CYCLE TIME PHASES

The catastrophe management cycle in 4 steps (e.g. preparation, prevention and protection, response and recovery, learning and adaption) as well as in 5 steps as used in Table 2 take advantage of a logic or time ordering of events with respect to disruption events (Häring et al. 2016a): (far) before, during, immediately (after). Another typical timeline as well as logic sequence example was given in section 2.

The first observation in Table 2 is that the analysis methods should be conducted, if considered relevant, mainly in the preparation phase. However, especially fast analytical simple methods can also be applied during actual conduction of response and recovery. For instance, during and post events, a PHA scheme could be used to identify further possible second-order events given a disruption.

The second observation in Table 2 comprises the coverage of resilience cycle phases. The suitability of method assessment stems from the fact that classical analytical approaches by definition cover prevention and protection when identified with frequency of event assessment and immediate (first order) damage assessment.

Table 2. Suitability of system modelling, analytical system analysis and selected development methods for resilience analysis along the 5 phases of the resilience cycle: Resilience event order logic or timeline.

Method\ Resilience timeline cycle phase	(1) Prepare	(2) Prevent	(3) Protect	(4) Respond	(5) Recover
SysML	++	++	++	++	++
UML	++	+	+	+	+
HL	++	++	++	++	++
PHA	++	++	++	+	+
SSHA, HA	++	++	++	+	+
O&SHA, HAZOP	++	++	++	+	+
FMEA, FMECA	++	++	++	+	+
FMEDA	++	++	++	+	+
RBD	++	++	++	+	+
ETA	++	++	++	+	+
DFM	++	++	++	+	+
FTA, TDFTA	++	++	++	+	+
Reliability prediction with standards	++	++	++	+	+
Methods for HW and SW development	++	++	++	+	+
Bit error correction methods	++	++	++	++	+

If system failure, in case of variations of HA, FMEA and FTA, is defined as failure of adequate response (e.g. absorption and stabilization), of recovery (e.g. reconstruction and rebuilding) or even of improving or bouncing forward using damage as optimization opportunity, these methods can be used with adaptions also for these resilience timeline phases. Similarly, RBD and ETA are assessed.

The system modelling and development methods for hardware and software (HW/SW) can be used for all resilience cycle phases. As in the case of classical system analysis methods, adaptions up to major new developments are believed to be necessary.

Even if especially the classical tabular system analysis methods were assessed as very relevant for resilience assessment, it is noted that they are part of established processes in practice. Therefore, even when adding only some additional columns, their modified best practice of use is expected to be challenging in company development environments. In this sense, Table 2 is a guideline for the expected usability of the listed methods.

5 METHOD USABILITY ASSESSMENT USING TECHNICAL RESILIENCE CAPABILITIES

Sensor-logic-actor chains are basic functional elements used for active safety applications in safety instrumented systems, especially within the context of functional safety as governed by (IEC 61508 Series). The technical resilience capabilities can be considered as a generalization of such functional capabilities.

They can also be related to the much more abstract and generic OODA (observe, orient, decide, act) loop, which has found much application also in the catastrophe response arena, see e.g. (Lubitz et al. 2008; Huang 2015). The technical resilience capabilities are also very close to capabilities to be expected from a general artificial intelligence (Baum et al. 2010) and related possible architectures (Goertzel et al. 2008).

Table 3 assesses the suitability for use of the selected methods along each technical resilience capability dimension attribute. Since the technical resilience capabilities are generic properties of (socio) technical systems, the realization of the properties in systems is prone to risks: e.g. external and internal; accidental and intentional; safety and security related; systematic (by construction) and statistic.

In Table 3, the more generic system modelling methods SysML and UML are rated better when compared to more specific methods. Table 3 expresses with the uniform distribution of "+" that any resilience analysis conducted using the methods has to take into account all the technical resilience properties. This shows that major adaptations and further developments are necessary to apply classical methods, since a cross-cutting task has to be covered by the methods.

Table 3. Suitability of system modelling, analytical system analysis and selected development methods for resilience analysis along the technical resilience capability dimension attributes.

Method\ Technical resilience capabilities	(1) Observation, sensing	(2) Representation, modeling, Simulation	(3) Inference. decision making	(4) Activation, action	(5) Learning, modification, adaption, rearrangement
SysML	++	++	++	++	++
UML	+	++	++	+	++
HL	+	+	+	+	+
PHA	+	+	+	+	+
HA	+	+	+	+	+
O&SHA, HAZOP	+	+	+	+	+
FMEA, FMECA	+	+	+	+	+
FMEDA	+	+	+	+	+
RBD	+	+	+	+	+
ETA	+	+	+	+	+
DFM	+	+	+	+	+
FTA, TDFTA	+	+	+	+	+
Reliability prediction with standards	+	+	+	+	+
Methods for HW and SW development	+	+	+	+	+
Bit error correction methods	o	o	o	o	o

For instance, columns or labels could be added to assess to which type of system resilience function a system failure belongs in case of HA, FMEA and ETAs. Also FTAs top level event formulations either have to address the functional steps separately or find sufficient generic top level formulations allowing for combinations of top events.

6 METHOD USABILITY ASSESSMENT USING SYSTEM LAYERS

Table 4 assesses the potential of application of the representative methods of section 2 with the help of system layers for socio technical systems. The often used 4 layers physical, information, cognitive, social, see e.g. (Fox-Lent et al. 2015), have been refined in the physical-technical domain and more specified in all attributes when compared to (Häring et al 2016a).

The strength of the selected representative very specific methods is in the domain of hardware and data integrity as well as HW/SW development. Also the classical tabular methods focus somewhat on electronics, especially FMEDA.

Table 4. Suitability of system modelling, analytical system analysis and selected development methods for resilience analysis along system layers or generic management domains.

Method\ System layer, Management domain	(1) Physical	(2) Technical, hardware	(3) Cyber, software-wise, protocols	(4) Operational, organizational	(5) Societal, economic, ethical
SysML	++	++	++	+	+
UML	++	++	++	+	+
HL	+	++	+	o	o
PHA	+	++	+	o	o
HA	+	++	+	o	o
O&SHA, HAZOP	+	++	+	+	+
FMEA, FMECA	+	++	+	o	o
FMEDA	+	++	+	o	o
RBD	+	++	+	o	o
ETA	+	++	+	+	+
DFM	+	++	+	+	+
FTA, TDFTA	+	++	+	o	o
Reliability prediction with standards	–	++	–	o	o
Methods for HW and SW development	+	++	++	o	o
Bit error correction methods	–	+	++	o	o

Table 5. Suitability of system modelling, analytical system analysis and selected development methods for resilience analysis along modified resilience criteria.

Method\ Modified resilience criteria	(1) Robustness: low initial damage	(2) Redundancy: system property of overall damage tolerance	(3) Resourcefulness: fast stabilization and response	(4) Rapidity: fast recovery and reconstruction
SysML	+	++	++	++
UML	o	++	+	+
HL	+	o	–	–
PHA	+	+	–	–
HA	+	+	–	–
O&SHA, HAZOP	–	+	–	–
FMEA, FMECA	+	o	–	–
FMEDA	o	–	–	–
RBD	+	++	–	–
ETA	+	+	–	–
DFM	o	++	–	–
FTA, TDFTA	+	++	+	+
Reliability prediction with standards	–	–	–	–
Methods for HW and SW development	–	–	–	–
Bit error correction methods	–	–	–	–

The general purpose methods ETA, DFM, RBD and FTA require educated application, often out of their classical domain of application, especially HAZOP. RBD diagrams and SysML/UML methods are expected to be of use for acquiring and documenting sufficient system understanding.

HA-type methods are well suited but need to be applied out of their typical technical domain also to operational and societal system layers.

7 METHOD USABILITY ASSESSMENT USING RESILIENCE CRITERIA

Table 5 assesses the potential of application of the representative methods of section 2 with the help of the often used 4 resilience criteria introduced by (Bruneau et al. 2003) and technically refined by (Pant et al. 2014). For the suitability of method assessment, the following modified working definitions are used in this paper:

1. Robustness: measure for low level of damage (vulnerability) in case of event; 'good' absorption behavior; 'good' protection.
2. Redundancy: measure for low level of overall system effect in case of local (in space, in time, etc.) disruption event; system disruption tolerance.
3. Resourcefulness: measure for capability of successful allocation of resources in the response phase to stabilize the system post disruptions.
4. Rapidity: measure for fast recovery of system.

The results are similar to the suitability assessment along the logic or timeline resilience cycle phases as conducted in section 4: classical analytical approaches do not focus beyond the damage events. Robustness, resourcefulness and rapidity are according to the working definitions strongly related to the resilience cycle phases absorption/protection, response and recovery.

Redundancy is understood in the classical way as an overall system property. Hence in all cases sufficient system understanding is required, which is supported by graphical modelling.

The (also) graphical approaches RBD, ETA and FTA are strong for redundancy and resourceful-

ness assessment. Especially time dependent FTA and underlying time dependent Markov models are believed to be key for resourcefulness and redundancy assessment, nevertheless with major adaptations.

8 SUMMARY AND CONCLUSIONS

In summary, each of the representative analytical system analysis methods as well as HW/SW development methods (techniques and measures in the sense of (IEC 61508 Series)) showed potential for resilience engineering, i.e. resilience assessment and development and optimization as defined in the introductory sections.

The classical tabular approaches HA and FMEA are assessed to be suited with minor up to major modifications for resilience analytics. Major advantages are expected by redefining and adding dedicated columns to cover resilience aspects. Also graphical methods like RBD, ETA and FTA are tools that by definition cover at least technical aspects of resilience of systems in case of very informed application.

In all cases, the extensions and adaptions need to carefully consider the initial background and application context of the methods. Therefore, in case of technical resilience engineering contexts, it is expected that the methods have to be newly established. This holds since all these methods are prone to routinely use, which is often very contrary to the out-of-the box thinking necessary for resilience engineering. For instance, established hazard lists for an application domain will not contribute to an as complete as possible disruption threat list.

The different resilience dimensions used for suitability assessment exhibited strengths and weaknesses for exploring the methods' potentials:

- The risk management cycle is a very generic process, allocating most analysis methods in the risk analysis step. Resilience objectives formulation is key and challenge.
- Resilience cycle (time or logic) phases allow to spread out assessments and activities. However, they are prone to 'divide et impera' effects of losing the overall picture.
- Technical resilience capabilities need to be covered for the operation of typical system (service) functions on overall system level allowing a technical approach. It is deemed challenging how to modify and extend classical methods to cover them.
- Traditional resilience criteria ("Resilience Rs") can be nicely linked to timeline/logic concepts as well as system redundancy assessments. They also link with performance-based resilience curve assessments. Challenges are expected when trying to translate the more abstract concepts into system analysis and development requests.

In summary, future work is expected to benefit from informed further development of classical system analysis methods for resilience analysis. Such resilience analytics is believed also to strongly support the development of resilient systems, in particular in industrial environments. Such informed applications are expected to ripe many of the benefits listed in the bullet list of the introduction.

ACKNOWLEDGEMENTS

This research has been conducted in the context of the Freiburg Sustainability Center of Excellence, a cooperation of the Fraunhofer institutes in Freiburg and the Albert-Ludwigs-University Freiburg. It is supported by grants from the Baden-Württemberg Ministry of Economics and the Baden-Württemberg Ministry of Science, Research and the Arts. In parts, the work has also been supported by the German BMBF Project "Windows for continuous academic education" within the Sub-Project "Resilient Technical Systems". Thanks goes also to master students of security and safety engineering of the Hochschule Furtwangen University.

REFERENCES

AS/NZS ISO 31000:2009: Risk management - Principles and guidelines.

Aven, Terje (2011): On the new ISO guide on risk management terminology. In *Reliability Engineering and System Safety* 96 (7), pp. 719–726. DOI: 10.1016/j.ress.2010.12.020.

Aven, Terje (2017): How some types of risk assessments can support resilience analysis and management. In *Reliability Engineering & System Safety* 167, pp. 536–543. DOI: 10.1016/j.ress.2017.07.005.

Baum, Eric; Hutter, Marcus; Kitzelmann, Emanuel (Eds.) (2010): Artificial general intelligence. Proceedings of the Third Conference on Artificial General Intelligence, AGI 2010, Lugano, Switzerland, March 5–8, 2010. Conference on Artificial General Intelligence; AGI. Amsterdam: Atlantis Press (Advances in intelligent systems research, 10).

Bruneau, Michel; Chang, Stephanie E.; Eguchi, Ronald T.; Lee, George C.; O'Rourke, Thomas D.; Reinhorn, Andrei M. et al. (2003): A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities. In *Earthquake Spectra* 19 (4), pp. 733–752. DOI: 10.1193/1.1623497.

Cai, Zhansheng; Hu, Jinqiu; Zhang, Laibin; Ma, Xi (2015): Hierarchical fault propagation and control modeling for the resilience analysis of process system. In *Chemical Engineering Research and Design* 103, pp. 50–60. DOI: 10.1016/j.cherd.2015.07.024.

Daniel M. Gerstein; James G. Kallimani; Lauren A. Mayer; Leila Meshkat; Jan Osburg; Paul Davis et al. (2016): Developing a Risk Assessment Methodology for the National Aeronautics and Space Administration. RAND Corporation (RR-1537-NASA). Available online at https://www.rand.org/pubs/research_reports/RR1537.html.

Fox-Lent, Cate; Bates, Matthew E.; Linkov, Igor (2015): A matrix approach to community resilience assessment. An illustrative case at Rockaway Peninsula. In *Environ Syst Decis* 35 (2), pp. 209–218. DOI: 10.1007/s10669-015-9555-4.

IEC 61508 Series, 2010: Functional safety of electrical/electronic/programmable electronic safety-related systems. Available online at http://www.iec.ch/functionalsafety/standards/page2.htm, checked on 12/27/2017.

Goertzel, Ben; Wang, Pei; Franklin, Stan (Eds.) (2008): Artificial general intelligence, 2008. Proceedings of the First AGI Conference. ebrary, Inc; AGI Conference. Amsterdam, Washington, DC: IOS Press (Frontiers in artificial intelligence and applications, v. 171).

Häring, Ivo; Ebenhöch, Stefan; Stolz, Alexander (2016a): Quantifying resilience for resilience engineering of socio technical systems. In Eur J Secur Res 1 (1), pp. 21–58. DOI: 10.1007/s41125-015-0001-x.

Häring, Ivo; Sansavini, Giovanni; Bellini, Emanuel; Martyn, Nick; Kovalenko, Tatyana; Kitsak, Maksim et al. (2017): Towards a generic resilience management, quantification and development approach. In: Linkov I., Palma-Oliveira J. (eds) Resilience and Risk. NATO Science for Peace and Security Series C: Environmental Security. Springer, pp. 21–80. https://link.springer.com/chapter/10.1007/978-94-024-1123-2_2.

Häring, Ivo; Scharte, Benjamin; Hiermaier, Stefan (2016b): Towards a novel and applicable approach for Resilience Engineering. In: 6-th International Disaster and Risk Conference (IDRC). Integrative Risk Management – towards resilient cities. 6-th International Disaster and Risk Conference (IDRC). Davos, 28.08-01.09.

Häring, Ivo; Scharte, Benjamin; Stolz, Alexander; Leismann, Tobias; Hiermaier, Stefan (2016c): Resilience Engineering and Quantification for Sustainable Systems Development and Assessment. In: Resource Guide on Resilience. Lausanne: EPFL International Risk Governance Center.

Huang, Yanyan (2015): Modeling and simulation method of the emergency response systems based on OODA. In *Knowledge-Based Systems* 89, pp. 527–540. DOI: 10.1016/j.knosys.2015.08.020.

2010: IEC 61508 - Functional safety of electrical/electronic/programmable electronic safety-related systems.

Linkov, Igor; Florin, M.-V. (Eds.) (2016): IRGC Resource Guide on Resilience. Edited Book. Available online at https://www.irgc.org/risk-governance/resilience/.

López-Cuevas, Armando; Ramírez-Márquez, José; Sanchez-Ante, Gildardo; Barker, Kash (2017): A Community Perspective on Resilience Analytics. A Visual Analysis of Community Mood. In Risk analysis: an official publication of *the Society for Risk Analysis* 37 (8), pp. 1566–1579. DOI: 10.1111/risa.12788.

Lubitz, D.K. von; Beakley, James E.; Patricelli, Frederic (2008): 'All hazards approach' to disaster management: the role of information and knowledge management, Boyd's OODA Loop, and network-centricity. In Disasters (32, 4), pp. 561–585. DOI: 10.1111/j.0361-3666.2008.01055.x.

Luko, Stephen N. (2013): Risk Management Principles and Guidelines. In *Quality Engineering* 25 (4), pp. 451–454. DOI: 10.1080/08982112.2013.814508.

Meyer, M.A.; Booker, J.M. (2001): Eliciting and Analyzing Expert Judgment. A Practical Guide: Society for Industrial and Applied Mathematics.

Mock, R.; Lopez de Obeso, Luis; Zipper, Christian (2016): Resilience assessment of internet of things. A case study on smart buildings. In Lesley Walls, Matthew Revie, Tim Bedford (Eds.): European Safety and Reliability Conference (ESREL). Glasgow, 25-29.09. London: Taylor & Francis Group, pp. 2260–2267.

Moret, B.M.E.; Thomason, M.G. (1984): Boolean Difference Techniques for Time-Sequence and Common-Cause Analysis of Fault-Trees. In IEEE Trans. Rel. R-33 (5), pp. 399–405. DOI: 10.1109/TR.1984.5221879.

Pant, Raghav; Barker, Kash; Ramirez-Marquez, Jose Emmanuel; Rocco, Claudio M. (2014): Stochastic measures of resilience and their application to container terminals. In *Computers & Industrial Engineering* 70, pp. 183–194. DOI: 10.1016/j.cie.2014.01.017.

Purdy, Grant (2010): ISO 31000. 2009—Setting a New Standard for Risk Management. In Risk analysis: an official publication of *the Society for Risk Analysis* 30 (6), pp. 881–886. DOI: 10.1111/j.1539-6924.2010.01442.x.

Thoma, Klaus (Ed.) (2014): Resilien-Tech: "Resilience by Design": a strategy for the technology issues of the future. München: Herbert Utz Verlag; Utz, Herbert (acatech STUDY).

Thorisson, Heimir; Lambert, James H.; Cardenas, John J.; Linkov, Igor (2017): Resilience Analytics with Application to Power Grid of a Developing Region. In Risk analysis: an official publication of *the Society for Risk Analysis* 37 (7), pp. 1268–1286. DOI: 10.1111/risa.12711.

Yodo, Nita; Wang, Pingfeng; Zhou, Zhi (2017): Predictive Resilience Analysis of Complex Systems Using Dynamic Bayesian Networks. In IEEE Trans. Rel. 66 (3), pp. 761–770. DOI: 10.1109/TR.2017.2722471.

Zhao, S.; Liu, X.; Zhuo, Y. (2017): Hybrid Hidden Markov Models for resilience metrics in a dynamic infrastructure system. In RESS 164, pp. 84–97. DOI: 10.1016/j.ress.2017.02.009.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Interdependent infrastructure network restoration from a community resilience perspective

K. Barker, D.B. Karakoc & Y. Almoghathawi
University of Oklahoma, Norman, OK, US

ABSTRACT: Many critical infrastructure networks that dot the global landscape often rely on each other in different ways for each to be functional. Government planning documents around the world recognize the interdependence of these infrastructure networks. But naturally infrastructure networks do not exist for their own operation but because society relies upon them for convenience, productivity, and health, among others. Recent large-scale disruptions to critical infrastructure, primarily due to natural disasters whose frequency appears to be increasing, have left communities devastated for extended periods. As such, planning for the resilience of critical cyber-physical-social networks should emphasize the social aspects of disruptions. In this work, we study the problem of the restoration of interdependent infrastructure networks after the occurrence of a disruptive event with a focus on the vulnerability of the society that interacts with the networks. We integrate (i) a resilience-driven multi-objective mixed-integer programming formulation that schedules the restoration of disrupted demand nodes in each network with (ii) a geographically distributed index of social vulnerability that measures the impact to the community surrounding the disrupted demand nodes. This model integration is illustrated with an example of community resilience in Shelby County, Tennessee.

1 INTRODUCTION

A *critical infrastructure network* is defined as a network of independent, mostly privately-owned, human-made systems and processes that function collaboratively and synergistically to produce and distribute a continuous flow of essential goods and services (The Report of the President's Commission on Critical Infrastructure Protection 1997). Such infrastructure networks, such as electric power, water distribution, natural gas, transportation, and telecommunications, operate on a daily basis to maintain the functioning of modern societies and to provide their essential needs.

With the continuous technological developments, infrastructure networks and their distribution systems become more dependent on each other's functionality to perform with higher efficiency (Rinaldi et al. 2001). However, this type of a complex coordination and interconnection on various aspects such as sharing components, utilizing one's output as another's input, transmitting information and much more, make critical infrastructure networks more vulnerable against possible disruptive events (Rinaldi et al. 2001). This (often bi-directional) relationship among these networks enhances the possibility of chain reactions between the disrupted and undisrupted components where one infrastructure network might lead the failure of another one due to their high interdependency (Little 2002, Wallace et al. 2003, Buldyrev et al. 2010, Eusgeld et al. 2011, Ouyang 2014, Danziger et al. 2016, Wu et al. 2016). Therefore, resilience planning in the form of restoration scheduling of these potentially highly vulnerable networks becomes more challenging especially when the increasing frequency of man-made or natural disruptive events considered.

In the literature, many different approaches have been introduced to quantify the resilience of a network where the ability to withstand, adapt to, and recover from a disruption is referred as resilience (Barker et al. 2017). As shown in Figure 1, consider two primary dimensions of resilience: vulnerability and recoverability (Henry and Ramirez-Marquez 2012, Barker et al. 2013). The vulnerability of a network is defined as the magnitude of damage

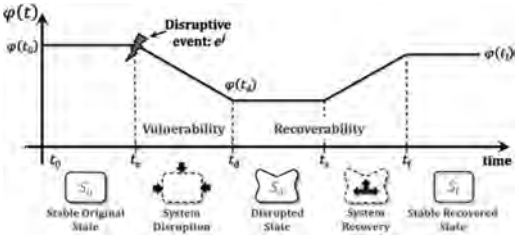


Figure 1. Network performance, φ(t), across state transitions before, during, and after the occurrence of a disruptive event.

in network performance due to a disruptive event (Jönsson et al., 2008), where the recoverability of a network describes the speed at which the network reaches to a desired performance level (Rose 2007). Resilience can be defined as the time dependent ratio of network recovery over its loss (i.e. $Я(t) = \text{Recovery}(t) / \text{Loss}(t)$where $Я_{\varphi}(t \mid e^{j}) = 1$ indicates that network is fully resilient [Henry and Ramirez-Marquez 2012]).

Recent research has explored the relationship between critical infrastructure networks are interrelated with the geographical vulnerabilities of the local regions where they are built and the social vulnerabilities of their surrounding communities (Cutter and Finch 2007). The specific and varying demographics of the communities that these critical infrastructures provide service to can be considered as the key factors to guide response and restoration operations.

In this initial study, we have considered a social vulnerability index as a community resilience measure which is a function of predetermined demographic factors of a community (Cutter et al. 2003). The weighted sum of these factors assigns relative vulnerability scores to disrupted components of the interdependent infrastructure networks according to the region in which they are located. In addition to the social vulnerability index, we have also considered the population density of the community whose interdependent infrastructure network service is disrupted.

2 SOCIAL VULNERABILITY INDEX

Different levels of socio-economic conditions and distinguishing properties of a community shape its resilience against disruptive events by either contributing to or counteracting its vulnerability. In their work on the Social Vulnerability Index (SoVI), Cutter et al. (2003) identify eleven key factors that contribute to measuring the vulnerability of a community including the age, gender, race, wealth, and occupation of members of the community.

A significant level for each of the factors has been identified and the percentage of population that is below these limits are considered as more vulnerable to disruptive events and, therefore, contributors to the overall social vulnerability. Due to their higher vulnerability, it is noted that through the recovery process, these subgroups would require more time and investment of resources to achieve a resilience level similar to the other communities (Cutter et al. 2003).

The SoVI-Lite technique (Cutter et al. 2011, Evans et al. 2014) is a reduced version of the SoVI that calculates a community's vulnerability score with the following steps:

1. Calculate the percentage of population that falls beyond the predetermined level of vulnerability for each factor,
2. Calculate the *z*-score of each factor by using mean and standard deviation of each factor, and
3. Sum the *z*-scores of all factors to find the total social vulnerability score of a specific community.

Furthermore, the SoVI-Lite score can be scaled between 0 and 1 using Eq. (1), where 0 represents the least socially vulnerable community and 1 is the most socially vulnerable community.

$$\frac{z - \min(x)}{\max(x) - \min(x)}, \forall\ z \in X \tag{1}$$

3 PROPOSED MODEL

In this study, we have proposed a multi-objective resilience-driven restoration optimization model using mixed-integer programming, where our main goal is to maximize the resilience of interdependent infrastructure networks while minimizing the total cost associated with the entire restoration phase (Almoghathawi et al. 2017). We have integrated social vulnerability index into the objectives to help guide interdependent infrastructure network restoration from a community resilience perspective.

Let K represent a set of infrastructure networks, $K = \{1, \quad, \kappa\}$, and T represent a set of available time periods, $T = \{1, \quad, \tau\}$. For each network $k \in K$, the sets of nodes and links are represented by N^k and L^k, respectively. The sets of source nodes and demand nodes are defined with $N_s^k \subseteq N^k$ and $N_d^k \subseteq N^k$, respectively. The sets of disrupted nodes and links are denoted by $N'k$ and $L'k$, respectively.

Let b_i^k be the maximum amount of supply at node $i \in N_s^k$ in network $k \in K$, considered to be the maximum flow from node $i \in N_s^k$ to all demand nodes in network $k \in K$. The amount of unmet demand at node $i \in N_d^k$ in network $k \in K$ in time $t \in T$ is denoted by s_{it}^k. Total unmet demand at all demand nodes in network $k \in K$ after recovery at time period $t \in T$ is $\sum_{i \in N_d^k} s_{it}^k$.

To introduce the social vulnerability index into the model, the SoVI-Lite score, $SoVI_i^k$, is calculated for node $i \in N'_d{}^k$ in network $k \in K$. To more effectively emphasize the social vulnerability index, Eq. (2) introduces an exponential effect to give more relative importance to nodes in socially vulnerable areas with V_i^k.

$$V_i^k = e^{b * SoVI_i^k}, \forall\ i \in N'_d{}^k, b \in Z \tag{2}$$

Additionally, population density was also included in the proposed model, where densities were assigned to demand nodes to more effectively place importance on them during the restoration process. P_i^k is the population density for demand node $i \in N_d^k$ in network $k \in K$, shown in Eq. (3).

$$P_i^k = \frac{\text{population of community where node } i \text{ is located}}{\text{total population of area being studied}}, \quad \forall i \in N_d'^k \tag{3}$$

The unmet demand in the network represents the system loss in the maximum flow which will be caused by disruptive event. In this manner, decreasing the total amount of unmet demand to a desirable level refers to the effectiveness of restoration process and a reasonable recoverability level of the network. Hence, the resilience of the system could be represented by Eq. (4) where it is the ratio of total unmet demand that is recovered over total amount of unmet demand after the disruption occurs. This equation represents cumulative recovery of the interdependent infrastructure networks over time $t \in T$ where Q_i^k is the unmet demand at demand node $i \in N_d^k$ in network $k \in K$ after a disruption and μ_i^k is the weight of demand node $i \in N_d^k$ in network $k \in K$ such that $\sum_{k \in K} \sum_{i \in N_d'^k} \mu_i^k = 1$.

$$\sum_{k \in K} \sum_{i \in N_d'^k} \frac{\mu_i^k}{\tau\left(Q_i^k V_i^k P_i^k\right)} \cdot \left[\sum_{t=1}^{T} \left[t\left(\left(Q_i^k V_i^k P_i^k\right) - \left(s_{it}^k V_i^k P_i^k\right)\right) - (t-1)\left(\left(Q_i^k V_i^k P_i^k\right) - \left(s_{i(t-1)}^k V_i^k P_i^k\right)\right)\right]\right] \tag{4}$$

The other objective of the restoration process minimizes the total cost associated with the restoration process. The fixed restoration cost for disrupted nodes and links are fn_i^k for $i \in N'^k$ and f_{lij}^k for $(i,j) \in L'^k$, respectively. The unitary flow cost through link $(i,j) \in L'^k$ is c_{ij}^k and p_i^k is the unitary unmet demand cost for node $i \in N'^k$. The binary decision variable z_i^k equals 1 if the node $i \in N'^k$ is restored and 0 otherwise, and y_{ij}^k is also a binary decision variable that equals 1 if link $(i,j) \in L'^k$ is restored and 0 otherwise. Finally, x_{ijt}^k is the non-negative decision variable that represents the total flow through link $(i,j) \in L'^k$ in network $k \in K$ at time $t \in T$. Therefore, the total cost of the restoration process can be represented as Eq. (5).

$$\sum_{k \in K}\left(\sum_{i \in N'^k} fn_i^k z_i^k + \sum_{(i,j) \in L'^k} fl_{ij}^k y_{ij}^k + \sum_{t \in T}\left[\sum_{(i,j) \in L^k} c_{ij}^k x_{ijt}^k + \sum_{i \in N_d^k} p_i^k s_{it}^k V_i^k P_i^k\right]\right) \tag{5}$$

In network $k \in K$, the restoration duration for node $i \in N'^k$ and link $(i,j) \in L'^k$ are dn_i^k and dl_{ij}^k, respectively. The link capacity is u_{ij}^k for link $(i,j) \in L'^k$. The binary decision variable β_{it}^k equals 1 if node $i \in N'^k$ is operational and 0 otherwise, where binary decision variable α_{ijt}^k is 1 if the link $(i,j) \in L'^k$ is operational and 0 otherwise in network $k \in K$ at time $t \in T$. For each network $k \in K$, R^k represents the available work crews or resources that are specific to network k (e.g., in terms of work crew expertise and restoration equipment). The scheduling variables are denoted by binary variables γ_{it}^{kr} and δ_{ijt}^{kr}, respectively, for node $i \in N'^k$ and link $(i,j) \in L'^k$, where they are equal to 1 if restoration of the related component is completed by work crew $r \in R^k$ at time $t \in T$ and 0 otherwise. Finally, the network interdependencies are denoted by $\left((i,k),\left(\bar{i},\bar{k}\right)\right) \in \Psi$ that node $\bar{i} \in N^{\bar{k}}$ in network $\bar{k} \in K$ depends the functionality of node $i \in N^k$ in network $k \in K$. The complete version of the proposed mathematical model is as follows.

$$\max \sum_{k \in K} \sum_{i \in N_d^k} \frac{\mu_i^k}{\tau\left(Q_i^k V_i^k P_i^k\right)}\left[\sum_{t=1}^{T}\left[t\left(\left(Q_i^k V_i^k P_i^k\right) - \left(s_{it}^k V_i^k P_i^k\right)\right) - (t-1)\left(\left(Q_i^k V_i^k P_i^k\right) - \left(s_{i(t-1)}^k V_i^k P_i^k\right)\right)\right]\right]$$

$$\min \sum_{k \in K}\left(\sum_{i \in N'^k} fn_i^k z_i^k + \sum_{(i,j) \in L'^k} fl_{ij}^k y_{ij}^k + \sum_{t \in T}\left[\sum_{(i,j) \in L^k} c_{ij}^k x_{ijt}^k + \sum_{i \in N_d'^k} p_i^k s_{it}^k V_i^k P_i^k\right]\right)$$

Subject to:

$$\sum_{(i,j) \in L^k} x_{ijt}^k \le b_i^k, \qquad \forall i \in N_s^k, k \in K, t \in T$$

$$\sum_{(i,j) \in L^k} x_{ijt}^k - \sum_{(j,i) \in L^k} x_{jit}^k = 0, \qquad \forall i \in N^k \setminus \left\{N_s^k, N_d^k\right\}, k \in K, t \in T$$

$$\sum_{(j,i) \in L^k} x_{jit}^k + s_{it}^k = b_i^k, \qquad \forall i \in N_d^k, k \in K, t \in T$$

$$x_{ijt}^k - u_{ij}^k \le 0, \qquad \forall (i,j) \in L^k, k \in K, t \in T$$

$$x_{ijt}^k - u_{ij}^k \beta_{it}^k \le 0, \qquad \forall (i,j) \in L^k, i \in N^k, k \in K, t \in T$$

$$x_{ijt}^k - u_{ij}^k \beta_{jt}^k \le 0, \qquad \forall (i,j) \in L^k, i \in N^k, k \in K, t \in T$$

$$x_{ijt}^k - u_{ij}^k \alpha_{ijt}^k \le 0, \qquad \forall (i,j) \in L'^k, k \in K, t \in T$$

$$\beta_{it}^{\bar{k}} - \beta_{it}^k \le 0, \qquad \forall \left((i,k),\left(\bar{i},\bar{k}\right)\right) \in \Psi, t \in T$$

$$y_{ij}^k = \sum_{r \in R^k} \sum_{t \in T} \delta_{ijt}^{kr}, \qquad \forall (i,j) \in L'^k, k \in K$$

$$z_i^k = \sum_{r \in R^k} \sum_{t \in T} \gamma_{it}^{kr}, \qquad \forall i \in N'^k, k \in K$$

$$\sum_{(i,j)\in L'^k}\int_{l=t}^{\min\left(\tau,t+dl_{ij}^k-1\right)}\delta_{ijl}^{kr}+\sum_{i\in N'^k}\sum_{l=t}^{\min\left(\tau,t+dn_i^k-1\right)}$$

$$\gamma_{il}^{kr}\le 1, \qquad \forall k\in K, r\in R^k, t\in T$$

$$\alpha_{ijt}^k\le\sum_{r\in R^k}\sum_{l=1}^{t}\delta_{ijl}^{kr}, \qquad \forall(i,j)\in L'^k, k\in K, t\in T$$

$$\beta_{it}^k\le\sum_{r\in R^k}\sum_{l=1}^{t}\gamma_{il}^{kr}, \qquad \forall i\in N'^k, k\in K, t\in T$$

$$\sum_{t=1}^{dl_{ij}^k-1}\alpha_{ijt}^k=0, \qquad \forall(i,j)\in L'^k, k\in K$$

$$\sum_{t=1}^{dn_i^k-1}\beta_{it}^k=0, \qquad \forall i\in N'^k, k\in K$$

$$\sum_{r\in R^k}\sum_{t=1}^{dl_{ij}^k-1}\delta_{ijt}^{kr}=0, \qquad \forall(i,j)\in L'^k, k\in K$$

$$\sum_{r\in R^k}\sum_{t=1}^{dn_i^k-1}\gamma_{it}^{kr}=0, \qquad \forall i\in N'^k, k\in K$$

$$s_{it}^k\ge 0, \qquad \forall i\in N^k, k\in K, t\in T$$

$$x_{ijt}^k\ge 0, \qquad \forall(i,j)\in L^k, k\in K, t\in T$$

$$y_{ij}^k\in\{0,1\}, \qquad \forall(i,j)\in L'^k, k\in K$$

$$z_i^k\in\{0,1\}, \qquad \forall i\in N^k, k\in K$$

$$\alpha_{ijt}^k\in\{0,1\}, \qquad \forall(i,j)\in L'^k, k\in K, t\in T$$

$$\beta_{it}^k\in\{0,1\}, \qquad \forall i\in N^k, k\in K, t\in T$$

$$\delta_{ijt}^{kr}\in\{0,1\}, \qquad \forall(i,j)\in L'^k, k\in K, t\in T, r\in R^k$$

$$\gamma_{it}^{kr}\in\{0,1\}, \qquad \forall i\in N'^k, k\in K, t\in T, r\in R^k$$

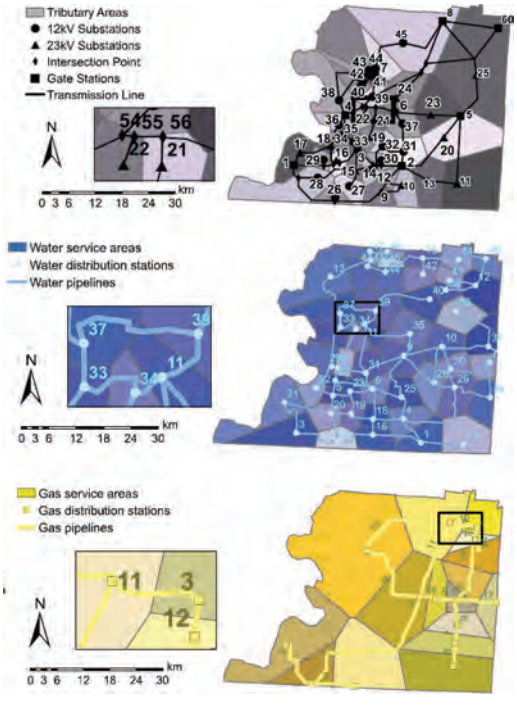


Figure 2. Critical gas, water, and power infrastructure networks of Shelby County, TN, respectively (González et al., 2016).

4 IUSTRATIVE EXAMPLE

In this study, the proposed model is illustrated with data collected for Shelby County, Tennessee in the United State, a location in the New Madrid Seismic Zone at risk of earthquake (González et al., 2016).

Three critical interdependent infrastructure networks, water, gas, and power distribution systems as shown in Figure 2, were examined. We consider a disruptive scenario consisting of a total of 43 disrupted components, 19 of which are demand nodes. We assigned two work crews for each network and time horizon of 23 periods in total.

Figure 3 represents five districts in Shelby County. To relate unmet demand in the three infrastructure networks to adverse community effects, SoVI-Lite was calculated for the five districts guide the restoration process with community resilience in mind. The demand nodes in each district were assigned the different V_i^k value that is specific to that district.

The SoVI-Lite algorithm (Cutter et al., 2011, Evans et al., 2014) was implemented using the available variables for Shelby County to cover the

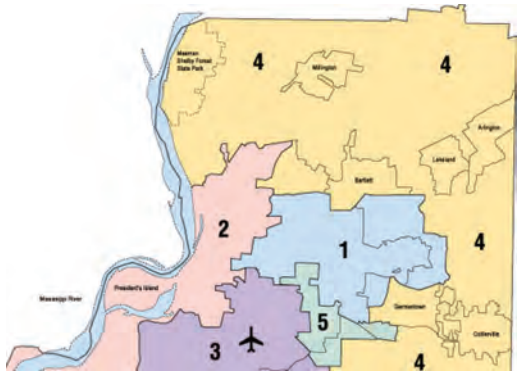


Figure 3. Representation of five districts in Shelby County, TN (www.smartcitymemphis.com).

eleven key factors, shown in Table 1. As shown in Figure 4, District 5 is assigned with the highest scaled SoVI, indicating that it may be given a higher priority in the restoration process. Figure 5 illustrates the exponential transform of the SoVI score, V_i^k, where District 5 stands out substantially from the other districts. Population density, P_i^k,

Table 1. SoVI-Lite algorithm variables for Shelby County, TN.

SoVI-Lite Variables
% of population that lives in poverty
% of population that is over 65
% of population that is under 5
% of population earning less than $75,000 per year
% of population lives in a single-mother household
% of population that is female
% of population that is Hispanic
% of population that is African-American
% of population that is Asian
% of population that isn't high school graduate
% of population that relies on food stamps
% of population that is unemployed
% of population works in low-skilled service jobs
% of population that speaks English as 2nd language

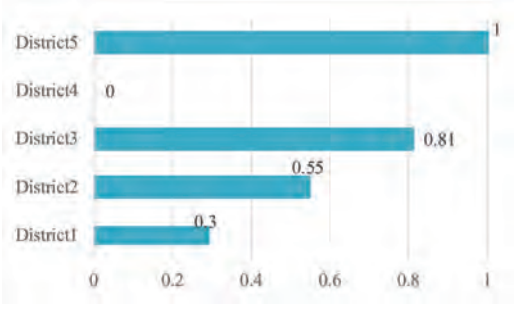


Figure 4. Social-vulnerability indexes for Shelby County, TN districts.

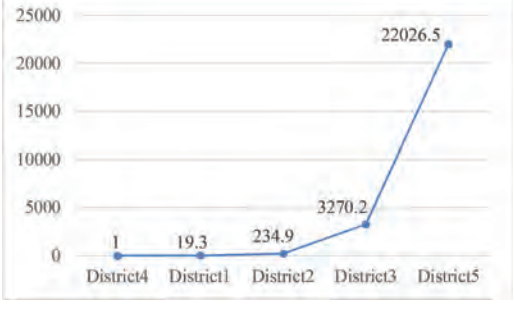


Figure 5. Exponentially increasing social-vulnerability scores for the districts of Shelby County, TN for b = 10.

was also calculated by and assigned to each district through the study.

The multi-objective problem was solved using the ε-constraint method where the resilient constraint was assigned values such that $\varepsilon \in [0,1]$ as in Eq. (6) (Almoghathawi et al., 2017).

$$\sum_{k\in K}\sum_{i\in N'^{k}_{d}}\frac{\mu_i^k}{\tau\left(Q_i^k V_i^k P_i^k\right)} \cdot\left[\sum_{t=1}^{T}\left[t\left(\left(Q_i^k V_i^k P_i^k\right)-\left(s_{it}^k V_i^k P_i^k\right)\right)-(t-1)\left(\left(Q_i^k V_i^k P_i^k\right)-\left(s_{i(t-1)}^k V_i^k P_i^k\right)\right)\right]\right] \geq \varepsilon \quad (6)$$

Tables 2, 3, and 4 are subset comparisons of the restoration optimization model results where second column represents the schedule without SoVI and population density factors and the third column represents the scheduling with consideration of those factors.

"Without SoVI and population density" refers to the removal of the V_i^k and P_i^k terms from the vulnerability objective function. Note how the ranking, that is, the order in which those particular components of the individual networks would be restored, differs when community resilience

Table 2. Water network restoration schedule comparison with and without considering social-vulnerability.

Water network components	Rank without SoVI	Rank with SoVI
Node 29	3	1
Node 37	2	2
Node 11	1	3
Node 36	4	4

Table 3. Power network restoration scheduling comparison with and without considering social-vulnerability.

Power network components	Rank without SoVI	Rank with SoVI
Node 13	3	1
Node 14	4	2
Node 57	2	3
Node 56	1	4

Table 4. Gas network restoration schedule comparison with and without considering social-vulnerability.

Gas network components	Rank without SoVI	Rank with SoVI
Node 9	4	1
Node 14	1	2
Node 8	2	3
Node 6	3	4

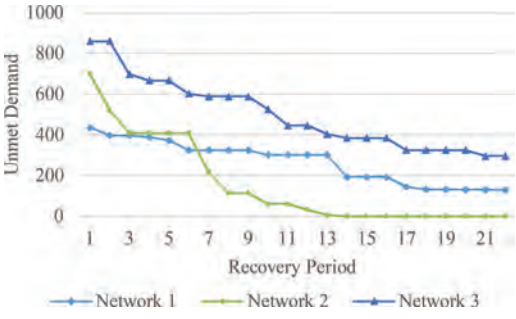


Figure 6. Unmet demand through recovery period, with the consideration of social-vulnerability scores.

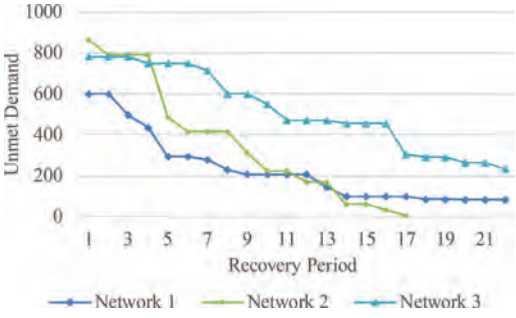


Figure 7. Unmet demand through recovery period, without considering the social-vulnerability scores.

measures are taken into account. The trajectory of recovery, as measured by unmet demand over time, is depicted in Figures 6 and 7, comparison of total unmet demand with and without consideration of the community resilience perspective. Naturally, the figures display a different trajectory as a different priority is given to meeting demand at the demand nodes of the different networks.

5 CONCLUDING REMARKS

Critical infrastructure networks often rely on each other. However, such dependent and interdependent relationships potentially result in these networks being more vulnerable to disruption. However, not only physical infrastructure networks are adversely impacted, as the communities that rely on those networks can also be significantly disrupted. As such, restoration planning and resource allocation should account for disrupted communities in addition to disrupted physical infrastructure networks.

In this paper, we have studied the restoration process planning and scheduling with (i) social vulnerability and (ii) population density in mind. Social vulnerability was calculated using an established index, a variation on the Social Vulnerability Index (Cutter et al. 2003, 2011), that accounts for several age, income, race, and educational attainment dimensions.

This community resilience perspective was added to a multi-objective resilience-driven restoration model using mixed-integer programming. The objective of the model was maximizing the cumulative community resilience of the interdependent networks over time while also the total cost associated with the restoration process is minimized.

For the results of our study, we have found that accounting for community resilience measures impacts the restoration schedule of disrupted infrastructure networks. While more future work will follow up this initial study, at a minimum we have demonstrated that different restoration priorities are found when we account for the vulnerability and the density of the population associated with unsatisfied demand in the interdependent networks. The community resilience perspective should be emphasized in future research.

REFERENCES

Almoghathawi, Y., K. Barker, and L. Albert. 2017. Resilience-Driven Restoration Model for Interdependent Infrastructure Networks. Submitted to *Reliability Engineering and System Safety*.

Barker, K., J.E. Ramirez-Marquez, and C.M. Rocco. 2013. Resilience-based Network Component Importance Measures. *Reliability Engineering and System Safety*, **117**(1): 89–97.

Barker, K., J.H. Lambert, C.W. Zobel, A.H. Tapia, J.E. Ramirez-Marquez, L. Albert, C.D. Nicholson, and C. Caragea. 2017. Defining Resilience Analytics for Interdependent Cyber-Physical-Social Networks. Submitted to *Sustainable and Resilient Infrastructure,* **2**(2): 59–67.

Buldyrev, S.V., R. Parshani, G. Paul, H.E. Stanley, and S. Havlin. 2010. Catastrophic Cascade of Failures in Interdependent Networks. *Nature*, **464**(7291): 1025–1028.

Cutter, S.L. & C. Finch. 2007. Temporal and Spatial Changes in Social Vulnerability to Natural Hazards. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(7): 2301–2306.

Cutter, S.L., B.J. Boruff & W.L. Shirley. 2003. Social Vulnerability to Environmental Hazards. *Social Science Quarterly*, **84**(1): 242–261.

Cutter, S.L., C.T. Emrich & D. Morath. 2011. Social Vulnerability and Place Vulnerability Analysis Methods and Application for Corps Planning: Technical Analyses. In *Social Vulnerability Analysis Methods for Corps Planning*, eds. C.M. Dunning & S. Durden, 74–88. Institute for Water Resources: U.S. Army Corps of Engineers.

Danziger, M.M., L.M. Shekhtman, A. Bashan, Y. Berezin, and S. Havlin. 2016. Vulnerability of Interdependent Networks and Networks of Networks. In *Interconnected Networks* (pg. 79–99). Springer International Publishing, Switzerland.

Eusgeld, I., C. Nan, and S. Dietz. 2011. "System-of-Systems" Approach for Interdependent Critical Infrastructures. *Reliability Engineering and System Safety*, **96**(6): 679–686.

Evans, J.M., D. Hardy, M. Hauer. 2014. Assessing Social Vulnerability using "SoVI-Lite:" A Demonstration Study at Glynn County, GA.

González, A.D., L. Dueñas-Osorio, M. Sánchez-Silva, and A.L. Medaglia. 2016. The Interdependent Network Design Problem for Optimal Infrastructure System Restoration. *Computer-Aided Civil and Infrastructure Engineering*, **31**(5): 334–350.

Henry, D., and J.E. Ramirez-Marquez. 2012. Generic Metrics and Quantitative Approaches for System Resilience as a Function of Time. *Reliability Engineering and System Safety*, **99**(1): 114–122.

Jönsson, H., J. Johansson and H. Johansson. 2008. Identifying Critical Components in Technical Infrastructure Networks. *Journal of Risk and Reliability*, **222**(2): 235–243.

Little, R.G. 2002. Controlling Cascading Failure: Understanding the Vulnerabilities of Interconnected Infrastructures. *Journal of Urban Technology*, **9**(1): 109–123.

Ouyang, M. 2014. Review on Modeling and Simulation of Interdependent Critical Infrastructure Systems. *Reliability Engineering and System Safety*, **121**: 43–60.

Rinaldi, S.M., J.P. Peerenboom & T.K. Kelly. 2001. Identifying, Understanding and Analyzing Critical Infrastructure Interdependencies. *IEEE Control Systems Magazine*, **21**(6): 11–25.

Rose, A. 2007. Economic Resilience to Natural and Man-made Disasters: Multidisciplinary Origins and Contextual Dimensions. *Environmental Hazards*, **7**(4): 383–398.

The Report of the President's Commission on Critical Infrastructure Protection. 1997. *Critical Foundation: Protecting America's Infrastructure*. USA.

Wallace, W.A., D.M. Mendoca, E.E. Lee, J.E. Mitchell, J.H. Chow Wallace and J.L. Monday. 2003. Managing Disruptions to Critical Interdependent Infrastructures in the Context of the 2001 World Trade Center Attack. In *Beyond September 11th: An Account of Post-Disaster Research*, Special Publication **39**: 165–198, Natural Hazards Research and Applications Information Center, University of Colorado.

Wu, B., A. Tang, and J. Wu. 2016. Modeling Cascading Failures in Interdependent Infrastructures under Terrorist Attacks. *Reliability Engineering and System Safety*, **147**: 1–8.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Resilience assessment of smart critical infrastructures based on indicators

K. Øien & L. Bodsberg
SINTEF Technology and Society, Trondheim, Norway

A. Jovanović
European Virtual Institute for Integrated Risk Management, Stuttgart, Germany
Steinbeis Advanced Risk Technologies GmbH, Stuttgart, Germany

ABSTRACT: The resilience of modern societies is to a large degree determined by the resilience of their Critical Infrastructures (CI). These infrastructures are critical because interruptions not only influence the infrastructures themselves, but loss of functionality has secondary effects on the society. The use of smart technologies makes these "Smart" CIs (i.e. SCIs) increasingly interdependent and vulnerable to various hazards, such as terror attacks, cyber-attacks and extreme weather. The EU H2020 research project SmartResilience has developed a baseline resilience assessment method, which measures the level of resilience indirectly through a selection of resilience indicators considered relevant by the user of the SCI in question. Other methods have also been developed in SmartResilience, but this paper focus on the development and application of the baseline resilience assessment method and the development and collection of resilience indicators used in the assessment method. The application is demonstrated using a production facility as a case.

1 INTRODUCTION

The power grid in Ukraine was cyber-attacked both in 2015 and 2017. The attack in 2015 was a complex and pervasive attack on three energy distribution companies, resulting in about 230 thousand people being left without electricity for a period from 1 to 6 hours (Wikipedia 2017). Energy supply systems, such as those attacked in Ukraine, are examples of critical infrastructures (CIs); *critical* because their functions are vital for the society.

Smart technologies are introduced in infrastructures to maximize the service they provide using intelligent systems. Thus, the term S*mart* Critical Infrastructure (SCI) is introduced. However, smart features may also make the SCIs more vulnerable, e.g. by providing a gateway for hackers and cyber-terrorists.

The need to defend these SCIs has been recognized for decades through e.g. Critical Infrastructure Protection (CIP) programs. However, in recent years, it has been realized that with increasingly complex and interdependent infrastructure systems, CIP is not enough (HSAC 2006). It is not enough to focus on protection of a CI from events like cyber-attacks, terror attacks and extreme weather, because the complexity and interdependencies makes it virtual impossible to foresee and prevent all scenarios, and when they occur—no matter how unlikely—it is vital for society that the loss of functionality is minimized, e.g. that the CIs are up and running as soon as possible after an event.

A shift of the focus from CIP towards CIR, i.e. Critical Infrastructure *Resilience* has been observed. "Overall, a resilience-based approach for CI is an approach that is gradually adopted by nations in order to face the challenges and costs of achieving maximum protection in an increasingly complex environment and to overcome limitations of the traditional scenario-based risk management approach, where the organization may lack capabilities to face risk from unknown or unforeseen threats and vulnerabilities" (Setola et al. 2016).

Resilience is not a straight-forward term. It has many different applications and a broad scope. A helpful review paper providing insights into the term and its history is Alexander (2013). Suffice to state here is that although the term was unfamiliar within risk of critical infrastructures in the US some ten years ago (HSAC 2006), it is now a well-recognized term. Resilience is also a familiar everyday term in English speaking countries, but it is not easily understood by lay people when translated to other languages. In addition, the CIR approach is relatively new in the EU compared to the US. This

gives some challenges for the implementation of CIR in EU and the single EU member states.

Recognizing the challenges with the term resilience, the questions are still: How can we make a system like the energy system in Ukraine, and other SCIs, resilient against cyber-attacks and other relevant threats? How can we know—and measure—the level of resilience of an SCI? These are the challenges that the EU H2020 project SmartResilience (2016) is set out to solve. It answers the DRS-14 call, which explicitly asks for an indicator-based approach.

Several methods and tools for assessing and monitoring resilience are developed in the SmartResilience project. In this paper, we present the baseline resilience assessment method measuring the Resilience Level (RIL) of SCIs through resilience indicators. We denote this as the "RIL method" in the following. It is based on review, adaptation and further development of relevant reference methods having their roots in high reliability theory (Wreathall 2006), resilience engineering (Woods 2006) and critical infrastructure resilience (Fisher et al. 2010).

The resilience indicators have been developed (identified and/or proposed) mainly by the case study partners in the SmartResilience project, covering a range of different critical infrastructures. They are stored in a database as "candidate" resilience indicators, i.e. the users select the most relevant indicators for their case from the candidates in the database, or add new indicators, when necessary.

Based on the selected set of resilience indicators, the RIL method provides a level of resilience on a scale from E (worst) to A (best) for one specific SCI, or several SCIs, within an area. In addition to an overall level of resilience, that can be trended periodically, the results point to areas where improvements are most needed. In this paper, the application of the RIL method is demonstrated for a production facility.

The description of the development of the RIL method and the resilience indicators are based on Øien et al. (2017a-c). Earlier versions of the Smart-Resilience methodology are also presented in Jovanović et al. (2017a; 2018).

1.1 *Concepts and definitions*

In the SmartResilience project, the *resilience* of an infrastructure is defined as: "The ability to anticipate possible adverse scenarios/events (including the new/emerging ones) representing threats and leading to possible disruptions in operation/functionality of the infrastructure, prepare for them, withstand/absorb their impacts, recover from disruptions caused by them and adapt to the changing conditions" (Jovanović et al. 2016).

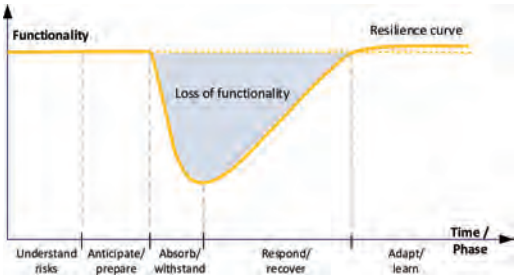


Figure 1. Resilience phases in the resilience curve/cycle.

Based on this definition, we derive at the following five *phases* of the resilience curve/cycle: understand risks, anticipate/prepare, absorb/withstand, respond/recover, and adapt/learn. The five phases, representing the main resilience attributes in SmartResilience, are illustrated in Figure 1.

Each of the phases are measured by indicators through the most important "issues" affecting each of the phases.

An *issue* is a very general term referring to anything (factors, conditions, functions, actions, capacities, capabilities, etc.) that is important in order to be resilient against severe threats such as terror attacks, cyber threats and extreme weather. It is *what* is important, and it is allocated to one of the five phases in the resilience cycle. E.g., it can be "training" performed in the anticipate/prepare phase.

An *indicator* is the description of *how* to measure an issue. Any type/form of indicators are considered appropriate in the RIL method, meaning that they can be yes/no questions, numbers, percentages, frequencies, or some other type. E.g., it can be "percentage of personnel in a certain response team taken a certain course".

2 METHOD DEVELOPMENT

The RIL method is an indicator based approach consisting of two main parts; the resilience assessment method itself and the indicators used to measure the resilience level. The development of the two parts are described in the following.

2.1 *Resilience assessment method*

The RIL method has its roots in high reliability organization theory (EPRI 2000, 2001) and resilience engineering (Øien 2010, 2012; Øien & Nielsen, 2012; Øien et al. 2012), but also more resent resilience developments within critical infrastructures, especially in the US (e.g. Petit et al. 2013, Linkov et al. 2014).

2.1.1 *The ANL method*

The Argonne National Laboratory (ANL) method for assessing a resilience index (RI) (Fisher et al., 2010), or a resilience measurement index (RMI), as it is termed in the most recent version (Petit et al. 2013), is structured in five (or six) levels, providing indicators on the lowest level. A similar hierarchy is used in the SmartResilience project for assessing resilience levels, entering the indicators on level 6. The structure is comparable in the two approaches, and many of the resilience attributes are the same; however, the level at which the various resilience attributes are found, differs between these two methods.

2.1.2 *The LIOH method*

The Leading Indicators of Organizational Health (LIOH) method focused on developing indicators for a set of seven themes important for the "health" of a nuclear power plant, some of which have their roots from the research on high reliability organizations (HRO) (Wreathall 2006). They also formed part of the basis for factors considered important in resilience engineering. In addition to *themes*, LIOH uses *issues* and *indicators* as the three levels in the structure of the method.

The LIOH method is a contributory-based method in which the users of the indicators take part in workshops and define their own issues (general and nuclear power plant—NPP—specific) for each theme, and for each issue they define indicators. There are no predefined examples of issues prior to the workshops, and no proposals or "candidate" indicators are in place prior to the workshops.

The case studies of the LIOH method show that there is often only one level of issues used, i.e. the issues are not divided into general and NPP issues (EPRI 2000, 2001). A second observation is that the results (the issues and indicators defined) from identical power plant units are very different. The reason for this difference is that there is no guidance with respect to issues and indicators (no a priori "candidates"), and that there have been different participants in the workshops in each of the case studies.

2.1.3 *The REWI method*

The idea of combining the issues into one common level was brought further to the Resilience-based Early Warning Indicator (REWI) method (Øien et al. 2010, 2012); using three levels to identify early warning indicators for resilience, i.e. starting with resilience attributes, followed by issues important for these resilience attributes, and finally developing indicators to measure the issues. In REWI, the level of resilience attributes is not termed themes as in LIOH, but rather *contributing success factors* (CSFs). Thus, the structure consists of *CSFs*, *issues* and *indicators*.

The CSFs are structured in two levels, of which the lowest level consists of eight factors, or resilience attributes. The CSFs at the first level are: risk awareness, response capacity, and support. The CSFs at the second level are: risk understanding, anticipation, attention, response, robustness (of response), resourcefulness/rapidity, decision support, redundancy (for support). The CSFs represent the REWI operationalization of the concept of resilience, similar as themes are used in LIOH and phases are used in the Smart-Resilience project. The CSFs are partly, but not entirely, sequential. For each CSF, there is a set of issues contributing to the fulfillment of the goals of the CSF. There is only one level of issues—denoted general issues—for which indicators are developed. The CSFs were developed based on a literature review and an empirical study on successful recovery of high-risk incidents; thus, the term *contributing success factors* (Størseth et al. 2009).

The REWI method consists of a predefined set of issues and a set of candidate indicators for each issue. This is a main difference compared to the LIOH method, and makes it less "open ended". However, it is still a contributory-based method and new issues may be added. The predefined set of issues and sets of candidate indicators "forces" the participants to assess the a priori set of general issues and candidate indicators. Thus, it counteracts the tendency to identify indicators during workshops just as random "indicators of the day".

The issues are just candidates, which may be considered appropriate or rejected, and additional issues may be included. After selecting the important issues, the next step is to consider how to measure them. How well are we doing with the selected issues? What would tell me that we are doing well (or have problems) with a specific issue? What information do we have about this? This is the role of the indicators.

The issues we try to measure, and the indicators we use to measure the issues, are two different things. The indicator will typically be described as a number, ratio, score on some scale, or similar. Without this type of specification or operationalization, we are left with just a theoretical issue. We cannot start with the indicators either, since we need to know what we want to measure (i.e. the issues) and why.

2.1.4 *The SmartResilience RIL method*

Like the LIOH method and the REWI method, the RIL method uses *issues* and *indicators* on the two lowest levels of the structure, whereas *phases* are used on the next higher level, compared to themes in LIOH and contributing success factors

in REWI. For each of the phases, issues that are important for them are identified, and indicators to measure the issues are developed.

In addition, the issues (and corresponding indicators) may be structured according to five *dimensions*, which are system/physical, information/data, organizational/business, societal/political, and cognitive/decision-making (Jovanović et al. 2016). The phases and dimensions forms what is denoted the Resilience Matrix, commonly used in several resilience assessment methods (e.g. Linkov et al. 2014). However, in the SmartResilience project, dimensions are only optionally used for structuring and triggering the identification of issues and indicators. Only phases are directly included in the quantification, i.e. it is the columns in the Resilience Matrix that are of interest, not the rows (or the single cells) in the matrix.

The SmartResilience RIL method has been developed through several iterations, including input from user requirements (Buhr et al. 2016), test case use, and feedback from case study partners in workshops and through a questionnaire (Jovanović et al. 2017b). A description of the resulting method is provided in Section 3.1.

2.2 *Resilience indicators*

The candidate issues and indicators collected in the SmartResilience project are to a large degree provided by the partners from existing standards, guidelines and reports within the areas of risk, safety, security, crisis management, business continuity and similar domains.

Resilience is considered an "umbrella" term (Setola et al. 2016), covering all the mentioned domains; thus, the term *resilience indicators* may include risk indicators, safety indicators, etc. The umbrella concept is illustrated in Figure 2.

In addition to standards, guidelines and reports, some indicators are based on what the case study providers already are using, and some indicators are developed as part of the project. Figure 2 also illustrates that the resilience concept in general and the resilience indicators, aim at capturing the unexpected, by using the metaphor "rain from a blue sky".

Candidate issues and indicators are stored in a database, and reported in Øien et al. (2017a), representing the status of the collected issues and indicators approximately half way through the project.

In addition, Øien et al. (2017c) present generic candidate issues (without indicators) covering more genuine resilience issues, i.e. capturing topics typically discussed in the resilience literature. The two main sources are the guideline for implementing the REWI method (Øien et al. 2012), and an emergency preparedness plan developed by SINTEF

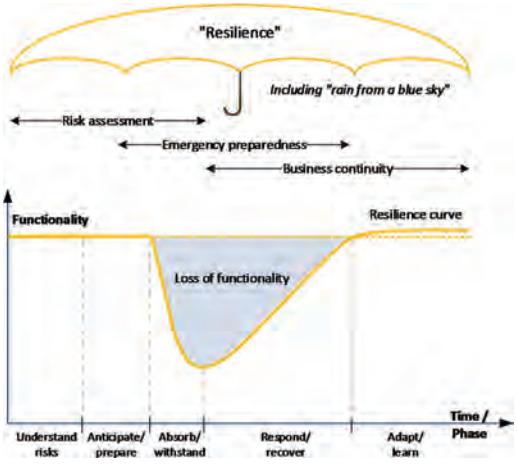


Figure 2. Resilience as an "umbrella" term.

(2014). Some issues are derived from IMPROVER (2016), a few from RESILENS (2016a, b), and the rest is based on input from SINTEF as part of the SmartResilience project. Some issues are taken directly from the original sources, whereas others are slightly adapted. Only for those generic candidate issues that are considered relevant for each user, indicators need to be developed.

A presentation of the collected candidate issues and indicators is provided in Section 3.2.

3 RESULTS

3.1 *The SmartResilience RIL Method*

3.1.1 *Model*

The three lower levels (level 4–6) of the hierarchical model are phases, issues and indicators, as described in Section 2.1. In addition, the overall structure consists of three more levels. The first level is the area level, e.g. a city. The second level consists of the smart critical infrastructures (SCIs), and the third level defines the threats. This is illustrated in Figure 3.

3.1.2 *Method steps*

At each level, the scores—alternatively combined with weights—corresponds to a certain resilience level (RIL) given by a character E-A, where E is worst, and A is best. A weighted score between 0–1 corresponds to resilience level E, a weighted score 1–2 corresponds to resilience level D, and so on.

The method steps are as follows:

Step 1: Select the area, e.g. a smart city
Step 2: Select the relevant SCIs for the area
Step 3: Select relevant threats for each SCI

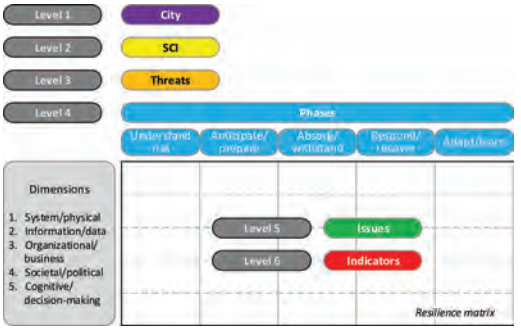


Figure 3. The six levels in the hierarchical model.

Step 4: Consider each phase for each threat
Step 5: Define the issues within each phase
Step 6: Search for the indicators for each issue
Step 7: Determine the range of values for each indicator (and optionally assign weights)
Step 8: Assign values to the indicators
Step 9: Perform the calculations (scores and RILs)
Step 10: Use the results and make decisions

The method steps have been described in Jovanović et al. (2017a) and we will only focus on the changes that have been made lately. This apply to Steps 7 and 9.

The indicators real values are collected and transformed to a *score* (or rating) on a scale from 0 (worst) to 5 (best). This requires the determination of best and the worst values for each indicator, i.e. Step 7. This part is simplified by using five categories, or value ranges (Øien et al. 2017b).

At every level, there is a possibility to give *weights*; however, we recommend being restrictive with the use of different weights. It is challenging to substantiate the assignment of weights (who and how), and the assignment itself can easily be criticized. Thus, equal weights are the default values at all levels. However, if different weights are considered necessary, we now propose using a simple type of pairwise comparison (Øien et al. 2017b). It can also be considered to include weights after gaining some experience, i.e. "tuning" the assessment.

In Step 8, the values are assigned to the indicators, i.e. the measurement itself is performed, and in Step 9 scores are calculated, first on the indicator level, and then aggregated upwards through all levels until the area level. On each level in the hierarchy, the scores can be transformed to resilience levels. This is new, and also the use of characters E-A is new; previously a scale 0–10 was used for RILs, and the transformation from scores to RILs only took place at the phase level (Øien et al. 2017b).

The use of the results, in Step 10, is described in Section 3.3.

3.1.3 *Special topics*

The way cascading effects, dependencies and interdependencies, interoperability, and smartness opportunities and vulnerabilities are treated in the RIL method is briefly described below. We strive for a good balance between the comprehensiveness of the analysis framework and the simplicity of understanding and using the framework. Thus, the specific topics have been addressed explicitly, but relatively simplistic.

Cascading effects where the SCI in question is affected from the outside should be *treated as a specific threat* e.g. toxic cloud, flooding, etc. If the effect is in the form of loss of service, then it is treated as dependencies as part of Step 5, i.e. explicitly as issues. Internal escalation of an event is also treated explicitly as issues (Step 5) reflecting the required safety systems or barriers needed to prevent escalation.

Critical infrastructures, or other infrastructures, services or systems that the SCI are dependent on, should be *addressed explicitly as issues* in the relevant phases for the relevant threats. This could e.g. be the need for redundant energy supply or communication networks. Interdependencies are treated in the same way. The difference is that the SCIs being dependent on "your" SCI, need to explicitly include this as issues in their resilience assessment.

If interoperability is an internal concern e.g. interoperable communication systems, then it should be *treated as an issue*. If it is related to external interoperability in the sense of external backup systems, e.g. "bus for train", then it should be included explicitly as an issue (e.g. cooperation agreements) if this is the responsibility of the SCI being assessed.

The relevance of smartness opportunities and smartness vulnerabilities related to smart features (sensors, gateways, processors, actuators, etc.) should be considered *explicitly as issues* in each phase.

3.2 *The collection of issues and indicators*

Øien et al. (2017c) describes candidate resilience issues and indicators to be used when assessing, predicting and monitoring resilience of Smart Critical Infrastructures (SCIs). A total of 233 candidate issues and 1264 indicators are provided for various threats, SCIs and the five phases of the resilience cycle.

Table 1 shows the number of issues and indicators in the five phases defined in the Smart-Resilience project. In addition, some issues and indicators are considered relevant for all phases.

Table 1. No. of issues and indicators in each phase.

Phase		Issues	Indicators
Phase I	Understand risks	46	226
Phase II	Anticipate/prepare	93	520
Phase III	Absorb/withstand	45	236
Phase IV	Respond/recover	39	180
Phase V	Adapt/learn	20	95
Relevant for all phases		10	182

Although a substantial number of issues and indicators have been collected, they will never be complete and they are just candidates. There will always be a need for additional and/or more relevant issues and indicators for each specific user; and in the end, it is always the user that is responsible for finding a relevant and complete set of issues and indicators for his/her own case study.

Issues are essential in order to focus on those aspects that are most important to measure. Therefore, issues are considered first, and then indicators to measure the selected issues are established. Focusing on indicators first may result in important aspects (issues) being missed and not measured.

The importance of issues is also reflected by the 143 generic candidate issues provided in Øien et al. (2016c).

3.3 *Results obtained by using the method*

From the overall result, i.e. the resilience level of an area or a specific SCI, we can "drill-down" through the levels 2–6 for detailed results, which can be used in Step 10, together with the overall result. We do *not* have "just one number" (the overall resilience level).

There are many possibilities for use of the results, including:

1. Following up own development over time (trending) and analyse status
2. Comparing with others (benchmarking)
3. Providing overview of strengths and weaknesses and point at improvement needs
4. Making any gaps visible (lack of relevant indicators)

3.4 *Example*

To explain the assessment and calculations performed, Table 2 shows an extract of an example RIL assessment of a production facility within the chemical industry. The threat considered is terrorist attack (threat 1), and only the first phase (phase I) is shown.

Issues and indicators (I & I) IDs are listed in the first column. The indicators for the first issue (I.1) are: Does a safety risk register exist? (I.1.1); Is this registry used in decision making? (I.1.2); Is a frequency for updating the registry defined? (I.1.3). The second issue (I.2) only have one indicator: Is a procedure for MOC established? (I.2.1). The third issue (I.3) has the following three indicators: Does an accident/incident register exist? (I.3.1); Frequency of communication about incidents (I.3.2); Percentage of employees informed about incidents (I.3.3).

Table 2. Calculations on indicator level (example).

Indicator scores, weights and RILs					
I & I	Real value	Score value	RIL	Weight	Weighted score
I.1 Safety risk registry					
I.1.1	Y	5	A	0,33	1,67
I.1.2	Y	5	A	0,33	1.67
I.1.3	N	0	E	0,33	0,00
I.2 Management of change—MOC					
I.2.1	N	0	E	1,00	0,00
I.3 Register of accidents/incidents					
I.3.1	Y	5	A	0,33	1,67
I.3.2	1/6 mth	1,5	D	0,33	0,50
I.3.3	80%	3,5	B	0,33	1,17

Each indicator is measured, i.e. providing the real values for the indicators, whether it is yes/no questions, frequencies, percentages, or some other type of indicator. Based on the real value and the predetermined range of values, from worst to best (not shown in Table 2), an indicator score value is calculated. This value can be transformed to an indicator resilience level, from E (worst) to A (best) according to a predefined scale. Weights are determined, and the default values are equal weights. By multiplying the indicator scores with the indicator weights, the indicator weighted score is obtained in the last column. The indicator weighted scores are brought to the next level in the calculations, i.e. the issue level (level 5), where similar calculations are performed obtaining issue weighted scores, and so on, all the way to the area level (level 1).

The calculations gave an overall score on area level of 3,06 corresponding to RIL = B (Øien et al. 2017b).

The overall result just represents one aggregated character or value, which provides limited information. We need to "drill down" in the levels beneath, to reveal more detailed information about the various contributions to the overall result. One example of results on level 2 (SCI level) is shown in Figure 4. Here it is revealed that the threats with the lowest scores are *Threat 1 – Terrorist attack* and *Threat 2 – Natural threats*, both with a score of 2,64, which would be natural to look further into to improve resilience.

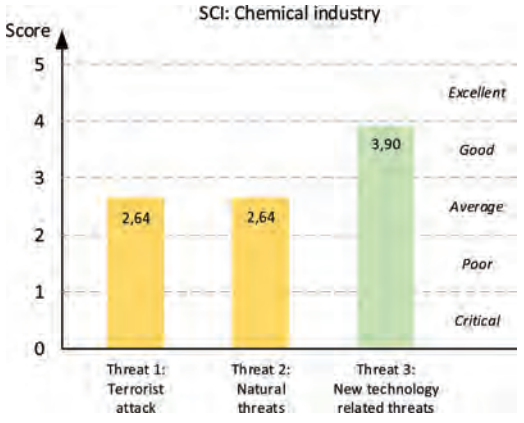


Figure 4. Resilience status at threat level (example).

4 DISCUSSIONS AND CONCLUSIONS

The SmartResilience RIL method helps to understand how resilient the SCIs are against specific types of threats and what measures could help improve their resilience. The results show the level of resilience (RIL) and where improvements are most needed ("drill-down"), emphasizing and fostering a continuous improvement mindset through regularly (typically yearly) updated assessments.

The resilience assessment uses a holistic ("umbrella") approach that goes beyond traditional risk of known events, emergency preparedness, crisis management, and business continuity. It covers e.g. preparing for the unforeseen, imagination, vigilance, flexibility, improvisation, recovery including business continuity aspects, and learning and adaptation.

4.1 *How to use the SmartResilience RIL method*

There are two main options for resilience assessment; internal self-assessment and external assessor audit. One main reason for using external assessments is the possibility for bench-marking between similar SCIs or even areas/cities with similar SCIs. To ensure comparability, it is important to use the same threats, issues and indicators, with the same range of indicator values, weights and similar requirements for collecting data for the indicators. This is possible to achieve (at least for a simple assessment/audit), but may not prove very useful for each individual user.

It is also possible to make user adaptation and customize the set of threats, issues and indicators, ranges of indicator values, weights and so on, e.g. by allowing to reject or add new indicators. However, the more the "dynamic checklists" (the tool used in the SmartResilience project) of threats, issues and indicators are adapted to take user requirements into account, the less comparable they will be.

Internal self-assessment can also be performed using similar checklists as an external assessor would use; however, if the focus is not on benchmarking and comparing with others, the assessment can be adapted to the specific requirements of each user. This will ensure a more relevant and accurate assessment useful for trending own development over time. A user customized self-assessment approach requires more engagement from the users. On one hand this is positive, since the users will take more ownership to the analysis framework and the results; however, on the other hand it will require more resources compared to an external assessment using a standardized framework.

4.2 *Usefulness of the SmartResilience RIL method*

The purpose of assessing resilience is to obtain a measure of how resilient a city or an individual SCI are against severe threats such as terror attacks, cyber-attacks and extreme weather. Assessing RIL provides a baseline assessment of resilience that gives insight on *status and improvement needs* to increase or maintain a high level of resilience.

A RIL assessment goes beyond traditional risk assessments by focusing on unknown and unforeseen events, and the capability to recover from events. This is achieved by capturing the time dimension through (five) distinct phases, incorporating e.g. emergency response and business continuity. A RIL assessment complements risk assessment; it is not a substitute for risk assessment. Risk assessments also provide valuable input to a RIL assessment, specially to phase I "Understanding risks".

An important purpose of a RIL assessment is to identify potential problems before they occur, so that risk reducing measures may be planned and implemented as needed, regardless of the likelihood of events. Most SCIs in the world have never, and will never, experience an extreme event. Still it is possible to assess the RIL, i.e. the level of risk understanding, anticipation and preparation, the capability to absorb and withstand, to respond and recover, and the abilities to learn and adapt. With a high RIL, it is less likely to experience adverse consequences due to an extreme event, and should it occur, then disruptions are likely to be less severe.

4.3 *Conclusions*

The SmartResilience project has developed a method for assessing resilience of SCIs with respect to specific type of threats on a scale from E

(worst) to A (best). An overall RIL is obtained by combining resilience levels for five main attributes/phases of resilience for each threat. For each phase, the user/analyst must identify the most important "issues" affecting SCI resilience and for each issue select relevant indicators, indicator range values, and perform calculations. The Smart-Resilience project has provided candidate issues and indicators for various SCIs that may be used as a starting point for identifying issues and indicators for resilience assessment of specific SCIs. This baseline resilience assessment can be used for trending as well as identifying improvement needs.

The resilience curve, describing the SCI functionality as a function of time, before, during and after an adverse event, is treated as a conceptual model, i.e. the method does not consider the exact shape, size or area of the curve directly. It is an indirect measurement. For direct assessment of SCI resilience, the SmartResilience project has developed a functionality assessment method with respect to specific threat scenarios. This alternative method provides a quantitative measure of loss of SCI functionality as a function of time addressing explicitly the resilience curve.

ACKNOWLEDGEMENTS

The contribution is based on the Grant Agreement No. 700621 supporting the work on the SmartResilience project provided by the Research Executive Agency (REA) ('the Agency'), under the power delegated by the European Commission ('the Commission'). This support is gladly acknowledged, together with the support from all the partners in the project.

REFERENCES

Alexander, E. 2013. Resilience and disaster risk reduction: an etymological journey. *Nat Hazard Earth Syst Sci* 13:2707–16.

Buhr, K., Karlsson, A., Sanne, J.M., Albrecht, N., Santamaría, N.A., Antonsen, S., … Warkentin, S. 2016. SmartResilience D1.3: *End users' challenges, needs and requirements for assessing resilience*, EU project SmartResilience, Project No. 700621 (2016–2019), Contact: EU-VRi, Stuttgart, Germany.

EPRI 2000. *Guidelines for trial use of leading indicators of human performance: the human performance assistance package*. EPRI (U.S. Electric Power Research Institute), Palo Alto, CA, 10000647.

EPRI 2001. *Final report on leading indicators of human performance*. EPRI, Palo Alto, CA, and the U.S. Department of Energy, Washington, DC, 1003033.

Fisher R.E., Bassett G.W., Buehring W.A., Collins M.J., Dickinson D.C., Eaton L.K., … Peerenboom J.P. 2010. *Constructing a resilience index for the enhanced critical infrastructure protection program*, Argonne National Laboratory, Decision and Information Sciences Division, ANL/DIS-10–9, Argonne, IL, USA.

HSAC 2006. *Report of the Critical Infrastructure Task Force*. Homeland Security Advisory Council.

IMPROVER Consortium 2016. Deliverable 2.2: *Report of criteria for evaluating resilience*. www.improver-project.eu/2016/06/23/deliverable-2-2-report-of-criteria-for-evaluat-ingresilience/.

Jovanović, A., Øien, K., Choudhary, A. 2018. An indicator-based approach to assessing resilience of smart critical infrastructures. In A. Fekete & F. Fiedrich (eds), *Urban disaster resilience and security: addressing risks in societies*. Springer.

Jovanović, A., Klimek, P., Choudhary, A., Schmid, N., Linkov, I., Øien, K., … Lieberz, D. 2016. SmartResilience D1.2: *Analysis of existing assessment resilience approaches, indicators and data sources: Usability and limitations of existing indicators for assessing, predicting and monitoring critical infrastructure resilience*, EU project SmartResilience, Project No. 700621 (2016–2019), Contact: EU-VRi, Stuttgart, Germany.

Jovanović, A., Choudhary, A., Tetlak, K., Albrecht, N., Roque, R., Klimek, P., … Bergfors, L. 2017b. SmartResilience D5.1: *Report on the results of the interactive workshop*, EU project SmartResilience, Project No. 700621 (2016–2019), Contact: EU-VRi, Stuttgart, Germany.

Jovanović A., Quintero, F., Choudhary, A. 2017a. *Use of safety-related indicators in resilience assessment of Smart Critical Infrastructures (SCIs)*, ESREL 2017 – European Safety and Reliability Conference, 18–22 June 2017, Portoroz, Slovenia.

Linkov, I. et al. (2014). Changing the Resilience Paradigm. *Nature Climate Change* 4(6), 407–409. Retrieved from http://www.nature.com/doifinder/10.1038/nclimate2227.

Øien, K. 2010. *Remote operation in environmentally sensitive areas; development of early warning indicators*. 2nd iNTeg-Risk Conference, Stuttgart, Germany, 15–16 June 2010.

Øien, K. 2013. Remote operation in environmentally sensitive areas: development of early warning indicators. *J Risk Res* 16(3–4):323–336.

Øien, K., Bodsberg, L., Hoem, Å., Øren, A., Grøtan, T. O., Jovanović, A., … Tuurna, S. 2017c. SmartResilience D4.1: *Supervised RIs: Defining resilience indicators based on risk assessment frameworks*, EU project SmartResilience, Project No. 700621 (2016–2019), Contact: EU-VRi, Stuttgart, Germany.

Øien, K., Jovanović, A., Grøtan, T.O., Choudhary, A., Øren, A., Tetlak, K., … Jelic, M. 2017a. SmartResilience D3.2: *Assessing resilience of SCIs based on indicators*, EU project SmartResilience, Project No. 700621 (2016–2019), Contact: EU-VRi, Stuttgart, Germany.

Øien, K., Jovanović, A., Bodsberg, L., Øren, A., Choudhary, A., Sanne, J., … Szekely, Z. 2017b. SmartResilience D3.6 draft report: *Guideline for assessing, predicting and monitoring resilience of Smart Critical Infrastructures (SCIs)*, EU project SmartResilience, Project No. 700621 (2016–2019), Contact: EU-VRi, Stuttgart, Germany.

Øien, K., Massaiu, S., Tinmannsvik, R.K. 2012. *Guideline for implementing the REWI method; resilience based Early Warning Indicators*. SINTEF report A22026, Trondheim, Norway.

Øien, K., Massaiu, S., Tinmannsvik, R.K., Størseth, F. 2010. *Development of early warning indicators based on Resilience Engineering*. International Conference on Probabilistic Safety Assessment and Management (PSAM10), Seattle, USA, 7–11 June 2010.

Øien, K. & Nielsen, L. 2012. *Proactive resilience based indicators: the case of the Deepwater Horizon accident*. SPE/APPEA international conference on health, safety and environment in oil & gas exploration and production, Perth, Australia, 11–13 September 2012.

Petit, F.D., Bassett, G.W., Black, R., Buehring, W.A., Collins, M.J., Dickinson, D.C., … Peerenboom J.P. 2013. *Resilience Measurement Index: An Indicator of Critical Infrastructure Resilience*, Argonne National Laboratory, Decision and Information Sciences Division, ANL/DIS-13-01, Argonne, IL, USA.

RESILENS Consortium 2016a. *Qualitative, Semi-Quantitative and Quantitative Methods and Measures for Resilience Assessment and Enhancement*. Ireland. Retrieved from http://resilens.eu/wp-content/uploads/2016/08/D2.2-Methods-for-Resilience-Assessment-Final.pdf.

RESILENS Consortium 2016b. *Resilience Management Matrix and Audit Toolkit*. Ireland. Retrieved from http://resilens.eu/wp-content/uploads/2016/06/D2.3-Resilience-Management-Matrix-and-Audit-Toolkit.pdf.

Setola, R., Luiijf, E., Theocharidou, M. 2016. Critical Infrastructures, Protection and Resilience. In R. Setola et al. (eds.), *Managing the Complexity of Critical Infrastructures, Studies in Systems, Decision and Control 90*, DOI 10.1007/978-3-319-51043-9-1.

SINTEF 2014. Emergency preparedness plan, Restricted.

SmartResilience 2016. *Smart resilience indicators for smart critical infrastructures* – the European Union's horizon 2020 research and innovation programme, grant agreement No 700621 (2016–2019). Coordinator: EU-VRi, www. smartresilience.eu-vri.eu.

Størseth, F., Tinmannsvik, R.K., Øien, K. 2009. *Building safety by resilient organization—a case specific approach*. The European Safety and Reliability Conference (ESREL '09), Prague, Czech Republic, 7–10 September 2009.

Wikipedia, 2017. https://en.wikipedia.org/wiki/December_2015_Ukraine_power_grid_cyberattack, Accessed on Dec 4, 2017.

Woods, D.D. 2006. Essential Characteristics of Resilience. In: N. Leveson, E. Hollnagel, and D.D. Woods (eds), *Resilience engineering: concepts and precepts:* 21–34. Aldershot: Ashgate.

Wreathall, J. 2006. Properties of resilient organizations: an initial view. In: N. Leveson, E. Hollnagel, and D.D. Woods (eds), *Resilience engineering: concepts and precepts:* 21–34. Aldershot: Ashgate.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Measuring infrastructure and community recovery rate using Bayesian methods: A case study of power systems resilience

H. Baroud & S. Murlidar
Vanderbilt University, Nashville, USA

ABSTRACT: With the increasing frequency and severity of disasters resulting especially from natural hazards and impacting both infrastructure systems and communities, thus challenging their timely recovery, there is a strong need to prepare for more effective response and recovery. Communities have especially struggled to understand the aspects of recovery patterns for different systems and prepare accordingly. Therefore, it is essential to develop models that are able to measure and estimate the recovery trajectory for a certain community or infrastructure network given system characteristics and event information. The objective of the study is to deploy the Poisson Bayesian kernel model developed and tested in earlier work in risk analysis to measure the recovery rate of a system. In this paper, the model is implemented and tested on a resilience modeling case study of power systems. The model is validated using a comparison to other count data models such as Poisson generalized linear model and the negative binomial generalized linear model.

1 INTRODUCTION

Recent disasters severely impacting both infrastructure systems and communities emphasize the need to prepare for more effective response and recovery. Communities have especially struggled in understanding the aspects of recovery patterns for different systems. Therefore, there is a strong need to develop models that are able to measure and estimate what are the recovery prospects for a certain community or infrastructure network given system characteristics and event information. In addition, the models need to account for uncertainty underlying the information that has been or being gathered before, during, and after the disruption.

Prior work on recovery rate modeling of infrastructure systems focuses on the time to recovery from power outages as a function of event attributes and impact of the disaster (Mackenzie & Barker 2013, Barker & Baroud 2014, Barabadi & Ayele 2018). In this research, the goal is to incorporate the uncertainty in estimating the resilience of systems after disruption. More specifically, the objective of the study is to analyze the recovery rate of a system or a community that has been impacted by a disaster. The response variable considered in this work is the average recovery rate computed based on the impact of the event and the total time to network recovery as well as other variables.

In order to integrate information from experts with data on the disruptive event and recovery process, this work proposes the use of a Poisson Bayesian kernel model which accommodates count data while accounting for prior information and uncertainty in the estimates. The model has been developed and tested using sample data in earlier work (Floyd et al. 2014) and has been applied to a risk analysis case study to predict the frequency of disruptive events in inland waterway (Baroud et al. 2013). However, the method has never been implemented in post-disaster scenarios, more specifically to model recovery rate. In this paper, the model is implemented and tested on a resilience modeling case study of power systems. More specifically, the recovery rate of a community from power outages is represented by a parameter following a Gamma distribution. This prior distribution is updated using historical data of disruptive events as well as a set of attributes that are represented by the kernel function, a measure of similarity between the new data point and the training set. The model performance is evaluated in comparison to other count data models such as the Poisson generalized linear model and the negative binomial generalized linear model.

Section 2 provides background literature on community resilience modeling and count data methods with an outline of the paper's contributions. Section 3 briefly describes the Poisson Bayesian kernel method and provides a structure to the model comparison and performance measures. Section 4 describes the case study with an overview of the data and a summary of the results of the models used in this work. Finally, concluding remarks are provided in section 5.

2 BACKGROUND AND CONTRIBUTION

2.1 *Community resilience modeling*

The ultimate goal of recovery measures after a disaster is to insure the society is able to bounce back from the losses incurred and reach normalcy as fast as possible, in recent studies this has been termed as "community resilience." One common definition for community resilience refers to the ability for a social system to respond and recover from a disaster. While vulnerability was previously used as an indicator, researchers and government policy have realized the advantages of utilizing resilience as an indicator to measure the ability of a community to not only recover during the post-disaster phase, but also advance beyond the pre-disaster state and adapt or transform to improve preparedness to future events. Furthermore, resilient communities are also less vulnerable to hazards than an equivalent less resilient community. Initially, community resilience modeling research focused on qualitative approaches founded in a set of metrics and indicators that describe the resilience of a community (Johansen et al. 2016). The concept of resilience can be useful when quantified and used as a decision-making tool, however, this can be challenging due to the uncertainty in many factors impacting resilience as well as the lack of data in recovery measures. As such, a number of research initiatives have focused on quantifying resilience ranging from stochastic modeling to simulation and data-driven approaches, among others.

Models of community resilience often include a variety of social factors. In one study, community resilience was modeled as categorical variables based on four primary sets of adaptive capacities-Economic Development, Social Capital, Information and Communication, and Community competence (Norris et al. 2008). It is proposed in this work that advancements within each category will aim to create a community that is more resilient to disasters as a whole. More specifically, one example of the hypothesis proposed in Norris et al. (2008) is the ability to measure infrastructure and economic resilience in terms of power restoration time which can therefore be used as a proxy to understand community resilience.

A more robust model for community resilience uses a composite index of social and geographical factors, the Baseline Resilience Indicators for Communities (BRIC) (Cutter et al. 2014). This relative value measure of resilience can point to counties and tracts within a specific geographic location that are particularly vulnerable to disasters and require more attention and more time to fully recover. This measure was found to have significant negative correlation with the previously established Social Vulnerability Index (SoVI).

Analysis has been performed to identify recovery rate specifically following a disaster. However, two relevant primary issues are dealing with missing data as well as homogeneity and heterogeneity across the data set and the fact that some models are so specific that they need to be adapted for different situations. In addition, most studies have aimed to provide restoration curves that give information on the number of customers with service over time. A lack of literature exists to model recovery rate specifically. One study focuses on the need to not only develop recovery rate plots but to be able to select the appropriate models based on the characteristics of a specific data set (Barabadi & Ayele 2018).

2.2 *Methods for modeling count data*

Modeling the recovery rate requires methods that can accommodate count data as the response variables in this case constitutes the number of recovered subjects per unit of time.

Generalized Linear Models (GLM) are widely used within regression models when count data is present. Within this class of models, the Poisson density function is often used with a log-link function, if the variance of the counts is higher than the mean of the counts, it is common to also use a negative binomial GLM. In certain special cases, extensions of these models can accommodate specific situations. For example, zero-truncated models and zero-inflated models can be used when there are excess zero counts (Shankar & Mannering, 1997), and both use an underlying Poisson distribution.

However, both Poisson and negative binomial lack the flexibility to handle data that is, for example, both underdispersed and overdispersed. As such, other models have been developed. One example is the Conway-Maxwell Poisson (COM) distribution GLM (Guikema & Goffelt, 2008). The model functions by having underdispersed data yield a Bernoulli distribution, overdispersed data yield a geometric distribution, and a Poisson distribution when the variance is equal to the mean.

Using a Bayesian framework to account for the uncertainty in the regression parameters, it is possible to improve on their accurate estimation by updating the parameter distributions with new data. Other approaches of analyzing count data using a Bayesian framework are conjugate priors. These methods are quite attractive as they offer the benefit of uncertainty modeling using Bayesian techniques without adding any computational cost. Given a specific prior distribution and a specific likelihood function, the posterior distribution will have the same form as the prior distribution but with updated posterior parameters. There are different forms of conjugate priors, one of which

is the Gamma conjugate prior used to model count data in the model presented in this paper. The method assumes that the rate of occurrence follows a Gamma prior and updates the distribution using information represented by a Poisson likelihood. The Gamma conjugate prior is the foundation of the Poisson Bayesian kernel model used in this paper and will be further discussed in the following section. This method allows the user to model and understand the uncertainty around each variable and estimate them by considering their probability distributions as opposed to point estimate.

2.3 *Contributions*

This paper presents new analysis for data-driven community resilience modeling. A Bayesian approach developed and tested in prior work is implemented and tested in a case study of community recovery from power outages. The work presented here constitutes a first step in advancing data-driven methods for applications in infrastructure and community resilience.

3 METHODOLOGY

3.1 *Poisson Bayesian kernel model*

Poisson Bayesian kernel methods estimate the rate of occurrence of the event rather than estimating a deterministic value for the number of times the event is estimated to occur. A common distribution to model count data within a Bayesian framework is the Gamma-Poisson conjugate prior. The development of the Poisson Bayesian kernel method discussed can be found in Baroud et al. (2013) and Floyd et al. (2014). The approach uses the Gamma conjugate prior as the basis of the model.

It is assumed that the parameter to be estimated is the rate of occurrence, $\lambda > 0$, which follows a Gamma prior distribution with parameters $\alpha > 0$ and $\beta > 0$, as shown in Eq. (1).

$$P(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{(-\beta\lambda)} \tag{1}$$

For the likelihood function, the product of the Poisson density function, shown in Eq. (2), is used, since this is a Gamma-Poisson conjugate prior approach.

$$\begin{aligned} L = \prod_{i=1}^{m} P(y_i) &= \prod_{i=1}^{m} \frac{\left(\lambda_i^{y_i} e^{-\lambda_i}\right)}{y_i!} \\ &= \frac{\lambda_i^{\sum_{i=1}^{m} y_i} e^{-m\lambda_i}}{\prod_{i=1}^{m} y_i!} \end{aligned} \tag{2}$$

Thus, the posterior distribution is the product of Eqs. (1) and (2). Rearranging the product of the likelihood function and the prior distribution function results in a Gamma posterior distribution where $\alpha^* = \sum_{i=1}^{m} x_i + \alpha$ and $\beta^* = m + \beta$.

$$\begin{aligned} &P(\lambda|x) \\ &= \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}\right)\left(\lambda^{\sum_{i=1}^{m} y_i} e^{-m\lambda}\right) \\ &= \frac{\lambda^{\left(\sum_{i=1}^{m} y_i + \alpha - 1\right)} e^{-\lambda(m+\beta)} (n+\beta)^{\sum_{i=1}^{m} y_i + \alpha}}{\Gamma\left(\sum_{i=1}^{m} y_i + \alpha\right)} \\ &= \text{Gamma}(\alpha^*, \beta^*) \end{aligned} \tag{3}$$

This result is the basic Gamma conjugate prior approach used in Bayesian analysis. This approach assumes the notion of exchangeability meaning that for different sets of training and testing data, the resulting posterior parameters will be similar since they are a function of the prior parameter, the size of the dataset, and the summation of all the data points. The characteristics of each outcome are not taken into consideration in this case, but rather the overall property of the dataset (MacKenzie et al., 2014).

The Poisson Bayesian kernel approach extends the notion of the conjugate prior such that the posterior parameters computation not only depends on the prior parameters and the historical data but also on the attributes through the kernel matrix. The parameters for the Bayesian kernel model for counts are expressed in Eqs. (4) and (5). **K** is the $m \times m$ kernel matrix, **Y** is an $m \times 1$ vector containing the output data associated with the m observations of **X**, and **V** is an $m \times 1$ vector containing ones. Each entry in the kernel matrix represents the similarity measure between the attributes of the testing set and the training set, respectively. As such, the new data point is compared with the training set and according to the similarities of the attributes, new values for the parameter of the posterior distribution are computed. Note that in this case, the training and testing sets are assumed to have the same size, m. However, when the model is deployed, the sets can be of different sizes, and in some cases, the testing set could include only one data point such as in a leave-one-out analysis.

$$\alpha^* = \mathbf{KY} + \alpha \tag{4}$$

$$\beta^* = \mathbf{KV} + \beta \tag{5}$$

As with other statistical and mathematical models, there are a few assumptions underlying the deployment of such modeling approach. Even though the form of the prior distribution is

known from the conjugate prior, the model user would still need to identify the values of the prior parameters. While there are formal ways to determine the prior parameters (Kass & Wasserman, 1996), the selection of such parameters might not always be considered (Montesano & Lopes, 2009; Mason & Lopes, 2011). Oftentimes, the priors are either assumed to be known or are assigned such that the prior distribution is non informative. In other cases, these parameters are estimated using data and prior knowledge by matching the sample mean and variance to those of the prior distribution (MacKenzie et al. 2014; Carlin & Louis, 2008). Another assumption to consider is the choice of the kernel function which depends on the application and the model user. This research uses the most popular kernel function, the Radial Basis Function (RBF) in Eq. (6), where $k\left(\mathbf{x}_i,\mathbf{x}_j\right)$ is one entry in the matrix **K** representing the kernel function between the attributes of the i^{th} and j^{th} data points.

$$k\left(\mathbf{x}_i,\mathbf{x}_j\right)=\exp\left(-\frac{\left\|\mathbf{x}_i-\mathbf{x}_j\right\|^2}{2\sigma^2}\right) \tag{6}$$

In addition to being commonly used in kernel methods, RBF has nice properties. The function has only one parameter, σ, to be tuned to an optimal value. This reduces computation efforts significantly in comparison to other kernel functions with two or more parameters requiring a grid search to estimate them. Also, the structure of the function is based on the Euclidean distance, whereby similar data points are closer to each other in the feature space. Finally, the kernel matrix of the RBF has full rank and the entries fall between zero and one resulting in kernel functions of the data points acting as weights in the computation of the posterior parameters (Schölkopf & Smola, 2002). More discussion on the impact of the RBF parameter, σ, on the performance of the model will follow in the case study presented in section 4.

The estimated rate for the new data point follows then a Gamma distribution with parameters α^* and β^*. As a point estimate for this parameter, the expected value of the posterior distribution is considered, shown in Eq. (7) as the ratio of the Gamma distribution parameters α^* and β^*.

$$\hat{\lambda}=\frac{\alpha^*}{\beta^*} \tag{7}$$

Note that a different point estimate for the rate can be used such as the median, the mode, or the variance, depending on the type of problem and the model users.

3.2 *Predictive accuracy measures*

The ultimate objective of developing and identifying predictive models is their application in risk and resilience analysis problems, such as predicting the frequency of disruptions in a particular network system or the recovery rate of infrastructure and communities. While the goodness of fit is important to assess whether the model is capturing the pattern and variability in the data, is it equally important to analyze the prediction power of a statistical model if it is going to be used for forecasting purposes. Prediction accuracy is assessed by the out-of-sample error, which accounts for the discrepancy between the estimated parameter and the actual observation of data points that were not in the set used to train the model. In order to validate the prediction power of the models, several metrics are evaluated to assess the out-of-sample error, and they are summarized in Table 1.

While RMSE and MAE are the most commonly used measurements of error, the normalized RMSE is also considered to account for the variability across different samples of training sets generated by the multi-iteration validation process. NRMSE can either be normalized based on the standard deviation of the observed values, $sd\left(Y_i\right)$, or the range of values in the testing set, $Y_{maximum}-Y_{minimum}$, and both cases are considered in this paper.

3.3 *Comparative analysis*

In order to assess the performance of the models, the predictive accuracy measures are used to evaluate the models. More specifically, Poisson Bayesian kernel model is compared to a Poisson generalized linear model and a negative binomial generalized linear model (Cameron & Trivedi, 1986, 2013).

Table 1. Prediction accuracy metrics formulae.

Prediction accuracy metrics	Formula
Root Mean Square Error (RMSE)	$\frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Y_i-\hat{\lambda}_i\right)^2}$
Normalized Root Mean Square Error (NRMSEM & NRMSED)	$\frac{\frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Y_i-\hat{\lambda}_i\right)^2}}{sd\left(Y_i\right)}$ $\frac{\frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(Y_i-\hat{\lambda}_i\right)^2}}{Y_{maximum}-Y_{minimum}}$
Mean Absolute Error (MAE)	$\frac{1}{n}\sum_{i=1}^{n}\left\|Y_i-\hat{\lambda}_i\right\|$

The Poisson GLM assumes that the rate to be estimated has an exponential relationship with a set of covariates representing coefficients for the different attributes, $\hat{\lambda}_{PGLM} = e^{\beta_i X}$, while the predicted rate for the PBK is equal to the expected value of the posterior probability distribution, $\hat{\lambda}_{PBK} = \frac{\mathbf{KY}+\alpha}{\mathbf{KV}+\beta}$.

4 CASE STUDY

A case study is presented in this paper to demonstrate the use of the Poisson Bayesian kernel model in assessing the resilience of communities. More specifically, the study is focused on major power outage events that happened in the US between 1999 and 2016. The goal is to compare the performance of the model against classical methods and assess its ability to predict, with a high level of accuracy, the recovery rate after these major events.

The ability to accurately measure and predict the recovery rate from power outages allows responders and recovery crews to improve their strategies and resource allocations before, during, and after a disruption.

4.1 *Data*

The data used in the case study is collected from the Energy Information Administration and includes information on the time, date and length of an outage occurred, the magnitude of the power outage (Megawatt Loss & Customers Affected) and the disturbance type (severe weather, equipment failure, among others). The dependent variable to be modeled is recovery rate which is the number of customers affected divided by the duration of outage. To model the rate using a Poisson linear model, an offset of duration was used. Recovery rate is modeled based on 10 regression coefficients that represent information on the cause of the outage, the severity, the location, the duration, and the time of the day and month.

Figure 1 is a scatterplot of all variables in the data set, each square represents a pairwise plot between the corresponding pair of variables on the x-axis and the y-axis, the red line represents a local regression line of the two variables. The numbers shown in the upper side of the scatterplot represent correlations of the pairs of variables which, in this data set, are not significant with the exception of a couple of variables. Examining Figure 1, it is

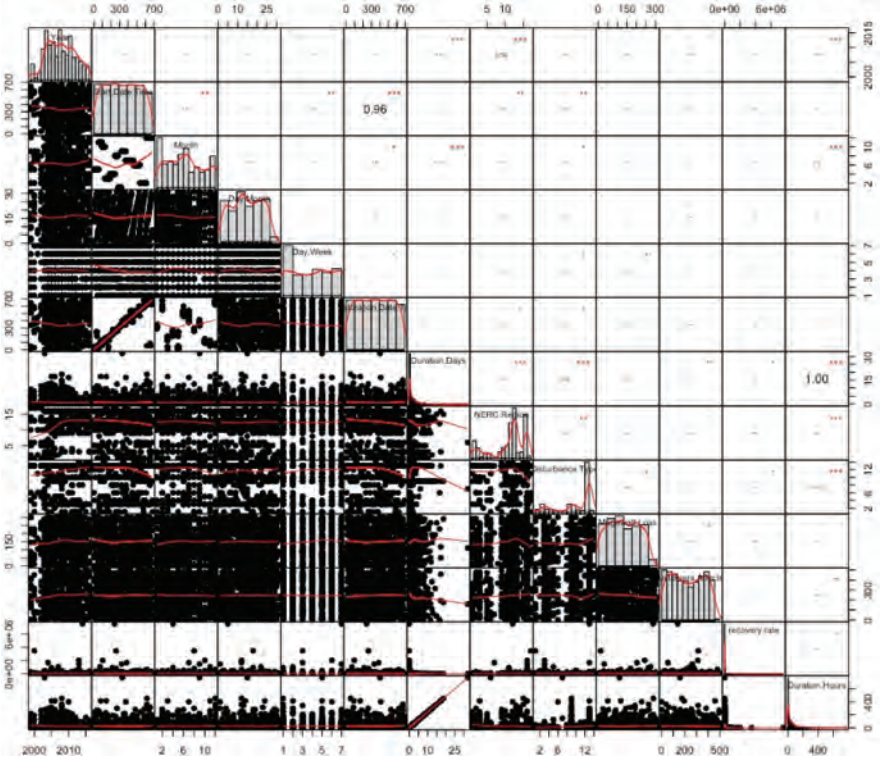


Figure 1. Pairwise scatter plots for all the variables in the data.

difficult to identify visually any particular relationships beyond the expected linear correlations due to multicollinearity such as start date and time with restoration time. The plot provides histograms for the different variables and it can be seen that there is a large variance for many predictors.

Further examination of the patterns in the data focus on the impact of seasonal variations and types of disturbance on the recovery process from power outages. Rates of recovery are generally slower in the winter than in the summer months (Figure 2).

While wide variations are observed in the recovery rate by the type of disturbance, outages due to load shed and fire/extreme heat experience the highest average recovery rate. Disasters such as flooding and hurricane, however, have much slower recovery rates (Figure 3). Also, Severe Weather events result in the largest number of outliers in the data.

4.2 *Results*

The Poisson Bayesian kernel model referred to as PBK, the Poisson GLM referred to as PGLM, and the negative binomial GLM referred to as NBGLM were used to model the data and predict the recovery rate as a function of the predictors related to the time, location, disruption, and other characteristics. The error measures discussed earlier and presented in Table 1 were calculated for each model and summarized in Table 2.

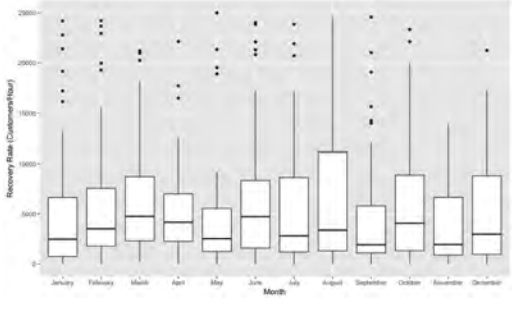


Figure 2. Recovery rate as a function of month.

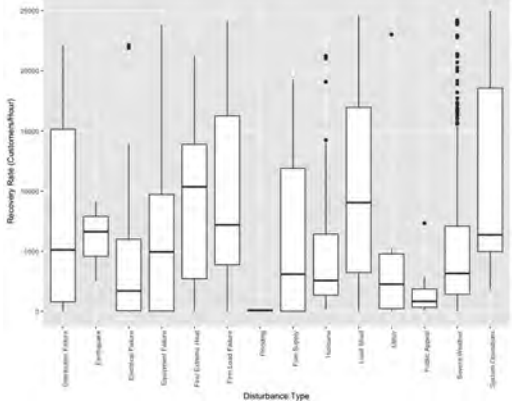


Figure 3. Recovery rate by disturbance type.

Across all predictive accuracy measures, PBK performs yields small errors overall. PGLM results in very large errors that could be driven by the extreme values under Severe Events for instance, whereas PBK is able to control for that and provide more stable estimates. For two of the predictive error measures, NRMSED and MAE, the PBK outperforms the NBGLM.

Overall the performance of PBK and NBGLM is comparable from a predictive accuracy standpoint. However, using PBK would provide an assessment of the uncertainty in the estimates through the prior and posterior distributions of the recovery rate, the outcome is a probability distribution of a comprehensive range of possible values for the recovery rate. As a result, it is possible for a decision maker to identify multiple point estimates based on their risk preference. For example, if the decision maker or infrastructure operator is risk averse, he/she will rely on a more extreme (lower) value than the expected value of the recovery rate posterior distribution since a more conservative mitigation and recovery strategy is preferred. However, if the decision maker is risk taking, the preference would be to save on cost of mitigation and recovery and the upper tail of the distribution will be considered as an optimistic measure of the recovery rate. The choice of the posterior point estimate is not the only way a decision maker is involved in this process. Stakeholders play an important role in identifying multiple initial parameters in the model.

As mentioned earlier, the definition of the prior is an important consideration for any Bayesian approach. In this case, a non-informative prior was assumed. However, another important consideration is the value of the parameter in the kernel function. The results in the table above were obtained based on an arbitrary value of sigma. In order to understand the effect of this parameter on the predictive accuracy, Figure 4 shows the value of the root mean squared error as a function of $1/\sigma$.

There is clearly an optimal value for this parameter valued at approximately 12. It would be ideal if the

Table 2. Prediction error values for all the models.

Model	RMSE	NRMSED	NRMSEM	MAE
PBK	2435	2.03	0.25	1258
PGLM	12961	10.88	1.34	5579
NBGLM	1706	2.06	0.17	2039

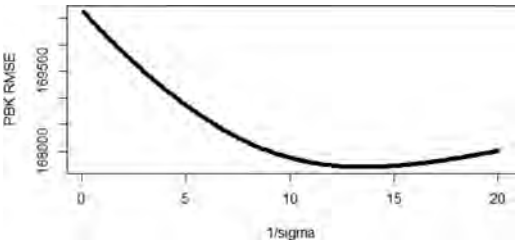


Figure 4. PBK RMSE as a function of different values for the tuning parameter of the kernel function.

parameter is tuned to minimize the error during the training process. The drawback of doing so is the additional computation time for tuning which would exponentially increase as more parameters are considered in other forms of the kernel function.

5 CONCLUSION

The work presented in this paper evaluates the use of Poisson Bayesian kernel models to measure and predict the rate of recovery. The ultimate goal of the research is to be able to quantify community resilience in order to inform resource allocation before, during, and after a disruption. The proposed approach to model the rate of recovery was compared to traditional count data models such as Poisson and negative binomial generalized linear models.

The advantage of using Bayesian techniques is their ability to provide probability distribution of the estimates, accounting for the uncertainty in resilience metrics. Another important benefit is the ability to update predictions as new information on the evolvement of the disaster and the corresponding response of the community becomes available.

An initial comparison to other methods shows that PBK provides a higher accuracy than traditional models with the added benefit of accounting for uncertainty and the decision maker's opinion and prior knowledge.

REFERENCES

Barker, K., & Baroud, H. (2014). Proportional hazards models of infrastructure system recovery. *Reliability Engineering & System Safety*, *124*, 201–206.

Baroud, H., Barker, K., Lurvey, R., and Mackenzie, C. (2013, January). Bayesian kernel model for disruptive event data. *In Proceedings of IIE Annual Conference*. (p. 1777). Institute of Industrial Engineers-Publisher.

Barabadi, A., & Ayele, Y.Z. (2018). Post-disaster infrastructure recovery: Prediction of recovery rate using historical data. *Reliability Engineering & System Safety*, *169*, 209–223. https://doi.org/10.1016/J.RESS.2017.08.018.

Cameron, A.C., & Trivedi, P.K. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, *1*(1), 29–53.

Cameron, A.C., & Trivedi, P.K. (2013). *Regression Analysis of Count Data* (Vol. 53). Cambridge university press.

Carlin, B.P., & Louis, T.A. (2008). *Bayesian Methods for Data Analysis*. CRC Press.

Cutter, S.L., Ash, K.D., & Emrich, C.T. (2014). The geographies of community disaster resilience. *Global Environmental Change*, *29*, 65–77. https://doi.org/10.1016/J.GLOENVCHA.2014.08.005.

Floyd, M.S., Baroud, H., & Barker, K. (2014). Empirical analysis of Bayesian kernel methods for modeling count data. In *Systems and Information Engineering Design Symposium (SIEDS), 2014* (pp. 328–333). IEEE.

Guikema, S.D., & Goffelt, J.P. (2008). A Flexible Count Data Regression Model for Risk Analysis. *Risk Analysis*, *28*(1), 213–223. https://doi.org/10.1111/j.1539-6924.2008.01014.x.

Johansen, C., Horney, J., & Tien, I. (2016). Metrics for Evaluating and Improving Community Resilience. *Journal of Infrastructure Systems*, *23*(2), 04016032.

Kass, R.E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association, 91*(435), 1343–1370.

MacKenzie, C. & Barker, K. (2013). Empirical Data and Regression Analysis for Estimation of Infrastructure Resilience with Application to Electric Power Outages. *Journal of Infrastructure Systems*, *19*(1), 25–35. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000103.

MacKenzie, C.A., Trafalis, T.B., & Barker, K. (2014b). A Bayesian Beta kernel model for binary classification and online learning problems. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *7*(6), 434–449.

Mason, M., & Lopes, M. (2011, March). Robot self-initiative and personalization by learning through repeated interactions. *In Conference on Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International* (pp. 433–440). IEEE.

Montesano, L., & Lopes, M. (2009, June). Learning grasping affordances from local visual descriptors. In *Development and Learning, 2009. ICDL 2009. IEEE 8th International Conference on* (pp. 1–6). IEEE.

Norris, F.H., Stevens, S.P., Pfefferbaum, B., Wyche, K.F., & Pfefferbaum, R.L. (2008). Community Resilience as a Metaphor, Theory, Set of Capacities, and Strategy for Disaster Readiness. *American Journal of Community Psychology*, *41*(1–2), 127–150. https://doi.org/10.1007/s10464-007-9156-6.

Schölkopf, B. & Smola, A.J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Contrasting critical infrastructure resilience from Swedish infrastructure failure data

J. Johansson
Division of Risk Management and Societal Safety, Lund University, Sweden
Centre for Critical Infrastructure Protection Research (CenCIP), Lund University, Sweden

R. Jonason Bjärenstam & E. Axelsdóttir
Division of Risk Management and Societal Safety, Lund University, Sweden

ABSTRACT: Critical technical infrastructures constitute the backbone of society by providing essential services to vital societal functions and the community at large, hence it is of essence that these are resilient. Critical infrastructures, e.g. power systems, telecommunication systems and railway systems, are designed and operated based on different philosophies of where to put the resilience emphasis, robustness, rapidity of recovery or a combination of the two. Here empirical failure data, such as duration and consequence of disruptions, from several critical infrastructures in Sweden are explored and analysed. To facilitate comparisons, a generic resilience assessment approach is also presented and applied. The results give insight to the resilience level of different infrastructures in Sweden and a basis for an exploration of its reasons, e.g. due to difference in regulatory schemes, design or risk cultures. It is concluded that there exist significant differences of infrastructures resilience levels and the factors shaping the resilience.

1 INTRODUCTION

The large scale societal consequences arising in past events involving critical infrastructures, such as the European blackout in 2006, the Eyjafjallajökull eruption in 2010 and Hurricane Sandy in 2012 (Johannsson et al., 2015), clearly indicates the need for approaches and measures aiming at increasing infrastructure robustness and rapidity of recovery, in essence addressing the resilience of critical infrastructures. These events have also revealed the complexities and uncertainties involved in assessing resilience of critical infrastructures.

During the last years there have been ample contributions of different, both academic and policy oriented, approaches for assessing critical infrastructure resilience. These include expert based methods, e.g. different types of index methods, modelling and simulation based methods and empirical data based methods. Limited attention has been drawn towards contrasting and comparing the resilience of different types of critical infrastructures through the use of empirical failure and interruption data that are normally gathered by regulatory authorities or internally by infrastructure owners.

Here we are presenting and applying a generic resilience assessment approach to this type of data, aiming at assessing and contrasting resilience levels as input to exploration of possible causes, e.g. underlying influential factors such as regulatory schemes, differing design philosophies or risk cultures.

In Chapter 2 a background is given for the generic resilience assessment approach as presented in Chapter 3 together with the data collection method. In Chapter 4 the data is presented for five critical infrastructures in Sweden, electric transmission, electric distribution, transport railway, transport road and water supply. In Chapter 5 the resilience levels of the infrastructures are contrasted and compared. The findings are then discussed in Chapter 6 and conclusions from the study are drawn in Chapter 7.

2 BACKGROUND

The most common approach in the scientific literature, related to critical infrastructures, is to assess resilience in terms of the system's functionality over a time period. Emphasis is normally on measuring a system's resilience towards a single specific event or scenario (e.g. Pursiainen et al., 2016, Hosseini et al., 2016, Panteli et al., 2017, Nan & Sansavini, 2017). The approaches also tend to follow the fundamental resilience conceptualisation proposed by Bruneau et al. (2003), sometimes

referred to as the "resilience triangle" and where the area of the triangle is defined as resilience loss. Several more detailed conceptualisations of this engineering oriented concept of resilience have lately emerged in the literature. For example, accounting for that the recovery behaviour is normally not linear but can e.g. be exponential (Cimellaro et al., 2010). There is also examples of where the resilience curve is divided into several different phases, e.g. original steady state, disruptive phase, system recovery state, and finally a stable end state (c.f. Henry & Ramirez-Marquez, 2012, Nan, C. & Sansavini, G., 2017). For these conceptualisations, the events in consideration seems to be for single large impact events, resulting in great stress on the measured system's functionality and high societal consequences. Further, the literature also presents a great variety of ways for defining or measuring the functionality of different infrastructures.

Methods for assessing the resilience of critical infrastructures can broadly be divided into three categories: expert based, modelling and simulation based, and empirically based methods. Expert based methods are generally an assessment based on system characteristics where the opinion from experts have been aggregated to produce some sort of index of infrastructure resilience (Hosseini et al., 2016, Hassel & Johansson, 2016). One example is Chang et al. (2014) that presents a resilience elicitation approach, where they estimate resilience levels based on expert opinions for a specific scenario in an interdependent critical infrastructure setting. Resulting in estimated service disruption levels, rated from no loss to severe disruption for different time lengths of disruption. Modelling and simulation based methods, in this context, are quantitative approaches aiming at observing modelled system behaviour during disruptive events (e.g. Hosseini et al., 2016, Ouyang, 2014). Some authors also include more general resilience metric conceptualisations, that could be applicable also for empirical studies, and utilize them in a modelling and simulation context (e.g. Nan & Sansavini, 2017, Panteli et al., 2017). Modelling and simulations efforts span e.g. from network theoretical, Monte Carlo, optimization to fuzzy logic approaches (e.g. Hosseini et al., 2016). Empirically based methods aim at deriving resilience levels of critical infrastructures during past events (see e.g. Johansson et al., 2015, Zorn & Shamseldin, 2015a). Normally it involves statistical analyses of the rapidity and robustness of one or several infrastructures during a specific disruptive event, sometimes also the quantification of interdependencies are addressed (Zorn & Shamseldin, 2015b). The empirical methods normally aim at measuring resilience during different phases of the disruptive event. These approaches are less applicable when addressing the overall resilience of an infrastructure towards the range from low impact to high impact events and specifically for assessing the resilience over time.

Here the aim is to provide such an approach that is based on assessing the resilience level of different types of critical infrastructures based on interruption data, which has not been covered in the scientific literature to our knowledge. It should be noted however that reliability oriented approaches for critical infrastructures (for power systems see e.g. Billinton & Allan, 1996) to a large extent deals with similar issues of assessing system functionality during disruptive events based on either modelling and simulation approaches or through empirical failure data, then however normally focusing on infrastructure specific indices and not towards a unified approach for the comparison of different types of infrastructures.

3 METHODS

3.1 *Data collection*

The data was collected through a three phase process: I) identifying technical critical infrastructures (CIs) within Sweden, II) identifying actors within these infrastructures, preferably at national level, that potentially collects interruption data for their respective CI, and III) contacting the actors and retrieving the data. This data then underwent an examination to gain further understanding of the quality and quantity of the data, e.g.: how the data is collected? (e.g. in the form of databases or in the form of incident reports?), what are the parameters given in the data?, and what is the delimitation and limitation of the data? In some instances, further contacts where made with the actors for clarifications or to gain additional data, resulting in an iterative data collection process. For each data set it was finally determined if it contained the desired parameters for the resilience assessment approach.

In the end data from the following five infrastructures and actors are included (out of 17 initially addressed infrastructure actors): a) Electricity transmission, Svenska Kraftnät (SVK), b) Electricity distribution, Energiföretagen, c) Transport Road, Trafikverket, d) Transport Railway, Trafikverket, and e) Water supply, Stockholm vatten och avfall (SVOA).

3.2 *Resilience assessment approach*

The aim of the Resilience Assessment Approach (RAA) is to present a generic method that is applicable for the resilience analysis of several different types of critical infrastructures, at a variety of hierarchical levels, based on empirical interruption data.

The approach is hence naturally bounded by the interruption parameters that are typically collected by infrastructure owners or regulatory authorities. The interruptions are typically reported as independent incidents, although there might be underlying correlations between the interruptions (e.g. during a storm). Furthermore, the collected interruption data of a single event generally lacks detailed information of the actual functionality during the recovery process for the recorded interruptions. Normally only start time, end time and maximum functionality loss is given. Hence the data does not reflect the widely spread "resilience curve" as normally depicted in the academic literature (e.g. Cimellaro et al., 2006, Hosseini et al., 2016, Panteli et al., 2017). If the interruption data would have had higher granularity, the approach presented here would however still be applicable.

The resilience assessment approach aims to measure the resilience of a critical infrastructure at a "national system level" to guide policy making and regulatory considerations at this level. As typically several sub-infrastructures together build up an overall critical infrastructure (e.g. in Sweden there are about 160 distribution system owners, 4 regional system owners and 1 national system owner of the electricity infrastructure) it is necessary to aggregate the data from several sub-infrastructures that covers different geographical areas to a national level. Each individual failure is considered to place a strain on the overall system with a negative effect on the functionality during a specific time period. In cases where two or more failures occur simultaneously or overlapping at any time period, these are summed up (stacked), resulting in a fictive national functionality loss of the system as a whole (F_L), see Figure 1. As the true behaviour of the system functionality during a single interruption is unknown, each individual reported interruption has been depicted with a box-shape profile.

The fundamental criteria for the approach is that each interruption is recorded with a start time, end time and one or several consequence parameters which can be transformed into a single functionality loss, F_L, by normalising with an appropriately chosen time-independent baseline.

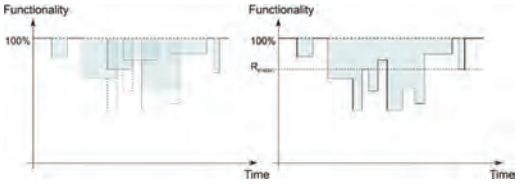


Figure 1. Illustration of how interruption data from several sub-infrastructures (left) is aggregated to a fictive "national system level" (right).

The total resilience, R_{Tot}, of a critical infrastructure over a specified time period is then derived as:

$$\mathbf{R_{Tot}} = \sum_{T_0}^{T_{end}} 1 - \mathbf{F_L(t)} \quad (1)$$

where:

F_L = Functionality loss during time period t,
T_0 = Start time of time period
T_{end} = End time of time period

In order to facilitate a comparison of different infrastructures and for varying time periods the mean resilience loss, R_{Mean} for a given time period is calculated as:

$$\mathbf{R_{Mean}} = \frac{\boldsymbol{R}_{Loss,Tot}}{N} \quad (2)$$

N = Total number of samples in time period

4 INFRASTRUCTURE DATA

Here the background and boundaries of the data for the infrastructures are presented. Many factors sets boundaries of the data, such as if its gathered due to regulatory obligations or for reasons of internal audits and improvements. If an interruption were lacking necessary information for the resilience calculations, these are labelled as "missing data".

4.1 *Electricity transmission*

Svenska Kraftnät (SvK) is the national operator and the Swedish authority responsible for the electric power transmission system. The transmission system transports high voltage power from generating sources, such as power plants and wind parks, to regional networks. The transmission system is also synchronously interconnected to other countries (Norway, Finland and Denmark), but the data is delimited only to the Swedish system. According to Swedish and EU regulations, all companies providing power supplies are bound by law to deliver reports of outages and disturbances since the year 2009 (ENTSOE, 2017, Regeringskansliet, 2009, SVK, 2017).

The data from SvK contained operational outages from the years 2006 to 2016. Each interruption has in total 14 different parameters. Of specific interest here is the start time, end time and the amount of Energy Not Supplied (ENS) due to the interruption. It should be noted that most interruptions in the transmission system leads to no loss of power supply (95%), due to high levels of redundancies and spare capacities.

The data set received from SvK is judged to be of very high quality and comes with only a few unusable data points each year, see Figure 2. Regarding data required for the resilience calculation, the data set is almost complete. In total 2468 interruptions are contained in the data, where only 113 (4.6%) were missing the relevant information.

The data used for the resilience calculations are the given start and end times, and the consequence is calculated by transforming the reported ENS to Power Not Supplied (PNS), averaged for the duration of the outage. The functionality loss is derived by dividing the PNS with the baseline of total Swedish power consumption for a specific year.

4.2 *Electricity distribution*

Energiföretagen is a relatively new Swedish trade association, formed in 2016 through the merging of existing trade organisations. Their main task is to ensure and maintain its members' commercial interests, but they also support the Swedish distribution actors to keep up to date with changing demands and environmental needs. Energiföretagen continuously gather and analyse failure data from their members, annually summarised and presented in the "DARWin-reports" (Energiföretagen, 2017). With almost 100 member companies, the data represents the distribution of electricity to nearly 90% of the end users in Sweden.

The data from Energiföretagen were anonymized (no reference to company) and contained outage data from the year 2005 to 2015, see Figure 3. Each interruption has in total 13 different parameters. As oppose to the data for the transmission system, disruptions at the distribution level to a much higher degree leads to consequences (only 0.75% of the interruptions had zero affected customers), due to the infrastructures limited redundancies and spare capacities.

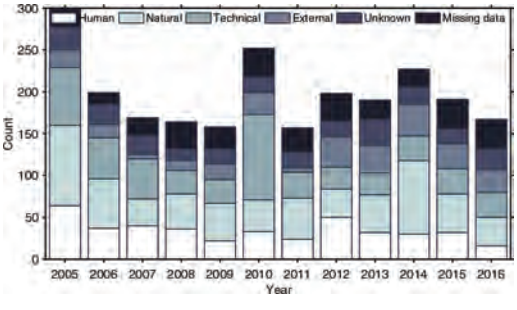


Figure 2. Interruption data for the electricity transmission infrastructure. Categorized into failure cause and missing data.

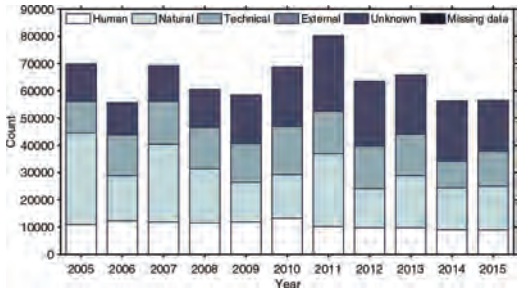


Figure 3. Interruption data for the electricity distribution infrastructure. Categorized into failure cause and missing data.

The data set contains about three hundred times more interruptions than the SvK data. The data is judged to be of high quality, where there is only one bad value, after correcting some obvious errors, of just over 700'000 interruptions in total.

The parameters used for the resilience calculations are the start and end times of the interruptions, and the consequence is calculated as the sum of affected low and medium voltage customers. The baseline for functionality is the total number of customers during a specific year for the reporting companies.

4.3 *Transport road*

Trafikverket is a public authority responsible for the long-term infrastructure planning of road-, railway-, shipping- and aviation-operations, and construction and maintenance of the road and railway in Sweden.

The data from Trafikverket regarding the road infrastructure comes from a project called "Total Traffic Stops" (TTS). This data started to be collected in 2016 and we have data until the first half of 2017. In TTS, unplanned and total stop of traffic for a given road section is reported. Hence e.g. planned interruptions or maintenance actions that give rise to considerable delays in the traffic is not included. On larger roads, with several lanes and the two directions completely separated, it is classified as a total stop only if all the lanes in one direction are closed. For roads where the two directions are not separated, all lanes in both directions must be in full stop to be classified and registered in the TTS.

In the data quite many parameters, 30 in total, are described for each interruption. Most of them for classification in accordance to TTS and considered less relevant here, e.g. coordinates of the accident and other location specific descriptions. The quality of the data is good. Of a total of 8295 recorded interruptions, no one where missing the

relevant data. 2730 interruptions had either zero duration or zero consequences.

As resilience metrics the given start and end times of the interruptions, and the consequence of each failure is given by the reported "Average daily traffic" in terms of vehicles affected. The baseline for calculating functionality is the average number of vehicles that is in the transport system on any given moment (about 600'000), derived from additional data provided by Trafikverket.

4.4 *Transport railway*

The data for the railway infrastructure is also from Trafikverket and covers the years 2012 to 2016. Failure data on railway system outages was scarce and according to Trafikverket there is no general data that sums up or connects railway outages or failures with a root cause. However, they do log train delays which can be used as proxy. If a train is more than 3 minutes delayed when passing a designated "measurement point" (usually a train station), it is registered. This is logged manually and if correctly logged, all train delays associated with the same root cause are given the same "delay ID". It is however quite uncertain how accurate and stringent these logs connect cause and effect. In total the data contains three million delays and about one million delay IDs. Less than 0.1% of the interruptions miss the relevant parameters. The quality of the data is judged to be fair.

Start and end time gives the duration, and the number of registered delays are averaged over the interruption duration with the unit of delays/hour. The baseline is derived from the total number of train departures in the system during the year 2012, in total 38 million. Hence, on average about 4 300 trains/hour are passing or departing from a measurement point on average. Due to lack of yearly data, this number is normalised and scaled with the total transported km in the system per year.

4.5 *Water supply*

For the water supply infrastructure, attempts to gather both national data and data from several different distribution system operators were made. However, only Stockholm Vatten och Avfall (SVOA), which is the main distributor of drinking water and operates the wastewater treatment for Stockholm city and the municipality of Huddinge, gathered interruption data in a format suitable for our purpose. In total, SVOA serves about 10% of the Swedish population and the data is used for internal audit, as no regulatory demands exist for the water sector with respect to disruption data. The data covers water outages caused by service and repairs of the infrastructure. The data set stretches from July 2009 to May 2017.

In total, each interruption has 9 different parameters, where those of specific interest here is start and end time, duration (given in "Hours without water") and consequence (given as "Number of users"). In total 2500 interruptions are recorded in the data, where quite many are missing the relevant information. About 1700 interruptions were in the end used. The baseline for calculating functionality

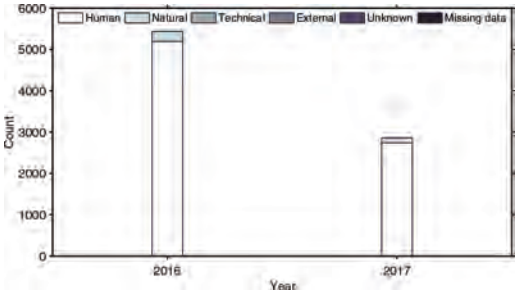


Figure 4. Interruption data for the transport road infrastructure. Categorized into failure cause and missing data.

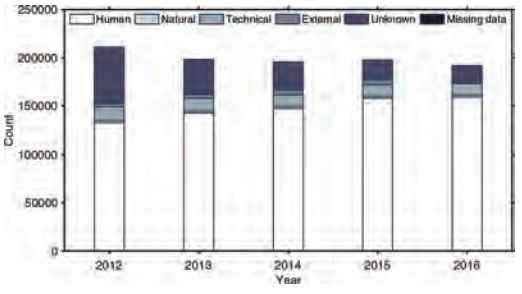


Figure 5. Interruption data for the transport railway infrastructure. Categorized into failure cause and missing data.

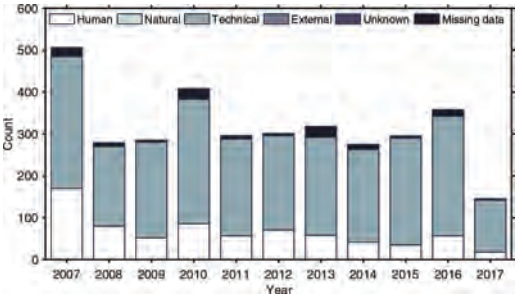


Figure 6. Interruption data for the water supply infrastructure. Categorized into failure cause and missing data.

is the approximate total number of customers that are served by SVOA (circa one million).

4.6 *Other infrastructures*

During the process of gathering interruption data attempts were also made to get data for telecommunication, electronic communication, maritime and aviation infrastructures. However, either due to reasons of sensitivity of data or incomplete/lacking data collection processes no usable data was possible to retrieve for these infrastructures.

5 RESULTS

5.1 *Comparison of resilience levels*

In Figure 7 the resilience level for the different infrastructures are presented. For each of the infrastructures, depending on the availability of data, the yearly mean resilience level for 2005–2017 is presented.

It is clear that there are rather large differences of the resilience levels when comparing the infrastructures. The worst performing infrastructure is the transport railway and the best is the electric transmission system. The resilience level for a given infrastructure also varies over the studied period. The electricity distribution system demonstrates quite fluctuation resilience levels over the years. The resilience levels of the transport railway system are declining over the years, with more or less similar resilience level during the last two years, 2015 and 2016.

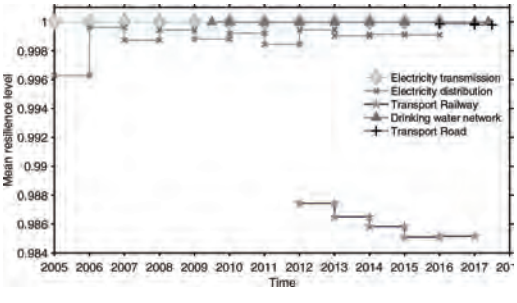


Figure 7. Annual resilience levels of the studied infrastructures.

Table 1. Resilience values for: ETr = Electricity Transmission, EDi = Electricity Distribution, TRo = Transport Road, TRa = Transport Railway, Wat = Water supply.

	Duration (h)			F_L			Resilience
Inf.	Mean	Var	Max	Mean (10^{-3})	Var (10^{-6})	Max (10^{-3})	Mean
ETr	2.23	76.5	67.3	0.51	8.00	44.5	1.0000
EDi	8.16	3420	8490	3.95	160	159	0.9989
TRo	5.87	3200	10500	0.24	0.229	7.48	0.9999
TRa	19.3	8170	42700	16.6	55.1	421	0.9899
Wat	4.76	130	350	0.17	0.155	7.55	1.0000

In Table 1 the overall mean resilience level are given, together with mean and variance of duration and functionality loss (respectively). The Electricity transmission scores the best result, closely followed by Transport road and Water supply. Then comes Electricity distribution followed by Transport railway that is by far the least resilient infrastructure.

5.2 *Comparison of duration and consequence*

In Figure 8 histograms of duration and functionality loss for each of the infrastructures are presented. All interruption data that meets the requirements for calculating the resilience metrics, in accordance with the data section, are included. Interruptions that either exceeds the duration threshold of 12 hours or the functionality loss threshold of 0.001 are binned. These binned data (i.e. the tails of the distributions) are also presented separately in insert figures.

Comparing the length of disruptions among the different infrastructures give that the water supply infrastructure seem to have slightly different shape with respect to the length of disruptions, where a typical interruption last for about 4–5 hours compared to 0–1 hours for the other infrastructures. Comparing functionality loss, the values are generally very small (maximum from 0.5% to 3%), except for the transport railway infrastructure that have experienced larger scale disruptions (up to 39%).

6 DISCUSSION

One of the complications of comparing critical infrastructure resilience levels based on empirical failure data is that the processes for gathering the data and the format of the data varies significantly between the studied infrastructures. To facilitate comparisons, the parameters that are of essence for assessing resilience levels of critical infrastructures at a national level needs to be defined, communicated and maybe even regulated. Some of the infrastructures are either legally obliged, or encouraged by official instances, to systematically gather interruption data, while for other infrastructures such incentives are lacking. If the latter, and if interruption data is gathered, these are generally more geared towards being suitable for specific internal usages, rather than for comparisons across different infrastructures.

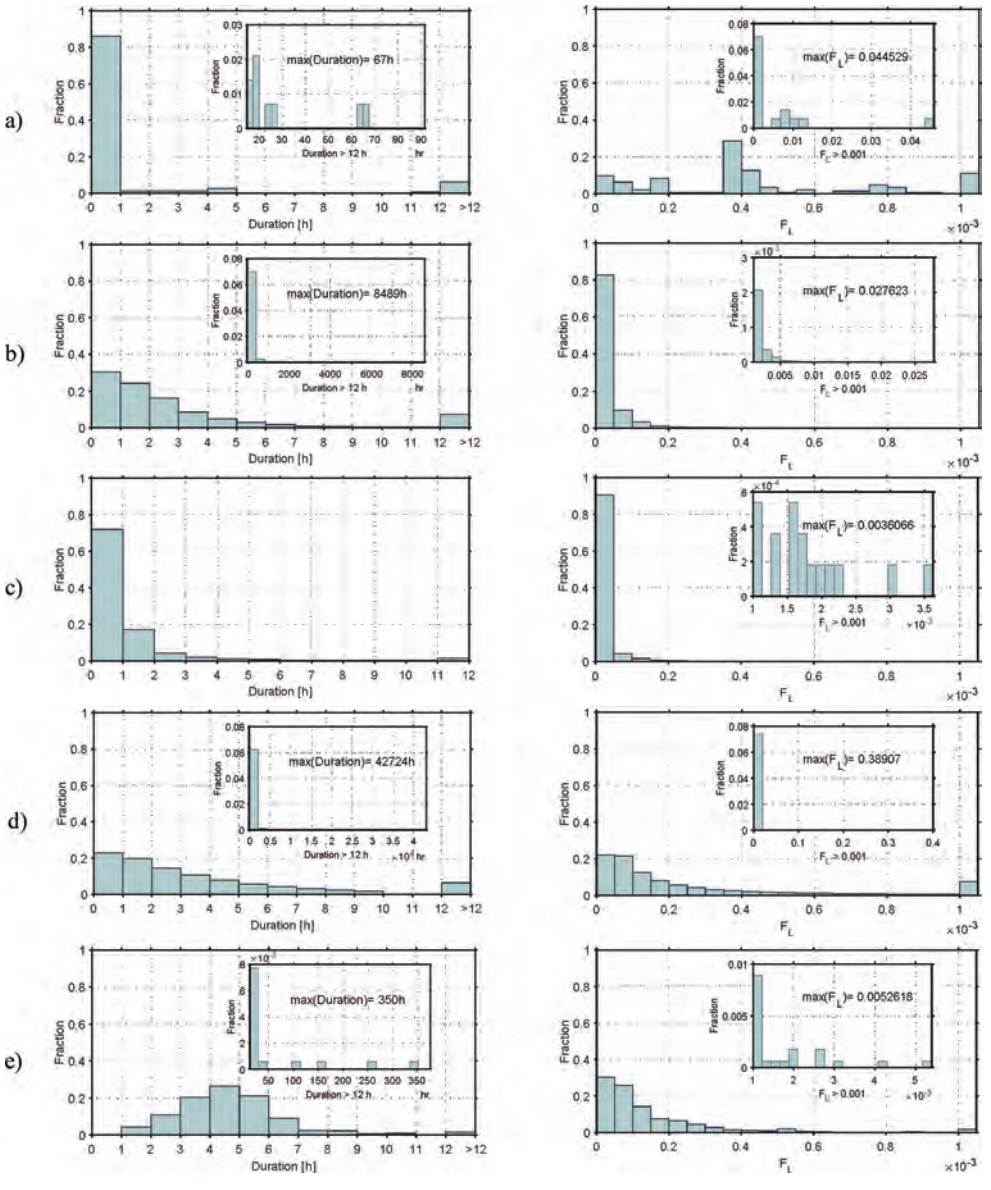


Figure 8. Histograms of duration (left) and functionality loss (right) for: a) Electricity transmission, b) Electricity distribution, c) Transport Road, d) Transport Railway, e) Water supply. In the plots the interruptions exceeding 12-hour duration or exceeding 0.001 functionality loss have been separately binned for reasons of clarity. In the inset figures these binned exceedance values are shown.

Striking a balance between these goals, national comparisons and internal usage, is hence of essence. For some infrastructures there seems to overall be a lack of structured data collection processes in place, e.g. water supply infrastructure and telecommunication/electronic communication infrastructures. For example, for the water infrastructure we found a document from the trade organisation Svenskt Vatten that is not encouraging collection of the kind of data that is necessary for empirical resilience estimations as carried out here. This document gives quite detailed suggestions of how to collect disruption data for the water distribution sector and what parameters to include, whereas only start time is proposed as being of importance and excluding end time. This is one of the causes why much of the data

we could receive from the water sector was not usable as it did not come with accounts of ending times for the interruptions.

There are many aspects that can potentially be explanatory of the found differences between the resilience levels of the infrastructures. For example, the infrastructure with the lowest resilience level, transport railway, is known to be a highly a congested system with few redundancies and a system that often operates close to the limits of maximum capacity, leaving no margins to absorb coincidental variations and stresses. In comparison, the transport road infrastructure has greater redundancies, where there are alternative routes available if a certain road is interrupted, lessening the overall consequences. Similarly, the infrastructure with the highest resilience level, the electricity transmission system, has a high degree of redundancy and more seldom are operating close to the limits of maximum capacity. The electricity transmission data was the only one with systematically collected zero consequence interruptions. These zero consequence interruptions are of high interest when trying to give an account of the level of flexibility and adaptability in the system for tolerating stresses, but unfortunately these type of interruptions are not systematically logged for the other infrastructures.

Since electricity transmission is one of the most critical infrastructures, for which many other infrastructures are dependent upon, (e.g. Johansson et al., 2015) there might be more incentives in place for this infrastructure that has led to the high level of resilience that we see in our data, which would be interesting to explore further.

There are also other structural properties and contextual factors that influence the resilience level as assessed from empirical failure data. For example, the electricity transmission system is a quite protected system, e.g. highly tree-secured and generally harder to access, compared to the electricity distribution system, where interruption due to falling trees and excavations are quite common and are, in general, more exposed systems. The water supply infrastructure was shown to be the second most resilient infrastructure, this is likely because even when failures occur, many times some degree of functionality can still be obtained (although loosing pressure in the system), where customers in many cases will not experience a complete loss of service.

The resilience assessment approach presented here is designed to utilise empirical failure data. We discovered that there are inherent limitations of the data in terms of resolution when it comes to assessing infrastructure resilience, such as that the behaviour during the disruptive- and recovery phase is unknown. As such the results are slightly pessimistic, as normally customers get incrementally reconnected during an interruption. Further, as with all empirical approaches it is a retrospective exercise, and hence have limited predictability of the resilience towards unknown or not yet experienced stresses. The approach however does allow for a robust resilience comparison of different critical infrastructures and how the resilience level changes over time. As such it provides relevant input for further explorations of causes and incentives for achieving resilience across infrastructures.

With further extensions of the presented work, it could also be possible to analyse the influence of infrastructure interdependencies on the resilience levels (c.f. Dueñas-Osorio & Kwasinski, 2012, Zorn & Shamseldin, 2015b). Given for example a power outage in the electricity transmission system it could be possible to quantify how that impacts other infrastructures, e.g. the electricity distribution system and the transport railway system. Further, if the interruption data is also combined with other data such as wind speed, rainfall and temperature, or average income or population densities given the outage area, further explorations into the causes and effects of infrastructure failures is possible.

The results presented here is valuable input for further studies of the underlying factors that are shaping and influencing the resilience levels of different infrastructures. In the end, giving guidance towards how underlying influential factors can shape the resilience of critical infrastructure, factors such as regulatory schemes, differing design philosophies or risk cultures. The approach can also be used for bench-mark type studies, e.g. how resilience strategies for one type of infrastructure might be possible, or impossible, to implement for another. It is also possible to explore different design philosophies in more depth and their impact on the resilience of critical infrastructures.

7 CONCLUSIONS

In the paper, a generic approach for the assessment of resilience levels of different types of critical infrastructures based on empirical interruption data is presented and applied to several Swedish critical infrastructures. It is concluded that the approach is applicable for a unified comparison of the resilience levels of different types of technical infrastructures at a national level. The most resilient infrastructures are Electricity transmission and Water supply, and the less resilient infrastructures are Transport railway and Electricity distribution. The results also reveal difference in how resilience seems to be achieved, where some infrastructure seems to focus on limiting the interruption times (such as Electricity transmission) and some focus on limiting the consequences that arise (Water supply). This type of information is valuable for

understanding the level of resilience and what is influencing and shaping the resilience of our critical infrastructures.

ACKNOWLEDGEMENT

This research has been financed by the Swedish Civil Contingencies Agency, by funding the establishment of the Centre for Critical Infrastructure Protection Research (CenCIP) at Lund University, and the EU Internal Security Fund which is gratefully acknowledged.

REFERENCES

Billinton, R., Allan, R.N., (1996). *Reliability Evaluation of Power Systems—Second Edition*. ISBN 978-0-306-45259-8, New York, Springer.

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., von Winterfeldt, D. (2003). A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities. *Earthquake Spectra*, 19(4), 733–752.

Cimellaro, G. P. (2006). *Quantification of seismic resilience of health care facilities*. Technical Report MCEER-09-0009, University at Buffalo, State University of New York.

Cimellaro, G. P., Reinhorn, A. M., & Bruneau, M. (2010). Framework for analytical quantification of disaster resilience. *Engineering Structures*, 32, 3639–3649.

Chang, S. E., McDaniels, T., Fox, J., Dhariwal, R., & Longstaff, H. (2014). Toward Disaster-Resilient Cities: Characterizing Resilience of Infrastructure Systems with Expert Judgments. *Risk Analysis*, 34(3), 416–434.

Dueñas-Osorio, L., & Kwasinski, A. (2012). Quantification of lifeline system interdependencies after the 27 February 2010 Mw 8.8 Offshore Maule, Chile, earthquake. Earthquake Spectra, 28(1), 581–603.

Energiföretagen. (2017). Leveranssäkerhet/DARWin. Retrieved from https://www.energiforetagen.se/statistik/elstatistik/leveranssakerhetdarwin/, 2017-10-11.

ENTSOE (2017). Guidelines for the classification of grid disturbances above 100 kV. Brussels: ENTSO-E Aisbl.

Henry, D., & Ramirez-Marquez, J. (2012). Generic metrics and quantitative approaches for system resilience as a function of time. *Reliability Engineering & System Safety*, 99, 114–122.

Hosseini, S., Barker, K., & Ramirez-Marquez, J. E. (2016). A review of definitions and measures of system resilience. *Reliability Engineering & System Safety*, 145, 47–61.

Johansson, J., Hassel, H., Cedergren, A., Svegrup, L., Arvidsson, B., (2015). Method for describing and analysing cascading effects in past events: Initial conclusions and findings, *ESREL 2015*, Zürich, Switzerland, September 7–10.

Hassel, H., & Johansson, J. (2016). Review of methods for measuring societal resilience and how they address critical infrastructures. *ESREL 2016*, 25–29 September 2016, Glasgow, Scotland, UK.

Nan, C., & Sansavini, G. (2017). A quantitative method for assessing resilience of interdependent infrastructures. *Reliability Engineering & System Safety*, 157, 35–53.

Ouyang, M. (2014). Review on modeling and simulation of interdependent critical infrastructure systems. Reliability engineering & System safety, 121, 43–60.

Panteli, M., Mancarella, P., Trakas, D., Kyriakides, E., & Hatziargyriou, N. (2017). Metrics and Quantification of Operational and Infrastructure Resilience in Power Systems. *IEEE Transactions on Power Systems*, 32(6), 4732–4742.

Pursiainen, C., Rød, B., Baker, G., Honfi, D., & Lange, D. (2016). Critical infrastructure resilience index, *ESREL 2016*, 25–29 September 2016, Glasgow, Scotland, UK.

Regeringskansliet. (2009). Lag om ändring i ellagen (1997:857). Regeringskansliet Retrieved from: http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/ellag-1997857_sfs-1997-857, 2017-11-20.

SVK. (2017). Svenska Kraftnät. Our activities. Retrieved from http://www.svk.se/en/about-us/our-activities/, 2017-08-24.

Zorn, C. R., & Shamseldin, A. Y. (2015a). Post-disaster infrastructure restoration: A comparison of events for future planning. *International Journal of Disaster Risk Reduction*, 13, 158–166.

Zorn, C. & Shamseldin, A.Y. (2015b). Quantifying Directional Dependencies from Infrastructure Restoration Data. Earthquake Spectra, 32(3), 1363–1381.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Working together towards Critical Infrastructure (CI) resilience

C. Lomba-Fernández, J.M. Sarriegi, P. Marana & L. Labaka
Tecnun, University of Navarra, Donostia-San Sebastián, Spain

ABSTRACT: Improving the resilience level of a critical infrastructure is vital to face crises and ensure its proper functioning. This paper describes the methodology followed to gather information for developing a methodology to assess the resilience level of Red Eléctrica de España (REE), which is the company responsible for the transmission and operation of the electricity system in Spain. The process was based on the three units of analysis of a crisis: peak of the crisis, lifecycle of the crisis, and learning process among crisis. In the peak of the crisis we identified the impact variables that allow to characterize a crisis, and the main key stakeholders. In the lifecycle of crisis, we analyzed how the preparation and prevention activities affect the response and recovery. Furthermore, a CIs' interdependencies analysis was carried out. Finally, the last step focused on understanding how the CI learns between one crisis and the next one.

1 INTRODUCTION

The welfare of society is highly dependent on the efficient performance of Critical Infrastructures (CIs). In general, CIs are those systems and companies that provide essential services that underpin, maintain and sustain vital societal functions in which relies societies' wellbeing (DHS 2017; EU 2017).

CIs are very complex systems with high degree of interconnections among them (Rinaldi 2001). These systems are designed to be reliable and robust but, at the same time, the inherent complexity increases their vulnerability since they become completely dependent of other CIs proper functioning. This vulnerability could be translated into impacts, direct and indirect, when one of those CIs fails.

Nowadays, as the potential threats affecting CIs are global, so are the effects of the crises they lead to. Trends like the technological advances of the last years, climate change or interdependencies among systems have modified how crises occur and evolve, making the events and their consequences more difficult to foresee and prevent. A crisis can be defined as a consequence of an unexpected triggering event that suddenly or by an accumulative process of near misses strikes the entire system (Coleman, 2004; Mitroff & Anagnos, 2000; Pearson & Claire, 1998).

Relevant crises in the las years, like Hurricanes Katrina in 2005, Sandy in 2012, the most recent Harvey in 2017, the earthquake in Japan that derived in the Fukushima nuclear accident in 2011 or the Ukraine Cyberattack in 2015 have something in common, all of them affected several CIs significantly causing important impacts in many sectors and making the recovery phase more difficult and longer (Comes & Van de Walle 2014; Chang et al. 2007). Sometimes, cross border effects and interdependencies make several nations being affected by crises, and require international cooperation. Dealing with crises requires a huge effort, and involves many stakeholders from different organizations of different nature and even from different countries.

It is clear that CIs must do all they can in order to anticipate and prevent crises but, at the same time, they should improve their capacity to act in a dynamic, flexible and creative way. Furthermore, CIs should develop skills and tools that lead them to a successful resolution of any type of crisis, predictable or not. Risk management is necessary but is not enough when facing unexpected crises. CIs must go beyond known risks and must be resilient (CSS 2011).

Resilience can be considered as a strategic property of the systems that enhances system's capacities to adapt to and face successfully any type of crisis, in a changing environment. Those capacities refer to: 1. – the ability to prevent, anticipate and then avoid a crisis; when the crisis occurs, 2. – the capacity of a system to resist the triggering event absorbing and mitigating the impacts; and 3. – the capacity to recover rapidly and efficiently, being able 4. – to learn, improve and prepare for future stressors (Ganin et al. 2016; Hosseini 2016; Francis & Bekera 2013; Hollnagel et al. 2007).

In the literature there can be found some approaches and proposals for measuring a CI resilience. The majority of them focus on their technical attributes such as the robustness of the physical assets or their redundancy level. Other methodologies

adopt a risk management approach, considering the system behavior against a specific risk. The organizational resilience perspective takes into account the sociotechnical attributes of the CIs but they don't pay special attention to the technical aspects. Furthermore, most of the methodologies lack to offer a more holistic view, and they don't provide tools to characterize both the event and the impacts (Cimellaro et al., 2010; Panteli et al. 2017; Brown et al. 2017; Labaka et al. 2016; Ouyang & Wang 2015; Aleksić et al. 2013; Petit et al. 2013; McManus et al. 2008).

To develop a methodology to evaluate the resilience level of a CI from a holistic approach, considering both the technical and the organizational aspects, it is necessary to get a deep knowledge about the crisis occurrence, analyzing how it happens and the effects it has in the CI. Furthermore, it is interesting to find the way to analyze the crisis evolution over time, to identify key cross-cutting capacities that, duly enhanced, will lead to a more resilient system.

Before defining the resilience assessment methodology, the challenge is to design an effective strategy to gather the most relevant information to create the method.

2 METHODOLOGY

Knowledge about CIs behaviour against crises is embebed into the practitioners' minds. Thus, to determine the CIs performance in a crisis, is essential to work together with all relevant stakeholders to agregate their knowledge and obtain a comprehensive vision of the whole crisis including all different points of view.

As the resilience building process needs the commitment and participation of all the stakeholders involved (Gimenez et al. in press), we can combine different methods to gather as much information as possible from all the sources available.

Individual interviews with key stakeholders are an interesting resource, as they provide precise data from their daily work that let get an idea about their personal perspectives, what kind of data they manage or how far they are familiarize with the topics under study.

If we want to collect information from a large group of people we could carry out a survey. A survey is a systematic method for gathering information from entities for the purpose of constructing quantitative descriptors of the attributes of the larger population of which the entities are members (Groves et al. 2011). The selection of a representative sampling, the adequacy of the questions and the processing of the data to reach the goals set will be crucial to get valid results from the survey.

Once the most significant issues related to resilience have been identified from a wide range of practitioners, we can work on the most relevant issues in workshops with reduced number of experts through collaborative methodologies.

Collaborative methodologies can be very useful, as people within the organization from different departments and with distinct levels of responsibility, share and discuss they knowledge and ideas to contribute to a common goal.

As a result, we will get a shared vision from the individual points of view and understanding.

Group model building (GMB) is an example of collaborative methodology that enables integrating fragmented knowledge, initially residing in the minds of different agents, into aggregated models (Scott et al. 2016; Andersen et al. 2007). GMB is based on workshops where modelers work on the problem jointly with multidisciplinary domain experts.

GMB has been employed successfully in many areas such as inter-organizational integration of information, health care system organization, organizational strategy changes, analysis of integrated operation strategy in a crude oil and gas company or CIP (Luna-Reyes et al. 2016; Hernantes et al. 2012; Ackermann et al., 2010; Rich et al., 2009).

The outcomes from the GMB process are essential to understand and identify the crisis behavior patterns, and then, to set the most representative variables to typify the resilience of the system.

This paper describes how the above explained methods have been used to identify the key elements in the CI under study for developing a methodology to assess its resilience level and improve it.

The process has been carried out within the context of a project with a CI from the energy sector in Spain called Red Eléctrica de España (REE).

3 INFORMATION GATHERING PROCESS

Crises can be analyzed from different perspectives. The analysis can only focus on the triggering event and subsequent direct impact, or it can have a broader perspective including the pre-crisis period, the peak of the crisis and the post-crisis. Furthermore, the analysis can focus on looking at how the system learns from one crisis to the next one, broadening even more the perspective.

In order to analyze the crises from different perspectives, this research uses the three units of analysis (see Figure 1) defined by Labaka et al. (2011): 1. Peak of the crisis. It focuses on the magnitude of the triggering event evaluating the immediate consequences due to the event and the impacts. This analysis provides relevant information about issues like what factors characterize the crisis, which are the most important impact variables, who are the

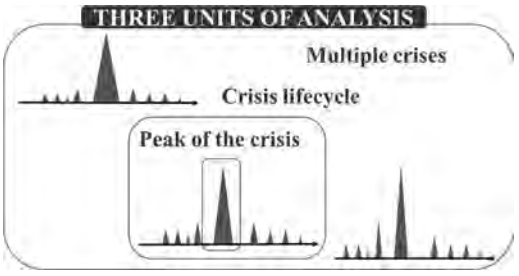


Figure 1. Three units of analysis (Labaka et al., 2011).

key stakeholders and which are the mechanisms of the CI to cope with the crisis. 2. Lifecycle of the crisis comprises the pre-crisis, the peak of the crisis and the post-crisis. The study of the complete cycle of one crisis allows to understand how the preparation and prevention activities carried out in the pre-crisis phase, affect the response and recovery activities. Finally, the last unit of analysis is 3. Multiple crises, where the period of time between one crisis and the next ones is analyzed. The objective in this case is to identify how the CI under study learns between one crisis and the next one.

This method has been successfully used in another project to analyze major power-cut crises in Europe. (Hernantes et al. 2013).

Based on the three perspectives defined, the process we followed was structured in the next three steps:

1. Analysis of the peak of the crisis.
2. Analysis of the lifecycle of the crisis.
   a. Analysis of the CI interdependencies.
3. Multiple crisis analysis.

Each of those three steps was performed for obtaining knowledge about specific objectives as it is explained bellow.

3.1 *Peak of the crisis*

The peak of the crisis covers the period of time between the moment when the crisis strikes, and the moment when the CI starts its recovery. It represents the most visible part of the crisis, in which most of the impacts appear. It is the moment where all the efforts we have made before to improve our capacity to face crises are tested.

The objectives for this first phase were: (1) To characterize the peak of a crisis in the CI under study, and (2) To determine the most important variables to characterize the impact of a crisis.

To reach the goals set we needed: to comprehend the CI under study, to understand what a crisis means for them and, finally, to know how they represent and measure their crisis.

With that aim, we asked the CI for information related to the organization itself and, in particular, everything related with crisis management, records and indicators.

The CI employees participating in the project were, firstly, interviewed individually. In particular, seven people were interviewed. The participants had technical profile and belonged to the transmision and operation divisions, more specifically to the following departments: renovation and facilities improvement, lines maintenance, substations maintenance, telecommunications, lines engineering, substations engineering and system operation.

All the interviews had the same structure. At the beginning general information was requested (company structure, departments, activities, business, …) and they were asked, in particular, about their function and responsibility in the organization. Secondly, they were asked about their knowledge regarding the crises related procedures in the organization (risk analysis, contingency plans, chains of command for crises management, reports of past crises and incidents, procedures and agreements with external stakeholders). After that, they were asked about the indicators and variables they managed to identify, report and register the crisis. We were interested, especially, in the indicators used to represent the crisis impacts.

In addition to the interviews with the participants in the project, we also interviewed the manager of the risk management department, to have the strategic view of the company about crises. The manager for the press office was also interviewed, to understand how crises are communicated externally, such as to the media, to other stakeholders and to the society.

The outcomes from this first step were:

1. A list of 21 indicators to quantify the impacts of an eventual crisis, independently of the triggering event and classified in four dimensions: structural impacts in the organization, social impacts derived from the CI failure, economic impacts, and environmental impacts.
2. A set of five representative indicators of the peak of the crisis, that will let us to represent the evolution of the crisis over time, taking into account both internal and external aspects for the CI. Those five indicators were chosen among the previously identified 21 indicators, and they are: percentage of their own infrastructures damaged (V1), percentage of users without electrical supply (V2), percentage of extra resources needed for the crisis resolution (V3), level of the company's knowledge about the situation (V4) and, finally, the public anxiety due to the crisis (V5).
3. The duration of a "standard" peak of the crisis in the CI was set in 48 hours.

4. A classification of the crises according to four extreme patterns defined based on two parameters: on one hand the resilience level of the CI and on the other hand, the magnitude of the event. Thus, we would be able to represent any kind of crises based on the resilience level of the system and the magnitude of the triggering event regardless of its origin (see Figure 2).

This first phase of interviews, highlighted some barriers that hindered the optimal data and knowledge collection. For that reason, we decided to apply, hereinafter, collaborative methodologies as GMB, to manage barriers like:

- Silo thinking: Each department is expert in a specific issue and they only focus in the aspects related to their specific field of knowledge.
- Confidentiality. Sensitive information is only available for a few people within the organization, usually "high" profiles and does not transcend to other operational levels.
- The existence of long chains of command (and levels of information) that derive in loss of perspective of the whole crisis and its impacts.
- Poorly established failure sharing culture.
- Reports about crises don't reflect a holistic view of the crisis as, in general, they focus on technical aspects and don't collect information about social or economic impacts.

3.2 *Lifecycle of the crisis*

In this unit of analysis, we extend the study taking into account the whole lifecycle of the crisis, that is: the pre-crisis, the peak of the crisis and the post-crisis.

Whilst in the peak of the crisis the focus was on the impacts, the lifecycle also considers the preparation and the recovery activities to manage the crisis. During the pre-crisis, CIs' activities are oriented towards anticipating and preventing the crisis occurrence identifying hazards, threats and vulnerabilities, and preparing plans to afford risks and reduce weaknesses. In the peak of the crisis all the efforts are concentrated on absorbing and minimizing the impacts, trying to reduce the human and material losses, avoiding the cascading effects and recovering the service, all in the shortest period of time. Finally, the post-crisis encompasses the recovery to the normal operation levels where all the equipment and physical systems damaged and affected must be repaired and restored. These activities can last several months or even years.

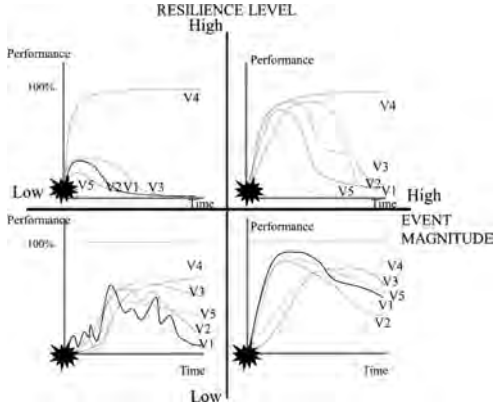


Figure 2. Four patterns for CI crises classification according to CI resilience level and the magnitude of the triggering event.

With the objective of obtaining information about the prevention, preparation, response and recovery activities in REE we carried out a table-top exercise with the participants. We put as an example a hypothetical catastrophe which damages considerably the physical systems, such as power lines and some important substations, as well as weaken the response capacity of REE leaving the system without the telecommunication service.

The session was structured in three parts.

In the first part the objective was to understand the mechanisms of CI for the response and recovery phases. With that aim, the hypothetical scenario of the catastrophe was introduced to the attendees and then, the participants were invited to describe, individually, the activities they would carried out to face the crisis to bounce back to their normal operational state. In addition, they should identify all the stakeholders involved in the crisis resolution. In particular, they were asked to match the stakeholders with the activities they should be involved in.

In the second part of the session the participants focused on the preparation phase. In this occasion they needed to think about the activities and measures that must be carried out to prevent a disaster and to be ready to solve it when it happens. We invited them not only to focus on technical issues but also to take into account management aspects. Moreover, they should identify strategic relationships with stakeholders within and outside the company for the satisfactory crisis resolution.

Finally, once the policies had been identified they recognized and analyzed the barriers and the difficulties to implement those policies.

The outcomes from this unit of analysis were:

1. The four dimensions of Resilience for REE, based on the literature (Labaka et al. 2016; McManus 2008; SMR 2016) and on the policies resulting from the table-top exercise. 1) Leadership: company management's commitment with resilience building process and its capability to promote and consolidate a cul-

ture, attitude and values based on it. 2) Preparedness: It refers to capabilities, knowledge, procedures and technical means of the company to afford crises. 3) Technical: it refers to system robustness and resistance capability in terms of impact reduction and loss of performance. 4) Cooperation: efficient management of human and material resources, internal and external, in major crises.

2. A set of policies to improve the resilience level of REE. Those policies refer to properties that the system should have and transversal activities that should be developed in order to enhance the resilience capacity of the CI.
3. A list of indicators for the preparation phase. For each of the policies identified at least one indicator is proposed, in order to be able to measure their evolution over time.
4. A list of indicators for the respond and recovery phase. As in the previous point, for each of the policies at least one indicator has been identified, to monitor and control the degree of implementation of each policy.
5. A tool to generate random scenarios of crisis (see Figure 3). Resilience is the capacity to withstand and deal with any kind of crises, predictable or not. With the idea of introducing a degree of uncertainty we proposed a tool that takes into account the potential risks but also trends, like Climate Change, demographic imbalances or interdependencies that can act as crisis enhancers. The core of the tool represents the resilience capacity of the CI.

3.2.1 *Analysis of the interdependencies*

CIs are tightly coupled than ever before constituting a very complex and strongly interconnected systems. The resilience approach takes into account the context of the CI under study and its interaction with other CIs, stakeholders, organizations, etc.

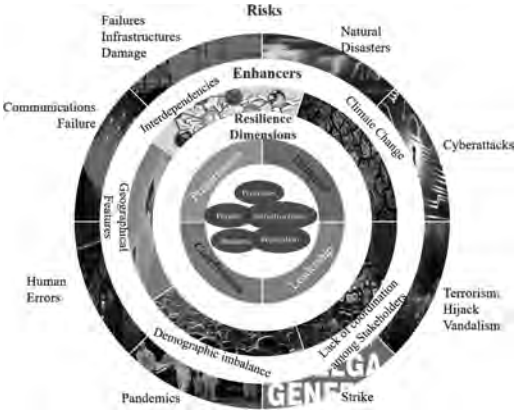


Figure 3. Random scenarios generator.

Interdependencies are the cause of many indirect impacts and cascading effects. Furthermore, they can be determinant for the recovery phase of one crisis. For that reason, when we want to evaluate the resilience level of a CI, an analisis of the interdependencies is needed.

As part of the crisis lifecycle analysis, we carried out the study of the CI interdependencies.

The objectives for this phase were:

1. Identify the CI's interdependencies:
   a. From which CIs they are more dependent.
   b. Which CIs are dependent from them.
2. Analyse in detail of the most critical interdependencies.

For the consecution of the objective 1, we decided to conduct an on-line survey within the company (CI) under study, based on the structure that Lauge et al. (2014) suggests, to analyze the CI interdependencies. We chose that method to obtain a broad view about interdependencies, and indirectly, about crisis management. In the survey 39 people participated, of whom 26 completed the entire questionnaire. They belonged to 15 different departments and most of them had technical profile.

The survey content focused on the dependencies that REE had with other CIs. We wanted to establish in which degree REE is dependent on other CIs' operation, in case of failure in any of them.

The survey proposes different scenarios of crisis due to a failure in one CI for different periods of time: less than 2 hours, from 2 to 8 hours, from 8 to 24 hours, from 24 hours to one week, more than one week.

With the obtained answers, the CIs from which REE is more dependant and the duration of the failure from which the situation becomes critical for REE were defined.

Furthermore, the analysis of the data from the survey also provided aditional information related to the perception of the employees about crisis and interdependencies. In general, there was not consensus on the issues asked. There was not a common view of the crisis and it was difficult for the employees to identify both the stakeholders and the dependencies.

To complete the interdependencies analysis (objective 2), a GMB workshop was carried out in order to share, discuss and then reach a consensus among the participants. In the first part of the workshop the results of the questionnaire were presented, explained and discussed. In the second part of the session, we focused on the analysis of the most critical scenarios according to the results of the survey.

We applied the tool dependency radar for the analysis (Laugé-Eizagirre 2014). The dependency radar provides a graphical representation of the level of dependency among CIs and helps to understand the CIs dependencies by identifying those dimensions that determine the dependency of a CI on others (see Figure 4). The tool sets out five dependency dimensions classified under two main areas. Failure area refers to how a CI can make another CI fail. Dependency for working, redundancy, and effect of external aggravating factors are the dimensions within this area. Recovery area refers to how the first CI can contribute or difficult the second CI's recovery. Recovery time and effects of resources sharing for recovery are the dimensions defined in this area.

Two scenarios of dependency were analysed with the dependency radar: the first one considered a power supply failure for 8 hours and the second one set out a 24 hours' communications failure.

They worked the cases in small groups of two and three people. Then, they were invited to explain and discuss the different radars to finally reach an agreement for each of the dependency scenarios.

The outcomes from this phase were:

1. The CIs from which the company is more dependent and the duration of the failure from which the situation becomes critical. This information is relevant as it lets REE to define actions to control and reduce, when possible, the dependency level with other CIs and, consequently, minimize the risks.
2. The period of time from which the situation becomes critical, when one of those previously identified CIs fails. This information is very important for characterizing the crises. Furthermore, it will help to design and choose appropriate prevention and mitigation actions.
3. A classification in four extreme patterns of dependencies, according to the CI dependencies to operate and to recover: FDRD (Failure Dependent Recovery Dependent), FDRI (Failure Dependent Recovery Independent), FIRD (Failure Independent Reovery Dependent, FIRI (Failure Independent Recovery Independent), (see Figure 5). This classification of the dependency relationships with other CIs by patterns allows to stablish similarities and differences among the scenarios analized.

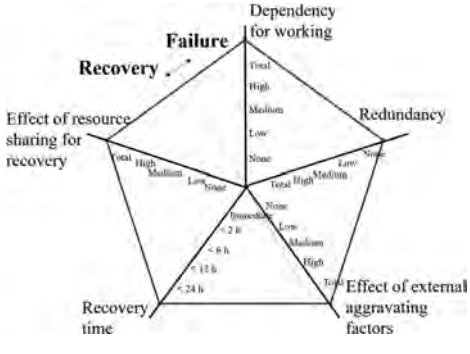


Figure 4. Dependency radar (Laugé-Eizagirre 2014).

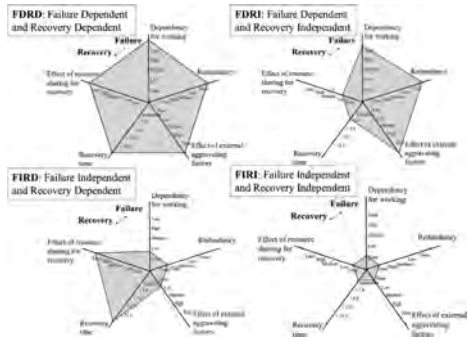


Figure 5. Dependency Patterns for dependency analysis.

3.3 *Multiple crisis analysis*

In the third unit of analysis we studied learning process from one crisis and the next one. When the learning process about the incidents or crisis is not systematize, organizations are condemned to repeat once and again the same mistakes.

Resilient systems learn from their own crises and also from others. One of the characteristics of the resilient systems is their situational awareness, an attribute that let them to maintain the warning level to detect small changes and signals that may prevent a crisis.

The objective of this third step was to understand how REE learns from its own crisis and from others'. With that aim we organized a workshop with the participants in the project. The session was structured in two parts.

In the first part we focused on the learning process and the knowledge they had about their own crisis. The exercise had four steps: 1. Identify relevant crisis of REE by heart, that is, without documentary support. The goal was to know the degree of knowledge, and awareness of the participants about their most relevant crisis. 2. To look up for information about those crises in the internal network. We wanted to get information about how they document their crises and incidents, which kind of information they report and the quality of the reports from a lessons' learned point of view. Moreover, how accessible was the information and how far they were familiarized with these searches. 3. Identify relevant changes in procedures and plans due to lessons learned from those crises. The aim was to see if the learnings turn into actions contribut-

ing to improve the resilience of the organization. 4. Identify indicators for the learning process in REE.

In the second part of the session, the objective was to analyze how REE learns from the crises that happen to others. With that aim, we asked them to repeat the four steps performed in the previous part of the session. Moreover, in this part, we asked them about their participation in national and international sectorial networks, related to crises management.

The outcomes resulting from this third unit of analysis were:

1. A list of indicators to monitor the learning process in REE. The indicators are classified in four dimensions of learning previously identified: documentation, analysis, dissemination and communities of practice. The documentation dimension assesses how far the organization reports the crises and incidents. In the analysis dimension we pay attention to the reports' quality and the reflection process and learnings derived from the incidents. Dissemination refers to how the organization spreads and shares the knowledge, lessons learned and good practices extracted from the crises analysis. And, finally, the communities of practice dimension studies how the organization promotes and participate in communities of practice about resilience topics within and outside the company, including international communities (see Table 1).
2. A set of good practices and lessons learned were identified extracted from the reports and from the workshop.
3. A list of barriers in learning identified by the particpants in the workshop and a proposal of actions for improvement.

Table 1. Dimensions and Indicators for the learning process.

Dimensions	Indicators
Documentation	Documented incidents in 1 year/total incidents in 1 year.
	Number of lessons learned from incidents in 1 year.
Analysis	Own incidents analized in 1 year/total incidents 1 year.
	Others' incidents analized in 1 year.
	Improvement actions from lessons learned in 1 year.
Dissemination	Number of incident analysis sessions in one year.
	Number of lessons learned sessions in one year.
	Man hours for incident analysis/total man hours per year.
	% of the staff whom receive direct communication of lessons learned.
	Level of knowledge of the staff about Company's incidents (0–10).
Communities of Practice	Quantity of communities of practice within the company, or conversations, related to crisis.
	Level of use of internal communities of practice (0–10).
	Company's level of presence in international communities of practice (0–10)

4 CONCLUSIONS, LIMITATIONS AND FUTURE RESEARCH

The process described in this paper has been successful for gathering the information needed to design a methodology to evaluate the resilience of REE.

Resilience building process requires of the participation and the commitment of all the stakeholders. In this way, it is very important to involve them from the early stages of the projects. While three units of analysis offers a wide view of the crisis cycle and establish an appropriate framework for the crisis analysis, the use of a combination of different methods to get the data and knowledge has created spaces for both the individual and the common reflection. Furthermore, the workshops themselves have provided an opportunity for encouraging the dialogue and the exchange of experiences around crisis management and resilience within the participants involved in the project.

However, the process has also highlighted some barriers. One of the most important is that information related to crises is treated as confidential, and only few people within the company have access to it. Related to this issue, in this case, it was not possible to interview or work with stakeholders outside the organization. Anyway the steps proposed consider also the participation of other external stakeholders.

Once the information has been gathered, the next step must be to analyze and manage all the gathered information, so it can be used to assess the resilience level of REE. Furthermore, the methodology should provide a guide to REE to improve its resilience level.

ACKNOWLEDGMENTS

This project has been promoted and supported for REE. The authors are thankful to REE for their implication in the project and their interest in disseminating the knowledge acquired.

REFERENCES

Ackermann, F., Andersen, D. F., Eden, C., & Richardson, G. P. (2010). Using a group decision support system to add value to group model building. System Dynamics Review, 26(4), 335–346

Aleksić, A., Stefanović, M., Arsovski, S., & Tadić, D. (2013). An assessment of organizational resilience potential in SMEs of the process industry, a fuzzy approach. *Journal of Loss Prevention in the Process Industries*, *26*(6), 1238–1245.

Andersen, D. F., Vennix, J. A., Richardson, G. P., & Rouwette, E. A. (2007). Group model building: problem structuring, policy simulation and decision support. Journal of the Operational Research Society, 58(5), 691–694.

Brown, C., Seville, E., & Vargo, J. (2017). Measuring the organizational resilience of critical infrastructure providers: A New Zealand case study. International Journal of Critical Infrastructure Protection.

Chang, S. E., McDaniels, T. L., Mikawoz, J., & Peterson, K. (2007). Infrastructure failure interdependencies in extreme events: power outage consequences in the 1998 Ice Storm. *Natural Hazards*, *41*(2), 337–358.

Cimellaro, G. P., Reinhorn, A. M., & Bruneau, M. (2010). Framework for analytical quantification of disaster resilience. *Engineering Structures*, *32*(11), 3639–3649.

Coleman, L. (2004) The Frequency and Cost of Corporate Crises, Journal of Contingencies and Crisis Management, Vol. 12, No. 1, pp. 2–13.

Comes, T., & Van de Walle, B. (2014, May). Measuring disaster resilience: the impact of Hurricane Sandy on critical infrastructure systems. In *Proceedings of the Eleventh International ISCRAM Conference, University Park, Pennsylvania, USA* (pp. 195–204).

DHS Department of Homeland Security (2017). Available at: https://www.dhs.gov/topic/critical-infrastructure-security [Last access: 2017, December].

EU Critical Infrastructures (2017). Available at: https://ec.europa.eu/home-affairs/what-we-do/policies/crisis-and-terrorism/critical-infrastructure_en [Last access: 2017, December].

Francis, R., & Bekera, B. (2014). A metric and frameworks for resilience analysis of engineered and infrastructure systems. *Reliability Engineering & System Safety*, *121*, 90–103.

Ganin, A. A., Massaro, E., Gutfraind, A., Steen, N., Keisler, J. M., Kott, A.,... & Linkov, I. (2016). Operational resilience: concepts, design and analysis. *Scientific reports*, *6*, 19540.

Gimenez, R., Labaka, L., & Hernantes, J. (in press 2017) Union means strength: Building city resilience through multistakeholder collaboration. Journal of Contingencies and Crisis Management.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). Survey methodology (Vol. 561). John Wiley & Sons.

Hernantes, J., Labaka, L., Laugé, A., Sarriegi, J. M., & Gonzalez, J. J. (2012). Group model building: a collaborative modelling methodology applied to critical infrastructure protection. International Journal of Organisational Design and Engineering, 2(1), 41–60.

Hernantes, J., Rich, E., Laugé, A., Labaka, L., & Sarriegi, J. M. (2013). Learning before the storm: Modeling multiple stakeholder activities in support of crisis management, a practical case. Technological Forecasting and Social Change, 80(9), 1742–1755.

Hollnagel, E., Woods, D. D., & Leveson, N. (2007). *Resilience engineering: Concepts and precepts*. Ashgate Publishing, Ltd.

Hosseini, S., Barker, K., & Ramirez-Marquez, J. E. (2016). A review of definitions and measures of system resilience. Reliability Engineering & System Safety, 145, 47–61.

Labaka, L., Hernantes, J., & Sarriegi, J. M. (2016). A holistic framework for building critical infrastructure resilience. *Technological Forecasting and Social Change*, *103*, 21–33.

Labaka, L., Hernantes, J., Laugé, A., & Sarriegi, J. M. 2011. Three units of analysis for crisis management and critical infrastructure protection. In *International Conference on Information Systems for Crisis Response and Management (ISCRAM), Lisbon (Portugal)*.

Laugé-Eizaguirre, A. (2014). Crisis Management Toolbox: The relevant role of Critical Infrastructures and their Dependencies.

Luna Reyes, L. F., Martinez Moyano, I. J., Pardo, T. A., Cresswell, A. M., Andersen, D. F., & Richardson, G. P. (2006). Anatomy of a group model building intervention: Building dynamic theory from case study research. System Dynamics Review, 22(4), 291–320.

McManus, S., Seville, E., Vargo, J., & Brunsdon, D. (2008). Facilitated process for improving organizational resilience. Natural Hazards Review, 9(2), 81–90.

Mitroff, I. and Anagnos, G. (2000) Managing Crises Before They Happen: What Every Executive And Manager Needs to Know About Crisis Management, AMACOM, New York.

Ouyang, M., & Wang, Z. (2015). Resilience assessment of interdependent infrastructure systems: With a focus on joint restoration modeling and analysis. *Reliability Engineering & System Safety*, *141*, 74–82.

Panteli, M., Trakas, D. N., Mancarella, P., & Hatziargyriou, N. D. (2017). Power Systems Resilience Assessment: Hardening and Smart Operational Enhancement Strategies. *Proceedings of the IEEE*.

Pearson, C.M. and Clair, J.A. (1998) Reframing Crisis Management, The Academy of Management Review, Vol. 23, No. 1, pp. 59–76.

Petit, F. D. P., Bassett, G. W., Black, R., Buehring, W. A., Collins, M. J., Dickinson, D. C.,... & Phillips, J. A. (2013). *Resilience measurement index: An indicator of critical infrastructure resilience* (No. ANL/DIS-13–01). Argonne National Lab.(ANL), Argonne, IL (United States).

Rich, E., Gonzalez, J. J., Qian, Y., Sveen, F. O., Radianti, J., & Hillen, S. (2009). Emergent vulnerabilities in integrated operations: a proactive simulation study of economic risk. International Journal of Critical Infrastructure Protection, 2(3), 110–123.

Rinaldi, S. M., Peerenboom, J. P., & Kelly, T. K. (2001). Identifying, understanding, and analyzing critical infrastructure interdependencies. *Control Systems, IEEE*, *21*(6), 11–25.

Risk and Resilience Research Group—Center for Security Studies. Resilience and Risk Management in Critical Infrastructure Protection Policy: Exploring the Relationship and Comparing its Use. 2011.

Scott, R. J., Cavana, R. Y., & Cameron, D. (2016). Recent evidence on the effectiveness of group model building. *European Journal of Operational Research*, *249*(3), 908–918.

SMR (2016) Smart Mature Resilience. Available at: http://smr-project.eu/home/ [Last access: 2017, December].

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A simulation-game to explore collective critical infrastructure resilience

Joeri van Laere
University of Skövde, Skövde, Sweden

Peter Berggren
Linköping University, Linköping, Sweden

Osama Ibrahim
Stockholm University, Kista, Sweden

Aron Larsson
Mid Sweden University, Sundsvall, Sweden

Susanne Kallin
Combitech AB, Sundbyberg, Sweden

ABSTRACT: Resilience of interdependent infrastructures increasingly depends on collaborative responses from actors with diverse backgrounds that may not be familiar with cascade effects into areas beyond their own sector. A simulation-game can enable societal actors to obtain a deeper understanding of the interdependencies between their infrastructures and their respective crisis responses. Following a design science approach, a simulation-game has been developed that combines role-playing simulation and computer simulation. The simulation-game challenges participants to address the interaction between payment disruptions, food and fuel supply, security problems (riots, robberies) and communication challenges (preventing hoarding). A number of crucial design choices were handled while developing the simulation-game. The main design challenges were: How to validate an unthinkable escalation scenario?; How to give the simulation a sufficient level of detail on all aspects and keep the complexity graspable so it can be played instantly?; and How much time should each playing round take?

1 INTRODUCTION

Resilience of critical infrastructures is a complex problem area. When societal actors with different backgrounds quickly need to orchestrate a collective crisis response, a deep understanding for the existing interdependencies between their respective infrastructures and crisis response strategies is required. Gaming-simulation can help to collaboratively develop a deeper understanding of such interdependencies and be a safe environment to explore the robustness of response strategies from a multi-sectorial perspective. In the context of tightly interrelated infrastructures a response strategy should not only be beneficial for individual organizations or sectors, but even mitigate consequences and limit escalations from a holistic multi-sectorial perspective.

Building a simulation game involves many design choices. Depending on which choices are made, consciously or unconsciously, very different simulations or simulation-games can be created for studying the same problem. It is important to build simulation-games of good quality and to understand how crucial design choices impact simulation-game design and simulation-game outcomes. Consequently, the contribution of this paper is a detailed description of our simulation-game design. Following a design science research approach, a simulation-game has been created that enables actors from a large variety of critical infrastructures to analyze and mitigate the cascading effects of payment disruptions on their respective infrastructures. Besides a presentation and motivation of the most important simulation-game design choices (Section 4), three major design challenges were identified: scenario validation, game complexity and length and number of playing rounds (Section 5).

2 THEORY BACKGROUND

Our research builds upon three research areas: critical infrastructures, resilience and gaming-simulation.

2.1 *Critical infrastructures and cascading effects*

Societies rely on well-functioning critical infrastructures such as Energy, Information and Communication Technology, Water Supply, Food and Agriculture, Healthcare, Financial Systems, Transportation Systems, Public Order and Safety, Chemical Industry, Nuclear Industry, Commerce, Critical Manufacturing, and so on (Alcaraz & Zeadally 2015). When one or more critical infrastructures break down or provide only limited service, large numbers of citizens, companies or government agencies can be severely affected (Boin & McConnell 2007, Van Eeten et al. 2011). Breakdowns can be caused by internal factors (human or technical failure), external factors (nature catastrophes, terror attacks) or by failures of other infrastructures as there are many dependencies between critical infrastructures (Van Eeten et al. 2011). Energy and Information Technology or Telecommunications are well-known event-originating infrastructures that generate cascading effects in many other infrastructures, as has been shown in different types of analyses (Van Eeten et al. 2011, Laugé et al. 2015). In times of increasing digitalisation and an ever increasing development towards a digitally interconnected society, security experts argue for more awareness for digital vulnerabilities, more attention for cyber security and a need to educate professionals and citizens on these matters (Hagen 2016).

Ansell et al. (2010) argue that resilience of interdependent infrastructures increasingly depends on collaborative responses from actors with diverse backgrounds that may not be familiar with cascade effects into areas beyond and outside their own organisation or sector. Boin & McConnell (2007) and Van Eeten et al. (2011) argue that there is limited empirical evidence of cascading effects across many infrastructures, which makes it hard to foresee which interactions may occur across sectors. Risk analysis, business continuity management and crisis management training are often performed within the context of a single organisation or sector and are seldom addressing the holistic analysis of multiple infrastructures (Van Eeten et al. 2011).

More research is needed to understand collective resilience in the context of critical infrastructure management. In this study, a contribution is made by focusing on one application area, i.e. how payment disruptions impact other critical infrastructures. Despite the long term efforts of public and private actors in the financial sector in Sweden to identify, analyse and understand risks and to develop routines for preventing and mitigating serious disruptions in the payment system in Sweden, there is still a lack of insight into how the proposed action plans exactly need to be executed and how numerous other actors in society (e.g. citizens, food stores, gas stations, voluntary organizations, governmental agencies and so on) will act in case of a temporary or complete breakdown of the payment system. For instance, several key actors in the payment system have in earlier studies expressed that they will take a larger responsibility than their formal responsibility (MSB-2009-3309 2010), but it is not clear what this implies and how these organizations actually will act when crisis hits.

2.2 *Resilience*

Lundberg & Johansson (2015) and Bergström et al. (2015) list that resilience amongst others can refer to: bouncing back to a previous state, or bouncing forward to a new state, or both; absorbing variety and preserve functioning, or recovering from damage, or both; and being proactive and anticipating, or being reactive (when recovering during and after events), or both. Given the variety of interpretations of resilience, resilience is hard to operationalize into measurable indicators (Lundberg & Johansson 2015).

Lundberg & Johansson (2015) made an effort to merge and compile different points of view in the field of disaster and crisis response resilience into one systemic model, the *Systemic Resilience Model* (SyRes). The model departs from the idea that the coping with an unwanted event can be seen as a downward spiral activating certain basic resilience functions (anticipation, monitoring, responding, recovery and learning) and their associated strategies (where the strategies are the actual manifestation of the functions, or their 'form', which may differ from system to system). Further, Lundberg & Johansson (2015) suggest that resilience is needed to protect *core values*, i.e. values central for the existence of the system in focus. In safety-critical systems, such core values usually take the form of maintaining safety, such as avoiding harm to humans or critical infrastructures. For a commercial business such as a grocery store, a petrol station or a bank, a core value is typical to create revenue, i.e. to assure a higher income than outcome. Without this profit, the business will seize to exist. This core value will manifest itself in a number of practical activities which usually take the form of different flows such as goods, money, services etc.

In line with the challenges to resilience suggested by Johansson & Lundberg (2010) comes the fact that most systems in society, such as the payment system, depend on several different actors to function properly. Therefore, resilience must be considered from a systems perspective. In the field of resilience, this is sometimes referred to as '*collective resilience*'. Weick & Sutcliffe (2007) argue that

loosely coupled systems relying on a '*sensemaking*' process generally are more resilient than tightly coupled systems based on the assumption that all system states can be predicted and safeguarded against possible threats. This resembles distinctions made in safety science between the paradigms labelled Safety I and Safety II (Hollnagel 2013) where Safety I is signified by the idea that safety can be designed into a system and Safety II is signified by the idea that human adaptability is the most important contributor to success despite inadequate design or insufficient predictive capacity of safety engineers. Weick & Sutcliffe (2007) argue that a dilemma exists in sensemaking: you can optimise for analysis or action, but not both. This dilemma seems contradictory to the requirements of resilience, because Weick & Sutcliffe argue for sensitivity to operations and reluctance to simplify (i.e. an interest in details and scrutinize the situation at hand) and simultaneous blunt and immediate action without thorough analysis. The solution suggested by Weick & Sutcliffe (2007) is that deep knowledge about the system should have been acquired earlier (long before the disruption) so that quick and blunt action based on deep understanding of the system's dynamics is possible in case of disruptions. As more actors may simultaneously initiate a quick and blunt response, a risk is that these responses counteract each other. Weick & Roberts (1993) discuss how attentiveness (heedful interrelating) is key in a resilient group response, i.e. while acting quick and blunt, various actors should pay close attention to how other actors respond and to what kind of system behaviour their collective response leads. Heedful interrelating has been demonstrated in small groups. Heedful interrelating becomes challenging when systems become larger, more interrelated and involve more and more decision makers that do not really know each other and do not understand the impact of their decisions on nearby systems, as in the case of large interdependent infrastructure systems (Ansell et al. 2010). Then these groups of stakeholders may lack swift trust (Weick & Roberts 1993) and may lack a shared understanding of the situation and a shared vision, which may lead to inferior performance (Berggren et al. 2014). Yet another risk might be organisations or companies who continue putting their own goals ahead of the common good, thus risking initiating counterproductive actions that may hamper the process of recovery from disruptions.

2.3 *Gaming-simulation*

Gaming-simulation is defined as a specific form of simulation. Simulation in general aims at designing a model of a system in a complex problem area in other to be able to experiment with the model. Deeper insight in the behavior of the system is created by evaluating various operating strategies against each other in one ore multiple scenarios. Gaming-simulation differs from other forms of simulation in that it incorporates roles to be played by participants and game administrators, implying that people and their (goal-directed) interactions become part of the simulation (Laere et al. 2006). In addition to role descriptions and interaction formats, simulation-games can also include a physical simulation model (a board game, a mock-up, a computer simulation, or any other representation of a physical reality) which the game participants need to interact with. It is important to understand that both the changes and impacts of changes to the physical simulation model in the simulation-game and the interaction between the participants (often negotiation processes about what to change and how to interpret changes in the physical simulation model) are part of the simulation-game and object of study (Mayer 2009). Gaming-simulation is especially relevant when the "*how and why*" of the interaction processes between the participants are of interest and when these interactions cannot easily be incorporated in computer simulation models. In addition, it creates a deeper learning opportunity, as simulation-game participants literally are active participants in the simulation, rather than passive observers of a computer simulation.

To design a high quality simulation-game, many design choices have to be taken into account, which often are not self-evident, but rather involve tricky cost-benefit analyses ending up with a dilemma (is the benefit worth the extra cost?). Examples of such design choices are for example (Laere 2003, Mayer 2009, Meijer 2009): defining a limited number of research or learning objectives, defining the number and content of roles, defining the scope of the modelled situation/problem, guaranteeing the validity of the simulation, defining rules and constraints, defining the load (difficulty), choosing the location/environment where the game will be played, selecting the type of participants to be invited, design of qualitative and quantitative data collection during the game, degree of realism of the scenario, degree of complexity of the game (often phrased as modelling internal complexity of the system to be modelled, but creating external simplicity, i.e. an easy to understand and easy to play game for the participants), degree of competition, degree of dynamics, macro cycle (preparation, playing, debriefing, follow-up), micro-cycle (number of playing rounds) and real-time or symbolic-time.

3 RESEARCH DESIGN

Our research design is based on an inductive research strategy and a qualitative research method.

A clear theory on how critical infrastructures exactly are related, and how the many actors involved collaboratively could manage disruptions that create cascading effects in many infrastructures, is lacking. As such, there is a need for theory building rather than theory testing, which leads us to an inductive research strategy (Eisenhardt & Graebner 2007). From an interpretative perspective, we are interested in exploring the many different interpretations of actors involved regarding what challenges disruptions can pose and how they could be handled collaboratively across the affected infrastructures. A simulation-game can be a safe environment where participating actors can experiment with different action alternatives, and through their participation and their choice of resilience strategies demonstrate the core values they hold.

For the design of the simulation-game a design science research strategy is adopted. The result of design science research is a purposeful artifact created to address an important organizational problem (Hevner et al. 2004). In our case, the problem is "*understanding critical infrastructure dependencies and exploring collective infrastructure resilience strategies*" and the artifact is "*a simulation-game that can serve as save analysis, learning and exploration environment*". As argued in Hevner et al. (2004) design science is an iterative search method aiming at identifying a creative solution for the problem at hand. Given our interpretative stance, our aim is not to design the best or an optimal simulation-game, but rather to design one appropriate simulation-game (amongst many alternatives), and developing a deep understanding what the benefits and drawbacks of our chosen design are. Design science addresses relevance by a strong interest the societal needs in the application environment studies, and aims simultaneously at rigor through reflecting on the design process and arguing how the produced solution informs the research front (where either the produced artifact and/or the insights regarding *how to design* such an artifact can be research contributions).

A first data collection phase consisted of document study of prior incidents (33 reports), 6 interviews with key representatives from each sector and two half-day workshops with respectively 26 national and 11 local actors in order to identify cascading effects, consequences, actors involved and potential mitigating actions which they could perform with regard to payment disruptions (Laere et al. 2017a). Mapping these characteristics of our problem environment contributed to identification of the elements to be simulated in our simulation-game. A second data collection phase aimed at analysing existing simulation-games for critical infrastructure resilience (Laere et al. 2017b). Here, six existing simulation-games where analysed in detail with the purpose of understanding how different design choices impact the capabilities of the learning environment and the learning experience of the participants.

Next, the collected data was analysed and transformed to elements of the envisioned simulation-game. During a series of six bi-monthly organised full day workshops with the project team of 10 researchers, different versions of the simulation-game were created, tested and refined. In between the workshops the involved researchers worked in smaller task forces on different elements of the simulation-game. During the last to full day workshops societal actors from the different sectors were involved to gather their feedback on the simulation-game design. The next two sessions summarize the main design choices and main design challenges that were identified and dealt with under this design process.

4 GAME DESIGN CHOICES

4.1 *Game overall structure*

When role playing simulation games and computer simulations are combined a powerful simulation environment is created. Actors, as game participants, can collaborate or compete with each other in different rounds, enter their decisions in the computer simulation and receive the output of the computer simulation as input in their next playing round. As such, participants can experience social interaction (role playing) and large scale system dynamics (impacts of their decisions over time, or on a large scale). The participating decision makers can compare intended consequences with unintended and unexpected consequences and create a deeper understanding of the system as a whole and the behavior of other game participants.

The main purpose of the simulation-game is to create a deeper understanding of the dynamics and interdependencies in the overall system. Alternatively or additionally, collaboration between the different actors involved could be a learning goal. When collaboration is a learning goal, actors may be placed in different rooms and different actors may have different information at hand. In such games sharing the right information with the right actor at the right time might be in focus. In our design became clear quite early that grasping the complexity of the overall societal system (i.e. all sectors that are impacted by payment disruptions) and their interactions is a challenge at such. It was decided that grasping this complexity created sufficient load and that additional collaboration challenges would adventure the main objective of understanding overall system dynamics. Therefore

it was decided that the players, who each can represent different societal roles (i.e. food sector, fuel sector, media etc.) would be placed in one team that in collaboration would try to manage payment disruptions.

Putting the participants in one team makes the use of simulation-game flexible. Teams could consist of either 3, 5, 7, 9 or 11 participants interacting as one team with the computer simulation. From a learning perspective it is preferable to have a larger group with a strong diversity in backgrounds, but from an execution perspective it is a benefit that a simulation-game session still can be performed even if two of the seven participants would not show up.

The team interacts with a fictive society represented in the computer simulation. The computer simulation is created with Anylogic simulation software. The main reason to choose this software package is that it enables to combine agent-based simulation, discrete event simulation and system dynamics simulation, which gives us a certain flexibility to implement different scenarios. The computer simulation covers a typical region with some cities and some countryside, where relevant societal infrastructures can be distinguished (see 4.2). The overall idea is that payment disruptions occur (see 4.3) in this fictive society and that the team can try out different combinations of actions strategies (see 4.4) to learn how they differ in impact on a number of performance criteria (see 4.5). An important characteristic of the simulation-game is that the participating teams can re-play the same scenario over and over again (see 4.6). By keeping the scenario conditions constant the participants can really compare their chosen action strategies and experience and learn how different combinations of actions give different impacts.

During the design process we have alternated between versions that could be played at a distance, or at one physical location. Playing at a distance allows for more elaboration time between playing rounds which might be beneficial for learning (i.e. making more thoughtful choices). While keeping the alternative of playing at a distance as a potential future development, our current impression is that the intense discussion and interaction between the participants in the team are of major importance (as the learning and creation of deeper insight occurs exactly there). Therefore physical presence at one location is to be preferred.

4.2 *Sectors represented in the computer simulation*

From the document studies and workshops with societal actors (Laere et al. 2017a) a number of societal actors, sectors and processes has been selected that are primarily vulnerable for payment disruptions and therefore form the core of computer simulation of the fictive society in the simulation-game.

The fictive society consists of a number of grocery stores of varying size, a number of fuel stations and a number of pharmacies (where medicine can be bought). For each store a customer flow is created. The number of customers, their demands, and the number of stores are balanced based on statistics for typical regions in Sweden. Stores offer one or several of the following payment options (card payment, cash payment, digital phone payments and delayed invoice payments). Individual customers have also one or more different payment options available. When customers collect goods in the store the store's payment options and their payment preferences need to match to create a transaction. Payment transactions are performed and accredited by the actors from the finance sectors (i.e. credit card companies and/or banks) and lead to account changes for stores and customers. When goods are sold new goods are order and delivered by transport companies. Customers and transport companies consume fuel, which in turn requires financial transactions when they buy new fuel. ATMs are available for those customers and transport companies who want to acquire cash and ATMs are refilled by certain transport companies. Security guards are present at the larger stores, and more could be hired when needed. Different media actors are represented who can spread news which in turn can influence consuming behavior.

In our current implementation there is a rather rough logic. The purpose in the development has been to quickly arrive at an implementation that can be played with actual representatives from different critical infrastructure managers. Given their feedback in early playing sessions the simulation-game will be further refined. Our aim is to perform 30 playing sessions in 2018 and 2019 and gradually improve the design science artifact under study.

4.3 *Payment disruption scenario*

Thus far one main scenario has been developed and implemented. During the course of our project (2016–2021) two additional scenarios will be created. Our current scenario is a 10-day card payment disruption at the store level. The other scenarios will be developed in such a way that they effect other parts of the payment system (i.e. disruptions in the transferring of money between accounts—or a long term scenario that covers multiple years rather than only a few days).

The current 10 day card-payment scenario is based on the fact that 90% of transactions in stores in Sweden is based on card payment, which makes

the Swedish society extremely dependent on that payment option as the other alternatives are not capable to instantly handle such large volumes of transactions. Although the scenario is much more detailed than presented here, the main elements of the scenario are as follows.

Day 1: Card payment disappears as payment option. The expectation of most actors is that it will take some hours. Stores close or offer digital phone payments or cash payments as alternatives. Chaotic scenes for those customers who are disappointed. Queues at stores and at ATMs.

Day 2–3: Banks and media announce that the disruption will take several days. Customers are confused where they can buy. Sales drop dramatically, use of cash and digital payments increase dramatically, some customers start hoarding, deliveries and logistics to stores are a mess as major fluctuations occur. A lot of cash in stores and in society at large increase robbery risks.

Day 4–5: Cash and digital payment options collapse as well as they cannot cope with the large volumes. Long queues, angry customers as they are running out of goods at home, customers become aggressive, a lot of stores close, those who are open experience massive hoarding. Perishable goods need to be thrown away as they cannot be sold. Logistics trouble increases.

Day 6–7: Government in collaboration with stores introduce a general "*buy based on your identity and pay later by invoice option*". Massive hoarding when stores open. Logistics collapse again as they have hard to adjust from total sales stop to massive hoarding.

Day 8–10: The general "*buy based on your identity and pay later by invoice option*" is too complicated and time consuming which creates enormous queues, frustration and aggression. Chaos and panic on more and more places. Police and army guard the few stores that still keep open.

The cascading effects that occur are not hard implementations, but do occur as cascading effects as a result of the initial card payment disruption. All other effects can be influenced when other actions are chosen by the players.

4.4 *Action alternatives to mitigate disruptions*

The team that plays the simulation-game in several rounds can select on ore more of the following actions. Besides these alternatives that are given (and prepared) we are open for creative ideas of the participants. When they come with a suggestion for an unforeseen action the game facilitators will try to simulate that action and its presumed impacts instantly in the simulation if possible.

Possible actions that the team can select are for example (note that each action can be implemented at any day in the scenario): offer more/less payment options at all or some stores; close or open stores; increase/decrease deliveries to stores; communicate information or instructions to customers; offer cash withdrawal in stores; limiting the amount of goods per purchase; increase/decrease the number of security guards for one or several stores; throw away perishable goods; give away perishable goods for free.

The design of the computer simulation involves an implementation of impacts of each and every action, based on interviews and discussions with key representatives from the different societal processes simulated. Even as we as designers know the approximate impact of individual action, the playing sessions need to reveal how the different actions in combination fall out. In addition, actions can be implemented on different moments in time (day one to ten in the scenario), which makes the number of alternative strategies near to infinite. Rather than experimenting with the computer simulation as such ourselves, the whole idea with involving real societal actors in role-playing is to let their expertise and value frames guide the selection and time-planning of combinations of actions. Moreover, not only the selection of actions as such is of interest, but also the motivation and reasoning behind. Therefore, the teams who play need to motivate the timing and selection of actions before they are implemented in various playing rounds and the collection of these motivations is seen as a crucial element of the simulation-game.

4.5 *Performance metrics*

Extensive discussions have been held at several of our design workshops and in intermediate work group meetings considering what indicators are most relevant and appropriate to visualize performance in the various sectors of society. Currently, three major performance areas have arisen: 1) payment options, 2) good flows, and 3) security

Available payment options are statistics on the actual use of each of the four different payment over time, or the amount of stores (in% of total stores) where they each option is available.

For good flows the main indicators is "disappointed customers" over time (the simulation counts the number of arriving customers that cannot fulfil their purchase for any reason). Additionally it is shown how many stores currently are closed (in%), which groups of goods currently are out of stock, how many perishable goods are destroyed over time, and how many planned deliveries that fail (due to fuel shortages).

Security related indicators are amount of cash in stores (implying increased robbery risk), number of shop lifting incidents, and the number of security guards per store.

A performance area which has been suggested but been hard to implement thus far is "trust". Although trust is a core value in society, it can be different kinds of trust (trust that you can obtain certain goods, trust in banks and stores, trust that you will be safe when being out in society). Our current interpretation is that trust depends on the other indicators and that is thus might be sufficient to only model them.

4.6 *Replay-ability*

After a short introduction into the learning goals, the computer simulation environment, the start scenario, and the way how the team can choose actions to influence the scenario, the team can play an optional number of rounds. When the start scenario is introduced the simulation is paused at day 1, day 3, day 6 and day 10 to show how the performance measures slowly deteriorate.

When the team later plays itself and chooses actions the simulation-games is initially paused at the same moments to be able to compare the new performance statistics with the earlier ones. Typically, it takes 10 to 20 minutes to discuss and decided on actions, so 1 to 1½ hour to play the full scenario once. Our expectation is that teams might succeed to play 3 rounds on a half-day (leaving time to sum up and debrief the whole playing sessions) and maybe 6–8 rounds one a full day (where the expectation is that playing speed slowly can be increased when the team plays more rounds as they get familiar with the simulation-game).

5 GAME DESIGN CHALLENGES

Most design choices have after some iterations and refinements evolved into more permanent choices where motivation why each respective choice was important gradually became more profound. Three design issues have been particularly challenging and are therefore interesting to highlight as potential areas for future research.

5.1 *Validation*

How to validate an unthinkable crisis escalation scenario? Many of the interactions that are simulated in the computer simulation are based on slightly related incidents and expectations of experts we have interviewed. It is however hard to translate observed effects of poorly related cases or judge the imaginary power of the experts. There might be certain interactions that are hard to imagine and which are not correctly represented in our current simulation. Normally, when building a simulation of an existing system, there is some kind of real data to validate against. As the purpose of crisis scenarios is to be far from the current equilibrium state, it is hard to foresee or imagine what relevant (new) elements and (new) interactions and dependencies are. An interesting future research area is therefore to develop methods and tools to improve the validation of crisis scenarios and simulations.

5.2 *Fidelity and playability*

A major concern in our current design is that players easily can get stuck in details. Multiplying 25 stores and several other actors with 3 decision points in time and roughly 15 different types of actions that each individual store can pick at each point in time results over 1000 potential actions which can be combined in infinite variations. Even though our simulation is a strong simplification of the actual complexity of our society, players might easily get lost here. It has particularly been clear that players easily can zoom in on individual decisions in individual stores and loose the "*overall society helicopter view*". This is a typical risk of introducing a detailed computer simulation in the role playing simulation.

In our current design discussions different options are explored to handle this issue. One is to develop facilitator strategies to keep the playing teams on track (while keeping the fine granularity of the computer simulation interface). Another is simplifying the computer simulation interface (i.e. limiting the amount or granularity of actions to be taken & decreasing the number of performance statistics). The latter has the danger that the simulation becomes to abstract and transferability between simulation-game learning and value of the lessons learned in real society is lost.

5.3 *Time per playing round and number of rounds*

A closely related concern is the number of playing rounds and the time per playing round for discussion in the team. More round is preferable, but they should not become so short that players quit discussing their motivations and just guess. On the other hand, teams might get stuck in endless discussions about which actions to choose without ever implementing them in the computer simulation.

Here, well-experienced facilitators are currently seen as the major viable option to fix this challenge. Alternative options could be to allow for playing the simulation independently at a distance after participating in the first facilitated team session.

6 DISCUSSION AND CONCLUSION

During the last two full day workshops where the latest version of the simulation-game was tested it was concluded (by designers and potential players, i.e. representatives from societal sectors) that the current design potentially can increase insight in collective critical infrastructure resilience. The main challenge is to make sure that the team who plays the game does not get stuck in details (due to complexity) and that the game facilitation is of such quality that a reasonable playing speed and number of playing rounds is achieved in a session, while at the same time team players experience to have sufficient time in each playing round to come to thoughtful and well-motivated action packages.

Researchers and practitioners can benefit from an increased insight into the challenges of designing simulation-games for critical infrastructure resilience analysis and training, as documented in this paper. Combining the insights from our design process with insights from alternative applications and approaches can increase the quality of our designs and thereby subsequently improve overall critical infrastructure resilience in society.

ACKNOWLEDGMENTS

This research was supported by Grant 2016-3046 of the Swedish Civil Contingencies Agency.

REFERENCES

Alcaraz, C. & Zeadally, S. 2015. Critical infrastructure protection: Requirements and challenges for the 21st century. *International Journal of Critical Infrastructure Protection* 8: 53–66.

Ansell, C., Boin, A. & Keller, A. 2010. Managing Transboundary Crises: Identifying the Building Blocks of an Effective Response System. *Journal of Contingencies and Crisis Management* 18: 195–207.

Boin, A. & McConnell, A. 2007. Preparing for critical Infrastructure Breakdowns: The Limits of Crisis Management and the Need of Resilience. *Journal of Contingencies and Crisis Management* 15(1): 50–59.

Berggren, P., Johansson, B., Baroutsi, N., Turcotte, I. & Tremblay, S. 2014. Assessing team focused behaviors in emergency response teams using the shared priorities measure. *Proceedings of the 11th International ISCRAM Conference*, *University Park, Pennsylvania, USA*: 130–134.

Bergström, J., van Winsen, R & Henriqson, E. 2015. On the rationale of resilience in the domain of safety: A literature review. *Reliability Engineering and System Safety* 141: 131–141.

Eisenhardt, K.M. & Graebner, M.E. 2007. Theory Building from Cases: Opportunities and Challenges. *Academy of Management Journal* 50(1): 25–32.

Hagen, J.M. 2016. Cyber security – The Norwegian way. *International Journal of Critical Infrastructure Protection* 14: 41–42.

Hevner, A., March, S., Park, J., & Ram, S. 2004. Design Science in Information Systems Research. *MIS Quarterly* 28(1): 75–105.

Hollnagel, E. 2013. A tale of two safeties. *Nuclear Safety and Simulation* 4(1): 1–9.

Johansson, B. & Lundberg, J. 2010. Engineering Safe Aviation Systems – Balancing Resilience and Stability. In: D.J Garland, J.A Wise & V.D. Hopkin (Eds.), *Handbook of Aviation Human Factors*: 6-1 to 6-8. Boca Raton: CRC Press.

Laere, J. van 2003. *Coordinating distributed work, Exploring situated coordination with gaming-simulation*. Doctoral dissertation, Delft, The Netherlands: Delft University of Technology.

Laere, J. van, Vreede, G.J. de., & Sol, H.G. 2006. A social simulation game to explore future coordination in knowledge networks at the Amsterdam Police Force. *Journal of Production Planning and Control* 17(6): 558–568.

Laere, J. van, Berggren, P., Gustavsson, P., Ibrahim, O., Johansson, B., Larsson, A., Lindqwister, T, Olsson, L. & Wiberg, C. 2017a. Challenges for critical infrastructure resilience : cascading effects of payment system disruptions. *Proceedings of the 14th International Conference on Information Systems for Crisis Response and Management Albi, France, 21–24 May 2017 (ISCRAM2017)*: 281–292.

Laere, J. van, Ibrahim, O., Larsson, A., Olsson, L., Johansson, B., & Gustavsson P. 2017b. Analyzing the implications of design choices in existing simulation-games for critical infrastructure resilience. *Proceedings of the International Simulation and Gaming Association's conference (ISAGA), Delft, The Netherlands, 10–14 July 2017*.

Laugé, A., Hernantes, J., and Sarriegi, J.M: 2015. Critical infrastructure dependencies: A holistic, dynamic and quantitative approach. *International Journal of Critical Infrastructure Protection* 8: 16–23.

Mayer I.S. 2009. The gaming of policy and the politics of gaming: A review. *Simulation & Gaming* 40(6): 825–862.

Meijer S.A. 2009. *The organization of transactions: Studying supply networks using gaming simulation*. Wageningen, The Netherlands: Wageningen Academic Publishers.

MSB 2009-3309 2010. Gemensamma rutiner, uppdrag inom SOES. (Swedish Civil Contingencies Agency, Shared routines, assignment within Collaboration Area Economic Security), available at: https://www.msb.se/Upload/Forebyggande/Krisberedskap/Samverkansomraden/Gemensamma%20rutiner_ver%201.0.pdf.

Van Eeten, M., Nieuwenhuis, A., Luijf, E., Klaver, M. & Cruz, E. 2011. The state and the threat of cascading failure across critical infrastructures: The implications of empirical evidence from media incident reports. *Public Administration* 89: 381–400.

Weick, K.E. & Roberts, K. 1993. Collective mind in organizations: Heedful interrelating on flight decks. *Administrative science quarterly* 38(3): 357–381.

Weick, K.E. & Sutcliffe, K.M. 2007. *Managing the unexpected. Resilient performance in and age of uncertainty*. San Francisco: Jossey Bass.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Resilient performance in response to the 2015 refugee influx in the Øresund region

H. Degerman, S. Bram & K. Eriksson
RISE Research Institutes of Sweden, Gothenburg, Sweden

ABSTRACT: September 2015 saw a sharp increase in the influx of refugees in the Øresund region. In this study, resilience defined as flexible adaptation was taken as a baseline to guide interviews with societal infrastructure actors and NGOs engaged in managing the situation. Different actors had different organisational preconditions that influenced their ability to adapt to the new situation. Among the strongest drivers behind resilient performance were the organisation's ways of relating to established rules, regulations, procedures and processes, the way relationships were formed between people and hierarchical layers within the organisations, and the perceived value of the human operator and the human contribution within the organisational whole. These values, in turn, determined how the organisations shaped many of the basic conditions that allowed resilient performance to develop. In the study it was found, for public actors in particular, that the criteria necessary to adapt to the situation were not met by organisational structures and processes.

1 INTRODUCTION

In the Øresund region, which consists of Denmark and the southernmost province of Sweden (Skåne), an increase in immigration was noticed during the spring and summer of 2015, but fluctuations were still within the normal range. However, in the beginning of September, the number of refugees reached unexpectedly high levels in just a few days, rising to the highest levels since the Second World War. In October the amount of asylum seekers doubled compared to the month before, and in November, with the argument that the large amount of refugees was threatening national safety and straining critical infrastructure functions to an unacceptable level (SOU 2017:12) the Swedish government decided to initiate border controls, which continued in steps during the rest of the year.

Refugees travelled to or through Denmark, Sweden, Norway and Finland, some reaching the Nordic countries by boat from Germany, but most travelling over the Øresund bridge, arriving in Malmö which is the third largest city in Sweden. Even though structures for the reception of refugees existed, the volume and rapid increase of refugees was a surprise for most of the organisations involved. The situation put a massive stress on infrastructure and vital societal functions, especially in the southern part of Sweden, and some organisations went into formal crisis management.

This article, which represents a limited part of a larger study, focuses on drivers and barriers for resilient performance within a group of organisations involved in immigration management in Malmö. The aim of this article is to identify areas of future organisational research and development that could contribute to better support for such performance.

2 STUDY OF CRITICAL INFRASTRUCTURE RESILIENCE IN THE ØRESUND REGION

This case study on critical infrastructure resilience was performed within the EU Horizon 2020 project IMPROVER. In the IMPROVER project, the concept "critical infrastructure" is defined in the following way:

Critical Infrastructure is an asset, system or part thereof located in Member States which is essential for the maintenance of vital societal functions, health, safety, security, economic or social well-being of people, and the disruption or destruction of which would have a significant impact in a Member State as a result of the failure to maintain those functions.

Resilience is a popular concept that is interpreted and applied in different ways (Bergström & Dekker, 2014; Woods, 2015). In this study, resilience defined as flexible adaptation was taken as a baseline. When studying resilience, researchers should strive to understand the factors that allow organisations to uphold their functionality despite changes in context, time constraints and workload, i.e. adaptability, how adaptability may manifest itself and evolve over time. This stance is exemplified by Woods (2015) saying that the search

for understanding rebound in past events should focus on understanding how the system changed for managing the new circumstances. This notion was used as the starting point for this case study. The rapidly increasing number of refugees within a short period of time meant that many organisations had to change their practices, particularly the public institutions on which society relies. In this study, the narratives provided by these actors were examined for signs of resilient performance.

2.1 *Research method*

A literature review of organisational aspects of resilience was conducted. In addition, several organisations were interviewed: DSB Trains (Danish national railway operator), the Danish police force, the Øresund bridge consortium (Danish and Swedish sides), the Swedish Migration Agency, Malmö Municipality, the Swedish Civil Contingencies Agency, the Swedish Armed Forces, Jernhusen (operator of Malmö central station), the Red Cross (Danish and Swedish) and Kontrapunkt (a Swedish autonomous NGO). In total, 28 semi-structured interviews were performed.

Since resilient aspects are not operationalised in most organisations today, informants could not be asked direct questions about resilience. Therefore a thematic analysis (Bowen, 2009) of the entire empirical and theoretical body was suitable, since such an analysis is directed towards the empirical data as a whole, not only towards already predetermined rules. Reflections over the empirical interview material were based on Alvesson's (2011) reflection approach and framework, where traditional notions of the interview have been complemented with eight metaphors with a social, psychological and linguistic nature.

An overarching goal of the IMPROVER project is to create an indicator hierarchy for resilience in critical infrastructure, but from a qualitative ethnographic perspective this could be problematic. Describing the performance of a complex system in terms of measurements may not provide the most useful information, because such system interactions typically require qualitative descriptions. The use of indicators is a logical non-contextual exercise, while real decision-making processes in everyday work or crisis is a complex analytical process (Salmon et al., 2014). Because of these difficulties, the abstraction hierarchy framework from Rasmussen (1985) was used as an inspiration when describing the identified themes from the thematic analysis in terms of indicators (or "things to look for in story-telling about past events in organisations"). The Rasmussen framework was originally intended for the design of technical decision support in complex systems, typically an automated system and a human operator combined. In this study we applied the framework to an organisational structure, where artefacts rather are procedural than technical.

2.1.1 *Result of thematic analysis*

The thematic analysis of the entire collection of empirical material resulted in four overarching themes. These themes represent areas where key examples of organisational resilient performance were identified.

1. Design of roles, tasks and processes
2. Artefact design: procedures and tools
3. Strengthening collaboration
4. Learning and re-design

Beyond these four areas, a fifth theme was identified which is referred to in the study as "Underlying values and interpretations", representing the way in which the other themes are interpreted. In the view of the authors, striving for resilience is not only about knowing what organisational abilities to enhance, but also about the way in which such abilities are sought—organisational values that may allow or deny developments that enable resilient performance. This article focuses on how design could support resilient performance.

2.2 *Focus of this article*

While the engineering community has spent many years looking for ways to measure and increase the reliability of isolated components in critical infrastructure, not nearly as much attention has been given to interactions within socio-technical systems. Paradoxically, this is precisely where the causes of many of the great disasters of our time can be found (Woods, Leveson, & Hollnagel, 2012). As time passed after the initiation of this study, official crisis evaluation reports were published (e.g. RiR 2017:4; SOU 2017:12) and measures were established. These measures have largely consisted of more administrative routines, plans and control processes, with little respect for the fact that similar structures caused problems when public organisations were faced with unexpected and rapidly evolving events. This article focuses on how design aspects within the five identified themes could equip these organisations with better adaptive capacities.

3 SIGNS OF RESILIENCE IN REFUGEE RECEPTION IN THE ØRESUND REGION

The results of this study reflect the notion in systems oriented safety research that because of system complexity, work is always to some extent

under-specified, and that humans in a socio-technical system should be seen as a unique asset instead of an unreliable and risky system component (Cook, 1998). Our study exposed several examples of resilient performance, i.e. adaptation to evolving circumstances and needs, both on a societal level and within each organisation.

3.1 *First response at Malmö central station*

The security guards of Jernhusen, the organisation that manages Malmö central station, made the first observations of the increasing amount of refugees at the station. Jernhusen informants said that they normally respond quickly to different circumstances, which explained their fast response in this situation. Jernhusen is a relatively flat organisation with few formal procedures and they have short paths of communication between hierarchical layers. Informants said decision-makers have a tradition of using the information provided by operative guards to build a picture of what happens at the station and to determine their response. Jernhusen functioned as a central hub during the entire autumn, organising space and functions within the station's walls and donating conference rooms where the involved organisations could meet twice a week. At an early stage, Jernhusen had to fight hard to get the attention of public agencies, and to get representatives of those agencies to visit the station and make their own assessment.

3.2 *Adaptations within public organisations*

The most affected public agencies were the Swedish Migration Agency and the Malmö municipality. In addition to the more obvious long-term responsibility of the asylum process, the Swedish Migration Agency was also responsible for the short-term housing of adults and families, while Malmö municipality was responsible for the housing of unaccompanied children. In the operative functions of these organisations, typically on or near the accommodation sites, the employees rapidly adjusted to the new circumstances.

As the workload quickly increased it was no longer possible to use the normal routines, which were deemed too time and resource consuming. Personnel at the accommodation sites understood that if they were to follow normal procedures, they would fail to meet the overall purpose of housing people in need. Instead they started to change the way they worked. Electronic registration forms were replaced with whiteboards, giving an overview of all the persons living there, their health status, their asylum process status and other important data. The formal way to procure materials and tools was too time consuming and had to be set aside. Instead managers themselves went to IKEA to pick up necessities like mattresses and kitchenware. These adaptations were crucial to fulfil the purpose of putting a roof over the head of every newly-arrived refugee. In all interviewed organisations, normal operations were abandoned in favour of new solutions, often based on the working experience and creativity of the personnel, or as Rasmussen (1985) states it, familiarity with the system's value structures. There is not only just one representation of how to operate, especially not when travelling towards the higher purpose in the abstraction hierarchy. The purpose could always be met in different ways.

3.3 *Focus on routines in public organisations*

Even though operative adaptations within Malmö Municipality and Swedish Migration Agency were necessary to handle the situation, these adaptations were not always facilitated by existing organisational structures or even condoned by management. Informants said that the abandonment of too rigid routines was to some extent contested by management and operative achievements were never fully acknowledged afterwards. Since the activities of operative personnel were not officially sanctioned, they were put in a sensitive situation in the case of negative outcomes.

Informants from the operative parts of public agencies said that during the whole event, no representatives of upper leadership came to the sites to form their own view of the situation. In the non-operative parts of these public organisations, there seems to have been a more wide-spread belief that the situation could be solved through ordinary processes, and this seems to have resulted in a lack of practical support for operations. It is known from safety research that when an outside observer attempts to describe the work of others, those descriptions tend to be more simplified and linear than actual work (Woods et al., 2012). In many organisations, decisions about procedures, plans and routines are made on a management level. If the gap between management and operations is too large, the assumptions guiding decisions about organisational structures may not reflect actual work needs, as seen in Malmö among the public actors.

3.4 *Internal organisational dynamics*

Judging from many examples described by informants, a key factor behind good internal collaboration and the ability to meet goals was a tight interface or likeness between those who detected and interpreted early signals of change and those who decided about actions. Signals of important changes are often sub-

tle and manifest in the course of operations, and if such information is not adhered to, the organisation's response may be delayed, which was the case for public agencies. Jernhusen on the other hand, where reactions were quick, generally depend on the input from their guards on the floor for decision-making. For Jernhusen surprises are normal for operations, which mean that reactions to these surprises have to be of the same flexible nature. In public organisations surprises are not desirable, which could be coupled to the demand on their processes for legal certainty. Paradoxically, in this case the focus on regular procedure combined with reluctance towards adaptations resulted in delays and problems in the process of fulfilling the basic needs of arriving refugees.

3.5 *Complementing NGO activities*

While public agencies were in some cases tied down by regulations, the results of this study showed that NGO's could sometimes fill the gaps created by slow official response. A large group of refugees arriving in Sweden were transit refugees, attempting to pass through Sweden on their way to Norway or Finland. In Swedish legislation, however, there is no such thing as a transit refugee. Any refugee arriving in Sweden has to seek asylum there, and the policies of the migration agency made it impossible for established NGO's to help these people. The Swedish Red Cross had a permanent barrack outside Malmö central station, on the square Posthusplatsen, which functioned as a welcome center and housed numerous different organisations. The organisations at Posthusplatsen had agreed to follow Swedish legislation, meaning that they would only help people who had the intention to seek asylum in Sweden. Instead, autonomous organisations stepped in to help the transit refugees. As one example, the cultural association Kontrapunkt arranged housing for 100–500 persons per night.

4 DISCUSSION

The analysis revealed a number of drivers for resilient performance within the studied organisations, such as different ways of relating to established rules, regulations, procedures and processes, the way relationships were formed between people and hierarchical layers within the organisations, and the perceived value of the human operator and of the human contribution to the organisational whole. These values, in turn, affect how the organisations shape many of the basic conditions that allow—or obstruct—resilient performance.

Different actors involved in the response to the 2015situation had different organisational formal and informal prerequisites that influenced their ability to adapt to the situation. This section deals with a number of such prerequisites such as tasks, roles, working environments, supporting tools and organisational structures.

4.1 *Heuristics for adaptation in organisational goals and values*

In the light of the present analysis, a likely challenge for public organisations will be to find ways of detecting when established rules and procedures are no longer appropriate and to find ways of adjusting them based on information from emerging events. In several of the organisations that were able to make successful adaptations, adjustments of normal procedures were made in relation to clearly defined and deeply rooted core goals. One way of approaching the issue of how to guide adaptations may be to look for such core goals or core functions which are central to the organisation and necessary for operations under any circumstances. Organisational goals may however have considerable room for interpretation. Because of that, it might be equally important to examine how organisations engage in dialogue around goals and values and who is allowed to participate when abstract goals are interpreted in terms of strategies and action.

4.2 *Adaptation supported by shared perceptions*

Examples of operative adaptations within the case could serve as an inspiration to more permanent and widespread design approaches within public organisations. Today, management typically has the final decision about artefacts such as plans, processes, procedures and work roles. For public actors a gap was observed between real work challenges and higher management's understanding of the circumstances and issues of operations. If organisations were to adapt supporting artefacts more consciously to operative needs, a first step may be to try to increase management knowledge about operative conditions, the real-life issues and difficulties faced by operators, so that solutions implemented by management do not run the risk of undermining operations. In terms of more profound changes to organisational practices, it could be rewarding to explore the implementation of design processes centred on system and user needs and to give employees a more active role in the constant evolution of artefacts. This could be an opportunity to bring the fields of safety science, organisational science and design science closer together.

4.3 *Lessons learned as re-design*

All of the interviewed organisations involved in the 2015 reception of refugees in the Øresund

region give examples of organisational lessons learned and adjustments that have been made with respect to these experiences. On the other hand, and for public organisations in particular, a large portion of learning and re-design seems to have been oriented towards formal organisational artefacts such as plans and procedures. For example, one of the main interventions has been to extend the practice of Risk and Vulnerability Analysis to the Swedish Migration Agency. These analyses are a national requirement for specific public agencies and are coordinated by the Swedish Civil Contingencies Agency (MSBFS 2016:7). As noted in the analysis, this could be interpreted as a reflection of the very focus on compliance and official doctrine that undermined adaptation during the crisis. From the perspective of systems oriented safety research, strict routines and procedures are in themselves no guarantees for safety and efficiency (Dekker, 2001). Yet another scenario is added to risk analyses, only to find that the next large disturbance has unexpected or unique qualities. The calibration of routines and procedures, or the addition of new documentation or similar symbolic barriers is a common response to negative events within organisations (Hollnagel, 2008). Here it must be acknowledged that procedures will never cover every possible scenario and may even limit the creative problem-solving abilities of professionals (Dekker, 2003). Every added barrier or new procedure will also increase system complexity, thus possibly increasing the demands on the people controlling the process, with potential negative effects on their performance (Praino & Sharit, 2016). On the other hand, procedures may provide stability and common ground, particularly in crisis response that involves a diverse set of actors. Just as with any organisational artefact, procedures can be consciously designed, informed by and adapted to their users. As noted above, it may also be possible to explore different ways for organisations to assess when normal procedures are not enough and adaptation is needed.

Another possible interpretation is that the problem does not lie so much in the procedures themselves, but rather in organisational perceptions and values around management and compliance. The process of gathering experiences and turning them into improvements should never focus only on administrative outputs. Firstly, future research within the public sector could explore the implementation of methods that allow for systems-oriented analyses of events and activities. This could give governmental organisations a more complete picture of past events and a better understanding of all the different pre-conditions that must exist to support operations. Secondly, event analysis could be tied more tightly to a general design process so that lessons learned are used to produce solutions that are suited to the needs of potential end users. This approach could counter the risk of adding to their administrative work burden.

4.4 *The role of employee experience and overarching organisational values*

Much of the above discussion centres on design issues, but the extent to which the studied organisations were able to perform resiliently also seems to depend on the way relationships were formed between people and hierarchical layers within the organisation, or on the perceived value of the human operator in the organisational whole. It is important to acknowledge that processes and tools do not in themselves guarantee good outcomes. Even though work-supporting artefacts from processes to decision-aids may have all the right attributes, an organisation can still prove too rigid and slow to adapt to the fast pace of real-world operations. For example, if an organisation is to reap the benefits of use-centred design, it also has to have a fundamental respect for the experience and practical knowledge of operative personnel. It is hard to imagine that the individual operator envisioned in resilience research -knowledgeable, creative and full of initiative—would be likely to exist within an organisation that does not have a fundamental appreciation of its employees, acknowledging the signals and interpretations emerging from operations. Furthermore, the existence of good adaptive designs that are achieved repeatedly presupposes a continuous dialogue around working conditions and developments within the organisation, so that designs can build on a good representation of reality. This concept calls into question the measure-centric, hierarchical paradigm of management and control found in many of today's organisations and raises the question of how future management could be steered towards more inclusive and systems-oriented principles. Purposeful designs cannot be reached without applying a systems perspective, making sure that solutions support not only individual actors within the system, but also activities that are distributed among people, mediated by technology, within an organisational context.

5 CONCLUSION

The 2015 increase of refugees arriving in Malmö meant a great challenge for the organisations involved in the response, and interviews within this study revealed a number of examples of resilient performance. One of the main findings of this study has been that resilience does not primarily

reside in simple organisational features, functions or resources, i.e. simple boxes to tick in an organisation's management system. Rather, some of the most important resilient behaviours in the observed case had to do with adaptations—trade-offs and judgments made by professionals under the sometimes harsh conditions of real-world operations.

In terms of resilient performance, although many examples of this emerged during the reception of refugees, it also became clear that disconnections between management and operative personnel may hamper adaptability and lead to designs of organisational structures that do not fully answer the needs of their users. For these reasons, it is suggested that future research investigates different ways of guiding adaptations e.g. from a basis of core organisational goals and values, and of creating joint perceptions between management and personnel.

Practices such as employee inclusion, increased local autonomy, user-centred design and systems-oriented organisational learning may require a new set of values within an organisation. These values include an understanding of the organisation as a socio-technical system and a fundamental respect for human experience, initiative, collaboration and problem-solving. While studies such as this one can provide positive examples of resilient performance, deeper changes may require a paradigm shift for both public and private organisations that in itself will require further research and interventions.

ACKNOWLEDGEMENTS

This study is part of the EU Horizon 2020 project IMPROVER (grant agreement no. 653390, www.improverproject.eu), where the Øresund region is included as a living lab.

REFERENCES

Alvesson, M. (2011). *Intervjuer: genomförande, tolkning och reflexivitet*. Liber.

Bergström, J., & Dekker, S.W.A. (2014). Bridging the Macro and the Micro by Considering the Meso : Reflections on the Fractal Nature of Resilience. *Ecology and Society*, *19*(4).

Bowen, G.A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40.

Cook, R.I. (1998). *How Complex Systems Fail*. Chicago IL.

Dekker, S. (2003). Failure to adapt or adaptations that fail: contrasting models on procedures and safety. *Applied Ergonomics*, *34*(3), 233–238.

Dekker, S.W.A. (2001). Follow the procedure or survive. *Human Factors and Aerospace Safety*, *1*(4), 381–385.

Hollnagel, E. (2008). Risk + barriers = safety? *Safety Science*, *46*(2), 221–229.

MSBFS 2016:7. (2016). MSBFS 2016:7 föreskrifter och allmänna råd om statliga myndigheters risk- och sårbaretsanalyser. Myndigheten för samhällsskydd och beredskap.

Praino, G., & Sharit, J. (2016). Written work procedures: Identifying and understanding their risks and a proposed framework for modeling procedure risk. *Safety Science*, *82*, 382–392.

Rasmussen, J. (1985). The role of hierarchical knowledge representation in decision making and system management. {*IEEE*} *Transactions on Systems, Man, and Cybernetics*, *SMC-15*(2), 234–243.

RiR 2017:4. (2017). Lärdomar av flyktingsituationen hösten 2015 - beredskap och hantering. Riksrevisionen.

Salmon, P.M., Goode, N., Archer, F., Spencer, C., McArdle, D., & McClure, R.J. (2014). A systems approach to examining disaster response: Using Accimap to describe the factors influencing bushfire response. *Safety Science*, *70*, 114–122.

SOU 2017:12. (2017). Att ta emot människor på flykt Sverige hösten 2015. Statens Offentliga Utredningar.

Woods, D.D. (2015). Four concepts for resilience and the implications for the future of resilience engineering. *Reliability Engineering and System Safety*, *141*, 5–9.

Woods, D.D., Leveson, N., & Hollnagel, E. (2012). *Resilience Engineering: Concepts and Precepts*. Ashgate Publishing, Ltd.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Improving resilience management for critical infrastructures—strategies and practices across air traffic management and healthcare

V. Cedrini & M. Mancini
ENAV S.p.A, Air Navigation Service Provider (ANSP), Rome, Italy

L. Rosi, G. Mandarino & S. Giorgi
Italy National Health Institute (ISS), Rome, Italy

I. Herrera & M. Branlat
SINTEF, Trondheim, Norway

J. Pettersson & C.-O. Jonson
Center for Disaster Medicine and Traumatology, and Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

L. Save & D. Ruscio
DeepBlue SrL, Rome, Italy

ABSTRACT: Recent natural and man-made disasters highlight that a more resilient approach to preparing for and dealing with such events is needed. To address this challenge, the main objective of the research and innovation H2020 project DARWIN is the development of European resilience management guidelines for Critical Infrastructures (CI). Based on a systematic literature survey with a world-wide scope and prioritization of resilience concepts, the guidelines have been developed taking into account everyday operations, contingency plans, training, etc. This paper describes insights gained from the adaptation of these guidelines in the domains of Air Traffic Management (ATM) and Healthcare (HC). A collaborative and iterative process has been defined involving relevant experts and practitioners. To ensure transnational, cross-sector applicability and uptake, a Community of Crisis and Resilience Practitioners (DARWIN DCoP) has been involved. The preliminary results indicate that a big step has been taken in moving from the resilience theory to practice.

1 INTRODUCTION

ATM and HC have a great track record of safe operations in challenging conditions, even if disruptions or occasional crises may happen routinely. While it can certainly be improved, both domains have already implemented a number of practices and methods, especially related to being able to handle such disruptions or to learning from them. Still, recent examples from disasters are reminders of the urgent need to improve our ability to reveal, assess and manage resilience, both in everyday operations and during crises (Hollnagel et al., 2011, Adini et al, 2017).

The overall objective and main result of the Horizon 2020 EC project DARWIN is the development of European resilience management guidelines. These guidelines are called DARWIN Resilience Management Guidelines (DRMG).

The DRMG consist of suggested interventions and guiding principles to help or advise any organization in the creation, assessment or improvement of its own reference guidelines, procedures and practices.

What is really important is that DARWIN results are useful for our end users namely the Critical Infrastructures that include ATM and HC.

For this purpose, the DARWIN Resilience management guidelines are designed to address disruptions, changes and opportunities; facilitate anticipation, adaptation, flexibility; and provide a foundation for an effective crisis response (Adini et al., 2017).

An initial set of generic DRMG was produced (DARWIN D2.1, 2016) and then adapted to ATM and HC to make the guidelines more operational and usable in these domains.

This paper presents the approach and methodology carried out to adapt the DRMG to both domains and discusses relevant results.

1.1 *Nature of the DARWIN guidelines*

The methodology to obtain the list of DRMG has been thoroughly defined: based on a world-wide systematic literature review carried out for the DARWIN project (DARWIN D1.1), 56 concepts, approaches and practices have been identified and evaluated (DARWIN D1.2).

The results of an evaluation following a modified Delphi process with practitioners and experts resulted in essential and important resilience concepts to be included in the resilience management guidelines. These conceptual as well as user requirements are input for the development of the DRMG (DARWIN D1.3).

The guidelines are developed as individual topics that address the conceptual requirements identified. Those topics are referred to as Concept Cards (CC). CCs propose interventions that organizations can implement (the *how*) to reach the resilience management capabilities captured in the conceptual requirements (the *what*). Through those interventions, the guidelines aim to help CI organizations in developing a critical view of their own crisis management activities (management of resources, procedures, training, etc.). The CC are structured in content blocks that contain information such as: purpose; interventions proposed; actors in charge; illustration; associated practices, methods and tools; etc. In addition, while they address specific aspects of resilience management, CCs are not independent and links between them are captured through various means.

DARWIN CCs, and in particular adapted CCs, could be complementary to guidelines, procedures and practices already present in the organizations of the two domains, fostering their revision, improvement or even creation of new guidelines.

Also, each CC includes a Minimum Viable Product (MVP) which is the smallest way to start using the interventions proposed in the CC. The MVP is the set of minimum set of features required to test or experiment a solution. Its purpose is to get through the "build-measure-learn" feedback cycle as quickly and efficiently as possible (Ries, 2011). The DARWIN project proposed this solution based on interactions with experts (managers and front-line operations). This approach contrasts the traditional product development of designing, performing preliminary and critical reviews, producing and testing and perfecting the product.

1.2 *Content of the DARWIN guidelines*

The DARWIN CCs are organized under the following themes:

Supporting coordination and synchronization of distributed operations

1. Promoting common ground in cross-organizational collaboration
2. Establishing networks for promoting inter-organizational collaboration
3. Ensuring that actors involved in resilience management have a clear understanding of their responsibilities and the responsibilities of other involved actors

Managing adaptive capacity

4. Enhancing the capacity to adapt to both expected and unexpected situations
5. Establishing the capacity for adapting during crises and other events that challenge normal plans and procedures

Assessing resilience

6. Identifying sources of resilience
7. Noticing brittleness
8. Assessing community resilience to understand and develop its capacity to manage crises

Developing and revising procedures and checklists

9. Managing policies involving systematically—policy makers and operational personnel for dealing with emergencies and disruptions

Involving the public in resilience management

10. Interacting with the public not yet affected by or involved in a crisis

2 METHODOLOGY

The established methodology is a systematic step by step approach strictly intertwined with the other DARWN activities. These include in particular those relevant to the development of generic guidelines, to their evaluation and to interaction with the DCoP.

The adaptation process consists of two main steps:

- **Step 1: Selection of adaptable CCs**, i.e. the assessment for the adaptability of the generic CCs
- **Step 2: Adaptation of adaptable CCs**, that is the adaptation of the generic CCs to ATM and HC domains, and the release of the adapted guidelines.

2.1 *Selection of adaptable CCs*

This phase has been performed by applying a methodology based on a quantitative and qualitative SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis that assessed if a CC was adaptable or not.

The SWOT analysis methodology is commonly used to develop a deep understanding of all kinds of situations in business, organizations, and for individuals, to support the decision-making process.

It is noteworthy that the SWOT analysis findings address relevant actions for Guidelines developers concerning the improvement of the generic guideline content also. In particular, during the development of DRMG, the CCs were simultaneously assessed with regard to a possible adaptation to the specific domains, possibly avoiding any gaps between development and later adaptation.

At the end of the adaptability assessment two lists are expected: the list of non-adaptable CCs including the rationale behind their non-adaptability or elements that can be improved, and the list of adaptable CCs including the rationale behind their adaptability.

The applied SWOT analysis has been defined combining the quantitative and qualitative approaches for a richer collection of data.

The quantitative SWOT analysis has been based on the definition of a set of Indicators (*I*) that were identified starting from the fields of the CC used for the process of adaptation.

Seventeen indicators (Table 2) have been established, each of them formulated as a specific statement and categorized according to the four areas of the SWOT:

- the Strength/Weakness (S/W) areas include indicators concerning internal aspects of the CC (i.e. specific contents of the CC fields).
- the Opportunity/Threat (O/T) areas include indicators whose assessment needs to take into account a more long-term perspective and the interdependency with external factors linked to the contexts of the CC application.

Table 1. Step 1 overview.

Input	Last available version of Generic CCs
Output	• List of Adaptable and non-Adaptable CCs • Information concerning content for CCs adaptation (from qualitative SWOT • Information concerning elements of the Generic CC improvement
Effort Required	• 1-day per each interview with each ATM/HC expert concerning each single CC SWOT • 3–4 days (per CC) to organize relevant information and perform additional research • 1 day to review the results with involved expert

Experts' opinions on the indicators have subsequently been collected through seventeen questions formulated as follows "*How much do you agree with the following statement (I_01, I_02, … I_17)?"*

The answers to each question were recorded using a 5-point Likert scale ranging from "Disagree" to "Very Strongly Agree", with "Somewhat agree" in the middle.

In order to obtain a quantitative figure for each indicator's assessment, a numeric value was assigned to each level of the scale, starting from 1 (="Disagree") to 5 (="Very Strongly Agree") and incrementing by one per level.

Thus, according to the *I* mean value, the assessment of each I has been classified according to the criteria described below.

The qualitative SWOT analysis of each generic CC was carried out by collecting comments and feedback from experts during the assessment of the quantitative SWOT analysis indicators.

After the quantitative assessment of each indicator, the expert was asked to explain the rationale of the scoring, indicator by indicator, while the interviewer was taking notes.

The interview started with the narration of the illustrative case or lesson learnt that, according to the expert, better supports the discussion on the contents of the specific CC applied to the domain.

The rationales provided during the interviews have been collected and grouped into four areas of the SWOT on the basis of the mean values calculated for each indicator (ref. Table 2).

It is noteworthy that also the rationales fully contrasting the average evaluation for the specific indicator have been kept and taken into account for the sake of richness of data.

In addition to the CCs adaptability assessment, the qualitative SWOT analysis results have been considered as one of the main sources of information used to adapt the CCs' content to the specific domain. Moreover, the collected information has been enriched using sources of information available online.

Some criteria were established to evaluate the adaptability of each CC to the specific domain. They were based on two mean scores of the SWOT analysis results:

- I-02 mean score—This indicator directly refers to the applicability of the CC to the local context in which the card will be used. The applicability is the condition *sine qua non* the CC can be used in real ATM/HC environment.
- Total mean score of all CC indicators—This value provides a synthetic measure of the "adequacy"

Table 2. List of SWOT indicators

Nr.	Statement
1	The CC overlaps with other CCs **SWOT category: S/W**
2	The CC is applicable to local ATM/HC contexts (where the card will be used) **SWOT category: O/T**
3	The CC can be complementary to local ATM/HC arte-facts (i.e. procedures, regulations) **SWOT category: O/T**
4	Actors, as described in the CC, are identifiable in the ATM/HC domain **SWOT category: S/W**
5	The roles and responsibilities of the actors, as described in the CC, are clear in the ATM/HC domain **SWOT category: S/W**
6	It is possible to identify actors, roles and responsibilities, as described in the CC, in case of sudden changes in the ATM/HC domain (i.e. regulatory bodies, etc.) **SWOT category: O/T**
7	It is possible to identify actors, roles and responsibilities, as described in the CC, in case of future changes in the ATM/HC domain (i.e. regulatory bodies, etc.) **SWOT category: O/T**
8	The implementation **before**, as developed in the CC, is rel-evant for the ATM/HC domain and adaptable **SWOT category: S/W**
9	The implementation **during**, as developed in the CC, is relevant for the ATM/HC domain and adaptable **SWOT category: S/W**
10	The implementation **after**, as developed in the CC, is rele-vant for the ATM/HC domain and adaptable **SWOT category: S/W**
11	Internal factors of the ATM/HC domain, facilitating or hindering the implementation of the contents of the CC, can be easily identified and explained **SWOT category: O/T**
12	External factors (cultural, social, economic environment), facilitating or hindering the implementation of the contents of the CC, can be easily identified and explained **SWOT category: O/T**
13	Expected results, that can be inferred from the CC, can be identified and explained within the ATM/HC domain **SWOT category: S/W**
14	Illustrative cases and/or lessons learnt, linked to the con-tents of the CC, are available in ATM/HC domain **SWOT category: S/W**
15	Practices, linked to the contents of the CC, are available in ATM/HC domain **SWOT category: S/W**
16	Methods, linked to the contents of the CC, are available in ATM/HC domain **SWOT category: S/W**
17	Tools, linked to the contents of the CC, are available in ATM/HC domain **SWOT category: S/W**

and maturity of the CC fields for the adaptation purposes.

The combination of these two mean scores—as explained in Figure 1 - defines the adaptability of each CC and specific issues to be addressed by CC developers.

Figure 1, Table 4 show the criteria applied to establish if a CC is adaptable or not, and the actions identified to handle the issue with guide-line developers.

2.2 *Adaptation of adaptable CCs*

Once a CC has been evaluated as adaptable, the second step of the adaptation process begins. The Adapted CCs have been developed by integrating several sources:

- The findings of the qualitative SWOT analysis performed in Step 1;
- The information collected during *ad-hoc* inter-views with domain specific experts;

Table 3. Criteria for the classification of SWOT results.

Indicator mean score	Classification of the Indicator I
$I > 3$	*I* classified as Strength or Opportunity The indicator is helpful to the CC adaptation to the ATM/HC domain
$I < 3$	*I* classified as Weakness or Threat The indicator is "harmful" to the CC adaptation to the ATM/HC domain
$I = 3$	Those Indicators whose mean value was =3 have been classified by taking into account the experts' comments collected by the qualitative SWOT analysis: • the Indicator has been classified as Strength or Opportunity if the majority of the comments mainly emphasized positive elements; • the Indicator has been classified as Weakness or Threat if the majority of comments highlighted lacks and missingpoints.

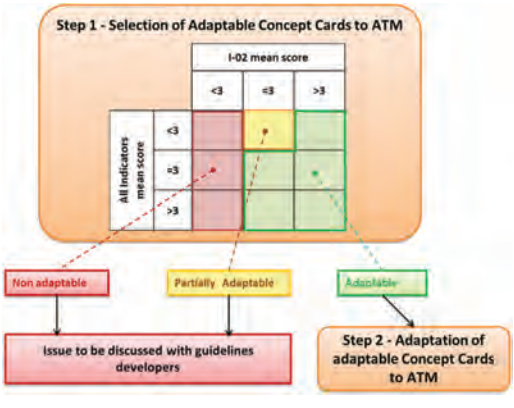


Figure 1. Adaptability Criteria.

– The information provided by the domain specific experts involved in the "initial evaluation of guidelines";
– The feedback provided by the DCoP during the workshop;
– The results collected during the implementation of Pilot exercises.

A strategic selection of participants from management as well as front line operators was performed. A template was prepared to follow a semi-structured interview with experts. The interview started with a narration of an illustrative case to better support the discussion of the context of the CC.

The topics covered concern actors involved, actions prior to, during and after the crisis. The template also includes context information, practices, methods and tools as well as other illustrative cases. The relevance of the content proposed by the CC, in particular concerning the interventions, was discussed.

Table 4. Rationale for CC classification (example).

CC Classific.	"Non-adaptable" or "Partially adaptable"
Rationale	• the concept is valid at a general level but difficult to adapt. It is not applicable in the local ATM/HC domain (i.e. due to type of organization and current policies of the local ATM/HC systems); or • the CC is not adequately developed to be adapted; or • so far, it has been particularly difficult to find specific ATM/HC content for the majority of the fields.
Action	Major amendments are needed and the issue has to be discussed with guideline developers
CC Classific.	**"Adaptable"**
Rationale	• the CC is applicable in the local ATM/HC domain; or • most of fields of the CC are adequately developed. However, in some cases some effort could be needed to make adjustments or to find specific ATM/HC content for some fields.
Action	The CC can be adapted and some issues to be discussed with guidelines developers

Table 5. Step 2 overview.

Input	• Last available version of Generic CCs • Output from Step 1: – List of Adaptable CCs – Information concerning content for adaptation of CCs (from qualitative SWOT)
Output	DRMG/Adapted CCs to ATM/HC
Effort Required	• 1 day (per CC) to interview one/two ATM/HC expert/s; • 5 days (per CC) to integrate Wiki with relevant information coming from SWOT, expert interviews, DCoP feedback, CC evaluations, feedback from pilots, additional research on internet.

3 RESULTS ON THE SELECTION OF ADAPTABLE CC

The adaptability assessment of the CCs has been gradually carried out during the development process of the generic CCs.

The SWOT analysis started as soon as the generic guidelines developers team considered the CC mature enough to be released and assessed for adaptation.

The information collected during the SWOT has been enriched using sources of information, suggested by the interviewee, available online and integrated in the DARWIN Wiki.

Overall review of expert feedback has provided guidance in terms of quality of the adaptable guidelines. It has been taken into consideration when updating the adaptable CC as well when elaborating new CCs.

Elements that have been mostly appreciated among experts are:

- The concepts developed by the CCs are relevant to the ATM and HC context,
- Actors in ATM and HC context are identifiable in a clear and concise manner; also roles and responsibilities are clear, being hierarchy well defined in ATM and HC,
- The list of actions/interventions give sufficient explanation of responsibilities making it easier to adapt to the ATM, and to HC, while taking into account the broader and different fields of HC,
- The triggering questions are useful and well grouped,
- The indications provided in the fields "implementation before/during/after" are sufficient to develop a CC adapted to the ATM and HC context,
- The level of provided information makes it easier to be integrated with local artefacts (procedures, plans),
- Useful examples, illustrative cases, practices and methods are available in the ATM and HC context.

Elements that need improvement:

- In some CCs, the information is very high level or too generic thus making it difficult to adapt,
- Some content concerning the "Triggering Questions" and "Actions" is redundant and needs to be simplified,
- No tools are provided in some of the current version of the CCs, thus, during the adaptation process, efforts should be spent, accordingly,
- Harmonization still needs to be reached among some CCs.

4 RESULTS FROM ADAPTATION OF CC

At first sight, ATM and HC seem to be very different contexts, but during meetings, the DARWIN team has discovered that they share many similarities and many common issues (i.e. criticalities of the infrastructures, impact on the public, etc.).

Notwithstanding that, the aviation domain in general is characterized by high level of standardization. The number of standards and regulations guarantee that ATM has a great track record of safe operations.

Regulatory bodies and concerned actors are well defined together with roles and responsibilities.

For example, the geographical limitation of an aerodrome makes this type of environment exposed to a relatively limited number of crisis types (e.g., aircraft accident during take-off or landing, disaster in the premises, loss of working resources, climatic event, etc.).

Although it is impossible to know when the crisis will occur, the characteristics and dynamics of crisis situations can be foreseen in advance to some degree. As a consequence, the concerned actors and the response procedures can be defined with sufficient accuracy before the crisis occurs.

On the other hand, other types of crises in ATM may be much more extended from a geographical point of view and less predictable in the way they evolve (as in the example of the Eyjafjallajökull volcano eruption in 2010).

As previously suggested, the HC domain exhibits common aspects with ATM that deals with criticalities and brittleness, and, in addition, they share the same scientific basis on which public health and HC tasks (i.e. care, surveillance, research, regulation and control) that are the same bases/criteria of Safety and Quality Assurance present in ATM.

What is, however, peculiar to HC is the individual/team resilience that HC workers (professional and operators) practice daily while performing their task and while coping with unexpected situations. In this case, the management of this resilient approach, i.e. systematically creating the conditions to bridge the gap between *work-as-imagined* and *work-as-done* (WAI vs. WAD), proves to be challenging.

Other relevant aspects include that this domain shows more complexity, and a variety of tasks, with many actors, ranging from surgeons to nurses, from regulatory bodies, providers, training organizations. These lead to an *variety* of systems, processes and outputs (protocols, documents, or records that could differ from hospital to hospital.

HC moves to innovation, however it should be recalled how professionals frequently support clinical decisions over standardization (a clinician sometimes stands on autonomous judgement provided for it is based on *knowledge and belief*).

One of the noteworthy outcomes of the adaptation process is that we discovered more uses than we expected at the beginning of the project. We found out that the guidelines are useful, they can

be adapted and adopted in many occasions such as training, workshops and meetings.

They help to start discussions and to deal with significant topics and they can be used to:

1. Check or update current procedures and guidelines, if already existing;
2. Define new procedures and guidelines if not existing;
3. Identify possible indicators and evaluation of trends (to do possible benchmarking);
4. Prepare plans;
5. Perform risk assessment and management.

During the interviews with ATM and HC experts concerning the CCs, some common aspects that play an important role in the resilient management of crisis emerged:

– **The CCs should concern all levels of organization**

Even if, at first sight, the DARWIN CCs may address only policy makers and management, being responsible for the modification of current procedures, it is noteworthy that all concepts address all levels of organization starting from senior management to front line operators.

– **The Roles and responsibilities of involved actors change according to the type of crisis and the related environment of operations**

In the ATM context, according to the type of crisis several actors are involved. According to 'ICAO Annex 14. Emergency and other services', An Airport Emergency Plan shall be established to coordinate the response and participation of all existing agencies which could assist in responding to an emergency.

Examples of possible agencies ON and OFF aerodrome are provided:

ON-aerodrome: air traffic control unit, rescue and firefighting services, aerodrome administration, medical and ambulance services, aircraft operators, security services, and police;

OFF-aerodrome: fire departments, police, health authorities (including medical, ambulance, hospital and public health services), military, and harbour patrol or coast guard.

– **The establishment of Just Culture and Safety Culture in all organizations**

With particular reference to the concept of "noticing brittleness", Just Culture and Safety Culture are the internal factors that could help in facilitating the identification of brittleness in each organization.

The concept of 'Just culture' is discussed in EUROCONTROL (2006) "*in recent years the concept of "Just culture" has become better understood and accepted by people employed in the aviation industry. However […] the need for a "just culture" is generally not understood by many legislators and therefore not accepted within their State judicial systems.*"

This issue causes "*increased fear of sanctions against the reporter, particularly if partly or fully responsible for the reported occurrence*."

"Furthermore, certain elements of the media may deal aggressively with apparent breaches of flight safety within certain airlines and ANSPs."

"These factors—punishing Air Traffic Controllers or pilots with fines or license suspension—may have the cumulative effect of reducing the level of incident reporting and the sharing of safety information. This hinders safety improvement and as a cascading effect resilience."

There could be concerns about possible misuse of information regarding brittleness in the organizations, since "*one of the major problems with collecting and analysing information is that such information can be a very powerful tool and, like any powerful tool, if used properly it will provide great benefit. However, it can also be used improperly and if that occurs considerable harm can be caused".*

In the last decade, many progresses have been made to encourage Just Culture in the European ATM context, mainly thanks to the efforts of EUROCONTROL: e.g. Air Navigation Service Providers are endorsing Just Culture policies and programmes, Task Forces have been created to promote, debate and discuss issues concerning safety and justice, meetings are organized to encourage interaction between safety and the judicial experts; special "just culture" courses for aviation experts and prosecutors have been organized, etc.

According to EUROCONTROL (2008), Safety Culture is "*the way safety is perceived, valued and prioritised in an organization. It reflects the real commitment to safety at all levels in the organization. […] It is not something you get or buy; it is something an organisation has. […] It can therefore be positive, negative or neutral.*"

Since 2006, there is an active involvement of EUROCONTROL, in collaboration with FAA and CANSO, in measuring and improving Safety Culture within ANSP organizations. Safety Culture surveys are continuously planned and performed, results and recommendations are taken into account and implemented to guarantee an effective SMS and a healthy Safety Culture.

– **The importance of planning, training and testing in advance**

The plan should include a clear definition of the agencies involved, the responsibility and role of each agency and the coordinates of offices/people to be contacted in case of emergency.

The **training** of the people allows to maintain the high level of preparedness for possible crisis events.

The **test** of the plan could be done in many different ways beginning with the organization of exercises, from a lower level to a higher level, with each one building on the concepts of the previous: discussion-based and operations-based exercises. The execution of the exercises allows to identify weaknesses in the plans and possibly improve them.

Discussion-based exercises are organized to discuss the plans for upcoming operations-based exercises, and to make everyone familiar with roles, procedures and responsibilities. They include: seminars, workshops, tabletop exercises, and games.

Operations-based exercises are used to validate and test plans and procedures that have been consolidated after the discussion-based exercises. They allow to better clarify roles and responsibilities of involved actors, identify gaps and limitations of the plan, and improve everyone's performance. They include drills, functional and full-scale exercises.

– **The importance of lesson learned dissemination**

The importance of the dissemination of the relevant information after the crisis events is fundamental in order to improve the resilience of the organization during crisis. In the ATM context and for this particular purpose, EUROCONTROL encourages the lesson learnt distribution and exchange of best practices though the website Skybrary. As well, the magazine Hindsight contains lot of useful case studies and provides the Air Traffic Controllers (ATCo) with a means to share their experiences concerning ATM-related safety occurrences. The objective is to "*broaden ATCOs understanding of the problems that may be encountered, learn more about possible solutions and be better prepared in the face of similar occurrences*."

Moreover, the presence of the "triggering questions" was particularly appreciated even if it may be difficult to use them during time-critical types of crisis as a checklist to be read step-by-step and to identify someone that checks their completion. On the other hand, it is important that all the actors involved in the management of the crisis are fully aware of the topics addressed.

For crises developing over a longer time (e.g. Icelandic volcano eruption or Ebola outbreak) it is possible to organize workshops and meetings to reflect with other colleagues on the possible sources of brittleness and use the triggering questions to support the reflection. The same approach can be used during a drill or a simulation by a facilitator to guide the simulation and stimulate participants to notice brittleness.

5 DISCUSSION AND CONCLUSIONS

The work performed so far confirms the intended readership as policy makers, front line operators, resilience engineering managers, crisis managers, critical infrastructures managers, methodologists, community of practice in ATM and HC. Stakeholders such as managers and policy makers can use this work as source of inspiration when adapting resilience guidelines to their domains.

In particular, applying the DARWIN resilience concepts, triggering questions, methods and tools, they will be able to:

- Apply the proposed interventions provided in the CCs to survey current practices, strategies, procedures and guidelines;
- Start to reflect on "what went well" and not only "what went wrong" when learning from events;
- Assess the effectiveness of roles and responsibilities during a crisis;
- Revise and/or define common action plans through periodical coordination activities and training;
- Identify brittleness in the system and the application of procedures and response to the crisis;
- Get to know practices, methods and tools applied by others;
- Test and improve their plan of communication with public during emergencies.

The collaborative method presented in this paper illustrates an iterative approach that brings theoretical concepts close to their practical implementation. We gathered information from other domains through workshops with members of the DCoP. The SWOT facilitated translation of resilience concepts into practical interventions. The methodologies proposed to adapt the concepts are defined in detail to ensure other concepts to be included in the future. The participation of experts is essential to ensure applicability, relate to the specific domain as well as enrich the cards with existing practices and methods.

At the beginning of the work, we planned separate guidelines for ATM and HC. The first results were cards that replicated the generic cards. This overlap in relevance and adaptation in CCs outcome in ATM and HC indicate the potential of the generic CCs to be applicable to other sectors. The current result combines generic fields with adaptations to HC and ATM as required. The results indicate the possibilities of similar adaptations to other domains.

A challenge is the achievement of consensus on the review process and iteration to achieve sufficient maturity. Another challenge as well as an opportunity is to merge different cultural perspectives across Europe when dealing with crises. We

found this as a window of opportunity to learn mapping recommended practices and methods within and across domains.

Further work includes evaluation of DRMG and associated CCs in relevant operational scenarios. We consider collecting feedback from ATM, HC as well as other domains from the DCoP.

ACKNOWLEDGEMENTS

The research leading to these results received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 653289. Opinions expressed in this publication reflect only the authors' views and that the Agency is not responsible for any use that may be made of the information it contains. The authors would like to thank the group of professionals—both researchers and practitioners from the DARWIN project as well to experts contributing to the work related to the adaptation and paper review including other members from CARR, ENAV, FOI, ISS, KMC and SINTEF.

REFERENCES

Adini, B., Cohen, O., Eide, A.W., Nilsson, S., Aharonson-Daniel, L., Herrera, I., (2017) Striving to be resilient: What concepts, approaches and practices should be incorporated in resilience management guidelines? *Journal Technological Forecasting & Social Change.* Vol 121, pp. 39–49.

DARWIN (2015). Deliverable D1.1. Consolidation of resilience concepts and practices for crisis management. Available at: http://www.h2020darwin.eu/project-deliverables.

DARWIN (2016). Deliverable D1.3. Practitioner and academic requirements for resilience management guidelines. Available at: http://h2020darwin.eu/project-deliverables/.

DARWIN (2016). Deliverable D2.1. Generic Resilience Management Guidelines. Available at: http://www.h2020darwin.eu/project-deliverables.

DARWIN (2017). Deliverable D2.2. Generic Resilience Management Guidelines Adapted to Healthcare Domain. By the end of 2017, available at: http://www.h2020darwin.eu/project-deliverables.

DARWIN (2017). Deliverable D2.3. Generic Resilience Management Guidelines Adapted to Air Traffic Management Domain. By the end of 2017, available at: http://www.h2020darwin.eu/project-deliverables.

EUROCONTROL ESARR advisory material/guidance document (EAM/GUI) – EAM 2/GUI 6 "Establishment of 'just culture' principles in ATM safety data reporting and assessment".

EUROCONTROL (2008) "Safety Culture in ATM—An overview".

EUROCONTROL Hindsight Magazine [https://www.eurocontrol.int/content/hindsight].

Hollnagel, E., Pariès, J., Woods, D.D., Wreathall, J., 2011. Resilience Engineering in Practice: A Guidebook. Farnham, (2011). Resilience Engineering Perspectives volume 3. Ashgate Publishing, UK: Ashgate., UK.

ICAO Annex 14 "Emergency and other services"

Reis, E. (2011). The Lean Start-up. How today's entrepreneurs use continuous innovation to create radically successful businesses.

Risk assessment

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

PSA modeling method for a safety critical DI&C system

Sung Min Shin & Jaehyun Cho
Korea Atomic Energy Research Institute, Yuseong-gu, Daejeon, Republic of Korea

ABSTRACT: I&C systems in NPPs are being digitalized by adopting new features, such as software, fault-tolerant techniques, and network communication. Although the risk caused by these new features should be analyzed in an appropriate framework, at present there is no consensus on PSA methods for them. In this study, a general frame of a PSA model for the automatic safety signal generation function in a DI&C system is proposed, in consideration of the representative safety features of this system and the linkage between them. Through the related literature, we identified the requirements to construct the DI&C PSA model, constructed a general frame reflecting its possible parts, and specified the assumptions and approaches applied in this process. Although this study has focused on a qualitative approach because an appropriate database cannot be obtained yet, important failure modes that are understood in this current phase, and the research topics that need to be considered for the development of the enhanced DI&C PSA model, are summarized.

1 INTRODUCTION

Numerous analog Instrumentation and Control (I&C) systems in Nuclear Power Plant (NPP) are now being replaced by digital systems owing to the obsolescence of safety-critical analog components. This shift entails the adoption of new features that did not exist in analog systems, such as software, a Fault Tolerant Technique (FTT), and network communication. Although these features are expected to contribute to the enhancement of both efficiency and economy, from a safety point of view, the risk caused by the new features should be analyzed in an appropriate framework to ensure the dependability of the entire NPP (Kang, 2009, Authen, 2012). In recent years, regulatory bodies in each country have been actually demanding that the reliability of DI&C systems be incorporated in the PSA model. In this regard, some studies have been conducted in relation to software reliability, which are the representative feature of DI&C system; however, a more comprehensive approach seems necessary, such as how to integrate it with other digital features and apply it to the actual NPP PSA model.

Therefore, in this study, an approach of the DI&C PSA model is suggested in consideration of the representative safety features of the DI&C system and the linkage between them. In this preliminary phase for the DI&C PSA, it is based on a qualitative approach, without considering the available database, and we focused on the reliability model related to the automatic signal generation part, excluding the part related to human behavior among the functions of the DI&C system.

2 AUTOMATIC SAFETY SIGNAL GENERATION IN DI&C SYSTEM

Although a more appropriate modeling method for the DI&C system can be developed during the research progress, in this phase, a conventional fault tree (FT) format is considered for the DI&C system reliability model in terms of the convenience of integration with the plant model.

From the risk perspective, the modeling factors caused by the introduction of the DI&C system can be roughly expressed through Figure 1.

- DI&C induced initial event (spurious operation of DI&C system)
- Hardware and software failure (failure mechanism)
- Fault tolerant failure (fail-safe mechanism)
- Human error in digital environment

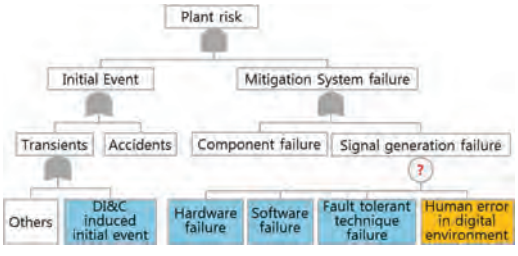


Figure 1. DI&C modeling factors from risk perspective.

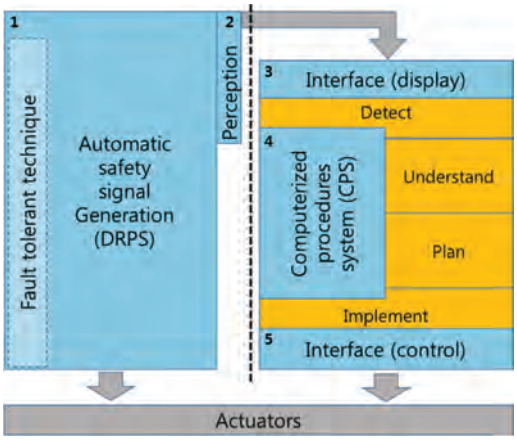


Figure 2. Functions implemented for safety signal generation.

Among the four categories, the initial DI&C induced event is not considered in this first phase, as this factor can have different effects on the system and is possible to be considered separately.

The remaining factors, in combination, implement certain functions for safety signal generation, as expressed in Figure 2. The functions in orange indicate human behavior. As can be seen from this figure, there are two methods for safety signal generation. One is a fully automated one in which the fault tolerant technique is applied. In another one, the safety signal is generated by a human operator using the information transmitted. This method also utilizes digitalized features, such as a display, CPS, and control, helping the decision making of the human operator. To develop a reliability model of this, information of the hardware and software failure in the functions corresponding to 2 through 5 (numbers in Figure 2), and the Human Error Probability (HEP) in this digitalized environment, should be analyzed.

Because the reliability model of the two methods can be treated separately, and the HEP has to be additionally obtained, research into the reliability model of human intervention is being conducted separately, with a plan to merge them later.

Therefore, this study focuses on only the automatic safety signal generation (1 in Figure 2) in consideration of the failure mechanism and the effect of the fault tolerant technique.

3 TYPICAL CONFIGURATION OF DRPS

The DRPS of NPP has some differences in a detailed configuration according to the type and reactor model. To develop a general frame, it is first necessary to confirm the typical configuration of the DRPS covering various types through the examination of many systems. For this purpose, the DRPS applied to the IDiPS-RPS (Integrated Digital Protection System-Reactor Protection system) developed through the KNICS (Korean Nuclear Instrumentation and control) project, the OPR-1000 (Optimized Power Reactor), and the APR-1400 (Advanced Power Reactor) are investigated, and a typical configuration, as shown in Figure 3, is confirmed. For reference, only the parts related to automatic safety signal generation are shown in this configuration except for the indication parts. The notations of each configuration are as follows.

- AIM: Analog Input Module
- DIM: Digital Input Module
- PM: Processor Module
- CM: Communication Module
- F: Fiber Optic Module (FOM)
- DOM: Digital Output Module
- AT: Automatic Test Module
- CPC: Core Protection Calculator

The more detailed functions are as follows.

- It consists of four physically isolated multiple channels (A, B, C, and D) and performs the same function independently in each channel.
- There are two sub-racks in each channel, sharing the inputs and performing the same function.
- Each sub-rack in each channel is composed of bistable logic (B) that generates Trip and Engineering Safety Feature (ESF) signals through comparison with internal set points, and coincidence logic (C) that receives signals generated from each PM in bistable logic in each channel and applies a voting.
- Bistable logic receives inputs from sensors using two AIMs and CPC values through one DIM, and the actual comparison with internal set points is proceeded using AS (Application Software) in PM.
- There are three PMs in each coincidence logic, two of which perform trip-related functions, one PM applies ESF-related functions.
- The two DOMs in the coincidence logic receive the trip signal from connected PM and transmit it to the function of selective 2/4 logic.
- The safety signal generated by the PM in the bistable is transmitted to the coincidence logics in the other channels. In this process, FOM is used to ensure one-way signal transmission and independence between channels.
- Each PM in the coincidence logic receives two values generated by PMs in the different rack in the different channel. Before performing 2/4 voting logic in the associated PM in the coincidence logic, 1/2 logic is performed first for the two values at each PM.

– Regarding the fault tolerant techniques, each input module and PM has a self-diagnostic function that is performed in real time at the operating system level, and the AT module, which exists in each channel, detects some portion of faults in each module within a channel periodically through the network and the CMs. The fault detection coverage of each technique for each module is different but partially overlapped.

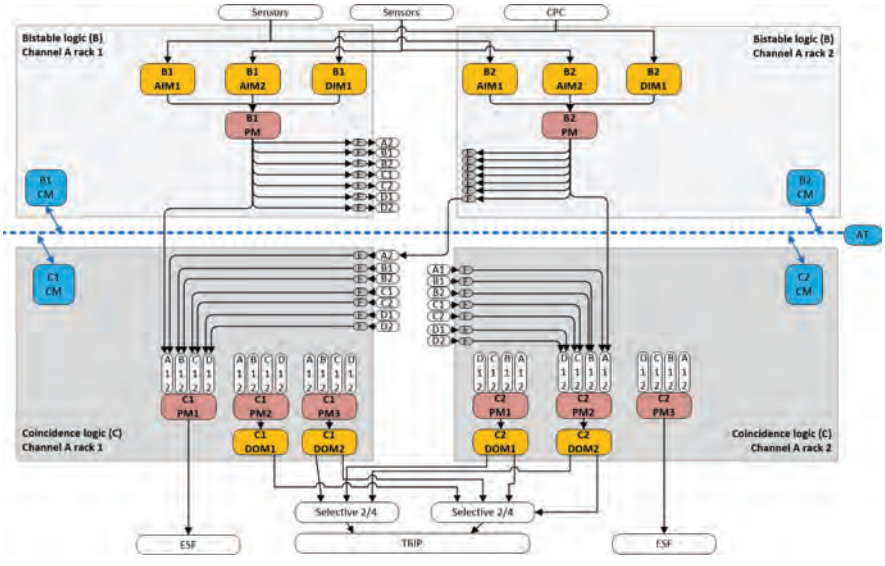


Figure 3. Typical configuration of DRPS for safety signal generation correspond to one channel (channel A).

Table 1. Proposed requirements [U.S.NRC, 2008].

	Proposed requirements	Note
R-1	Review the level of PSA proportional to the use of results and insights.	B
R-2	Identify how systems can fail and what these failures can effect.	A
R-3	Identify CCF events.	A
R-4	Address the uncertainties in modeling and data.	C
R-5	Confirm the capability of its safety function.	B
R-6	Address the impact of external events.	C
R-7	Model the failure of control room indication.	C
R-8	Determine and evaluate the scope, boundary condition, and modeling assumptions.	A
R-9	Model the recovery actions taken for loss of DI&C function.	C
R-10	Quantify the contribution of software failures.	A
R-11	Verify the credit for defensive design.	A
R-12	Review the DI&C data.	B, C
A-1	Verify that physical and logical dependencies are identified and their bases provided.	A
A-2	Evaluate the spurious actuations of diverse backup systems or functions.	C
A-3	CCFs can occur in areas where there is sharing of design, application, or functional attributes.	A
A-4	Evaluate the credit that should be given for defensive design features.	A
A-5	If a DI&C system shares a communication network, the effects on all systems due to failures of the network should be modeled.	A
A-6	Calculations, their bases, and the modeling assumptions used in standard methods may be warranted.	B
A-7	Review of applicant claims regarding data should be proportional to the use made of the PRA results.	B
A-8	Confirm the suitability of data based on the suggested criteria.	B
A-9	Interactions 1.between plant system and physical process, and 2.within a DI&C system.	C
A-10	Target reliability and availability specifications should be described (2)	B

4 REQUIREMENTS FOR DI&C PSA

The US is also carrying out research on a reliability evaluation of a digitalized I&C system in relation to the second-lifetime extension issue. However, because the recent guidance (10 CFR part 52) does not provide sufficient details regarding the DI&C system, a separate ISG (Interim Staff Guidance), which is consistent with the existing guidance, is issued to provide the reviewers with much more detailed requirements to be identified in the DI&C PSA (U.S.NRC, 2008). To consider the requirements to the general frame of the DI&C reliability model, the 12 review guidelines (R-x) and the 10 additional steps (A-x) in this reference are checked and marked according to the applicability to this study. In Table 1, the marks in the note column mean the follows: "A" means the requirements selected for consideration in this study, "B" indicates what is required in the actual review process, and "C" shows the requirements that are difficult to consider or apply in this phase of the study.

5 GENERAL FRAME OF RELIABILITY MODEL FOR SAFETY SIGNAL GENERATION

To implement the requirements selected in chapter 4 to the general frame, the DI&C system is basically divided into standard units, and within each unit, the characteristics of the HW/SW failure and the effects of fault tolerant techniques are considered. Some assumptions are made in this standard unit based approach and a general frame is developed on it.

5.1 *Assumptions and approaches*

The top event of the DI&C reliability model is defined as the failure of automatic safety signal generation for each mitigation system in each accident scenario.

To effectively analyze the DI&C system, this study took the module level, such as the input, output, and processor module, as the standard unit, as the different sets of modules are utilized for different safety signal generation.

Within a single module, the categories of failure are largely divided into an HW (hardware) failure, OS (operating system) failure, and AS (application software) failure. Although one failure among them can affect another, it is assumed that the original cause is independent of each other. As for the AS and OS, because the development process, developer, and functions of each is different, the two need to be considered separately. Furthermore, a different AS can be applied to the same base (HW and OS), and thus the AS and OS should be distinguished in order to consider the CCF in a more realistic way.

The effect of a fault tolerant technique can be reflected in such a way that the technique detects some portion of each HW, OS, and AS failures in each module and treats them in a fail-safe way so as to not lead to a module failure, and the reliability of the technique can be considered in this approach together.

5.2 *General frame of reliability model for a module failure*

Figure 4 shows the general frame of the reliability model for a module failure. This frame is applicable to all modules as a basic structure, but only some of them can be modeled according to modules. For example, the processor module has a specific AS for comparing the set points and input value, but the input/output modules are composed of only an HW and an OS, and do not have an AS.

The safety signal is generated by a sequential process in which a value generated in one module is passed to another module and a value generated in that module is passed to the other module. Therefore, the linkage between the modules can be modeled such that the failure of the previous module is input as the 'signal transmission failure' and the failure of that module is connected to the 'signal transmission failure' to the next module reliability model.

Failures of the HW, OS, and AS in each module leads to a module failure when they are not detected by the FTT or are detected but cannot be handled properly by the FTT. Based on the HW failures illustrated in Figure 4, if an HW failure is not detected by the FTT, it simply leads to a failure of the module. On the other hand, when an HW failure is detected by the FTT, it can be treated in a fail-safe way when the FTT works properly. Therefore, the reliability of

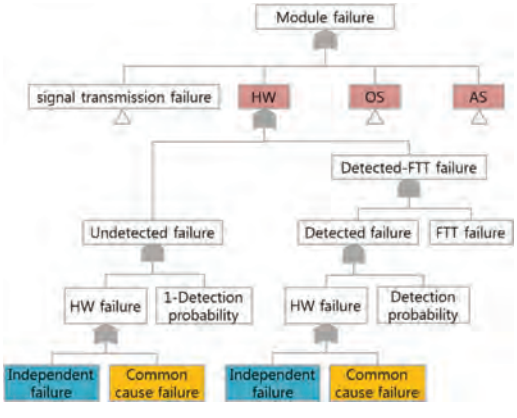


Figure 4. General frame of reliability model for a module failure.

the relevant FTT needs to be considered together in this case. The reliability of the FTT itself proportionally affects the reliability of the entire DI&C, and the degree of the influence depends on the magnitude of the fault detection coverage that the FTT has for each fault.

In certain cases, various FTTs are applied to a specific module simultaneously, in which case, regarding the detection probability and FTT failure, it is necessary to integrate the effects of the various FTTs or to model the effect of each in detail. In either approach, it can be extended based on this structure.

5.3 *Example application: Steam generator low water level trip*

In order to verify the feasibility of the proposed frame, as an example, a fault tree was developed for the configuration of the DRPS shown in Figure 3. As a result of analyzing which trip signal is required first for 18 initiating events of the OPR 1000 nuclear reactor model, the steam generator (SG) low water level trip signal was most likely to occur first [Cho et al., 2016]. The following assumptions were made during the development of the reliability model for the failure of this trip signal generation. The sensor for the SG water level is connected to each AIM1 in each rack, FOM integrity of both the transmitting and receiving sides should be guaranteed for the signal transmission from the bistable logic to the coincidence logic, and the failure of hardwired connection is ignored.

Figure 5 shows a part of the reliability model for the typical configuration of the DRPS shown in Figure 3, and Figure 6 shows an example of the

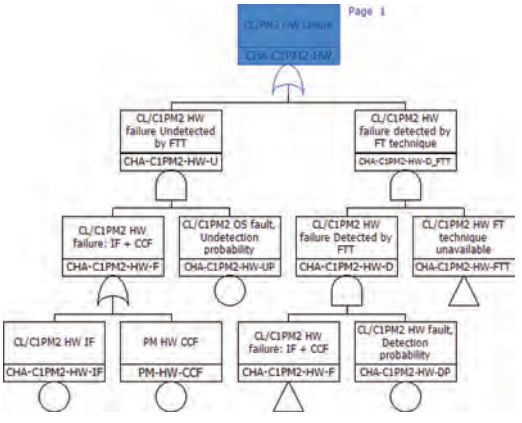


Figure 6. Example of application of the module reliability model (Processor module failure in figure 5).

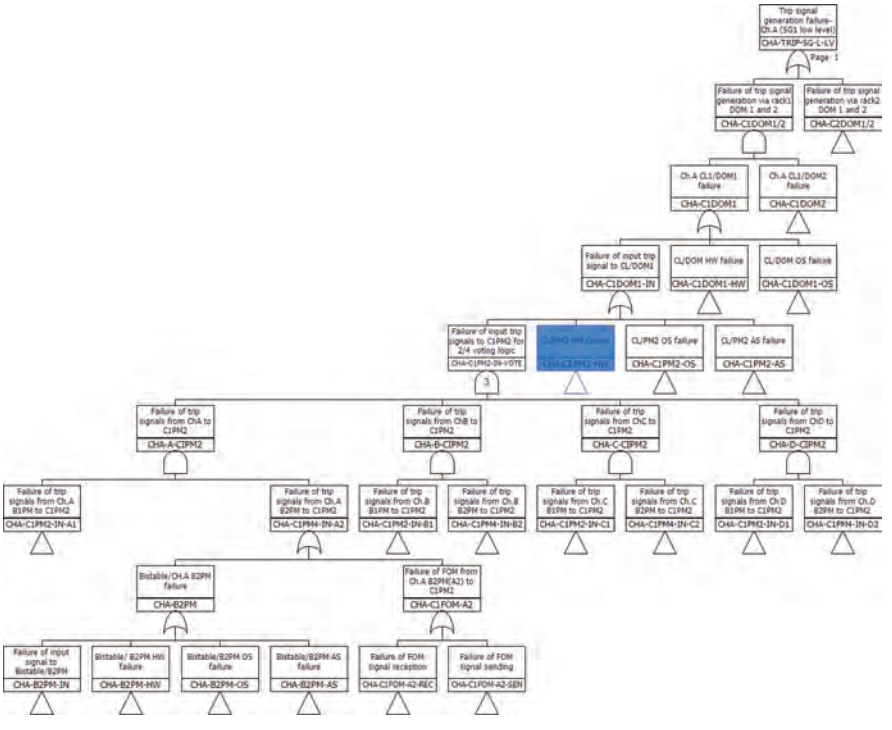


Figure 5. Example of reliability model of the general configuration of DRPS.

application of the general frame (Figure 4) for a module failure regarding the processor module HW failure shown in Figure 5.

6 DISCUSSION OF PROPOSED APPROACH

Although the failure information needs to be obtained to develop the actual reliability model, in the first phase of this study, we focused on the development of the general reliability frame for the DI&C system. On the other hand, the frame of the DI&C PSA developed through this study may suggest the required characteristics of the reliability data to be collected or the direction for securing them.

The important failure types identified based on the example application are basically the form of CCF, which can disable the independence and redundancy in the following cases, resulting in the failure of the entire system.

- Bistable logic input module HW/OS CCF
- Bistable logic PM HW/OS/AS CCF
- FOM HW CCF and OS CCF
- Coincidence logic PM HW/OS/AS CCF
- Coincidence logic DOM HW/OS CCF

Because it is difficult to compare the objective significance of the above cases through this qualitative analysis, it is necessary to reconfirm its importance by comparing the impacts on the overall system by developing/applying appropriate reliability data. Although the reliability data that can be applied at present have not been obtained yet, we can consider the importance of priority under the following conditions expected.

- Regarding the CCF Group, OS and HW are expected to be grouped into the same size according to each module, and the size of the AS CCF group is expected to be relatively small (CCF group size: HW = OS> AS).
- For the CCF parameter, the Alpha factor can be applied for the HW CCF, but the beta factor is more likely to be valid for the OS and AS CCF (value of CCF parameter: OS = AS>>HW).
- The reliability data value is expected to be large in the order of HW, AS, and OS (value of reliability data: HW>>AS > OS).

Finally, in addition to the above topics (CCF group, CCF parameter, and reliability data), for the development of the more advanced DI&C PSA, the following points should be further investigated or confirmed.

- The self-reliability of various FTTs and the combined fault detection coverage for each module
- Interdependency between SW and HW

7 CONCLUDING REMARKS

In this study, a method for DI&C PSA modeling for the automatic safety signal generation based on a general reliability model frame of a module. Although there are some assumptions and things to check to confirm the validity of this methodology, authors think that the complex relationship between elements composing the DI&C system can be objectively modeled by using frames suggested.

REFERENCES

Authen, S., Holmberg, J. 2012. Reliability Analysis of Digital Systems in a Probabilistic Risk Analysis for Nuclear Power Plants. *Nuclear Engineering and Technology.* 44. 471–482.

Cho, J.H., Shin, S.M., Lee, S.J., Jung, W.D., 2016. Exhausted Test Case for Software Reliability of Nuclear Digital Systems. *PSAM* 13. 2016.

Chu, T.L., Yue, M., Martinez-Guridi, M., Lehner. J.R., 2010. *Review Of Quantitative Software Reliability Methods*, Brookhaven National Laboratory, BNL-94047-2010.

Kang, H.G., Kim, M.C., Lee, S.J., Lee, H.J., Eom, H.S., Choi, J.G., Jang, S.C. 2009. An overview of risk quantification issues of digitalized nuclear power plants using static fault tree. *Nuclear Engineering and Technology*. 41. 849–858.

U.S. NRC. Review of new reactor digital instrumentation and control probabilistic risk assessment-revision0, DI&C-ISG-03.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Air traffic safety in relation to visualization systems reliability

J. Skorupski
Faculty of Transport, Warsaw University of Technology, Warsaw, Poland

P. Ferduła
Polish Air Navigation Services Agency, Warsaw, Poland

ABSTRACT: Complex technical systems assist air traffic controllers in management of air traffic in the airspace. Their very important task is to provide the controller with visualization of the traffic situation, thereby enabling situational awareness and making right decisions. In practice, errors sometimes occur, despite the efforts to improve the reliability of visualization systems. The purpose of this paper is to present a method to determine the impact of error types on the air traffic safety. The assessment of the threat level is influenced by subjective factors and cannot be expressed in a precise way. Therefore, the fuzzy reasoning theory has been used in the paper. The developed fuzzy model has been used to obtain a tool enabling the experimental simulation of the impact of various factors on traffic safety assessment. As the results of experiments indicate, the most important determinants of safety are: the time when air traffic controller remains unaware of the breakdown and the total time he/she does not have full knowledge of the traffic situation. It can be therefore stated that the key role for the proper operation of the air traffic visualization system and thus the restoration of full situational awareness is played by self-diagnostic systems that can restore the visualization system's correct functioning without even the controller being aware of the error occurrence. Their role in assuring safety might be even greater than redundancy which is commonly used.

1 INTRODUCTION

The airspace is divided into smaller volumes called sectors in every of which a radar controller is responsible for the safety of aircraft. One of the most important determinants of ability to issue appropriate clearances is precise information about the traffic situation. The controllers are supported by Air Traffic Control (ATC) systems. Their subsystem of key importance is the Traffic Situation Visualization System (TSVS) that is responsible for representation of the aircraft's positions. ATC systems are constructed with the awareness of their role in air traffic safety. Nonetheless, software bugs and hardware failures still occur sometimes. One of the most severe effects are visualization systems operation errors.

An error in the visualization system causes a partial loss of controller's situational awareness (Endsley & Smolensky 1998). Depending on the complexity of the situation and the error type, various risks of air traffic safety may be caused. In this paper different types of visualization systems' errors are analyzed, their impact on the air traffic safety is assessed quantitatively, and finally, some case studies are conducted, vulnerability is examined and opportunities to reduce the risk are sought.

There are many studies examining the various types of factors affecting air traffic safety, as well as evaluating it. Many authors suggest a quantitative approach to the analysis and assessment of safety and security (Lee 2006, Ali et al. 2015, Vismari & Camargo 2011, Skorupski 2015, Skorupski 2016; Patriarca et al., 2017; Stroeve et al., 2015).

In typical ATC systems, information about three-dimensional (3D) scenery is displayed with a two-dimensional representation. Bagassi et al. (2010) presented an innovative concept based on a four-dimensional (4D = 3D space + time) visualization. A new working environment containing special information was proposed by Rohacs et al. (2016).

Many papers have been devoted to analyzing controllers' information perception (Moehlenbrink & Papenfuss 2011, Inoue et al. 2012). Ahlstrom (2005) analyzed the results of improper construction of visualization systems, on the probability of causing a threat to air traffic safety. Kesseler & Knapen (2000) highlighted the need to consider the interactions between the controllers, ATC systems and functions offered by individual systems.

Many papers suggest using the fuzzy logic methods and tools in the area of air traffic management (Hadjimichael 2009, Teodorović & Lučić 1998, Lower et al. 2016). The research by Xianfeng &

Shengguo (2012) and Skorupski & Uchroński (2015, 2016) includes the attempt of airport security assessment where the human factor was taken into consideration. The other examples of fuzzy methods utilization in air traffic management can be found in Babić & Krstić (2000) and Netjasov (2004).

The literature review indicates the need to analyze the visualization systems' errors that occur in air traffic controllers' practice. This analysis will be the basis for assessing the risk caused by different types of errors. This assessment has a significant degree of subjectivity and is impossible to be quantified unequivocally. In such situations, methods which deal well with uncertain and imprecise information are applied. In our paper, fuzzy logic, more precisely fuzzy reasoning systems are used to develop an expert advisory system that will categorize different error types to hazard classes.

It is also important to look for the kinds of errors that have the greatest impact on safety. Several distinct factors are considered, such as the ability to identify the error quickly or the availability of backup resources. The rest of the paper is organized as follows. Section 2 outlines the essence of visualization systems' errors. Section 3 provides a brief introduction to the theory of fuzzy reasoning systems. Section 4 describes a fuzzy model for assessing the hazard caused by TSVS errors. Section 5 shows the results of several simulation experiments using a computer tool created in the SciLab environment. Section 6 provides a summary and final conclusion.

2 AIR TRAFFIC CONTROL SYSTEMS

The main task of ATC systems is to assist air traffic controllers in ensuring a safe and effective flow of air traffic, but their range of applications is much wider. That is why they are often called Air Traffic Management (ATM) systems.

2.1 *General structure of ATC systems*

The main part of an ATM system is usually a server processing data from multiple sources. Following main data processing modules can be listed:

- surveillance data processing module,
- tracker—its task is to follow objects based on surveillance data,
- flight data processing module,
- decision supporting modules.

2.2 *Air traffic controllers work technology*

Actions performed by radar controllers depend on many factors, like type and distribution of traffic streams, traffic volume or airspace availability. To ensure safety and efficiency of air traffic, the controller needs information. The most important source of traffic information is the visualization system. It allows determining the position of the aircraft related to different objects.

Anticipating future positions of the aircraft is an essential element of the controller's work. A function of displaying routes according to current flight plans and so-called vectors (predicted trajectories of the aircraft) is used for this purpose.

Another action of the controller interacting with the TSVS is to read the predicted time necessary for the aircraft to arrive at a specific point. A major part of the radar controller's work is also monitoring the aircraft maneuvers to ensure that they are consistent with the expectations and do not jeopardize safety.

A cardinal part of the work is verifying of the data displayed by the system, such as the flight level, its speed, the heading, the altitude selected by the crew in FMS (Flight Management System). These data are essential for making the right decisions.

2.3 *The role of the visualization system*

The actions performed by the controller, which are presented in Section 2.2, show that the visualization system is a fundamental component of the air traffic control system. Functionally, the most important role of the TSVS is to assist the controller in creating an image of the current and future traffic situation. As this is the basis for decision making, the visualization system is the controller's essential tool. Its role is further enhanced by integrating the TSVS with some executive features, for example transfer of control to an adjacent sector.

2.4 *Errors in ATC systems*

Experience in using TSVS shows that some errors may occur. This section contains their classification. All the described errors have been observed during operational work at the air traffic control position at approach (APP) and area control (ACC) units (one of this paper's co-authors is an active air traffic controller).

2.4.1 *Incorrect indication of aircraft position*

This error lies in showing the aircraft position symbol at a distance from its actual position. The most likely cause is a malfunction of the tracker algorithm due to an internal error or erroneous input. The significance of this error is determined by the number of incorrectly positioned symbols and the type of misrepresentation. The most dangerous situation occurs when the symbols on display are spaced apart while, in fact, aircraft are close together. In case of significant divergences, the

error is relatively easy to spot, but it generates stress and high workload to clarify the situation. Furthermore, this situation is dangerous also because of distraction from observing other traffic.

2.4.2 *Flight plan—track linking error*
This type of error may appear as a total or partial lack of flight plan data, that should be available when aircraft's position symbol is indicated. Another variant of this error is the inability to update the current flight plan. This function is integrated with the visualization system as standard. The threat to safety is dependent on the number of aircraft affected by the error, the period that the issue exists and the number of flight plan parameters which are incorrect.

2.4.3 *Disappearance of aircraft position symbol*
Threat level caused by this error depends on the time when the symbol is not displayed. During a straight and level flight, a temporary disappearance of the symbol of a single aircraft makes a relatively small trouble. However, when more complicated maneuvers are performed or when the disappearance prolongs, the threat created by such a failure is much higher. Also, aircraft symbols' disappearance prevents the controller from monitoring the aircraft's maneuvers which in some cases can substantially increase the safety risk.

2.4.4 *Delayed aircraft position update*
It takes some time from the moment of measuring the aircraft's position to the moment of displaying it in the TSVS. It consists of the data transmission time, processing time, hardware delay. As long as the delay is approximately constant, the error may be compensated. The problem occurs when this value is variable. In such a situation, an aircraft movement may be displayed unrealistically. The error of this kind is easy to spot and is not a serious problem as long as the image is constantly visible and the deviations are not large. The problem, however, arises when the crew performs an incorrect maneuver or crosses the final approach track.

2.4.5 *Conflict warning systems malfunctions*
ATC systems are equipped with the functions of short- and medium-term conflict detection (STCA and MTCD). All of the above-mentioned positioning algorithms' shortcomings and errors such as information delay and incorrect input data may indirectly lead to conflict warning systems malfunctions. The most serious is the lack of necessary STCA or MTCD action or its too late activation. However, this type of error is extremely rare. The so-called false alarms are much more widespread. False alarms may also be caused by errors in STCA and MTCD algorithms. Whatever the cause, false alarms can pose a threat.

2.4.6 *Total loss of image*
An error of this type may be caused by a restart of the workstation resulting from an internal system error, a major technical failure or even a terrorist activity. This error can be divided into two kinds: the image disappears entirely or just stops refreshing. The latter situation is obviously better from a safety point of view, as it allows the controller to use historical data to build a picture of the traffic situation for some time. Depending on the number and type of the protection systems (additional power sources, additional data lines) the period for which the image is lost may differ. Lack of image on several workstations is even more critical as there is no opportunity to use the picture at the workstation nearby.

3 FUZZY REASONING SYSTEMS

The problem discussed in this paper is distinct by two major characteristics. On the one hand, we consider a socio-technical system where the role of the human factor is crucial. The result is an intense subjectivism of opinions. That is since the controllers are not equally vulnerable to make mistakes arising from a faulty indication in the TSVS.

On the other hand, high ambiguity and lack of precision characterize the problem. Errors in the visualization system are concerned, that are unpredictable. Not only we cannot predict the time of their appearance, but also we are unable to define their type precisely. That is caused by the fact that errors can result from numerous causes.

In such situations, literature recommends using the tools and methods suitable for problems of epistemic uncertainty, i.e. ones in which full knowledge of the phenomenon is unavailable. In such cases, it is required to use expert opinions. It is a known fact that very often they are formulated in a descriptive and an imprecise way.

Among the possible approaches, we have chosen to use the fuzzy logic, in particular, fuzzy reasoning systems. Zadeh (1965) created the basis for modern applications of fuzzy logic.

A fuzzy set A will denote a set of

$$A = \left\{ \left(x, \mu_A(x) \right) : x \in X, \mu_A(x) \in [0,1] \right\}$$

where μ_A is the membership function of this set and X is a set of considerations.

A linguistic variable is a variable whose values are words or sentences in a natural or artificial language. These words or sentences will be called the linguistic values of a linguistic variable.

Within the scope of the reasoning process, we will use the input value fuzzification block, reasoning block using some fuzzy rules and the defuzzification block. The rule sets will be created using experts' opinions, in particular, air traffic

controllers. We will use the so-called compositional method of reasoning introduced by Zadeh (1973) which uses a generalized "modus ponens" fuzzy reasoning rule.

4 FUZZY REASONING SYSTEM FOR THREAT LEVEL ASSESSMENT CAUSED BY ERRORS IN VISUALIZATION SYSTEMS

In this section, the fuzzy reasoning system, which allows for an assessment of the threat caused by errors in the visualization system. Factors influencing threat assessment are introduced as well as their representation by linguistic variables that are inputs to the fuzzy inference system. The knowledge base plays a critical role in this system. We have gained it from experts—air traffic controllers. A computer application created in SciLab environment allows for assessment of specific breakdown situations.

4.1 *Factors influencing the threat level assessment*

4.1.1 *Degree of situational awareness loss*

The concept of 'situational awareness loss' generally describes all situations when a controller or pilot is not entirely aware of what the current traffic situation is. Visualization errors cause safety risks only if they cause a loss of situational awareness for the air traffic controller. Obviously, the level of threat depends on the degree of loss of situational awareness, that is how the image of the traffic situation created in the controller's mind differs from reality.

4.1.2 *Awareness of errors in the situation image*

The problem of the loss of situational awareness is linked with the issue of the controller's belief that the picture of the traffic situation he/she has created in mind is correct. In some cases, he/she may be convinced that the image is proper, while reality is different. That is the most dangerous case because the controller will continue to work based on the wrong image without taking any verification action.

Knowledge of the existence of a malfunction may appear after some time, which depends on the obviousness of this error. The period from the moment the error occurs until the controller learns about it will be used as an evaluation criterion for this factor.

4.1.3 *Time of situational awareness loss*

Another factor affecting the degree of threat is the time interval in which the controller, because of a system error, does not have full situational awareness. It may range from a few seconds to dozens of minutes. The longer the time, the greater the safety threat. For this factor, we will use a judgment based on the anticipated time of situational awareness loss.

4.1.4 *Backup resources availability*

When a controller is aware of the error, especially if the error persists for an extended period, he/she will try to use other available resources to keep him/her aware of the situation. A straightforward and efficient manner is to use:

- an image in another workstation,
- a backup system,
- another data source of aircraft positions.

An important remedy is the use of traditional flight progress strips, which can be employed when no traffic picture is available.

Backup resources availability should be considered in two ways. Firstly, it is necessary to determine whether backup resources are available at all and, secondly, to assess their quality.

4.1.5 *Human factor*

Maybe the most important, factor influencing the assessment of the safety threat in case of visualization system failure is the human factor. One of the key components is the level of training. Certain standards and requirements must be met by all controllers, but the differences between individuals may affect their ability to deal with an emergency.

Besides, the experience of the controller, which can be expressed both by the number of years of work or by the number of hours worked at the position, affects the level of threat.

On the other hand, the perception abilities of the human being fall with age, so that the older person is less able to perceive and remember, and this can have an adverse impact on actions in a particular situation. Another, equally important component of the human factor that affects the level of threat is the psychophysical condition of the controller.

4.2 *General structure of the Threat level fuzzy inference system*

The scheme of the fuzzy model for assessing the level of threat caused by the error in the visualization system is shown in Figure 1. The output variable *Threat level* (z_t) depends on five input variables. These are: *Loss of situational awareness* (x_{sa}), *Time without knowledge of error* (x_{wk}), *Time of situational awareness loss* (x_{lsa}), *Quality of remedies* (x_q) and

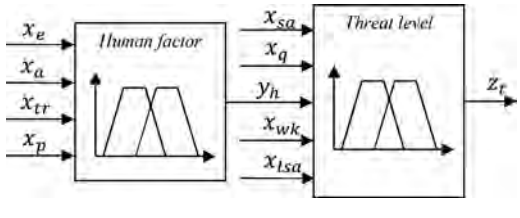


Figure 1. General scheme of the *Threat level* fuzzy model.

Human factor (y_h). The last of these input variables is the output of the local fuzzy reasoning system, with four input variables: *Experience* (x_e), *Age* (x_a), *Training* (x_{tr}), *Psychophysical condition* (x_p).

The form of membership functions of linguistic variables values, the basis for their determination and the knowledge bases of both fuzzy reasoning models will be presented in subsequent sections.

4.3 *Input linguistic variables of fuzzy reasoning system*

4.3.1 *Loss of situational awareness*

The degree of divergence of the traffic situation picture in the controller's mind relative to reality can be assessed based on the difference between the actual position of the aircraft and the position indicated (incorrectly) by the visualization system. We refer this value to the separation minimum that is obligatory in a given airspace. Based on expert knowledge, it has been assumed that the linguistic variable *Loss of situational awareness* will take one of the six values: *slight*, *small*, *significant*, *serious*, *large*, *total*. We propose the use of an integrated indicator with the following form to determine the value of a linguistic variable:

$$d = max\left(n \cdot e^{\frac{\delta_r}{S_r}}, n \cdot e^{\frac{\delta_v}{S_v}} \right)$$

where:

d – integrated indicator determining the degree of situational awareness loss,

n – number of aircraft, which visualized positions do not correspond to their actual locations,

δ_r, δ_v – the amount of deviation of the imaged position from the actual position of the aircraft in the horizontal and vertical plane respectively,

S_r, S_v – separation minimum obligatory in the considered airspace in the horizontal and vertical plane respectively.

Membership functions of values of linguistic variable *Loss of situational awareness* are shown in the logarithmic scale in Figure 2.

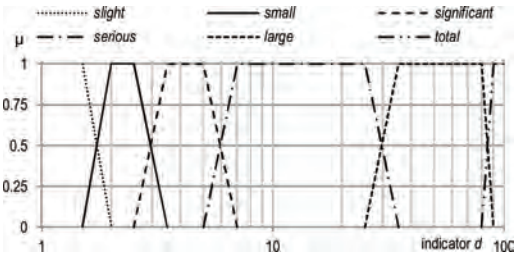


Figure 2. Membership functions of values of *Loss of situational awareness* linguistic variable.

4.3.2 *Time without knowledge of error*

The level of safety threat depends on whether the controller is aware of the occurrence of an error or, more specifically, how long he/she is not. Accordingly, we have used the linguistic variable *Time without knowledge of error*, which can take five values: *very short, short, average, long, very long*. The membership functions of the values of this linguistic variable were adopted based on expert knowledge, and their logarithmic form is shown in Figure 3.

4.3.3 *Time of situational awareness loss*

Depending on the type of error, the period in which the situational awareness of the controller is disturbed differs. In case of a short-term disappearance of the track of individual aircraft, controller's situational awareness is usually maintained all the time. In case of major system malfunction, the time of situational awareness loss may be long and is not necessarily the same as the time when the system works incorrectly.

The linguistic variable *Time of situational awareness loss* may take five values: *very short, short, average, long, very long* (Figure 4).

4.3.4 *Quality of remedies*

A controller who is aware of the dysfunctional operation of the TSVS will seek to use other available means to ensure air traffic safety. As already

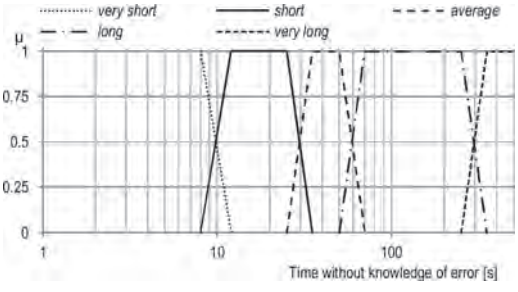


Figure 3. Membership functions of values of *Time without knowledge of error* linguistic variable.

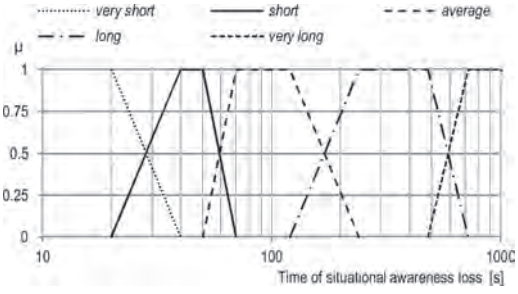


Figure 4. Membership functions of *Time of situational awareness loss* linguistic variable values.

Table 1. Fuzzy inference rules for the local model *Threat level*.

Rule number	Loss of situational awareness (x_{sa})	Time without knowledge of error (x_{wk})	Time of situational awareness loss (x_{lsa})	Quality of remedies (x_q)	Human factor (y_h)	Threat level (z_t)
1	*total*	any	≠ *very short*	*none*	any	*very high*
9	*large*	any	*short*	*none*	any	*high*
27	*significant*	*very short*	*average*	*none*	*very high*	*average*
30	*significant*	*average*	*average*	*average*	*average*	*high*
40	*small*	any	*average*	any	*very high*	*low*
45	*slight*	any	*very short*	any	any	*very low*

mentioned, the possibility of using them in each situation will be considered within two categories—availability and quality of available remedies. Linguistic variable *Quality of remedies* will take four values: *none*, *low*, *average*, *high*.

4.3.5 *Human factor*

For an assessment of the influence of the human factor on the degree of safety threat in the event of a TSVS failure, we will use combined information about the professional experience, age, the level of training and psychophysical condition of the controller. The linguistic variable *Human factor*, which describes the ability to cope with a failure of a visualization system in general, will take five values: *very low*, *low*, *average*, *high*, *very high*. It will be determined by the result of the local fuzzy reasoning system with four inputs – *Experience*, *Age*, *Training*, *Psychophysical condition*.

The linguistic variable *Experience* will take three values: *low*, *average* and *high*, and will be determined by the number of years of operation at the radar control position. The linguistic variable *Age* will take three values: *young*, *middle* and *old*. The linguistic variable *Training* will take one of three values: *poor*, *average* and *good*, and will be determined by the controller's level of training in procedural control. The linguistic variable *Psychophysical condition* will take one of three values: *poor*, *average*, *good*.

4.4 *Output variables of the fuzzy reasoning systems*

Both local models *Human factor* and *Threat level* are Takagi-Sugeno-Kang models with singleton output values. The *Human Factor* variable, discussed in Section 4.3, is also the input variable for the *Threat level* model. In turn, this variable will take five values: *very low*, *low*, *average*, *high* and *very high* (Figure 5). At the output of the fuzzy reasoning system, we evaluate the level of safety threat caused by the visualization system error as a real number from the interval [1,5].

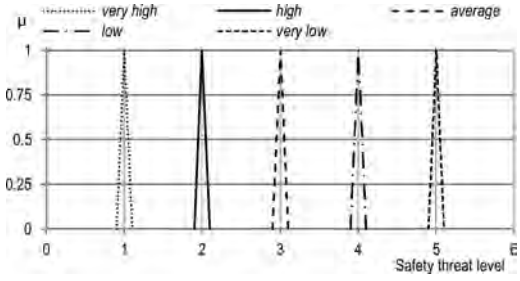


Figure 5. Membership functions of values of *Threat level* linguistic variable.

4.5 *Knowledge base of the fuzzy reasoning system*

The knowledge in inference systems describing complex socio-technical systems is subjective to some degree and as such impossible to quantify precisely. Therefore, fuzzy inference rules based on expert knowledge have been applied. One of the authors of the paper is an active air traffic controller, but for more credibility of the knowledge base, the rules have been verified with other field experts.

In the *Human factor* fuzzy reasoning system, 81 fuzzy inference rules are defined. In the *Threat level* fuzzy reasoning system, 49 rules have been defined, some of which are shown in Table 1.

The fuzzy model for evaluation of the safety threat to air traffic caused by visualization system errors has been implemented in the SciLab 5.4 environment with Fuzzy Logic Toolbox package.

5 SIMULATION EXPERIMENTS

The developed model together with its computer implementation allows us to assess the influence of errors in TSVS on traffic safety. Simulation experiments have been carried out to check the usability of this software tool for threat assessment and also for the selection of the most critical system components that influence the safety

of air traffic control. Some of these experiments are described in this section. For all experiments, it was assumed that the controller who encountered an error has been characterized by the average values of parameters related to their individual characteristics (age, experience, training, psychophysical condition).

5.1 *Scenario S1 – minor failure*

The analyzed scenario (S1) can be described as follows. Due to an anomaly in the tracker subsystem, one of the tracks representing an aircraft stops and does not change its position. Because of heavy traffic, the air traffic controller does not notice an error, and after one minute the system automatically switches to a backup tracker. The maximum deviation of the radar position shown in relation to the actual aircraft location was 7 NM. The aircraft involved was performing a level flight. Remedies were not available.

The input parameters of the fuzzy inference system and the results of the experiment for the scenario S1 are given in Table 2.

For the scenario being analyzed, the result obtained from the fuzzy inference system places the emergency in a *low* threat area with a slight shift towards the *average* rating (Figure 5).

5.2 *Scenario S2 – major failure*

The scenario S2 can be described as follows. Because of power failure, the controller completely loses indications from TSVS and the monitor goes blank. At the time, there are 10 aircraft in the control sector. The backup display located nearby is available, so after one minute the air traffic controller starts to work with its use. That finishes the emergency. The controller's characteristics are the same as in the scenario S1.

The input parameters of the fuzzy reasoning system and the results of the experiment for the S2 scenario are given in Table 3.

For the scenario being analyzed, the result obtained from the fuzzy reasoning system places the situation in a *high* threat area with a slight shift towards the *average* rating.

Table 2. Results of the experiment for the scenario S1.

Parameter	Value	Threat level
Human factor	3.0	
Loss of situational awareness [indicator *d*]	2.7	
Time without knowledge of error [s]	60	3.8
Time of situational awareness loss [s]	60	
Quality of remedies	*none*	

Table 3. Results of the experiment for the scenario S2.

Parameter	Value	Threat level
Human factor	3.0	
Loss of situational awareness [indicator *d*]	27.2	
Time without knowledge of error [s]	0	2.3
Time of situational awareness loss [s]	60	
Quality of remedies	*none*	

Table 4. Results of the experiment in scenario S1a.

Parameter	Value	Threat level
Human factor	3.0	
Loss of situational awareness	8.5	
Time without knowledge of error [s]	90	2.0
Time of loss of situational awareness [s]	120	
Quality of remedies	*none*	

5.3 *Sensitivity analysis in scenario S1*

The scenario S1 is characterized by a relatively small deterioration of safety because TSVS quickly switches to a backup tracker. At this point, we will assume that the switch to the backup tracker does not take place, and the track does not move for a few minutes. We will mark it as scenario S1a. After about 90 seconds the controller notices the error and, using the available previous generation ATC system, continues the work after about two minutes. The difference between the displayed and the actual position of an aircraft is 15 NM.

The input parameters of the fuzzy inference system and the results of the experiment for scenario S1a are set out in Table 4.

As we can see extending the duration of the failure causes the threat level to fall into the *high* rating area. That clearly shows how dangerous these errors can be, and how important it is to implement effective self-diagnostic means in TSVS, that are responsible for detecting, for example, a tracker error and switching to a backup.

6 SUMMARY AND FINAL CONCLUSIONS

Traffic situation visualization modules are essential elements of air traffic control systems. They constitute a basis for building situational awareness for air traffic controllers. At the same time, they focus all the hardware failures and software errors which, despite the use of technology with a very high level of reliability, can happen in practice.

Regardless of the origin of malfunctions of the system, they can result in several typical situations

that have been categorized in Section 2.4. The essence of this paper has been to analyze the level of threat to air traffic resulting from errors of each category, taking into account factors such as the controller's experience or the volume of traffic at which the error occurred.

Experiments have shown that one of the most important factors influencing threat assessment is the amount of time a controller does not have full knowledge of the traffic situation. The time depends on the awareness that we are handling an abnormal image of the TSVS. That, in turn depends on the type of error. The results of experiments carried out using the created computer tool confirm these observations. In addition, they allow for quantitative assessment. It is worth noting that the results indicate a crucial role of diagnostic modules built into ATC systems. Waiting for the controller to notice an error in TSVS and take any corrective action can significantly increase the time spent without complete knowledge of the traffic situation. It is therefore possible to provide a general recommendation to extend and further develop such systems. They can be even more important than redundancy that is usually used for increasing reliability of the system. In the case of redundancy, duplication of the same error can occur on all backup devices. In contrast, self-diagnostic systems can restore system performance even without the controller being aware of the malfunction.

REFERENCES

Ahlstrom, U., 2005. Work domain analysis for air traffic controller weather displays. *Journal of Safety Research*, 36 (2), 159–169.

Ali, B.S., Ochieng, W.Y., Schuster, W., Majumdar, A., and Chiew, T.K., 2015. A safety assessment framework for the Automatic Dependent Surveillance Broadcast (ADS-B) system. *Safety Science*, 78, 91–100.

Babić, O. and Krstić, T., 2000. Airspace daily operational sectorization by fuzzy logic. *Fuzzy Sets and Systems*, 116 (1), 49–64.

Bagassi, S., De Crescenzio, F., and Persiani, F., 2010. Design and evaluation of a four-dimensional interface for air traffic control. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 224 (8), 937–947.

Endsley, M.R. and Smolensky, M.W., 1998. Situation awareness in air traffic control. *In: Human factors in air traffic control*, Academic Press, SanDiego.

Hadjimichael, M., 2009. A fuzzy expert system for aviation risk assessment. *Expert Systems with Applications*, 36 (3, Part 2), 6512–6519.

Inoue, S., Furuta, K., Nakata, K., Kanno, T., Aoyama, H., and Brown, M., 2012. Cognitive process modelling of controllers in en route air traffic control. *Ergonomics*, 55 (4), 450–464.

Kesseler, E. and Knapen, E.G., 2000. Interactions : Advanced controller displays, an ATM essential. *In*: *3rd USA/Europe Air Traffic Management R&D Seminar Napoli*. 1–15.

Lee, W.-K., 2006. Risk assessment modeling in aviation safety management. *Journal of Air Transport Management*, 12 (5), 267–273.

Lower, M., Magott, J., Skorupski, J., 2016. Analysis of Air Traffic Incidents using Event Trees with Fuzzy Probabilities. *Fuzzy Sets and Systems*, 293, 50–79.

Moehlenbrink, C. and Papenfuss, A., 2011. ATC-monitoring when one controller operates two airports: Research for remote tower centres. *In*: *55th Human Factors and Ergonomics Society Annual Meeting, HFES 2011*, 76–80.

Netjasov, F., 2004. Fuzzy expert model for determination of runway in use case study: Airport Zurich. *In*: *1st International Conference on Research in Air Transportation ICRAT 2004*. Zilina, Slovakia, 59–64.

Patriarca, R., Di Gravio, G., and Costantino, F., 2017. A Monte Carlo evolution of the Functional Resonance Analysis Method (FRAM) to assess performance variability in complex systems. *Safety Science*, 91, 49–60.

Rohacs, J., Rohacs, D., and Jankovics, I., 2016. Conceptual development of an advanced air traffic controller workstation based on objective workload monitoring and augmented reality. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 230 (9), 1747–1761.

Skorupski, J., 2015. The risk of an air accident as a result of a serious incident of the hybrid type. *Reliability Engineering & System Safety*, 140 (140), 37–52.

Skorupski, J., 2016. The simulation-fuzzy method of assessing the risk of air traffic accidents using the fuzzy risk matrix. *Safety Science*, 88, 76–87.

Skorupski, J. and Uchroński, P., 2015. A fuzzy model for evaluating airport security screeners' work. *Journal of Air Transport Management*, 48, 42–51.

Skorupski, J. and Uchroński, P., 2016. A fuzzy system to support the configuration of baggage screening devices at an airport. *Expert Systems With Applications*, 44, 114–125.

Stroeve, S., Doorn, B. Van, and Bakker, B., 2015. A risk-based framework for assessment of runway incursion events. *11th USA/Europe Air Traffic Management Research and Development Seminar ATM 2015*, 1–11.

Teodorović, D. and Lučić, P., 1998. A fuzzy set theory approach to the aircrew rostering problem. *Fuzzy Sets and Systems*, 95 (3), 261–271.

Vismari, L.F. and Camargo jr, J.B., 2011. A safety assessment methodology applied to CNS/ATM-based air traffic control system. *Reliability Engineering and System Safety*, 96 (7), 727–738.

Xianfeng, L. and Shengguo, H., 2012. Airport Safety Risk Evaluation Based on Modification of Quantitative Safety Management Model. *Procedia Engineering*, 43, 238–244.

Zadeh, L.A., 1965. Fuzzy sets. *Information and Control* 8 (3), 338–353.

Zadeh, L.A., 1973. Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Transactions on Systems Man and Cybernetics* 3: 28–44.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Automated driving on steel and rubber

H. Schäbe
TÜV Rheinland InterTraffic GmbH, Köln, Germany

ABSTRACT: In this paper, we provide a comparison between principles and experience of autonomous or automatic systems on rails and on the street. An automatic metro operates in a controlled and well-defined environment that makes automatic driving possible. Passengers are separated from moving systems, e.g. by using platform screen doors that allow access only directly into the train, which is at standstill. In addition, passengers and third persons are separated from driving trains by fences, tunnels, etc. For road vehicles, currently a large number of assistance system is available that are able to handle specific situations. This leads to the impression that these vehicles can move autonomously. However, these assistance systems are developed in such a manner that the driver must always be able to interfere. There are only some exclusions with genuine autonomously moving vehicles. In general, the environment, in which a road vehicle operates, is much more complex than that of a train, mainly caused by unforeseen situations. We describe differences regarding approval for automated metros, road vehicles and so called Automated Guided Vehicles (AGV). Legal requirements for homologation of road vehicles according to the convention on road traffic are discussed and the implication for the system and the behavior of the driver. We sketch the current technical possibilities for automated driving and the existing technical solutions.

1 INTRODUCTION

Autonomous driving on the street has become more and more popular and the first demonstrator systems are operational. On the other hand, automatic metros and people movers are already successfully working for many years.

In this paper, we provide a comparison between principles and experience of autonomous or automatic systems on rails and on the street.

We compare the different levels of automation as defined by UITP and SAE and their meaning for the system. In addition, manual fallback modes are considered.

An automatic metro is located in a controlled and well-defined environment that makes automatic driving possible. Passengers are separated from moving systems, e.g. by using platform screen doors that allow access only directly into the train.

For road vehicles, currently a large number of assistance system is available that are able to handle specific situations. This leads to the impression that these vehicles move autonomously.

In general, the situation for a road vehicle is much more complex than that of a train.

We describe differences regarding approval for automated metros, road vehicles and so called Automated Guided Vehicles (AGV). Legal requirements for homologation of road vehicles according to the convention on road traffic are discussed and the implication for the system and the behavior of the driver.

We sketch the current technical possibilities for automated driving and the existing technical solutions. Especially, we discuss the possibilities and restrictions of artificial intelligence. We briefly describe a roadmap of possible next steps.

2 THE STATUS WITH METROS AND PEOPLE MOVERS

In many cities in the meanwhile automated metros and automated people movers are working

Examples are

- On the New York City Subway, the BMT Canarsie Line.
- On the London Underground, the Central, Northern, Jubilee, and Victoria lines run with ATO.
- On the Nuremberg U-Bahn, existing U2 and new U3 lines converted to ATO.
- On the Barcelona Metro, the L9 (as the Europe's longest driverless line), L10 and L11 runs with ATO.
- The Rio Tinto Group has the iron ore railway driverless go-ahead.
- The Tren Urbano, has an Siemens ATC system that allows for fully automatic operation.
- The Vancouver SkyTrain.

- Frankfurt Airport Skyline.
- Copenhagen Metro.
- On the Milan Metro, the M1 Red Line runs with ATO.

On the Mass Rapid Transit (Singapore), all lines operating currently run with ATO since 1987

For metros and people movers, a principle of separation has been applied: The automated trains are separated from all other traffic, running in the tunnels, open track is separated by fences, platform screen doors are used to separate the trains from passengers. This simplified the exploitation conditions significantly.

The Automated Train Protection system (ATP) is used to prevent collision and derailment. This allows also manually operated trains to use the same network.

The normal safety requirement for the ATP is a safety integrity level SIL 4. Nevertheless, manually operated fallback modes exist. Partially stewards are present to assist the passengers, especially in case in case of evacuation.

For metros and people movers, the UITP (2017) has established 5 levels of automation. That means, the picture is not black and white, knowing either manual or automated driving. Automation is a stepwise process. The following five levels are established, UITP (2017).

GoA 0 is on-sight train operation, similar to a tram running in street traffic. (No automation at all)

GoA 1 is manual train operation where a train driver controls starting and stopping, operation of doors and handling of emergencies or sudden diversions.

GoA 2 is Semi-automatic Train Operation (STO) where starting and stopping is automated, but a driver operates the doors, drives the train if needed and handles emergencies. Many ATO systems are GoA 2.

GoA 3 is Driverless Train Operation (DTO) where starting and stopping are automated but a train attendant operates the doors and drives the train in case of emergencies.

GoA 4 is Unattended Train Operation (UTO) where starting and stopping, operation of doors and handling of emergencies are fully automated without any on-train staff.

As a conclusion, automatic metros and automatic people movers can be seen as established systems. However, one needs to note that they operate in a controlled and simplified environment.

Road vehicles: The general impression on how autonomous driving works is mainly dominated by vehicles as the Google vehicle or the Tesla and other systems that have shown up in the meanwhile. Simpler systems are those for automated parking, which is carried out using the mobile phone, the driver being outside. Studies for autonomous driving have been carried out with a driver on board for testing purposes or for demonstration. Automated Guided Vehicle on closed areas or transport systems in workshops are also applied. The latter systems are strictly speaking not road vehicles but moving machines.

As an example, just consider the Google vehicle. This is a Smart-like vehicle with two seats and one can read that it drives autonomously, with no driver action being necessary.

Alas, an accident has been reported and Google said it bears "some responsibility" after the car struck the municipal bus in Mountain View, Google (2016). That means that the Google vehicle caused a crash. In that case, the car would be responsible, i.e. finally its manufacturer. However, also the driver and his responsibility need to be discussed.

Another example is a Tesla vehicle that crashed into a trailer. The driver did not react since he relied on automated driving and died as a consequence of the crash. In fact, the technical driving system of the Tesla was not able to detect the trailer. Then the question arises on the responsibility for the accident. Surely, the automatic systems needed permanent supervision by the driver and the question arises whether the driver was sufficiently instructed. Also, it needs to be discussed whether the driver had the possibility to stop the vehicle or take over the steer. This includes reaction time as well as features of the technical systems.

Some statistics might be done for the Tesla S at the time of the accident to illustrate the problems. One fatality has occurred after 210 Million km that have bene accumulated by the Tesla S model, see Focus (2016). This yields a rate of approximately $5\ 10^{-9}$ fatalities/km = $1/(2.1\ 10^{8}$ km). We need to admit that this "rate" has been computed from just one fatality and a serious statistical investigation would indicate that this figure contains a lot of spread.

What does this rate practically mean? We demonstrate this with some data on traffic accidents in Germany.

In 2013, in Germany private cars travelled 494 080 million kilometers, see Statista (2016). Let us now estimate the number of additional fatalities to be expected. We get $5\ 10^{-9}$ fatalities/km $*5\ 10^{11}$ km = 2500 additional fatalities per year, provided everyone would use the Tesla S model.

Note that a driver would have noticed the trailer and have reacted.

Moreover, in the statistics only possible additional accidents have been computed under the provision that the Tesla S model will be rolled out as it is now and that all drivers would behave as the one who died in the accident. On the other hand, accidents that would have been prevented by the Tesla S are taken into account.

This short computation shows the complexity and the problems connected to autonomous driving. Mainly this is caused by the very complex situation on the streets.

By the SAE (2016) and the UN (2017) the following levels have been defined.

- 0 No automation
- 1 Driver assistance
- 2 Partial automation
- 3 Conditional automation
- 4 High automation
- 5 Full automation

Detailed information on the levels is shown on the following Figure 1.

The currently present systems are mainly systems for assisted driving. The assistant helps in simple situations, however, the driver has always full responsibility. Examples are

- Distance assistant,
- Platooning,
- Lane assistant,
- Highway pilot for trucks.

A short glance on the approval systems shows the differences:

- Automated metros are approved according to EN 50126, EN 50128, EN 50129 and local laws on metros, that differ per country,
- Road vehicles are approved by a European approval based on ECE rules. In Germany this institution for approval is the KBA, in Netherlands this is the RDW,
- AGVs are not road vehicle and not a train, they are considered as automated machines and approval is according to Machine directive (2006) and IEC 61508 (2011).

A new law for homologation of road vehicles in Germany allows automated driving in specific cases—note that this is not assisted driving—but driver must be able to overrule the technical system.

This is in line with Convention (1973) on Road Traffic, which says:

- article 8,1: "Every moving vehicle or combination of vehicles shall have a driver",
- article 8, 3: "Every driver shall possess the necessary physical and mental ability and be in a fit physical and mental condition to drive.",
- article 8, 5. "Every driver shall at all times be able to control his vehicle or to guide his animals."

Currently, these principles are implemented in the law of the countries.

What does this mean for automated driving? The requirement "… driver must always be able to overrule the system" leads to the following requirements:

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	*Monitoring* of Driving Environment	Fallback Performance of *Dynamic Driving Task*	System Capability *(Driving Modes)*
***Human driver* monitors the driving environment**						
0	No Automation	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	Driver Assistance	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes
2	Partial Automation	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/ deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	**System**	Human driver	Human driver	Some driving modes
***Automated driving system* ("system") monitors the driving environment**						
3	Conditional Automation	the *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	**System**	Human driver	Some driving modes
4	High Automation	the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	**System**	Some driving modes
5	Full Automation	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	**All driving modes**

Figure 1. Overview of automation levels.

1. Technical possibility: brake, accelerator pedal, steering wheel must have a priority over automatic systems. Must be implied in safe controllers (ASIL C or ASIL D)
2. Driver must be present and vigilant, i.e. no sleeping, no messaging on the smartphone, no games including Pokémon GO etc.
3. Driver must have the necessary time to react. If there is a failure of the system or an unwanted reaction.

In fact, the last point leads to the following requirements for automatic driving.

- Braking: braking by automatic systems must be with a smaller acceleration than the driver could apply, the difference in accelerations (vehicle, driver) must still allow for a reaction time of the driver (braking curves),
- Steering: the distance from dangerous objects (other vehicles, border of the lane etc.) must be large enough to allow for drivers reaction, together with a limit of the steering angle. This might lead to speed restrictions.
- Perhaps the driver needs special training.

Figure 2 shows an example of a brake curve. Speed (m/s) versus distance is shown. There are two curves, one for automatic braking (deceleration 3 m/s**2) versus braking by driver (5 m/s*2), where a reaction time of 1.3 s has been taken into account for the driver. The initial speed is 20 m/s.

In this example, the driver is still able to come to a standstill in time, if he detects that the automatic system fails to brake. Of course, the driver must react and be able to react with 1.3 s.

For steering, similar requirements must be taken into account: Driver must have necessary reaction time. This reaction time depends on the distance to shoulder or adjacent lane, the speed and the reaction of the system. The latter includes maximal angular velocities and accelerations with which the system might show a faulty reaction.

The current technical solutions are supported by the following existing equipment:

- Different controllers or safe computers are available that are qualified according to up to ASIL D/SIL 4,
- Sometimes even “intelligent sensors” with a SIL available.
  - Different, diverse sensors (no SIL), which are cross-validated by the safe computer. Examples of such sensors are cameras, lasers, radar, infrared, ultrasonic etc.
- Multiple, diverse actors; safety relays as electric actors, the use of proven mechanical systems is also possible.

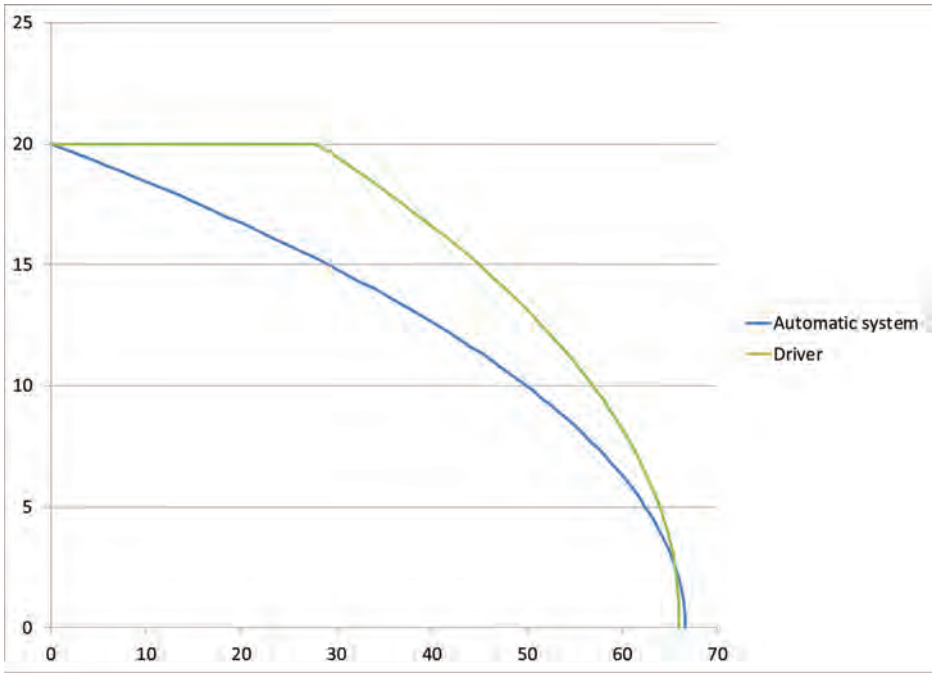


Figure 2. Example for a brake curve.

3 POSSIBLE NEXT STEPS

Based on the current status one can imagine the following future steps for road vehicle.

- Safe guidance (lane keeping) could be implemented, e.g. using differential GPS together with good update service of precise maps. All work on the road and all temporarily blocked roads need to be present on these maps.
- Stopping before traffic lights enforced by a wireless transmission of information between traffic lights and vehicles. Nevertheless, the driver needs to watch out for violators, e.g. cyclists even if he has a green lights.
- Speed limit enforcement, e.g. the speed limit is transmitted in a wireless manner form a sign broadcasting the speed limit or the sign is read by a camera, alternatively a map is used as source.
- Handling of simple traffic situations as e.g. on motorways following the lane, without overtaking maneuvers.
- Vehicles on separated areas and on separated road networks.

Further development leads to a following scenario, which include:

- The road or lane might be separated by two fences forming a controlled environment and on this environment a vehicle can run automatically, with steering, braking, driving implemented according to ASIL D.
- Vehicles drive with very short distances using platooning.
- At certain places entry and exit to this network of roads is allowed. There, the driver takes over the automatic vehicle and drives it manually to the destination.
- The necessary information as maps, position, speed limits, communication with other automated vehicles would be implemented on the vehicle, rather than on the road.
- The infrastructure would be rather cheap, consisting of the road and fences. Comparing this with a railway, the infrastructure is more flexible, no signals, no switches, no ballast and sleepers are necessary.

In all these cases, the relevant technical systems would need to be safe life systems with a safety level up to ASIL D/SIL 4.

A safe life system is a system, in contrast to a fail-safe system, does not switch itself off in case of a failure, but where the safety function is ensured even in case of one (or sometimes several) failures.

The safety integrity levels (SIL/ASIL) are defined in standards for functional safety. IEC 61508 and EN 50129 define SIL 1 to SIL 4. ISO 26262 defines the automotive SIL (ASIL) A to D.

The SIL/ASIL consists of two essential requirements:

- Maximum tolerable rate of dangerous failure which cannot be exceeded
- Measures against systematic failures (verification, traceability of requirement, specific techniques)

Regarding future development, also possible problems need to be considered, that an automatic or autonomous vehicle driving on the road need to face to become comparable with a human driver. First of all, such a system needs to distinguish objects as persons or animals from unmoving objects. Another example would be to distinguish vehicles on high wheel from bridges etc. Another problem is that sometimes intentions of a person or animal need to be guessed: does the person or the animal intend to cross the road and step on the road? A typical example would be a child with a ball standing on the sidewalk, having dropped the ball and this has moved on the street. There are a lot of such tasks would require intelligence and one would tend to use artificial intelligence for such a task.

Assume now that artificial intelligence should be implemented for autonomous driving. Then requirements for SIL 4/ASIL D would need to be implemented in full rigor in the software and the hardware. On the other hand, the algorithms for artificial intelligence are voluminous and complex. If then e.g. traceability needs to be shown from a requirement as e.g. "The algorithm must distinguish human beings from other objects" one might imagine the complexity of such a task. This would only be one requirement. The entire complex of requirements to the software would have to take into account a lot of driving situations, in the environment etc. If the algorithm is a self learning algorithm, one needs to ensure that it has learned in a certain time enough and this must be proven in the light of the standards /IEC 61508/and/or/ ISO 26262/. Another possibility would be to use a proven in use argument and accumulated 3 10^9 hours in service, see /IEC 61508/part 7 annex D. With 600 hours of driving that would mean to have 5 000 000 vehicles driving an entire year under controlled circumstances, i.e. with trained drivers that can override the system and that would also register all events—or the vehicle has to do this. One can decrease the number of vehicles by increasing the number of driving hours per year, e.g. up to 6,000, which would mean driving in shifts. Nevertheless, still 500,000 vehicles would be necessary. In addition, each change of the software would require to repeat this approval process

The conclusions is that solutions for the safety relevant software must be simpler, without

guessing intentions etc. in order to overcome these problems. Artificial intelligence would be good for assistance systems.

4 CONCLUSIONS

In this paper we have provided some considerations on automatic (or autonomous) driving for rail and road vehicles. It turns out that for road vehicles, the environment is much more complex than for rail vehicles. Therefore, the experience from e.g. automatic metros cannot be directly used.

Most of the existing systems are either pure assistance systems or they are dedicated to simplified traffic situations

It has to be expected that the first safe solutions for autonomous driving would come for situations with a simplified environment, especially where the environment is controlled or even adapted to the task of autonomous driving. Here, a special solutions are AGV (automatic guided vehicles) that are just moving in an environment fully adapted to them, but not on an open road.

REFERENCES

Bouwman R. Schäbe, H., Vis, H. (2009), Application of safety principles for a guidance system in public transport, *ESREL 2009, Proceedings Reliability, Risk and Safety*, vol. 3, p. 2275–2278.

Breitinger M. (2016), Kabinett erlaubt teilautomatisiertes Fahren, http://www.zeit.de/mobilitaet/2016-04/autonomes-fahren-gesetzentwurf-verkehrsrecht-alexander-dobrindt, published 13.4.2016, retrieved on 19.10.2017.

Convention 1973 Convention on Road Traffic, 8.11.1968, European Additional Treaty from 1.5.1971 and Protocol 1.3.1973.

Daimler 2017 The Mercedes-Benz Future Bus The future of mobility, *https://www.daimler.com/innovation/autonomous-driving/future-bus.html*, retrieved on 19.10.2017.

EN 50126 Railway applications—The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS), 1999 with correction 2011.

EN 50128 Railway applications—Communication, signaling and processing systems—Software for railway control and protection systems, 2011, correction 2014.

EN 50129 Railway applications—Communication, signalling and processing systems—Safety related electronic systems for signalling, 2003.

Focus (2016) Todesfall im selbstfahrenden E-AutoUS-Verkehrsaufsicht prüft Teslas "Autopilot", *http://www.focus.de/auto/elektroauto/todesfall-im-selbstfahrenden-auto-us-verkehrsaufsicht-prueft-teslas-autopilot_id_5687341.html*, 1.7.2016.

Frog 2017, Website, www.frog.nl, retrieved on 19.10.2017.

Google car (2016) Google self-driving car hits public bus near Mountain View headquarters *http://www.mercurynews.com/2016/02/29/google-self-driving-car-hits-public-bus-near-mountain-view-headquarters/*, retrieved on 19.10.2017.

IEC 61508 Functional safety of electrical/electronic/programmable electronic safety-related systems, part 1–7, 2011.

ISO 26262 Road vehicles—Functional safety, parts 1–10, 2011.

Machine Directive (2006) DIRECTIVE 2006/42/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast).

Nahverkehrspraxis (2017), Weltpremiere: Daimler Buses präsentiert autonom fahrenden Stadtbus", *http://www.nahverkehrs-praxis.de/news/nahverkehrspraxis-top-news/article/weltpremiere-daimler-buses-praesentiert-autonom-fahrenden-stadtbus/*, retrieved on 19.10.2017.

SAE 2016 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, SAE J3016, September 2016.

Statista (2016) Fahrleistung der vorhandenen Personenkraftwagen in privaten Haushalten in Deutschland in den Jahren von 2005 bis 2015 (in Millionen Kilometer), *http://de.statista.com/statistik/daten/studie/484040/umfrage/fahrleistung-pkw-in-privaten-haushalten-in-deutschland/*, retrieved 19.10.2017.

UITP 2017, *International Association of Public Transport.* "A global bid for automation: UITP Observatory of Automated Metros confirms sustained growth rates for the coming years". *Belgium*, retrieved 19.10.2017.

UN 2017 Economic Commission for Europe, Inland Transport Committee, World Forum for Harmonization of Vehicle Regulations, *Consolidated Resolution on the Construction of Vehicles*, (R.E.3), Revision 6, 11.7.2017.

Vogelpohl, T., Vollrath, M. (2016) UDV (Unfallforschung der Versicherer) *Takeover times in highly automated driving Compact accident research*, Nr. 57, 07/2016.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Probabilistic analysis of faults affecting multiple trains of the electrical power supply system of nuclear power plants

B. Brück, G. Gänßmantel, A. Kreuser, C. Müller, E. Piljugin & J.C. Stiller
Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) gGmbH, Germany

ABSTRACT: Faults simultaneously impairing multiple redundant trains of the electrical power supply system of NPPs recently received growing attention by the nuclear community. This was triggered by events at several different NPPs including Byron in the U.S. or Forsmark in Sweden. Such events have generally not been included in PSAs of NPPs yet. Therefore, GRS has initiated a research project aiming at a comprehensive and in-depth analysis of events characterized by fault states of multiple trains of the electrical power supply system (including but not limited to open phase conditions) and at the development of modelling and quantification methods to include them in PSAs. The project consists of different interacting efforts. Firstly, the possible causes of faults affecting multiple trains of the electrical power supply system and their consequences are assessed from an operating and modelling perspective. The second step comprises the development of a detailed dynamic model of the electrical power supply system of a German PWR and the investigation of the cause and the propagation of such faults. Then, a current PSA model of a German PWR is extended to allow for the modelling of the phenomena identified in the previous steps. This includes adding relevant equipment not modelled before and new failure modes of equipment already modelled. The additional reliability parameters and frequencies of initiating events required to quantify the extended PSA model are estimated. Finally, the additional failure mechanisms considered in the extended PSA model are evaluated quantitatively.

1 INTRODUCTION

Faults simultaneously impairing multiple trains of the electrical power supply system of Nuclear Power Plants (NPPs) have recently received growing attention by the nuclear community (Brück 2016). This was triggered by events that involved so-called asymmetrical faults at several different NPPs where such faults occurred. An asymmetrical fault results from the degradation (e.g. an interruption) of one or two of the three phases in a three-phase alternating current system. For example, at the Byron NPP in the U.S., asymmetries in the power supply system arose from a single failure of an insulator in the switchyard of the plant. The asymmetry failed to cause the Reactor Protection System (RPS) to initiate the isolation of the emergency bus bars and the operation of the emergency diesel generators. As another example, at the Forsmark NPP in Sweden, the failure of one pole of a breaker to open led to an open phase condition that was also not detected by the RPS. In both cases the electrical consumers remained connected with the fault and were exposed to an asymmetric voltage supply, leading to unavailabilities and even destruction of electrical equipment.

Such events have generally not been included in Probabilistic Safety Analyses (PSAs) of NPPs yet. Therefore, GRS has initiated a research project aiming at a comprehensive and in-depth analysis of events characterized by fault states of multiple trains of the electrical power supply system, including—but not limited to—open phase conditions, and at the development of modeling and quantification methods to include them in PSAs.

The electrical power supply system is particularly susceptible to faults affecting multiple trains since during normal power operation there is no separation between the redundant trains. As shown in Fig. 1, failures that occur on or above the generator bus bars will affect all underlying bus bars simultaneously.

2 PROJECT

This project consists of different interacting efforts:

Initially, the possible causes of faults affecting multiple trains of the electrical power supply system and their consequences are assessed from an operating and modeling perspective. To achieve the project goals, first a detailed analysis of international

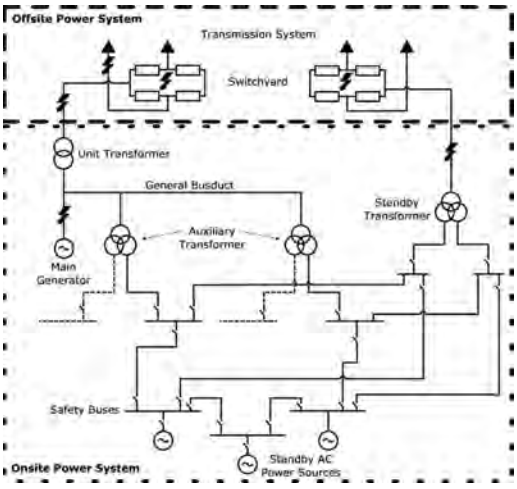


Figure 1. Typical electrical systems of a NPP (inspired by IAEA 2016).

operating experience is carried out with respect to actual and potential faults affecting multiple trains of the electrical power supply system of NPPs, complementing the German operating experience that had also previously been analyzed by GRS with respect to this topic in the course of the general monitoring and analysis of German operating experience (Mildenberger 2016).

The second step comprises the development of a detailed dynamic model of the electrical power supply system of a generic modern German Pressurized Water Reactor (PWR) and the investigation of the cause and the propagation of such faults.

Then a current PSA model of a German PWR is extended to allow for the modeling of the phenomena identified in the previous steps. This includes adding relevant equipment not modeled before and new failure modes of equipment already modeled.

The additional reliability parameters and frequencies of initiating events required to quantify the extended PSA model are estimated. New approaches and procedures to achieve this are developed as needed.

Finally, the additional failure mechanisms considered in the extended PSA model are evaluated quantitatively.

3 OPERATING EXPERIENCE ANALYSIS

The events at the Byron and Forsmark NPPs highlighted the importance of the grid connections and the associated equipment for the reliability of the plants' safety system. These events also revealed the importance of a systematic evaluation and analysis of their operating experience; while the physical and electro-technical effects that led to the failures were all well-known and well understood in theory, this theoretical knowledge was not used in the design of the safety system of NPPs worldwide until operating experience made these problems obvious.

Therefore, an evaluation of operating experience with a special focus on effects that might impair multiple trains of the electrical power supply system is performed. By doing so, two targets are pursued:

- Identification of failure mechanisms that might lead to multi-train impairments and that are not yet covered by the design assumption of the plant. Also, events where so far no (multiple) failures have been observed but where the effective failure mechanism may cause such failures in case of other circumstances are relevant.
- Development of failure scenarios that describe how such failure mechanisms would affect modern German KWU type NPPs. With the development of the scenarios it is intended not only to "copy" the event to the KWU type plant, but also to develop variations of the actual event.

3.1 *Events with asymmetrical faults*

As a first step, the systematic evaluation of international operating experience with a focus on asymmetrical faults, which was conducted by GRS after the events in Byron and Forsmark, was taken as basis to develop failure scenarios. This evaluation (NRC 2007) revealed that such faults can be observed regularly in NPP operating experience. Ten events were revealed where the active grid connection of the plant was affected by an asymmetrical fault. In four events such faults were discovered in the standby grid connection. The identified events where active grid connections were affected are presented in Table 1.

The systematic analysis of asymmetric faults showed that a single failure mechanism might lead to various different failure scenarios.

In case of an asymmetrical failure event, the following set of features that had a significant influence on the extent of the degradation of the onsite power system could be identified:

- Type of failure: Failures of one and of two breaker poles have been observed in operating experience and need to be analyzed.
- Location of the failure: Asymmetrical failures may occur in the main grid connection, the auxiliary grid connection and the generator bus duct, each with different consequences. In case of grid side asymmetries, the different distances between the location of the failure and the plant

Table 1. Events with asymmetrical faults.

Date	Plant	Failure cause
1994-05-13	Kalinin	Collapse of a transformer duct, OPC in one phase
1997-02-25	Balakovo	Unintended closure of a single breaker pole
2001-03-31	South Texas	One breaker pole in the switchyard failed to close
2005-11-11	Koeberg	One breaker pole in the switchyard failed to close
2006-07-26	Vandellos	Mechanical failure of a disconnector
2007-05-14	Dungeness-B	One pole of a HV-transformer breaker failed to close
2012-01-30	Byron	Collapsed insulator caused a line interruption
2012-12-01	Bruce	Mechanical line failure during severe weather (storm)
2013-05-30	Forsmark	Failure to open on command of a single breaker pole
2014-04-27	Dungeness-B	Open breaker pole in the switchyard

have to be considered as well as parallel grid connections that are not impaired.
- Neutral point treatment: Operating experience has shown that the treatment of the neutral point of the main or auxiliary transformers has a crucial effect on the propagation of a grid side asymmetry into the plant.
- Load of the onsite power system: Both the load and its characteristics (inductive or ohmic) have an influence on the asymmetry and need to be evaluated carefully.

In total, more than 500 combinations of features can be derived from the list above. Currently, methods are being developed how this number can be reduced to a practicable amount of failure scenarios. Once this reduction is achieved, the identified scenarios will be analyzed with the simulation model described in Section 4.

3.2 *Extended scope*

Beside the asymmetrical faults, several other phenomena that might affect multiple redundant trains of the electrical power supply system are already known from operating experience, both from the International Reporting System for Operating Experience (IRS) and from German operating experience. Among these phenomena are the Forsmark event of 2006-07-25 (NRC 2007), where combined voltage and frequency fluctuations in the 400 kV grid caused multiple impairments of Uninterruptible Power Supply (UPS) units necessary for the startup of the emergency diesel generators (EDGs), and an event in a German NPP (RSK 2015) where four inverters (each in a separate redundant train) failed because of a single failure in a 660 V breaker of a residual heat removal pump.

In the light of these insights it was concluded to extend the scope of the analysis from asymmetrical faults to all failures that are capable of affecting more than one of the redundant trains of the electrical power supply system. This includes non-redundant components or systems like the grid connections, the generator or the generator bus duct, but also redundant components inside and outside of the safety systems that have caused impairments in more than one train during a failure event.

Since German operating experience with NPPs is limited to about 800 reactor years, it was decided to extend the scope of the analysis. Based upon an evaluation of the technical comparability of the plants, the accessibility of the necessary information and the amount of available data, it was decided to use the operating experience of the U.S. NPP as it is provided through the Licensee Event Reports (LER) as additional information source. Although all these events have already been analyzed in depth by the U.S. NRC, it was concluded that additional insights from the events could be gained by an analysis focused on possible effects on modern KWU type plants, which have a safety system that relies primarily on redundancy rather than diversity.

3.3 *Methodology*

Effects with the potential to affect multiple redundant trains simultaneously are rare since extensive precautionary measures are taken to avoid such events. Therefore, a substantial amount of operating experience has to be analyzed to identify some of these types of failures. To achieve this, all 3466 LERs with events in PWRs and BWRs from the beginning of the year 2000 to the end of 2009 were included into the scope of the analysis. To cope with this high number of events in a reasonable amount of time, a four stage process was developed to identify the relevant fault scenarios:

1. Initial screening of the events; based upon an event summary, all available events are screened to filter out all those events that are obviously not relevant for a further analysis.
2. Thorough analyses of the remaining events to further reduce the number of events; at this stage all information included in the LER is used for the assessment.
3. The remaining events are analyzed and described in depth by taking into account all available information.

4. Based upon the results achieved in step 3, failure scenarios (see Section 3.1) are developed.

The first step of the process resulted in 250 potential events, which were reduced to 29 events in step two.

They cover a wide range of effects like external impacts (severe weather or grid fluctuations), component failures inside the plant or in the associated switchyard, or events due to human error. Therefore, it may be expected that a comprehensive set of relevant failure scenarios will result from this effort. Up to now, three scenarios have been developed.

4 DETAILED DYNAMIC MODEL

A generic model of the auxiliary power system of modern German KWU type NPPs has been developed using the software NEPLAN (NEPLAN), see Fig. 2 and Fig. 3. Using this model, different calculations and analyses can be performed, such as load flow calculations, short circuit calculations, harmonic analysis, and dynamic simulations.

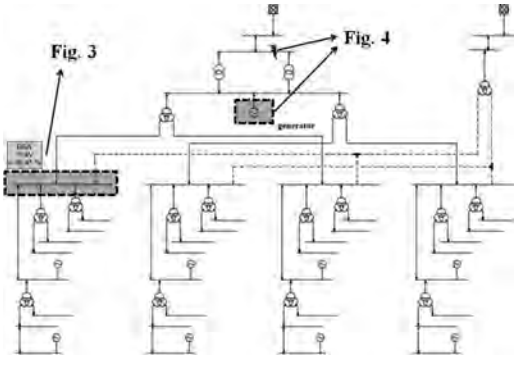


Figure 2. Generic model of the auxiliary power system of a German NPP. On each bus bar, several individual electrical consumers are modelled (see Fig. 3).

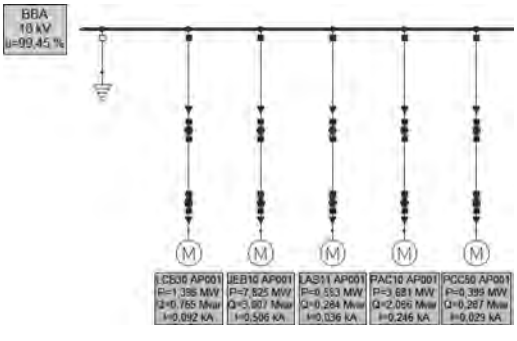


Figure 3. One of the four main 10 kV bus bars with electrical consumers.

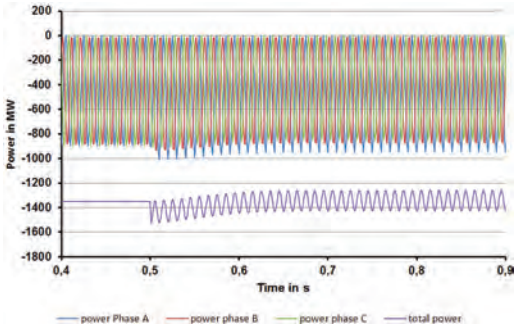


Figure 4. Total power and power of the three phases at the generator terminal after a single-phase interruption occurring at t = 0.5 of the connection to the grid (see Fig. 2).

The model currently consists of 733 elements (including 114 asynchronous machines as consumers) so far and is currently being further developed. It is already suitable for estimating the impact of different scenarios on the NPP's safety system.

As an example, Fig. 4 shows the voltage and current of the generator for all three phases as a function of time for a three phase disturbance as marked in Fig. 2. The next step will be the detailed investigation of the scenarios to examine the failure modes and effects.

5 PROBABILISTIC MODEL

First analyses have shown that the appropriate modelling of realistic single phase failure scenarios and other scenarios simultaneously impairing multiple trains of the electrical power supply system in a PSA requires extensive augmentation and modification of present PSA models. This comprises modeling the complex impacts of such phenomena on the electrical equipment, including parts of the electrical power supply system not important to safety, in the PSA model and adding additional failure modes of electrical components already modeled. To efficiently and systematically integrate such modifications into existing PSA models, GRS has developed and continuously improves the software tool "pyRiskRobot" for modifying complex fault tree topologies in an automated and traceable manner (Berner 2017). This tool will facilitate modifying and enhancing the PSA model.

6 QUANTIFICATION

To quantify the model, the frequencies of the new initiating events and additional reliability parameters need to be estimated including e.g. failure

probabilities and failure rates for electrical equipment that is or has been exposed to single phase failure conditions. While the extension of the PSA model is expected to be more or less straightforward, quantification will pose a major challenge since reliability parameters for the equipment under the respective conditions or analyses of relevant operating experience that would be suitable to base estimations of these parameters on are not readily available.

To achieve this nonetheless, information from existing databases, from operating experience and expert knowledge will be utilized. Models, methods and procedures to extract and combine the available information will be developed and evaluated as needed. This may include modeling of the emergence of relevant initiating events by fault trees in order to estimate their rate, modeling of the failures of components under appropriate boundary conditions as a result of failures of their piece parts by fault trees, quantitative analysis of operating experience using Bayesian statistical methods, and quantitative assessments by experts.

As first step, the rates of the initiating events single-phase and dual-phase failures in the active grid connection have been assessed utilizing the operating experience analysis described in section 3. The rate of single-phase failures in the active grid connection is comparable to the rate of small Loss of Coolant Accidents (LOCAs) while the rate of dual-phase failures is approximately one order of magnitude smaller.

7 CONCLUSIONS

Faults affecting multiple redundant trains of the electrical power supply system of NPPs—including but not limited to open phase conditions—may pose a significant threat to the safety of NPPs. Such failures have generally not been appropriately considered in PSAs despite the fact that the rate of such events is comparable to small LOCAs. An on-going project of GRS to research this subject comprises the systematic analysis of national and international operating experience, the development of a detailed dynamic model of the electrical power supply system and the extension and modifications of an existing PSA model and the quantification of the enhanced model. The advancements made so far suggest that substantial new insights will be gained in the frame of this project. However, the estimation of the reliability parameters and rates of initiating events needed to quantify the enhanced PSA model will still pose a challenge.

REFERENCES

Berner, N. & Herb, J. 2016. Generic framework for the automated integration of impacts from hazards in PSA models, pp. 2645–2649, *Risk, Reliability and Safety: Innovation Theory and Practice,* Walls, Revie and Bedford (Eds.).

Brück, B. et al. 2016. Implications of Open Phase Conditions in the Electrical Grid Connections for the Safety System of NPPs, Munich: EUROSAFE 2016.

IAEA Safety Standards Series No. SSG-34 *Design of Electrical Power Systems for Nuclear Power Plants*, International Atomic Energy Agency (2016).

Mildenberger, O. (Ed.) 2017, *Vertiefte Untersuchungen von Betriebserfahrungen aus Kernreaktoren,* GRS-458 (in German with an English abstract), Cologne: Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) gGmbH.

NEPLAN, NEPLAN AG, Küsnacht, Switzerland, http://www.neplan.ch/.

NRC Information Notice 2006-18, Supplement 1: *Significant loss of safety-related electrical power at Forsmark Unit 1 in Sweden,* U. S. Nuclear Regulatory Commission (2007), https://www.nrc.gov/reading-rm/doc-collections/gen-comm/info-notices/2006/in200618sup1.pdf.

RSK-Stellungnahme *Scheibenübergreifende Unverfügbarkeiten aufgrund elektrischer Kopplungen zwischen redundanten Scheiben des Notstromsystems deutscher Kernkraftwerke*, 472. Sitzung der Reaktorsicherheitskommission 2015-01-14 (in German), http://www.rskonline.de/sites/default/files/reports/epanlagersk472hp.pdf.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

A scenario-based risk analysis oriented to manage safety critical situations in autonomous driving

A. De Galizia & A. Bracquemond
VEDECOM Institute, Versailles, France

E. Arbaretier
APSYS—Airbus Group, Elancourt, France

ABSTRACT: Risk analysis is very useful for identifying hazardous events affecting the functional limits of complex systems as autonomous vehicles. To ensure the robustness of an autonomous vehicle architecture, new approaches should be investigated providing the dimensioning parameters related to functional scenarios. Research is currently ongoing at the VEDECOM Institute to identify and analyze safety critical situations including accidents. For this reason, a set of functional scenarios has been defined as an abstraction of real life driving situations. The aim is to explore how autonomous vehicles evolve and behave under various environmental and traffic conditions. Thus, this paper presents a qualitative risk analysis approach taking into account the evolution of a scenario by means of the concept of transition between successive scenes. Then, combinations of hazards are identified potentially leading to safety critical situations. Finally, the feasibility of our proposal is shown on an application case examining a level 4 autonomous vehicle behavior in high traffic driving.

1 INTRODUCTION

During the last decade, focus of research in the automotive industry has shifted to the development of highly if not even fully automated (autonomous) driving (Bengler et al. 2014). Recent progress in this field showed that nowadays is possible to provide the driver with useful assistance systems, *e.g.* lane departure warning (Dickmanns 2002), lane change assistant (Ruder, M. et al. 2002), Adaptive Cruise Control (ACC), or even autonomous driving over long distances on highways (Dagli et al. 2004).

Nevertheless, in order to achieve highly robust autonomous driving systems, there are still challenging use cases that must be mastered (Schmidt et al. 2015, Geyer et al. 2014). In fact, mastering these cases is not only substantial for the Society of Automotive Engineers (SAE) level 4 and 5 automated vehicles (Donges 1999), which are capable of sensing its environment and navigating without human inputs, but relatedly on the rate of triggering driver take-over requests in case of level 3, conditional automation. Outside of these areas or circumstances, the vehicle must be able to safely abort the trip, i.e. park the car, if the driver does not retake control. Increasing the automation level requires remarkable improvements on the existing perception, prediction and planning algorithms but also on the system architectures (Weiss et al. 2004). Although the former have been an active field of research, focus on the system (functional) architecture has been limited so far. In our point of view, system architecture, and in particular its behavioral planner layer, is a key element for future autonomous vehicles (Bertolazzi et al. 2004, Bonic et al. 2017).

In order to set up a robust decision module in an autonomous systems, a number of functions have to be realized in an integrated architecture. As processing tasks for perception, situation assessment, behavioral module and actual control of the vehicle have tight constraints, it is necessary to focus on the environment around (Chan et al. 2004, Dickmanns 2003, Bertozzi et al. 2000).

Developing a safe functional architecture is one of the main objectives of the 'Robustness of architectures and systems' program at VEDECOM (Bonic et al. 2017).

Today, risk and safety analysis has not sufficiently addressed the problem of explore the behavioral and decision module of an autonomous system architecture taking into account the surroundings (Donges 1999, Figlewski & Levich 2002).

In this sense, risk analysis should provide for relevant design insights concerning the overall robustness of an autonomous vehicle, *i.e.* its faculty to adapt and operate properly under a large

diversity of operating conditions (e.g. infrastructure, weather, illumination, traffic etc.). However, most of existing risk analysis approaches are rather specialized in assessing system reliability and usually they focus on system and components failure modes (Kontio & Basili 1997). Moreover, in these approaches information is more or less used directly on a quantitate level, with an uncertainty that have to be considered.

In this paper, we present a risk analysis approach which aims to qualitatively identify hazardous patterns and by this way the underlying critical situations. In that way, we want to focus on the limits of behavioral performance of a SAE level 4 autonomous vehicle. The aim is to address the robustness of the autonomous system architecture by focusing on the evolution of a driving scenario.

Then, the risk is manly seen as discussed in (Polychronopoulos et al. 2004) as the approaching critical situations resulting from the combination of several hazardous events which might endanger the autonomous vehicle and its passengers or other traffic participants. Concretely, the paper addresses risk analysis qualitatively and thus allows for describing the abstract critical situations in a comprehensible way. Based on an approach oriented to the dynamic modelling, it allow for integrating risk analysis insights on an experimental evaluation (simulation) platform for a scenario-based design of autonomous system architecture (Go & Carroll, 2004).

The rest of the paper is structured as follows: Section 2 gives an overview of the global framework on which our risk analysis methodology relies on. Section 3 introduces the conceptual formalization of scenes and scenarios along with some notions relating with dynamic modeling of autonomous driving. In Section 4 an application case is proposed. The paper close with some conclusions and an outlook for further developments to adapt the proposed technique to specifics for autonomous systems design.

2 FROM REAL WORLD DRIVING TO SAFETY CRITICAL SCENARIOS

The risk analysis approach that we propose is based on a global framework allowing to assess risks in critical driving situations generated by the occurrence of *hazardous events* combined with critical operational modes. We make the choice to use the adjective *hazardous* since it makes the tradeoff between the semantic usage specific to dependability studies and the standardized expressions used in a research area such as that related to autonomous driving.

As illustrated in Figure 1, this risk analysis methodology has been proposed to the need of analyzing nominal functional scenarios designed starting

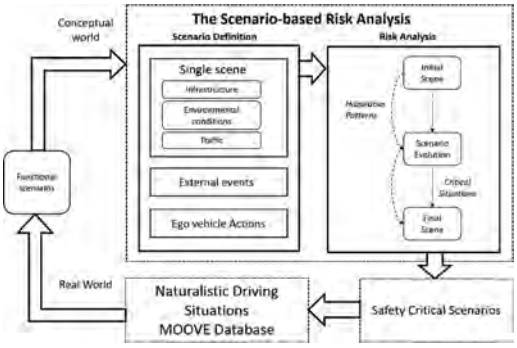


Figure 1. Scheme of the global approach.

from real world driving data collected in the context of the MOOVE project (Bonic et al. 2017).

The parameterization of the functional scenarios makes it possible to do queries in the MOOVE database and search "virtual" aggravating situations in driving records.

In our risk analysis approach, we focus on the hazardous events leading to major risks situations linked to safety issues affecting the autonomous vehicle performance (behavioral and decision).

After this introduction to the main concepts and the framework of development in which this method fits, in Section 2 we present in detail the different parts that compose it and the underlying approach aiming to make the link with dynamic modelling.

3 A SCENARIO-BASED RISK ANALYSIS FOR AUTONOMOUS DRIVING

3.1 *Defining the main terms*

In the literature several definitions exist which are used in different application fields. We have considered those that fit better with the choices made in the context of the MOOVE project. Thus, the following terms have been defined as they constitute the basic components of our approach.

First of all, we need to define the term scene as below:

> *A scene describes a snapshot of the ego vehicle and the surrounding including the infrastructure, the environmental conditions as well as all the interacting traffic participants, and their relationships with the ego vehicle.*

The objects in the scene can be characterized in different ways:

– Dynamic or static (*i.e.* "potentially" mobile or "always" static objects);

– Environmental elements;
– Etc.

Overall, the *scene* describes the current state of the system according to the *traffic conditions* (position, speed, etc. of the ego and the other vehicles), the *environment* (weather, road surface, visibility, etc.) and *infrastructure* (type of road, number of lanes, slope, presence of working zones and mobiles objects, etc.).

The second term which has to be defined is *scenario*:

> *A scenario describes the evolution of the system by a succession of scenes. The (temporal) succession starts with an initial scene and ends with a final scene.*

The mechanisms for moving from one scene to another are mainly of two types:

– *Actions, i.e.* the ego vehicle state changes;
– *External events, i.e.* other vehicles or the environment state changes.

After having formalized the current situation (initial scene), how it evolve and the final state of the ego vehicle, the nominal development of the scenario can be modified or degraded by adding sequential hazardous event (Bengler et al. 2014). These can be maneuvers of traffic participants interacting with the ego vehicle can be considered along with an unsafe behavior of the ego itself. This allow for identifying *critical patterns*, and by consequent, risks of a missed control about deviations of a specific scenario.

So, we have to define two last concepts, *hazardous event*:

> *A hazardous event is an event that may increase the risk likelihood or severity. It is any condition that can degrade the development of the ongoing scene towards a negative evolution of the scenario, with more or less serious consequences on the achievement of objectives directly related to safety.*

and *critical pattern*:

> *A critical pattern is a set of hazardous events which combined with actions or external events leads to a critical situation.*

Based on the nominal scenario, we have to deal with a large combinatorial explosion concerning the scenario branching as in sequential critical patterns (Lattner & Herzog, 2004). This combinatorial explosion essentially depends on:

– The *actions* and/or *external events* that may take different values and occur at different times;
– The number of *hazardous events* and their combination to produce *critical patterns*.

This representation of a driving scenario is naturally oriented to the dynamic modelling of an autonomous system functional architecture. Next section provides for a further formalization supporting the dynamic modelling.

3.2 *Formalization of transitions for dynamic modelling*

Following the definition of the concepts that we have proposed in previous section about this risk analysis approach, *i.e.*:

– *Scene*, which in turn is based on the 'infrastructure', 'environmental conditions' and 'traffic' descriptors,
– *Scenario*, which start with an initial scene, evolves and ends with a final scene,
– *Critical patterns,* which are introduced in a scenario by considering a combination of hazardous events, then we give some elements to formalize the evolution of a driving scenario for dynamic risk assessment. This formalization is intended to provide a way to deal with the dynamic modelling of autonomous driving functions.

To do that, we have referred to the concept of *transition* which is well known in graph models for risk assessment, as Petri Nets (Reschka et al. 2015; Ulbrich et al., 2014). So, in this risk analysis methodology a transition is defined as follows:

> *A transition is any action of the ego vehicle or external events occurring during the evolution of the scenario and leading to the successive scene.*

Otherwise said, scenes are connected and triggered by transitions. Graphically, a driving scenario can be then represented as a discrete event dynamic system. As Petri nets, it is a directed bipartite graph, in which the nodes represent transitions (i.e. events that may occur, represented by bars) and places (*i.e.* conditions, represented by circles) (see Figure 2).

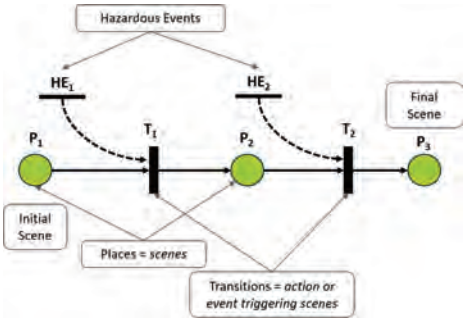


Figure 2. Illustration of the evolving driving scenario as a state-transition system.

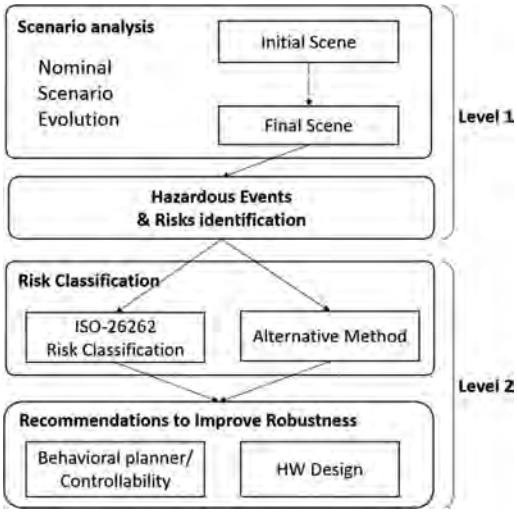


Figure 3. Illustration of the proposed risk analysis methodology: level 1 focuses on the scenario evolution and how harzards can lead to critical situations; level 2 allows for identifiyng risks bsed on the critical situations and classify them.

To resume, the risk analysis methodology proposed in this paper can be depicted as a top-down approach (Matthaei & Maurer, 2015) (Figure 3).

In Section 4, an illustrative case is presented as application of the level 1. In particular, the analysis of a relevant scenario is discussed for the decision and behavioral module of an autonomous vehicle.

4 ILLUSTRATIVE CASE

As application cases, we tested the approach on a set of scenarios of interest in the context of the MOOVE project. Among these, a very interesting and illustrative one is the "Inconsistent yellow and white lane marking" scenario. It allows for showing the behavioral layer of the ego vehicle architecture in the presence of a working zone (WZ). In particular, it is a WZ characterized by two types of lane markings simultaneously: white marking for lanes before the WZ, and yellow marking for lanes after the WZ. The superposition of both white and yellow markings on the road makes lane detection problematic for the perception system. Consequently, the ego vehicle trajectory planner depends on tracking (*i.e.* following the target vehicle ahead). The absence, or in any case the loss of the target vehicle, significantly compromises the possibility for the ego vehicle to go beyond the WZ.

So, we proceed to identification of all the critical patterns potentially leading to critical situations. As illustrated in Fig. 4, level 1 proceeds by analyzing the scenario evolution as described above, *i.e.*:

- Initial scene;
- Evolution (*i.e.* transitions);
- Final scene.

This first part is followed by the identification of hazardous patterns and the critical situations.

Let's consider the nominal scenario evolution analysis. The scenario that we are considering starts with an initial scene, illustrated in Figure 4. The ego vehicle drives on the right lane and follows the green car ahead (the target vehicle). Let's name this latter T1.

The initial scene is described in our methodology as shown in Table 1.

An event occurring after the initial scene marks the evolution of the scenario. This event is the beginning of a working zone (WZ), as illustrated in Figure 5.

T1 detects the WZ and changes lanes.

Figure 4. Snapshot of the initial scene.

Table 1. Initial scene as analyzed in the methodology.

Level 1 – Initial scene	
	Description
Infrastructure	– Road with separate carriageways; – Two or more lanes for the same direction; – Dashed lines in the middle;
Environmental Conditions	Nothing to report (NR)
Traffic	– The ego vehicle (**E**) drives on the rightest lane at a constant average speed; – The vehicle (**T1**) in front of E advances at the same speed; – Other vehicles drive on the left lane.

Figure 5. Evolution of the scenario: An event occurs consisting on the presence of the WZ.

Based on this event, ego vehicle performs actions basically consisting in T1 tracking in order to go over the WZ (see Figure 6 below).

The scenario closes with the final scene (Figure 7).

Table 3 shows how the final scene is described in the risk analysis it before to proceed to

Figure 6. Evolution of the scenario: Ego vehicle do some actions in response to the events occurred in the scenario.

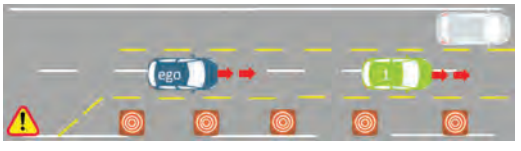


Figure 7. Final scene of the scenario: Action.

Table 2. Scenario evolution.

Level 1 – Evolution	
	Description
Infrastructure	– In the right-hand lane there is a working area indicated by warning signs; – New traffic lanes are marked in yellow on the ground, and white markings of initial lanes are still present.
Environmental conditions	Nothing to report (NR)
Traffic	– **T1** turns on the left indicators and follows the lines of the new yellow marking to go beyond the working zone; – **E** detects that **T1** changes lanes to the left, detects the new marking lines, turns on the left indicators and follows **T1**.

Table 3. Final scene.

Level 1 – Final scene	
	Description
Infrastructure	– Two tighter lanes are available
Environmental conditions	Nothing to report (NR)
Traffic	– **E** continues to drive behind C1 in the new lane on the right (yellow marking) and goes beyond the working zone.

Table 4. Hazardous patterns & critical situations.

Level 1 – Hazardous patterns & critical situations	
Hazardous patterns	Critical situations
1. **T1** suddenly changes lanes without turn signals AND E brakes too late;	2. Collision in the working zone;
3. **T1** stops abruptly;	4. Accident between **E** and **T1**/Exit of lane or road;
5. **E** does not detect or detects too late the work area (panels placed too close to the work) AND does not brake sufficiently before;	6. Accident in the work area with injured workers;
7. **E** brutally shifts to the right (bad weather conditions: e.g. rain, side wind, loss of grip: e.g. slippery ground).	8. Accident with one or more injured pedestrians.

identification of the critical patterns and the resulting critical situations (Table 4).

5 CONCLUSIONS

In this paper we presented a risk analysis approach adapted to the specifications of the autonomous vehicle. For a SAE level 4 autonomous vehicle the driver only provides a destination or navigation instructions.

Then, a robust behavioral planner should manage safely a large number of critical situations. Then, critical situations need to be well identified and mastered to deal with the design of the functional architecture of an autonomous vehicle. The aim is to evaluate the performance limits of a level 4 autonomous vehicles.

To identify these limits, in this paper we proposed a risk analysis methodology which focuses on a specific scenario and identify possible hazardous events. In that way, we tried to contribute to two main research axes relevant for autonomous driving:

1. Exploring the ego vehicle behavior qualitatively in functional scenarios and identifying the safety critical situations dimensioning the ego vehicle architecture without taking into account failure modes;
2. Mastering critical situations by proceeding with a contextual semantic adapted to autonomous vehicles operational behavior.

In order to validate these contributions, this methodology has been applied to one of the

functional scenarios defined in the MOOVE project basing on our real life driving database. Scenarios have been designed to explore the ego vehicle behavioral limits (safe functional) and identify critical situations which could face in real life. These critical situations arise by combining hazardous events. As hazardous events, we do not considered those related to failure modes but we focus on hazards arising from extreme or critical operational modes that no longer allow the ego vehicle to safely behave.

Based on the application case, some perspectives have been identified to extend this approach to fully deal with autonomous driving safety:

- On the one hand, in the first part of the analysis, the integration into the scenario analysis of failure modes concerning perception and control;
- On the other hand, in the second part, the authors would like to simulate a deeper discussion on the risk classification new exigencies for autonomous vehicles compared to the classical ISO-26262 standard (International Organization for Standardization, 2011).

REFERENCES

Bengler, K. et al. 2014. Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent Transportation Systems Mag.*, vol. 6, no. 4, pp. 6–22.

Bertolazzi, E. et al. 2004. Future advances driver assistance systems based on optimal control: The influence of "risk functions" on overall system behavior and on prediction of dangerous situations. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 386–390, Parma, Italy, June 14–17.

Bertozzi, M. et al. 2000. Vision-based intelligent vehicles: State of the art and perspectives. *Journal of Robotics and Autonomous Systems*, vol. 32, no. 1, pp. 1–16.

Bonic, L., et al. 2017. Identification of real world driving scenarios for the functional safety of autonomous vehicles. In *Proceedings of the 30th International Electric Vehicle Symposium & Exhibition,* October 9–11, Stuttgart, Germany.

Chan, C.Y. et al. 2004. Threat assessment of traffic moving toward a controlled intersection. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 931–936, Parma, Italy.

Dagli, I. et al. 2004. Cutting-in vehicle recognition for ACC systems—towards feasible situation analysis methodologies. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* pages 925–930, Parma, Italy.

Dickmanns, E. D. 2002. The development of machine vision for road vehicles in the last decade. In *Proceedings of the IEEE Intelligent Vehicles Symposium 2002 (IV'2002)*, pages 268–281, Versailles, France, June 17–21.

Dickmanns, E. D. 2003. An advanced vision system for ground vehicles. In *Proceedings of 1st International Workshop on In-Vehicle Cognitive Computer Vision Systems (IVCCVS)*, pages 1–12, Graz, Austria.

Donges, E. 1999. A Conceptual Framework for Active Safety in Road Traffic. *Vehicle System Dynamics: International Journal of Vehicle Mechanics and Mobility,* vol. 32, no. 2–3, pp. 113–128.

Figlewski, S. & Levich, R. M. (ed.), 2002. Risk management: the state of the art. *In New York University Salomon Center series on financial markets and institutions,* no. 8. Kluwer Academic, Boston, USA.

Geyer, S. et al. 2014. Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance. *IET Intelligent Transport Systems,* vol. 8, no. 3, pp. 183–189.

Go, K. & Carroll, J. M. 2004. The blind men and the elephant: Views of scenario-based system design, interactions, vol. 11, no. 6, pp. 44–53.

Hunter, D. R. 2002. Risk perception and risk tolerance in aircraft pilots. *Technical report, Federal Aviation Administration, Office of Aerospace Medicine,* Washington, DC, September.

International Organization for Standardization (ISO), 2011. ISO 26262 Road vehicles—Functional safety, Geneva, Switzerland.

Kontio, J. & Basili, V. 1997. Empirical evaluation of a risk management method. In *Proceedings of the SEI Conference on Risk Management,* Pittsburgh, PA.

Lattner, A. D. & Herzog, O. 2004. Unsupervised learning of sequential patterns. In *ICDM 2004 Workshop on Temporal Data Mining: Algorithms, Theory and Applications (TDM'04),* Brighton, UK.

Matthaei, R. & Maurer, M. 2015. Autonomous Driving—A Top-Down Approach. *In Automatisierungstechnik,* vol. 63, no. 3, pp. 155–167.

Polychronopoulos, A. et al. 2004. Dynamic situation and threat assessment for collision warning systems: The EUCLIDE approach. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* pages 636–641, Parma, Italy.

Reschka, A. et al. 2015. The ability and skill graphs for system modeling, online monitoring, and decision support for vehicle guidance systems. In *2015 IEEE Intelligent Vehicles Symposium (IV),* Seoul, Korea.

Ruder, M. et al. 2002. Highway lane change assistant. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* pages 240–244, June 17–21, Versailles, France.

Schmidt, M. T. et al. 2015. A Novel Goal Oriented Concept for Situation Representation for ADAS and Automated Driving. In *18th IEEE International Conference on Intelligent Transportation Systems (ITSC),* pp. 886–893, Las Palmas, Spain.

Ulbrich, S. et al., 2014. Graph-Based Context Representation, Environment Modeling and Information Aggregation for Automated Driving. *In 2014 IEEE Intelligent Vehicles Symposium* (IV), Dearborn, MI, USA, pp. 541–547.

Weiss, K. et al. 2004. Multiple model tracking for the detection of lane change maneuvers. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* pp. 937–942, Parma, Italy, June 14–17.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A whole system approach to managing defective on-train equipment

A.J. Gilchrist
RSSB, London, UK

ABSTRACT: This paper describes recent risk analysis work undertaken to determine the safest operational responses to on-train equipment failures on the mainline railway in Great Britain. The current rules and guidance for managing these failures have focused on minimising the risk to passengers onboard the train with defective equipment. In more recent years, the rail network has become significantly more congested and it has become increasingly necessary to also consider the safety impact any control measures applied to an individual train will have on the rest of the network. For different operational responses two types of safety risk were calculated: the immediate risk from a train not being able to use the defective on-train equipment; and the knock-on risk resulting from any impact on train performance. To illustrate the methodology, results are presented for an Automatic Warning System (AWS) failure on a passenger train.

1 INTRODUCTION

Each train operator on the mainline railway in Great Britain (GB) is required to produce a contingency plan outlining how they will manage situations where certain items of on-train equipment fail. The pieces of equipment for which contingency plans should be made includes a wide range of safety-critical equipment, such as train protection systems and cab radio failures.

The safety impact from an on-train equipment failure will depend on the equipment which has failed and the role of that equipment onboard the train. The loss of on-train equipment will generally result in an increased risk from train accidents such as train collisions, derailments and buffer stop collisions. This is called the 'immediate risk' and may be reduced by imposing restrictions on the train, such as reducing the maximum permissible speed or detraining passengers.

Over the last decade, passenger usage on the GB mainline railway has increased significantly. The annual number of passenger journeys in 2006 was 1.2 million whilst this has increased to 1.7 million in 2016 (Office of Rail and Road 2017). This increased passenger usage has resulted in increased congestion on the rail network, with both trains and stations becoming busier. Any control measures which impact train performance will further increase this congestion and has a resulting safety impact. This is called the 'knock-on risk' and includes personal accident risk resulting from extra boarding, alighting, and crowding at stations, as well as train accident risk caused by miscommunication and additional red signal approaches.

In order to effectively manage on-train equipment failures both the immediate and knock-on risks must be taken into account so that the immediate risk to the train with defective equipment is reduced whilst minimizing the knock-on risk to the rest of the rail network. The management of defective on-train equipment is currently covered by a rail industry standard (RSSB 2016) and associated guidance note (RSSB 2015) as well as Rule Book module TW5 (RSSB 2017). These documents outline the basis by which a train operator should produce their contingency plans, including speed restrictions and maximum distances travelled with the defective equipment. The current rules and guidance were determined to reduce the immediate risk as low as possible but did not fully consider the knock-on risk associated with these operational responses.

Quantitative risk models have been developed to calculate both the immediate and knock-on risk for a range of different operational responses to defective on-train equipment. Using these risk models, the operational response resulting in the lowest total risk (both immediate and knock-on) has been determined. This paper will describe the risk modelling undertaken to determine the most effective way of managing an Automatic Warning System (AWS) failure on a passenger train.

2 AUTOMATIC WARNING SYSTEM (AWS)

2.1 *Introduction*

The primary safety role of the AWS is to provide warnings to the driver of potentially hazardous

situations that are approaching. This might be signals where the train is required to slow down and be prepared to stop, or of approaching severe speed reductions.

When the driver is approaching a potentially hazardous situation:

- The driver will receive a warning horn.
- The driver must acknowledge the warning within a set time period.
- The AWS will then change visually to indicate that the driver has acknowledged the warning.

An emergency brake application will be applied on the train if the driver does not respond correctly to an AWS warning.

Fitment of AWS at signals became standard for British Rail in 1956. Figure 1 shows the historical number of accidents caused by Signals Passed At Danger (SPADs) for the years 1950–1980 (Evans 2003). It can be seen that in the years immediately after AWS fitment became standard, the number of accidents caused by SPADs was significantly reduced (approximately 3 to 10 times). Whilst other factors such as the change from steam to diesel and electric trains and the introduction of colour light signals will have contributed to this decrease, the introduction of AWS is thought to have been a major factor contributing to this decrease.

In the years after 1956, AWS has been introduced at locations other than signals, primarily as a response to major accidents. Following the Morpeth derailment in 1969, a recommendation was made to provide AWS for Permanent Speed Restrictions (PSRs) requiring a third reduction (or more) in speed. Following a derailment at Nuneaton in 1975, temporary AWS magnets were provided on the approach to Temporary Speed Restrictions (TSRs) and, since 1987, they are also provided on the approach to Emergency Speed Restrictions (ESRs). AWS started to be provided on the approach to certain locally monitored level crossings in 1981.

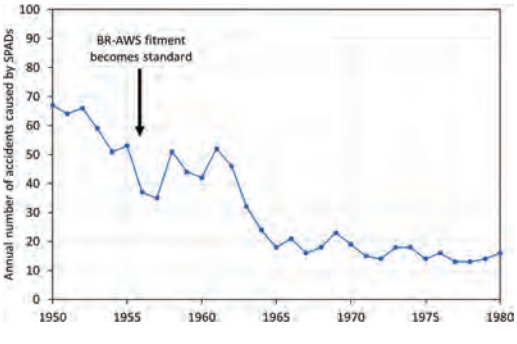


Figure 1. Historical number of accidents caused by SPADs for the years 1950–1980.

3 CALCULATING THE IMMEDIATE RISK

3.1 *Risk model structure*

The immediate risk from a train travelling without AWS is principally from the train either travelling too fast or too far at locations where AWS is fitted. The hazards which have been explicitly quantified in the risk model are therefore:

- Train collisions (caused by SPADs).
- Derailments (caused by overspeeding at speed restrictions and from SPADs).
- Collisions at level crossings.

All frequencies and consequences have been quantified based on established risk models. The main risk model used to determine the immediate risk in this work was RSSB's Safety Risk Model (SRM).

The SRM consists of a series of fault tree and event tree models representing 131 hazardous events, which collectively define the overall level of risk on the GB mainline railway. It provides a structured representation of the causes and consequences of potential accidents arising from railway operations and maintenance on railway infrastructure as well as other areas where the industry has a commitment to record and report accidents. These risk estimates are for the current level of residual risk on the railway, which is the level of risk remaining with the current risk control measures in place and with their current degree of effectiveness. The SRM is calculated assuming that all events are independent and can be attributed to a single cause, reflecting how safety event data has been historically recorded in GB. More information on the SRM and how it calculates risk may be found in the Risk Profile Bulletin (Dacre 2014).

The SRM has been designed to take account of both high-frequency, low-consequence events (occurring routinely, and for which there is a significant quantity of recorded data) and low-frequency, high-consequence events (occurring rarely, and for which there is little recorded data). For each of the low-frequency, high-consequence train accidents considered in this work the SRM has a specific fault and event tree structure. For example, for train collisions, the national frequency of SPADs manually coded by type of signal and cause is used as a precursor. Fault trees are then used to estimate a predicted frequency of train collisions for each type of SPAD, based on the probability the train will reach a potential conflict point and whether there is another train at this location. Event trees are subsequently used to determine the average consequence from a train collision, considering escalation factors such as the probability of train fires and secondary collisions. Where there is sufficient data, all

probabilities and frequencies in the SRM fault and event trees are derived from historical event data by determining average rates and trending analysis. Since all available event data is used to determine these probabilities, the SRM provides national average risk estimates. Because of the network-wide nature of the SRM, it is necessary to make average assumptions that represent the general characteristics of the network when calculating the risk values.

To determine the effect of operational responses on the immediate risk, risk estimates were required for trains travelling at different speeds and passenger loadings. This required significant modification of the fault and event trees contained within the SRM, including estimating the change in effectiveness of the Train Protection and Warning System (TPWS) in stopping a train from reaching a potential conflict point following a SPAD. TPWS is a system which is fitted in certain high-risk locations and automatically applies the brakes on a train if it passes a signal at danger, or if the train's speed is excessive when approaching a signal at danger, permanent speed restriction or buffer stop. Note that not all signals are fitted with TPWS and of those that are, only some are additionally fitted with overspeed protection on approach. By analysing the historical data from trains which have been stopped following a TPWS intervention, the effectiveness of TPWS fitment at signals and speed restrictions can be estimated (Harrison 2007). This methodology was incorporated into the SRM fault and event trees to determine the effect of a speed restriction on the train's immediate risk. The TPWS effectiveness calculations have assumed Network Rail's standard Overspeed Sensor (OSS) and Train Stop Sensor (TSS) loop positions and set speeds, including the addition of an additional TPWS+ loop for line speeds of 75 mph and over.

3.2 *Risk without AWS*

The risk values in the SRM are based on recent historical data and therefore only estimates the risk for trains with AWS fitted and working. In order to calculate the increase in risk during an AWS failure the increase in driver error probabilities when AWS is no longer available as a reminder was estimated using Railway Action Reliability Assessment (RARA) (Gibson 2012). RARA provides a consistent approach to human error quantification for the rail industry and may be used to determine high-level estimates of the increase in driver error rates without AWS for different causes of train accidents.

To estimate the increase in train collision and derailment risk at signals, the causes of SPADs most likely to be affected by AWS were determined. These were found to be:

- The driver failing to check the signal aspect.
- The driver failing to react correctly to a cautionary aspect.
- The driver misreading a signal (either misreading the correct signal or reading the incorrect signal).
- Miscommunication between the driver and signaller.

Of these causes, failing to check the signal aspect and not reacting correctly to a cautionary aspect were determined to be those where AWS would have the greatest influence since these are the situations where AWS can directly prevent the error. For these causes of SPADs, RARA analysis gives two estimates of the increase in SPAD frequency without AWS:

- 20 times increase—drivers who have driven the route many times before and have a good knowledge of where signals are so that, even without AWS, approaching signals is still a simple, routine task.
- 150 times increase—drivers who are less familiar with the route or are particularly reliant on AWS to let them know where they are and react correctly to signals. This may only be a very initial increase for a short distance whilst the driver becomes used to driving without AWS.

In reality, this increase will not be constant and will vary with different signal approaches and drivers. Some signal approaches will be simple and correspond to the lower increase whilst the more complex approaches may correspond to the higher increase. Therefore, an average increase of 85 times has been used as a best estimate for this increase.

For the other causes of SPAD identified as being influenced by AWS, the initial error will not be prevented by AWS but the AWS sunflower will act as a visual reminder to the driver, providing an opportunity to rectify the initial error before it results in a SPAD. For these causes the effect of AWS being unavailable will be less and RARA analysis suggests that there should be a 6.25 times increase in these causes of SPADs without AWS.

As well as accidents caused by SPADs, the calculated error rate increases were also used to estimate the increase in drivers overspeeding at speed restrictions. Since AWS is not generally fitted at buffer stops, no increase in the frequency of buffer stop collisions is predicted without AWS.

The total calculated immediate risk for different line speeds is illustrated in Figure 2. The figure shows the estimated immediate risk without AWS, including the upper and lower bounds as calculated from the RARA analysis. For reference, the baseline level of risk calculated by the risk model for when AWS is working is also shown. The drop in risk at

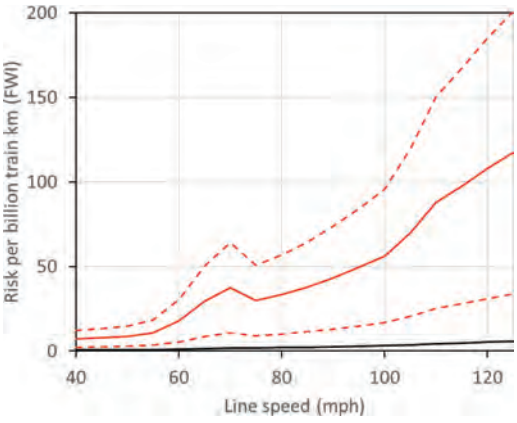


Figure 2. Immediate risk per billion train km in FWI as a function of line speed. The black and red lines show the average calculated risk with and without AWS respectively. The dashed lines show the upper and lower bounds of the calculated risk values without AWS using the two different estimates for the increase in driver error from the RARA analysis.

75 mph is due to the addition of a second TPWS+ loop, providing extra protection at signals. The risk results are given in units of Fatalities and Weighted Injuries (FWI) per billion train km. This is an aggregate measurement of safety risk, using weightings for fatalities, major injuries, minor injuries and shock/trauma events which have been agreed for use by the GB rail industry (Jones-Lee & Loomes 2008).

Analysis of the historical data in Figure 1 can also give an estimate of the risk without AWS. This will, however, not take into account any reliance on AWS that a driver may have developed by driving regularly with AWS. The level of risk calculated from historical data can be therefore be thought of as the level of risk in areas where AWS has never been introduced or if AWS has been unavailable for a long period of time. The values calculated from historical data analysis are almost identical to the lower bound of the RARA estimate. This may be expected since in areas where AWS has never been provided drivers will not have developed any reliance on AWS to warn them of signals and approaching signals without it will be a routine task.

4 KNOCK-ON RISK

Analysis of data shows that there is a strong correlation between certain types of hazardous events and train performance. Any operational response which affects train performance will therefore result in an increase in the risk from these hazardous events. This risk increase is called knock-on risk.

4.1 *Relationship between train performance and risk*

In order to calculate the knock-on risk, the amount of risk associated with train performance needs to be determined. This is achieved by estimating the percentage of annual risk, as calculated by the SRM, which is attributable to train delay minutes and cancellations. These percentages may be derived by investigating the correlation between event frequency and train performance data. Four main areas of risk were considered:

- SPADs
- Staff assaults (both physical and verbal)
- Boarding and alighting incidents
- Passenger slips, trips & falls at stations.

The daily frequency with which each of these events occur may be plotted against Public Performance Measure (PPM) data (Office of Rail and Road 2017) to determine the percentage of these events which are attributable to performance. PPM is a measure of the percentage of passenger trains which were delayed or cancelled on a particular day. An example plot for SPADs is shown in Figure 3. In this example, it can be shown that each delayed train approximately results in an additional 1 in 10,000 chance of a SPAD and that 34% of current SPADs are related to train delays. The knock-on risk from SPADs is subsequently determined by adding together all of the risk in the SRM associated with SPADs, calculating the fraction of the annual delay minutes accumulated during the failure and multiplying this by the portion of the risk identified as resulting from delays. This process was then repeated for each of the four main risk areas.

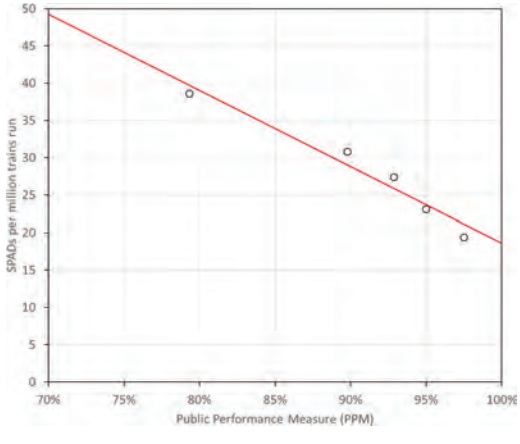


Figure 3. Number of SPADs per million trains run as a function of Public Performance Measure (PPM). The red line is a linear best-fit to the data.

In addition to the risk areas where the percentage associated with delays and cancellations have been determined through the analysis of current data, other risks have also been included in the knock-on risk calculations. These were identified in a previous risk assessment for the Interim Voice Radio System (IVRS) through expert judgment in workshops (Harris 2006).

4.2 *Calculating the knock-on risk*

Delay minutes and the number of cancelled trains are calculated for each possible operational response to an on-train equipment failure. The factors considered when calculating these delay minutes are:

- Any delays (both reactionary and primary) from running at reduced speed.
- Delays accrued from part or full cancellation of trains.

In addition to delay minutes, any extra boarding or alighting resulting from cancelling a train mid-journey is also calculated. These delay minutes and extra boarding and alighting are then used with the relationships described previously to determine a knock-on risk for each operational response.

5 OPERATIONAL RESPONSES

For AWS failures, once a failure has been detected the train should travel to the next available location where the train can be dealt with. At this location it is assumed one of the following will occur:

1. The train is taken out of service and is replaced by one with working AWS for the rest of the day.
2. The affected cab is "boxed in" so that it is not required to be used for the rest of the day.
3. If the rear cab has working AWS, then the train enters service driven from the working rear cab, and at the end of this subsequent journey, either 1 or 2 above occur.

Whilst travelling to the next available location, the train should proceed at a maximum speed of either 40 mph or 60 mph. If passengers are onboard the train, they may either be detrained at the next suitable station or remain onboard whilst the train travels to the next available location in order to complete their journey. Therefore, the total risk during four possible operational responses has been explicitly calculated:

1. The train proceeds with a maximum permitted speed of 40 mph. Passengers may remain onboard the train for the duration of the distance to the next available location whilst they finish their journey.
2. The train proceeds with a maximum permitted speed of 40 mph. Passengers are detrained at the next suitable station and the train then proceeds without passengers to the next available location.
3. The train proceeds with a maximum permitted speed of 60 mph. Passengers may remain onboard the train for the duration of the distance to the next available location whilst they finish their journey.
4. The train proceeds with a maximum permitted speed of 60 mph. Passengers are detrained at the next suitable station and the train then proceeds without passengers to the next available location.

6 RESULTS

The calculated risk values, in units of Fatalities and Weighted Injuries per million trains (FWI/mt), for each of the possible operational responses are given in Figure 4. The risk is calculated for each train from the point of failure (assumed to be halfway through the day), to the end of the day (when the failure is assumed to be fixed). The baseline is the risk from a train running from the middle of the day till the end of the day with no failure. The risk values have been calculated by assuming the train is operating under approximately network average conditions and using the average risk increase from

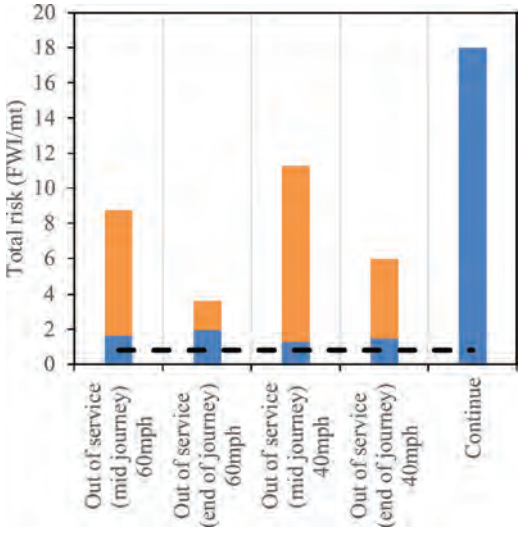


Figure 4. Risk in FWI/mt for different operational responses to an AWS failure. The blue and orange bars represent the immediate and knock-on risks respectively. The black dotted line is the baseline level of risk for a train with AWS working.

the RARA analysis. For comparison, the level of risk if a train continues until the end of the day without AWS is also shown.

As can be seen in Figure 4, for a train operating under network average conditions the operational response which results in the lowest total risk is when:

- The train proceeds to the next available location with a maximum permitted speed of 60 mph.
- Passengers remain onboard the train for the duration of the distance to the next available location whilst they finish their journey.

6.1 *Effect of speed restrictions*

There is only a very small safety benefit in the immediate risk from reducing a train's speed further from 60 mph to 40 mph. The main reason for this is due to the effectiveness of TPWS at stopping a collision. TPWS is designed at signals to automatically apply the brakes if trains are overspeeding on the approach to a signal as well as if they go past the signal at danger. The majority of overspeed sensors for TPWS systems are set so that they only intervene if trains are travelling over speeds in the range 42–46 mph on the approach to a signal at danger. Reducing the train's speed to 40 mph therefore removes most of the protection provided by the overspeed sensor. Whilst the effectiveness of applying the brakes once the train has passed the signal will increase as speed is reduced, the overall effect is that the effectiveness of the TPWS system is very similar at 40 mph and 60 mph. The overall predicted effectiveness of the TPWS system is illustrated in Figure 5 for a train travelling with different speed restrictions on a track with a maximum permitted line speed of 70 mph. Whilst this only applies to signals where TPWS is fitted, since these junction signals represent a large percentage of the overall train collision risk they have a large influence on the overall results.

Reducing a train's speed does, however, cause a large number of extra delay minutes and consequently an increase in the knock-on risk. Reducing the train's speed from 60 mph to 40 mph results in an increase in the knock-on risk from 1.6 FWI/mt to 4.5 FWI/mt for an initial line speed of 70 mph. Since any safety benefit in the immediate risk from reducing the speed is very small, it is outweighed by the approximately three times increase in knock-on risk as delay minutes increase. It is found that for any initial line speed, applying a speed restriction of 60 mph always provides a lower total risk than 40 mph.

6.2 *De-training passengers*

The knock-on risk from de-training passengers immediately and not letting them remain onboard to complete their journey has been calculated to be 5.5 FWI/mt. This risk is mainly due to the risk from extra boarding and alighting at unscheduled stations, where the platform may not be as suitable as a terminal station to accommodate a train full of passengers.

The immediate risk for a train operating without passengers is still significant since an empty train could have a collision with another passenger train. Whilst the knock-on risk from de-training passengers will only depend on the loading of the

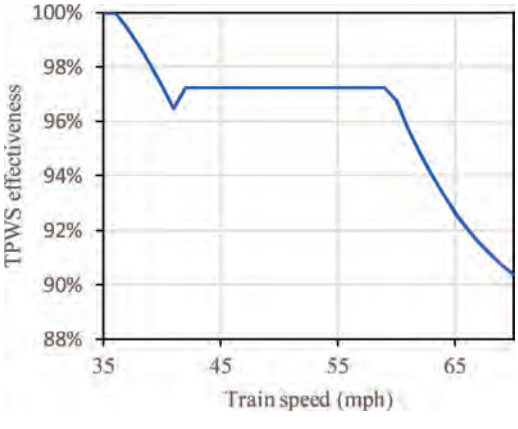


Figure 5. Predicted effectiveness of TPWS at stopping a train collision at a signal for different trains travelling with different maximum speeds on a track with a maximum permitted line speed of 70 mph.

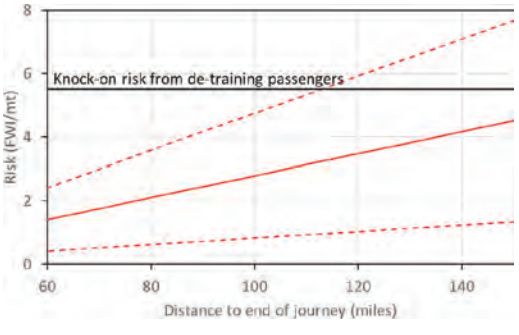


Figure 6. Immediate risk associated with allowing passengers to remain onboard a train to complete their journey or knock-on risk from de-training them, given by the red and black lines respectively. Results are for a train travelling with a speed restriction of 60 mph on a section of track with a line speed of 125 mph. The dotted lines represent the upper and lower bounds of the immediate risk estimate using the two different estimates for the increase in driver error from the RARA analysis.

train when it is cancelled, the immediate risk will depend on the distance passengers would need to remain on the train in order to finish their journey. Figure 6 illustrates how the difference in immediate risk (with and without passengers) increases with journey distance for a train travelling with a speed restriction of 60 mph on a section of track with a line speed of 125 mph.

From Figure 6 it can be seen that, even for a track with a line speed of 125 mph, the additional immediate risk of passengers remaining onboard in order to complete their journey is less than the knock-on risk of de-training them immediately for distances of up to approximately 100 miles. For distances of over 100 miles, the upper bound of the risk estimate starts to outweigh the knock-on risk. Therefore, if the remaining journey distance is longer than 100 miles, passengers should be de-trained before this distance is reached.

7 CONCLUSIONS

Following the risk analysis work, the operational response which was found to provide the lowest total risk during an AWS failure was determined to be:

- The train to proceed to the next available location at 60 mph once the failure has been identified.
- If passengers are onboard, they should be allowed to remain onboard in order to finish their journey whilst the train proceeds to the next available location.
- If the distance to the next available location is greater than 100 miles, the distance passengers remain onboard should be limited to 100 miles. The train should proceed for the remainder of the distance to the next available location without passengers.

Another possible mitigation for AWS failures is to provide a competent person in the driver's cab to monitor the driver and ensure they react correctly to speed restrictions and signal aspects. The risk from operating with a competent person was reviewed separately and all results presented in this paper have assumed no competent person has been provided.

These operational responses represent a change from the current rules whereby the train proceeds at 40 mph and passengers are de-trained at the next suitable station. These rules were previously determined by only considering the immediate risk to the train without AWS and before the national fitment of TPWS was completed. The proposed increase in speed restriction therefore seems reasonable and is consistent with more recently updated rules for other items of defective on-train equipment. The sensitivity of the chosen operational responses to the various assumptions and input parameters in the model has been assessed and the results are found to be robust.

The changes proposed in this work should provide a more effective way of dealing with an AWS failure, considering both the immediate risk to the train operating without AWS and the knock-on risk an operational response has on the rest of the rail network. These changes should therefore provide a benefit both in terms of safety and performance for the GB mainline railway.

REFERENCES

Dacre et al. 2003. *Safety Risk Model Risk Profile Bulletin version 8*. http://www.safetyriskmodel.co.uk.

Evans, A.W. 2003. *Fatal railway accidents in Great Britain: 1946–2002.* University College London.

Gibson, H. 2012. *Railway Action Reliability Assessment user manual: A technique for the quantification of human error (T270 Manual)*. RSSB.

Harris et al. 2006. *Risk Assessment of Failure of the Interim Voice Radio System (IVRS)*. RSSB.

Harrison, C.H. 2007. *TPWS Effectiveness Spreadsheet Methodology, Issue 2.0.* RSSB.

Jones-Lee, M. & Loomes, G. 2008. *The weighting of non-fatal injuries. Fatalities and weighted injuries (T440 Report)*. RSSB.

Office of Rail and Road, 2017. *Passenger and Freight Rail Performance.* http://dataportal.orr.gov.uk/.

Office of Rail and Road, 2017. *Passenger journeys by year*. http://dataportal.orr.gov.uk/.

RSSB, 2015. *Guidance on Defective On-Train Equipment, GO/GN3637, Issue 2.* Railway Industry Guidance Note.

RSSB, 2016. *Defective On-Train Equipment, RIS-3437-TOM, Issue 1.* Railway Industry Standard.

RSSB, 2017. *Preparation and movement of trains: Defective or isolated vehicles and on-train equipment, GERT8000-TW5, Issue 8.* Rule Book Module TW5.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Multi-risk and L.U.P.: A methodology to evaluate neglected risks and risk interactions. An Italian case study

E. Pilone & M. Demichela
Politecnico di Torino, Torino, Italy

G. Camuncoli
ARIA s.r.l., Torino, Italy

ABSTRACT: The paper presents a semi-quantitative methodology developed to help Italian local authorities in facing multi-risk aspects in their Land Use Planning practices. The methodology acts as a pre-screening of the risks present on the territory, highlighting the areas more exposed to risk and risk interactions, also taking into account aspects neglected by the sectorial plans. A quick overview of the methodology is provided, together with a significant Italian case study: a small town in Piedmont, for which neither the land use planning related to major risk plants, nor the supra-regional plans for flood preventions were sufficient to obtain a detailed representation of the overall risk. The proposed methodology analyzed the context and evaluated the possible interactions, identifying possible environmental consequences, and then addressed further studies and interventions to the critical situations. A dedicated questionnaire was developed for the plants, to examine in depth the assets more exposed to NaTech risks.

1 INTRODUCTION

Land Use Planning (LUP) procedures improve and program the use of territories, therefore they have to deal with several types of risks, starting with the natural ones deriving by the territory itself (flood, earthquake) and arriving to the risks generated by men (Technological risks, climate change, etc.). Risks are faced through dedicated sectorial plans, that have a hierarchical development and application: i.e. in Italy, they are usually drafted by regional or supra-regional authorities and then applied by the Municipalities. However, the multiplication of tools dealing with risks (City plans, Emergency plans, supra-local plans) can sometimes bring to lose some important information; also, climate change is varying the reliability of the calculations of return times for events influenced by climate. Most of all, currently an integrated plan containing all the risks does not exist, therefore also the possible risk interactions are neglected.

Several projects related to Multi-risks have been developed in recent years, proposing different types of methodologies to deal with the problem of interactions. Besides qualitative approaches, that are mainly adopted at a wider general scale, many projects proposed quantitative analyses aimed at taking into account and harmonize the different probabilities of occurrence of the risks. However, these methodologies can be affected by the lack of data, and are very long, costly and difficult for the final users and stakeholders, that means LUP decision-makers (Menoni et al, 2006; Nadim & Liu, 2013). This is particularly evident for Italy: Municipalities, as final LUP planners, have not the right expertise and financial resources to apply any multi-risk approach, most of all if it is not made mandatory by law. The lack of integration between risk plans and the non-consideration of risk interactions, summed up to the increasing effects of climate changes, already brought to several disasters (i.e., the repeated floods in Geneva, caused by a creek whose dangerousness was well known but not adequately represented in the Municipal emergency plan and City plan, or the Rigopiano hotel tragedy, where an avalanche caused by a earth shake invested an hotel built in an area where constructions should not have been permitted. The dangerousness of this area was correctly identified by an old map produced by Abruzzo region, that

however was not reported in the updating of the local City plan).

In order to help the Municipalities in taking into account their territorial risks in an integrated way, the authors proposed a semi-quantitative methodology for the local scale, acting as a pre-screening instrument to rapidly identify the areas more exposed to risks and their possible interactions. Further studies, resources and planning actions shall be primarily addressed to these areas. A semiquantitative methodology, based on indexes, and intended for a direct use by the Municipality technicians, was developed by the authors; it is briefly summarized in the following Paragraph 2; an in-depth explanation can be found in (Pilone et al, 2017). This paper mainly focuses on the application for an Italian case study, a Municipality in the Piedmont region, where possible interactions between flood and industrial risks were identified.

2 METHODOLOGY FOR RISK INTERACTION

The steps of the proposed methodology were partially inspired by the ERIR, a Plan for the safe LUP around Seveso plants, that in Italy is mandatory for the Municipalities with a Seveso plant inside their territory. While ERIR only considers Industrial risk, the objective of the proposed methodology was the identification of the impact of several territorial risks and of their possible interactions, on the basis of a semi-quantitative rating scale, going from 0 to 3 onwards:

$0 < I \leq 0.99$: Negligible
$1 < I \leq 1.99$: from Low to Moderate
$2 < I \leq 2.99$: from Moderate to High
$I \geq 3$ onwards: from High to very high.

The methodology was explicitly designed for a direct use by the Municipality technicians.

2.1 *Risk characterization*

The most relevant territorial risks have to be described and investigated according to three Macro-categories, that express peculiar aspects of the risk analyzed and determine its impact: 1) *HE Historical and recent events*: recurrence of the risk events analyzed; 2) *PM Protection measures*: protection and preventive measures that could reduce the impact of the risk analyzed; 3) *SE Strengthening effects*: Local characteristics increasing the risk effects. The latter was explicitly introduced to consider risks not only on the basis of their probability of occurrence, but also in relation to all the intrinsic factors, sometimes neglected in the sectorial plans, that could enhance the final risk impact: i.e., for earthquakes, the quality of the soil; for floods, the section reductions and flow obstructions, for Seveso industries, the quantity and type of substances detained and type of items etc.

The macro-categories are rated in accordance to the local variations of the risk, in compliance with the scale above-mentioned. A dedicated guideline was developed to help the Municipality technicians in this procedure; at the moment, the guide is related to the most diffused risks in Italy and in Europe: industrial, flood and seismic risks (Table 1 reports the first two).

2.2 *Risk interaction*

The macro-categories constitute the basis to evaluate possible risk interactions, because they accurately describe each risk and its possible impact. Therefore, when risks overlay, the effects of one risk on another one (binary interaction) can be assessed through an average sum of the ratings assumed by the two risks in the analyzed point of the territory, following Equation 1 below:

$$Interaction = \left[\left(HE_{R1} + HE_{R2} \right) * 2 + \left(SE_{R1} + SE_{R2} \right) * 1 + \left(PM_{R1} + PM_{R2} \right) * 0.5 \right] / 6 \quad (1)$$

Equation 1 also shows the different weights assigned to the Risk macro-categories for the interaction assessment; in fact, they have different reliability in terms of available data and capacity to influence the final interaction value. The weights assigned following these criteria were validated through expert judgement.

2.3 *Risk compatibility and planning phase*

The values obtained for binary interactions and risks have to be superimposed to the territorial and environmental vulnerabilities, whose identification follows the Italian legislation for ERIR (Ministerial Decree 09/05/2001). If the values of the interactions or of the macro-categories overcome the threshold of 2.5 (medium-high impact) in areas where relevant or extreme vulnerable elements are present, a potential incompatibility is identified. The Municipality shall conduct further studies and investigations to verify the situation and adopt opportune planning measures.

An optional step was introduced in this phase of the methodology to help the Municipalities in the assessment of the interactions involving industrial risks: thanks to two modelling software (ALOHA® and HSSM®), the possible spatial extension of the damage areas can be hypothesized and taken into account for the planning phase.

Table 1. Rating guideline—industrial and flood risks.

Category	Rating		
	1 < I ≤ 1.99	2 < I ≤ 2.99	I ≥ 3 onwards
INDUSTRY			
SE	Few items with Na-tech risk; scenarios related to flammable substances, with a reduced area of impact.	Items with NaTech risk; scenarios related to flammable and environmental substances	Huge quantities of hazardous substances. Items with NaTech risk. Toxic scenarios and/or with a great extension.
HE	No relevant or NaTech accidents occurred.	Low impact events (NaTech/ with external repercussions)	High impact events (NaTech/ with external repercussions)
PM	No dedicated measures for NaTech; lack of protective measures towards the environment	Good safety level, partially effective also towards NaTech accidents	Preventive measures adequate for avoiding NaTech risk and domino effects
FLOOD			
SE	Interaction with other rivers/ creeks with low or reduced criticalities; hydraulic devices in good state; no or few critical points	Interaction with and hydraulic control devices with moderate criticalities; critical points; the river/creek/etc. analysed contains key element for the safeguarding of the general safety of the system	Problematic interaction, recognized high critical areas, reported in Flood plans. Hydraulic devices in bad conditions, with recognized criticalities
HE	Rare main flood events, return time of Flood management plans is confirmed (zones classified as C, or Em, Cn—if recent events do not evidence different distributions/timing of the floods)	Floods of moderate impact, and/or in areas not included in Plans, with a short return time (≥50 years) (zones classified as B, or Eb, Cp—if recent events do not evidence different distributions/timing of the floods)	Events with return time > than that of the Flood management plan worst zone. (zones classified as A, or Ee, Ca—if recent events do not evidence different distributions/timing of the floods)
PM	No water regulation artefacts/ systems or insufficient number/way. Criticalities and inadequate safety level	Water network/river/creek is properly controlled, the artefacts do not show relevant criticalities	The management of the water network/river/creek is well coordinated, evidencing no criticalities

3 CASE STUDY

The application of the proposed methodology is showed through an Italian case study: a little Municipality near Turin with 16000 citizens, interested by flood and industrial risks provoked by "minor sources", which are not adequately considered in the sectorial planning.

The town raises on a flat land crossed by several artificial channels, derived from Stura river and used in the past for irrigation purposes. Urbanization and industrialization completely altered the functioning of the water network: many channels were deviated, interrupted or undergrounded, and their maintenance completely ceased, while the waterproof surface dramatically increased. Besides this minor water network, the northern portion of the municipal territory is crossed by a creek, connoted by several reduced sections of the water flow and banks with low height.

Banna-Bendola and the water network produced extensive floods in 1994, 2000 and 2008; the water height reached 80–100 cm. The flooded areas were mapped in detail by (Regione Piemonte, 1998) and (Provincia di Torino, 2009), but the new PGRA Piano Gestione del Rischio Alluvioni—Plan for Flood management (AD.B.Po, 2016), only reported the potential flooding areas of the creek. The dangerousness of the secondary water network and its combined effects with the creek in case of intense rainy events were not analyzed and neither mapped. At the same time, the return times assigned to the creek buffer-zones were not so in line with the recurrence of the recent events, as demonstrated also by (Politecnico di Torino, 2009).

The industrial risk of the Municipality analyzed is ascribed to a single Seveso plant, a former plating factory named 'X' for this paper, that was closed in 2010 for breach of obligations related to the

AIA—Integrated Environmental Authorization. Since it was never ascertained if the underground platin basins had been emptied in a safe way, the Municipality had to prepare ERIR plan. ERIR plan was drafted in compliance with (Regione Piemonte, 2010) and (Provincia di Torino, 2010) that require to include in the analysis the so-called Seveso Sub-threshold plants (detaining the 20% of the hazardous substances necessaries to be classified as Under-tier Seveso plant) and all the potential hazardous activities. Thanks to this request, not applied in other Italian regions, two potential Seveso plants (named 'Y' and 'Z'), not signalled by the Authorities in charge, were identified. Probably, since ERIR was drafted immediately after the approval of Legislative Decree 105/2015 (Italian implementation of Seveso III Directive), these plants had not adequately checked the new classification of hazardous substances imposed by the decree. The hazardous plants identified are majorly located close to the water network; they were repeatedly interested by flooding, even if there are no testimonies on the consequences of these events (see Figure 1 in the following page).

Since the current sectorial plans neglect some sources of risk (i.e. the secondary water network) and do not allow to consider the danger deriving from the interaction Flood/Industry in an integrated way, the methodology was applied to assess possible unforeseen consequences of the compresence of the risks on the Municipal territory.

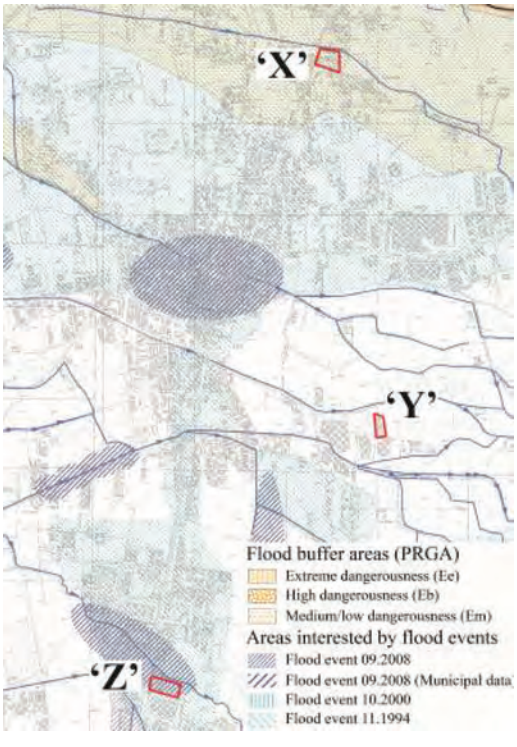


Figure 1. Plants and areas interested by flood risk (Provincia di Torino, 2009).

3.1 *Risk characterization*

FLOOD: In order to proceed with the rating assignment for flood risk macro-categories, rivers and creeks have to be divided in portions connoted by homogenous characteristics and behavior. For the creek, a unique portion was identified in the Municipal territory, because according to (Politecnico di Torino, 2009), this is a unique sub-basin, uniformly connoted by very small slopes of both riversides and river course. The secondary water network was also considered as a unique element, given the complexity of the interactions and interdependencies between the canals, and also because all the main canals equally produced overflowing during the past events.

Table 2 shows the rating attributed to the macro-categories of these two hydraulic elements, responsible of the municipal flood risk. As showed in Table 1, as far as it concerns *HE* ratings, the areas with higher probability of flood obtain a higher rating. However, the return times of the flood events that interested the municipal area in the last 20 years were higher than those defined by PGRA, in particular for areas interested by medium-low flood hazard. The necessity to re-assess the buffer zones for the creek because of this mismatch was recognized also by (A.D.B.Po, 2016). As a consequence, *HE* rating of the areas with medium-low flood hazard (Em) was raised to 2, instead of 1 (see Table 2, fourth column of 'Creek').

For the secondary network, no flood return times were available, but since the network participated in all the major flood events of the last 20 years, being responsible in 2008 of a proper breakdown (Politecnico di Torino, 2009), a *HE* value equal to 2 was assigned.

SE ratings considered all the possible criticalities encountered both for the creek and the water network, while unfortunately *PM* ratings reflect the total absence of protection measures, in spite of the interventions proposed by (Provincia di Torino, 2009).

INDUSTRY: The ratings for the plants 'X,' 'Y' and 'Z' were assigned on the basis of the following information: for the first plant, officially recognized as Seveso → Emergency plan (Prefettura di Torino, 2007), Environmental Authorization—AIA (Provincia di Torino, 2007) and Notification of the plant, that were the last documents drafted before the closure; for the other plants → questionnaires compiled by the owners during ERIR draft.

Table 2. Flood risk—rating assignation.

SE		PM	HE
Interaction with other elements	Criticalities of the artefacts, sections	Hydraulic artefacts, levees etc.	Recurrence
Creek			
Possibilities of inverted flow from the creek to the tributaries or upstream	5 critical sections producing hydric insufficiencies were identified in correspondence of bridges	2 areas for flood expansion and a stone riverbank were planned; only the last was realized, after the flood in 2000.	2 zones: Extreme flood hazard Ee (20–50 yrs), Medium-low flood hazard Em (300–500 yrs). Floods in Em more frequent than the assigned Return time, water height 1 m.
3		**0**	**Ee 3 Em2**
Secondary water network			
Water intakes from Stura cannot be regulated; The creek can feed the channels network during flood events	Reduced slope of soil and canals, scarce maintenance, obstructions, riverbeds not defined, inadequate crossing artefacts; covered portions, diversions. Raised roads block the natural flow	5 floodway channels were planned to return the exceeding flows to Stura, but no interventions were executed.	No flood buffer zones and return times assigned, except for a little portion in an agricultural area (Em). Recurrent overflowing in 1994, 2000 e 2008, water heights between 30–80 cm in the city centre, other areas.
3		**0**	**2**

However, essential information related to case history, storage conditions and preventive and protection measures were missing for all the plants.

Table 3 shows the rating assignation to the macro-categories of each industrial plant: for the macro-categories *SE* and *PM*, Google maps and Google street view allowed to partially integrate the missing data on items exposed to NaTech risk, waterproof aprons, etc., but no alternative sources of information were available for *HE*. Therefore, a common indicative value of 1.5, corresponding to a low-medium impact, was assigned if the plant had been involved by past flood events. A negligible *HE* value, equal to 0.5, was given to the plants not hit by flood events located in proximity of canals, to take into account the possibility overflowing water. The ratings assigned to the macro-category *PM* were maintained generally low because of cautionary reasons.

3.2 *Risk interaction*

The industrial and flood characterizations made clear that the analyzed Municipality is not interested by extremely high risks, due to huge plants or important rivers; the risks are generated by little lower-tier Seveso plants and low energy flood events. However, the interaction of plants, mostly detaining toxic and environmental hazardous substances, with the recurrent overflowing events, could produce unexpected and severe conditions for people and environment, because of the lack of adequate protection and prevention measures. In order to verify possible consequences, Equation 1 was applied to each area where flood areas and plants overlay, through dedicate Binary Interactions tables (Tables 4, 5 and 6). They report the ratings attributed to the risks in a specific point of the territory (in this case, in the areas of plants 'X', 'Y' and 'Z') and allow to repeatedly apply Equation 1.

The possible interaction was verified also for plant 'Y', even if it was never interested by flood, because of a cautionary reason: in fact, it is adjacent to canals whose slope, maintenance and riverbed conditions are not different from those of the other canals. Indicative low values were attributed to Flood risk in this area: SE = 1, HE = 0, PM = 0.

3.3 *Compatibility assessment*

The vulnerable elements were investigated according to the requests of national and regional legislation. For the territorial vulnerability, related to urban density and people density, the majority of the residential areas was included in the classes C and D of Ministerial Decree 09/05/2001 (building ratio index $\leq 1.5\ m^3/m^2$). Some punctual

Table 3. Industrial risk—rating assignation.

SE		PM		HE
Assets items at risk	Substance	General	NaTech and pollution	Recurrence
Plant 'X' Seveso				
5 plating basins	Chromium trioxide (T, N)	The plant did not adopt the recommended measures and was closed for its not compliance. POLLUTION: measures for environmental protection not adopted NATECH: information not available		No information. Plant included in (Em) buffer zone.
5 plating basins	Nickel (N)			
3.5 t storage barrels	Chromium trioxide, nickel (T, N)			
Other SE elements: plant closed under unsafe conditions, situation not monitored				
3		**0**		**1.5**
Plant 'Y' potential Seveso				
4.6 t storage barrels	Diisocyanates (T)	Waterproof apron; storage and area for loading/unloading under cover. POLLUTION: Plant subjected to AIA; Plan for the management of rainy water (addressed to a canal). Collection system for accidental spills; different drainage lines for rainy and process water; sedimentation basin, shutter. NATECH: information n.a.		No information on the plant accidents are available. The area is very close to those interested by the flood in 1994 and 2008.
5 t storage barrels	Isoforon diisocyanate (T, N)			
20.4 t storage barrels	Formic acid (T)			
	DMAE (T, F)			
	Propylene diamine (F)			
	Zinc oxide, derivates (N)			
	DPG dust (N)			
	Acetone (F)			
5 t. bags	Sodium fluosilicate (F)			
Other SE elements: Toxic substances ≥ lower tier Legislative Decree 105/2015. Plant not compliant with the regulation.				
2.5		**–2**		**0.5**
Plant 'Z' potential Seveso				
Tank, 27 t	Phenol (T)	AIA authority recommended adopting for the tanks: level alarms, containment basins. POLLUTION: Plant subjected to AIA; Plan for the management of rainy water (addressed to a canal). Collection system for accidental spills; different drainage lines for rainy and process water; emergency basin. NATECH: Information n.a.		No information on the plant accidents are available. The plant was repeatedly interest by the flooding of the adjacent canal
Tank, 50 t	Formaldehyde 24% (T)			
Tank, 25 t	Acrylic acid (F, N)			
Tank, 27 t	Acetic acid (F)			
Bags, 22 t	Ammonia (F)			
Other SE elements: Toxic substances ≥ lower tier Legislative Decree 105/2015. Seveso plant not compliant with the regulation. Outdoor unprotected storage areas, some tanks seem to have the containment basin.				
2.8		**–1.8**		**1.5**

buildings were categorized as A (mostly schools) and B (discotheque, bowling center), because of their high frequentation. In relation to Environmental vulnerability, no elements with Extreme vulnerability were encountered, but Relevant vulnerable elements typical of flat land areas were present: 1) depth of the aquifer between 0 and 3 meters; 2) Land use capacity of soil between 1 and 2. These features made the Municipal territory very vulnerable towards possible pollution events.

The compatibility assessment was verified in the areas of risk interactions (flood → industry), corresponding to the zones of plants 'X', 'Y' and 'Z': a buffer zone of 500 m. was drafted around each

Table 4. Binary interaction—Plant 'X'.

			Flood risk			Industrial risk		
			SE	HE	PM	SE	HE	PM
Plant 'X' area			3	2	0	3	1.5	0
Flood risk	SE	3						
	HE	2	No interaction			*2.17*		
	PM	0						
Industrial risk	SE	3						
	HE	1.5	No interaction			–		
	PM	0						

Table 5. Binary interaction—Plant 'Y'.

			Flood risk			Industrial risk		
			SE	HE	PM	SE	HE	PM
Plant 'Y' area			1	0	0	2.5	1.5	–2
Flood risk	SE	1						
	HE	0	No interaction			*0.58*		
	PM	0						
Industrial risk	SE	2.5						
	HE	1.5	No interaction			–		
	PM	–2						

Table 6. Binary interaction—Plant 'Z'.

			Flood risk			Industrial risk		
			SE	HE	PM	SE	HE	PM
Plant 'Z' area			3	2	0	2.8	1.5	–1.8
Flood risk	SE	3						
	HE	2	No interaction			*1.98*		
	PM	0						
Industrial risk	SE	2.8						
	HE	1.5	No interaction			–		
	PM	–1.8						

plant, projecting here the values of the Industrial macro-categories and of F/I interaction.

The condition of compatibility can be considered satisfied if no A and B vulnerable elements are included in buffer zones where H.E., S.E. or Interaction values are higher than 2.5, a threshold corresponding to a medium-high impact.

Figure 2 shows an example of Territorial compatibility analysis for the plant 'X': inside the buffer zone, residential areas classified as C and D, and E areas (building ratio index I ≤ 0.5 m^3/m^2) are identified.

Figure 2. Plant 'X' buffer zone with territorial vulnerable elements.

The threshold of 2.5 is adopted for the environmental vulnerability too, but the specific relation between the threats and the environmental vulnerable element has to be investigated (not all the elements are equally sensitive to risks). The environmental vulnerable elements identified for the case study were sensitive both to Industrial risk and its combined effects with flood.

The assessment of the territorial and environmental compatibility for each plant is reported in Tables 7, 8, 9.

3.4 *Results and planning steps*

The application of the methodology to the case study demonstrated that the simultaneous presence of Industrial and Flood risk can produce unexpected interactions, connoted by low-medium impacts (plant 'X' area = 2.17, plant 'Z' area = 1.98), which are reasonably in line with the verified low energy of the flood events in the areas (water height between 30–80 cm).

The Interaction values do not overcome the alert threshold of 2.5, however the plants analyzed are subjected to potential incompatibility related to their Industrial macro-category SE, whose values are high because of particular conditions (abandon of plant 'X', not compliance with Seveso regulation of plants 'Y' and 'Z'). HE received low ratings only as a consequence of the unavailability of data.

The overcoming of the 2.5 threshold signals to the Municipality that further investigations are needed, in order to: 1) confirm or not the

Table 7. Compatibility—Plant 'X'.

Territorial vulnerabilities inside 500 m.	*Environmental vulnerabilities inside 500 m.*
1) C and D Residential areas. 3 productive areas (E) destined to future commercial function. Two are interested by Flood HE = 3; *2)* Few C punctual elements are included; *3)* No linear elements and strategic areas/building/ infrastructures	Land use soil capacity 1st and 2nd classes; Water table depth <3 m; historical urban areas. Canal for irrigation adjacent to the northern border of the plant, probably used in the past to drain the rainy water. Presence of a well inside the plant.
Territorial compatibility	*Environmental compatibility*
HE and SE ratings for Flood and Industrial risks ≥ 2.5 threshold; no manifest incompatibility because of low people density. However, the state of abandon of the plant represents a potential threat for the territorial elements, particularly in case of flood events (medium value of interaction). Further analysis should be carried out, in particular to verify the state and filling of the containment basins. Areas addressed to future transformations: considering the High Flood risk level, avoiding high density of people and adopting specific constructive parameters.	A potential incompatibility is detected: Industrial SE and HE ratings ≥ 2.5 threshold, in an area where the environmental elements are particularly sensitive to pollution. The interaction value is medium: flood events, even with their low energy, could cause unexpected consequences of spreading and diffusion of pollutants towards the underground water and superficial water. No prevention and protective measures for the environment have never been adopted. An onsite visit is recommended to verify the actual conditions of the plant, and to organize a recovery procedure.

Table 8. Compatibility—Plant 'Y'.

Territorial vulnerabilities inside 500 m.	*Environmental vulnerabilities inside 500 m.*
1) E productive areas; *2)* No A and B punctual elements; *3)* No linear elements and strategic areas/buildings/ infrastructures	Water table depth <3 m; land use soil capacity 1st and 2nd classes (agricultural areas around the plant); two canals for irrigation are close to the plant
Territorial compatibility	*Environmental compatibility*
No incompatibilities were encountered with respect to the territorial vulnerabilities.	Potential incompatibility (SE ≥ threshold) in a highly sensitive area. The plant declared adequate Protection measures, however, since it is not in line with the Seveso regulation, at least an in-depth analysis on the storage methods, and protection and preventive measures should be carry out.

Table 9. Compatibility—Plant 'Z'.

Territorial vulnerabilities inside 500 m.	*Environmental vulnerabilities inside 500 m.*
1) C residential areas + 2 E productive areas for future commercial function; *2)* 2 punctual elements in B (commercial centre, bowling; church); *3)* Energetic lines	Water table depth <3 m; presence of a canal for irrigation adjacent to the northern of the plant
Territorial compatibility	*Environmental compatibility*
Potential incompatibility: threshold for SE > 2.5 with two punctual elements classified as B. An in-depth analysis is recommended for: 1) specific activities of the 2 vulnerable elements; 2) the storage methods and protection and preventive measures of the substances classified as toxic (H2)	Potential incompatibility: SE = 2.8 overcomes the compatibility threshold; the interaction with flood events, even if connoted by a low-medium value (1.98), could enhance the threat. Further analysis on the possible pollution scenarios and prevention and protective measures against flood should be carried out.

incompatibility; 2) plan possible LUP actions, taking into account the actual conditions of the plants and their possible interactions. However, collecting further information on the plants could be a difficult task: Sub-threshold plants and abandoned plants indeed represent a potential threat for population and environment, but they have no legal obligation to provide information about substances detained, or possible external risks deriving from their activities. This lack of obligation and monitoring could in some cases enhance the level of risk in comparison to a Seveso plant.

In order to verify the actual hazardousness of the plants, and establish their compatibility, it is essential to know at least the type of storage, the prevention and protection measures adopted and the case history. For this reason, a detailed questionnaire, reported in Table 10, was proposed by the authors; the portion related to the environmental

Table 10. Questionnaire for in-depth investigation of plants.

A. STORAGE CONDITIONS & NA-TECH ITEMS
1) With reference to the hazardous substances detained, please indicate in detail the storage conditions of each hazardous substance, describing type, capacity, quantity and containment measures adopted:
Hazardous substance:
Stored in (container type):
Number of containers and/or total capacity:
Single Container Capacity:
Position (Inside, outside, outside under coverage, underground, etc.):
Containment measure adopted for the container (basin, waterproof ground etc.):
2) Please report if the following items are present:
Underground pipelines, pipelines passing on not-waterproofed soil *Description (length, width, substance transported, protection measure):*
Long and slim structures (torches, chimneys, cooling and distillation towers etc.) *Description of the structure and its function*:
Open-air water treatment basin/liquid waste storage. *Description of the installation and related preventive measures*
B. CASE HISTORY
3) Please report a list of the accidents occurred in the last 20 years that have provoked release of hazardous materials
Date Item interested Accident description
4) Please signal eventual damages provoked by: flood events, extreme climate events, earthquake.
Date Item interested Accident description
C. ENVIRONMENTAL ANALYSIS[1]
5) For the environmental protection, the owner shall demonstrate to have adopted the protective and preventive measures recommended by Turin Province Guidelines;
OR
5) Proceed with a vulnerability analysis of the conditions of water and soil around their plants:
➢ Depth and the direction of the phreatic aquifer nearby the plant, in a sector with 30° degrees of amplitude and 3 kilometres of extension, measured from the possible point of release in the direction of the aquifer flow;
➢ Presence of wells inside the same sector, within an extension of 500 metres
➢ Presence of drains in superficial creeks or canals.

analysis is extracted from Provincia di Torino Seveso guidelines (Provincia di Torino, 2010).

1. Turin Province guidelines; If the three conditions reported are all verified, the owner shall adopt all the measures of points 1, 2, 3 (although the Municipality could in some cases relieve the owner of the application of point 3).

4 CONCLUSIONS

The application of the proposed methodology to the case study quickly identified the areas more exposed to risk, returning feasible results in terms of possible risk interaction impact, in line with the initial risk values. The risk pre-screening allows to take into account in an integrate way the risks information contained in the various sectorial plans, and at the same time, the Municipality technicians can employ their direct and enriched knowledge of the Municipal territory. Therefore, the methodology can create an increased awareness about risks and a correct risk and LUP management.

Many possible developments and further steps could be carried out: the proposed framework, till now elaborated for 3 risks (Industrial, Flood and seismic), can be extended to more territorial threats and the methodology could be exported to other countries simply adapting the criteria for rating assignation. The authors are currently working on the development of participative practices to facilitate the approach of the technicians to the methodology, and some contacts are in course with Municipalities to directly experiment the proposed approach.

REFERENCES

A.D.B.Po – Autorità di bacino del Po (2016). Piano per la valutazione e la gestione del rischio di alluvioni (Art. 7 della Direttiva 2007/60/CE e del D.lgs. n. 49 del 23.02.2010). Programma di misure del Piano. http://www.adbpo.it/PDGA_Documenti_Piano/PGRA2015/Sezione_A/Relazioni/Programma_di_misure_del_Piano/PROGRAMMA_MISURE.pdf.

Menoni S., Galderisi A., Ceudech A., Delmonaco G., Margottini C., Spizzichino D. 2006. FP6 ARMONIA PROJECT—Applied multi-risk mapping of natural hazards for impact assessment. Deliverable 5.1, Harmonised hazard, vulnerability and risk assessment methods informing mitigation strategies addressing land-use planning and management, http://forum.eionet.europa.eu/eionet-air-climate/library/public/2010_citiesproject/interchange/armonia_project.

Ministero Lavori Pubblici, 2001. D.M. 9 maggio 2001 – Requisiti minimi di sicurezza in materia di pianificazione urbanistica e territoriale per le zone interessate da stabilimenti soggetti agli obblighi di cui agli articoli 6, 7 e 8 del decreto legislativo 17 agosto 1999, n. 334.

Nadim, F. & Liu Z. 2013. MATRIX D5.2 – Framework for multi-risk assessment: New methodologies for multi-hazard and multi-risk assessment methods for Europe. http://matrix.gpi.kit.edu/Deliverables.php.

Pilone E., Demichela M., Camuncoli G., A semi-quantitative methodology to evaluate the main local territorial risks and their interactions, in Cepin R. & Radis B. (ed.), *ESREL 2017* (Portoroz, Slovenia, 18–22 June, 2017), CRC Press.

Politecnico di Torino. 2009. Relazione finale. Esame funzionale e valutazione di efficienza di interventi per la messa in sicurezza idraulica del torrente xxxxx. http://www.idrologia.polito.it/web2/progetti/conclusi/banna-bendola/.

Prefettura di Torino. 2007. Piano di Emergenza Esterno ditta Galvanica. http://www.cittametropolitana.torino.it/cms/risorse/ambiente/dwd/rischio-industriale/pee/bertola/Piano_Emergenza_Esterno.pdf.

Provincia di Torino. 2007. Determinazione del Direttore Area Risorse idriche e Qualità dell'aria n. 115-1461827/2007. Autorizzazione integrata ambientale, e s.m.i. http://aia.minambiente.it/impiantiperterritorio.aspx?t=3&id=001130.

Provincia di Torino (2009). Messa in sicurezza del reticolo idrografico del territorio posto tra il T. Stura di Lanzo e il T. Banna. Analisi di fattibilità. http://www.provincia.torino.gov.it/territorio/file-storage/download/pdf/dif_suolo/news/messa_sicurezza/relazione_geologica.pdf.

Provincia di Torino. 2010. Variante al PTC. Requisiti minimi in materia di pianificazione urbanistica e territoriale per le zone interessate da stabilimenti a rischio di incidente rilevante.

Regione Piemonte (1998). Eventi alluvionali in Piemonte: 2–6 novembre 1994, 8 luglio 1996, 7–10 ottobre 1996, Torino 1998. http://www.arpa.piemonte.gov.it/approfondimenti/temi-ambientali/geologia-e-dissesto/pubblicazioni/immagini-e-files/ev9496/ev9496.

Regione Piemonte (2010). D.G.R. n. 17-377 del 26 luglio 2010. Linee guida per la valutazione del rischio industriale nella pianificazione territoriale, B.U. del 5 agosto 2010, n 31.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Emergency assessment in case of hazardous substance leakage at Czech Republic freight rail transport in 2008–2016

Š. Hošková-Mayerová
Department of Mathematics and Physics, Faculty of Military Technology, University of Defence, Czech Republic

ABSTRACT: Emergency occurrence during the freight rail transport does not necessarily have to be and frequently is not affected by the type of the material transported; however, in case that such an emergency occurs, the category of the substance transported may indicate an increased risk; in particular, the risk of explosions, fire, and significant threat to property and people. Leakage of any substance, in particular a hazardous substance, represents a significant part of all emergency cases at rail transport. Despite the seemingly decreasing number of these risky threats, every single incident has to be investigated and analysed. After examining every case in question, it was found out that the leaked substance was not classified hazardous because it was either plain water or frequently leaking operating fluids. Nevertheless, every leakage has to be thoroughly investigated because it is not clear beforehand what category the substance belongs to. Our aim was to identify crucial factors resulting in emergency cases occurrence on the railway in the Czech Republic through a detailed examination of past incidents. Most of these factors are the same for railway transport in general; however, there may be particular local specificities and organizational faults in individual transport units. Another goal of research consisted in revealing these specifics.

1 INTRODUCTION

Emergency occurrence during the freight rail transport does not necessarily have to be and frequently is not affected by the type of the material transported; however, in case that such an emergency occurs, the category of the substance transported may indicate an increased risk; in particular, the risk of explosions, fire, and significant threat to property and people.

Leakage of any substance, in particular a hazardous substance, represents a significant part of all emergency cases at rail transport. Despite the seemingly decreasing number of these risky threats, every single incident has to be investigated and analysed. After examining every case in question, it was found out that the leaked substance was not classified hazardous because it was either plain water or frequently leaking operating fluids. Nevertheless, every leakage has to be thoroughly investigated because it is not clear beforehand what category the substance belongs to. Správa železniční dopravní cesty (Management of Railway Network Company) provided us the access to its database, which was thoroughly studied, and data from the 9-year period (2008–2016) could be selected and processed. Unfortunately, a longer data period was not available for further use and processing due to the different data archiving method. In 2008–2016, 597 leakages were recorded in the Czech Republic. In accordance with the Regulations for international rail transport RID, dangerous substances are classified into categories depending on their hazard class. The most hazardous substance leakage according to this categorization was in the hazard class 3 – flammable liquids, class 8 – corrosive substances, and class 2 – gasses.

Our aim was to identify crucial factors resulting in emergency cases occurrence on the railway in the Czech Republic through a detailed examination of past incidents. Most of these factors are the same for railway transport in general; however, there may be particular local specificities and organizational faults in individual transport units. Another goal of research consisted in revealing these specifics.

Using the statistical processing of available data, a model of accidents distribution based on the leakage location has been developed (Hasilova, et al. 2017b). The research also identified the operating units with the highest number of emergency cases, the risk of emergency occurrence and the development trend. The unambiguous research objective was to contribute to increasing the safety of hazardous substances transport by rail. All the data obtained were passed to the transportation company for further assessment and incorporation into internal safety regulations.

Presence of hazardous substances in rail transport poses a higher risk effect on the individuals, critical infrastructure an environment.

The most frequent factors affecting the emergency cases occurrence on the railway are as follows:

1. *Technical condition:* state of the track, state of safety devices, state of a notification device, state of the crossing security equipment, state of railway rolling stock, conditions for equipment used for the railway infrastructure maintenance. Technical state is frequently affected by vandalism: equipment and various devices are often damaged, destroyed or stolen. Therefore, frequent monitoring of all equipment connected to the transmission system is necessary.
2. *Human factor:* Poor competence, fatigue, negligence, failed thinking, inattention, scattered attention, non-observance of working/technological procedures.
3. *Climate conditions*: snow, rain, fog, floods, calamities.

In recent years, many incidents and accidents have occurred on the railways. The impact of these accidents could have been much worse if hazardous substances were involved. Therefore, the issue covering transport of hazardous substances on the railways is still urgent and critical. Our effort is aimed at reducing and eliminating consequences of possible accidents to minimum. In order to prevent emergency cases and reduce negative effects, it is necessary to examine thoroughly both causes and consequences of all accidents which had occurred; in addition, knowledge of current situation in terms of hazardous substances transport has to be available as well. The scope of consequences of a possible accident is related both to the type of material transported and the extent of residential zones which are close the site of the accident. Selecting an appropriate route as well as the right time is one of crucial safety factors. Quality analysis of the current state and the identification of the busiest corridors and stations, determination of the level of safety at work in particular operating units and understanding the overall transport context will be of great help in the subsequent proposing measures in terms of ensuring the maximum transport safety.

The developed analysis specifies in the first phase the most loaded corridor sections (operating units); in the second phase, there are specified their risky segments, potential sources of threat at typical activities, and predicted the occurrence frequency of emergency cases.

The Czech Republic belongs to the countries with the densest railroad network used for both domestic and international transport. The chemical industry in the Czech Republic is the third largest industrial sector with basic chemistry, petroleum processing (petro-chemistry), pharmaceutical industry (drug production), rubber industry, industry producing plastics and paper producing industry. Production of basic chemicals (64% of total sales) and drug production (17%) are dominating. Other five sectors are represented in lower quantities: production of specific chemical products and fibres (9%), cleaning and cosmetic agents (5%), production of coating material and paints (4%), production of pesticides and agrochemicals (1%) (Jenerálová, 2011).

The company ČD Cargo, a.s. is dominating in freight transport by rail and a substantial part of transport is made up of transport of hazardous substances.

Table 1 shows the most important producers of chemicals in the Czech Republic that use the company ČD Cargo, a.s. for transporting commodities. There is also listed the number of carriages with hazardous substances, which had been transported by the company ČD Cargo, a.s. within the Czech Republic territory.

The presented data are analyzed for a 3-year period (2014–2016) according to data provided by the ČD Cargo, a.s. database. There are considered data covering only the transport within the Czech Republic territory. The assessment of transported hazardous substances comprises the total of 141,229 railway carriages for a monitored period. The authors also prepared an overview of all companies using the service of the company ČD Cargo, a.s. considering both individual hazard classes and number of railway carriages transported. However, the table is too extensive and cannot be presented here. Therefore, a map had been created to visualize a current situation: there are shown 4 transit corridors that are used by the

Table 1. The most significant companies and number of transported carriages with hazardous substances.

Nr.	Company	2014	2015	2016	Sum
1.	CESKÁ RAFINÉRSKÁ, a.s.	12 656	13 490	3 842	29 988
2.	TERMINAL OIL a.s.	4 224	5 554	1 883	11 661
3.	METRANS, a.s.	5 512	4 471	4 846	14 829
4.	DEZA, a.s.	2 979	3 688	3 807	10 474
5.	BorsodChem MCHZ, s.r.o.	1 367	1 451	1 600	4 418
6.	Ceské dráhy, a.s.	1 015	1 022	1 010	3 047
7.	Czech Airlines Handling, a.s.	0	1 870	1 088	2 958
8.	ArcelorMittal Ostrava a.s.	0	1 246	1 321	2 567
9.	Synthesia, a.s.	1 447	0	1 026	2 473
10.	Lovochemie, a.s.	1 216	0	1 108	2 324

Source: Processed by the author (Internal materials CD Cargo, 2017).

Figure 1. Map of the most significant companies utilizing the company ČD Cargo, a.s. depending on operating units. Source: (Becherová, 2017).

company ČD Cargo, a.s., which transports hazardous substances produced by the most significant companies. The map also shows the sites where hazardous substances are loaded depending on the operating units (main stations).

In 2009–2016, the system of data storage changed: some transport units were excluded (Nymburk, Olomouc and Plzeň), and the available data do not show where they were subsequently included. Therefore, the analysis has to be divided into two time periods.

In the first period, the highest number of emergency cases at transport of hazardous substances belongs to Ústí nad Labem and Praha. In the second period, it is Praha again, Ostrava takes the second position. The main reason of high number of emergency cases consists in the highest density of shipment in these regions; in terms of transport, the operating unit Praha is the busiest. The operating unit Praha also comprises the former operating unit Nymburk and operating unit Plzeň. In 3 years (2014–2016), this operating unit dispatched 142,079 railway carriages with hazardous substances, i.e., approx. 130 carriages a day. At this station, hazardous substances of hazard class 3 prevail, i.e. flammable liquids. The analysis was complicated by the fact that the company ČD Cargo, a.s. does not file all defects; therefore, it was necessary to get additional data (incompatible with the original ones) from the Management of Railway Network Company (hereinafter referred to as SŽDC-MRNC). Having carried out a detailed analysis of these data as well as manual comparison of the individual RID for 2006–2016 periods (inland shipment of hazardous substances), the following data were obtained. Due to the amount of information presented in the Table 2, the table is processed using years and figures. Explanation of particular figures is provided below the table (Table 2).

Table 2. Number of defects at shipment of hazardous substances reported by the company ČD Cargo, a.s.

Type of defect	2006	2007	2008	2009	2010	2011
1.	13	10	13	13	19	35
2.	256	175	172	149	171	294
3.	39	33	50	68	35	51
4.	20	22	8	8	10	10
5.	57	39	21	39	46	15
6.	0	0	3	1	1	0
7.	0	0	0	0	0	0
Total	385	279	267	278	282	405

Fault type	2012	2013	2014	2015	2016	Total
1.	31	0	3	0	1	138
2.	166	69	85	102	41	1680
3.	45	1	5	4	2	333
4.	7	0	3	6	3	97
5.	22	6	5	6	4	260
6.	0	0	0	0	0	5
7.	0	0	0	1	0	1
Total	271	76	101	119	51	2514

Source: Processed by the author.

Explanatory notes in terms of recorded defects:

1. A column not marked with a cross,
2. Incorrect or missing entry necessary for RID transport,
3. Incorrect entry "EMPTY TANK", "LAST LADING" or alternatively "EMPTY, UNCLEANED", or "RESIDUES, LAST CONTENT",
4. Leakage of the transported agent—tanks are not sealed properly, leakage through some device/valve/fittings,
5. Faults without hazardous substance leakage—improperly set boards with labels, missing blind flanges, screws,
6. Relevant specification for transport and inscription of the stored goods are inconsistent (gasses of class 2 are in tanks),
7. Overfilling of the transport unit

2 EMERGENCY CASES DUE TO HAZARDOUS SUBSTANCE LEAKAGE

Leakages of hazardous substance at transport by rail represent a significant share of all emergency cases. The following tables illustrate the leakages of hazardous substances that occurred at the operation of the company ČD Cargo, a.s. in the Czech Republic. SŽDC-MRNC made us possible

to access the database; after detailed examination, the data were processed to get the following results. (Bekesiene, et al., 2016).

In a 9-year period 2008–2016, 597 leakages of hazardous substances occurred at transport by rail in the Czech Republic. There are recorded all emergency cases where various amounts of hazardous substances leaked. In compliance with the Regulations for international rail transport RID, the highest number of hazardous substances leakage occurred in hazard class 2 i.e., flammable liquids. Considerable number of leakages also occurred in class 8, i.e. corrosive substances, and in class 2, i.e. gases. Table 3 presents the number of leakages of hazardous substances according to hazard class that occurred within the Czech Republic territory. (Becherová & Hošková-Mayerová, 2017).

From the presented table becomes evident that the highest number of leakages in the Czech Republic occurred at the operation of the company ČD Cargo, a.s.; they belong to class 3 in compliance with RID, i.e., flammable liquids. The aim of the analysis consisted in the emergency cases distribution, particularly leakages, within the Czech Republic territory.

The following Table 4 is focused on the number of accidents at transport of hazardous substances in compliance with a specific UN code. Data presented cover a 9-year period 2008–2016. Presented data come from the database provided by SŽDC-MRNC.

From Table 4 and supplement N becomes evident that the most frequently leaked hazardous substances in a 9-year 2008–2016 periods are the following agents:

- class 2 – carbon dioxide (UN 1013) and propane butane (UN 1965);

Table 3. Number of leakages at transport of hazardous substances in 2008–2016 periods.

Year	Number of leakages in compliance with relevant hazard class								Total
	2	3	4.1	4.2	5.1	6.1	8	9	
2008	25	107	1	1	1	2	31	5	173
2009	9	82	0	0	5	0	3	1	100
2010	15	73	1	0	0	2	9	1	101
2011	8	94	0	0	2	1	16	1	122
2012	2	45	1	0	4	0	9	0	61
2013	4	33	0	0	0	0	9	1	47
2014	2	18	2	0	5	0	6	3	36
2015	11	12	0	0	1	0	4	0	28
2016	4	14	0	0	5	1	6	0	28
Total	80	478	5	1	23	4	93	12	696

Source: Processed by the author (Internal materials CD Cargo, 2017).

Table 4. Emergency cases due to leakage of hazardous substance in compliance with hazard class.

Year	2	3	4.1	4.2	4.3	5.1	5.2	6.1	8	9	Total
2008	25	107	1	1				2	31	5	173
2009	9	82				5			3	1	100
2010	15	73	1					2	9	1	101
2011	8	94				2		1	16	1	122
2012	2	45				4			9		61
2013	4	33							9	1	47
2014	2	18	2			5			6	3	36
2015	11	12				1			4		28
2016	4	14				5		1	6		30
Total	80	478	4	1		22		6	93	12	698

Source: Processed by the author.

Table 5. Number of leakages—The entire leakage sites.

Year	ČT	BR	ČB	NY	OL	OS	PL	PR	UL	Sum
2008	9	14	11	12	28	12	13	26	48	173
2009	3	15	12	4	9	15	4	19	19	100
2010	5	18	12	11	6	11	5	15	18	101
2011	11	33	6	2	4	17	5	28	16	122
2012	3	7	8	2	2	10	5	10	15	62
2013	6	11	1	1	1	9		8	10	47
2014	1	3	2	4	1	8	4	4	9	36
2015	2	2	1	2	3	10	1	4	3	28
2016		8			3	9	2	5	3	30
Sum	40	111	53	38	57	101	39	119	141	699

Source: Processed by the author.

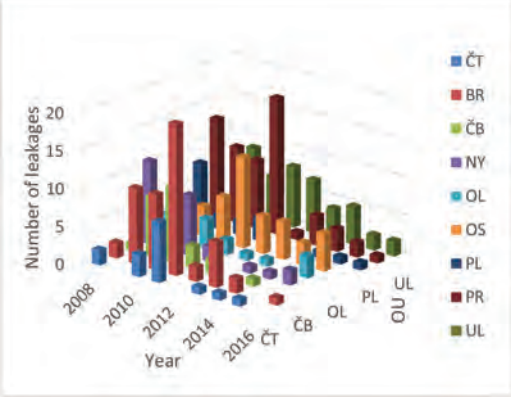


Figure 2. Accident distribution—leakage of hazardous substances according to operating units. Source: processed by the author.

- class 3 – diesel (UN 1202); class 8 – sodium hydroxide (UN 1824) and hydrochloric acid (UN 1789);
- class 9 – liquid tar (UN 3257).

From the Table 5 and graph illustration becomes evident that the accident distribution depends on the site of the leakage. This graph results totally from the operating units, i.e., the number of emergency cases occurred both in Praha, Ústí nad Labem and Ostrava.

3 DISCUSSION AND CONCLUSION

The analytical part examines in detail main factors affecting the railway traffic fluency and the emergency cases occurrence at shipment of hazardous substances. Having applied detailed identification and risky segment exploration involved in shipment, potential sources of threat for typical activities were identified; therefore, the most loaded operating units could have been selected (Hošková-Mayerová et al., 2017, Kass, 1980).

Having solved the specified area of interest, a crucial factor consisted in exploring the cargo company ČD Cargo, a.s. in terms how the company operates, which companies it cooperates with, what goods and in what volumes are cargoes transported, etc. The developed analyses showed what hazardous substances are transported within the Czech Republic territory most frequently, which operating units are the busiest, how many railway carriages and how many tons of particular substances were transported, what the accident rate is during the transport itself (Hošková-Mayerová, 2016, Vališ et al., 2017).

In 2009–2016 periods, 2,384 emergency cases occurred on the Czech Republic railways at the operation of the company ČD Cargo, a.s. In a 3-year period (2014–2016), 3,809,266.262 tons of hazardous substances were shipped; most of them belonged to hazard class 3, i.e., flammable liquids. In the monitored periods 2014–2016, 12,114 loadings of hazardous substances were carried out. The analysis showed that the highest number of loadings was in the operating unit Ústí nad Labem. In terms of number of railway carriages, 142,079 carriages were loaded with hazardous substances in that period: most of them were from the operating unit Praha. Number of reported defects and failures at hazardous substances shipment within the company ČD Cargo, a.s. was monitored as well. In a 11-year period (2006–2016), 2,514 defects and failures were reported; most of them belonged to the category: 'Incorrect or missing record required by RID transport'. In terms of hazardous substance leakages, 696 more or less significant leakages occurred in 2008–2016 periods. The analysis showed that the highest number of leakages occurred in class 3 (in compliance with RID), i.e., flammable liquids at the company ČD Cargo, a.s. operation within the Czech Republic territory (Procházková, 2016, Procházková & Procházka, 2016).

4 SUGGESTED SYSTEMIC MEASURES TO IMPROVE THE PREDICTING EMERGENCY CASES, THEREBY INCREASING RAILWAY SAFETY

Due to the employee fluctuation, the database with data record is not unambiguous. The categorization of emergency cases changed every year; therefore, the data processing was very complicated. In particular, it is essential to

- have data available in uniform and specified form,
- obtain the data in the same way,
- categorize them using the same method which is not modified.

These systemic measures lead to accurate and practical data handling, therefore to more accurate analysis of a possible emergency causes and thereby increasing the safety of persons and property (Hasilová, 2017a, b).

In case that both data categorization and record face the inevitable change, the employees have to be retrained in a particular type of categorization (Woch, 2015, 2017).

The practical benefits are a thorough review of emergency cases in which hazardous substances were present. The paper provides the assessment of hazardous substances transported within the Czech Republic territory in 2008–2016 periods. The phenomenon of exact time information identifying where a particular hazardous substance is present, might contribute to increase the population safety resident close the operating unit area. Close proximity of particular hazardous substance found close residential zones might result in serious consequences; therefore, the introduction of accurate records of particular substances transported by rail would encourage preventing such situations and eliminating possible consequences of emergency cases. (Málek, et al., 2017, Otřísal, et al., 2017, Rosicka, 2006).

The interlink between the information system of the company ČD Cargo, a.s. and SŽDC-MRNC, and between the company ČD Cargo, a.s. and the Integrated rescue system would result in fast notification of employees and further to immediate warning of the population.

ACKNOWLEDGEMENT

The work presented in this paper was supported within the projects "Development of basic and applied research developed in the long term by

the departments of theoretical and applied bases FMT" (Project code DZRO 217) and "Development of the methods for increasing mobility of military vehicles" (Project code: MOBAUT) by the Ministry of Defence the Czech Republic.

The work presented in this paper was also supported by MŠMT ČR, research project no. *SV17-FVL_K106-BEN*: Identification and security of places with high population movement.

REFERENCES

Bekesiene, S., Hošková-Mayerová, Š., & Becherová, O. 2016. Accidents and Emergency Events in Railway Transport while Transporting Dangerous Items. In *Proceedings of 20th International Scientific Conference. Transport Means 2016*. Kaunas: Kaunas University of Technology, 2016, 936–941.

Becherová, O. 2017. Prediction of emergency events on the railways, Thesis, University of Defence 2017, 140.

Becherová, O., & Hošková-Mayerová, Š. 2017. Rail infrastructure as a part of critical infrastructure. In Epin & Briš (eds), *Safety and Reliability—Theory and Applications*, London: Taylor & Francis Group, 2017, 1615–1619.

Hasilová, K. & Vališ, D. 2017a. Hazard Function Modelling of the Composites Reinforced by Natural Fibres for Safety, Security and Defence Applications. In *2017 International Conference on Military Technologies (ICMT)*. Brno: University of Defence, 136–141.

Hasilová, K. & Vališ, D. 2017b. Hazard Function Modelling of the Composites Reinforced by Natural Fibres for Safety, Security and Defence Applications. In: *2017 International Conference on Military Technologies (ICMT)*. Brno: University of Defence, 2017, 136–141.

Hošková-Mayerová, Š., Hubáček, M., Bekesiene, S. & Bureš, M. 2017. Vehicle movement modelling possibilities for defense and crisis management. In Epin & Briš (eds). *Safety and Reliability—Theory and Applications*, London: Taylor & Francis Group, 2017, 3035–3039.

Hošková-Mayerová, Š. 2016. Education and Training in Crisis Management, In: *The European Proceedings of Social & Behavioural Sciences EpSBS*, XVI, Future Academy, 2016, 849–856.

Internal materials of ČD Cargo (Czech Railwais Cargo), 2017.

Jenerálová, I. 2011. Chemický průmysl v ČR [online]. 2011, Available from: http://www.czech.cz/cz/Podnikani/Firmy-v-CR/Chemicky-prumysl-v-CR. (In Czech).

Kass, G.V. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 29(2). 1980.

Málek, Z., Lukaskova, E., Pitrova, K, & Rosicka, Z. 2017. Foodstuff Transportation and HACCP. In Soliman S.K. (ed.), *29th IBIMA Conference: Innovation Management and Education Excellence Vision 2020: From Regional Development Sustainability to Global Economic Growth*, 2017, 823–81.

Otřísal, P., Florus, S. Švorc, Ľ., Barsan, G. & Mosteanu, D. 2017. A New Colorimetric Assay for Determination of Selected Toxic Vapors and Liquids Permeation through Barrier Materials Using the Minitest Device. *Revista de Materiale Plastice*, 54(4), 748–751.

Rosicka, Z. (2006). You Spill, You Dig. Comments on Reliable and Safe Materials Management. In: *Conference Proceedings Degradation, dependability, diagnostics*. UO Brno, ISBN 80-723165-4, 223–228.

Procházková, D. 2016. Environmental disasters' management and detection of priority problems for future research, *Journal of Environmental Protection, Safety, Education and Management*. 2016, 3(5): 76–82.

Procházková, D., & Procházka, J. 2017. Problems Connected with Determination of Size of Maximum Expected Disaster in Selected Site, In Walls, Revie, Bedford (eds), *Risk, Reliability and Safety: Innovating Theory and Practice,* London: Taylor & Francis Group, 2017, 1443–1450.

Vališ, D., Hasilová, K., Leuchter, J. 2017. Assessment and estimation of energy power sources availability. In: *Risk, Reliability and Safety: Innovating Theory and Practice.* London: Taylor & Francis Group, 2017, 2054–2060. Doi. 10.1201/9781315374987-311.

Woch, M. 2015. Risk analysis of the point on aging aircraft structure, In Nowakowski, et al. (eds.) *Safety and Reliability: Methodology and Applications 2014*, Wroclaw, 2355–2360.

Woch, M., Valis, D. 2017. Comparison of different methods for calculation of aircraft structure failure probability per single flight, In: *Risk, Reliability and Safety: Innovating Theory and Practice,* Walls, Revie, Bedford, (eds), London: Taylor & Francis Group, 2017, 1406–1410.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A study on the influence of uncertainties in physical security risk analysis

D. Lichte & K.-D. Wolf
Institute for Security Systems, University of Wuppertal, Velbert, Germany

ABSTRACT: In security risk analysis and assessment, uncertainties regarding occurring threats, consequences and the capabilities of security systems to mitigate vulnerability are enormous. Although some quantitative approaches exist in security risk analysis that allow the consideration of these uncertainties, most practical assessments are based on expert knowledge in semi-quantitative or qualitative models. This paper presents a study on the influence of uncertainties in physical security risk analysis using the example of a semi-quantitative risk assessment of a notional production infrastructure. Therefore, a procedure is suggested as a systematic approach to transfer differing expert ratings into a pdf-based description for a quantitative approach. The influences of uncertainties on the exemplary assessment are calculated and discussed regarding the validity of the results. To visualize these results and to support the decision-making process, a three-dimensional risk matrix is proposed.

1 INTRODUCTION

In security risk analysis and assessment, uncertainties regarding occurring threats, consequences and the capabilities of security systems to mitigate vulnerability are high. Although some quantitative approaches exist in security risk analysis that allow the consideration of these uncertainties, most practical assessments are based on expert knowledge in semi-quantitative or qualitative models.

As uncertainties are hardly or not completely considered, these models do not allow an estimation of their influence on the validity of a conducted assessment. Additionally, the influence of occurring events with a very low probability of occurrence and disastrous consequences, e.g. black swan events, are possibly not considered in these models, as they are very uncertain.

This paper investigates the influence of uncertainties on physical security risk analysis for the example of a semi-quantitative risk assessment of a notional production infrastructure.

To demonstrate the described lack of consideration of uncertainties, the state of the art regarding security risk assessment and vulnerability assessment is outlined. Existing approaches are briefly introduced and the occurring difficulties regarding the use of expert knowledge are described.

For analysis purposes a simple semi-quantitative model for security risk assessment is set up and a transition to a quantitative approach based on (Lichte & Wolf 2017) is introduced. This approach uses probability density functions (pdfs) to describe the abilities of security measures as well as the probability of occurrence of a threat and also the level of possible consequences. The variance of the security measure-related pdfs can be interpreted as the uncertainty concerning the capabilities of the considered measures that are part of the security system. At the same time, differing ratings by experts reflect a rising level of uncertainty that consequently lead to a higher level of variance of the generated pdfs.

Besides the intended purpose of the analysis, the transition is developed to suggest a systematic approach to transfer differing expert ratings to the pdf-based description for the quantitative appSroach. Subsequently, the influences of the resulting uncertainties on the exemplary assessment are calculated and discussed regarding the validity of the results. To visualize these results and to support the decision-making process, a three dimensional risk matrix is proposed. Finally, the study is summarized and discussed in the general context of security risk assessment and the consideration of uncertainties.

2 STATE OF THE ART

2.1 *Security risk analysis*

Security risk analysis is conducted by experts from different fields of expertise, as security is a holistic term (Harnser Group 2010). The protection of infrastructures against intentional physical attacks is covered by the subdomain of physical security

(Beyerer et al. 2010). In detail, the protection of infrastructures is implemented by security measures intended to prevent attackers from reaching their targets or assumed infrastructure assets. The measures include different means of protection, detection and intervention and additionally resilient structures to mitigate the consequences of successful attacks (Garcia 2008).

A possible definition of security risk is formulated as a function of the triplet threat, vulnerability and consequence (Contini et al. 2012, McGill et al. 2007):

$$\text{Risk} = \text{Threat} \times \text{Vulnerability} \times \text{Consequence}$$

This definition is the common basis for mostly all semi-quantitative methods for risk assessment that are used in the field. Recently, this classic definition was discussed e.g. in (Amundrud et al. 2017), regarding its lack to consider uncertainties inherent to its parameters. If the above introduced formulation for security risk is considered as a multiplication of quantitative values, it is apparently only valid for stochastically independent discrete probabilities. Usually, these preconditions are not given, as at least threat and vulnerability are fraught with uncertainty. It may even be discussed whether the description of these parameters as probabilities in the sense of classical frequentist probability theory is justified. However, probability theory is an established concept to handle uncertain quantities. Additionally it can be extended in accordance to the axioms of Kolgomorov in a suitable manner to use it according to the above outlined constraints in security risk assessment. For example, the interpretation of the risk definition based on Bayesian probability theory enables the modeling of the parameter triplet by the formulation as degree-of-belief-densities, thus enabling the consideration of inherent uncertainties formally compliant with probability theory (Beyerer & Geisler 2016).

Thus, the definition combines in a quantitative manner consequences of attacks and probabilities of threat scenarios with the risk of success of individual attacks defined as vulnerability. The above quantitative definition of risk may help to deduce acceptable risks and necessary measures to reduce risks (Broder & Tucker 2012). Inherent uncertainties regarding the three risk factors should be cautiously considered (Campbell & Stamp 2004).

Various approaches to security risk assessment have been developed, which may be divided into qualitative, quantitative and hybrid methods (Meritt 2008). Qualitative methods are mostly based on expert knowledge, while existing quantitative methods use discrete probabilities. Additionally, some quantitative methods aiming at cost-benefit analysis have been developed. Typically, cost-benefit analyses of security measures compute potential financial losses as a result of an attack, the probability of occurrence of various attack scenarios and the vulnerability of the security system (Flammini et al. 2009). This analysis yields accurate results but raises the complexity compared to qualitative methods (Landoll 2011).

2.2 *Vulnerability assessment*

Quantitative vulnerability analysis as part of the quantitative risk analysis is mostly based on methods adapted from reliability and general risk analysis. Here, the considered model is dependent on given attack scenarios (French & Gootzit 2011). This dependency is detrimental to a comprehensive analysis as knowledge about the behavior of a potential attacker may be insufficient (Cox Jr. 2009). The different modeling approaches can be further split up into mainly analytical but also formal methods. An overview of approaches is given by Nicol et al. (Nicol et al. 2004). Analytical methods are often based on attack trees, which can be seen as a derivative of the fault trees known from reliability analysis. Attack trees were first used by Schneier (Schneier 1999) for IT-security analysis and since then have been further developed by different authors, summarized e.g. by Vintr et al. (Vintr et al. 2012).

Contini et al. have introduced incoherent attack trees to characterize the dynamic behavior of the considered system (Contini et al. 2008). Additionally, they integrated simple probability distributions for protection into attack trees to investigate the chronologic sequence of attacks. Hence, it is possible to analyze the security system's ability for an attack intervention by comparing the probabilities of residual protection and system's response (Contini et al. 2012).

Garcia describes this relation (Garcia, 2008), where feasible attack paths as part of different attack scenarios and corresponding barriers are used. The model is time-based and introduces the critical detection point, which is the latest possible point of detection that ensures a successful intervention against the potential attacker.

A quantitative model that further develops the approaches of Contini et al. (Contini et al. 2012) and Garcia (Garcia, 2008) was introduced in (Lichte et al. 2016) and applied to an infrastructure in (Lichte & Wolf 2017). The suggested vulnerability model uses pdfs to describe the characteristics of security measures and uses a path-based barrier model. Additionally it uses the principle of the weakest path. The developed approach in this paper uses this vulnerability model as a basis.

Summarizing, the different existing approaches to analytical modeling and analysis of vulnerability as well as security risk analysis are lacking the consideration of uncertainties in the system parameters and overall behavior. Additionally a great

number of these approaches relies on expert knowledge, causing various problems that are described in the next paragraph.

2.3 *Expert knowledge and uncertainty in security risk assessment*

The elicitation and use of expert knowledge is an often used approach in risk assessment, e.g. to determine point estimates for unknown parameters (Kaplan 1992). Especially, this approach is used to develop and parametrize models for risk assessment in case the available database is very small or there is a lack of objective data (Bolger & Wright 2017).

Therefore, methods for the elicitation of expert knowledge are developed to minimize subjectivity and uncertainty in parameter estimation. Though, the use of expert knowledge itself generally raises the problem of uncertainties in probabilistic risk assessment that should cautiously be analyzed (Rausand 2013) (Aven & Zio 2013). Latest developments in general risk assessment are focusing on uncertainty as a key concept of risk assessment (Aven 2016). Especially in the analysis of rare events that may lead to catastrophic consequences with large uncertainties, different approaches to deal with uncertainty have been suggested and analyzed, e.g. in (Flage et al. 2014).

Especially security risk assessment is often accompanied by great uncertainties, as there is a lack of evidence of threats, consequences and the abilities of security measures. Thus, qualitative or semi-quantitative models that strongly rely on expert knowledge are often used, although these models can lead to misleading or even wrong results (Landoll 2011). Consequently, an analysis of the described resulting uncertainties should be conducted for security risk assessment. This paper discusses an approach of a transition from a semi-quantitative to a quantitative model, which enables further analysis in this direction.

3 APPROACH

To analyze the influence of uncertainties especially emerging from the dependence on expert knowledge in semi quantitative risk assessment methods, an approach is introduced that allows the transition to a risk model that uses triangular pdfs. Both methods are gradually applied to an exemplary infrastructure depicted in Figure 1. The results of both risk models are analyzed for deviations and possible causes. Therefore, a simplified semi-quantitative approach to security risk assessment is presented first. It is based on existing methods, e.g. (Harnser Group 2010). The part of vulnerability assessment reflects four basic assumptions presented in (Lichte & Wolf 2017):

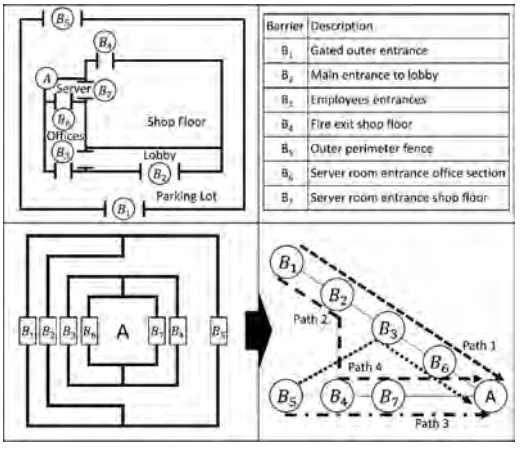


Figure 1. Outline of infrastructure and security system.

1. The weakest path of the security system determines the system's vulnerability as the chosen path of the attacker is uncertain.
2. The combination of protection and observation at barriers is necessary as an attacker is always able to break through a barrier given infinite time without being detected.
3. The detection of an attack is possible only if the protection is sufficient to prevent a breakthrough under observation until detection.
4. After detection, an attack can be stopped only if the residual protection along the remaining attack path lasts long enough to prevent the attacker from reaching the asset until intervention is completed.

In a second step, the developed approach is extended to be able to consider various contributions from expert knowledge. Subsequently, a transition is presented that uses triangular pdfs to represent the various semi-quantitative ratings of the experts. Thus, the differences of the experts' assessments are treated as uncertainties. In the last step, possible results for the semi-quantitative model are compared to the transformed model.

3.1 *Exemplary production infrastructure*

A basic drawing of the considered production infrastructure is outlined on the upper left of Figure 1. The model only considers security measures located along paths that lead to the server room, which is assumed as the asset of a possible attack. The upper right of Figure 1 denotes the security measures marked in the drawing.

The model depicted on the lower left of Figure 1 (onion layer model) represents the structure and feasible attack paths by means of barriers. Four individual attack paths lead to the asset A via barriers B_1-B_7.

All feasible attack paths of the exemplary infrastructure are extracted into a path-based model. The result of the extraction of the attack paths is shown on the lower right of Figure 1. It includes all four attack paths directed towards the asset.

3.2 *Semi-quantitative risk assessment*

The presented semi-quantitative method for security assessment is divided into three submodels. In analogy to the risk definition there are submodels for threat, vulnerability and consequence assessment. The security risk is then indicated in a two dimensional risk matrix, in which the ordinate shows the probability of occurrence of a threat incorporating the vulnerability of the infrastructure against the threat. The axis of abscissas maps estimated consequences.

Threat and consequence assessment are simplified to the estimation of threat probability and level of consequences in ranking scales with five steps, which are described in Table 1.

The presented semi-quantitative model for vulnerability assessment is based on a barrier-oriented view of the infrastructure. Therefore, the commonly used parameters protection, observation and intervention have to be estimated at every security barrier of the infrastructure (Lichte et al. 2016). To realize this, the model uses a five step ranking scale to estimate the time based capabilities of the considered security measures. The ranking scale that is chosen depending on the considered infrastructure as well as the corresponding estimations is shown in Table 2 and 3.

First, the probability of a detection of the attacker needs to be modeled—Based on the four basic assumptions a high level of protection and observation enables a high probability of a triggered alarm A_i at a barrier, so that:

Table 1. Ranking scale threat & consequence.

1	2	3	4	5	Expert score
Scale threat [probability]					
0.0–0.2	0.2–0.4	0.4–0.6	0.6–0.8	0.8–1.0	2
Scale consequence [100k $]					
0–1	1–2	2–3	3–4	4–5	4

Table 2. Score descriptions P,O,I.

1	2	3	4	5
Scale P,O [s]				
0–90	90–180	180–270	270–360	360–450
Scale I [s]				
0–180	180–360	360–540	540–720	720–900

Table 3. Ranking scale threat & consequence.

	Expert score		
Barrier i	P	O	I
1	1	5	3
2	4	1	1
3	2	2	4
4	2	5	5
5	1	5	4
6	5	3	2
7	5	2	2

$$A_i = \frac{P_i + \left(O_{max} - O_i\right)}{P_{max} + O_{max}} \tag{1}$$

Herein P_{max} and O_{max} represent the highest possible score in the ranking scale. For the first barrier B_1 of path 1 of the example infrastructure we get:

$$A_1 = \frac{1 + \left(5 - 5\right)}{10} = 0.1 \tag{2}$$

In a second step, the level of security of a barrier is described by the possibility of an intervention taking place before the attacker reaches the asset. This is described by merging the residual protection of a barrier with the estimated duration of an intervention. The residual protection R_i can be interpreted as the sum of all remaining *n* protection times along the considered attack path. Using a floor function to obtain an integer as a result we obtain:

$$R_i = \left\lfloor \frac{\sum_{i=1}^{n} P_i}{n} \right\rfloor \tag{3}$$

$$R_1 = \left\lfloor \frac{1 + 4 + 2 + 5}{4} \right\rfloor = 3 \tag{4}$$

The residual protection is then merged with the duration of intervention I_i of the barrier to obtain the possibility of a timely intervention T_i.

$$T_i = \left(\frac{R_i - I_i}{R_{max} + I_{max}} \right) + 0.5, \; T \in \left(0.1, \ldots, 0.9\right) \tag{5}$$

$$T_1 = \left(\frac{3 - 3}{10} \right) + 0.5 = 0.5 \tag{6}$$

Using the results for the possibility of a triggered alarm and a timely intervention the vulnerability of the feasible attack paths depicted in Figure 2 is calculated by:

$$V_{Path} = \prod_{i=1}^{n} 1 - \left(A_i \cap T_i\right) \tag{7}$$

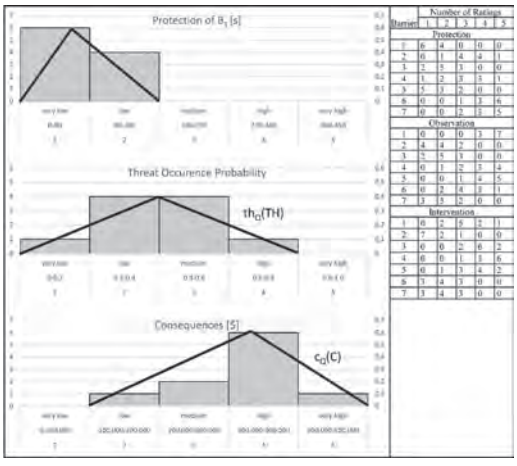


Figure 2. Exemplary histograms and rankings according to aggregated expert knowledge.

For path 1 this yields:

$$V_{Path,1} = (1-0.3)(1-0.35)(1-0.25)\cdot(1-0.35) = 0.147 \quad (8)$$

3.3 *Consideration of multiple expert knowledge*

In order to analyze the influence of uncertainties in threat, vulnerability and consequence assessments, judgements of multiple experts are now introduced. Hypothetical assessments of k = 10 experts are introduced and aggregated in a histogram that shows the frequency of the scores given by the experts. As an example, the histogram of the absolute and relative frequencies for protection measures at B_1 as well as for threat and consequence are shown in Figure 2.

On the right-hand side of Figure 2 the ranking of all experts for all barriers of the security system of the example infrastructure is included.

3.4 *Transition to the usage of probability density functions*

In this step, the semi-quantitative models are transformed into quantitative models. Both types of models use similar operations based on conditional probabilities to describe the security risk of an infrastructure.

Therefore we develop triangular pdfs on the base of the above found histograms that result from the differing ratings of multiple experts. Triangular distributions are widely used in different fields of risk analysis to estimate probability distributions under uncertainty (Haimes 2015). The characteristic parameters of the function are lower limit *a*, upper limit *b* and mode *c*. The determination of the parameters for all input parameters of the risk model are based on the respective histograms of the relative frequency (see Fig. 2).

Table 4. Distribution parameters for th_Q & c_Q.

	a	b	c
th_Q	0	0.8	0.4
c_Q	1	5	3.33

The lower limit *a* and the upper limit *b* for the pdfs of threat and consequence are determined by the minimum and maximum value given in the ranking score for probability of occurrence and costs of a successful attack, respectively. The median interval of the respective histogram is assumed as mode *c*. The resulting fitted triangular pdfs are outlined in Figure 2.

The obtained values for the pdf parameters for threat $th_Q(TH)$ and consequence $c_Q(C)$ are listed in Table 4.

In contrast to the simplified threat and consequence submodels, the transition of the vulnerability submodel is more complex due to the relations between the input parameters. To manage transition, the five ranking scores are transformed to time-based intervals depending on the chosen time-based ranking scales (see Tab. 2) to describe the characteristics of the security measures according to [Lichte & Wolf 2017]:

- Protection is characterized by an estimated time needed for a break-through.
- Observation is the time span needed for a detection.
- Intervention is the period of time until an intervention is completed.

The resulting intervals mark reasonable time steps to describe the time based characteristics and at the same time reflect the assessment of the experts.

In order to ensure comparability within the model, the J = 5 time intervals for all protection and observation measures Δt_{ij} at the i barriers of the infrastructure lie in the time span $[0,t_{max})$. The reasonable choice for the value of t_{max} depends on the considered infrastructure. The intervals Δt_{ij} are all of the same size. For the example infrastructure $t_{max} = 450$ [s]

Due to the modeled relations, the time span of the intervention measures can possibly reach higher values. Therefore the upper considered time span for intervention measures is set to $t_{max,I} = 900$ [s]. As the number of intervals does not change, the intervals $\Delta t_{I,ij}$ grow proportionally.

Now, a triangular pdf can be fitted into the histograms for protection, observation and intervention measures of the security system. The intended fitting of the function is outlined in Figure 2.

Therefore, the lower limit *a* and the upper limit *b* are defined for protection and observation as the minimum and maximum value for t within the intervals comprised in the scoring of the experts. The lower interval l of the j = 5 overall intervals containing expert scores can be found via the following conditions:

$$\sum_{j=1}^{l} k_j > 0 \quad \text{and} \quad \sum_{j=1}^{l-1} k_j = 0 \tag{9}$$

The conditions for the upper interval u are:

$$\sum_{j=1}^{u} k_j > 0 \quad \text{and} \quad \sum_{j=u}^{j} k_j = 0 \tag{10}$$

Following, the lower limit a and the upper limit b are determined by:

$$a = t_{\min} = \min(\Delta t_{il}) \tag{11}$$

$$b = t_{\max} = \max(\Delta t_{iu}) \tag{12}$$

The definition of intervention is equal using $\Delta t_{I,ij}$. The mode *c* for all security measures is determined by means of calculating the median of grouped data. First the median interval of the grouped data can be defined as the interval *m*, which contains the median of the dataset:

$$\sum_{j=1}^{m-1} k_j < \frac{k}{2} \quad \text{and} \quad \sum_{j=1}^{m} k_j \geq \frac{k}{2} \tag{13}$$

Following, the median within the found interval *m* is estimated by linear interpolation assuming an equal distribution as no further information about the distribution within the interval is available. With the lower boundary of the median interval l_m and the upper boundary u_m we obtain:

$$c = t_m = l_m + \frac{\frac{k}{2} - \sum_{j=1}^{m-1} k_j}{k_m} \cdot (u_m - l_m) \tag{14}$$

For the derivation of the parameters of the triangular function of the protection measure at barrier B_1 of the example system we obtain:

$$\sum_{j=1}^{l} k_j > 0 \quad \text{and} \quad \sum_{j=1}^{l-1} k_j = 0 \rightarrow l = 1 \tag{15}$$

$$\sum_{j=1}^{u} k_j > 0 \quad \text{and} \quad \sum_{j=u}^{J} k_j = 0 \rightarrow u = 2 \tag{16}$$

$$a = t_{\min} = \min(\Delta t_{11}) = 0 \tag{17}$$

$$b = t_{\max} = \max(\Delta t_{12}) = 180 \tag{18}$$

$$\sum_{j=1}^{m-1} k_j < \frac{k}{2} \quad \text{and} \quad \sum_{j=1}^{m} k_j \geq \frac{k}{2} \rightarrow m = 1 \tag{19}$$

$$c = t_m = 0 + \frac{\frac{10}{2} - \sum_{j=1}^{1-1} k_j}{6} \cdot (90 - 0) = 75 \tag{20}$$

As all parameters for the triangular pdfs for the vulnerability submodel p(t), o(t) and I(t) are now defined, the calculation of the path vulnerability $V_{Q,Path1}$. by applying the quantitative model is possible. The characteristic parameters and their relation to each other in the quantitative vulnerability submodel are taken from (Lichte& Wolf 2017).

Inserting the obtained pdfs for the barriers of path 1 of the example system into the quantitative model we obtain for the path vulnerability $V_{Q,Path1}$

$$V_{Q,Path1} = 0.028 \tag{21}$$

3.5 *Results and analysis*

As the vulnerability of both models is based on the principle of the weakest path, the other three feasible attack paths also have to be calculated. The results for the semi-quantitative and the quantitative model are summarized in Table 5:

The resulting risk is entered into a security risk matrix for both models. The risk matrix for the semi-quantitative model considering the vulnerability of the weakest path 4 is shown in Figure 3.

The security risk R_S estimated by the semi-quantitative model is calculated to:

Table 5. Vulnerability V_S & V_Q for all attack paths.

Path	V_S	V_Q
1	0.147	0.028
2	0.141	0.014
3	0.328	0.209
4	0.341	0.430

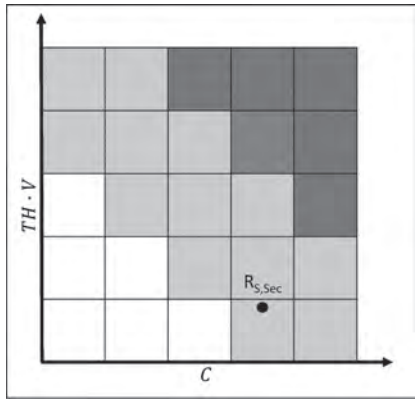


Figure 3. Security risk matrix for the semi-quantitative model.

$$R_{S,Sec} = \left(TH_S \cdot V_{S,Path4}\right) \cdot C_S = 59{,}752\ \$ \tag{22}$$

As the quantitative model uses pdfs the risk matrix is extended to a three dimensional risk matrix. Basically, the matrix maps the bivariate pdf of the risk $R_{Q,Sec}$. It incorporates the probability of occurrence of a threat considering the vulnerability of the infrastructure and the consequences. The distribution-based quantitative risk matrix of the example infrastructure is depicted in Figure 4.

The security risk estimated by the use of the quantitative model is received by the cumulated bivariate pdf. For the example infrastructure we obtain:

$$R_{Q,Sec}\left(c_Q, th_{V,Q}\right) = \int_0^{c_Q} \int_0^{th_{V,Q}} f_{C_Q,TH_{V,Q}}\left(C_Q, TH_{V,Q}\right) \times dC_Q\, dTH_{V,Q}, \boldsymbol{TH}_{V,Q} = TH_Q \cdot V_{Q,Path4} \tag{23}$$

Thus it is possible to calculate the overall security risk.

$$R_{Q,Sec} = 133{,}344\ \$ \tag{24}$$

Other results for interesting boundaries of the distribution variables can also be obtained For example we chose two especially important risk cases to be considered:

- Very high estimated consequences C > 400,000 $
- Black swan events with a very low probability of occurrence of the threat TH < 0.05 and very high consequences C > 400,000 $

When calculating according to (23) we are able to obtain a probability P for the special case, given the overall scenario as well as the resulting risk in terms of consequence:

$$R_{Q,1} = 29{,}920\$,\ P_{Q,1} = 0{,}07 \tag{25}$$

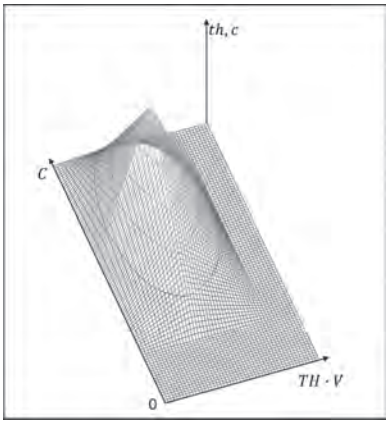


Figure 4. 3D security risk matrix.

$$R_{Q,2} = 1029\$,\ P_{Q,2} = 0.0024 \tag{26}$$

The analysis shows that the semi-quantitative model covers a smaller window of possibly occurring security risks and estimates a significantly lower overall risk. As a result, the model only provides a risk point estimate. Therefore, information about certain cases, e.g. rare events, is neither visible nor can be considered in the risk assessment.

This is especially shown by the two risk cases applied to the considered scenario and the example infrastructure. In contrast to the quantitative model, the information about a probability of occurrence of such events is not comprised in the results of the semi-quantitative model. This information is visualized in the three dimensional risk matrix of the quantitative model (see Fig. 4). It shows the probability of occurring risk cases at the boundaries of the considered scenario besides an average risk, e.g. black swan scenarios like in $R_{Q,2}$.

Furthermore, this is reflected in the differing results of the vulnerability analysis. Although the trend for the vulnerability of the single attack paths of the considered infrastructure is similar, the absolute results differ. The reason for this difference is again a result of the consideration of uncertainties within the vulnerability assessment resulting from the elicitated expert knowledge.

Both findings underline the importance to consider uncertainties to be aware of possible boundary cases, e.g. of rare events with great consequences in a complete risk analysis and decision-making process.

4 CONCLUSION AND OUTLOOK

The paper shows the influence of uncertainties on physical security risk assessment and the resulting need for a consideration of uncertainties. Therefore the paper outlines the state of the art in the field of security risk assessments and expert knowledge especially regarding resulting uncertainties. It is shown that semi-quantitative modeling using expert knowledge is a common used practice though these models lack a consideration of uncertainties so far. To tackle this problem, an approach is introduced that enables a transition towards quantitative modeling using expert knowledge as a basis.

Following, the approach is further detailed and step by step applied to a notional production infrastructure. Therefore a simplified semi-quantitative security risk assessment method is introduced and the transition using triangular pdfs is shown. In a last step a three dimensional risk matrix based on a bivariate pdf resulting on threat, vulnerability and consequences is set up. Subsequently, a comparison of the results of the example infrastructure for

both modeling approaches is conducted. Hereby, differing results for the risk analysis and the possible influence of uncertainties are revealed.

The presented approach shows the influence of uncertainties to physical risk analysis and proposes a three dimensional risk matrix that visualizes possible rare events at the boundaries of estimated threats. Additionally it presents a method for elicitation and use of expert knowledge in the context of physical security risk assessment. Continuously this approach can be further developed to support profound decision-making based on a comprehensive security risk assessment.

Nevertheless, the approach needs further enhancement especially in detailing the input parameters for threats and consequences. The further analysis of the vulnerability is needed also. Here, the influence of uncertainties on the vulnerability assessment should be analyzed, e.g. by Monte-Carlo simulation and sensitivity analysis. The applicability of other modeling approaches like info-gap models should be investigated.

REFERENCES

Amundrud, O. & Aven, T. & Flage, R. 2017. How the definition of security risk can be made compatible with safety definitions. In: Proceedings of the Institution of Mechanical Engineers, Part O: *Journal of Risk and Reliability* 231(3): 286–294.

Aven, T. & Zio, E. 2014. Foundational Issues in Risk Assessment and Risk Management. *Risk Analysis* 34 (7):1164–1172.

Aven, T. 2016. Risk assessment and risk management: Review of recent advances on their foundations. *European Journal of Operational Research* 253 (1): 1–13.

Beyerer, J. & Geisler, J & Dahlem, A. & Winzer, P. 2010. Sicherheit: Systemanalyse und Design. In: *Sicherheitsforschung—Chancen und Perspektiven.* Berlin: Springer.

Beyerer, J. & Geisler, J. 2016. A Framework for a Uniform Quantitative Description of Risk with Respect to Safety and Security. In: *European Journal for Security Research* 1: 135–150.

Bolger, F. & Wright G. 2017. Use of expert knowledge to anticipate the future: Issues, analysis and directions. *International Journal of Forecasting* 33 (1): 230–243.

Broder, J.F. & Tucker, E. 2012. *Risk Analysis and the Security Survey, 4th ed.* Waltham: Butterworth-Heinemann.

Campbell, P.L. & Stamp J.E. 2004. A Classification Scheme for Risk Assessment Methods. Albuquerque: Sandia National Laboratories.

Contini, S. & Cojazzi, G.G.M. & Renda, G. 2008. On the use of non-coherent fault trees in safety and security studies. *Reliability Engineering and System Safety 93* (12): 1886–1895.

Contini, S. & Fabbri, L. & Matuzas, V. & Cojazzi, G. 2012. Protection of Multiple Assets to Intentional Attacks. A Methodological Framework. In: *11th Probalistic Safety Assessment 2012,* Proc. intern. conf., Helsinki.

Cox Jr., L.A. 2009. *Risk Analysis of Complex and Uncertain Systems.* New York: Springer.

Flage R. & Aven, T. & Zio, E. & Baraldi, P. 2014. Concerns, Challenges, and Directions of Development for the Issue of Representing Uncertainty. In: *Risk Assessment-. Risk Analysis 34* (7): 1196–1207.

Flammini, F. & Gaglione, A. & Mazzocca, N. & Pragliola, C. 2009. Quantitative security risk assessment and management for railway transportation infrastructures. In: *Critical Information Infrastructure Security*. Berlin: Springer.

French, G.S. & Gootzit, D. 2011. Defining and Assessing Vulnerability of Infrastructure to Terrorist Attack. In: *Vulnerability, Unvertainty and Risk: Analysis, Modeling and Management,* Proc. conf., Hyattsville.

Garcia, M.L. 2008. *The Design and Evaluation of Physical Protection Systems. 2nd ed.* Burlington: Butterworth-Heinemann.

Haimes, Y. 2015. Risk modeling, assessment, and management. In: *Wikey series in systems engineering and management, 4th ed*. Hoboken:Wiley.

Harnser Group (Ed.) 2010. A Reference Security Management Plan for Energy Infrastructure. Brussels: European Commission.

Kaplan, S. 1992. 'Expert Information' versus 'expert opinions'. Another approach to the problem of eliciting/combining/using expert knowledge in PRA. *Reliability Engineering & System Safety* 35 (1): 61–72.

Landoll, D.J. 2011. *The Security Risk Assessment Handbook: A Complete Guide for Performing Security Risk Assessments, 2nd ed.* Boca Raton: CRC Press.

Lichte, D. & Marchlewitz, S. & Wolf, K.-D. 2016. A Quantitative Approach to Vulnerability Assessment of Critical Infrastructures With Respect to Multiple Physical Attack Scenarios. In: *Future Security* 2016, Proc. intern. conf., Berlin, Germany.

Lichte, D. & Wolf, K.-D. 2017. Quantitative Multiple-Scenario Vulnerability Assessment Applied to a Civil Airport Infrastructure. In: *27th European Safety and Reliability Conference ESREL 2017*, Proc. intern. conf., Portoroz, Slovenia.

McGill, W.L. & Ayyub, B.M. & Kaminskiy, M. 2007. Risk Analysis for Critical Asset Protection. *Risk Analysis*, 27 (5), 1265–1281.

Meritt, J.W. 2008. A Method for Quantitative Risk Analysis. In: *22nd National Information Systems Security Conference*, Proc. nat. conf., Arlington.

Nicol, D.M. & Sanders, W.H. & Trivedi, K.S. 2004. Model-based evaluation: from dependability to security. *IEEE Transactions on Dependable and Secure Computing 1* (1): 48–65.

Rausand, M. 2013. *Risk Assessment Theory, Methods, and Applications.* New York:Wiley.

Schneier, B. 1999. Attack Trees. In: *Dr. Dobbs Journal* 24 (12): 21–29.

Solano, E.: Methods for Assessing Vulnerability of Critical Infrastructure, Institute for Homeland Security Solutions.

Vintr, Z. & Valis, D. & Malach, J. 2012. Attack tree-based evaluation of physical protection systems vulnerability. In: *2012 IEEE International Carnahan Conference on Security Technology (ICCST)*, Proc. intern. conf., Carnahan.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Use case-based consideration of safety and security in cyber physical production systems applied to a collaborative robot system

D. Lichte & K.-D. Wolf
Institute for Security Systems, University of Wuppertal, Velbert, Germany

ABSTRACT: An increasing number of Cyber Physical Systems is used in different areas of application like smart grid, smart factory or smart home. This paper outlines a first approach for an integrated consideration of safety and security for Cyber Physical Production Systems in the so-called Industry 4.0 context which can be interpreted as Systems of Systems. The approach is based on a use case-based model for application in the context of Industry 4.0. To realize a safe and secure operation of Cyber Physical Production Systems in System of Systems a high number of elements, relations and functions have to be taken into account. A Systems Engineering-based approach will be introduced in this paper to deal with this complexity. The approach consists of a SysML-based model which is associated with a procedure that ensures the safe and secure design of Cyber Physical Systems. Specified safety use cases will be used in the following security analysis and assessment. By harmonizing security assessment and safety use cases the integrated consideration is accomplished. The results can be used for technically solution-neutral designs in early development phases.

1 INTRODUCTION

Cyber Physical Systems (CPS) have become increasingly important. As mechatronic systems they consist of sensors, actuators, an embedded intelligence and the ability to communicate with other CPS (Anderl et al. 2013). Applications of CPS are diverse, e.g. advanced automotive systems, environmental control or smart structures (Lee 2008), like e-health, smart home, smart factories, micro grids etc. (Geisberger & Broy 2012).

This paper focuses on applications in industrial environments, where CPS are an important part of so-called Industry 4.0 (Jazdi 2014), where intelligent manufacturing systems (IMS) are communicating and collaborating within production networks (PN) which are connected via internet of things (IoT). These systems are defined as cyber physical production systems (CPPS). Basically, CPPS are systems of systems (SoS), as they consist of autonomous and cooperative elements and sub-systems that are interacting with each other depending on predefined (production) goals (Monostori 2014).

Besides autonomous machine to machine communication (M2M), the collaboration of humans and machines, so-called human machine interaction (HMI), is essential in Industry 4.0 scenarios. Collaborative robots (Cobots), on which a strong emphasis is put on in this paper, are a representative example for these interactions. Cobots collaborate with humans without special safety barriers and use sensors and intelligence to avoid collisions with co-workers. They are connected to other machines in a CPPS for e.g. collecting data and updating production procedures or steering software.

A primary challenge is to maintain safe and secure operation of these CPPS for industrial applications (Lee et al. 2008), since safety and security requirements are key factors to reduce operational risks. The design of a safe and secure system may be a difficult task because of inherent tradeoffs (Lichte et al.) like for example a safe shutdown of a cobot in case of an imminent collision with the collaborating human and the protection against an interruption of production by an intentionally precipitated malfunction of the safety system. If several CPPS are considered simultaneously, this task is getting even more difficult, the result is a high number of CPS combinations and associated use cases.

Frequently, CPS research focuses on interfaces or technical standardization approaches for communication. Yet these approaches are not harmonized so far to provide safe and secure interoperability between the great variety of connected collaborating devices (Kim et al. 2014, Knight 2007, Sarijari et al. 2014). The detailed level and discipline specific focus of these approaches do not allow to adopt the design of the SoS sufficiently.

In this context, this paper demonstrates a first approach for an integrated consideration of safety

and security for CPS in a CPPS by a use case-based model. To realize a safe and secure operation of a large number of elements, relations and functions have to be taken into account (Banerjee et al. 2012, Axelrod 2013).

A systems engineering-based approach will be introduced in this paper in order to deal with this complexity. The approach consists of a SysML-based model which is combined with a procedure to ensure the safe and secure design of Cyber Physical Systems. The procedure is then applied to the cobot system and a first simplified model focusing on the CPS in the CPPS based on use cases is developed. Overlapping use cases (by time and location) are investigated throughout the analysis and supported by the model for a safe and secure design. Finally, results are summarized and discussed.

2 STATE OF THE ART

In a scientific context, safety and security are often defined as a deliberate threat (security) and an unwanted hazard (safety) (Beyerer et al. 2010). Safety functions are designed to protect users from hazards, e.g. an accident. Security functions protect the system and its contents against attacks like intentional misuse. The variety of components and their IT-based networking lead to a growing number of safety and security requirements, which have to be fulfilled by functions.

Regarding CPS in SoS, the variety and diversity of requirements, components and functions is growing as industrial applications are increasingly integrated into the Industry 4.0 context. As the general development in the field of CPPS is similar, CPPS mainly face the same challenges.

The diverse functions are often subject to a fundamental goal conflict. For reasons of safety, redundancies are designed to ensure safety in dangerous situations. Simultaneously these redundancies should not be implemented for reasons of security, because they result in additional attack vectors. Consequently, safety and security functions influence each other.

Additionally, the system's complexity, which is defined by the number and diversity of elements, relations as well as dynamics (Meyer 2007), is increasing, e.g. due to networked systems. Results are for example the requirement of additional security functions to avoid an intrusion into the SoS. Besides the mentioned diversity of elements, complexity is also described, by the high number of participating systems. In turn systems, which carry out tasks independently of each other, as well as together for a limited period of time, can be considered as a SoS (Holt & Perry 2014). According to this characteristics CPPS, which are considered to be SoS of CPS, can also be defined as virtual SoS:

- No central management and no overarching agreed-upon purpose.
- No consistent configuration or maintenance of the SoS as a whole system.
- The individual constituent system will be configured and managed.

These constituent systems consist of a variety of components. For instance, a Cobot includes components for movement, steering, control and positioning, which implement various safety—and security-related functions. As there is no central management or consistent configuration of such a virtual SoS—which is composed of such systems—an integrated safety and security considering model is needed.

Use cases of CPS allow an extensive description of their safety-related behavior. These use cases do not describe the behavior of a SoS consisting of different collaborating CPS. To focus on a comprehensive description of safety and security aspects and resulting goal conflicts, intersection points between the CPS in a SoS have to be investigated. These intersection points have to be defined by CPS specific use cases, which can be postulated (Cockburn 2015). Nevertheless, these use cases do not contain the required information on consequent safety and security goal conflicts.

In order to reach a defined level of safety and security, different methods and concepts may be used, e.g. TSM or GlobalPlatform for security architectures or risk analysis to estimate a safety level. Although specific methods for safety or security exist, an integrated, simultaneous consideration of both aspects is not possible yet (Lichte et al. 2016) or only for software related aspects (Axelrod 2013).

Safety and security aspects need an interdisciplinary understanding for CPS as well as CPS in SoS. Many existing approaches lack a common understanding.

While focusing on complexity, Systems Engineering (SE) can handle these challenges (Mamrot et al. 2014) as it is about creating effective solutions to problems and managing the technical complexity of the resulting developments. SE includes a system model for handling complexity with an interdisciplinary procedure. However many different SE-based approaches were developed. In (Marchlewitz et al. 2015) a first common model for SoS was developed and combined with a procedure. This new Generic Systems Engineering (GSE)-based procedure consists of a standardized procedure using the modules "analysis" (problem identification and system analysis), "target definition" (problem localization) and "design" (recommendations) (Winzer 2015).

The order and structure of these modules is depending on the specific problem under consideration.

Different GSE-based approaches are used, e.g. for requirements engineering (Nicklas 2015) or for the design support of autonomous robots (Mamrot et al. 2014, Marchlewitz et al. 2015). However, the existing system model which was introduced in (Mamrot et al. 2014) does not support the specific combination of CPS in CPPS which is needed for an integrated safety and security consideration as well as a standardized notation. Therefore, a SysML-based approach will be used. Based on its diagrams and standardized notations, an integration in the existing common GSE model of thinking can be realized.

In summary, the following challenges were identified:

- No standardized model for SoS or CPPS (of CPS) for safety and security aspects already exists.
- Missing description of the virtual SoS and its behavior by the CPS specific use cases.
- Difficulties in handling diverging high level use cases caused by inherent complexity.

To deal with these challenges, the approach for an integrated consideration of safety and security aspects for smart home applications will be developed and introduced in the following section.

3 APPROACH

In order to analyze the described complex systems and enable a further development, a Systems Engineering-based approach is introduced [24]. This new GSE-based approach is shown in Figure 1. The proposed procedure is combined with a safety and security integrated system model for applications in the context of CPPS as SoS of CPS. In step 1 the CPPS and its scope are analyzed by using the GSE-module "analysis". This analysis is initially realized by the CPS use cases. Hazardous behavior of the CPPS is then identified by combining relevant use cases, e.g. by time and location intersection points. Step 2 represents the safety use case definition based on SysML notation and diagrams. This is then combined with the GSE target definition module. In the following step 3, resulting safety use cases are investigated to identify related attack scenarios. The security analysis based on the derived safety use cases is necessary as safety use cases possibly create new attack vectors. This ensures the extensive analysis of goal conflicts between safety and security. Finally, the harmonization of safety and security is carried out in step 4. As a result, design recommendations can be derived based on the GSE-module "design".

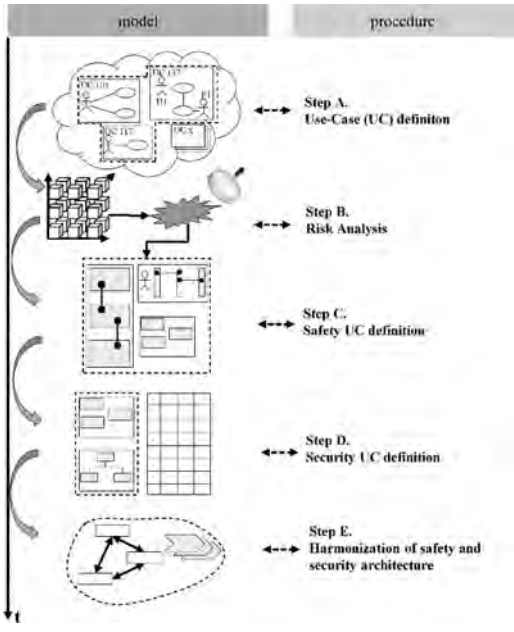


Figure 1. Approach.

Following, the four-level approach will be explained in detail.

3.1 *CPPS definition*

In a first step being based on systems thinking the scope of system has to be limited in order to handle the complexity of SoS [25]. Therefore the proposed CPPS is divided into its subsystems and users. With this limitation, the focus is placed on systems and interacting users. In this article an example based on four different systems will be used:

- Cobot (CB),
- Smart wristband (SWB),
- Intelligent emergency reaction unit (IER),
- Communication hub (CH).

The smart wristband worn by the co-worker is used for position tracking while working with the cobot to avoid hazardous collisions and to monitor the health status in case of potential work accidents. In addition, the communication hub is understood as a part of the CPPS and not as a central management. It only enables the communication between the systems. With the help of these systems the identified challenges and use cases have to be derived. A typical use case description includes preconditions, postconditions, primary flow and an alternative or exception flow (Friedenthal et al. 2015). Therefore, a predefined template is suitable. Different suitable use cases are shown in Figure 2 which have to be

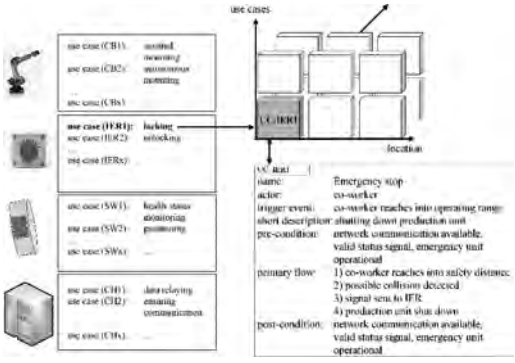


Figure 2. Use cases and combination.

Safety UC: AC	
name:	avoid collision
actor:	co-worker
trigger event:	co-worker falling below safety distance
short description:	avoid collision between co-worker and „CB“
pre-condition:	co-worker wearing „SW“, „CB“ operational, network-communication available
primary flow:	1) recognize direction of co-worker movement 2) adapt dmovement of „CB“ to regain safety distance
Alternative flow:	1) recognize direction of co-worker 2) adapt movement 3) collision
post-condition:	co-worker wearing „SW“, „CB“ operational, network-communication available

Figure 3. Safety use case “AC”.

analyzed regarding safety risks. The challenge is to identify every hazard resulting from an interaction of two or more CPS. This interaction is depicted by the intersection points with regard to time and location. For example, the use case analysis determines that the use cases “CB1-assisted mounting” and “CB2-autonomous mounting” (see Fig. 2) cannot overlap.

In the following exemplary application these four use cases will be used:

- Use case “CB1-assisted mounting”
- Use case “SW1-health status monitoring”
- Use case “SW2-position tracking”
- Use case “CH2-ensuring communication”

By combining the use cases the virtual CPPS is formed out of the cobot, the smart wristband and the communication hub.

For the identified intersecting use cases respectively the new CPS in CPPS a risk analysis has to be performed. This risk analysis is state of the art and therefore not further focused on in this paper. Here, the risk is defined in a quantitative or qualitative way as a function of the severity, the exposure, the occurrence and the controllability (ISO 2011).

By this procedure the risk is assessed for the corresponding use case (Step 2). As a result potential risks are identified for the following steps (step 2–4) to achieve a sufficient safe and secure CPPS.

3.2 *Safety use case definition*

With the result of step 2 safety use cases are defined. The goal of the safety use cases is derived from the risk analysis in step 2. In the example of the combined use cases “CB1” and “SW2” the collision between the user and the AVC should be avoided. Therefore, a new use case “AC” (avoid collision) is defined. The following Figure 3 shows the storyline of this use case.

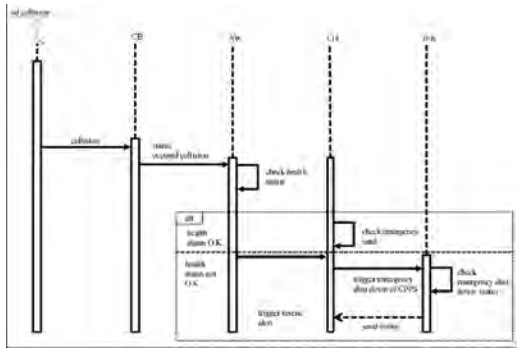


Figure 4. Sequence diagram for “collision” from safety UC: AC and safety UC: emergency.

Unlike “IER1” (see Fig. 2) the use case “AC” has an alternative flow to include a possible collision. Therefore it is necessary to consider a safety use case that reflects the resulting hazard of the collision of the alternative flow. In consequence the safety use case “emergency” is equally defined and documented.

Consequently, a sequence diagram is used to describe the interaction of the safety use cases. Sequence diagrams are based on the predefined use cases (Cockburn 2000). The use cases will be depicted and considered in the safety sequence diagram, which shows the required exchange of messages to describe the functionality of the safety scenario (SysML 2015).

In the example the alternative flow from safety use case “AC” is represented by the first two sequence steps of the diagram. The other part illustrates the steps of the subsequent safety use case “emergency”. An emergency alert will be triggered and “IER” will shut down the whole CPPS if the monitored health status of the user is not okay after the incident (see Figure 4).

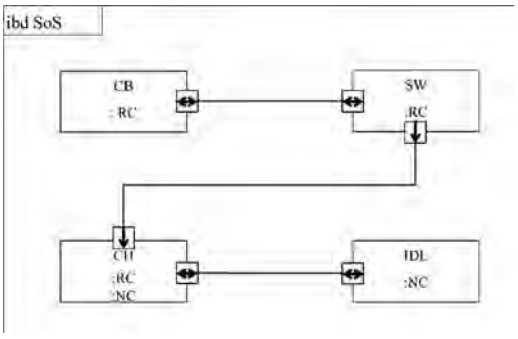


Figure 5. Internal block diagram.

The technical safety relevant information flow of the involved CPS "CB", "SW", "CH" and "IER" is determined by an internal block diagram (see Fig. 5). In addition to the logical task-orientated sequence the internal block diagram allows the description of information flow. Hereby, design support and further security analysis of the involved CPS are prepared. Figure 5 shows the information flow through the radio ("RC") and network ("NC") communication ports and its direction. Likewise, communication redundancies can be defined.

Based on the internal block diagram a security analysis and assessment is prepared in step 3.

3.3 *Security analysis and assessment*

In step 3 the needed security measures are defined to prevent the intended occurrence of threats as results of the safety use case by an outside attacker. The goal is to describe barriers between the components that show necessary limitations of information flow and encryption of communication between the components of the CPPS. Both items of information can be used to extend the solution structure of the safety use case by adding security barriers and hierarchic structures.

Therefore, the CPPS is analyzed by means of a security assessment based on attack scenarios. The most important results of these scenarios are goals and methods of the attack. The safety use case derived in step 2 is used to define the goal of the attack. The attack goal that results from the exemplary safety use case is achieving access to the home. Feasible attack paths and methods are deduced by the diagrams of the CPPS defined in step 2, which show involved CPS and information flows between them. The description of use cases and attack paths may require the integration of further CPPS components. The resulting simplified attack scenarios are summarized in attack trees, which were introduced by (Schneier 1999).

Figure 6 shows five resulting scenarios defined by the CPPS information flow of the use cases. Following, a qualitative assessment is conducted on the CPPS considering the developed security scenarios. The assessment includes a ranking regarding the probability of occurrence (PO) and goal achievement (PG) based on the attack trees shown in Figure 6. The scenarios S5 and S6 are excluded from further analysis as they are very unlikely to occur in terms of PO and PG.

As a result of the security analysis the attack vectors of the probable scenarios (S1-S4) have to be investigated. This shows where barriers are needed to secure the considered CPPS for the specific use case. The simplified block diagram in Figure 7 depicts this.

The above shown barriers describe the limitation of flowing information or needed encryption. Additionally, a simplified hierarchic model is established by analyzing the proposed limitation of

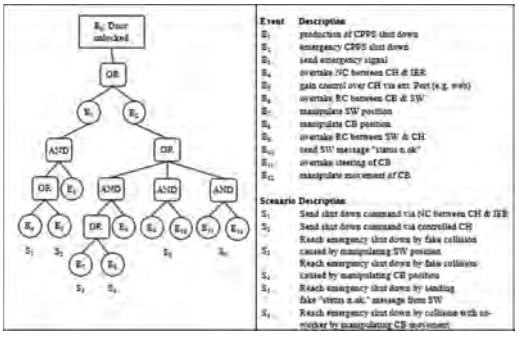


Figure 6. Attack tree.

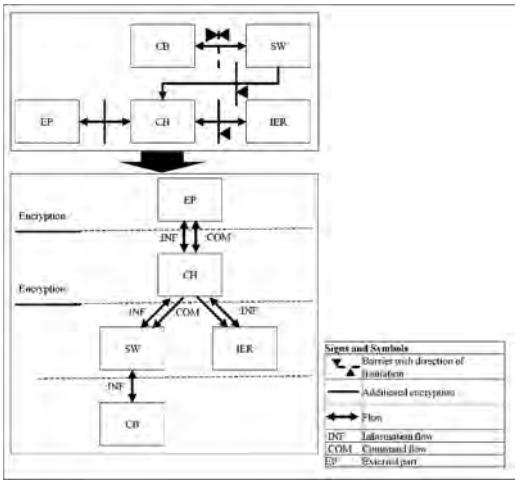


Figure 7. Security ibd and communication flow.

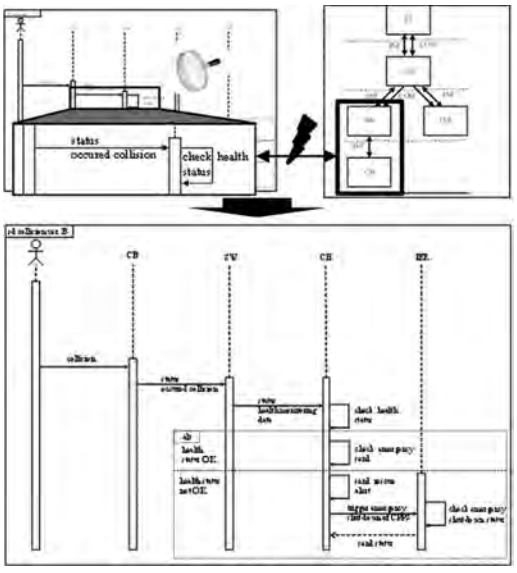


Figure 8. Sequence diagram for "collision" from safety UC: AC and safety UC: emergency.

direction of command and information flow signals between the components of the CPPS.

3.4 *Harmonization of safety use cases and security scenarios*

In the last step, the results of step 2 and 3 will be matched to expose and solve the safety-security goal conflicts related to the analyzed safety use case. As a result, Figure 8 shows a harmonized sequence diagram to achieve an adequate safety and security level. The analysis of information and command flow leads to changed connections in the sequence diagram. In the explained example, the connection of "SW" and "CB" is identified as critical for security. Therefore, the tasks "check health status" and "check emergency send" have to be executed by the "CH". The activity "send rescue alert" is additionally realized by the "CH". As a result, the sequence diagram "sd collision ver. B" is recursively adjusted.

As a result an integrated safety and security consideration for the CPPS based on the use cases is derived. On the one hand possible goal conflicts between safety and security functions are revealed. On the other hand the method enables a designing process that solves these conflicts. Due to the high degree of abstraction, early and technically solution neutral design can be planned.

4 CONCLUSION AND OUTLOOK

In this article a first use case-based approach is developed, which integrates safety use cases and resulting security scenarios for a widespread overview. First the problem of goal conflicts between safety and security was outlined and the state of the art regarding Cyber Physical Production Systems (CPPS) that can be considered as SoS and especially cobots as CPS was summarized. It was shown that existing models and approaches regarding SoS do not focus on an integrated safety and security perspective. Simultaneously the security of CPPS in the context of Industry 4.0 has to be considered in a more detailed way with regard to users and experts. Hence this article proposes an approach based on Systems Engineering to analyze and harmonize safety and security at the same time. The four individual steps of the approach include use case definition, safety use case definition, security scenario analysis and harmonization. Additionally, an example illustrates how the concurrent single steps may contribute to a safe and secure model of the CPPS, which is enhanced in every step of the procedure. In the first step, use cases are defined that may overlap in time and space and combined by time and location for the identified systems of the CPPS. These combinations are analyzed. Step 2 comprises the definition of the resulting safety use cases to avoid risks. They are described by storyline, internal block diagram and sequence diagram in SysML-based diagram types (Alt 2012). The security analysis in step 3 identifies attack goals as a result of the safety use cases and establishes attack scenarios based on attack trees. The probabilities of occurrence and goal achievement of the attack scenarios are qualitatively assessed and security structures containing the limitation of communication and encryption are derived. The resulting security structure is compared to the safety use case in step 4. Occurring goal conflicts are solved by adapting the sequence diagram of the safety use cases.

REFERENCES

Anderl R. & Picard R. & Albrecht K. 2013. Smart Engineering for Smart Products. In: 23rd CIRP Design Conference: Smart Product Engineering, Proc. intern. conf., Bochum, Germany.

Alt, O. 2012. Modellbasierte Systementwicklung mit SysML. München: Carl Hanser Verlag.

Axelrod, C.W. 2013. Managing the Risks of Cyber-Physical Systems. In: IEEE Long Island Systems, Applications and Technology Conference (LISAT), Proc. intern. conf., USA.

Banerjee, A. & Venkatasubramanian, K.K. & Mukherjee, T. & Gupta, S.K.S. 2012. Ensuring Safety, Security, and Sustainability of Mission—Critical Cyber-Physical Systems. In: Proceedings of the IEEE 100 (1).

Beyerer, J. & Geisler, J. & Dahlem, A. & Winzer, P. 2010. Sicherheit: Systemanalyse und—Design. In:

acatech diskutiert: Sicherheitsforschung—Chancen und Perspektiven, Springer Verlag: Berlin.

Cockburn, A. 2000. Writing Effective Use Cases. Addison Wesley.

Friedenthal, S. & Moore, A. & Steiner, P. 2015. A Practical Guide to SysML. *The Systems Modeling Language*. 3. Ed. Waltham, MA: Elsevier.

Geisberger, E. & Broy, M 2012. agenda CPS. Integrierte Forschungsagenda Cyber-Physical Systems, Acatech STUDIE, München: acatech.

Holt, J. & Perry, S. 2014. SysML for Systems Engineering. A Model-Based Approach," *IET Professional Applications of Computing. 2nd Edition. Stevenage, UK*: IET.

ISO (International Standards) 2011. ISO 12100:2011-3, Safety of machinery—General principles for design—Risk assessment and risk reduction.

Jazdi, N. 2014 Cyber Physical Systems in the Context of Industry 4.0. In: *2014 IEEE International Conference on Automation, Quality and Testing, Robotics,* Proc. intern. conf., Cluj,Napoca, Romania.

Kim, S. & Hong, J.-Y. & Kim, S. & Kim, S.-H. & Kim, J.-H. & Chun, J. 2014. RESTful Design and Implementation of Smart Appliances for Smart Home. In:*2014 IEEE 11th International Conference on Ubiquitous Intelligence & Computing and 2014 IEEE 11th Conference on Autonomic & Trusted Computing* and 2014 IEEE 14th International Conference on Scalable Computing and Communications and Associated Symposia/Workshops, IEEE:2014: 717–722.

Knight; M. 2007. Wireless security—How safe is Z-wave?. IET& Computing & Control Engineering Journal. December/January 2006/2007: 18–23.

Lee, E.A. 2008. Cyber Physical Systems: Design Challenges, Electrical Engineering and Computer Sciences, University of California at Berkeley. Technical Report No. UCB/EECS-2008–8.

Lee, J. & Bagheri, B. & Kao, H.-A. 2015. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters* 3 (2015): 18–23.

Lichte, D., Marchlewitz, S., Schlüter, N., Wolf, K.-D.: An Approach to Holistic Safety and Security Risk Assessment Considering Contradictory Requirements under Uncertainty. In: *27th European Safety and Reliability Conference ESREL 2017*, Proc. intern. conf., Portoroz, Slovenia.

Mamrot, M. & Marchlewitz, S. & Nicklas, J.-P. & Winzer, P. 2014. Using Systems Engineering for a Requirement-Based Design Support for Autonomous Robots. In: *IEEE International Conference on Systems, Man, and Cybernetics*, October 5–8, 2014, San Diego, CA, USA.

Mamrot, M. & Winzer, P. 2013. Approach for Structuring the Product Environment for a Systematic Analysis of Field Data. In: *IEEE 8th International Conference on System of Systems Engineering (SoSE*). Maui, Hawaii, USA.

Marchlewitz, S. & Nicklas, J.-P. & Winzer, P. 2015. Using Systems Engineering for Improving Autonomous Robot Performances. In: *IEEE 10th International Conference on System of Systems Engineering*, San Antonio, TX, USA.

Meyer, C.M. 2007. Integration des Komplexitäts managements in den strategischen Führungsprozess der Logistik, Haupt.

Monostori, S. 2014. Cyber-physical production systems: Roots, expectations and R&D challenges. Procedia CIRP 17: 9–13.

Nicklas, J.-P & Winzer, P. 2014. Approach for Using Requirements Engineering in Collaborative Networks. In: *Proceedings of the 17th QMOD-ICQSS International Conference on Quality and Service Sciences,* ICQSS 2014.

Sarijari, M.A. & Abdullah, M. S & Lo, A. & Rashid, R.A. 2014. Experimental Studies of the ZigBee Frequency Agility Mechanism in Home Area Networks. In: *3rd IEEE International Workshop on Global Trends in Smart Cities. goSMART 2014.* Proc. intern. conf., Edmonton, Canada.

Schneier, B. 1999. Attack Trees. *Dr. Dobbs Journal* (24) 12, 21–29.

SysML 2015. SysML V1.4 Specification Release.

Winzer, P. 2015. Generic System Description and Problem Solving in Systems Engineering. In: *IEEE Systems Journal*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Anti-icing expected heat loss as a risk indicator for arctic offshore logistics operations

M. Naseri
UiT The Arctic University of Norway, Tromsø, Norway

E.M. Samuelsen
Norwegian Meteorological Institute, Tromsø, Norway
UiT The Arctic University of Norway, Tromsø, Norway

ABSTRACT: The aim of this paper is to present a mathematical framework for estimation the expected heat loss due to the implementation of de-icing measures for vessels operating in the Arctic offshore. Sea-spray icing on vessels operating in Arctic waters imposes financial and safety risks, such as loss of vessel stability, safety risks for vessel crew, as well as delays in maintenance and operations of offshore units. In Arctic offshore logistics, efficient planning of platform supply vessel operations should be performed while accounting for the risk of spray icing associated with selected voyages. Although Arctic vessels are equipped with a range of anti-icing and de-icing options, an optimum design and estimation of energy consumption for winterisation purposes remains a challenging task especially due to the uncertainties associated with the temporal-spatial variation of meteorological and oceanographic parameters contributing to ice accretion. However, long-term forecasts of such parameters are hardly available during logistics planning phase. Thus, this study uses 3-hourly reanalysis hindcast data (Norwegian Reanalysis 10 km data: NORA10) for estimation of icing rate and develops a probabilistic framework for estimation of expected heat loss, due to implementation of winterisation measures, over sea voyages for long-term logistics plans. Expected icing rate and winterisation-related heat loss can be used as safety and financial risk indicators in Arctic offshore logistics operations and their involved long-term decision-making processes. The framework is illustrated by a case study in the Arctic-Norwegian waters.

1 INTRODUCTION

Sea-spray icing is considered the most severe icing type due to its potentially high accretion rate (Ryerson, 2008; Ryerson, 2011) and a major safety concern for vessels operating in the Arctic offshore as the weight of the ice negatively affects vessel's stability and manoeuvrability. Spray icing can threaten the safety of crew on-board and structural reliability of platforms and vessels. It can interrupt routine on-board maintenance and operations activities due to safety concerns. Heavy icing events and its following de-icing operations can delay the delivery of goods, spare parts, and services (Jones and Andreas, 2009; Naseri and Barabady, 2016; Samuelsen et al., 2015). Figure 1 shows a severe spray icing on KV Nordkapp vessel on 26.02.1987 in the Barents Sea.

There are numerous works on planning offshore logistics operations such designing and optimising offshore fleet size and scheduling a number of platform supply vessels to support the needs of a cluster of offshore platforms; for example, see

Figure 1. 110 tons of ice accumulating during a 17 hours period on KV Nordkapp on 26.02.1987, while sailing from Tromsø to waters between Bjørnøya and Hopen in the Barents Sea (Samuelsen et al., 2015).

(Fernández Cuesta et al., 2017; Halvorsen-Weare et al., 2012; Maisiuk and Gribkovskaia, 2014; Stålhane et al., 2016). Some other works tackle the issue of reducing CO2 emission and vessel voyage cost by lowering fuel and energy consumption (Norlund and Gribkovskaia, 2013; Norlund et al., 2015). However, application of these research findings and developed frameworks and techniques for Arctic offshore logistics may be faced with a great deal of uncertainties due to the contribution of icing risks, its related safety concerns, and induced operational delays.

In order to tackle the issue of icing, vessels and platforms operating in the Arctic offshore are equipped with a range of anti-icing and de-icing techniques such as using heat tracers, insulations, shelters, semi-enclosures, or chemical ice protection options (DNV, 2013; Farzaneh M., 2015; Rashid et al., 2016; Ryerson, 2008; Ryerson, 2011). Implementation of such techniques, on the other hand, increases capital and operations costs, energy usage, greenhouse gas emissions, and fuel consumption. In addition, one should make sure of the reliability of winterisation techniques, whether or not they manage to perform their required functions as expected, and if malfunctioned, they can be maintained and repaired within an acceptable timeframe.

In this regard, efficient planning and execution of platform supply vessel operations is crucially important in terms of timely supply of goods and services to Arctic offshore units while taking the risk of spray icing associated with selected voyages into consideration. To this aim, long-term and real-time estimation of spray icing rate over the voyages is vital, especially due to the temporal-spatial variation of meteorological and oceanographic parameters contributing to ice accretion.

However, modelling sea-spray icing rate, is in general very challenging. This is mainly due to the uncertainties related to accurately estimating the spray amount during wave-ship interaction, the turbulent heat transfer between the atmosphere and wetted surfaces on the ship, and the freezing temperature of the brine water. The brine water on the wetted surfaces of a ship is namely lower than the temperature of the incoming sea water due to salt expulsion during the freezing process (Samuelsen, 2017a). In spite of such challenges and issues, some researchers have proposed highly sophisticated models to estimate spray-icing rate (e.g., (Horjen, 2013; Kulyakhtin and Tsarau, 2014)). However, these models have been little verified, and their complexity are therefore not justified by observations. A major drawback is for instance the fact that they assume that the wave height may be estimated directly from the wind speed, which is rarely the case in observed icing events (Samuelsen et al., 2017). In this study, MINCOG model developed by Samuelsen et al. (2017) is adopted for estimating spray icing rates. A brief description on the MINCOG model, is given in Section 2.

Long-term prediction of icing rates or the parameters contributing to spray-ice formation are hardly available during logistics planning phase. Therefore, by employing MINCOG model, and using 3-hourly reanalysis hindcast data (Norwegian Reanalysis 10 km data: NORA10), icing rates are estimated for a sufficiently long period in the past. By the use of such estimates and applying a non-sequential Monte Carlo simulation technique (Zio, 2013), a probabilistic representation of icing rate for a specific sea voyage and time period is developed to simulate the icing events and their rates in future for that voyage. This is further used as a fundamental input for estimation of expected icing rates and expected amount of energy required for winterisation purposes associated with selected sea voyages during certain time intervals.

This framework and its provided information on expected icing risk and winterisation-related energy consumption can be used as a safety and financial risk indicator for long-term decision-making processes in Arctic offshore logistics operations. In addition, by applying short-term weather prediction data, the presented framework can help vessel crew for making short-term decisions with respect to icing risks along the selected sea voyage. The rest of this paper is organised as follows. In Section 2, after reviewing spray-icing process, the MINCOG model is briefly discussed. Section 3 describes the proposed mathematical framework for estimating expected icing rate and its related energy consumption for a given voyage. The case study and conclusions are presented in Sections 4 and 5, respectively.

2 SEA-SPRAY ICING RATE MODELLING

In this study, a newly developed physics-based ship-icing model, known as Marine Icing model for the Norwegian COast Guard (MINCOG), is adopted from (Samuelsen et al., 2017) to predict the sea-spray icing rate. This model uses the Norwegian coast guard ship class named "KV Nordkapp" as a reference ship type for ship-icing calculations. Samuelsen (2017b) shows how this model provides higher verification scores than previously-applied ship-icing models and nomograms, like the commonly-applied Overland (1990) model, when the models are verified against ship-icing data from Arctic-Norwegian waters, outside Alaska, and at the east coast of Canada.

Sea-spray generated from the interaction between ships and waves are considered the most

dominating water source in ship-icing events (e.g. (Samuelsen, 2017a) and references therein). The MINCOG model is, therefore, based on the modelling of wave-ship interaction icing. Firstly, the sea-spray flux is calculated based on spray data derived from (Borisenkov et al., 1975). Icing rate r, can be calculated from the average sea-spray flux by taking into account the different heat fluxes involved in the icing process on a fixed position in the front of the ship. The heat balance is given by (Samuelsen et al., 2017),

$$q = q_c + q_e + q_d + q_r \quad (1)$$

where q is the energy that is released by freezing process, during which, salt is expelled making freezing temperature lower than that of incoming sea water. In Equation (1), q_c is the convective cooling from the air to the freezing brine, q_e is the evaporative cooling of the brine, q_d is heating (or cooling) from the sea water to the brine, and q_r is the incoming or outgoing longwave and shortwave radiative heat fluxes. Once icing rate is computed, the amount of energy released by freezing, i.e., q, is used for estimating the amount of energy to avoid icing (see Section 3.3).

Six model-input parameters including wind speed, air temperature, relative humidity, mean-sea level pressure, significant wave height, and significant wave period, are all derived from NOrwegian Reanalysis 10 km data (NORA10) (Reistad et al., 2011). Constant values of ship speed, sea-surface or water temperature, and incoming sea-water salinity, are applied.

As illustrated in (Samuelsen, 2017b – Figure 12 and 13), the sensitivity to these latter parameters in the normal range considered in marine-icing studies are relatively low compared to the former parameters, and thus, their medians are chosen as fixed model inputs. Incoming short-wave radiation, is here neglected, and incoming longwave radiation is parametrized by assuming that the atmosphere is radiating as a black body with a temperature equal to the air temperature at the level of the ship.

Furthermore, it is assumed that the winds and waves are coming from the same direction, and a constant angle is applied for the direction between the ship and wind equal to the median value of this angle derived from the icing reported by Samuelsen et al. (2017).

For simplicity, the trajectory model of water droplets used in is Samuelsen et al. (2017) skipped, and the droplet velocity is calculated from the relative velocity of the wind and ship in the horizontal direction, and an assumed terminal velocity of uniform droplets with a constant spherical size with a diameter of 2 mm. A detailed discussion on the MINCOG model, its underlying assumptions and constant values of input parameters are given in (Samuelsen, 2017a; Samuelsen, 2017b; Samuelsen et al., 2017).

3 LOGISTICS RISK INDICATOR MODEL: ANTI-ICING EXPECTED HEAT LOSS

In this work, energy consumption for anti-icing purposes and heat anti-icing heat loss are used equivalently. In order to calculate expected heat loss, a systematic procedure based on a non-sequential Monte Carlo simulation is suggested in this study, as shown in Figure 2. First, the trajectory is decomposed into several segments and coordinates of the origin and destination of each segment is determined. According to the vessel speed and the length of each segment, the time interval the vessel is sailing along segment is computed. In the second step, corresponding to those time intervals and for each segment, the occurrence of icing event and its rate is determined using the hindcast data, which will be used in the next step to extract the statistics of icing rates for those locations. Using determined distributions of icing events and rates for each segment, the occurrence of icing events and their associated icing rates are simulated for each segment at given times. This procedure is repeated for a sufficiently large number of times in order to estimate the statistics and determine the distribution of icing rate and heat loss for the whole trajectory.

3.1 *Vessel trajectory decomposition*

Let α and β denote, respectively, latitude and longitude, in degrees, of a geographical location on a map. Thus, vessel trajectory AB, is recognised by the coordinates of its origin A and destination point B, i.e., $A = (\alpha_A, \beta_A)$ and $B = (\alpha_B, \beta_B)$.

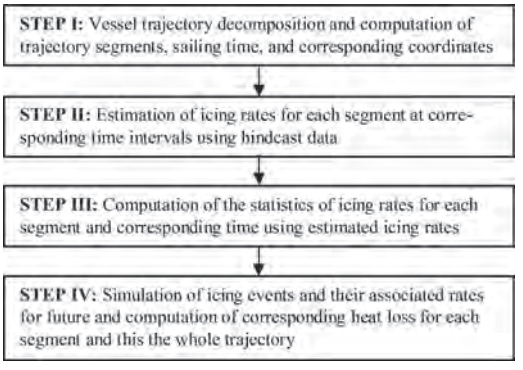


Figure 2. A systematic procedure for estimating expected heat loss for a given vessel trajectory.

Further, let trajectory $\overline{AB}$ be divided into N not-necessarily-equal segments $S_i = n_i n_{i+1}, i = 1, \ , N$, with length d_i and coordinates $n_i = (\alpha_{n_i}, \beta_{n_i})$, where,

$$n_1 = A \to (\alpha_{n_1}, \beta_{n_1}) = (\alpha_A, \beta_A) \tag{2}$$

$$n_{N+1} = B \to (\alpha_{n_{N+1}}, \beta_{n_{N+1}}) = (\alpha_B, \beta_B) \tag{3}$$

Note that, $\sum_{i=1}^{N} d_i = D$, with D being the rhumb line distance (Snyder, 1987) taken as the length of trajectory $\overline{AB}$ in km, given by (Alexander, 2004),

$$D = \frac{\pi R}{180} |\alpha_B - \alpha_A| \left| \frac{1}{\cos \omega} \right| \tag{4}$$

where, R = 6371 km is the earth's radius ω, in degrees, is the constant heading of trajectory $\overline{AB}$, i.e., the heading of rhumb line connecting A to B, clockwise from north. The constant heading is given by (Snyder, 1987),

$$\omega = \frac{180}{\pi} \tan^{-1} \left(\frac{x_B - x_A}{y_B - y_A} \right) \tag{5}$$

with,

$$x_A = \frac{\pi R}{180} \beta_A \tag{6}$$

$$y_A = R \ln \left[\tan \left(45 + \frac{\alpha_A}{2} \right) \right] \tag{7}$$

being rectangular coordinates of $A = (\alpha_A, \beta_A)$. While x_A lies along the equator, increasing towards east, y_A lies along the central meridian increasing towards north. Similar definitions stand for x_B and y_B.

In order to determine the coordinates of segment vertices, i.e., $n_i = (\alpha_{n_i}, \beta_{n_i}), i = 2, \ , N$, we can use the length of segment S_{i-1}, i.e., d_{i-1}, $i = 2, \ , N$, and the constant heading of trajectory $\overline{AB}$, ω, in degrees.

The latitude of the destination point of segment S_{i-1}, n_i, is given by (Kaplan, 1995),

$$\alpha_{n_i} = \alpha_{n_{i-1}} + \frac{180}{\pi} \left(\frac{d_{i-1}}{6371} \right) \cos \omega \tag{8}$$

where, d_{i-1} in km is the rhumb line distance between n_{i-1} to n_i. The longitude of the destination point of segment S_{i-1}, n_i, is given by (Kaplan, 1995),

$$\beta_{n_i} = \beta_{n_{i-1}} + \frac{180}{\pi} \left(\frac{d_i - 1}{6371 \cos \alpha_{n_{i-1}}} \right) \sin \omega \tag{9}$$

Once the coordinates $n_i, i = 1, \ , N+1$ are determined, the coordinates of midpoint of segment S_i, denoted by $n_i^* = (\alpha_{n_i^*}, \beta_{n_i^*}), i = 1, \ , N$, are determined by substituting half of segment's length into Equations (8) and (9), i.e., $d_{i-1} \leftarrow d_{i-1}/2$.

3.2 *Probabilistic representation of icing events and rates*

Probabilistic representation of icing events for a location includes representing two stochastic processes. In the first stochastic process, one predicts whether an icing event occurs, and during the second process, the amount of icing rate is predicted should the icing event occur. To this aim, assume that there exists sufficiently large number of icing event observations for a given location, using which both stochastic processes can be simulated.

For modelling the first process at a given location and time instant, let the probability of the occurrence of icing event be denoted by $\Pr(O = 1) = p$. Thus the complementary event (i.e., not occurrence of icing event) occurs with a probability of $\Pr(O = 0) = 1 - p$. For a given location and time instant, this process can be represented by a Bernoulli distribution, whose cumulative distribution function (CDF) is given by (Rausand and Høyland, 2004):

$$O \sim F_O(o) = \begin{cases} 1 - p & \text{for } O = 0 \\ 1 & \text{for } O = 1 \end{cases} \tag{10}$$

In order to simulate this process using a non-sequential Monte Carlo simulation, one can sample a random number form the CDF given by Equation (9). In other words, let $\zeta \sim U[0,1)$, then, icing event occurs if $\zeta \geq 1 - p$, and it does not occur, otherwise.

The second stochastic process initiates once the icing event occurs, i.e., $O = 1$. Once sufficiently large number of icing rate observations are collected, one can determine the empirical CDF of icing rate for a given location and time instant, i.e., $r \sim F_R(r)$, using which the amount of icing rate can be predicted.

3.3 *Heat loss estimation*

To estimate the amount of heat required for anti-icing, one can calculate the equivalent amount of heat released due to freezing, denoted by q_f in $\mathrm{Jm^{-2}\ s^{-1}}$ (Samuelsen et al., 2017),

$$q = L \cdot r \cdot \rho \tag{11}$$

where, ρ = 890 $\mathrm{kgm^{-3}}$ is ice density, taken constant, r is the ice accretion rate in $\mathrm{ms^{-1}}$, and the

$L = 2.3 \times 10^5$ Jkg^{-1} is the latent heat of freezing for saline-water ice (Samuelsen et al., 2017).

3.4 *Anti-icing expected heat loss for a given vessel trajectory*

Assume that the vessel departs from the origin A at time t_0. The vessel sails segment S_i with constant speed of V_i. By using the length of each segment and time of departure from origin, the time at which the vessel is sailing along segment S_i can be calculated accordingly.

By assuming that the vessel is exposed to the same level of icing continuously for the period it sails segment S_i, the amount of heat loss for segment S_i can be calculated by:

$$e_i = o_i \cdot q_i \cdot \frac{d_i}{V_i} \tag{12}$$

with $o_i = 1$ if icing event occurs, and 0 otherwise. By repeating the presented procedure for a sufficiently large number of times, an empirical CDF of e_i, $E_i \sim F_{E_i}(e)$, can be determined.

The amount of heat loss for the whole trajectory $\overline{AB}$, thus, will be given by:

$$e_{\overline{AB}} = \sum_{i=1}^{N} e_i \tag{13}$$

Mean, median, standard deviation, and quantiles of heat loss due to anti-icing can be extracted from empirical distribution of $e_{\overline{AB}}$ to represent the associate uncertainties with the occurrence and rate of icing events along the given trajectory. Another application of presented framework is identification of the most critical segment of the trajectory, from safety or heat loss viewpoint.

4 ILLUSTRATIVE CASE STUDY

Consider an operational site in the northern Barents Sea, 75.40N and 24.46E degrees, and Hammerfest 70.66N and 23.68E degrees; points B and A in Figure 3. A vessel leaves Hammerfest at 00:00 01.01.2018 towards location B. The aim is to simulate the occurrence of possible icing events along this trajectory and, in addition, simulate the expected heat loss associated with implementation of anti-icing measures.

4.1 *Data*

As discussed in Section 2, the MINCOG model is used in this study to model the sea-spray icing rate. As suggested by Samuelsen et al. (2017), median values of vessel speed $V = 4$ ms^{-1}, water temperature 2.5°C, and water salinity 35 ppt are chosen as

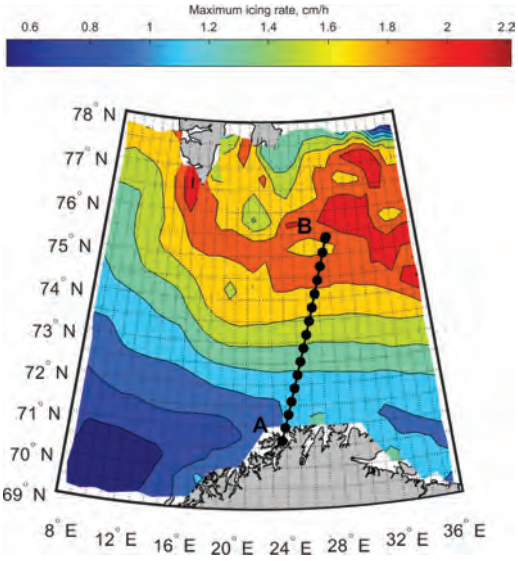


Figure 3. Maximum icing rate over the Barents Sea in January during 1980 to 2012. The solid black line shows the rhumb line of vessel trajectory $\overline{AB}$, decomposed into some segments showed by dark circles.

fixed input parameters. The main six input parameters including wind speed, air temperature, relative humidity, mean-sea level pressure, significant wave height, and significant wave period, are all derived from NORA10, obtained every 3 hours from 01.01.1980 00:00:00 to 31.12.2012 21:00:00.

4.2 *Analysis, results, and discussion*

The occurrence of icing events and their rates are subject to temporal and spatial variations over the Barents Sea due to changes in meteorological and oceanographic conditions. Figure 3 shows the maximum icing rate occurred in January during the period 1980 to 2012. As illustrated, the icing rate in the region is highest in the northeastern part. As shown, the maximum icing rate in January over the selected trajectory is also subject to considerable variations due to temporal and spatial changes in meteorological and oceanographic parameters.

Icing rate is lowest in the southwest due to higher air temperatures. This is only partly a result of the relatively high sea-surface temperature associated with the North Atlantic Current. However, the main reason is that these areas are located some distance away from the cold lands or sea ice, leading to the fact that the air has time to be sufficiently heated from below to avoid the high icing rates apparent in the other areas due to strong vertical mixing in weather situations in which icing occurs (Samuelsen and Graversen, 2017).

The icing rate is highest in the north and north-east due to lower air temperatures associated with strong winds and high waves that may arise during cold-air outbreaks from the ice. Since the ice edge normally is located just north of these areas exposed to severe icing, the air does not have time to be sufficiently heated from below in these areas to avoid severe icing. Sea-surface temperature alone does not have a large effect on icing (Samuelsen, 2017a). An interesting finding in Figure 3 is the maximum values obtained outside some of the fjords near the coast of Northern Norway. Such maximum values are associated with strong and cold gap winds out some of the large fjords generated in mountain wave situations during offshore flow from the Scandinavian archipelago described by Samuelsen and Graversen (2017). If using hindcast reanalysis data for atmosphere and ocean variables with higher spatial resolution than 10 km between the grid points, most likely this effect would have been more apparent. Thus, the maximum icing rates in these sea areas are probably underestimated.

In order to discuss the expected icing rate and heat loss over trajectory $\overline{AB}$, one needs to decompose it into different segments. The heading and length of trajectory $\overline{AB}$ is obtained using Equations (5) and (4), respectively, $\omega = 2.732$ degrees and $D = 653$ km. Thus, total sailing time of a vessel with speed of $V = 4$ ms^{-1} will be 45.35 h. Since the model input data are available every 3 hours, we divide the trajectory into equal intervals of d_i = 43.2 km, $i = 1,$,15, and d_{16} = 5 km (i.e., N = 16), and assume that the meteorological and oceanographic conditions remain constant during this time interval and along the segment. Coordinates of the vertices and midpoint of each segment are then computed using Equations (8) and (9). According to the vessel speed, one can determine at what time the vessel enters each segment and continues sailing towards destination.

MINCOG input data corresponding to each midpoint location, if existed, and the nearest location otherwise, are extracted from NORA10 database. Over the period 1980 to 2012, in total, 33 sets of input data are extracted corresponding to each coordinate and arrival time. Later, their corresponding icing rate is computed using MINCOG model. Figure 4 shows the frequency of the occurrence of an icing event, in per cent, at given locations and arrival times. These results are obtained based on the reanalysis hindcast data over the period 1980 to 2012 corresponding to certain arrival times at those specific locations. We assume that the probability of the occurrence of icing event, i.e., $\Pr(O = 1)$, in future is equal to its frequency, in per cent, over the past years, although this approach suggests that, for example

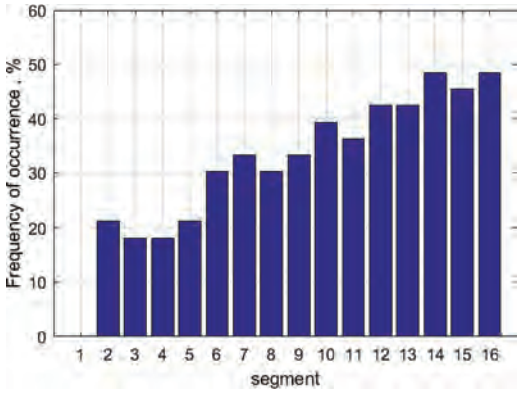


Figure 4. Frequency (probability of icing) in each segment at certain arrival times.

in segments S_{12} to S_{16} icing event occurs with certainty for specific arrival times to those segments. Thus, for instance, for segment S_2, the Bernoulli distribution of the occurrence of icing event can be given by,

$$O_2 \sim F_{O_2}(o) = \begin{cases} 1 - 0.2121 & \text{for } O = 0 \\ 1 & \text{for } O = 1 \end{cases}$$

Not that the overall trend of increasing icing rate in Figure 3, is also verified by the frequency of icing occurrence that increases from location *A* towards location *B*.

Empirical distributions of icing rates in each segment at given times, $F_R(r)$, is obtained by considering the occasions where $r \geq 0.05$ cmh^{-1} (Samuelsen et al., 2017). The modified box-plots in Figure 5 show the minimum, the 5th quantile, median, the 95th quantile, and maximum of icing rates in each segment. Segment S_1 is disregarded due to lack of weather data in its nearby locations.

The amount of heat loss for each segment is then calculated using Equations (11) and (12). For this purpose, a non-sequential Monte Carlo simulation is used, wherein, for each section, a realisation of icing event is sampled to simulate whether the icing event occurs.

Once an icing event occurs, another stochastic process is simulated and a random icing rate is sampled form the corresponding icing rate distributions, using which the required amount of energy for anti-icing applications is computed accordingly. This procedure is repeated for a sufficiently large number of times to obtain the statistics of the energy consumption for each segment, as the median, the 5th upper quantile and maximum values are presented in Table 1.

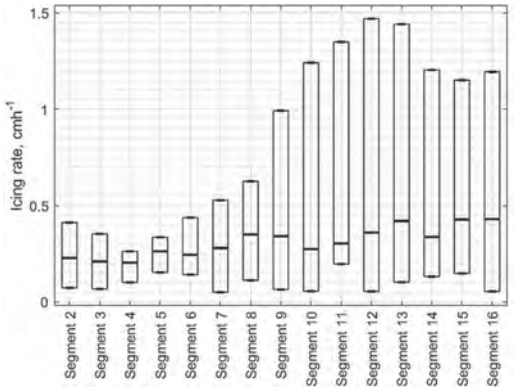


Figure 5. Box-plots showing the minimum, lower 5th quantile, median, upper 5th quantile and maximum values of icing rates for each segment at arrival times.

Table 1. Maximum, mean, and the 95th quantile of anti-icing energy consumption per area for each segment.

Segment, S_i	Energy consumption for each segment, e_i, MJm^{-2}		
	Maximum	Mean	95th Quantile
2	1.7283	0.2221	1.4708
3	1.6353	0.1745	1.3452
4	1.5429	0.1897	1.4747
5	1.7001	0.2966	1.6411
6	2.2691	0.4373	2.2318
7	2.0671	0.5098	2.8852
8	3.7993	0.6128	3.1980
9	4.0705	0.6325	3.0958
10	4.1505	0.6671	3.0741
11	4.0949	0.7177	2.9959
12	3.7782	0.7983	3.2998
13	4.0377	0.8963	4.0040
14	5.2738	1.0631	4.4079
15	5.0115	1.0447	4.0969
16	0.5971	0.1276	0.5574

Finally Equation (13) is employed to compute the overall expected energy consumption for all segments, i.e., trajectory AB, whose CDF is shown in Figure 6. Mean, median, the 5th upper quantile and maximum amount of energy consumption along trajectory AB are 8.40, 8.13, 15.00 and 29.24 MJm^{-2}, respectively. Depending on the decision-maker's risk perception approach and required safety guidelines, one may choose either maximum or some other statistics like the 95th quantile of expected energy consumption.

The maximum amount of icing rate and thus energy consumption, can be computed by adding up the required energy associated with maximum

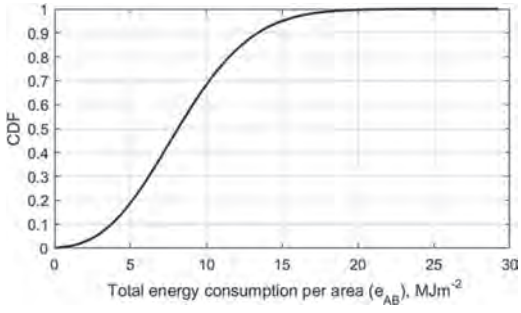


Figure 6. CDF of total expected energy consumption for anti-icing along trajectory AB.

icing rates computed using the data available for previous years. However, from a probabilistic viewpoint, the probability of having such a scenario is extremely low. Alternatively, a Monte Carlo simulation can be used to sample a sufficiently large collection of scenarios for icing events along the trajectory, based on which maximum amount of energy consumption for anti-icing is computed as around 29.24 MJm^{-2}.

5 CONCLUSIONS

In this study, a sea-spray icing rate prediction model, known as MINCOG, is employed to estimate the rates of icing events using reanalysis hindcast data. Such estimates are used to represent icing events and their rates probabilistically. Based on this procedure, a mathematical framework is proposed to simulate the icing events and their associated rates along a sea voyage through two stochastic processes. It further estimates expected energy consumption required to avoid icing on vessels. The results of this paper, illustrated by a case study, can be used in long-term decisions to be made in Arctic offshore logistics operations, such as scheduling and routing platform supply vessels, while reducing icing risks, operational costs, fuel consumption and CO2 emission. Future research on this area can be revolved around improving the simulation process of meteorological and oceanographic parameters, and thus icing events, by taking into account the effects of short-term patterns present in such parameters.

REFERENCES

Alexander, J., 2004. Loxodromes: A rhumb way to go. *Mathematics magazine*, 77(5): 349–356.

Borisenkov, Y.P. & Zablockiy, G.A. & Makshtas, A.P. & Migulin, A.I. & Panov, V.V., 1975. On the approximation of the spray-cloud dimensions (In Russian).

In: L.I. Romanova (Editor), Trudy Arkticeskogo, vol. 317. Antarkticheskii Nauchno-Issledovatelskii Institut, Gidrometeoizdat Leningrad, pp. 121–126.

DNV, 2013. DNV-OS-A201: Winterization for Cold Climate Operations. Det Norske Veritas As (DNV).

Farzaneh M., V.C.& Leblond A., 2015. Anti-icing and de-icing techniques for overhead lines. In: M. Farzaneh (Editor), *Atmospheric icing of power networks*. Springer, Dordrecht.

Fernández Cuesta, E. & Andersson, H. & Fagerholt, K. & Laporte, G., 2017. Vessel routing with pickups and deliveries: An application to the supply of offshore oil platforms. Computers & Operations Research, 79(Supplement C): 140–147.

Halvorsen-Weare, E.E. & Fagerholt, K. & Nonås, L.M. & Asbjørnslett, B.E., 2012. Optimal fleet composition and periodic routing of offshore supply vessels. *European Journal of Operational Research*, 223(2): 508–517.

Horjen, I., 2013. Numerical modeling of two-dimensional sea spray icing on vessel-mounted cylinders. Cold Regions Science and Technology, 93(Supplement C): 20–35.

Jones, K.F. & Andreas, E.L., 2009. Sea Spray Icing of Drilling and Production Platforms, US Army Engineer Research and Development Centre, New Hampshire, USA.

Kaplan, G.H., 1995. Practical Sailing Formulas for Rhumb-Line Tracks on an Oblate Earth. Navigation, 42(2): 313–326.

Kulyakhtin, A. & Tsarau, A., 2014. A time-dependent model of marine icing with application of computational fluid dynamics. Cold Regions Science and Technology, 104–105(Supplement C): 33–44.

Maisiuk, Y. & Gribkovskaia, I., 2014. Fleet Sizing for Offshore Supply Vessels with Stochastic Sailing and Service Times. Procedia Computer Science, 31(Supplement C): 939–948.

Naseri, M. & Barabady, J., 2016. On RAM performance of production facilities operating under the Barents Sea harsh environmental conditions. *International Journal of System Assurance Engineering and Management*, 7(3): 273–298.

Norlund, E.K. & Gribkovskaia, I., 2013. Reducing emissions through speed optimization in supply vessel operations. Transportation Research Part D: Transport and Environment, 23(Supplement C): 105–113.

Norlund, E.K. & Gribkovskaia, I. & Laporte, G., 2015. Supply vessel planning under cost, environment and robustness considerations. Omega, 57(Part B): 271–281.

Overland, J.E., 1990. Prediction of vessel icing for near-freezing sea temperatures. Weather and forecasting, 5(1): 62–77.

Rashid, T. & Khawaja, H.A. & Edvardsen, K., 2016. Review of marine icing and anti-/de-icing systems. *Journal of Marine Engineering & Technology*, 15(2): 79–87.

Rausand, M. & Høyland, A., 2004. *System reliability theory: models, statistical methods, and applications*. John Wiley & Sons, Hoboken.

Reistad, M. & Breivik, Ø. & Haakenstad, H. & Aarnes, O.J.& Furevik, B.R. & Bidlot, J.-R., 2011. A high-resolution hindcast of wind and waves for the North Sea, the Norwegian Sea, and the Barents Sea. *Journal of Geophysical Research: Oceans (1978–2012)*, 116(C5): C05019

Ryerson, C.C., 2008. Assessment of Superstructure Ice Protection as Applied to Offshore Oil Operations: Safety Problems, Hazards, Needs, and Potential Transfer Technologies, US Army Engineer Research and Development Center, New Hampshire, USA.

Ryerson, C.C., 2011. Ice protection of offshore platforms. *Cold Regions Science and Technology,* 65(1): 97–110.

Samuelsen, E.M., 2017a. Prediction of ship icing in Arctic waters – Observations and modelling for application in operational weather forecasting - PhD Thesis, UiT The Arctic University of Norway, Tromsø.

Samuelsen, E.M., 2017b. Ship-icing prediction methods applied in operational weather forecasting. *Quarterly Journal of the Royal Meteorological Society,* Accepted Manuscript Online 11 October 2017: n/a-n/a.

Samuelsen, E.M. & Edvardsen, K. & Graversen, R.G., 2017. Modelled and observed sea-spray icing in Arctic-Norwegian waters. *Cold Regions Science and Technology,* 134: 54–81.

Samuelsen, E.M. & Graversen, R.G., 2017. Weather situation during observed ship-icing events off the coast of Northern Norway and the Svalbard archipelago. Submitted to Tellus A: Dynamic Meteorology and Oceanography.

Samuelsen, E.M. & Løset, S. & Edvardsen, K., 2015. Marine icing observed on KV Nordkapp during a cold air outbreak with a developing polar low in the Barents sea, *Proceedings of the 23rd International Conference on Port and Ocean Engineering under Arctic Conditions (POAC)*, June 14–18, Trondheim.

Snyder, J.P., 1987. Map Projections—A wrking Manual, U.S. Geological Survey (USGS), Washington D.C.

Stålhane, M. & Vefsnmo, H. & Halvorsen-Weare, E.E. & Hvattum, L.M. & Nonås, L.M., 2016. Vessel Fleet Optimization for Maintenance Operations at Offshore Wind Farms Under Uncertainty. Energy Procedia, 94(Supplement C): 357–366.

Zio, E., 2013. *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*. Springer, London.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Safety of machinery—risk analysis and requirements for safety of gravity loaded axes

Luca Landi
Department of Engineering, University of Perugia, Perugia, Italy

Heinrich Mödden
VDW, Frankfurt am Main, Germany

Iuri Betti
SCM Group S.p.A., Rimini, Italy

Martin Kohnle
MAG-IAS, Eislingen, Germany

Rüdiger Knorpp
Heller Maschinenfabrik, Nürtingen, Germany

Armin Bornemann
Deckel Maho Pfronten GmbH, Pfronten, Germany

Peter Steger
Grob-Werke, Mindelheim, Germany

ABSTRACT: One of the design measures adopted for the risk reduction of machinery is the de-energization of all the components of the machine for assuring a "safe stop state". So, for gravity loaded axes, the risk of unintended gravity descent in the de-energized state has to be considered in the risk assessment. In case of power failure, the gravity loaded axes are held solely by the brake/counterweight systems, which are installed in the machine. The gravity loaded axes may drop down, if the existing brakes/counterweights do not provide adequate protection against unintended descent due to gravity.

In the paper first, a full analysis for this underestimated risk of today's very complex machines is presented. Than a description of the problem is illustrated and the resulting requirements are presented in terms of: design measures, information to be given to the end user and testing procedures to be assured for the new machines.

1 INTRODUCTION

While it can be assumed that during horizontal movements in the automatic production no hazards to persons occur due to gravity in the de-energized state, for vertical movements. However, the risks of unintended gravity descent have to be considered in the risk assessment, see DGUV (2012). These hazards particularly become obvious with linear robots for the handling of heavy parts (Fig. 1), but also with jointed-arm robots or inside machines, e.g. at vertical axes of any machine tools. If the existing brakes/counterbalancing systems do not provide sufficient protection against unintended gravity descent, control measures can contribute to reduce the risk of hazard in any case.

So, when the machine systems are in an energized state in any mode of operation, it can be usually assumed that the risk assessment performed by the machine builder assures a proper safety of the machine itself.

The aim of this article is to give a guide to machine builders for design of "state of the art" Gravity Load Axes (GLA) for machine tools. The risk assessment and the design solutions presented were mainly discussed during ISO standardization works of ISO 16090-1 (2017), but the results are of general applicability for all the types of machine tools.

The typical iterative three step method used for risk reduction, see ISO 12100 (2010) assures, in conjunction with the hypothesis of full disposal

Figure 1. Robot arm used for handling of heavy parts, from DGUV (2012).

of machine energy, the proper reduction of risk of uncontrolled descend of all the gravity loaded axes.

In case of power failure or energy removal (for example during maintenance), gravity loaded axes (weight-loaded, vertical, slant axes) are held solely by their brake/counterbalancing system installed in the machine.

The gravity loaded axes could descent in case of failure of the retention system and, usually, also the control/warning system, could be off-line due to the complete de-energization of the machine.

As is simple to understand, the risk reduction for this potentially unsafe condition cannot be satisfied also though some of the typical step two risk reduction means: "other protective measures", because they are also usually ineffective (e.g. light barriers, emergency stop buttons.…).

As an example, for vertical axes with braking systems the mechanical wear or oil-fouling may cause the braking torque/force of the brakes to fall below its nominal value which may result in an unintended descend of the gravity loaded axes.

In de-energized state the risk reduction cannot be achieved by machine builder through safety functions: only inherently safety design measures, physical guards, and information for the user are available. It is to say that, for the risk reduction of this potentially unsafe "state", the end user behaviour is essential during the utilization of the machine: correct procedures of maintenance and frequent inspections are essential for reduce probability, severity and occurrence of the risk of descend of GLA in de-energized state.

In the following paragraphs first, the risk analysis performed for safety standardization will be presented. Than the "state of the art" and future coming technical possible solution needs for risk reduction will be shown, using also an easy to be read table format.

This table format was preferred by the authors also during ISO standardization works to assure excellent clarity/effectiveness/brevity necessary to be understood to machine builder design departments.

Depending on the technical application and the risk to be reduced, different technical safety devices are suitable to prevent the unintended gravity descent of gravity-loaded axes.

Safety functions related to gravity loaded or slant axis will be also presented and discussed. At the end some examples of already in the marked design solutions will be presented for clarity. The presented design measures will be introduced in annexes of safety standards for machine tools such as the ISO 16090-1 (2017). This latter new ISO standard was published in December 2017.

2 RISK ANALYSIS FOR GLA

For the risk assessment the tables in new annex G of the ISO 16090-1 (2017) shall be used. The annex will help the designer to find the "perfect solution" for the necessary risk reduction. The tables cover the foreseeable operation of the machines as well as maintenance, cleaning and repair of the machines.

The risk for the GLA hazard started from precondition of machine design, i.e. the different safety condition of operators between all operational mode and maintenance. For more information see the table G.1 and G.2 on annex G of ISO 16090-1 (2017).

During maintenance no energy supply is present, so Numerical Control (NC) and Safety Function (SF) are not generally available. The safety of the workers it is accomplished during maintenance through inherent safe measures, such as mechanical support (e.g. using struts) of the axes loaded by gravity or equivalent measures.

Sometimes a direct safe support of the GLA is not the best technical solution because, as an example, it might not be possible, due to space limitation, to have adequate space left for the worker in the maintenance zone with the supported axis. In this case the mechanical locking of a different component directly connected to the GLA must be assured.

It is possible that:

- an appropriate mechanical support equipment (i.e. a strut) shall be designed and provided by the manufacturer of the machine, or
- a proper locking system (position) shall be designed and provided by the manufacture or
- at least, if no "special equipment/locking system is provided by the manufacturer, information on how to provide a proper support shall be given (see Table G.2 situation G1.3 and G1.4).

For all others operative modes, it was assumed that the most critical situation for hazard is full body access in the work area under a GLA (see G1.1).

Using as an example ISO/TR 14121-2 (2012) for risk assessment the authors stated that if only the upper limb stays under the GLA of the machine (G1.2) no serious (S2) injury is expected for the worker. Even if sometime the hypothetical severity is S2 than the velocity of the GL descending axis is not high, and the worker have not to fully move the body to escape to the risk. So, it is possible to avoid/reduce the harm and use A1 for possibility of avoidance/reduce the harm for risk the risk graph, see again Figure 3 of ISO/TR 14121-2 (2012).

Moreover, even if for some machines the hypothetical severity can be higher (S2), the necessity for cycling testing of "single channel/unsafe systems", such as a single brake (design V1 of Table G.1.), leads always to low occurrence of the hazardous situation (O1).

The hazard become a real risk only if:

- the machine goes in de-energized state and
- the worker is currently under the GLA and
- there is a not already evident failure on braking system, so we are between two cyclic tests. As one can see in table G.3. if the workpiece is manually loaded (i.e. frequent exposure to the risk without possibility to avoid it) the cycle testing time is 8h, otherwise for automatic loading of workpiece, so short presence, testing time is 48h.

It should be mentioned that, for correctly maintained/tested systems, usually the typical failure of braking of GLA are due, as an example, to wear and/or leakage (respectively for mechanical and/or pneumatic systems). So, no free falling of the GLA is expected.

At the end of this paragraph authors wants to emphasize that: a correct utilization/maintenance of the machine is a crucial point for safety of the workers for GLA risk. The defeating of the safety systems by the user is very dangerous for those systems.

2.1 *Safety functions for GLA*

It ought to be said clearly that it is not possible to design any SF for GLA in de-energized state of the machine, which is the main hazardous situation in operational field. As mentioned before, when the energy is lacking from the system only inherent safe measures shall be used if required.

When the machine is in de-energized state (often called the weak state) or by disturbance of the energy supply, only the mechanical braking system can work properly (e.g. brake, clamping etc.).

The faultless function of this mechanical braking system has to be checked in periodical distances (cyclic testing). However, this diagnosis function has no real SF according ISO 13849-1 (2008) but guarantees the faultless function for the weak state.

So, all the SF covering the hazard of GLA are designed in energized state of the machine. The matching safety functions (SF) are defined in the annex J, Table J.3 of ISO/FDIS 16090-1. For information on this annex see Bornemann A. & al. (2015a, 2015b).

As an example a "safe" stop function for a GLA, when only the upper limbs are under the GLA, can be accomplished in performance level c (PL_r = c, see ISO 13849-1 (2008)) with, at least, two different designs: the system is brought to a standstill, for example by the opening of an interlocked guard and is then held in position by a de-energized clamping or by an integrated SF category 2 stopping with a monitored SOS (see IEC 62061:2005 + A1 (2012)).

For more examples such as the prevention of unexpected start-up see Table J.3 of ISO 16090-1 (2017).

3 DESIGN OF BRAKING SYSTEM

The authors considered a lot of different braking systems for GLA, see the different design solutions in Table G.1. column 1, from V1 to V7.

The state of the art for machining centres is:

- single or redundant systems based on pure mechanical brakes. The redundant brake can be internal to the electrical motor or external to it. The internal redundant brake can be used for axis movements if no additional risks is foreseen. As an example, a fault of the electrical motor shaft due to fatigue. The mechanical parts of power transmission shall be at least designed with double weight load to withstand the occurring static and dynamic stresses. Moreover, if the same shaft is used for redundant internal brake the situation of full braking with only the redundant system (the added one) have to be taken into account for mechanical strength calculation, if a longer shaft is expected in this case (resulting in a greater torsional deflection). In order to prevent unnecessary wear of the

brakes, it is preferable to decelerate with the electrical drive controller instead of stopping with mechanical brakes.

- "hybrid systems" based on mechanical brakes in conjunction with counterweights. In this case the authors stated a difference between pure mechanical counterweights, usually designed by proper balancing masses connected to the GLA, and hydraulic/pneumatic counterbalancing systems. The reliability of pure mechanical counterweight is based on static/fatigue calculation of mechanical components, whose failure analysis is quite simple to be taken into account. There is a lot of literature for "usable safety". Safety coefficients, can be derived, as an example, from the design/safety of mechanical lifting systems. Conversely in pneumatic/hydraulic counterweight systems more non-return valves and other components were introduced in the field of "safety" system recently. Also "failure and effect methods" such as FMEA, see Stamatis, D.H, (2003), for those components is more complex with respect of simple balancing masses systems. So, it was decided that, if the manufacturer cannot explicitly justify the fault exclusion for hydraulic/pneumatic system (V6 of Table G.1), a hydraulic counterweight cannot be used when the worker whole body can be exposed to the hazard. An example of a proper hydraulic counterweight system for GLA where fault exclusion can be managed properly will be presented in paragraph 5 below;
- redundant system with one brake and one external clamping device (see V7 of Table G.1.). In this case the external clamping system is designed completely separate from the motor braking system. If a dangerous fault exclusion can be assessed from the designer the cyclic test can be avoided. It needs to be mentioned that, for external clamping, the cleanliness of every part of the clamping device is a key factor for a proper braking in a coolant environment of machine tools. In this case the cleanliness instruction should be properly defined in the instruction manual.

As one can see the V1 design is not suitable if whole body access is foreseen (situation G1.2). Due to simple tabular format it is not explained that this is not suitable due to frequent access during automatic mode because of frequent full body access without other measures possible.

For repair and maintenance additional measures are necessary, e.g., underpinning, mechanical locking, hanging as we will see in the next paragraph.

4 ADDITIONAL DESIGN MEASURES

As all safety engineers know that the inherent safe measures (see step 1 of risk reduction process of ISO 12100) usually are not sufficient to reduce the risk to a tolerable level. So other additional measures have to be assigned during the subsequent step 2 of risk reduction process.

Depending from the design of the braking system additional measures have to be selected. In Table G.2, the additional measures are described depending from the same situation (Gx.x) of the Table G.1.

The authors defined those measures using some guiding factors such as: minimize the gravity stored energy, maximize the stability of the system especially during maintenance. At least, information and measures for preventing the misuse during manual intervention on the system have to be provided by the machine manufacturer.

The locking of the GLA with a mechanical lock during maintenance is not only assuring that no intentional restart of the system is done, but also is a "guide" for the end user to stop the axis in the correct position (i.e. the correct maintenance location chosen by the designer of the machine).

It should not be forgotten at this point that parking the axis in the "lowest position" is the safest measure at all, because the axis cannot fall under this position.

Finally, according to step 2 of the risk reduction process, one or several warning signs shall be visibly fixed at the machine pointing out to hazards due to GLA and suspended loads. As an example, "Do not stay underneath the vertical axis!".

The same should be reported in the instruction manual giving also advice for safe working practices.

4.1 *Instruction for use, key role of the end user*

As all the national assurance of the workers reports shows, see as an example INAIL (2015), a lot of injuries are caused by workers defeating and/or misuse of the machine.

In the particular case of hazards related to GLA the user has a key role to maintain the safety of the system over the time. If the correct maintenance of the system is not performed also the safer redundant braking systems can by ineffective, especially for problem related to clean and wear. Also, small fluid leaks of hydraulic circuits that are not removed can cause an ineffective braking torque (force) of the system.

So, the machine manufacturer has to define, how normal operation, repair, cleaning shall be carried out safely by the machine user. It has to be remembered that maintenance, cleaning and repair works

are carried out at or next to the gravity-loaded axis. Usually a safe mechanical support of the gravity-loaded axis is easily feasible and consequently it has to be done for the sake of safety.

Operating instructions shall describe measures to protect the operator from a fall-down of GLA. These instructions shall also point out to hazards due to gravity-loaded axes and suspended loads. Also, the required skill level of the operators needs to be considered. If the brake is removed for maintenance, a support or a manual mechanical lock shall be used also in the designed system where fault exclusion can be done (see V3-V6-V7 in Table G.1).

For extensive additional measures required for the different operative/inoperative modes, see Table G.2.

4.2 *Cyclic testing and testing of torque/force*

The need of this diagnosis function is considered in the column "Requirement for cyclic test", see Table G.1. The maximum tolerable time span between tests depends mainly on frequency of exposure to the GLA risk for the worker, see Table G.3.

As mentioned before it is very important to understand that cyclic testing is not a SF itself and that it can be done with NC. The cyclic testing is always performed in safety condition (e.g. with machine doors closed). Cyclic testing is performed for the brake system, when fault exclusion cannot be done, at a predefined time span.

The test has to be able to measure the braking performance of the system over the time: a test torque/force is applied to the brake, e.g. motor brake or the clamping device.

Because the test condition of the ISO 13849-1 (2015) is not applicable for cat. 2 systems of braking test (i.e. mainly because we cannot test the brake 100 times before it is used again), a proper specification is defined in ISO/FDIS 16090-1 (2017). A sudden complete failure of a brake with the force actuated by a spring can be excluded because of the basic principles of the mechanical brake design of ISO 13849-2 (2012).

For torque/force testing the following requirements apply:

- 1 motor/1 brake (or clamping device) systems. The brake or the clamping device is charged with 1,3 times the maximum gravitational load for at least 1 s by the electric drive. If also a permanently present counterweight system is installed, the braking device is charged with 1,3 times the maximum gravitational weight minus the counterbalanced weight.
- 1 motor/2 brakes (or clamping device) systems. The braking devices are tested separately one after the other on 1,0 times the maximum gravitational load.
- 2 motors/2 brakes (or clamping device) system mechanically connected. The braking devices are tested together on 2,0 times the maximum gravitational load or one after the other on 1,0 times the maximum gravitational load.

All the requirements for brake test are defined in annex G of ISO 16090-1 v(2017), it is important to note that, again, the safety of the worker during the test is mainly defined by the GLA position:

- *before the test,* the GLA must be placed in a proper position where no hazard for the worker is foreseen, even if the test is failed,
- *during the test,* no additional hazard should arise from the failure of the test. As an example, the designer should take care of possible tools/parts breaking for undesired falling of the GLA during the test with sharp object contours,
- *after the test,* the machine must be placed in a safe state before any worker can enter inside the machine and further operation of the machine shall only be possible after a new successful cyclic test.

In case a fault detection occurs during the cyclic test the NC shall inform on the screen for a brake repair. In case of guards with closed and protective doors, a safe position shall not be approached until an unlock demand signal has been given. Again, further operation of the machine shall only be possible after a new successful cyclic test.

5 EXAMPLES OF EXISTING SYSTEMS

In this paragraph some design examples for GLA braking system are presented.

5.1 *Example of a single brake system—with or without fault exclusion*

In Figure 2 the typical V1 system of Table G.1 is presented. This system can be used only if the whole-body exposure to the GLA is not feasible during operative phases. Usually during setting, repairing works and so on, sometimes a staying for short time under the gravity loaded axis is necessary. Then a single brake with cyclic test is also a suitable design measure. It is to be remembered that during setting, as an example, usually additional measures have to be taken, such as, an operator pendant and/or reduced axis speed (leads also to shorter braking distance). Due to those latter additional measures a greater probability of avoidance of the accident is expected.

In Figure 3 the typical V2 system of Table G.1 is presented. A redundant brake system is safer with

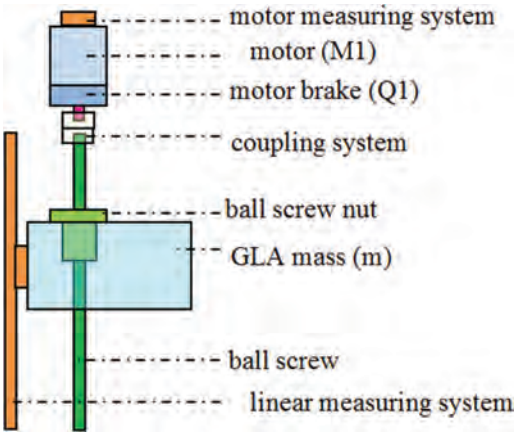


Figure 2. Single brake system without fault exclusion (V1).

respect to V1, even if no fault exclusion is made as in the redundant system V2. As one can see, all the parts of the braking system are redundant in this figure also the clamping stick attached to the fixed part of the machine. This is so, because all the possibilities of failures of braking cannot simply be excluded (as an example due to leaking of coolants lowering coefficient of friction).

It is very important that the risk assessment needs to be done for the V3 design under the condition of fault exclusion e.g. ISO 13849-2 (2010). However, if a single fault to a component cannot be excluded than a partial redundancy of the system is necessary.

5.2 *Redundant brakes system*

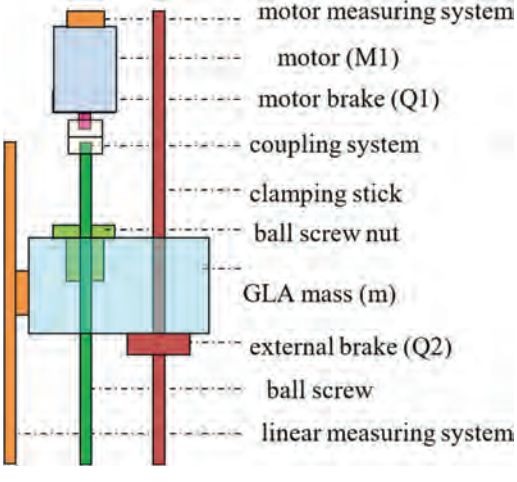


Figure 3. Redundant brake system without fault exclusion (V2).

5.3 *Counterweight systems*

As mentioned above a counterweight system is often used in machine GLA tools technology also to improve the dynamic capabilities of heavier axes. A GLA counterbalancing with a proper system results in small unbalanced masses to be moved. So, smaller motors can be used for axis movement and/or greater accelerations is foreseen with respect to the same GLA without counterbalancing system.

In Figure 4 a typical single counterweight system with external clamping is presented as an example of system V5.

Looking again to Table G.1. from V4 to V6 some simple conclusions can be derived:

- if it is not possible to prevent the staying of the full body of the worker under the GLA, it is not possible to use simple hydraulic/pneumatic system without fault exclusion, even if cyclic testing is performed,
- a mechanical counterweight without fault exclusion can be used in conjunction with a braking motor, but, in this case, a cyclic test is required. Basic and well-tried safety principles in the mechanics have to be used. Also, classical safety factors in the interpretation of the mechanic components, e.g., cable (rope), clamping device, bearing system have to be done. There is not safety control system for the counterweight that is required to be assembled only by mechanical parts. In this case the full system has 2 different and huge inertias to be supported, the system inertia and the counterbalancing inertia. The

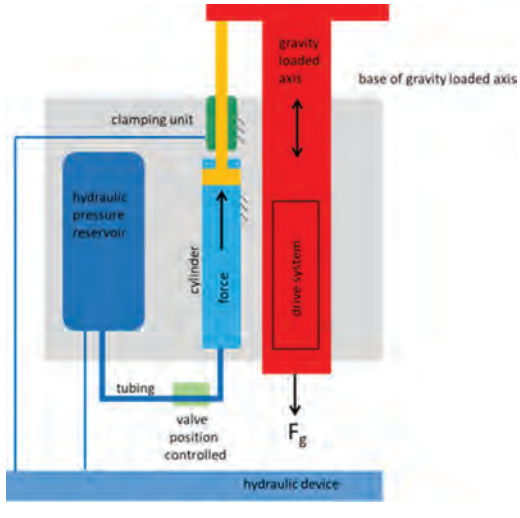


Figure 4. Single counterweight system based on hydraulic/pneumatic (V5).

validation has to be done, as usual, according to ISO 13849-2 (2012), annex "A" for mechanical systems.

For hydraulic counterweight systems the most important aspect to be considered is the requirement of fault exclusion for the non-return valves of the counterbalancing system.

Because it is considered possible a pressure leakage due to different events (breaking on a pipe, leakage of a valve, …), the piston(s) shall be sufficiently protected from non-closure of non-return valves.

Typical design measures to avoid dangerous failures (or dangerous interference on correct operational behaviour of the system) are:

- devices for fluid temperature monitoring (if hydraulic),
- device for pressure monitoring,
- devices for fluid pollution monitoring (filters up to 1 micron), because complete valves closure can be prevented by pollution particles dispersed in the fluids. For the faults related to problems of valves aging, see Schuster, U., (2004).

In Figure 5, a fully redundant counterweight system, with also redundant clamping unit, is shown.

If correctly designed this system could be appropriate for V6.

It has to be said that, during operative phases, the drive systems and the correct pressure of the fluid, assures GLA movements with counterbalancing. For safety aspects, the load capacity of the system with closed non-return valves is the crucial argument: during de-energization the balancing force is not acted through hydraulic system, but directly in the pressurized pistons by closed valves.

Depending on the applied principle for fault exclusion some of the double component of the system in Figure 5 can be avoided, such as the double reservoir.

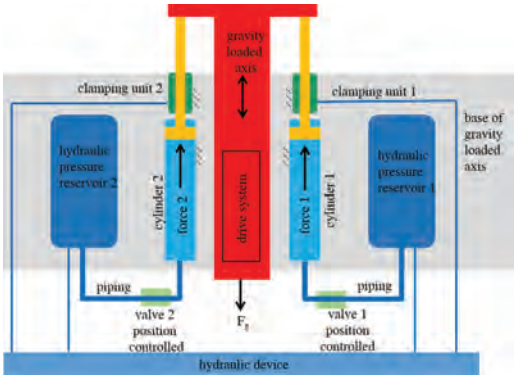


Figure 5. Fully redundant counterweight system based on hydraulic/pneumatic (V6).

As examples, the counterbalancing system with fault exclusion to non-return valves can be:

- a single piston, system capable to support with the non-return valves closed (with no motion/limited motion at low speed) the 1,3-full weight of the GLA. In this case, even if the motor or external clamping has a fault, the hydraulic system can stop the descending axis during the de-energized state. The valves closure shall be double monitored with cross monitoring. Moreover, it must be possible to make a fault exclusion to all relevant mechanic/hydraulic components of the system safety related (see below). Even if theoretically possible at the current state of the art a complete fault exclusion for this system is still difficult to be done.
- a redundant piston/brake system, each of them capable to support with the non-return valves closed (with no motion/limited motion at low speed) 1,3 the full weight of the GLA. Each valve closure shall be monitored with safety functions in $PL_r = c$.

In relation to non-returning valves, a fault exclusion of the mechanical spring of the valve shall be done by the valve manufacturer (the spring can be considered a well-tried component also for hydraulic, see Table B.6 of ISO 13849-2 (2012)).

For the validation of hydraulic system, e.g., failure exclusions for pipes related with safety according to ISO 13849-2:2012 Table "C.7" and failure exclusions for hydraulic connecting elements of Table "C.9" are necessarily in any case. Safety valves shall be firmly connected with the cylinders in order to avoid safety problems with connection pipes.

6 CONCLUSIONS

The risk analysis and the design measures presented in this article were initially developed as a part of the standardization works of ISO/TC 39/SC 10/WG4, "Safety of machining centres, milling machines, transfer machines for cold metal materials". The authors have been involved in this standard and they think that this risk assessment can be adopted also for other types of machine tools.

Because of its innovative subject for standardization purpose, the DGUV (2012) document has been taken as a basis for the discussion. This document needed to be clarified by the authors also with examples foreseeing the publication of a normative annex on GLA of a new standard. The authors believe that the risk analysis and technical

solutions presented in the paper provide an objective view of the state of the art and design solutions that can be used to effectively reduce the risk due to GLA.

The authors encourage the drafting of similar regulatory annexes for the C-type standards of forthcoming publication in the field of machine tools safety.

REFERENCES

Bornemann A., Froese Y., Landi L., Mödden H., 2015a. Probabilities in safety of machinery-Part 1: Risk profiling and farmer matrix, Safety and Reliability: Methodology and Applications—Proceedings of the European Safety and Reliability Conference, ESREL 2014, CRC Press/Balkema, pp. 1933–1942.

Bornemann A., Froese Y., Landi L., Mödden H., 2015b. Probabilities in safety of machinery-Part 2: Theoretical and practical design, Safety and Reliability: Methodology and Applications—Proceedings of the European Safety and Reliability Conference, ESREL 2014, CRC Press/Balkema, pp. 1943–1950.

Directive 2006/42/EC, 2006. Machinery directive of the European Parliament.

DGUV Division information sheet No. 005, 2012. Edition 09/2012 Page 3/7 Gravity-loaded axes - vertical axes.

EN ISO 12100, 2010. Safety of machinery – General principles for design – Risk assessment and risk reduction.

EN ISO 13849-1, 2015. Safety of machinery – Safety-related parts of control systems – Part 1: General principles for design.

EN ISO 13849-2, 2010. Safety of machinery – Safety-related parts of control systems—Part 2: Validation.

IEC 62061:2005 + A1 2012. Safety of machinery—Functional safety of safety-related electrical, electronic and programmable electronic control systems. International Organization for Standardization, Geneva, Switzerland.

INAIL, 2015. 8° Rapporto INAIL sulla sorveglianza del mercato per la direttiva machine, in italian language.

ISO 13849-1 2008. Safety of machinery—Safety-related parts of control systems—Part 1: General principles for design. International Organization for Standardization, Geneva, Switzerland.

ISO 13849-2, 2012. Safety of machinery – Safety-related parts of control systems – Part 2: Validation. 2nd Edition. General principles for design, International Organization for Standardization, Geneva, Switzerland.

ISO 16090-1, 2017. Machine tools safety—Machining centres, Milling machines, Transfer machines—Part 1: Safety requirements.

ISO/TR 14121-2, 2013. Safety of machinery—Risk assessment—Part 2: Practical guidance and examples of methods. International Organization for Standardization, Geneva, Switzerland.

Schuster, U., 2004. Untersuchung des Alterungsprozesses von hydraulischen Ventilen BIA-Report 6/2004. Hrsg.: Hauptverband der gewerblichen Berufsgenossenschaften (HVBG), Sankt Augustin 2004. ISBN: 3-88383-672-9 (only in German).

Stamatis, D.H, 2003. Failure mode and effect analysis: FMEA from theory to execution, second edition, ASQC Quality Press Milaukee, WI.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Risk significance assessment with operational events of Korea nuclear power plants

Seunghwan Kim, Sun Yeong Choi, Sang Hoon Han & Jaewhan Kim
Korea Atomic Energy Research Institute, Daejeon, South Korea

ABSTRACT: A handbook was published to document methods and guidance that NRC staff should use to achieve more consistent results when performing risk assessments of operational events by the U.S. NRC. Korea Atomic Energy Research Institute (KAERI) launched a research project to develop a regulatory purpose Level 1 PSA (Probabilistic Safety Assessment) model and framework for use in risk-informed regulation. To this end, we designed a regulatory risk model reflected regulatory purposes based on the real conditions and developed a regulatory software called RYAN (Risk Analysis for ASP/SDP of NPP) that enables the regulatory body to perform the overall safety assessment such as ASP/SDP (Accident Sequence Precursor/Significance Determination Process). In order to verify and validate the RYAN software, we investigated operational events occurred in domestic NPPs from databases such as OPIS (Operational Performance Information System for Nuclear Power Plant) and KRDB (Korean Integrated Reliability Database). From those nuclear event databases, we selected some component failures and IEs for the software verification and validation. We performed a sensitivity analysis for the various cases with the selected operational events. Based on the framework, it is expected that the regulatory staff can identify a nuclear power plant or SSC (Structure, System, and Component) for which safety performance has been decreased. For a further study, we are confirming the applicability of RYAN by performing a sensitivity analysis with more event data. In addition, we are to analyze the significance of each case with the AIMS-PSA to compare the results from RYAN and AIMS-PSA for ensuring the accuracy of the analysis results with RYAN.

1 INTRODUCTION

US.NRC provides the Risk Assessment of Operational Events Handbook (RASP Handbook) to assist NRC staff to achieve more consistent results when performing risk assessments of operational events and licensee performance issues. The methods and processes described in the RASP handbook can be primarily applied to risk assessments for Phase 3 of the SDP (Significance Determination Process), the ASP Program, and event assessments under the NRC's Incident Investigation Program. For example, Figure 1 depicts the criteria to determine the level of safety significance to characterize the safety significance of inspection findings for the NRC ROP by assigning a color to the inspection findings (U.S.NRC 2013).

Korea Atomic Energy Research Institute (KAERI) launched a research project to develop a regulatory purpose Level 1 PSA (Probabilistic Safety Assessment) model and framework for use in risk-informed regulation (J. Kim et al, 2017). The purpose of this research is to estimate the risk significance of initiating events and degraded conditions occurred in domestic NPPs (Nuclear Power Plants) and characterize the significance of an inspection finding consistent with regulatory response thresholds such as SDP (Significance Determination Process) of the US NRC.

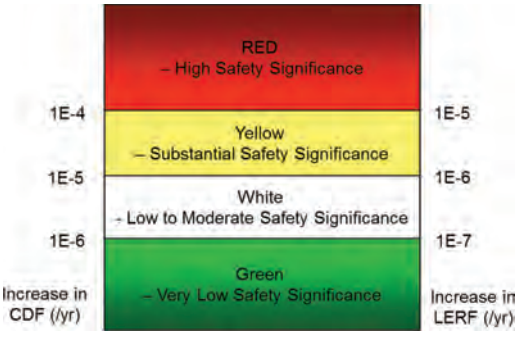


Figure 1. Criteria to determine the level of safety significance.

2 DEVELOPMENT OF A REGULATORY RISK ASSESSMENT ALGORITHM

As this research aims to estimate the risk significance and to characterize the significance of an

inspection finding consistent with regulatory response thresholds, it is essential to develop a regulatory risk assessment logic to satisfy the requirements of risk estimation for regulatory purpose.

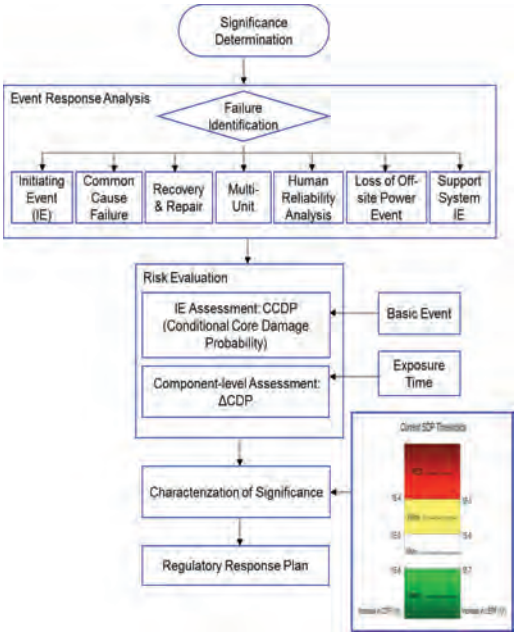


Figure 2. Schematic diagram of algorithm for regulatory risk assessment.

Figure 2 represents the schematic diagram of the algorithm for regulatory risk assessment. The analysis of the risk significance consists of the following procedures. First, the type of failure is determined by performing failure identification.

The types of failures include 'Initiating Event', 'Common Cause Failure', 'Recovery and Repair', and 'Human Reliability Analysis'. For example, the failure caused by one of initiating event or component failure. After the failure analysis is completed, a risk analysis is performed in which the associated failure calculates the Conditional Core Damage Probability (CCDP) for the initial event or evaluates the component level in case of failure in the unit of equipment.

At this time, the risk assessment is performed by modifying the basic event probability (BEP) for the initial event and the inoperable (out of service) time when the related component is exposed to the fault. Finally, the risk assessment is performed using the risk assessment results and a regulator response plan is determined.

Figure 3 shows the procedure for evaluating the risk increase due to reactor shutdown and SSC failure: 1) Calculate the delta CDF (Risk increase by Degrade SSC) by performing a PSA model modification to change the BEP and CCF values of the equipment for the SSC degraded condition including the SSC fault or out of service (OOS) due to maintenance. 2) Calculate the delta CDF (Risk increase by Degrade SSC) by performing a

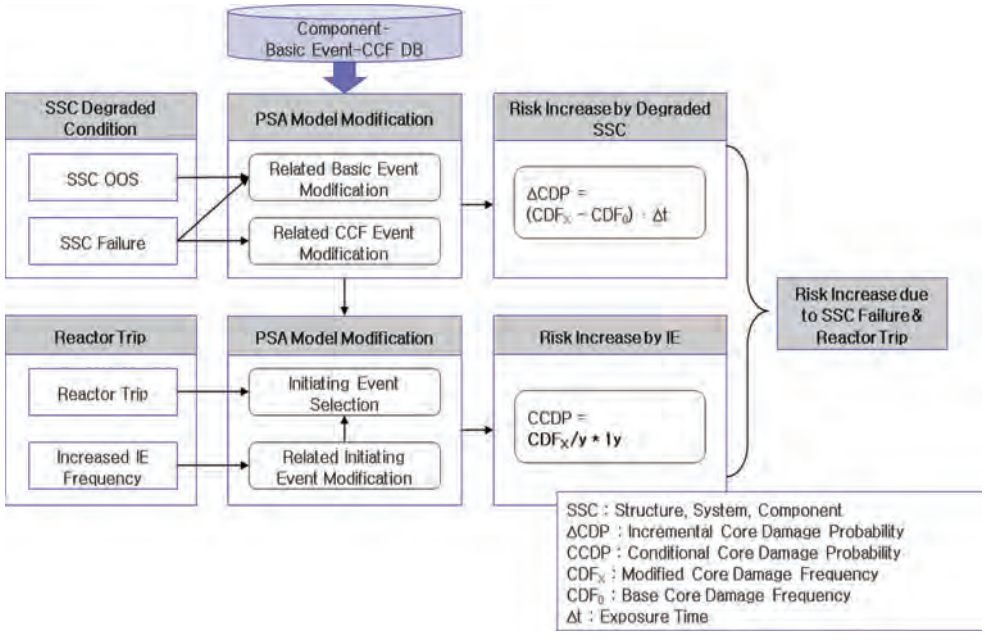


Figure 3. Process of increased risk evaluation due to SSC failure and reactor trip.

Table 1. CCDP calculation for sensitivity analysis.

Case	Calculate CCDP (Conditional Core Damage Probability)
Case 1: IE only	[1] CCDP Calculation – By setting the observed IE to 1.0 and all other IEs to 0.0
Case 2: IE & mutually exclusive SSC Unavailability (SDP only)	[1] CCDP Calculation [2] $\Delta CDP\ [(CDF_x - CDF_0)\Delta t]$ Calculation for the SSC unavailability – By setting the basic event associated with the SSC unavailability to TRUE [3] Total Risk Calculation, $\Delta CDP_{Total} = CCDP + [CDF_x - CDF_0]\Delta t$
Case 3: IE & Mutually Inclusive SSC Unavailability	[1] CCDP Calculation for the combined IE and SSC unavailability – By setting the observed IE to 1.0 and all other IEs to 0.0 – By setting the basic event associated with the SSC unavailability to TRUE [2] ΔCDP Calculation for the SSC unavailability only [3] Choose the highest of the CCDP or ΔCDP result
Case 4: SSC Unavailability Increases the IE frequency (No IE Occurred)	[1] Baseline system failure prob. estimation by solving an applicable FT [2] Calculate the system failure probability factor (or ratio) – By setting the basic event to TRUE – Calculation of system failure probability factor (new value/baseline system failure probability) [3] The modified initiating event frequency calculation – By multiplying system failure probability factor with the baseline IE frequency [4] Calculate ΔCDP for degraded condition

PSA model modification to change the BEP and CCF values of the equipment for the SSC degraded condition including the SSC fault or OOS due to maintenance.

In order to carry out these procedures, the following cases are evaluated. From those nuclear event databases, we selected some component failures and IEs for the software verification and validation. We performed a sensitivity analysis for the four kinds of cases with the selected operational events. Table 1 shows the CCDP calculation method for each case

- Case 1: IE only
- Case 2: IE & mutually exclusive SSC Unavailability (SDP only)
- Case 3: IE & Mutually Inclusive SSC Unavailability
- Case 4: SSC Unavailability Increases the IE frequency (No IE Occurred)

3 SDP ASSESSMENT SOFTWARE (RYAN)

3.1 *Development of RYAN Software*

The SDP importance assessment can be performed using full-scale quantification programs such as AIMS-P. However, the importance evaluation using AIMS-P is possible by PSA experts who have knowledge of PSA model. Therefore, it is necessary to develop software that provides Non-PSA Expert with the ability to roughly estimate the risk increase due to incidents / accidents. To solve this problem, this study developed RYAN, which is a SDP evaluation program for regulatory verification.

RYAN provides a simple sensitivity analysis interface that can easily assess RISK changes to events/accidents resulting from plant shutdowns and equipment failures. In the case of detailed evaluation, PSA Expert can evaluate using AIMS-PSA. By using this, the regulatory body can establish regulatory standards for the change of the risk of the power plant. We called the software for a regulatory PSA model RYAN (Risk Analysis for ASP/SDP of NPP) which is for a significance assessment of incidents/accidents in domestic NPPs (Nuclear Power Plants).

The assessment of the importance of incidents/accidents is based on the evaluation of the increase in risk due to them, and is divided into the following: risk increase by an IE (Initiating Event) assessment and risk increase by a conditional assessment due to damaged SSC. Figure 4 shows the risk assessment concept for RYAN (S. Y. Choi et al, 2017) (S. H. Han, 2017). RYAN is a user-friendly interface tool developed under Windows environment using the PSA model provided by the AIMS-PSA (Advanced Information Management System for PSA) which is a software for integrating various types of PSAs including typical external and shutdown PSAs (S. H. Han, 2016).

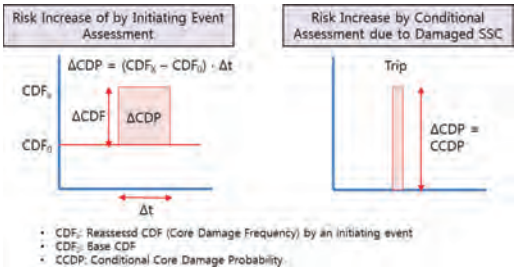


Figure 4. Risk assessment with RYAN.

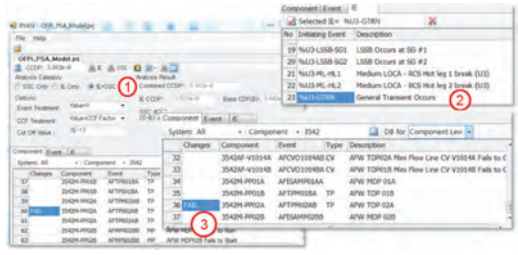


Figure 6. An example of SDP calculation with RYAN.

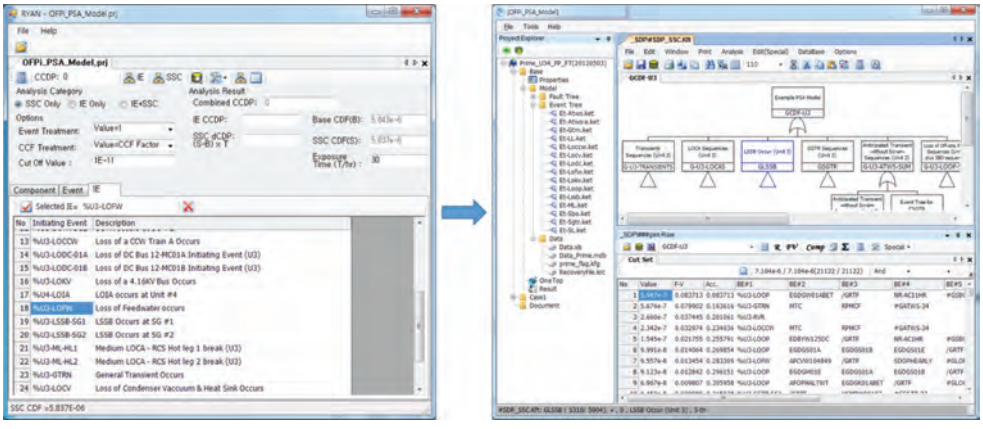


Figure 5. An example of RYAN analysis interface.

Figure 5 shows an example of RYAN analysis interface to quantify risk increase of an accident. To evaluate the increased risk due to an accident, the user should input damaged SSC information and IE information when the damaged SSC affects an IE. Then RYAN automatically changes the PSA model to quantify the increased risk.

3.2 *Example of SDP Calculation with RYAN*

In order to verify and validate the RYAN software, we investigated operational events occurred in domestic NPPs from databases such as OPIS (Operational Performance Information System for Nuclear Power Plant) and KRDB (Korean Integrated Reliability Database) (OPIS, 2017) (S.Y.Choi et al, 2005). Figure 6 shows an example of SDP analysis using RYAN.

Assumptions for the analysis are as follows.

– IE: General Transients
– Failed SSC component: Fail to Start of AFWS-PP02 A (T/D Pump)
– Exposure Time: 720 hours

As shown in the figure, the analysis is performed in the following order.

Table 2. Result of sample calculation.

Option	Component	IE CDP	SSC CDP	Total
Event Level	Comp = Fail Value = 1	2.52E-06	1.39E-06	3.91E-06
Comp Level	Comp = Fail Value = 11	4.05E-06	3.15E-06	7.20E-06

1. Select the analysis type as the initial event + SSC analysis.
2. Select the initial event as General Transient.
3. Change the BE of the failed SSC to Fail.
4. Set the OOS Time to 720 hr.

Table 2 shows the results of SDP analysis in this example and indicates that the SDP analysis using RYAN has been performed properly.

4 CONCLUSION

In this research, KAERI developed a regulatory purpose Level 1 PSA model and framework for use in risk-informed regulation. The purpose of this research is to estimate the risk significance of

initiating events and degraded conditions occurred in Korean NPPs and characterize the significance of an inspection finding consistent with regulatory response thresholds. To this end, we designed a regulatory risk model algorithm reflected regulatory purposes based on the real conditions and developed a RYAN (Risk Analysis for ASP/SDP of NPP) that enables the regulatory body to perform an overall safety assessment such as ASP/SDP of U.S.NRC. To verify the applicability of RYAN, we investigated operational events occurred in domestic NPPs from databases such as OPIS (Operational Performance Information System for Nuclear Power Plant) and KRDB (Korean Integrated Reliability Database), and performed a sensitivity analysis for the selected operational events. Based on the framework, it is expected that the regulatory staff can identify a nuclear power plant or SSC (Structure, System, and Component) for which safety performance has been decreased. For a further study, we are confirming the applicability of RYAN by performing a sensitivity analysis with more event data. In addition, we are going to analyze the significance of each case with the AIMS-PSA to compare the results with RYAN for ensuring the accuracy of the analysis results of RYAN.

ACKNOWLEDGEMENTS

This work was supported by Nuclear Research & Development Program of the National Research Foundation of Korea (NRF) grant funded by the Korean government, Ministry of Science, Ict & future Planning (MSIP).

REFERENCES

Choi S.Y. and S.H. Han, Analysis of Component Reliability of Korean Standard NPPs, KAERI/TR-2749/2004, 2005.

Choi S.Y. et al, Software Module Development for a Regulatory Risk Assessment, WiN-Global Annual Conference, 2017.

Han, S.H. MPAS-SDP Quick User Guide, KAERI-ISA-Memo-MPAS_SDP-01, 2017.

Han S.H. et al., AIMS-PSA: A Software for Integrated PSA, PSAM13, 2016.

http://opis.kins.re.kr/opis?act=OPISMAIN, September, 2017.

Kim J. et al., Annual Project Report (Development of Basic Regulatory Framework for APR1400 Level 1 PSA Model), 2017.

U.S. NRC, Risk Assessment of Operational Events Handbook, Vol. 1-Internal Events, 2013.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Risk dimensions of fish farming operations and conflicting objectives

S.M. Holen, I.B. Utne & X. Yang
Institute of Marine Technology, NTNU, Trondheim, Norway

ABSTRACT: Operations at sea-based fish farms can be challenging, and several risk dimensions are of concern during operations. Sea lice represent a challenge for the fish farmers who are required to perform delousing when the infestation levels rice above a set value. Delousing operations are frequently performed and require the use of heavy machinery operated from service vessels moored to the net-cages. Operators are exposed to hazards that may cause severe injuries and fatalities. Escape of salmon, which is a substantial environmental risk, has occurred in relation to delousing operations. Chemicals used during the operations may cause negative environmental consequences. Other safety related issues are the fish health and welfare. In this paper, a delousing operation on a fish farm is discussed with respect to different dimensions of risk, and potential conflicting objectives are discussed.

1 INTRODUCTION

The operators on fish farm localities have to navigate and make decisions in an environment where their own safety is lined up against other factors, such as fish welfare and prevention of escape of salmon. The workplace is exposed to forces from the environment, such as waves, current and wind, and maintaining focus on safety is crucial in all operations. Authorities with different regulatory responsibilities require risk assessment of prevention of fish escape, environmental impact and fish welfare (Holmen et al., 2017). Identification of hazards and risk assessments are measures implemented to avoid accidents. Holistic and systematic risk management is a prerequisite for safe operations, however, the fragmented regulation might work against this (Utne et al., 2017).

Projects related to the evaluation of risks in fish farms have identified critical operations, such as lice counting, well boat operations and operations involving cranes (Sandberg et al., 2012). Technology, the physical working environment, work-load, work pressure and safety management are found to be among the factors influencing escape events (Thorvaldsen et al., 2015). External pressures on operations, such as time, costs and weather conditions also puts constraints on operations.

Lice infestations has become a major sustainability challenge in Norwegian fish farming, and has also become the main delimiting factor for future growth in the industry (Svåsand et al., 2017, Norwegian Ministry of Trade Industry and Fisheries, 2017). The fish farming industry in Norway uses up to NOK 4,5 billion in anti-lice measures (DN, 2017). Treatments to remove lice are decreed in regulations (Norwegian Ministry of Trade Industry and Fisheries, 2012), and has become an operation frequently performed in fish farms. Delousing is an operation where several factors identified as critical or risk-influencing are present, see Table 1.

In this paper, the first three risk dimensions are presented and compared with the purpose of identifying examples of potential conflicting objectives in the fish farming operation delousing. Conflicting objectives is an accident perspective, and highlighting consequences of the different pressures the human operators are exposed to in aquaculture, risk-reducing measures can be developed.

2 RISK DIMENSIONS IN A CONFLICTING OBJECTIVES' ACCIDENT PERCPECTIVE

The concept of conflicting objectives is described by Rasmussen's migration model (Rasmussen, 1997a). It explains how accidents may happen when decisions in an organization are made based on different objectives and constraints. One example is the decisions made by management to minimize costs, while operators may focus on making the operations as efficient as possible. These sometimes competing, or conflicting, objectives may eventually lead to a migration towards the boundary of a functionally acceptable performance. As the decisions are made local at separate levels, the side effects of the decisions may eventually set the stage for an accident (Rasmussen, 1997b). The operators can be seen to be at the sharp-end, close to the hazard sources, while management can be seen to be at the blunt end, removed from the hazards (Rosness, 2001, Rosness et al., 2010a).

Table 1. Risk dimensions present in the fish farming operation delousing. Adapted from (Yang et al., 2017).

Risk dimension	General description	Relation to delousing operations
Risk to personnel	The Norwegian fish farming industry has one of the highest fatality and accidents rates when compared to similar industries (Aasjord, 2010). Accident statistics show that the fish employ ees are among the most exposed workers with regards to injuries and fatalities (Holen et al., 2017a).	Frequent use of safety critical equipment during delousing operations.
Risk to environment	The escape of salmon represents a hazard for the stock of wild salmon living in the rivers and fjords of Norway (Svåsand et al., 2017). The use of chemicals in delousing operations and on the net-cage to avoid fouling may affect the environment around the fish farm. Waste that accumulate under the fish farms due to fodder spill and organic matter and may have benthic impacts and on species living around the fish farm (Holmer, 2010).	Risk of net-tear is present during delousing operations. Medical treatment chemicals are released after operation.
Risk to fish welfare	Fish welfare in fish farms are under pressure due to sea lice and diseases (Hjeltnes et al., 2017).	Delousing operations require handling of the fish and may cause harm. The chemicals used in delousing may cause discomfort and wounds.
Food safety	Food safety is a general concern due to the accumulation of toxins in the fish meat.	Chemicals used for treatment of fish are not seen as critical for food safety (Norwegian Veterinary Institute, 2016).
Risk to material assets	Risk to material assets (e.g., net-cages, service vessels, workboats etc.) in fish farm operations may have severe economic consequences, mainly to the fish farm company. This risk dimension has not gotten much attention in the literature (Xue, Yang et al. 2017).	Structural damages of net during delousing may lead to escape of salmon which is a risk dimension already included.

Safety is an emergent property of a system and risk should be considered in a systems perspective where all factors that can influence safety, should be analyzed. Control can be made by increasing the safety margin, increase awareness of the boundary, or make the boundaries explicit. Making visible the limits on acceptable risk by establishing criteria for critical decisions or other ways of establishing clear lines as to when the safety margin is small should encounter challenges with conflicting objectives. Managers should also communicate openly about the existence of conflict of interest (Rosness et al., 2010b).

Fish farming is an industry dealing with production of livestock, thus requiring knowledge about biology, welfare, and diseases. In addition, operations are increasingly resource demanding and large production equipment requires special expertise for safe handling. The fish farms are mainly placed in the fjords where the operations may impact the fauna and wild animals living around the fish farm. These are all risk dimensions of concern for the operators at the sharp-end, and in some situations trade-offs between the risk dimensions must be made. In this paper these are seen as conflicting objectives. An example of a situation where operators are faced with having to choose between prioritizing risk objectives is provided by Størkersen (2012). The operators have to choose between fixing a net cage damage immediately after discovery, or use valuable time to provide the appropriate safety equipment to do the repair according to safety procedures. In the case presented, the operators do not hesitate to improvise and make the repair without the required safety equipment. Thus, the risk of escape is reduced, while the operators face a greater personal risk by down prioritizing their own safety (Størkersen, 2012).

The management at the blunt end is also making choices that affect the risk in operation, by allocating resources, like personnel, equipment and timeslots to operations. Management decisions influenced one of the biggest single escape event in Norway, which happened in relation to a delousing operation in 2011 where 176 000 salmon escaped (Soknes, 2012). The delousing operation had been ongoing

for two continuous days in order to finish the operation as quickly and efficiently as possible. The company later claimed that the responsible operator was disloyal to the company when breaching procedures to get the job done. However, a court case ruled that the employee had loyally tried to fulfill the management's expectations and that there had been a great time pressure on the employees, and no willingness from the company to compensate economically for extra personnel (Soknes, 2012).

Time pressure is a risk-influencing factor mentioned by personnel at fish farms in relation to both escape events, fish welfare and personnel safety (Thorvaldsen et al., 2015, Hjeltnes et al., 2017, Fenstad et al., 2009). Time pressure is not only created by allocation of resources by management, but also unforeseen weather changes puts this constraint on operations.The regulation of the fish farming industry is characterized by being fragmented and the authorities have developed separate regulations to ensure the different values being protected (Holmen et al., 2017). Fish farmers state that the focus in planning for safety in operations will be towards the area were they experience pressure from the authorities (Skjærvik, 2017). In line with Rasmussen's framework of distanced decision-making, some unforeseen consequences might be the result. For example, the strict regulations on delousing according to infestation levels may lead to both unsafe situations concerning escape and reduced welfare for the fish.

3 THE DELOUSING OPERATION

3.1 *Anti-lice measures*

The sea lice, or salmon lice, is a parasite, which only have salmonids as hosts. The last five stages of the life cycle of the sea lice are parasitic to the salmon, when it feeds of the mucus, skin and blood. The sea lice may cause fish welfare problems both to farmed and wild salmonids, and may ultimately cause fish death. The sea lice has become a major issue in the fish farming industry where large outbreaks of the parasite is made possible by the high density of salmon in the fish farms along the coast. The sea lice is sensitive to temperature, and infestation levels change according to the season; the lowest levels are registered in the spring and the levels increase during summer and fall (Svåsand et al., 2017). This have led to frequent delousing in periods of the year.

As the sea lice mainly lives in the higher levels of the sea some preventive measures to sea lice have been developed, e.g., a skirt placed around the net cages with a depth up to 3 meters preventing the sea lice to enter in the area where the salmon are (Lien et al., 2014). The skirts around the net are the most used preventive measure (Svåsand et al., 2017). Also a "snorkel"-solution, where the fish are held in an semi-enclosed net cage, only with access to water air through a "snorkel" with a diameter around 6 meters (Stien et al., 2016). In 2017 over 27 million wrasse was captured, mainly used for delousing purposes in fish farming (Directorate of Fisheries, 2017).

The main mode of combating the sea lice have been medicinal products. These can either be introduced to the fodder, or the salmon are exposed to the medicament in bath-treatments. In bath treatments, the salmon are exposed to medicinal products added to the seawater after the salmon is gathered in an enclosed area, in either a well vessel or using a tarpaulin around the net cage. The bath treatments require major resources and is one of the most demanding operations that is carried out in fish farming.

In addition to the bath-treatments, some new technologies have been developed to remove sea lice from farmed salmon. These methods have been developed mainly due to resistance in the salmon lice of the medicaments used. The new treatments use mechanical aids, such as water jets, higher temperature and brushes. These new methods are seen as the main cause in the large drop in prescribed anti-lice treatment medicaments from 2015 to 2016. There is a concern that the new methods might be a risk to fish welfare, and that they have not been sufficiently tested for welfare before they have been put to use (Hjeltnes et al., 2017). In addition, signs of possible resistance to these new anti-lice treatments have been discovered.

3.2 *Steps of a bath treatment operation*

Figure 1 show the steps of a fish farm operation using tarpaulin. This approach is representative for all methods of delousing, only the step "Perform delousing" differs according to the method and technologies used.

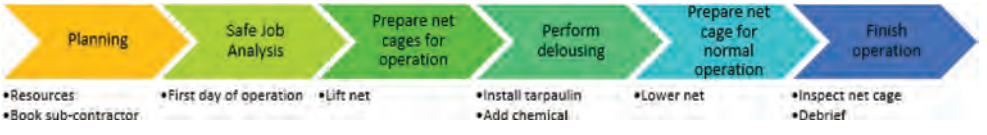


Figure 1. Steps of a delousing operation using tarpaulin.

- Planning

Delousing must be carried out when the critical level of lice is reached. The operation is planned by the operations manager on the fish farm, sometimes in cooperation with higher level onshore area managers. The operations are in most cases performed by or in co-operation with service providers who have both required equipment and expertise.

- Safe Job Analysis (SJA)

Most fish farmers conduct a preparation meeting the day the operation starts. An integral part of this meeting is to perform a SJA, where hazards in the operation are identified and responsibilities for tasks during the operation are assigned.

- Prepare net cage for delousing (Lift net)

It is necessary to make the volume of the net cage smaller so that the fish is easily accessible in the upper layers of the sea. Lifting is demanding and time-consuming, and requires the use of crane and winches from work vessels. If a well vessel or a type of barge is used in the treatment, a "crowding" of the fish is also necessary. This is done by using an extra net to push the fish together in an even more confined area.

- Perform delousing

Bath-treatment are either performed with tarpaulin in the net cage or in a well vessel. New types of mechanical treatments are performed on specialized barges.

- Prepare net cage for normal operation (Lower net)

After the treatment, the fish is put back in the net or the tarpaulin is removed, depending on the type of treatment. Then the net needs to be lowered to its normal position. This is done in a reverse manner to the lifting of the net. Careful lowering of the net and ropes are necessary to avoid any damage.

- Finish operation

After the operation is finished, an underwater inspection should be made by either divers or a ROV. Debrief-meetings will ensure that any adverse events during the operation are discussed and subsequent changes implemented in safety management systems.

4 RISK DIMENSIONS OF DELOUSING OPERATIONS

In this section, the three first dimensions of risk in Table 1 (Yang et al. 2017) are presented and discussed for the delousing operations.

4.1 *Risk to personnel*

Delousing operations are demanding operations where the operators on fish farms are exposed to several hazards. Most of the delousing techniques require use of cranes when preparing for the operation. In accident statistics from the fish farming industry, the use of cranes are found to contribute to several of the blow by object and entanglement injuries (Holen et al., 2017a). Work operations are also an increasing contributor to fatalities in the fish farming industry (Holen et al., 2017b). As service vessels are an important part of the operation, also man over board accidents is in important risk to consider. In addition, the chemicals used in delousing operations may present a hazard to the operators. In some delousing operations, extra oxygen is used, and explosions may happen.

4.2 *Risk to the environment*

In general, two types of hazards to the environment should be assessed in relation to delousing operations; (i) the effects from escaped farmed salmon, and (ii) the release of treatment chemicals, which may have an effect on organisms around the fish farms.

4.2.1 *Risk of escape*

The main causes to escape from fish farms are due to structural failures including net tearing. Net tearing can happen during operations and from abrasion from related components (Jensen, Dempster et al. 2010). Abrasion from the sinker tube chain is the most common cause for net tearing, while handling of net weights, including the sinker tube is the second largest cause (Føre and Thorvaldsen, 2017). Handling of net weights must be done in all delousing operations, as part of the preparation before the operation, and after the operation has been completed. Organizational factors influencing escape events are found in Thorvaldsen, Holmen et al. (2015).

The consequences of escaped salmon are related to introgression of genes and the spreading of diseases, which both may influence the wild salmon. Introgression of farmed salmon genes is unwanted because of the genetic differences in farmed salmon and wild salmon (Taranger et al., 2015). The long term consequences of introgression may lead to "changes in life-history traits, reduced population productivity and decreased resilience to future changes" (Glover et al., 2017).

4.2.2 *Risk of treatment chemicals on surrounding environment*

The medical chemicals used for bath treatment of sea lice may affect other animals, especially crustaceous animals as the sea lice belongs to this type of animals. The chemicals used for bath treatments are Azametifos, Deltametrin, Cypermetrin and Hydrogrenperoxid; the three first chemicals are mainly used in tarpaulin treatments, while the last

is used in well vessels. When the bath treatment is made with tarpaulin, the chemicals are directly released into the sea at the fish farm; when well boats are used the chemicals can be transported away (Svåsand et al., 2017). The different chemicals have different levels of toxicity, where Deltametrin have been shown to be very toxic for some non-target organisms, such as plankton, and may also be bound up in seaweeds. Hydrogen peroxide have the least effect on organisms in the surroundings of the fish farm (Svåsand et al., 2017). In a five-year study of effects of sea lice medicine to the receiving environment in Scottish sea lochs, no long-term effects could be found (Scottish Association for Marine Science, 2005). Chemical release in the case of vessel capsizing may also be a risk.

4.3 *Risk to fish welfare*

Fish welfare is affected by the salmon louse itself and the anti-lice treatments carried out to remove the lice. Normally the damage to the farmed salmon is not high because treatment is required before a critical number of lice is reached (Svåsand et al., 2017, Norwegian Ministry of Trade Industry and Fisheries, 2012). However, substantial injuries in some areas where the salmon lice infection pressures have not been possible to control have been reported (Hjeltnes et al., 2017). The larger wounds caused by sea louse may lead to dehydration, electrolyte balance and increased influence on physiological functions with the fish (Svåsand et al., 2017).

Anti-lice treatment represents a significant negative welfare challenge to the fish (Hjeltnes et al., 2017). Especially handling and crowding of fish, which is done in relation to the treatment, will have an impact on welfare of the fish. The stress and fear-levels increase in the fish during these operations and if the fish is weak, heart failure may occur. Open wounds, scale and mucus-loss and stress are factors caused by handling which might also increase the risk of other infections in the fish (Svåsand et al., 2017). The chemicals used in treatment may be overdosed and give toxic effects. Observed fish behavior during delousing operations may indicate that the fish experience the treatment chemicals as uncomfortable (Oppedal et al., 2011).

Bath treatments have been the primary method of delousing, but new methods and technologies, which does not use chemicals, are increasingly in use, mainly due to resistance of chemicals in the salmon louse. Mechanical delousing using heated water, water jets or a combination of water jets and brushes are reported to give welfare issues related to reduced appetite, eye injuries, reduced mucus production and poor skin health, amongst others. These new methods of anti-lice treatment are of great concern to fish welfare as they are not sufficiently tested for effectiveness and welfare (Hjeltnes et al., 2017). Heated water treatment has caused mass-fatalities of salmon (Heraldscotland, 2016).

4.4 *Conflicting objectives of the risk dimensions*

Some examples of how each risk dimension may influence the others during the delousing operation are presented below. Especially, risk to personnel safety, risk of escape, and risk to fish health may come in conflict. All these dimensions are also under the constraints introduced by management decisions like allocation of resources, such as personnel, equipment and timeslots to operations.

4.4.1 *Prioritizing personnel safety*

Personnel safety has been given increasing focus in the fish farming industry. Major hazards for personnel are especially present during operations using heavy machinery. For delousing operations, this type of machinery is used in preparation of the delousing, and after the operation when net is lifted and lowered. Handling of the net also involves hazards with regards to tearing of net and following escape. In stressful situations, due to limited attention span, there could be a need to focus on one of the risk factors. Situations where focusing on personnel safety may cause higher risk with regards to escape may also occur after operations when inspections of the net cages should be done to ensure that the nets have been correctly lowered. Inspections by divers or cameras must be done so that potential holes caused during operation are discovered. In cases where there may be risk of injuries because of, e.g., weather conditions, personnel safety must be prioritized over prevention of escape.

Stopping operation too soon or too late in cases of risk to personnel may cause the delousing treatment not to work adequately. The operation must then be repeated later, which represents an extra strain to fish welfare, which must undergo handling again in a short time. If operation is not completed the net may not be lowered in between operations which also means that the fish must be kept “crowded”.

4.4.2 *Prioritizing fish welfare*

Fish welfare has traditionally been given high priority. Fish welfare is important to management as it affects earnings. Cases when fish welfare may influence personnel safety or prevention of escape during operations, may occur if delousing with tarpaulin must be abruptly stopped due to, e.g., too low oxygen levels in the net cage. Stressful

situations and a focus on fish welfare may lead to hazardous situations by personnel. The choice of delousing methods may also have an influence on the fish welfare. Bath-treatments with tarpaulin include some more hazardous tasks using crane compared to bath-treatments in well vessels. Whereas well vessels may include more welfare issues due to "crowding" and pumping of the fish in and out of the vessel.

4.4.3 *Prioritizing prevention of escape*

Prevention of escape is a major focus of the fish farming industry. This focus may also have been at the sacrifice of personal safety in procedures and risk assessments. An inadequate focus on hazards that may cause personnel injuries when planning operations and in safe job analysis performed before delousing may contribute to accidents.

If a hole in the net is discovered, fish may be kept crowded longer than normally to keep the fish away from hole in the net. This will be at the expense of fish welfare.

4.4.4 *Prioritizing limited consequences for environment*

When using well vessels for delousing operations, chemicals used during operations may be transported out of the fjords into designated "drop zones". The choice of using well-vessels may have an influence on fish welfare in operations. Some of the new delousing methods do not use chemicals and, in this regard do not represent a challenge to the environment. Emissions to the environment of the chemicals used in delousing operations are an integrated part of the operation, especially when using the tarpaulin. The consequences of the release of chemicals into the fjords is a controversial issue between the stakeholders.

5 DISCUSSION

Several risk issues are present during fish farm operations, and delousing is no exception. Accidents, such as escape, serious personal injuries and major fish deaths, have happened in relation to the activities in delousing operations. In the accident perspective of conflicting objectives, one of the measures towards avoiding accidents is to make visible the limits of acceptable performance. It is important to assess how prioritizing one risk issue may affect other risk aspects and dimensions. During delousing operations both personnel safety, fish welfare and fish escape are concerns, which require attention. It is not possible to eliminate the conflicting objective as they, in today's methods available for delousing, are inherent in the operation. However, means to avoid accidents due to conflicting objectives are to highlight the conflicts themselves and the possible consequences of giving priority to one aspect in operations. Visualizing the different risk dimensions, which may give rice to hazardous situation, gives an opportunity for operators and management to gain awareness of possible hazards in the operation.

Possible risk mitigating actions could be to assign some operators the main responsibility to follow whether one risk issue is given an unbalanced focus. The different steps of the operation may also be more hazardous with regards to one type of risk. For example, the beginning of preparation of the net cage is hazardous related to tearing of the net, while the last part of preparation may be more hazardous to personnel injuries because of excess chains suspended from the crane. In addition, correct lowering of the net after operation is a critical part of the operation concerning escape events.

Risk avoidance of some of the measure might also have mutual positive effects. One example of this is to not starting delousing treatment in harsh weather, as this might present hazards to both personnel and fish welfare (Størkersen, 2012, Fenstad et al., 2009).

In almost all situations during operations where one risk issue might be prioritized over a different one, management decisions, such as time pressure, costs and weather may influence the decisions made during an operation. Stress due to time limits and limited resources will affect how choices are made, and violation of procedures might be done if that is what seems most rational in the moment. When evaluating how risk mitigating measures might work, one should be aware off the mechanisms of the socio-technical system where different actors will make decisions according to their respective constraints and options, and that some interpretation of rules will be made at lower levels of the organization (Rasmussen, 1997b). Within the aquaculture company, the operators are in the sharp-end in close proximity to the hazard, and they make decisions within different frames of what the land based organization with higher level of authority and distance to the hazard do. Without the possibility of always seeing the whole picture decisions on both ends are made on "local rationality". Often, it is explicitly said that safety should be prioritized, but tacitly opposite messages are sent through planning, follow-up and resource allocation. Measures should be implemented in and continuously monitored by the management systems to ensure that safety is not compromised.

In this paper, the immediate risk issues that arise during an operation due to conflicting objectives has been in focus. In a broader perspective, other risk issues would also be relevant to consider with regards to conflicting objectives such as resistance

of sea lice to the different treatments and the influence of regulations on the different risk issues. The decisions made on higher level may have more impact to the risk picture, than the decisions made by operators during operations. The regulations that specifies the limits for the acceptable lice level may challenge fish welfare as it leads to frequent delousing (Hjeltnes et al., 2017). This is seen as a challenge to the welfare and in some cases the levels of lice might be more acceptable to welfare than performing repeated treatments which cause strain and stress to the fish. Repeated delousing operations will also increase the possibility of escape due to handling of the net.

6 CONCLUSION

Sea lice is a major challenge to the fish farming industry and delousing is decreed by the authorities. The delousing operation involves risk dimensions with regards to personnel safety, the environment and fish welfare, all issues including severe consequences. Conflicting objectives may arise during the operation. Prioritizing one risk dimension at the expense of others may lead to situations, such as: (i) focusing on personnel safety may hinder the discovery or repairing holes in the net, or (ii) operator stress to finish operation due to fish welfare, may cause hazardous situations for personnel. Higher-level management decisions also influence the risk during operations through, e.g., timely allocation of resources. Unforeseen accidents may happen if conflicting objectives are not visible to management and operators, and they should be openly discussed to ensure safety in operations.

ACKNOWLEDGEMENT

This article is funded as part of the research project "Towards Sustainable Fish Farming at Exposed Marine Sites—SustainFarmEx" supported by the Norwegian Research Council (project no. 210794/O70). References. We are grateful to the Directorate of Fisheries Norway, who provided information about escape events.

REFERENCES

Aasjord, H. (2010) Den norske fiskeflåten—HMS-status pr. 2010. Sintef Fisheries and Aquaculture.

Directorate of Fisheries (2017) Wrasse in Trade fishing (In Norwegian).

DN (2017) Kan ha nådd kostnadstoppen.

Fenstad, J., Osmundsen, T. & Størkersen, K.V. (2009) Danger on the net-cage? Needs for change in safety work at Norwegian fish farms (in Norwegian). NTNU Samfunnsforskning AS.

Føre, H.M. & Thorvaldsen, T. (2017) Causes for escape of farmed salmon and trout in the period (2010–2016) (In Norwegian). Trondheim, Sintef Ocean.

Glover, K.A., Solberg, M.F., McGinnity, P., Hindar, K., Verspoor, E., Coulson, M.W., Hansen, M.M., Araki, H., Skaala, Ø. & Svåsand, T. (2017) Half a century of genetic interaction between farmed and wild Atlantic salmon: Status of knowledge and unanswered questions. *Fish and Fisheries,* 18, 890–927.

Heraldscotland (2016) Oops: fish farm firm kills 175,000 of its salmon by accident.

Hjeltnes, B., Bornø, G., Jansen, M.D., Haukaas, A. & Walde, C.E. (2017) The Health Situation in Norwegian Aquaculture 2016. Norwegian Veterinary Institute.

Holen, S.M., Utne, I.B., Holmen, I.M. & Aasjord, H. (2017a) Occupational safety in aquaculture—Part 1: Injuries in Norway. *Marine Policy*.

Holen, S.M., Utne, I.B., Holmen, I.M. & Aasjord, H. (2017b) Occupational safety in aquaculture—Part 2: Fatalities in Norway 1982–2015. *Marine Policy*.

Holmen, I., Utne, I., Haugen, S. & Ratvik, I. (2017) The status of risk assessments in Norwegian fish farming. *ESREL*.

Holmer, M. (2010) Environmental issues of fish farming in offshore waters: perspectives, concerns and research needs. *Aquaculture Environment Interactions,* 1, 57–70.

Lien, A.M., Volent, Z., Jensen, Ø., Lader, P. & Sunde, L.M. (2014) Shielding skirt for prevention of salmon lice (Lepeophtheirus salmonis) infestation on Atlantic salmon (Salmo salar L.) in cages—A scaled model experimental study on net and skirt deformation, total mooring load, and currents. *Aquacultural Engineering,* 58, 1–10.

Norwegian Ministry of Trade Industry and Fisheries (2012) Regulation on the combating of salmon louse in aquaculture farms (In Norwegian). *FOR-2017-03-06–275*.

Norwegian Ministry of Trade Industry and Fisheries (2017) Rules for new fish farming system ready (In Norwegian). Norwegian Ministry of Trade, Industry and Fisheries.

Norwegian Veterinary Institute (2016) Use of chemical agents against salmon lice in Norwegian Aquaculture on behalf of Norwegian Seafood Council. The Norwegian Veterinary Institute.

Oppedal, F., Dempster, T. & Stien, L.H. (2011) Environmental drivers of Atlantic salmon behaviour in sea-cages: A review. *Aquaculture,* 311, 1–18.

Rasmussen, J. (1997a) Risk management in a dynamic society: a modelling problem. *Safety Science,* 27, 183–213.

Rasmussen, J. (1997b) Risk Management in a Dynamic Society: A Modelling Problem.

Rosness, R. (2001) OM jeg hamrer eller hamres, like fullt så skal der jamres" Målkonflikter og sikkerhet.

Rosness, R., Grøtan, T.O., Guttormsen, G., Herrera, I.A., Steiro, T., Størseth, F., Tinmannsvik, R.K. & Wærø, I. (2010a) Organisational Accidents and Resilient Organisations: Six Perspectives. Revision 2.

Rosness, R., Grøtan, T.O., Guttormsen, G., Herrera, I.A., Steiro, T., Størseth, F., Tinmansvik, R.K. &

Wærø, I. (2010b) Organisational Accidents and Resilient Organisations: Six Perspectives. Revision 2. SINTEF Technology and Society.

Scottish Association for Marine Science (2005) Ecological effects of sea lice medicines in Scottish sea lochs.

Skjærvik, A.J. (2017) Safety management on small ships. *Presentation from the Norwegian Maritime Authority Seminar on safety management, TEKMAR.* Trondheim.

Soknes, B. (2012) Bot på 2 millioner for lakserømming. *Miljøkrim.* http://www.okokrim.no/www/okokrim/resource.nsf/files/wwww933csy-miljokrim_32012/$FILE/miljokrim_32012.pdf.

Stien, L.H., Dempster, T., Bui, S., Glaropoulos, A., Fosseidengen, J.E., Wright, D.W. & Oppedal, F. (2016) 'Snorkel' sea lice barrier technology reduces sea lice loads on harvest-sized Atlantic salmon with minimal welfare impacts. *Aquaculture,* 458, 29–37.

Størkersen, K.V. (2012) Fish first: Sharp end decision-making at Norwegian fish farms. *Safety Science,* 50, 2028–2034.

Svåsand, T., Grefsrud, E.S., Karlsen, Ø., Kvamme, B.O., Glover, K., Husa, V., Kristiansen, T.S. & (red) (2017) Risk report Norwegian Fish Farming 2017 (in Norwegian). Institute of Marine Research.

Taranger, G.L., Karlsen, Ø., Bannister, R.J., Glover, K.A., Husa, V., Karlsbakk, E., Kvamme, B.O., Boxaspen, K.K., Bjørn, P.A., Finstad, B., Madhun, A.S., Morton, H.C. & Svåsand, T. (2015) Risk assessment of the environmental impact of Norwegian Atlantic salmon farming. *ICES Journal of Marine Science,* 72, 997–1021.

Thorvaldsen, T., Holmen, I.M. & Moe, H.K. (2015) The escape of fish from Norwegian fish farms: Causes, risks and the influence of organisational aspects. *Marine Policy,* 55, 33–38.

Utne, I.B., Schjølberg, I., Holmen, I.M. & Bar, E.M.S. (2017) Risk Management in Aquaculture: Integrating Sustainability Perspectives. V07BT06 A054.

Yang, X., Utne, I.B. & Holmen, I.M. (2017) MIMACHE: a Methodology for the Identification of Major ACcident hazards and Hazardous Events in Norwegian aquaculture. *Submittet Safety Science*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

The future of driver training and driver instructor education in Norway with increasing ADAS technology in cars

G.B. Sætren, J.P. Wigum, R. Robertsen, P. Bogfjellmo & E. Suzen
Road Traffic Section, Business School, Nord University, Stjørdal, Norway

ABSTRACT: On average, more than two people are killed or severely injured every day in Norway in road traffic. Hence, elements that benefit a decrease in this number will be welcomed, such as "Advanced Driver-Assist System" (ADAS) technology. However, increasing technology in cars might require new driving skills compared to those taught today and the transition to more and new technology could potentially increase the accident rate. In the safety industry, it is well known that training for new and more automated technology is important. This raises a question: How does the transition to new, more complex and more automated technology affect driver training and the education of driver instructors? At the present time, there are no clear answers to this question. However, it seems that there is a need for a discussion and potentially a redefinition on which driver skills should be required, and how to implement these skills. This is what we attempt to discuss in this paper.

1 INTRODUCTION

In 2016, there were 135 road deaths in Norway. The number for 2015 was 117, and for 2014 it was 147. However, if you include the number of accidents resulting in severe injuries, the number was 791 in 2016, 810 in 2015, and 821 in 2014 (SSB 2018a). This means that, on average, more than two people are killed or severely injured every day due to traffic accidents in Norway. Compared to any other high-risk sector, the number is high, but the trend over the past decades is that the number is decreasing. The Norwegian government bases the National Transport Plan (NTP) on a vision of zero. This vision means zero dead and zero severely injured in road traffic. The objective for this period of NTP (2014–2023) is to halve the number of road deaths and severe injuries, and that, in 2020, there should be no more than 775 killed and severely injured in road traffic in Norway, that is about two people per day on average. Strategies to achieve this objective in Norway are, for instance, to design safer roads, to encourage safer behaviour from road users, and to encourage the development of technology to produce safer vehicles (NTP 2014–2023). Norway is not alone in such objectives as this is also in accordance with the EU objective, which is to halve the number of people killed in road traffic during the period 2010–2020 (European Commission 2015a). In order to achieve this, the EU has developed seven strategies: (1) improve education and training of road users, (2) Increase enforcement of road rules, (3) safer road infrastructure, (4) safer vehicles, (5) promote the use of modern technology to increase road safety, (6) improve emergency and post-injuries service, and (7) protect vulnerable road users (European Commission 2010). Technological innovation is one of the seven strategies, in addition to improving education and training of road users. However, what we know from other industries regarding humans interrelating with increased automation (e.g., Lee 2006; Sætren and Laumann 2015), it is not a certainty that the numbers of killed and severely injured will continue to decrease with an increase in technological solutions. Reasons such as lack of standardisation in technological solutions, mode confusion, lack of situational awareness, overreliance, complacency and so forth, could all be reasons why the interrelation between humans and technology have a possibility of not going according to plan (Young and Stanton 2007). One of the reasons is a lack of focus on training for automation (Sætren and Laumann 2015). Regarding the technological development in cars, new technology is implemented at a fast tempo, but little attention is given to training in using the technology to new and existing drivers. Research shows for instance that only 24% of buyers were given instructions from the car dealer when cars with an "Advanced Driver-Assist System" (ADAS) were bought in The Netherlands (Harms and Dekker, 2017). Even less attention seems to be placed on teaching driver instructors how to teach driving skills with this vast variety of technology. In addition, we have found no literature on this topic from a pedagogy aspect. For this

reason, we would like to look at training for automation when it comes to driving cars.

How will ADAS technology in cars potentially affect driver training and driver instructor education, and which new skills might be needed for a driver?

In order to answer this, the driver training program and the driver instructor education in Norway will be presented first, before we present issues regarding automation and training. After this we discuss ADAS technology in cars and how it would affect driver training and driver instructor education. Next, we look at which new skills a driver might need. Then, we present our conclusion.

2 DRIVER LEARNING PROGRAM AND DRIVER INSTRUCTOR EDUCATION IN NORWAY

The Norwegian driver education model is very comprehensive and systematic (Rismark and Sølvberg 2007) and it normally takes about two years to become a driver with the program which contains detailed curricula for content, progression, and teaching methods (NPRA 2013). This two-year education is a module based training program consisting of four modules that include both individual and group tutorials that are both theoretical and practical. In addition, accompanied driving with someone who has had their driving license for a minimum of five years is highly recommended and thus it is common, from the age of sixteen to drive with a parent as a passenger.

Driver instructor education in Norway is also an extensive education as it is a two-year university education with an emphasis on traffic pedagogy, road traffic law, and traffic psychology in addition to physics and technology (Nord universitet 2017). This two-year education includes both theory and practice and emphasises operational, tactical and strategic driving skills (Michon 1985), and the GDE framework (Peräaho, et al. 2003). However, in the future we might see a reduced need for an extensive focus on these elements, which until now have been viewed as basic. As future in-car technology might replace some of the information retrieval, assessments and decisions previously made by the driver, we might see a shift in which are the knowledge and skills that are important for driving instructors to develop.

3 AUTOMATION AND TRAINING

There are a number of different systems, ranging from basics such as automatic windscreen wipers to more advanced technology such as lane departure tracking, automatic braking systems and even more enhanced levels of automated driving functionality. Such systems are for instance autopilot (Tesla), distronic plus steering assist (Mercedes), and intellisafe (Volvo).

Increased automation in cars will probably lead to an eventual decrease in the numbers of accidents (Elvik and Høye 2015; Wilmink et al. 2008). Some reports indicate that traffic fatalities could be reduced by as much as 90% (Bertanocelli and Wee 2015). Further, levels of automation in cars will most probably increase as a result of the increased digitalisation of the transport sector, and brands such as Volvo, BMW, and Tesla, all popular brands in Norway, expect to have self-driving cars on the roads within the next five years (TechEmergence 2017). However, it is expected that the leap from where we are today to all cars being self-driven, is remote, and that semi-automation with in-built ADAS technology seems to be a reality for some time to come, considering that age of the motor vehicle population in Norway in 2016 was, on average, 10.6 years (SSB 2018b). The number for Europe is 10.7 years (ACEA 2017).

There are several different taxonomies trying to capture the essence of the development of advanced technology in cars, and the most common seems to be the SAE's levels of automation (SAE 2014). This approach is based on six levels of automation ranging from "No automation" (level 0) to full automation (level 5). In levels 0–3 the human driver has the responsibility for the driving, and in levels 4–5 the car takes on this responsibility. Examples of technology at each level, according to Banks et al. (2017) are for instance level 1: Adaptive Cruise Control (ACC), level 2: Tesla Autopilot, level 3: Audi A7 prototype, level 4: Toyota Highway Teammate, and level 5: Google self driving car. Today, most ADAS technology equipped cars are at level 1. Furthermore, seen from a drivers perspective (Banks and Stanton 2017), there are different roles for the driver within automated systems. As an example, a Driver Driving (DD) is defined as an operator responsible for completing basic operational, tactical, and strategic tasks (Michon 1985). However, the Driver Not Driving (DND) would expect an automated system to have full control of these tasks. That being said, the transition is not straightforward, and during the middle phases of automation, Driver Monitoring (DM) should be assumed. A challenge is that in level 2 the driver operates the vehicle, which assumes a transition between DD and DM and in level 3 the driver, to a larger degree, supervises the vehicle but needs to intervene if needed assuming a transition between DM and DND (Banks and Stanton 2017). The cars with the most advanced

driver assist systems will be additional to the many cars that have less advanced technological equipment on the roads. However, this middle phase is, according to human factors and safety research, a phase where the human interference is relied upon, but the human is not very reliable (Wickens et al. 2016; Son and Park 2017). Human interrelationship with semi-automated technology is known to potentially result in serious unwanted incidents in a wide range of sectors such as petroleum (Sætren and Laumann 2015), aviation (Billings 1997; Parasuraman and Byrne 2003), and road transport (NTSB 2017). Research has found there are several causes for this, for instance the issue of trust, overreliance, or complacency (Sætren and Laumann 2015; NTSB 2017), situational awareness (Kaber and Endsley 2007), mode confusion (NTSB 2014), or lack of optimal training (Salas et al. 2006; Sætren and Laumann, 2015). Additionally, news items concerning ADAS technology in cars seems to share a common misperception that when more automation is introduced, human error will disappear (e.g. NRK 2017). This gives rise to the idea that training is not necessarily needed. Human factors research advises against not training for the use of new complex technology (Lee 2006; Salas et al. 2006; Sætren and Laumann 2015), as there will always be a human in the technology loop, for instance in use, maintenance or design.

It might even be an issue that increased automation might increase the level of competence required for the operator, as an operator must know both how to handle the system more or less manually, for instance if the sensors in a car turn off due to bad weather, and additionally know how to handle and supervise the advanced technology.

So, as driving skills decrease, the need for potentially taking over the car will occur in more difficult scenarios such as in bad weather conditions like slippery roads, heavy snow, and so forth, because such conditions could be difficult for ADAS technology to handle. One example is Adaptive Cruise Control (ACC). A driver who uses ACC, that works most of the time, does not get much training in driving without it. Then, when it is time for the driver to take over control, for instance because the weather conditions are too harsh for the system to operate, the driver might lack optimal skills to handle the driving. Research has indicated that ACC technology leads to a reduction in mental workload and thus problems with regaining control of the vehicle in failure scenarios (Stanton and Young 1998). ACC is one of the technologies that might be turned off in, for instance, heavy rain without advance warning, implying in that the driver must be skilled in handling bad weather conditions while driving, and be able to take control of the car straight away.

During a transition period where there will be cars on the roads with very little to no ADAS technology in combination with cars with a large variety of ADAS technology. There is the important question of which skills should be taught in a driver training program and in driver instructor education. The introduction of more automation in cars will lead to a change in the skills needed for the driver, and hence will bring about a need for a change in the competence of the driver instructor. This, in turn, will probably affect driver instructor education.

4 AUTOMATED AND ADVANCED NEW TECHNOLOGY IN CARS IN REGARD TO DRIVER TRAINING AND DRIVER INSTRUCTOR EDUCATION

There are some obvious strengths regarding more automation in cars as opposed to fully manual cars. First of all, the workload will decrease for the human driver. With more technology taking over tasks such as changing gears, keeping the speed stable, avoiding collisions with pre-crash systems, navigation, and so forth, the driver can pay attention to other aspects. However, it is commonly known that when humans supervise a system as opposed to being an active participant, attention seem to fall (e.g. Yerkes and Dodson 1908). Even though there are many benefits such as the probability of a lower accident rate, there are also several concerns regarding automation. Most of these concerns are about when the driver needs to take over a vehicle, for instance in critical conditions (Son and Park 2017) or intention to use/user resistance (Kyriakidis, et al. 2015; König and Neymar 2017). When technology takes over many of the tasks, and works most of the time, driver skills will decrease. This is because maintaining skills without practice is probably not possible. However, very little information exists on driver training in regard to how to learn to drive with new technology as a new driver, or driving cars with new technology as an experienced driver (Harms and Dekker 2017). The topic of learning to use the technology is not even mentioned when opportunities and barriers on a societal level are considered (Fagnant and Kockelman 2015). However, the use of the technology on the market today, such as, for instance, lane assist and Adaptive Cruise Control (ACC) should perhaps be taught after proper driver skills are acquired. For instance, technical driving using lane assist could be perceived as uncomfortable as the technical reaction of the car is generally slower compared to a driver. When turning, for instance, the car is often too far out the curve before the turn is performed and this can

be repeated several times during the turn. If this was the behaviour of a learner driver during a lesson, the instructor would not have considered the technical driving skills to be adequate. This means that the driver must be skilled in order to understand that the car's behaviour is not adequate, and respond accordingly. The driver requires both good driver skills and an understanding of how the technology works, together with its advantages and limitations. On the other hand, technology such as lane assist could probably be of support in the event that an unexpected incident occurs and the driver loses control of the car. As single vehicle off the road together with head-on accidents are the most frequent accidents with the highest death rate in Norway for the past decades (SSB 2018a), this technology could potentially save lives. However, perhaps, it should not be trusted for use on a regular basis. In driver instructor education today, the teaching is that when driver assistance systems take over, the driving is not optimal. Thus, the systems could be there as a backup, but not trustworthy enough to be used regularly. The driver should drive the car. Furthermore, if such systems are to be used while driving, there are other considerations involved. For instance, regarding ACC, it is a technological system that perhaps works better in some driving conditions than in others. As an example, on icy roads, or in higher density traffic in a more complex driving environment, it might be a better solution to control speed manually. Making the correct decisions on when to use, and when not to use, technology while driving requires good driving skills.

Regarding driver instructor education and driver training, it seems that the introduction of ADAS technology requires that elements are added to the education and training rather than removed. Additionally, operating these technologies should perhaps be a larger part of driver training, driver testing, and hence driver instructor education.

Technology has always had an impact on the content of the Norwegian driver education curricula. For instance, driving on slippery roads has been a mandatory part of driver training in Norway since 1975. In the early days, the learner drivers were trained to manually adjust the brake pedal in different ways to minimise the braking distance, on ice and snow, as much as possible. After the ABS braking system was introduced and became common in most cars, the content of driver training on slippery roads changed and focused more on letting the learner drivers experience that the ABS system enabled them to brake as hard as they could and to simultaneously use the steering wheel to control the car (NPRA 1995; NPRA 2005). However, the main difference between the ABS brakes transition and the present technological transition, is that ABS brakes became common in many cars and used the same way of braking in all brands of car. The driver needed to change how to move the foot while braking, but the brakes were in the same place, the basic movement was the same, and most brands of car had the same system. Nowadays, new technological solutions such as ACC, are different in different makes of car where some brands for instance have a switch on the right side of the steering wheel, while others have a button on the front or on the left side of the steering wheel. This lack of standardisation could be confusing and hence could distract the driver. All kinds of different solutions such as these, and different software solutions in touchscreens in new cars may have as a result that it may not be as easy as previously to drive a car that the driver has not driven before, due to a wide variety of technological solutions. It could be difficult to know which technological solutions are included in the car, and difficult to know how to use the technology. Currently, distractions for the driver are about to increase due to in-vehicle devices. This runs counter to the necessity of keeping an eye on road (Wickens et al. 2004).

There is a possibility that the answer to this is to have differentiated driving licenses and not a standardised license such as we are used to today, because technology in cars is too varied and unstandardised. It should be a matter for discussion as to when cars are so different from each other that a standardised driving license is no longer good enough.

Increased technology has affected the training situation for a long time, and, in Norway, one example of an aspect that is in a transitional phase, is the trend that new cars are not equipped with manual gears. There are two important aspects to this situation. First, we see that the educational system does not keep up with the speed of technological development. Toyota for instance, sold more than 99% of new cars equipped with automatic gears so far in 2017, in Norway (Korsvoll 2017). Thus, the driver will not need to learn how to use manual gears as automatic gears will most probably become the new normal. However, in Norway, driver training is based on manual gears, and the education of driver instructors is based on vehicles equipped with manual gears. Perhaps the driver instructor program should focus instead on other tasks rather than teaching new drivers how to drive with manual gears. If a technology as basic as gears is hard to keep up with regarding a transition from manual to more automation, it could be a challenge when now even more technologically equipped cars enter the market.

Second, the gearing system is an example where different technological equipment in cars requires

different types of license for the driver. In Norway, as in the EU, you are allowed to drive an automatic car if your license is for manual gears, but not the other way around. You are not allowed to drive a car with manual gears if your license is for automatic driving (FOR 2017). For this reason, many driving schools only have manual gears in their cars, as for instance learner drivers know that they will probably buy a cheaper car with manual gears when they have their license. A solution such as this might also include more ADAS technology in the years to come. There could be different licenses based on the technology in the car you drive.

The rapid speed of introducing new technology seems to be happening faster than the changes in the educational system. Furthermore, if you have received your class B driving license, there is no retesting or system to update your driving skills, so there is a question as to how these drivers should learn how to operate new technology properly. Additionally, for driver instructors who are already authorised, there are no mandatory courses for updating their competence, so another question could be how they should get the necessary skills to teach new and existing learner drivers. If the two-year university education to become a driver instructor in Norway adjusts today, the market will not change completely for many years. Nevertheless, the rapid speed of technological progress will continue.

5 NEW DRIVER SKILLS REQUIREMENTS

In order to know which skills a driver must have, we need to know how the car works. For example, the GDE matrix has been the basic understanding of the driving skills that is necessary for a driver to have and thus, one of the central elements in the driver instructor education. The GDE-matrix consists of five levels, where the lowest level is vehicle manoeuvring, the second level is mastering traffic situations, the third level is goals and context of driving, the fourth level is goals for life and skills for living (Keskinen 1996 in Hatakka et al. 2002), and the fifth level is social skills (Keskinen 2014; Keskinen et al. 2010). However, the situation regarding new technology in the car is also changing the skills needed for a driver. It seems to be time to redefine which competence a driver must hold, and the GDE matrix may not be the optimal way to define the necessary skills in the future. If cars become more or less self-driving and automated, perhaps the lower stages of the GDE matrix might not correspond with the actual skills that are needed to drive a car.

Another example is the driving process, which might be explained using a basic information processing model (e.g. Wickens and Carswell 2006). This model assumes that information is *perceived*, then *processed* before *decisions* are made based on how the information is processed and *action* is then taken. In regard to the driving process, the question is who is collecting the information and who is responsible for collecting which information, the car or the human? As an example, when driving with ACC, the driver needs to monitor the environment and collect information on driving conditions as the car does not collect information, for instance, on the road conditions such as rain or ice or dry asphalt. Furthermore, the system does not correspond with any other systems in the car, so, for instance, if the car skids and the traction gets the car back on track, the ACC does not take the slippery road condition into consideration, and will only work to get the car back to the required speed or distance from the car in front. This assumes Driver Driving and Driver Monitoring with this technology (Banks and Stanton 2017). Regarding another technology, lane assist, the same aspect occurs as, for instance, lane assist will not work without proper road markings. Therefore the driver must pay attention to whether the road is properly marked or not. This is information a driver normally would not need to pay that much attention to if driving the car, as the driver would most likely hold the steering wheel and stay on her/his side of the road regardless of the quality of the road marks. The technology could thus make the driver pay attention to the road closer to the vehicle rather than paying attention to the road traffic environment further ahead. Additionally, regarding decision making, it could be questioned as to whether it is the car or the driver that makes the decisions. For instance, with ACC, if the car does not collect information on the road conditions, it cannot be responsible for making decisions in this regard. The driver must monitor and make decisions based on the information gathered and processed. Finally, the question is who takes action based on the information and decisions? If the car does not make decisions or gather relevant information, it probably cannot take appropriate action, meaning this would be the driver's responsibility.

So, what do we hand over to the car and what is left to the driver? The question will have different answers for different technologies. If, for instance, using the same scenario as with the ADAS technology, adaptive lighting, there is a different situation. Here the car gathers information on for instance the light conditions in the environment, and oncoming cars, and makes decisions based on the information gathered and takes action to turn lights on or off or chooses the degree of brightness. Thus, the driver will not need to use as much cognitive capacity for this operation.

Another issue that complicates which driving skills are needed is the lack of standardisation between the car manufacturers on how the new technology should interact with the driver. For instance, there are several different solutions to touchscreen software in cars. So, the skills of the driver need to correspond to the actual car the driver will be driving, the technological solutions in the car and how the technological solutions interact with the driver.

It seems that training needs adjustment in order to meet the new digitalisation of the future. However, in order to change what we teach to the learner drivers, we probably need to start with the educational institutions who educate the driver instructors. In addition, there is the question of if and how to re-educate driver instructors who are already certified as driver instructors for traditional driving. In Norway alone, there are more than 1,000 driving schools, and providing courses for the instructors in all of these schools will take time and effort. This time does not seem to be available at the speed at which changes are happening today.

One solution could be that the manufacturers are responsible for the specific technological training for drivers, and license drivers for their technology. A solution such as this also requires consideration as to what training and testing for such a license should involve, in addition to who is responsible for the training and testing. Today it is the National Road Authorities in Norway who conduct the testing of learner drivers in order for them to qualify for a driving license. Therefore, to maintain the driver skills requirements, the testing could be the responsibility of the authorities. This testing could include drivers ability to drive and supervise the systems in addition to how to respond to alarms and warnings. Therefore, one could think of driver education that comes in two levels. In that case, a standard learner driver could learn how to handle a manual car as level one in a standard driving school, but also learn how to operate and supervise a car of the future as level two. How to drive a car with technological solutions could, with this system, be up to the manufacturers to teach properly to all drivers, and be tested by the road authorities.

We see in aviation, for instance, that pilots are trained in simulators in order to uphold the required skill level to fly an aircraft. This is partly because flying with a high level of automation decreases flying skills. This could be a solution for drivers as well. In order to keep their driving license, drivers could be required to have a certain amount of simulator training in order to uphold driving skills because their cars have ADAS technology. However, this will require an increase in simulators for one, and in Norway today there are between 5–10 simulators for driving license B. Furthermore, retraining to uphold skills requires a system where everyone holding a driving license in Norway has training. A system will also be needed to deal with the bureaucratic aspects. Thus, there are some obvious obstacles to such a solution in regard to costs and resources in addition to the issue of how society would respond to it and if there will be public acceptance for such a system, and the political will to implement it.

6 CONCLUDING REMARKS

How will ADAS technology in cars potentially affect driver training and driver instructor education and which new skills might be needed for a driver? This was the question we wanted to examine closely in this paper.

We must be honest and admit that today, we do not know how to provide general training for more technology equipped cars, or even for self-driving cars. To be able to assess a good training program, it is essential that we know what we are training for. Today, however, with the vast variety of technological solutions on the roads, the lack of standardisation of the software and devices in cars, in addition to a future which seems to have new technological solutions happening quickly, it seems difficult for the driver instructor industry to prepare and come up with an optimal solution in the short run.

Hence, we recommend that the content of driver training and driver instructor education should preferably be increased, not decreased, as good driving skills are still needed in addition to good understanding on how to operate the technology. This is because, as of today, ADAS technology in cars seems to result in more rather than less work for the driver.

REFERENCES

ACEA, 2017. Average vehicle age. ACEA European Automobile Manufacturers Associations. http://www.acea.be/statistics/tag/category/average-vehicle-age.

Banks, V & Stanton, N. 2017. Analysis of driver roles: Modelling the changing role of the driver in automated driving systems using EAST, *Theoretical Issues in Ergonomics Science.* doi.org/10.1080/1463922X.2017.1305465.

Banks, V. et al. 2017. Is partially automated driving a bad idea? Observations from an on road study. *Applied Ergonomics, 68,* 138–145.

Bertanocelli, M. & Wee, D. 2015. *Ten ways autonomous driving could redefine the automotive world.* McKinsey and Company. Downloaded November 25th from https://www.mckinsey.com/industries/automotive-

and-assembly/our-insights/ten-ways-autonomous-driving-could-redefine-the-automotive-world.
Billings, C. 1997 *Aviation automation: The search for a human-centered approach.* Englewood Cliffs, NJ: Erlbaum.
Elvik, R. and Høye, A. 2015. *Hvor mye kan drepte og hardt skadde i trafikken reduseres?* (English translation: How much can the number of those killed and severely injured in road traffic be reduced?) Transportøkonomisk Institutt: TOI Report 1417/2015.
European Commission 2010. *Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions. Towards a European road safety area: policy orientations on road safety 2011–2020.*
Fagnant, D. J. & Kockelman, K. 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A, 77,* 167–181.
Forskrift om trafikkopplæring (FOR) 2017. https://lovdata.no/dokument/SF/forskrift/2004-10-01-1339?q=trafikkopplæringsforskrift [Norwegian Regulations for traffic education] The Norwegian Ministry of Transport.
Harms, I.M & Dekker, G-M. 2017. ADAS: from owner to user. Insight in the conditions for a breakthrough of Advanced Driver Assistance Systems. *Connecting Mobility NL.*
Hatakka, M. et al. 2002. Goals for driver development, published in EU MERIT Project (2004). *Driving instructors' education in Europe: A long-term vision*, Working Paper for Workshop 1, (21 January, 2005).
Hatakka, M. et al. 2002. From control of the vehicle to personal self-control; broadening the perspectives to driver education. *Transport Research Part F, 5,* 201–215.
Kaber, D. B. & Endsley, M. R. 2007. The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomic Science, 5,* 113–153.
Keskinen, E. 2014. Education for older drivers in the future. *IATSS Research, 38,* 14–21. doi.org/10.1016/j.iatssr.2014.03.003.
Keskinen, E. et al. 2010 *GDE-5PRO and GDE-5SOC: Goals for driver education in a wider context–professional and private drivers in their environment* Unveröffentlichtes Manuskript, Universität Turku, Finland.
König, M. & Neymar, L. 2017. Users' resistance towards radical innovations: The case of self-driving cars. *Transportation Research Part F, 44,* 42–52.
Korsvoll, R. 2017. Nå forsvinner det manuelle giret. (Now the manual gear disappears. Our translation) *Motor:* Downloaded December 12th from https://www.motor.no/artikler/2017/oktober/na-forsvinner-det-manuelle-giret/.
Kyriakidis, M. et al. 2015. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F, 32,* 127–140.
Lee, J.D. 2006. Human factors and ergonomics in automation design. In G. Salvendy (Ed.) *Handbook of human factors and engineering* (3rd ed.). Hoboken, NJ: John Wiley and Sons.
Michon, J. A. 1985. A critical view of driver behaviour models: What do we know, what should we do? In L. Evans and R. C. Schwing (Eds.), *Human behavior and traffic safety*, pp. 485–502. New York: Plenum Press.
NTP 2014–2023 Stortingsmelding 26 *Nasjonal Transportplan* Det Kongelige Norske Samferdelsdepartment.
NRK 30.04.2017 https://www.nrk.no/norge/null-omkomne-i-fire-fylker_-forerlose-biler-kan-halvere-ulykkene-1.13494298
NTSB 2014. *Descent below visual glidepath and impact with Seawall, Asiana Airlines Flight 214, Boeing 777-200ER, HL7742, San Francisco, California July 6, 2013.* Washington, DC: Author. Retrieved November 25th from https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1401.pdf
NTSB 2017. *Driver errors, overreliance on automation, lack of safeguards, led to fatal Tesla crash.* National Transportation Safety Board. Retrieved November 20th 2017 from https://www.ntsb.gov/news/press-releases/Pages/PR20170619.aspx
Nord universitet 2017. Studieplan for Trafikklærer, høgskolekandidatstudium. (English translation: Driving Instructor Education for License B. Nord University.) Received December 15th 2017 from: https://www.nord.no/no/studier/trafikklarer-hogskolekandidatstudium#.
Norwegian Public Road Administration (NPRA) 1995. Curricullum for driver training category B, BE and code B 96.
Norwegian Public Road Administration (NPRA) 2005. Curricullum for driver training category B, BE and code B 96.
Norwegian Public Road Administration (NPRA) 2013. Curricullum for driver training category B, BE and code B 96. www.vegvesen.no.
Parasuraman, R., and Byrne, E. A. 2003. Automation and human performance in aviation. In P. Tsang and M. Vidulich (Eds.), *Principles of aviation psychology* (pp. 123–155). Lanham, MD: Rowmand Littlefield.
Peräaho, M, Keskinen, E, Hatakka, M. 2003. Driver competence in a hierarchical perspective; implications for driver training. University of Turku, Traffic research Rismark, M., and Sølvberg, A.M. 2007. Effective dialogue in driver education. *Accident Analysis and Prevention, 39,* 600–605.
Rismark, M., and Sølvberg, A.M. 2007. Effecive dialogues in driver education. *Accident Analysis and Prevention, 39,* 600–605.
SAE 2014. *Automated driving levels*, SAE International Standard J3016.
Salas, E. et al. 2006. Design, delivery, and evaluation of training systems. In G. Salvendy (Ed.) *Handbook of human factors and engineering* (3rd ed.). Hoboken, NJ: John Wiley and Sons.
Sætren, G. B. & Laumann, K. 2015. Effects of trust in high-risk organizations during technological changes. *Cognition, Technology and Work, 17,* 131–144.
Son, J. & Park, M. 2017. Situation awareness and transition in highly automated driving: A framework and mini review. *Journal of Ergonomics 7*, DOI: 10.4172/2165-7556.1000212.
SSB.no 2018a. Veitrafikkulykker med personskade (Road traffic accidents involving personal injury).
SSB.no 2018b. Personbiler alder (Passenger cars, age).
Stanton, N. and Young, M.S. (1998). Vehicle automation and driving performance. *Ergonomics, 41*(7), 1014–1028.

SVV 2017. Statens Vegvesen (Norwegian Road Authorities) https://www.vegvesen.no/

TechEmergence.com 2017. The self-driving car timeline—predictions from the top 11 global automakers. Consulted on November 5th, 2017 at: https://www.techemergence.com/self-driving-car-timeline-themselves-top-11-automakers/

Young, M.S. & Stanton, N. A. 2007. What's skill got to do with it? Vehicle automation and driver mental workload, *Ergonomics*, 50:8, 1324–1339, DOI: 10.1080/00140130701318855.

Wickens, C.D. & Carswell, C.M 2006. Information processing. In G. Salvendy (Ed.) *Handbook of human factors and engineering* (3rd ed.).Hoboken, NJ: John Wiley and Sons.

Wickens, C.D. et al. 2016. *Engineering psychology and human performance* (4th ed.). New York: Routledge

Wickens, C.D., . et al. 2004. *An introduction to human factors engineering* (2nd ed). New Jersey: Pearson Prentice Hall.

Wilmink I. et al. 2008. *Impact assessment of intelligent vehicle safety systems*. eIMPACT Deliverable D4. Version 1.0.

Yerkes, R.M., and Dodson, J.D. 1908. The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology & Psychology,* 18, 459–482. doi.org/10.1002/cne.920180503.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A method to evaluate an aircraft operational risk

Š. Hošková-Mayerová
Department of Mathematics and Physics, University of Defence, Brno, Czech Republic

M. Zieja & M. Woch
Division for IT Support of Logistics, Air Force Institute of Technology, Warsaw, Poland

J. Tomaszewska
Faculty of Aviation, Polish Air Force Academy, Dęblin, Poland

M. Matyjewski
Faculty of Power and Aeronautical Engineering, Division of Fundamentals of Machine Design, Warsaw University of Technology, Warsaw, Poland

ABSTRACT: Improving performing and safety of the aircraft operation is one of the most important issues addressed by experts. Such improvement can result not only in the less frequent loss of equipment but primarily in the protection of health or saving lives of both crew members and others involved. Reducing such risks or minimizing impacts is possible by analyzing events, which had already occurred. In this paper, our main motivation consists in developing an effective and intelligent decision support system based on data mining techniques. In this context, data mining classifying algorithms with large datasets have been utilized to assess and analyse the risk factors statistically related to aircraft incidents in order to compare the performance of the implemented classifiers such as decision tree, discriminant and random forest. To underscore the practical cost, i.e., effectiveness of our approach, the selected classifiers have been implemented using statistical programming tools with datasets taken from the operation process. This analysis is expected to find the algorithm, which can support the decision taking.

1 INTRODUCTION

1.1 *Formulation of the problem*

Following the increasing requirements of flight safety and cost reduction, the problem of finding the optimum between the economic demands and acceptable level of risk arises in terms of reliability. Modern control systems equipped with computerized processes and extensive diagnostic tools do not often use all the information collected from the hardware level (Tloczynski 2017b). Moreover, some of the relations between events are often ignored or neglected. The article presents a new approach to increasing reliability of aircraft operations by predictive data analysis and increasing acceptable levels of safety.

2 FORMULATION OF THE PROBLEM - A STATISTICAL APPROACH TO SAFETY AVIATION PREDICTION

This article deals with the needs of the usage of statistical tools and methods of artificial intelligence, which enable to discover the relations between events stored in the database.

2.1 *Decision trees*

Decision trees are a class of predictive data mining tools which predict either a categorical or continuous response variable. They get their name from the structure of the models built. A series of decisions are made to segment the data into homogeneous subgroups. This is also called recursive partitioning. If presented graphically, the model can resemble a tree with branches (StatSoft Inc. 2013).

A decision tree is composed of nodes and splits of the data. The tree starts with all training data residing in the first node. An initial division is made using a predictor variable, segmenting the data into 2 or more child nodes. Divisions can then be made from the child nodes. A terminal node is the one where no more divisions are made. Predictions are made based on the behaviour of terminal nodes (Mueller et al. 2017).

Decision trees offer many advantages. One important advantage is the ease of interpretation

of a decision tree. While the tree can be complex, involving a large number of splits and nodes, users can interpret the model (Sedlacik & Cechova 2016). Additionally, making model predictions does not involve mathematical calculations as in General Linear Models. The predictions are based on decision rules. In classification problems, the user can specify misclassification cost. Decision trees tend to give good predictive accuracy and can allow for missing data in deployment (Hinz et al. 2017a).

2.2 *Logistic regression*

The statistical methods have been used so that the safe performance of the aviation operation could be determined. The logistic regression and the decision trees have been implemented as the most promising methods to reach this goal. The regression logistics model is based on the similar assumption as the model of linear regression; however, the former can be used in case the predicted variable is in binomial form (Babiarz 2016).

Logistic regression is a mathematical modelling approach that can be used to describe the relationship of several independent variables X_s to a dichotomous dependent variable, such as outcome - decisions D (Kleinbaum & Klein 2010). Other modelling approaches are possible as well; however, logistic regression is by far the most popular modelling procedure used to analyse epidemiologic data when the illness measure is dichotomous (Vintr & Valis 2011).

Formally, the model logistic regression model is as follows (Valis, Zak, & Pokora 2014):

$$\log \frac{p(x)}{1-p(x)} = \beta_0 + x \cdot \beta \tag{1}$$

where
p – logistic function of the probability,
β_0, β – constant terms representing unknown parameters.

Solving for p, this gives:

$$p(x) = \frac{e^{\beta_0 + x \cdot \beta}}{1 + e^{\beta_0 + x \cdot \beta}} = \frac{1}{1 + e^{-(\beta_0 + x \cdot \beta)}} \tag{2}$$

It should be noted that the overall specification is a lot easier to fathom in terms of the transformed probability that in terms of the untransformed probability (Shalizi 2013, Hinz et al. 2017b).

3 EXPERIMENTAL STUDY OF STATISTICAL METHODS

In order to carry out experimental studies, flights performed on a third generation fighter aircraft of two types were considered. The data were derived from the last 8 years of operation exploitation process in Poland. Flights were analysed in terms of incidents or undesirable events occurrences.

Table 1. Flight parameters.

Variable name	Unit	Type
Time after sunrise	min	Continuous
Time after sunset	min	Continuous
Month		Continuous
Aircraft type		Categorical
Age of aircraft	day	Continuous
Atmospheric conditions		Categorical
Name of the military department		Categorical
Real time in air	min	Continuous
Number of crew members		Continuous
Flight-hour of the first pilot	hour	Continuous
Flight-hour of the first pilot performed on a given aircraft type	hour	Continuous
Year of the promotion of the first pilot		Continuous
Subsequent departure of the first pilot on a given day		Continuous

Table 1 presents variables, which represent the examined flight parameters.

3.1 *Decision trees results*

Evaluation of the performance of a classification model is based on the counts of test records correctly and incorrectly predicted by the model. These counts are tabulated in a table known as a confusion matrix. Although a confusion matrix provides the information needed to determine how well a classification models perform, summarizing this information with a single number would make it more convenient (Bořil & Čičmanec 2016). This can be done using a performance metric such as accuracy, which is defined as follows:

$$\text{accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

$$\text{error rate} = \frac{\text{Number of wrong predictions}}{\text{Total number of predictions}}$$

Figure 1 shows the obtained results with the use of decision tress.

In this analysis, 80% of the cases were selected as the testing samples.

Tables 2 and 3 shows the percentage of correct decisions for the training sample and the learning sample, respectively.

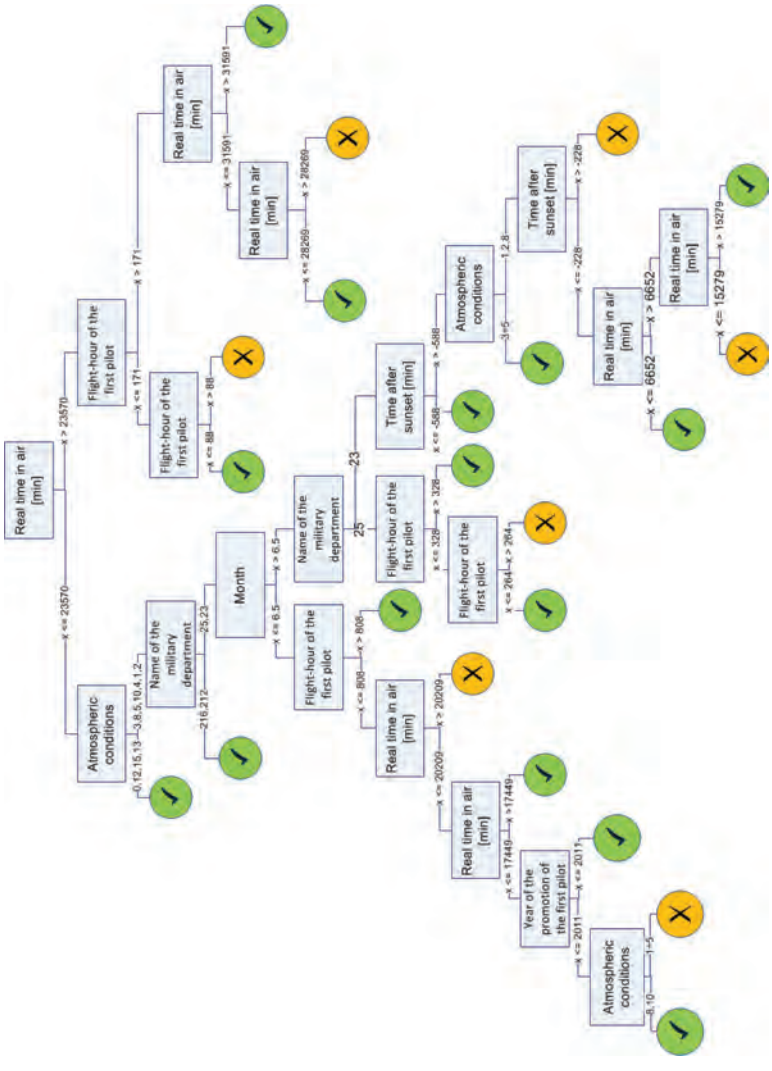


Figure 1. Sketch of the decision tree.

Table 2. Cross tabulation for training sample.

Reality \ Model	No incident	Incident
No incident	62.34%	31.46%
Incident	0.29%	5.91%

Table 3. Cross tabulation for testing sample.

Reality \ Model	No incident	Incident
No incident	60.15%	34.21%
Incident	3.01%	2.63%

3.2 *Logistic regression results*

In order to analyse the significance of parameters, not all available flight data were included. For the analysis, the entire data from the operation process with the recorded incidents were used, and the same number of randomly selected flights without an incident was considered.

In the proper statistical model all of the parameters should be quite different from zero (Koucky & Valis 2007). There are cases when the model has to comprise parameters that are not statistically significant however, the decision has to take them into account. Such a justification may be either theoretical knowledge of a phenomenon or experience from another, similar analysis. Wald statistics

Table 4. Cross tabulation for generalized linear model results.

Reality \ Model	No incident	Incident
No incident	31%	18%
Incident	23%	29%

Table 5. Cross tabulation for generalized linear model results – backward elimination.

Reality \ Model	No incident	Incident
No incident	33%	18%
Incident	25%	25%

can be used for the predictors determination and the associated test probability level p. The simplest way to obtain a model with all statistically relevant parameters (Table 7) consists in removing those parameters from the model which are not significant; this model is presented in Table 6. The strategy of construction of the model was achieved by backward elimination. The elimination starts with the entire model and then the most probable variables (test probability level p) are eliminated. The elimination finishes at the moment when the model comprises only statistically significant variables.

Tables 4 and 5 show the percentage of correct decisions for the model with all parameters and for the model with all statistically relevant parameters, respectively.

In the present analysis, the logit node with implementation was applied, and the choice of predictors has been made with automatic execu-

Table 6. Assessment of the parameters significance.

Variable name	Value	Value of the parameter	Std. err.	Wald's stat.	p
Free term		–6.86E+01	3.99E+01	2.95E+00	8.58E-02
Real time in air		2.76E-05	7.53E-06	1.34E+01	2.47E-04
Time after sunrise		5.87E-04	3.95E-04	2.22E+00	1.37E-01
Time after sunset		–1.22E-03	4.19E-04	8.53E+00	3.49E-03
Month		3.80E-02	1.84E-02	4.24E+00	3.95E-02
Flight-hour of the first pilot		4.48E-04	2.28E-04	3.86E+00	4.94E-02
Flight-hour of the first pilot performed on a given aircraft type		–4.10E-04	3.18E-04	1.67E+00	1.97E-01
Year of the promotion of the first pilot		3.65E-02	1.89E-02	3.76E+00	5.26E-02
Subsequent departure of the first pilot on a given day		–1.16E-01	1.13E-01	1.06E+00	3.04E-01
Age of aircraft		–2.64E-04	3.10E-04	7.27E-01	3.94E-01
Aircraft type	103	7.14E-02	1.43E-01	2.48E-01	6.19E-01
Aircraft type	102	1.96E-01	1.68E-01	1.36E+00	2.44E-01
Aircraft type	101	–3.66E-01	1.51E-01	5.85E+00	1.55E-02
Atmospheric conditions	0	2.30E+00	8.30E-01	7.65E+00	5.69E-03
Atmospheric conditions	1	2.91E+00	7.47E-01	1.51E+01	9.99E-05
Atmospheric conditions	2	2.76E+00	7.54E-01	1.34E+01	2.52E-04
Atmospheric conditions	3	3.19E+00	7.74E-01	1.70E+01	3.73E-05
Atmospheric conditions	4	3.12E+00	7.56E-01	1.70E+01	3.76E-05
Atmospheric conditions	5	2.81E+00	8.28E-01	1.15E+01	7.00E-04
Atmospheric conditions	6	–1.84E+01	7.59E+00	5.90E+00	1.51E-02
Atmospheric conditions	8	2.83E+00	7.94E-01	1.27E+01	3.62E-04
Atmospheric conditions	9	2.43E+00	1.44E+00	2.85E+00	9.16E-02
Atmospheric conditions	10	3.00E+00	8.72E-01	1.18E+01	5.84E-04
Atmospheric conditions	12	3.92E+00			
Atmospheric conditions	15	–1.41E+01			
Type of flight	0	1.81E-01	3.77E-01	2.31E-01	6.31E-01
Type of flight	1	4.90E+00	3.20E-01	2.35E+02	0.00E+00
Type of flight	2	4.86E+00	2.35E-01	4.28E+02	0.00E+00
Name of the military department	23	–5.01E+00	2.60E-01	3.71E+02	0.00E+00
Name of the military department	25	–4.73E+00			
Name of the military department	216	–1.00E+00			

Table 7. Assessment of the parameters significance—backward elimination.

Variable name	Value	Value of the parameter	Std. err.	Wald's stat.	p
Free term		–4.85E-01	2.25E-01	4.66E+00	3.08E-02
Real time in air		3.06E-05	6.39E-06	2.30E+01	1.64E-06
Name of the military department	23	–2.73E-01	1.82E-01	2.25E+00	1.34E-01
Name of the military department	25	1.46E-01	1.85E-01	6.20E-01	4.31E-01
Name of the military department	216	–6.94E-01	3.57E-01	3.77E+00	5.21E-02
Time after sunset		–5.39E-04	2.38E-04	5.12E+00	2.37E-02
Month		4.18E-02	1.77E-02	5.55E+00	1.84E-02
Aircraft type	103	2.49E-02	1.30E-01	3.64E-02	8.49E-01
Aircraft type	102	1.91E-01	1.55E-01	1.52E+00	2.18E-01
Aircraft type	101	–4.78E-01	1.42E-01	1.13E+01	7.61E-04

tion. The logit model is linear and its complexity is determined by the number of independent variables included in the design (Valis et al. 2016, Tloczynski 2017a). The more variables there are, the greater the risk of failure of the model is.

4 CONCLUSIONS

The work was carried out by the three research groups: University of Defence, Air Force Institute of Technology and Polish Air Force Academy as a part of the activities of the recently international cooperation between military universities and the institute.

In this article the comparison between the results of two models, regression model and decision trees, is presented. It was shown that both of them are giving similar results around 60% of accuracy. Results calculation should consider the fact they are slightly different for various models; the process of the training sample selection has to be consider as well.

The algorithms give us, step by step, importance of predictors, which can be used in the processes of decision taking.

This particular analysis resulted in finding the first, general algorithm, which can support the decision taking. Moreover, this algorithm is general, and could be used for different type of aircraft. In order to reach better accuracy, further calculations are expected.

ACKNOWLEDGEMENTS

The work of first author presented in this paper was supported within the project for "Development of basic and applied research developed in the long term by the departments of theoretical and applied bases of FMT (Project code: DZRO K-217) supported by the Ministry of Defence of the Czech Republic.

REFERENCES

Babiarz, B. (2016, Jul). Reliability analysis in subsystem of heat supply. In IEEE (Ed.), *2016 International Conference on Information and Digital Technologies (IDT)*, Rzeszow, Poland, pp. 11–16.

Bořil, J. & L. Čičmanec (2016). Tactical mission execution on flight simulators with evaluation of measurement data. In *Transport Means 2016*, Juodkrante, Lithuania: Kaunas University of Technology, pp. 264–267.

(Electronic Version): StatSoft, Inc. (2013). *Electronic Statistics Textbook*. Tulsa: OK: StatSoft.

Hinz, M., F. Hienzsch, & S. Bracke (2017a). Detection of distinctions in car fleets based on measured and simulated data. In *RAMS 2017 63rd Annual Reliability and Maintainability Symposium*, Orlando, Florida, U.S.A.

Hinz, M., F. Hienzsch, & S. Bracke (2017b, Sep). Development of two methods for the characterisation of an automotive fleet behaviour based on the simulation of single car rides. In *Risk, Reliability and Safety: Innovating Theory and Practice*, Glasgow, Scotland, pp. 1593–1598.

Kleinbaum, D.G. & M. Klein (2010). *Logistic Regression A SelfLearning Text*. New York Dordrecht Heidelberg London: Springer.

Koucky, M. & D. Valis (2007, June). Reliability of sequential systems with a restricted number of renewals. In T. Aven and J.E. Vinnem (Eds.), *Proceedings and Monographs in Engineering, Water and Earth Sciences 2007*, Stavanger, Norway, pp. 1845–1849.

Mueller, A., M. Hinz, & S. Bracke (2017, Sep). Optimization of the dental implant testing based on fem simulation of fatigue and accelerated life. In *Risk, Reliability and Safety: Innovating Theory and Practice*, Glasgow, Scotland, pp. 16–22.

Sedlacik, M. & I. Cechova (2016). Computer-adaptive testing: Item analysis and statistics for effective testing. In N. Jancarik (Ed.), *Proceedings on the European Conference of e-Learning*, pp. 650–656.

Shalizi, C.R. (2013). Lecture 12, Logistic regression. http://www.stat.cmu.edu/ cshalizi/uADA/12/.

Tloczynski, D. (2017a). Air transport service in academic research at polish airports. In G. Sierpinski (Ed.), *Advances in Intelligent Systems and Computing*, Volume 505, Katowice, Poland, pp. 23–32. Conference: 13th Scientific and Technical Conference on Transport Systems. Theory and Practice.

Tloczynski, D. (2017b). Security as a determinant of choice of air transport service and air carrier on the basis of research. *Scientific Journal of Silesian University of Technology-Series Transport 95*, 213–222.

Valis, D., L. Zak, & O. Pokora (2014). Engine residual technical life estimation based on tribo data. *Eksploatacja i Niezawodnosc Maintenance and Reliability 16*(2), 203210.

Valis, D., L. Zak, O. Pokora, & P. Lansky (2016). Perspective analysis outcomes of selected tribodiagnostic data used as input for condition based maintenance. *Reliability Engineering & System Safety 145*, 231–242.

Vintr, Z. & D. Valis (2011). A tool for decision making in kout-of-n system maintenance. *Applied Mechanics and Materials 110–116*, 5257–5264.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Evaluating models for the inclusion in a safety assessment framework for efficient transport

P. Karpati & A.A. Hauge
Institute for Energy Technology, Halden, Norway

T. Sivertsen
Bane NOR SF, Oslo, Norway

B.A. Gran
Institute for Energy Technology, Halden, Norway
NTNU, Trondheim, Norway

ABSTRACT: This paper presents the experiences from applying SysML models as support for establishing the safety requirements specification of a new safety-related railway application. The new railway application is a software-based system for securing work areas, meaning it prevents railway traffic in areas along the track allocated to maintenance. The experiences are collected within the Safety Assessment Framework for Efficient Transport (SafeT) project managed by Bane NOR. Bane NOR is the government agency that owns, operates and develops the Norwegian railway infrastructure. The objective of the SafeT framework is to offer a systematic, reusable way for creating system wide conceptual design models and based on them, creating a common risk model, which in turn will facilitate safety assessment, establishing the requirements specification, and safety demonstration of the system under consideration. The paper introduces the SafeT project as context of the work and presents experiences on the application of SysML for the conceptual system design of the new securing work areas application. The paper also discusses whether SysML models fit the SafeT framework's objectives.

1 INTRODUCTION

The SafeT project aims at developing a framework that supports the implementation of EN 50126 (CENELEC, 2017) and thereby of the Common Safety Methods for Risk Assessment (CSM RA) (EU, 2013). Figure 1 illustrates which phases of EN 50126 that is within the scope of the current SafeT work and this paper, annotated by a dark grey rectangle.

The current focus is on the development phases 1 to 4 of EN 50126. In these phases of a systems life cycle, Bane NOR takes a lead role in the development while successive development phases to a large extent are outsourced. The SafeT framework intends to support the development of the core artefacts within the system life cycle. In the early stages of the life cycle, in the part of the framework that concerns the in-house conceptualisation, the core artefacts are: 1) the conceptual system design model; 2) common risk model; and 3) requirements specification.

The main objective of the SafeT framework is to offer a systematic, reusable way for creating system wide conceptual design models and based on them, creating a common risk model, which in turn will facilitate the safety assessment, requirements specification and safety demonstration of the system under consideration, throughout the system's lifetime.

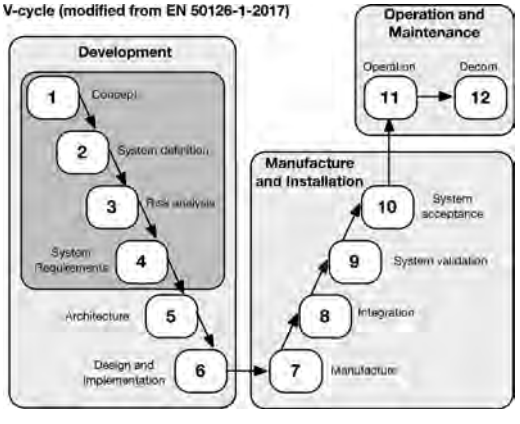


Figure 1. Scope of paper and relationship to EN50126.

2 RELATED WORK

International safety standards, such as EN 50126, provide requirements and guidance on how to carry out safety demonstration and assessment. Although most safety standards often view the safety of a system as a function of the reliability of its components, little guidance is provided on how to derive safety requirements and acceptable risk for components whose failure rates are not known. Particularly, it is often difficult to derive safety requirements for logical components such as the software. The problem can be formulated from a consideration of the following two important tasks in the development of safety critical systems: (1) *establishing the requirements to the system*, and (2) *ensuring that the system fulfils these requirements.* The safety requirements should be established through risk assessment and hazard analysis, and fulfilled through the use of techniques and measures adequate for the risk level. The framework proposed in SafeT has much of its inspiration from theoretical aspects of international safety standards such as IEC 61508 (IEC 61508). The novel part of the framework is fivefold: reusability, modularity, unification, transparency and argumentation.

Next, many past projects that relate to the topics of SafeT are briefly introduced. The OPENCOSS project provides a common language for both safety-case and standards-based approaches for certification. The CHESS project seeks to improve Model Driven Engineering practices and technologies to better address safety, reliability, performance, robustness and other extra-functional concerns while guaranteeing correctness of component development and composition for embedded systems, and offers a modelling language and editor. The CHESS modelling language and editor is a collection-extension of subsets of standard OMG languages such as UML (UML), MARTE (MARTE) and SysML (SysML).

The EU funded project MODSafe provides a risk analysis method purposed to combine potential hazards, safety requirements and functions, and link these elements to a generic functional and object-oriented structure of a guided transport system. The SaferCer project (Björnander, 2012) provides a generic process model for integrated certification and development of component based systems, including an overall picture of the development and verification of components and systems. ASCOS (Roelen, 2014) focuses on safety and certification of new aviation operation and systems, including advices on methods and tools for safety based design. ModelMe! (Falessi, 2011) provides a tool-supported traceability framework where the tool automatically extracts the safety-related slices of SysML design models.

Another approach is the AltaRica Language (Griffault, 1998). AltaRica is an object-oriented modelling language dedicated to performance evaluation of complex systems. The main motivation for its creation was the difficulty to design, to share and most importantly to maintain safety and reliability models such as fault trees, event trees, Markov chains or stochastic Petri nets. The application and further development of the language is a continuous research activity at NTNU (Legendre 2017).

Of relevance is also CORAS (Lund, 2013; Gran, 2004) which provides a methodology for model-based risk assessment, integrating aspects from partly complementary risk assessment methods and state-of-the-art modelling methodology.

The SafeT project has also reviewed a number of ongoing and past industrial experiences among the project partners related to the use of design and risk models to facilitate the safety assessment and demonstration of complex systems. Some of the challenges observed in these projects have also been reported earlier within aviation (Gran, 2007). Finally, the CHASSIS method (Raspotnig, 2018) utilizes UML use cases and sequence diagrams with HAZOP guidewords to integrate safety and security considerations for early requirements determination.

3 CONCEPTUAL MODELLING

3.1 *The role of models*

An important aspect of SafeT is the role of system modelling in the RAMS process defined in EN 50126, in particular for supporting the risk assessment process and the identification of safety requirements. An example of a modelling task related to the RAMS life-cycle phases is the introduction of the system under consideration in a model at the railway system level (phase 1). In phase 2, the model can be refined as necessary to support the description of system objective, mission profile, boundaries and external interfaces and interactions. In phase 3, the model can be further refined to support the establishment of the risk model, followed by a refinement in phase 4 to support the specification of requirements and application conditions for the system under consideration. In addition to the system models, there is also a need to establish risk models that capture the relations between the different hazards, causes, barriers, accidents, and consequences identified in the hazard identification performed at the different system levels. SafeT looks into the possibilities to enhance the system and risk modelling tasks by the appropriate application and combination of techniques evaluated against a set of criteria derived from the relevant standards.

SafeT intends to support the implementation of EN 50126 by giving guidance on what kind of models can be used, and how they can be utilized, in the life cycle phases within the standard. In this paper, we focus on the application of models. In another paper, we focus on the risk assessment part (Skogvang, 2018). An important research problem in the SafeT project is how the use of models throughout the life cycle of a system can be integrated in a way that facilitates the overall safety demonstration and assessment. The models will serve different needs, related to the analysis of system, risk, requirements, etc. SafeT aims at arriving at a set of techniques that covers the modelling needs in the different life cycle phases, with a current focus on the first four phases aimed at establishing the requirements specification. Some examples of the prospective use of models are:

- describing and analysing the static structure of a system and its constituent parts, down to the system level and the level of detail necessary to support analysis, independence demonstration, etc.;
- describing and analysing the behaviour of a system, internally as well as through its boundaries;
- describing and analysing a system's interaction with its environment, and how it affects, and is affected by, agents involved in its operation;
- supporting the activities involved in risk assessment and hazard control, including the identification of hazards at all system levels, their causes and possible consequences;
- supporting the derivation of the safety requirements needed to handle the hazards at the overall system level as well as technical hazards at any system level; and
- communicating the different design and risk aspects, as well as the safety argumentation as such, to the different stakeholders involved.

3.2 *Requirements to models*

To facilitate the selection of design and risk models, an initial set of 58 requirements to be fulfilled by the models is established within SafeT. The requirements were derived by reviewing the process requirements in the CENELEC standards EN 50126, 50128 and 50129 (CENELEC 2017, 2011 and 2003). The set of requirements acts as the evaluation criteria supporting the selection of techniques to be used in the development of the desired models. The identified modelling needs were reformulated in terms of requirements to the models as such and categorised as requirements concerning

- Structure: to model the static aspects of a system at any system level, e.g. the possibility to support any hierarchy of system levels, and describe any system level at the appropriate level of detail without introducing unnecessary detail and complexity at other system levels;
- Behaviour: to describe the dynamic aspects of a system at any level, e.g. the possibility to show how the behaviour and state of a system depends on, and changes with, the functionality of its sub-systems and components;
- Interaction: to describe the reciprocal impact between a system and its environment, e.g. the possibility to show how the environment can influence, or be influenced by, the system, including anything to which the system connects mechanically, electrically or by other means;
- Risk: to carry out the risk assessment and hazard control, e.g. the possibility to facilitate the identification of hazards associated with the system and events leading to these hazards, the determination of the risk associated with the hazards, and the identification of possible further safety requirements needed to reduce the risk to an acceptable level, at any system level;
- Requirements: to identify and specify safety requirements, e.g. the possibility to provide the details necessary to explain and understand the requirements to the functions to be provided by the system, as well as any additional requirements that are necessary to ensure proper functioning, including contextual and technical requirements;
- Design: to analyse the safety aspects of a design, e.g. the possibility to identify the need for, and analyse the effectiveness of, safety functions or any other barrier; and
- Quality: to assure clarity, unambiguity, consistency, etc., e.g. the possibility to review the models for completeness of the identified safety requirements.

For each requirement, SafeT provided an explanation to guide the application of the requirement on models to be used in the RAMS life cycle. An example is shown in Figure 2.

Requirement: The models must support the breakdown of a system into its constituent parts, in terms of system, sub-systems, and components.
Explanation: A system generally consists of a hierarchy of subsystems and components, each of which can be understood as a system itself. It is therefore meaningful to speak about the different levels of a system, and represent these levels in such a way that the details presented for each level are adequate for this level. Furthermore, it should be possible to study the details at any system level by recursively opening up the system model down to the subsystem or component of interest.

Figure 2. Example of a requirement and its explanation.

3.3 *The use of models in the RAMS life cycle*

The 58 requirements to models reflect needs identified from an analysis of the tasks to be performed in the different phases of the RAMS life cycle. The requirements can therefore relatively easily be interpreted in this context by describing how they apply to the modelling needs in the first ten RAMS phases. The different requirements were gradually introduced along with possible procedures and flow charts. Concerning modelling, the concept phase can be carried out in accordance with the following procedure:

1. Describe the needs and how these are met today without the system.
2. Make a first informal description of the system and its environment.
3. Make a first model of the system and its environment.
4. Define the aspects to be analysed, including the aspects defined in EN 50126.
5. Select an aspect for analysis.
6. Analyse the aspect, refining the model to make it adequate for the analysis.
7. If necessary, refine the model to make it represent the analysis result adequately.
8. Repeat from step 5 for the remaining aspects.

The RAMS life cycle is initiated with the concept phase. The main objective of the phase is to investigate the overall system and its environment, confined to (1) scope, context and purpose, as well as (2) physical, interface, legislative and economic issues. This means that there already is some idea of a "system under consideration", and some idea of the functionality that shall be offered, and most likely some constraints. The purpose of a model in this phase would therefore be to facilitate this investigation. Even if the system has not yet been defined in a proper sense, it will usually be possible to introduce the system as a black box, and concretize the aspects to be investigated. It might already in this phase even be possible to decompose this black box into a set of connected subsystems, each with its specific scope, context and purpose.

Requirements posed to models in this phase demand the ability of the models to support different needs, for example:

- support the breakdown of a system into its constituent parts, in terms of system, sub-systems, and components;
- facilitate the treatment of systems, sub-systems and components as black boxes, for which the details on architecture, design and implementation can be kept out of consideration, evaluating functions and hazards only at the boundaries;
- describe the system as contained in its operational environment;
- show how the environment can influence, or be influenced by, the system, including anything to which the system connects mechanically, electrically or by other means;
- show how man and organization can affect, or be affected by, the operation of the system;
- use clear and intelligible means of description, such as formal notation for logical functions, natural language for introductions, justifications and representations of intentions, graphical representations of examples, semantic definition of graphical elements, and directories of specialised words;
- be possible to communicate to the different stakeholders;
- be understandable in themselves;
- be understandable to the prospective user.

4 APPLYING SYSML

The Concept phase and System definition phase are focused on preparing the conceptual system model. The model acts as an input to the Risk analysis phase (see Figure 1). The first activity of the Risk analysis phase is the Hazard Identification (HI). This was the focus of a workshop in the SafeT project (see section 4.5) using the model-based description of an example case described in section 4.1.

Related to the use of models, two questions were investigated: (1) whether the modelling technique selected on the basis of theoretical considerations (the identified requirements to models based on the standards) is also practical for phase 1 and 2, and (2) whether the model-based description prepared is practical for the hazard identification activity.

4.1 *The securing work areas case*

In order to realistically evaluate existing techniques and develop the SafeT framework, the project chose a case example based on a concept of a new solution for securing work areas (Sivertsen, 2014). The problem concerns the need to protect maintenance workers from accidents caused by the interference with the railway traffic. The concept involves the development of a software-based system for securing the work areas from such interference. The basic requirements to such a system are to identify the workers' position correctly, effectively block the correct work area, and prevent a premature unblocking of this work area.

In the proposed solution (see Figure 3), a safety guard uses a smartphone both for the interaction with the train dispatcher and for identifying the works areas under consideration. The smartphone contains a dedicated application with functionality

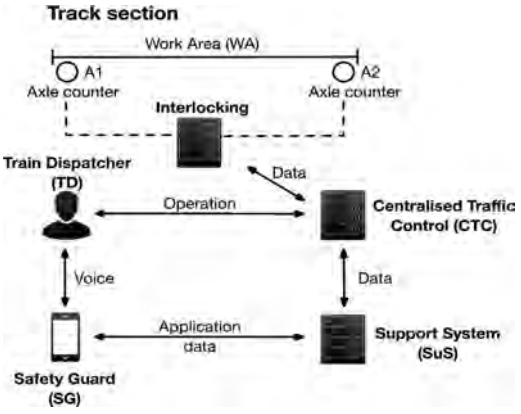


Figure 3. The securing work areas case.

to manage the securing and releasing of the work areas. Some of the characteristics of the functionality are:

- The main Safety Guard (SG) selects the functions from the application on his smart phone, e.g. secure a Work Area (WA).
- The scanning of the associated QR-code of a WA identifies both the SG and the WA.
- The application communicates with the Support System (SuS), which communicates with the Centralised Traffic Control (CTC) and other applications.
- The SuS supervises the associated protocols.
- The SuS supervises the secured WAs, and prevents the Train Dispatcher (TD) from prematurely unblocking them.

4.2 *SysML*

UML (Unified Modeling Language) was initially selected to be applied for modelling in phases 1 and 2 as it fulfils all of the related requirements to models in these phases. However, UML's focus is on supporting software analysis and design, while the system in our example case is not limited to software. Another important consideration was that the first RAMS phases are carried out at a higher system level ("the railway system level"), requiring a focus on the system as such and not merely on its software. Hence, we used SysML (Systems Modeling Language) instead which supports system engineering.

SysML is an extension of a frequently used subset of UML, and thus is expected to comply with most of the requirements to models that UML complies with. A SysML model is usually developed in a tool that stores the model entities with their characteristics and relations. The model entities can then be used in diagrams to present graphical views on specific aspects, e.g. structural or behavioural aspects.

Because of this unified model in the core of UML and SysML, they can be considered as a single but complex modelling technique. Furthermore, they offer different kinds of diagrams where each kind can be considered as a modelling technique in itself.

4.3 *Modelling the conceptual design*

Within the concept phase, modelling of the system and its environment with respect to the following aspects are required: (1) scope of the system, (2) (application) context of the system, (3) purpose of the system, and (4) environment of the system (anything that could influence, or be influenced by, the system, including people and procedures). All of these aspects are expected to be considered in the context of RAMS performance. The system definition phase requires extending the model with:

- functions and elements which need to be considered in the risk assessment;
- interfaces and interactions with the physical environment, other systems, humans, and other organisations;
- operational requirements influencing the system, including a description of conditions, constraints, logistics;
- existing safety measures and assumptions that determine the limits for the risk assessment.

The modelling was performed by an IT and dependability specialist with some experience in UML modelling, using a tool. A short textual description of the proposed system was the input to the modelling, and was analysed according to the needs of the two phases described above. The models were developed in an iterative process including consultation with the system owner.

Diagrams were prepared for a HAZOP workshop, e.g. Block Definition Diagrams (**BDD**) about Work Area and related concepts (see Figure 4), Internal Block Diagrams (**IBD**) about the internal communication and interfaces of the Support System (see Figure 5), Use Case diagrams (**UC**) of the main functions of the Securing Work Area application, State Machine diagrams (**STM**) about the registerable states of a Work Area (see Figure 6), and Sequence Diagrams (**SD**) about the main functions of the application.

4.4 *Using the models to meet the needs of the concept phase and the system definition phase*

The concept phase modelling needs can be mainly fulfilled by using BDDs and IBDs since those needs

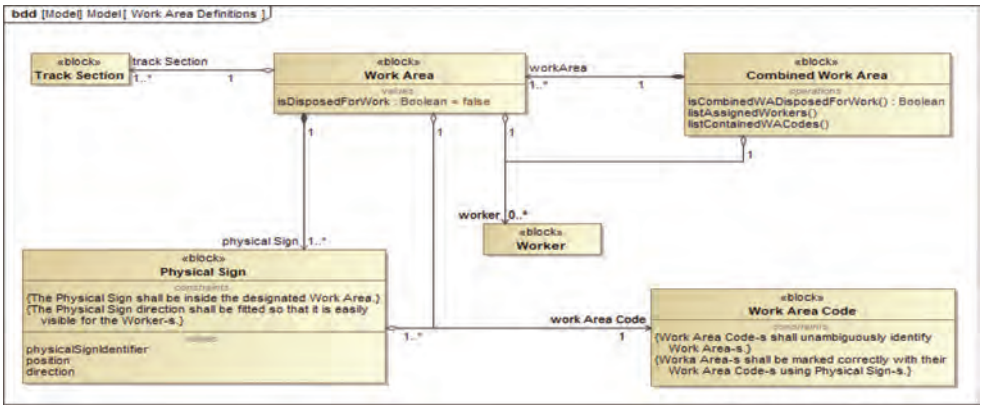


Figure 4. Example BDD defining the work area and related concepts.

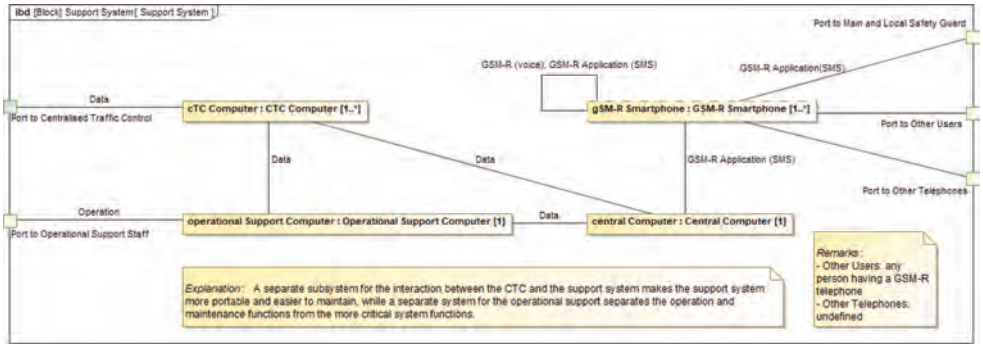


Figure 5. Example IBD of the internal structure of the support system.

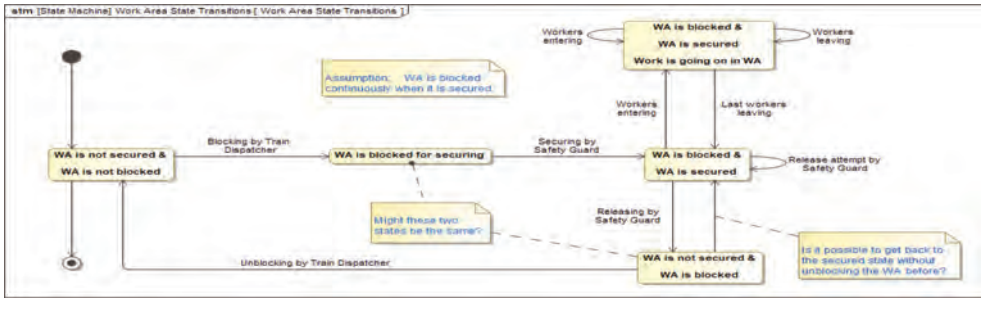


Figure 6. Example STM with the different states of a work area from the securing point of view.

require to represent the system with its elements and environment, and their static, conceptual relations. BDDs can depict an ontology. For example, the BDD in Figure 4 identifies the main concepts connected to work area in the proposed solution, their relevant characteristics, and their relations. IBDs can visualize internal structure, lines of communication and interfaces. For example, the IBD in Figure 5 depicts the internal structure of the Support System with it interfaces. The central computer communicates with the applications via its GSM-R receiver and transmitter; the operational support computer is used by the operational support staff to operate the system; the CTC computer ensures the correct interaction between the central computer and the CTC system. Another SysML diagram type not utilized by us is the Requirement Diagram (REQ) which could have

been useful for embedding the requirements in the model if a structured requirement specification had been available.

The system definition phase modelling needs can be partially fulfilled by using all 5 mentioned diagram types. BDDs and IBDs can be used when the system is further detailed (i.e. elements and interfaces in the system definition description). UC and STM diagrams as well as SDs are useful for depicting the dynamic, behavioural aspects (i.e. functions and interactions). For example, Figure 6 presents an STM with the different states of a Work Area from the securing and releasing functions point of view. This diagram shows for example that the states of the Work Area (as seen by the system) were unclear between "before securing"/"after releasing" and when it was "secured". Whether these states ("WA is blocked for securing" and "WA is not secured & WA is blocked") are the same, whether a transition from the second directly back to the state of "WA is blocked & WA is secured" is possible, triggered lots of discussions in the workshop.

Operational requirements were mostly not included in the model, but they can be added through REQ diagrams and by defining constraints. Existing safety measures were not specifically identified as such in the model, they were depicted as regular parts of the diagrams. However, there are suggestions in this direction, for example extending UC and SD for safety and security considerations (e.g. Misuse Cases, Failure Sequence Diagrams, Misuse Sequence Diagrams; an overview can be found in (Raspotnig, 2014)). Assumptions were either depicted as notes in the diagrams (e.g. see Figure 6), or as constraints. In summary, SysML has the potential to fulfil the needs of the concept and the system definition phases with respect to the requirements to models connected to these phases.

One experience was that the modelling process *helped identifying unclear and missing parts* of the case description which were necessary to develop an understanding for persons not familiar with the planned system. It is quite hard for a person involved in a task to evaluate what pieces of information are necessary for understanding the task by another person with different expertise working on another aspect of the task. The necessary amount of information is usually underestimated, which is also reflected by the related system descriptions. Modelling helps overcoming this gap but it does not guarantee the completeness of the information provided.

Another experience was that modelling with SysML sometimes demands more details than available or expected in the conceptual design phase. In other words, it *might be hard to draw the line between the conceptual design* (defining the "what") *and detailed design* (defining the "how"). For example, the conceptual design might stop at the level where the actors and systems of the New Solution are identified, maybe including the subsystems of the Support System. However, including the SWA App in the model required some further details since it resides in the software part of the Smartphone, which is a subsystem of the Support System. SafeT will need to specify clear criteria or guidelines regarding the detailing of models at the different phases of the development of a planned system.

4.5 *Using the models in a HAZOP workshop*

Two workshops utilizing HAZOP for hazard identification (HI) were organized, one using only a textual description as input and the other using a model-based description as input. The hazard identification related experiences of the workshops are presented in paper (Skogvang, 2018). Here, we focus on the modelling related experiences from the model-based workshop. A description utilizing the diagrams with limited text and explanation of the modelling language, was sent out one week before the workshop.

Even though modelling helped identifying unclear and missing parts from the modeller's perspective, it gave no guarantee that these identifications covered every necessary detail for HI. This became clear since the workshop participants had many questions outside the scope covered by the model but important nevertheless for their understanding of the context and for identifying hazards. A conclusion is that, for a better coverage of the hazard identification, relevant details in the model and the diagrams are desired. This could be achieved for example by a preparatory workshop focusing on eliciting such information, or by involving a RAMS expert in the modelling.

Constructs in models can become complex, and so their visualization. According to the experiences in the workshop, after a certain level of visual complexity (e.g. when not the whole diagram can be shown at once or if it is shown then it becomes unreadable), understanding of the diagram and following the track of thought becomes cumbersome. One related problem was following the flow of logic in SDs when branches and parallel activities were involved. Modularization might help with this issue.

During the workshop, an example of the physical outline was drawn ad hoc as an illustration which was used a lot in the discussions. This suggests that a physical outline diagram could be part of the model. SysML has no obvious means for this, therefore another modelling technique might be required as support. Another consideration is that modelling specific, representative cases (e.g. application of the planned system at a specific work area) might be a necessary supplement to the

general model of the planned system. In our case, a specific, representative train station could be considered. The model-based description also missed some information, e.g. preconditions of the main functions of the software application, necessary for understanding how the system was intended to work. A question related to this is whether the workshop would have been able to process and utilize the information requested by the participants (defined terminology and roles, description about the old and current solutions, etc.). This needs to be taken into account when considering the use of models with other techniques. SafeT needs to prepare guidelines on how to use HAZOP in combination with specific SysML diagrams.

5 CONCLUSIONS

In this paper we have elaborated on the experiences on using SysML diagrams as support for the concept and system definition phases. To the question of whether the modelling technique selected on the basis of theoretical considerations is also practical for the two first phases, we can answer affirmatively based on the experiences. The concept phase modelling needs can be fulfilled by using BDDs and IBDs since those needs require representing the system with its elements and environment, and their static, conceptual relations. BDDs can depict an ontology. The system definition phase modelling needs can be partially fulfilled by using all five mentioned diagram types.

However, further investigations and fitting guidelines will be necessary. Whether the model-based description prepared was practical for the hazard identification activities were not concluded, but the HAZOP workshop suggests that the use of SysML models requires good preparation of the HAZOP, and the participants should be familiar with such modelling to benefit from the models.

ACKNOWLEDGMENT

The SafeT project is funded by the Norwegian Research Council (project number 257167/O80) and Bane NOR, and has beside participation by Bane NOR and IFE, also participation from Indra Navia AS, Avinor, Solvina AB, Safetec Nordic AS, NTNU, VTT and Beijing Jiaotong University.

REFERENCES

AltaRica project, https://altarica.labri.fr/wp/ (Accessed Apr 10, 2017).

ASCOS project: https://www.ascos-project.eu/ (Accessed Apr 10, 2017).

Björnander, S., Land, R., Graydon, P., Lundqvist, K., Conmy. P. 2012. A Method to Formally Evaluate Safety Case: Arguments against a System Architecture Model. *IEEE Computer Society. WoSoCER2012.*

CENELEC, EN 50126–1:2017. Railway applications—The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS).

CENELEC, EN 50128:2011. Railway applications—Communication, signalling and processing systems—Software for railway control and protection systems.

CENELEC, EN 50129:2003. Railway applications—Communications, signalling and processing systems—Safety related electronic systems for signalling.

CHESS project: http://chess-project.ning.com (Accessed Apr 10, 2017).

EU, 2013. EU COMMISSION IMPLEMENTING REGULATION (EU) No 402/2013 of 30 April 2013 on the common safety method for risk evaluation and assessment and repealing Regulation (EC) No 352/2009.

Falessi, D. et al. 2011. Safeslice: a model slicing and design safety inspection tool for SysML. *In SIGSOFT FSE, pages 460–463.*

Gran, B. A. et al. 2007. Some challenges and solutions assessing the safety of ATM systems, *In Risk, Reliability and Societal Safety, ESREL 2007, Aven & Vinnem (eds), Taylor & Francis Group, pp 2113–2120.*

Gran, B.A. et al.2004. An Approach for Model-Based Risk Assessment. *In Proc. Computer Safety, Reliability, and Security (LNCS 3219). Heisel, M. Liggesmeyer, P., Wittmann, S. (Eds). Pp 311–324.*

Griffault, A. et al. 1998. The AltaRica Language. *In Lydersen and Hansen and Sandtorv ed., Proceedings of European Safety and Reliability Conference, ESREL'98.*

IEC 61508:2010. Functional safety of electrical/electronic/programmable electronic safety related systems.

Legendre, A. et al. 2017. Toward model synchronization between safety analysis and system architecture design in industrial contexts. *In LNCS 10437.*

Lund, M. S. et al. 2010. Model-Driven Risk Analysis. The CORAS Approach. *Springer.*

MARTE project, http://www.omgmarte.org/.

MODSafe project: http://www.modsafe.eu (Accessed Apr 10, 2017).

OPENCOSS project, http://next.opencoss-project.eu/node/2 (Accessed Apr 10, 2017).

Raspotnig, C. 2014. Requirements for safe and secure information systems. *PhD thesis.*

Raspotnig, C. et al. 2018. Coordinated Assessment of Software Safety and Security—An Industrial Evaluation of the CHASSIS Method. *To be published in Journal of Cases on Information Technology (JCIT) Vol. 20, Is.1.*

Roelen, A.L.C. et al. 2014. Risk models and accident scenarios in the total aviation system.

SafeCer project: http://www.safecer.eu/ (Accessed Apr 10, 2017).

Sivertsen, T. 2014. Concept of a New Solution for Securing Work Areas. *EHPG 2014, Røros, Norway, 2014.*

Skogvang, Ø. et al. 2018. Evaluating approaches for hazard identification for the inclusion in a Safety Assessment Framework for Efficient Transport. *To be presented at ESREL 2018.*

SysML, http://www.omgsysml.org/.

UML, http://www.uml.org/.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

A framework for modeling of multiple system failures—recoveries through multi-dimensional distributions in dynamic event trees

C. Picoco
The Ohio State University, Columbus, Ohio, USA
Électricité de France, EDF R&D, Palaiseau, France

V. Rychkov
Électricité de France, EDF R&D, Palaiseau, France

T. Aldemir
The Ohio State University, Columbus, Ohio, USA

ABSTRACT: Dynamic Event Tree (DET) methodology has been developed to overcome the limitations of the traditional Event Tree approach by taking timing of events explicitly into account through communicating with the system model that describes its dynamic behavior in event sequence construction. In addition, more rigorously accounting for process/hardware/software/human interactions, this capability allows including recoveries within the sequence analysis. Furthermore, particularly for long term scenarios, DET would be able to model multiple failures and recoveries for a given system with this capability. From probabilistic point of view, modeling multiple failures and recoveries introduces a major challenge since failure and recovery distributions for a given system can be correlated. Use of a multidimensional distribution is proposed to address this challenge.

1 INTRODUCTION

Dynamic Probabilistic Risk/Safety Assessment (DPRA/DPSA) methodologies are those methodologies that, by using simulator for the system under analysis, are able to explicitly model time dependent system evolution along with its stochastic behavior under accident conditions (Aldemir, 2013). Among these methodologies Dynamic Event Tree (DET) is perhaps the most popular one, given its similarity to the static event-tree (ET) approach, but with capability to model the interaction between stochastic events (e.g., failures, recoveries, etc.) and the dynamic evolution of the system as a consequence of these events.

A DET consists of an initiating event (e.g., station blackout, loss of offsite power, etc.), and a set of events that initiate system evolution in different directions, called branching conditions. Each branching point is defined by the analyst, and consists of a stochastic event described by (Alfonsi, 2013):

- values of process variables or thresholds, and,
- corresponding probability distributions (e.g., the exponential distribution of the failure time).

When, during the simulation, the thresholds corresponding to the branching points are met, branches are generated that give rise to the typical ET structure. Figure 1 shows a typical DET structure with thresholds defined for values 0.33 and 0.66 of the relevant cumulative distribution function (Cdf).

The scheme presented in Figure 1 refers to one of the codes currently used for generating DETs (i.e., RAVEN (Rabiti, 2016)).

Since the timing of the events in the sequence is explicitly modeled with a DET, failure recovery can be included within the analysis by assigning a probability distribution to the recovery time. The modeling of system failure and recovery becomes particularly important when long term scenarios

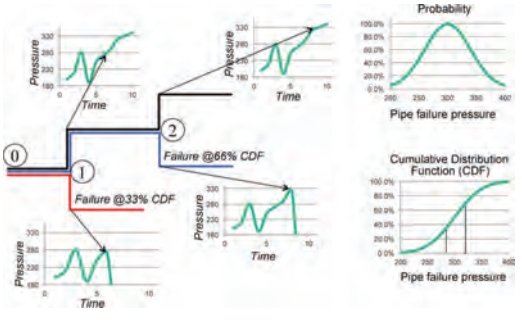


Figure 1. DET scheme (Alfonsi, 2014).

are considered. In these cases, multiple failure/recovery cycles for a given system become possible within the mission time.

The modeling of recoveries, however, introduces two main issues to consider. On one hand, as previously mentioned, failures and recoveries for a single system can be multiple along the accident sequence. On the other hand, recovery and failure time distributions can be interdependent since, from a causality point of view, a recovery can occur only after a failure.

Let us consider a system made up by different components, each with a characteristic failure and recovery distribution. The recovery time distribution for the overall system will be dependent on the failure time distribution for the system itself, through the failed components. Each event from the first recovery afterwards will depend on the previous one. Second recovery time will depend on second failure time, second failure time will depend on first recovery time and finally, first recovery time will depend on when the system has first failed. Therefore, in order to realistically represent the behavior in this example, we need to use a 4-dimensional distribution to be sampled according to the branching conditions.

The case of multiple failures and recoveries presents a unique challenge. In terms of physical states of the system, the multiple failures and recoveries can be represented as a two-way transition between two discrete states: an ON state and an OFF state (Picoco, 2017a). However, from a probabilistic point of view, each single transition (e.g., first failure, first recovery, second failure, etc.) could correspond to a different probability distribution.

In this work, we address these two aspects of DET generation: failure/recovery behavior (potentially multiple) and use of multi-dimensional distributions.

The paper is organized as follows. In Section 2, the framework for DET generation with multi-dimensional distributions is presented. In Section 3 some conclusions are drawn.

2 FRAMEWORK TO MODEL MULTIPLE FAILURE/RECOVERY BEHAVIOR IN DET

In order to generate a DET, a driver coupled with a simulator is needed. Different couples driver—simulator have been presented in literature such as ADAPT—MELCOR (Hakobyan, 2006), ADAPT—MAAP4 (Rychkov, 2015), RAVEN—RELAP7 (Alfonsi, 2013), MCDET-MELCOR (Hofer, 2002), ADAPT-SAS4-SASSYS-1 (Jankovsky, 2015), RAVEN-MAAP5 (Picoco, 2017b).

In the coupling, the driver is the code responsible for taking care of the probabilistic aspects. Generally, the driver requires for each branching condition the definition of a probability distribution. The branching conditions are expressed as a grid of thresholds, defined in either values of the process variables/system configuration or Cdfs. The driver is responsible for generating the different branches and managing their run. Driver also collects the results from the simulator, and, for some drivers, perform desired post-processing analysis, if any, set by the analyst.

The simulator provides the branch consequences and simulates the plant evolution as the different branchings occur. In nuclear field, the simulator (e.g., MELCOR (Summers, 1981), MAAP5 (MAAP5, 2015), RELAP7 (Anders, 2012)) is able to predict the behavior of the reactor under several accident conditions as the different branches occur. The control logic of the simulator is often used to model operator actions, and possible interactions among the different branches.

The driver and the simulator usually exchange information during the DET generation in order to create the different branches, based on the plant state.

Overall, the DET generation process is the following (the process can slightly vary depending on the driver-simulator couple used):

1. Simulation starts: the first branch is run.
2. The simulation stops when a branching point is met.
3. Two (or more) branches are created by the driver, each corresponding to a different new simulator instance, and run in parallel.
4. The simulation of each branches progresses until the next branching point is met.
5. When all the branches simulated are completed, the DET is generated.

In order to define each branch, as previously mentioned in Section 1, a probability distribution is needed.

In case of multiple failures and recoveries, a first approach would suggest to treat these conditions independently, defining a different distribution (and corresponding branching points) for each event. However, this type of modeling will not account for their intrinsic correlated behavior.

Since most of the drivers, so far, require definition of branching points in terms of 1-dimensional distributions and 1-dimensional grid for each variable as a first approach to face the use of N-dimensional distribution, we propose the following framework:

1. *From the N – dimensional distribution, define the corresponding N 1-dimensional marginal distributions*. It is worth recalling that, starting from a joint *N*-dimensional distribution, the marginal of a variable describes the corresponding probability distribution of that variable only. Formally, given the N – dimensional probability distribution function $f(x_1, x_2, \ldots, x_n)$, then the marginal for the generic x_i is

$$f_{X_i}(x_i) = \int \ldots \int f(x_1, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n \quad (1)$$

   The general concept of multidimensional distribution and corresponding marginal is shown in Figure 2, referred to the case of a bivariate normal distribution. In Figure 2, the two marginal distributions for x_1 and x_2 are $f(x_1)$ and $f(x_2)$, respectively.
2. Once 1-dimensional distributions for each variable are obtained, it is possible to use them with the driver and define for each variable the grid of the branching points independently, as is current practice with DET generation.
3. Run the simulation and generate the DET.
4. Once the DET has been generated, post-process the results to recalculate the probabilities of each history based on the values of the multidimensional distribution rather than on the marginal.

This approach allows to account for correlated probabilities without necessity to re-run the DET, and by defining the branching conditions as is the case for 1-dimensional distributions.

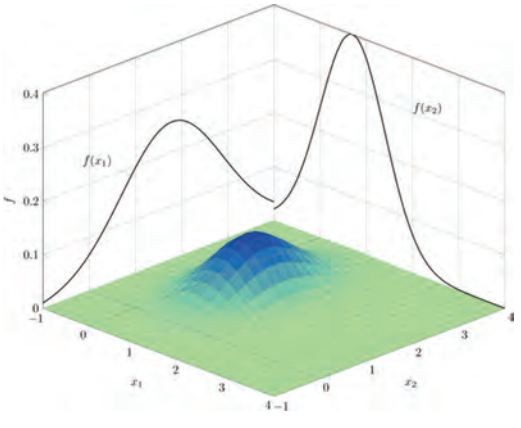


Figure 2. Two-dimensions distribution and corresponding marginal for a bivariate normal distribution.

3 CONCLUSIONS

DETs are currently used to analyze the possible evolution of an accident scenario starting from a given initiating event. In general, as in traditional ET, only failures of the different systems involved in the accident evolution have been considered. By explicitly modeling time, DET has the capability to include recoveries within the model. Including recoveries, and even multiple failures and recoveries for the same system, becomes particularly important when long mission times are considered.

In this paper, we have presented a framework for dealing with the DET generation in case of multiple failures and recoveries for a given system modeled by multidimensional distribution. In summary, the approach proposed in this work consists in the following: a) variables are sampled starting from their marginal distributions, and, b) the DET is generated using the marginal distributions. Then history probabilities are recalculated based on the values of the *N*-dimensional distributions.

This framework represents a first approach for the use of multidimensional distribution in DET. The approach is valid in regular cases (e.g., multidimensional uniform, multivariate normal), however, the sampling from the marginal can potentially lead, in some cases, to weak density of sampling points in probabilistically relevant regions of the multi-dimensional distribution.

REFERENCES

Aldemir, T. 2013. A survey of Dynamic Methodologies for Probabilistic Safety Assessment of Nuclear Power Plants. *Annals of Nuclear Engineering* 52: 113–124.

Alfonsi, A. et al. 2013. Dynamic Event Tree Analysis through RAVEN. ANS PSA 2013 International Topical Meeting on Probabilistic Safety Assessment and Analysis, Columbia.

Alfonsi, A. et al. 2014. Adaptive dynamic event tree in RAVEN code. Transactions of the American Nuclear Society, 111, pp. 924–926.

Anders, D. et al. 2012. RELAP-7 Level 2 Milestone Report: Demonstration of a Steady State Single Phase PWR Simulation with RELAP–7. INL/EXT-12–25924.

Hakobyan, A. et al. 2006. A methodology for generating dynamic accident progression event trees for level-2 PRA. PHYSOR-2006 - American Nuclear Society's Topical Meeting on Reactor Physics, 9 p.

Hofer, E. et al. 2002. An approximate epistemic uncertainty analysis approach in the presence of epistemic and aleatory uncertainties. *Reliability Engineering and System Safety* 77, 229–238.

Jankovsky, Z.K. et al. October 2015. Interim Status Report for Risk Management for SFRs, SAND2015–8872. Sandia National Laboratories, Albuquerque, NM.

Modular Accident Analysis Program 5 (MAAP5) Applications Guidance: Desktop Reference for Using MAAP5 Software—Phase 2 Report. EPRI, Palo Alto, CA: 2015. 3002005285.

Picoco, C. et al. 2017. Dynamic Event Tree generation with RAVEN—MAAP5 using Finite State Machine system model, American Nuclear Society Probabilistic Safety Assessment 2017 Topical Meeting, Pittsburgh, September 2017.

Picoco, C. et al. 2017. Coupling of RAVEN and MAAP5 for the Dynamic Event Tree analysis of Nuclear Power Plants, Proceedings of the European Safety and Reliability Conference, Portoroz, June 2017.

Rabiti, C. et al. 2016. RAVEN User Manual, INL/EXT-15–34123, Idaho National Laboratory, 2016.

Rychkov, V. & Kawahara, K. 2015. ADAPT-MAAP4 coupling for a dynamic event tree study. International Topical Meeting on Probabilistic Safety Assessment and Analysis, PSA 2015, 1, pp. 140–143.

Summers, R.M. et al. 1981. MELCOR 1.8.0: A Computer Code for Nuclear Reactor Severe Accident Source Term and Risk Assessment Analyses.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Using an enterprise architecture model for assessing the resilience of critical infrastructure

Gonçalo Cadete & Miguel Mira da Silva
Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

ABSTRACT: Assessing the resilience of Critical Infrastructure (CI) is a complex problem. Complexity becomes especially high when hybrid threats are considered and, additionally, crisis span interdependent sectors and sovereign borders, with time-variable cascading effects. Models may help address complexity, since they are simplified representations of systems, intended to promote understanding within some domain of discourse. Furthermore, Enterprise Architecture (EA) allows for managing and visualizing integrated model repositories. In this paper, we propose an EA model to assist resilience assessments, performed using a Resilience Assessment Framework (RAF). To ensure that the framework was fit for purpose according to the evaluators' needs, a new version of an existing RAF was designed and tested, using a Design Science Research Methodology (DSRM) process model. To measure the value of the EA model, we performed a comparative evaluation of the new RAF's usefulness—i.e. with and without assistance of the EA model. We conclude that the proposed EA model is useful for assisting resilience assessment initiatives for CI. The main scientific contributions of this paper are the validation of the EA model's usefulness, a set of EA viewpoints for assisting resilience assessments of CI, as well as an evaluation of the new version of the RAF.

1 INTRODUCTION

Assessing the resilience of Critical Infrastructure (CI) presents both conceptual and implementation challenges. At the conceptual level, frameworks and standards are required to align terminology (ISO 2009, 2012), provide reference models (ISACA 2013a), enable consistent audit and assurance programmes (NIST 2014, ISACA 2013b), as well as to facilitate communication, cooperation, and collaboration. At the implementation level, adequate methods and tools are required to enable effective and efficient assessment initiatives.

Models are important to address the complexity of resilience assessments, since they are simplified representations of the system of interest. Also, as we demonstrate in this paper, models may help clarify conceptual and methodological ambiguities. Furthermore, Enterprise Architecture (EA) allows for managing and visualizing integrated model repositories (Lankhorst 2013), thus enhancing the performance of assessment initiatives – when compared with the exclusive use of spreadsheet-like artifacts, informal diagrams, and natural-language descriptions.

In this paper, we propose an EA model to assist the implementation of resilience assessment initiatives. The EA model's usefulness was evaluated using a demonstration, evaluation questionnaires, and group sessions, to measure the efficacy, generality, consistency, simplicity, and clarity of the artifacts. For the demonstration and evaluation, we used a new version of an existing Resilience Assessment Framework (RAF) – from Cadete et al (2017).

We conclude that the proposed EA model is useful for assisting resilience assessment initiatives, by helping to achieve three framework objectives: provide a logical link between management and operational indicators for resilience, clarify conceptual and methodological ambiguities, and facilitate the implementation of resilience assessment initiatives.

2 METHODOLOGY

In this paper, we performed two iterations of a Design Science Research Methodology (DSRM) process model (Peffers et al 2014), for guiding the construction and evaluation of the architectural artifacts.

DSRM incorporates principles, practices, and process models which are adequate to conduct design science research in applied research disciplines, whose cultures value incrementally effective solutions (Hevner & Chatterjee 2010). The design science paradigm seeks to create and evaluate

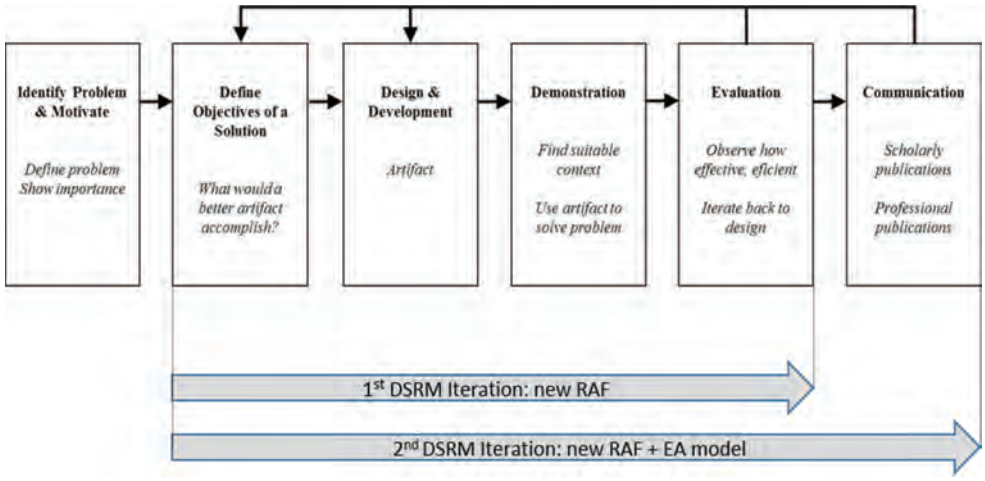


Figure 1. DSRM process model used in this paper. Two DSRM iterations were performed: in the first DSRM iteration, a new conceptual model for the RAF was designed and tested; in the second DSRM iteration, an EA model for the new RAF was designed and tested.

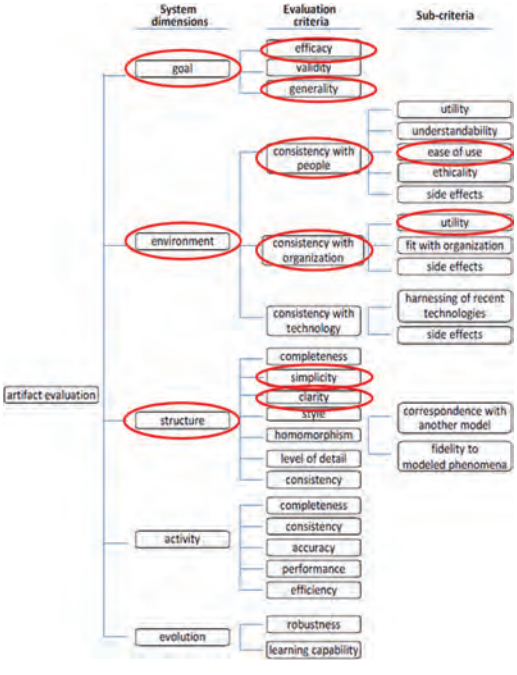


Figure 2. Evaluation criteria, taken from Prat et al (2014). The evaluation criteria used for evaluating the new RAF are highlighted with ellipses.

"what is effective" in the problem space (Hevner et al. 2004).

The first DSRM iteration (Fig. 1) was performed to ensure that the framework was fit for purpose according to the evaluator's needs. Had

Table 1. Evaluation objectives for the new RAF.

System dimension	Evaluation criteria/ sub-criteria	Objectives for the new RAF
Goal	Efficacy	Integration of disaster risk management aspects
		Risk is associated to effect/ impact on relevant objectives
		Indicators are relevant for risk management
		Management indicators relate to operational indicators
	Generality	May be tailored for any CI organization
		Is not overly prescriptive
		Cross-sector generality
		Cross-border generality
Environment	Consistency with organization/ utility	Useful for CI organizations
	Consistency with people/ ease of use	Easy to implement for CI managers
Structure	Simplicity	Is simple to communicate and understand
	Clarity	The concepts and methods are clear and unambiguous

this iteration been omitted, the evaluators' ratings might have been negatively affected by perceived deficiencies in the reference framework design (Cadete et al. 2017). In the second DSRM iteration, an EA model was designed, for modeling the new RAF that resulted from the first DSRM iteration. This strategy allows for comparing the two sets of DSRM results, ensuring that the evaluation rating differences (i.e. between the first and second DSRM iterations) are a reasonably good measure of the benefits of using the EA model.

As shown in Fig. 1, the DSRM process model includes an evaluation activity. For both DSRM iterations, we adopted the same evaluation criteria (see Fig. 2), as well as the same evaluation objectives (see Table 1), that were used for evaluating the RAF from Cadete et al (2017) – based on the evaluation taxonomy from Prat et al (2014).

3 RESEARCH PROBLEM

In Cadete et al (2017), the RAF evaluation results showed relatively weak ratings, regarding the achievement of three objectives:

- Consistency with people, ease of use: easy to implement for CI managers;
- Structural clarity: the concepts and methods are clear and unambiguous;
- Generality: may be tailored for any CI organization.

In related work, EA artifacts were used successfully to assist business continuity planning (Gomes et al. 2017) as well as to represent process-based frameworks (Vicente et al. 2013). Also, EA is recommended as best practice for guiding the creation and maintenance of the governance and management enablers for information systems and related technologies (ISACA 2012a). Finally, it is important to note that when holistic frameworks are represented, complex graph-like structures of entities and relationships emerge due to the variety of concerns (e.g. many sectors, many countries, and many areas of expertise), as well as due to the complex networks of dependencies. Such complex graph-like structures may be represented and managed using EA methods and tools.

These facts lead to the hypothesis that EA techniques and artifacts might help address some of the implementation issues identified previously by the RAF evaluators in Cadete et al (2017).

The research problem is therefore to find and validate an EA solution to address the RAF conceptualization and implementation shortcomings, namely regarding goal efficacy, environmental consistency, and structural simplicity and clarity.

4 DESIGN AND DEVELOPMENT

To help improve the design of the new RAF, we selected senior evaluators from the defense sector,

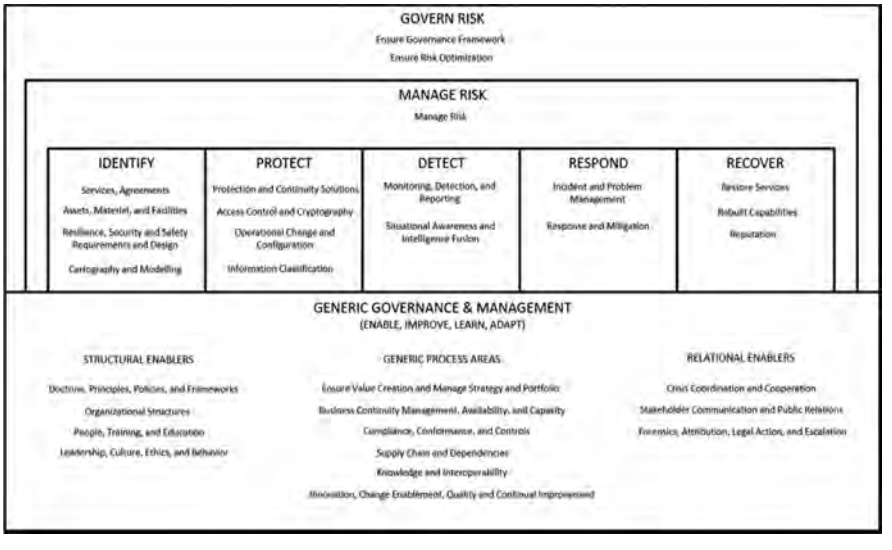


Figure 3. The new RAF Process Reference Model (PRM) design, resulting from the first DSRM iteration. This artifact was used in both DSRM evaluations (first and second). This new design ensured that the RAF framework was sufficiently fit for purpose, according to the evaluator's needs –thereby minimizing negative bias in the evaluator's ratings, regarding the EA model's usefulness.

with cyber-physical expertise. This selection differs from the evaluators selected for evaluating the reference RAF, who were experts from civilian sectors (ICT, water and water waste, and financial sectors).

To improve the fit for purpose of the RAF artifacts, according to the evaluator's needs, the first DSRM design and development activity was dedicated to producing a new RAF design, shown in Fig. 3. The differences between the reference RAF design and the new RAF design concern the Process Reference Model (PRM), and are the following (see Fig. 3):

- 6 new process areas were added to the PRM:
  - Doctrine, Principles, Policies, and Frameworks;
  - Organizational Structures;
  - People, Training, and Education;
  - Leadership, Culture, Ethics, and Behavior;
  - Crisis Coordination and Cooperation;
  - Stakeholder Communication and Public Relations.
- The generic governance and management areas were sub-divided according to:
  - Generic process areas;
  - Structural enablers;
  - Relational enablers.

Essentially, these additions relate to standard defense planning capabilities (DOTMLPF-I, the acronym standing for Doctrine, Organization, Training, Materiel, Leadership, Personnel, Facili-

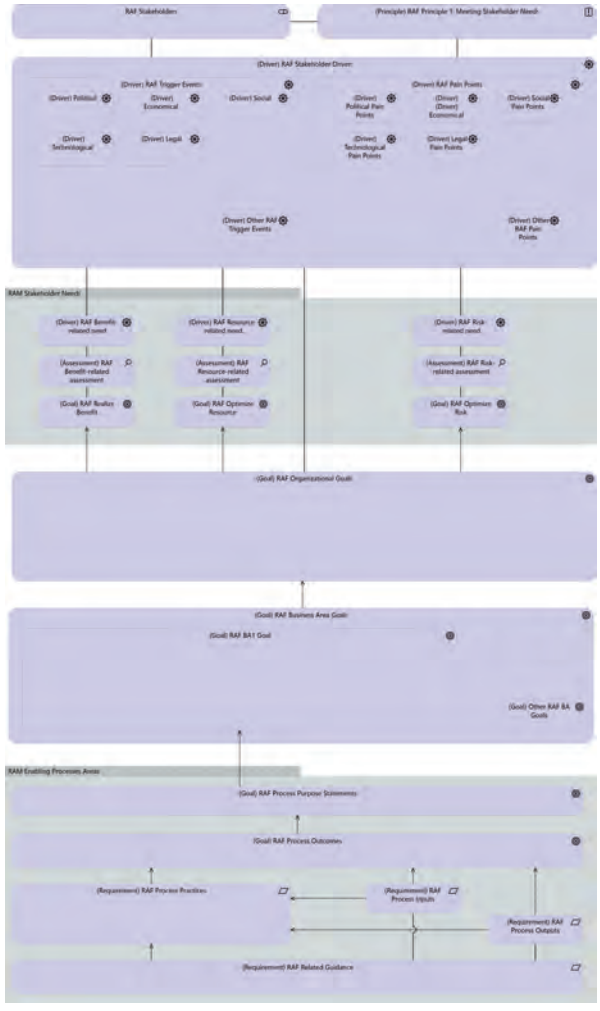


Figure 4. The RAF Goals Cascade and PAM model (modeled using ArchiMate). The organizational drivers, needs, and goals cascade to a representation of the PAM elements (process purpose, outcomes, practices, inputs, outputs, and related guidance).

ties, and Interoperability), as well as to COBIT5 enablers (ISACA 2012b).

The Process Assessment Model (PAM), the Process Measurement Model (PMM), and the Goals Cascade Methodology (GCM) remained unchanged in the new RAF design.

For the second DSRM iteration, an EA model was created, to model the new RAF framework that resulted from the first DSRM iteration. The ArchiMate (The Open Group 2016) modeling language was used for the EA representations. An open source EA tool (Archi 2017) was used for providing an integrated EA repository, as well as the ArchiMate views. The viewpoint representing the goals cascade and the PAM is shown in Figure 4. Note that this viewpoint clearly shows the relation between the goal cascade elements (stakeholder drivers and needs, organizational goals, business area goals, and enabling process goals), as well as the relations between all their subcomponents. Also, note that the graphical arrangement intuitively conveys the notion of top-down and bottom-up alignment – important for relating management and operational concerns.

5 DEMONSTRATION

For demonstration purposes, we used Fig. 4 (goals cascade and PAM) as well as ArchiMate artifacts taken from Gomes et al (2017) (Figs. 5, 6). These artifacts provide an EA model that instantiates the RAF for the following resilience assessment scenario:

- RAF goals cascade Business Area (see Fig. 4):
  - COBIT5, information and related technologies.
- RAF PRM process area:
  - "Business Continuity Management, Availability, and Capacity" (see Fig. 3).
  - Corresponding COBIT5 process: DSS04 Manage Continuity (see Fig. 5).

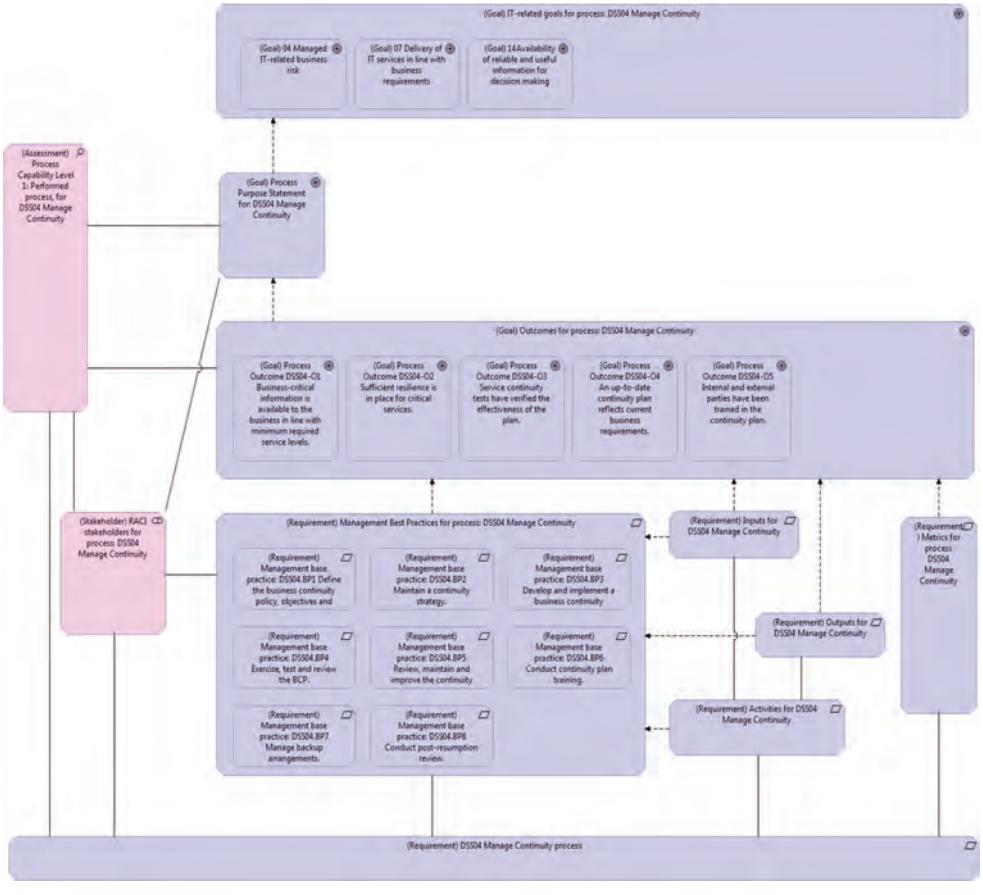


Figure 5. Viewpoint for the COBIT5 manage continuity process, for process capability level 1 assessments. Taken from Gomes et al (2017).

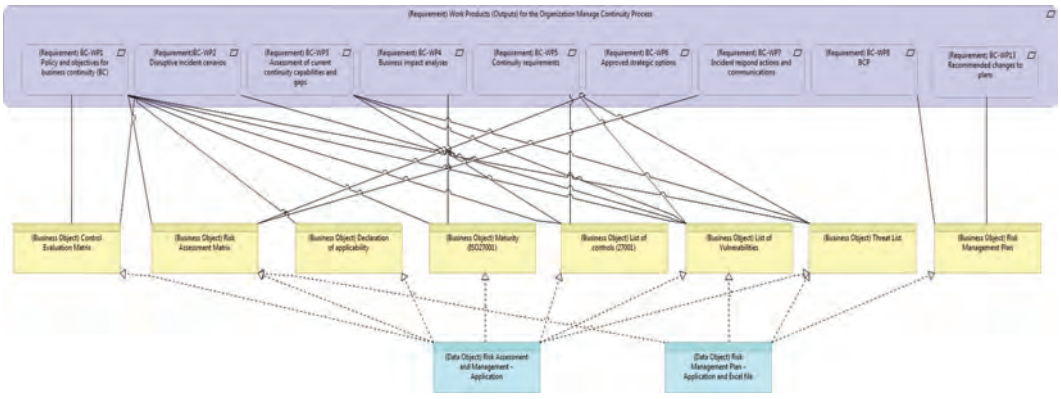


Figure 6. Resilience assessment evidence (middle-row and lower-row elements) mapped to outputs (upper-row elements) of the Manage Continuity COBIT5 process. Taken from Gomes et al (2017).

– RAF PAM process capability level:
  – Process capability level 1 (see Fig. 5).

In other words, in this scenario we are assessing process performance (i.e. achieving process goals) for the process area "Business Continuity Management, Availability, and Capacity", focusing on information and related technologies' concerns. For adequate coverage of such informational concerns, we have used the state-of-the-art COBIT5 governance and management framework.

Real-world formal assessments must be justified and documented using evidence. We demonstrated the instantiation of resilience assessment evidence using the ArchiMate view in Fig. 6, that represents the mapping between the evidence (middle-row and lower-row elements) and the outputs of the Manage Continuity COBIT5 process (upper row elements). This demonstration artifact was taken from Gomes et al (2017).

6 EVALUATION

The evaluation results are shown in Table 2. For rating the achievement of RAF objectives, we used a standard ordinal scale that is used for assessing outcomes (ISO 2015):

– "FA" = fully achieved
– "LA" = largely achieved
– "PA" = partially achieved
– "NA" = not achieved

In the right column "*Gain*" we show the added value of using the EA model (i.e. difference between the second DSRM and first DSRM ratings). Each "+" sign accounts for one rating improvement (e.g. from PA to LA, or from LA to FA). A change from PA to FA is thus represented with two "+" signs. Where no rating changes occurred, an "=" sign was used.

In the "Research Problem" section, we reported three issues that were found in the reference RAF:

– Consistency with people, ease of use: easy to implement for CI managers;
– Structural clarity: the concepts and methods are clear and unambiguous;
– Generality: may be tailored for any CI organization.

From the results presented in Table 2, we can observe only a minor improvement in achieving the objective "*may be tailored for any CI organization*". However, higher gains were obtained for the objectives "*management indicators relate to operational*", "*easy to implement for CI managers*", and "*concepts/methods: clear and unambiguous*".

Using an EA model (introduced in the second DSRM iteration), we have thus obtained significant gains in the system dimensions:

– Goal efficacy: management indicators relate to operational;
– Environmental consistency: easy to implement for CI managers;
– Structural clarity: the concepts and methods are clear and unambiguous.

For all the remaining objectives, no achievement degradation was observed.

These results are consistent with the hypothesis that EA models are useful to assist resilience assessment initiatives, since the only difference between the first and second DSRM iterations is, precisely, the use of the proposed EA model.

Table 2. Evaluation ratings for the two DSRM iterations. The column "*Gain*" shows the benefits of using the EA model: each "+" accounts for one rating improvement, e.g. from PA to LA, or from LA to FA. A change from PA to FA is thus accounted for using two "+" signs.

Objectives for the RAF	1st DSRM	2nd DSRM	Gain
Disaster risk management aspects	FA, FA, FA, FA	FA, FA, FA, FA	=
Risk is associated to effect on objectives	FA, FA, FA, FA	FA, FA, FA, FA	=
Indicators are relevant for risk management	FA, FA, FA, FA	FA, FA, FA, FA	=
Management indicators relate to operational	FA, LA, LA, PA	FA, FA, FA, FA	++++
May be tailored for any ci organization	FA, LA, LA, LA	FA, LA, LA, FA	+
Not overly prescriptive	FA, FA, FA, FA	FA, FA, FA, FA	=
Cross-sector generality	FA, LA, LA, FA	FA, LA, LA, FA	=
Cross-border generality	FA, LA, LA, FA	FA, LA, LA, FA	=
Useful for ci organizations	FA, FA, FA, FA	FA, FA, FA, FA	=
Easy to implement for ci managers	LA, LA, PA, PA	FA, FA, LA, LA	++++
Simple to communicate and understand	FA, FA, FA, FA	FA, FA, FA, FA	=
Concepts/methods: clear and unambiguous	LA, LA, LA, LA	FA, LA, FA, FA	+++

Interestingly, the efficacy objective "*management indicators relate to operational*" received maximum ratings in the second DSRM iteration (i.e. with EA model), a significant upgrade from the first DSRM ratings (i.e. without EA model). These results reflect the expressiveness benefits of the EA model, that provided an integrated representation including all levels the assessment rationale (from high-level organizational drivers and needs, down to low-level assessment evidence) in a graph-like conceptual structure.

Note that these EA representations may be stored in an integrated EA repository, which means that they can be reused in several assessment initiatives, as well as integrated in the larger EA landscape of the CI organization.

However, during the group sessions, the evaluators commented that additional ontological artifacts – such more thorough formal definitions for entities and relationships, as well as ontological mappings – are needed to further clarify the framework's semantics, as well as to ease real-world implementation initiatives in critical infrastructure operators.

7 CONCLUSION

Assessing the resilience of Critical Infrastructure (CI) is a complex conceptual and implementation challenge. Modeling artifacts such as languages, methods, and tools, are instrumental to address such complexity. Furthermore, EA models, methods, and tools allow for managing and visualizing integrated model repositories, thus providing a powerful complement to representations based on spreadsheet-like artifacts, informal diagrams, and natural-language descriptions.

It is important to note that this work does not prove that the new RAF is an improved version of the reference RAF from Cadete et al (2017). Also, no claims are made in relation to the relative benefits of the proposed ArchiMate artifacts used in the demonstration, vis-à-vis other EA modeling languages. Future work may address optimization of the RAF and related EA models, to assist actual assessment initiatives.

However, the evaluation results are consistent with the hypothesis that EA models are useful to assist resilience assessment initiatives. Such results are also consistent with the informal feedback elicited during the group sessions.

Regarding limitations, the evaluators commented that additional ontological artifacts are needed, namely for achieving higher ratings for the generality goals (such as achieving cross-sector, cross-border, and tailoring for any CI organization), as well as to improve the framework's conceptual clarity and ease of implementation.

Also, EA models may not be as useful for frameworks that are based on simple checklists or matrices of indicators. For these cases, spreadsheet-like artifacts and natural-language descriptions may be sufficient to assist the assessment initiatives. Note, however, that such simple artifacts may not provide the optimal solution for assisting holistic frameworks that comprise several points of view (i.e. many related concerns, many sectors, many countries, and many areas and levels of expertise) and account for complex networks of dependencies and interdependencies. In these cases, a complex graph-like structure emerges and may be successfully be addressed with adequate EA models, methods, and tools.

A secondary contribution of this paper is the new version of the resilience assessment frame-

work, as well as its set of evaluation ratings. This new version and evaluation ratings may be used to inform design, development, and testing for future DSRM iterations.

The main contributions of this paper are the validation of the EA model's usefulness for assisting resilience assessment initiatives, as well as a set of EA viewpoints that may be reused, improved, or adapted for actual resilience assessment initiatives.

REFERENCES

Archi 2017. *Archi – The Free ArchiMate Modelling Tool, version 4*. Available at: http://www.archimatetool.com/.

Cadete, G. et al 2017. A Conceptual Framework for Assessing the Resilience of Critical Infrastructure. *Safety and Reliability – Theory and Applications*, ISBN 978-1-138-62937-0, Taylor & Francis Group, London.

Gomes, P. et al. 2017. Using Enterprise Architecture to Assist Business Continuity Planning in Large Public Organizations. *CBI 2017 – 19th IEEE International Conference on Business Informatics*.

Hevner, A. & Chatterjee, S. 2010. *Design Research in Information Systems*. Springer.

Hevner, A. et al. 2004. Design Science in Information Systems Research. *MIS Quarterly*, vol. Vol. 28 No.1.

ISACA 2012a. *COBIT 5 Implementation*, Rolling Meadows, IL, USA, ISACA.

ISACA 2012b. COBIT 5: *A Business Framework for the Governance and Management of Enterprise IT*, Rolling Meadows, IL, USA, ISACA.

ISACA 2013a. COBIT *Process Assessment Model (PAM): Using COBIT 5*, Rolling Meadows, IL, USA, ISACA.

ISACA 2013b. *COBIT 5 for Assurance*, Rolling Meadows, IL, USA, ISACA.

ISO 2009. *ISO 31000:2009 – Risk management—Principles and guidelines*.

ISO 2011. *ISO/IEC/IEEE 42010:2011 Systems and software engineering—Architecture description*.

ISO 2012. *ISO 22300:2012 Societal security—Terminology*.

ISO 2015. *ISO/IEC 33020 – Information Technology—Process assessment—Process measurement framework for assessment of process capability*.

Lankhorst, M. 2013. *Enterprise Architecture at Work: Modelling, Communication and Analysis, third ed*. Springer-Verlag.

NIST 2014. *Framework for Improving Critical Infrastructure Cybersecurity – Version 1.0*. NIST – National Institute of Standards and Technology.

Peffers, K. et al 2007. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, vol. Vol. 24 No.3.

Prat, N. et al. 2014. Artifact Evaluation in Information Systems Design-Science Research – a Holistic View. *PACIS 2014 Proceedings*, Paper 23.

The Open Group 2016. *The ArchiMate 3.0 Specification*.

Vicente, M. et al. 2013. Using ArchiMate to Represent ITIL Metamodel. *Proceedings of the 2013 IEEE 15th Conference on Business Informatics (CBI)*, Washington, DC, USA, IEEE Computer Society, pp. 270–275.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Application of systems-theoretic process analysis to a subsea gas compression system

H. Kim & M.A. Lundteigen
Norwegian University of Science and Technology, Trondheim, Norway

A. Hafver & F.B. Pedersen
DNV-GL, Oslo, Norway

G. Skofteland
Statoil, Trondheim, Norway

C. Holden & S.J. Ohrem
Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: The life and recovery factor of already existing subsea gas fields and infrastructure may be increased by installing boosting facilities to compensate for declining well pressures. The installation of such boosting facilities subsea has often been identified as more cost-efficient than installation topside. A recent example is the Åsgard Subsea Gas Compressor installed and started up in 2016 on the Norwegian Continental Shelf. The compressor system is highly complex, involving, beyond the compressor itself, numerous pipes, valves, sensors, a liquid removal facility and a liquid pump. The design of control and safety systems is based on requirements in regulations and key standards, many of which build on topside philosophies for process safety and protection. An ongoing research in the Centre on Subsea Production and Processing (SUBPRO) is to investigate if new requirement formulation methods can verify if the current subsea safety and control philosophy is adequate. A motivation is to investigate areas of improvement for future subsea installations or similar systems. One such method is the Systems-Theoretic Process Analysis (STPA), a method that has been developed specifically for hazard identification in system control architectures. The main advantage of STPA over other hazards identification techniques is its ability to capture system failures that may arise from the communication between equipment in the control architecture, and this insight can be used to build more robust and reliable systems. STPA has already been adopted in many different sectors and domains, but has not yet been tested for subsea processing systems. The main objectives of this paper are: (1) to apply STPA to a subsea processing system; a subsea compression system; (2) to discuss opportunities and challenges of applying STPA to subsea compression systems, and; (3) to extend the discussion to the general use of STPA and necessity to improve the method.

1 INTRODUCTION

1.1 *Background*

The life and recovery factor of a subsea gas reservoir depends on the reservoir pressure and pressure loss in the production system. The reservoir pressure is typically higher than the pressure loss in the first period of gas production, so that the production rate can be maintained (Monsen et al., 2012). However, at some point during the field life, installation of a boosting facility may be needed to compensate for declining well pressure and extend the plateau production (Baggerud et al., 2007). In other cases, long distance transport increases pressure loss, and consequently, the gas production can be sustained at lower pressure.

One traditional and proven solution in these cases is topside gas compression, but subsea gas compression has often been identified as more cost-efficient than topside gas compression. Subsea gas compression can sustain higher production rates with lower power consumption, because the compressor is closer to the well (Lima et al., 2011). In addition, unmanned operation of subsea gas compression reduces operation costs (Lima et al., 2011). On the other hand, the application of subsea gas compression has been technically challenging due to large electrical power consumption, the

need for fast acting control and use of complex equipment (Baggerud et al., 2007).

To prevent hazardous events of subsea gas compression systems, the control and safety systems of subsea gas compression are designed in accordance with regulations and key standards. However, many of the regulations and key standards of subsea gas compression systems build on topside philosophies for process safety and protection (Kim et al., 2016). Research is currently underway in the Centre on Subsea Production and Processing (SUBPRO) to investigate whether new requirement formulation methods can help verify adequateness of subsea safety and control philosophies (SUBPRO, 2017). One such method is the Systems-Theoretic Process Analysis (STPA), a hazard identification method that was recently developed based on the Systems-Theoretic Accident Model and Processes (STAMP).

STPA has widely been adopted and used in cyber security (Young and Leveson, 2013, Salim, 2014, Young, 2014, Schmittner et al., 2016), aerospace (Ishimatsu et al., 2010, Nakao et al., 2011, Leveson, 2014), aviation (Leveson et al., 2014, Chen et al., 2015, Allison et al., 2017), medical device (Antoine, 2013, Samost, 2014, Proctor et al., 2015, Zhang et al., 2017) and so on. However, there are limited number of studies on STPA application in the oil and gas industry, and only one conference paper investigated the application of STPA to subsea systems (Rachman and Ratnayake, 2015). To the best of our knowledge, no study has conducted STPA on subsea processing systems and discussed opportunities and challenges of applying STPA to subsea processing systems.

1.2 *Objectives*

The main objective of this paper is to apply STPA to a subsea gas compression system and discuss associated opportunities and challenges. This main objective is further developed into three sub-objectives:

- To conduct STPA analysis on a general subsea gas compression system and summarize the results
- To discuss opportunities and challenges applying STPA to subsea processing systems, based on the results of the analysis
- To extend the discussion to the general use of STPA and necessity to improve the method

1.3 *Structure of the paper*

The remainder of this paper is organized as follows: subsea gas compression system is introduced in Section 2, and STPA is applied to a typical subsea dry gas compression system in Section 3. Summary of the analysis results and discussions follow in Section 4.

2 SUBSEA GAS COMPRESSION SYSTEM

2.1 *Subsea processing*

Any handling and treatment of the produced hydrocarbon fluids prior to reaching the platform or onshore can be defined as subsea processing, e.g. subsea boosting, subsea separation and subsea gas compression (Bai and Bai, 2012). Compared with topside processing, the advantages of subsea processing are (Bai and Bai, 2012):

- *Accelerated and/or increased production and/or recovery;*
- *Enabling marginal field developments, especially fields at deepwater/ultra-deepwater depths and with long tie-backs;*
- *Extended production from existing fields;*
- *Enabling tie-in of satellite developments into existing infrastructure by removing fluid;*
- *Handling constraints;*
- *Improved flow management;*
- *Reduced impact on the environment.*

2.2 *Subsea gas compression*

Since the late 1980s, several oil companies and research institutions tried to develop and commercialize subsea gas compression technology (Vinterstø et al., 2016), because a well can produce at lower wellhead pressures with subsea gas compression, thereby accelerating gas production and/or increasing recovery rate (Kuhnle et al., 2015).

However, it was considered that subsea gas compression requires *extensive further technology maturing* until 2005, while the other subsea processing were classified as *mature technology* or *high technical maturity level* (Fantoft, 2005).

On 16th September 2015, the world's first commercial subsea gas compression station was started-up on the Åsgard field, and it was followed by the Gullfaks (Vinterstø et al., 2016, Wadel-Andersen and Moe, 2016).

There are currently two different solutions for subsea gas compression: dry gas compression and wet gas compression (Tønnessen and Romanello, 2017). The former was applied at the Åsgard subsea gas compression station, while the latter is the concept for the Gullfaks subsea compression project. In dry gas compression, gas and liquid in the well stream are separated and boosted by a compressor and a pump respectively. In the wet gas compression, on the other hand, the well stream is boosted directly by a multiphase wet gas compressor without

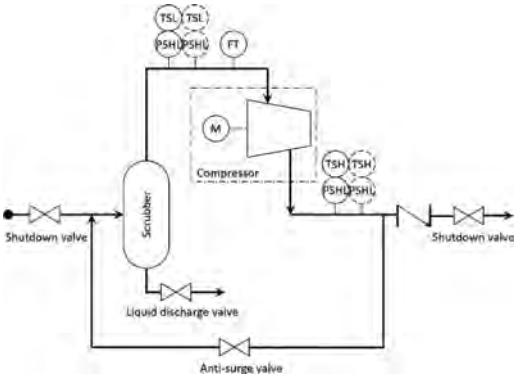


Figure 1. A typical subsea dry gas compression system (API RP 17V, 2015).

separation (Tønnessen and Romanello, 2017). Dry subsea compression is considered the standard solution, because it adopted some common principles from conventional topside compression (Dettwyler et al., 2016, Tønnessen and Romanello, 2017). Similarly, wet compression is called *well-stream compression* (Dettwyler et al., 2016).

2.3 *General configuration of subsea dry gas compression system*

API RP 17V (2015) provides a diagram that includes subsea dry gas compressors with typical safety devices, and a liquid discharge valve and a flow transmitter were added for the analysis of this paper as shown in Figure 1.

Abbreviations in Figure 1 are

- M: Motor
- FT: Flow Transmitter
- TSL: Temperature Safety Low
- TSH: Temperature Safety High
- PSHL: Pressure Safety High and Low

3 STPA FOR SUBSEA GAS COMPRESSION

3.1 *STAMP and STPA*

Leveson (2012) proposed a new accident causation theory, called Systems-Theoretic Accident Model and Processes (STAMP), whose main idea is that major accidents in today's complex, software-intensive, and sociotechnical systems are mainly caused by control problems rather than reliability problems. The three main concepts of this theory are safety constraints, hierarchical control structures, and process models.

Based on the STAMP theory, Leveson (2012) developed a new approach to hazard analysis, called Systems-Theoretic Process Analysis (STPA). The main reasons for developing STPA were to include new causal factors of STAMP that are not identified by traditional hazard identification techniques and to provide guidance to the users in getting good hazard identification results. STPA can identify more causal factors and hazardous scenarios, which are related to software, system design, and human behavior, than the other methods (Leveson and Thomas, 2013).

3.2 *STPA procedure*

The STPA procedure consists of one preparatory step (step 0) and two main steps (step 1 and 2) as described below (Leveson and Thomas, 2013):

- Step 0: *Establishing the system engineering foundation*
- Step 1: *Identifying unsafe control actions (UCAs)*
- Step 2: *Identifying the causes of the unsafe control actions*

Sub-steps with associated outcomes of each step are summarized in Figure 2. This paper follows this procedure to conduct STPA analysis for a subsea gas compression system.

3.3 *STPA analysis for subsea gas compression system*

3.3.1 *STPA Step 0*

In this step, we first identified system-level accidents (SLA), system-level hazards (SLH), and system-level safety constraints (SLSC) to define the scope of the analysis. Any unsafe control actions not relevant with defined SLHs were excluded in the further analysis.

The definition of a *hazard* in STPA is significantly different from traditional definition. The accidents and hazards in STPA are defined as below (Leveson and Thomas, 2013):

- *An accident is an undesired and unplanned event that results in a loss, including a loss of human life*

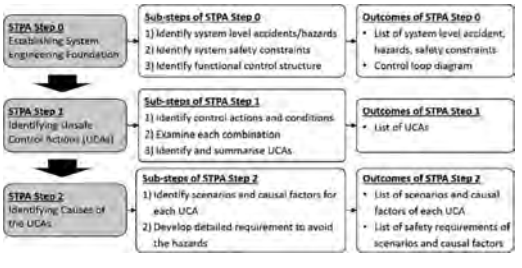


Figure 2. STPA procedure.

or human injury, property damage, environmental pollution, mission loss, financial loss, etc.
– *A hazard is a system state or set of conditions that together with a worst-case set of environmental conditions, will lead to an accident (loss).*

In subsea gas compression, large amounts of gas release can lead to loss of human lives (due to explosion on topside installations or loss of buoyancy of a vessel on the surface during a gas leak) or environmental pollution. In addition, significant economic loss may occur due to damage of costly subsea component or reduced gas production rate. The SLAs, SLHs, and SLSCs of subsea gas compression are summarized in Table 1. It is assumed that the compression system is designed inherently safe, so that the system can endure the pressure of the well stream and the compressor.

The next step was to identify functional control structure of the system. The high-level functional control structure of subsea gas compression system can be illustrated as shown in Figure 3. The system consists of Human Operator, Control System, Subsea Gas Compression (SGC) Unit,

Table 1. System-level accidents, hazards, and safety constraints of subsea gas compression.

System-level accident	System-level hazard	System-level safety constraints
SLA1: People die or are injured due to large amount of gas release	SLH1: SGC unit continues to supply gas when gas leaks to the environment	SLSC1: SGC unit must stop compressing gas when gas leaks to the environment
SLA2: The sea is polluted due to large amount of gas release		
SLA3: Valuable subsea components are damaged	SLH2: Compressor operates outside normal operation conditions	SLSC2: Compressor must be protected from extreme operating conditions that can damage the compressor
SLA4: Production is reduced or interrupted unnecessarily	SLH3: SGC unit stops compressing gas when not necessary	SLSC3: SGC unit must never stop compressing gas when not necessary
	SLH4: Compressor operates outside optimal conditions	SLSC4: SGC must be operated within optimal conditions

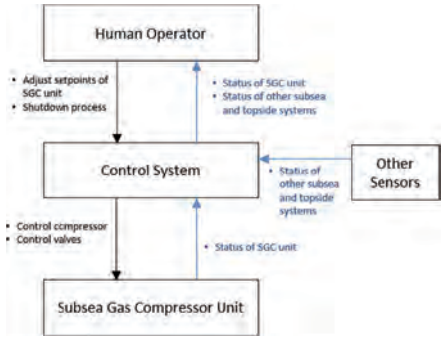


Figure 3. High-level functional control structure of subsea gas compression system.

and Other Sensors. The Human Operator provides *adjust setpoint of SGC unit* and *shutdown process* commands to the Control System, and the Control System provides commands to the compressor and valves of Subsea Gas Compressor Unit. The Control System receives feedbacks about *status of SGC unit* and *status of other subsea and topside systems* from Subsea Gas Compressor Unit and Other Sensors respectively, and these feedbacks are finally delivered to the Human Operator.

Based on the high-level functional control structure, a detailed model can be further developed as shown in Figure 4. The *Control System* consists of Variable Speed Drive (VSD), Process Control System (PCS), Process Shutdown (PSD) System, Subsea Control Unit (SCU), Subsea Control Module (SCM), and Subsea Electronic Module (SEM). VSD, PCS, PSD System, and SCU are located topside, while SCM and SEM are installed subsea. The Subsea Gas Compressor Unit is composed of Subsea Gas Compressor (SGC), Shutdown Valves (SDVs), Anti-Surge Valve (ASV), Liquid Discharge Valve (LDV), and Sensors. Responsibilities and process models of each controller are summarized in Table 2.

3.3.2 *STPA Step 1*

After establishing the functional control structure with responsibilities and process models, we could identify unsafe control actions (UCAs) by examining combinations of control actions and associated process models identified in Step 0.

Table 3 shows an example of how to identify UCAs of the system. One of responsibilities of PCS is to automatically control LDV depending on the level inside the scrubber. If the level inside the scrubber is too high, then liquid may flow into gas compressor, resulting in severe damage of the compressor. PCS should therefore open or close LDV to control the level of liquid in the scrub-

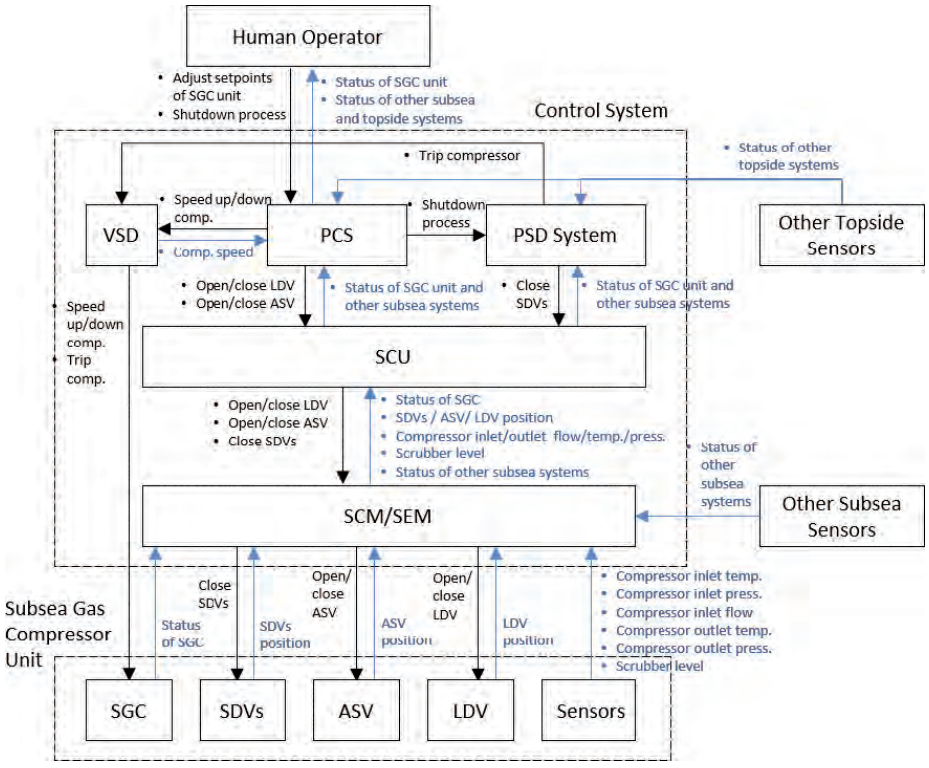


Figure 4. Detailed functional control structure.

Table 2. Responsibilities and process models of each controller.

Controller	Responsibilities	Process models
Human operator	– Adjust setpoint to maximize the efficiency of SGC unit – Shutdown process when needed	– Compressor inlet temperature/pressure/flow (low/normal/high) – Compressor outlet temperature/pressure (low/normal/high) – Status of other subsea and topside systems (normal/gas leak)
PCS	– Deliver *shutdown process* command from human operator to PSD system – Automatically adjust compressor speed – Automatically open/close LDV – Automatically open/close ASV	– Setpoints (optimal/not optimal) – Compressor inlet temperature/pressure/flow (low/normal/high) – Compressor outlet temperature/pressure (low/normal/high) – Scrubber level (low/normal/high)
VSD	– Deliver *speed up/down* and *trip* command from PCS and PSD to SGC	– Control command from PCS (speed up/down) – Control command from PSD (trip compressor)
PSD	– Trip compressor and close SDVs based on shutdown command from human operator – Automatically shut down process when needed	– Control command from PCS (shutdown) – Status of other subsea and topside systems (normal/gas leak)
SCU	– Deliver control commands from PCS and PSD system to SCM/SEM	– Control commands from PCS (open/close LDV, open/close ASV) – Control commands from PSD (close SDVs)
SCM/SEM	– Distribute control commands to each component	– Control commands from PCS (open/close LDV, open/close ASV, close SDVs)

Table 3. Identifying UCAs of open/close LDV command.

Controller: PCS								
No	Control action	Scrubber level	Not provided	Provided	Too early	Too late	Too short	Too long
1	Open LDV	High	***Unsafe [SLH2]***	Safe	Safe	***Unsafe [SLH2]***	***Unsafe [SLH2]***	Safe
2		Normal	Safe	Safe	Safe	Safe	Safe	Safe
3		Low	Safe	***Unsafe [SLH2]***	N/A	N/A	N/A	N/A
4	Close LDV	High	Safe	***Unsafe [SLH2]***	N/A	N/A	N/A	N/A
5		Normal	Safe	Safe	Safe	Safe	Safe	Safe
6		Low	***Unsafe [SLH2]***	Safe	Safe	***Unsafe [SLH2]***	***Unsafe [SLH2]***	Safe

Table 4. Scenarios, causal factors, and safety constraints of a UCA.

UCA.PCS001: Open LDV command is not provided when scrubber level is high		
Scenario	Associated causal factors	Safety constraints
PCS receives wrong measurement of scrubber level	Drift of scrubber LT	SC.PCS001.01 Scrubber LT must be calibrated periodically SC.PCS001.02 Scrubber LT must have 2003 configuration
PCS receives no measurement of scrubber level	No power supply to scrubber LT	SC.PCS001.03 PCS must generate an alarm when no signal is received from scrubber LT SC.PCS001.04 Scrubber LT must have UPS
	Broken signal wires from scrubber LT to PCS	SC.PCS001.03 PCS must generate an alarm when no signal is received from scrubber LT SC.PCS001.05 Signal wires must be inspected periodically
PCS receives correct measurement, but PCS does not provide open LDV command	Wrong logic inside PCS	SC.PCS001.06 PCS logic to generate open LDV command must be fully demonstrated during commissioning period

ber. In this case, control actions are *open LDV* and *close LDV*, and the process model is *scrubber level (high/normal/low)*.

UCAs can be identified by examining combinations of control actions and associated process models. For instance, if the *open LDV* command is not provided when the *scrubber level is high*, then this is a UCA that can cause SLH 2 identified in Step 0. On the contrary, it is a safe control action if *open LDV* command is not provided when the scrubber level is low. UCAs identified in Table 3 are

UCA.PCS001 Open LDV command is not provided when scrubber level is high
UCA.PSC002 Open LDV command is provided too late when scrubber level is high
UCA.PSC003 Open LDV command is provided too short when scrubber level is high
UCA.PSC004 Open LDV command is provided when scrubber is low
UCA.PSC005 Close LDV command is provided when scrubber level is high
UCA.PSC006 Close LDV command is not provided when scrubber level is low
UCA.PSC007 Close LDV command is provided too late when scrubber level is low
UCA.PSC008 Close LDV command is provided too short when scrubber level is low.

Similarly, other UCAs can be identified by combining control actions and process models of each controller.

3.3.3 *STPA Step 2*

The last step of STPA was to identify scenarios, associated casual factors, and safety constraints of each UCA. STPA provides less help for this step, and therefore, the analysts must rely on brainstorming with their own background knowledge and prior experiences (Leveson and Thomas, 2013).

An example of identified scenarios, causal factors, and safety constraints of an UCA are summarized in Table 4.

4 RESULTS AND DISCUSSION

4.1 *Results*

In this study, a total of 129 high-level UCAs have been identified for a subsea gas compression system. SCU and SCM/SEM have the largest number of UCAs, while the Human Operator is associated with the smallest number of UCAs. 66 out of 129 UCAs are related with SLH2, *compressor operates outside normal operation conditions*, while only nine UCAs can cause SLH3, *SGC unit stops compressing gas when not needed*. The number of UCAs of each controller and system-level hazard are summarized in Table 5.

Human operator's main responsibility is to adjust setpoints, so most of UCAs of human operator are related with SLH4, *compressor operates outside optimal conditions*. On the other hand, the main function of PSD system is to shut down the process when there occurs a gas leak, so the greater part of UCAs of PSD system are connected to SLH1, *SGC unit continues to supply gas when gas leaks to the environment*.

The main responsibility of PCS is to automatically control LDV and ASV to prevent the compressor from being damaged due to operating outside normal operation conditions, and therefore, UCAs of PCS are mostly relevant with SLH2, *compressor operates outside normal operation conditions*.

Table 5. Results of the analysis.

Controller	Total UCAs	UCAs to SLH1	UCAs to SLH2	UCAs to SLH3	UCAs to SLH4
Human Op.	10	2	0	1	7
PCS	30	2	10	1	17
PSD	12	8	0	4	0
VSD	15	2	0	1	12
SCU	31	2	28	1	0
SCM/SEM	31	2	28	1	0
Sum	129	18	66	9	36

SCU and SCM/SEM have the same number of UCAs, because the responsibility of SCU is to deliver control commands from PCS and PSD system to SCM/SEM.

Most of the control actions in the subsea gas compression system are to prevent compressor damage and/or operate the compressor in optimal conditions, rather than to prevent or stop gas leaks, because these functions are already covered by emergency shutdown (ESD) system with shutdown valves. Therefore, the subsea gas compression system has a large number of UCAs related to SLH2 and SLH4 compared to SLH1 and SLH3.

4.2 *Discussion and concluding remarks*

When applied to a subsea gas compression system, STPA provided a systematic and structured way of identifying UCAs in STPA Step 0 and 1. STPA provides, on the other hand, less help to identify scenarios and causes of UCAs, and relies on brainstorming in STPA Step 2. However, this is not only a problem of STPA. Other traditional hazard identification methods also have the same limitation. For instance, hazard and operability (HAZOP) studies provides guide words and parameters to identify hazards, structured what-if (SWIFT) uses checklists and what-if questions to identify hazards, and failure modes, effects, and criticality analysis (FMECA) utilizes system breakdown and functional analyses to identify hazards. However, once hazards are identified, all these methods rely on brainstorming to identify causes of the hazards. As Leveson and Thomas (2013) mentioned, STPA can provide more help for STPA Step 2 in the future, because there are common flaws that lead to accidents.

One of the distinct characteristics of subsea systems is the remoteness of control actions. The control commands for subsea systems are provided from topside installations or onshore that is sometimes hundreds of kilometers away from the subsea facilities. Therefore, there are some equipment that collect control commands and distribute the commands to associated components, like SCU and SCM/SEM in Figure 4. Accordingly, appropriate transmission and handling of control signals becomes important for the operation of the subsea systems, and STPA is well suited for addressing hazards in such distributed control structures.

However, only looking at the system from the perspective of the control units and flow of control commands and signals during operation can lead to omission of important hazards. To be fair, STAMP does consider a wider range of accident causes, including unhandled environmental disturbances or conditions, unhandled or uncontrolled component failures, unsafe interactions among

components and inadequately coordinated control actions by multiple controllers (Leveson and Thomas, 2013). Also, Leveson uses the term control in a broad sense, to include control by design (i.e. to prevent or protect against component failures or unsafe interactions), control processes (including developmental, manufacturing maintenance and operational processes) and social controls (i.e. legal requirements, cultural norms, or other interests that constrain behaviour) (Leveson and Thomas, 2013). However, although STPA may in principle cover a wide range of hazards, it is not necessarily the best method or easiest method to apply for identifying or analyzing all types of hazards. Other methods, such as HAZOP and FMECA also have their advantages and may be more familiar to safety engineers and risk managers. For example, while STPA takes the working system as point of departure and then identify flaws that could cause hazards, FMECA starts from the failure modes of subsystems to identify system effects, and HAZOP focus on deviations from normal operations that could cause hazards. Often these failure modes and possible deviations are known and STPA is not needed to identify them. Rather, the question is how critical they are for the system operation and safety, and if there are system hazards *not* covered by these subsystem-oriented approaches. In the latter case STPA may add important additional insight into system hazards.

The oil and gas industry has widely adopted risk-based safety philosophies, where safety is regarded as tolerable risk. In contrast, STAMP and STPA views safety as a control problem, where hazard can be eliminated by imposing control measures and constraints, or by modifying the system. STPA does not provide a method for describing, ranking or comparing the risk associated with identified unsafe control actions. In reality, imposing constraints and controls to eliminate unsafe control actions will have a cost and will not be perfectly reliable or remove hazards completely. While STPA is good for identifying inadequate control and suggesting additional controls, it does not provide stop criteria for when the control is sufficient. Eventually, the need for prioritization of resources will necessitate an evaluation of the associated risk to ensure that resources are spent optimally and that safety of the system represents a tolerable risk.

In conclusion, if the goal is to identify as many hazards, unsafe control actions and dangerous scenarios as possible, it is useful to view a system from several different perspectives, and not insist on using only one method. Hence, rather than replacing other methods, STPA should be used as a supplement, providing a control perspective on safety.

ACKNOWLEDGEMENTS

This paper has been written under the Norwegian Centre for Research based Innovation on Subsea Production and Processing (SUBPRO). The authors would like to thank the Research Council of Norway, as well to the industrial partners involved in this project.

REFERENCES

Allison, C.K., Revell, K.M., Sears, R. & Stanton, N.A. (2017) Systems Theoretic Accident Model and Process (STAMP) safety modelling applied to an aircraft rapid decompression event. *Safety science,* 98, 159–166.

Antoine, B. (2013) Systems Theoretic Hazard Analysis (STPA) applied to the risk review of complex systems: an example from the medical device industry. Massachusetts Institute of Technology.

API RP 17V (2015) Recommended Practice for Analysis, Design, Installation, and Testing of Safety Systems for Subsea Applications. American Petroleum Institute.

Baggerud, E., Halvorsen, V.S. & Fantoft, R. (2007) Technical Status and Development Needs for Subsea Gas Compression. *Offshore Technology Conference.* 30 April-3 May, Houston, Texas, U.S.A.

Bai, Y. & Bai, Q. (2012) *Subsea engineering handbook*, Gulf Professional Publishing.

Chen, J., Zhang, S., Lu, Y. & Tang, P. (2015) STPA-based hazard analysis of a complex UAV system in take-off. *International Conference on Transportation Information and Safety (ICTIS)*. IEEE.

Dettwyler, M., Büche, D. & Baumann, U. (2016) Subsea Compression-Current Technology and its Use to Maximize Late Life Production. *Proceedings of the 45th Turbomachinery Symposium.* Turbomachinery Laboratories, Texas A&M Engineering Experiment Station.

Fantoft, R. (2005) Subsea Gas Compression—Challenges and Solutions. *Offshore Technology Conference.* 2–5 May, Houston, Texas.

Ishimatsu, T., Leveson, N.G., Thomas, J., Katahira, M., Miyamoto, Y. & Nakao, H. (2010) Modeling and hazard analysis using STPA.

Kim, H., Lundteigen, M.A. & Holden, C. (2016) A Gap Analysis for Subsea Control and Safety Philosophies on the Norwegian Continental Shelf. *13th International Conference on Probabilistic Safety Assesment and Management.* 2–7 October, Seoul, South Korea.

Kuhnle, T.I., Myhrvold, T., Grande, Ø., Pedersen, F.B., Yang, Y., Jafar, M., Sewraz, D. & Irvine, M. (2015) All Subsea—Creating Value from Subsea Processing. Norway, DNV GL.

Leveson, N. (2012) *Engineering a safer world: Systems thinking applied to safety,* MIT press.

Leveson, N. & Thomas, J. (2013) An STPA primer. Cambridge, MA.

Leveson, N., Wilkinson, C., Fleming, C., Thomas, J. & Tracy, I. (2014) A Comparison of STPA and the ARP 4761 Safety Assessment Process. MIT PSAS Technical Report.

Leveson, N.G. (2014) Extending the human controller methodology in Systems-Theoretic Process Analysis (STPA). Massachusetts Institute of Technology.

Lima, F.S., Storstenvik, A. & Nyborg, K. (2011) Subsea Compression: A Game Changer. *Offshore Technology Conference Brasil.* 4–6 October, Rio de Janeiro, Brazil.

Monsen, B., Rongve, K.S., Lægreid, T. & Gutscher, C. (2012) Åsgard subsea gas compression-Technology qualification testing with high-speed VSD and very long step-out cable. *Petroleum and Chemical Industry Technical Conference.* 24–26 Sept. 2012, Chicago, IL, USA.

Nakao, H., Katahira, M., Miyamoto, Y. & Leveson, N. (2011) Safety guided design of crew return vehicle in concept design phase using STAMP/STPA. *Proc. of the 5: th IAASS Conference.* Citeseer.

Proctor, S., Hatcliff, J., Fernando, A. & Weininger, S. (2015) Using stpa to support risk management for interoperable medical systems. *2015 STAMP Workshop Presentations.*

Rachman, A. & Ratnayake, R.C. (2015) Implementation of system-based hazard analysis on physical safety barrier: A case study in subsea HIPPS. *IEEE International Conference on Industrial Engineering and Engineering Management.* 6–9 Dec. 2015, Singapore, IEEE.

Salim, H.M. (2014) Cyber safety: A systems thinking and systems theory approach to managing cyber security risks. Massachusetts Institute of Technology.

Samost, A. (2014) Evaluating Systems with Multiple Processes Using STPA: A Case Study in a Medical Intensive Care Unit. *GI-Jahrestagung 2014.* 25 Sep. 2014, Stuttgart, Germany.

Schmittner, C., Ma, Z. & Puschner, P. (2016) Limitation and Improvement of STPA-Sec for Safety and Security Co-analysis. *International Conference on Computer Safety, Reliability, and Security.* 22 Sep. 2015, Delft, Netherlands, Springer.

SUBPRO (2017) SUBPRO Annual Report 2017. Trondheim, Norway, NTNU.

Tønnessen, L.A. & Romanello, P. (2017) Future Subsea Compression. *Offshore Mediterranean Conference and Exhibition.* 29–31 March, Ravenna, Italy, Offshore Mediterranean Conference.

Vinterstø, T., Birkeland, B., Ramberg, R.M., Davies, S. & Hedne, P.E. (2016) Subsea Compression – Project Overview. *Offshore Technology Conference.* 2–5 May, Houston, Texas, USA, Offshore Technology Conference.

Wadel-Andersen, F. & Moe, H. (2016) Qualification and Implementation of a Subsea Wet Gas Compressor Solution. *Offshore Technology Conference.* 2–5 May, Houston, Texas, USA, Offshore Mediterranean Conference.

Young, W. & Leveson, N. (2013) Systems thinking for safety and security. *Proceedings of the 29th Annual Computer Security Applications Conference.* ACM.

Young, W.E. (2014) STPA-SEC for cyber security mission assurance. *Eng Syst. Div. Syst. Eng. Res. Lab*.

Zhang, Y., Jones, P., Campos, J.C., Masci, P. & Campos, J.C. (2017) A Hazard Analysis Method for Systematic Identification of Safety Requirements for User Interface Software in Medical Devices. *15th International Conference on Software Engineering and Formal Methods.* 4–8 Sep. 2017, Trento, Italy, Springer.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Enhanced condition monitoring of the machining process using wavelet packet transform

L. Mao & L.M. Jackson
Department of Aeronautical and Automotive Engineering, Loughborough University, UK

P. Goodall & A. West
Department of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, UK

ABSTRACT: Tool wear in machining processes can have a detrimental impact upon the surface finish of a machined part, increase the energy consumption during manufacture and potentially, if the tool fails completely, damage incurred may require the part to be scrapped. Monitoring of the tools condition can therefore lead to preventative steps being taken to avoid excessively worn tools being used during machining, which could cause a part becoming damaged. Several studies have been devoted to condition monitoring of the machining process, including the evaluation of cutting tool condition. However, these methods are either impractical for a production environment due to lengthy monitoring time, or require knowledge of cutting parameters (e.g. spindle speed, feed rate, material, tool) which can be difficult to obtain. In this study, we aim to investigate if tool wear can be directly identified using features extracted from the electrical power signal of the entire Computer Numerical Control (CNC) machine (three phase voltage and current) captured at 50 KHz, for different cutting parameters. Wavelet packet transform is applied to extract the feature from the raw measurement under different conditions. By analyzing the energy and entropy of reconstructed signals at different frequency sub-bands, the tool wear level can be evaluated. Results demonstrate that with the selected features, the effects due to cutting parameter variation and tool wear level change can be discriminated with good quality, which paves the way for using this technique to monitor the machining process in practical applications.

1 INTRODUCTION

Tool wear and subsequent failure of tools during the manufacturing process will have a significant impact on the economics of machining, and about 25% of machine down time can be attributed to the direct results of tool wear failure (Altintas & Yellowley 1989). Moreover, the development of tool wear will give rise to inconsistencies in surface finishes and geometric tolerances, affecting the quality of manufactured products. Therefore, a series of studies have been devoted to monitoring systems detecting underperforming tooling and improving machining efficiency and productivity.

The monitoring techniques for tool wear can be divided into two categories, direct and indirect methods (Bhattacharyya & Sengupta 2009, Teti et al. 2010). With direct methods, tool wear is evaluated by analyzing the cutter itself, such as measuring the surface roughness and flank wear, etc. On the other hand, indirect methods apply either model-based or data-driven techniques to the measurements like cutting force, tool vibration and output power for evaluating tool wear conditions. It should be noted that due to the restrictions of direct methods such as stopping requirements during production, indirect methods are more suitable for industrial applications (Zhu et al. 2009).

With indirect methods, different measurements can be collected and analyzed to evaluate the tool condition, including acoustic emission (Prickett & Johns 1999, Karimi et al. 2013, Hass et al 2013), cutting force (Dimla & Lister 2000, Li et al. 2006, Deng et al. 2013, Lee et al. 2006), vibration (Yesilyurt & Ozturk 2007, Zhang & Chen 2007, Lamraoui et al. 2014), temperature (Byrne 1987, Davoodi & Hosseinzadeh 2012), spindle power/current (He et al. 2017, Li et al. 2000, Simoneau & Meehan 2013), etc. However, several of these methods often require expensive sensing equipment (Nouri et al. 2015) and can be difficult to install due to the need for close proximity to the cutting tool and workpiece, meaning they can be impractical for large production environments. Additionally, the classification of tool wear from the collected data is challenging due to the high sensitivity of data to the cutting parameters (i.e. spindle speed, feed rate, depth and width of cut, material, tool type). Thresholding of time domain data has been used as a method of classifying tool wear (Shao et al.

2004), however, this requires large amounts of calibration and training data which is time consuming to collect and reduces the robustness of the system to a limited set of cutting conditions. Several studies of investigated frequency and time-frequency domain analysis to reduce the sensitivity to classification to cutting parameters (Kuljanic et al. 2009, Liao et al. 2007, Huang et al. 2010, Lauro et al. 2004], however, most of these methods require specialist monitoring equipment which pose the challenges described above.

Within this research we investigate the potential of a low cost, non-invasive sensing approach which is also cutting parameter agnostic to the problem of tool condition monitoring, which has so far not been identified within existing literature. The investigated solution uses current and voltage sensors across electrical three phase input to the machine to monitor the overall machine power consumption, whilst classification of the signal is conducted though time-frequency analysis using wavelet packet transform.

In Section 2 the diagnostic approach using the wavelet analysis is described. Section 3 details the experimental methodology and results, and Section 4 concludes the findings and highlights limitations and future work.

2 DIAGNOSTIC APPROACH

Although several studies have been performed for condition monitoring of the milling process using wavelet transform [Choi et al. 2004, Li et al. 2008, Zhong et al. 2010], these have mainly used vibration or cutting force measurements in the analysis instead of electrical power consumed by the machine. Moreover, the effectiveness of wavelet transform in discriminating tool wear level operated at varying cutting parameters still requires further investigation.

In this study, wavelet packet transform (WPT) is selected to evaluate the tool wear level. The reason of using WPT is that compared to wavelet transform, which only filters the signal to get the low-pass results (approximation), WPT can filter the signal to obtain both low-pass and high-pass (detailed) results (depicted in Figure 1). Therefore, more information can be extracted from the original signals using WPT [Torrence & Compo, 1995]. The extracted wavelet coefficients $C_{j,k}$ can be expressed using Eq. (1).

$$C_{j,k} = \int f(t)\psi_{j,k}(t)dt \tag{1}$$

where $f(t)$ is the original signal, $\psi_{j,k}$ is the wavelet function, j and k are the scale and shift parameters, this can be expressed in Eq. (2).

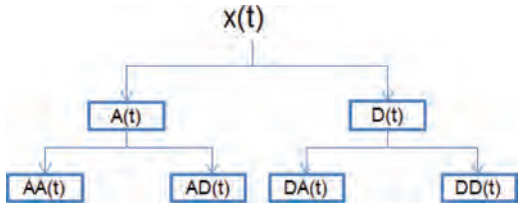


Figure 1. Two-level wavelet packet transform, where A and D are the approximation and detail by filtering the signal at the previous level.

$$\psi_{j,k}(t) = \frac{1}{\sqrt{j}}\psi\left(\frac{t-k}{j}\right) \tag{2}$$

It can be seen from Figure 1 that the application of WPT provides a sub-band filtering of the original signal into progressively finer equal-width intervals with the extracted packets of wavelet coefficients, i.e. the i^{th} packet of wavelet coefficients at j^{th} level represent the information of original signal within the frequency sub-band of $\left[iF_s/2^{j+1}, (i+1)F_s/2^{j+1}\right]$, where F_s is the sampling frequency.

With wavelet coefficients, the time-history of signals at different frequency sub-bands can be reconstructed using Eq. (3).

$$f_j(t) = C\sum_k C_{j,k}\psi_{j,k}(t) \tag{3}$$

where C is a constant independent of signals.

With these constructed signals, energy and entropy are calculated from each signal using Eqs. (4) and (5).

$$E_s = \int \left| f(t)f(t) \right| dt \tag{4}$$

$$H_s = -\sum_{i=1}^{N} p(x_i)\log_{10} p(x_i) \tag{5}$$

where E_s and H_s are the energy and entropy of the signal, $p(x_i)$ in Eq. (5) is a probability of the signal with value of x_i.

The energy of reconstructed signals represents the amount of information within different frequency sub-bands, while entropy of reconstructed signals can indicate the signal disorganization at the frequency sub-bands. It is expected that these two features would be sensitive to the change of cutting parameters and cutting tool wear level, thus can be used for the discrimination of cutting tool levels. This will be further investigated in the following section.

It can be seen that with the use of WPT, the original signals can be decomposed and reconstructed

at different frequency ranges, from which the frequency information can be related to the signals in the time domain, and better used for the feature extraction and fault diagnosis.

3 PERFORMANCE OF DIAGNOSTIC APPROACH

3.1 *Experiments*

In the study, HSS-Co8 is selected as the end milling tool due to its ease of wear measurement, which is a high speed steel containing 8% cobalt with 4 flutes. Two end milling tools with different diameters are selected herein for the analysis. Table 1 lists the characteristics of these end milling tools, where LOC refers to the tool's length of cut.

In the experiments, each end mill was assigned a work piece of dimension 150 mm × 120 mm × 30 mm, and the plate material was selected as commercial aluminum grade 6082 T651, which is a common alloy used in manufacturing.

Cutting parameters used in the tests were selected according to the manufacture's recommendation, which are listed in Table 2.

For the duration of each cutting session the energy monitoring device was connected to the system, which collected the current and voltage measurements at a sampling frequency of 50 kHz. During each session, the tools were used to perform climb milling on the work pieces. The number of passes, cut depth and cutting radius are selected as 10, half of the cutter diameter, respectively.

It should be mentioned that after each cutting session, the tools were used to machine the carbon steel to induce wear (40 min initially, subsequently 20 min), and this process was repeated until 100 min, where full tool wear was observed. Table 2 lists the cutting parameters used for different cuts and corresponding wear measurements.

Table 1. Characteristics of end milling tools.

Mill Dia. (mm)	Shank Dia. (mm)	LOC (mm)	Overall length (mm)	No. teeth/ flutes
8.0	10	19	69	4
10.0	10	22	72	4

Table 2. Cutting parameters and corresponding wear measurements.

Dia. (mm)	Cut no.	Time (min)	Spindle speed (RPM)	Feed rate (m/min)	Localized tool wear (mm)
8	1	0	4000	580	0
8	2	40	4000	580	0.182
8	3	60	4000	580	0.279
8	4	80	4000	580	0.327
8	5	100	4000	580	0.493
10	1	0	3100	600	0
10	2	40	3100	600	0.185
10	3	80	3100	600	0.434
10	4	100	3100	600	0.582

3.2 *Discrimination of different cutting tools with different wear level*

In this section, the current and voltage from two end milling tools with 8 mm and 10 mm cutting diameters are collected at 0 min and 100 min, which represents the intact and fully worn tools. The reason of selecting these measurements is that the collection process will not interrupt the machining process, and the installation of sensors will not add complexity of monitoring systems, thus the results can be better applied in the practical machining process.With current and voltage measurements, the instantaneous power can be calculated using the following equation.

$$P_{ins}(t) = v(t)i(t) \tag{6}$$

where $v(t)$ and $i(t)$ are the collected voltage and current measurements at time t.

Figure 2 depicts the instantaneous powers for 8 mm and 10 mm tools at intact and fully worn conditions. It should be mentioned that only power from single pass cutting is illustrated herein, as the powers of 10 passes have a similar trend. In the current study, only the power from a single pass is analyzed. Table 3 lists the average and maximum instantaneous power at each condition. It can be seen from the table that the instantaneous power will be increased with cutting tool wear.

From the instantaneous powers shown in Figure 2, the cutting tools with different diameters and cutting parameters, and the same tool with different wear levels cannot be discriminated easily in the time-domain, as the signals from different conditions have a similar shape, thus the four different conditions cannot be discriminated using only the power amplitude variation.

As described in section 2, WPT is applied to the instantaneous power to extract wavelet coefficients and reconstruct signals at different frequency sub-bands. In the current study the WPT is used to decompose the original signal over 8 levels. This decomposition level is selected by considering both the range of frequency sub-band and computational time. In the current study the Shannon wavelet

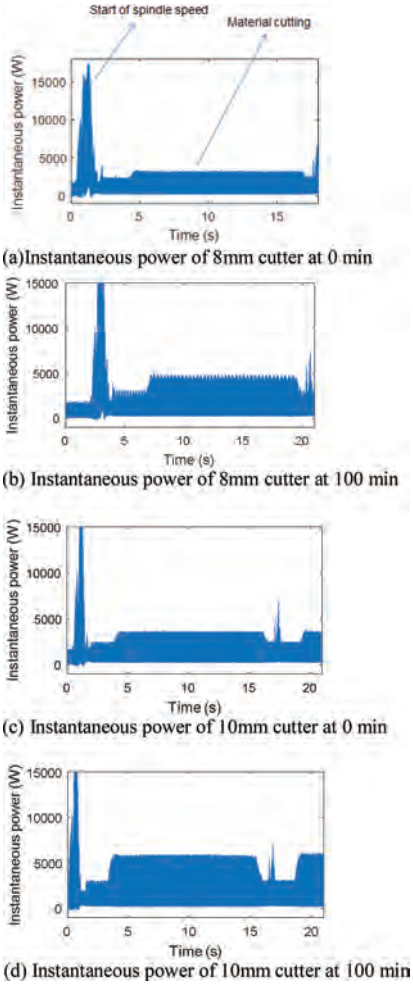


Figure 2. Instantaneous power of two cutters at intact and fully worn conditions.

Table 3. Average and max instantaneous power.

Condition	Average power (W)	Max power (W)
8 mm tool at 0 min	1593.7	3277.4
8 mm tool at 100 min	1989.6	4890.1
10 mm tool at 0 min	1712.3	3827.7
10 mm tool at 100 min	2537.5	5971.5

function is used in the WPT analysis, which can be written as follows:

$$\Psi_{j,k}(t) = \sqrt{j}\,sinc(jt)e^{2\pi ikt} \tag{7}$$

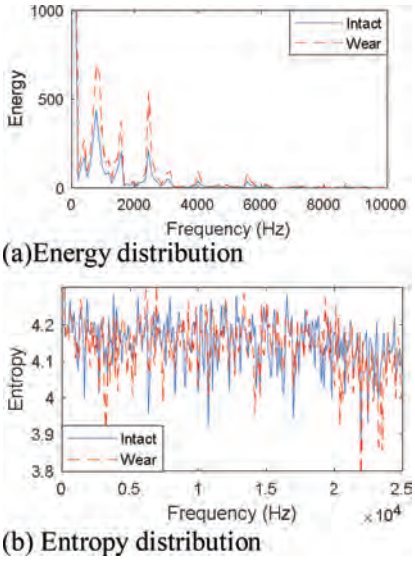


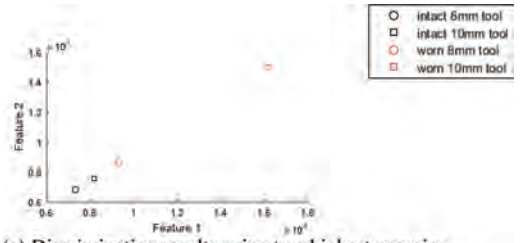
Figure 3. Energy and Entropy distributions of 8 mm cutter at intact and worn conditions.

Energy and entropy are then calculated from each reconstructed signal. Figure 3 depicts the distribution of energy and entropy over the whole frequency range. It should be noted that as the distributions are similar for the two end milling tools, only energy and entropy distribution from the 8 mm tool are illustrated herein.

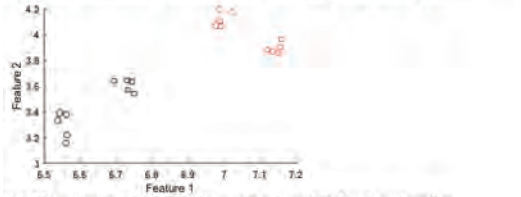
It can be seen from Figure 3 that the energy distribution shows similar trends for both intact and worn conditions, and the maximum energy is concentrated at around 700 Hz. However, the entropy distributes well along the whole frequency range, and the entropy distribution at intact and worn conditions shows clear variation. This indicates that the energy features can provide more consistent results, while entropy features are more sensitive to the change in the cutting parameters.

In this study, the two highest energies and entropies at the intact condition are selected for the discrimination, as they represent the most information and disorganization in the original signal. Figure 4 depicts the discrimination results. It should be noted that each point in Figure 4 represent the feature calculated with a two-second length instantaneous power signal.

From Figure 4, it can be seen that with selected energy and entropy features, all four different states, i.e. two end milling tools with two wear levels, can be discriminated with good quality, indicating that not only the worn condition can be identified clearly for the same cutting tool, but the different cutting tools with similar worn levels can also be separated accurately.



(a) Discrimination results using two highest energies



(b) Discrimination results using two highest entropies

Figure 4. Discrimination results using two highest energies and entropies.

When this approach is used in practical applications, the state of the end milling tool can be determined with the minimum Euclidean distance between features (two highest energies or entropies) of instantaneous power from the unknown state and the features shown in Figure 4. It should be mentioned that as the analysis is computational efficient (only taking about 20 seconds to gain the results), this approach can be used in the practical application for on-line monitoring purposes.

3.3 *Discrimination of different cutting tools with different wear level and similar instantaneous power*

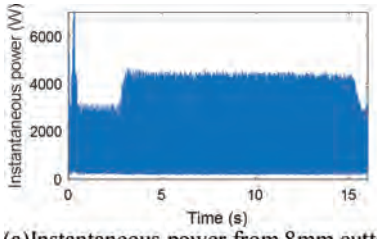
The performance of WPT in discriminating cutting tool conditions is further investigated using the data from the end milling tools at different cutting parameters but having similar instantaneous power, which makes it extremely difficult for discrimination using time-domain techniques.

In this study, two sets of data are used for the analysis, including end milling tools with diameters of 6 mm and 10 mm at different wear levels and cutting parameters. Table 4 lists the cutting parameters of these two end mill tools and the corresponding wear measurements.

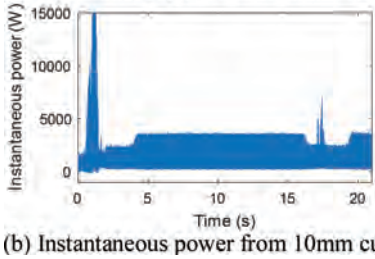
Figure 5 depicts the instantaneous powers from these two end milling tools. Similar instantaneous power can be observed due to the combination of different wear levels and cutting parameters. The average instantaneous powers from these conditions are listed in Table 5. It can be seen that these two conditions will provide similar average instantaneous power, while clear variation is observed in the wear level, which is listed in Table 4.

Table 4. Cutting parameters and corresponding wear measurements.

Dia. (mm)	Spindle speed (RPM)	Feed rate (m/min)	Time (min)	Wear measurement (mm)
8	4000	580	60	0.279
10	3100	600	0	0



(a)Instantaneous power from 8mm cutter



(b) Instantaneous power from 10mm cutter

Figure 5. Instantaneous powers from 6 mm and 10 mm diameter end mill tools.

Table 5. Average and max instantaneous power.

Condition	Average power (W)	Max power (W)
8 mm tool at 60 min	1810.1	4461.2
10 mm tool at 0 min	1712.3	3827.7

WPT described in section 2 is applied to extract the wavelet coefficients over 8 levels, and signals at different frequency sub-bands are reconstructed. The two highest energies and entropies are then selected for the discrimination. Results are depicted in Figure 6.

It can be seen from Figure 6 that with the selected energy and entropy, the two end milling tools can be discriminated with good quality, indicating the effectiveness of the proposed approach in identifying the states of different end mill tools at varying cutting parameters.

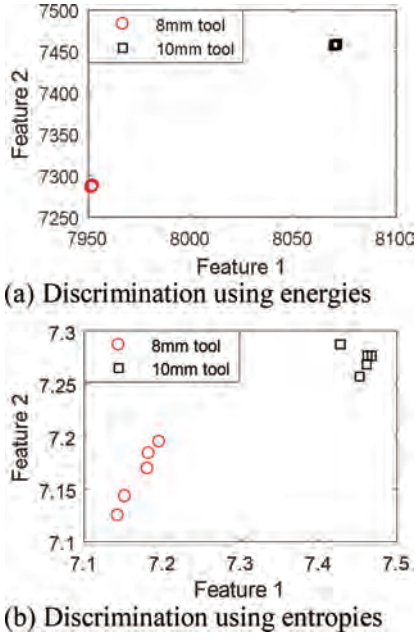


Figure 6. Discrimination results of 6 mm and 10 mm tools at different conditions.

4 CONCLUSIONS

In this paper, the discrimination of end mill tools with different diameters, wear levels, and cutting parameters is investigated. Wavelet packet transform is applied to extract wavelet coefficients from the original signal. Signals at different frequency sub-bands are then reconstructed using wavelet coefficients from which energy and entropy are calculated. The two highest energies and entropies are selected to discriminate different cutting tool states.

Two cases are used in this study to investigate the performance of the proposed method; cutting tools with different diameters and wear levels, and cutting tools with different diameters and wear level but similar instantaneous powers. Results demonstrate that with the proposed approach, the state of the cutting tool can be discriminated with good quality, both the tool wear level and cutting parameters can be discriminated.

Whilst these initial results are promising further work is required to expand the analysis over a wider range of cutting parameters to establish if the methodology holds. Additionally, refinement of the sensor measurement and tool monitoring service is required. At present signal analysis is performed off-line, whilst data is captured at higher frequency (50khz) increasing the cost of equipment and time of analysis. Optimization of this methodology is required in order to enable on-line monitoring.

ACKNOWLEDGEMENT

The work is supported by grant EP/K014137/1 for Loughborough University from the UK Engineering and Physical Sciences Research Council (EPSRC). The authors also acknowledge the industrial and academic collaborators of the AI2M project (Adaptive Informatics for Intelligent Manufacturing).

REFERENCES

Altintas, Y., Yellowley, A. 1989. In-process detection of tool failure in milling using cutting force models. *Journal of Engineering for Industry* 111(2): 149–157.

Bhaskaran, J., Murugan, M., Balashanmugam, N., Chellamalai, M. 2012. Monitoring of hard turning using acoustic emission signal, *Journal of Mechanical Science and Technology* 26(2): 609–615

Bhattacharyya, P., Sengupta, D. 2009. Estimation of tool wear based on adaptive sensor fusion of force and power in face milling. *International Journal of Production Research* 47(3): 817–833.

Byrne, G. 1987. Thermoelectric signal characteristics and average interfacial temperatures in the machining of metals under geometrically defined conditions. *International Journal of Machine Tools and Manufacture* 27(2): 215–224.

Choi, Y., Narayanaswami, R., Chandra, A. Tool wear monitoring in ramp cuts in end milling using the wavelet transform. *International Journal of Advanced Manufacturing Technology* 23(5–6): 419–428.

Davoodi, B., Hosseinzadeh, H. 2012. A new method for heat measurement during high speed machining. *Measurement* 45(8): 2135–2140.

Deng, W.J., Li, Q., Li, B.L., He, Y.T., Xia, W., Tang, Y. 2013. Study on the cutting force of cylindrical turning with novel restricted contact tools. *The International Journal of Advanced Manufacturing Technology* 69(5–8): 1625–1638.

Dimla, D.E., Lister, P.M. 2000. On-line metal cutting tool condition monitoring I: force and vibration analyses. *International Journal of Machine Tools and Manufacture* 40(5): 739–768.

Hase, A., Wada, M., Koga, T., Mishina, H. 2013. The relationship between acoustic emission signals and cutting phenomena in turning process. *The International Journal of Advanced Manufacturing Technology* 70(5–8): 947–955.

He, Y.V., Leung, L.C., Linn, R. 2017. Pareto fronts of machining parameters for trade-off among energy consumption, cutting force and processing time. *International Journal of Production Economics* 185: 113–127.

Huang, L., Kemao, Q., Pan, B., Asundi, A.K. 2010. Comparison of Fourier transform, windowed Fourier transform, and wavelet transform methods for phase

extraction form a signal fringe pattern in fringe projection profilometry. *Optics and Lasers in Engineering* 48(2): 141–148.

Karimi, N.Z., Heidary, H., Minak, G., Ahmadi, M. 2013. Effect of the drilling process on the compression behavior of glass/epoxy laminates. *Composite Structures* 98: 59–68.

Kono, D., Matsubara, A., Yamaji, I., Fujita, T. 2008. High-precision machining by measurement and compensation of motion error. *International Journal of Machine Tools and Manufacture* 48(10): 1103–1110.

Kuljanic, E., Totis, G., Sortino, M. 2009. Development of an intelligent multi-sensor chatter detection system in milling. *Mechanical Systems and Signal Processing* 23(5): 1704–1718.

Lamraoui, M., Thomas, M., Badaoui, M.El. 2014. Cyclostationarity approach for monitoring chatter and tool wear in high speed milling. *Mechanical Systems and Signal Processing* 44(1): 177–198.

Lauro, C.H., Brandao, L.C., Baldo, D., Reis, R.A., Davim, J.P. 2014. Monitoring and processing signal applied in machining processes—A review. *Measurement* 58: 73–86.

Lee, K.J., Lee, T.M., Yang, M.Y. 2006. Tool wear monitoring system for CNC end milling using a hybrid approach to cutting force regulation. *International Journal of Machine Tools and Manufacture* 48(3–4): 371–379.

Li, X., Djordjevich, A., Venuvinod, P.K. 2000. Current-sensor-based feed cutting force intelligent estimation and tool wear condition monitoring. *IEEE Transactions on Industrial Electronics* 47(3): 697–702.

Li, X., Quyang, G., Liang, Z. 2008. Complexity measure of motor current signals for tool flute breakage detection in end milling. *International Journal of Machine Tools and Manufacture* 48(3–4): 371–379.

Li, H.Z., Zeng, H., Chen, X.Q. 2006. An experimental study of tool wear and cutting force variation in the end milling of Inconel 718 with coated carbide inserts. *Journal of Materials Processing Technology* 180(1–3): 296–304.

Liao, T.W., Ting, C.F., Qu, J., Blau, P.J. 2007. A wavelet-based methodology for grinding wheel condition monitoring. *International Journal of Machine Tools and Manufacture* 47(3–4): 580–592.

Nouri, M., Fussell, B.K., Ziniti, B.L., Linder, E. 2015. Real-time tool wear monitoring in milling using a cutting condition independent method, *International Journal of Machine Tools and Manufacture* 89: 1–13.

Prickett, P.W., Johns, C. 1999. An overview of approaches to end milling tool *monitoring. International Journal of Machine Tools and Manufacture* 39(1): 105–122.

Scheffer, C., Heyns, P.S. 2001. Wear monitoring in turning operations using vibration and strain measurements. *Mechanical Systems and Signal Processing* 15(6): 1185–1202.

Shao, H., Wang, H.L., Zhao, X.M. 2004. A cutting power model for tool wear monitoring in milling. *International Journal of Machine Tools and Manufacture* 44(14): 1503–1509.

Simoneau, A., Meehan, J. 2013. The impact of machining parameters on peak power and energy consumption in CNC end milling. *Energy and Power* 3(5): 85–90.

Teti, R., Jemielniak, K., Donnell, G.O., Dornfeld, D. 2010. Advanced monitoring of machining operations. *CIRP Ann. – Manuf. Technol.* 59(2): 717–739.

Torrence, C., Compo, G.P. 1995. A wavelet packet approach to transient signal classification. *Applied and Computational Harmonic Analysis*, 2: 265–278.

Yesilyurt, I., Ozturk, H. 2007. Tool condition monitoring in milling using vibration analysis. *International Journal of Production Research* 45(4): 1013–1028.

Zhang, J.Z., Chen, J.C. 2007. Tool condition monitoring in an end-milling operation based on the vibration signal collected through a microcontroller-based data acquisition system. *The International Journal of Advanced Manufacturing Technology* 39(1–2): 118–128.

Zhong, W., Zhao, D., Wang, X. A comparative study on dry milling and little quantity lubricant milling based on vibration signals. *International Journal of Machine Tools and Manufacture* 50(12): 1057–1064.

Zhu, K., Wong, Y.S., Hong, G.S. 2009. Wavelet analysis of sensor signals for tool condition monitoring: a review and some new results. *International Journal of Machine Tools and Manufacture* 49(7–8): 537–553.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Risk from cyberattacks on autonomous ships

Jan Erik Vinnem & Ingrid Bouwer Utne
Department of Marine Technology, NTNU, Norway

ABSTRACT: The vulnerability of technological and administrative systems to cyberattacks has been shown to be high in several cases, which has led to different unwanted consequences. Autonomous ships will also be exposed to the threat of cyberattacks, due to their need for connecting to operational, management and administrative systems onshore. The most critical hazards are possibly not associated with consequences for the ship itself or its cargo, but the threat to infrastructure along the coast and offshore if a ship under alien command is used as a "battering ram" to cause major structural damage. Even relatively small autonomous ships may pose a real threat, and ships sailing in international waters may come from distant locations. This implies that all autonomous ships may be considered as possible threats. This paper outlines the risk for some infrastructure systems. Even though the probability may be low, such events cannot be ruled out in the future, and the design of autonomous ships must involve a series of risk reducing actions and designs.

1 INTRODUCTION

Maritime security has come on the agenda the past decade. In 2004, the U.S. presented a national maritime security policy. The Sept. 11th attacks also put maritime terrorism on the agenda. The increase in piracy attacks in 2008 and 2011 outside the coast of Somalia contributed to even more attention to maritime security globally. In 2011, maritime security became one of the objectives in The North Atlantic Treaty Organization's (NATO) Alliance Maritime Strategy. The UK, EU and the African Union proposed maritime security strategies in 2014 (Bueger, 2015). The Maritime Safety Committee (MSC) in the International Maritime Organization has recently published guidelines on maritime cyber risk management (IMO, 2017a).

There is an increased focus on developing autonomous ships. A motivation is reduced building and operational costs, because the ships can be redesigned. Research projects, such as the Maritime Unmanned Navigation Through Intelligence in Networks (MUNIN) (Rødseth & Tjora, 2014) and Advanced Autonomous Waterborne Applications (AAWA, 2016) focus on the development of technological specifications and designs for autonomous ships. Industry projects aim at realizing the first autonomous ships in the next 1–3 years, e.g., Yara Birkeland (Kongsberg Maritime, 2017).

Autonomous ships will be exposed to the threat of cyberattacks, due to their need to connect to operational, management and administrative systems onshore. The most critical hazards are possibly not associated with consequences for the ship itself or its cargo, but the threat to infrastructure along the coast and offshore if a ship under alien command is used as a 'battering ram' to cause major structural damage. Even relatively small autonomous ships represent a high kinetic energy when travelling at full speed and may thus pose a real threat to infrastructure systems. Ships sailing in international waters may come from distant locations. This implies that all autonomous ships may be considered as possible threats. It will not be sufficient to ensure that the high-quality classification societies have stringent requirements; all classification societies or IMO need to focus on such threats.

We may think that the probability of cyberattacks may be low, but such events cannot be ruled out in the future. We therefore believe that it is important, before autonomous ships are built and commissioned, that the marine and maritime industry at large, consider this threat and takes necessary actions to implement sufficient risk control actions.

A cyber-attack may have some parallels with the terrorist attack on USS Cole, the United States Navy guided-missile destroyer, on 12th October 2000, while it was being refueled in Yemen's Aden harbor (US Navy, 2001). 17 sailors were killed and 39 injured, due to the attack from a small fiberglass boat carrying explosives and two suicide bombers. The boat approached the port side of the destroyer in bright daylight, and exploded, creating a 12 by 18 m gash in the ship's port side from what was estimated to 180–320 kg of explosives.

The objective of the paper is to discuss the implications of the vulnerability of autonomous ships to cyber-attacks, the threats that a ship under alien control may represent for infrastructure systems, and how such risk should be mitigated in general. There are also other activities and sectors in the society where cyber-attacks may be a potential threat. One incident known from the petroleum industry is described in Section 2.1. Some incidents in the energy sector are briefly mentioned in Section 2.3. Autonomous cars are another such sector, see further descriptions in Section 2.2. Experiences from other sectors can be used as a basis for assessing risk and developing relevant risk mitigation measures for autonomous ships.

Traditional risks to ships, which also apply to autonomous ships, such as collision, grounding, foundering, etc. are outside the scope of the paper, and are therefore not discussed. These risks are still important, and are subject to attention by several researchers. The risks to infrastructure systems are special in the sense that catastrophic consequences may cascade outside the industry itself.

The paper considers unmanned autonomous ships primarily, but differences between unmanned and manned autonomous ships are also considered.

2 REVIEW OF CYBER THREATS IN COMPARABLE SYSTEMS

2.1 *Petroleum industry*

It is not easy to collect experience data about cyber-attacks. Statoil corporate management was invited to give a university lecture about cyber threats to their systems and operations in October 2016 (Statoil, 2016). The incidents presented during this lecture are presented in Section 2.3 below. No incidents were mentioned in the lecture from Statoil's own operations. Three weeks later it was revealed through media that there had been a serious unintended incident at Statoil's Mongstad refinery in May 2014, as described in the following. Through the subsequent handling of this incident, it became clear that Statoil has had many more incidents of probably different severity. What was revealed by media a short while after the guest lecture puts the lack of openness in the university lecture in a special light.

The most well-known incident in the petroleum industry is from the downstream part, where maintenance on a server by an IT specialist in Hindustan Computers Ltd. (HCL) in India disrupted the loading of a gasoline tanker at the Statoil operated Mongstad refinery just outside Bergen in Norway on 21st May 2014. An input error by the operator gave him access to a server he should not be able to access. It should not be possible to stop the server in question remotely, but the HCL specialist inadvertently accessed the server through a 'back door', according to media.

The operations of certain IT systems, including the Mongstad refinery, was outsourced by Statoil to HCL in India in 2012, after a risk assessment. The incident referred to here did not affect safety directly, but could potentially have affected safety functions and barriers, according to the audit report by Petroleum Safety Authority (PSA, 2017).

The NRK broadcasting company in Norway found 29 incidents where information and communication technology (ICT) employees from India had accessed servers they should not have access to in Statoil. Anonymous sources in Statoil have commented that the problem was more extensive than what the journalists found.[1]

The PSA audit was initiated after the incident was known in the public domain, almost 2.5 years after it occurred. The audit considered the handling of incidents associated with ICT and information security by Statoil in general. PSA considered several ICT related incidents, as well as Statoil's technical requirements to information security for industrial automation and control systems. The wording of the PSA audit report is such that it indicates that other incidents have occurred that are unavailable in the public domain.

Statoil was criticized by PSA for failing to notify the authorities about the incident at the time it occurred, which according to Statoil's own assessment could have had consequences, such as failure of safety functions or barriers, according to media reports (see footnote[1] above).

Statoil informed in mid 2017 that they had cancelled all outsourcing contracts that affected safety critical systems. They had concluded that the outsourcing of these systems represented too high risk for unwanted influence on the systems.

From the media, it is known that Statoil was the target of a massive attack over three days in 2013, where hackers tried to install dangerous code into Statoil computers[2], apparently an unsuccessful attack.

2.2 *Autonomous vehicles*

Autonomous cars are expected to become an important part of the transportation system within the next decade. Self-driving vehicles will

[1]https://www.nrk.no/norge/xl/tastefeilen-som-stoppet-statoil-1.13174013.

[2]http://www.newsinenglish.no/2014/08/28/statoil-held-off-hacker-attack/.

be shared by several users (Lyche, 2017). Autonomous buses may be realized in the near future with operators in control centers remotely overseeing several buses. In specific circumstances, the operators may take over control and remotely operate the buses if needed (Lyche, 2017). This means that the autonomous buses will operate in different autonomy levels, with shared control.

A major challenge is the increasing interconnection that may expose safety-critical systems to security threats. Cars are no longer physically isolated machines controlled mechanically and locally (Macher et al, 2017). They have become computers with various electronic control units (ECU) and hackers may take control over brakes, engine, the steering wheel, radio, and lights. Recently, it was discovered that one million cars could be hacked simultaneously (Kibar, 2017; Slovik, 2017).

A car's vulnerability to hacking depends on what kind of remotely connection the car has, the configuration of the car's internal computer network, and how external digital commands may affect physical components (Kibar, 2017). Press (2017) discusses how cars can become weapons of mass destruction on the road. It will not be sufficient to install firewalls or intrusion detection systems. The UK Government states that Wi-Fi connected cars along with autonomous cars are getting increasingly vulnerable to hacking and data theft. They recently published key principles of vehicle cyber security for connected and automated vehicles to support the industry (GOV.UK, 2017). These principles are (quote):

1. Organizational security is owned, governed and promoted at board level.
2. Security risks are assessed and managed appropriately and proportionately, including those specific to the supply chain.
3. Organizations need product aftercare and incident response to ensure systems are secure over their lifetime.
4. All organizations, including sub-contractors, suppliers and potential 3rd parties, work together to enhance the security of the system.
5. Systems are designed using a defence-in-depth approach.
6. The security of all software is managed throughout its lifetime.
7. The storage and transmission of data is secure and can be controlled.
8. The system is designed to be resilient to attacks and respond appropriately when its defences or sensors fail.

The connectivity means that the vehicle is integrated in a global ad-hoc network system where external information are important for decision making. Security has become an important aspect to include in systems safety engineering. The development of these novel transportation systems means that systematic approaches taking both safety and security aspects into consideration are needed (Macher et al, 2017).

Standards relevant for the automotive domain are increasing their focus on security. IEC 61508:2010 mentions that security threats may be identified during hazard analysis. Nevertheless, the security threat analysis is not specified or detailed. The SAE J3061:2016 is a guideline for cybersecurity engineering. Among other things, it focuses on defining a process for implementing cybersecurity in the design, considering a vehicle's lifecycle and providing basic guiding principles on cybersecurity.

2.3 *Energy sector*

Some other cyber-attacks on the energy sector that are known in the public domain are the following (Statoil, 2016):

- Attacks on Technical network
  - Stuxnet: Iran's uranium enrichment facility 2010
  - German Steel Mill 2014
  - Ukrainian power network 2015
  - German nuclear plant 2016
- Attacks on Office network
  - Shamoon incident: Saudi Aramco office network 2012
  - Energetic Bear: Energy industry in the US and Europe 2012 →
  - Cleaver (recon): Energy infrastructure several countries around the globe 2012 →

The Gundremmingen nuclear power plant in Germany, located about 120 km northwest of Munich, is run by the German utility company RWE. It was found to be infected with computer viruses, but they appeared not to have posed a threat to the facility's operations because it is isolated from the Internet, according to press reports.[3] The viruses, which included "W32.Ramnit" and "Conficker", were discovered at Gundremmingen's B unit in a computer system retrofitted in 2008 with data visualisation software associated with equipment for moving nuclear fuel rods, RWE said. Malware was also found on 18 removable data drives, mainly USB sticks, in office computers maintained separately from the plant's operating systems. W32.Ramnit is designed to steal files from infected computers and targets Microsoft

[3]http://www.telegraph.co.uk/news/2016/04/27/cyber-attackers-hack-german-nuclear-plant/.

Windows software, according to the security firm Symantec. Conficker has infected millions of Windows computers worldwide since it first came to light in 2008. It is able to spread through networks and by copying itself onto removable data drives, Symantec said.

The 'Energetic Bear' is a Russian virus that let hackers take control of power plants. Over 1,000 energy firms have been infected, according to media reports.[4]The hackers obtained access to power plant control systems, and could have disrupted energy supplies in affected countries, if they had used the sabotage capabilities open to them, according to Daily Mail.

In October 2017, it has been revealed by media[5] that Russians have jammed the GPS signals in Northern Norway in September 2017, as a deliberate action by Russian militaries during a cyber warfare exercise.

3 CYBER RISKS FOR AUTONOMOUS SHIPS

3.1 *Hacking of autonomous ships*

The technological advancements towards ships operating without an onboard crew is enabled by the developments in ICT in recent years. ICT provides data connection and on-board intelligence and data connection capabilities. The ships may operate in different levels of autonomy. In a high level of autonomy, ships may be supervised by human operators in Shore Control Centres (SCC). Whenever necessary, the operator (supervisor) may intervene. A SSC could take responsibility for overseeing specific phases of a ship's operation or voyage, for example, maneuvering in and out of port, which then means that the ship would operate in a lower level of autonomy. The connectivity between the ship and SSC must have high capacity and availability and is crucial for the realization of autonomous ships (AAWA, 2016; MUNIN, 2015).

The increasing usage of networked ICT technology makes it possible to access systems through net-work interfaces and gain unauthorized remote capability to control ship systems in undesired manners (AAWA, 2016). Security threats that are relevant for ships are piracy and highjacking, smuggling of goods, human trafficking, damaging of ship or port facility, vandalism, sabotage, such as intentional jamming or spoofing of the ship automatic identification system (AIS), GPS signals and communication systems, and use of the ship as weapon for terrorist activity (AAWA, 2016; MUNIN, 2015).

The security challenge of shipping has been addressed by the International Maritime Organization (IMO) Maritime Safety Committee and The Facilitation Committee, who recently issued guidelines on maritime cyber risk management (IMO, 2017a). The guidelines give high-level recommendations on security risk management to protect shipping from current and emerging threats. Five functional elements are presented consisting of identification, protection, detection, responding and recovering. Vulnerable systems that are mentioned in the guideline are bridge systems, cargo handling and management systems, machinery and propulsion systems, control systems, passenger servicing and management systems, passenger public networks, crew welfare systems, and communication systems. IMO states that cyber risk management should be integrated into ship safety management within 2021 (IMO, 2017b).

To protect a ship against cyber threats means that vulnerabilities in the ICT infrastructure need to be eliminated and effective measures for intrusion prevention must be implemented. It is also necessary to consider that hackers may become more skillful over time with more advanced techniques available. This means that cyber security needs to be dynamic and proactive. Classification and encryption of data, user identification, authentication, authorization, protection of data integrity and connectivity, as well as activity logging and auditing are examples of typical cyber security methods that are expected to be needed (AAWA, 2016).

MUNIN (2015) presents a risk matrix, including both safety and security aspects. The highest ranked threats are found to be jamming, spoofing or hacker attacks of AIS, GPS signals, or communication systems, leading to collision with other ships, or ship grounding in critical areas.

3.2 *Autonomous ships used as threat to infrastructure systems*

The control over an unmanned ship which is hacked may be lost completely, which is the most severe situation. It is assumed that complete loss of control is impossible if there is a small crew onboard. It is assumed that a small crew may be able to deactivate external control and take over control locally. If this fails, they should at least be able to shut of power and let the ship drift until they may be able to take back control locally.

But without local crew such possibilities are not available, and control may be lost completely, at least for some time. In theory, control may be

[4]http://www.dailymail.co.uk/sciencetech/article-2675798/Hundreds-European-US-energy-firms-hit-Russian-Energetic-Bear-virus-let-hackers-control-power-plants.html.

[5]https://www.nrk.no/finnmark/e-tjenesten-bekrefter_-russerne-jammet-gps-signaler-bevisst-1.13721504.

reestablished by boarding the ship, for instance by helicopter, such as police helicopter or naval helicopter. This will take time in any case, and if the vessel is far from shore, a helicopter may not be able to reach the ship until it comes closer to shore, and then it may be too late.

It is therefore possible that an unmanned, autonomous ship that has been hacked may be used to ram into infrastructure systems. This is discussed further below. A similar scenario could also occur with a conventional manned ship, if the ship is highjacked, but this is outside the scope of the present discussion.

Let us first consider if a hacked, unmanned ship may be a threat to other ships in open seas. This may be possible in principle, but if the other ships are conventional, manned ships, they may be able to avoid the hacked, unmanned ship through maneuvering away from the threat. This may fail if the threat is not observed, but should normally be successful. It the second ship is an unmanned, autonomous ship, control from shore should be able to observe the threat in a similar manner.

A special case occurs if other ships represent potential extreme catastrophic consequences, for instance if the other ship is a cruise ship with many thousands of cruise passengers. Or if the other ship is a very (or ultra) large crude carrier, capable of transporting in order of 2,000,000 bbls of crude oil. These ships would not be autonomous, and should normally be able to avoid attack.

But infrastructure installations are usually stationary and not able to relocate to avoid the threat. By infrastructure systems in this context one may first of all think of bridges crossing fjords and bays and other seawater open areas which are found in almost all coastal areas worldwide. Other systems may be offshore petroleum installations, which are found far away from shore in several parts of the world; the North Sea, Gulf of Mexico, Atlantic Sea off the coast of Brazil, several African countries, Newfoundland, Shetland as well as the Pacific in some areas off Australia and the South China Sea.

There are considerable differences with respect to impact resistance to external impact in the various types of infrastructure systems. In Norway for example, there has been a study project ongoing to establish possible concepts for fjord crossing of some of the largest fjords on the West coast of Southern Norway. For a possible fjord crossing of the Sognefjord, a floating bridge concept has been specified to have 1563 MJ kinetic energy resistance, corresponding to a ship of about 31,500 tdw, travelling at a full speed of 17.7 knots (Statens Vegvesen, 2013). Smaller bridges along the coast are believed to have resistance at least one order of magnitude lower, but the consequences of a collision against a smaller bridge may be less extensive.

When it comes to offshore structures, the traditional resistance has been designed to take the impact from a drifting service vessel. Typically, this was a value of 14 MJ for many years (Vinnem, 2013), but is in recent years increased to around 50 MJ (Yu & Amdahl, 2018), due to increasing size of service vessels used for these installations. The largest offshore structures, the concrete gravity based structures (so-called Condeep structures), which we commonly installed in the North Sea some 20 years ago, have a push-over resistance about 200 MJ (Vinnem, 2013). This is almost an order of magnitude lower than the specified resistance of the bridge for the fjord crossing of the Sognefjord. Most of the offshore structures have capacity in the order of 50 MJ or less.

Floating offshore structures may in theory move away, if threat is detected sufficiently early. If the hacked ship is used with the intention to ram into a structure, it may be able to follow the movements of the offshore installation.

Even a small ship with a mass of 5,000 tons, travelling at a speed of 12 knots, has a kinetic energy of roughly around 200 MJ, which is excessive in relation to structural capabilities of most offshore structures; only the Condeep structures could be expected to survive. Larger ships will be a threat to all offshore structures.

The largest offshore structures are usually manned with up to a few hundred persons, implying that many lives are at risk. In addition, comes the blowout potential. Here the fixed installations are the most vulnerable, because the equipment to isolate the wells are mainly on deck. If the installation is wiped out, very long-lasting blowouts may occur as a result, in addition to the death toll.

4 FEASIBILITY OF RISK REDUCTION

4.1 *Approach to risk control*

The previous sections have shown that hacked autonomous unmanned ships may be a considerable threat to offshore installations, and to infrastructure elements along the coast unless particularly strengthened.

It is considered that further strengthening of constructions is not relevant. First of all, this is impossible for existing structures, and further strengthening of future structures is not relevant due to excessive costs. The risk control actions will need to be focused on prevention of the threats to cause incidents.

Traffic surveillance is one of the solutions adopted by the offshore oil and gas industry for protection of offshore installations against collision threats by passing vessels. For the Norwegian sector, there are several centers; two operated by off-

shore companies and several government operated centers along the coast. The main principle is to detect a ship on collision course as early as possible, to give the possibility to communicate with the ship and warn it to alter its course. If contact is not established, the approach implies to warn the installation sufficiently early, such that safe evacuation of all personnel may be completed. In addition, available resources may be used to try to establish contact with the vessel, if communication fails.

But the approach in this case assumes that the vessel does not want to collide. If on collision course, this is due to lack of knowledge, or in some cases with intent for a certain period, with a planned future course change. This approach is not correspondingly well suited if the ship is on collision course by intent. Communication is not going to change anything, nor the use of vessels or other resources to achieve physical contact. Still, the detection of a ship on collision course will imply that evacuation of personnel may be possible, if the procedures to start evacuation in a timely manner are adhered to. This will not protect the installation, though.

Keeping a small crew onboard is the most effective risk control action. It was assumed above that a small crew may be able to deactivate external control and take over control locally and mechanically. If this fails, they should at least be able to shut of power and let the ship drift until they may be able to take back control locally or assistance from shore has arrived.

A small crew would not need to be onboard al the time, the duration could be limited to where there are critical infrastructures.

If keeping a small crew onboard is infeasible, then the only option is to ensure as far as possible that there are no possibilities for hackers to gain access to the control of an autonomous unmanned ship.

Another option would be to limit the operational area of an unmanned autonomous ship for instance by limiting the available fuel stored onboard. This is to some extent used for aircrafts, although the main approach in this case is to limit the weight the aircraft is carrying. But this would be an option with some other risks. If the ship due to weather or other unforeseen events is significantly delayed, it could run out of fuel, if this is limited. If such risks are judged to be tolerable, however, it may provide an effective manner to avoid that hackers turn a ship into a threat to goals far away from the intended route. A battery powered ship will have such limitations in any case.

4.2 *Principles of prevention of threats*

If the ship is completely unmanned, it will be essential to avoid any opportunities any vulnerabilities in the control and communication systems onboard that may be used in a cyber-attack to gain control over the ship. This implies that complete control over the construction, procurement, management, operation and maintenance of autonomous ships without manning of the ship for any purpose is necessary. At all times, no unauthorized organizations nor individuals should get the opportunity to install software or hardware which may provide a "backdoor" into the control system and software available to hackers.

4.3 *Responsibilities*

Even though the probability for a cyber-attack against an unmanned autonomous ship may be low, such events cannot be ruled out in the future, and the design of autonomous ships has to involve a series of risk reducing actions and designs. Requirements to completely non-vulnerable control and communication systems may pose extreme restrictions to the construction, management, operation and maintenance of a completely unmanned ship, perhaps to the extent that the advantage of zero manning by far is overridden by costs increases associated with such restrictions.

4.3.1 *Role of ship owners*

It will be the responsibility of the ship owner who is commissioning the construction of an unmanned, autonomous ship that no alien software or hardware is allowed on board, which may be used in a cyberattack.

This will imply that every aspect of construction, procurement, management, operation and maintenance of such ships is controlled in extreme detail. All suppliers, vendors and component manufacturers and all their personnel will have to be scrutinized in order to ensure that no one has illegitimate purposes. This would be an extreme control system.

In the late 1970s, the possibility to construct nuclear power plants in Norway was considered by specialists and politicians. For a lot of the people who were against, the most fundamental argument was that there would need to be so strong requirements to control of the personnel who would operate and maintain nuclear power plants. Such very strong restrictions and surveillance of personnel were completely unacceptable to many persons.

To prevent successful cyber-attacks to autonomous ships, it will be crucial to maintain control and sufficient quality assurance over the whole software development process. This might become costly and reduce some of the expected cost savings related to autonomous ships.

4.3.2 *Role of designers and ship builders*

It is still the responsibility of designers and ship builders to implement the very strict control outlined above.

4.3.3 *Role of classification societies*

The classification societies will have to provide assurance that no alien software or hardware has been installed at any time during construction. This will require quite extreme housekeeping and control activities. It will not be sufficient to ensure that the high-quality classification societies have stringent requirements, all classification societies (high quality and low quality) need to focus on such threats.

Such assurance will need to maintained also after commissioning, due to software updates, etc. verification of software and software updates therefore becomes even more important and challenging.

4.3.4 *Role of IMO*

It is required to establish very stringent international requirements to control the risk of cyberattacks on autonomous ships. Any ship from anywhere in the world can travel international waters all over the globe and become a threat in very distant waters, provided it has sufficient amount of fuel (or operates on solar power!). All ships will therefore need to follow strict requirements.

It would be expected that the following were high-level IMO requirements for two alternative categories of autonomous ships, with and without manning:

1. Autonomous ships that always require a small crew onboard to operate
   a. Ships to have function which deactivates mechanically external control and replaces it with local control
   b. Ships to have a global power off function which as a last resort gives a dead ship
2. Autonomous ships that may operate without any crew members onboard
   a. Ships to have a function which limits the stored fuel to the distance between ports with a small margin, or
   b. Take steps to ensure fully that nobody has opportunity to install hardware or software that may be used in cyberattacks against the ship.

5 CONCLUSIONS

This article discusses cyber-attacks and its potential threat to autonomous ships. Experience from other sectors are presented and discussed. A hacked autonomous ship may be used as a weapon and ram offshore oil and gas systems, infrastructure systems along the coast, or collide with, cruise ships or oil tankers.

Infrastructure systems along the coast may be considerably more robust against collision impact compared to offshore structures. Typical offshore structures may have a resistance up to 200 MJ, which corresponds to a 5,000 tons ship with a speed of 12 knots, and are thus quite vulnerable.

Keeping a small crew onboard is the most effective risk control action, assuming that the crew may be able to deactivate external control and take over control locally and mechanically.

An option to keeping a small crew onboard is to ensure, as far as possible, that hackers cannot gain access to the control of an autonomous unmanned ship. This implies that there will have to be complete control over the construction, procurement, management, operation and maintenance of autonomous ships.

Another option would be to limit the operational area of an unmanned autonomous ship, for instance, by limiting the available fuel stored onboard. But this would be an option involving some other risk: if the ship due to weather or other unforeseen events is significantly delayed, it could run out of fuel, if this is limited. If such risks are judged to be tolerable, however, it may provide an effective manner to avoiding that hackers turn a ship into a threat for objectives far away from the intended route.

It is required to establish very stringent international requirements to control the risk of cyberattacks on autonomous ships. As ships may travel all over the globe and become a threat in very distant waters, all ships will therefore need to follow strict requirements. There will have to be different requirements to ships which require a small crew onboard than those without any crew.

REFERENCES

Advanced Autonomous Waterborne Applications (AAWA). 2016. "Remote and autonomous ships. The next steps", Position paper. http://www.rolls-royce.com/~/media/Files/R/RollsRoyce/documents/customers/marine/ship-intel/aawa-whitepaper–210616.pdf (Accessed: 2017–02–14).

Bueger, C. 2015. What is maritime security? Marine Policy 53, 159–164.

GOV.UK. 2017. Key principles of vehicle cyber security for connected and automated vehicles, guidance. Department of Transport, 6.8.2017. (Accessed: 12.11.2017: https://www.gov.uk/government/publications/principles-of-cyber-security-for-connected-and-automated-vehicles/the-key-principles-of-vehicle-cyber-security-for-connected-and-automated-vehicles).

IEC61508:2010. Functional safety of electrical/electronic/programmable electronic safety-related systems, Part 1 – General requirements.

International Maritime Organization (IMO), Maritime Safety Committee, 2017b. Maritime Cyber Risk Management in Safety Management Systems (Resolution MSC. 428 (98)).

International Maritime Organization (IMO). 2017a. Facilitation Committee and Maritime Safety Committee. Guidelines on maritime cyber risk management (MSC-FAL.1/Circ.3, 5.7.2017).
ISO. 2011. Road vehicles—Functional safety, ISO26262:2011
Kibar, O. 2017. The car hacker (in Norwegian). Teknologi. Magasinet, Dagens Næringsliv, 5.8.2017.
Kongsberg. 2017. Autonomous ship project, key facts about Yara Birkeland. https://www.km.kongsberg.com/ks/web/nokbg0240.nsf/AllWeb/4B8113B707A50A4FC125811D00407045?OpenDocument
Lyche, K. 2017. The driverless car is already here (in Norwegian). Magasinet, Dagens Næringsliv, 2.9.2017.
Macher, G., Messnarz, R., Armengaud, E., Riel, A., Brenner, E., Kreiner, C. 2017. Integrated Safety and Security Development in the Automotive Domain, SAE Technical Paper 2017–01–1661, doi:10.4271/2017–01–1661.
Maritime Unmanned Navigation through Intelligence in Networks (MUNIN). D9.2. Qualitative assessment. Report, 30.9.2015.
Petroleum Safety Authority. 2017. The handling of incidents associated with ICT and information security and associated barrier management by Statoil, Audit report, PSA, 30.1.2017.
Press, G. Stopping self-driving cars from becoming cybersecurity weapons. Forbes, 19.7.2017 (Accessed: 12.11.2017: https://www.forbes.com/sites/gilpress/2017/07/19/stopping-self-driving-cars-from-becoming-cybersecurity-weapons/#4e49e2a06723)
Rødseth, Ø.J. & Tjora, Å. 2014. "A risk based approach to the design of unmanned ship control systems". In: Maritime-Port Technology and Development—Ehlers et al. (Eds), Taylor & Francis Group, London.
SAE. 2016. Cybersecurity Guidebook for Cyber-Physical Vehicle Systems, J3016_201601
Slovic, M. 2017. Security issues could still crimp the selv-driving car. Electronic Design, 28.6.2017. (Accessed: 12.1.2017: http://www.electronicdesign.com/automotive/security-issues-could-still-crimp-self-driving-car)
Statens Vegvesen. 2013. Sognefjord feasibility study floating bridge (in Norwegian only), Main report 11258–03, Statens Vegvesen, Region Vest, 15.2.2013
Statoil. 2016. Statoil's global risk management including IT Security, lecture by Monica Solem at NTNU Marin Technology Dept, October 2016
Vinnem, J.E. 2013. Offshore risk assessment, 3rd Edition, Springer, LondonUS Nayv, 2001. https://www.history.navy.mil/research/library/online-reading-room/title-list-alphabetically/t/terrorist-attack-on-uss-cole-background-and-issues-for-congress.html
Yu, Z. & Amdahl, J. 2018. Analysis and design of offshore tubular members against ship impacts. Marine Structures. vol. 58.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

An evaluation of the Functional Resonance Analysis Method (FRAM) as a practical risk assessment tool within a manufacturing environment

S. Albery, S. Tepe & D. Borys
RMIT University, Melbourne, Australia

ABSTRACT: The objective of this research was to evaluate a risk assessment process based on the Functional Resonance Analysis Method (FRAM) (Hollnagel, 2012) as a practical tool within a manufacturing environment. Instead of focusing on the activities and events which can cause adverse outcomes, known as a Safety-I approach, this study uses risk assessment for detecting the sources of variability within activities and how this may lead to both negative and positive outcomes within the system; Hollnagel (2012) refers to this as a Safety-II approach. Four examples of work process from an upholstered seat manufacturer were assessed using a risk assessment process involving consultation with workers to determine sources of variability in the 'work-as-done'. The data was mapped onto the FRAM six aspects to envision instantiations. The method provided a means of clearly articulating gaps in the system design impacting safety and productivity. FRAM was found to be an effective mechanism for revealing aspects of variability but required a greater resource commitment over regular risk assessment tools.

1 INTRODUCTION

Despite their wide spread use, risk assessment tools such as risk matrices have documented pitfalls (Gadd et al., 2004). Due to restricted requisite variety, their effectiveness is limited in the identification and control of hazards (Conant and Ross, 1970) and as a result, limit the benefit they provide to both employees and organisations. Further, the use of a linear causal relationship to describe hazards generates a concentration on negative outcomes and lower order controls, limiting stakeholder learning and cross disciplinary engagement (Cox, 2008).

This research proposed to challenge this Safety-I archetype (Hollnagel, 2014) and replace it instead with the perspective of Safety-II (Hollnagel, 2014), principally to look beyond isolated hazard management to the consideration and optimisation of the greater system within which the hazards exist. The objective was to investigate if the pitfalls associated with linear risk assessment methods could be mitigated by using a tool with greater requisite variety.

The research evaluated four work systems within a manufacturing environment. The systems were selected based on work-as-imagined descriptions of characteristics key to both Safety-II and FRAM assessments, being: (i) variability in functions, (ii) the level of acknowledgement and control of that variability, and (iii) the couplings between functions within the systems, as well as (iv) couplings to upstream and downstream systems.

1.1 *Safety-I and Safety-II*

Hazard management is based on the definition of risk, the perception of risk, the tools employed to measure risk, and the respective controls identified as a result of the risk assessment process (IEC, 2009). The type of tools used for risk assessment will also influence the type and number of hazards identified, the controls for management of those hazards, and the capacity for organisational learning (Lundberg et al., 2009).

In line with this practice, the risk matrix is a tool used for the assessment of risks based on the likelihood and consequence of negative outcomes (IEC, 2009). The risk matrix is widely accepted as an industry tool because of its simplicity to use and understand, and it is recommended by regulators such as Australian Governments (Worksafe, 2014).

In contrast to current Safety-I based practice, Safety-II seeks to understand why things go right even when there is variability in the system which would otherwise lead to error states (Hollnagel, 2012b). This creates a focus on 'work-as-done' (the way work is actually done, as opposed to the documentation which describes work-as-imagined) and how the system may adapt to emerging situations to prevent the occurrence of negative outcomes.

The adoption of a Safety-II perspective does not mean that the principles of Safety-I should be discarded. Safety-II provides a different lens to question the way work-as-done is understood and Safety-I is a subset of Safety-II (Hollnagel, 2014). The trade-off of looking at everything that goes right, as opposed to just that which goes wrong, is the need to look at all activity not just specific outliers.

Further to this line of thought, the Safety-II concept recognises that workers adapt to situations that do not function as intended, interpreting the conditions of the environment and adjusting their own output accordingly (Hollnagel, 2012b). Recognising that a system is unlikely to be completely defined or closed, humans play an important role in bringing together elements that could not otherwise manage the variability of an open system (Woods et al., 2010).

1.2 *The Functional Resonance Analysis Method (FRAM)*

The degree of variation that a Safety-I tool such as a risk matrix can describe is insufficient to characterise the degree of variation needing to be understood or controlled in the physical system being interrogated (Hollnagel and Woods, 2005). Conversely, the FRAM may have a greater requisite variety due to its ability to better model the environment being investigated and thus its regulation (Conant and Ross, 1970).

Unlike a risk matrix assessment which only looks at likelihood and consequence, a FRAM assessment requires four steps. The first two steps focus on understanding and defining work-as-done, the third examines emergent system states that result from system variability and the fourth considers managing these unwanted system states (Hollnagel, 2012a).

In conducting a FRAM assessment, the system is divided into key functions which either directly or indirectly affect the outcome of the activity as it would usually be performed. In FRAM, functions have up to six aspects which define their characteristics and how they are coupled within the system: inputs, outputs, preconditions, resources, time constraints and controls. Potential variability is then described within the functions in terms of the way a function would typically vary in normal working conditions, and from where that variability originates.

This information constitutes a FRAM model which may then be used to recognise emergent states or instantiations of the system. When variability leads to system states which are not considered normal activity, the variability is construed as uncontrolled, and resonance is said to have occurred.

By observing the relationship between instantiations and emergent system states, it is possible to identify sources of variability and how they may be managed. These may be tabulated to qualitatively aggregate all sources of variability within an instantiation (Hollnagel, 2012a), thus forming a diagram of the relationship between functions through coupled aspects (Hill, 2015).

From the perspective of risk assessment, the Safety-II objective is to identify and assess the sources of variability within a system, so they may be damped to stop the occurrence of resonant system states which otherwise would generate negative outcomes.

A small automotive manufacturing site was selected which had a range of different material processing and fabrication systems, all involving different degrees of worker and technology adjustments, based on work-as-imagined descriptions. By definition, systems with low technological precision required more human adjustment and where thus expected to contain higher variability. Comparatively, systems with a high technological precision required less human adjustment and where thus expected to contain lower variability.

Based on work-as-imagined descriptions from written procedures, four manufacturing systems were selected which contrasted both technological and human precision. The systems were selected as two coupled pairs so that the upstream and downstream relationship of variability could be observed on the overall systems (Table 1).

System 1A required two workers to move carpet stock through a number of automated stations to convert raw carpet stock into a moulded three-dimensional shape with plastic-welded components. The output of System 1A was fed into System 1B which required another two workers to move the

Table 1. Perceived level of adjustment (precision) in manufacturing activities based on work-as-imagined descriptions of technology and human action; resulting in potential output variability.

		Perceived precision	
System		Technology	Human
1A	Carpet forming and welding	Precise	Imprecise
1B*	Carpet foaming and cutting	Precise	Imprecise
2A	Track welding	Precise	Imprecise
2B*	Track assembly	Acceptable	Acceptable

Precise = no adjustment for success;
Acceptable = approximate adjustment for success;
Imprecise = high adjustment for success;
*System B is downstream of System A.

stock through further automated stations, initially adding a foam backing before waterjet cutting the stock to a required shape for stillaging.

System 2A required a worker to assemble many individual pre-welded steel components into a fixture inside a robotic weld station, producing a welded sub-assembly. The welded sub-assembly of System 2A was then fed into System 2B which required a second worker to construct a finished structure from several different sub-assemblies.

At each of these workstations, four question sets were employed (Albery et al., 2016) to discuss each of the systems with the workers thus collecting the necessary data required to perform FRAM analyses.

Workers, supervisors and quality engineers were asked the four sets of questions. All participants had been in their respective roles greater than a year and were skilled in their roles. The question sets were applied as a semi-structured interview process, focusing discussions on the precision of each of the systems in the workers' own language (Louise Barriball and While, 1994).

Aggregate variability was determined qualitatively from participant responses to the question sets. Alignments and misalignments between work-as-imagined and work-as-done were recorded within the precision of respective aspects.

Notes from the data collection were subsequently transcribed into a template adapted from the FRAM (Hollnagel, 2012a). This qualitative coding formed the basis of the aggregation of variability described for each system in Tables 2 through 5.

For each system the variability associated with normal work (N) was recorded as instantiation "0" for each aspect of each function. Normal work was considered an alignment between work-as-imagined and work-as-done where the required adjustments were considered in the system design.

When the response to the questions revealed either an increase (I) or reduction (R) in variability for the same aspect, this was construed as a unique instantiation and "1, 2, …" was recorded, signifying that a different level of precision was required to manage the variability within the system.

Finally, visualisations of the systems then constructed based on this information.

2 RESULTS

The first instantiation of System 1A, System 1A.0 was considered "normal work". It revealed that the shape of the carpet through the process from start to finish was heavily dependent on its temperature. If the carpet temperature dropped two low at any stage in the process it would become incompatible with the other functions. This placed a reliance on human precision to carefully maintain the system cycle time (Table 2. and Fig. 1.).

Instantiation 1A.1 identified time delays moving the stock between stations in upstream functions; this caused a temperature drop in the carpet

Table 2. System 1A, aggregation of variability.

Variability	Normal	Unexpected		
Instantiation	1A.0	1A.1	1A.2	1A.3
Heat carpet				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	N	I	N
Control	N	N	N	N
Time	N	N	N	N
Move carpet				
Input	N	N	N	N
Output	N	N	N	I
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	N	I	I
Time	N	I	I	N
Form carpet				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	I	N	N
Time	N	N	N	N
Hand trim carpet				
Input	R	N	N	R
Output	R	N	N	R
Precondition	R	N	N	R
Resource	R	N	N	R
Control	R	I	N	R
Time	R	I	N	R
Weld carpet				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	N	N	N
Time	N	I	I	N
Foam backing to System 1B				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	N	N	N
Time	N	I	I	I

N Normal variability.
I Increased variability.
R Reduced or no variability (function not active).

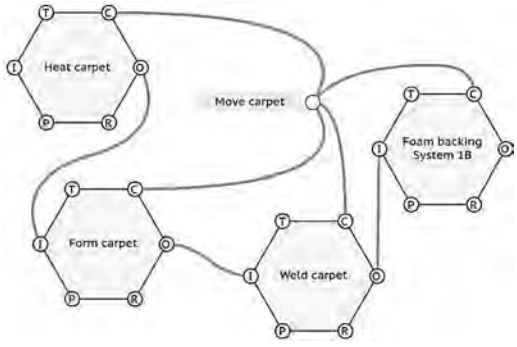


Figure 1. System 1A/Instantiation 0 (1A.0), normal work.

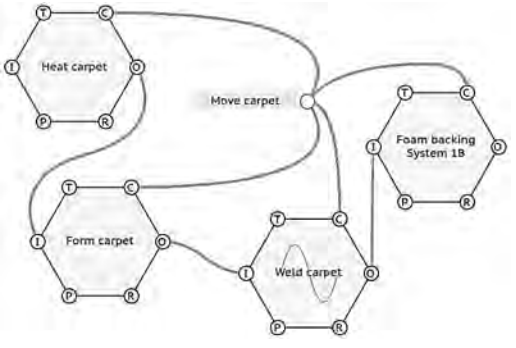


Figure 2. System 1A/Instantiation 1 (1A.1), unexpected variability welding carpet.

stock which led to the carpet border, or perimeter, being geometrically incompatible with the downstream functions.

Specifically, the geometric border change meant that the carpet stock was incompatible with machine fixtures, meaning (i) it could not be correctly welded and (ii) the foam backing material could not be correctly applied as the last stage before entering the next system (Fig.2).

To force the welding and foam backing functions work successfully, workers compensated by hand cutting the carpet borders as a hidden upstream sub-function.

Instantiation 1A.2 identified that approximately 50% of carpet stock was a recycled material which was stiffer than the normal carpet and incompatible with the system temperature range. The recycled material was imprecise and required additional heating time as well as a hidden hand cutting operation to remain compatible with the other functions (Fig. 3).

Instantiation 1A.3 identified that workers were exposed to minor burns manually handling the hot carpet between each of the functions (human, imprecise). This resulted in the carpet becoming soiled and rejected during the foam backing process (Fig. 4).

The first instantiation of System 1B, System 1B.0 was considered "normal work". Workers advised that when the system was functioning correctly, a sufficient carpet temperature was maintained and there was no rejected stock from the previous System 1A (Table 3, Fig. 5). If the workers in System 1A could achieve the 'minimum' temperature meant the carpet stock could successfully pass through each function without hidden worker adjustments, such as also cutting the boarder as described in System 1A (Fig. 2).

Instantiation 1B.1 resulted from a delay in instantiation 1A.1. This resulted in the introduc-

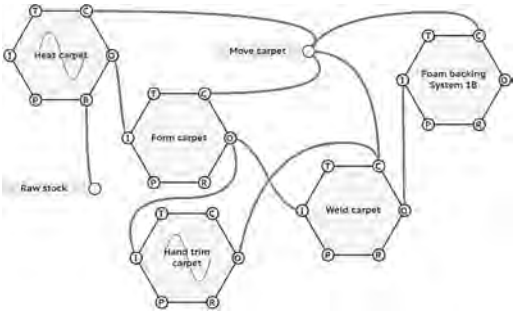


Figure 3. System 1A/Instantiation 2 (1A.2).

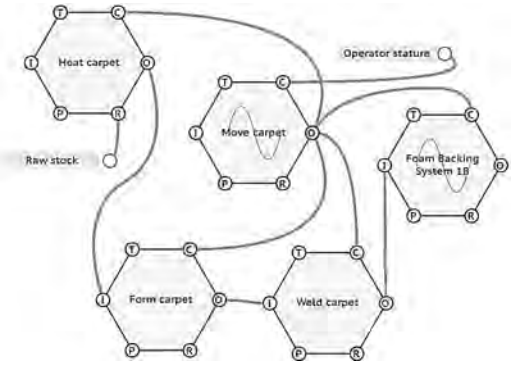


Figure 4. System 1A/Instantiation 3 (1A.3).

tion of a second precise hidden function required to hand cut the carpet for the waterjet function to succeed (Fig. 6).

Instantiation 1B.2 was also caused by the upstream instantiation 1A.1. As the carpet was at the wrong temperature (technology imprecise) when entering instantiation 1B.1 the foam backing

Table 3. System 1B, aggregation of variability.

Variability	Normal	Unexpected		
Instantiation	1B.0	1B.1	1B.2	1B.3
Foam backing from System 1A				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	I	I	N
Time	N	N	N	N
Move carpet				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	N	N	I
Time	N	N	N	N
Waterjet cut carpet				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	I	N	N
Time	N	N	N	N
Hand trim carpet				
Input	R	N	R	R
Output	R	N	R	R
Precondition	R	N	R	R
Resource	R	N	R	R
Control	R	I	R	R
Time	R	I	R	R
Stillage finished carpet				
Input	N	N	N	N
Output	N	N	N	I
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	N	I	N
Time	N	N	N	N

N Normal variability;
I Increased variability;
R Reduced or no variability (function not active).

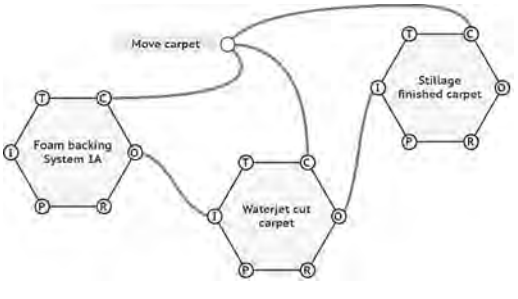


Figure 5. System 1B/Instantiation 0 (1B.0). Normal work.

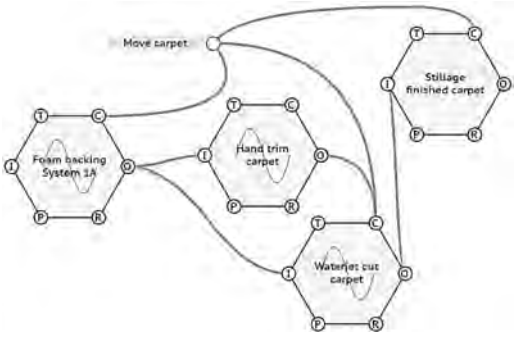


Figure 6. System 1B/Instantiation 1 (1B.1).

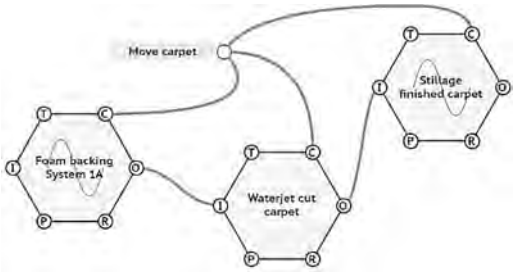


Figure 7. System 1B/Instantiation 2 (1B.2).

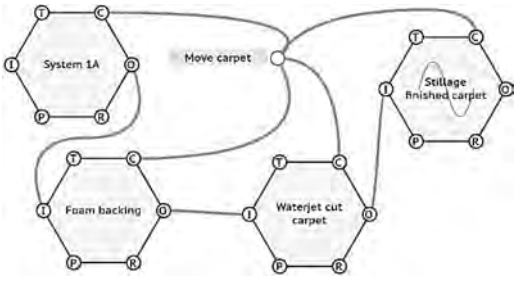


Figure 8. System 1B/Instantiation 3 (1B.3).

was incorrectly applied, and it was rejected when stillaged (Fig. 7).

Instantiation 1B.3 resulted from the carpet becoming soiled when manually handled between functions (human imprecise), and rejected when stillaged (Fig. 8).

The first instantiation of System 2A, 2A.0 was considered "normal work". The processes and assemblies which incorporated the stamping stock were dependent on its inputs being within their correct geometric tolerances (Table 4, Fig. 9).

Instantiation 2A.1 resulted from variability in a previous upstream function (geometric toler-

Table 4. System 2A, aggregation of variability.

Variability	Normal	Unexpected
Instantiation	2A.0	2A.1
Move stock		
Input	N	N
Output	N	N
Precondition	N	N
Resource	N	N
Control	N	N
Time	N	N
Assemble stamping in weld station		
Input	N	N
Output	N	N
Precondition	N	N
Resource	N	I
Control	N	N
Time	N	N
Weld assembly		
Input	N	I
Output	N	I
Precondition	N	N
Resource	N	N
Control	N	N
Time	N	N
Remove welded assembly from weld station TO System 2B		
Input	N	N
Output	N	N
Precondition	N	N
Resource	N	N
Control	N	N
Time	N	N

N – Normal variability. I – Increased variability.

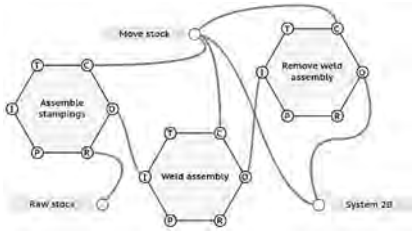


Figure 9. System 2A/Instantiation 0 (2A.0). Normal work.

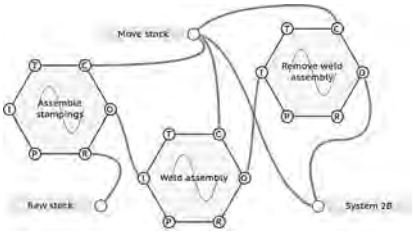


Figure 10. System 2A/Instantiation 1 (2A.1).

Table 5. System 2B, track assembly, aggregation of variability.

Variability	Normal	Unexpected		
Instantiation	2B.0	2B.1	2B.2	2B.3
Move parts FROM System 1A				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	N	N	N
Time	N	N	N	N
Assemble upper and lower parts in station #1				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	I	I	I
Control	N	N	N	N
Time	N	N	N	N
Install hardware in station #1				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	N	N	N
Time	N	N	N	N
Install floating washers in station #2				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	N	N	N
Time	N	N	N	N
Torque hardware in station #2				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	N	N	N
Time	N	N	N	N
Install cable				
Input	N	N	N	N
Output	N	N	N	N
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	N	N	N
Time	N	N	N	N
Check/mark finished assembly				
Input	N	N	N	N
Output	N	I	N	I
Precondition	N	N	N	N
Resource	N	N	N	N
Control	N	N	N	N
Time	N	N	N	N

N – Normal variability; I – Increased variability.

ances, human or technology imprecise), impacting all functions within this system, and those downstream from it (Fig. 10). This meant stock was either rejected as it could not be welded, or was welded and rejected in a downstream function.

The first instantiation of System 2B, System 2B.0 was considered "normal work". The system required workers to manually assembly a number of sub-assemblies into a station, including sub-assemblies from System 2A. These were then bolted together (human precise). The system was reliant on workers continually adjusting their control of the processes and timing between each of the functions to ensure the total aggregate variability was successfully managed (Table 5, Fig. 11).

Instantiation 2B.1 resulted from welded stock from instantiation 2A.1 entering this system. Workers could not confirm if the 2B.1 part was geometrically incorrect (human, imprecise) until checking the final assembly at which point it was rejected and must be returned to the first system input for re-assembly (Fig. 12).

Instantiation 2B.2. resulted from using stock from instantiation 2A.0 that had misaligned hardware. Like Instantiation 2B.1, the assembly was returned to the first system input for re-assembly (Fig. 13).

Finally, instantiation 2B.3 resulted from using stock from instantiation 2A.1 that was welded correctly but was geometrically incorrect. However, the assembly was not detected and rejected initially;

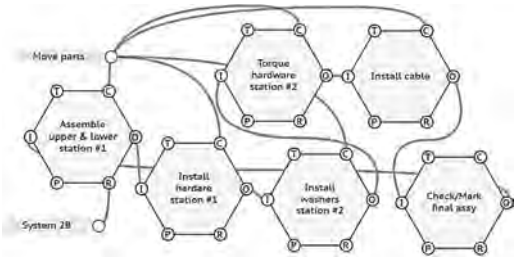


Figure 11. System 2B/Instantiation 0 (2B.0). Normal work.

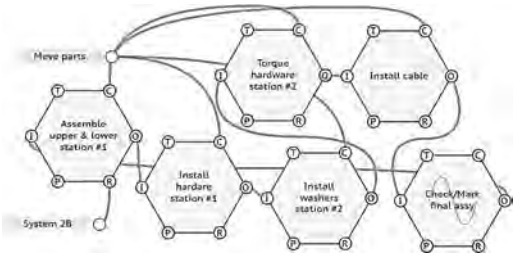


Figure 12. System 2B/Instantiation 1 (2B.1).

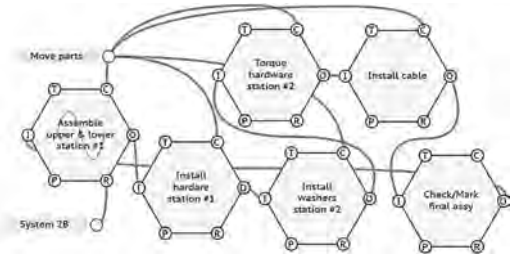


Figure 13. System 2B/Instantiation 2 (2B.2).

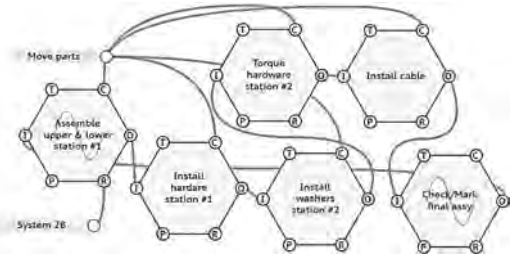


Figure 14. System 2B/Instantiation 3 (2B.3).

so the variation was carried into downstream functions and then rejected in the last function.

3 DISCUSSION

The key learning associated with each instantiation was the identification of variability introduced by upstream plant process where technological precision was expected but absent. Subsequently reactive human precision was introduced to manage variability and maintain successful outcomes (Dekker, 2006). The adjustments to the introduced variability led to the subsequent creation of hazards to workers (Table 6).

Further upstream variability was observed to have downstream effects on quality and productivity, as demonstrated by any increased output variability described within tables of the results section. This was best typified as reject stock output from each of the systems. These learnings contrasted the initial work-as-imagined evaluations of each of the four systems (Table 7).

Acknowledging that each of the systems has a reliance on human precision to manage variability a summary of the main source of variability in each of the systems is provided (Table 8). Interestingly, downstream functions could be improved with no action other than better managing unwanted variability in the upstream functions.

Two specific interventions which would has an impact on the aggregate variability of the systems

Table 6. Potential safety impacts of unwanted variability.

System	Safety	Potential
Instantiation	Hazard	Injury type*
1A.1	Manual handing	Musculoskeletal disorder
	Use knife	Cut/laceration
1A.2	Manual handing	Musculoskeletal disorder
	Use knife	Cut/laceration
	Hot surfaces	Burn
1A.3	Manual handing	Musculoskeletal disorder
1B.1	Using knife	Cut/laceration
1B.2	Manual handling	Musculoskeletal disorder
1B.3	Manual handling	Musculoskeletal disorder
2A.1	Manual handling	Musculoskeletal disorder
2B.1	Manual handling	Musculoskeletal disorder
	Pinch point	Crush/bruise/ laceration
2B.2	Manual handling	Musculoskeletal disorder
	Pinch point	Crush/bruise/ laceration
2B.3	Manual handling	Musculoskeletal disorder
	Pinch point	Crush/bruise/ laceration

*Observed or described by workers.

Table 7. Perceived (work-as-imagined) and evaluated (work-as-done) control of variability (precision) in manufacturing activities based on descriptions and subsequent assessments of technology and human action resulting in actual output variability.

Work-as-imagined		Perceived precision	
System		Technology	Human
1A	Carpet forming and welding	Precise	Imprecise
1B	Carpet foaming and cutting	Precise	Imprecise
2A	Track welding	Precise	Imprecise
2B	Track assembly	Acceptable	Acceptable
Work-as-done		Evaluated precision	
System		Technology	Human
1A	Carpet forming and welding	Imprecise	Acceptable
1B	Carpet foaming and cutting	Imprecise	Acceptable
2A	Track welding	Acceptable	Precise
2B	Track assembly	Imprecise	Precise

Table 8. Actual sources of variability identified from analysis.

System Instantiation	Aspect actual variability	Source of variability
1A.1	Time, Control	Carpet temperature
1A.2	Resource, Time, Control	Recycled carpet raw stock
1A.3	Output, Control, Time	Manual handling carpet
1B.1	Time, Control	Instantiation 1A.1
		Instantiation 1A.2
1B.2	Time, Control	Instantiation 1A.1
1B.3	Output, Control	Manual handling carpet
2A.1	Resource, Input, Output	Geometric variation in raw stock
2B.1	Resource, Output	Instantiation 2A.1
2B.2	Resource, Output	Instantiation 2A.1
2B.3	Resource, Output	Instantiation 2A.1

discussed above are (i) for system 1A, ensuring the system cycle time is compatible with the required raw material forming temperature, and (ii) for system 2A, tightening the tolerances on raw stock prior to welding sub-assemblies.

It is hoped that both of these simple examples serve to emphasise the importance of addressing or damping unwanted variability at its source.

Further to these points, the FRAM analysis of each of the functions aligns each of the sources of unwanted variability with potential hazards to workers. This is significant as traditional risk assessment tools do possess the requisite variety to understand the detail of the systems they seek to control (Gadd et al., 2004). This said, the cost and training required to develop a level of analysis similar to that described within this paper maybe potentially greater than many standard risk assessment tools (IEC, 2009), possibly deterring some safety practitioners.

4 CONCLUSION

Using a Safety-II lens it was found that for each of the systems analysed, work-as-imagined was reliant on worker adjustments for success. These adjustments were conscious actions made in response to introduced variability in each of the systems. From a safety viewpoint, this highlights that collaboration with other stakeholders is required to identify systemic solutions which look beyond the prevention of only localised hazards to those which emerge from the system design (i.e. systemic upstream issues that are not proximal to the worker).

It is concluded that the use of the FRAM provides deeper learnings of system performance in the management of variability as well as the impact of precise and imprecise control. The Safety-II perspective potentially involves greater time and practice to develop more meaningful output than a traditional Safety-I approach; however, this perspective is necessary if it is desirable to increase requisite variety and thus knowledge of the system.

These conclusions are limited by the sample size and findings of the systems considered in this research. However, the case studies presented indicate that a degree of variability and adjustment exists in all of the systems investigated. Further to these learnings, it would be of value to extend this type of analysis to (i) a greater sample size and (ii) coupled systems in other disciplines.

REFERENCES

Albery, S., Borys, D. & Tepe, S. 2016. Advantages for risk assessment: Evaluating learnings from question sets inspired by the FRAM and the risk matrix in a manufacturing environment. *Safety Science,* 89, 180–189.

Conant, R.C. & Ross, A.W. 1970. Every good regulator of a system must be a model of that system. *International journal of systems science,* 1, 89–97.

Cox, L.A. 2008. What's Wrong with Risk Matrices? *Risk Analysis,* 28, 497–512.

Dekker, S. 2006. *The Field Guide to Understanding Human Error*, Ashgate Publishing.

Gadd, S.A., Keeley, D.M. & Balmforth, H.F. 2004. Pitfalls in risk assessment: examples from the UK. *Safety Science,* 42, 841–857.

Hill, R. 2015. *FRAM Model Visualiser (FMV)* [Online]. Available: http://functionalresonance.com/FMV/index.html 2017].

Hollnagel, E. 2012a. *FRAM Modelling Complex Sociotechnical Systems,* Farnham, Ashgate Publishing.

Hollnagel, E. 2012b. A tale of two safeties. *Nuclear Safety and Simulation,* 4, 1–9.

Hollnagel, E. 2014. *Safety-I and Safety-II The past and future of safety mangement*, Farnham, Ashgate Publishing.

Hollnagel, E. & Woods, D.D. 2005. *Joint cognative systems: Foundations of cognative systems engineering*, Boca Raton FL, Taylor & Francis.

IEC 2009. Risk management - Risk assessment techniques. *(IEC/ISO 31010)*. Retrieved from http://www.saiglobal.com.

Louise Barriball, K. & While, A. 1994. Collecting Data using a semi-structured interview: a discussion paper. *Journal of advanced nursing,* 19, 328–335.

Lundberg, J., Rollenhagen, C. & Hollnagel, E. 2009. What-You-Look-For-Is-What-You-Find–The consequences of underlying accident models in eight accident investigation manuals. *Safety Science,* 47, 1297–1311.

Woods, D.D., Dekker, S., Cook, R., Johannesen, L. & Sarter, N. 2010. *Behind Human Error,* Farnham, Ashgate.

Worksafe 2014. Guidance note: Safety assessment of a major hazard facility. *In:* WORKSAFE (ed.).

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Evaluating approaches for hazard identification for the inclusion in a safety assessment framework for efficient transport

Ø. Skogvang, R.K. Opsahl & S. Solibakke
Safetec Nordic AS, Norway

P. Karpati & A.A. Hauge
Institute for Energy Technology, Halden, Norway

T. Sivertsen
Bane NOR SF, Oslo, Norway

B.A. Gran
Institute for Energy Technology, Halden, Norway
Norwegian University of Science and Technology (NTNU), Trondheim, Norway

M.A. Lundteigen
Norwegian University of Science and Technology (NTNU), Trondheim, Norway

ABSTRACT: This paper presents the experiences from applying hazard and operability analysis (HAZOP) as support for establishing the safety requirements specification of a new safety-related railway application. The new railway application is a software based system for securing work areas, meaning it prevents railway traffic in areas along the track allocated to maintenance. The experiences are collected within the Safety Assessment Framework for Efficient Transport (SafeT) project managed by Bane NOR. Bane NOR is the government agency that owns, operates and develops the Norwegian railway infrastructure. The objective of the SafeT framework is to offer a systematic, reusable way for creating system wide conceptual design models and based on them, creating a common risk model, which in turn will facilitate safety assessment, establishing the requirements specification, and safety demonstration of the system under consideration. The experience collected on applying HAZOP is done through two workshops with different formats on the documentation. The objective was to collect guidance on how HAZOP can be supported in the SafeT framework.

1 INTRODUCTION

The project "Safety Assessment Framework for Efficient Transport" (SafeT) aims at developing a framework that supports the implementation of EN 50126 (CENELEC, 2017) and thereby of the Common Safety Methods for Risk Assessment (CSM RA) (EU, 2013) in the railway industry, in particular how the railway infrastructure may support efficient transport.

This paper presents ongoing results from the case studies, while the results from the modelling is presented in another paper (Karpati et al, 2017). Figure 1 illustrates which phases of EN 50126 that is within the scope of the current SafeT work and both papers, annotated by a dark grey rectangle. Some of the related work (chapter 2) is therefore relevant for both papers, and the case (chapter 4)

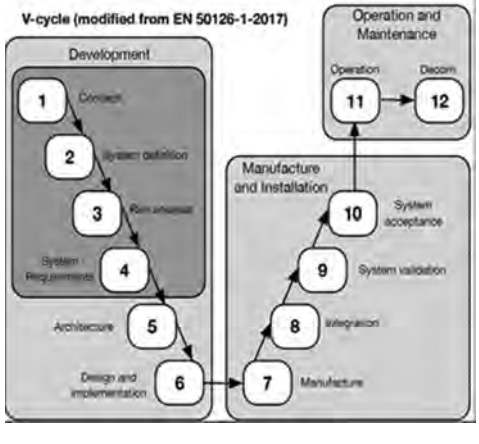


Figure 1. Scope of paper and relationship to EN50126.

is applied throughout the SafeT project. The aim of the paper is to show differences, advantages and disadvantages for two different approaches for hazard identification, applied on the new safety-related railway application.

The current focus in the SafeT project is on the development phases 1 to 4 of EN 50126. In these phases of a systems life cycle, Bane NOR takes a lead role in the development while successive development phases to a large extent are outsourced. The SafeT framework intends to support the development of the core artefacts within the system life cycle. In the early stages of the life cycle, in the part of the framework that concerns the in-house conceptualisation, the core artefacts are: 1) the conceptual system design model; 2) common risk model; and 3) requirements specification.

The main objective of the SafeT framework is to offer a systematic, reusable way for creating system wide conceptual design models and based on them, creating a common risk model, which in turn will facilitate the safety assessment and safety demonstration of the system in focus, throughout the system's lifetime.

2 RELATED WORK

International safety standards, such as EN 50126, provide requirements and guidance on how to carry out the assessment process. Although most safety standards often view the safety of a system as a function of the reliability of its components, little guidance is provided on how to derive safety requirements and acceptable risk for components whose failure rates are not known. Particularly, it is often difficult to derive safety requirements for logical components such as the software. The problem can be formulated from a consideration of the following two important tasks in the development of safety critical systems: (1) *establishing the requirements to the system*, and (2) *ensuring that the system fulfils these requirements.* The safety requirements should be established through risk assessment and hazard analysis, and fulfilled through the use of techniques and measures adequate for the risk level. The framework proposed in the project has much of its inspiration from theoretical aspects of international safety standards such as IEC 61508 (IEC 61508). The novel part of the framework is fivefold: reusability, modularity, unification, transparency and argumentation.

In the following, a number of past projects that relate to the topics of SafeT are briefly introduced. However, most of them relates to the need of establishing models and providing support for a safety case, see related work presented in the other Safe-T paper (Karpati et al, 2017).

The EU funded project MODSafe provides a risk analysis method purposed to combine potential hazards, safety requirements and functions, and link these elements to a generic functional, and object structure of a guided transport system. ASCOS (Roelen, 2014) focused on safety and certification of new aviation operation and systems, and included among other advices on methods and tools for safety based design. ModelMe! (Falessi, 2011) provides a tool-supported traceability framework where the tool for example automatically extracts the safety-related slices of SysML design models (SysML).

The AltaRice Language (Griffault, 1998) is an object-oriented modelling language dedicated to performance evaluation of complex systems. The main motivation for its creation was the difficulty to design, to share and most importantly to maintain safety and reliability models such as fault trees, event trees, Markov chains or stochastic Petri nets. The application and further development of the language is a continuous research activity at NTNU (Legendre, 2017).

Of relevance is also CORAS (Lund, 2011; Gran, 2004) which provides a methodology for model-based risk assessment integrating aspects from partly complementary risk assessment methods and state-of-the-art modelling methodology.

The SafeT project has also reviewed a number of ongoing and past industrial experiences among the project partners related to the use of design and risk models to facilitate the safety assessment and demonstration of complex systems. Some of the challenges observed in these projects have also been reported earlier within aviation (Gran, 2007). Finally, the CHASSIS method (Raspotnig, 2018) utilizes UML use cases and sequence diagrams with HAZOP guidewords to integrate safety and security considerations for early requirements determination.

3 APPLYING DIFFERENT APPROACHES FOR HAZARD IDENTIFICATION

3.1 *The role of the hazard identification*

There may be a number of different motivations for performing a hazard identification. Among them are avoiding loss of value, life and property, optimizing performance and reducing costs. The motivation for studying hazard identification in the SafeT project is to make sure that relevant hazards associated with development and use of software are evaluated, risk mitigation is in place, and the methods used for hazard identification are applicable and useful, with a basis in case studies that are carefully selected together with Bane NOR.

The purpose and method of a hazard identification and operability study (HAZOP-study) is well described in the literature, for example in *Risk assessment* (Rausand, 2011) and *IEC 61882:2016 (HAZOP studies)* (IEC 61882). The hazard identification and operability study is performed by a group review using structured brainstorming to identify and assess potential hazards. The group of experts starts with a list of tasks or functions, and next uses keywords such as none, reverse, less, later than, part of, more. The aim is to discover potential hazards, operability problems and potential deviations from intended operation conditions. Finally, the group of experts establishes the likelihood and the consequences of each hazard and identifies potential mitigating measures. The analysis covers all stages of project life cycle. In practice, the name HAZOP is sometimes (ab)used for any "brainstorming with experts to fill a table with hazards and their effects". Many variations or extensions of HAZOP have been developed.

Hazard identification can be defined as the process of identifying and listing the hazards and accidents associated with a system (DEF-STAN 00-56, 2007). There are numerous different definitions of the term *hazard* described in standards and the literature. In the following, we will combine the definitions used in EN 50126 and EN 50129 and define hazard as "*a physical situation or a condition that can lead to an accident*".

3.2 *Hazard identification in the RAMS lifecycle*

Throughout the European Union, railway signalling and interlocking projects are carried out on the basis of the CENELEC standards EN 50126 (CENELEC 2017), 50128 (CENELEC 2011) and 50129 (CENELEC 2003). The set of standards provide a consistent, European approach to the management of reliability, availability, maintainability, and safety, denoted by the acronym RAMS. In order to demonstrate that a technical system is safe to take into use and suitable for its intended application, the CENELEC standards require that the system under consideration is described and analysed in its intended context, in particular with respect to its relationship to hazards that can occur in this context and how these hazards can be controlled through the system design. This requires good models of both system design and risk that capture the relations between the different system levels and between hazards, causes, barriers, accidents, and consequences. Of particular importance to the safety demonstration is the utilization of common risk models that include the results from the hazard identifications at the different system levels, from an overall railway system down to the separate subsystems (Sivertsen, T. 2016). The use of models to support the safety management is central to SafeT, which therefore focuses on criteria for the choice of modelling techniques and how they can be combined, adapted and further developed to satisfy the modelling needs. These needs are associated to the analyses at the different system levels and its context, the risk associated to the application, and the requirements established to control this risk.

Hazard identification, operability studies, analysis and evaluation of the risks are key activities in phase 3, but they are also relevant for all the following RAMS-phases, shown in Figure 1, and in accordance with 50126-1 (CENELEC 2017):

#	Phase
1	Concept
2	System definition and operational context
3	Risk analysis and evaluation
4	Specification of system requirements
5	Architecture and apportionment of system requirements
6	Design and implementation
7	Manufacture
8	Integration
9	System validation
10	System acceptance
11	Operation and maintenance
12	De-commissioning and disposal

As part of continuous improvement work as described in the ISO 9000-family of standards (ISO 9001), identification and evaluation of potential hazards should also be done as a continuous activity throughout the system's whole life cycle. For all steps and phases, there may be numerous hazards that can compromise the RAMS performance of the system.

4 CASE EXAMPLE DESCRIPTION

4.1 *Introduction to the case example of securing work areas*

The introduction of axle counters for train detection necessitates a new solution for securing work areas. The current solution, on track sections without axle counters, is to use a contact magnet to induce a short circuit in a manner similar to how an axle of a train induces a short circuit and thereby is detected. The short circuit induced by the contact magnet triggers a state change in the interlocking that prevents the train dispatcher from locking routes through the affected section until the contact magnet is removed by the safety guard.

In the proposed solution for securing work areas (see Fig. 2 and Fig. 3), a safety guard uses a

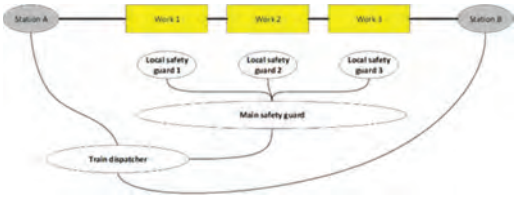


Figure 2. Work areas and roles.

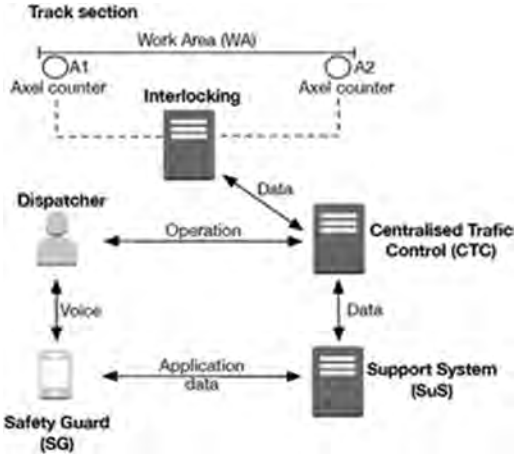


Figure 3. The securing work area case.

smartphone to interact with the train dispatcher. Besides allowing voice communication with the train dispatcher, the smartphone also contains a dedicated application with functionality to manage the securing and releasing of work areas.

In the Norwegian infrastructure, the train detection has usually been performed with different variants of track circuits. Axle counters were introduced in the infrastructure just a few years ago, and gradually replace the existing track circuits.

Irrespectively of the train detection system used, there is a need to protect workers along the track from trains unintentionally moving into the work area. A work area is a track section (possibly more than one track) that can be disposed for work, without any trains entering or leaving the area (Figure 2). The work area and the surrounding tracks can be protected by points, derailers, main signals, shunting signals, and regulations.

While the train dispatcher in either case has the possibility to block the work area, a basic safety principle in Norwegian railway operation is that the workers should be able to prevent the train dispatcher from unblocking the work area before the work is finished. Basically,

- the workers' position must be correctly identified;
- the correct work area must be effectively blocked; and
- the work area must not be unblocked prematurely.

One of the challenges with the introduction of axle counters has been that the existing methods to secure the work area no longer worked. This applies both to the correct identification of the workers' position and to the barriers against hazards caused by premature unblocking of the work area. Since a track circuit short-circuits when a train is present in the track section, the presence of trains can be imitated by short-circuiting the track circuit with other means, viz. the *contact magnets*. In this way, the workers along the track can indicate their position to the train dispatcher, who can block the section to prevent trains from entering. The contact magnet furthermore works as a barrier to hazards caused by premature unblocking of the area (trains entering the work area), since the track section is considered occupied by the interlocking.

The current solution in Norway for securing the work area when axle counters are used for the train detection involves removing a physical key for the relevant work area from its lock when the train dispatcher has blocked the work area and released the key. The train dispatcher is prevented from unblocking the work area until the safety guard has put the key back. While this certainly works, the solution is both expensive and inefficient due to the need for additional physical equipment along the track, and physically interlocking this with the signalling system. There is therefore a need for a system that can replace the current use of physical keys.

This is the background for the invention of the concept described in the next section.

4.2 *A concept of a new solution for securing work areas*

The concept involves the development of a software based system for safe interaction and supervision related to the protection of maintenance workers from accidents caused by the interference with the railway traffic. The solution is planned to require no other physical measures in the infrastructure than simple marking along the track in terms of a barcode or QR-code identifying the work area.

The proposed solution for securing work areas (see Fig. 3) consists of a software-based solution whereby a safety guard uses a smartphone to interact with the dispatcher. Besides allowing voice communication with the dispatcher, the smartphone also contains a dedicated application with functionality to manage the securing and releasing of work areas.

The safety guard identifies the work area by scanning the code on site. This identification of the work area is required at certain steps in the operation. Some of the characteristics of the functionality are:

- The main safety guard selects the functions from the application on his smart phone.
- Scanning the work area identifies both the safety guard and the work area.
- The application communicates with the support system, which communicates with the CTC and other applications.
- The support system supervises the protocol associated to each function.
- The support system supervises the secured work areas, and prevents the train dispatcher from prematurely unblocking the work area.

The solution gives several advantages, like less intervention in the infrastructure, no physical key to be kept and replaced, more convenient inspection, improved safety locally, additional functionality, larger flexibility, and simpler maintenance.

For simplicity, the interfaces between the operational support staff and the other roles are not shown in the figures. The operational support is not mentioned in the descriptions of the main functions, but a separate analysis of the support functions should be part of a complete analysis of the system. The responsibilities of the operational support include

- correcting errors or operational problems;
- keeping the support system updated with respect to information about known faults or operational problems; and
- keeping the support system data updated.

For the purpose of the risk assessment at the railway system level, all the functions can be described by considering only the interfaces between the applications and the safety guards, between the applications and the support system, and between the support system and the CTC.

Twelve main functions have been specified for the system (T. Sivertsen, 2014):

1. Log in: Logging into the system, thereby getting access to the other main functions.
2. Log out: Logging out of the system, thereby being prevented from using other functions before a new login.
3. Join: Enrolling in a work area, thereby preventing the safety guard in charge to release the work area.
4. Resign: Withdrawing from a work area, thereby allowing the safety guard in charge to release the securing of the work area.
5. Secure: Securing a work area, thereby preventing the work area from being unblocked.
6. Release: Releasing a secured work area, thereby allowing the work area to be unblocked.
7. Set time: Setting the time available for work in a work area, thereby allowing an automatic countdown of the time available.
8. Time: Reading the time available for work in a work area, thereby facilitating management of work in the work area.
9. Status: Reading the status a work area, thereby facilitating management of work in the work area.
10. Takeover: Requesting takeover of responsibility for a work area.
11. Full takeover: Requesting takeover of another safety guard's responsibilities.
12. Overview: Overview of the work areas the safety guard is in charge of or enrolled in.

For each of these functions there is a list of tasks that is performed by one or more of the involved actors in the process of securing and releasing the work areas, as showed in Figure 3.

5 TESTING TWO ALTERNATIVE APPROACHES FOR HAZARD IDENTIFICATION ON THE CASE

In order to evaluate the importance of the system description in relation to the result of an analysis, two alternative system descriptions were applied in two different HAZOP workshops with different participants.

The aim was to evaluate if different ways of presenting the system would result in different findings. In the first workshop, the basis for preparation and discussion was a graphical model of the system, while the other used a textual description. The same type of competence was present in both workshops, however, not represented by the same individuals.

The participants in the two workshops were mainly academics, with theoretical knowledge of the new and current system and of different approaches for risk assessment. There were no participants with practical experience with using the existing system for securing work areas, or other roles involved when performing such tasks. Most participants were familiar with the railway infrastructure in general and had experience with the HAZOP technique. All participants in the workshop where familiar with the new concept for securing work areas, through either the graphical representation of the system or the textual description.

5.1 *HAZOP based on a graphical model*

As preparation, a description of the case utilizing SysML diagrams with limited text and explanation

of the modelling language was sent out to the participants one week before the workshop. In the workshop the participants had many questions outside the scope covered by the model, there were also questions related to the meaning of some of the modelling symbols. During the workshop, an example of the physical outline was drawn ad hoc as illustration, and it was used a lot in the discussions. The facilitator had guidewords on hand, but they were not applied actively, as the participants constantly came up with new questions related to system achitecture or potential problems. The HAZOP resulted in the identification of two hazards, a large number of potential hazards and potential situations leading to down-time. The large number of the identified potential hazards was due to uncertainty and lack of detailed system procedures.

5.2 *HAZOP based on a textual description*

A textual description of the case was provided in advance as input to the HAZOP workshop (a summary of the textual description is given in chapter 4.1). The participants had one week to familiarize themselves with the textual description of the system before the workshop.

The following guide words were used in the meeting: early, late, before, after, wrong place, missing and wrong. The guidewords were not used actively for each function, but were presented on a separate marker board throughout the whole workshop. Each of the main functions was discussed in the HAZOP workshop, in accordance to the order given in chapter 4.2.

5.3 *Experiences from testing the two approaches*

A textual description is, compared to a model description, a well-known and common way of presenting systems for most people. A textual description may therefore be less time consuming to understand and is easy to present in a meeting. However, the textual description was not detailed enough to present the system logic and all the preconditions in depth. Hence, an illustration including the sequence of main functions and roles involved in each function was made by one of the participants in the workshop.

The illustrations were found to be useful complements, and indicated that the textual descriptions alone were not able to provide sufficient information. In specific, it was found that understanding the correct sequence of functions performed by the different roles was critical to the hazard identification, and this was not easily covered and captured by the textual description.

Constructs in models can become complex and thus their visualization as well. According to the experiences in the workshop based on graphical models, the models became difficult to understand after a certain level of visual complexity (e.g. when it is no longer possible to present the whole system in *one* single and readable screen diagram), it becomes more difficult to find support in the visual representation). One specific related problem was following the flow of logic in diagrams when branches were involved. Modularization of the visual representation added to the textual descriptions (if meaningfully possible) might help here.

Both workshops included a physical description in addition to the text or models provided on beforehand. This suggests that a physical outline diagram could be part of the models, or an addition to textual descriptions. Another consideration is that modelling or describing specific, representative cases (e.g. application of the planned system at a specific work area) might be a necessary supplement to the initial descriptions of the planned system. In our case, a specific, representative train station could be considered.

Even though participants in both workshops helped identifying unclear and missing parts, both workshops pointed to a number of potential hazards due to uncertainty about how the system was intended to work. Some of these details were contained in only one of the descriptions, but a number of descriptions were missing in both workshops, for example: preconditions of the main functions of the securing work area app, defined terminology and roles, description about the old and current solutions etc. A question related to this is whether the workshops would have been able to process and utilize the information requested by the participants. This needs to be taken into account when considering the use of HAZOP. In particular, there is a need to find models supporting the balance between the two considerations: giving sufficient descriptions, but not drowning the participants in details.

Based upon one workshop with models, we cannot conclude on the question of whether the model-based description prepared is practical for the hazard identification. There were, as described above, many other influences in the workshop independent from the modelling. However, it is clear that SafeT will need to prepare guidelines on how to use HAZOP in combination with specific SysML diagrams. Another question is if other models could have provided the same.

The two workshops came up with the same hazards. The only differences lay in how they were identified in the two workshops. This is in accordance to what one should expect. Since the textual and graphical descriptions were based upon the same source of knowledge within Bane NOR, differences in the assessment would typically point to

flaws in one of the descriptions. Another reason for having the same results is that the two workshops had rather homogenous group of knowledge and experiences. None of the groups had participants with practical experience, such as train dispatchers or safety guards. This is also illustrated by the high number of potential hazards. It is assumed that by having additional competence in the workshop, some of these potential hazards would be closed as not possible, while others would be confirmed. One interesting observation is that most of the potential hazards are not closed by just adding the graphical and the textual description. The uncertainty lies in what is not presented in any of the two workshops. If the experiment would have included only one HAZOP, we could falsely have concluded that the solution was simply to add the graphical or the textual description.

Both for the model based and the text based descriptions there is a need to supplement the descriptions by all the following different visualisations, to compensate for their inherent advantages and disadvantages:

- High-level visualisations—everything on one drawing.
- Modularised visualisations—to explore the details where and when needed.
- Sequences—to get necessary understanding on the order and timing of activities and tasks.
- Visualisation of interactions: man–machine/technology–organisation–environment.

A conclusion from this is that a better coverage of relevant details for the hazard identification could have been included in the model and the diagrams. This could also have been achieved by a preparatory workshop focusing on eliciting such information, or by involving a RAMS expert in the modelling beside the system modeller and the system owner.

There are several sources of uncertainty in the conclusions relating to relevant hazards identified in the two workshops. The uncertainty related to the sum of competencies covered by the participants in the workshop is crucial. That means that whatever approach, the sum of competencies is of great importance. It is not possible to compensate for lack of competence by choosing the other approach, or adding more time for each participant's preparations.

Applying these two approaches to the case identify basically the same hazards. This means that the conclusion is not that one of the approaches is preferable. On the contrary, both approaches give different nuances and different perspectives, resulting in a broader risk picture, which may be useful when it comes to communicating, evaluating and mitigating the risk.

How sensitive these findings may be to the chosen case is not investigated. This means that if the case was a totally different one, we do not know whether the two approaches would end up with similar hazards. Anyway, the findings in the HAZOPs from the two approaches, and the findings from the comparison of the two approaches, both indicate that the case is complex enough for an experiment like this.

When introducing new technologies or new applications of existing technologies, it is important to assess the risk by using not only one approach, but rather apply different approaches to get a broader understanding of the potential hazards.

6 CONCLUSIONS

In this paper we have elaborated on the experiences on using a graphical model presented as SysML diagrams in comparison with an ordinary textual description as a basis for hazard identification.

The model-based description is a practical and useful supplement for the hazard identification activities, but the HAZOP workshops point out that the use of SysML models requires good preparation of the HAZOP. SafeT will need to prepare guidelines on how to use HAZOP in combination with specific SysML diagrams. The participants should be familiar with such modelling to benefit from the models. A textual description is a mode of communication that most of the potential participants in the HAZOP workshop will be familiar with and trained in on beforehand. Graphical models, pictures and drawings are necessary and useful supplements for getting a broader understanding on the case that is subject for analysis.

ACKNOWLEDGMENT

The SafeT project is funded by the Norwegian Research Council (project number 257167/O80) and Bane NOR, and has participation by Bane NOR and IFE, also participation from Indra Navia AS, Avinor, Solvina AB, Safetec Nordic AS, NTNU, VTT and Beijing Jiaotong University.

REFERENCES

AltaRica, https://altarica.labri.fr/wp/ (Accessed Apr 10, 2017).

ASCOS project: https://www.ascos-project.eu/ (Accessed Apr 10, 2017).

CENELEC, EN 50126-1:2017. Railway applications – The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS).

CENELEC, EN 50128:2011. Railway applications – Communication, signalling and processing systems – Software for railway control and protection systems.

CENELEC, EN 50129:2003. Railway applications – Communications, signalling and processing systems – Safety related electronic systems for signalling.

DEF-STAN 00-56, 2007. Safety management requirements for defence systems, parts 1 and 2.

EU, 2013. EU COMMISSION IMPLEMENTING REGULATION (EU) No 402/2013 of 30 April 2013 on the common safety method for risk evaluation and assessment and repealing Regulation (EC) No 352/2009.

Falessi, D., Nejati, S., Sabetzadeh, M., Briand, L. and Messina, A. 2011. Safeslice: a model slicing and design safety inspection tool for SysML. *In SIGSOFT FSE, pages 460–463.*

Gran, B.A., Hauge, A., Winther, R., Lavik, L. 2007. Some challenges and solutions assessing the safety of ATM systems, *In Risk, Reliability and Societal Safety, ESREL 2007, Aven & Vinnem (eds), Taylor & Francis Group, pp 2113–2120.*

Gran, B.A., Fredriksen, R., Thunem, A.P.-J. 2004. "An Approach for Model-Based Risk Assessment". *In Proc. Computer Safety, Reliability, and Security (LNCS 3219). Heisel, M. Liggesmeyer, P., Wittmann, S. (Eds). Pp 311–324.*

Griffault, A., Point, G., Rauzy, A., Signoret, J.P. and Thomas, P. 1998. The AltaRica Language. *In Lydersen and Hansen and Sandtorv ed., Proceedings of European Safety and Reliability Conference, ESREL'98.*

IEC 61508:2010. Functional safety of electrical/electronic/programmable electronic safety related systems.

IEC 61882:2016. Hazard and operability studies (HAZOP studies) – Application guide.

ISO 9001:2008 Quality management systems – Requirements.

Karpati, P., Hauge, A.A., Sivertsen, T., Gran, B.A. 2017. Evaluating models for the inclusion in a Safety Assessment Framework for Efficient Transport. *To be presented at ESREL 2018.*

Legendre, A., Lanusse, A., Rauzy, A. 2017. Toward model synchronization between safety analysis and system architecture design in industrial contexts. *In LNCS 10437.*

Lund, M.S., Solhaug, B., and Stølen, K. 2011. Model-Driven Risk Analysis. The CORAS Approach. *Springer-Verlag Berlin Heidelberg.*

MODSafe project: http://www.modsafe.eu (Accessed Apr 10, 2017).

Raspotnig, C. 2014. "Requirements for safe and secure information systems". *PhD thesis.*

Raspotnig, C., Karpati, P., Opdahl, A. L. 2018. Coordinated Assessment of Software Safety and Security – An Industrial Evaluation of the CHASSIS Method. *To be published in in Journal of Cases on Information Technology (JCIT) Vol. 20, Is. 1.*

Rausand. 2011. Risk Assessment: Theory, Methods, and Applications, ISBN: 978-0-470-63764-7.

Roelen, A.L.C., Verstraeten, J.G., Speijker, L.J.P., Bravo Muñoz, S., Heckmann, J.P., Save, L., and Longhurst, T. 2014. Risk models and accident scenarios in the total aviation system.

Sivertsen, T. 2014. Concept of a New Solution for Securing Work Areas, EHPG 2014, Røros, Norway.

Sivertsen, T. 2016, Validation of safety requirements within railway signalling and interlocking, Enlarged Halden Programme Group Meeting, Scandic Fornebu Hotel, Norway, 8th – 13th May, 2016.

SysML, http://www.omgsysml.org/ (Accessed Dec 12, 2017).

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Analysis of the risk of pipe breaks based on hydraulic model

E. Bartkiewicz & I. Zimoch
Institute of Water and Wastewater Engineering, Silesian University of Technology, Gliwice, Poland

ABSTRACT: The Water Supply System (WSS) distributes and supplies water to customers, which is used for consumption, production, maintenance of sanitary conditions and extinguishing fires, i.e. activities necessary for life. For this reason, hydraulic conditions of the WSS and the water quality must be maintained. One of the most common problems encountered in underground infrastructure is pipe failure, which has a major impact on WSS performance and water quality. To avoid such incidents, a safety plan should be developed. The WSSs safety plans requires some steps for risk assessment and decision-support systems. This system should use all the information resources stored in the databases. Currently, water supply companies have a number of network monitoring systems, however these are more warning rather than preventing systems. The risk is defined as undesirable situations that may occur, and therefore the preventing system should be more complex and based on historical operational events. The basic tools for building such a system are mathematical models that simulate the operational states of the WSS at different time intervals and during accidently situations. There are two kinds of models—hydraulic to simulate operating condition of the water pipe network such as pressure and velocity of water flow and- quality model, that show changes in water quality at different time intervals. The article presents a hydraulic model of a real WSS, created in WaterGEMS. This software uses historical data of pipe breaks in Pipe Break Analysis function, to compute a pipe break score for each pipe. The article presents the results of the analysis and risk assessment, presented in matrix form.

1 INTRODUCTION

Water Distribution Systems (WDS) are technical systems with a spatial structure, which the purpose is the continuous supply water to customers with appropriate quality, quantity and pressure. Unfortunately, water supply companies are increasingly faced with aging and unreliable network infrastructure, which contributes to water supply interruptions. These interruptions may be caused by pipe damages, pump failure and equipment failures [1]. Pipeline failures are a common problem depending on many internal and external effects that cause the pipe break. These effects include pipe factors (material, age and diameter), operational factors (pressure, corrosion, external stresses) and environmental factors (temperature, rainfall, soil conditions) [2, 3, 4, 5, 6]. The Figure 1 shows a few examples of conditions that may cause pipes break. Pipes failures cause economic losses (leakages) and threats to consumer health [7, 8, 9]. In small systems (often branched), pipe failures cause interruptions in water supply, which causes risk in public health places (hospitals, hotels, kinder gardens, schools), whereas in large systems reveal itself as water leaks. Determining the type of recipient is an important element in risk assessment. Every year, 32 to 48 billion cubic meter of water is lost all around the world through leaks that correspond to the total annual cost of water loss of more than 14 billion US dollars [10, 11]. Through cracked pipes, the system gets contaminated soil (by pesticides or animal and human waste) that causes water borne diseases outbreaks [12, 13]. For these reasons, the analysis of pipe failures should be included in the

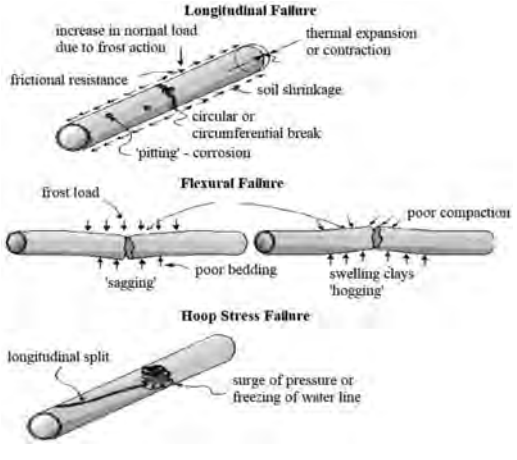


Figure 1. Causes for pipe break [2].

Water Safety Plans, as one of the elements enabling the determination of water contamination areas. World Health Organization (WHO) wrote in the Water Safety Plans: "*Understanding the nature of sources of contamination and how these may enter the water supply is critical for assuring water safety*" [14], which means that every risk analysis is crucial to the protection of human health.

There are two categories of pipe failure modeling: physical and statistical. Physical methods are designed to describe the physical mechanisms underlying the failure of pipelines. However, the mechanism of physical pipe cracks is often very complex and difficult to determine accurately. The reason for this is the fact that the pipes are buried, and the records from the failure are insufficiently accurate. Acquiring data to determine the physical causes of failure requires expensive investments, which is why statistical methods are more often used. Statistical methods use different characteristics of water networks, usually data that can be easily obtained, as pipe age or the number of failures [2]. The aim of statistic approach was to determine the deterioration of the pipes condition and creating a forecasting model to assess the risk of failure [11, 15]. Statistic methods can be divided into three categories: deterministic, probabilistic multi-variates and probabilistic single-variate. Models considering uncertainties related to input data are probabilistic models, otherwise, these are deterministic models, while models based on more parameter then pipe age are multidimensional [16]. The deterministic methods ware described by Shamir and Howard as a prediction model that relates a pipe's breakage to the exponent of its age [17]. Examples of probabilistic methods include, among others models defining the time of failure occurrence, such as the Cox proportional hazard model, accelerate failure models [18]. Table 1 presents sample models of pipe failure estimation. Analysis of the risk of pipe breaks is related to the concept of reliability. Reliability of WDS systems is defined as a property that relies on the system's ability to perform its functions under certain conditions of existence and exploitation and within the assumed time. Which means that the reliability of WDS is determined as the probability of supplied all demand nodes [4, 19, 20]. Based on the actual behavior of the network, study the reliability of water distribution network is planning to accurate operation and management of WDS. Taking into account pipe failure prediction models and models of WDS reliability, a Water Safety Plans supporting system can be created.

Forecasting pipe failure requires the collection and analysis of historical data. One of the most commonly used system to collect data is the GIS (Geographic Information System) database, which enables the integration of multiple data sources (including location data of dangerous locations, geodemographic features and vulnerable populations), the use of spatial analysis techniques (including buffering and overlap), potential spatial integration models and geographical presentation of complex data in a cartographic format [22]. However, the GIS database is used to collect data

Table 1. Pipe failure models.

Author	Year of study	Objective	Explanatory variables	Model type
Shamir Howard	1979 [17]	pipe breaks and other structural degradation	pipe length, pipe age, breakage history	Deterministic
Walski Pellicia	1982 [21]	pipe breaks and other structural degradation	pipe length, pipe age, breakage history	Deterministic
Kettler Goulter	1985 [16]	pipe breaks and other structural degradation	pipe length, pipe age, breakage history	Deterministic
Cox	1972 [18]	time between failures or time to the next failure	pipe length, operating pressure, pipe age, break rate, soil corrosivity	Probabilistic
Constantine Darroch	1993 [16]	pipe breaks and other structural degradation	operating pressure, pipe diameter, soil type, overhead traffic conditions	Probabilistic
Accelerated failure models	[18]	time between failures or time to the next failure	pipe age, pipe diameter, pipe length, pipe material, traffic loading, soil acidity, soil humidity, number of breaks	Probabilistic

and not to acquire data, for this purpose various devices and systems enabling reading information from these devices are using. SCADA (Supervisory Control and Data Acquisition) belongs to such systems. The WDS commonly used SCADA systems to transmit data flow and pressure in real time, which is why it also using as early warning systems (e.g. in the event of pressure/flow drop or increase). In combination with a telemetry system, SCADA can be used to find water leaks and detect the place of failure. These systems enable transferring and storing data, network models are used to predict (simulate) failures. Mathematical models of water distribution network allow the simulation of water flow and pressure in the system during "normal" and partially failed system as well it can generates failure rates and repair events according to specified probability distributions [23]. Using the collected historical data and supporting systems (e.g. hydraulic models), decision systems can be created, which can then be used in the Water Safety Plans. Water Safety Plans are based on risk assessment and risk management. Risk issues provide answers to three questions: What can happen? How likely is it to happen? and Given that it occurs, what are the consequences? [24]. To answer these questions, a risk analysis should be carried out using the matrix method. In order to answer these questions, a risk analysis should be carried out using a matrix method in which probabilities and consequences weights are assumed, and the product of these weights will determine the risk [25].

2 RESEARCH OBJECT

The subject of the study is the selected subsystem of the biggest collective Water Distribution System (WDS) in Poland, which is located in the southern-west of Poland in Silesian region. Analyzed WDS is composed of four local Water Treatment Plants (WTP A, B, C and D) with a total average daily production of 72 577 m^3, and four storage tanks (E, F, G and H) with the total capacity of 155 200 m^3 (Figure 2). The area under consideration is additionally fed by a pumping station (located outside the research area), which supplies water in the amount of 60 000 m^3 per day. The average daily water consumption of this area is 102 000 m^3. The study area is characterized by high altitude variability from 240 m to 364 m above sea level. The central point of the subsystem is the storage tank E (345 m above sea level), which are supplied from two directions (WTP A and Pump Station I) and delivers water to the largest number of customers in north area. Tanks G (315 m above sea level) and H (364 m above sea level) collect water and if necessary provide a water supply. WTP C and B

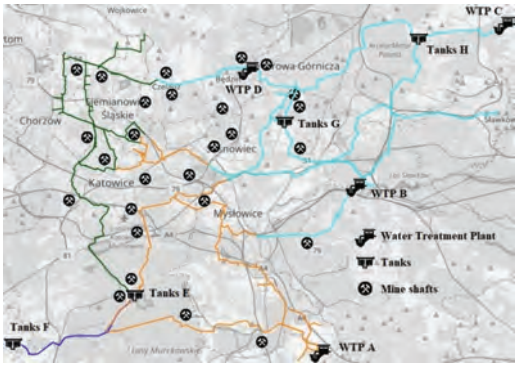


Figure 2. Scheme of the water supply network with marked mine shafts.

Figure 3. Material structure of water pipe network.

Figure 4. Age structure of water pipe network.

supply the smallest part of north-west area, while WTP D works occasionally in a case of higher water demands.

Analyzed subsystem of WDS is a widespread system covering about 100 km^2 of the area with a total water pipe length of 256 km. Water pipe network is made mainly of steel, as well as polyethylene (PE) and ductile iron (Figure 3) with pipe diameters from 55 mm to 1600 mm. The oldest pipes that build this subsystem come from 1929 (Steel) and the latest from 2016 (PE) (Figure 4).

The analyzed network is located in the area of intensive exploitation mining areas where there are mining damages and shocks. The reason for this is the collapse of the terrain and cracks in the urban infrastructure. Figure 2 shows the location of existing and inactive mine shafts in the considered area.

3 RESEARCH METHODOLOGY

The hydraulic model of the subsystem was used for the analysis. The network topology was exported

from the GIS database to the WaterGEMS software. The model was calibrated to average values of water demand from 2016 (102 000 m^3). While the validation was carried out for the operation conditions during three days (17–19.10.2016). For this model we obtained a high correlation of computed values and observed values, for pressure 99.4% and for flow 99.0%. Failure data form mentioned area has been entered into the software from the 5-year period (2012–2016). Table 2 shows the total number of pipe cracks for a given year.

Based on length, number of breaks and break history (duration of break history) WaterGEMS calculate predict rate of failure. The results that are calculated by the Pipe Break Analysis include:

– failure rate λ_{IB} (breaks/yr/km)- number of breaks for individual pipe, according to the pattern:

$$\lambda_{IB} = \frac{N}{L \cdot t'}, \text{break/yr/km} \quad (1)$$

where N = number of breaks; L = length of pipe (kilometer); t = period of analysis (year).

– pipe group failure rate λ_{GB} (breaks/yr/km) – number of breaks for a given pipe group, according to the pattern:

$$\lambda_{GB} = \frac{N}{L \cdot t'}, \text{break/yr/km} \quad (2)$$

where N = number of breaks for the group; L = total length of pipe in the group (kilometer); t = period of analysis (year).

Based on the total length of the network, the number of failure and the length of the failure time, one group break rate was created for calculations for all pipes.

– projected failure rate λ_{PB} the product of the scaled break rate, the projection period and the length of pipe. Estimate of the number of breaks over the projection period assuming that past break rates persist, according to the pattern:

$$\lambda_{PB} = a \cdot \lambda_{IB} + (1-a) \cdot \lambda_{GB} \quad (3)$$

where a = index of the share of damage intensity.

Due to the creation of only one pipe break group, the index value *a* was chosen at 0.8.

The studied network is a main network supplying water to cities and industry, so every water customer will belong to a critical group (hospitals, food production, etc.). Any interruption in water supply can have a major impact on the functioning of customer group. The task of this work is to determine the impact of pipe breakage on the efficiency of water supply. For this purpose, the water flows calculated in the hydraulic model and the predicted number of pipe failure were used. The results of the analyzes were presented in the form of a risk matrix.

Table 2. Number of pipe failure for a given year.

Year	2012	2013	2014	2015	2016
Number of pipe failure	313	226	232	303	210

4 RESULT AND DISCUSSION

Figure 5 shows the average water flow in the water supply system, the lowest flow values (below 50 m^3/h) are marked in dark blue color, while the highest values by red color (above 600 m^3/h).

The number of failure for individual pipe was determined, zero values were obtained for 2533 sections (sections without a failure during the considered period), the minimum value of failure rate was 0.098 (break/yr/km) and maximum 452. Failure rate for pipe group was obtain for all section in analyzed network and was 1.06 (break/yr/km). Based on these rates, projected failure was calculated. The result of simulation a base map representation of the failure risk of pipes in the network. Figure 6 is a map of these pipes and their corresponding risk levels. The largest number of expected pipes failures was achieved for the section located east of the Tanks E and the section located north of the Tanks G. They constitute 2.5% of the analyzed network. The biggest part (90%) is the probability of occurrence of up to 4 failures per year, while the smallest part is the probability of the number of failures between 12 and 16 (1%).

The values of the predicted pipe fractures were calculated for each segment. The smallest value of failure rate is 0.001, while the highest is 20.0. Based on the obtained simulation results, a risk matrix

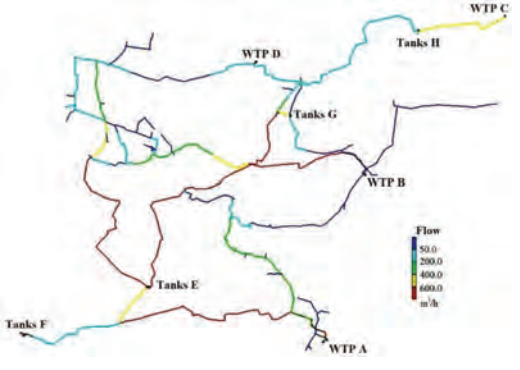


Figure 5. Hydraulic simulation result.

Figure 6. Network map with marked ranges of projected breaks.

Weight of projected failure rate W_2	Weight of consequences of flow rate W_1				
	1	2	3	4	5
1	1	2	3	4	5
2	2	4	6	8	10
3	3	6	9	12	15
4	4	8	12	16	20
5	5	10	15	20	25

Figure 7. Risk matrix of pipe failure.

Table 3. Flow rate division into weight classes.

	Categorization due to the flow rate				
Water flow [m^3/h]	0–50	50–200	200–400	400–600	>600
Weight scale	1	2	3	4	5

Table 4. Projected breaks division into weight classes.

	Categorization due to the intensity of failure rate				
Projected failure rate	0–4	4–8	8–12	12–16	>16
Weight scale	1	2	3	4	5

was created (Figure 7). Based on the analysis, a five-stage categorization of the effects of damage based on the volume of water supplies was made. The ranges of water flow values and failure rate were assigned to individual weight classes from 1 to 5 (Tables 3 and 4).

For risk analysis, a risk matrix was developed, which takes into account the consequences of failure occurrence expressed by the flow rate and the intensity of failures. Defined risk is expressed by the formula (4):

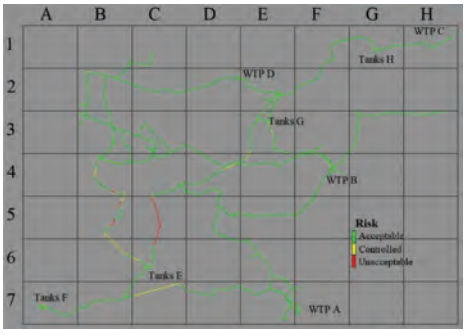


Figure 8. The distribution of risk on the network map.

$$R = W_1 \cdot W_2 \quad (4)$$

where W_1 = weight of flow rate; W_2 = weight of projected failure rate.

Risk value varies from 1 to 25. For this risk, a three-class classification was made: acceptable risk, control and unacceptable risk.

Three risk classes have been classified: Acceptable risk – weight 1–4 (green color), Controlled risk – weight 6–10 (yellow color) and Unacceptable risk – weight 12–25 (red color). Green color indicates a low impact of pipe failure on network operation, yellow – medium, and red – high. Figure 8 shows the distribution of risk classes in the network. The result of the risk assessment of damage to the pipe section allows for easy identification and classification of pipes, which need to be modernized. Based on the impact levels, areas with the highest risk of lack of water supply/reduction of water supply to the largest number of customers were designated. In areas located in cells 4B, 5B, 5C and 6C there is the highest impact of pipe failure on water supply, so the pipes in these areas must have the highest priority for modernization. Pipes in the mentioned cells distribute water to the largest number of customers and the failures of these pipes will cause water shortages in the largest area. Pipes located in cells 4B, 5B, 6B, 6C, 7C, 4D, 3E and 4E belong to the medium impact, therefore rehabilitation should be considered later. The remaining pipelines are the least affected and do not pose a threat.

5 CONCLUSION

The water distribution system is an extremely important infrastructure enabling the functioning of various social zones. Lack of water supply, reduced amount of water supply or contamination of water can lead to the risk of human life. An important element of network management is the determination and assessment of risk. These analyzes may concern various aspects related to the operation of the water

supply network, among others chlorine disappearance, secondary water contamination or pipe failures. Data sets and decision support systems are important for risk management. These systems certainly include the GIS and SCADA databases that acquire and provide data, but also a useful tool are hydraulic models that allow to perform various simulations. Based on the conducted simulations, so-called "sensitive" areas can be identified, for which different types of risks can then be determined. Knowing the risks, waterworks companies can save money and time to plan future investments.

ACKNOWLEDGEMENT

This work was supported by Ministry of Science and Higher Education Republic of Poland within statutory funds as well as BKM-554/RIE-4/2017 research.

REFERENCES

[1] Z. Kapelan, D. Savic & H. Mahmoud, A Response Methodology for Reducing Impacts of Failure Events in Water Distribution Networks, Procedia Engineering 186 (2017) 218–227.

[2] J.C. Devera, Risk Assessment Model for Pipe Rehabilitation and Replacement in a Water Distribution System (2013).

[3] R. Farmani et al. Pipe Failure Prediction in Water Distribution Systems Considering Static and Dynamic Factors, *Procedia Engineering* 186 (2017) 117–126.

[4] M. Tabesh et al. Assessing Pipe Failure Rate and Mechanical Reliability of Water Distribution Networks Using Data Driven Modelling, Journal of Hydroinformatics 11 (2009) 1–17.

[5] A. Mailhot, G. Pelletier, JF. Noel, JP. Villeneuve, Modeling the evolution of the structural state of water pipe networks with brief recorded pipe break histories: Methodology and application, Water Resources Research, 36 (2000) 3053–3062.

[6] P. Rajeev et al. Rajani, Factors Contributing to Large Diameter Water Pipe Failure as Evident from Failure Inspection, Strategic Asset Management of Water and Wastewater Infrastructure: Leading Edge Strategic Asset Management (LESAM13) (2012).

[7] A. FioriniMorosini, P. Veltri, F. Costanzo, D. Savic, Identification of leakages by calibration of WDS model, Procedia Engineering 70 (2014) 660–667.

[8] S. Yamijala et al, Statistical models for the analysis of water distribution system pipe break data, *Reliability Engineering & System Safety* 94 (2009) 282–293.

[9] R.A. Francis et al, Bayesian Belief Neworks for Predicting Drinking Water Distribution System Pipe Breaks, *Reliability Engineering and System Safety* 130 (2014) 1–11.

[10] S. Bebiasi, C.M. Giorgio Bort, A. Bosoni, P. Bertola, M. Righetti, Influence of hourly water consumption in model calibration for leakage detection in a WDS, Procedia Engineering 70 (2014) 467–476.

[11] C. Hua Tian, J. Xiao, J. Huang, F. Albertao, Pipe Failure Prediction, Service Operations, Logistics, and Informatics (SOLI) (2011) 121–125.

[12] K. Nygard, E. Wahl, T. Krogh, O.A. Tveit, E. Bohleng, A. Tverdal, P. Aavitsland, Breaks and maintenance work in the water distribution systems and gastrointestinal illness: a cohort study, *International Journal of Epidemiology* 36 (2007) 873–880.

[13] M.C. Besner, M. Prevost, S. Regli, Assessing the public health risk of microbial intrusion events in distribution systems: Conceptual model, available data, and challenges, Water Research 45 (2011) 961–979.

[14] World Health Organization, Water Safety Plans, Managing drinking-water quality from catchment to consumer, Geneva (2005).

[15] O. Giustolisi, D. Laucelli, D.A. Savic, Development of rehabilitation plans for water mains replacement considering risk and cost-benefit assessment, *Civil Engineering and Environmental Systems* 23 (2005) 175–190.

[16] A. Scheidegger, J.P. Leitão, L. Scholten, Statistical failure models for water distribution pipes—A review from a unified perspective, Water Research 83 (2015) 237–247.

[17] U. Shamir, C.D.D. Howard, An analytic approach to scheduling pipe replacement, J. AWWA 71 (1979) 248–258.

[18] A. Debón, A. Carrión, E. Cabrera, H. Solanoc, Comparing risk of failure models in water supply networks using ROC curves, *Reliability Engineering & System Safety* 95 (2010) 43–48.

[19] S.Sægrov, J.F.M. Baptista, P.Conroy, R.K. Herz, P. LeGauffre, G. Moss, J.E. Oddevald, B. Rajani, M. Schiatti, Rehabilitation of water networks: Survey of research needs and on-going efforts, Urban Water 1 (1999) 15–22.

[20] R. Farmani, G.A. Walters, D.A. Savic, Trade-off between Total Cost and Reliability for Anytown Water Distribution Network, Journal of Water Resources Planning and Management 131 (2005) 161–171.

[21] T.M. Walski, A. Pellicia, Economic Analysis of Water Main Breaks, *Journal of American Water Works Association* 74 (1982) 140–147.

[22] R.B. McMaster, H. Leitner, E. Sheppard, GIS-based Environmental Equity and Risk Assessment: Methodological Problems and Prospects, *Cartography and Geographic Information Systems,* 24 (1997) 172–189.

[23] J.M. Wagner et al. Water Distribution Reliability: Simulation Methods, Journal of Water Resources Planning and Management 114 (1988) 253–275.

[24] T. Bedford, R. Cooke, Probabilistic Risk Analysis: Foundations and Methods, Cambridge University Press (2001) 9–13.

[25] R.L. Murray, K.E. Holbert, Nuclear Energy—An Introduction to the Concepts, Systems, and Applications of Nuclear Processes (7th Edition), Elsevier (2015) 359–362.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Understanding and including the dynamics of extreme natural hazard event uncertainty within the overall offshore wind farm project risk assessment using a causality-based graphical modelling approach

R. Zamora, J. Qin, A.S. Kristensen, S. Mehmood & S. Ahmed
Danish Center for Risk and Safety Management, Aalborg University, Esbjerg, Denmark

S. Cuthbert
Ørsted, Fredericia, Denmark

ABSTRACT: Offshore wind structures are subject to the combined action of wind and wave loads. A change of these loads may significantly affect the integrity of the structural elements. Increased instabilities in the Earth's climate system could increase the frequency of extreme events (e.g. rogue waves) well beyond the frequency values currently recommended within structural design standards. Inherent to extreme event modelling is the need to use expert (subjective) judgement and sparse data sets. In this context, a Bayesian Belief Network (BBN) can be applied to describe the effect of these changes on the frequency of rogue waves within wind farms located in shallow water depths of 20–60 metres. This graphical modelling approach provides the structure to effectively communicate, among others, parameter uncertainty, causality across multiple risk factors, quantitative definition of assessment subjectivity or potential impact of a change in rogue wave frequency relative to that described in current design standards.

1 INTRODUCTION

The term "rogue", "freak", "abnormal" or "giant" wave commonly refers to waves that are very steep and large in absolute measures and, at the same time, significantly larger than the surrounding waves in the sea state, and are thus unexpected (Bitner-Gregersen, 2017). They are statistically unlikely to occur in a given sea state (either low, intermediate or high), based on averaged properties of that sea state (Bitner-Gregersen & Gramstad, 2015).

This physical phenomenon is not fully understood, but increasing reliable measurements and records, as well as the significant increase in computational power and numerical modelling capacity, allow to explore these extreme events with greater accuracy.

There are several motivations to reduce the risk of wave-related incidents. First, because they clearly represent a current threat to marine installations. Second, because more severe sea state conditions may be expected in some ocean regions associated with climate change and global warming (IPCC Panel, 2014). Third, because understanding and forecasting waves under various conditions is essential with respect to design and operation of offshore structures.

Based on these initial premises, addressing these extreme events as potential risk and including them in the customary Risk Assessment process of a company that operates physical assets in an offshore environment is entirely justified, despite its complexity and the high number of uncertainties involved.

In the present work a causality-based probabilistic graphical modelling methodology is proposed to assess the risk associated with rogue waves in offshore wind farm projects at the final design stage. The methodology includes the impact of future climate change and provides the structure in which to effectively communicate: a) parameter uncertainty; b) correlation across multiple risk factors (i.e. "Systems of Systems" (SoS) complexity mapping/analyses); c) definition of assessment subjectivity; d) and potential impacts of low probability catastrophic events (i.e. extreme events). The methodology provides a holistic framework that can be integrated into existing decision-making processes currently defined within a large capital project execution process.

In brief, the method studies the probability of a rogue wave impacting an offshore structure situated in a predefined location of the Northern North Sea, between 20 and 60 m depth, and includes 3 main stages: risk understanding, qualitative bow-tie creation; and transformation to a Belief Bayesian Network.

2 RISK UNDERSTANDING

Risk assessment is to a large extent about gaining 'risk understanding' in the sense of knowledge—justified beliefs, by producing a risk description (C_0,Q,K), where C_0 are the specified consequences of the activity studied, Q a measure of uncertainty, and K is the background knowledge on which C_0 and Q are based (Amundrud & Aven, 2015). According to these authors, these justified beliefs are based on data, information (relevant processed data) and models. The uncertainty judgments about C_0 using Q can also be seen as justified beliefs.

K is a limiting aspect in the proposed methodology, due to the lack of understanding of the physical process of creation of rogue waves. For example, describing the wave phenomenon is the result of a set of uncertainties. The random model for ocean waves is constructed by representing the sea surface as a sum of elementary waves with different wavelengths, frequencies, and directions of propagation (Bitner-Gregersen & Gramstad, 2015). However, in reality ocean waves are not described exactly by a linear formulation or second-order theories, and therefore require a set of increasingly accurate formulations. The more accurate, the more mathematically complex and more difficult the model will be. As a result, the logical functions and equations included in the proposed graphical model are based on the linear theory, the most tractable approach for the graphical model under design.

Uncertainty related to environmental phenomena may be divided in aleatory uncertainty (natural randomness) and epistemic (knowledge) uncertainty; and the latest in: data uncertainty, statistical uncertainty, model uncertainty and climatic uncertainty (Bitner-Gregersen et al. 2013).

Assessing data uncertainty is out of the scope of this study, so available data are assumed to be appropriate. To minimize the statistical and climatic uncertainty, a long-term data source was selected. The European Centre for Medium-Range Weather Forecasts' ERA-Interim is a global atmospheric reanalysis from 1979, publicly accessible and continuously updated in real time (European Centre for Medium-Range Weather Forecast, 2017). After establishing a geographical location in the Northern Sea, 4 measurements per day were obtained between 1979 and 2017 (about 55.000 values per variable) for 30 different variables. Only 9 of them were considered relevant for the project: model depth (d), zero-crossing mean period (Tz), wave spectral directional width (σ_θ), significant wave height (Hs), mean wave direction (θ), mean direction of wind waves (θ_1), mean direction of swell (θ_2), and Benjamin-Feir index (BFI).

There are other relevant variables, such as the wave length (λ), that are not independent. In these cases, formulae given by the Recommended Practice DNVGL-RP-C205 have been used (DNV GL, 2017).

Finally, defining and managing the model is the core part of this work and a main responsibility of the risk analyst (designing, building, assigning probability, running simulations, reporting and maintaining). It reflects the limitations of the previous factors and adds new uncertainties, due to failed assumptions in physical process formulations, or choices of probability distribution types for representation of uncertainties. In this regard, the method tries to register and track all the detected uncertainties. To limit this effect, all the variables were fitted to a probability distribution using the software tool, ModelRisk (Vose Software, 2018) only when the best fit was not supported by the Bayesian Network software (OpenBUGS).

3 BOW-TIE CREATION

The bow-tie is a graphical approach frequently used to represent a Risk Event, its Causes (Drivers), Prevention Barriers (Controls), Mitigation Barriers, and its Consequences (Impacts) in a visual and logical manner. Centered on a critical (risk) event, it is composed of a simplified fault tree on the left-hand side and an equally simplified event tree on the right-hand side showing the possible consequences of the critical event based on the failure or success of safety functions (Khakzad et al. 2013). To understand the relations and dependences among factors involved in the creation and impacts of rogue waves and climate change on offshore wind structures, a qualitative bow-tie is proposed. The first step consists of formulating the critical event: impact of a breaking rogue wave on an offshore wind structure (named "IMPACT" in the graphical model). This step seems to be obvious, but in complex or emerging risks it is essential to organize and plan the following phases of the method.

In this case, due to the complexity of the analyzed physical phenomena, the bow-tie focuses on the left-side, or analysis of causes (drivers) and barriers (controls). The event tree of consequences is reduced to one: the failure of the structure (F).

After a deep review of the state-of-the-art related to rogue waves and climate change impacts on the study area, as well as the available data, the drivers and controls are analyzed individually and placed in the bow-tie, establishing the appropriate connections and causal relations. The graphic is continuously updated until it gets its final shape, shown in Figure 1.

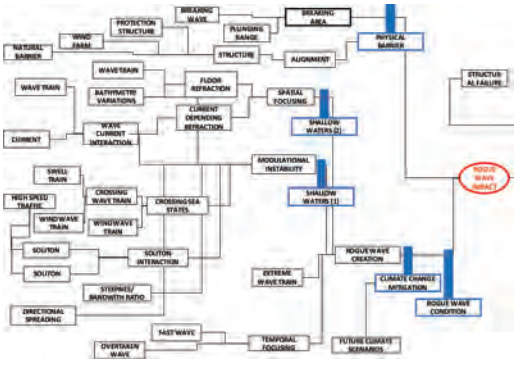


Figure 1. Final bow-tie.

Several different mechanisms may be responsible for generating rogue waves such as linear focusing of energy (spatial and dispersive: K_f, K_c and T_f), wave-current interactions (CI), crossing seas (wind sea and swell or two swell systems, CS), quasi-resonant nonlinear interactions (modulational instability, BFI), shallow water effects (SWC_i), solitons interactions (SO), directional spreading (DS), and wind forcing (W).

Atmospheric forcing has not been considered in the bow-tie as a cause of waves to simplify the visual understanding of the process. The relevant variables obtained from the dataset are included in the bow-tie as primary events. Other relevant variables, as slope (SL), angle between the wave crest and depth contours (α_0), angle between target and protection structure (β), Ursell number (U_R) or maximum height (H_m) are added as primary events, when statistical data are not available but are required for a consistent explanation of an intermediate event. Some of them are calculated in future steps or treated as assumptions. The bow-tie shows two main controls: protective structure (O), as a physical barrier to avoid the impact of a breaking rogue wave against the offshore wind structure (OWS); and climate change (C). C is placed as control, assuming its barrier effect is focused on limiting or preventing the CO_2 emissions caused by humans, where the key assumption is that the accumulation of CO_2 and other greenhouse gases are the primary drivers for climate change and that the human population is largely the driver for the significant increase in the atmospheric concentrations of those gases in the past 200 years. Other controls are related to shallow water restrictions or used for reversing interactions of separated subsystems (current, ship traffic, etc.) over the wave fields or between drivers of different nature, when needed. The other three are natural controls: shallow water conditions and rogue wave conditions.

4 BAYESIAN NETWORK

The bow-tie graphical model is used in this method as a primary tool to understand the risk and locate the critical event in its cause-effect framework. However, it presents a static picture of the problem. Besides, no causal relation can be established between primary events or other events of different branches of the fault or event tree. These problems are solved with its transformation to a Belief Bayesian Network (BBN).

A BBN is an explicit description of the direct dependencies between a set of variables, in the form of a directed graph and a set of nodes linked to a probability. This structure offers the following benefits (Fenton & Neil, 2013): modelling causal factors explicitly, reasoning from effect to cause and vice versa; updating the probability distributions for every unknown variable whenever an observation is entered into any node; reducing the burden of parameter acquisition; overturning previous beliefs in the light of new evidence (explaining away); making predictions with incomplete data; combining diverse types of evidence including both subjective beliefs and objective data; and arriving at decisions based on visible, auditable reasoning.

The conversion of a bow-tie into a BBN is summarized in Figures 2 and 3.

The BBN includes different interacting systems besides the waves system, and includes the current, seabed, wind, climate, ship traffic and artificial structure. Figure 4 shows this graphical model.

The fitted distributions are included as parent nodes, because are the basic parameters of the model. There are 13 "parent distributions", whereas only four of them are not obtained from available data. In these cases, a uniform distribution is assumed. One variable relies on the seabed conditions and would be subject to a better characterization with the consideration of a bathymetry

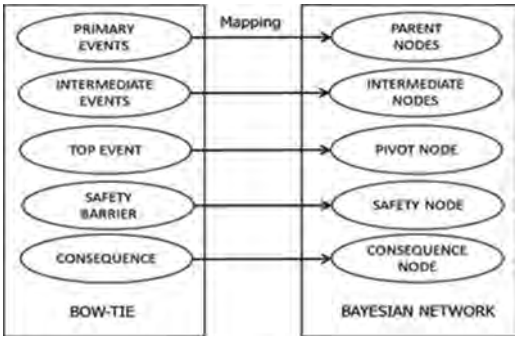


Figure 2. Mapping algorithm from bow-tie to Bayesian Network (Khakzad et al., 2013).

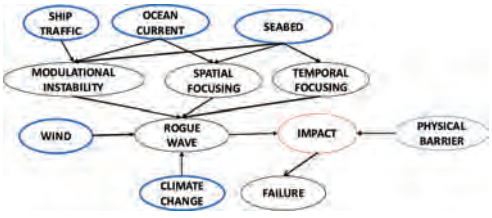


Figure 3. Bayesian Network related to bow-tie elements.

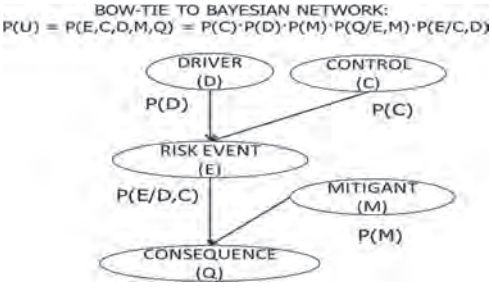


Figure 4. Overall Bayesian Network with interacting systems.

model: the angle between wave crests and depth contours (α_0). Another one (Froude Number Fd) depends on the ship traffic around the offshore wind turbine structure (OWS), but its assessment is out of the scope of this work.

Following the conclusions of the bow-tie analysis, the failure of the OWS occurs when a rogue wave impacts on it. The probability of this impact is "*the probability of a rogue wave breaking in front of the OWS within the plunging range without a protective structure in between*". When the wave breaks just at the location or behind, the plunge distance is not relevant for the targeted OWS. By contrast, when the wave breaks in front of the structure, this distance is relevant, because it defines the area where the wave is dangerous. However, given that the available data are restricted to the selected location, further spatial considerations (i.e. defining a breaking point or a plunge distance in front of the OWS) are out of the scope of this study.

Therefore, for this event to happen or not, it is necessary the presence of a rogue wave that breaks without an opposing protective structure in between.

In the graphical model, an extreme wave is considered a rogue wave R when the height doubles the significant height H_s ($R > 2H_s$) (Bitner-Gregersen & Gramstad, 2015). The wave height is limited by breaking. The maximum wave height H_b condition is based on the Recommended Practice DNVGL-RP-C205 (DNV GL, 2017):

$$H_b = \lambda 0.142 tanh \frac{2\pi d}{\lambda} \quad (1)$$

where λ is the wave length corresponding to water depth d.

The accompanying structure may be natural or artificial. If the structure is artificial, it can be either of floating type with a mooring to the seafloor or a solid anchored structure that is submerged or slightly above the surface. In an offshore wind farm, another OWS may protect the selected structure from the impact of a rogue wave. The condition to be protective is being total or partially aligned with the OWS in the mean wave direction. This condition happens, as shown in Figure 5, when the angle between the wave and the segment that links both structures (β) is between θ+90º and θ+270º (no other physical phenomena, i.e. refraction or diffraction, are included).

The critical assumption of the model is that the extreme wave heights (H_m) calculated from the available data are generated exclusively by the wave focusing under the action of wind. The final wave height (W) is then the result of an increase over the value of H_m due to the causes explained through the bow-tie, as expressed in Eq. (2):

$$W = H_m \cdot (1+C_{5\,m} \cdot SWC_1 \cdot M + C_{6l} \cdot C + step (0.6–U_R) \cdot C_{7t} \cdot TF + C_{3f} \cdot SWC_2 \cdot step (K_f–1) \cdot (K_f –1) + C_{1c} \cdot step (CI) \cdot K_c) \quad (2)$$

where

H_m = extreme value of height;

K_c = height increase proportion due to current refraction;

K_f = height increase proportion due to floor refraction;

M = height increase proportion due to modularity;

C = height increase proportion due to the climate change; and

TF = height increase proportion due to temporal focusing.

It may be argued that the measured heights are already the result of these causes or, at least, the

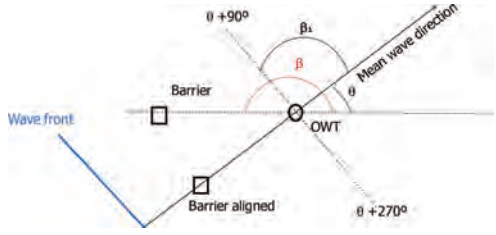


Figure 5. Alignment of the protective structure.

linear causes, i.e. spatial and temporal focusing. To deal with such complications, each driver has one control node (constant), so that the unexpected cause or interacting system can be eliminated from the model: C_{1c} for the current; C_{2s} for the ship traffic; C_{3f} for the seabed refraction; C_{4o} for the protective structure; C_{5m} for the modularity instability; and C_{6t} for the temporal focusing.

There are also natural controls (step(0.6-U_R), step(CI), SWC1, SWC2) that cancel the drivers due to natural conditions. These natural conditions can be modelled.

Only height increases are considered, so the condition to take refraction into account is $K_f > 1$: step ($K_f - 1$).

H_m is calculated following the extreme value theory and fitting the results to a Gumbel distribution.

K_f is calculated based on the Recommended Practice DNVGL-RP-C205 (DNV GL, 2017):

$$K_f = K_s \cdot K_r \tag{3}$$

$$K_r = \left[\frac{1 - sin^2\alpha_0 tanh^2(kd)}{\cos^2\alpha_0} \right]^{-1/4} \tag{4}$$

$$K_s = \sqrt{\frac{c_{g,0}}{c_g}} \tag{5}$$

where

K_s = shoaling coefficient;

K_r = refraction coefficient;

α_0 = the angle between the wave crest and the depth contours at the location;

k = wave number;

d = depth; and

C_g = group velocity.

K_c is a good example of the difficulties found to model some of the drivers involved in the process. The first approach to define the variable K_c was based on the analysis of this phenomenon presented by Sorensen (Sorensen, 2006). Figure 6 shows how a wave propagating with speed C from still water to water having a current velocity U, changes its direction.

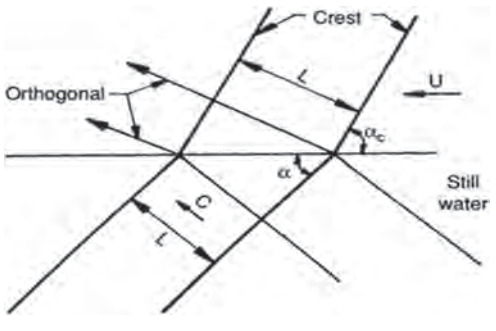


Figure 6. Definition sketch for wave refraction by a current (Sorensen, 2006).

In mathematical terms, these equations are obtained:

$$K_c = \frac{H_c}{H} = \left(\frac{L_c}{L}\right)^2 \frac{cos\alpha}{cos\alpha_c} \left[\frac{\left(1 - \frac{U}{C} sin\alpha\right)^6}{1 + \frac{U}{C} sin\alpha} \right] \tag{6}$$

$$L_c = L \frac{sin\alpha_c}{sin\alpha} \tag{7}$$

$$sin\alpha_c = \frac{sin\alpha}{\left(1 - \frac{U}{C} sin\alpha\right)^2} \tag{8}$$

where

H_c = Height after refraction;

H = Height before refraction;

L_c = Wave length after refraction;

L = wave length before refraction;

U = current velocity;

C = wave velocity;

α = angle between the current and the crest front;

α_c = angle between the current and the crest front after refraction.

Considering $K_c = H_c/H$, an expression of K_c as function of α and α_c can be obtained, but introducing it in the model was impossible and always led to a systematic software error. Other equations were checked, such as those presented by Iwagaki et al. (1977). A different approach was finally selected based on the work by Mathiesen (1987), which is derived from the computer model to measure the refraction of ocean directional wave spectra and applied it to a circular current whirl typical in the Norwegian coastal current. This model found that the relative changes in wave heights were within ±20% as compared with the wave height of the incoming waves.

M is calculated as the average probability of the nonlinear modularity drivers, which are: solitons interactions (SO), variable bathymetry (SL), crossing seas (CS), Benjamin-Feir interaction (BFI), directional spreading (DS) and wave-current interaction (CI). M is limited to a maximum value (M_{max}) of 0.20. This value is defined considering several systematic studies which shows that effects of modulational instability can enhance the crest height for long-crested waves by up to 20%, at lower probability levels, while the troughs become about 20% deeper than second–order troughs (Kharif et al. 2009).

C is calculated based on the CO_2 emissions originating from the socio-economic scenarios (A1B, A2, B1 and B2) proposed by The Intergovernmental

Panel on Climate Change IPCC and the values of emissions currently estimated for the North Sea.

T_f is calculated as a function of the Ursell number U_R, with a maximum value to be established at the moment:

$$U_R = \frac{H\lambda^2}{d^3} \tag{9}$$

Kharif et al. (2009) stated that this number characterizes the ratio of nonlinearity to dispersion. When the Ursell parameter is small, the nonlinearity can be neglected, and the wave is a linear dispersive wave. In real situations of wind waves, the values of U_R parameters are not too large, and the dispersive trains contribute significantly to the statistical wave characteristics. Based on these authors, a value of $U_R < 0.6$ is selected to consider the impact of the temporal focusing as relevant.

There are two restrictions related to the shallow waters which must be considered, and are given the variable names, SWC_1 and SWC_2. Water is considered shallow when the surface waves are noticeably affected by bottom topography (Bitner-Gregersen & Gramstad, 2015). This condition occurs when the depth, d, becomes less than half the wavelength, λ.

Modulational instability becomes weaker with decreasing depth and it is suspected to play a less important role in shallow water (Bitner-Gregersen & Gramstad, 2015). Benney & Roskes (1969) estimated that modulational instability disappears when $2\pi d/l < 1.363$ for unidirectional waves. Under this threshold, the model cancels the driver M. This is the restriction with the variable name, SWC_1

Similarly, the seabed related refraction (K_f) is canceled when the shallow water condition is not accomplished (restriction SWC_2).

4.1 *Modulational instability drivers*

Seven drivers are involved in the creation of nonlinear instability. Their inclusion, conditions and limits are discussed in the following sections.

4.1.1 *Solitons interaction (SO)*

Solitons interaction has been suggested as a source of nonlinearity in shallow water (Kharif et al., 2009). Peterson et al. (2003) linked this mechanism to relatively shallow coastal areas with high ship traffic density, particularly high-speed ships when they sail with critical or supercritical speeds. These speed levels rely on a value of the Froude number, F_d, which is the ratio of the ship speed and the maximum phase speed of gravity waves, equal or higher than 1. Therefore, the model constrains the impact of this driver to this threshold. It is out of scope of this study to analyze the traffic in the vicinity of the location, so a uniform distribution has been used for the variable F_d.

4.1.2 *Variable bathymetry (SL)*

Recent works have shown that the probability of rogue waves may increase on the shallow side of an underwater slope. Sergeeva et al. (2011) linked the probability of rogue waves to the wave steepness, which is characterized in terms of the Ursell parameter. Both variables increase when the depth decreases (water shallowing), and the wave state deviates from the Gaussian. Based on previous research, the condition for nonlinearity due to the interactions with a variable bottom has been fixed when $U_R > 0.6$ (Kharif et al., 2009).

4.1.3 *Crossing seas (CS)*

When two wave systems (wind sea and swell or two swell systems) are separated in direction or frequency and cross, the modularity increases depending on the angle between them. Both wave trains are assumed to be narrow banded and weakly nonlinear (Kharif et al., 2009).

Onorato et al. (2010) suggested that an increased probability of rogue waves was associated with angles between 40° and 60°. This is the condition used in the graphical model. ERA INTERIM database offers separated information about the mean wind waves directions (θ_1) and mean swell direction (θ_2), so the possibility of crossing wind seas is not considered.

4.1.4 *Benjamin-Feir interaction (BFI)*

A key parameter controlling the importance of the nonlinear wave-wave interactions is the Benjamin–Feir Index (BFI) which is the ratio of the wave steepness to the spectral bandwidth (Kharif et al., 2009).

$$BFI = \frac{\varepsilon\sqrt{2}}{\delta_\theta} \tag{10}$$

where:

ε = wave steepness; and

δ_θ = spectral directional width.

Instability condition is given by Eq. (11), and is used as a condition in the graphical model (Bitner-Gregersen, 2017):

$$\sqrt{2}BFI > 1 \tag{11}$$

4.1.5 *Directional spreading (DS)*

Onorato et al. (2002) showed that the probability of occurrence of rogue waves depends not only on BFI, but also on the directional spreading of the waves. Waseda et al. (2011) found evidence that occurrence of rogue waves was associated with sea states with directional spreading of less than about 30°, sug-

gesting that sea states with increased occurrence of rogue waves may occur in realistic ocean conditions. This has been the condition used in the model.

4.1.6 *Nonlinear wave-Current Interaction (CI)*

There are theoretical, experimental, and numerical evidences to support that in some situations the combined effect of wave nonlinearity and currents can lead to an increase in rogue wave occurrence (Nakicenovic et al., 2000). Janssen & Herbers (2009) first discovered that initially stable narrow banded wave fields could become unstable when the nonlinearity was increased due to linear focusing. Toffoli et al. (2015) experimentally showed that realistic random waves propagating in opposing currents could destabilize, with a resulting increase in the occurrence of rogue waves, even for waves with directional spread that normally obey near-Gaussian properties. The probability of a current opposing to a wave field depends then on the angle between wave and current. The opposing condition is addressed by the model as the probability of the mean current direction (θc) between the values of θ + 90º and θ + 270º, with a maximum when θc is equal to θ + 180º, as shown in the Figure 7.

4.2 *Addressing climate change*

For the estimation of the climate change impact on the frequency of occurrence of a breaking rogue wave within the location proposed for an offshore wind farm, several assumptions have been made. It is accepted that there is a stochastic dependence between levels of CO_2 in the atmosphere and the ocean wave climate. On the other hand, only CO_2 is considered as a factor of climate change, although it is just one of the components of the greenhouse gas group (GHG).

The projections of future climate change scenarios are based on the four marker scenarios (A1B, A2, B1 and B2) proposed by The Intergovernmental Panel on Climate Change IPCC, over the twenty-first century (Quante & Colijn, 2016). Each emission scenario reflects different assumptions on future socioeconomic development. Scenario A2 is the worst, followed by A1, B2 and B1.

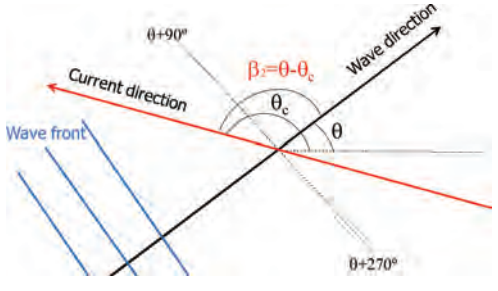


Figure 7. Wave-current interaction.

Regarding the study area, Grabemann et al (2015) analyzed a set of ten wave climate to estimate the possible impact of anthropogenic climate change on mean and extreme wave conditions in the North Sea. The projections were based on different IPCC emission scenarios, included different global and regional models starting from different initial conditions.

They found a solid pattern for the increase in median and severe significant wave height in the eastern North Sea (parts of the southeastern North Sea and large parts of the Dutch, German, and Danish coasts up to the Skagerrak) towards the end of the twenty-first century, while a decreasing trend in the western North Sea was detected. However, the magnitude of this increase was much more uncertain and oscillates between about −10 and 15% relative to the reference Hs. These numbers are consistent with other relevant studies in the area, which establish the increase between 6–8%, or up to 10% (Kharif et al., 2009).

Therefore, in the model the increase on the wave height has been defined as:

$$C = c_0 \cdot \sum_j^{10} x_{ij} \cdot s_i \tag{12}$$

where c_0 = maximum emission factor in decimal fraction; x_{ij} = reduction factor for the emission scenario i during the decade j in decimal fraction; and s_j = emission scenario.

Based on the abovementioned data, a value of 0.10 has been assigned to c_o. It corresponds to the value for the worst scenario (A2). The values of s_{ij} have been calculated based on the projections of the IPCC simulated with model AIM in the OCDE region, as stated in Table 1.

Table 1. Emission Reduction factors based on IPCC scenarios (x_{ij}).

Decade	Scenario s_i			
	A2	A1	B2	B1
1990	1	1	1	1
2000	1	1	1	1
2010	1	0.97	0.93	0.89
2020	1	0.91	0.86	0.81
2030	1	0.83	0.78	0.71
2040	1	0.77	0.73	0.64
2050	1	0.72	0.68	0.58
2060	1	0.64	0.58	0.48
2070	1	0.57	0.49	0.4
2080	1	0.49	0.41	0.31
2090	1	0.4	0.33	0.23
2100	1	0.32	0.26	0.16

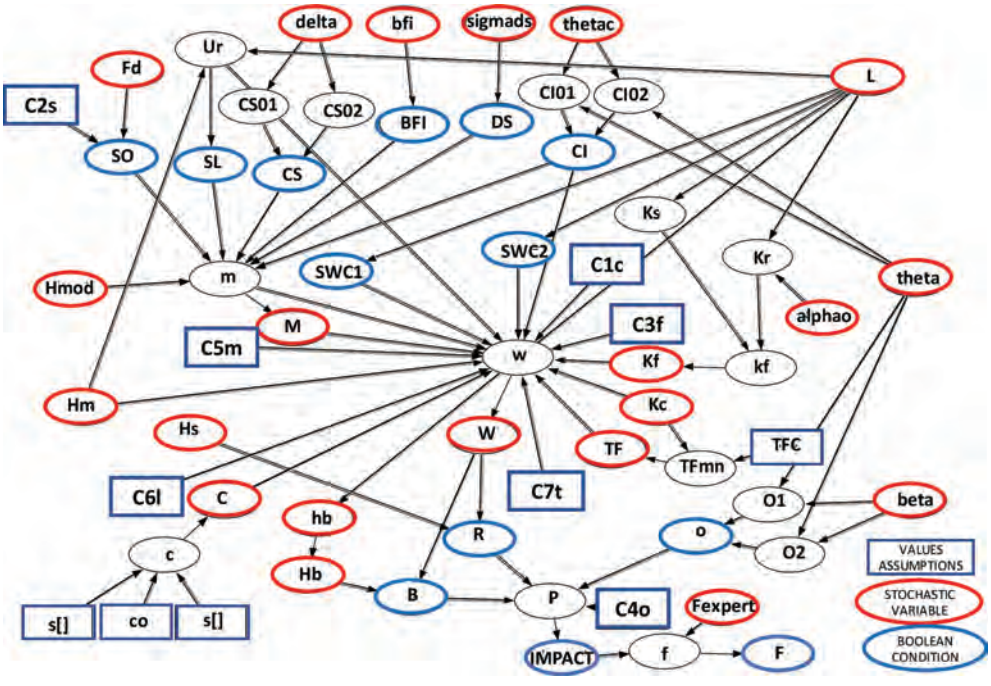


Figure 8. Final version of the Bayesian Network.

The final structure of the proposed Bayesian Network is presented in Figure 8.

5 DISCUSSION AND CONCLUSIONS

In the current study, a method for assessing a complex risk associated to physical phenomena not fully understood has been presented. The high level of complexity results in a high number of uncertainties which necessarily must be faced by the risk analyst. The focus of this study has been on understanding the physics behind the rogue wave phenomenon and determining the conditions under which such waves can be expected to occur more frequently when considering specific temporal and spatial ranges. Without the right outcome from this stage, the aim at creating a Bayesian Network would have been impossible. The role of the analyst in a decision-making process is to create a model as efficient as possible. This requirement includes its running speed, computational calculation and memory requirements, maintenance effort, file size, the least amount of assumptions, and finally, the ability to communicate the risk and the utility for the decision makers.

Due to the complexity of the risk analyzed, the number of assumptions in the model is remarkable, but a considerable effort has been made to manage those assumptions via: tracking for awareness and future improvement; and defining the model with multiple options for isolating and simulating only a partial number of individual drivers.

Currently, the BNN is being tested under different scenarios and limitations. Further conclusions will arise with the coming analysis of the results. In the worst scenario, the method will serve as a learning tool to understand the risk and its consequences in a deeper way. It will also be used to perform sensitivity analysis of the different drivers involved in the critical event. The optimal implementation would be reached when the model is used as a part of the strategic decision-making process. However, several limitations have been already detected. The output interface of the BBN software (OpenBUGS) complicates the presentation of results. There are other products in the market that seem to be more prepared for sharing results with the management in a visual way. On the other hand, spatial considerations cannot be addressed by the graphical model, i.e., the analysis of the plunging distance and the location of a potential breaking point of the wave in front of the structure.

REFERENCES

Amundrud, O., & Aven, T. (2015). On how to understand and acknowledge risk. *Reliability Engineering and System Safety*, 142, 42–47.

Benney, D.J., & Roskes, G.J. (1969). Wave Instabilities. Studies in *Applied Mathematics*, 48(4), 377–385.

Bitner-Gregersen, E. (2017). Rethinking rogue waves. Towards better modelling, insight and action. DNV. GL.

Bitner-Gregersen, E., & Gramstad, O. (2015). Rogue waves. Impact on ships and offshore structures. DNV GL STRATEGIC RESEARCH & INNOVATION POSITION PAPER 05–2015.

Bitner-Gregersen, E.M., Eide, L.I., Hørte, T., & Skjong, R. (2013). Ship and Offshore Structure Design in Climate Change Perspective.

DNV GL. (2017). Edition August 2017 Environmental conditions and environmental loads. Recommended practice DNVGL-RP-C205 (August 201).

European Centre for Medium-Range Weather Forecast. (2017). Who we are | ECMWF. Retrieved 15 December 2017, from https://www.ecmwf.int/en/about/who-we-are.

Fenton, N., & Neil, M. (2013). *Risk Assessment And Decision Analysis with Bayesian Networks.* CRC Press Taylor & Francis Group.

Grabemann, I., Groll, N., Möller, J., & Weisse, R. (2015). Climate change impact on North Sea wave conditions: a consistent analysis of ten projections. *Ocean Dynamics*, 65(2), 255–267.

IPCC Panel. (2014). Climate Change 2014: Synthesis Report.

Iwagaki, Y., Sakay, T., Tsuda, T., & Oka, Y. (1977). Wave refraction and wave height variation due to current. Bulleting of the Disaster Prevention Research Institute, 27(2), 73–91.

Janssen, T.T., & Herbers, T.H.C. (2009). Nonlinear Wave Statistics in a Focal Zone. J*ournal of Physical Oceanography,* 39(8), 1948–1964.

Khakzad, N., Khan, F., & Amyotte, P. (2013). Dynamic safety analysis of process systems by mapping bow-tie into Bayesian network. *Process Safety and Environmental Protection*, 91(1–2), 46–53.

Kharif, C., Pelinovsk, E., & Slunyaev, A. (2009). Rogue Waves in the Ocean. Springer.

Mathiesen, M. (1987). Wave refraction by a current whirl. Journal of Geophysical Research: Oceans, 92(C4), 3905–3912.

Nakicenovic, N., Alcamo, J., Davis, G., De Vries, B., Fenhann, J., Gaffin, S., … others. (2000). Emissions scenarios.

Onorato, M., Osborne, A., & Serio, M. (2002). Extreme wave events in directional, random oceanic sea states. P*hysics of Fluids,* 14(4), L25–L28.

Onorato, M., Proment, D., & Toffoli, A. (2010). Freak waves in crossing seas. T*he European Physical Journal-Special Topics*, 185(1), 45–55.

Peterson, P., Soomere, T., Engelbrecht, J., & Groesen, E. Van. (2003). Soliton interaction as a possible model for extreme waves in shallow water. Nonlinear Processes in *Geophysics*, 10, 503–510.

Quante, M., & Colijn, F. (2016). North Sea Region Climate Change Assessment. SpringerOpen.

Sergeeva, A., Pelinovsky, E., & Talipova, T. (2011). Nonlinear random wave field in shallow water: variable Korteweg-de Vries framework. *Natural Hazards and Earth System Sciences,* 11(2), 323–330.

Sorensen, R.M. (2006). *Basic Coastal Engineering* (3rd ed., p. 331). Boston: Springer.

Toffoli, A., Waseda, T., Houtani, H., & Cavaleri, L. (2015). Rogue waves in opposing currents: an experimental study on deterministic and stochastic wave trains. *Journal of Fluid Mechanics*, 769, 277–297.

Vose Software. (2018). Risk Analysis Software for Excel | Vose Software. Retrieved 13 January 2018, from https://www.vosesoftware.com/products/modelrisk/.

Waseda, T., Hallerstig, M., Ozaki, K., & Tomita, H. (2011). Enhanced freak wave occurrence with narrow directional spectrum in the North Sea. Geophysical Research Letters, 38(13).

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Risk of crack formation in power grid wooden poles and relationship with meteorological conditions: A Norwegian case study

Michael Pacevicius
eSmart Systems, Halden, Norway
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology NTNU, Trondheim, Norway

Davide Roverso
eSmart Systems, Halden, Norway

Pierluigi Salvo Rossi
Kongsberg Digital, Trondheim, Norway

Nicola Paltrinieri
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology NTNU, Trondheim, Norway

ABSTRACT: Predicting the occurrence of failures in power grids through specific outage risk predictors is a primary concern for utilities nowadays. Wooden poles represent core items to focus on in this process. Millions of them are used worldwide and they are all subject to the risk of crack formation. Analyzing the evolution of pole cracks is particularly relevant in reliability analyses of power grids for two main reasons. First: the cracks might highlight previously unconsidered or changing factors, such as unusual local weather conditions (e.g. overload of ice and/or wind). Second: as cracks provide an access for external threats (e.g. humidity, fungi, insects) to potentially non-treated internal parts of the poles, they might in turn accelerate the occurrence of further failures. Evaluating the role of crack formation is thus essential for estimating the risk of outages in power grids. As climatic variations are known to be among the most influencing factors in the initiation and propagation of cracks in wooden poles, we address this topic by suggesting a method combining open-access weather-data sources with information provided by new technologies, such as drones. We first highlight the influence of climatic factors on the reliability of wooden poles by reviewing studies describing the physical properties of wood. We then focus our research on a Norwegian case study and show how we can combine up to 60 years of meteorological information with the information provided by 17,352 geo-localized aerial pictures of cracked and non-cracked wooden utility poles. We finally discuss the way an indicator constructed on this combination can be used to predict the formation of cracks and optimize the allocation of decision-maker resources for inspection procedures.

1 INTRODUCTION

The modernization of the society has led to a global increase of power consumption over the last 50 years (Refsnæs, Rolfseng, Solvang, & Heggset, 2006; Shiu & Lam, 2004; Yoo & Kwak, 2010). As numerous businesses, public infrastructures and private households rely on the provision of power for their daily tasks, there is a need for companies in charge of the power supply to maximize their capacity and reliability in delivering power.

Predicting outage risks and avoiding downtime is crucial to ensure customer satisfaction. Moreover, anticipating unwanted events directly enables power utilities to significantly reduce losses and costs. Finally, it also enables them to optimize resource allocations for the inspection of their infrastructures after natural disasters (e.g. storms, flooding) or during scheduled maintenance procedures.

Ensuring this quality of service requires utilities to use reliable components, from the power source, through the transmission lines and to the consumption nodes. Wooden poles are widely used for the distribution part of the power grid (from regional substations to local substations and from local substations to end-users) (Eurelectric, 2010).

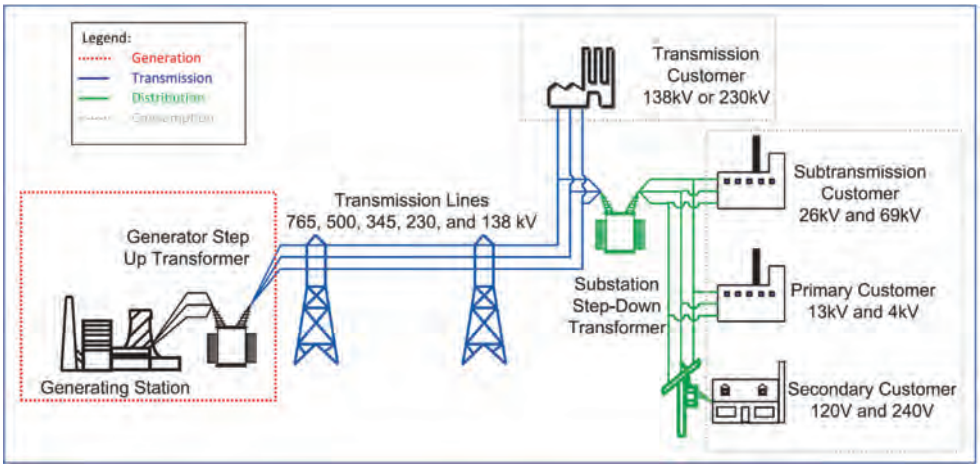


Figure 1. Outline of the transmission and distribution of power in a power grid, going from the production sites to the consumption nodes. Adapted from (U.S.-Canada Power System Outage Task Force, 2004).

Identifying the principal factors responsible for the apparition of cracks in wooden poles represents thus a main objective for predicting their failures. For this purpose, we suggest a method enabling to evaluate the effects of potential predictors. The contribution identifies the way forward for this research topic and presents preliminary findings, representing the basis for future research.

The rest of the paper is constructed as follows. Section 2 provides an overview on wooden poles characteristics and failures. Section 3 mentions various studies summarizing the main properties of wood on microscopic level. On this basis, it highlights the influence climatic variations can have on the physical structure of wooden poles. It furthermore shows how the variations can affect the reliability of the pole and thus of the transmission line. Section 4 describes the strategy applied to provide values of a crack-apparition likelihood using a Norwegian case study. It explains the choices made in the selection of the different datasets and the methods used to acquire them. Section 5 discusses the pros and the cons of the method used and shortly describes plans for future research. The last section finally concludes our work by summarizing and suggesting additional research possibilities.

2 WOODEN POLES CHARACTERISTICS AND FAILURES

Figure 1 shows schematically how power is delivered from a generating station, through transmission and distribution lines (respectively maintained by Transmission System Operators (TSO) and Distribution System Operators (DSO)), to different categories of end customers. Wooden utility poles used in the power grid exist in different shapes and configurations, depending on the physical requirements of the power lines, on the geographical conformation of their location, and on their position in the transmission or distribution line (see Figures 2–4 as illustrations).

Figure 2. First example of the shape of a wooden utility pole.

Despite the variety of the existing shapes and configurations, the number of elements basically composing an electrical pole is relatively limited. A wooden utility pole is generally composed of one or more wooden poles, one or more crossarms and multiple insulators responsible for the junction between the electrical cables and the pole. Figure 5 schematizes this assembling.

Using wooden utility poles has multiple advantages in comparison to concrete or steel utility poles (Bolin & Smith, 2011; SEMCO, 1992; Stewart, 1996)

Figure 3. Second example of the shape of a wooden utility pole.

Figure 4. Third example of the shape of a wooden utility pole.

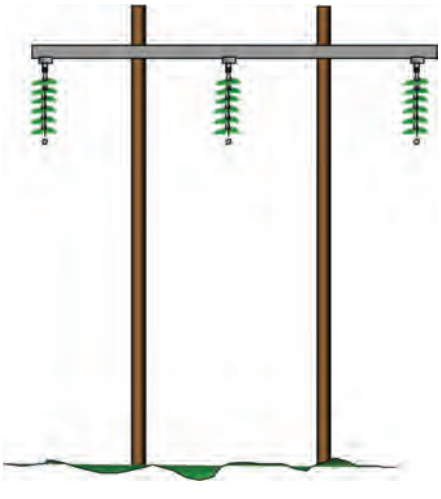


Figure 5. Basic components of a wooden utility pole: poles (brown), cross-arm (grey) and insulators (green) (Refsnæs, 2008).

– They are lighter and easier to transport on mountainous fields.
– They do not require earthing, which makes them interesting when lightning occur.
– They are easy to produce in wooded areas (e.g. Canada, Norway).
– They generally have a reduced environmental impact.
– They have interesting lifetimes, possibly going up to 75 years in favorable conditions.

Identifying the main threats for wooden utility poles enables to look for root causes of failures. This gives the possibility to estimate their effective remaining lifetime and optimize their replacement before any outage.

In their review on power line inspection procedures, Nguyen et al. (Nguyen, Jenssen, & Roverso, 2018) summarize some of the main common faults of power line components. They identify the apparition of cracks in the wooden poles as being one of the main failure to identify during visual inspection procedures. An additional review of the literature shows that there is need for inspection protocols enabling to recognize and assess cracks in timber structures in general (Dubois, Chazal, & Petit, 2002; Riahi, Moutou Pitti, Dubois, & Chateauneuf, 2016) and in wooden poles in particular (Morrell, 2012).

Identifying cracks is fundamental for two main reasons:

– First, as "stresses perpendicular to grain induce cracks which propagate longitudinally" (Coureau & Morel, 2005), we can consider multiple apparitions of significant cracks as being indicators of the presence of stress factors. This can for example suggest the existence of a localized area subject to harsher weather conditions (e.g. overload of ice and/or wind) (Wong & Miller, 2010) and prompt deepened analysis of the concerned region.
– Second, as cracks provide an access for external threats (e.g. fungi, insects, humidity) to potentially non-treated internal parts of the poles, their existence might accelerate the apparition of decay (Morrell, 2012; Refsnæs et al., 2006; SEMCO, 1992). This permanently alters the structural resistance of the pole and considerably increases its probability of failure.

3 WOOD PROPERTIES AND POTENTIAL INFLUENCE OF CLIMATIC VARIATIONS ON CRACK APPARITION

The theory of fracture mechanics has mainly been developed since the first half of the 20th century. Initiated by A.A. Griffith in 1920 (Griffith, 1921), it has then been popularized by G.R. Irwin in

1958 (Irwin, 1958) and is since being widely used to analyze the origins and consequences of crack apparition in physical objects. Focusing on the microscopic level, it enables to provide models describing the "mechanical behavior of cracked materials subjected to applied load" (Perez, 2017).

Multiple studies use this theory as a basis for the evaluation of crack growth in wooden structures (Barrett, Haigh, & Lovegrove, 1981; Coureau & Morel, 2005; Dubois et al., 2002; Riahi et al., 2016). A characterization of the structure is initially made on microscopic level to understand how wood behaves when it is subject to a modification of its external environment (load variation, climatic variation, etc.). Figure 6 shows the structure on microscopic level of a typical softwood. It highlights the anisotropic characteristic of wood and intuitively shows that cracks are more probable to occur parallel to the direction of growth of a three (longitudinal direction).

Wood being furthermore a viscoelastic material, its physical properties (e.g. modulus of elasticity, volume) are directly influenced by their environment. This is due to the hygroscopic behavior of wood (i.e. tendency to absorb humidity) and implies that physical properties of wood are highly sensitive to the meteorological properties of its surrounding (especially temperature and humidity) (Chaplain & Valentin, 2010; Hamdi, Moutou Pitti, & Saifouni, 2017; Lamy, 2016; Morrell, 2012; Refsnæs et al., 2006; Saifouni, 2014; Thybring, Lindegaard, & Morsing, 2009).

Because of the former functionalities of their cells during their living period and because of the variations in their environment during their growth, mechanical properties of timber-based structures can furthermore be locally modified. This includes structure modifications due to natural defects such as knots, rotten knots holes or cracks due to freezing lifeblood. Combined with the application of external loads (e.g. wind, ice on the wires in the case of wooden poles) and the modification of its internal structure due to temperature and humidity variations, there is a fertile ground for the apparition of cracks.

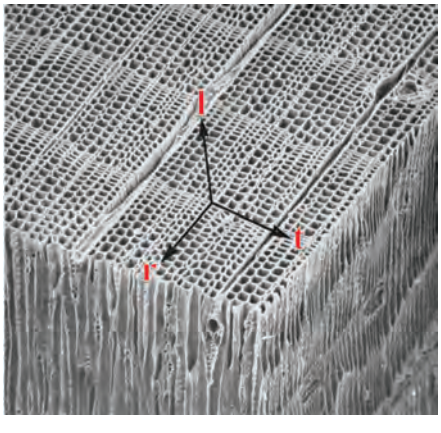


Figure 6. Typical softwood structure showing orientation of longitudinal (l), radial (r) and tangential (t) directions (Barrett et al., 1981).

4 DATA ACQUISITION AND PREDICTION METHODS

Utility companies in Norway use over 3.5 million wooden poles in their power grids to support over 25,400 km of electrical overhead lines (Eurelectric, 2010; Refsnæs et al., 2006). The Norwegian IT company eSmart Systems[1] is specialized in digital intelligence and uses artificial intelligence to support Statnett, Norway's TSO, as well as some of the main Norwegian DSOs (e.g., Lyse Elnett, Ringeriks-Kraft Nett, Troms Kraft Nett, Hafslund Nett). In particular, the algorithms used by eSmart Systems automatically identify specific objects and recognize pre-defined faults, such as cracks on wooden poles (see Figure 7 as an illustration). This enabled us to access a database of 17,352 geo-localized aerial pictures of wooden utility poles, from which 5383 are classified as cracked.

In most of the cases, two to three pictures of a unique utility pole were taken from different angles. This was done to ensure having accurate information for each of the observed poles without suffering from hidden information. We merged this information with the exact geographical coordinates of the electric poles, made available by the Norwegian Water Resources and Energy Directorate (NVE)[2]. We could thus analyze a dataset of 7653 geo-localized wooden utility poles, either classified as cracked or not.

Figure 7. Wooden pole where a crack has been localized on the mast (see rectangle).

1. eSmart Systems: www.esmartsystems.com.
2. NVE: www.nve.no.

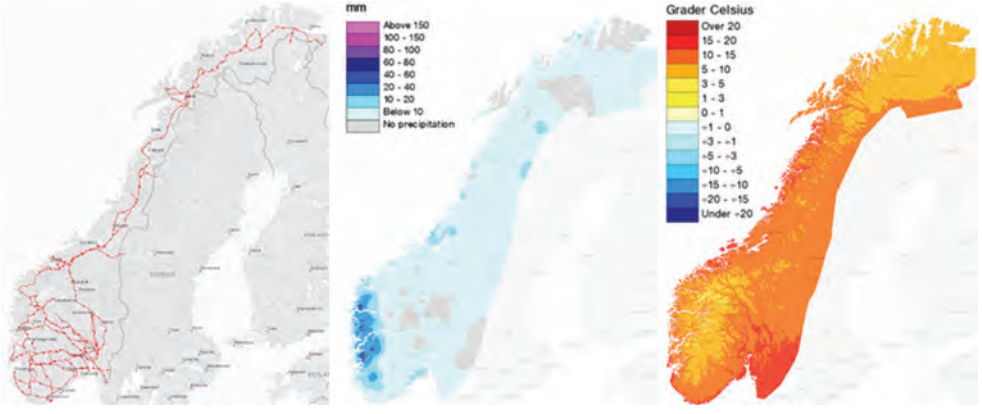


Figure 8. From left to right: main axes of the Norwegian electrical grid[6]; map of the precipitation in Norway on the 1st of August 2017[7]; map of the temperatures in Norway on the 1st of August 2017[7].

In parallel, seNorge[3] (created in collaboration between the NVE, the Norwegian Meteorological Institute[4] and the Norwegian Mapping Authority[5]) enables us to access daily observed (or interpolated) records of the climatic conditions in Norway. Especially, it enables us to access temperature and precipitation measures going as far back as 1957.

Figure 8 illustrates the type of information made available by NVE and seNorge. Using a scroll up/down feature of the websites, it is possible to move from a global and national overview up to a specific geo-localized point (in our case, the localization of the wooden utility poles).

Different approaches are considered in our work. The purpose is to create an indicator for the likelihood of crack apparition on wooden poles.

In order to benefit from the high granularity offered by the webservices used, we plan to use daily records of temperature and precipitation as potential predictors for a binary classification problem (labeling as cracked or not-cracked). Predictive features can be designed, that summarize at different granularities the daily weather data and extract relevant indicators that correlate with crack appearance. Considering an extreme reduction, we can for example summarize the intensity of the meteorological variation on a localized point into, e.g. a temperature coefficient and a precipitation coefficient. This would lead to a method using only two predictors when focusing on this classification problem.

Equation (1) provides an example of the type of coefficient c that can be used when focusing on a specific pole.

$$c = \sum_{i=2}^{n} \frac{|X_i - X_{i-1}|}{X_{max} - X_{min}} \quad (1)$$

Where n is the number of daily records since the installation of the wooden pole observed; i the enumeration index; X_i the value of the meteorological phenomenon observed on the specified location on day i (here in millimeters or in degrees Celsius); X_{i-1} the record of the same phenomenon on the same location on the previous day; X_{max} (resp. X_{min}) the maximum (resp. minimum) value of the observed phenomenon that has been recorded over the entire timestamp of observation on the specified location.

Alternatively, predictive features can be automatically learned from the raw temperature and precipitation time series using deep learning techniques. Such techniques, belonging to the class of artificial intelligence methods (and more especially, to the class of machine learning methods) are based on recursive analyses of data over time and/or over space, from which they identify and highlight step by step the most relevant characteristics.

High temperatures favor the proliferation of fungus, which weakens the structure of the wood. Furthermore, high humidity levels on extended periods might soften the wood and make it more sensitive to sudden external loads (e.g. wind or ice rain). Finally, the intrinsic properties of wood

3. seNorge: www.senorge.no.
4. Norwegian Meteorological Institute: www.met.no.
5. Norwegian Mapping Authority: www.kartverket.no.
6. https://temakart.nve.no/link/?link=nettanlegg.
7. http://www.senorge.no/index.html?p=senorgeny&st=weather.

lead it to easily accept slow variation of external loads and environmental conditions but make it particularly sensitive to sudden variations. These approaches will thus enable us to identify meteorological patterns favoring the apparition of cracks, as well as located regions where the likelihood of crack apparition will be higher.

An increase in the period of exposition to external factors leads to a rise of the probability of crack apparition. This implies that the age of the poles plays a big role in the suggested methods. However, part of this information might be missing. In such a case, we could consider a generic day of installation depending on the period of installation of the power line in the observed region.

5 DISCUSSION

The suggested methods enable to evaluate the role that temperature variations and precipitations have on the formation of cracks on wooden poles. These methods have the advantage to be flexible and easily integrated when accessing additional data sources, such as daily records of wind intensity and direction, humidity variations, clouds presence, etc. They are nevertheless highly dependent on two main facts:

- First, the initial classification of the poles as cracked or not. This is an important topic as the size of the cracks directly affects its detection by the algorithm used to classify the poles. There is thus a need for utility companies to define what should be considered as a problematic crack or not.
- Second, the information initially available on the poles themselves (e.g. age, maintenance tasks carried out). This information might be difficult to access because not necessarily well reported in the first phases of the grid installation.

Despite using relatively simple techniques and being highly dependent on initial parameters, the proposed methods represent a first approach in the analysis and handling of cracks in wooden poles. This information may in turn be useful for decision makers in the prioritization of additional inspection procedures and future maintenance tasks.

It is to mention that our paper only highlights preliminary results of an ongoing research, as the described methods have not yet been fully applied. Further work will thus focus on the extensive application and validation of these approaches and provide an in-depth analysis of the phenomenon of crack apparition on wooden poles by using additional real data from the Norwegian network.

6 CONCLUSION

Our paper highlighted the importance for utilities of early detection and analysis of cracks on wooden poles. We summarized how environmental conditions can directly affect the physical properties of wood and thus favor or limit the apparition of cracks on wooden poles. In order to better understand and predict their occurrence, we then suggested two approaches using pre-classified and geo-localized aerial pictures of cracked and non-cracked poles in combination with up to 60 years of meteorological measurements. Further, we saw that, despite being highly dependent on initial information, our approach might provide useful information for the generation of maintenance policies. This approach might finally be a good starting point for researchers wanting to combine fields of expertise such as structural study of wood on microscopic level and crack detection methods using image analysis.

REFERENCES

Barrett, J.D., Haigh, I.P., & Lovegrove, J.M. (1981). Fracture Mechanics and the Design of Wood Structures. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *299*(1446), 217–226. https://doi.org/10.1098/rsta.1981.0020.

Bolin, C.A., & Smith, S.T. (2011). Life cycle assessment of pentachlorophenol-treated wooden utility poles with comparisons to steel and concrete utility poles. *Renewable and Sustainable Energy Reviews*, *15*(5), 2475–2486. https://doi.org/10.1016/j.rser.2011.01.019.

Chaplain, M., & Valentin, G. (2010). Effects of Relative Humidity Conditions on Crack Propagation in Timber: Experiments and Modelling. In *World Conf. on Timber Engineering* (pp. 1–8). Retrieved from http://support.sbcindustry.com/Archive/2010/june/Paper_438.pdf?PHPSESSID=ju29kfh90oviu5o371pv47cgf3.

Coureau, J.L., & Morel, S. (2005). Non-Linear Fracture Mechanics Applied To Wood In Mode I. In *ICF11* (pp. 1–6). Italy. Retrieved from http://www.gruppofrattura.it/ocs/index.php/ICF/ICF11/paper/viewFile/10698/10044.

Dubois, F., Chazal, C., & Petit, C. (2002). Viscoelastic crack growth process in wood timbers: An approach by the finite element method for mode I fracture. *International Journal of Fracture*, *113*(4), 367–388. https://doi.org/10.1023/A:1014203405764.

Eurelectric. (2010). *EURELECTRIC's views on the use of creosote for impregnation of wooden poles in electricity networks*. Brussels, Belgium. Retrieved from http://www.eurelectric.org/media/44303/eurelectric_comments_on_creosote_2010-11-16-2010-030-1024-01-e.pdf.

Griffith, A.A. (1921). The phenomena of rupture and flow in solids. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 221*, 163–198. Retrieved from http://www.jstor.org/stable/91192.

Hamdi, S.E., Moutou Pitti, R., & Saifouni, O. (2017). Moisture driven failure monitoring in wood material: numerical analysis based on viscoelastic crack growth approach. In *CompWood 2017 – ECCOMAS Thematic Conference on Computational Methods in Wood Mechanics – from Material Properties to Timber* (pp. 187–198). Retrieved from https://www.researchgate.net/profile/Rostand_Pitti/publication/317329144_Moisture_driven_failure_monitoring_in_wood_material_numerical_analysis_based_on_viscoelastic_crack_growth_approach/links/59328dac0f7e9beee791a678/Moisture-driven-failure-monitoring-in-wood-material-numerical-analysis-based-on-viscoelastic-crack-growth-approach.pdf.

Irwin, G.R. (1958). Fracture. In F.S. (Ed.), *Elasticity and Plasticity/Elastizität und Plastizität* (Vol. 3/6, pp. 551–590). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/https://doi.org/10.1007/978-3-642-45887-3_5.

Lamy, F. (2016). *Crack analysis in wood under mechanical and climatic loadings: Contribution of Acoustic Emission*. Université de Limoges. Retrieved from https://tel.archives-ouvertes.fr/tel-01364070.

Morrell, J.J. (2012). *Wood Pole Maintenance Manual: 2012 Edition*. Oregon State University. Forest Research Laboratory. https://doi.org/http://ir.library.oregonstate.edu/concern/technical_reports/ft848r69b.

Nguyen, N. Van, Jenssen, R., & Roverso, D. (2018). Automatic Autonomous Vision-based Power Line Inspection: A Review of Current Status and the Potential Role of Deep Learning. *International Journal of Electrical Power & Energy Systems*.

Perez, N. (2017). *Fracture Mechanics*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24999-5.

Refsnæs, S. (2008). *Lineoppheng*. Retrieved from http://docplayer.me/36970970-Lineoppheng-sintef-energiforskning-as.html.

Refsnæs, S., Rolfseng, L., Solvang, E., & Heggset, J. (2006). Timing of wood pole replacement based on lifetime estimation. In *9th International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2006* (pp. 1–8). https://doi.org/10.1109/PMAPS.2006.360286.

Riahi, H., Moutou Pitti, R., Dubois, F., & Chateauneuf, A. (2016). Mixed-mode fracture analysis combining mechanical, thermal and hydrological effects in an isotropic and orthotropic material by means of invariant integrals. *Theoretical and Applied Fracture Mechanics*, *85*, 424–434. https://doi.org/10.1016/j.tafmec.2016.06.002.

Saifouni, O. (2014). *Modeling of rheological effects in materials: application to the mecanosorptive behaviour of wood*. Universit_e Blaise Pascal—Clermont-Ferrand II. Retrieved from https://tel.archives-ouvertes.fr/tel-01069026/.

SEMCO. (1992). *Wood pole maintenance*. *Bureau of Reclamation, Facilities Instructions, Standards, and Techniques* (Vol. 4–6). Retrieved from https://www.usbr.gov/power/data/fist/fist_vol_4/vol4-6.pdf.

Shiu, A., & Lam, P.-L. (2004). Electricity consumption and economic growth in China. *Energy Policy*, *32*(1), 47–54. https://doi.org/10.1016/S0301-4215(02)00250-1.

Stewart, A.H. (1996). *How long do wood poles last?* Fort Collins. Retrieved from http://www.americanpoleandtimber.com/wp-content/uploads/how-long-do-wood-poles-last.pdf.

Thybring, E.E., Lindegaard, B., & Morsing, N. (2009). Service Life Prediction of Wood Claddings by in-situ Measurement of Wood Moisture Content: Status after 5 years of Outdoor Exposure. In *40th Annual Meeting of the International Research Group on Wood Protection*. Beijing, China. Retrieved from https://www.researchgate.net/profile/Emil_Thybring/publication/262258436_Service_life_prediction_of_wood_claddings_by_in-situ_measurement_of_wood_moisture_content_status_after_5_years_of_outdoor_exposure/links/00b7d5372224cd1f31000000/Service-life-prediction-of-wood-claddings-by-in-situ-measurement-of-wood-moisture-content-status-after-5-years-of-outdoor-exposure.pdf.

U.S.-Canada Power System Outage Task Force. (2004). *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations* (Vol. 40). Washington, DC, US. Retrieved from https://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/BlackoutFinal-Web.pdf.

Wong, J.C., & Miller, M.D. (2010). *Guidelines for Electrical Transmission Line Structural Loading*. Reston, Virginia: American Society of Civil Engineers. https://doi.org/10.1061/9780784410356.

Yoo, S.H., & Kwak, S.Y. (2010). Electricity consumption and economic growth in seven South American countries. *Energy Policy*, *38*(1), 181–188. https://doi.org/10.1016/j.enpol.2009.09.003.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Improvement of the risk-based approach for evaluation of permanently plugged and abandoned oil and gas wells

H. Langdalen
University of Stavanger, Stavanger, Norway

E.B. Abrahamsen & J.T. Selvik
University of Stavanger, Stavanger, Norway
International Research Institute of Stavanger, Stavanger, Norway

H.P. Lohne
International Research Institute of Stavanger, Stavanger, Norway

ABSTRACT: A risk-based approach has been demonstrated to be capable of evaluating the quality of permanently plugged and abandoned petroleum wells. The quality measure in this context is the leakage risk, expressed in terms of probability of barrier failure and the leakage rate (consequence), where associated uncertainties are dealt with by means of probability distributions. In complex engineered systems, such as a barrier system in a permanently plugged and abandoned oil or gas well, it is reasonable to question whether the probability distributions provide adequate representations of the uncertainties. To improve the risk-based approach for plug and abandonment, and to contribute to more informed decisions by better reflecting the uncertainties upon evaluation of the leakage risk, in this paper, we propose an approach to assess the assumptions made and reflect the strength of knowledge, to complement the probability distributions. An example is included to illustrate the approach.

1 INTRODUCTION

The final phase in the life cycle of a petroleum well is permanent plug and abandonment, in order to e.g. prevent hydrocarbon leakage to the surface and pressure breakdown of the formations (Liversidge et al., 2006). On the Norwegian continental shelf (NCS), a significant number of offshore oil and gas wells will be entering their final phase in the coming decades and need to be permanently plugged and abandoned. Plug and abandonment (P&A) designs on the NCS are governed by the requirements and guidelines in NORSOK Standard D-010 (Standards Norway, 2013). In line with these requirements, P&A operations are considered prescriptive "one-size-fits-all" approaches (Arild et al., 2017). A criticism of the prescriptive approach is that it neglects the well-specific characteristics, which can differ from well to well, such as geological formations, reservoir qualities, wellbore schematics and surrounding marine environments, and is thus not cost-effective.

A quantitative risk-based approach for evaluating the containment performance of permanently plugged and abandoned oil and gas wells has been established as an alternative to the prescriptive approach (Arild et al., 2017). The risk-based approach is considered a "fit-for-purpose" alternative, which incorporates the well-specific characteristics. Here, the quality of the plugged and abandoned wells is measured by the leakage risk, which is expressed in terms of the probability of barrier failure within a given time period and the associated leakage rate at the seabed. The risk-based approach, demonstrated by Arild et al. (2017), builds on the leakage calculator presented by Ford et al. (2017) and can be seen as a risk assessment tool applicable in the recommended practice of "fit-for-purpose" well abandonment assessment proposed by Buchmiller et al. (2016).

In order to evaluate and assess the leakage risk for a permanently plugged and abandoned well, the risk-based approach identifies potential failure modes of each barrier element in the well, and assesses the failure probabilities and consequences. Thus, leakage risk can be considered as the two-dimensional combination of probability and consequence, which is widely criticized in the literature as being too narrow (see e.g. Aven (2014), pp. 28).

The workflow is described in detail by Arild et al. (2017). Simply stated, the assessment of failure probability is performed by means of

Bayesian analysis, to create a lifetime distribution of the well. Assessment of the consequences is done by calculating the leakage rates through identified leakage pathways, to provide a distribution of overall leakage rates at the seabed. For both assessments, uncertainties are dealt with by means of probability distributions (propagated through the use of Monte Carlo simulations).

A risk-based approach for evaluating the P&A of oil and gas wells should provide a broad, informative and balanced description of the leakage risk, as a means to support decision-making. To achieve this, proper treatment and communication of uncertainties are necessities (Flage & Aven, 2009). Looking at the current practice in the risk-based approach for evaluation of P&A, where uncertainties are treated by probability distributions, it is reasonable to ask whether all relevant uncertainties are fully reflected. Probability distributions are based on assumptions which may be more or less reasonable, and knowledge which may be strong or weak (or in-between). These two aspects are somewhat ignored, or not adequately reflected, by the probability distributions.

The present paper intends to improve the risk-based approach by implementing a semi-quantitative assessment of the uncertainties, in addition to the probability distributions, to incorporate the aspects mentioned above. The improved approach goes beyond probabilities to assess and express uncertainties, by focusing on the knowledge, which forms the basis for the assessment. A strength-of-knowledge categorization, in line with the scoring used by Flage & Aven (2009), will be the starting point of the semi-quantitative analysis to reveal uncertainties hidden in the background knowledge. In addition, the concept of assumption deviation risk, introduced by Aven (2013), will be applied as a tool to reveal the effect of potential deviations in the assumptions. Assumption deviation risk assessment has been demonstrated in recent time to be a useful tool in the context of highlighting critical assumptions (see e.g., Berner & Flage (2016), Khorsandi & Aven (2017)). The improved approach for risk-based P&A intends to provide a more complete risk description, such that the decision-makers have greater decision support when reasoning and deliberating upon the decisions to be made (Aven & Zio, 2011).

The remainder of this paper is organized as follows: In Section 2, we briefly introduce the probability and consequence assessments in the risk-based approach, respectively. In Section 3, we discuss the weaknesses associated with the current practice in the risk-based approach. Section 4 presents the suggested improvements. Section 5 illustrates the improved approach through an example, while Section 6 discusses and evaluates the improved approach. Section 7 concludes.

2 INTRODUCING THE CURRENT RISK-BASED APPROACH

The current risk-based approach for evaluating the containment performance of plugged and abandoned oil or gas wells is based on the assessments of probability of failures and consequences. Here, we briefly introduce the two assessments. See Arild et al. (2017) and the references therein for further details.

2.1 *Probability assessment*

A key quantity of interest when evaluating the quality of a P&A design is the lifetime of the well, i.e. reflecting how long it will take before the barrier system starts to leak (Arild et al., 2017). Commonly used lifetime distributions are the exponential and Weibull distributions (Singpurwalla, 2006). In order to establish a lifetime distribution, historical data are often used as a basis to perform statistical inference, such as maximum likelihood estimation (MLE), to obtain estimates of the parameters which best explain the distribution of the observed data.

To be applicable in estimating the parameters of a lifetime distribution, traditional methods such as MLE require that some failures have occurred (Singpurwalla, 2006). On the NCS, however, none of the 334 wells which are assumed plugged and abandoned are said to have failed since they were abandoned; they are considered censored observations (Arild et al., 2017). One approach which is deemed feasible when working with completely censored data is Bayesian analysis (Singpurwalla, 2006).

In Bayesian analysis, the objective is to use known quantities along with a specified parametric expression to make inferences about the unknown quantities or parameters (Singpurwalla, 2006). Since the parameters are unknown, we assign prior distributions for the parameters to reflect our lack of knowledge about the parameter values a priori any evidence. After some data are observed, Bayes' formula, expressed in terms of probability distributions, is used to update the prior beliefs into posterior distributions of the parameters. The quantity of interest in the probability assessment is the lifetime of the well and not the parameters estimated by Bayes' formula. By drawing parameters of interest from the posterior distribution, a posterior predictive distribution of the lifetimes can be generated, which is a distribution of unobserved observations (predictions), conditional on the observed data.

2.2 *Assessing the consequences of a failure*

If a well barrier fails, leakage of hydrocarbons to the seabed can occur. Ford et al. (2017) have developed a simple leakage calculator to assess the leakage potential from a well barrier system. This

calculator estimates the cumulative leakage rate through potential leakage pathways for each well barrier. We refer to Ford et al. (2017) for detailed explanations of the leakage calculations. The leakage pathways considered in this paper are: leakage through bulk cement given by Darcy's law (e.g. Godøy et al. (2015)); leakage through fractures or cracks in the cement (e.g. Sarkar et al. (2004)); and, leakage through the micro-annuli (e.g. Aas et al. (2016)). These pathways are represented by the following equations, respectively:

$$Q_1 = \left(\frac{kA}{\mu L}\right)\left(\Delta P - \rho g L \cos\theta\right) \quad (1)$$

$$Q_2 = \left(\frac{h^3 \cos\alpha}{12\mu}\right)\left(\frac{\Delta P}{L}\right)W \quad (2)$$

$$Q_3 = \left(\frac{\pi R_c \Delta P}{6\mu L}\right)\delta R^3 \quad (3)$$

where Q = flow rate; k = cement permeability; A = cross-section of the cement plug and/or annulus; μ = reservoir fluid viscosity; L = length of the plug or annular cement; ΔP = pressure difference over the cement plug and/or annuli; ρ = density of the reservoir fluid; g = gravitational acceleration; θ = inclination of the well at the depth of the plug; h = fracture aperture; α = orientation of the fracture; W = fracture width; R_c = casing diameter; and δR = micro-annuli gap.

3 ISSUES WITH THE CURRENT RISK-BASED APPROACH

In the risk-based approach, as in any risk assessment, we make a number of assumptions that are more or less explicitly stated. A single assumption may be formulated as $X = x_0$, where the value x_0 is fixed, such as an increasing failure rate ($X = x_0 = \beta = 1.5$, in a Weibull function). Since the result of the leakage risk assessment is conditional on the assumptions made being true, it is important to understand the uncertainties related to these assumptions. We can classify the assumptions as either probability-influencing assumptions or consequence-influencing assumptions (which is a broad interpretation of the three areas of assumptions listed by Khorsandi & Aven (2017)).

3.1 *Weaknesses with the probability assessment*

By virtue of Bayes' theorem, the posterior predictive distribution is based on the posterior distribution of some parameters θ. The true values of such parameters are unknown. By using an alternative prior distribution, or probability distribution, a different posterior predictive distribution would be generated. Paramount for the assessment is to understand the basis for the resulting predictive distribution. Other technical difficulties with Bayesian analysis are summarized by e.g. Ferson (2005).

The historical data extracted from the Norwegian Petroleum Directorate (2017) database can be questioned, regarding the validity and value of information with respect to when an abandoned well will fail. Censored data, as we have here, imply that we have not detected any failures (leakages) since the wells were abandoned. For a leakage to be detected, a failure must have taken place. In that sense, it is reasonable to assume that no detected leakages imply zero failures. A failure, on the contrary, does not automatically imply a detectable leakage. So the true survival times of the data are actually unknown.

The data contain survival times of abandoned wells, designed according to NORSOK Standard D-010. As the risk-based approach intends to incorporate well-specific characteristics, such as flow potential, it is contradictory to use all available data as a basis for the probability assessment. Further investigation of the data shows that observations differ with respect to well-specific properties, such as geology and flow potential. It is reasonable to question whether a survival time from a well in a reservoir with, say, limited flow potential is relevant when evaluating a P&A design for a well in a reservoir with high flow potential. The risk-based approach aims to justify alternative P&A designs, which are "fit-for-purpose". Utilizing the whole sample when evaluating the leakage risk is therefore a radical assumption, as it takes into account censored lifetimes from, most likely, stricter P&A designs.

3.2 *Weaknesses with the consequence assessment*

Challenges with the leakage rate assessment relate to uncertain input values, as the models (Equations 1–3) are established on strong knowledge, according to criteria in e.g. Aven (2014), pp. 139.

There are two categories of inputs in the consequence approach: (1) uncertain inputs that can be described by probability distributions, and (2) known inputs related to design variables, scenarios or low importance inputs (Arild et al., 2017). The former is the focus of the present paper. With respect to Equations 1–3, Table 1 summarizes the uncertain inputs of interest. Both the values and distributions (here, a triangular distribution) given in the second column are assumed. These values are difficult to measure exactly, and they impose uncertainties. The degree of uncertainty is case-specific

Table 1. Uncertain parameters (Arild et al., 2017).

Uncertain parameters	Values (min, most likely, max)
Cement permeability	0.1, 0.5, 5.0 μD
Micro-annuli gap	3, 20, 70 μm
Fracture aperture	10, 50, 200 μm
Inclination of fracture	0, 30, 70°
Fracture width	1, 2, 3 mm

and subject to judgment by the assessor (Flage et al., 2013).

4 DISCUSSION OF SOME POSSIBLE IMPROVEMENTS FOR THE RISK-BASED APPROACH

The common denominator for the aspects discussed in the previous section is lack of knowledge when quantitatively assessing the leakage risk. A limited amount of information is available on the subject matter, and determining whether the parametric functions and input parameters are appropriate is challenging. Decision-making under such conditions is also challenging. Not all uncertainties can be fully expressed or transferred into quantitative formats. The background knowledge, on which the probabilities of failure and leakage rates are based, can hide uncertainty. The assumptions made are known to potentially deviate in reality, and any deviation that could affect the risk picture needs to be highlighted and assessed. We believe the following approaches can improve the treatment of uncertainty in the risk-based approach: (1) crude strength-of-knowledge assessment (SoK) and (2) assumption deviation risk assessment.

4.1 *Crude strength of knowledge assessment*

In the probability and consequence assessments, uncertain inputs are used as the basis for the assessments. Subjective probabilities (distributions) are assigned to these inputs, expressing our degree of belief about the values of these parameters, conditional on our background knowledge. To reflect the strength of this background knowledge, a crude SoK assessment is recommended. A categorization in line with the scoring by Flage & Aven (2009) is used as a basis for an assessment of the SoK. Here, the background knowledge is categorized as strong if all the following criteria are met (Flage & Aven, 2009):

- The assumptions made are seen as very reasonable (s1)
- A large amount of reliable data is available (s2)
- There is broad consensus among experts (s3)
- The phenomena involved are well understood (s4).

If, on the other hand, at least one of the following criteria is true, the background knowledge is classified as weak (Flage & Aven, 2009):

- The assumptions made are strong simplifications (w1)
- Data are non-existent or unreliable (w2)
- There is a lack of consensus among experts (w3)
- The phenomena involved are not well understood (w4).

An in-between background knowledge is classified as moderate. The SoK assessment intends to capture uncertainties which are not easily transformed or expressed quantitatively, e.g. by probabilities.

4.2 *Assumption deviation risk assessment*

The concept of assumption deviation risk, introduced by Aven (2013), highlights the risk with respect to the (main) assumptions on which the quantitative risk assessment is based. In general, Aven (2013) suggests focusing on the magnitude of the deviation, the degree of belief in this magnitude occurring, how this deviation will influence the risk, and the SoK related to the phenomena affecting the assumption (as described above).

Thus, the concept of assumption deviation risk goes beyond traditional sensitivity and uncertainty analysis. Sensitivity analysis, such as asking "what if" questions, is informative, as it produces a range of outcomes based on different input values. Assumption deviation risk extends this type of analysis, as it assesses the risk of deviations, covering an assessment of the consequences of the deviations, uncertainties related to the outcomes and a crude SoK assessment. The concept of assumption deviation risk aims to better communicate the uncertainties related to the leakage risk to the decision-maker, by explicitly assessing and identifying the influence of any deviation in the assumptions. The assessment is systematic, as it intends to identify critical assumptions, analysing the risk of deviations in those assumptions, and is a means to understanding assumptions which may deviate far ahead in the future (Khorsandi & Aven, 2017). Requirements, such that the well barrier should perform its purpose for eternity, make it important to understand how the leakage risk may change over the lifetime of an abandoned well. This is better understood when we understand how deviations in the assumed states can occur. The main concerns in the leakage rate assessment are the negative consequences; hence, we can restrict our focus to deviations that worsen the outcomes.

Often, as the decision-making problem becomes more complex, the number of assumptions has a tendency to increase and the assumption deviation risk assessment becomes tedious and time-consuming. To make the improved approach practical, we suggest handling each assumption systematically, in line with guidance suggested by Berner & Flage (2016), who focus on six so-called settings that the decision-maker faces when making assumptions in a risk assessment. These settings are classified by the degree of belief in deviation from the initial assumption, the effect of such deviation and the SoK related to the assumption. When we know what setting we are facing, a general approach to handle the assumption is suggested. We must emphasize that the guideline should not to be used as a mechanistic framework and must be adapted to the case of interest.

In addition, the assumption deviation risk assessment will increase the likelihood of detecting surprises. Two categorizes are of particular concern with respect to surprises: a moderate/high belief in deviation, and when the SoK is moderate/weak. Assessment of the assumptions will highlight such potential surprises. This is a way of revealing black swans (Aven, 2014).

5 A CASE STUDY

A synthetic vertical gas well is considered to be plugged and abandoned. The well, well barrier and reservoir characteristics are based on Arild et al. (2017) (except the inclination of the well). For simplicity, we refer to the values in Table 1 for the uncertain parameters and to Arild et al. (2017) for an overview of the known parameters of interest. The secondary barrier, which functions as a backup to the primary, is placed right on top of the primary barrier, and the combination of the two is treated as one barrier (plug 1+2). In addition, there is a surface barrier (plug 3) located at 300 m TVD (true vertical depth).

Let us say that the well appears to have a low leakage risk (before any assessment is conducted). Negligible issues, such as scaling and cross flows, were observed during the production phase. The only concern is an unconsolidated sandstone formation around the plug 1+2, indicated by an acoustic log. However, the overall impression is that the NORSOK Standard D-010 requirement of having 100 m thick barriers is too strict. In other words, the requirement is questioned in terms of its cost-effectiveness. A more cost-effective design, with shorter cement plug lengths, is considered as an alternative.

Reducing the cement plug lengths intuitively increases leakage risk and needs to be justified before implementation. A deviation in some of the assumptions, which form the basis for the assessment, can be critical, as the consequence threshold will be further reduced. The aim of the analysis is to illustrate how assumptions can be assessed according to the methods presented in this paper, to achieve better decision support. The case study is divided into three parts. First, the current risk-based approach is applied to assess the leakage risk. Then, the suggested improvements discussed in this paper are applied to the same case. Finally, we compare the current risk-based approach with the improved approach, in terms of decision support.

5.1 *The current risk-based approach*

The alternative P&A design, with shorter cement plug lengths needs to be justified before implementation. This is done by the leakage risk, which depends on the probability of failure and the associated leakage rates. These are estimated by the procedures introduced in Section 2.

5.1.1 *Probability assessment*

Following the approach described in Section 2.1 and the example of Arild et al. (2017), we establish a Weibull distribution to reflect the probability of failure within a given time for this well. We assume that the shape parameter is 1.5, indicating that the failure rates increase with time. The scale parameter is assumed to have a uniform distribution, between 0 and 1000 years. Basically, this means that whether failures are likely to occur in 10 years or 1000 years is unknown. The data used to update our prior belief about the scale parameter are the 334 censored observations of wells assumed to be permanently abandoned. Probability of failure within a certain time period is then predicted from the posterior predictive distribution in Figure 1. The probability of failure within the first 100 years is approximately 5%, if the assumptions hold true.

5.1.2 *Consequence assessment*

If a failure occurs, the quantity of interest is the leakage rate. Here, we follow the calculations from

Figure 1. Posterior predictive probability of failure times given the initial assumptions.

Equations 1–3 for both the surface barrier and the combined primary and secondary barrier. Leakage to the seabed is determined by the minimum leakage rate through the surface barrier or the combined plug. The combined plug has a significantly lower flow rate, due to having twice the thickness of the surface barrier, presented with sensitivity to different plug lengths in Figure 2, where the solid lines represent our 90% certainty that the flow rate is below these flow rates, the dashed lines are the most likely rates, and the stippled lines reflect our 90% certainty that the flow rates are above these flow rates.

When the plug lengths are above the prescriptive length of 100 m, there is a negligible increase in the flow rates from reducing the plug lengths. However, a reduction in plug lengths which are shorter than 100 m, imposes a relatively significant increase in flow rate.

5.1.3 *Results from the current risk-based approach*
The probability and consequence assessments resulted in low leakage risk. The probability of failure within the next 100 years is predicted to be 5%. With plug lengths of 50 m, there is 80% probability that the leakage rate will be in the range of 1.303×10^{-6} m^3/s to 5.087×10^{-5} m^3/s, with a most likely leakage rate of 1.829×10^{-5} m^3/s. If we assume that the reservoir fluid composition is mostly methane, these flow rates correspond to a yearly release in the range of 0.021 to 0.831 ton, with a most likely yearly release of 0.298 ton. With the prescriptive plug lengths of 100 m, there is an 80% probability that the flow rate is in the range of 7.667×10^{-7} to 2.780×10^{-5} m^3/s (0.012 to 0.454 ton/year), with a mean flow rate of 1.021×10^{-5} m^3/s. (0.457 ton/year). The leakage rate is almost doubled by reducing the plug lengths from 100 to 50 m. These results represent the base case of this analysis. As the assumptions made in the probability and consequence assessment may deviate, we need to assess their influence on the leakage risk, in order to justify a plug length of 50 m.

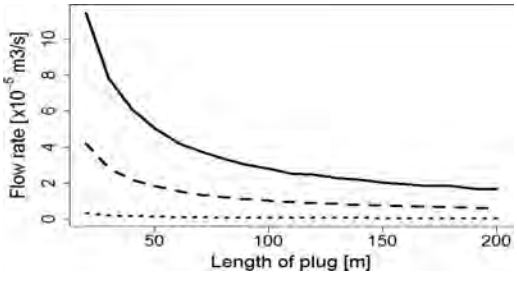


Figure 2. Flow rates through the primary and secondary plug for different cement plug lengths.

5.2 *The improved risk-based approach*
Following the approach described in Section 4, we intend to provide a more detailed understanding of the assumptions, and therefore the leakage risk, before making a decision. For simplicity, we have only considered a few assumptions in this paper. The assumption deviation risk assessment is summarized in Table 2, and complements the results presented in Section 5.1. All the assumptions are assessed in terms of degree of belief in deviation, effect of such deviation on the leakage risk and the SoK regarding the assumption. Some of the assumptions show significant influence on the leakage risk.

Assumptions classified by a moderate or high belief in deviation from the initial assumption and moderate or high influence on the leakage risk, based on weak background knowledge, are the most critical in the leakage risk assessment (assumption Nos. 1, 2 and 3). For assumption No. 1, the SoK is weak. This is controversial, as weak knowledge discourages the establishment of probability distributions and raises questions regarding the final result. A uniform prior distribution is the most appropriate option, since it intends to reflect the epistemic uncertainty to some degree. The goodness of this distribution is, however, open for debate, as we do not know with certainty which prior distribution or interval to choose. A deviation in the range of the uniform prior distribution greatly affects the probability of failure.

Assumption No. 2, regarding the validity of the historical data, is also critical. As the historical data are based on P&A designs in line with NORSOK Standard D-010, the data are poor representations of this well. There is a trade-off between large sample and relevant data, which the assessors need to consider and reflect. Based on sound engineering thinking, it is likely that this well, with shorter plug lengths, will have a higher probability of failure within the first 100 years, compared to the values in Figure 1.

The assumption deviation risk assessment revealed that most of the assumptions are believed to deviate to some degree (Nos. 4, 6, 7 and 8), in addition to having moderate influence on the leakage risk. As the SoK is not weak for these assumptions, probability distributions may be established. These assumptions may deviate, reflected by not being based on strong background knowledge or by the fact that we believe in a deviation; and the number of such assumptions is of great concern. If deviations from more than one of the assumptions take place simultaneously, the cumulative influence on the leakage risk can be significant, despite a deviation in each assumption, separately, showing little influence. This is highly relevant for the

Table 2. Summary of the simplified assumption deviation risk assessment for the case study in Section 5.

No.	Assumption influence	Assumption ($X = x_0$)	SoK	Belief in deviation	Deviation	Effect of deviation	SoK evaluation
1	P	Scale parameter has a uniform prior distribution on the interval [$x_{min} = 0$, $x_{max} = 1000$], resulting in: $P(T < 100) = 5\%$, $P(T < 1000) = 76\%$.	Weak	Moderate	$x_{max,1} = x_{max} + 1000$ $x_{max,2} = x_{max,1} + 1000$	$P(T < 100 \mid x_{max,1}) = 2\%$ $P(T < 1000 \mid x_{max,1}) = 49\%$ $P(T < 100 \mid x_{max,2}) = 1\%$ $P(T < 1000 \mid x_{max,2}) = 35\%$	w1, w2, w3, w4, Critical
2	P	Historical data are representative for this well. $X = x_0$ = survival times.	Weak	Moderate/ high	Failure times could increase, $X > x_0$.	Moderate/high	w1, w2, w4 Critical
3	C	Triangular distribution for all uncertain parameters, $X = x_0 = T$ (min, mode, max).	Weak	Moderate/ high	High	High	w1, w2, w4 Critical
4	P	Failure rate (shape parameter) is constant, $X = x_0 = 1.5$.	Moderate	Moderate/ high	$X \in [1.0, 2.0]$	High $\Delta P(T100\mid X) \in [121\%, -36\%]$	S1, w2, w4 Critical
5	P	Failure rate is increasing, $X = x_0 = \beta > 1.0$ (see No. 4).	Moderate	Low	Low	High, but not likely to have a decreasing failure rate.	s1, s3, w4
6	C	Cement permeability has a max value of $X = x_0 = 5.0$ μD.	Moderate	Moderate/ high	$X \in [0.5, 2.0]$ μD	Low, flow through bulk cement is negligible.	s2, s4, w1, w3
7	C	All three leakage pathways occur at the same time (with probability of 1) if a failure occurs.	Moderate	Strong	Moderate/high. (i.e. only one of the three can occur)	Moderate	Conservative assumption
8	C	All the barriers have the same properties.	Strong	Low	Low	Moderate	s1, s2, s4
9	C	Maximum pressure difference over the plug or annulus.	Moderate	Low	Low	Low (not likely to be worse)	Conservative assumption
10	C	Micro-annuli gap is given as a triangular distribution, T(3, 20, 70) μm.	Weak/ moderate	Moderate	Mode and max can deviate with + 20-70 μm (due to cement shrinkage potential).	Moderate/high	w2, w4

P: Probability assessment; C: Consequence assessment; SoK: Strength-of-knowledge. Sok Evaluations: see Section 4.1; $\Delta P(T) = [P(T|xi) - P(T|x0)]/P(T|x0)$.

evaluation of P&A designs but beyond the scope of the present paper.

5.3 *Evaluating the results*

Based on the assumption deviation risk summarized in Table 2, it is now possible to argue that a reduction in the plug length from 100 to 50 m is not justified. If a decision were made exclusively on the probability and consequence assessment, the conclusion would most likely be different. The two-dimensional leakage risk of probabilities and consequences is low. Seeing beyond such a narrow perspective on risk, we see that there are some uncertainties which should not be ignored. An incorrect choice of the prior distribution can greatly affect the probability of failure, for example. In addition, the data may provide too optimistic information about expected survival times if the prior distribution's ability to support longer lifetimes is poor. The leakage rate can also be significantly increased if the assumed value of the micro-annuli gap is slightly too low, compared to the true future value. The latter can take place if the annuli cement has expanding properties, such that the cement, over time, may expand radially towards the formation and cause a greater inner micro-annuli gap (Baumgarte et al., 1999).

Ideally, the justification should be made with reference to some risk acceptance criteria. This has not been established at the time being. The final decision should be made after a managerial review and judgment process, according to the idea of risk-informative decision-making (e.g. Aven & Zio (2011)).

6 DISCUSSION

The main aim of this paper was to illustrate how the risk-based approach for evaluating of the containment performance of a plugged and abandoned well can be improved by assessing uncertainties beyond probabilities. The methodology is systematic in its treatment of the assumptions. It should be mentioned that no effort was conducted to quantify what a high degree of sensitivity actually corresponds to, although this would affect the number of critical assumptions. Does a deviation in an assumption, resulting in a 20% change in the leakage rate indicate high sensitivity? The challenge is related to how to justify what increase in the leakage risk is acceptable, which is beyond the scope of this paper.

The quantitative approach for assessing the leakage risk is attractive. This type of approach in plug and abandonment evaluation is new to the industry and has seen little practical experience. One of the main challenges that remains to further develop the approach is to establish criterion which can be used in the justification of a specific design. Then the lifetime distribution and calculated leakage rates complemented by an assumption deviation risk assessment would provide a powerful tool for making risk-informed decisions.

Many of the assumptions made are based on relevant literature on the subject matter. The key point is that these values may be more or less relevant for the specific case. Assessing the assumptions helps to declare whether the assumptions are appropriate or not. However, it requires the assessor to have a sound understanding of what it takes for an assumption to occur. If the knowledge is weak, we cannot say something with certainty about the deviation in an assumption and the effect of such a deviation. We should communicate this fact to the decision-maker. The assumption deviation risk assessment provides a deeper understanding of the phenomena involved in a P&A operation, thus revealing uncertainties beyond the capability of the current risk-based approach. The fact that barriers should perform their functions to eternity requires that assumption deviations are taken into consideration when justifying an alternative P&A design.

A deviation in an assumption which increases the leakage risk is undesirable. To avoid such negative outcomes, it is common to apply conservative assumptions, which are often strong simplifications. In the current risk-based approach conservative assumptions are often made, such as assumption Nos. 7 and 9 in Table 2. One reason why we often resort to conservative assumptions is lack of knowledge. Rather than searching for increased knowledge, it is less time- and resource-consuming to make a conservative assumption. A danger of such a practice is that a too high risk picture is presented to the decision-maker (Aven, 2016). Basic safety management principles would then imply an implementation of costly risk-reducing measures. Thus, the careless use of conservative assumption works against the purpose of the risk-based approach to provide decision support regarding an alternative, more cost-effective P&A design.

7 CONCLUSION

In this paper, we have shown how assumption deviation risk assessment can improve the treatment and reflection of uncertainties. The assumptions are assessed separately with respect to the analyst's degree of belief in deviation, the sensitivity of the leakage risk to such deviation and the SoK related to the assumption. The fact that assumptions are

known to potentially deviate more or less should not be ignored. Not all uncertainties are easily transformed and expressed quantitatively, and they require a more semi-quantitative approach to be fully reflected, to ensure appropriate decision support. By complementing the leakage risk with an assumption deviation risk assessment, more informed decisions can be made.

A case study was performed to illustrate how the improved risk-based approach provides more informed decision support than the current approach. In this case, an alternative P&A design was evaluated. The assumption deviation risk assessment highlighted some critical assumptions, which led to a decision that differed from what we would decide upon if it were made exclusively on the leakage risk.

REFERENCES

Aas, B., Sørbø, J., Stokka, S., Saasen, A., Godøy, R., Lunde, Ø. & Vrålstad, T. 2016. Cement Placement with Tubing Left in Hole during Plug and Abandonment Operations, SPE-178840-MS. *IADC/SPE Drilling Conference and Exhibition 2016.* Forth Worth, TX, USA: Society of Petroleum Engineers.

Arild, Ø., Lohne, H.P., Majoumerd, M.M., Ford, E.P. & Moeinikia, F. 2017. Establishment of a Quantitative Risk-Based Approach for Evaluation of Containment Performance in the Context of Permanently Plugged and Abandoned Petroleum Wells, OTC-27711-MS. *Offshore Technology Conference 2017.* Houston, TX, USA: Offshore Technology Conference.

Aven, T. & Zio, E. 2011. Some considerations on the treatment of uncertainties in risk assessment for practical decision making. *Reliability Engineering & System Safety,* 96, 64–74.

Aven, T. 2013. Practical implications of the new risk perspectives. *Reliability Engineering & System Safety,* 115, 136–145.

Aven, T. 2014. *Risk, surprises and black swans: Fundamental ideas and concepts in risk assessment and risk management*, Routledge.

Aven, T. 2016. On the use of conservatism in risk assessments. *Reliability Engineering & System Safety,* 146, 33–38.

Baumgarte, C., Thiercelin, M. & Klaus, D. 1999. Case Studies of Expanding Cement To Prevent Microannular Formation, SPE-56535-MS. *SPE Annual Technical Conference and Exhibition.* Houston, TX, USA: Society of Petroleum Engineers.

Berner, C. & Flage, R. 2016. Strengthening quantitative risk assessments by systematic treatment of uncertain assumptions. *Reliability Engineering & System Safety,* 151, 46–59.

Buchmiller, D., Jahre-Nilsen, P., Sætre, S. & Allen, E. 2016. Introducing a new Recommended Practice for Fit for Purpose Well Abandonment, OTC-27084-MS. *Offshore Technology Conference 2016.* Houston, TX, USA: Offshore Technology Conference.

Ferson, S. 2005. Bayesian methods in risk assessment. Technical report: http://www.ramas.com/. Unpublished Report prepared for the Bureau de Recherches Geologiqueset Minieres (BRGM): Applied Biomathematics.

Flage, R. & Aven, T. 2009. Expressing and communicating uncertainty in relation to quantitative risk analysis (QRA). *Reliability & Risk Analysis: Theory & Applications,* 2, 9–18.

Flage, R., Baraldi, P., Zio, E. & Aven, T. 2013. Probability and Possibility-Based Representations of Uncertainty in Fault Tree Analysis. *Risk Analysis,* 33, 121–133.

Ford, E.P., Moeinikia, F., Lohne, H.P., Arild, Ø., Majoumerd, M.M. & Fjelde, K.K. 2017. Leakage Calculator for Plugged and Abandoned Wells, SPE-185890. *SPE Bergen One Day Seminar 2017.* Bergen, Norway: Society of Petroleum Engineers.

Godøy, R., Fontan, M., Capra, B., Kvalsund, R. & Poupard, O. 2015. Well Integrity Support by Extended Cement Evaluation – Numerical Modeling of Primary Cement Jobs, SPE-177612-MS. *International Petroleum Exhibition and Conference 2015.* Abu Dhabi, UAE: Society of Petroleum Engineers.

Khorsandi, J. & Aven, T. 2017. Incorporating assumption deviation risk in quantitative risk assessments: A semi-quantitative approach. *Reliability Engineering & System Safety,* 163, 22–32.

Liversidge, D., Taoutaou, S. & Agarwal, S. 2006. Permanent Plug and Abandonment Solution for the North Sea, SPE-100771-MS. *SPE Asia Pacific Oil and Gas Conference and Exhibition 2006.* Adelaide, Australia: Society of Petroleum Engineers.

Norwegian Petroleum Directorate. 2017. *Norwegian Petroleum Directorate FactPages* [Online]. Norwegian Petroleum Directorate. Available: http://factpages.npd.no [Accessed October 2017].

Sarkar, S., Toksöz, M.N. & Burns, D.R. 2004. *Fluid Flow Modeling in Fractures*, Massachusetts Institute of Technology. Earth Resources Laboratory.

Singpurwalla, N.D. 2006. *Reliability and risk: A Bayesian perspective,* Chichester, England, John Wiley & Sons, Ltd.

Standards Norway 2013. NORSOK Standard D-010 Rev. 4 – Well integrity in drilling and well operations.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

The use of bond graph modelling in polymer electrolyte membrane fuel cell fault diagnosis

A. Vasilyev & J. Andrews
Department of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, UK

L. Mao & L.M. Jackson
Department of Aeronautical and Automotive Engineering, Loughborough University, UK

ABSTRACT: As a possible alternative energy source, hydrogen fuel cells, especially Polymer Electrolyte Membrane (PEM) fuel cells, have received much more attention in the last few decades, which have already been equipped in many applications. A series of studies have been devoted to PEM fuel cell fault diagnosis to ensure its reliability during its lifetime, but due to the complexity of PEM fuel cell systems and incomplete PEM fuel cell test protocols, it is difficult to test various PEM fuel cell failure modes, thus the performance of fault diagnostic techniques cannot be fully investigated. On this basis, it is necessary to develop a reliable PEM fuel cell model with capability of simulating various PEM fuel cell faults. In this study, a hybrid model is developed to represent the behavior of PEM fuel cells in both continuous and discrete-time domains. With a continuous-time domain sub-model, various aspects of PEM fuel cell behavior can be simulated, including fluid, thermal, and electro-chemical dynamics. Moreover, the PEM fuel cell failure modes are implemented with stochastic Petri nets in the discrete-time domain. Based on the developed hybrid model, various PEM fuel cell failure modes can be simulated and their effects on the system performance can be observed. With the simulated data under different conditions, the performance of fault diagnostic techniques can be better evaluated by studying their performance in different failure mode scenarios.

1 INTRODUCTION

Due to the characteristics such as zero-emission and high efficiency, the PEM fuel cell has attracted more attention as an alternative energy source. In the last few decades, PEM fuel cells have been equipped in several systems, including automotive, consumer devices, and stationary power systems.

However, the reliability of PEM fuel cell during its lifetime is still a main barrier for further commercialization. To address this, several studies have been devoted to PEM fuel cell fault diagnosis, which could detect and isolate PEM fuel cell abnormal performance, thus mitigation strategies can be taken to recover and extend the fuel cell performance. Based on the methods adopted, these studies can be loosely divided into two categories, model-based techniques and data-driven approaches [Petrone et al. 2013, Zheng et al. 2013].

In model-based techniques, a PEM fuel cell model should be developed to express the system behavior, and the fault can be identified by calculating the residual between the model outputs and actual measurements [Kamal and Yu 2011, Ohs et al. 2011, Zeller et al. 2010]. With data-driven approaches, the features indicating the fuel cell condition would be extracted from the measurements, and the fuel cell state can be determined by applying pattern recognition algorithms to the extracted features [Mao et al. 2017, Placca et al. 2010, Rubio et al. 2010, Steiner et al. 2011, Zhongliang et al. 2015].

From the previous studies, data-driven approaches are more widely used in PEM fuel cell fault diagnosis [Zheng et al. 2013]. The main reason is that the PEM fuel cell contains physical interactions consisting phenomena from fluidic, thermal and electrical domains, making it difficult to develop an accurate model, where data-driven approach can perform fault diagnosis using only measurements from PEM fuel cell system.

However, due to the incomplete protocol of the PEM fuel cell failure tests, only limited PEM fuel cell failure mode conditions can be tested in the lab, [Yuan et al. 2011, Miller and Bazylak 2011], which cannot fully investigate the effectiveness of data-based fault diagnostic techniques. Therefore, further studies for the performance of these approaches indiagnosing more PEM fuel cell failure modes are still required.

In this study, a PEM fuel cell model is developed based on the bond graph technique, which

can represent multiple physical phenomena in a unified graphical notation. With the developed model, various PEM fuel cell failure modes can be simulated, and simulated data can be used in data-based fault diagnostic techniques to investigate their effectiveness. In section 2, the knowledge of the PEM fuel cell and development of its bond graph model will be presented. The performance of the developed model in representing PEM fuel cell performance will be validated in section 3. In section 4, the model will be used to simulate the fuel cell dehydration phenomenon, and the performance of data-based fault diagnostic techniques will be studied using the simulated data. From the results, some conclusions are given in section 5.

2 PEM FUEL CELL BOND GRAPH MODEL

2.1 *Introduction of PEM fuel cell*

A typical PEM fuel cell includes several components, i.e. anode and cathode electrodes, gas diffusion layer, catalyst layer, and polymer electrolyte membrane, which are depicted in Figure 1.

During operation, hydrogen and air/oxygen are injected into the anode and cathode sides, respectively. Hydrogen is divided into protons and ions with Eq. (1), protons can pass through the membrane, while ions can only arrive at the cathode via the external circuit, where current is generated. At the cathode side, protons, ions, and oxygen will react to produce heat and water (Eq. 2), which can be removed from the cathode side.

$$H_2 \rightarrow 2H^+ + 2e^- \tag{1}$$

$$\frac{1}{2}O_2 + 2H^+ + 2e^- \rightarrow H_2O \tag{2}$$

2.2 *Bond Graph method*

The core principle of the bond graph (BG) method is energy conservation, i.e. the total energy in a closed system is never destroyed or lost, but converted from one form to another. With this method, systems involving multiple physical domains can be unified.

In a BG the rate at which energy is transferred between components is power, which is denoted as a half arrow as shown in Figure 2. It can be seen that power flow is characterized by two power variables: effort (e) and flow (f), where

$$e \cdot f = \text{power} \tag{3}$$

Table 1 depicts some commonly used analogies for the meanings of effort and flow.

Elements in the BG are located at the BG nodes, and represent different energy manipulation mechanisms. Sources of effort (Se) and flow (Sf) are active elements and provide inputs to the system. Such elements, controlled by an external signal, are called 'modulated' and denoted by a prefix 'm', e.g. mSe. Energy dissipation and storage phenomena are implemented via resistive (R), capacitive (C) or inductive (I) elements. Detectors of effort (De) and flow (Df) are shown with a full arrow to emphasize that they do not participate in energy exchange, but rather simply act as sensors and measure corresponding power variables.

Multiple power bonds can meet at one of two junction types, 0- and 1- type, which enforce the laws of energy conservation within the system. Another junction structure called Transformers (TF) act as energy transducers converting the transferred power from one physical domain to another. TF elements can only have two bonds connected. Figure 3 shows the different junction types, and corresponding equations are written in Eqs. (4)–(6).

$$\text{1-junction} \quad f_1 = f_2 = \ldots = f_N \sum e_N = 0 \tag{4}$$

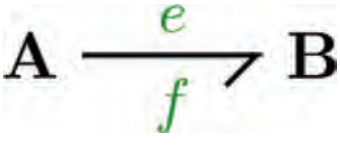


Figure 2. Power bond between objects A and B.

Table 1. Physical analogies for power variables.

Domain	Effort	Flow
Electrical	Voltage	Current
Mechanical	Force	Velocity
Pneumatic	Pressure	Volumetric flow
Chemical	Chemical potential	Molar flow
Thermal	Temperature	Entropy flow

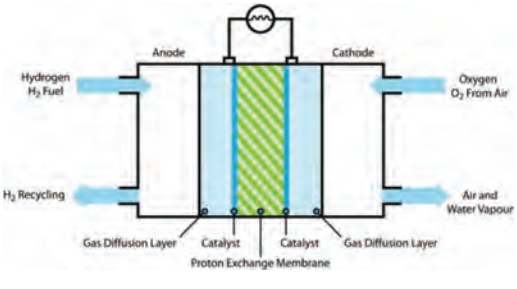


Figure 1. A typical PEM fuel cell.

0-junction $e_1 = e_2 = \ldots = e_N \sum f_N = 0$ (5)

Transformer $e_1 = me_2 f_2 = mf_1$ (6)

2.3 *EM fuel cell BG*

The hierarchy of the PEM fuel cell BG includes basic bond graphic elements describing energy storage and transfer mechanisms, which are at the base. A set of BG elements describing pneumatic and heat transfer phenomena are constructed for the two bipolar plates and for the anode and cathode sides, and BG elements describing electrochemical, transport and thermal phenomena representing the membrane electrode assembly [Saisset et al. 2006]. Additionally, cooling channels and the end plates are implemented as separate components [Vasilyev et al. 2017].

Figure 4 depicts the blocks resembling physical components of the PEM fuel cell and bonds connecting them representing power flows between components. Anode/cathode inlet and outlet blocks correspond to mass flow controllers or valves and regulate the flow of matter in or out of the cell. This is shown by bond labelled with P, T as efforts and $\dot{m}, \dot{H}$ as flows. The source of electric current (mSf) represents the load demanded from the fuel cell. Electrochemical phenomena within the membrane electrode assembly components calculate the rates of reactants and product consumption $\dot{m}_{H_2}, m_{O_2}$ and $\dot{m}_{H_2O}$. Transport phenomena determine the diffusion flows through the membrane electrode assembly, while thermal effects evaluate heat flows $\dot{Q}$ between the bipolar plates and the membrane.

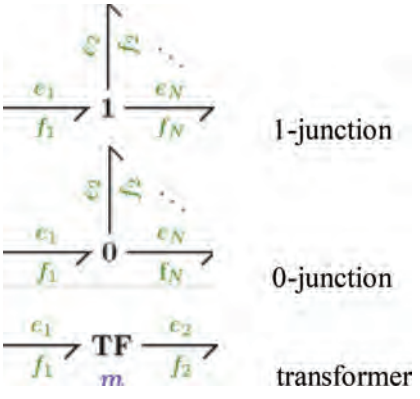


Figure 3. Bond graph junctions.

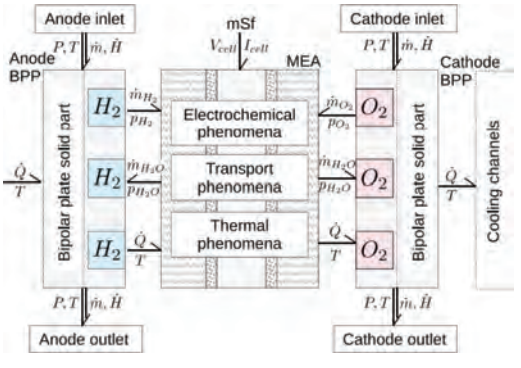


Figure 4. PEM fuel cell BG [Vasilyev et al. 2017].

A set of equations are used to develop the PEM fuel cell BG, including the computation of mass flow rates of gasses in and out of the cell, and thermal and pneumatic activities within bipolar plates. More details about the modelling procedures for the PEM fuel cell BG can be found in previous studies [Gawthrop and Bevan 2007, Vasilyev et al. 2017].

With results from different PEM fuel cell components, the single cell voltage can be calculated using Eq. (7).

$$V_{cell} = E_{Nernst} - \eta_{act} - \eta_{ohm} - \eta_{con} \quad (7)$$

where E_{Nernst} is the reversible potential, η_{act}, η_{ohm} and η_{con} are the activation loss, ohmic loss, and concentration loss, respectively.

Figure 5 depicts the single PEM fuel cell BG by putting the developed individual components together. Figure 6 shows the PEM fuel cell system BG including single cell BG and cooling loops, where inlet mass flows are regulated by Rth1 and Rth2. Each cooling loop is comprised of a single

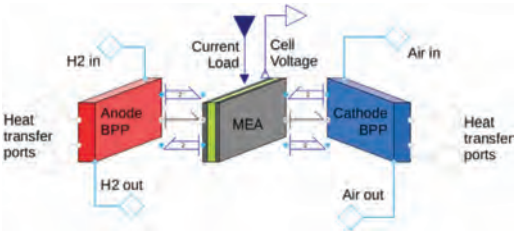


Figure 5. Single PEM fuel cell BG [Vasilyev et al. 2017].

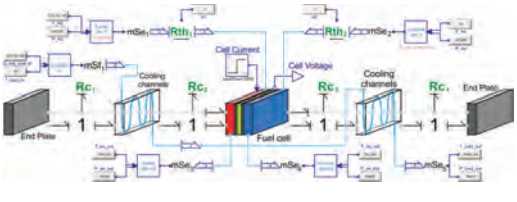


Figure 6. PEM fuel cell system BG [Vasilyev et al. 2017].

Cth-element and elements RC1-4 calculating the heat transfer rate.

3 VALIDATION OF PEM FUEL CELL BG

Before using the developed PEM fuel cell BG in fault diagnosis, the performance of the developed BG should be validated.

In this study, the Electrochemical Impedance Spectroscopy (EIS) is obtained from the test to determine the model parameters including electrical resistance and double layer capacitance. With the determined model parameters, the polarization cure is obtained from the model and compared with those from the test, the comparison results are depicted in Figure 7.

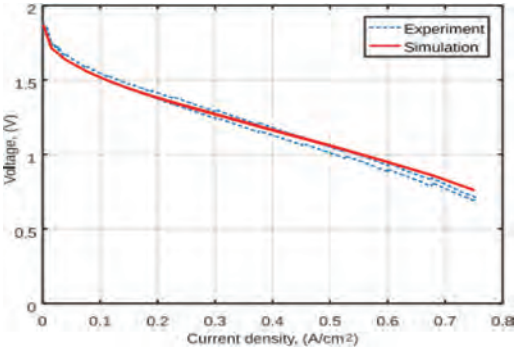


Figure 7. Comparison results of polarization curves between the model and test [Vasilyev et al. 2017].

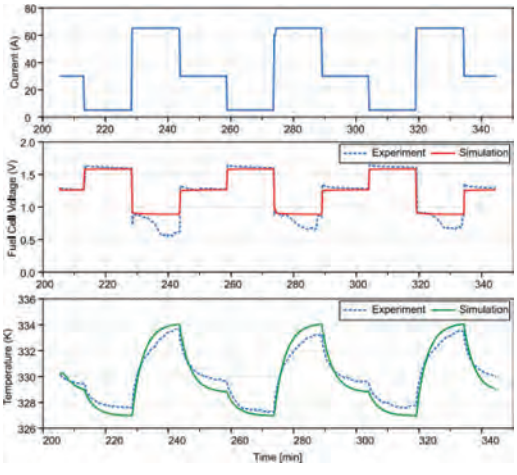


Figure 8. Comparison results of cell voltage and temperature curves between the model and test [Vasilyev et al. 2017].

It can be seen from Figure 7 that the overall polarization from the model can match the tested data with good accuracy, and the deviation becomes slightly larger in the region of concentration loss with current density higher than 0.55 A/cm^2. The reason is that the model doesn't fully consider the electrode porosity and effects of liquid water formation within the cell.

Furthermore, a test with varying current densities is performed, and the cell voltage and temperature are obtained and compared with those from the developed model. Results are shown in Figure 8. It can be seen that the developed model can capture the PEM fuel cell behavior with good quality, which paves the way for using the developed model for the following analysis.

4 USE OF FUEL CELL BG MODEL FOR FAULT DIAGNOSIS

4.1 *Simultion of PEM fuel cell failure mode*

In this study, dehydration is simulated and the simulated data is used to test the performance of data-driven fault diagnostic approaches. The reason for selecting dehydration is that it is a commonly experienced failure mode in PEM fuel cell systems due to unbalanced water management. Moreover, dehydration is not usually performed in testing as it will cause permanent damage of the membrane. Therefore, with the developed PEM fuel cell BG model, the performance of data-driven approaches can be investigated more efficiently in terms of both computational time and financial cost.

In the simulation, the constant current (70 A herein) is applied to the developed model, and after normal operation of a certain time, the relative humidity at the anode side is reduced from 100% to 50% at 500h, which can cause decreased water contents within the cell and thus dehydration. Figure 9 depicts the variation of anode relative humidity, voltage, and stack temperature.

It can be seen from Figure 9 that when operated at the constant condition, PEM fuel cell voltage will decay linearly, representing the degradation phenomena due to fuel cell aging. Moreover, with decrease of anode relative humidity, stack voltage shows a more steep decrease, and the increased stack temperature can be observed more clearly, this is due to the reduced water content within the cell from the reduced relative humidity of inlet gas.

4.2 *Data-driven fault diagnostic approaches*

In this study, several fault diagnostic approaches have been applied to the simulated data for both normal and dehydration conditions. As data from

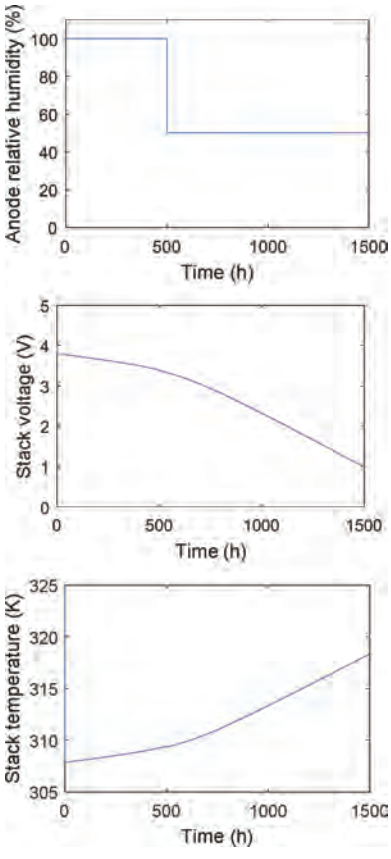


Figure 9. Simulation results of anode relative humidity, fuel cell stack voltage, and stack temperature.

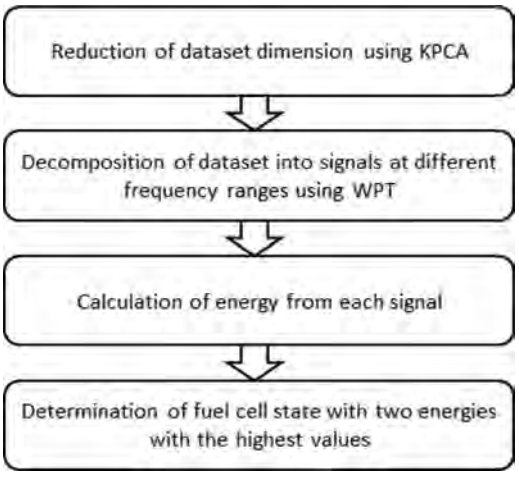


Figure 10. Flowchart of PEM fuel cell fault diagnosis using data-driven approaches.

Table 2. Sensor measurements used in the analysis.

Sensor	Unit	Sensor	Unit
Voltage	*V*	Air inlet flow	*l/min*
H2 inlet flow	*l/min*	Air inlet pressure	*bar*
H2 inlet pressure	*bar*	Air outlet pressure	*bar*
H2 outlet pressure	*bar*	Air inlet temperature	*K*
H2 inlet temp	*K*	Stack temperature	*K*

multiple sensors are simulated, the approaches reducing the size of dataset is applied, Kernel Principal Component Analysis (KPCA) is selected herein due to its better performance in non-linear systems. After that, wavelet packet transform (WPT) is applied to decompose the original signal into different frequency ranges, from which the features are constructed. The features with the highest values are used in this case to discriminate the PEM fuel cell states. This flowchart is illustrated in Figure 10. It should be noted that the selected diagnostic framework is effective in identifying various failure modes in PEM fuel cell systems [Mao et al. 2017]. More details about these approaches can be found in previous studies [Mao et al. 2017, Placca et al. 2010].

4.3 *Diagnostic results*

In the analysis, simulation data from multiple sensors are used, which are listed in Table 2. The reason of using multiple sensors is that multiple sensors can provide complementary information about the PEM fuel cell performance, which should be included in order not to lose useful information, without further interpretation of the sensor measurements.

KPCA is applied to the dataset including measurements from sensors (listed in Table 2) to project the dataset into the two principal directions. It should be noted that the simulated data shown in Figure 9 is divided into 2 parts representing different states (normal and dehydration), and each part is further divided into several segments for the following analysis.

WPT is then applied to each segment data from KPCA over 3 levels, the extracted wavelet coefficients are used to re-construct the signals at different frequency ranges, from which the signal energies are calculated. Figure 11 depicts the energy distribution at both normal and dehydration states.

It can be seen from Figure 11 that the energy shows similar distribution at different PEM fuel cell states, and the first few highest energies are

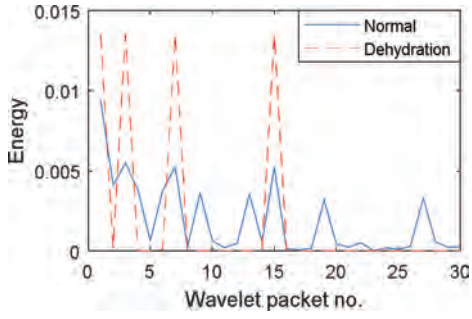


Figure 11. Distribution of energy from signals at different frequency ranges.

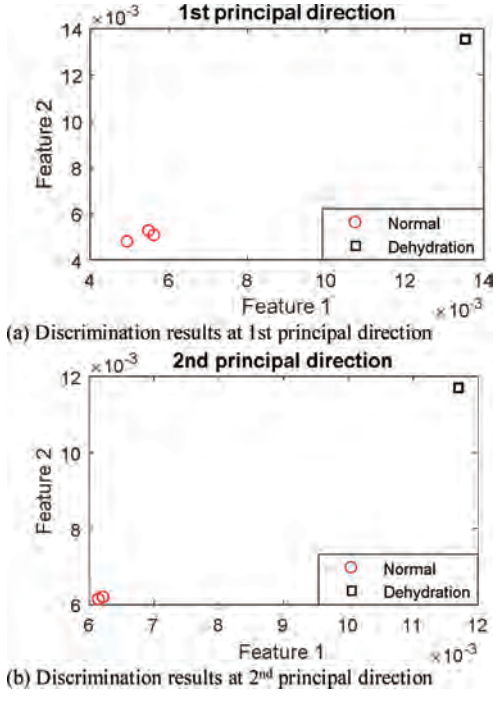


Figure 12. Discrimination results using selected features.

located at the same wavelet packet. Therefore, the first two highest energies are used as features herein for the discrimination. The results are shown in Figure 12. Since KPCA is used to project the original dataset into the two principal directions, the discrimination results at these two directions are depicted. It can be observed that the dehydration state can be discriminated with good quality from the normal state, indicating the applied data-driven approaches can identify the PEM fuel cell dehydration accurately.

5 CONCLUSIONS

In this paper, the PEM fuel cell model is developed using the bond graph technique, which can represent the various behaviours in PEM fuel cell system. The model parameters are determined using the collected EIS from the PEM fuel cell, and the performance of developed model is validated using the fuel cell test data at different conditions.

With the developed PEM fuel cell bond graph model, the data from different fuel cell failure scenarios can be simulated. In this study the fuel cell dehydration is simulated, and the simulated data is used in the data-driven fault diagnostic approaches. Results demonstrate that the used approaches can discriminate the dehydration state with good quality. In the future study, more fuel cell failure modes will be simulated using the developed model, and the capability of various fault diagnostic approaches in identifying these fuel cell faults can be fully investigated.

ACKNOWLEDGEMENT

The work is supported by grant EP/K02101X/1 for Loughborough University from the UK Engineering and Physical Sciences Research Council (EPSRC). The authors also acknowledge the industrial and academic collaborators of the RESILIENCE project (Robust Lifecycle Design and Health Monitoring for Fuel Cell Extended Performance).

REFERENCES

Gawthrop, P., Bevan, G. 2007. Bond-graph modelling, *IEEE Control Systems Magazine* 27(2):24–45.

Kamal, M.M., Yu, D. 2011. Model based fault detection for proton exchange membrane fuel cell systems, *International Journal of Engineering, Science and Technology* 3(9): 1–15.

Mao, L., Jackson, L.M., Dunnett, S.J. 2017. Fault diagnosis of practical polymer electrolyte membrane (PEM) fuel cell system with data-driven approaches, *Fuel Cells* 17(2):247–258.

Miller, M., Bazylak, A. 2011. A review of polymer electrolyte membrane fuel cell stack testing, *Journal of Power Sources* 196:601–613.

Petrone, R., Zheng, Z., Hissel, D., Pera, M.C., Pianese, C., Sorrentino, M., Becherif, M., Yousfi-Steiner, N. 2013. A review on model-based diagnosis methodologies for PEMFCs, *International Journal of Hydrogen Energy* 38(17): 7077–7091.

Placca, L., Kouta, R., Candusso, D., Blachot, J.F., Charon, W. 2010. Analysis of PEM fuel cell experimental data using principal component analysis and multi linear regression, *International Journal of Hydrogen Energy* 35(10):4582–4591.

Ohs, J.H., Sauter, U., Maass, S., Stolten, D. 2011. Modelling hydrogen starvation conditions in pro-

ton exchange membrane fuel cells, *Journal of Power Sources* 196(1):255–263.

Rubio, M.A., Urquia, A., Dormido, S. 2010. Diagnosis of performance degradation phenomenon in PEM fuel cells, *International Journal of Hydrogen Energy* 35(7):2586–2590.

Saisset, R., Fontes, G., Turpin, C., Astier S. 2006. Bond graph model of a PEM fuel cell, *Journal of Power Sources* 156(1):100–107.

Steiner, N.Y., Hissel, D., Mocoteguy, P., Candusso, D. 2011. Non intrusive diagnosis of polymer electrolyte fuel cells by wavelet packet transform, *International Journal of Hydrogen Energy* 36(1):740–746.

Vasilyev, A., Andrews, J., Jackson, L.M., Dunnett, S.J., Davies, B. 2017. Component-based modelling of PEM fuel cells with bond graphs, *International Journal of Hydrogen Energy* 42(49):29406–29421.

Yuan, X., Li, H., Zhang, S., Martin, J., Wang, H. 2011. A review of polymer electrolyte membrane fuel cell durability test protocols, *Journal of Power Sources* 196:9107–9116.

Zeller, A., Rallieres, O., Regnier, J., Turpin, C. 2010. Diagnosis of a hydrogen/air fuel cell by a statistical model-based method, *Vehicle Power and Propulsion Conference (VPPC)*, Lille, France.

Zheng, Z., Petrone, R., Pera, M.C., Hissel, D., Becherif, M., Pianese, C., Steiner, N.Y., Sorrentino, M. 2013. A review on non-model based diagnosis methodologies for PEM fuel cell stacks and systems, *International Journal of Hydrogen Energy* 38():8914–8926.

Zhongliang, L., Outbib, R., Giurgea, S., Hissel, D., Li, Y. 2015. Fault detection and isolation for polymer electrolyte membrane fuel cell systems by analysing cell voltage generated space, *Applied Energy* 148:260–272.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Risk assessment and the influence of new information

Tor Stålhane
Norwegian University of Science and Technology, Trondheim, Norway

Stig Ole Johnsen
Norwegian University of Science and Technology, Trondheim, Norway
SINTEF Safety and Reliability, Norway

ABSTRACT: We have run experiments with 17 groups with three participants—to see how they assess hazards, probability and possibility to avoid dangers. We considered two scenarios taken from ISO 15998-2. We used a three-step process—each participant read the scenario and assesses consequence, probability of the event and the possibility of avoiding the hazard. Then they wrote down their rational for the assessments and adjusted. They then presented their rationales to the other two members and made final assessments.

We discuss the results: what rationale made participants change their assessments, and which parameters were changed? As in earlier experiments, risks are overestimated. We need to improve the way we describe hazardous scenarios so that we can get more realistic risk assessments.

There exist several guidelines for writing scenarios, and they are followed by ISO 15998-2 but this did not seem to help. We suggest some new guidelines for scenario description.

1 INTRODUCTION

Risk assessment is usually done by one or more persons, based on a scenario description. No feedback or rational is needed. We would like to examine how the assessment process will work if the participants need to write rationales for their decisions and how the assessments process will be affected by hearing other persons' rationales.

By studying how the assessment process is influenced by other peoples' rationales and studying which parameter's value is influenced, we can get a better understanding of the involved persons' assessment process and come up with a better way to construct scenarios and procedures for risk assessment.

2 RELATED WORK

There is a lot published around problems pertaining to risk assessment. A large part of these papers appears in the field of psychology. The papers cited below is just a small sample of the published material.

Kahneman (2011) has done research on human reasoning. One result is of special interest here—the strong tendency always to assume the worst consequences. His observation is worth quoting in full: "*The danger is increasingly exaggerated as the media compete for attention-grabbing headlines. Scientists and others who try to dampen the increasing fear and revulsion attract little attention, most of it hostile: anyone who claims that the danger is overstated is suspected of association with a 'heinous cover-up*'".

Sandeman et al. (1998) have done research on risk communication and found that you get a better assessment of risks by enabling the respondent to relate it to a "normal" situation—a situation that the respondents already were comfortable with.

Li and Wang (2010) report valuable experience with using the Delphi process in a risk analysis process. A more complete evaluation of the Delphi method used in risk assessment is given by Zaloom and Subhedar (2009) who report an experiment using the method to assess the likelihood of 10 identified risk scenarios. Unfortunately, after the second round of the process, there was still only one risk scenario where the participants agreed and they ended up using the score averages as the result.

Scenario construction is an important part of risk assessment. One way of writing and analysing scenarios is presented by Whitney and Thompson (2009). They argue that a useful scenario should contain information related to who, what, why, when, where and how. In addition, it is not enough to assess consequences, probability and how easy it is to avoid the danger. It is also necessary to assess the probability of the scenario.

Last, but not least, it is important to remember that all experiments and case studies into the

domain of hazard assessment have shown that there is no significant difference between laypersons and experts. See e.g. Rowe and Wright (2001) for an extensive discussion of this topic.

3 THE EXPERIMENT

3.1 *The experiment layout*

Input to the experiment was two scenarios for earth-moving machinery. These scenarios, and seven others, have earlier also been used in a set of experiments reported by Stålhane and Malm (2016). The scenarios are taken from ISO 15998-2:2012. According to the standard, the two chosen scenarios—later referred to as case 4 and case 5—should be assessed as needing performance level (PL) c and e respectively. Average probability of dangerous failure per hour (1/h) for PL c is $> 10^{-6}$ to $< 3 \times 10^{-6}$; for PL e $> 10^{-8}$ to $< 10^{-7}$.

The scenarios are described as follows:

- Case 4: Articulated Wheeled Loader. Machine boom moves without command. Operator is not compelled to be in the operator station. Operator may be greasing machine or otherwise near moving parts. Operator typically in harm's way much less than 10% of time. If operator is near moving part, it may be very difficult to get away quickly enough to prevent injury.
- Case 5: Articulated Wheeled Loaders < 40 km/h. Complete loss of Primary Steering and Emergency Steering (Either steers un-commanded or not at all while propelling). Operator has braking to stop the machine. Operator is not warned prior to loss of steering. Potential to hit higher speed vehicle with multiple passengers' Multi-passenger vehicles in the path of machine is much less than 10% of time. Operator can stop the machine. Vehicle may be able to avoid the loader.

To check the quality of the two scenario descriptions, we run a set of readability checks—Adobe (2007). Most of the readability scores—e.g., the Gunning Fog index and the Flesch Kincaid Grade level gave approximately the same results for both scenarios. The main difference was observed in the Flesch Reading Ease index where 70 to 60 indicates that the text is easy to read while 50 to 40 is difficult. Case 4 scored 41 and case 5 scored 59 indicating the case 4 is more difficult to understand.

The experiments were performed with 51 participants in 17 groups with three persons in each group. Each group did the experiment under the supervision of the researcher. The participants were given the scenario descriptions shown above and went through a three-phase process which was a simplified version of the Delphi process—see Linestone and Turoff (1975). The process used was as follows

1. Phase 1: Read the scenario and the scoring rules for the following three parameters: event consequence (S), event probability (F) and the possibility of avoiding the danger caused by the event (P). Descriptions for scoring of the parameters were taken from ISO 13849-1:2015. Based on the scenario description, they decided which value to assign to each of the three parameters.
2. Phase 2: Write down the rational for the three values assigned. During this process, they were free to change their assessment if they felt that the rational did not fit their previous assessments.
3. Phase 3: Read the rationales aloud to the rest of the group. Each participant was then free to change his scores. At last, they wrote their new assessments on the form and returned it to the researcher. There was no pressure or incentives for the group to reach a final agreement.

The process described above gave us total of 51 answers, which were registered in an Excel sheet. Each answer contained the participant's assessment of S, F and P for each phase—a total of 153 values. The response set for each participant contains three sets of assessed S, F and P values plus the rationales for each choice.

3.2 *The assessment model*

The assessment model—from risk parameters (S, F and P) to performance level (PL) or safety integrity level (SIL) are defined by a decision tree. For each of the risk assessment parameters S, F and P, an assessor (e.g., the participants in our experiment) can choose between two values as shown below:

- *Severity of injury*. S1 Slight injury – bruise. S2 Severe injury – amputation or death.
- *Frequency of exposure to injury*. F1 Seldom. F2 Frequent to continuous.
- *Possibility of avoiding the hazard*. P1 Possible. P2 Less possible. Based on the speed of approach of the hazard and the ability of the operator to avoid the hazard. (If the operator can avoid the hazard you would choose P1).

4 DATA ANALYSIS

We will need to discuss whether differences are statistically significant when discussing differences between persons or groups. We have approached this by using a simple approximation to the standard uncertainty for probability differences. The equation that we start with is

$$d^2 = u_{\alpha/2}^2 p(1-p)/N \quad (1)$$

Here p is the probability of the event under consideration and N is the size of the sample. Since p(1-p) is always less than or equal to ¼ and $u_{\alpha/2}$ for 5% significance is approximately 2, we get the simple approximation

$$d^2 = 1/N \quad (2)$$

Thus, for differences between the 17 groups, the uncertainty is 0.24 (4 groups) and for differences between the participants, it is 0.14 (7 participants).

4.1 *An exploratory case*

The experiments where not done to accept or reject one or more predefined hypotheses. Neither did we start with a set of research questions. What we did was to run the experiments to collect data and then perform an explorative data analysis—looking for what the data might tell us. For this reason, the results are not final. However, they will serve as a starting point for hypothesis and subsequent experiments. Throughout this paper, when we say, "the data shows that…" we mean "the data indicates that…"

We got the following data from the experiment (1) the participants' scores for S, F and P, (2) their rational for choosing the values, and (3) how these values changed when they write down a rational for their decisions and share this information with the other persons in their group when assessing the three parameters. Since we have the rationales we connect them to the information provided by the scenario descriptions—e.g., which parts did they use and what was used in addition.

4.2 *Different processes—different results*

Earlier, a total of nine cases—including case 4 and case 5 used in this experiment—have been assessed earlier, see Stålhane and Malm (2016). Most of assessments for all cases, irrespective of the results required by the standard, were S = 2, F = 1 and P = 1.

For case 4 we get S = 2 and F = 1, while the split between P = 1 and P = 2 is 67% versus 33%.

When we introduced the need to write a rational and use it as feedback to the other participants in the group, most of assessments, both for case 4 and case 5 changed to S = 2, F = 1 and P = 1 or P = 2. As we see, the split between P = 1 and P = 2 is 44% versus 56%. The results for case 5 is almost identical.

The feedback cannot be the only cause of this change since the decision tree structure is almost the same for phase 1 and phase 3. The only case 4 change between phase 1 and phase 3 are that we in phase 1 has 45% 2, 1, 2 and 55% 2, 1, 1, while we in phase 3 have 57% 2, 1, 2 and 43% 2, 1, 1. These differences are not significant at the 5% level. Thus, the main reason for the changes is the requirement that the participants must write rationales. In addition, we should note that this change only affects the P-parameter.

4.3 *Documentation of the rationales used*

Different participants used various parts of the scenario description when writing their rationales. If they found the scenarios to lack information, they used the information in the scenarios plus their own experience to make their own rationales. Tables 1 and 2 show the used information from the scenarios while Tables 3, 4 and 5 show information used but not found in the scenario descriptions.

For case 5, we see that two thirds of the participants used one or more terms from the

Table 1. Part of scenario information used in case 4.

Case 4				
Scenario info used	S	F	P	Sum
Machine boom moves without command	4	1	2	7
Operator may be near moving parts	17	9	15	31
Operator in harm's way less than 10% of time	2	14		16
Near moving parts = > difficult to get away quickly	3	1	21	25
Persons (#)	24	23	26	
Persons (%)	48	48	52	

Table 2. Part of scenario information used in case 5.

Case 5				
Scenario info used	S	F	P	Sum
Speed less than 40 km/h	27	2	9	38
Complete loss of primary and secondary steering		11	5	16
Breaks are working OK	2	3	22	27
No warning prior to loss of steering	1			1
May hit high speed vehicles with multiple passengers	10	1	1	12
Multi-passenger vehicle in path less than 10% of time	1	10		11
Operator can stop machine	4	3	4	11
Vehicle may be able to avoid the loader		3	15	18
Persons (#)	38	31	39	
Persons (%)	73	61	75	

Table 3. General information used to assess S in cases 4 and 5.

S – event consequences		
Rationales	Case 4	Case 5
Safety instructions	1	
Crushed limbs	14	4
Machine turned off	1	
Strong hydraulics	2	
Heavy parts or machine	8	6
Person run over	3	
Hit someone		3
Persons (#)	29	13
Persons (%)	57	25

Table 4. General information used to assess F in cases 4 and 5.

F – event probability		
Rationales	Case 4	Case 5
Happens seldom	8	11
Operator mindful of danger	10	1
Regular maintenance	2	2
Safety mechanisms	4	
Harm other vehicles		3
Persons (#)	25	18
Persons (%)	52	36

Table 5. General information used to assess P in cases 4 and 5.

P – escape possibility from hazardous event		
Rationales	Case 4	Case 5
Safety instructions	5	2
Operator mindful of danger	7	1
Keep away from moving parts	1	1
Easy to avoid	5	2
Severe injuries	2	4
Cannot stop the machine		3
Persons (#)	24	13
Persons (%)	48	26

scenario descriptions when writing a rational for their choice of scores for S, F and P. As for case 4, there are no statistically significant differences between the three parameters at the 5% level. The differences between the two cases are, however, statistically significant at the 5% level. The only part of the case 5 scenario that is not used is "*Either steers un-commanded or not at all while propelling*".

For case 4, we see that half the participants used one or more terms from the scenario descriptions when writing a rational for their choice of scores for S, F and P. There are no statistically significant differences between the three parameters at the 5% level. The only part of the scenario that is not used is "*Operator is not compelled to be in the operator station*".

Several parts of the scenario descriptions are used for several parameters—see Tables 1 and 2. For case 4 the statement "*Operator may be near moving parts*" is used for all three parameters.

The Tables 3 to 5 show the rationales which are not taken from the scenario descriptions. The terms used have been extracted from the rationales and grouped under a set of generic terms—e.g., all rationales related to safety instructions and safety procedures have been grouped under the term "Safety instructions".

That three participants used "Cannot stop the machine" as a problem for case 5 is strange, since the scenario explicitly states that "The operator can stop the machine".

We went through all rationales to look for rationales used for the wrong parameter—i.e., the whole or a part the rational indicates one parameter while it is used for another. Examples of wrong / misplaced rationales:

- **S** (P): If the operator is in the danger zone, he may be seriously injured if he is crushed by the machine
- **S** (F): If an accident occurs with a big and heavy machine—e.g. get caught in moving parts—the probability for severe damages is large
- **F** (P): The operator is not close to dangerous parts of the machine very often
- **P** (F): Daily inspection of hydraulics will enable us to discover problems before an accident. Routines

When we went through all the rationales, we found the following:

From the table, we see that the most frequent wrong or mixed rationales are related to F and P, while the next most frequent is between S and P.

4.4 *Parameter changes*

The last piece of information that we need is the changes—when and where. This is shown in the two tables below.

There are 41 changed assessments from phase 1 to phase 2 and 32 changers in assessment from phase 2 to phase 3. The difference between the two cases is statistically significant at the 5% level.

The agreement for all parameter assessments improved from phase 1 to phase 3. However, the only statistically significant differences are between

Table 6. Wrong/mixed parameter rationales for cases 3 and 5.

Case 4		Case 5	
S	P = 3 SP = 7	S	P = 4 SP = 7
F	P = 20 FP = 13	F	P = 9 FP = 14
P	S = 1	P	F = 1
Sum	44	Sum	35
%	28	%	22

Table 7. Parameter assessment changes in case 4.

Case 4			
Phase 1 to Phase 2		Phase 2 to Phase 3	
S	5	S	0
F	3	F	3
P	7	P	9
Sum	15	Sum	12
Percentage of total	30	Percentage of total	24

Table 8. Parameter assessment changes in case 5.

Case 5			
Phase 1 to Phase 2		Phase 2 to Phase 3	
S	2	S	0
F	18	F	18
P	6	P	2
Sum	26	Sum	20
Percentage of total	52	Percentage of total	40

phase 1 and 3 for S and P in case 4. For case 5, there are no statistically significant differences.

4.5 *Summary of the data analysis*

Based on the tables above, we formulate the following exploratory research questions—ERQ:

1. Which parts of the scenario descriptions are used for the risk assessment?
2. From where do the participants who do not use the scenarios take their information?
3. Which assessments are stable and which do change?
4. How often do the participants assess two or more factors using the same rational?
5. Based on the findings from these ERQ, how can we improve the assessment process and the standard's scenario descriptions so that they are used in a more efficient way?

ERQ1: Roughly 50% of the case 4 assessments and 75% of the case 5 assessments used only terms from the scenario descriptions—see Tables 1 and 2. The most commonly used terms are:

- Case 4, S: "Operator may be near moving parts", F: "Operator in harm's way less than 10% of time" and P: "Near moving parts = > difficult to get away quickly"
- Case 5, S: "Speed less than 40 km/h", F: "Speed less than 40 km/h" and "Multi-passenger vehicle in path less than 10% of time" and for P: "Breaks are working OK".

One possible reason why the case 5 assessments used more scenario description terms that the case 4 may be that this description is longer – 76 words—while case 4 has only 58 words.

ERQ2: Those who also used other terms than those used in the scenario, used terms related to the machine, the environment and the machine's behaviour. The same terms are considered important for both cases e.g., "Crushed limbs" or "Heavy parts or machine" for consequences.

ERQ3: For case 4, approximately a third of the assessments have been changed. For case 5 approximately half of the participants changed their mind one or more times during the three phases. There are significant differences between the two cases. For case 4, P is changed most often while for case 5, F is changed most often. The differences between case 4 and case 5 are statistically significant at the 5% level.

ERQ4: We see from the data that more than half of all assessment rationales for S, F or P mixed two parameters—most commonly, they mixed F and P. However, a considerable amount of the participants also mixed S and P. The participants' logic is clear and can be split into the following:

- For S: since it is easy to avoid the hazard, the consequences are not severe
- For F: since it is easy to avoid the hazard, it cannot happen so often

ERQ5: Based on this summary we suggest three temporary conclusions:

- Differentiating between P and the two other parameters is difficult for the participants. The possibility to avoid the danger is used to influence the assessment of the event frequency and the seriousness of the consequences.
- A considerable part of the participants did not think the information given in the scenarios is sufficient or relevant for risk assessment.
- Most participants always consider the worst consequences.

These temporary conclusions lead us to the next conclusion, namely that standards should use a

different scenario description and a different decision tree to help the participants to arrive at more consistent hazard assessments.

The terms used in the assessment rationales but not present in the case descriptions, are the following:

- Case 4 and 5: crushed limbs (S), heavy parts (S), heavy machinery(S), happen seldom (F), safety instructions (P), easy to avoid (P)
- Case 4: operator mindful of danger (P)
- Case 5: cannot stop the machine (P)

Some of these rationales are just derived from the information found in the case description—e.g., crushed limbs, heavy parts and heavy machinery. These words are warning signal words—Hellier (2000)—to add emphasis to the case description. Others are contradicting the case description—e.g., cannot stop the machine.

5 ASSESSMENT CHANGES

The most important observation related to the assessments done in this experiment is the misplaced assessment rationales, especially for F and P. The discussions in chapter 4.2 and 4.3 lead us to the conclusion that people are not good at separating dangerous event probability and the possibility of avoiding dangers.

Thus, instead of the current three-level decision tree used in the standard we suggest that the standard should use a two-level decision tree. As before, S is the consequences of the dangerous event while P should be defined as the probability that a person is harmed. This will include both the event probability and the escape opportunities. Based on this, we suggest the following ratings:

- P = 1: we will almost surely avoid the consequences
- P = 2: we might avoid the consequences
- P = 3: we will most likely suffer the consequences

As noted earlier, there are more changes to assessment between phase 1 and phase 2 than between phase 2 and phase 3. The significant difference between the two transitions is that in the first phase, each participant first writes down his assessment and then wrote down the rational.

There are presumably two mechanisms at work when a participant changes one or more parameter assessments: (1) one or two of the other participants use a negative term when describing their parameter choice rational or (2) he sees that the rational he wrote does not support his assessment. In many cases, both for laypersons and for experts, the assessment is done based on intuition—see Freeman et al. (2012). The rational is written after the decision and is thus an attempt to justify this decision—see e.g., Pigozzi et al. (2009).

26 of the changes caused a parameter to be changed from 1 to 2 (a parameter was changed from 2 to 1 in only six cases). In these six cases, the changes occurred after phase-2 and in all cases the parameters were changed back to 2 again in phase-3. Thus, in no case did any participant in the final case move any parameter assessment from 2 to 1. This agrees with the general observation that one negative statement wins over several positive ones—see e.g., Baumeister et al. (2001). The research of e.g., Shang et al. (2015) show that "people are likely to direct more attentional resources toward high-hazard stimuli compared to low-hazard ones", thus supporting Baumeister's conclusions.

The P parameter (probability of avoiding the hazard) is the one changed most frequently – 11 times for case 4 and seven times for case 5. This fits well with our observation that P is the parameter that the participants had the most problem with. The mixing of F and P assessment gets even more problematic if we consider it together with the effect of a negative assessment. This may cause a negative P assessment to change a positive F assessment to a negative assessment and vice versa.

6 ANOTHER WAY TO DESCRIBE SCENARIOS

6.1 *The scenarios revisited*

If a scenario shall be useful in risk assessment, it must include information on the event– what went wrong– possible consequences– and how can we avoid or reduce these consequences? To check this against the scenarios used here and the rationales used by the participants we will first restructure the scenarios so that each argument is separated and tagged with the relevant part of the risk assessment – S: event consequences, F: event probability or P: avoidance possibility. The result is shown in the two lists shown below.

Case 4: Articulated Wheeled Loader.

- Machine boom moves without command – S, F
- Operator is not compelled to be in the operator station – P
- Operator may be greasing machine or otherwise near moving parts – P
- Operator typically in harm's way much less than 10% of time – P
- If operator is near moving part, it may be very difficult to get away quickly enough to prevent injury – P

Case 5: Articulated Wheeled Loader

- Speed < 40 km/h – S, P

– Complete loss of Primary Steering and Emergency Steering (Either steers un-commanded or not at all while propelling) – F
– Operator has braking to stop the machine – P
– Operator is not warned prior to loss of steering – P
– Potential to hit higher speed vehicle with multiple passengers – S
– Multi-passenger vehicles in the path of machine much less than 10% of time – P
– Operator can stop the machine – P
– Vehicle may be able to avoid the loader – P

The most important thing to note is that most statements/sentences are about getting away or preventing harm—four out of five for case 4 and five out of eight for case 5. Thus, there is little focus on the event probability (F) and much focus on the possibility for getting out of harm's way (P). This is another factor that can explain the large portion of P-related rationales for assessment of the F parameter.

6.2 *How were the scenarios used in the standard?*

The scenarios used in the standard can be used in two ways. They can be considered as reference scenarios – if your case is like this, then the PL should be the same as the PL for this scenario in the standard. The other way is to use it to indicate which factors are important in the risk assessment process – see Tables 1 and 2. Thus, the information listed in Tables 3–5 is indicators that the participants found the information in the scenarios to be wanting. Our experiments have shown that the participants in many cases need more or different information than what the scenarios provide. The most used pieces of information not in the scenario descriptions and not derived from this are

– Consequences: person run over – used three times
– Equipment descriptions: safety mechanisms – used four times
– Operator description: mindful operator, safety instructions – used 20 times

The escape or avoid part (P) of the risk assessment process will be greatly improved if the scenarios were accompanied by a preamble giving information about operator courses, protective gear and safety instructions.

6.3 *A better way to construct scenarios*

A good method for construction scenarios is the GMA approach. However, using GMA may lead to many possible scenarios and it is not always clear how we can reduce this to a practical volume. Another, event-driven method for constructing scenarios is the STEP process—Hendrick (1987). However, this method is difficult to use with the scenarios used in the relevant standard – e.g., it is difficult to model statements such as "*Operator typically in harm's way much less than 10% of time*".

A doable alternative is to use the work of Whitney and Thompson (2009). They have suggested a checklist for scenarios, which runs as follows:

– Who: groups, individuals and organizations involved
– What: activity, objectives and targets
– Why: motivation of person or organizational
– When: triggering events, time requirements, opportunity
– Where: city, building, institutions or road
– How: approaches required related to technology, funding or know-how

We might think that the quality of the scenarios increases with the amount of available information. However, Oskamp (1965) has published both own results and meta-studies that seem to confirm that "*Beyond some early point in the information-gathering process, predictive accuracy reaches a ceiling*". Oskamp's studies are related to psychological diagnoses but the results are general enough to apply also to our case. Thus, we need to keep the amount of information low so that we do not water down the few, important pieces of information.

Since we need to assess three parameters, we need to provide sufficient information for the assessment of each of them. Sticking to the schema of Whitney and Thompson (2009), we must decide which three of the checklist items that are most important for each parameter. We suggest the following

S: Event consequence:

– **Who** is or could be involved?
– **What** may happen or is happening?

F: Event probability:

– **What** may happen or is happening?
– **Where** is the event happening?
– **How** – which approach is taken or required to initiate the event?

P: Possibility to escape consequences:

– **Who** is or could be involved?
– **How** – which approach is taken or required?
– **Where** is the event happening?

When we go through the two scenarios used in our experiments, we see that both scenarios lack information needed to assess probability – i.e., how probable this scenario is.

If we apply the checklist of Whitney and Thompson to the two cases we have considered in this experiment, we get the following results:

Case 4: Articulated Wheeled Loader.

- Machine boom moves without command – S, F **What.**
- Operator is not compelled to be in the operator station. Not used during assessment.
- Operator may be greasing machine (implicates "near moving parts") or otherwise near moving parts – P **What, Who**
- Operator typically in harm's way much less than 10% of time – P **Who, When**
- If operator is near moving part, it may be very difficult to get away quickly enough to prevent injury – P **Who, Where, When**

Case 5: Articulated Wheeled Loader

- Speed < 40 km/h – S, P **What**
- Complete loss of Primary Steering and Emergency Steering (Either steers un-commanded or not at all while propelling – not used during assessment) – F **What**
- Operator has braking to stop the machine—P **Who, What**
- Operator is not warned prior to loss of steering – P **Who, What, When**
- Potential to hit higher speed vehicle with multiple passengers – S **What, Who**
- Multi-passenger vehicles in the path of machine much less than 10% of time – P **What, Where, When**
- Operator can stop the machine – P **Who, What**
- Vehicle may be able to avoid the loader – P **What**

Note that F and P share the keywords "where" and "how". In addition, S lacks "who"-information, which is important to understand the consequences. F does in both cases lack information on "where" and "how", which are important to understand the failure mechanism and thus assess the event probability. P lacks information on "what" and "how", which makes it difficult to assess the escape probability.

Tables 3, 4 and 5 shows information added by the participants during their assessment. The information added are the same for both cases. If we apply Whitney and Thompson's key words to the information in the Tables 3, 4 and 5 we find that for S we can add "what" and "who", while we for F and P can add "when", "who" and "what". Thus, for both cases we still miss "how" for F and P and "where" for P.

Suggested additions to take care of the "where" aspects for the two cases are e.g., "on a construction site" for case 4 and "on a public road" for case 5.

An additional challenge when writing scenarios is the choice of words. Our will have an important influence on how much and what kind of attention each part of a scenario description will get. In the end, it will also decide the assessment of PL or SIL.

7 THREATS TO VALIDITY

The main threats to validity are the small sample and participant motivations. We will give a brief discussion on how these two threats may influence our temporary conclusions.

We chose the two cases 4 and 5 for the experiment to get a large distance in protection levels. Case 4 should have PL c while case 5 should have PL e—both according to ISO 13849-1:2015. However, just two cases are way too small a sample for making a statistically significant statement about the Delphi method when it comes to PL assessment. On the other hand, 51 participants assessing two cases is still enough to say something about the general trends, such as mixing P and F, and the influence of the general observation that "bad is stronger than good".

The questions related to participants' motivation is more serious. People who do this for real have a strong incentive to get it right while the students are just motivated to get the job done because that is what they promised the researcher. Thus, we cannot be sure that they "really put their soul into it".

Even so, it is our opinion that the data are sufficient for doing some exploratory data analysis and to come up with some issues that should be considered more seriously in future research.

8 CONCLUSIONS

Based on the results of this small experiment, we have tentatively identified the following potential problems with the way hazard analyses are performed:

- Whatever the scenario description, people mostly goes for the worst case because "***it just might happen***" and "Bad is stronger than good". Thus, we should ***not*** use the Delphi process to assess risks. Relevant empirical data might improve this situation.
- Many persons have problems with keeping separate the dangerous event probability and the probability of avoiding harm.
- Those who write scenarios for risk assessment – e.g., for standards – should consider available rules for writing efficient scenarios – e.g., considering the rules of Whitney and Thompson (2009) and readability checks, Adobe (2007).

These problem areas need to be addressed when writing a scenario intended for risk assessment. The first problem is the most serious and difficult one because it is related to psychology and not to any technical problem. The two other ones can be solved by providing relevant information to the assessors.

REFERENCES

Adobe: Silicon Prairie Software. Readability Tools, Adobe Systems Incorporated, 2007.

Baumeister, R.F. et al. Bad Is Stronger Than Good. *Review of General Psychology*, vol. 5, no. 4, 2001.

Freeman, D., Evans, N. and Lister, R.: Gut feelings, deliberative thought, and paranoid ideation: A study of experimental and rational reasoning. *Psychiatry Research,* no. 197, 2012.

Hendrick, K. & Benner, L. *Investigating accidents with STEP.* Marcel Dekker Inc. New York, 1987.

Kahneman, D.: *Thinking, Fast and Slow.* Penguin Random House, UK, 2012.

Li, Y and Wang, C. Based on the Delphi Method of Deep Excavation Safety Risk Analysis. *International Conference on Artificial Intelligence and Education (ICAIE), 2010*.

Linestone, H.A. and Turoff, M.: *The Delphi Method. Techniques and Applications.* Addison Wesley Publishing Company, 1975.

Oskamp, S. Over-Confidence in Case Study Judgement. *Journal of Consulting Psychology,* 1965, vol. 29, no. 3.

Pigozzi, G. et al. Formal ex-post rationalization – A complete conclusion-based procedure for judgement aggregation. *Proceedings of the Cinquièmes Journées Francophones Modèles Formel de l'Interaction,* 2009.

Rowe, G. and Wright G. Differences in expert and lay judgment: Myth or reality. *Risk Analysis.* vol. 21, no 2, 2001.

Sandeman, P.M. et al. Communication to reduce risk underestimation and overestimation. *Risk Decision and Policy*, vol. 3, 1998.

Stålhane, T. and Malm, T. Risk Assessment – Experts vs. Lay People. *ESREL* 2016.

Whitney, P. and Thompson, S *Risk Assessment for Scenarios.* Pacific Northwest, National Laboratory, 2009.

Wullt, T. Safety Aspects of Product Limitation. Behaviour under non fault conditions. *Third Scandinavian Conference on System & Software safety*. March 25, 2015.

Zaloom, V. and Subhedar, V. Use of the Delphi method to Prioritize Event Impacting Operations in the Maritime Domain. Lamar University Working Paper, 2009.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Risk assessment in construction projects with the use of neural networks

L. Giannakos
School of Science and Technology, Hellenic Open University, Patra, Greece

Y. Xenidis
Department of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

ABSTRACT: The inherently complex nature of risks interdependencies in construction projects coupled with incomplete data records during projects development often results to inaccurate assessments. This paper showcases the use of neural networks for risk assessment in construction projects. A detailed literature review identifies the different types and training methods of neural networks as well as the respective tools applicable to construction projects risk management. Based on these findings, the paper presents the development of a specific neural network that partially assesses occupational risk in a construction engineering project. The proposed neural network is trained with metadata from previous risks assessments. The modeling of the network is realized through two software tools, in order to identify potential difficulties in the modeling process as well as potential deviations in the assessments' outputs. The main conclusion is that neural networks are reliable for conducting risks assessments that realistically integrate risks interdependencies in complex problems.

1 INTRODUCTION

Neural Networks (NNs) are inspired by the biological neural network of human brain (Haykin, 2008). According to Haykin (2008): "*A Neural Network is a massively parallel distributed processor made up of simple processing units that has a natural propensity for storing experimental knowledge and making it available for use. It resembles the brain in two respects: 1) Knowledge is acquired by the network from its environment through a learning process. 2) Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge*."

Three basic types of NNs' architecture are recognized (Haykin, 2008): 1) single-layer NN, 2) Multi-Layer NN (MLNN) and 3) recurrent NN. The learning processes through which NNs function can be categorized are (Haykin, 2008): 1) supervised learning, 2) reinforcement learning and 3) unsupervised learning. Usual basic problems that NNs are capable of dealing with are fitting (or function approximation), pattern recognition and association, and clustering and prediction, all applicable to many scientific fields such as automotive, financial, medical, robotics, telecommunications and management (Prieto et al., 2016).

The ability to use NNs either individually or in combination with other Artificial Intelligence (AI) techniques, such as Expert Systems, in construction industry and construction project management has been recognized since the beginning of 1990 (Moselhi, 1991). Currently, alternative tools of traditional methods based on AI techniques, such as NNs, are widely applied in engineering and construction industry (Paliwal & Kumar, 2009) and with a variety of ways in construction project management such as assessing project success and identifying critical success factors of the project, planning, estimating time and cost and managing risks (Magaña Martínez & Fernandez-Rodriguez, 2015).

Construction projects are featured by complexity and incomplete data records during development, thus affecting the capacity of a construction organization to conduct a credible risk assessment for every new project in hand. Furthermore, the inherently complex issue of risks interdependencies has led so far to approximate assessment approaches or to assessments based on simplistic assumptions that only remotely represent reality. This paper proposes the use of NNs for risk assessment in construction projects through the presentation of a specific application of a NN that partially assesses occupational risk in a construction engineering project.

2 METHODOLOGY

The paper, first, presents very briefly the basic characteristics of NNs to demonstrate the advantages of using them for risk assessment. Then, it investigates the existing applications of NNs in construction project risk management, in order to document their appropriateness for use in complex nonlinear problems such as construction projects. To this end, a detailed literature review identifies the different types and training methods of NNs that are applicable in construction projects risk management as well as of the respective tools that are available for real applications.

Based on the findings from the literature review, the paper, then, presents a specific application of a neural network that partially assesses occupational risk in a construction engineering project. The training of the proposed NN is based on metadata retrieved from a compilation of available data from previous risks assessments. The modeling of the network is realized with the help of two different software tools, namely Palisade's NeuralTools (Palisade Corporation, 2015) and Mathwork's Neural Network Toolbox, in order to identify potential difficulties in the modeling process as well as potential deviations in the assessments' outputs.

3 REVIEW OF NEURAL NETWORK APPLICATIONS IN CONSTRUCTION PROJECT RISK MANAGEMENT

As mentioned in Section 1, the ability to use NNs in the construction industry and construction project management has been recognized since the beginning of 1990. One of the first NN applications in construction project management is authored by Boussabaine (1996) who recognized the usefulness of NNs in construction project management and in risk analysis specifically. Especially in construction project risk management, previous NN applications focus on various topics as presented in Table 1. It is noted that even more applications than those presented in Table 1 were identified on the analysis of claims from legal disputes as well as of contractor's performance.

Especially for construction project risk management, the first application of NN was, probably, carried out by Sanchez (2005) with the aim of quantifying total risk in economic terms. Wen (2010) developed a total risk assessment model, embedded with NN, Genetic Algorithm (GA) and Rough Set Theory (RST) techniques, for construction projects. Zhu et al. (2011) implemented NN to perform analysis and evaluation of project cost risk and identification of critical factors. Chenyun & Zichun (2012) developed a Back-Propagation (BP) NN to assess risks in the construction phase of an expressway.

Concerning claim causing assessment, Al-Sobiei et al. (2005) developed two models for predicting the risk of contractor default in construction projects utilizing NN and GA techniques. Chau (2007) predicted the outcome of construction claims through the adaptation of a Particle Swarm Optimization (PSO) NN, trained with data from cases and past court decisions. Hosny et al. (2015) developed a NN-based predictive and decision-awareness framework for construction claims using backward optimization. Gholhaki et al. (2016) used Radial Basis Function (RBF) NN to predict claims' causes and control and minimize claims.

Manik et al. (2008) investigated the use of NNs to predict pavement construction payment-risk based on the quality of the construction. El-Sawalhi et al. (2008) developed a hybrid BP NN and GA model for predicting contractor's performance in terms of cost, time and quality in a process of contractor's prequalification. Jin & Zhang (2011) used NNs to model risk allocation decision-making process in PPP projects, mainly drawing upon transaction cost economics. Goh & Chua (2013) performed NN analysis in quantified occupational safety and health management system audit with accident data obtained from the Singaporean construction industry to predict accidents and identify critical factors.

Gajzler (2013) developed a method for supporting the decision-making process for the selection of materials and technology for repairing industrial building floors using Knowledge-based NN, Hybrid model of BP NN, RBF NN and Fuzzy Logic (FL). Gajzler & Konczak (2015) investigated the possibility of applying NN in the analysis of observational data on the issue of simulation concrete supplies in the construction industry. Patel & Jha (2016) developed an application of BP NN for the prediction and evaluation of employees' work behavior in construction projects using the constructs of the safety climate. Shahrara et al. (2016) used NNs to model the relationship between important project parameters and risk variables in the process of negotiating the financial parameters and uncertainties of a Build-Operate-Transfer project with the use of NN.

Regarding the software programs used in the abovementioned studies, Neural Network Toolbox of Mathwork's Matlab was used by Manik et al. (2008), Wen (2010), Jin & Zhang (2011), Li et al. (2012), Chenyun & Zichun (2012), Shahrara et al. (2015), and Gholhaki et al. (2016). Haykin (2008) also used it for the example applications of his work. Al-Sobiei et al (2005) and Goh & Chua (2013) deployed NeuroShell Predictor and Neu-

Table 1. Applications of NN in construction project risk management.

Author/ Researcher	Year	Innovative idea	Risk management processes
Sanchez	2005	NN application in construction project risk management with the aim of quantifying total risk in economic terms.	Quantitative risk analysis
Al-Sobiei et al.	2005	Development of two models for predicting the risk of contractor default in construction projects utilizing NN and GA techniques.	Qualitative risk analysis / Risk response analysis
Chau	2007	Adoption of a PSO NN model, provided with characteristics of cases and the corresponding past court decision, for predicting the outcome of construction claims. Comparison with BP NN model.	Qualitative risk analysis / Risk response analysis
Manik et al.	2008	Investigation of the use of NN to predict pavement construction payment-risk based on the quality of the construction.	Qualitative risk analysis / Risk response analysis
El-Sawalhi et al.	2008	Development of hybrid BP NN and GA model for predicting contractor's performance in terms of cost, time and quality in a process of pre-qualification.	Qualitative risk analysis / Risk response analysis
Wen	2010	Development of a total risk assessment model, embedded with NN, GA and RST techniques, for construction projects.	Quantitative risk analysis / Risk response analysis
Zhu et al.	2011	Analysis and evaluation of project cost risk and identification of critical factors based on BP NN.	Quantitative risk analysis
Jin & Zhang	2011	Modeling risk allocation decision-making process in PPP projects, mainly drawing upon transaction cost economics, using NNs. Comparison with multiple regression technique.	Risk response analysis
Chenyun & Zichun	2012	Application of BP NN to assess expressway construction phase risk.	Quantitative risk analysis
Goh & Chua	2013	NN analysis in quantified occupational safety and health management system audit with accident data obtained from the Singaporean construction industry in order to predict accidents and identify safety critical factors.	Qualitative risk analysis / Risk response analysis
Gajzler	2013	Developing a method for supporting the decision-making process for the selection of materials and technology for repairing industrial building floors using Knowledge-based NN. Hybrid model of BP NN, RBF NN and FL.	Qualitative risk analysis / Risk response analysis
Gajzler & Konczak	2015	Investigating the possibility of applying NN in the analysis of observational data on the issue of simulation concrete supplies in construction industry.	Qualitative risk analysis / Risk response analysis
Hosny et al.	2015	Development of a NN-based predictive and decision-awareness framework for construction claims using backward optimization.	Risk identification / Quantitative risk analysis / Risk response analysis
Patel & Jha	2016	Application of BP NN for prediction and evaluation of employees' work behavior in construction projects using the constructs of the safety climate.	Qualitative risk analysis / Risk response analysis
Shahrara et al.	2016	Modeling the relationship between important project parameters and risk variables in the process of negotiating the financial parameters and uncertainties of a BOT project with the use of NN.	Qualitative risk analysis / Risk response analysis
Gholhaki et al.	2016	Application of RBF NN to predict claims' causes and control and minimize claims.	Risk response analysis

roShell2 respectively. El-Sawalhi et al. (2005) used Neuro Genetic Optimizer for their research. Gajzler & Konczak (2015) used STATISTICA Data Miner.

Other software programs that were used in the general field of construction project management are NeuralWorks Professional II/Plus, Neurosolutions, SPSS and FANN Library.

A research concerning software use in NNs showed that Matlab is the most widely used software for NN implementation accumulating 28% users' choice (Baptista et al., 2013). The reasons for this preference are that it is a complete, flexible and easy to program software, has strong and fast computational power and several types of NNs available (Baptista et al., 2013). In the same research, "self-created code" (24%) and "other software" (36%) including NeuroSolution (1%), also accrue large percentages. Additionally, Matlab responds better to the needs of the NN research community with 29%, while SPSS collects 1% of the question "what other software are you using" (Baptista et al., 2013).

It is obvious that NNs can find application in a variety of ways in construction project risk management. Applications primarily introduce qualitative and quantitative risk analysis and risk response analysis addressing the problem with different approaches either as function approximation or as pattern recognition. In most of the cases, the use of NNs reproduces credible results, which is the reason behind suggesting the use of NNs in construction project risk management. BP algorithm has a prominent place in supervised learning and the development of robust and credible models. Furthermore, the combination of NNs with other techniques seems to present even more potential in the development of such models. Research into the implementation of NNs in construction project risk management is still a dynamic scientific field given the small number of scientific studies of the last decade in the field.

4 DEVELOPMENT OF A NEURAL NETWORK MODEL FOR RISK ASSESSMENT IN CONSTRUCTION PROJECTS

4.1 *General context*

A common and acceptable formula of risk assessment is shown in Equation (1):

$$R = P \times S \tag{1}$$

where R = risk; P = probability index; and S = severity of harm index or importance of effect index.

For several risk factors i with different probability of occurrence and consequences, equation (1) is extended to the forms shown in Equations (2) and (3):

$$R_i = P_i \times S_i \; i = 1, 2, \ldots, n \tag{2}$$

$$R = [R_1, R_2 \ldots, R_n] = [P_1 \times S_1, P_2 \times S_2, \ldots, P_i \times S_n] \tag{3}$$

Assuming that each risk R_i is independent of other risks, the total risk R can be assessed according to Equation (4):

$$R = \Sigma R_i = P_1 \times S_1 + P_2 \times S_2 + \ldots + P_n \times S_n \tag{4}$$

Nevertheless, considering individual risks as independent and using their algebraic sum to assess total risk is considered to be an extremely simplistic approach as well as mathematical modeling of risk. The complex and often unpredictable nature of construction projects and their risks inevitably contribute to complex nonlinear risks interdependencies.

NNs can approximate every function for the assessment of total risk in construction projects given that a sufficient volume of valid and reliable historical data is available (Haykin, 2008); however, as already mentioned, such a historical record is lacking in most of construction projects as former research efforts for applications of NNs in construction projects risk management evidently show.

A NN trained with data of similar projects to a project in hand can capture the inherent relation between individual risks and total risk of the projects. This is possible through the appropriate modeling of the project's development parameters in combination with accurate historical data on previous risks occurrences and their impact. The rightly designed NN can approximate the value of the total risk in an indirect, yet practical manner since the historical record used to train the network while it does not analytically present the interdependencies between risks, it inherently contains them; therefore, the trained NN possesses the respective knowledge on this aspect (i.e. total risk) and its assessment can be considered reliable and credible. The historical data can also serve for: a) deriving distributions of the probability of occurrence of each risk, and b) quantifying each risk's severity index. Both are useful in assessing the total risk of construction projects using modern computational tools such as NNs.

4.2 *Data compilation for training the neural network*

Occupational safety and health is a key factor towards a successful delivery of construction projects. Proper assessment of occupational risk in project's design stage can significantly contribute to avoiding accidents (Pinto et al., 2011). The importance of occupational risk assessment is evidenced by the increasing rate of published safety and health studies in the construction sector and the development of relevant risk assessment models (Zhou et al., 2015).

The training of the proposed NN is based on metadata retrieved from a compilation of available data from previous risk assessments. Available data concern the possible risk sources that may occur during the execution of construction projects, as well as the ranges of their probability of occurrence (P) and their severity of harm (S) (Argyriou, 2016). Risk sources are categorized into general risk categories based on the grouping of risk sources according to the Labor Inspection Body of Greece. In Table 2, the risk range $R_i = [R_{imin}, R_{imax}] = [P_{imin} * S_{imin}, P_{imax} * S_{imax}]$ for each source of one category and the total risk range $R_t = [\Sigma R_{imin}, \Sigma R_{imax}]$ of the category are evaluated, using Equations (2) and (4) and the minimum and maximum values of P and S, and presented. Totally, 300 scenarios of construction projects risks in the ranges described above were randomly generated. Each scenario was assigned 100 random values of three overlapping ranges. Values were generated from a uniform distribution for both individual risks and total risk. The use of the overlapping ranges was performed towards a more credible reflection of reality considering that they introduce more interdependencies to the modeling of a NN that assesses total risk. Two similar scenarios of data sets as described above and depicted in Figure 1 were used for training the NN. The difference between them lies in the extent of the overlap between risks ranges, as the second data set introduces a more extensive overlap. The reasoning behind this differentiation was to address more individual risks in the scenarios reflecting in this way the complex and nonlinear relationships which are met between risks in construction projects.

Table 2. An example of a risk category, risk sources and risk ranges.

Category	Risk source	Risk range
Fall of person	Fall from height—moving ladder	0,182–3,171
	Fall from height—steady ladder	0,182–3,171
	Fall from height—stairs or steps	0,061–3,171
	Fall from height—moving scaffolding or staging	0,091–3,171
	Fall from height—steady scaffolding or staging	0,091–3,171
	Fall from height—assembly/disassembly of scaffolding	0,091–3,171
	Fall from height—roof	0,151–3,171
	Fall from height—floor	0,061–3,171
	Fall from height—platform	0,151–3,171
	Fall from height—through floor openings	0,182–3,171
	Fall from height—mobile platform	0,182–3,171
	Fall from height—moving vehicle	0,182–1,269
	Fall from height—work at height without protection	0,091–2,230
	Fall at the same height—sliding / obstacle impact	0,061–2,262
	Fall down by stairs or ramp	0,242–2,262
Total risk range		2,001–42,904

4.3 *Neural network tools selection*

Several software products were investigated for modeling the developed NN. In order to identify potential difficulties in the modeling process as well as potential deviations in the assessments' outputs, a decision was made to test with two different software tools. The first choice was to use the Neural Network Toolbox of Mathwork's Matlab, which is considered as the most widely used software in NN modeling (Baptista et al., 2013).

The second choice was Palisade's NeuralTools because of its ease of use as an extension of Microsoft's widely used MS Excel and the ability to combine it with other management-oriented software programs of Palisade (e.g. @Risk). A final reason for selecting NeuralTools was that the conducted literature review revealed its limited use in research studies contrary to its wide application in practice.

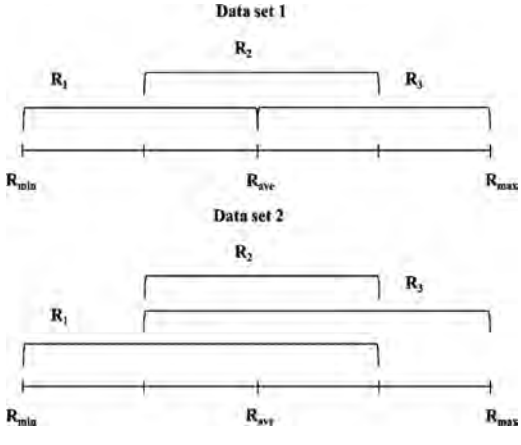


Figure 1. Data sets for training the NN.

This finding urged for application of NeuralTools to allow an insight and useful conclusions regarding to its appropriateness in NN modeling.

4.4 *Neural network training and results*

Supervised learning is the learning process used to train MLNNs developed both in NeuralTools and Neural Network Toolbox. MLNNs were chosen due to their ability to approximate any function satisfactorily thanks to the speed of learning and the plethora of options for their training as long as sufficient and reliable data is available. It is noted that the conjugate-gradient back-propagation algorithm is exclusively used in NeuralTools, while the Levenberg-Marquardt back-propagation algorithm is the default option in Neural Network Toolbox with the possibility of modification. In this research the default option was selected.

The first data set was randomly separated in 85% for training and 15% for testing the MLNN in NeuralTools. A 15-4-1 MLNN was finally chosen as the one with the best achieved performance. It is noted that available stopping criteria for the training can be a) training time, b) number of trials and c) training error decrease below a user-defined limit for a period of time, while there is also possibility of combining the above criteria. In this case, the training time was set to two hours; the number of trials was 10.000.000 and the error decrease less than 1% for five minutes. The third criterion was the one to cause training stop. The performance measures were: a) Root Mean Square Error (RMSE) equal to 3,873 and 5,046, b) Mean Absolute Error (MAE) equal to 3,019 and 4,261 and c) Standard Deviation of Absolute Error (SDAE) equal to 2,426 and 2,703 for the training and testing samples, respectively. Then, a regression analysis between the results of the network and the desired output was used to evaluate the reliability of the data. For that purpose, StatTools of Palisade was used and the results are depicted in Figure 2. The values of R (0,8920) and R^2 (0,7957) suggest a good relationship between results and desired outputs. The accumulation of extreme prediction values of the testing sample in one value, as seen in Figure 2, possibly indicates that the test data fall outside the range of the training and validation data and the network may be extrapolating. A better separation of the data set could lead to more accurate and reliable results.

The same 15-4-1 MLNN was then developed with Neural Network Toolbox. The data set was randomly divided into three samples; the training sample (70%), the validation sample (15%) and the test sample (15%). Training stopping criterion used in this case was the increase of the validation sample error for six iterations. The increase of the validation

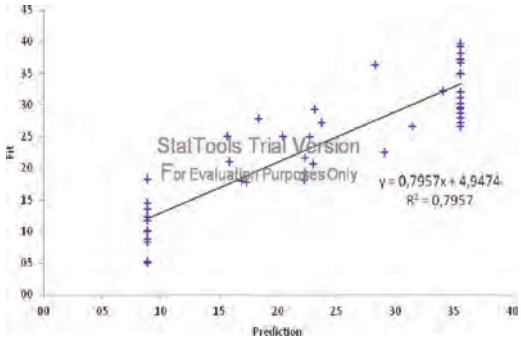


Figure 2. Regression analysis for the testing sample of 15-4-1 NN (NeuralTools).

sample error indicates the NN's inability to "explain" the data and in conjunction with the simultaneous reduction of the training sample error is a sign of memorization of the training sample. As a result, generalization, which is the NN's ability to deal with new data adequately, cannot be achieved. NN was trained for 11 epochs, while the best performance was achieved in the fifth epoch. Performance measure was RMSE equal to 4,362, 5,483 and 4,968 for training, validation and testing, respectively. Regression analysis between results and desired outputs suggests a good relationship between them for all three samples as depicted in Figure 3. Results with both software tools seem to be relatively close. Finally, a multiple regression analysis was carried out with the use of StatTools in order to compare NNs' results with those of a common and widely used traditional method. Aggregate results are presented in Table 3. Results are close enough to come to an exclusive conclusion. Furthermore, all models probably accommodate improvement.

The use of wider overlapping ranges used in the second data set, as mentioned in Section 4.2, is expected to match values of independent and dependent variables with more randomness resulting in even more complex relationship between individual risks and total risk.

In this case, the development of MLNN was performed with Neural Network Toolbox because regression analysis, which is a performance measure of the reliability of data division, is automatically generated, while using NeuralTools requires the processing of the results using another MS Excel extension, such as StatTools. Best performance was achieved with a 15-12-1 MLNN. The Levenberg-Marquardt BP algorithm was used for training the NN, while training stopping criterion was also the increase of the validation sample error for six iterations. Performance measure was RMSE and best value (5,645) was achieved at the third epoch for the validation sample. For the

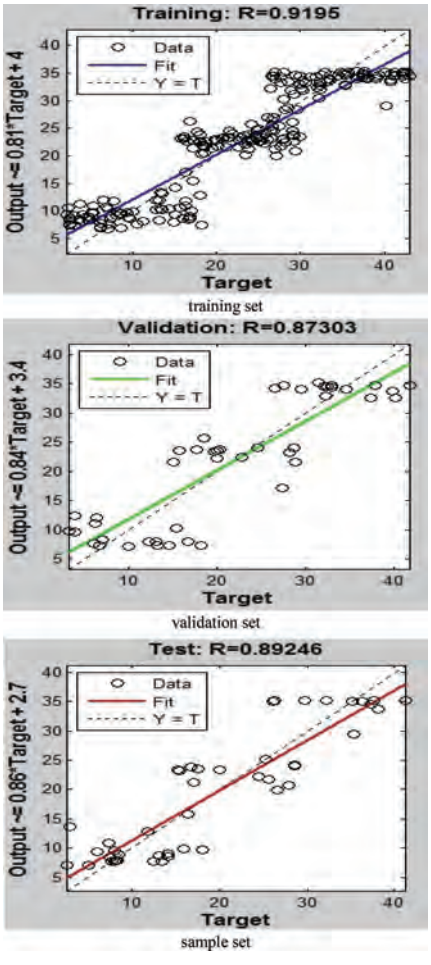


Figure 3. Regression analysis for training, validation and testing sample of 15-4-1 NN (Neural Network Toolbox).

Table 3. Regression analysis' R values for the testing sets of the models developed.

Model	R
NN 15-4-1 (Neural Tools)	0,8920
NN 15-4-1 (Neural Network Toolbox)	0,89246
Multiple regression analysis	0,8922
NN 15-12-1 (Neural Network Toolbox)	0,85649
Multiple regression analysis	0,837

training and testing sample, RMSE was equal to 5,676 and 6,451, respectively. Regression analysis between results and desired outputs delivered values of R equal to 0,861, 0,879 and 0,856 for training, validation and testing respectively, as depicted in Figure 4. Multiple regression analysis was also conducted and delivered R equal to 0,837. The relationship between variables is better depicted with the NN in relation to the regression model, as R values suggest. Aggregate results concerning training set R values of all the models developed are presented in Table 3.

Similarly, NNs can be developed to assess the risk of all risk categories, as well as the total risk of a construction project by combining those NNs. Generally, the results of this research demonstrate the computational power of NNs as a function approximation tool and confirm their usefulness in construction projects risk management and particularly in quantitative risk analysis, provided that reliable and sufficient historical data is available.

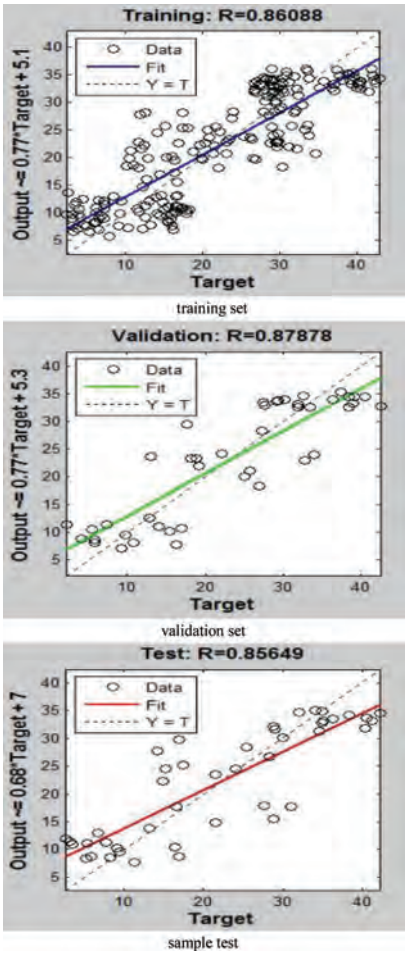


Figure 4. Regression analysis for training, validation and testing sample of 15-12-1 NN (Neural Network Toolbox).

5 CONCLUSIONS

NN applications in construction project risk management primarily concern qualitative and quantitative risk analysis and risk response analysis addressing the problem with different approaches either as function approximation or as pattern recognition. In most of the cases, the use of NNs reproduces credible results, as evaluated in the studies suggesting the use of NNs in construction project risk management. Similarly, results of the present research work demonstrate the computational power of NNs as function approximation tools and confirm their usefulness in construction projects risk management and particularly in assessing total risk. Neural networks are reliable for conducting risks assessments that realistically integrate risks interdependencies and complexities stemming from non-linearity in problems modeling. The availability of training data constitutes a prerequisite towards this goal. In this context, it is necessary to collect historical data concerning construction projects features and risks. Further research into the field can be addressed in improving the models developed, deploying other techniques and incorporating them in NN models for construction project total risk assessment.

REFERENCES

Al-Sobiei, O.S., Arditi, D., & Polat, G. 2005. Predicting the risk of contractor default in Saudi Arabia utilizing artificial neural network (ANN) and genetic algorithm (GA) techniques. *Construction Management and Economics*, 23(4): 423–430.

Argyriou, M. 2016. Occupational risk modeling per site activity in construction projects. *MSc thesis*, Patra: Hellenic Open University.

Baptista, D., Abreu, S., Freitas, F., Vasconcelos, R. and Morgado-Dias, F.A. 2013.A survey of software and hardware use in artificial neural networks. *Neural Computing and Applications*, 1: 1–9.

Boussabaine, A.H. 1996.The use of artificial neural networks in construction management : a review. *Construction Management and Economics*, 14(5):427–436.

Chau, K.W. 2007. Application of a PSO-based neural network in analysis of outcomes of construction claims. *Automation in Construction*, 16 (5): 642–646.

Chenyun, & Zichun, Y. 2012. The BP Artificial Neural Network Model on Expressway Construction Phase Risk. *Systems Engineering Procedia*, 4: 409–415.

El-Sawalhi, N., Eaton, D., & Rustom, R. 2008. Forecasting contractor performance using a neural network and genetic algorithm in a pre-qualification model. *Construction Innovation: Information, Process, Management*, 8(4): 280–298.

Gajzler, M., & Konczak, A. 2015. The Possibility of Using Neural Networks in Data Analysis Connected with Observation in the Construction Process Simulation. *Procedia Engineering*, 122: 228–234.

Gajzler, M. 2013. The idea of knowledge supplementation and explanation using neural networks to support decisions in construction engineering. *Procedia Engineering*, 57: 302–309.

Gholhaki, M., Kheiroddin, A. &Ghorbani, A. 2016.Claim causing assessment in construction projects in Iran using artificial neural networks model: Radial Basis Function (RBF). *Journal of Engineering and Applied Sciences*, 11(5): 1122–1127.

Goh, Y.M., & Chua, D. 2013. Neural network analysis of construction safety management systems: a case study in Singapore. *Construction Management and Economics*, 31(5): 460–470.

Haykin, S. 2008. *Neural Networks and Learning Machines, 3rd Edition*. USA, Prentice Hall.

Hosny, O.A., Elbarkouky, M.M.G., & Elhakeem, A. 2015. Construction Claims Prediction and Decision Awareness Framework using Artificial Neural Networks and Backward Optimization. *Journal of Construction Engineering and Project Management*, 1(5): 11–19.

Jin, X.-H., & Zhang, G. 2011.Modelling optimal risk allocation in PPP projects using artificial neural networks. *International Journal of Project Management*, 29(5): 591–603.

Li, M., & Chen, W. 2012. Application of BP Neural Network Algorithm in Sustainable Development of Highway Construction Projects. *Physics Procedia*, 25: 1212–1217.

Magaña Martínez, D., & Fernandez-Rodriguez, J.C. 2015. Artificial Intelligence Applied to Project Success: A Literature Review. *International Journal of Interactive Multimedia and Artificial Intelligence*, 3(5): 77–82.

Manik, A., Gopalakrishnan, K., Singh, A., & Yan, S. 2008. Neural networks surrogate models for simulating payment risk in pavement construction. *Journal of Civil Engineering and Management*, 14(4): 235–240.

Moselhi, O., Hegazy, T. & Fazio, P. 1991. Neural Networks as Tools in Construction. *Journal of Construction Engineering and Management*, 117(4): 606–625.

Palisade Corporation 2015. *NeuralTools. User's Guide*. Retrieved fromhttp://www.palisade.com/downloads/documentation/7/EN/NeuralTools7_EN.pdf

Paliwal, M., & Kumar, U.A. 2009.Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, 36(1): 2–17.

Patel, D.A., & Jha, K.N. 2016. Evaluation of construction projects based on the safe work behavior of co-employees through a neural network model. *Safety Science*, 89: 240–248.

Pinto, A., Nunes, I.L., & Ribeiro, R.A. 2011.Occupational risk assessment in construction industry – Overview and reflection. *Safety Science*, 49(5), 616–624.

Prieto, A., Prieto, B., Ortigosa, E.M., Ros, E., Pelayo, F., Ortega, J., & Rojas, I. 2016. Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, 214: 1–27.

Sanchez, P.M. (2005). Neural-risk assessment system for construction projects. *Construction Research Congress 2005: Broadening Perspectives - Proceedings of the Congress*, 1405–1414. Retrieved from https://www.engineeringvillage.com/share/document.url?mid=cpx_18a9.

Shahrara, N., Çelik, T., & Gandomi, A.H. 2016.Risk analysis of BOT contracts using soft computing. *Journal of Civil Engineering and Management*, 3730(2007): 1–9.

Wen, G. 2010. Construction project risk evaluation based on Rough Sets and Artificial Neural Networks. *Proceedings of 2010 Sixth International Conference on Natural Computation*, 1624–1628.

Zhu, B., Zhang, H., & Wang, X. 2011. Analysis and Evaluation of Project Cost Risk Based on BP Algorithm. *Systems Engineering Procedia, 1*: 264–270.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Analysis of domino scenarios in chemical and process facilities operating in harsh environmental conditions

M. Bucelli
Alma Mater Studiorum—University of Bologna, Bologna, Italy
Norwegian University of Science and Technology NTNU, Trondheim, Norway
Safetec, Trondheim, Norway

G. Landucci
University of Pisa, Pisa, Italy
University of Leiden, Den Haag, The Netherlands

S. Haugen & N. Paltrinieri
Norwegian University of Science and Technology NTNU, Trondheim, Norway

V. Cozzani
Alma Mater Studiorum—University of Bologna, Bologna, Italy

ABSTRACT: In the framework of chemical and process industry, accidental fires may lead to damages to equipment with severe consequences and possible domino effects. The availability and effectiveness of safety measures, aimed at reducing the risk associated with this type of events, may be strongly affected and decreased if the facility is located in harsh environment, due to complicating meteorological factors and extreme temperatures. The present work is aimed at defining a structured approach to the quantitative assessment of fired domino events accounting for the influence of harsh environment conditions on safety barriers performance. A specific metric is defined in order to consider the external factors related to harsh environments on the determination of hardware and emergency safety barriers availability and effectiveness, with a specific focus on the evaluation of the time-scale of emergency response. A dedicated event tree analysis is then applied implementing the obtained performance values of the safety barriers, in order to support the quantitative assessment of accident frequency associated with domino scenarios. The present method is applied to the analysis of a chemical facility located in harsh environmental conditions.

1 INTRODUCTION

In the last decades, interest has been increasing for cascading events and the assessment of their possible risks. The chemical process industry has been hit by major accidents worldwide, some of which were completely disregarded by hazard identification techniques (Paltrinieri et al., 2010; Paltrinieri and Reniers, 2017). Among them, several domino events have been documented (Abdolhamidzadeh et al., 2011; Darbra et al., 2010; Delvosalle, 1996; Kourniotis et al., 2000; Lees, 1996; Rasmussen, 1996).

One of the most destructive cascading event disasters is the one that happened in Mexico City in 1984 (Pietersen, 1988). Europe recognized the hazard posed by domino events and specific requirements are stated in the article 9 of the latest Seveso Directive (European Commission, 2012). According to these, the risk of propagation of primary hazardous scenarios to nearby units is required to be assessed.

Different safety barriers are used and monitored in chemical process plants (Paltrinieri and Khan, 2016), such barriers defined to prevent escalation scenarios. These include active, passive and procedural protections. Examples include the water deluge system (WDS), fireproofing coating, pressure safety valves (PSVs) and the site emergency response plan. Different performance parameters in terms of availability (expressed as probability of failure on demand) and effectiveness are associated to every safety barrier.

However, barriers are subject to deterioration and depletion of their performance. Meteorological and climatological conditions are factors that can

enhance these phenomena. For instance, cold temperatures, extreme wind and snowfall may either cause deterioration of hardware plant components or lead to difficulties for operators performing routine tasks and/or in emergency contingency situations (Bercha et al., 2003; Gao et al., 2010). The Arctic and sub-Arctic regions experience extremely unique weather conditions that may be challenging for technical barrier components as well as human intervention. However, a dedicated framework for the analysis of safety barriers performance degradation in harsh environment is still missing.

This work is aimed at investigating the safety barrier performance of chemical and process facilities operating in harsh environmental conditions, in order to evaluate the frequency and probability of escalation scenarios triggered by fire.

The paper is organized as follows: Section 2 provides a detailed overview of the methodology applied to assess the frequency of cascading events addressing the effect of severe environment on protection devices; Section 3 describes the reference case considered for the present analysis; the results of the application of the methodology to the reference case are shown in Section 4, while Section 5 provides room for their discussion. The paper ends with conclusions in Section 6.

2 METHODOLOGY

2.1 *Overview*

Figure 1 shows the flowchart of the methodology adopted in the present study. The methodology was developed for the oil and gas sector (Landucci et al., 2017) and it is hereby extended to chemical process industry. A detailed description of the methodology is provided in sections 2.2–2.5.

2.2 *Identification of reference safety barriers*

The first step of the methodology consists of a preliminary characterization of the safety barriers performance, with particular reference to the prevention and mitigation of cascading events triggered by fire. According to CCPS—Center of Chemical Process Safety (2000), barriers are classified as:

- Passive, which are in place and do not require external activation;
- Active, which require automatic and/or external activation;
- Procedural and emergency measures, which involve the intervention of operators and emergency teams.

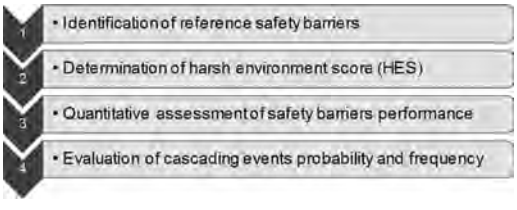


Figure 1. Flowchart of the methodology.

This step is based on the application of a previously developed methodology (Landucci et al., 2016) in which the evaluation of safety barriers performance in the framework of escalation is aimed at quantifying:

- availability, defined as the probability of failure on demand (*PFD*) of the safety barriers;
- effectiveness (η), defined as the probability that the safety barrier, once successfully activated, will be able to prevent the escalation.

Once the parameters needed to support the quantitative evaluation of safety barriers are defined, the influence of harsh environmental conditions on their performance is inferred in the following steps.

2.3 *Definition of Harsh Environment Score (HES)*

The Harsh Environment Score (*HES*) is a preliminary metric aimed at describing the harshness of the environment and it is used to assess the influence of weather conditions on safety devices performance. *HES* consists of a combination of different site-specific environmental parameters, such as, for instance, temperature and wind velocity.

The approach for the *HES* evaluation is based on the identification of stressors. They are factors that mostly affect the human performance during operations in extreme weather conditions (Section 2.4.1) but are adopted in the present study also to address the influence of extreme weather conditions on hardware barriers performance (Section 2.4.2).

Musharraf et al. (2013) identify the significant stressors for harsh environment as coldness, ice slippery, difficulty in breathing, combined weather effect, low visibility and remoteness. The present approach associates one or more external factors (EFs) to each stressor. EFs are climate or environmental conditions that can be measured and/or quantified. To each EF, a non-dimensional penalty, namely a score S_i, is assigned. Scores represent the distance from favorable conditions. They vary from 0 to 1, where 0 represents good favorable conditions and 1 the worst ones. Table 1 lists the EFs and relative scoring system applied in the present study.

Table 1. Summary of external factors and scores adopted for HES evaluation (adapted from Landucci et al., 2017).

External factor	ID	Range	S_i
Temperature (°C)	1	>45	0.4
		4 to 45	0
		–4 to 4	0.2
		–10 to –4	0.6
		–30 to –10	0.8
		<–30	1
Extreme wind speed (m/s)	2	0 to 3.3	0
		3.3 to 5.5	0.2
		5.5 to 8	0.4
		8 to 10.8	0.6
		10.8 to 13.9	0.8
		>13.9	1
Snowfall (m/year)	3	0 to 0.125	0
		0.125 to 0.5	0.2
		0.5 to 1	0.4
		1 to 1.5	0.6
		1.5 to 2	0.8
		>2	1
Visibility (fog/snow) (m)	4	<50	1
		50 to 200	0.8
		200 to 500	0.6
		500 to 1000	0.4
		1000 to 2000	0.2
		>2000	0
Sunlight hours (h/year)	5	<1200	1
		1200 to 1600	0.8
		1600 to 2000	0.6
		2000 to 2400	0.4
		2400 to 3000	0.2
		>3000	0
Remoteness	6	Low	0
		Medium	0.5
		High	1

More detailed information about the scores assignment process and the EFs may be retrieved in a previous study (Landucci et al., 2017). The scores are assigned according to extensive literature surveys about the effects of different physical factors on technical and human behavior (American Petroleum Institute, 2000; DOA—Department of Army, 1982; Kunkel et al., 2007; Landsberg and Pinna, 1978; Musharraf et al., 2013; Shaw and Austin, 1919).

Finally, *HES* is obtained as a weighted summation of the assessed scores, as follows:

$$HES = \sum_{i=1}^{N} w_i S_i \tag{1}$$

where S_i and w_i are respectively the score and the weight associated to the *i*-th EF. In the present analysis, a preliminary set of weights is assigned by using the Zipf's law (Zipf, 1949).

2.4 *Barrier performance assessment*

2.4.1 *Hardware barriers*

According to Gao et al. (2010), extreme environmental conditions may affect hardware barrier availability but they have no significant effect on their effectiveness. The depletion of barrier performance is strictly related to environmental temperature. Recommended Practices 581 by American Petroleum Institute (2000) identify a threshold value of –6.7°C for considerable effect on protection performance. This value corresponds to a penalty $S_1 = 0.6$ or higher according to Table 1. This framework addresses the depletion in barrier availability using the proportional hazard model (Cox, 1972) as suggested by Gao et al. (2010). The failure rate of a generic component, λ, increases in harsh environment according to the following relationship:

$$\lambda(z) = \lambda_0 \, e^{-1.409 z_1 - 1.013 z_2} \tag{2}$$

where λ_0 is the failure rate in normal environment (namely, the baseline value), assumed hereby as constant during the entire lifecycle of the facility. The factors z_1 and z_2 are the named covariates; z_1 describes the protection conditions and z_2 the equipment quality, respectively. Covariates are considered as binary and they can assume the value +1 or –1. The positive value is associated with good quality of protections and equipment. The base relationship for the estimation of tested component unavailability (Lees, 1996) is applied to obtain the barrier *PFD* describing, from this analysis perspective, the barrier availability.

The present work considers that the effectiveness of the barriers is not affected by environmental conditions. Once activated, hardware barriers perform as in the case of normal environment (Landucci et al., 2016).

The reference active safety barriers analyzed in the present study are water deluge systems (WDS) aimed at attenuating heat radiation from fires affecting process units. According to different experimental studies (Hankinson and Lowesmith, 2004; Roberts, 2004a, 2004b; Shirvill, 2004), the heat-load reduction on a target due to presence of WDS is about 50% compared to the unmitigated case. Hence, Q_{WDS} (the heat load received by a fired target in case of available WDS) is expressed as follows:

$$Q_{WDS} = 0.5 \, Q_{HL} \tag{3}$$

where Q_{HL} represents the heat-load affecting the target due to the primary fire scenario.

Passive safety protections include the PSV and the fireproofing coating. Birk (2006) proved that the presence of the PSV alone does not delay significantly the time to failure (*TTF*) of the target equipment. In that case, the PSV effectiveness is considered as unitary but the TTF is evaluated assuming that the vessel is unprotected (Landucci et al., 2009). Fireproofing coatings are instead able to delay the vessel failure. Their effectiveness is set as 1. The *TTF* of the target vessel in case of presence of protective coatings is evaluated by adding a further term, TTF_C, as shown in Eq. (4), which represents the delay action of the coating:

$$TTF = TTF_{unprotected} + TTF_C \tag{4}$$

The TTF_C is evaluated according to a simplified approach considering the quality of the materials used as coating. For high performance materials (intumescent, vermiculite spray, fibrous mineral wool) the TTF_C is set conservatively as 70 minutes. TTF_C is equal to 0 minutes in case of use as coatings of common insulating materials (glass wool, rock wool).

2.4.2 *Procedural barriers*

Human reliability may be significantly affected by extreme weather (Musharraf et al., 2013).

A customized version of the Success Likelihood Index Methodology (SLIM) (Embrey, 1986) is adopted in the present framework to evaluate the deterioration of emergency response availability (e.g. in terms of *PFD*). *HES* is considered as a simplified ranking of performance shaping factors affecting the emergency response in harsh environment (Landucci et al., 2017). The higher the *HES* the lower the probability of success of the emergency team intervention. The *PFD* is then evaluated as:

$$\log_{10} PFD = a(1 - HES) + b \tag{5}$$

where *a* and *b* are –0.954 and –0.046 respectively. They have been determined by setting the *PFD* equal to 0.1 in case of favorable environmental conditions (*HES* = 0) and by setting the *PFD* as 0.9 in worst case environmental conditions (*HES* = 1) (Landucci et al., 2016).

The evaluation of the emergency response effectiveness is carried out by following the approach suggested by Landucci et al. (2017). The evaluation is based on the comparison between the *TTF* of the target equipment and the Time for Final Mitigation (*TFM*) required to the emergency team to extinguish the primary fire. The *TFM* is defined as the sum of different times for emergency operations as follows:

$$TFM = \sum_{j=1, j\neq 2}^{6} \tau_j \tag{6}$$

The times are defined according to Table 2 (Landucci et al., 2017), where also the different relationships applied to account for the delay due to harsh environment are shown. The effectiveness of the emergency response is set equal to 1 or 0 by comparison between *TFM* and *TTF* of the target equipment. When *TFM* is lower than *TTF* of the target equipment, the emergency response effectiveness is set as unitary, otherwise it is zero.

2.5 *Evaluation of escalation probability*

A customized Event Tree Analysis (ETA) is adopted in order to evaluate the frequency (and probability) of domino escalation triggered by fire. The availability and effectiveness of barriers evaluated as described in Section 2.4 are addressed in the ETA by using dedicated logic gates, as shown in Table 3.

Further detailed information about gate definitions may be retrieved elsewhere (Landucci et al., 2016).

Gate A represents a simple composite probability. In this case, the availability (expressed in terms of PFD) is multiplied by a single probability value expressing the probability of barrier success in the prevention of the escalation.

Gate B represents a composite probability distribution. In this case, the PFD is multiplied by a

Table 2. Time scale for emergency operations and simplified relationship for the estimation of time increment due to harsh environment (adapted from Landucci et al., 2017). The baseline is the time required in normal environment (HES = 0).

ID	Name	Baseline (min)	Simplified relationship (τ in min)
τ_1	Time to alert	5	$\log_{10}\tau_1 = -0.3(I - HES) + 1$
τ_2	Time to onsite mitigation	20	$\log_{10}\tau_2 = -0.3(I - HES) + 1.6$
τ_3	Time for external team intervention	12	$\log_{10}\tau_3 = -0.3(I - HES) + 1.38$
τ_4	Time for equipment deployment	7	$\log_{10}\tau_4 = -0.3(I - HES) + 1.15$
τ_5	Time for extra set-up operations	8	$\log_{10}\tau_5 = -0.3(I - HES) + 1.2$
τ_6	Additional time in case of need of interregional assistance	30–60[a]	$\log_{10}\tau_1 = -0.3(I - HES) + 2.08$

[a]Depending on the type of location.

Table 3. Summary of gates introduced in the ETA to account of barrier performance (adapted from (Landucci et al., 2016)).

Gate type	Graphical representation
A	IN → A → $OUT_1 = IN \cdot [PFD + (1-\eta) \cdot (1-PFD)]$; $OUT_2 = IN \cdot (1-PFD) \cdot \eta$
B	IN → B → $OUT_1 = IN \cdot [PFD + (1-\eta) \cdot (1-PFD)]$; $OUT_2 = IN \cdot (1-PFD) \cdot \eta$
C	IN → C → $OUT_1 = IN \cdot PFD$; $OUT_2 = IN \cdot (1-\eta) \cdot (1-PFD)$; $OUT_3 = IN \cdot (1-PFD) \cdot \eta$
D	IN → P_D → $OUT_1 = IN \cdot P_D$; $OUT_2 = IN \cdot (1-P_D)$

probability distribution expressing the probability of barrier success in the prevention of escalation, thus obtaining a composite probability of barrier failure on demand. In this work, the integrated probability is adopted, obtaining the rule for gate quantification reported in Table 3.

Gate C is associated with a discrete probability distribution.

Finally, Gate D incorporates equipment vulnerability models based on probit approaches for the estimation of P_D (the probability of vessel failure). The effect of harsh environmental conditions has been addressed in the probit models in describing the vessel resistance behaviour. More details on vessel fragility models are extensively described in previous works (Landucci et al., 2009).

3 CASE STUDY

3.1 *Overview*

The reference case study refers to a production plant for the production of personal and home hygiene products. The plant uses as main raw materials ethanol and propane and, for the quantities stored, it is subject to fulfill the Seveso Directive requirements concerning hazardous materials (European Commission, 2012). The field is located in harsh environment (see Section 3.2). The methodology described in Section 2 is applied to estimate the frequency of domino events triggered by fire and thus providing a more complete risk picture of the facility.

Figure 2 shows the layout considered in the analysis of the case study. Ethanol is stored in three underground tanks (T1, T2, T3) with an overall volume of 90 m^3 and kept at 15°C. Ethanol is transferred to the processing area (see Fig. 2)

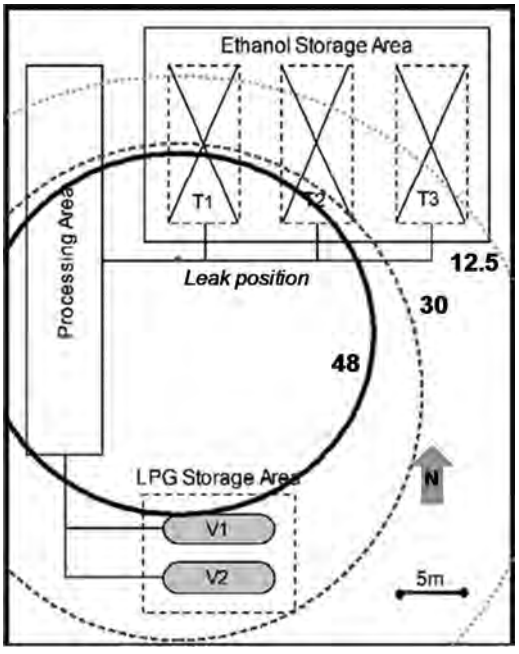


Figure 2. Layout defined for the case study associated with a non-confined pool-fire following the rupture of the process ethanol pipeline.

through a pipeline featuring 20 m length and a nominal diameter of 100 mm. Full-bore rupture of the pipeline is considered to derive the features of the primary scenario potentially triggering the domino escalation. In particular, a non-confined pool fire following immediate ignition of the spilled ethanol is taken into account. A standard frequency of $3.9 \cdot 10^{-7}$ y^{-1} has been assumed from literature analysis for pool-fire. The physical effects associated with the pool fire have been analyzed applying the conventional literature integral models implemented in the DNV GL Phast 7.11 commercial software. According to consequence assessment results, the pool fire affects the target propane storage tank (V1, see Fig. 2), which is exposed to about 48 kW/m^2.

The safety barriers in place to protect V1 are listed in Table 4. They are defined on the basis of different regulations for fire protection of liquefied petroleum gas storage units (American Petroleum Institute, 1996; National Fire Protection Agency (NFPA), 2018, 2017). The results of their performance assessment (in normal and harsh environments) are shown in Section 4. The quality of both the target equipment V1 and its protection devices is assumed as low following a conservative approach.

Table 4. Summary of fire protection devices for horizontal LPG storage tank (American Petroleum Institute, 1996; National Fire Protection Agency (NFPA), 2018, 2017).

Target	Active barriers	Passive barriers	Procedural barriers
V1	Water deluge system (WDS-V1)	Pressure safety valve (PSV-V1) Fireproofing coating (PFP-V1) (2 h rating)	Emergency response (ER-01)

3.2 *Environmental and meteorological conditions*

The reference production plant is located in an industrial site close to Bodø just North of the Arctic Circle, in Norway. The climatic conditions in the reference area can be characterized as severe. Table 5 summarizes the meteorological and climatological conditions experienced in that area and adopted for the determination of *HES* and, thus, to derive performance data in harsh environment.

4 RESULTS

4.1 *Performance assessment of safety barriers*

Adverse meteorological conditions significantly affect the protection effect of safety devices. In order to account for this effect, the methodology described in Section 2 has been applied to the reference chemical processing plant described in Section 3.

According to the meteorological and climatological data summarized in Table 5 and to the scoring system described in Section 2.3, the estimated *HES* is 0.43 for the considered case. This value is implemented to evaluate the performance of the safety barrier protecting the target tank V1. Since the score associated with the external temperature $S_1 = 0.6$, a degradation of hardware barrier availability must also be considered (see Section 2.4.1).

Data were also calculated for normal environmental conditions for sake of comparison (thus, featuring *HES* = 0). The time for external emergency response is calculated according to the guidelines described in Section 2.4. It increases from 77 minutes (normal environmental conditions, HES = 0) to 124 minutes (harsh environmental conditions, HES = 0.43).

Table 6 summarizes the results of performance assessment in normal and harsh environment and it shows the gates associated with each barrier.

4.2 *Evaluation of escalation probability*

The customized ETA approach for the evaluation of the escalation probability and frequency has been carried out starting from the frequency and consequence assessment of the primary scenario (ethanol non-confined pool-fire).

Figure 3 shows an extract of the ETA developed for harsh environmental conditions (*HES* = 0.43). Each branch in the event tree is quantified according to the rules described in Section 2. A similar event tree is derived for the normal environment case.

Three different scenarios arising from unconfined pool fire are analyzed in both normal and harsh environment. These scenarios are:

Table 5. Summary of meteorological and climatological conditions experienced in Bodø.

Factor	Meteorological data	Reference
Temperature	Coldest month: January Minimum average temperature: –11.8°C Typical value: –2.2°C	(Norwegian Meteorological Institute, 2017)
Wind speed	Harsh month: January Maximum wind speed: 24.4 m/s (10 m above sea level) Annual range: 8.9 m/s	(Norwegian Meteorological Institute, 2017)
Snow	Duration: 6 months (October-April) Average snowfall per day: 2.54 cm	(weatherspark.com, 2017)
Fog/snow effect	Visibility lower than 2000 m	(ISO-International standardization organization, 2010)
Sunlight hours	1200–1600 h/year	(Landsberg and Pinna, 1978)
Remoteness	The plant is located in an industrial site close to cities and amenities. The remoteness is considered to be low.	(Suedfeld and Steel, 2000)

Table 6. Summary of data adopted for the quantification of the ETA in the present case study. HES = 0: normal environment; HES = 0.43: harsh environment.

Barrier	Gate type	PFD		Effectiveness	
		HES = 0	HES = 0.43	HES = 0	HES = 0.43
WDS-V1	A	$4.33\ 10^{-2}$	$5.57\ 10^{-1}$	1	1
PSV-V1	A	$1\ 10^{-2}$	$1.29\ 10^{-1}$	1	1
PFP-V1	A	$1\ 10^{-3}$	$1.29\ 10^{-2}$	1	1
ER-01	C	$1\ 10^{-1}$	$2.57\ 10^{-1}$	0; 1[a]	0; 1[a]

[a]Depending on the comparison between TFM and TTF.

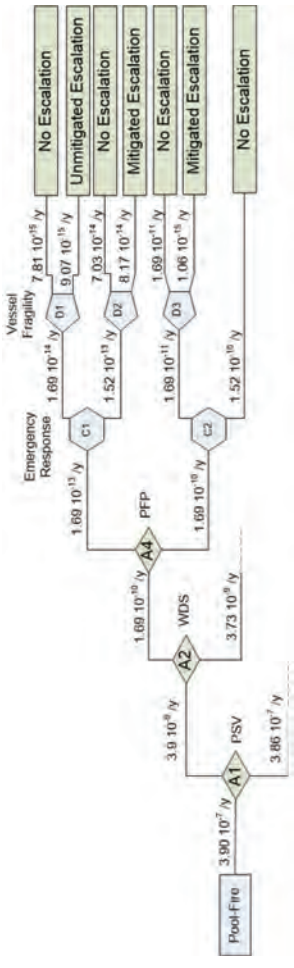


Figure 3. Extract of the ETA for the evaluation of cascading event probability/frequency for the target tank V1. It refers to the case of harsh environment, with $HES = 0.43$.

1. Unmitigated domino (not effective activation of safety barriers);
2. Mitigated domino (partial or ineffective activation of one or more safety barriers);
3. No domino scenario (barriers effectively mitigate/suppress the primary fire and avoid escalation).

The target equipment V1 may withstand the fire even in the absence of barrier activation. Also in these cases, escalation is excluded.

Figure 4 shows the result of the analysis in terms of frequency and probability of the three examined scenarios. The "No safety barrier" scenario has been considered for sake of comparison, e.g. based on the method developed in a previous work (Landucci et al., 2009).

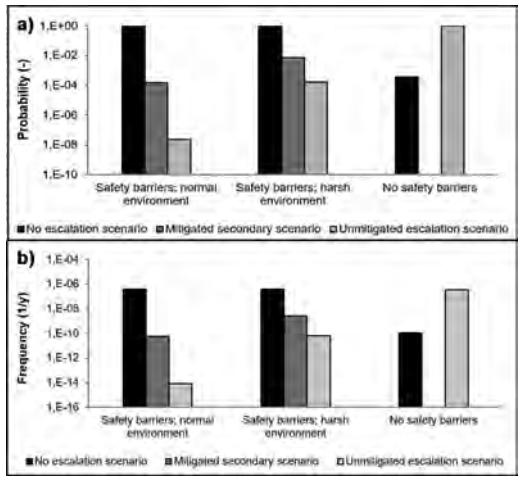


Figure 4. a) Probability and b) frequency of secondary scenarios.

5 DISCUSSION

The analysis of the case study demonstrates the potentialities of the methodology in the assessment of domino scenarios for chemical facilities located in harsh environments. As shown in Figure 4, a significant increase in escalation probability and frequency is predicted in harsh environment operation with respect to normal environment. When safety barriers are considered, unmitigated domino scenario is the less credible, both in normal and harsh environments. Anyway, the degradation of barrier performance in harsh environment leads to higher frequency values. In particular, reduction of four orders of magnitude with respect to the case without protection is obtained for harsh environment. In normal environment, the reduction is of eight orders of magnitude.

This is due to the depletion in the barrier performance in harsh environment, as documented in the analysis shown Section 4.1. In particular, procedural and emergency measures are significantly affected by cold environmental conditions. In fact, the time for external emergency response increases about 60% compared to the value in normal environment. This is due to delays and difficulties in carrying out emergency actions.

The escalation frequency results obtained from the ETA analysis shown in Section 4.2 may be implemented in detailed quantitative risk assessment studies. In this way, a more detailed risk picture of the facility may be evaluated, thus including escalation scenarios. The necessary input to apply the method, as exemplified in Sections 3 and 4, is normally available from conventional risk analysis

studies and therefore no additional work needs to be carried out for collecting input data. The meteorological and climatological data for the HES assessment are site-specific, but easily retrievable from national institutes (see the example dataset gathered in Table 5).

It is worth mentioning that the methodology addresses human factor and deterioration of barrier phenomena in a very simplified way, despite these issues featuring relevant complexity. For that reason, the so evaluated escalation probabilities and frequencies should be considered on the safe side.

The methodology allows room for further refinement of data and for using different available methods. In particular, for human reliability, more advanced techniques may be implemented supporting the evaluation of operators' performance and error probability given the environmental stressors; on the same time, emergency response analysis may be improved with site specific response time data for a more accurate effectiveness estimation.

Finally, for hardware barriers, further review of the methodology should be considered when site-specific performance data will be available from facilities operating in harsh cold environments.

6 CONCLUSIONS

The present contribution shows a systematic approach for the quantification of domino event frequency and probability for chemical facilities operating in harsh environmental conditions. The approach accounts for the deterioration of safety barriers performance due to extreme climate conditions. A dedicated metric is used as preliminary index to assess the influence of environmental conditions on barrier performance, thus allowing for a modification of barriers availability and effectiveness. The modified values of barrier performance data allow for a more detailed probability and frequency assessment of cascading scenarios triggered by fire.

The outcomes of the methodology may drive the design of hardware barrier components and improvement of emergency procedures in order to decrement the risk of severe accidental scenarios in chemical facilities operating in harsh environments.

REFERENCES

Abdolhamidzadeh, B., Abbasi, T., Rashtchian, D., Abbasi, S.A., 2011. Domino effect in process industry—an inventory of past events and identification of some patterns. J. Loss Prev. Process Ind. 24, 575–593.

American Petroleum Institute, 2000. API Publication 581. Risk-based inspection base resource document.

American Petroleum Institute, 1996. Fire-protection considerations for the design and operation of liquified petroleum gas (LPG) storage facilities—Standard API2510A.

Bercha, F., Brooks, C., Leafloor, F., 2003. Human performance in arctic offshore escape, in: Proceedings of the Thirteenth International Offshore and Polar Engineering Conference. International Society of Offshore and Polar Engineers (ISOPE), Cupertino, Ca. USA, Honululu, Hawaii, USA, pp. 2755–2762.

Birk, A.M., 2006. Fire testing and computer modelling of rail tank-cars engulfed in fires: literature review. Ottawa, Canada.

CCPS—Center of Chemical Process Safety, 2000. Guidelines for chemical process quantitative risk analysis. American Institute of Chemical Engineers—Center of Chemical Process Safety, New York.

Cox, D.R., 1972. Regression Models and Life-Tables. J.R. Stat. Soc. Ser. B 34, 187–220.

Darbra, R.M., Palacios, A., Casal, J., 2010. Domino effect in chemical accidents: main features and accident sequences. J. Hazard. Mater. 183, 565–573.

Delvosalle, C., 1996. Domino effect phenomena: definition, overview and classification, in: Proceedings of European Seminar on Domino Effects. Leuven (B), pp. 5–10.

DOA—Department of Army, 1982. FM 6-16-2 Tables for Artillery Meteorology (Visual) Ballistic Type 3 and Computer Messages and Limited Surface Observations. Washington DC.

Embrey, D.E., 1986. SLIM-MAUD: a computer based technique for human reliability assessment. Int. J. Qual. Reliab. Manag. 3, 15–12.

European Commission, 2012. European Parliament and Council Directive 2012/18/EU of 4 July 2012 on control of major-accident hazards involving dangerous substances, amending and subsequently repealing council directive 96/82/EC. Off. J. Eur. Communities L197, 1–37.

Gao, X., Barabady, J., Markeset, T., 2010. An approach for prediction of petroleum production facility performance considering Arctic influence factors. Reliab. Eng. Syst. Saf. 95, 837–846.

Hankinson, G., Lowesmith, B.J., 2004. Effectiveness of area and dedicated water deluge in protecting objects impacted by crude oil/gas jet fires on offshore installations. J. Loss Prev. Process Ind. 17, 119–125.

ISO-International standardization organization, 2010. ISO 19906:2010 Petroleum and natural gas industries—Arctic and offshore structures.

Kourniotis, S.P., Kiranoudis, C.T., Markatos, N.C., 2000. Statistical analysis of domino chemical accidents. J. Hazard. Mater. 71, 239–252.

Kunkel, K.E., Palecki, M.A., Hubbard, K.G., Robinson, D.A., Redmont, K.T., Easterling, D.R., 2007. Trend identification in twentieth-century U.S. snowfall: the challenges. J. Atmos. Ocean. Technol. 24, 64–73.

Landsberg, H., Pinna, M., 1978. L'atmosfera e il clima. Torino, Italy.

Landucci, G., Gubinelli, G., Antonioni, G., Cozzani, V., 2009. The assessment of the damage probability of storage tanks in domino events triggered by fire 41, 1206–1215. doi:10.1016/j.aap.2008.05.006.

Landucci, G., Necci, A., Antonioni, G., Argenti, F., Cozzani, V., 2017. Risk assessment of mitigated domino scenarios in process facilities. Reliab. Eng. Syst. Saf. 160, 37–53. doi:10.1016/j.ress.2016.11.023.

Landucci, G., Argenti, F., Spadoni, G., Cozzani, V., 2016. Domino effect frequency assessment : The role of safety barriers. J. Loss Prev. Process Ind. 44, 706–717. doi:10.1016/j.jlp.2016.03.006.

Lees, F.P., 1996. Loss prevention in the process industries, 2nd ed. Butterworth—Heinemann, Oxford.

Musharraf, M., Khan, F., Veitch, B., Mackinnon, S., Imtiaz, S., 2013. Human factor risk assessment during emergency condition in harsh environment, in: Proceedings of the ASME 2013 32nd International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2013), June 9–14, 2013, Nantes, France. American Society of Mechanical Engineers, New York, NY, pp. 1–9. doi:DOI: 10.1115/OMAE2013-10867.

National Fire Protection Agency (NFPA), 2018. Utility LP-gas plant code—NFPA 59.

National Fire Protection Agency (NFPA), 2017. Liquified petroleum gas code—NFPA 58.

Norwegian Meteorological Institute, 2017. www.yr.no [WWW Document].

Paltrinieri, N., Cozzani, V., Wardman, M., Dechy, N., Salzano, E., 2010. Atypical major hazard scenarios and their inclusion in risk analysis and safety assessment, in: Reliability, Risk and Safety: Back to the Future.

Paltrinieri, N., Khan, F., 2016. Dynamic Risk Analysis in the Chemical and Petroleum Industry: Evolution and Interaction with Parallel Disciplines in the Perspective of Industrial Application, 1st ed. Elsevier Science.

Paltrinieri, N., Reniers, G., 2017. Dynamic risk analysis for Seveso sites. J. Loss Prev. Process Ind. 49, 111–119. doi:10.1016/J.JLP.2017.03.023.

Pietersen, C.M., 1988. Analysis of the LPG-disaster in Mexico City. J. Hazard. Mater. 20, 85–107.

Rasmussen, K., 1996. The experience with the Major Accident Reporting System from 1984 to 1993. Luxemburg, Luxemburg.

Roberts, T., 2004a. Directed deluge system designs and determination of the effectiveness of the currently recommended minimum deluge rate for the protection of LPG tanks. J. Loss Prev. Process Ind. 17, 103–109.

Roberts, T., 2004b. Effectiveness of an enhanced deluge system to protect LPG tanks and sensitivity to blocked nozzles and delayed deluge initiation. J. Loss Prev. Process Ind. 17, 151–158.

Shaw, N., Austin, E.J., 1919. Manual of meteorology: comparative meteorology. Cambridge, UK.

Shirvill, L., 2004. Efficacy of water spray protection against propane and butane jet fires impinging on LPG storage tanks. J. Loss Prev. Process Ind. 17, 111–118.

Suedfeld, P., Steel, G.D., 2000. The environmental psychology of capsule habitats. Annu. Rev. Psychol. 51, 227–253.

weatherspark.com, 2017. www.weatherspark.com [WWW Document].

Zipf, G.K., 1949. Human Behaviour and the Principles of Least Effort. Addison-Wesley, Cambridge, MA.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Site risk analysis for nuclear installations—Nordic method developments and pilot studies

J.-E. Holmberg
Risk Pilot AB, Espoo, Finland

O. Bäckström
Lloyds Register Consulting, Sundbyberg, Sweden

E. Cederhorn & C. Sunde
Risk Pilot AB, Stockholm, Sweden

T. Tyrväinen
VTT Technical Research Centre of Finland Ltd., Espoo, Finland

ABSTRACT: Major part of the nuclear power sites house more than one reactor unit and other nuclear facilities such as spent fuel pool storage. Currently, multi-unit risks have not typically been adequately accounted for in risk assessments, since the licensing is based on unit-specific PSA with focus on a reactor accident. This paper presents an approach to site risk analysis, taking into account various dependences between the units. The dependences can be caused by external hazards, which can affect multiple units at the same time; shared operational and safety systems at the site; common staff who should manage the situations. The site risk assessment approach has been developed with aid of two pilot studies made for two Swedish sites. Preliminary results from the pilot studies are presented in the paper.

1 INTRODUCTION

After the Fukushima Daiichi accident in March 2011 general interest in site level Probabilistic Safety Assessment (PSA) has increased. Major part of the nuclear power sites house more than one reactor unit and other nuclear facilities such as spent fuel pool storage. Currently, multi-unit risks have not typically been adequately accounted for in risk assessments, since the licensing is based on unit-specific PSA with focus on a reactor accident.

The methodology for a site level risk analysis needs to consider the dependences between the units. By "unit" we mean here not only reactors but also other relevant sources for radioactive release such as spent fuel pools and storages. The dependences can be caused by external hazards, which can affect multiple units at the same time; shared operational and safety systems at the site; common staff who should manage the situations. Site risk analysis is not only a matter of extending current risk analyses to properly cover inter-unit dependences in the risk assessment, but it should also provide risk insights for the site level safety management, e.g., w.r.t., severe accident management, emergency preparedness, design, operation and maintenance of shared systems.

In 2017, the Swedish nuclear utilities Forsmark Kraftgrupp and Ringhals Ab and the Swedish Radiation Safety Authority financed together with the Finnish Nuclear Safety Research Programme SAFIR2018 a project, called SITRON (SITe Risk of Nuclear installations). The objective with SITRON is to develop methods and requirements for a nuclear power plant site risk analysis, driven by a performance of pilot studies. The paper will summarise the developed method for the site risk analysis and preliminary results from the pilot studies.

2 OVERALL APPROACH

2.1 *Definitions*

This section introduces main concepts and definitions used in this paper. They are adopted from the SITRON project reports (Holmberg 2017; Bäckström et al. 2018).

A "single-unit PSA" means a PSA made for a nuclear facility such as the reactor facility and the interim storage for spent fuel. A single-unit PSA is assumed to cover all fuel locations within the facility. For a reactor facility, the reactor and the fuel pool are the relevant locations from the risk assessment point of view. Single-unit PSA can be

also understood to refer the types of risk analyses that currently have been prepared for licensing of nuclear facilities.

A "multi-unit PSA" means a PSA or a set of PSAs made to cover accident scenarios related to all fuel locations at the site, including spent fuel transportations. Multi-unit PSA can be also understood to be an extension of a single-unit PSA which can be used to quantify multi-unit risk metrics. The aim of this paper is to discuss one approach towards this direction.

The SITRON project is limited to level 1 and level 2 PSA reflecting the current state-of-the-practice for nuclear power plant applications and licensing requirements in most countries. Level 1 PSA assesses the risk of a reactor core damage or more generally the risk of a fuel damage. The main risk metric of level 1 PSA is the Core Damage Frequency (*CDF*) or generally the Fuel Damage Frequency (*FDF*).

Level 2 PSA assesses the risk of radioactive release to the environment as a consequence of a fuel damage. In level 2 PSA, it is typical to use several risk metrics following the categorisation of releases. Internationally commonly used risk metrics are the large release frequency (*LRF*) and the large early release frequency (*LERF*) (OECD 2009). Meaning of "large" may vary depending on the regulatory framework, e.g., in Finland it is a release larger than 100 TBq of Cs-137 (STUK 2013). "Early" release means an accident where the release occur before sufficient time for offsite protective measures. As a general term for level 2 PSA risk metrics, "release category frequency" (*RCF*) is used in this paper.

The structure of PSA model can be defined to consist of a number of initiating events (IE) and plant response models that are used to quantify in level 1 the conditional probabilities of core damage (*CCDP*) with respect to each IE and in level 2 the conditional probability of certain release given the Plant Damage State (PDS). Plant damage states are interface states between level 1 and 2 PSA to facilitate more compact modelling of level 2 PSA.

When analysing and modelling multi-unit scenarios, it is necessary to extend the above concepts into those involving a single-unit impacts and those involving multi-unit impacts. For instance, initiating events are grouped into single-unit initiating events (SUIE) and multi-unit initiating events (MUIE).

Risk metrics for level 1 PSA can include the following *CDF* metrics:

- Single-Unit Core Damage Frequency (SUCDF) – frequency of a reactor accident involving core damage on one and only one reactor unit per site calendar-year.
- Multi-Unit Core Damage Frequency (MUCDF) – frequency of an accident involving core damage on two or more reactor units concurrently per site calendar-year
- Site Core Damage Frequency (SCDF) – frequency of a reactor accident involving core damage on one or more reactor units concurrently per site calendar-year.

For a level 2 PSA, the significance of the release can be characterised by two components: magnitude of released radionuclides (Cs-137 is typically used as a representative isotope) and the timing of the release. Unlike to level 1, in level 2 there is no need to count the number of units or sources that contributes to the release. Therefore, the release categorisation in multi-unit scenarios can be based on the following metrics:

- Release magnitude = sum of release magnitudes from the units having fuel accidents
- Release timing = time point when the magnitude criterion for the release category is exceeded.

The Site-level Release Category Frequency (*SRCF*) is the sum of frequencies of the single-unit and multi-unit scenarios leading to a certain release category.

It should be noted that the SITRON project has not proposed numerical criteria for release categorisation due to national differences in the risk criteria for level 2 PSA. It is, however, recommended that release magnitudes are counted in absolute units (e.g. TBq of Cs-137) rather than in relative units (e.g., x% of core inventory is released).

2.2 *Basic assumptions for site risk analysis*

A key assumption for the SITRON project is that the site risk analysis does not need to start from scratch. It is assumed that the nuclear power plants have rather complete and well-developed PSAs for the units at the site. The site risk analysis is expected to complement the existing PSA-studies by addressing multi-unit scenarios and unit dependences.

The impact of site risk analysis is two-fold. Firstly, it should lead to improved single-unit PSAs, by ensuring that multi-unit scenarios and unit dependences are properly accounted for. Secondly, site risk analysis should provide a representation of risk at the site-level, i.e., it enables the quantification of site-level risk metrics.

One important principle in the method development of the SITRON project is that it should be possible to quantify the site-level risk metrics with the single-unit PSAs, without a development of dedicated multi-unit risk models. This idea follows the assumption that approximative quantifications are sufficient for site-level PSA application purposes.

Another important principle of the site risk analysis is that effective screening will be applied to

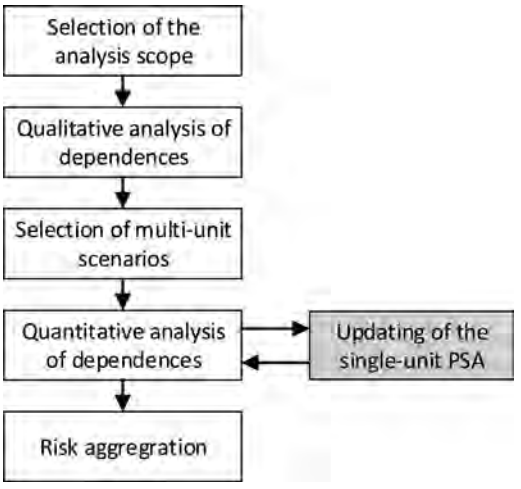


Figure 1. Site risk analysis procedure.

identify relevant multi-unit scenarios. In principle, one could postulate a huge number of multi-unit scenarios, but most of them can be shown to have an insignificant risk contribution and could be thus screened out from further analysis. For instance, it can be shown that a combination of two simultaneous independent initiating events (events occurring within a certain short time window) has an insignificant risk importance and therefore can be screened out (Bäckström et al. 2018).

2.3 *Analysis steps*

Figure 1 depicts a general procedure for a site-level PSA. As the first step, the scope of the analysis is determined, including a selection of sources of radioactive releases considered in the study and which level of PSA is considered. Section 3 of the paper will describe the elements of the qualitative analysis of dependences, which results in the selection of multi-unit scenarios for the quantitative analysis and finally the risk aggregation. These topics are discussed in Section 4.

Figure 1 also includes a grey box, "updating of the single-unit PSA", which can be seen an excursion in the procedure. This analysis step is needed to cope with identified deficiencies in the single-unit PSA. It also facilitates the quantification approach that is based on an effective utilization of single-unit PSAs.

3 QUALITATIVE ANALYSIS OF INTER-UNIT DEPENDENCES

Qualitative analysis of inter-unit dependences should be a systematic and comprehensive assessment of possible inter-unit-dependences to identify important factors in multi-unit scenarios. The purpose is to ensure that the dependences that are considered likely to be relevant are captured correctly in the quantitative analysis, but also to screen out dependences that do not require further analysis.

The analysis should cover various dependences topic by topic, as will be discussed in the following sub-sections. The identification of relevant initiating events is the basis for further analysis. The analysis of dependences due to shared systems and structures is a step that can be performed in a general manner without postulating any particular multi-unit scenario (by a multi-unit scenario we practically mean a scenario initiated by a multi-unit initiating event). The other topics (identical components, human and organisational dependences, and plant operating states) are more practical to assess when the relevant multi-unit scenarios have been selected.

3.1 *Identification of initiating events*

The primary way of identifying important multi-unit scenarios is based on the analysis of initiating events which can have multi-unit impacts. There are two types of such events: 1) a multi-unit initiating event that has more or less simultaneous impact on multiple units, and 2) a propagating initiating event, which has first impact on a single unit and then it propagates to other units.

The identification of multi-unit initiating events should be a straight-forward task since practically all external hazards form this group of events. Screening can be carried on the frequency basis, but this could have been done already in the single-unit PSAs.

The identification of relevant propagating initiating events may require further analysis compared to the analysis made in the context of the single-unit PSA. An initiating event that causes a disturbance on one unit could potentially propagate to another unit, either by creating an initiating event during the accident progression, for example loss of offsite power, or through an accident scenario (core damage) that ultimately affects the other unit. Another example of an initiating event that propagates would be a fire in one unit that spreads to another unit. An assessment of propagating hazards (fire and floodings) may require plant walk-downs to judge the likelihood for a hazard propagation.

3.2 *Common systems, buildings and structures*

There are different types of shared connections in a nuclear power plant. These connections can be categorized according to the approach used in (Muhlheim & Wood 2007), where the two main

categories "structures and facilities" and "systems and equipment" are used.

Examples of shared structures and facilities include for example service water intake structures and different types of storage tanks. There are also plant designs where, for example, turbine or auxiliary buildings are shared.

Shared of systems and equipment can be grouped into three categories

- Systems that can support several units simultaneously. Systems in this sub-category include for example station blackout gas turbines and common fire protection systems.
- Independent systems at each unit that can be cross-connected to support another unit or single systems able to fully support only one single unit at a time. Systems in this sub-category could for example include demineralized water distribution. Emergency diesel generators may also be configured to support only one unit at a time.
- Independent systems at each unit sharing standby or spare equipment. Systems in this sub-category include for example portable pumps for independent cooling.

Which connections that are shared differ widely between plants, even between plants with the same vendor. Many of the shared connections are, however, not important from a PSA point of view, e.g. shared office buildings and shared communication systems.

3.3 *Identical components*

Identical components at different units form a potential group of components which can fail due to Common Cause Failures (CCF). In (Schroer & Modarres 2013) strong evidence is presented that dependent failures occur with a relatively high frequency involving multiple units. The OECD/NEA CCF data project ICDE has also made a study on multi-unit CCF events, which indicate that such events happen (Håkansson 2017), and therefore they cannot be categorically ruled out from a multi-unit PSA. The CCF candidates are selected by studying the scenarios for the relevant multi-unit initiators.

3.4 *Human and organisational dependences*

Human and organizational dependences related to the multi-unit scenarios should be identified and covered in the human reliability analysis (HRA). In general, multi-unit HRA will need to put more emphasis on organizational and management aspects in the analysis. These factors need to be included in not only quantification, but also task analysis and modelling.

Multi-unit accidents pose additional challenges on operators which are not modelled in a single-unit PSA. These challenges may arise from constrained human resources, additional complexity in managing multiple scenarios from a common location, shared system prioritization, prioritizing the deployment of portable equipment, etc. A radioactive release from one unit in case of a multi-unit accident might affect critical operator actions that have to take place outside the main control room of another unit.

The degree of added complexity for multi-unit accidents will depend greatly upon the amount of interdependence between the individual units. This interdependence may come from the nature of the initiating event, the amount of shared systems/equipment or the amount of shared resources.

3.5 *Plant operating states*

PSA models for nuclear power plants shall take into account various Plant Operating States (POS) of the facility, since list of relevant initiating events, status of safety systems and system success criteria can vary strongly between POSs. Usual POS categorisation includes states full-power operation, reactor shutdown (from full-power to outage), reactor up-rate (from outage to full-power) and a number of POSs during the maintenance outage period.

A realistic multi-unit scenario assessment has to account for the units' various combinations of POSs. However, a complete consideration of all possible combinations of POSs between several units could lead to a large number of "site level" POSs. For instance, in a Hungarian pilot study (Bareith et al. 2016), 123 distinct site level POSs were identified for a site with four reactors and four spent fuel pools.

In SITRON, it is assumed that the need to consider various POS combinations can be considerably reduced by screening of irrelevant combinations and by merging together similar POSs. This should be true at least for various outage period POSs which have short time windows or for which the time to fuel damage is very long due to the high capacity of water pool to keep the fuel cooled even without active cooling system. In any case, the final screening of relevant POS combinations must be carried out specifically for each multi-unit initiating event.

4 QUANTIFICATION OF MULTI-UNIT SCENARIOS

4.1 *Basic approach*

The modelling and quantification approach followed in SITRON assumes that the single-unit PSA

is properly addressing the multi-unit scenarios, i.e., the impacts of dependences are modelled in such manner that the quantification of the model provides "correct" risk metrics from the single-unit point of view.

4.2 *Generation of scenarios*

Generation of scenarios means studying the Minimal Cut Set (MCS) lists for the identified multi-unit initiating events. Minimal cut set (MCS) lists are generated from the single-unit PSA. Basic events that can be associated with multi-unit dependences are identified for further quantification of dependences (see Section 4.3).

At this stage, unimportant dependences can be quantitatively screened out. One approach is to study the maximum contribution from potential multi-unit sequences for each relevant dependence (represented by selected basic events). If the sequence has a frequency below a screening criterion (e.g. 1E-8/yr) – even if full dependence is assumed—then the dependence can be screened out. Another approach is to select MCSs which are above a screening criterion and to restrict the examination of basic event dependences into that MCS list.

4.3 *Quantification of dependences*

For those dependences and associated basic events that are not screened out, a quantification of the degree of dependence must be performed. Basically, it means the evaluation of the conditional probability of an event at another unit given that a dependent event has occurred at one unit.

For multi-unit initiating events, a full dependence is assumed. For propagating events and for partial multi-unit events, case specific assessment needs to be made.

For shared systems that are common a full dependence is assumed. For shared systems, which have partly common sections and partly unit-specific sections, an assumption needs to be made which unit takes credit for the common section.

The assessment of inter-unit CCFs is crucial issue from the quantification point of view. When the units are identical, there are several important CCF groups that dominates the risk. Since typically full CCF dominates the results, a conservative approach is to assume that the components of two units form a joint CCF group. The event of interest will be a complete CCF of the full group given that a specific half of the components have failed. In SITRON, several approaches to assess the conditional CCF probabilities have been tested, e.g., using the CCF parameters of the single-unit model, ICDE operating experience data (Håkanson 2017) and generic U.S. CCF data (U.S.NRC 2016).

Post-initiating event operator actions can be divided into three groups from the dependence assessment point of view. Firstly, there are actions that can be considered unit-specific without dependences. This group mostly includes actions required in the short time window when the units need to manage the disturbance individually. Secondly, there are actions in longer term, for which partial dependence could be assumed due to shared resources. Thirdly, there are actions for which full dependence should be assumed.

4.4 *Computation of multi-unit CDF*

The general formula for two-unit CDF for a multi-unit initiating event i is

$$MUCDF_i = f_i \cdot \sum_j p(d_j) \cdot P(CD1 \mid IE_i, d_j) \cdot P(CD2 \mid IE_i, d_j), \quad (1)$$

where f_i is the initiating event frequency; $p(d_j)$ is the probability of the dependence event(s) j; and $P(CD1|\cdot)$ and $P(CD2|\cdot)$ are the conditional core damage probabilities given an initiating event i and dependence event(s) j.

There are several possibilities how to handle the dependent events. One approach used in the pilot studies was to re-quantify the joint MCS list generated from the MCS-lists of single-unit models. In this approach, a basic event, A, associated with a dependence is partitioned into a "common basic event", cA, and a "unit-specific event", iA,

$$A = cA * iA. \quad (2)$$

When the joint MCS list is created as a Boolean product of MCS lists, the dependences will be explicitly taken into account by the common basic events. The probabilities for the common respectively unit-specific basic events are obtained from the previous analysis step (Section 4.3).

5 PILOT STUDIES

5.1 *Pilot study scope*

In the SITRON project, two Swedish pilot studies are made, one for the Forsmark nuclear power station (Cederhorn et al. 2018) and second for the Ringhals nuclear power station (Bäckström et al. 2018). Forsmark pilot study is limited to reactor units 1 and 2, and the Ringhals pilot study to reactor units 3 and 4. Forsmark 1 and 2 are boiling water reactors (BWR) of Asea-Atom design and

Ringhals 3 and 4 are Pressurised Water Reactors (PWR) of Westinghouse design.

In both cases, the two units are practically identical reactors located close to each other and have several common systems and structures such as sea water intake. For both cases, there exist complete level 1 and 2 PSAs covering all initiating event categories (internal events, internal hazards, external hazards) and plant operating states (power operation, shutdown, outage, power up-rate).

In 2017, the pilot studies have included a qualitative analysis of unit dependences and a quantitative analysis of the multi-unit initiating event Loss-Of-Offsite Power (LOOP). The pilot studies were limited to level 1 PSA.

5.2 *Findings from the qualitative analyses*

In this section, a summary of findings from the qualitative analysis of the two pilot studies are presented. It can be noted that the identified dependences are very similar even though the other study concerns with two BWRs and the other with two PWRs. Therefore, the discussion given below is valid for both studies.

For initiating events, both PSA-studies include a comprehensive analysis of external hazards. The list of external hazards can be directly taken as a list of potential multi-unit initiating events, including events like loss of offsite power and organic material in sea water. Assessment of propagating initiating events was left out-of-the-scope of the pilot studies, since this task would require plant visits and walk-downs. It was however identified that there are few common buildings for which fire and flooding hazards may be considered as propagating events. Later when the pilot studies will be extended to level 2 PSA, propagating effects of severe accident situations, e.g. increased radiation level at the site, may need to be considered, too.

Both pilot cases have almost same important system and building dependences. Examples of important common systems are the offsite grid connections and sea water intake. There are also several less important common systems such as the fire water system, and the demineralized water system.

Since in both pilot studies the units at the site are identical, practically all common cause failure groups could be considered potential inter-unit CCF groups. Assessment of relevant CCF groups was limited to the example scenario, LOOP.

Both pilot studies consider a full scope of plant operating states. The average time share that the twin-units are simultaneously at-power is about 90%. Since maintenance outages are not carried out in parallel, it can be assumed that the other possible POS-combinations include one unit being at-power and the second unit being at some shutdown state.

5.3 *Results of the quantitative analyses*

In both PSAs, LOOP initiating events are divided into several sub-cases. In the pilot study, the multi-unit LOOP, leading to simultaneous loss of external grid for twin-units is considered. This initiating event has rather high risk importance in both PSA studies.

LOOP event has been considered for all POSs. When quantifying the time shares of POSs and risk importances of LOOP during various POSs, the result was that only both units being at-power is a significant POS combination. The reason for this is that other POSs are very short except one longer POS during maintenance outage during which the core/fuel damage risk is very low due to long time window to recover the situation. Also, the POSs immediately after and before at-power POS are from the PSA-modelling point of view very similar to the at-power scenarios. Same inter-unit dependences are important for those POSs as for at-power POS.

Most important minimal cuts sets for the multi-unit LOOP have been analysed qualitatively to group similar minimal cuts sets together and to characterize the cut sets from the time window and system failures point of view. There are about ten groups of minimal cut sets that dominate the result.

In almost all cases, the core damage happens due to loss of power supply to systems required for core cooling. A common feature is that house turbine operation fails after which the safety functions are dependent on Emergency Diesel Generator (EDG), Gas Turbines (GT) or mobile diesel generators (MDG). Recovery of external grid is also a possibility.

From the timing point of view, there are two main categories for the loss of 500 V AC power supply:

- Immediate loss of power supply. This is caused by various combinations of failures to start EDGs, GTs or to connect them to supply the bus bars.
- Later loss of power supply. This is caused by various combinations of failures where EDG start succeeds but stops later. From the battery capacity point of view, these minimal cut sets could be further divided into those occurring before or after the battery depletion time.

In the assessment of twin-unit CDF, the following events have large importance

- Failure of house turbine operation. House turbine operation is rather unreliable and a high probability is assumed that it is failed in both units (2×2 turbines). This event is included in all dominating MCSs
- Inter-unit CCF of batteries (two systems), which are vital for successful power supply from

EDGs. Over 80% of *MUCDF* can be eliminated if both CCFs can be eliminated. Inter-unit CCF for batteries have been assessed conservatively assuming that they form a joint CCF-group. The conditional probability for a full CCF given that half of the batteries have failed is high due to the assumed CCF-model parameters.
- Failure to recover 400 kV grid which is a common event for both units. *MUCDF* is decreased by almost 50% if LOOP is only a short-term event.
- Unavailability of gas turbine, which is a common system for both units. Gas turbine events contribute about 50% to *MUCDF*.

Regarding operator actions, multi-unit dependent actions are important only in later phase of scenarios and they do not have large risk importance.

The *MUCDF* assessment has been very simplified, and includes several uncertainties. Conditional probability of a double-unit core damage given one core damage is 0,1–0,2. The most important uncertainty is the assessment of the probability of the inter-unit CCF. If no inter-unit CCF is assumed, *MUCDF* decreases by a factor more than 100.

6 CONCLUSIONS

Qualitative analysis of multi-unit dependences is a rather straight-forward task and should be included already in the single-unit PSA. For instance, multi-unit IEs can be rather easily identified since they are practically equal to the list of external hazards.

In the first hand, it can be assumed that external hazards are complete multi-unit IEs. To judge whether hazards should be considered partial multi-unit IEs may require considerable more effort, plant visits and statistical analyses or expert judgements. The same applies for the identification of propagating IEs for whose relevance from multi-unit risk point of view cannot be judged without plant visits.

Identification of common systems and structures is also a straight-forward task and should have been already considered in the single-unit PSA. Relevant operator action dependences can be assumed to exist mainly in long term action since in the beginning of scenarios, the units are designed to manage the emergency situations independently.

Relevant multi-unit dependences and scenarios are related to events impacting safety functions core cooling and residual heat removal. From the initiating event point of view, the common disturbances can be classified in the general groups 1) loss of power supply or 2) loss of ultimate heat sink. Some external hazards can cause both plant impacts.

From the POS combinations point of view, it is likely that the only POS combination which needs to be considered is both units being at-power. The other combinations can be screened out either by their very short time duration or by the very long recovery times for which reasons such events have negligible contribution to the multi-unit risk. This conclusion is based on the assessment of the multi-unit LOOP initiating event.

The assessment of inter-unit CCF is crucial issue from the quantification point of view. When the units are identical, there are several important CCF groups that dominates the risk. The applied quantification principle in the pilot study is presumably very conservative, which can lead to high conditional probability for a multi-unit core damage given a single-unit core damage.

The pilot study case, loss of offsite power (both 400 kV and 70 kV) did not reveal any such dependences that would suggest revisions in the single-unit PSA. There are few dependent operator actions which have some importance, but they are treated in the current PSA quite simplified manner. The functional role of gas turbine in the multi-unit scenario might also require some more detailed analysis.

From the *MUCDF* assessment point, a rather simple quantification can be performed using dominating minimal cut sets and basic events. Two quantification approaches have been tried out in pilot studies, both providing practical approaches to assess the multi-unit risk metrics and risk importances of various items of the model.

It should be noted that the conclusions made here are based on a single scenario and on level 1 PSA. It could be expected that some weather-related hazards impacting the sea water intake and the extension of the analysis to level 2 may bring up further issues, e.g., the role of dependent operator actions may be more important.

ACKNOWLEDGEMENTS

The work has been financed by SAFIR2018 (The Finnish Research Programme on Nuclear Power Plant Safety 2015–2018), Forsmark Kraftgrupp AB, Ringhals AB and Swedish Radiation Safety Authority (SSM).

REFERENCES

Bäckström, O., et al. 2018. SITRON—WP2 — Method development. Site level risk assessment. Report 212634-R-001, Lloyd's Register Consulting, Sundbyberg.

Bäckström, O., Häggström, A., He, X. 2018. SITRON—Pilot study Ringhals 3&4, Report 212634-R-002, Lloyd's Register Consulting, Sundbyberg.

Bareith, A., Hollo, D., Karsa, Z., Siklossy, P., Siklossy, T. A pilot study on developing a site risk model. In Proc. of 13th International Conference on Probabilistic Safety Assessment and Management (PSAM 13), 2–7 October, 2016, Seoul, Korea. Paper A-420.

Cederhorn, E., Holmberg, J.-E. Sunde, C. 2018. SITRON—Pilot study Forsmark 1&2. Report 14124_R007, Risk Pilot Ab, Espoo.

Håkansson, M. 2017. Action item 43-09 (42-05, 41-04, 40-16): Summary of workshops on Multi-unit events. ICDE Work note. Draft 2017-03-10 (limited distribution).

Holmberg, J.-E. 2017. SITRON—Risk metrics. Report 14124-R005, Risk Pilot AB, Espoo.

Muhlheim, M.D., Wood R.T. 2007. Design Strategies and Evaluation for Sharing Systems at Multi Unit Plants Phase I. ORNL/LTR/INERI-BRAZIL/06-01, Oak Ridge National Laboratory.

OECD/NEA. 2009. Probabilistic Risk Criteria and Safety Goals, NEA/CSNI/R(2009)16, OECD/NEA, Paris.

Schroer, S., Modarres, M. 2013. An Event Classification Schema for Evaluating Site Risk in a Multi-Unit Nuclear Power Plant Probabilistic Risk Assessment, *Reliability Engineering and System Safety* 117, 40–51.

STUK. 2013. Probabilistic risk assessment and risk management of a nuclear power plant, Guide YVL A.7, Radiation and Nuclear Safety Authority in Finland, Helsinki.

U.S.NRC. 2016. CCF Parameter Estimations, 2015 Update.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Risk-informed safety classification of components of auxiliary systems for emergency diesel generators in nuclear power plants

J.-E. Holmberg
Risk Pilot AB, Espoo, Finland

ABSTRACT: Safety classification of structures, systems and components of nuclear power plants shall be based on the functional importance of the items. One challenge with the nuclear safety classification is that different classification systems are used in different countries and standards. Even if certain classification scheme can be agreed upon, it is not straight-forward how the classification should be carried out for various components, e.g., in electric and automation systems. At higher plant and system level, the safety importance of functions can be defined directly, but the classification of smaller components requires further assessments. In principle, a component's safety class follows its functional importance. Downgrading is possible if mitigative factors and reliability arguments can be shown. The importance of finding correct safety class is that it determines the QA requirements for the component. The paper will outline a risk-informed safety classification approach for the components, based on both probabilistic and deterministic assessments. Emergency diesel generator system will be used as an example.

1 INTRODUCTION

Safety classification of Structures, Systems and Components (SSC) of nuclear power plants (NPP) shall be based on the functional importance of the items. One challenge with nuclear safety classification is that different classifications are used in different countries and standards. International standards and guidelines give some common guidance to the classification, but in practice the licensees and vendors need to adapt the classification into the national system (WNA 2015). Even if certain classification scheme can be agreed upon, it is not straight-forward how the classification should be carried out for various components.

At higher plant and system level, the safety importance of functions can be defined directly, but the classification of smaller components requires further assessments. In principle, a component inherits its safety class from its functional role based on the functional impact of its failure. This principle can be considered a deterministic approach to safety classification since the functional importance is determined by the deterministic safety analysis and the defence-in-depth concept of the design.

The strength with the deterministic approach for the safety classification is the link with the strong safety design principles such as defence-in-depth, safety margin, redundancy, diversity and independence (Ahn et al. 2010). This approach has limitations since it does not systematically and explicitly consider the risk importance of items, which could be assessed by Probabilistic Safety Assessment (PSA). Deterministic approach generally only considers worst case scenarios (Kirschsteiger 1999).

The importance of finding correct safety class is that it determines the Quality Assurance (QA) requirements for the component. Too low safety class may imply a system reliability concern; too high safety class can be a significant cost factor, and it can be even a problem of finding a component supplier for nuclear market, which has specific QA requirements.

Since the development of PSA methods and applications for NPP safety management, there have been attempts to implement risk-informed approach to safety classification. A well-known and applied approach is the U.S.NRC (2011) guide to risk-informed decision making. Shortly, the PSA is used to determine the risk importance of a component and the risk importances are compared to the deterministic safety class. Risk importance measures such as Fussell-Vesely and Risk Achievement Worth are used for this purpose. Such studies and discussions can be found, e.g., in (Jänkälä 2002; Holmberg & Männistö 2008).

Safety Integrity Levels (SIL) introduced in IEC-61508 (IEC 2010) and related branch-specific standards are also examples of risk-informed safety classification. The reasoning behind SIL is, however, quite different from the nuclear safety classification principles for which reason it would

be difficult to apply it as such in nuclear context, though associations can be made between SIL and nuclear safety classes, see e.g. Annex D of IEC 61513 (IEC 2011).

This paper discusses an approach to the risk-informed safety classification. The primary role of the deterministic safety classification is acknowledged, especially the assessment of the functional importance of items. The issue to be resolved is how the safety classification of components can be reassessed based on their risk importance. Electric and automation systems are the main intended application area since these systems consist of large number of components with different functional role, failure modes and risk importance. Section 2 describes the safety classification system. Section 3 presents the emergency diesel generator (EDG) system used as an example. Section 4 outlines the safety classification approach, and Section 5 concludes the paper.

2 SAFETY CLASSIFICATION SCHEME

2.1 *Defence-in-depth and plant condition categories*

The safety philosophy of NPPs builds on the Defence-in-Depth (DiD) principle. Here, we consider DiD as successive levels of protection, which rely on the application of safety principles of multiple barriers, physical separation, redundancy and diversity. The standard nuclear five level classification is taken as the basis (IAEA 1996). DiD levels can be mapped one-to-one with "plant condition categories" or "design basis categories", which are fundamental elements of deterministic safety analyses, see Table 1.

The relationship between defence-in-depth and safety classification is immediate in the sense that each DiD level is assigned to certain safety class, and all items belonging to one level have the same safety class. It follows that all items within one level have the same requirements for design, qualification, regulatory review and QA procedures during all life cycle phases. Different DiD levels can belong to different safety class, and this is considered beneficial both from the diversity point of view and from the optimal resource allocation point of view.

2.2 *Safety classification systems in nuclear field*

In nuclear field, there is both national and international variations in the safety classification systems, though all systems have a link to the DiD principle. For example, the International Electrotechnical Commission categorization (IEC 2009) defines three safety categories A, B and C, whereas the American standards of Institute of Electrical and Electronics Engineers uses a classification that only distinguishes between safety and non-safety systems (IEEE 2003).

Table 1. Defence-in-depth levels and plant condition categories.

DiD level	Objective	Plant condition category
1	Prevention of abnormal operation and failures	Normal operation
2	Control of abnormal operation and failures	Anticipated operational occurrences (AOO)
3a	Control of accident to limit radiological releases and prevent escalation to core melt conditions	Design basis accidents (DBA)
3b*		Design extension conditions (DEC)**
4	Control of accidents with core melt to limit off-site releases	Postulated core melt accidents or severe accidents (SA)
5	Mitigation of radiological consequences of significant releases of radioactive material	—

*3b controls postulated common cause failures in level 3a; **DEC has three subcategories (STUK 2013a): a) AOO/DBA & common cause failure in level 3a, b) significant events identified in PSA and c) rare external events.

The International Atomic Energy Agency (IAEA) has adopted the following three-level safety category system (IAEA 2016):

- Safety category 1: Any function that is required to reach the controlled state after AOO or DBA and whose failure, when challenged, would result in consequences of 'high' severity.
- Safety category 2: Any function that is required to reach a controlled state after AOO or DBA and whose failure, when challenged, would result in consequences of 'medium' severity.
- Safety category 3: Any function that is actuated in the event of AOO or DBA and whose failure, when challenged, would result in consequences of 'low' severity. Any function that is designed to reduce the actuation frequency of the reactor trip or engineered safety features in the event of a deviation from normal operation, including those designed to maintain the main plant parameters within the normal range of operation of the plant.

In addition to the overall safety classification systems, there are more technically-oriented clas-

sification standards, such as IEC 60709 for the physical separation of I&C equipment (IEC 2004) and IEEE 384 for independence requirements of electric circuits and equipment (IEEE 2008). In this respect, the treatment of electric and I&C equipment is quite straight-forward. With regard to mechanical equipment in electric and I&C systems, there are not so strict requirements, which leaves room for an interpretation.

2.3 *Risk-informed safety classification system*

In this paper, a risk-informed safety classification system is outlined, in which the system functions and associated components are first classified based on their functional importance. Refined assessment of the component safety classes is made based on the assessment of mitigative factors and reliability considerations. The procedure is illustrated in Figure 1.

In the first step, the functions, sub-function, sub-sub-functions, etc. of the system are defined down to a sufficient level of details so that the functional importance of each component can be defined. Here "component" is associated with the spare parts level itemization of a system, which is the level of details that needs to be achieved in the safety classification from the QA requirements point of view.

Functional importance of components can be analysed, e.g., using Failure Modes, and Effects Analysis (FMEA). The functional impact of the component's failure mode determines the preliminary safety class.

In the final step, the component's safety class can be reassessed based on two more criteria: the reliability of the component and the existence of mitigative factors. Low failure probability and mitigative factors can be used as arguments to downgrade the safety class (see Section 4 for details).

A three-level safety class system is proposed. Three classes are considered practical for the categorisation of components and yet sufficient to be acceptable in various national regulatory frameworks.

Safety Class 1 (SC1) represents the highest safety category. It is assigned to functions, which are required to cope with DBAs, i.e., they belong to DiD-level 3, which is the most important from safety point of view and has thus highest QA requirements.

Safety Class 2 (SC2) is assigned to other safety related functions than those critical in DBA scenarios. This includes, e.g., functions related to DiD levels 2, 3b and 4. In addition, SC2 can be assigned to manual back-up of SC1 functions, monitoring and surveillance of SC1 functions and functions whose failure can impact the long-term reliability of SC1 functions.

Safety Class 3 (SC3) is assigned to non-safety related functions. This includes, e.g., functions related to DiD level 1.

Compared to the safety classification systems of IEC and IAEA, the proposed classification system has one class less. Practically, SC1 corresponds with the highest safety category of the nuclear classification systems, SC2 corresponds with other safety related classes (e.g. Cat. B and C in IEC 61226). The reason to merge the other safety related classes into a single class is that it is not practical to have too many safety classes for the categorisation of components.

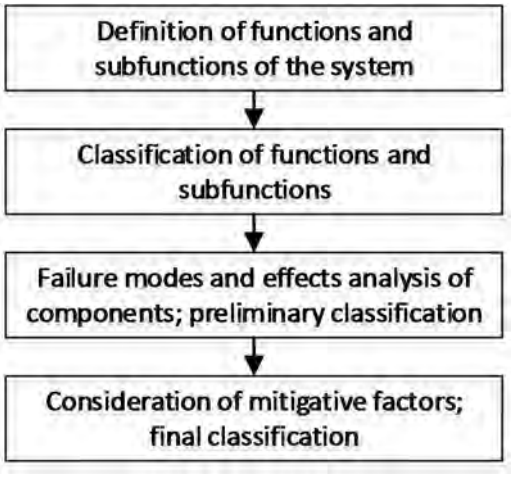


Figure 1. Procedure for risk-informed safety classification.

3 FUNCTIONAL DESCRIPTION OF EMERGENCY DIESEL GENERATORS

3.1 *Main functions and subsystems*

In this paper, Emergency Diesel Generators (EDG) will be used as an example. The main safety function of EDGs for nuclear power plants is to provide power supply to safety-critical electric bus bars in case of Loss Of Offsite Power (LOOP) initiating event. Typically, there is one EDG per one safety train or per two safety trains, i.e., there are two to four EDGs per reactor providing emergency power supply. In this paper, "EDG system" refers to the complete set of redundant EDGs (2 to 4 EDGs), and "EDG" refers to a single-train EDG.

For new designs, the requirement is to have a diverse back-up for the EDGs to cope with the DEC scenario where LOOP occurs in combination with a CCF of EDGs. For this purpose,

the plants may have diverse DGs, called Station Black-Out (SBO) DGs, gas turbine or/and mobile DGs. Even cross-connections between reactor units may be a solution, though this is not allowed in all countries.

To analyse EDG from the safety and reliability point of view, it needs to be broken down into subsystems and functions as well as a system boundary must be defined. In practical EDG applications, system boundaries vary, but here we divide the whole system into three major parts. The EDG generates the electric power from diesel fuel, and include subsystems, such as generator, diesel engine, fuel oil system, cooling system, lubrication system, starting air system, combustion air system, exhaust system, and I&C system. Important auxiliary or support systems of EDG include systems such as fuel oil storage and supply, cooling water system, cooling air and ventilation system for the engine room, and electric power system for the control and protection system. In addition, the EDG function needs electric power transmission lines and control logic to the bus bars dependent on the EDG.

3.2 *Functional classification*

For the sake of simplicity, we consider two types of LOOP events:

- Design Basis Accident (DBA) LOOP for which case the fuel stored in the day tank is sufficient. This is called short-term LOOP and the corresponding EDG function belongs to the DiD level 3a, receiving SC1.
- Design extension condition (DEC) LOOP for which cases the fuel oil from the storage tank is also needed. This is called long-term LOOP and the corresponding EDG function belongs to the DiD level 3b, receiving SC2.

Sub-functions of EDG can be classified depending on their criticality to the above main safety functions. As an example, the sub-functions of the fuel oil subsystem are provided in Table 2. Figure 2 depicts a simplified flow diagram. The fuel oil subsystem is responsible for the feeding of the diesel engine with fuel oil. Feeding is arranged from the day tank, which can be loaded from a larger storage.

The fuel oil subsystem has a functional safety class 1 since it is necessary for the DBA safety function. Most subsubsystems of the fuel oil subsystem have the same safety class. Oil storage and transfer is not needed during the DBA case but is needed in the DEC case. Therefore, it belongs to SC2. Fuel unloading from the oil storage belongs to SC3 since it is a maintenance action with no safety relevance.

In the functional analysis, the sub-functions are broken down into a level that facilitates the classification of components. In Table 2, the sub-function 4 "Fuel oil feeding and circulation" has been broken down into seven sub-sub-functions. The next step is to perform an FMEA to assign preliminary safety classes for the components. Table 3 shows an example for some components contained in the

Table 2. Functional classification of the fuel oil system.

Sub-function	Class
Fuel oil system	1
1. Fuel unloading	3
2. Oil storage and transfer	2
3. Fuel oil day tank	1
4. Fuel oil feeding and circulation	1
4.1 Fuel feeding	1
4.2 Fuel oil impurity removal (filters)	2
4.3 Fuel injection to engine	1
4.4 Leak fuel handling	1
4.5 Emergency cut-off	2
4.6 Overpressure protection	2
4.7 Drain	3
5. Fuel oil cooling	1
6. Leak fuel handling	1

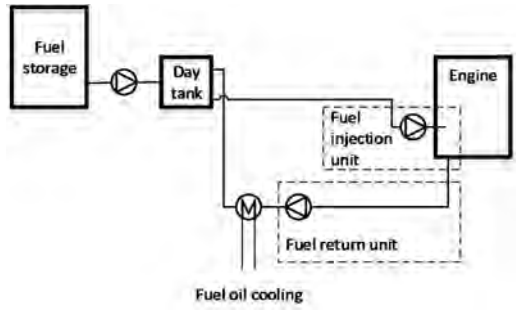


Figure 2. Simplified flow diagram of the fuel oil system for EDG.

Table 3. Failure modes and effects analysis example. Columns F = Function, C = Failure cause and D = Failure detection are left undeveloped in the example.

Component	F	Failure mode	C	D	Impacted function	Prel. class
Pipe x	.	Rupture	.	.	4.1	1
		Leakage	.	.	4.1	1
		Blocking	.	.	4.1	1
Manual shut-off valve y	.	Wrong position	.	.	4.1	1
Pump z	.	Failure to start	.	.	4.1	1
		Spurious stop	.	.	4.1	1
Strainer s	.	Clogging	.	.	4.1	1
		Rupture	.	.	4.1	1

subsubsystem "Fuel oil feeding and circulation". The primary classification is derived from the safety class of the impacted function.

4 RISK-INFORMED SAFETY CLASSIFICATION

The risk-informed safety classification account for two criteria in addition to the functional importance of the component. These are the reliability of the components and the existence of mitigative factors.

Reliability of a component is principally measured by the frequency or probability of the critical failure modes. The role of the component reliability assessment is two-fold: 1) to demonstrate the fulfilment of the system reliability target, 2) to justify down-grading for components whose unavailability can be demonstrated to be insignificant.

Mitigative factor is a feature of the system or component design which can eliminate or mitigate the impact of the component failure. Mitigative factor can be also seen means to improve the system reliability. Effectiveness of mitigative factors can be thus measured probabilistically.

Figure 3 depicts a principal scheme for the reclassification. It should be noted that this scheme is only applied to SC1 and SC2 components. For SC3 components, there is no need to consider reclassification from safety point of view. In the following subsections, a further interpretation for the boxes of the scheme will be provided.

4.1 *Safety goal based reliability targets*

4.1.1 *System reliability target*

In the risk-informed safety classification, the functionally derived classification of components can be revised based on the risk importance of the component. The idea is *not* to exactly quantify the risk importances, but to use probabilistic reasoning in an indicative manner to define reliability levels that can be used to define rules for the reassessment. Further, the derived reliability targets for equipment can be used in the argumentation on reasonable level of required reliability. In addition, the discussion can be used to support the assessment of effectiveness of possible mitigative factors and to derive an interpretation for negligible risk contributors.

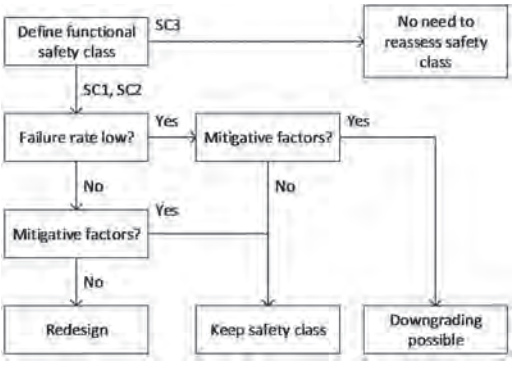


Figure 3. Principal scheme for final classification of components.

We assume certain generic risk criteria and design features of a nuclear power plant to obtain system reliability targets. In a specific project, these assessments need to be re-evaluated. Typical risk criteria for are (OECD/NEA 2009) 1E-5/yr for the Core Damage Frequency (CDF) and 1E-6/yr for the Large Release Frequency (LRF).

EDG-failure related accident sequences may not contribute more than 1% to the CDF criterion (1E-7/yr). This is a hard requirement, but station black-out sequences have also a high potential to a large release (LRF criterion).

The frequency of LOOP is typically of order 1E-1/yr (Johnson & Schroeder 2016) but most LOOP events are very short. The frequency for LOOP when recovery of offsite power cannot be credited is assumed to be about 1E-2/yr.

The plant has a diverse back-up for EDGs in case of station black-out. The unavailability of the diverse back-up is 0.1. This is a conservative value. The above probability numbers yield a failure probability target for the EDG-system

$$U_1(\text{system}) = \frac{CDF_{\text{target}}}{f(\text{LOOP}) \cdot p(\text{EDG back-up})} = \frac{1\text{E-}7/\text{yr}}{1\text{E-}2/\text{yr} \cdot 0.1} = 1\text{E-}4. \quad (1)$$

This target is for a set of redundant EDGs. Typically, an NPP has 2 to 4 EDGs. To derive a target for one EDG, the probability of Common Cause Failure (CCF) must be estimated. Probability of CCF is dependent on many factors, e.g., number of EDGs and the testing scheme of EDGs. Using CCF-parameters of EDGs estimated from data from US NPPs (U.S.NRC 2016), it can be estimated that the fraction of total CCF for two-redundant respectively four-redundant systems is about (order of magnitude)

$$\text{CCF2/2} \approx 5\% \quad (2)$$

$$\text{CCF4/4} \approx 0.5\% \quad (3)$$

Assuming a four-redundant system, an unavailability target for one EDG (out-of-four components) can be defined as follows

$$U_1(1\text{oo}4\ \text{EDG}) = \frac{U_1(\text{system})}{\text{CCF4/4}} = \frac{1\text{E-4}}{0.005} = 2\text{E-2}. \quad (4)$$

This is a first estimate for a single-train EDG reliability target, and it will be compared to experience based values for EDGs in the next subsection to define final choices for a reasonable reliability target.

4.1.2 *Experience based reliability of EDG*

As a reference for the reliability targets for EDGs, experience based reliability of EDGs is examined. The estimates will be made using T-book (TUD Office 2015) and U.S. NPP EDG reliability data (U.S.NRC 2015). Repair and maintenance time unavailabilities are omitted in the estimation.

Using the parametrization of T-book, the unavailability of EDG can be expressed as follows

$$U \approx q + \tfrac{1}{2}\lambda_s \cdot TI + \lambda_d \cdot TM, \quad (5)$$

where q = constant unavailability (failure to start); λ_s = standby failure rate (failure to start); λ_d = mission time failure rate (spurious stop); TI = test interval; and TM = mission time.

In T-book version 8, the generic parameters for the failure rates and probabilities are q = 3.9E-4; λ_s = 4.0E-3/h and λ_d = 3.9E-2/h. Assuming a test interval TI = 672 h and a mission time TM = 8 h, the mean unavailability is

$$U_{\text{T-book}}(1\text{EDG 8 h}) = 1.8\text{E-2}. \quad (6)$$

In the U.S. NPP reliability database, the parametrization is slightly different, as follows

$$U \approx q + \lambda_1 \cdot 1\text{ h} + \lambda_2 \cdot (TM - 1\text{ h}), \quad (7)$$

where q = probability of failure to start; λ_1, = first hour (load run) failure rate; and λ_2 = mission time failure rate after first hour. The parameter values for EDG are q = 2.88E-3 (table EDG-FTS); λ_1 = 3.72E-3/h (table EDG-FTLR); λ_2 = 1.52E-3/h (table EDG-FTR). These parameter values yield

$$U_{\text{US}}(1\text{EDG 8 h}) = 1.7\text{E-2}, \quad (8)$$

which is remarkably close to the T-book's estimate.

One should note that the EDG system boundaries used in T-book and US data may vary. In addition, the EDG system boundary applied by the above references do not include all support systems and connected systems, which are also critical to the overall EDG function. The conclusion is thus that the experience based unavailability figure of EDG is higher than the target derived in Section 4.1.1 (formula (4)). A difference is that the reliability based values reflects old systems while the target value derived in Section 4.1.1. is suited for new systems. Therefore, this target value is chosen for further development of the component reliability targets, i.e.,

$$U_1(\ 1\text{EDG}) = 2\text{E-2}. \quad (9)$$

If this value can be demonstrated for a single-train EDG including auxiliaries, the system should be at least as good as current EDGs.

In case of the DEC target (SC2 function), the mission time is longer and the fuel storage and transfer to day tank are included in the consideration. If the mission time is changed to 72 h (as an example), the following experience based unavailabilities are obtained (excluding auxiliaries)

$$U_{\text{T-book}}(1\text{EDG 72 h}) = 1.2\text{E-1}, \quad (10)$$

$$U_{US}(1\text{EDG 72 h}) = 1.1\text{E-1}. \quad (11)$$

Considering possible trending and missing auxiliaries in the above estimates, one could nevertheless define the following design extension condition reference value for the EDG

$$U_2(1\text{EDG}) = 1\text{E-1}. \quad (12)$$

The above unavailability targets may seem to be rather high, but it should be noted that EDGs are backed-up by diverse power supply in all new NPPs and in all modernised NPPs, e.g., by SBO DGs, mobile DG, unit cross-connections or by gas turbine. Therefore, from risk point of view, there is usually no need to demonstrate better reliability for EDGs.

4.1.3 *Derivation of reliability targets for EDG items*

The next step is to derive reliability targets for the items of the EDG. This cannot be done straightforwardly, since it depends on the way EDG is decomposed into items and the way reliability targets should be distributed between the items. For some items, a higher unavailability can be allowed if, at the same time, others can be shown to be very reliable. In any case, a fault tree analysis should be used to demonstrate that the overall system reliability target is achieved.

It is assumed that the items form a serial system so that the EDG unreliability is sum of items' unavailabilities. The impact of redundancy and other mitigative factors will be considered separately.

Using a kind of ALARP-approach (As Low As Reasonably Practicable), a limit and a target value

is defined for the reliability. The limit value must be achieved, and the target value is a reference for the interpretation of a negligible contribution.

We assume that there are, from the reliability point of view, three groups of components: a) unavailability is close to limit value, u^a, b) unavailability is clearly below the limit value but not negligible, u^b, and c) unavailability is negligible, u^c. The relationship between the unavailability values u^a, u^b, and u^c can be defined as follows

$$u^a = 10u^b = 100u^c. \tag{13}$$

The total unavailability will be then

$$U \approx n_a u^a + n_b u^b + n_c u^c \approx n_a u^a + n_b u^b, \tag{14}$$

where n_a = number of components having unavailability close to the limit u^a; n_b = number of components having unavailability well below the limit; and n_c = number of components having negligible unavailability, i.e., the term $n_c u^c$ is insignificant.

Next, we assume that the number of components that have unavailability close to the limit value, n_a, is much smaller than the number of component that have lower unavailability, n_b. We can thus allocate the system unavailability target about evenly between these two groups, i.e., 50% of the target value to a-components and 50% for b-components. Further, as an order of magnitude estimate, we may say that there are about 100 components per single EDG (in the spare parts level of itemization) in each safety-critical class, SC1 and SC2. n_a could be about 10 and n_b about 100. This reasoning leads to a relative item-level unavailability limit 5%, which corresponds with the following absolute unavailability limits

$$u_1^*(\text{EDG-item}) = 1\text{E-}3, \tag{15}$$

$$u_2^*(\text{EDG-item}) = 5\text{E-}3. \tag{16}$$

Consistently, the target value based on formula (13) should be a factor 100 lower,

$$u_1^0(\text{EDG-item}) = 1\text{E-}5, \tag{17}$$

$$u_2^0(\text{EDG-item}) = 5\text{E-}5. \tag{18}$$

The unavailability limits/targets include the following unavailability contributions of an item,

- latent failures occurring during the standby period or in connection to the previous test, maintenance or operation moment. This can be split into a time-independent part and a time-dependent part, c.f., formula (5),
- mission time failures.

The maintenance and repair related unavailability contribution can be omitted in this context since it is stipulated by Safety Technical Specifications.

Since for almost all items, the failure modes can be related either to the system standby time or to the operational time, there is no need to further split the item-specific reliability limits/targets into latent respectively mission time reliability targets.

Table 4 provides item-specific limits and targets. These numbers have been derived using specific

Table 4. Indicative unavailability limits and targets for items of a serial system with reliability targets 2E?2 (SC1) and 1E-1 (SC2). Assumed number of items per function ~100.

Failure mode time dependency	SC1 function		SC2 function	
	Limit	Target	Limit	Target
Latent, time-independent failure probability, q	1E-3	1E-5	5E-3	5E-5
Latent, time-dependent failure rate, λ_s*	1E-6/h	1E-8/h	5E-6/h	5E-8/h
Mission time failure rate, λ_d**	1E-4/h	1E-6/h	1E-4/h	1E-6/h

*1-month test interval assumed;
**8 h/72h mission time assumed for SC1/SC2 functions.

Table 5. Classification of mitigative measures.

Mitigative measure	Description, examples	Mitigation factor
Practical elimination of the failure mode	Inherent feature that eliminates the failure mode	~10^{-4}
Reliable elimination of the failure mode	Fail-safe behaviour, e.g., instrumentation failure causes an actuation or it does not stop the function if already actuated	~10^{-3}
	Diverse back-up, automated function, very reliable switch function, negligible possibility for CCF	
Redundancy	Duplication with an identical item, automatic reliable switch function, small (but non-negligible) possibility for CCF	~10^{-2}
Manual back-up, manual recovery	Back-up function exists, but it is not automated. Success conditions for the action exist. Human reliability analysis is needed to verify this.	~10^{-1}

Table 6. Reconsideration of component's safety class accounting for reliability and mitigative factors.

Mitigative measure	Component unavailability u_i		
	$u_i >$ limit	limit $> u_i >$ target	target $> u_i$
Practical elimination of the failure mode	safety class can be reduced		
Reliable elimination of the failure mode	keep safety class*	safety class can be reduced	
Redundancy	keep safety class	keep safety class*	safety class can be reduced
Manual back-up, manual recovery	keep safety class		keep safety class *
None	redesign	keep safety class	

* For SC2 components, the option is "safety class can be reduced".

assumptions about the system reliability target and numbers of items between which the reliability target must be allocated. Proposed numbers should be considered indicative values to support qualitative argumentation in the reassessment of the classification.

4.2 *Assessment of mitigative measures*

The reliability of equipment can be improved by various mitigative measures. These are classified in Table 5 regarding their effectiveness. The mitigation factors given in the last column of the table should be regarded as indicative probability numbers, with some correspondence with typical failure probability numbers used in risk and reliability studies for technical systems.

Combining the original component's unavailability with additional mitigative factors, the functional safety class of an item can be reconsidered. A proposal for such an approach is outlined in Table 6.

5 CONCLUSIONS

Safety classification of structures, systems and components has an important governing role for the definition of QA requirements for various items. Safety classification is fundamentally based on deterministic safety analysis thinking where an item's functional importance determines the safety class. At higher plant and system level, the safety importance of functions can be defined straightforwardly, but the classification of components requires further assessments. This is especially true for electric and automation systems, which consist of a large range of components whose importance can vary a lot.

The paper presents a risk-informed approach to safety classification where the safety classification is carried out in two phases: 1) deterministic, functional classification, 2) re-assessment considering of mitigative and reliability factors of the item. In the functional classification, the system functions are broken down into sub-functions and sub-sub-function until a level of details is reached so that the component's functional importance can be determined. FMEA is considered a practical tool to assess the functional importance which is determined by the functional impact of the component failure.

A challenge with the nuclear safety classification is that different classifications are used in different countries and standards. Besides the international standards organizations, almost every nuclear safety authority has local requirements. Thus, there can be inconsistencies between international and national codes and standards, which is problematic when local regulation must be combined with standards and applicable in vendor home countries.

The paper suggests a three-level safety classification system, which is considered sufficient for component level classification and yet applicable with respect to various national and international systems. Highest safety class, SC1, is applied functions belonging to DiD level 3. Other safety-related functions not critical to DiD level 3 functions are assigned to the second safety class, SC2. Non-safety-related functions are assigned to SC3.

Reassessment of component safety classes can be based on argumentation on the reliability of the components and existence of mitigative factors. The paper outlines an approach to judge which levels of reliability together with mitigative factors can justify downgrading. The basic idea is to control that the overall reliability target for the system will be reached, which also provides a reference for the judgment of unavailability contributions that are negligible. In any case, a fault tree analysis or equivalent quantitative system reliability assessment need to be performed to verify the fulfilment of the system reliability targets.

REFERENCES

Ahn, S.A., Kim, I.S. & Oh, K.M. 2010. Deterministic and risk-informed approaches for safety analysis of

advanced reactors: Part I, deterministic approaches. *Reliability Engineering and System Safety* 95, 451–458.
Holmberg, J.-E. & Männistö, I. 2008. Risk-informed classification of systems, structures and components. *Rakenteiden mekaniikka* (*Journal of Structural Mechanics*) Vol. 41, No 2, 90–98.
IAEA. 1996. Defence-in-depth in nuclear safety. INSAG–10. Vienna: International Atomic Energy Agency.
IAEA. 2016a. Safety Classification of Structures, Systems and Components in Nuclear Power Plants, Specific Safety Guide No. SSG-30. Vienna: International Atomic Energy Agency.
IEC. 2004. Nuclear power plants—Instrumentation and control systems important to safety—Separation, IEC 60709, Rev. 2.0. Geneva: International electrotechnical commission.
IEC. 2009. Nuclear power plants—Instrumentation and control important to safety—Classification of instrumentation and control functions, IEC 61226, ed. 3.0. Geneva: International electrotechnical commission.
IEC. 2010. Functional safety of electrical / electronic / programmable electronic safety-related systems (E/E/PES), IEC 61508, Rev. 2.0. Geneva: International electrotechnical commission.
IEC. 2011. Nuclear power plants—Instrumentation and control important to safety—General requirements for systems, IEC 61513, ed. 2.0. Geneva: International electrotechnical commission.
IEEE. 2003. IEEE Standard for Qualifying Class 1E Equipment for Nuclear Power Generating Stations, IEEE Std. 323–2003, Institute of Electrical and Electronics Engineers.
IEEE. 2008. IEEE Standard for Independence of Class 1E Equipment and Circuits, IEEE Std. 384–2008, Institute of Electrical and Electronics Engineers.
Jänkälä, K. 2002. A risk informed safety classification for a Nordic NPP, NKS-72, Nordic nuclear safety research, Roskilde.
Johnson, N. & Schroeder, J.A. 2016. Analysis of Loss-of-Offsite-Power Events 1987–2015, INL/EXT-16–39575. Idaho Falls: Idaho National Laboratory.
Kirchsteiger C. 1999. On the use of probabilistic and deterministic methods in risk analysis. *Journal of Loss Prevention in the Process Industries* 12, 399–419.
OECD/NEA. 2009. Probabilistic Risk Criteria and Safety Goals, NEA/CSNI/R(2009)16, OECD/NEA, Paris.
STUK. 2013a. Safety design of a nuclear power plant, Guide YVL B.1. Helsinki: Radiation and Nuclear Safety Authority.
STUK. 2013b. Classification of systems, structures and components of a nuclear facility, Guide YVL B.2. Helsinki: Radiation and Nuclear Safety Authority.
TUD Office. 2015. T-Book. Reliability Data of Components in Nordic Nuclear Power Plants, 8th edition. Stockholm: The TUD Office.
U.S.NRC. 2011. An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the licensing Basis, Regulatory Guide 1.174, Rev. 2. Washington, D.C.: United States Nuclear Regulatory Commission.
U.S.NRC. 2015. Summary of SPAR Component Unreliability Data and Results. 2015 Parameter Estimation Update. http://nrcoe.inel.gov/resultsdb/AvgPerf/
U.S.NRC. 2016. CCF Parameter Estimations, 2015 Update. Washington, D.C.: United States Nuclear Regulatory Commission.
WNA. 2015. Safety Classification for I&C Systems in Nuclear Power Plants—Current Status & Difficulties. CORDEL Digital Instrumentation & Control Task Force. London: World Nuclear Association.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Development of a qualitative framework for analysing high-impact low-probability events in power systems

I.B. Sperstad
SINTEF Energy Research, Trondheim, Norway

E.S. Kiel
Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: High-Impact Low-Probability (HILP) events in power systems historically involve a multitude of aspects, including diverse and disparate threats, failures and sequences of events. Each of these aspects are associated with different types of uncertainties. In practice, the analyst has to make trade-offs between computational efficiency and accuracy in the different aspects that are included in the analysis. Without a clear understanding of the specific problem to be solved and which aspects that are important to capture, elaborate quantitative analysis may be of limited value. This paper presents the development of a *qualitative* framework for analysing HILP events in power systems. By mapping aspects of power system HILP events to a bow-tie model, it provides a framework for defining, decomposing and delimitating decision problems related to such events. The framework may guide the analyst in the development and application of methods for quantitative analysis and for considering different types of uncertainties.

1 INTRODUCTION

A High-Impact Low-Probability (HILP) event, also referred to as an extraordinary event, is an event with a high societal impact and a low probability to occur. In power systems, such events are often understood as blackouts, i.e. wide-area power interruptions. A number of such major blackout events have occurred in the last few decades (Bompard et al. 2013, Hillberg 2016), each resulting in critical consequences to society. Such events therefore receive great attention both by power system operators and other stakeholders, such as researchers and the general public, despite their low probability of occurrence. Partly due to this low probability, these events typically are not captured in conventional reliability and risk analyses, which calls for analysis approaches specific to HILP events.

HILP events historically involve a multitude of diverse and disparate threats and complex sequences of events, which present the analysts and researchers studying them with numerous uncertainties. Relevant aspects that can be taken into account in quantitative modelling of HILP events include: failure bunching due extreme weather (Panteli and Mancarella 2015), other natural hazards, cascading outages (Vaiman et al. 2012, Dobson and Newman 2017), dynamic phenomena, system protection schemes (Hillberg et al. 2012), corrective actions (Vadlamudi et al. 2016), and valuation of the societal impact. Different approaches and methodologies exist for quantitatively analysing these events (Gjerde et al. 2011), including methods of identifying unwanted events, causal analysis, consequence analysis, and risk and vulnerability evaluation. Such methods typically focus on one or a subset of all potentially relevant aspects. The realization is that there is no single methodology covering all these aspects that is suitable for analyzing HILP events in power systems (Kjølle et al. 2013), and the full set of aspects is too comprehensive to analyse quantitatively. Without a clear understanding of what specifically is the problem to be solved or decision to be supported, and consequently which aspects are important to capture, elaborate quantitative analysis may be of limited value.

In this paper, we take a broader view on HILP events and present the development of a qualitative framework for analysing HILP events in power systems. A qualitative framework provides the analyst with a more complete overview of the set of problems and a starting point for detailed analysis. Previous work on HILP events largely focus on methods of detailed, quantitative analysis (Vaiman et al. 2012), but some work on the more conceptual level also exists. For instance, (Watson et al. 2014) developed a framework for resilience metrics for energy infrastructures. In (Veeramany et al. 2016),

an overarching modelling framework is formulated under which different models can be integrated for an multi-hazard risk assessment of power system HILP events. The cascading aspect of some HILP events is discussed conceptually in (Vaiman et al. 2012, Dobson and Newman 2017).

The qualitative framework presented in this paper is based on an existing framework for power system vulnerability analysis (Kjølle et al. 2013, Kjølle and Gjerde 2015). The present paper advance previous work and attempts to consolidate relevant aspects of HILP events in a consistent and all-encompassing mapping. This framework explicitly discusses and structures uncertainties related to different decision problems. The framework is presented in Section 2, which forms the bulk of this paper. Subsection 2.1 shows how mapping relevant aspects and their relationships to a bow tie model provides a more complete overview of HILP events. Subsection 2.2 to Subsection 2.4 presents an approach to defining, delimitating and decomposing decision problems related to HILP events. This provides a starting point for quantitative analysis, as discussed in Section 2.4, and a basis for taking into account uncertainties, which is discussed in Section 2.5. Throughout these subsections, concrete examples of problems are discussed to illustrate the application of the framework. Finally, Subsection 3 concludes the paper and indicates future work in refining and applying the framework.

2 QUALITATIVE FRAMEWORK FOR HILP EVENTS

The qualitative framework presented in this paper is based on the conceptual bow tie model and a previously developed framework for power system vulnerability analysis (Kjølle et al. 2013, Kjølle and Gjerde 2015). The bow tie model describes the relationship between causes and consequences of unwanted events, which are here defined as power system failures. Note that the unwanted event in the centre of the bow-tie is not by itself a HILP event, but it could be the initiating event of a sequence of events with critical consequences that constitutes the HILP event.

2.1 *Getting a better overview of relevant aspects*

The bow tie model can be used as a visual aid in structuring the causes and consequences of unwanted events as illustrated in Figure 1. This figure gives a comprehensive overview of aspects relevant to HILP events in power systems and how these relate to each other. Such an overview is useful when structuring an analysis of HILP events.

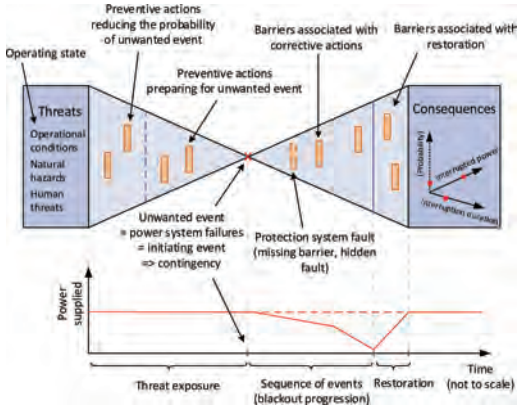


Figure 1. Overview of relevant aspects of HILP events in power systems mapped to a bow-tie model.

The left-hand part of the figure shows schematically how the exposure of the power system to different threats can cause power system failures, and the right-hand part shows how power system failures can result in consequences external to the power system, i.e. societal impact. The criticality of the consequences can be measured along different dimensions, but for the illustrations in this paper we will consider total end-user power interruption (MW) and interruption duration (hours) as the two principal dimensions. Each HILP event could, in principle, also be associated with a probability. Other relevant factors include the types of end-users affected and the dependence of the society on electricity supply; for further discussion of the definition of "critical", we refer to (Kjølle et al. 2013, Kjølle and Gjerde 2015).

Relevant threats on the left-hand side include conditions related to the operating state of the power system (e.g. challenges related to the power import/export situation, prior outages, etc.), natural hazards such as major storms and human threats. Barriers on the left-hand side of the bow tie reduce the susceptibility of the power system to threats. These barriers reduce the probability of unwanted events through preventive actions such as condition monitoring, preventive maintenance and vegetation management. Some barriers also preemptively increase the coping capacity of the system to reduce the probability of critical consequences in case an unwanted event does occur. This category of barriers includes preventive scheduling, grid reconfiguration and islanding in preparation for a major storm.

Barriers on the right-hand side of the bow-tie are intended to reduce the consequence of power system failures and correspond to the coping capacity of the power system with respect to these

unwanted events. Examples of such barriers are corrective actions such as emergency generation rescheduling, controlled load shedding, controlled islanding, and various system protection schemes. Other barriers are associated with the restoration of system operation after power has been interrupted, for instance the black-start capability of generators and the availability of spare parts, equipment and competent personnel.

To illustrate the distinction between these two types of barriers, we have in Figure 1 superimposed a timeline with an example of how the interrupted power could develop as a function of time throughout the course of the HILP event. The sequence of events after the occurrence of the initiating event can be broadly separated in a blackout progression phase and a restoration phase. Corrective action barriers are associated with the blackout progression phase and primarily intended to reduce the amount of interrupted power, whereas barriers associated with the restoration phase generally intended to reduce the restoration time and thus the interruption duration.

2.2 *Defining and framing the problem*

The analysis of HILP events in power systems is a broad problem area involving different decision problems as well as more fundamental research problems. The question one needs to ask is why one is interested in analyzing HILP events the first place. It is necessary with a clear definition the problem and a clear understanding of the motivation and purpose of solving the problem.

Figure 2 shows two dimensions that can be used to frame problems related to HILP events: The time scales for power system-related decisions and relevant stakeholders or decision makers. The figure also indicates the motivation of the stakeholders with regards to HILP events. The two dimensions in Figure 2 determine what information is available to the analyst and thus what uncertainties must be taken into account. This will be discussed in more detail in Section 2.5.

Here we will distinguish between operational, tactical and strategic decisions by the time scale of the planning horizon that is considered. Following the classification in (GARPUR Consortium 2016), these three time scales correspond to system operation (including both real-time operation and day-ahead operational planning), asset management, and system development or planning, respectively. Note that other references may use other terms and definitions for the time scales. For instance, (Watson *et al.* 2014) distinguishes between system planning decisions and policy decisions, and (Yang and Haugen 2015) defines both strategic and operational decisions as planning decision, which are in turn distinguished from instantaneous or emergency decisions.

Stakeholders can be differentiated in terms of their influence over power system related decisions, and since system operators have the most direct influence, we will in the following take the perspective of the system operator as a decision maker. Furthermore, we will focus on transmission system operators (TSOs) since distribution system operators (DSOs) have less influence over decisions relevant for wide-area power interruptions. In practice, decisions will be taken by different departments and at different levels in the organisation, but in the following we simply refer to the decision maker as "the system operator".

To put the more general problem of analysing HILP events in a decision-making context, Figure 3 shows some examples of relevant decision problems for system operators, sorted by time scale. These decision problems will be defined in broad terms below and be used in the following sections to illustrate the qualitative framework. Although we do not define the decision problems formally in terms of their objective function etc. as done e.g. in (GARPUR Consortium 2016), it

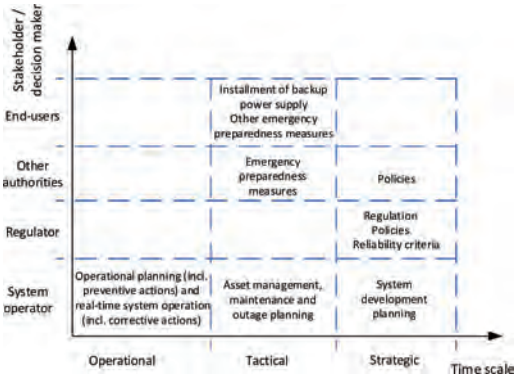


Figure 2. Two dimensions relevant for framing problems related to HILP events: The stakeholder or decision maker, and the time scale of relevant decision problems.

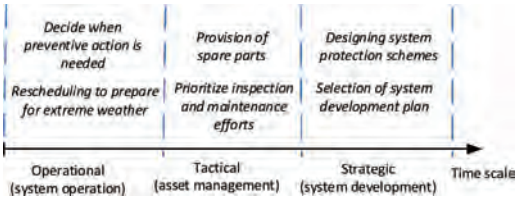


Figure 3. Examples of decision problems for transmission system operators with relevance for the analysis of HILP events.

is important to keep in mind that these reliability management decisions typically involve some form of trade-off between costs and reliability of supply. The value of reliability of supply is sometimes monetized in the form of expected interruption costs, i.e. the cost of energy not supplied.

Selection of system development plan: An example of a strategic decision problem is the evaluation of candidate system development plans (e.g. for new transmission lines) and selection of the best candidate. Regulation may dictate that a socio-economic cost-benefit analysis of the candidates is performed. Ideally, the cost of energy not supplied associated with possible HILP events should be included in such an analysis.

Designing system protection schemes: System protection schemes (SPSs) are important examples of barriers on the right-hand side of the bow-tie, and the system operator has to plan which SPSs to implement. The motivation of implementing an SPS could be to increase the transmission capacity of the system as well as to increase the coping capacity of the system with respect to the occurrence of contingencies that would otherwise result in critical consequences (Hillberg *et al.* 2012).

Prioritize inspection and maintenance efforts: The system operator has to decide how to best allocate limited resources for preventive actions such as intensified inspection and maintenance and improved condition monitoring of power system components. Mitigating certain susceptibilities could help reduce the risk of HILP events as well as more ordinary events.

Spare parts etc. for critical components: If the power system is vulnerable to the loss of certain component, e.g. a transformer, the decision can be made to provide for spare parts to reduce the duration of potential power interruptions.

Decide when preventive action is needed: During operation, preventive actions such as generation rescheduling may be needed e.g. due to the development of threat exposure and/or the operating state. The first step for the system operator is to correctly assess the situation and decide whether or not to effectuate preventive actions.

Rescheduling generation e.g. to prepare for extreme weather: During an extreme weather event the near-simultaneous failure of multiple transmission lines (failure bunching) is more likely. In this case, one relevant preventive action is to reschedule generation in a way that makes the power system better able to cope with failures on one or several transmission lines.

2.3 *Defining and delimiting the analysis*

Decision making for problems as exemplified above can be supported by the analysis of HILP events. One way of defining and delimitating "analysis of HILP events" is to consider sub-problems distinguished by the objective of the analysis. One possible classification is:

1. identifying critical contingencies
2. identifying critical operating states
3. identifying critical barriers
4. assessing the contributions to the overall reliability of supply

Each of these sub-problems can be associated with different parts of the bow-tie model as illustrated in Figure 4. In practice, the objectives may be overlapping and the sub-problems may be combined in one of the same analysis. The classification may nevertheless be useful in discussing specific decision problems and the underlying motivation.

2.3.1 *Identify critical contingencies*

A critical contingency is here understood as a failure or unplanned outage of a power system component that may potentially result in critical consequences. One purpose of identifying critical contingencies is to identify critical power system components with the motivation to strengthen or introduce appropriate barriers, cf. Section 2.3.3.

One example of a system operation decision involving the identification of critical contingencies is the (optimal) preventive rescheduling of generation in preparation for an extreme weather event. In this case, the system operator should ideally know which (critical) higher-order contingencies to take into account when rescheduling. In the context of system development, one would like to identify critical contingencies in the candidate development plans to reduce the vulnerabilities of the development plan that is selected. Another purpose of identifying critical contingencies can be to screen contingencies to be considered as input to more detailed (e.g. dynamic) analysis.

2.3.2 *Identify critical operating states*

We here understand a critical operating state as an operating state which in combination with a

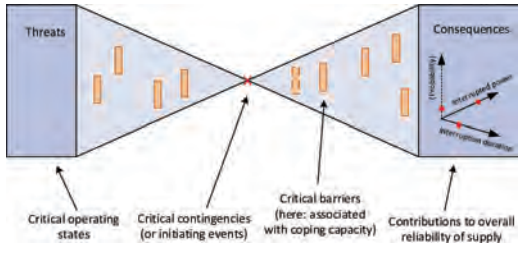


Figure 4. The placement in the bow tie model of different criticalities and sub-problems relevant in the analysis of HILP events.

critical contingency potentially result in critical consequences. The motivation for identifying these could be to increase the situational awareness of the system operators, which has previously been identified as being crucial to avoid HILP events (Johansson, E. *et al.* 2010). Situational awareness is relevant for operational decisions on which corrective actions to carry out after a contingency has occurred. Identifying critical operating states prior to contingencies may also be important to be able to decide when preventive action is needed.

2.3.3 *Identify critical barriers*

The identification of critical barriers may be used in selecting barriers to strengthen, and the identification of critical barriers that are missing may be used in proposing new barriers to put in place. This involves corrective barriers such as well-designed system protection schemes, or preventive barriers such as inspection and maintenance. For the latter example, the decision of which components to prioritize also depends on the identification of critical contingencies.

2.3.4 *Assessing the contributions to the overall reliability of supply*

An underlying premise of this work is that conventional power system reliability analysis methods do not fully capture HILP events. The reliability of a power system can be defined as "the probability of its satisfactory operation over the long run. It denotes the ability to supply adequate electric service on a nearly continuous basis, with few interruptions over an extended time period" (Kundur *et al.* 2004). The overall reliability of supply may be quantified by reliability indices such as the expected annual energy not supplied. Over the long run, HILP events do contribute to these reliability indices, but their contribution may be underestimated by conventional reliability analysis methods. For instance, this may happen when the methods do not capture failure bunching, protection system failures, or any of the other aspects and dependencies that may conspire to result in a HILP event. Furthermore, the short-term impact of a HILP event may be disproportional to their long-run visibility in expected values of reliability indices and therefore warrant separate treatment (Vaiman *et al.* 2012). These are some of the reasons why methods of vulnerability analysis focusing on HILP events have been advocated to complement traditional risk and reliability analysis methods (Johansson *et al.* 2013, Kjølle and Gjerde 2015).

Nevertheless, estimates of reliability indices are used by system operators as part of their reliability management processes also for decisions relating to HILP events. An example is the selection of system development plans for a given region, supported by a socio-economic cost-benefit analysis including expected interruption costs. If the region is exposed to strong winds, this could motivate capturing the contribution of HILP events due to failure bunching effects in the estimated interruption costs.

2.4 *Decomposition in quantitative analysis*

After defining the purpose of the analysis, one needs to consider which quantities the analysis method needs to estimate and which of them is most important to estimate accurately. Here we will consider three primary output parameters: 1) The probability of an event and its consequence in terms of 2) power interrupted and 3) interruption duration. As illustrated in Figure 5, these output parameters are broadly speaking associated with different parts of the bow-tie model. To assess the consequences of an unwanted event, it is sufficient to consider the right-hand side of the bow-tie: The interrupted power is primarily determined by the sequence of events within the phase labelled "blackout progression", and the interruption duration is primarily determined by the events in the restoration phase. On the other hand, to determine the probability of a HILP event, characterized by a given consequence, one has to consider both the left-hand side (with the label "threat exposure" in Figure 5) and the right-hand side of the bow-tie.

To approach more quantitative analysis and consideration of different uncertainties, we overlay the bow tie model with a schematic data flow diagram for the analysis in Figure 6. A cause analysis is depicted on the left-hand side of the bow tie that gives as output the failure rate (or the probability of failure during a certain time interval) for a given unwanted event (i.e. a given power system failure). Such a module could for instance be based on a fault tree. Failure bunching effects, for example due to major storms, could be incorporated in this step using existing tools for estimation of wind-dependent failure rates, as done in (Solheim *et al.* 2016).

The consequence analysis on the right-hand side of Figure 6 is divided in two modules representing

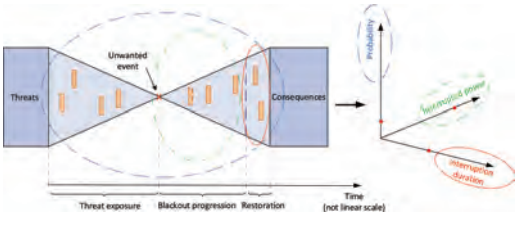


Figure 5. Illustration of how the problem of analysing extraordinary events can be decomposed and delimitated based on what quantity one is focusing on estimating.

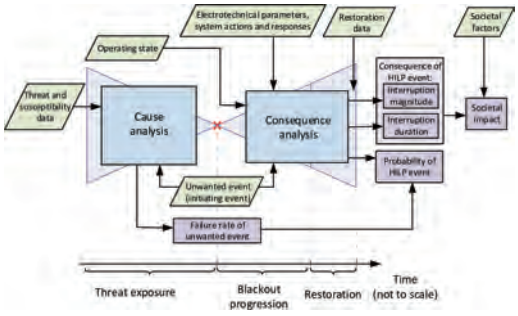


Figure 6. Schematic of quantitative analysis (blue, within the bow-tie) with input data (green parallelograms) and output data (purple).

the blackout progression phase and the restoration phase, respectively. The module for the blackout progression phase models system responses and resulting power interruptions. It could be based on an event tree model, power flow analysis, dynamic analysis, etc. This module can take as input electrotechnical parameters describing the power system and its operational limits as well as parameters describing the actions and responses in the system. For instance, if the analysis method is based on an event tree accounting for corrective action failures (Vadlamudi *et al.* 2016), input parameters can be conditional probabilities determining the probability of different sequences of events. The restoration phase module represents the restoration process. For instance, the restoration time could be modelled by average outage times of the components involved, in which case such outage times are needed as input. Alternatively, the restoration process could be modelled in more detail, which would require additional input parameters.

When analyzing system protection schemes to identify critical barriers for certain unwanted events, it may not be important for the purpose of the analysis to consider what caused these unwanted events. For such an analysis, one could omit the left-hand side of Figure 6 and focus on the first part of the consequence analysis, e.g. using dynamic analysis to estimate the power interrupted. On the other hand, if the objective is to assess the contribution to the overall reliability of supply, one would typically also have to represent power system restoration in the analysis.

In the determination of the consequences illustrated in Figure 6, the consequence analysis stops after finding the interruption magnitude and duration. However, as mentioned in Section 2.1, the societal impact of an HILP event is not determined by these two parameters alone. The box labeled societal factors in Figure 6 represent other factors determining the societal impact, such as the type of customers (end-users) and the criticality of the loads that are interrupted. Consequences of power interruptions are typically monetized using interruption cost functions determined by customer surveys, but these interruption costs give only a lower bound for the total socio-economic costs of the power interruption (GARPUR Consortium, 2016). Estimating quantitatively the impact on society more widely might involve modelling of the interactions between the power system and other infrastructures (Johansson et al. 2015).

2.5 *Taking into account uncertainties*

HILP events can be argued to be inherently associated with uncertainties (Taleb 2010, p. xxviii). Factors such as the operating state, the technical condition of components and failure bunching effects due to adverse weather all have their own individual uncertainties. HILP events are often the results of multiple, interacting factors and circumstances. As such, their combined uncertainty is larger than the uncertainty of the individual factors.

First, it is common to classify uncertainties as either aleatory, i.e. associated with random variability, or epistemic, i.e. associated with a lack of knowledge. Given that HILP events are characterized by a scarce experience base and severe lack of knowledge, epistemic uncertainties are especially important to consider. Next, following a similar classification as in (Rausand 2013), we will broadly distinguish between three types of uncertainties:

- Input data uncertainties
- Modelling uncertainties
- Completeness uncertainties

For the analysis of HILP events in power systems, these types of uncertainties can be related to Figure 6 as follows. Input data uncertainties and modelling uncertainties are related to green and blue boxes, respectively. The additional category that we have here chosen to label "completeness uncertainty" represents uncertainty associated with the completeness of the models of the system. Although there are different ways to understand this term (Rausand 2013, Aven 2016), and "completeness uncertainty" may not be unambiguously distinguished from "modelling uncertainty", we find the term useful to describe uncertainty associated with aspects omitted and/or outside the scope of the analysis. As an example, a consequence analysis starting from a given set of contingencies (i.e. covering only the right-hand side of Figure 6) does not explicitly consider what might have caused the contingencies. If the problem was to identify effective system protection schemes, for instance,

threat and susceptibility aspects may not have been within the scope of the analysis.

Sources of incompleteness in the analysis can be either known or unknown to the analyst (Aven 2016). If the analyst is unaware that an aspect is not considered in the analysis, this uncertainty can be labelled an "unknown unknown" (Feduzi and Runde 2014). Here, we use this term in a wider sense to refer to lack of knowledge that is implicit, i.e. a form of epistemic uncertainty associated with "what we don't know we don't know". Furthermore, we focus on "unknown unknowns" that are "knowable", i.e. that can in principle be transformed into "known unknowns" (Feduzi and Runde 2014).

Another way to classify uncertainties related to an analysis of HILP events that is more specific to the domain of power systems is to consider uncertainties related to the aspects discussed in Section 2.1. An example of such a classification is illustrated in Figure 7. Here, each of the categories along the vertical axis corresponds to one of the components of quantitative analysis that were illustrated in Figure 6. This shows how a domain-specific classification can be combined with the generic uncertainty classification discussed above: For each category, a given analysis is associated with uncertainty (indicated along the horizontal axis) related to the accuracy of modelling assumptions and the input data.

This multi-dimensional classification of uncertainties can be used to structure a qualitative assessment of the strength of background knowledge (Aven *et al.* 2014, p. 87) underlying a given analysis: If an aspect is modelled in a simplified or inaccurate manner, the knowledge of this aspect that is represented in the analysis is weak and the uncertainty is correspondingly high. Even if the modelling of an aspect is accurate, the uncertainty is still high if the associated input data represented in the analysis is inaccurate.

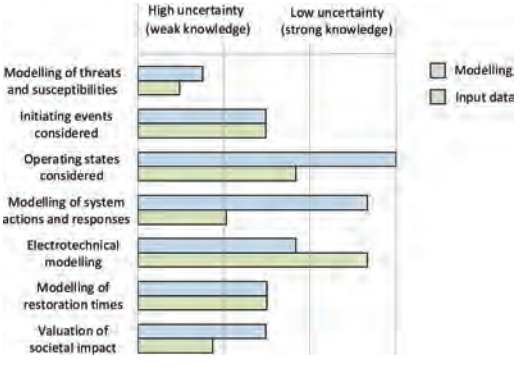


Figure 7. Example of classification and assessment of uncertainties associated with analyses of HILP events.

Such a structured assessment of the uncertainties of a HILP event analysis can be used by the analyst to rank which uncertainties are most important (Aven *et al.* 2014) to improve the overall accuracy and suitability of the analysis. More accurate modelling of an aspect often implies longer computation times. In practice, a trade-off must therefore be made between computational efficiency and accuracy, and trade-offs must be made between the modelling accuracy for the different aspects considered in the analysis.

An explicit qualitative assessment of uncertainties can also be used as a basis for comparing different analyses and informing the decision maker of their uncertainties (Aven *et al.* 2014). As an example, one can consider methods designed to analyse cascading outages. A number of such methods have been developed, each focusing on different subsets of the mechanisms and aspects involved in cascading outages. Considerable efforts have already been devoted to reviewing and validating such methods (Vaiman *et al.* 2012, Bialek *et al.* 2016), but there are still many open questions that may limit their credibility in decision making. More explicit classification and assessment of their uncertainties, scope and purpose could help inform system operators of which methods are most suitable for different problems.

Completeness uncertainty is not included as a separate dimension in Figure 7, but if an aspect is not covered in an analysis, the modelling uncertainties related to this aspect can be regarded as high. However, to fully characterize the completeness uncertainty dimension of the analysis one needs to identify and uncover "unknown unknowns". It has been argued that to do so, the analysis needs to be placed in a sufficiently broad framework and avoid starting out with a too narrow view of the problem (Feduzi and Runde 2014, Aven 2016). A qualitative mapping of relevant aspects to the analysis as proposed in this paper can contribute to transforming "unknown unknowns" to "known unknowns", or in other words making implicit assumptions and uncertainties explicit. Communicating such uncertainties associated with the completeness of the analysis can change, from the perspective of the decision maker, a "unknown unknown" to a "known unknown". To give a simple example: When deciding on system protection schemes to mitigate cascading outages and the analysis does not model the dynamics of rotor angle stability, the decision maker should be aware that the type of cascading events characterized by generators losing synchronism is omitted from the analysis.

As mentioned in Section 2.2, the time scale of the decision problem is relevant for what information is available during the analysis and hence what is uncertain and what is known. For instance,

the system operator knows the operating state to a good approximation during real-time system operation, whereas this information is not available for an analysis for long-term planning purposes (Vaiman et al. 2012). For the example of cost-benefit analysis including the contributions of wind-related failures, the analyst needs to assume a selection of operating states expected to be representative of the future, and this is associated with additional uncertainties. For the example of preventive rescheduling in preparation of a major storm, more information is available on the operating state over the planning horizon, although this is still imperfect information as one may have to consider the forecast uncertainties.

3 CONCLUSIONS AND FUTURE WORK

This paper proposes a qualitative framework for analysing HILP events in power systems that may complement or guide more quantitative analysis. Mapping relevant aspects of such HILP events to a bow tie model provides the analyst with a broad overview of the set of problems at hand and a starting point for detailed analysis. Although the full set of aspects is too comprehensive to analyse quantitatively, the qualitative framework provides a basis for decomposing and delimitating the problem: Defining precisely the purpose of the analysis, one can then choose what aspects need to be modelled accurately and which aspects one is choosing to omit. Omitting and neglecting aspects of the overall problem introduce uncertainties in the analysis, but by being explicit about what is omitted and assumed one reduces the amount of "unknown unknowns" in the analysis and may thus support more well-informed decisions.

Further work will test the applicability of the framework in case studies of real problems related to HILP events. The approach for defining the purpose of an analysis and delimitating the problem presented will also be used to guide the development and application of methods for quantitative analysis of HILP events. Furthermore, the classification of models and input data for the analysis may form the basis for considering which methods are most appropriate for handling different types of uncertainties related to modelling choices and input data.

REFERENCES

Aven, T., 2016. Ignoring scenarios in risk assessments: Understanding the issue and improving current practice. *Reliability Engineering & System Safety*, 145, 215–220.

Aven, T., Zio, E., Baraldi, P., & Flage, R., 2014. Uncertainty in Risk Assessment: The Representation and Treatment of *Uncertainties by Probabilistic and Non-Probabilistic Methods.* Chichester, UK: Wiley.

Bialek, J., Ciapessoni, E., Cirio, D., Cotilla-Sanchez, E., Dent, C., I. Dobson, P. Henneaux, P. Hines, J. Jardim, S. Miller, M. Panteli, M. Papic, A. Pitto, J. Quiros-Tortos, & D. Wu, 2016. Benchmarking and Validation of Cascading Failure Analysis Tools. *IEEE Transactions on Power Systems*, PP, 1–14.

Bompard, E., Huang, T., Wu, Y., & Cremenescu, M., 2013. Classification and trend analysis of threats origins to the security of power systems. *International Journal of Electrical Power & Energy Systems*, 50 (Supplement C), 50–64.

Dobson, I. and Newman, D.E., 2017. Cascading blackout overall structure and some implications for sampling and mitigation. *International Journal of Electrical Power & Energy Systems*, 86, 29–32.

Feduzi, A. & Runde, J., 2014. Uncovering unknown unknowns: Towards a Baconian approach to management decision-making. *Organizational Behavior and Human Decision Processes*, 124 (2), 268–283.

GARPUR Consortium, 2016. *D2.2: Guidelines for implementing the new reliability assessment and optimization methodology.*

GARPUR Consortium, 2016. *D3.2: Recommendations for implementing the socio-economic impact assessment methodology over the pan-European system in a tractable way.*

Gjerde, O., Kjølle, G.H., Detlefsen, N.K., & Brønmo, G., 2011. Risk and vulnerability analysis of power systems including extraordinary events. Presented at the PowerTech 2011, Trondheim.

Hillberg, E., 2016. Perception, Prediction and Prevention of Extraordinary Events in the Power System. PhD thesis. Norwegian University of Science and Technology, Trondheim.

Hillberg, E., Trengereid, F., Breidablik, Ø., Uhlen, K., Kjølle, G., Løvlund, S., & Gjerde, J.O., 2012. System integrity protection schemes—Increasing operational security and system capacity. Presented at the CIGRE Session, Paris.

Johansson, E., Uhlen, K., Nybø, A., Kjølle, G., & Gjerde, O., 2010. Extraordinary events: understanding sequence, causes, and remedies. Presented at the European Safety & Reliability Conference (ESREL) 2010, Rhodes.

Johansson, J., Hassel, H., Cedergren, A., Svegrup, L., & Arvidsson, B., 2015. Method for describing and analysing cascading effects in past events: Initial conclusions and findings. Presented at the European Safety & Reliability Conference (ESREL) 2015, Zürich, Switzerland.

Johansson, J., Hassel, H., & Zio, E., 2013. Reliability and vulnerability analyses of critical infrastructures: Comparing two approaches in the context of power systems. *Reliability Engineering & System Safety*, 120, 27–38.

Kjølle, G.H. & Gjerde, O., 2015. Vulnerability analysis related to extraordinary events in power systems. Presented at the PowerTech 2015, Eindhoven.

Kjølle, G.H., Gjerde, O., & Hofmann, M., 2013. *Vulnerability and security in a changing power system—*

Executive summary. Trondheim: SINTEF Energy Research, Report No. TR A7278.

Kundur, P., Paserba, J., Ajjarapu, V., Andersson, G., Bose, A., Canizares, C., Hatziargyriou, N., Hill, D., Stankovic, A., Taylor, C., van Cutsem, T., & Vittal, V., 2004. Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions. *IEEE Transactions on Power Systems*, 19, 1387–1401.

Panteli, M. & Mancarella, P., 2015. Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies. *Electric Power Systems Research*, 127, 259–270.

Rausand, M., 2013. *Risk assessment: theory, methods, and applications.* John Wiley & Sons.

Solheim, Ø.R., Kjølle, G., & Trötscher, T., 2016. Wind dependent failure rates for overhead transmission lines using reanalysis data and a Bayesian updating scheme. Presented at the 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Beijing: IEEE.

Taleb, N.N., 2010. *The black swan: The impact of the highly improbable*. Revised edition. London: Penguin Books.

Vadlamudi, V.V., Hamon, C., Gjerde, O., Kjølle, G., & Perkin, S., 2016. On Improving Data and Models on Corrective Control Failures for Use in Probabilistic Reliability Management. Presented at the 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Beijing: IEEE.

Vaiman, M., Bell, K., Chen, Y., Chowdhury, B., Dobson, I., Hines, Papic, Miller, & Zhang, 2012. Risk Assessment of Cascading Outages: Methodologies and Challenges. *IEEE Transactions on Power Systems*, 27, 631–641.

Veeramany, A., Unwin, S.D., Coles, G.A., Dagle, J.E., Millard, D.W., Yao, J., Glantz, C.S., & Gourisetti, S.N.G., 2016. Framework for modeling high-impact, low-frequency power grid events to support risk-informed decisions. *International Journal of Disaster Risk Reduction*, 18, 125–137.

Watson, J.-P., Guttromson, R., Silva-Monroy, C., Jeffers, R., Jones, K., Ellison, J., Rath, C., Gearhart, J., Jones, D., Corbet, T., Hanley, C., & Walker, L.T., 2014. *Conceptual Framework for Developing Resilience Metrics for the Electricity, Oil, and Gas Sectors in the United States*. Albuquerque, New Mexico and Livermore, California: Sandia National Laboratories, Report No. SAND2014–18019.

Yang, X. & Haugen, S., 2015. Classification of risk to support decision-making in hazardous processes. *Safety Science*, 80 (Supplement C), 115–126.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Safety assessment: Perspectives for next generation nuclear plants

A. Carpignano, S. Dulla & A.C. Uggenti
NEMO group, Dipartimento Energia, Politecnico di Torino, Italy

ABSTRACT: Safety assessment and risk analysis are recognized as a priority in the development of next generation nuclear systems (Generation-IV reactors and full-scale fusion reactor—DEMO-) and demand a reconsideration of the safety philosophy currently applied to the existing nuclear stations. Since their innovative physics and technology and the preliminary design phase of some of the concepts, their safety assessment has to rely on the basis of nuclear safety and technological neutral methodology. In order to satisfy this necessity, a bibliographic survey on nuclear and non-nuclear international standards and best practices is performed. By comparing them, this work tries to reach a new and more systematic approach, based on functional safety, suitable for dealing with the unique challenges of the innovative nuclear facilities, in order to guarantee that safety achievement is intended to be "built-in" rather than "added-on" by influencing the concept evolution from its earliest stages.

1 INTRODUCTION

The contemporary research activity in the nuclear field is focused on the development of nuclear facilities able to satisfy the four goal areas identified by the Generation IV International Forum (GIF, 2014) in its Technological Roadmap in order to advance nuclear energy in its next generation: sustainability, safety and reliability, economic competitiveness, proliferation resistance and physical protection. The attempt to answer this request with a fully innovative technology is the rationale associating all Generation IV reactor designs and the proposed concepts for a full-scale fusion reactor (EUROfusion website).

The nuclear energy systems must be designed so that, during normal operation or anticipated transients, safety margins are adequate, accidents are prevented and off-normal situations do not deteriorate into severe plant conditions (RSWG of the GIF, 2008). Therefore, safety assessment and risk analysis, in both operational and accidental conditions, are recognized as an essential priority in the development of these next generation nuclear systems. Because of their innovative physics and technology and the preliminary design phase of some of the concepts, their safety assessment has to rely on the basis of functional safety and technological neutral methodologies. This demands a reconsideration, a modernization and an adaptation of the safety philosophy currently applied to the existing nuclear stations and a constant innovation and development of safety assessment methods to continue to advance the state of the art and improve their adequateness.

In 2002 GIF selected six systems from nearly 100 concepts as the Generation-IV fission nuclear plants: the Gas-cooled Fast Reactor (GFR), Sodium-cooled Fast Reactor (SFR), Lead-cooled Fast Reactor (LFR), Molten Salt Reactor (MSR), Very-High-Temperature Reactor (VHTR), Supercritical-Water-cooled Reactor (SCWR) (GIF, 2014); on the other hand the DEMOnstration fusion power reactor (DEMO) is foreseen to follow the advancements of ITER (International Thermonuclear Experimental Reactor) by 2050 (EUROfusion website). These systems present a wide range of new technologies that create issues if the traditional safety approach adopted for Light Water Reactors (LWRs) is considered: for example, the MSR design is characterized by a liquid nuclear fuel, therefore the evaluation of the Core Damage Frequency (CDF) in terms of core melting as an indication of severe accident is no longer applicable. Moreover, due to the online refueling envisaged for MSR, also in normal operation conditions the fuel is not localized in the core (as it happens for the LWR) but it is spread in several subsystems and occupies different positions in the reactor, making inconsistent the traditional definition of physical barriers (cladding, primary circuit, containment building). A general comment, valid for all these innovative nuclear systems including fusion machines, is that, in many cases, the design is still in development therefore a safety assessment performed at the components level is not useful since their architecture will evolve in time: instead, a functional approach allows to identify the functional deviations challenging the system since the early design and, consequently, to include safety features in a holistic optics.

This work starts investigating the safety challenges of the new generation of nuclear plants and performing a bibliographic survey on nuclear and non-nuclear safety international standards and best practices; the objective of the paper is to present an iterative methodology that is coherently applicable since the conceptual phase of the design and aims at influencing the direction of the concept and design development from its earliest stages; hence the safety will be intended to be "built-in" rather than "added-on".

2 SAFETY CHALLENGES FOR NEW GENERATION NUCLEAR PLANTS

The majority of current nuclear safety regulatory requirements is based on LWRs technology and necessitates changes to suit to a new spectrum of novel, advanced, next generation plants (Southern Company, 2017). In Probabilistic Safety Assessment (PSA), the risks associated with the reactor accidents are highly design, plant and site specific; this is demonstrated for any kind of reactor. In particular, dealing with next generation nuclear plants implies a much larger range of risks variability with respect to an LWR: fundamental differences in the physical processes are present, as well as in the plant responses associated with the reactor transients and accidents. This is due both to the use of different materials for the reactor fuel, moderator and coolant and to different safety design approaches for the implementation of radionuclides barriers (Southern Company, 2017). Because of these differences, the LWR risk metrics, for instance the Core Damage Frequency (CDF) and the Large Early Release Frequency (LERF), are neither relevant nor useful for many advanced nuclear reactors; some plants, in fact, may not involve the core damage state that was defined for LWR and, even in the case, its meaning and risk framework can be fundamentally different from LWR (INL, 2011). Consequently, PSA for advanced reactors may be structured differently than the traditional Level 1, 2 and 3 model for LWR PSA: it is expected to include out of core sources of radioactive material (especially in the case of online refuel, as for the MSR) and to adopt adequate and more general risk metrics (INL, 2011); the latter may lead to an appropriate definition of severe accident, detached for the core melting concept. Additionally, while the traditional LWR risk assessment was developed following the "one-reactor-at-a-time" approach, in next generation nuclear plants the risk associated to multi-unit sites becomes certainly relevant and, especially after the Fukushima Daichi accident, even dominant (Fleming, 2017). Advanced non-LWRs are expected to be constituted by several modules, located in the same site: this increases the possibility of common cause failures/domino effects, due to the potential for sharing of systems and structures or hazards involving more than one reactor (e.g. external hazards). This influences the traditional frequency-consequence tolerability criteria.

Lastly, a major difference between the risk assessment methodologies (e.g. PSA) of LWRs and next generation nuclear plants is the following: the former were introduced after the plants were designed and licensed, limiting the risk-informed applications to additional systems or provisions for plants that were already built and operated; on the other hand, the latter are primarily used as tool to support the design and to expand the range of the risk-informed decisions (Southern Company, 2017).

3 OVERVIEW OF SAFETY ASSESSMENT METHODOLOGIES AND STANDARDS

A huge set of prior activities, policies, standards, practices and requirements support the design and the licensing of LWRs.

IAEA (International Atomic Energy Agency) standards provide the fundamental principles, requirements and recommendations to ensure nuclear safety. They serve as a global reference for protecting people and the environment and contribute to a harmonized high level of safety worldwide (IAEA, 2006): as stated in the fundamental safety principles of the IAEA Safety Standard for protecting people and environment, (IAEA, 2006), "the fundamental safety objective is to protect people and the environment from harmful effects of ionizing radiation"; this fundamental safety principle is detailed in ten safety principles on the basis of which safety requirements are developed and safety measures are implemented in all nuclear facilities and activities, and for all stages over the lifetime of a facility or a radiation source. These principles inspire the "General Safety Requirements" and the "General Safety Guide" that, for each technical area, are declined into a number of "Specific Safety Requirements" and of "Specific Safety Guides" that provide all the guidance necessary for implementing the general principles. While the ten safety principles are general enough to be applicable also to non-LWRs, all the other documents and standards are referred specifically to LWRs. Similarly, the Nuclear Regulatory Commission (NRC) regulations and in particular the Title 10 of the Code of Federal Regulations Part 50 (10 CFR 50) (USA NRC, 2017) establish Principal Design Criteria (PDC) derived from the General Design Criteria (GDC) that are specifically

referred to LWRs (Appendix A of 10 CFR 50). Considering the fact that the last USA commercial non-LWR was shut down in 1989 (Fort St. Vrain, a High-Temperature Gas-cooled Reactor—HTGR), the update of these documents is on-going but is especially challenging because of lack of specificity in the technology/designs that will be ultimately submitted to NRC for review, of lack of maturity of design and of the unavoidable technical skills gap (Lee, 2016).

Traditionally, the PSA is performed only after the definition of the detailed design and of the site: in this case, if the tolerability criteria are not fulfilled, it could be necessary to modify also the preliminary design.

Nowadays a widely accepted approach in the process industry is the one described in the IEC EN 61508, whose major idea is that the safety of systems must be studied and pursued from the early design by risk analysis tools; one of its main activities is to define the Safety Instrumented Functions (SIFs) that must be further and deeply analysed in order to understand the effective risk reduction needed and the necessity to implement them in terms of safety systems and in terms of additional safety requirements. Functional safety assessment in the context of IEC EN 61508 constitutes a milestone for safety to drive the design (IEC EN 61508, 2005). The IEC EN 61513 provides requirements and recommendations for the overall I&C architecture of a Nuclear Power Plant (NPP) which may contain both hard-wired and computer-based technologies; it aims at translating the general requirements of 61508–1, 61508–2 and 61508–4 for nuclear application sector and, similarly to the IEC EN 61508, it introduces the concept of a safety life-cycle for both the whole architecture and the individual system, highlighting the relations between the safety objectives of the NPP and the requirements for the I&C architecture (IEC EN 61513, 2013). Nevertheless, the need to maintain the traditional safety approach for nuclear applications makes the 61513 misrepresenting the nature of the 61508; instead, it represents an intermediate step included into a rigid process that was developed for and it is still suitable to LWRs, but difficult to apply to concepts of the next generation. Hence, the philosophy of the 61508 can inspire the safety assessment of advanced nuclear plants, but redefining the strict framework defined for LWRs.

A schematic representation of the two approaches is shown in Figure 1.

Other methodologies, such as the Integrated Safety Assessment Methodology (ISAM) and the International Project on Innovative Nuclear Reactors and Fuel Cycles (INPRO) are always inspired by the IAEA general principles but at the same

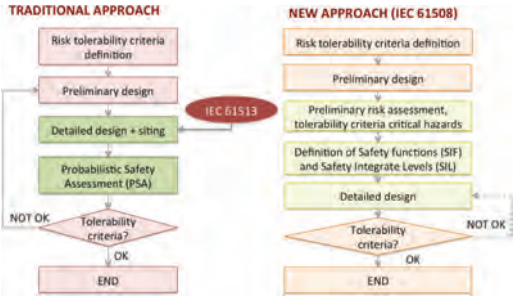


Figure 1. Schematic of the traditional and the new approaches to safety assessment.

time aim at implementing the concept of a safety driven design. The ISAM (RSWG of the GIF, 2011) is meant to combine both probabilistic and deterministic tools, both quantitative and qualitative methods and evaluations, some focusing on high-level issues, others on more detailed issues. It aims at providing a robust guidance, based on a good understanding of risk and safety issues, contributing to the achievement of Generation IV safety objectives. The INPRO assessment (IAEA, 2008) is a stepwise approach with a hierarchic structure: Basic Principles (BP), User Requirements (UR) and Coordinated Criteria (CC), which must be fulfilled by an Innovative Nuclear System (INS) to determine if the system is sustainable or not. This approach aims at providing a tool to analyze a nuclear installation in order to:

- Evaluate if it is compatible with the objective of sustainable energy development,
- Compare different plants or components to find a preferred or an optimum solution tailored to the needs of a specific region or a State and
- Identify possible improvements.

These two approaches represent guidelines that must be reviewed, completed and adapted, when needed, also using traditional risk analysis tools in order to better suit the unique case of each of the next generation nuclear plants. They define an inspiring philosophy but do not constitute an operational framework, which still has to be defined for advanced concepts through tailored criteria, requirements and consolidated operational safety assessment methods.

4 PERSPECTIVES OF SAFETY ASSESSMENT METHODOLOGIES

4.1 *Risk metrics*

Each nuclear plant must fulfil Quantitative Health Objectives of individual risk (QHO), used as a

basis for determining whether a level of safety ascribed to a plant is consistent with the safety goal policy (Whipple, 2012). In the LWR framework these objectives are embodied by LERF and CDF, which may not be consistent for some of the new generation nuclear plants, therefore they shall be reviewed (INL 2011). Although ISAM tries to adapt the CDF definition to all kinds of reactors, in some cases, especially those precluding the core damage states defined for LWRs, this results problematic and it may be appropriate to use a set of the risk-metrics that have the capability to define the significant contributions to risk and provide information to demonstrate defense in depth adequacy. The proposal for advanced reactors needs to be TI-RIPB (INL, 2017):

– Technological-Inclusive (TI), namely applicable to any design independently from the implemented processes;
– Risk-Informed (RI), since each decision must be an opportune derivation of both probabilistic and deterministic principles;
– Performance-Based (PB) because the risk and safety analysis lead to the formulation of performances requirements of Structures, Systems and Components (SSCs) in order to avoid accidents, or at least mitigate them.

Some proposed indexes include the frequencies of event sequences grouped in accident families having the same plant response and the same offsite radiological consequences, the integrated risk of a given consequence (e.g. site boundary dose), the individual fatalities (as compared to the existing limits for LWR) and the cumulative frequency of an early or latent effect. Moreover, some specific risk metrics can be defined for each reactor, depending on the specific characteristics.

These values can be expressed in the form of mean values and uncertainties percentiles (5th and 95th percentiles) and compared to the frequency-consequence evaluation criteria, as the one defined in Figure 2.

Furthermore, the integrated risk evaluation of the entire plant is performed taking into account four evaluation criteria (INL, 2017):

– the total frequency of exceeding a site boundary dose of 100 mrem shall not exceed 1/plant-year according to the annual exposure limits in 10 CFR 20;
– the total frequency of a site boundary dose exceeding 750 rem shall not exceed 10^{-6}/plant-year according to NRC Safety Goal Policy Statement on limiting the frequency of a large release;
– the average individual risk of early fatality within 1 mile of the Exclusion Area Boundary (EAB) shall not exceed 5×10^{-7}/plant-year according to the NRC Safety Goal QHO for early fatality risk;
– the average individual risk of latent cancer fatalities within 10 miles of the EAB shall not exceed 2×10^{-6}/plant-year according to NRC safety goal QHO for latent cancer fatality risk.

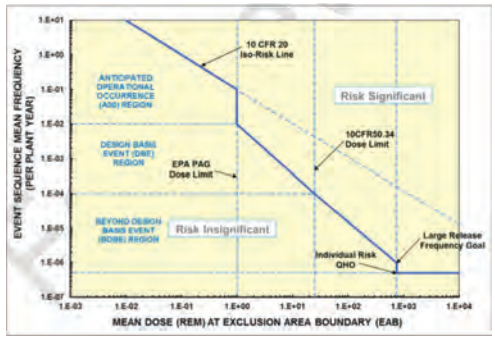


Figure 2. Frequency/consequence evaluation criteria (INL, 2017).

It is worth to note that the traditional classification of PSA Level 1, 2 and 3 starts from the concept of CDF and LERF, therefore the update of the risk metrics implies a new modernized PSA concept.

According to the "Basis for the Safety Approach for Design & Assessment of Generation IV Nuclear Systems" (RSWG of the GIF, 2008), one of the objectives to be pursed for the advanced designs optimization is their rationalization by the deliberate adoption of the ALARP (As Low As Reasonable Practicable) principle applicable to the full spectrum of design conditions. The UK Health and Safety Executive (HSE) defined an ALARP region (or tolerability region) between the acceptable and unacceptable risk regions: the comparison between advantages and disadvantages (i.e. between the reduction in risk and the cost of achieving it) on a quantitative or qualitative basis establishes what is "reasonably practicable" to be carried out in order to reduce the risk (HSE, 2001). This optimal risk reduction is translated in the implementation of innovative provisions looking for further risk reduction (prevention of the initiators and consequences mitigation) on a cost-benefit basis (RSWG of the GIF, 2008). In a frequency-consequence graph the ALARP area is usually represented by a range of values, but in Figure 2 it degenerates in a line. Since the uncertainties characterizing both the considered designs and their analyses, the ALARP principle should represent a key point for the definition of the acceptability criteria.

4.2 *Iterative risk assessment*

The risk assessment process for an advanced nuclear plant is proposed to be iterative rather than serial (as for the PSA Level 1,2,3) so that it can be introduced at an early stage of design and be consistent with the level of detail of the evolving design and successively with the site characteristics. Moreover, it is supposed to provide a logical and structured method to guide the design and evaluate its safety characteristics in a systematic and exhaustive manner. A very preliminary PSA is introduced when the reactor design is still conceptual: it is focused on internal events involving radioactive sources and it is simplified according to the level of knowledge regarding the definition of the design and physical phenomena occurring in the reactor. The challenges for the reactor are defined in terms of Initiating Events (IEs), with the correspondent plausible causes and consequences. The traditional list of IEs defined for the LWRs can inspire the advanced reactors ones, but it cannot be exhaustive and sometimes it is not even coherent; different technologies and phenomena have to be analyzed and can produce a completely new list of accident initiators, as in the case of fusion device (Pinna, 2017). The events leading to similar "reactor end-state" will be grouped together and the event involving the worst consequences will be selected as Postulated Initiating Event (PIE) to represent the entire group. The first group of PIEs is identified at a sufficiently early stage of the design to enable the designer to select the events worth to be considered to enhance the plant safety. At this level, methods following a functional approach can be used, whose suitability is assessed also in non-nuclear standards (IEC EN 61508, 2005). As the design matures and more design details become available, the set of PIEs will be updated and broadened to gradually address other plant systems and operational states. At the same time, the selected events will be studied through deterministic analyses in order to define more accurate events sequences. When the deterministic inputs are modified, the design changes and the PSA model evolves as well. Progressively all the internal hazards will be included (not only radioactive releases but also, for example, internal fire and floods) and, when the site characteristics become available, also external hazards (e.g. earthquake) can be taken into account. At the end, the analysis can be refined introducing information about human factor (Southern Company, 2017). This approach is expected to converge faster to a successful design rather than try to adapt and satisfy the LWR requirements.

4.3 *Preliminary PSA*

The PSA model begins with a systematic search of initiating events. Since the preliminary design stage of some of the new generation plants, a functional approach has been selected, suitable to define possible accident initiators when a sufficient design detail is not yet available to allow more specific evaluations at the component level. This methodology has already been applied on fusion devices, in particular to analyse the Primary Heat Transfer System (PHTS) of EU DEMO, with a WCLL (Water Cooled Lithium Lead) and a DCLL (Dual Coolant Lithium Lead) breeding blanket (Pinna, 2017), (Carpignano, 2016).

In order to identify functional deviations able to compromise system safety (in terms of Postulated Initiating Events, PIEs) as completely as reasonable, two approaches can be implemented at the same time: the Functional Failure Mode and Effect Analysis (FFMEA), a bottom-up approach, that focuses on the identification of the functions of the system and on the analysis of the consequences of the loss of each of them and the Master Logic Diagram (MLD), a top-down approach, that after the selection of a top event identifies its possible elementary causes. A list of PIEs is completed and for each of them a brief description of plausible causes, consequences, involved components, preventive and mitigation actions is provided. In addition to the identification of PIEs, this approach allows identifying lack of information on some systems, procedures or phenomena, to point out the potential limitations of the design and to make suggestions to enhance the safety of the concept. An example of application of this methodology is shown in (Uggenti, 2017). The complexity of the application of this methodology is reflected in the number of listed Postulated Initiating Events, between 25 and 30 for the three analysed systems, derived from an FFMEA of around 1000–1200 lines.

Successively, each accidental scenario has to be classified into frequencies and consequences severity macro-categories. Accordingly to (INL, 2017) the event sequences include relatively frequent events classified as Anticipated Operational Occurrences (AOO, with a frequency higher than 10^{-2} events per plant year), infrequent events classified as Design Basis Events (DBE, with a frequency between 10^{-4} and 10^{-2} events per plant year) and rare event classified as Beyond Design Basis Events (BDBE, with a frequency lower than 10^{-4} events per plant year). The severity of the consequences can be evaluated in terms of release of radioactive material (since the preliminary design, it can be evaluated in percentage with respect to the total amount) or in terms of damages to the asset (taking into account the possibility to restart the system immediately or after a while or its impossibility).

One or more risk matrices can be built using consistent definitions of technological-inclusive

risk metrics and severe accidents: according to the risk level of each unprotected accidental sequence, a number of provisions (or lines of defence) needs to be defined. This number is then compared to the number of existing barriers already present in the preliminary design and eventually suggesting new provisions to accomplish the requirements and to help sketching the final architecture of the system, using a more traditional approach.

5 CONCLUSIONS

For non-LWRs, the frequencies of accidents involving release of radioactive material may be very small and even those accidents with releases may involve very small source terms compared with releases of LWRs core damage accident. Therefore, the total risk may be very small (Southern Company, 2017). Nevertheless, it is necessary to understand the principal risk contributors in order to try to reduce the risk sources at the early design phase (Van der Börst, 2001): risk importance measures can be defined for any kind of risk metrics and it may be useful to calculate the risk significance both in relative and absolute basis, comparing it against risk goals rather than only against baseline risks.

The evaluation of the sources of uncertainty has to be performed without delay: uncertainties need to be evaluated both for frequencies and consequences through the performance of quantitative uncertainty analysis, where information is available to perform this function, and sensitivity analyses, to address other sources of uncertainty that are more difficult to quantify. To this aim, these uncertainties have to be considered in the frequency-consequence evaluation criteria. This uncertainty treatment then becomes an input to a risk-informed evaluation of Defence in Depth (DID).

A major limitation of preliminary nuclear risk assessment is due to the fact that all the efforts are concentrated on the nuclear island, while the remaining "traditional" components, for example all the components constituting the Balance of Plant, and the siting are usually only sketched: it is common to ignore the precise architecture of a system or the number of its redundancies, increasing the source of uncertainties in the risk evaluation: a design process to connect the research approach to a more engineering approach would be necessary to increase the realism and the accuracy of the safety evaluations.

In conclusion, it is worth to highlight that, due to their standardization, the LWRs safety assessment, and consequently their safety architecture, is prescriptive (what to do) or proscriptive (what to avoid doing), since historically the safety process standards are rules-based. The variegate nature of the next generation nuclear facilities imposes safety process standards to be simple and based on stable, general principles, as suggested by the IEC EN 61508, already implemented in some process industry sectors involving many diversified plants and technologies (e.g. Oil & Gas, chemical plants). Goal-based standards focus on the final objective of the safety assessment (what is necessary to achieve) and suits to new technologies (not only nuclear ones) better and more cost-effectively, by exploiting all the potentialities and the versatilities of the risk analysis.

REFERENCES

Carpignano A. et al., 2016, Safety issues related to the intermediate heat storage for the EU DEMO, Fusion Engineering and Design, 109–111 (Part A): 135–140.

Center for Chemical Process Safety (CCPS), Guidelines for Hazard Evaluation Procedures (2nd Edition), American Institute of Chemical Engineers, New York: 1992.

GIF, 2014, Technology Roadmap Update for Generation IV Nuclear Energy Systems.

EUROfusion, Programme preparing for ITER and developing DEMO, EUROfusion website available on https://www.euro-fusion.org/programme/.

Fleming K., Removing a Blind Spot in Our Safety Culture, presented at the to the International Topical Meeting on Probabilistic Safety Assessment and Analysis (PSA2017), Pittsburgh PA, USA.

IAEA, 2006, IAEA Safety Standard for protecting people and environment, Fundamental Safety Principles, No SF-1.

IAEA, 2008, Guidance for the Application of an Assessment Methodology for Innovative Nuclear Energy Systems, INPRO Manual-Overview of the methodology.

IEC EN 61508, 2005, Functional safety of electrical/electronic/programmable electronic safety-related systems, Parts 1÷7.

IEC EN 61513, 2013, Nuclear Power Plants—Instrumentation and control important to safety—General requirement for systems.

INL, 2011, Next Generation Nuclear Plant Probabilistic Risk Assessment Whit Paper, IN/EXT-11-21270.

INL, 2017, Modernization of Technical Requirements for Licensing of Advanced Non-Light Water Reactors, Selection of Licensing Basis Events.

Kelly T. & McDermid J. & Weaver R., Goal-Based Safety Standards: Opportunities and challenges, York University website available on https://www-users.cs.york.ac.uk/tpk/ISSC23.pdf.

Lee S., 2016, New Reactor Regulatory Activities Current Status for LLWR, SMR and non-LWRs, USA Nuclear Regulatory Commission.

Pinna T. et al., 2017, Identification of Accident Sequences for DEMO Plant, *Fusion Engineering and Design*, 124: 1277–1280.

RSWG of the GIF, 2011, An Integrated Safety Assessment Methodology (ISAM) for Generation IV Nuclear Systems, version 1.1 2011.

RSWG of the GIF, 2008, Basis for the Safety Approach for Design & Assessment of Generation IV Nuclear Systems.
Southern Company, 2017, Modernization of Technical Requirements for Licensing of Advanced Non—Light Water Reactors Probabilistic Risk Assessment Approach, SC-29980–101 Rev A.
Uggenti A.C. et al, 2017, Preliminary functional safety assessment for molten salt fast reactors in the framework of Samofar project, presented at the to the International Topical Meeting on Probabilistic Safety Assessment and Analysis (PSA2017), Pittsburgh PA, USA.
UK HSE, 2001, Reducing Risk, Protecting People, Sudbury: HSE books.
USA NRC, 2017, Part 50 – Domestic Licensing of Production and Utilization facilities.
Van der Börst & Shoonaker H., 2001, An Overview of Risk Importance Measures, *Reliability Engineering and System Safety*, 72: 241–245.
Whipple C., 2012, De Minimis Risk, New York: Springer Science & Business Media.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Scenario dependency of safety targets for platform doors

B. Hulin
NTC-Systems GmbH, Gilching, Germany

ABSTRACT: Platform barriers for railways shall protect passengers from different events like being crushed by a train or falling off the platform onto the track. Passenger can access a train through automatically operating platform doors that are integrated into the platform barriers. From a safety perspective, platform doors are an electronically controlled system, whose risks need to be analysed and reduced to an acceptable level. One of the most discussed points, in that relation, is the allocation of safety (design) targets for different functions of platform doors. This paper proposes the application of SIRF for the determination of safety targets for functions of platform doors. Beside a theoretical reasoning for using SIRF the paper gives examples for its application to platform doors. Especially, the hazard 'vehicle starts moving and doorways are open' is analysed for its criticality and the related functions are assigned with safety targets. It is shown, that this process highly depends on the scenario even if the starting conditions are equal. This leads to the conclusion that all scenarios for the same situation have to be analysed. The conclusion is that SIRF can be applied to platform doors easily, and delivers reasonable results.

1 INTRODUCTION

Access points of many recent transportation systems are limited by special barriers like platform barriers with platform doors. There is especially the need for such barriers within automated transportation systems like unmanned people movers or metro systems (see EN 62267 (CENELEC 2009)). Platform doors are usually implemented as screen doors that are automatically opened for boarding a train.

Platform barriers as well as their integrated platform doors are for protecting passengers from falling off the platform or being struck or crushed by a train. Insofar, platform barriers including their platform doors shall reduce the risk of injuries or deaths.

From a safety perspective, platform doors are an electronically controlled system that needs to be assessed for risk. In this context, the safety related functions with their safety targets are to be determined.

Safety targets (see CLC/TR 50451 (CENELEC 2007)) or design targets (see Regulation 2015/1136 (EU 2015)) can be defined amongst others as TFFR, THR, MTBF or SIL. A safety target is allocated to a functional safety requirement. The safety target of a functional safety requirement is the maximum criticality of all hazards related to it.

For the determination of safety targets for functional safety requirements there is available a huge amount of methods such as risk graph, risk matrix, and so on (Summers 1998). Each method has its own pros and cons and is best suited for some domains or some applications.

A good method for the determination of safety targets for railway vehicles is SIRF (EBA 2012)[1]. SIRF (Sicherheitsrichtlinie Fahrzeug) is a German tailoring of the EN 50126 (CENELEC 1999) for safety assessment for functions of railway vehicles. It was first released in June 2011 by the German national railway safety authority and has been applied successfully in many projects in Germany and Austria for main line and urban railway vehicles (e.g. metros, tramways and people movers).

As this method is used for the determination of safety targets of functions all over the railway vehicle including vehicle doors, the author argues that this method can be applied for platform doors, too, even if they do not belong to the structural subsystem vehicle.

This paper discusses this argumentation and concludes that an application of SIRF for platform doors is possible and reasonable. The main part of this discussion are examples of the hazard 'vehicle starts moving and doorways are open' with different situations and accident scenarios.

Since this is the first publication that applies SIRF to platform doors the method is described, first of all.

1. SIRF is freely available at www.eba.bund.de.

2 METHOD

2.1 *Generic process*

SIRF assumes initially a well defined system in which the main functions of this system are defined[2]. Starting from these functions a Functional Hazard Assessment (FHA) is conducted (Milius & Gayen 2004). For this, railway experts combine function failures with different operational situations and scenarios (Einer & Käser 2004). The result of the FHA are hazards with an estimation of its criticality. Then, a function receives the highest criticality of all hazard that are related to this function.

2.2 *Terminology of SIRF*

For the determination of safety targets, SIRF uses five parameters[3].

- S_A – number of affected persons
- S_V – degree of injury
- W – probability of the occurrence of the expected severity[4] after a function failure
- E – mean duration of exposure to a hazard
- V – possibility of avoidance of the severity of a harm by the person at risk, after the occurrence of the primary hazard

As described in SIRF, parameters S_A and S_V shall be estimated for a realistic worst-case outcome of a primary hazard within the considered scenario. Their combination $S = S_A \cdot S_V$ is an estimate of the expected outcome expressed by a severity of harm.

Parameter W is alternatively often referred to as inevitability of the transition from a function failure to the related severity of harm. Since the severity is scenario dependent and parameter W refers to a severity of harm, W is scenario dependent, too. An ontological analysis of the parameters can be found in (Hulin et al. 2016). Note, that SIRF-parameter W has a different meaning to the parameter W of the risk graph.

2.3 *Estimation of hazard criticality*

The parameters mentioned above are estimated qualitatively by experts. For that, SIRF defines certain values for each parameter.

The combination of these values results in a value which is called indicator. The calculation of this indicator I is carried out using equation 1.

Table 1. Mapping table.

Interval of indicator I	Safety target
]0; 21[	SIL 0
]21; 36[	SIL 1
]21; 72[	SIL 2
]72; 122[	SIL 3
]122; 281[	SIL 4

$$I = \frac{S_A \cdot S_V \cdot W \cdot E}{V} \tag{1}$$

Then the mapping of the inidicator to a safety target is done according to Table 1.

3 APPLICATION TO PLATFORM DOORS

3.1 *Justification*

Even if SIRF is defined only for railway vehicles some projects have shown its suitability for some other railway subsystems, especially platform doors and emergency lighting in tunnels. The reason for the application of SIRF to non-vehicle-functions was to use as less different risk assessment methods as possible in projects in which a complete railway system (consisting of vehicles, tracks, platform barriers, power supply and so on) was installed.

From a theoretical point of view, the functions of vehicle access doors are nearly the same as for platform doors. Amongst others these are "provide external access" to the vehicle (see EN 15380-4), "ensure exiting" the vehicle (see DIN 25002) and "provide passenger emergency exits from inside the vehicle". A special function for vehicle doors is "keep doors closed between two stations". A special function of platform doors could be "provide emergency exit from the track to the platform". Moreover, there exist some interface functions like "align train doors with platform doors".

Most of potential accidents that could happen with platform doors are similar to those of vehicle access doors. For example 'being crushed be the door leaves' or 'falling down through an open door' are an potential accidents for both types of doors.

Moreover, hazards like 'unintended door opening' or 'too fast closing of doors' are equal for both vehicle access doors as well as platform doors.

3.2 *Example 1: One open doorway*

3.2.1 *Initial situation*

All scenarios in this section start from the same operational situation: A driverless passenger train is standing at the platform. At the edge of the

2. A good overview of functions for railway vehicles can be found in EN 15380-4 and E DIN 25002.
3. Since SIRF is a German directive without an official and agreed English translation, we translated the definitions of terms for safety target determination on our own.
4. Severity S is the product of S_A and S_V.

platform a barrier containing platform doors is installed. All doors of the platform as well as of the vehicle are sliding doors with a passage width of 0.8 *m*. With this width it is possible that at most two persons can stay simultaneously in one door. Due to efficiency of air conditioning in the railway vehicle all doors are closed independently of each other as soon as possible and a door is opened only on passenger request. All platform doors as well as all train doors are closed except one doorway[5]. Travellers are boarding the train through this one open doorway.

3.2.2 *Hazard and potential function failure*

The hazard which is considered in this section is 'one doorway is open and vehicle starts moving'[6].

In this case, the potential accident 'shearing of persons' can occur. For the accident it does not matter which potential function failure is the cause for the hazard—e.g. the 'vehicle starts moving with one open doorway', 'one doorway remains open while the vehicle starts moving', 'omission to prevent persons passing through the one doorway while vehicle accelerates' or a 'door interlock signal is sent while one doorway is open'.

These potential function failures can be part of the vehicle, of the platform door system, of the control system or of a combination of some of these systems.

For this example, the potential function failure 'vehicle starts moving with one open doorway' is analysed.

3.2.3 *Analysis of accident scenarios*

The probability of shearing a person between stationary installations and vehicle parts is higher with than without platform barriers since only small vehicle movements are necessary to shear a person.

There are several accident scenarios with shearing of persons within the afore mentioned context. While one person looses just his arm another person could be killed. As mentioned before, due to the doorway width of 0.8 *m* it is imaginable that one or two persons are staying at the doorway simultaneously. This variability of accident scenarios can cause different outcomes. For each outcome (specified as a severity of harm) the conditional probability W of this outcome after the potential function failure 'vehicle starts moving with one open doorway' has to be estimated separately.

For the severity 'death of person' after the assumed potential function failure, parameter W is estimated conservatively with 'high' since in the largest part of the time period in which the one doorway is open we assume persons passing through this doorway. The reason for this assumption is that the doors are just opened on passenger request and passengers yield such a request for boarding or exiting and doors will be closed right after the doorway is cleared. Moreover, the time period for a person entering the vehicle usually increases with the number of persons standing in the corridors and the entrance area of the vehicle.

In the scenario that several persons are killed by this potential function failure we estimate parameter W with 'low' since it is very seldom that two persons are staying in the doorway at the same time. An affecting of many persons by only one doorway is impossible and, consequently, for these cases parameter W is assigned as 'incredible' ($W = 0$). A complete overview of all estimations of probability W of a certain outcome can be found in Table 2.

For all accident scenarios that relate to the hazard 'one doorway is open and vehicle starts moving' the exposure to the hazard is estimated with 'low' ($E = 1$) and the avoidance is seen as 'not or nearly not possible' ($V = 1$). The reason for the estimation of parameter E is that persons are just exposed to the hazard during boarding and exiting. The estimation of parameter V results from the consideration of the time period for an effective human reaction. It can be that due to the starting train movement persons are shocked or fall down and thus are suddenly limited in the ability to react effective.

The graph of the SIRF indicator is shown in Figure 1. Neither direct nor indirect proportionality between severity and the indicator is given. There can exist several local minima and maxima in such a graph. Consequently, a certain SIL can result for different severities. Moreover, it can be that the highest SIL of all credible scenarios starting

Table 2. Potential outcomes of hazard 'one doorway is open and vehicle starts moving'.

Number of persons	Degree of injury	S	Prob.	W	SIRF Indic.
No	none	0	high	3	0
One	minor injury	6	high	3	18
Several	minor injury	10	low	1	10
One	serious injury	12	high	3	36
Many	minor injury	16	incredible	0	0
Several	serious injury	20	low	1	20
One	death	27	high	3	81
Many	serious injury	32	incredible	0	0
Several	death	45	low	1	45
Many	death	72	incredible	0	0

5. A doorway consist of both a platform door and a corresponding vehicle access door. A doorway is called open if both doors are open.

6. The term 'vehicle' is more generic than the term 'train'.

from one situation to one potential accident is not the scenario with the highest severity. In Figure 1 the highest credible severity is 45 when several persons are killed since a severity $S = 72$ with many deaths is not credible with one open doorway. For a severity $S = 45$ SIRF returns a SIL 2 while for a severity $S = 27$ it returns SIL 3.

3.3 *Example 2: Variable number of open doorways*

3.3.1 *Description*

Let us now analyse the values of the SIRF indicator in dependence on the number of open doorways. The basis for this example is the same initial situation as described in section 3.2.1 for example 1 except that the number of doorways is variable. Therefore, the hazard is 'doorways are open and vehicle starts moving' with the function failure 'vehicle starts moving with open doorways'.

The potential accident 'shearing of persons' is the same as for example 1.

3.3.2 *Analysis of accident scenarios*

Of course, with increasing number of open doorways the worst case severity increases, too. More interesting is the relation of the number of open doorways to SIRF parameter W and the SIRF indicator for an invariant severity.

For that, we analysed three different severities: severity $S = 72$ is for deaths of many persons, severity $S = 45$ for deaths several persons, and severity $S = 32$ for many seriously injured persons. SIRF parameters E and V remain equal to example 1 with $E = 1$ and $V = 1$. Parameter W is the only parameter that changes with the number of open doorways for a constant severity. If less than 6 doorways are open it is incredible that many persons[7] are killed and, therefore, W is equal to 0 and the SIRF indicator is equal to 0, too. With 22 or more open doors the probability for the affection of 'many' persons is very probable and thus parameter W is assigned with 3. The death of 'several' persons[8] is classified as 'very probable' ($W = 3$) for 4 or more open doorways. This results in a SIRF indicator of 135 which can be mapped to SIL 4.

The relation of the indicator and the number of open doorways are shown in Figure 2.

The graphs that show the relations for many affected persons climbs at the same positions (see Figure 2) whereas the graph which is based on the deaths of several persons climbs in other positions. These different gradients produces an intersection

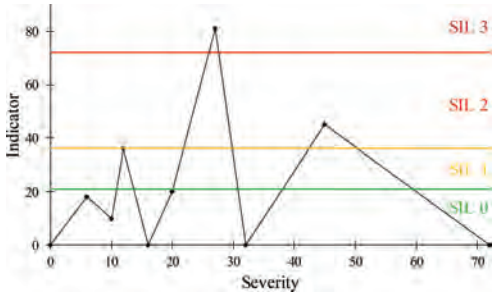


Figure 1. Relation between SIL an severity.

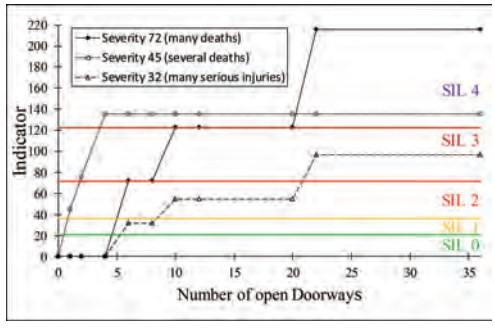


Figure 2. Relation between SIL and number of open doorways.

between graphs for severity $S = 72$ and for severity $S = 45$.

In this example, this intersection is meaningless for the SIL allocation since SIL does not change. However, for other graphs or other scenarios such intersections are indications to a change of priority of accident scenario for the determination of criticality and thus for SIL allocation.

4 COMPARISON OF RESULTS

The French national railway safety authority (called EPSF) claims the safety target TFFR $\leq 10^{-7}$ for the function failure 'authorized traction with one or more open doors' (see SAM C 305 (EPSF 2013)). This safety target which corresponds to a SIL 2 is for main lines without platform doors. Since, the situation is quite different without platform doors, safety targets are not easily comparable to the results of the examples of this paper.

A safety target of SIL 3 for platform doors functions whose failures can cause deaths is presented in Lecomte (2008). However, he just does not deduce it.

Due to the European Regulation 2015/1136 (EU 2015) the function failure 'vehicle starts moving

7. SIRF defines 'many persons' with 'more than 10 persons'.
8. SIRF defines 'several persons' as more than 1 but at most 10 persons.

with one open doorway' is assigned to a target frequency TFFR ≤ 10^{-7}. The reason for that is that the resulting potential accident is classified as 'critical accident' which is 'typically affecting a very small number of people and resulting in at least one fatality' (see Article 1 of (EU 2015)). For this function failure the safety target determined with SIRF in example 1 (see section 3.2) is a little bit more restrictive.

Moreover, Regulation 2015/1136 (EU 2015) defines 'catastrophic accidents' that are 'typically affecting a large number of people and resulting in multiple fatalities'. For function failures that could lead to such accidents 'an occurrence of failure at a frequency less than or equal to 10^{-9} per operating hour' is required. The regulation, however, does not define what a 'large number' and a 'very small number' of people means. This is defined in Jovicic (2017). In this guideline Jovicic (2017) estimates for the function failure 'train moves off at station with one bodyside door open' for the situation 'more than one door open' during passenger transfer TFFR ≤ 10^{-9}.

This safety target of Jovicic (2017) is equal to SIRF for trains that start moving with four or more open doors. For two or three open doors, however, the safety target of Jovicic (2017) is more restrictive than that of SIRF with SIL 3 (see Figure 2 and the explanation of example 2).

5 CONCLUSIONS

In this paper SIRF is applied to platform doors for safety target determination. It is shown that a determination is possible and useful with SIRF. Compared to safety targets of railway doors and European Regulations results are similar.

Some specialities of SIRF for safety target determination has been discussed in this paper. One very helpful speciality to better shape severity is the consideration of two parameters S_A and S_V for it.

The safety target is highly dependent on the initial situation and the way an accident happens. The consequence is that the worst case scenario for a class of potential accidents does not lead necessarily to the highest safety target. Therefore, all scenarios within a class of potential accidents have to be considered.

REFERENCES

CENELEC (1999, Sep.). EN 50126-1, Railway applications – The specification and demonstration of reliability, availability, maintainability and safety (RAMS). Part 1: Basic requirements and generic process.

CENELEC (2007, May). CLC/TR 50451, Railway applications – Systematic allocation of safety integrity requirements.

CENELEC (2009, Dec.). EN 62267, Railway applications – Automated urban guided transport (AUGT) – Safety requirements.

EBA (2012, June). Sicherheitsrichtlinie Fahrzeug (SIRF).

Einer, S. & P. Käser (2004, Sep.). Sicherheitsanalyse von Bahnbetriebsprozessen. *SIGNAL + DRAHT 96*, 18–23.

EPSF (2013, Dec.). Sam c 305 – Systéme d'accés voyageurs.

EU (2015, July). Regulation 2015/1136 – common safety method for risk evaluation and assessment.

Hulin, B., H. Kaindl, T. Rathfux, R. Popp, E. Arnautovic, & R. Beckert (2016). Towards a Common Safety Ontology for Automobiles and Railway Vehicles. In *European Dependable Computing Conference*.

Jovicic, D. (2017, May). *Guideline for the application of harmonised design targets (CSM-DT) for technical systems as defined in (EU) Regulation 2015/1136 within the risk assessment process of Regulation 402/2013*. ERA.

Lecomte, T. (2008). Safe and reliable metro platform screen doors control/command systems. In *FM 2008: Formal Methods, 15th International Symposium on Formal Methods, Turku, Finland, May 26–30, 2008, Proceedings*, pp. 430–434.

Milius, B. & J.-T. Gayen (2004, Oct.). Functional Hazard Assessment der Luftfahrt im Vergleich zu Risikoanalysen der Eisenbahn. *Signal + Draht 96*, 23–31.

Summers, A.E. (1998). Techniques for assigning a target safety integrity level. *ISA Transactions 37*(2), 95–104.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

A general framework for integrated risk assessment of nuclear/non-nuclear combined installations on market-oriented nuclear industry

K. Kowal, S. Potempski & Pawel M. Stano
National Centre for Nuclear Research, Poland

ABSTRACT: Development of new nuclear technologies tends toward decentralized small-medium size and modular installations with parameters tailored to a specific applications, resulting from the needs of the market wider than the energy sector, e.g. the production of process heat, hydrogen or hydrazine, which is of great importance for chemical industry. The High Temperature Reactors (HTR) and Dual Fluid Reactors (DFR) are the examples of the attempts for building such industrial applications. However, the implementation of these concepts poses a challenge for safety assessment due to the interfaces between nuclear and non-nuclear parts of the installation, which were not taken into account within the hitherto completed safety studies. This is a driven force for development of new framework for integrated risk assessment of nuclear/non-nuclear combined installations. This article is an attempt to sorting out the most demanding problems related witch this issue and to indicate possible paths for the solutions.

1 INTRODUCITON

Seeing the intensive development of new nuclear reactors technologies over recent years, one can expect major changes in the widely understood nuclear industry. So far, the innovation in nuclear reactors has been induced mostly by the technology-push (i.e., public R&D expenditures) and the demand-pull (i.e., NPPs construction) incentives (Berthélemy 2012), but the main stakeholders were focused mostly on the nuclear power generation.

Current trends in development of the new reactor technology tend toward decentralized small-medium size and modular installations with parameters tailored to a specific applications (Locatelli, Bingham & Mancini 2014). These solutions are considered rather as an energy source at site for different industrial processes than the way to electricity production on an industrial scale. The main driving force behind these concepts is the reduction of the emission of greenhouse gases from industrial processes. It seems that the nuclear market will soon evolve towards greater fragmentation and wider field of applications from which the non-electric services will play an increasingly important role.

The general concept and key technological solutions for non-electric nuclear applications have been already developed. However, they have not reached the same industrial maturity as for the generation of electricity. Nevertheless, expecting the progression in this type of nuclear technology applications, the International Atomic Energy Agency performed the initial target market analysis which showed that there is an increased interest in non-electric applications facilitated by the recent development of advanced reactor concepts (IAEA 2002).

The market oriented restructuring in the nuclear industry requires, however, an accurate estimation of the costs and benefits of nuclear applications in comparison with the non-nuclear suppliers of similar services and, what is much more important, appropriate frameworks, methods and tools for integrated risk assessment of nuclear and non-nuclear installations combined together in one complex structure. The High Temperature Reactors (HTR) or Dual Fluid Reactors (DFR) are the examples of the attempts for building industrial applications based on the Generation IV technologies.

For example, in the nearest future the best opportunities for cogeneration will be application of HTR for the chemical industry (Jackowski et al. 2017). In this respect within NC2I-R (Nuclear Cogeneration Industrial Initiative—Research and Development Coordination 2015) a review has been made taking into account the following main processes compatible with HTR capabilities:

- refinery distillation steam,
- refinery distillation superheated steam,
- petrochemicals—reaction enthalpy,
- steam as utility for industrial complex,
- and paper steam (drying).

Use of DFR, in turn, is expected in the following industrial processes (Huke et al. 2015):

- mixed process heat and electricity generation,
- medical isotope production with high efficiency,
- the hydrogen-based chemistry e.g., production of synthetic fuels suitable for the vehicles,
- and radiotomic chemical production—utilization of intensive radiation for radiotomic induction of chemical reactions requiring high doses (Stannet & Stahel 1971).

Implementation of new technologies encounters, however, new problems, among others with safety demonstration (inadequate core damage definition, interfaces between nuclear and non-nuclear installations, etc.), licensing processes (inadequate legislation), and social acceptance (nuclear technology in the place of work, close to industrial centers). The paper aims to discuss these kind of issues as a part of general framework for the integrated risk assessment of nuclear/non-nuclear combined installations.

2 THE LICENSING PROCESS ORGANIZATION

Nowadays, the licensing process of the newly designed reactors, seems to be one of the most burning challenges. It have to be developed with respect to all specific features of the nuclear technology and related chemical installation. Preliminary analysis of the HTR licensing issues has been made within NC2I-R and HTR-PL projects with consideration of the Next Generation Nuclear Plant guideline (NGNP 2010). However, development of a new framework for integrated licensing process for joint nuclear-chemical installations is highly expected.

Figure 1 shows two models of the licensing process organization for joint nuclear-chemical installations. One of them assumes separation of paths leading to receive the operation permission for nuclear and chemical parts of the installation, while the second one is a proposal for integration. The structure of regulatory bodies, their competences and communication, as well as the scope of the safety report or reports must be specified in details to make it applicable and this is a challenge for further studies.

3 RISK ASSESSMENT OF NUCLEAR/NON-NUCLEAR COMBINED INSTALLATIONS

Implementation of the integrated approach to the licensing process requires, among others, the integrated risk assessment for the whole installation consisting of a nuclear and chemical parts. This requires, in turn, consideration of insights coming from the Quantitative Risk Assessment (QRA) developed for the chemical part of the installation and Probabilistic Risk Assessment (PRA) developed for the nuclear one, including analysis of interfaces, mutual reactions and interdependencies.

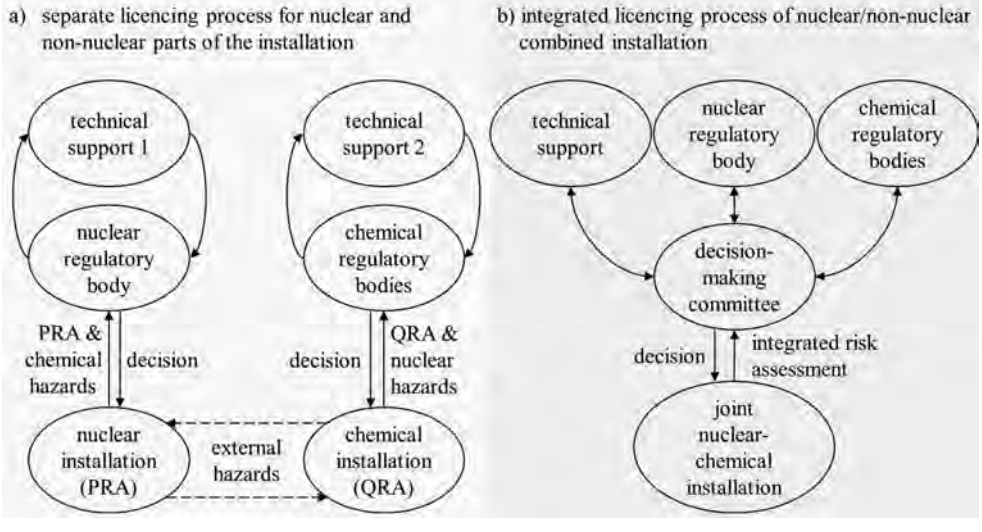


Figure 1. Two general models of the licensing process organization for the nuclear/non-nuclear combined installation; a). separation of the licensing processes for nuclear and chemical installations assuming different regulatory bodies and separate safety reports where the impact of the neighboring installation is treated as a set of specific (nuclear or chemical) external hazards; b). integrated licensing process where the object is defined as a joint nuclear-chemical installation and the integrated risk assessment is expected.

3.1 *Chemical QRA and nuclear PRA integration*

As a result of initiating event within the joint nuclear-chemical installation different systems on both parts (nuclear and chemical) can fail immediately or with a time delay. The time-sequence of the failures, however, is quite complex due to the interfaces between systems within each part of the installation separately and between the nuclear and chemical plants. Independently developed chemical QRA and nuclear PRA models do not describe properly the real state of the whole installation and thus need to be integrated. The integration of QRA and PRA models within the overall risk assessment framework can be proceeded with the following steps:

1. identification of postulated initiating events for the nuclear and chemical parts of the installation;
2. identification of systems which would be directly affected by the initiating events immediately after their occurrence or with a time delay;
3. identification of all possible interactions between the considered nuclear and chemical systems i.e.:
   a. internal interactions between systems within the nuclear installation;
   b. internal interactions between systems within the chemical installation;
   c. nuclear/chemical interactions (posing a challenge for safety of the chemical installation after failure of one or more nuclear systems);
   d. chemical/nuclear interactions (posing a challenge for safety of the nuclear installation after failure of one or more chemical systems);
4. specification of time-frames in which the whole installation remains in the specific states characterized by the systems affected and interactions within and between nuclear and chemical plant;
5. identification of all safety functions that must be performed in each time-frame and determination of their success/failure probability.

Figure 2 presents a simple example on how to deal with chemical QRA and nuclear PRA models developed for both parts of the installation. The following time-frames were established to define the periods in which a specific functionality is required:

- δt_0 – after initiating event occurrence and before the notification of effects on the nuclear systems;
- δt_1 – failure of nuclear system n_1 with possible or conditional effect on the chemical system ch_1;
- δt_2 – failure state of nuclear system n_1 and chemical system ch_1 (nuclear/chemical interaction);
- δt_3 – failure state of chemical system ch_1 and nuclear system n_2 (due to the internal interaction);
- δt_4 – failure state of chemical system ch_1 and nuclear system n_3 (due to the internal interaction);
- δt_5 – failure states of n_4 and n_5 nuclear systems due to the chemical/nuclear interaction witch system ch_1 or internal interaction with system n_3;

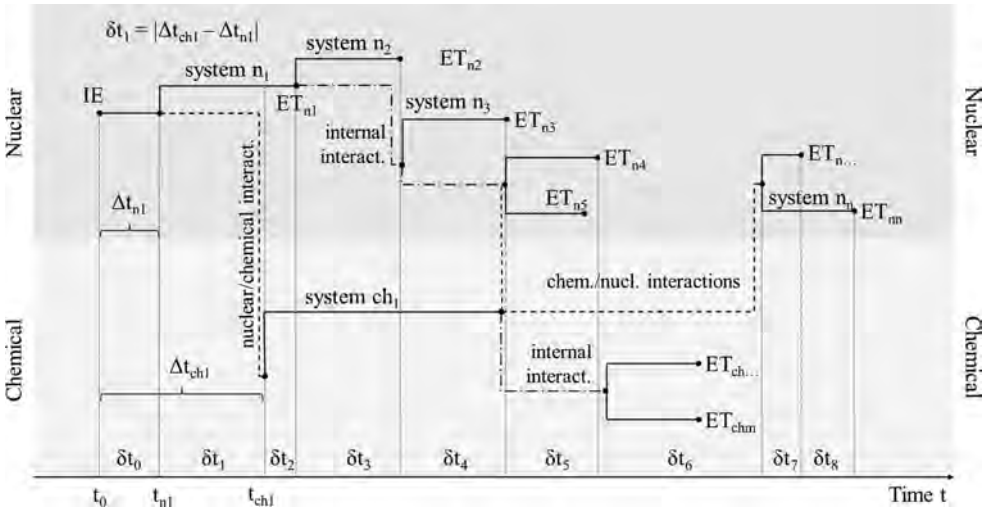


Figure 2. Example of integration of chemical QRA and nuclear PRA studies within the overall risk assessment framework for nuclear/non-nuclear combined installations: IE – Postulated Initiating Event; ET_{ni} – Event Tree for *i*-th system of nuclear installation; ET_{chi} – Event Tree for *i*-th system of chemical installation; δt_i – *i*-th time-frame to be considered in the integrated risk analysis.

- δt_6 – failure states of chemical systems due to the internal interactions in the chemical installation;
- δt_7 – failure state of nuclear system $n_{n...}$ due to the chemical/nuclear interaction with system ch_1;
- δt_8 – failure state of nuclear system n_n due to the chemical/nuclear interaction with system ch_1.

This concept has many advantages, among which the most important is possibility of modelling of a wide spectrum of failures sequences, including both nuclear and chemical parts of the installation, in response to the initiating event that occurred in one of them. Such approach, however, is not devoid of weaknesses among which the following should be mentioned here:

- numerically ineffectiveness of calculations based on large and complex failure three structures,
- multiple modelling of the same sequences of events appearing at different time frames,
- difficulties in adding new systems, interactions or time frames to the existing models.

3.2 *Block framework*

Apart from the failure tree approach presented above it is reasonable to consider alternative approach to modelling of failure sequences for combined nuclear-chemical installations that is based on the applications of Bayesian networks to risk assessment. An example of such a block framework is presented in Figure 3. This network corresponds one-to-one with the failure tree model discussed in Figure 2, and it is constructed from the following types of blocks:

- a collection of N initial events $(IE_1^n,\ldots, IE_N^n))$ defined for a nuclear plant and M initial events $(IE_1^{ch},\ldots, IE_M^{ch})$ defined for the chemical plant;
- a collection of n nuclear and m chemical systems potentially affected by previously defined initial events;
- logical gates that allow to introduce complex Boolean expressions.

The interactions between any two blocks i and j are defined by a pair of transition probabilities p_{ij}^{nch}

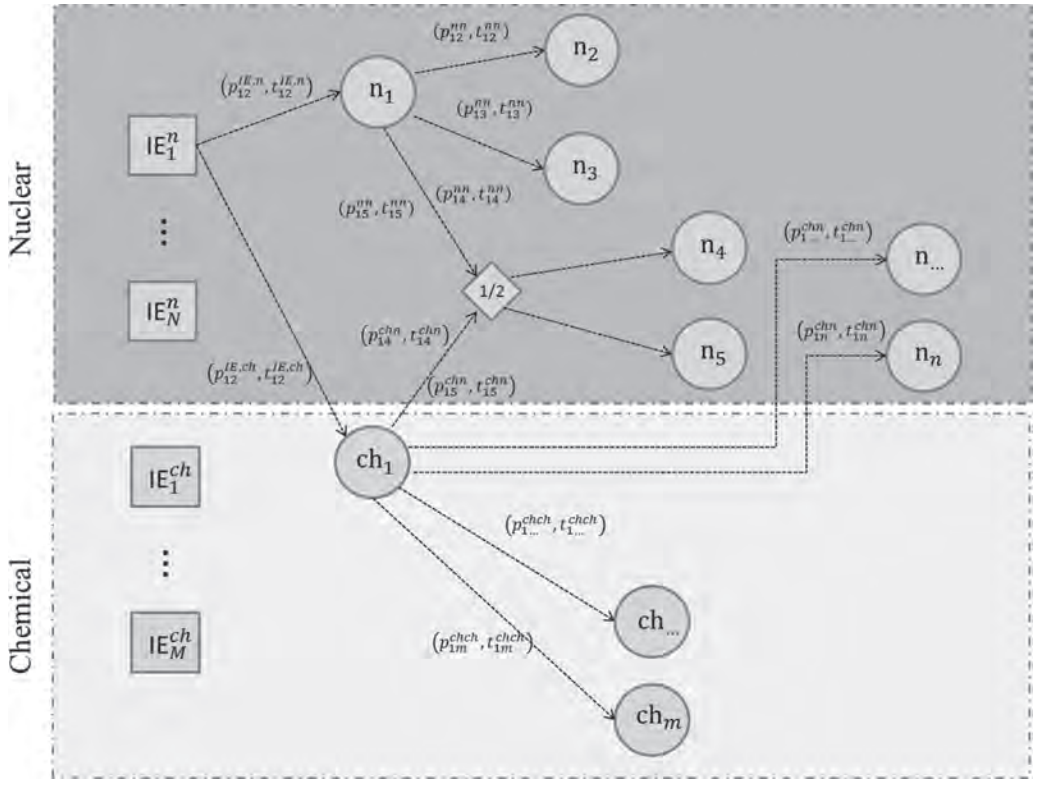


Figure 3. Block model corresponding to the failure tree described in Figure 2. For deterministic case, all the transition probabilities are set to 1. The transition times are set to match times in Figure 2, e.g., $t_{11}^{IE,n} = \delta t_0$; $t_{12}^{nn} = \delta t_1 + \delta t_2$; $t_{14}^{chn} = t_{15}^{chn} = \delta t_2 + \delta t_3 + \delta t_4$, etc.

and transition times t_{ij}^{nch} where the indexes in lower script indicate the direction of the connection (in this case from *i* to *j*) and the indexes in the upper script indicate to which type of installation the blocks belong (in this case *i* belongs to nuclear plant, whereas *j* belongs to chemical plant).

Such block framework has several advantages that might be particularly well suited to model joint nuclear-chemical installations. First of all, with such a model it is possible to model virtually all the combination of interactions between the systems in all the directions (back and forth). Secondly, merging independent block models into a single block model is relatively easy as it only requires defining new connections and transition parameters between systems belonging to distinctive block models and does not disturb the networks within the models.

This property is especially important for our considerations because thanks to it, it is possible to design safety models for nuclear plant, chemical plant, and interaction between plants independently and then combine them together only at the final stage of the analysis to obtain quantitative risk indicators. Furthermore, this property is scalable down, i.e., either plant can be broken down into smaller units, for which block models can be developed independently following the philosophy described above.

Finally, simulating block networks is very efficient numerically because the complete information about network can be stored in a simple matrix form. Consequently, computations of probabilistic risk measures require matrix algebra only. A general structure for the matrix of transition probabilities is presented below. The matrix of transition times is defined similarly.

$$\begin{bmatrix} \begin{bmatrix} p_{11}^{nn} & \cdots & p_{1n}^{nn} \\ \vdots & \ddots & \vdots \\ p_{n1}^{nn} & \cdots & p_{nn}^{nn} \end{bmatrix} & \begin{bmatrix} p_{11}^{nch} & \cdots & p_{1m}^{nch} \\ \vdots & \ddots & \vdots \\ p_{n1}^{nch} & \cdots & p_{mm}^{nch} \end{bmatrix} \\ \begin{bmatrix} p_{11}^{chn} & \cdots & p_{1n}^{chn} \\ \vdots & \ddots & \vdots \\ p_{m1}^{chm} & \cdots & p_{mn}^{chn} \end{bmatrix} & \begin{bmatrix} p_{11}^{chch} & \cdots & p_{1m}^{chch} \\ \vdots & \ddots & \vdots \\ p_{m1}^{chch} & \cdots & p_{mm}^{chch} \end{bmatrix} \end{bmatrix} \quad (1)$$

3.3 *Global uncertainty and sensitivity analysis*

In the context of HTR or DFR, where a nuclear facility is connected to chemical, from the perspective of safety analyst the structural differences between PRA (or QRA) models for both plants might be immense (in complexity, accuracy, purpose, etc.). Therefore, the uncertainty that is inherently associated with either of the model is further enlarged by the uncertainty about the way the plants, or their systems interact with each other. Consequently, for risk informed decision-making, a proper analysis of PRA models' results (as well as models' failures!) requires that the analysis of safety outputs should be accompanied by the identification of relevant uncertainties and the assessment of their impact on the final results. This is done by conducting uncertainty analysis (UA) and sensitivity analysis (SA) that aim to support decision-making by quantification of model uncertainties. The fundamental difference between the two is that the UA adopts a forward-looking approach, which is focused on investigating how the uncertainty in input variables (external uncertainty) and parameters (internal uncertainty) can affect the uncertainty of output variables. The SA adopts backward-looking approach, which is focused on investigating how sensitive the output variable is to fluctuations in uncertain input variables and parameters. Thus, in the face of prevailing un-certainties in models' parameters and input-output variables, the UA and the SA complement each other by looking at the same problem from two opposite directions.

In recent years a variety of methods have been developed to analyze models' uncertainty, from both UA and SA perspectives specifically to be used in the PSA (Borgonovo, Apostolakis, Tarantola, & Saltelli, 2003). At the same time, a number of spectacular failures in management of complex real-life processes, associated with nuclear industry, due to unrealistic uncertainty assessments have seen the light of day. Over four decades long stalemate around the Yucca Mountain nuclear waste depository is one example of the so-called "wicked problems" (Saltelli, Stark, Becker, & Stano, 2015), where uncertainty and disagreement about values affect the very framing of what the problem is and how to model it. Another example is handling the Fukushima Daiichi nuclear disaster, which provides a vivid illustration of how, in the face of "unknown unknowns" (Logan, 2009), safety assessments can become worthless in the blink of an eye. Better methods of ascertaining and managing model uncertainty are needed to realistically re-evaluate the safety features of the existing installations. Most importantly by paying more attention to structural uncertainty, which investigates issues such as: the selection of variables and processes to include in the model, how the variables and processes are described mathematically, how they interact, etc. When modelling complex phenomena, the structural uncertainty is likely larger source of uncertainty than the formerly mentioned uncertainty in input/output variables. Another open topic in uncertainty quantification is the assessment of human errors, especially in

highly stressful critical conditions during the accidents progressions.

In both PRA and QRA, which have been developed specifically to quantify various risks derived from operation of nuclear and chemical plant, respectively, the quantification of uncertainty is of crucial importance. However, despite much attention being devoted to studying uncertainty in the PRA context there exists no universally accepted standard for handling various types of uncertainties in a systematic way. Furthermore, in case of joint nuclear and non-nuclear installations all the aforementioned uncertainties are inherited from respective installations and further elevated by the model of interactions between the systems.

Although, many methods addressing structural uncertainty have been developed in recent years, this field of study is far from becoming mature. This is because the developed methods are usually highly subject-specific and it is not clear how they can be extrapolated and reliably applied to problems other than specified. For example, in the analysis of nuclear plants much focus is on prevention of accidents leading to core meltdown, while in the analysis of chemical plants the main attention is put on preventing fires and explosions. Each of these critical scenarios has its own typology, with different time scale, undesired effects, etc. and consequently with different types of uncertainties taken into consideration. It is not a surprise then, that each of these fields of scientific inquiry follows its more or less unique path of dealing with structural uncertainty and very little work have been done towards developing a global approach that would link structural uncertainty assessment with assessment of aleatory and epistemic ones in a synthetic manner.

Thus, to assure safety of joint nuclear and non-nuclear plants, a systematic approach needs to be develop that would allow to perform global UA and SA, i.e., taking into consideration the main sources of uncertainty for both plants jointly, but also investigate what are new sources of uncertainty that are due to interactions between the installations.

4 CONCLUSIONS

Many safety problems concern the mutual dependence of nuclear and chemical parts of the installations where the nuclear reactors are considered to be used as an energy source for various chemical processes. In order to enhance safety of such joint nuclear and non-nuclear installations, a systematic approach needs to be developed that would allow to perform integrated licensing process. Many efforts has to be made to accomplish this challenge. Development of a new framework for integrated risk assessment is one of them. In this article, two methods for integration of chemical QRA and nuclear PRA were proposed as an contribution to this task. The first one, based on the failure tree structure, is very informative, but indeed, not perfect and thus needed to be improved. The alternative approach in a form of block framework has more advantages and the authors believe that it can be applied in real studies.

REFERENCES

Berthélemy M., (2012) What drives innovation in nuclear reactors technologies? An empirical study based on patent counts, Cerna Working Paper Series 2012–01.

Borgonovo, E., Apostolakis, G.E., Tarantola, S., & Saltelli, A. (2003). Comparison of global sensitivity analysis techniques and importance measures in PSA. *Reliability Engineering & System Safety,* 79(2), pp. 175–185.

Brinkmann G. et al. (2006): Important viewpoints proposed for a safety approach of HTGR reactors in Europe. *Nucl. Eng. Des.*, vol 236, pp. 463–474.

Giorgio Locatelli, Chris Bingham, Mauro Mancini, (2014) Small modular reactors: A comprehensive overview of their economics and strategic aspects *Progress in Nuclear Energy* 73 (2014) 75–85.

Huke, A., et al. (2015) The Dual Fluid Reactor – A novel concept for a fast nuclear reactor of high efficiency. *Ann. Nucl. Energy.,* vol. 80, pp. 225–235.

International Atomic Energy Agency (2002) Market potential for non-electric applications of nuclear energy, IAEA Technical reports series No. 410, Vienna.

Jackowski, T., K. Kowal, S. Potempski (2017) Cogeneration: technologies, possibilities, challenges, *Proceedings from 53th ESReDA Seminar.*

Logan, D. (2009). Known knowns, known unknowns, unknown unknowns and the propagation of scientific enquiry. *Journal of experimental botany*, 712–714.

NGNP Licensing Basis Event Selection White Paper, September 2010.

Nuclear Cogeneration Industrial Initiative – Research and Development Coordination, Periodic Report, (2015).

Saltelli, A., Stark, P., Becker, W., & Stano, P. (2015). Climate Models as Economic Guides: Scientific Challenge or Quixotic Quest*? Issues in Science and Technology*, 31(3), 79–84.

Stannet, V.T., Stahel, E.P., (1971) Large scale radiation-induced chemical processing. *Ann. Rev. Nucl. Sci.*, vol. 21, pp. 397–416.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Assessment and management of ageing of critical equipment at seveso sites

M.F. Milazzo, G. Ancione & G. Scionti
Università degli Studi di Messina, Messina, Italy

P.A. Bragatto
INAIL, Monteporzio, Italy

ABSTRACT: The Directive Seveso III points towards the introduction of plans for a safe management of ageing of critical facilities at major-risk. Such plans have to cover all phases of the life cycle of the equipment and take into account current deterioration mechanisms (i.e. internal and external corrosion, erosion, thermal and mechanical fatigue, etc.). Due to this requirement, there is a need of procedures to check the equipment conditions, especially at the final stages of its life cycle, and evaluate the adequateness of actions for its control. Currently, managers adopt Risk-Based Inspection (RBI) standards, nevertheless it is essential to demonstrate the integration of ageing management within the overall management of major hazard plants. This paper discusses the adequateness of the measures, usually adopted to control the ageing phenomenon in primary containment equipment, whose deterioration could generate a major accident. In order to evaluate the status of such items, a shortcut method has been developed. It represents a first attempt to develop a tool for ageing monitoring. The first release of the method is static and appropriate for independent auditors and inspectors acting on behalf of Seveso Authorities. The second release, which is currently under development, is considered dynamic as information about process variables, external data, inspection information and etc. are continuously collected and processed, in order to provide an overall picture of the ageing of systems in a form of an index. These indexes allow the real-time forecasting of the equipment deterioration process and its management based on the industrial risk acceptance levels. The core of the method is a dynamic model of the strengths that accelerate the degradation processes and factors that slow down them. Based on this model a "digital twin" of a complex plant can be built, by integrating smart sensors and other smart devices.

1 INTRODUCTION

The new requirements of the last European legislation on the control of major accident hazard, the "Seveso III" Directive, include the monitoring of the risk due to equipment ageing (EU Council, 2012). The introduction of plan for the safe management of ageing has to cover all steps of the lifecycle of critical equipment, for this reason there is a need of procedures to verify the status of facilities and evaluate the adequateness of actions made by plant managers.

1.1 *The issue of ageing in process industries*

Recently, the safe ageing of equipment has become the latest hot issue for several industries, in particular those at major accident hazard. The term *ageing* does not refer to the time elapsed from the date of production, testing or commissioning of the equipment, but it is related to its condition and how it changes over the time (Wintle et al., 2006). Ageing of a component reveals itself as a general form of deterioration that is usually associated with the in-service time and reduces its reliability (Horrocks et al., 2010). Ageing increases the risk of loss of containment and other failures and has been proven to be a determining factor in many accidents in process industry (Wood et al., 2013).

Nuclear industry started a decade ago to pay attention to this problem, when it has been realised that the age of most in-service reactors was exceeding the designed lifetime. The guideline, published by IAEA (2009), defined the basic principles for managing ageing of the equipment in nuclear plants, with the aim to safely extend their life beyond the limits defined during the plant's design phase. According to IAEA definitions, there are two terms to be distinguished, i.e. ageing and obsolescence. Both terms basically refer to effects of the time on complex technical systems. In this frame, ageing includes processes that gradually change the physical characteristics

of the equipment over the time or with the use; whereas obsolescence refers to its becoming out of date by comparing with current knowledge, standards and regulations and technology, this makes the equipment inadequate. The consequences of obsolescence include the incompatibility between old and new equipment and the non-compliance of old equipment. Amongst degradation processes, corrosion plays a primary role, thus in many cases, ageing and corrosion are confused each other. The word ageing is often used with a negative connotation, and understood as degradation, even though the concept of "ageing management" clearly implies the idea that ageing processes may be controlled, in order to slow-down and minimise their effects. In the present paper, the focus is just on ageing; obsolescence is out of the scope of this research, as well as consequences of ageing of workers, managers and organisation.

In chemical industry and in the oil & gas sector, the issue of ageing is particularly relevant, as most European refineries have already been in service for forty or more years and it is supposed they will have to continue to be operating, given the difficulties to build new ones. A study promoted by the European Commission a few years ago analysed a hundred worldwide major accidents in oil refineries, which were due to inadequate management of ageing and corrosion, this investigation revealed the relevance of the problem (Wood et al., 2012). A more recent study, promoted by OECD (2017), outlined the impact of ageing also in process industry, including the chemical sector. A fundamental guideline for ageing management in chemical industry has been published by the HSE (Wintle et al., 2006).

A keystone for ageing management is the replacement. In many cases, deteriorated or damaged systems may be dismissed and replaced by new ones having equivalent features. As an example, in a typical process plant there are thousands of valves, which can lose their functions, due to deterioration processes; these components usually have an affordable cost and comply with standard rules, thus replacement may be reasonably proposed. The case of larger items, which have an unsustainable replacement cost as well as complex authorisation procedures, is different; replacement is very difficult and discourages the executives in a way that these items are considered practically "no-replaceable", thus, to extend the in-service life as long as possible becomes a priority for the maintenance engineers. In order to comply with the definitions of common engineering practices, i.e. HSE document (Wintle et al., 2010), these no-replaceable facilities are denoted as "*static primary containment systems*", i.e. systems for which the concern of ageing is much higher than other equipment, such as rotating machinery (e.g. pumps) and control systems.

1.2 *The issue of ageing in the framework of the European Seveso legislation*

The previously mentioned report of the European Commission (Wood et al., 2012) showed the relevance of ageing in refining industry. This oriented the EU Council to add, into the new Directive on major accident prevention, the requirement to define a management program for a safe ageing of critical equipment for all Seveso establishments.

For about a decade and more, the oil and gas industry has trusted in the popular Risk-Based Inspection (RBI) practice, as defined by the recommended documents API 580 (API, 2016) and API 581 (API, 2016a). As discussed by Bragatto et al. (2012), the traditional RBI approach is valuable, but it must be integrated within a dynamic management system for major accident prevention. The main limit of these American standards is that several industries, other than refineries, are classified at major accident hazard according to the Seveso legislation, but they are not included in the field of application of API580/581 and, thus, lack of clear guidelines. As discussed by Bragatto and Milazzo (2016), the Seveso II Directive stresses also the need to share with controlling authorities some aspects of risk management and consequently increases the need to reduce the uncertainties of RBI models. Such uncertainties are associated with different aspects: firstly, uncertainties derive from an inadequate knowledge about the failure modes and related probabilities; further uncertainties are introduced in the following steps of the application of the method (Milazzo & Aven, 2012).

1.3 *Audit of ageing programs*

The integrated safety management system, required by the new Seveso legislation, has to be verified by competent authorities in order to judge about the overall adequateness from the point of view of the previously discussed ageing issue. In order to meet the needs of establishment executives and controllers (auditors and inspectors), a shared model for the adequateness verification is essential.

This paper aims at discussing the main elements to be integrated in an effective approach for both monitoring and inspecting critical equipment at Seveso establishments. In the following text, a piece of equipment is defined critical, if it is involved in a sequence of events leading to a top-event, as identified by the fault tree analysis or equivalent methods; top-events could escalate and give major consequences. Equipment, containing an amount of hazardous

substance equal at least the 20% the threshold indicated by the Seveso III Directive, are also included in the category of "critical items". Therefore, not all equipment in a Seveso establishment has to be considered "critical", but only systems that are explicitly identified in the risk assessment as defined by Seveso III Directive (Safety Report).

This paper is structured as follows: Section 2 discusses scope and objectives of the research. Section 3 describes the ageing model, i.e. a list of factors that affect the phenomenon is given as well as correlations amongst them. Section 4 summarises a short-cut method for external audits, which supports to verify the adequacy of ageing management plans. Section 5 is focused on the use of the ageing model by its integration in a more sophisticated tool so-called ageing sensor. Finally, conclusions and a short discussion about further developments are given in Section 6.

2 SCOPE AND OBJECTIVES

The scope of this research is to discuss how to integrate the main factors, affecting the ageing of facilities, in an effective tool for monitoring and inspecting critical equipment at Seveso establishments. The quantitative analysis of top-events, through the fault tree technique or equivalent methods, shows those that should be considered credible and, by referring to a frequency threshold, allows selecting those to be analysed from the consequence point of view. Even if the likelihood threshold is set at 10^{-6} event/year, it could be possible that, due to deterioration processes, the failure probability does not respect the traditional bath curve trend, but shows an increasing trend when the equipment is close to end of life. Thus, it could be possible that some top events, which were not considered credible because having frequency $<10^{-6}$ event/year, become credible as failure probabilities are higher than expected due to ageing.

The main objectives of this research are summarised below: (1) to provide auditors and inspectors with a trustable short-cut method for ageing controlling at Seveso establishments; (2) to easily achieve the update of the conditions of the equipment when required. The short-cut method is based on a previous preliminary study (Bragatto et al., 2017), from which a simple and effective approach has been developed. The tool is useful for the management of equipment operability and is currently under testing in a few Italian establishments. Concerning the elaboration of the ageing status, the most ambitious future goal is to have a day by day monitoring. It must be pointed that, even if the ageing of machinery is also important, in this paper it has not been discussed, since the focus is only on major accidents.

3 AGEING MODEL

3.1 *Factors affecting equipment ageing*

The ageing status of a plant is described by a model that collects a number of factors that contribute to the equipment deterioration. The management of ageing aim at understanding these factors, assigning the proper weight to their contributing to the deterioration and finally understanding the relationships amongst them. According to these considerations, the safe management of ageing appears a complex issue, which needs three essential elements:

1. *Knowledge* (K) – it is the understanding of all deterioration mechanisms, affecting the equipment during its lifetime;
2. *Information* (I) – it refers to the collection of the documents that describe the past of the equipment starting from the early stage of lifetime, i.e. design criteria, materials of construction and each change made during lifetime;
3. *Data* (D) – it represents information collected by means of non-destructive tests, i.e. measurements that have to be processed in order to contribute to the monitoring of equipment integrity and functionality.

These key-elements allow depicting a complete picture of equipment. It is clear that, if knowledge about deterioration mechanisms is poor or information over the entire lifetime is lost, the measurement data are not enough for a good decision-making, because past actions could have earlier brought the facility to compromising conditions.

To describe the *ageing model*, factors affecting the phenomenon have been grouped in two categories *hardcore* and *softcore*. *Hardcore* includes direct factors (managerial factors), those that are linked to measurements and corrective actions contrasting deterioration mechanisms that are known (i.e. it is correlated to *Data*). *Softcore* includes indirect factors (physical factors), i.e. those elements that are correlated to *Information* and *Knowledge* and have the maximum control in predicting when the equipment has to be removed from service.

As shown in Figure 1, the core of the model is represented by the deterioration processes, which have either a physical or a chemical nature and sometimes both them. The scheme gives the dependence of these processes by a number of accelerating and slowing-down factors over the time, respectively, placed downward and upward. It can be also observed that, beyond those mentioned above, there are other factors that had an effect in the past (i.e. *design criteria*, *processes*, *materials*, *environment, repairs* and *age/in-service time*), their consequence is due to choices that were previously made, whose effects cannot be corrected anymore.

It must be pointed that *repairs* is an accelerating factor that refers to modifications, which are not included in the management of changes; whereas *modifications* has to be intended as a factor that influences the equipment age and, in some cases, has the power to reset it (*age re-conditioning*).

3.2 *Relationship amongst factors*

Accelerating and retarding factors, shown in Figure 1, are not independent in their actions on the degradation phenomenon. In this section a description about how they are related to each other is given. Relationships are summarised in Figure 2 in the form of arrows connecting related factors.

It should be noted that *defects*, *damages*, *failures* and *accidents/near-misses* are a direct consequence of the *deterioration mechanisms*. Defects refer to a structural damage, identified by inspection, which does not compromise the operating of the system, thus, its repair is not necessary; whereas damages refer to something that compromises the operability and compels repairs or replacement. Failure is the end of the containment capacity of the system; therefore, it manifests itself through a loss of containment. Finally, accidents are scenarios occurring after the loss of containment (fire, explosion or dispersion of a toxic substance). The factors discussed above represent different modalities of manifestation of the ageing phenomenon, which proceeds according to a precise sequence of evolution, which is the following: defects → damages → failures → accidents/near misses. All four factors contribute in turn increasing the degradation rate (accelerating factors). The *number of unplanned stops* of the plant directly contributes in accelerating the deterioration as it makes the equipment subject to various types of stress. Amongst the factors, which act directly by slowing-down the deterioration, there are *physical protections* (e.g. cladding and lining), *maintenance*, *inspection program* and *inspection results*. The inspections allow the identification of defects, which may be the quicker the more effective their planning. *Inspection policy* (or program) should be based on *risk assessment* and the knowledge of *inspection technique* and *scheduling*. A risk-based inspection system is influenced by the *inspector qualification* and the *audit of management system* (*SMS*). This last factor includes *change of property* (or ownership) and *experienced personnel loss* due to staff change (accelerating at the softcore level) and *documentation along lifecycle and risk assessment* (retarding at the softcore level), and *maintenance* (retarding at the hardcore level). The *process control* is a factor that acts on the monitoring and control of operating parameters of equipment, for which the choice was made during the process design phase (see Figure 1).

Finally, other factors that directly contribute to ageing are *environment*, *materials*, *processes* and *design criteria*. Their action, as discussed in the previous section, is independent and cannot be eliminated because associated with past choices.

3.3 *Model simplifications*

The model, described from the relational point of view in Section 3.2, can be simplified as shown in Figure 3. A short discussion on simplifications is given below.

A comprehensive knowledge of deterioration mechanisms and control techniques is the sound basis for recognized inspection practices, such as API 581 (2016a), which is the distillate of decades of scientific research and experience in oil and allied industries. Where there is a lack of knowledge, such as in a few chemical industries, inspection strategy is a relevant element; this is

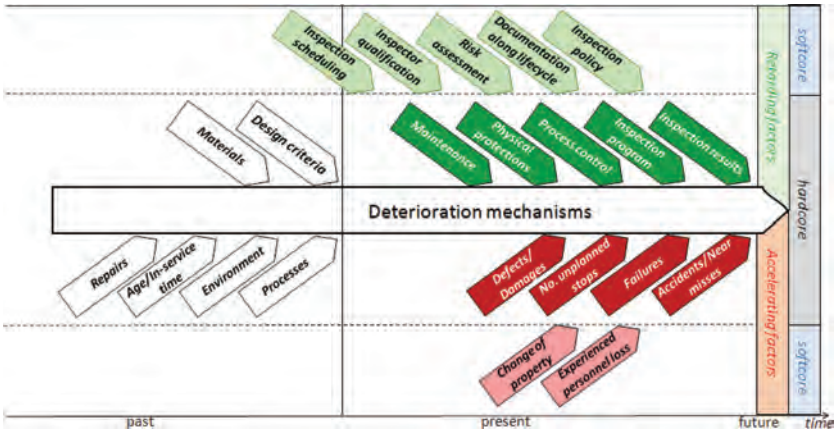


Figure 1. Ageing model.

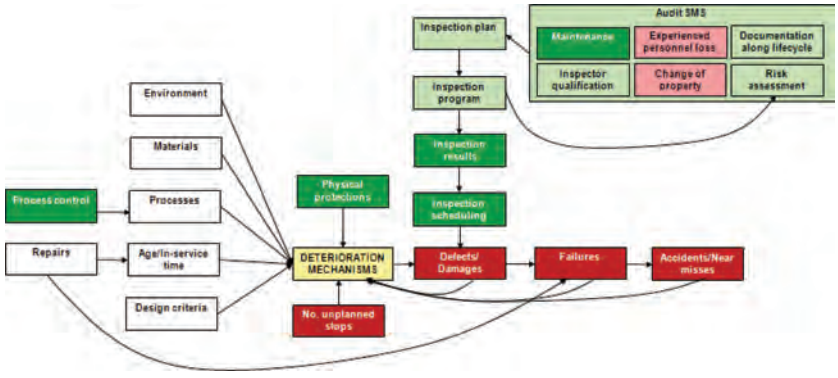


Figure 2. Relationship amongst accelerating and retarding factors.

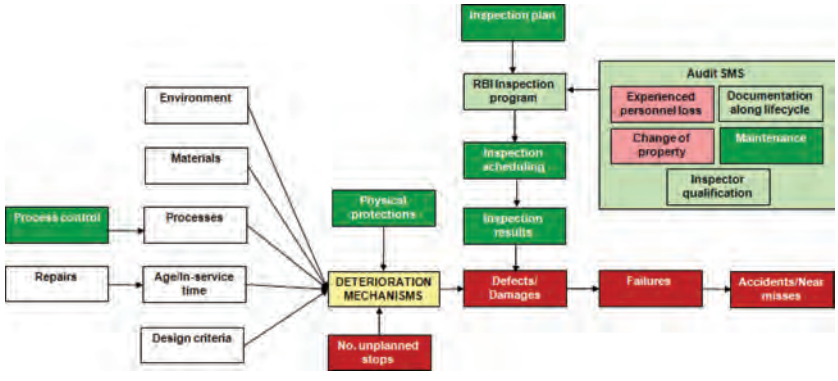


Figure 3. Simplified model for ageing.

particularly important for the knowledge if the inspection technique (*Inspection program*). If the *inspection program* is risk-based, a periodic update is certainly considered in the inspection planning; the approach could ideally be dynamic, especially if the use of an *ageing sensor* is exploited. In this case, *risk assessment* and *inspection program* are merged in the factor *RBI inspection program*.

The *failures* factor has previously been linked to *repairs* because usually a failure requires the need for an intervention. In the hypothesis of intervening before a possible loss of containment, given the effectiveness of the dynamic system for inspection planning, it is possible to not consider the link between the two factors.

4 SHORT-CUT METHOD TO ASSESS AGEING MANAGEMENT PLANS

To support the auditors in performing the assessment of the adequacy of ageing management plans, a short-cut method has recently been proposed (Bragatto et al., 2017). It is an index approach, which is simple and easy-to-use.

The method consists in the assignment scores to accelerating and retarding factors with respect to the ageing. These scores can be in the form of penalty for accelerating factors and of compensation for retarding ones. If the cumulated compensations are greater or equal to the cumulated penalties, the activities that are in place for the ageing management are adequate.

On the contrary, the ageing management system must be improved by the adoption of some technical and/or managerial solutions that increase the scores for retarding factors.

This approach is characterised by proportionality in applying countermeasures to the ageing phenomenon. Hence, if penalties are low, little compensation is required, whereas if they increase, it is necessary also to increase prevention activities to get a higher compensation. The industrial manager can choose technical and/or managerial solutions to be applied in order to offset the penalties, according to his/her preferences. As an advantage of the method, compared to traditional check-lists,

there is a greater clarity in the evaluation process and a quantification of the weight to be attributed to each factor. Unfortunately, as other index methods, it introduces uncertainties. The method and its main steps are represented in Figure 4. Each score (penalty or compensation) is assigned by referring a four-level scale. Such levels are identified as: 1 = low; 2 = medium; 3 = medium-high; 4 = high. A sign will be also associated with the score that will be negative for penalties and positive for compensations.

Accelerating and slowing down factors are those given in Figure 1. To simplify the work of the auditor and/or the industrial manager, some factors have been grouped in a new one and, thus, these have been considered sub-factors of the new factors. The new factors are: (i) *audit of SMS* includes *risk assessment*, *documentation along lifecycle*, *change of property*, *experienced personnel loss and maintenance*, (ii) *inspection management* refers to *inspection program*, (iii) *inspection effectiveness* includes *inspection program* and *inspector qualification* and, finally, (iv) *inspection results* includes *inspection scheduling*.

Tables 1 and 2 show the criteria for assigning scores to accelerating and retarding factors, these are based on the following definitions:

- *Age*/*In-service time* = ratio "current age/maximum designed age" or "current operating hours/maximum in-service hours".

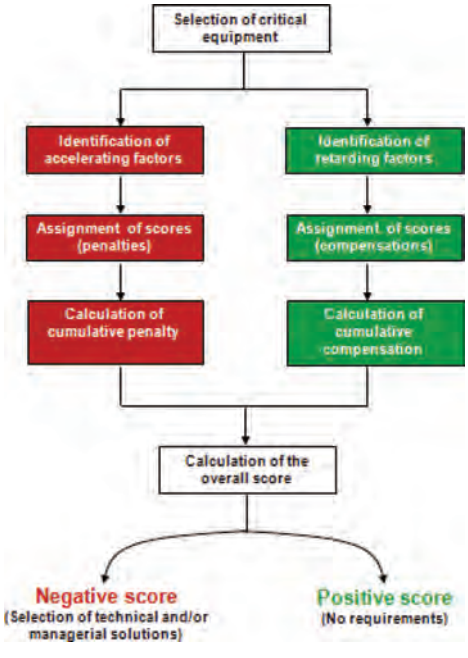


Figure 4. Flow-chart illustrating the method for the assessment of ageing management plans.

- *No. unplanned stops* = ratio "no unexpected stops/total stops" over a reference period.
- *Failures* = actual failure rate over the reference period (f) compared to the failure rate of data from international databases (f_{ref}).
- *Accidents*/*Near-misses* = ratio number of incidents and near misses due to ageing and the total number of registered events over a reference period.
- *Deterioration mechanisms* = average value among three scores related to the consequences of the degradation (i.e. dimension of leakage), the ability to detect the main damage mechanisms (by an inspection technique) and the velocity of propagation of the phenomenon- To quantify the score indications are given by Bragatto et al. (2017).
- *Defects*/*Damages* = percentage of serious damage, detected over the reference period, compared to the number of critical equipment.
- *Audits SMS* = average value between two scores related to audits conduction (internal and external) on the SMS and their results.
- *Inspection management* = main characteristics of the structure of the inspection management.

Table 1. Criteria for the assignment of scores for accelerating factors.

Factor	Score	Value
Age/In-service time	1	≤ 90 %
	2	90 ÷ 100 %
	3	100 ÷ 120 %
	4	120 %
No. of unplanned stops	1	≤ 10 %
	2	10 ÷ 25 %
	3	25 ÷ 60 %
	4	> 60 %
Failures	1	$f \leq 0.5 \cdot f_{ref}$
	2	$0.5 \cdot f_{ref} < f \leq f_{ref}$
	3	$f_{ref} < f \leq 2 \cdot f_{ref}$
	4	$f > 2 \cdot f_{ref}$
Accidents/ Near-misses	1	≤ 5 %
	2	5 ÷ 15%
	3	15 ÷ 35 %
	4	> 35 %
Deterioration mechanisms	1 2 3 4	Average score accounting for: (i) consequences, (ii) ability to detect mechanisms, (iii) propagation velocity
Defects/ Damages	1	≤ 1%
	2	1 ÷ 3 %
	3	3 ÷ 5 %
	4	> 5%

Table 2. Criteria for the assignment of scores for retarding factors.

Factor	Score	Value
Audits SMS	1	Average score accounting for: (i) % of minor non-compliances, (ii) % of greater non-compliances
	2	
	3	
	4	
Inspection management	1	compliant with the legislation
	2	risk-based integrated with inspection plan
	3	updated after changes
	4	periodically updated
Inspection results	1	Average score accounting for: (i) system functionality test results, (ii) system integrity test results, (iii) inspections planning (scheduling)
	2	
	3	
	4	
Inspections effectiveness	1	Average score accounting for: (i) effectiveness of inspections, (ii) inspector qualification
	2	
	3	
	4	
Process control	1	unregistering local control system
	2	control system with data recording
	3	data recording system with automatic blockage
	4	control system with data recording + automatic blockage + certified blockage
Specific protections	1	Average score accounting for: (i) inspection intervals, (ii) protection's conditions
	2	
	3	
	4	

- *Inspection results* = average value among three scores accounting for the inspections planning and the results of tests that verify the functionality and integrity of the systems.
- *Inspections effectiveness* (it is partly included in *inspection program*) = average value among three scores accounting for extension and degree of coverage of techniques, likelihood of damage detecting and qualification of inspectors. Reference should be done to UNI 11325-8 (UNI, 2013) and API 581 (API, 2016a).
- *Process control* = main characteristics of the installation control systems of process variables.
- *Physical protections* = average value among three scores accounting for the type of coating, the frequency of the controls and the actual condition of the material.

In Table 1 and 2, the score for factors, which takes into account various sub-factors, is calculated by averaging the scores of the various subfactors (see Bragatto & Milazzo, 2016).

5 VIRTUAL SENSOR FOR AGEING

The first version of the ageing model, which does not take into account the relationships amongst accelerating and retarding factors, is implemented in the short-cut method for the ageing monitoring and control at major hazard establishments presented in Section 4. To evaluate the status of critical equipment, a static model appears the most appropriate for auditors acting on behalf of Seveso Authorities.

Nevertheless, a dynamic model, which also accounts for the interaction amongst factors, could be more effective for industrial managers. For this reason, a second release of the method is currently

under development within a system, called *virtual sensor* for ageing and made up by hardware and software. The model is considered dynamic as information about process variables, external data, inspection information and etc. are continuously collected and processed, in order to provide an overall picture of the system's status in a form of an index (overall ageing score). This index allows the real-time forecasting of the equipment deterioration process and its management based on the industrial risk acceptance levels. Based on this model, a *digital twin* of a complex plant can be built, by integrating sensors and other smart devices collecting information from the equipment and the establishment.

Therefore, the *digital twin* is made up of measured data, managed information and models for the physical evolution of equipment. It simulates the real evolution of equipment, anticipating possible failures. This set of "data-information-knowledge" can be recalled from every "location" (e.g. cloud, DCS, etc.) through a device that constitutes the interface with the user.

6 CONCLUSIONS

Due to the legislation requirement, in the context of major hazard establishments, inspections planning and maintenance activities are essential elements to guarantee a safe ageing of installations. These must be based on in-depth knowledge of all damage mechanisms and backed up by appropriate controls. The proposed model (in its various versions), by accounting for the interaction amongst accelerating and slowing-down factors and its dynamics, is aimed at supporting the auditing activity and promoting an ageing management based on *knowledge*, *information* and *data*. At the present the second release of the method is under development and implementation within a system, called *virtual sensor*. It will work based on a *digital twin* of a complex plant and achieve a dynamic monitoring of the ageing status of the overall establishment.

ACKNOWLEDGMENT

This work is part of an Italian research project entitled "SmartBench" that is supported by INAIL (Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro) and funded within the call BRIC 2016.

REFERENCES

API American Petroleum Institute 2016. Risk-based Inspection. *API recommended practice API RP 580*.

API American Petroleum Institute 2016a. Risk-Based Inspection Methodology. *API recommended practice API RP 581*.

Bragatto, P., Della Site, C. & Faragnoli, A. 2012. Opportunities and threats of risk based inspections: The new italian legislation on pressure equipment inspection. *Chemical Engineering Transactions* 26, 177–182.

Bragatto, P. & Milazzo, M.F. 2016. Risk due to the ageing of equipment: Assessment and management. *Chemical Engineering Transactions* 53: 253–258.

Bragatto, P., Delle Site, C., & Milazzo, M.F. 2017. Audit of Ageing Management in Plants at Major Accident Hazard. *Proceedings 2nd International Conference on System Reliability and Safety ICSRS*, pp. 400–405.

EU Council, 2012. Directive 2012/18/EU on the control of major-accident hazards involving dangerous substances. *Official Journal of the European Union* L197/1-37.

Horrocks, P., Mansfield, D., Thomson, J., Parkerv, K. & Winter, P. 2010. Plant Ageing Study Phase 1 Report. *Health and Safety Executive Report no. RR823*. Available on-line: www.hse.gov.uk.

IAEA 2009. Ageing management for nuclear power plants. *IAEA Safety Standards Series No. NS-G-2.12*. Available on-line: http://www-pub.iaea.org.

OECD Organisation for Economic Cooperation and Development 2017. Ageing of hazardous installations. *OECD Environment, Health and Safety Publications* – Series on Chemical Accidents, no. 29. Available on-line: http://www.oecd.org.

Milazzo, M.F. & Aven, T. 2012. An extended risk assessment approach for chemical plants applied to a study related to pipe ruptures. *Reliability Engineering and System Safety* 99, 183–192.

Ente Nazionale Italiano di Unificazione (UNI) 2013. Pianificazione delle manutenzioni su attrezzature a pressione attraverso metodologie basate sulla valutazione del rischio (RBI). *Document UNI 11325, Part 8* (in Italian).

Wintle, J., Moore, P., Henry, N., Smalley, S. & Amphlett, G., 2006. Plant ageing. Management of equipment containing hazardous fluids or pressure. *Health and Safety Executive Report no. RR509*. Available on-line: www.hse.gov.uk.

Wood, M.H., Arellano, A.V. & Van Wijk, L. 2013. Corrosion Related Accidents in Petroleum Refineries. *European Commission Joint Research Centre Report no. EUR, 26331*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Failure prognosis of discrete events systems based on extended Petri Nets

R. Kanazy & S. Chafik
Pluridisciplinary Research and Innovation Laboratory, EMSI, Casablanca, Morocco

E. Niel
Department of Industrial Engineering, Ampere Laboratory, National Institute of Applied Science Lyon, Villeurbanne, France

ABSTRACT: Fault prognosis has become a major scope for complex and interconnected systems. Such significant events as fault events can cause partial or total stop of attempted functionalities. Prevention failure events are an issue to preserve performance, availability and safety of both operators and equipment. The aim of prognosis is to prevent fault events before their occurrence. Fault/repair management refers to event control, and so it is relevant to the domain of Discrete Events Systems (DES), for which stochastic finite state automaton and Petri Nets (PN) have been used to prognosticate fault state. They are based on predictions of fault event at least m-steps in advance.

The proposal is based on the time notion, which is crucial for fault prognosis. Indeed one can give the remaining time before the occurrence of fault event. The goal is to prevent the occurrence of a fault event at τ – *timeunits* in advance. This approach is based on labeled and T-temporal Petri Net, which has the advantage of a formal character for the assessment of properties and of sufficiently generic in order to apprehend a high level of complexity.

1 INTRODUCTION

The availability of complex systems can be ensured solely by control fault events, which can occur at any time triggering partial or total down of the system. The criticality of complex systems requires failure report before its occurrence beside the detection of failure and report of dysfunction alarms, avoiding the accidental down of the system. Prognosis allows meeting these requirements and giving visibility on the evolution of the system, thus, allowing the prediction of future failures. Several fault prognosis methods have been developed; some have adopted a stochastic approach (Ammour et al. 2017) (Dutta and Biswas 2015) (Chen and Kumar 2014), while others have chosen non-stochastic (Takai 2012) (Kumar and Takai 2010) (Takai 2015), one for state automaton or Petri Net. These approach are interested in prediction of failure m-steps in advance, based on a stochastic process that cannot predict in time. The challenge of each community working on this topic is to predict perfectly the future reality.

For visibility of the future, one must master the present and have enough information about the past. The verification of time constraints an extension of Petri Nets (PN) (Chen et al. 2017) called temporal Petri Nets (Berthomieu 2001), is part of these modeling methods. Tow modeling methods are proposed, the first one, combine T-temporal PN with labeled PN (Yin and Lafortune 2017) to give an extension of PN, called extended labeled T-temporal PN. In this extension, clock and label are associated with transitions. The second method, integer Watchdog techniques (Kovacs et al. 2007), it generates alarms to indicate the existence of a fault in the system. However, the goal is to exceeds detection and aims to prevent fault and hence comes the interest an improvement of the method; that allows for more expressivity of the model in order to determine τ-time units in advance and the occurrence or not of a failure in the system.

The first proposal, introduces the notion of INIT and EXC events, which represent respectively the initialization and exceeding time of the clock, thus, allowing the distribution of the system to operating modes (Nominal, Degraded and Failed), in order to control risks. Following this distribution, we can determine the relevant, non-relevant and critical places. From the first relevant place in degraded mode, one can prevent the occurrence of a fault event τ – time units in advance. Considering that the system cannot remain in degraded mode indefinitely, the notion of cycle execution of degraded mode is introduced. By exploiting

the temporal constraints of transitions, it would be possible to calculate the minimum time of a mode execution, yet it would be interesting to be able to calculate the maximum time of a mode execution. Then, the notion of execution cycle can help to calculate a temporal estimate of execution at the latest by a mode, which allows to have a time estimation interval for a mode execution at the earliest and latest. If the extended PN is not safe but bounded (places contain more than one token), the multi-clock is introduced. The second method, represents the watchdog approach based of the labeled T-temporal PN (Ru and Hadjicostis 2009); the alarm places generated by the watchdog become indicator places from which one can prevent a fault occurrence by computing the earliest and latest time of its occurrence.

The paper is organized as follows: the first section is dedicated to the representation of Petri nets and their extensions integrating time constraints. In the second section, the rules of transition from the T-temporal PN to the extended T-temporal PN are determined. The third section will be devoted to modeling the system using the various methods mentioned above. The fourth section will detail the proposed prognosis approach. The fifth section will push the boundaries of the proposed method by introducing multi-clocks in the model. The paper will be concluded with a conclusion.

2 PRELIMINARY

2.1 *T-temporal Petri Nets*

T-Temporal PN (Berthomieu 2001) (Sadou and Demmou 2009) (Zuberek 1991) (Jiacun 1998) is a tuple: $TR = (P, T, Pre, Post, M_0, I^S)$, where: I^S is the static interval function. It is represented by temporal constraints that can be associated with places, arcs or transitions.

This differentiation in the representation does not influence semantics. In this paper, the temporal constraints in T-temporal PN are associated with transitions by rational bounded intervals. $I^S(t) = [min, max]$ avec $0 \le min \le max$ and max can be ∞. The firing of a transition can only occur after a minimum of time units (*min*) and at the latest a maximum of time units (*max*). For example in Figure 1; T_1 is fired at earliest a UT and at latest b UT.

2.2 *Labeled Petri nets*

Labeled PN (Yin and Lafortune 2017) (Li 2017) (Jiacun 1998), adds to PN a label alphabet and a function for labeling transitions with these labels.

Labeled PN is a quadruplet $LR = (R, M_0, \Sigma, \lambda)$, with:

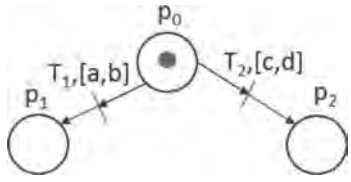


Figure 1. T-temporal Petri Nets.

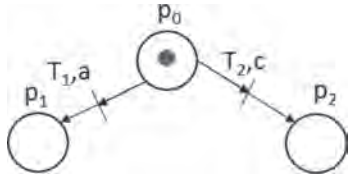


Figure 2. Labeled Petri Net.

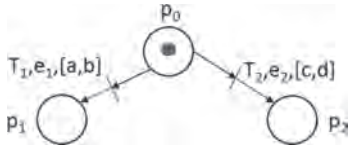


Figure 3. T-temporal labeled Petri Net.

- (R, M_0): Marking PN.
- Σ: finite Set of events.
- $\lambda : T \to \Sigma$ The function for labeling transitions that assigns a label (*event*) to each transition.

The model of Figure 2 can be considered as the execution of PN in Figure 1 by identifying events a and c respectively at transitions t_1 and t_2.

2.3 *Labeled T-temporal Petri nets*

Labeled T-temporal PN (Peres et al. 2011), is a n-uplet $LTR = (P, T, Pre, Post, M_0, I^S, \Sigma, \lambda)$, where:

- $\lambda : T \to \Sigma$ is the labeling function;

Figure 3 represents a T-temporal labeled PN:

3 FAULT PRONOSTICS BASED ON EXTENDED PN

The prognostic method proposed in this paper is based on the temporal estimation of the occurrence of a future event. Modeling the system on extended T-temporal labeled PN. In order to analyze the risks, the system must be distributed into operating modes (nominal, degraded, faulty); Once the modes of operation are identified, it is necessary to determine the places which are relevant or not and which are critical. The aim of prognosis is to calculate the time estimate of the execution of the degraded mode in order to determine in advance the earliest possible occurrence of a failure or entering in the failure mode.

3.1 *Extended PN*

The proposed modelisation in this section is based on T-temporal labeled PN, which are adapted to our approach, that consider critical systems such as real-time systems. An extended PN is a 5-tuplet $(P,T,\sum,Init,Exc,X,C)$, with:

- Init: the time initialization function of the clock.
- Exc: the time exceeding function of the clock.
- X: Set of Clocks.
- C: Set of temporal constraints.

The semantics differ from that of T-temporal labeled PN, Firing of a transition in extended T-temporal PN is possible by checking both the time constraint and the occurrence of the event. Firing transition allows knowing if the event occurred before, simultaneously or after the time constraint.

The integration of temporal constraints in the PN is expressed by their associations with places, arcs or transitions. Extended T-temporal labeled PN is modeled by associating temporal constraints with transitions; the goal is to determine if during the firing, an event has occurred on the intended time or not. Two notions have been introduced: initialization and exceeding clock noted, Init and Exc respectively. This modeling method gives more expressiveness, in order to predict the future evolution of the system. The Init sets the time clock when firing transition. The Exc checks whether the event occurred before or after time clock exceeding. To introduce the method, the rules of transition are modeled from T-temporal PN to extended PN (see Figure 10).

In the above model the firing of the transition will occur once the event appears; the reset of the clock x is automatic done after the firing transition. *Init* is associated with the first transition to reset the x time clock to 0 and set the guard. The transition is associated with $\{b,Exc(x,30)\}$. It is fired after the occurrence of event b, before or simultaneously with the expiration of time clock.

The expression $\{Exc(x,30),b\}$ allows the firing of the transition after checking the occurrence of event b, after the expiration of the time clock.

The equivalent model does not include an Init in the modeling, since the clock initialization is not expressed in the T-temporal PN. The expression $\{c,Exc(x,30)\}$ checks whether event c occurred before or simultaneously with the expiration of the time clock.

The only condition for firing the transition in the first model is based on the occurrence of event c; no guard and no initialization of the clock are indicated. The equivalent model will remain the same constraint. The notions of Init and Exc refer to the research realized by KHOUMSI on timed PN (Ouédraogo et al. 2006) (Khoumsi 2009) (Khoumsi 2005). Which consists of reformulating the prognosis problem in real time in a non-real time form, by transforming the timed automaton into finite state automaton called SEA (for Set-Exp-Automaton).

3.2 *Example of model*

3.2.1 *Modeling of T-temporal labeled PN*

In order to implement the rules of passage from the T-temporal PN to the extended PN, the model T-temporal labeled is used in (Figure 8). It is assumed that the PN of the Figure 9 is a safe one and event c always respects the temporal constraints with which it is associated in the model (to

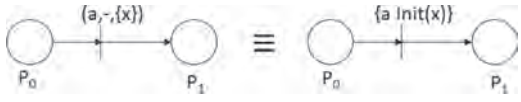


Figure 4. Rule 1.

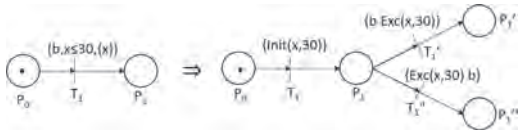


Figure 5. Rule 2.

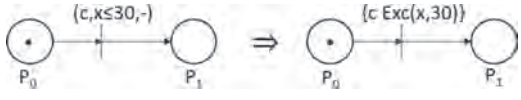


Figure 6. Rule 3.

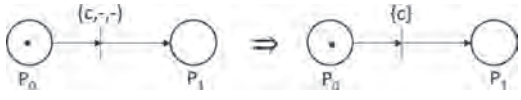


Figure 7. Rule 4.

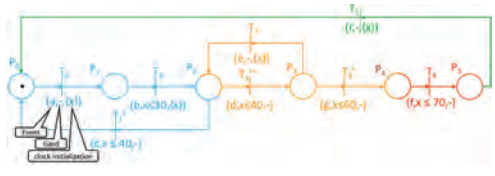


Figure 8. Example of T-temporal labeled PN model.

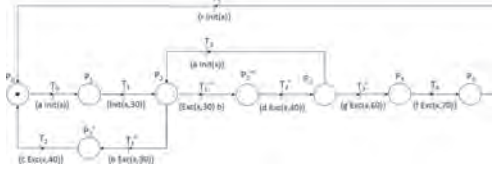


Figure 9. Extended PN model.

avoid the presence of a non-nominal behavior). Each transition is associated with a triplet *event, guard, reset*; The guard represents a temporal constraint that can be >, ≥, <, ≤ then a specified number of time units. It is possible that the guard may not be mentioned if the firing of the transition depends solely on the occurrence of an event.

- $P = \{P_0, P_1, P_2, P_3, P_4, P_5\}$ Set of places.
- $T = \{T_0, T_1, T_1', T_1'', T_3, T_3', T_4, T_5\}$ Set of transitions.
- $\Sigma = \{a, b, c, d, e, f, g, r\}$ Set of events, where f is a failure event and r a repair event.
- $\Sigma = \Sigma_n \cup \Sigma_d \cup \Sigma_f \cup \Sigma_r$.
  - $\Sigma_n \{a, b, c\}$: Set of nominal events
  - $\Sigma_d \{d, e, g\}$: Set of degraded events
  - $\Sigma_f \{f\}$: Set of failure event
  - $\Sigma_r \{r\}$: Set of repair
- x: Clock.

The clock reset is mentioned by x, an uninitialized clock is represented by –.

3.2.2 *Extended PN model*

Figure 10 represents system modeling with extended PN. Two notions have been introduced in this modeling; the Init, which is used to define both the time constraint and the initialization of the time clock when necessary. The notion of Exc allows checking the exceeding of the time clock; it is formed by two parameters: the first one indicates the clock and the second one indicates the temporal constraint to check. The model below represents the extended PN. It is assumed that this PN is safe:

3.3 *Description of the prognostic method*

3.3.1 *Steps of the prognosis method*

The prognostic approach is based on the verification of the future occurrence of a failure event at the earliest and latest. The first step is to provide temporal modeling of the system on extended PN (see next section) or on PN with Watchdog technique (see section 3.4), in order to exploit the time constraints represented in the model. To analyze the risks, the second step indicate the possible operating modes in the system (Nominal, Degraded, Failed). The nominal mode corresponds to operating mode

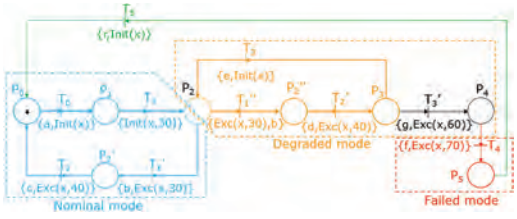


Figure 10. Operating modes of the system.

of the system without time constraints disrespect, while the degraded mode corresponds to system operation with degradation (presence of partial failure) but without stopping the system and failed mode represents the failure state. The third step determines whether places belong to the operating modes to identify which places are relevant or not and which are critical places. The relevant places can be nominal, degraded, critical or failed, depending on the mode in which they are represented. When a place is represented in both nominal and degradation mode, it is referred to as a non-relevant place. The critical place is an intermediate place between two modes: degraded and failed. The firing of transition leads to the failed mode.

Definition 1:

$P = P_r \cup P_{nr}$, where:

P_{nr} is the set of non-relevant places P_r is the set of the relevant places of the extended PN with $Pr = P_{rn} \cup P_{rdeg} \cup P_{rfail}$ and T the set of its transitions, where:

- P_{rn} is the set of relevant place p_r which belongs to nominal mode. p_r is called the nominal place, denotes p_{rn}. All p_{rn} places constitute the set of places in nominal mode, $\forall p_{rn}, p_{rn} \in P_{rn}$
- P_{rdeg} is the set of relevant place p_r which are only for degraded mode. p_r is called the degraded place, denote p_{rdeg}. All p_{rdeg} places constitute the set of places in degraded mode, $\forall p_{rdeg}, p_{rdeg} \in P_{rdeg}$
- P_{rdef} is the set of relevant place p_r which are only for failed mode. p_r is called the failed place, we denote p_{rfail}. All p_{rfail} places constitute the set of places in failed mode, $\forall p_{rfail}, p_{rfail} \in P_{rfail}$

Definition 2:

A place is said to be an relevant if and only if it belongs to one of the sub-sets $P_r n, P_{rdeg}, P_{rfail}$. A place is said to be critical if and only if it is between the degraded and faulty mode.

The introduction of time-constraints into transitions in modeling provides a time estimation of the execution time of a degraded mode, which is essential for calculating the earliest future failure occurrence. Thus, the number of execution of the degradation mode will make it possible in step four to evaluate the time occurrence of the future failure.

3.3.2 *Prognosis based on extended PN*

Step 1: Model of the system based on extended T-temporal PN.

Modeling by extended T-temporal PN is based on Init/Exc which are associated with transitions, as show in Figure 9. This association will tie the event to the clock to determine whether the event occurred at the right time. *Init*(x); means that the clock x will be initialized to zero without assigning a time-constraint. *Init*(x,30) means that clock x will be initialized to zero with a time-constraint

of 30 TU(*time – units*). *Exc*(*x*,30); verify if clock x has exceeded the time-constraint of 30 TU. Note that a transition with an Exc is not necessarily preceded by a transition with an Init. $\{b, Init(x)\}$; means that the initialization of the clock is dependent on the occurrence of event b. For the function Exc, associating the event with the time- constraint can be expressed in two ways: $b, Exc(x,30)$ means that the appearance of event b must occur before or simultaneously with the expiration of 30 TU, however $\{Exc(x,30), b\}$ means that the event will certainly occur after the exceeding time of 30 TU.

Step 2: Designation of the operating modes: nominal, degraded and faulty.

The repartition of the operating modes of system is done three modes, as show in Figure 10.

Step 3: Determination of relevant places: nominal, degraded, critical and failed. This nomination of places is intended to reinforce the proposed prognostic method (see Figure 11). Following the identification of operating modes, the places of the model are labeled either as relevant or non-relevant places.

Step 4: Evaluation of the execution time of the degraded cycle.

Let us take the example of Figure 11. The firing of T_1", indicates the change from nominal to degraded mode. this firing allows to calculate the execution time of the degraded mode (τ), which is equal to the summing of the time constraints associated with the transitions in this mode, it is possible to determine the earliest time to the failure event f occurs. $\tau = 10TU + 20TU; \tau = 30TU$. Indeed, since the clock is not initialized when firing T_2', the time constraint associated with event d is equal to $10TU = 40TU - 30TU$. However, it is possible that the degraded mode runs n-fold before changing to the failed mode. Using the concept of counter execution cycle.

3.3.3 *Counter of execution cycle*

A system cannot remain in a degraded mode indefinitely. It becomes necessary to compute the number of executions of this mode in advance. In fact, if the execution of a degraded mode is repeated more that n times, it can be sure that the system will be leaded to the failed mode. The problem is to how to compute the number of cycle executions. To do this, we add a place, associated to the degraded mode, in the model. This place (P_6) in the Figure 12 will contain a number of tokens equal to the number of possible cycle execution of the degraded mode, before leaded to the failed mode. It assumed that this number is fixed by expertise (equal to 4) in the example.

In this example, the firing of T_3, means that a cycle of degraded mode has been executed, and a token has been consumed in P_6. At this level, there are 3 execution possibilities before reaching the failed mode. Suppose that we repair the partial failure after the first execution cycle of degraded mode, the system will be leaded to the nominal mode. The next future partial failure will lead again the system to degraded mode. at this stage, the number of tokens in P_6 is equal to 3, while it should be 4. In this case the prediction of failure cannot be accurate. To avoid this problem, the notion of reset arcs, will maintain the exact number of tokens (4 tokens) in P_6.

3.3.4 *Notion of reset arcs*

To introduce the notion of reset arc, (Akshay et al. 2017) which represents the model of an access code verification system, the number of tokens, represents the number of possible attempts. After T_1 (Enter Code) is fired, if the entered code is correct (T_2 reached), the model reset the number of tokens in the place P_0 (Trials). However, the Arc reset allows this withdrawal action. PN with reset arcs (Comlan et al. 2015) is a 4-uplet $N_R = \langle P, T, W, AR \rangle$ with $\langle P, T, W \rangle$ a PN as defined in Definition 1 and $AR : P \times T \to 0,1$ is the set of reset arcs (AR) if there is a reset arc that connects p to t, otherwise $R(p,t) = 1$. In the example in Figure 13.

- P = {*Trials, InputCode, SystemAccess, Retry Code*};
- T = {*EnterCode, CorrectCode, WrongCode*};
- AR = {*CorrectCode*};
- M_0 = (3, 0, 0, 0).

The integration of the reset arc in the model is intended to initialize the counter for the execution cycle of the degraded mode once the fault has been repaired.

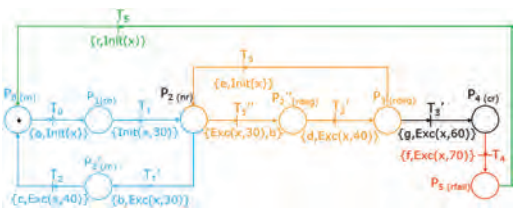


Figure 11. Relevant places on extended PN.

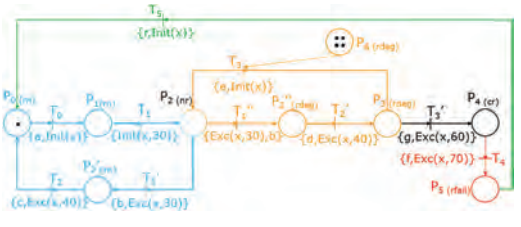


Figure 12. Execution cycle counter on the extended PN model.

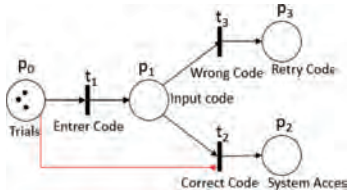


Figure 13. Reset arcs.

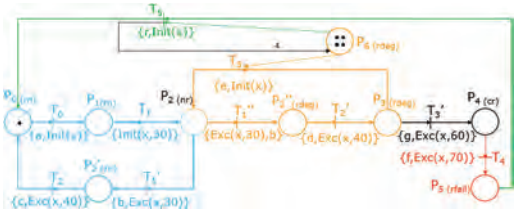


Figure 14. Reset arcs on Extended PN.

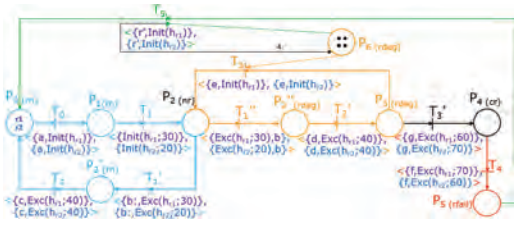


Figure 15. Modeling on Extended colored PN.

In the model of Figure 14, the firing of the T_5 transition represents the repair of the system, which can occur either before completing the possible execution cycles of the degraded mode, or after the occurrence of a failure event. In both cases, the reset arc ensures the emptying of the P_6 place. After firing T_5 transition, the P_6 place can be initialized with tokens equivalent to the number executions of the degraded mode. Assuming that the extended PN is safe, to predict a possible failure event f. The firing of the first transition of the degraded mode gives a visibility of the evolution of the system and makes it possible to determine a temporal estimation, at the earliest and at the latest of a failure events occurrence.

3.4 *Multi-clock approach*

When the assumption that the PN is safe is not verified, by introducing the notion of colored and timed PN (Soares 2017) (Jiacun 1998). The aim is to associate a clock to each token, in order to check their evolution in the system separately. Thus, in the same transition, several temporal constraints can be set depending on the clock associated with each one.

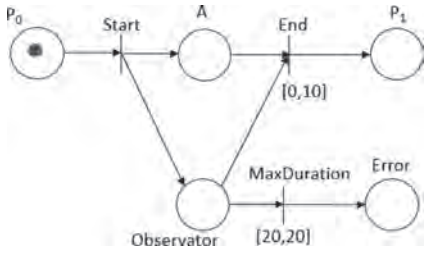


Figure 16. Watchdog on PN.

In the case of multi-token/multi-clock, the temporal estimation of the failure occurrence depends on the clock related to tokens and the time-constraints associated with transitions (see Figure 15. Let's take the case of token r1; if it ever changes from nominal to degraded mode following the firing of T_1", the temporal estimation at the earliest, of the occurrence of a failure event. $\tau_{min}(r1) = 30TU$, but for the token r2, $\tau_{min}(r2) = 50TU$. Taking into account the execution cycle of the degraded mode, the estimation at the latest of the failure occurrence is: $\tau_{max}(r1) = 150TU$ for the token r1 and $\tau_{max}(r2) = 170TU$ for the token r2. Thus, using the concept of multi-token/multi-clock assures prognosis, in advance the occurrence of a failure at the earliest and at the latest, although of parallel execution if the PN is not safe.

4 FAULT PROGNOSIS BASED ON WATCHDOG TOOL

4.1 *Watchdog concept*

In order to overcome a perplexity in diagnosis, T-temporal PN can be used to model alarms by integrating the watchdog mechanism. Based on experience and considering that the risks are known a priori, the watchdog mechanism (Kovacs et al. 2007) (Kovács et al. 2006) (Jerbi et al. 2006) verifies that an action has occurred before a given deadline, and signals the presence of an error in the system if a delay is exceeded. Figure 16 shows the watchdog mechanism on a T-temporal PN. It checks the maximum duration of a task, represented by the place A. The transition Start corresponds to the beginning of the task. Its firing creates a token in the place Observer, so the transition *MaxDuration* will be sensitized and triggers the countdown of his firing interval [20.20].

The place Observer will be marked either by the firing of the transition Start, i. e. when the task is completed, or when the firing interval of the transition *MaxDuration* is completed before the end of A: "*MaxDuration*" is then fired and the place Error is marked. This marking means that the execution time of task A is longer than 20 Time units.

4.2 *Application of watchdog on T-temporal PN*

The model in Figure 9 incorporates the watchdog mechanism. In order to exceed the limits of watchdog (Kovacs et al. 2007, Combacau 1991) and exploit it to the failure prognosis, a new method is proposed to interpret alarms. For example, to signal the change from nominal to degraded mode, the exceeding time of the occurrence of event b is checked when the transition T_2' is fired.

Upstream place of T_2', represents a waiting place (P_0') Downstream place (P_1'') of T_2', represents an indicator of the evolution from nominal to degraded mode. This makes possible the prediction time of the future failure occurrence; the temporal estimation of the occurrence failure on the model in Figure 17 can be predicted from firing the transition T_2'. The latest estimate of the occurrence of a failure is computed by summing the possible execution time of the degraded mode (possible number of cycles multiplied by run-time of the degraded mode) plus the temporal constraints related to transitions T_3'' and T_7. When transition T_3'' is firing, failure f is certain.

4.3 *Prognostic method with watchdog*

If the permitted number of the cycle execution of the degraded mode $n = 4$ (see Figure 18),then the temporal estimation at the earliest of the occurrence of a failure $\tau_{min} = 30TU$ and at the latest $\tau_{max} = 120TU$. So the time interval in which a failure can occur after T_2' firing is $[30UT, 120UT]$.

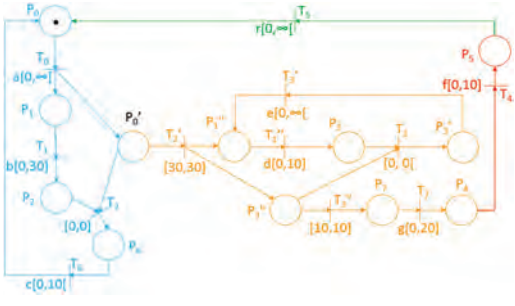


Figure 17. Watchdog on T-temporal PN.

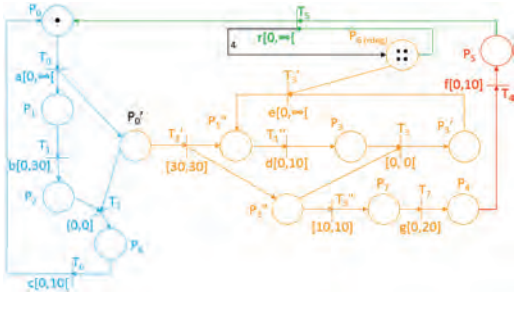


Figure 18. Watchdog on T-temporal PN with reset arc.

5 WATCHDOG VS EXTENDED PN

Following the two modeling represented in sections III and V, it is possible to follow evolutions of the system. In the model based on the Watchdog, the watchdog mechanism is extended, which was limited to fault detection, determining the switching from nominal to degraded mode. Thus, it would be possible to predict the future occurrence of a failure event. Extended PN model reduce the state space of the PN model compared to the Watchdog while maintaining the same properties. The two methods give a visibility on the evolution of the system and provide a temporal estimation of the failure at the earliest and at the latest time. The advantage of Watchdog method over extended PN is that there exists a modelisation and simulation tool called Little Parametric Tools (LPT) proposed by Karen GORDARY in (Godary-Dejean 2008). This advantage does not discriminate against the prognostic method with extended PN, but allows us to integrate into our perspectives the development of a modeling and simulation tool adapted to the extended PN proposed.

6 CONCLUSIONS

Existing prognostic methods are based on a stochastic or non-stochastic approach, to verify the occurrence in the future of a failure event, modeled by Petri net or automaton. The method of prognosis in this paper is based on a temporal approach; which allows the prediction of the occurrence of a failure in advance, by a temporal estimate of the appearance at the earliest and at the latest. Two modeling methods were proposed: The first is a modeling on extended Petri nets, integrating the concepts of initialization and expiration of clocks, the second method, proposes to integrate the watchdog technique in order to adapt it to the prognosis.

The prognosis approach begins with a distribution states of the system on operating mode (nominal, degraded, failed), in order to identify correct localization in each place. Thus, it would be easy to determine which places are relevant or not. Following this identification and from the first entrance date in degraded mode, one can predict the time at the earliest and at the latest of the failure occurrence.

After verification of the approach on both modeling methods, it is concluded that the two modeling methods give the same temporal estimate. the only difference is that the state space in PN with watchdog technique is more important than

the extended PN, which reduces system modeling complexity. That's why extended PN modeling is retained. Future works will consist to formalize the partial prognosability (Prognosis is possible for some parts of the system but not for others) on extended PN by verifying their property and the construction of the local prognoser.

REFERENCES

Akshay, S., S. Chakraborty, A. Das, V. Jagannath, & S. Sandeep (2017). On petri nets with hierarchical special arcs. *arXiv preprint arXiv:1707.01157*.

Ammour, R., E. Leclercq, E. Sanlaville, & D. Lefebvre (2017). Fault prognosis of timed stochastic discrete event systems with bounded estimation error. *Automatica 82*, 35–41.

Berthomieu, B. (2001). Laméthode des classes détats pour lanalyse des réseaux temporels. In *3e congres Modélisation des Systemes Réactifs (MSR2001)*, pp. 275–290.

Chen, J. & R. Kumar (2014). Failure prognosability of stochastic discrete event systems. In *American Control Conference (ACC), 2014*, pp. 2041–2046. IEEE.

Chen, Y., Z. Li, K. Barkaoui, N. Wu, & M. Zhou (2017). Compact supervisory control of discrete event systems by petri nets with data inhibitor arcs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems 47*(2), 364–379.

Combacau, M. (1991). *Commande et surveillance des systèmes à événements discrets complexes: application aux ateliers flexibles*. Ph. D. thesis.

Comlan, M., D. Delfieu, & M. Sogbohossou (2015). Processus de branchement des réseaux de petri à reset arcs. In *GPL GDR*.

Dutta, C.B. & U. Biswas (2015). Failure diagnosis in real time stochastic discrete event systems. *Engineering Science and Technology, an International Journal 18*(4), 616–633.

Godary-Dejean, K. (2008). Lpt: Little parametric tool, outil pour la validation d'une borne temporelle paramétrée. In *CIFA: Conférence Internationale Francophone d'Automatique*.

Jerbi, N., S.C. Dutilleul, E. Craye, & M. Benrejeb (2006). Time disturbances and filtering of sensors signals in tolerant multiproduct job-shops with time constraints. *International Journal of Computers Communications & Control 1*(4), 61–72.

Jiacun, W. (1998). Timed petri nets: theory and application.

Khoumsi, A. (2005). Complete test graph synthesis for symbolic real-time systems. *Electronic Notes in Theoretical Computer Science 130*, 79–100.

Khoumsi, A. (2009). Fault prognosis in real-time discrete event systems. *DX 9*, 259.

Kovács, G., B. Kiss, & E. Niel (2006).Watchdog–a practical approach of fault detection. *IFAC Proceedings Volumes 39*(3), 343–348.

Kovacs, G., L. Piétrac, B. Kiss, & E. Niel (2007). On the formalisation of integrating watchdogs into discrete event controller structures. In *Control Conference (ECC), 2007 European*, pp. 5522–5529. IEEE.

Kumar, R. & S. Takai (2010). Decentralized prognosis of failures in discrete event systems. *IEEE Transactions on Automatic Control 55*(1), 48–59.

Li, B. (2017). *Diagnosis and Diagnosability of Complex Discrete Event Systems Modeled by Labeled Petri Nets*. Ph. D. thesis, Ecole Centrale de Lille.

Ouédraogo, L., A. Khoumsi, & M. Nourelfath (2006). Méthode de transformation dautomates temporisés avec invariants de localités. In *Conférence francophone de modélisation et simulation (MOSIM), Rabat, Morroco*.

Peres, F., B. Berthomieu, & F. Vernadat (2011). On the composition of time petri nets. *Discrete Event Dynamic Systems 21*(3), 395.

Ru, Y. & C.N. Hadjicostis (2009). Fault diagnosis in discrete event systems modeled by partially observed petri nets. *Discrete Event Dynamic Systems 19*(4), 551.

Sadou, N. & H. Demmou (2009). Reliability analysis of discrete event dynamic systems with petri nets. *Reliability Engineering & System Safety 94*(11), 1848–1861.

Soares, J. a. A.C. (2017). Automatic model transformation from uml sequence diagrams to coloured petri nets.

Takai, S. (2012). Robust failure prognosis of partially observed discrete event systems. In *American Control Conference (ACC), 2012*, pp. 6077–6082. IEEE.

Takai, S. (2015). Robust prognosability for a set of partially observed discrete event systems. *Automatica 51*, 123–130.

Yin, X. & S. Lafortune (2017). On the decidability and complexity of diagnosability for labeled petri nets. *IEEE Transactions on Automatic Control*.

Zuberek,W. (1991). Timed petri nets definitions, properties, and applications. *Microelectronics Reliability 31*(4), 627–644.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

A probabilistic risk assessment method for the security of supply in gas networks supported by physical models

B. Gjorgiev, A. Antenucci & G. Sansavini
Reliability and Risk Engineering Laboratory, Department of Mechanical and Process Engineering, Institute of Energy Technology, ETH Zurich, Zurich, Switzerland

A. Volkanovski
Reactor Engineering Division, Jožef Stefan Institute, Ljubljana, Slovenia

ABSTRACT: The paper presents a Probabilistic Risk Assessment (PRA) method for the security of supply of a gas network. The method is based on a procedure for automatic generation of fault trees, which estimate the probability of disruption of the gas delivery from terminals/storages to each consumer nodes in the gas network. The method allows probabilistic analyses of the availability of the demand nodes and of the overall availability of the gas network. To assess the importance of each network component, risk achievement worth and risk reduction worth importance measures are utilized. The aim of the developed method is to assess potential weakness in the gas network as well as to be used as an analysis tool during expansion planning and maintenance scheduling activities. The framework developed in the paper leverages on steady-state analysis of the gas network performed using a physical flow/pressure model. The impact of a component failure on the gas supply interruption at different demand nods is assed and contrasted to the PRA results. The framework is exemplified with reference to the reduced UK gas network. The results provide insights to support a robust reliability assessment of the gas network. Moreover, the probabilistic mapping of the most important components in the gas network provides the means for assessing optimal strategies for maintenance schedule as well as to prioritize improvements of the gas network aiming for effective risk reduction.

1 INTRODUCTION

In the past decades, the consumption of gas in Europe increased significantly (Weisser, 2007). Natural gas is considered of the essence for the energy security in the European Union, by comprising a quarter of the primary energy supply to electrical power generation, households, feedstock for industry and fuel for transportation. Considering the decrease of domestic gas production and thus the higher import dependence, the need to address security of supply has increased (EU, 2010). As a response, the Regulation (EU) No. 994/2010 of the European parliament and council of 20 October 2010 concerning measures to safeguard security of gas supply and repealing Council Directive 2004/67/EC was released.

The International Energy Agency (IEA) shows that, in many IEA member countries, the electricity generation sector is especially dependent on natural gas, with tendency to grow. In 14 countries, gas accounts for over 20% of the electrical energy generation and more than 30% in nine countries, while in five countries more than half of the generation is dependent on gas (Simpson and Min, 2011). A sense of comfort exists over the gas markets capability to successfully adjust to demand or supply shocks due to low gas prices, along with expectations for continued well-supplied gas markets over the medium term. However, the IEA global security review shows that the current situation about gas security comfort may change suddenly as market conditions change (IEA, 2016).

Researchers have placed significant effort in modeling gas networks to provide accurate assessments of the network behavior, which can be of significance for the security of supply. In (Osiadacz, 1987), a thorough description of steady state and transient simulation models for gas network analysis is presented. Security of supply is investigated during conflicts and crisis in Europe in (Carvalho et al., 2014), and resilient response strategies to supply disruption events under relevant scenarios are provided. Furthermore, (Antenucci and Sansavini, 2017) identifies the contingencies that can jeopardize the coupled gas and power systems security under increasing gas demand scenarios.

A probabilistic model that studies the security of supply in a gas network is presented in (Praks et al., 2015). The model utilizes Monte-Carlo simulations along with graph theory to perform analyses on a real size gas network, i.e. an unspecified part of the EU gas network. Even though the model provides comprehensive simulation of the security of supply in a gas network, it lacks the support of a physical model. A fault-tree based analyses of the security of supply in a gas network is shown in (Praks et al., 2014). The paper argues that the application of a fault tree method on a real size gas network requires an automatic fault trees generation algorithm.

The limited choice of methods for the assessment of the gas network security of supply is the main motivation behind this research. The framework presented herein aims at assessing the reliability of a gas network by gaining insights from the application of both probabilistic and deterministic analyses. A PRA method based on the fault tree technique is developed. The method uses a procedure for the automatic generation of fault trees for a selected gas demand node. The unwanted event, i.e. the top event is defined as the "node not supplied with gas". The risk of having gas demand not supplied is a function of the gas network architecture and of the failure probabilities of the components of which the network is comprised. A global risk measure is introduced to evaluate the overall gas network security of supply. Different importance measures are employed to assess the importance of each network component with respect to each individual demand nodes, as well as to entire gas network. A physical analysis model is employed to assess the behavior of the gas flows, thus the gas properties in the system. Combining the PRA model analyses with the physical model analyses results in a complimentary platform capable of performing robust risk analyses of the security of supply of gas networks. Thus, the obtained results, besides providing various risk measures, bring together the complimentary insights from two conceptually different models.

The paper is structured as follows: Section 2 describes the PRA methodology used to assess the gas network reliability; Section 3 presents the physical model for simulating the gas network operational conditions; Section 4 presents the performed analyses and the obtained results; Section 5 gives conclusion remarks.

2 PRA METHOD FOR GAS NETWORKS RELIABILITY ANALYSES

The PRA supported by the fault tree and event tree analyses has been extensively used for risk evaluations in various engineering domains, i.e. nuclear power plant safety, aerospace design, power system reliability, etc.. The fault tree is a deductive analytical method, where an undesired system state is specified and the system is then assessed in relation of its operations and environment to find all relevant ways in which the undesired event can occur.

In this paper, a method for probabilistic risk assessment of gas networks is presented. The method exploits a procedure for the automatic generation of fault trees that was originally proposed for power system reliability analyses (Volkanovski et al., 2009). Herein we adapt the method for its application to real-size gas networks. The graphical representation of the procedure for the automatic generation of fault trees is shown in Figure 1.

The procedure starts by defining the adjacency matrix (Figure 1) of the corresponding network graph. The adjacency matrix of a graph, also known as the connection matrix, is a square matrix A (v_i, v_j) such that the element $A_{i,j}$ is equal to one when there is an edge from vertex i to vertex j, and equal to zero otherwise (Biggs, 1993). The adjacency matrix is used to determine all possible flow pats that connect a demand node with all the source nodes in the gas network. For example, if demand node 1 is connected to the source node 4 through node 2 and node 3, two possible paths can lead from node 1 to the source node 4. In the process of identifying a path from a demand node to a source node, no single component can be used twice in the same path. This constraint prevents the path from looping, i.e. repeating the same set of components in the same connectivity path. In addition, a path is only completed if it starts with a demand node and ends with a source node.

The two identified pats (Figure 1, Connectivity Paths: path one [1 2 3 4] and path two [1 3 4]) through which the demand at node 1 can be served from the source node are the foundation for the creation of the fault tree. It is a top-down procedure where the fault tree is created starting from the demand node and continues by unfolding the identified paths. The demand at node 1 is not supplied with gas if node 1 fails or the interruption of gas delivery to node 1 occurs, which is the first gate in the fault tree which is being written by the procedure for automatic generation of fault trees and represented by Boolean OR. The interruption of gas to node 1 occurs if both the gas pipeline between node 1 and node 2, which is on the first path, and the gas pipeline between nodes 1 and node 3, which is on the second path, do not deliver gas to node 1. This is the second gate in the fault tree and it is represented by a Boolean AND. The gas pipeline between node

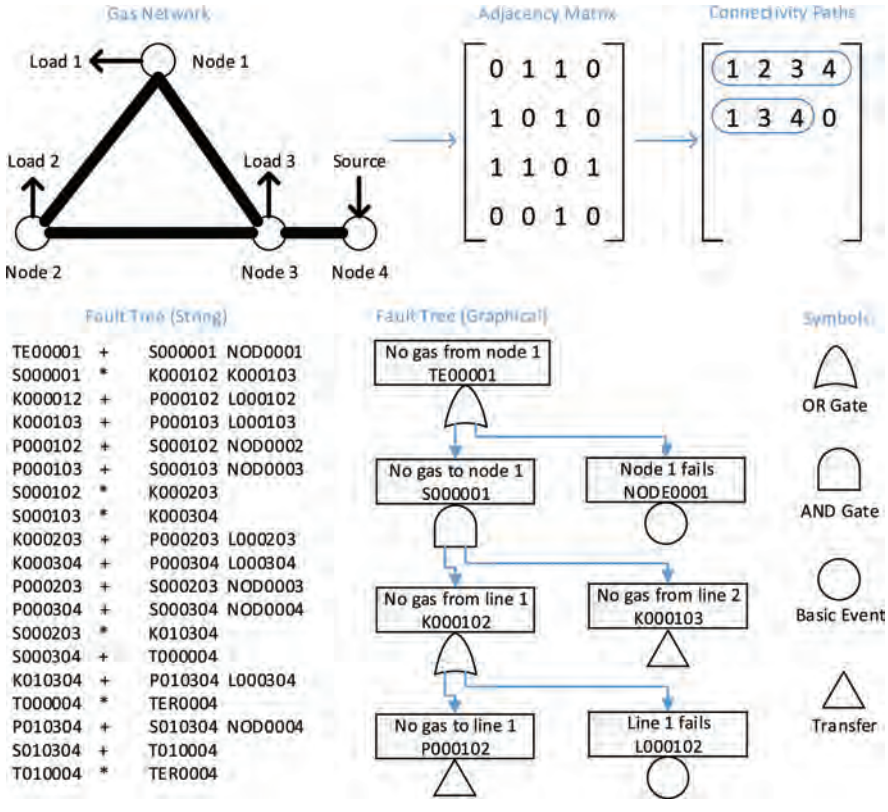


Figure 1. Graphical representation of the procedure for automatic fault tree creation.

1 and node 2 does not deliver gas to node 1 if either the pipeline fails or there is no gas delivered to the pipeline from elsewhere, which is represented by another OR gate. The process of unfolding the first path ends when the source at node 4 is reached. The same operation is repeated for the second path. The output of the algorithm for the automatic generation of fault trees for node 1 is a string-based fault tree given in the lower left part of Figure 1. The "*" denotes AND gate, while "+" denotes OR gate. The first line "TE00001 + S000001 NOD0001" denotes the occurrence of the top event "no gas from node 1" (TE00001) due to either there is "no gas to node 1" (S000001) or "node 1 fails" (NOD0001). The rest of the lines follows the same rationale. The graphical equivalent of the fault tree is presented in the lower mid part of Figure 1. The representation of each symbol is given in the lower right part of Figure 1.

2.1 *Fault tree quantification*

The method, when applied on a real-size gas network, results in large fault trees, comprising tens of thousandths to few hundred of thousandths lines. Solving a large fault trees efficiently is known to be a challenging problem (Contini and Matuzas, 2011). In general, fault trees can be converted into equivalent set of Boolean logical equations. The quantitative analyses of a fault tree is represented by the Rare Event Approximation method (Roberts et al., 1981):

$$Q_L = \sum_{i=1}^{n} Q_{MCS_i(BE_1,\ldots,BE_m)} = \sum_{i=1}^{n} \prod_{j=1}^{m} Q_{BE_{ij}} \quad (1)$$

where Q_L is the top event probability of occurrence, n is the number of minimal cut sets (MCS), i.e. the smallest set of basic events that induce the top event when occurring simultaneously, $Q_{MCS_i(BE_1,\ldots,BE_m)}$ is the probability of occurrence of ith MCS, which is comprised of m basic events, and $Q_{BE_{ij}}$ is the probability of occurrence of the jth basic event within the ith MCS.

2.2 *Network risk measure*

In this paper, a unique fault tree for each demand node in the gas network is created by employing

the proposed procedure. The fault trees are solved using the Rare Event Approximation method (Equation 1), and the probability of each demand node not being supplied with gas is obtained.

The risk of a gas demand not being supplied is a function of the gas networks architecture and of the probabilities of its components. A global risk measure is defined (Volkanovski et al., 2009) to assess the overall gas network security of supply:

$$U_{GS} = \sum_{i=1}^{LN} Q_{L_i} \frac{L_i}{\sum_{i=1}^{LN} L_i} \tag{2}$$

where U_{GS} is the gas system unavailability, Q_{L_i} is the probability of failure of gas supply to the ith demand node, LN is the number of nodes with load demands, L_i is the gas demand at the ith node.

2.3 *Importance measures*

The performance of a system (e.g. gas network, power system) depends on its components. Some components contribute more to the failure of the system than others. Therefore, the concept of importance plays a major role in the quantification of risk in engineered systems. In this paper two of the most frequently exploited importance measures in PRA are utilized, i.e. the Risk Achievement Worth (RAW) and Risk Reduction Worth (RRW). The RAW estimates the value of risk increase if the failure probability of a basic event is equal to one (component out of service):

$$RAW_j = \frac{Q_L\left(Q_{BE_j} = 1\right)}{Q_L} \tag{3}$$

where RAW_j is the risk achievement worth of basic event j, the $Q_L\left(Q_{BE_j} = 1\right)$ is the top event probability when the probability of occurrence of basic event j, Q_{BE_j}, is equal to one. The RAW determines the maximum increase of risk due to occurrence of the basic event j, i.e. identifies the components that need to be efficiently maintained such that the reliability of the system will not decrease (Volkanovski et al., 2009).

The RRW estimates the value of risk decrease if the failure probability of the basic event is equal to zero (the component never fails):

$$RRW_j = \frac{Q_L}{Q_L\left(Q_{BE_j} = 0\right)} \tag{4}$$

where RRW_j is the risk reduction worth of basic event j, the $Q_L(Q_{BE_j} = 0)$ is the top event probability when the probability of occurrence of basic event j, Q_{BE_j}, is equal to zero. The RRW determines the maximum reduction of risk due to perfect reliability of the component associated to the basic event j, i.e. identifies the redundancy level of a component associated with the basic event j (Volkanovski et al., 2009). In most cases, the basic event is associated with component unavailability (van der Borst and Schoonakker, 2001).

2.4 *Network importance measures*

A fault tree is created for each node with gas demand L_i, and the probability of failure of gas supply (Q_{L_i}) to the ith node is calculated. The RAW and RRW importance measures are calculated for each demand node, resulting in a unique set of importance values for each component (basic event). The importance values of the same basic event with respect to different nodes may significantly differ among each other. Hence, a component may be very important for some demand nodes and irrelevant for other demand nodes. In order to estimate the importance of each component on a global level, i.e. for the overall gas network, the network importance measures are introduced, i.e. the network risk achievement worth (NRAW) and the network risk reduction worth (NRRW) as shown in (Volkanovski et al., 2009).

3 PHYSICAL MODEL SIMULATION OF GAS NETWORKS

For a reliable representation of the gas system, several components need to be modelled, including pipelines and non-pipe elements, such as compressors, terminals and storages. Gas flow within a pipeline is represented with a steady state model. Pressure-drops in the network are modelled via the Panhandle "A" equation (Osiadacz, 1987):

$$p_1^2 - p_2^2 = 18.43 \cdot \frac{L \cdot Q^{1.854}}{E^2 \cdot D^{4.854}} \tag{5}$$

where p_1 and p_2 represent the pressure at the beginning and at the end of a pipeline, L is the pipeline length, Q is the volume flow rate through the pipeline, E is the efficiency factor and D is the pipeline diameter. The newton-node loop method is employed for solving the system of equations which define the entire network (Osiadacz, 1987). This iterative method is based on the Kirchhoff's second low, which states that the sum of the pressure-drops around any closed loop in the network is zero.

Compressor stations are modelled as fictitious branches, and a constant pressure ratio between the pressures at the sending and receiving node

of each compressor is considered. Terminals and storages injections into the network are proportional to their delivering capacity. The gas flow is allowed in both directions in the modeling of the network components.

3.1 *Physical importance measure: Total gas curtailment*

The importance of a component is evaluated by the induced change in the operation of the gas network. In fact, when a component is removed, the way the gas flows into the network varies and pressures change accordingly. However, in case pressures exceed the minimum or maximum safety pressures, actions are implemented in order to restore the normal operating conditions. In particular, in case of minimum pressure violation, gas curtailments are enforced in the location of the violation. The Total Gas Curtailment (TGC), which follows the basic event in order to bring the system back within safety margins, is considered as importance measure. The chosen matrix addresses the impact of a single component removal on the entire network, and it is formally expressed as:

$$TGC_j = \sum_{n=1}^{N} GC_n\left(Q\left(BE_j\right)=1\right) \tag{6}$$

where TGC_j is the importance measure for the basic event *j, N* is the number of nodes in the system and GC_n is the gas curtailment at node *n.*

4 ANALYSES AND RESULTS

The framework developed in this paper is applied to the UK gas network. The simplified transmission grid is constituted by 61 pipelines and 21 compressor stations that work with a constant pressure ratio. Safety operations are bounded in the pressure range of [38 85] bars. The gas system counts 9

Table 1. Top event probabilities for all demand nodes in the UK gas network.

Node	Demand (m^3/h)	Failure probability	Weighting factor	Weighted failure probability
3	2.85E+05	7.15E-05	1.50E-02	1.08E-06
5	7.63E+05	7.72E-06	4.02E-02	3.11E-07
7	2.68E+05	4.36E-06	1.41E-02	6.16E-08
9	7.19E+05	3.74E-08	3.79E-02	1.42E-09
11	2.44E+04	9.11E-06	1.29E-03	1.17E-08
12	2.37E+05	1.25E-07	1.25E-02	1.56E-09
14	1.49E+06	2.67E-07	7.86E-02	2.09E-08
15	5.10E+05	1.20E-10	2.69E-02	3.23E-12
17	1.88E+05	3.90E-11	9.94E-03	3.87E-13
18	1.27E+06	4.72E-06	6.72E-02	3.17E-07
22	2.04E+05	4.92E-09	1.07E-02	5.29E-11
24	2.29E+04	4.46E-06	1.21E-03	5.38E-09
26	3.80E+05	1.56E-06	2.00E-02	3.13E-08
28	9.45E+04	2.02E-09	4.98E-03	1.01E-11
30	9.28E+05	8.88E-08	4.89E-02	4.34E-09
31	2.08E+05	6.27E-05	1.10E-02	6.89E-07
32	5.97E+05	1.89E-06	3.15E-02	5.96E-08
33	1.43E+06	2.58E-04	7.52E-02	1.94E-05
34	1.45E+06	3.57E-08	7.62E-02	2.72E-09
36	3.43E+05	1.06E-05	1.81E-02	1.92E-07
37	5.92E+05	4.41E-05	3.12E-02	1.38E-06
38	1.88E+05	4.47E-06	9.92E-03	4.43E-08
39	6.30E+05	2.64E-05	3.32E-02	8.76E-07
40	2.62E+05	4.99E-06	1.38E-02	6.89E-08
41	1.77E+06	2.22E-06	9.32E-02	2.07E-07
44	1.25E+06	3.71E-08	6.60E-02	2.45E-09
46	6.04E+05	1.81E-08	3.18E-02	5.76E-10
48	6.59E+05	1.24E-05	3.48E-02	4.30E-07
52	1.39E+05	7.30E-06	7.32E-03	5.34E-08
53	9.41E+05	6.24E-06	4.96E-02	3.10E-07
54	3.14E+05	1.95E-06	1.66E-02	3.23E-08
55	2.05E+05	4.98E-04	1.08E-02	5.38E-06

terminals and 9 storage facilities. The UK gas network topological data is adopted from (Qadrdan et al., 2010), including the maximum supply capacities of different terminals and the characteristics of the storage facilities.

The failure probability data is adopted from (Praks et al., 2015). Considering one-year inspection time, the failure probability (complete rapture) of gas pipeline in European gas transmission system is 3.5E-5 per kilometer, the failure probability of a compressor is 2.5E-1, the failure probability of the gas terminal is 1.5E-1, and the failure probability of a gas storage is 1E–1. The failure of a compressor does not necessarily interrupt the gas flow through the compressor station, i.e. the gas flows through the compressor bypass. The failure probability of the bypass is based on the failure probability of a disconnection valve, i.e. "valve fail to open". Due to lack of specific data, for valves used in the bypasses at compressor stations a value of 1.6E-3 is taken from (IAEA, 1988).

4.1 *Probabilistic risk analyses of the UK gas network*

The fault tree method is applied for each load demand node in the UK gas network, thus the failure probability of gas supply at each node is calculated and the obtained results are presented in Table 1. In all of the fault tree analyses, the maximum number of basic events in a minimal cut set is truncated at 10, while the minimum failure probability of minimal cut set is 1E-15.

The first column from Table 1 represents the nodes with gas demands, while the second column represents the average gas demand per node in m^3/h. Node 41 is the node with the highest demand of gas, with average requirement of 1.77E+06 (m^3/h). The total average gas demand in the network is 1.90E+07 (m^3/h), while the total gas supply capacity is 3.78E+07 (m^3/h) including the gas storages which can provide gas for only limited number of time. The third column represents the top event probability of each demand node. The fourth column represents the weighting value of each demand node. The fifth column represents the weighted failure probability of gas supply at each node, calculated as the product of the top event probability and the weighting factor of the respective node. The total gas network failure probability is 3.00E-5 and is calculated with Equation 2 based on the individual demand node failure probability. Table 1 shows that in general the nodes with highest gas demands have the largest weighted probability of failure.

The network importance measures, NRAW and NRRW, for the UK gas network are calculated and given in Figure 2 a) and b), respectively. Furthermore, the RAW and the RRW results from the fault tree analyses performed for the largest load demand node 41 in the network are given in Figure 3.

Figure 2 shows the most important components in the gas network, according to the network importance measures presented in Section 2.4. The color defines the importance of each of the components shown in the figure, i.e. brighter colors are associated to components with higher importance values. According the NRAW, the most important gas pipelines to keep operational in the system are the pipeline between node 31 and node 32 with NRAW of 2726.8, and the pipeline between node 1 and node 3 with NRAW of 397.7. The pipeline between node

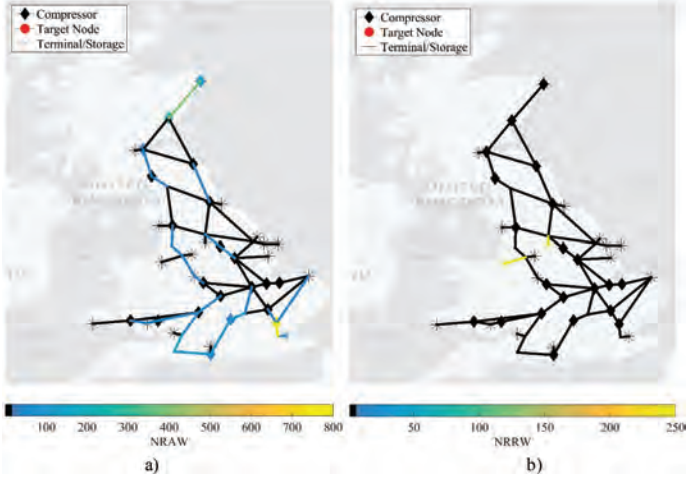


Figure 2. Representation of the most important components in the network according to: a) NRAW and b) NRRW.

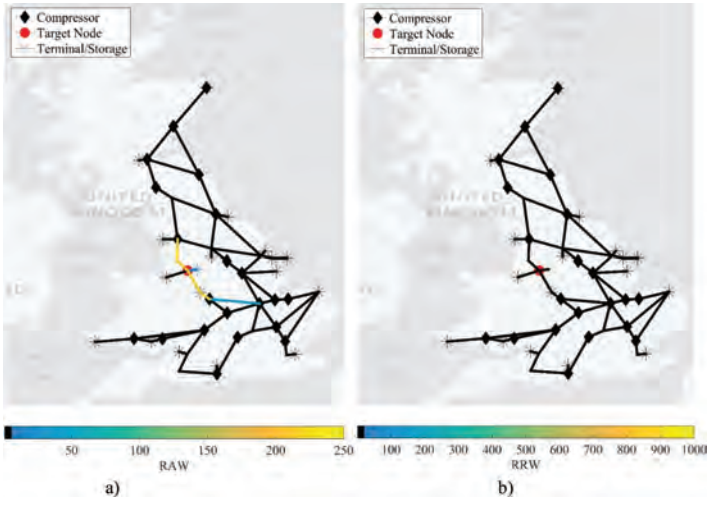


Figure 3. Node 41 component importance according to: a) RAW and b) RRW.

31 and node 32 is providing connection between the network and the gas terminal at node 59, IOG, which is positioned in the South of the island and it is the only terminal in the southern region besides the storages at node 68 and node 69 and the Bacton terminal in the South-East. The pipeline between node 1 and node 3 is connecting the network to the terminal at node 56, St. Fergus, which is the second largest gas source in the system and is positioned in the far North of the island. Furthermore, among the most important components in the gas network are the compressor bypasses between node 31 and node 32 with NRAW of 786.9, and between nodes 2 and 3 with NRAW of 242.4. Both bypasses are part of the compressor stations that provide connectivity to the terminals at node 59 and node 56, respectively. From the presented results, it can be deduced that performing maintenance simultaneously on any of the above components may significantly decrease the security of supply to the consumers in the gas network. For example, the simultaneous unavailability of the gas pipeline between node 31 and node 32 and the compressor station between node 2 and node 3, due to maintenance activities may have a significant impact to the gas network reliability. Based on the conducted analyses, it is possible to optimally prioritize the maintenance activities in the gas network, thus maximizing the security of supply.

On the other hand, Figure 2 b) shows the components with the highest risk reduction importance, i.e. the redundancy level of each of the components in the gas network. According to NRRW the most important components in the gas network are the gas terminal at node 61 and the pipeline connecting this node with node 41 and thus to the rest of network, as well as the gas storage at node 66 and the pipeline connecting this node with node 15 and thus the rest of the network, all with NRRW value of 243.7. Decreasing the failure probability of these components (e.g. by installing more reliable components) will result with increased security of supply. In other words, the obtained NRRW results can help in prioritizing feature improvements on the gas network, leading to the largest risk reduction.

Figure 3 a) and b) show the RAW and RRW values, respectively, for all gas network components when considering node 41 not supplied with gas as top event. The pipeline connecting node 41 to node 42 and the pipeline connecting node 42 to node 44 have the highest RAW of 214.6 and 214.5, respectively, making them the most important components with respect to the gas supply to node 41. On the other hand, the storage at node 65 is by far the most important component according to the RRW importance measure with value of 953.4, and the second most important component is the pipeline connecting node 13 to node 40 with RAW of 3.8.

4.2 *Steady-state simulations of the UK gas network based on the PRA results*

The physical model simulates the gas network behavior, and pressures and mass flows in the pipelines are computed. A steady-state analysis is performed for the loss of each network component and the consequent total gas curtailment is calculated (Table 2). Gas curtailments are specified for each gas demand node and for the entire network.

Table 2. The TGS effect of the 20 most important gas network components obtained using NRAW.

Component	Probability of failure	NRAW	TGS (m^3/h)	Risk
L 31-33	1.30E-03	2726.8	0	0
B 31-32	1.60E-03	786.9	0	0
L 01-03	2.45E-03	397.7	0	0
B 02-03	1.60E-03	242.4	0	0
L 39-54	2.10E-03	160.9	6.18E+04	1.30E+02
B 01-55	2.50E-01	105.1	0	0
L 34-36	1.51E-03	103.0	4.12E+06	6.20E+03
L 37-39	3.40E-03	92.2	7.17E+05	2.43E+03
L 35-38	5.08E-03	83.9	1.38E+06	7.03E+03
L 18-21	1.79E-03	49.9	4.53E+05	8.09E+02
L 47-49	2.10E-03	45.3	1.13E+05	2.38E+02
L 45-53	2.28E-03	43.2	1.59E+06	3.62E+03
L 33-59	8.05E-04	39.5	3.18E+05	2.56E+02
L 15-19	2.24E-03	36.1	5.71E+05	1.28E+03
TER-59	1.50E-01	34.8	3.18E+05	4.77E+04
L 47-54	2.45E-03	33.9	0	0
L 50-52	1.58E-03	30.1	0	0
L 48-50	1.54E-03	30.1	0	0
L 40-41	1.23E-03	29.5	1.41E+06	1.73E+03
B 04-05	1.60E-03	28.6	0	0

The first column from Table 1 represents the component/basic event name, such that the letters denote the component type (i.e. L stands for pipeline, B stands for compressor bypass, C stand for compressor, TER stands for gas terminal and TES stands for gas storage), while the numerical digits represents the nodes where the respective components are connected. The second column gives the failure probabilities of the 20 most important elements according to NRAW, and the third column represents NRAW values of these components. The fourth column gives the calculated TGS for the respective components represented by the first column. The fifth column gives the product of the component probability and its TGC impact on the gas network. In the physical model the compressor station failure is represented only by a compressor failure without losing the flow of gas through the respective branch in the gas network, instead the compressor station ratio is equal to 1, i.e. the gas pressure before and after the compressor station is remaining the same. Therefore, the TGS of zero caused by compressor bypass failures is by default, since no such failure is simulated. The obtained results show that the TER-59 (i.e. the IOG terminal connected at node 59) is the one with the highest risk impact on the gas network. Furthermore, a high risk impact is expected by the loss of the pipeline between node 25 and node 38 (L 35-38) which is one of the main links providing connectivity between the South and the South-West part of the gas network, and from the pipeline between node 45 and node 53 (L 45-53) which is one of the main links providing connectivity between the South-West and the central part of the gas network. Remarkably, the loss of some elements, such as the pipeline between node 31 and node 33 (L 31-33) or the pipeline between node 1 and node 3 (L 01-03), induce no pressure violations in the network, despite their large NRAW values. Therefore, irrespective of the relevance of these components from a probabilistic and network connectivity perspective, gas network operations manage efficiently the gas re-routing via different terminals and storages to sustain gas supply.

5 CONCLUSIONS

A framework for gas network security of supply analyses is presented. The framework is based on a PRA method which employs a procedure for automatic generation of fault trees related to a specific top event, i.e. gas demand node not supplied. The risk achievement worth and risk reduction worth importance measures are utilized to estimate the importance of the network components for the security of supply of the individual demand nodes and the overall gas network. Furthermore, a physical model for the simulation of the gas network behavior is developed. The model is capable of computing gas pressures throughout the network and of enforcing curtailments if single or multiple contingencies occur. The physical model is utilized

to supplement the PRA method by providing accurate estimates of the gas network conditions.

The PRA results show the most important components, based on risk increase and risk reduction measures, for the security of supply to each individual node and the entire network. The NRAW importance measures provides us with results that can be used to prioritize the maintenance activities in the overall gas network, while the RAW can be used to schedule the maintenance activities with respect of each individual node. The NRRW importance measure provides us with results that can be used to prioritize future improvements in the overall gas network thus increasing the security of supply, while the RRW importance measure can be used to schedule potential improvements that will increase the security of supply of each individual gas demand node. Furthermore, the physical model results show that even though some components are identified as one of the most important, according to the PRA importance measures, their failure may not have significant effect on the gas supply, because of the capability of the gas network to perform efficiently and provide gas re-routing via different terminals and storages to sustain gas supply.

REFERENCES

ANTENUCCI, A. & SANSAVINI, G. 2017. Adequacy and Security Analysis of Interdependent Electric and Gas Networks. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, DOI: 10.1177/1748006X17715953.

BIGGS, N. 1993. *Algebraic Graph Theory, Cambridge Mathematical Library (2nd ed.)*, Cambridge University Press.

CARVALHO, R., BUZNA, L., BONO, F., MASERA, M., ARROWSMITH, D.K. & HELBING, D. 2014. Resilience of natural gas networks during conflicts, crises and disruptions. *PloS one,* 9, e90265.

CONTINI, S. & MATUZAS, V. 2011. Analysis of large fault trees based on functional decomposition. *Reliability Engineering & System Safety,* 96, 383–390.

EU 2010. Regulation No. 994/2010 of the European Parliament and of the Council of 20 October 2010 concerning measures to safeguard security of gas supply and repealing Council Directive 2004/67/EC. *Official Journal of the European Union*.

IAEA 1988. TECDOC-478, Component Reliability Data for Use in PSA. Vienna.

IEA 2016. Global Gas Security Review: How Flexible are LNG Markets in Practice? *International Energy Agency (OECD/IEA)*.

OSIADACZ, A. 1987. *Simulation and analysis of gas networks*, Gulf Publishing Company, Houston, TX; None.

PRAKS, P., CONTINI, S. & KOPUSTINSKAS, V. Monte Carlo and fault tree aproaches in reliability applications of gas transmission network. Proceedings of the 2014 15th International Scientific Conference on Electric Power Engineering (EPE), 12–14 May 2014. 69–74.

PRAKS, P., KOPUSTINSKAS, V. & MASERA, M. 2015. Probabilistic modelling of security of supply in gas networks and evaluation of new infrastructure. *Reliability Engineering & System Safety,* 144, 254–264.

QADRDAN, M., CHAUDRY, M., WU, J., JENKINS, N. & EKANAYAKE, J. 2010. Impact of a large penetration of wind generation on the GB gas network. *Energy Policy,* 38, 5684–5695.

ROBERTS, N.H., VESELY, W.E., HAASL, D.F. & GOLDBERG, F.F. 1981. Fault Tree Handbook. *NUREG-0492, US NRC. Washington*.

SIMPSON, J. & MIN, K.-S. 2011. Gas Emergency Policy: Where do IEA Member Countries Stand? *International Energy Agency (OECD/IEA)*.

VAN DER BORST, M. & SCHOONAKKER, H. 2001. An overview of PSA importance measures. *Reliability Engineering & System Safety,* 72, 241–245.

VOLKANOVSKI, A., ČEPIN, M. & MAVKO, B. 2009. Application of the fault tree analysis for assessment of power system reliability. *Reliability Engineering & System Safety,* 94, 1116–1127.

WEISSER, H. 2007. The security of gas supply—a critical issue for Europe? *Energy Policy,* 35, 1–5.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A risk-based approach for the analysis of LNG carriers port operations

F. Ovidi
University of Pisa, Pisa, Italy

G. Landucci
University of Pisa, Pisa, Italy
University of Leiden, Den Haag, The Netherlands

L. Picconi & T. Chiavistelli
Chemical Controls Srl, Livorno, Italy

ABSTRACT: Liquefied Natural Gas (LNG) as ship fuel is considered a viable solution to marine environmental issues, due to the significant reduction in emissions with respect to conventional fuel oil. At the same time, safety issues may arise in port areas when LNG is used as a fuel due to its high flammability. The present work focuses on the safety assessment of LNG carriers approaching a bunkering terminal trough port channels located in an industrial area. A risk matrix approach is adopted to evaluate the risk level associated with the carrier approaching the harbour, considering the vulnerability of surrounding territory and potential interactions with industrial facilities located in the area. The methodology is applied to a case study of industrial interest showing the potential of the tool in supporting risk-based decision making.

1 INTRODUCTION

Liquefied Natural Gas (LNG) as ship fuel is considered as a viable solution to marine environmental issues, due to the significant reduction in emissions with respect to conventional fuel oil (Bittante et al. 2017). Therefore, due to the potential benefits related to this technology, several projects have been proposed for the realisation of LNG bunkering terminals in harbour areas, contributing to the development of LNG infrastructure network.

However, safety issues may arise in port areas when LNG is used as a fuel due to the high flammability of this substance, with the potential of severe fires and explosion scenarios (Jeong et al. 2017). Thus, in the stages of early development and selection of LNG bunkering and ship supply technologies in port areas, safety aspects will become crucial to develop sustainable and reliable technologies involving LNG as marine fuel.

Several studies were reported in the literature concerning the safety of LNG distribution chain, addressing the analysis of LNG regasification terminals, bunkering stations, ships fuel systems. Yun et al. (2009) proposed a risk assessment methodology for LNG terminals by incorporating Bayesian and LOPA (Layers of Protection Analysis) approaches. An inherent safety based approach was proposed in (Tugnoli et al. 2012) to estimate the safety aspects or alternative LNG regasification technologies.

Concerning the analysis of LNG for marine fuel application, several studies were presented (ABS 2014, ADN Administrative Committee 2014). Lee et al. (2015) compared the fire risk assessments of two types of LNG fuel gas supply systems. DNV (2012) conducted a site-specific quantitative risk assessment of LNG bunkering in an effort to determine a safe distance for passing ships at the Port of Rotterdam.

However, a structured methodology to perform feasibility studies for access of LNG supply ships in harbour areas close to sensitive urban areas and industrial parks is still lacking in the literature.

The present work shows a risk-based approach to support the feasibility study of LNG ships access to harbour areas. The approach integrates geometrical and ship size considerations, legislative framework and safety aspects, and is applied to a reference case study, located in the Port of Venice (Italy).

The paper is structured as follows: in Section 2 the case study object of the feasibility study is described; in Section 3, the overview of the methodology is presented; in Section 4 details on the safety and risk assessment are provided; Section 5 shows the results of the analysis, which are discussed in Section 6. Conclusions and recommendations are given in Section 7.

2 DESCRIPTION OF THE REFERENCE CASE STUDY

The case study is located in the Port of Venice, that is one of the most important industrial sites in Italy (Zonta et al. 2007). The port is strongly interconnected with the Marghera industrial area, known as "Porto Marghera". Figure 1 shows the overview of Porto Marghera area. The area is accessible through the Malamocco channel (see paths 3 and 4 in Fig. 1), which connects Porto Marghera with an artificial channel leading to the Adriatic Sea (paths 1 and 2 in Fig. 1).

Porto Marghera is the site selected for a new bunkering terminal for LNG storage and distribution. The bunkering station will be located in the Southern part of the industrial channel (path 5 in Fig. 1); in the following the channel is labelled as "SIC".

2.1 *Description of the SIC*

SIC is accessible through the basin located in the Malamocco-Marghera channel; a detailed view of SIC is shown in Figure 2. The channel length is about 4 km and the draft ranges between 6 and 10.1 m.

The LNG bunkering terminal will be served by LNG carriers accessing from the Northern Adriatic Sea through the path shown in Figure 1 towards the terminal located in the SIC. This ship traffic may increase the risk level of the surrounding area due to possible accidents involving the spill and consequent ignition of LNG from the carriers.

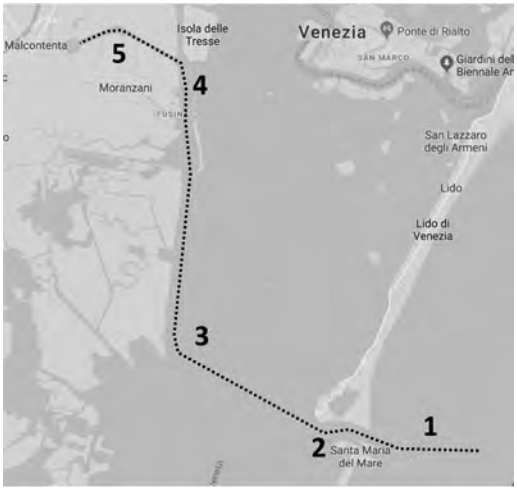


Figure 1. LNG carrier planned route through Porto Marghera: 1) roadstead in the Northern Adriatic Sea; 2) harbour inlet and artificial canal; 3) Malamocco-Marghera channel (before the industrial area); 4) Malamocco-Marghera channel (inside the industrial area), 5) SIC (Southern part of the industrial channel).

Figure 2. Detail view of SIC. The squared symbols indicate the location of the industrial berths, and the round symbols indicate the location of civil/commercial installations.

The critical areas along the LNG ship route are summarized in Figure 2. They are constituted by benches (either industrial or civil/commercial), industrial facilities, and the areas dedicated to civil activities or other services.

For what concerns the benches, along the route there are: 4 commercial berths located near the Fusina area (labelled with C in Fig. 2); 14 industrial berths (labelled with U in Fig. 2), 7 of which serving O&G facilities. The industrial activities are mostly chemical and petrochemical plants, and some smaller construction companies. There are some civil areas located close to the Malamocco-Marghera channel.

The civil installations of interest for this study are: two camping sites, e.g., Fusina and Darsena Fusina, the Venice Ro Port for passengers' transport (P01, P02, and P03 in Fig. 2, respectively), and smaller civil areas (beaches, little dockyards, etc.) located in the artificial canal connected to the Northern Adriatic Sea (path 2 in Fig. 1).

The new bunkering terminal location is foreseen at the end of the SIC, before the SIC final enlargement (e.g., between U4 and U5 in Fig. 2).

3 METHODOLOGY

3.1 *Overview*

The aim of the present work is to provide a methodology to perform feasibility studies for the access of LNG carriers to harbour areas. As shown in Figure 3, the methodology is based on two main parts.

The first part focuses on the nautical accessibility of the gas carriers with respect to both geometrical issues and legislative aspects. This part is aimed providing accessibility evaluations,

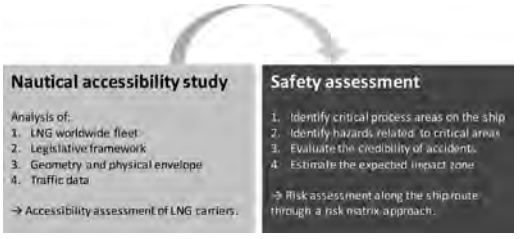


Figure 3. Overview of the methodology.

also considering possible mitigation or compensatory measures to eventually reduce suboptimal interactions of the LNG carriers with the current port configuration. In this way, potential events leading to collision or other manoeuvre upsets are identified and analysed in order to feed the safety assessment carried out in the second part of the methodology (see Fig. 3).

The safety assessment is aimed at investigating possible interaction between the LNG carriers and the neighbouring areas, either civil/commercial or industrial. Potential hazardous events affecting humans or process units in the surrounding industrial areas are considered. A risk register is compiled to determine the most critical scenarios and the associated vulnerability of the surrounding territory and industrial areas. Mitigation actions for critical scenarios are proposed to eventually reduce the risk level.

3.2 *Nautical accessibility study*

The nautical accessibility of LNG carriers is evaluated considering both geometrical aspects and the legislative framework (see Fig. 3).

Firstly, the type of gas tankers selected for the bunkering activities are analysed considering the worldwide LNG fleet. Secondarily, regulations and ordinances on the entry and the exit of port channel are analysed to point out the case specific legislative issues. Finally, the ordinary port traffic and the average residence time of each ship at the quay are estimated through historical data, in order to evaluate the geometrical compatibility with the gas carriers.

3.2.1 *LNG carrier selection for the case study*

The considered size of LNG carriers of interest in the present application feature nominal capacity ranging from about hundred cubic meters up to about 40,000 m^3. The current worldwide fleet of ships presenting these features consists of about 50 ships, considering those in activity and those in delivery by 2017 (Lloyd's Register Marine 2015).

In the present analysis, two reference carriers are considered (namely, ship A and B), which features are summarized in Table 1.

3.2.2 *Legislative framework and port data analysis*

There are no specific regulations for LNG carriers access to the port provided from the North Adriatic Sea Port Authority. Whereas, there are three ordinances from the Venice Harbour Masters of Italian Cost Guard concerning gas carriers sailing into the Venice Port areas (Ordinance 2009, 2010, 2016). The requirements specify minimum manoeuvre speed, draft in the area (between 7.95 m and 10.4 m), limitations for the navigation of dangerous cargo (e.g. interdiction in case of fog), mandatory tug service depending on gross tonnage and envelopment required, and maximum width of the convoy (never exceeding 1/3 of the minimum width of channels to be covered).

Given the legislative framework, port data concerning ship fluxes are analysed in order to evaluate the maximum time allowable for the LNG carrier transit and operations, according to the physical free space left by standard historical traffic. Thus, the most critical berths are identified as those which are located in correspondence of the smaller breadth of SIC. The three potentially critical areas and the corresponding channel breadth are reported in Table 2.

Data about critical berths are part of the physical compatibility assessment. In fact, the latter is the result of matching the data collected for berths above listed and the geometry of the LNG carriers defined in section 3.2.1.

The geometrical analysis of critical points along the SIC supports the risk assessment evaluating the likelihood of accidental scenarios not related to process failures, namely evaluation of ships

Table 1. Characterization of the LNG carriers for the case study.

Item	Units	Ship A	Ship B
Reference	–	Wartsila (WSD50)	ENI fleet (IGU 2016)
Capacity	m^3	30,000	65,000
Length	m	170	216
Breadth	m	29.5	34
Draught	m	8	9.5

Table 2. SIC width in correspondence of critical berths, see Figure 2 for berth location.

Berth ID	Corresponding channel width (m)
U09, U10, U11	120
U08, U12	140
U05, U06, U07, U13	160

collision scenarios. Risk linked to accidental scenarios following conventional upsets/failures in the carrier LNG is object of the safety assessment summarized in Section 3.3.

3.3 *Overview of the safety and risk assessment*

The second part of the method focuses on safety aspects related to the ships approaching a bunkering station in port area (Fig. 3). The assessment of possible interaction between the LNG carrier and the neighbouring areas, both civil and industrial, follows a risk-based approach adopting risk matrix analysis. The first step is the evaluation of the critical areas in standard LNG carriers to perform hazards identification and to determine potential release events. Secondarily, the likelihood of scenarios following an accident leak (expressed in annual probability) is evaluated through standard failure frequency databases and event tree analysis (ETA). Then, the consequences evaluation is carried out through the use of standard literature models (Mannan 2005), implemented on the software DNV PHAST 7.1, and following a threshold-based approach (see Section 4.4.3). Finally, the credibility of the accidental scenarios and their consequences are combined through a risk matrix and the results are summarized in a risk register (see Section 5.2).

4 RISK-BASED ANALYSIS OF LNG CARRIERS APPROACHING HARBOUR AREAS

4.1 *Evaluation of critical areas on the LNG carrier*

The safety assessment is based on the identification of critical areas on the LNG carrier in order to determine the potential accidents. Collision with other ships moving in the channel is excluded from the present analysis, according to the results of the likelihood assessment and geometrical considerations (see Section 5.1). Thus, all the accidental release events are associated to process units exposed to the external environment, such as equipment and pipelines on the open deck. This is due to the fact that the structural failure of the main LNG storage vessels due to process failures is not considered as a credible event (Uijt de Haag & Ale 1999).

The reference LNG carriers shown in Section 3.2.1 and considered for the analysis, despite featuring relevant differences in the total inventory, share the same types of process equipment on deck with similar geometries. Thus, since the structural failure of the storage vessels is excluded, the potential release events are the same for both types of carriers.

During navigation in port areas, there are three critical zones onboard located on deck, from which an accidental release may develop. The possible leak sources are summarized in Table 3.

4.2 *Identification of release scenarios*

The Purple Book (Uijt de Haag & Ale 1999) guidelines for quantitative risk assessment of inland waterway transport are adopted to identify the release scenarios and for the estimation of related frequencies. The structural failure of one or more tanks is not considered credible, while possible damage to the connections is taken into account.

For "gas tanker" ships category, the Purple Book considers two release diameters (3″ = 76.2 mm and 6″ = 152.4 mm equivalent diameter) and it provides the related occurrence frequencies. However, it is worth to consider the following limitations about the applicability of the guideline to the present case:

- the approach was determined for navigable channels with length greater than 1 km, and it is suggested to perform an area-specific study for more reliable data;
- data refer to ships with maximum capacity of 4000 m^3, which are smaller than the ships of interest in the case study. Despite this, excluding the structural failure of the storage vessels, the equipment and the pipeline systems over-the-top shall be roughly identical regardless of the ship's capacity;
- in the present case study, major release diameters (6″) are excluded, considering the typical piping configurations.

Table 3. Critical areas characterization. T = temperature; P = pressure.

	Operating conditions		
Description	Phase	T (°C)	P (bar)
Liquid piping system on deck	liquid	−161 ÷ −141	1 ÷ 4
Boil Off Gas (BOG) piping/over-the-top vapour connections	vapour	−161 ÷ −141	1 ÷ 4
Steam piping manifold on deck	vapour	40	2

The reference release diameters selected in the present analysis are summarized in the following:

1. Large size release: equivalent diameter of 3″ (= 76.2 mm)
2. Small size release: equivalent diameter of 10 mm.

The latter rupture type is the reference minor rupture in for fixed process vessels (Uijt de Haag & Ale 1999).

In order to estimate the duration of the release scenarios, the liquid and vapor connections on the deck are not considered directly opened to the storage vessels (e.g., the lines are isolated through shut down valves in closed position). Thus, leakages from those sources will last since the entire inventory inside the pipes has been released. Whereas, the BOG piping and the over-the-top vapour connections are assumed in open connection with the storage tanks. In this case, the release will last since the emergency shutdown (ESD) system will close the valves. Following the indication of the IGC code (IMO 2016), ESD valves in liquid piping systems shall close fully and smoothly within 30 s of actuation. In this analysis, further 30 s are considered for detection and actuation, reaching a total release time of 60 s, in normal ESD operating conditions. In case of ESD system failure, the total release time is extended to a maximum time of 30 min.

4.3 *Frequencies evaluation*

The evaluation of the annual probability or accidental frequency linked to a single scenario is carried out based on the indications of Purple Book (Uijt de Haag & Ale 1999). The frequency (f_i) of the i-th accidental event (fire, explosion, dispersion, etc.) is calculated as follows:

$$f_i = F \cdot P_r \cdot P_{e,i} \quad (1)$$

where F = initial accident frequency (1/y), P_r = probability of having the release following the accident; $P_{e,i}$ = probability of having the i-th event given the considered release.

F is a function of the frequency of damage to a ship per unit distance (events/year per vessel per km) which depends on the type of channel. A conservative value of 1.4×10^{-6} events/(year × vessel × km) is considered in the present analysis (Uijt de Haag & Ale 1999); a single vessel is assumed to be involved in case of damage, and 20 km are considered as length of the ship route obtaining $F = 2.8 \times 10^{-5}$ 1/y.

The probability P_r expresses the possibility of having a release following a serious accident to the ship. This value depends on the type of ship and on type of size of the release; indications reported in the literature (Spouge 2005, Uijt de Haag & Ale 1999) allowed determining $P_r = 0.025$ and $P_r = 0.2025$ for large and small size release respectively.

Once the incidental release has occurred, the LNG can ignite immediately giving rise to a pool/jet fire. Otherwise, it may spread out forming a pool on the surface of the channel. The pool evaporation generates a vapor cloud, which can ignite resulting in a flash fire or even a vapour cloud explosion (VCE). Each event described above has a probability of occurrence, expressed through the term $P_{e,i}$; where i is an identifier of the type of scenario. Standard ETA (Mannan 2005) is carried out to quantify $P_{e,i}$ for each scenario.

The accidental scenarios related to the BOG piping and to the over-the-top vapour connections may be mitigated by ESD activation (see Section 4.2). In case of ESD system failure the resulting hazardous scenario is not mitigated. In this study, the ESD system is assumed as a SIL 2 "low-demand-mode" level. Thus, ESD failure probability is derived from the IEC 61508 standard (IEC 2010) and set equal to 10^{-2} and implemented in the ETA. More details on the ETA are reported elsewhere (Chemical Controls 2017).

4.4 *Consequences evaluation*

Consequence assessment is based on standard literature models for physical effect analysis (Mannan 2005) implemented in DNV GL Phast 7.1 software package. The main settings and assumptions are shown in the following.

4.4.1 *Schematization of LNG composition*

LNG is a liquid mixture of hydrocarbons composed mainly of methane, with small amounts of ethane, propane, nitrogen and other typical components of natural gas. In the present work, the presence of other compounds in addition to methane is neglected since it is not significant for the purpose of evaluating the consequences. The physical properties of LNG are taken from (Lentner et al. 2017, Mannan 2005).

4.4.2 *Meteorological conditions*

Two reference meteorological conditions are assumed in this study:

- F/2 – Pasquill stability class "stable" and wind speed of 2 m/s,
- D/5 – Pasquill stability class "neutral" and wind speed of 5 m/s.

Other relevant atmospheric parameters are summarized in Table 4.

4.4.3 *Threshold based approach for physical effects assessment*

The maximum damage distances (or vulnerability radii, r_{vul}) for the accident scenarios considered in the present study are evaluated through conservative threshold values, which are summarized in Table 5. Threshold values are derived from Italian legislation on land use planning (DM 2001). Either damages to humans (in terms of irreversible effects) or industrial equipment are considered.

4.5 *Definition of the reference risk matrix*

Once the likelihood/probability and consequences are quantitatively evaluated, the risk associated with each LNG accidental scenario is estimated through a reference risk matrix. The matrix is built upon likelihood and consequences classification following criteria derived from a previous study related to the oil and gas sector (Petrone et al. 2011).

Table 6 shows the criteria chosen for the classification of likelihood, based on the evaluated annual probability of each considered scenario.

Table 7 summarizes the criteria considered for the consequence assessment. Consequences are classified upon the comparison of damage distances (r_{vul}) against a set of reference distances associated with the position of sensitive targets. Two categories of sensitive targets are considered, namely:

Table 4. Atmospheric parameters set up for the consequences evaluation.

Parameter	Units	Value
Air temperature	°C	20
Water temperature	°C	20
Pressure	kPa	101.3
Relative humidity	%	50
Surface roughness length	mm	0.2
Solar radiation	kW/m^2	0.4

Table 5. Threshold values implemented in the present study, derived from (DM 2001). LFL = lower flammability limit; VCE = vapor cloud explosion.

	Threshold value	
Event	Humans	Industrial equipment
Flash Fire	LFL/2	–*
Pool Fire	5 kW/m^2	12.5 kW/m^2
VCE	0.07 barg	0.3 barg
Flare/Jet Fire	5 kW/m^2	12.5 kW/m^2

*Escalation is not credible.

Table 6. Probability/likelihood qualitative classification criterion.

Probability/likelihood (F)	Qualitative rating
$f_i < 10^{-6}$ 1/y	Practically non credible occurrence
$10^{-6} \le f_i < 10^{-4}$ 1/y	Rare occurrence
$10^{-4} \le f_i < 10^{-3}$ 1/y	Unlikely occurrence
$10^{-3} \le f_i < 10^{-1}$ 1/y	Credible occurrence
$10^{-1} \le f_i < 1$ 1/y	Probable occurrence
$f_i \ge 1$ 1/y	Likely/Frequent occurrence

Table 7. Consequences qualitative classification criterion, r_{vul} = vulnerability radius, S_w = ship width.

Consequence severity	Qualitative rating
$r_{vul} < 1$ m	Slight effect
$1 \text{ m} \le r_{vul} < S_w$	Effects internal to the source (ship)
$S_w \le r_{vul} < d_U$	Effects external to the source (ship) & no interaction with targets
$d_U \le r_{vul} < d_P$	Damages to other units & possible single fatality
$r_{vul} \ge d_P$	Multiple fatalities

- P—installations or activities characterized by the presence of people (ferries, campsites, buildings, etc.)
- U—other units characterized by the presence of dangerous goods (ground-based plants, cargo ships, etc.)

The minimum distances between the source of the accidental release (e.g., the LNG carrier) and the sensitive targets are defined as follow:

- d_P – minimum distance between the LNG carrier and the type "P" installations;
- d_U – minimum distance between the LNG carrier and the type "U" installations.

The comparison against r_{vul} and the reference distances (e.g., d_P and d_U) allows for the classification of consequences according to the criteria summarized in Table 7.

Risk associated with each accidental scenario is finally assessed as the combination of the likelihood and the severity in the reference risk matrix shown in Figure 4 (see Section 5).

The matrix is divided into three zones:

- Low risk level: continuous improvement and acceptable risk;

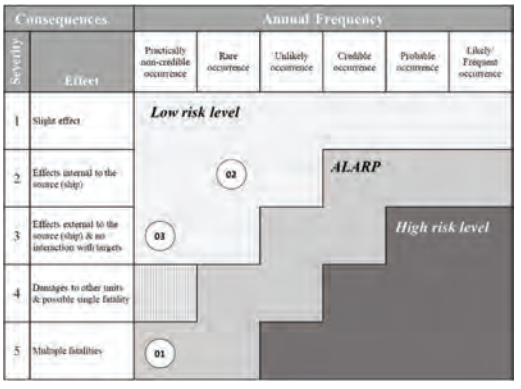


Figure 4. Example of risk matrix application, showing risk results for the scenarios listed in Table 8 (indicated by the numbered circles in the matrix).

- Medium risk area or ALARP (As Low AS Reasonably Practicable) zone;
- High or intolerable risk level: mitigation and prevention measures are mandatory to reduce the risk at acceptable levels.

Events with severity in class 4 (damages to other units & possible single fatality, see Table 7) and likelihood in class 1 (practically non-credible occurrence, see Table 6) may be associated with low or ALARP risk level (see Fig. 4) depending, respectively, on the absence or presence of dangerous goods in the impacted units. In fact, physical effects due to accidental release from the target unit will only be economic damages, if there are not hazardous substances in the installation; otherwise human may be affected by the release of hazardous substance from the target units, which is damaged by the primary accidental scenario associated with the LNG carrier.

The results obtained from the risk matrix are then summarized in the risk register. An example of risk register is shown in Section 5.2.

5 RESULTS

5.1 *Geometrical assessment and legislative framework*

The legislative framework (see Section 3.2.2) showed no particular limitations or restrictions for what concerns the LNG carrier access to the harbour, thus preliminarily supporting the feasibility of the LNG supply to the future terminal.

The results of the geometrical envelop analysis show that there are not suboptimal interactions for LNG carrier of small size (e.g., Ship A in Table 1). On the other hand, suboptimal geometrical interactions occur in about the 50% of the year-scale time, for large LNG carrier (e.g., Ship B in Table 1), in accordance with the legislative framework and the historical traffic data. Thus, it may be needed the selection of small size gas carriers (e.g., Ship A in Table 1) to avoid any poor or limiting interactions with the standard/historical harbour operations.

It may be concluded that ships collision, impact, grounding, and impact with berths are not credible scenarios. Explanation is that: piloting, tugs service, and speed reduction within the channels are mandatory (Ordinance 2009, 2010) for the cargo ships of interest; and that minimum distance between ships in the convoy and the transit of the latter in one direction (Ordinance 2009) ensure minimum or no interaction between ships. Moreover, impact of the LNG carrier with other moored ships has a practically non-credible likelihood in the range of 10^{-8} to 10^{-9} 1/y (Uijt de Haag & Ale 1999). Thus, the risk analysis focuses on the scenarios following a random process failure (see Section 4.2), excluding all other scenarios.

5.2 *Safety assessment results*

In this Section, the results of the risk-based analysis described in Section 4 are shown for a set of representative scenarios.

A total number of 45 scenarios is obtained from hazard identification, including fire and explosions events, following the release of either liquid or vapor natural gas. The risk register is compiled including all the mentioned scenarios, providing an ID to each identified event, the description of the event and the risk-based classification; this includes frequency and consequence class evaluation, and finally, the risk level. An example is shown in Table 8 for three most critical events associated with the LNG carrier.

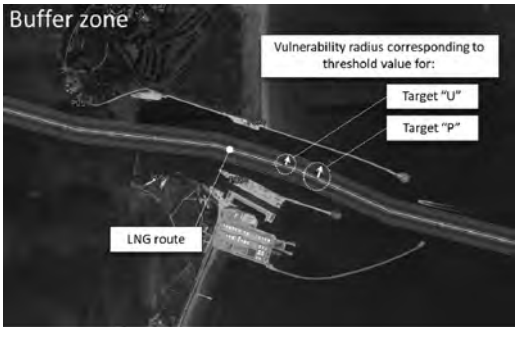


Figure 5. Example of buffer zone. For the definitions of threshold values and target type refer to Sections 4.5. The LNG route refers to path 2 in Figure 1. The points labelled with "P" are the minor civil installations in the area.

Table 8. Example of risk register showing risk analysis results. F = frequencies class; C = consequences class; R = risk level defined as low (L); medium (M); high (H) (see Section 4.5 for further details).

Source	ID	Scenario	F	C	R
Liquid piping system on deck	01	VCE	1	5	M
BOG piping/over-the-top vapour connections	02	Flare/Jet Fire	2	2	L
Steam piping manifold on deck	03	Flash Fire	1	3	L

Results are also reported in the risk matrix (see Fig. 4) in order to drive the strategy for risk reduction as discussed in Section 6.

The most critical scenarios evaluated in the present analysis are associated with major liquid releases, leading to pool spread and evaporation, with potential large fires and explosions. The consequences are represented through the use of buffer maps in order to trace the maximum extension of the scenarios. Figure 5 shows an example of buffer map.

The scenario showed in Figure 5 is a flare/jet fire following the accidental release of LNG from the liquid pipeline, at maximum operational pressure, from a minor hole of 10 mm, and F/2 meteorological conditions. The damage distance r_{vul} from the LNG route extends for 75 m and 62 m for targets type "P" and "U" (see Section 4.5), respectively. The LNG ship route is located in the middle of the channel, with an uncertainty of 15 m from the channel centreline (path 2 in Fig. 1). Thus, the same ±15 m uncertainty is applied to extend the consequences zone. The vulnerability radii are then moved along the LNG route obtaining the buffer zones, which help visualizing the possible targets with respect to the threshold values defined in Table 7.

6 DISCUSSION

The outcomes of the analysis demonstrate that the LNG carrier access induce a relevant risk level for the industrial and civil installations close to the channel, thus the analysis may constitute a preliminary driver to enhance safety measures and procedures in the development of the LNG terminal with a dual purpose: i) reducing possible suboptimal interactions between the LNG carrier and the current port configuration; ii) reducing the risk level by lowering likelihood and/or impact of the most critical events. Some prevention and mitigation actions are listed in the following in order to provide an example of utilization of the risk results obtained with the present methodology.

Prevention measures are aimed at reducing the credibility of the accidental scenarios. In the present case, the sequence of operations on LNG carrier before entering the harbour area is crucial to prevent the occurrence of critical scenarios. In fact, liquid is present on board because of the cooling down activities, which normally precede the loading/unloading operations. These activities are usually carried out in the roadstead. Carrying out the cooling down activities at berth removes (thus prevents) the accidental scenarios linked to the liquid pipelines system along the ship route. Implementing this prevention action, however, shifts the hazards from the ship route to the berth. Thus, a specific analysis should be performed to evaluate the risk of carrying out the cooling down operations at berth.

Mitigation measures reduce the impact of an accidental scenario by lowering the consequences. An example of mitigation action applicable to the case study is the utilization of fire-fighting tugs to reduce the effects of fire and explosion following LNG releases from the carrier. However, careful selection of fire-fighting tugs should be performed, accounting for the water-mist demand to effective fire-fighting, water suction pumps capacity, and depth of sea in the working area. Thus, it requires further studies for the most critical scenarios.

7 CONCLUSIONS

In the present work, a feasibility study was carried out to evaluate the geometrical, legislative and safety aspects associated with the access of LNG carriers in the port of Venice. The carriers supply LNG to a future bunkering terminal which is under development.

A specific risk-based analysis supported the identification and evaluation of potential accidents associated with the transit of LNG carriers in the harbour area through a risk matrix. The most critical scenarios were identified, providing indications for risk control, in terms prevention and mitigation actions.

The present method may support the planning of industrial harbour areas development in the perspective of a wider implementation of LNG bunkering and distribution terminals.

ACKNOWLEDGEMENTS

The authors wish to thank the financial and technical support of Venice Port Authority, which allowed developing and testing the present the methodology.

REFERENCES

ABS. 2014. *Bunkering of Liquefied Natural Gas-fuelled Marine Vessels in North America.* Houston, TX: ABS.

ADN Administrative Committee. 2014. *Proposed text of a derogation regarding the use of LNG for propulsion for a push boat to be built by Kooiman Marine.* Geneva: ADN Administrative Committee.

Bittante, A., Jokinen, R., Krooks, J., Pettersson, F. & Saxén, H. 2017. Optimal Design of a Small-Scale LNG Supply Chain Combining Sea and Land Transports. *Industrial & Engineering Chemistry Research* 56(45): 13434–13443.

Chemical Controls 2017. *Relazione Tecnica—Analisi e studi connessi all—accessibilità nautica delle navi che trasportano LNG nonché alla definizione dei rischi connessi alla navigazione e alle attività di bunkeraggio nel canale Sud del Porto di Venezia (in Italian).* Livorno: Chemical Controls srl.

DM 2001. *Decreto Ministeriale 09/05/2001. Requisiti minimi di sicurezza in materia di pianificazione urbanistica e territoriale per le zone interessate da stabilimenti a rischio di incidente rilevante.* Rome: Italian Ministry of Public Works.

DNV. 2012. *Port toolkit risk profile LNG bunkering—Port of Rotterdam, Ministry of Infrastructure & Environment. Port of Antwerp, Port of Amsterdam and Zeeland Seaport.* Oslo, Norway: DNV Det Norske Veritas.

IEC 2010. *IEC 61508:2010. Functional safety of electrical/electronic/programmable electronic safety-related systems. International Standards and Conformity Assessment for all electrical, electronic and related technologies.* International Electrotechnical Commission.

IGU 2016. *IGU World Gas LNG Report – 2016 Edition.* LNG 18 Conference & Exhibition Edition.

IMO 2016. *The International Code of the Construction and Equipment of Ships Carrying Liquefied Gases in Bulk (IGC Code).* London: IMO publication.

Jeong, B., Lee, B.S., Zhou, P. & Ha, S.m. 2017. Evaluation of safety exclusion zone for LNG bunkering station on LNG-fuelled ships. *Journal of Marine Engineering & Technology* 16(3): 121–144.

Lee, S., Seo, S., Chang, D. 2015. Fire risk comparison of fuel gas supply systems for LNG-fuelled ships. *Natural Gas Science and Engineering* 27: 1788–1795.

Lentner, R., Richter, M., Kleinrahm, R., Span, R. 2017. Density measurements of liquefied natural gas (LNG) over the temperature range from (105 to 135) K at pressures up to 8.9 MPa. *The Journal of Chemical Thermodynamics* 112: 68–76.

Lloyd's Register Marine 2015. *Small-scale LNG ships report*. Lloyd's Register Marine.

Mannan, S., 2005. *Lees' Loss Prevention in the Process Industries, 3rd edition.* Oxford: Elsevier.

Ordinance 2009. *Ordinance 175/09.* Venice Harbour Masters of Italian Cost Guard.

Ordinance 2010. *Ordinance 155/10*. Venice Harbour Masters of Italian Cost Guard.

Ordinance 2016. *Ordinance 36/16*. Venice Harbour Masters of Italian Cost Guard.

Petrone, A., Scataglini, L., and Cherubin, P. 2011. *B.A.R.T (Baseline Risk Assessment Tool): A Step Change in Traditional Risk Assessment Techniques for Process Safety and Asset Integrity Management.* Denver: Presented at the SPE Annual Technical Conference and Exhibition 30 October–2 November.

Spouge, J. 2005. New generic leak frequencies for process equipment. *Process Safety Progress* 24(4): 249–257.

Tugnoli, A., Landucci, G., Salzano, E. & Cozzani, V. 2012. Supporting the selection of process and plant design options by Inherent Safety KPIs. *Journal of Loss Prevention in the Process Industries* 25: 830–842.

Uijt de Haag, P.A.M. & Ale, B.J.M. 1999. *Guidelines for quantitative risk assessment (Purple Book)*. The Hague (NL): Committee for the Prevention of Disasters.

WSD50. *WSD50 30K, 30,000m³ LNG Carrier Datasheet.* Wartsila.

Yun, G., Rogers, W.J., Mannan, M.S. 2009. Risk assessment of LNG importation terminals using the Bayesian-LOPA methodology. *Journal of Loss Prevention in the Process Industries* 22:91–96.

Zonta, R., Botter, M., Cassin, D., & Pini, R., Scattolin, M., & Zaggia, L. 2007. Sediment chemical contamination of a shallow water area close to the industrial zone of Porto Marghera (Venice Lagoon, Italy). *Marine Pollution Bulletin* 55: 529–542.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A framework for aggregating risk information across organisational levels—the case of Swedish municipalities

H. Hassel
Division of Risk Management and Societal Safety, Lund University Centre for Risk Assessment and Management (LURCAM), Center for Critical Infrastructure Protection Research (CenCIP), Lund University, Lund, Sweden

ABSTRACT: Performing risk assessments for hierarchical, multi-functional systems, such as a municipality, is an activity that requires input from a multitude of actors. In such systems risk assessments can be performed at many system levels and support different types of decisions. For issues that are constrained to a specific sub-system, such as a municipal department, decisions can be preferably taken at sub-system level. However, for other issues, such as those crossing many sub-systems and system levels decisions should preferably be taken at higher system levels, e.g. at the municipal level. At the same time, these decisions require extensive information from the sub-systems. The aim of the present paper is therefore to outline a framework for how risk information can be aggregated—with application in the context of Swedish municipalities. The research builds on previous work by the authors where a method for performing risk and vulnerability assessments in municipal departments has been developed using an action research approach. The method will soon be implemented in each municipal department in the municipality of Malmö, Sweden, and the next step is to develop the aggregation of these assessments. It is argued that this aggregation is facilitated by ensuring that key aspects of the risk assessments in the municipal departments are harmonized. At the same time, too much standardisation may also reduce the utility of the assessments for the municipal departments.

1 INTRODUCTION

Performing risk assessments for hierarchical, multi-functional systems, such as a municipality, is an activity that requires input from a multitude of actors. In such systems, risk assessments can be performed at many system levels and support different types of decisions. For issues that are constrained to a specific sub-system, decisions may preferably be taken at sub-system level. However, for other issues, such as those crossing many sub-systems and system levels, decisions should preferably be taken at higher system levels. At the same time, these decisions require extensive information from the sub-system level which has to be aggregated and synthesized to become meaningful and possible to use as a basis for risk reductions at system level.

However, limited research has been conducted in research field of risk assessment and management on this topic. The issue is addressed by Ayyub et al. (2008) who argue that to facilitate aggregation of risk to higher levels of abstractions all levels "should share a common analytical framework" (Ayyub et al., 2008, p. 791). Furthermore, the UK Cabinet Office has a similar argument when claiming that a benefit of having a standardized risk assessment approach is that it "facilitates regional aggregation of local risk assessments" (UK Cabinet Office, 2005, p. 41). Klaver et al. (2008) provide similar arguments but also stress the need for using consistent scales for impact, probability and risk evaluation as well as using a common list of threat classes. Furthermore, David (2009) argues that one must be cautious when attempting to aggregate "risks" that are not independent, since a simple "summation" would not capture dependencies that may exist between them.

Although not specifically in the context of RVA, Kramer argues that information sharing, which is a precondition for risk aggregation, can be "impeded by differences in how information is coded and categorized" (Kramer, 2005), thus stressing the need for common ways of describing risk-related information. On the other hand, Vaughan (1997) argues that there are many obstacles in knowledge and information sharing founded in the very nature of complex organizations that have multiple, specialized units—creating so called "structural secrecy". She warns that too much information sharing in too standardized ways may actually result in units simply getting less knowledge about other units in the organization, e.g. due to information overload.

Other studies presented in Månsson et al. (2015) and Månsson et al. (2017) have shown that inconsistencies in how risk information is expressed in the Swedish crisis management system reduce the possibilities of aggregating risk information. In addition, an experimental study (Månsson et al. 2017), showed that semi-quantitative or quantitative ways of expressing risk information may facilitate risk aggregation compared to qualitative ways. Although, these two studies provide some clues on how to accomplish successful aggregation of risk information, they mainly focused on aggregation of likelihood and consequence information; however, in a risk and vulnerability assessment there is potentially additional information that should be aggregated. An example is information about interdependencies between critical societal functions.

In none of the referred studies above, however, a comprehensive analysis has been made regarding what the challenges actually are for aggregations/syntheses and what processes and methods are needed to overcome these challenges in order to accomplish an appropriate synthesis of RVAs. Hence, there is very little guidance for how to go about in establishing and implementing such a process in practice. This is troublesome since establishing a process able to generate a consistent, high-quality picture of risk and vulnerability at a higher level, which as much as possible utilizes the information from lower level RVAs, is far away from straightforward.

In the Swedish crisis management system, which is the context of the present paper, several public actors are obliged to perform Risk and Vulnerability Assessments (RVAs) (SFS 2006:544). In this system, aggregation of risk information is intended to play a key role. Information from municipal RVAs should be used as input at regional RVAs; and regional RVAs should be used as an input to the national RVA. However, even though this system has been in place for more than 10 years, there are still many needs for improvements before risk aggregation can be successfully accomplished across societal levels.

Rather than attempting to solve all these problems in the Swedish crisis management system, the present paper delimits its focus to the municipal and municipal department level. The aim of the RVAs carried out in the Swedish municipalities is to increase risk awareness and knowledge of decision makers and to implement effective risk reduction measures (SFS 2006:544; MSB 2015:5). The scope of the analysis is primarily the *municipal organization*, where the aim is to ensure that critical activities can continuously be performed also in times of crises, but secondarily also to create a risk picture for the municipality as *a geographic region*. The present paper mainly focuses on the organizational perspective.

In order to establish a successful risk and vulnerability assessment process, many municipalities in Sweden, especially larger ones, have decided to push extensive analytic activities down to the municipal department levels. The main reason for this is that much of the opportunities, mandate and budget to implement improvements occur at this level. In addition, the local ownership of the analysis process is important and the expertise necessary for good quality assessments exist at the department level. See Cedergren et al (forthcoming) for further discussions of challenges and success factors related to municipal risk and vulnerability analyses in Sweden.

At the same time, even though relevant risk information can be acquired at municipal department level, some critical risk information will probably not be apparent until risk information from departments is somehow put together, aggregated and synthesized at the municipal level. This aggregated risk information can then both be used at the municipal level to make decisions that concern the municipal organization as a whole as well as be fed back to the municipal department level in order to improve the next iteration of the department level assessments.

The aim of the present paper is to describe a general framework for how risk aggregation can be accomplished in order to support system-level decisions. This model is then applied in the context of the municipality of Malmö. The research builds on previous work by the authors where a method development process has been initiated in collaboration with the municipality of Malmö in southern Sweden, focusing on RVA in municipal departments. The method development process, grounded in action research and design science, has been described previously in Cedergren and Hassel (2017) and a first version of the method has been presented in Hassel and Cedergren (2017). The developed method is currently being implemented in each municipal department in the municipality of Malmö, Sweden, and this paper will provide a basis for the next step of the RVA process in Malmö.

The outline of the paper is as follows: in Chapter 2, a generalised model for aggregation of RVAs is described, in Chapter 3 the model is applied in the context of the municipality of Malmö to outline initial ideas regarding aggregation of risk information, and in Chapter 4 the results are discussed and conclusions drawn.

2 A GENERALISED MODEL FOR AGGREGATION

The general model for aggregation of RVAs is presented in Figure 1. It was developed in an iterative process where the point of departure has been the

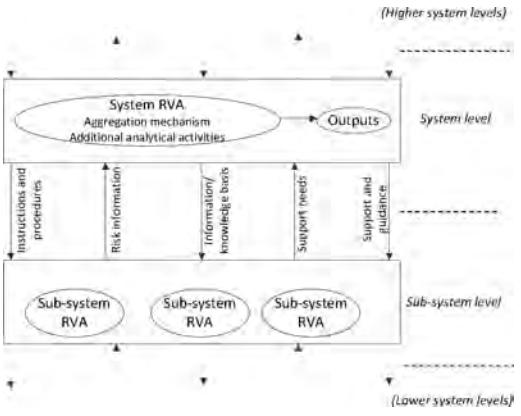


Figure 1. An overview of the model for aggregation of risk information.

research referred to above and the author's experience from the Swedish RVA-system. The author has interacted with practitioners in meetings and workshops who have expressed their needs and challenges related to the aggregation process.

The model should be seen as recursive, meaning that the displayed relationships between system and sub-system levels could be repeated between the system level (e.g. municipal) and higher system levels (regional, national or international) as well as between sub-system level (e.g. municipal departments) and even lower system levels (e.g. specific units). The model will be briefly described in what follows.

2.1 *Instructions and procedures*

First, *instructions and procedures* for the sub-system RVAs need to be formulated. The purpose is to ensure and facilitate the subsequent aggregation of risk information from the sub-system RVAs at the system level. It is especially the form of the risk information that needs to be harmonized and made consistent since aggregation would otherwise be extremely difficult and time-consuming. In addition, the method and analysis processes as a whole could also be harmonized although this is not strictly necessary. At the same time, since information across different RVAs must be using commensurate scales these must be either absolute/quantitative or be using common scale descriptions, definitions, categorizations, etc.

2.2 *Risk information to the system level*

Secondly, *risk information* must be submitted from the sub-system, i.e. RVAs must be conducted at the sub-system level and presented in a way that is in accordance with the instructions and procedures (see above). Of special importance here is not to compromise with information security. Hence, creating trust between the involved actors and organizations is a crucial precondition for successful aggregation—and a potential large issue especially when private actors are involved and sensitive information are being processed. Of course, in order to handle the information in way that do not compromise information security some type of information infrastructure need to be in place where information can be shared in a secure and efficient way.

2.3 *Aggregation mechanism*

Thirdly, there has to be an *aggregation mechanism* at the system level that can collect, store and process the risk information from the sub-systems in order to produce the desired *outputs* at the system level. In addition, in most cases it is likely that *additional analytical activities* need to be performed at the system level as aggregation is not likely to be a mechanical, additive process due to interconnectedness between the sub-systems (David, 2009). In special circumstances, where sub-systems do not have any significant relationships/interactions, the aggregation process can be mechanistic—e.g. aggregating/adding risk metrics from the sub-systems into risk metrics for the system, or comparing importance of sub-system elements across sub-systems.

These three aspects are the necessary components of an aggregation process; however, they should be complemented by three additional aspects in order to ensure high quality RVA processes at both the system and sub-system level. Hence, the following three aspects described below can be added to the model.

2.4 *Information and knowledge bases to the sub-system level*

Fourthly, *information and knowledge bases* should be provided from the system level to the sub-system levels. This can concern information that all sub-system RVAs are in need of but which would require extensive data collection efforts. This activity can then be more efficiently performed by the RVA coordinating actor at the system level which also would ensure a consistent knowledge basis concerning aspects that ideally should not vary across the sub-system RVAs (such as the likelihood of external events, such as floods, hurricanes, etc.).

2.5 *Support needs and guidance*

Fifthly, in order to develop the quality of the RVA processes, both at sub-system and system level *support needs* should be communicated from the sub-system; and sixthly *support and guidance* should

then be provided by the RVA coordinating actor at system level. Again, the reason is primarily efficiency-related—if several sub-system RVAs struggle with the same problems then a coordinated effort to reduce these problems is likely to be preferable compared to each sub-system working with quality improvements *only* on their own. Note that this has both to do with improving the quality of the assessment on the sub-system level and increasing the possibilities of a successful aggregation on the system-level.

3 APPLICATION OF THE GENERAL MODEL OF AGGREGATION FOR MUNICIPAL RVA IN SWEDEN

Below the general model for accomplishing system-level aggregation will be applied in the context of municipal RVA in Malmö, Southern Sweden. The authors have been extensively involved in the RVA development process there for more than 1.5 years. Currently RVAs are conducted in the municipal departments and the next step is to aggregate the output from these assessments. The description below is the results from the first round of discussions about how the aggregation could be accomplished. It is likely that the aggregation process will be further developed as more experience from the RVA processes have been gained.

3.1 *Instructions and procedures for the sub-system RVAs*

It is not always that organizations at higher system level has legal mandate to provide instructions that the lower system levels have to oblige to. In Sweden, the Swedish Civil Contingencies Agency has developed regulations for municipal RVAs (MSB 2015:5); however these are only stipulating *what* should be included in the assessment not *how* the assessment should be performed. In addition, a number of reports have been developed by MSB providing some general guidance (MSB 2011, 2013, 2014).

In the present paper the focus is on municipal RVA where the municipality of Malmö decided to develop a method that all municipal departments should use. The authors of this paper participated in the development of this method and the points of departure for the method development have been presented extensively in previous papers by the authors (Hassel and Cedergren, 2017; Cedergren and Hassel, 2017; Cedergren et al. forthcoming).

Note that the aim of the method is to assist municipal departments to create robustness in their organisation by identifying the main sources of risk and vulnerability which should be targeted by risk reduction measures. It is only a secondary aim that the output of the RVAs should be aggregated at municipal level (which is the focus of the present paper).

To support the municipal departments in conducting the assessments a method handbook has been developed, educational activities have been and will be arranged, and computer software developed where the analysis can be documented. In addition, the municipality of Malmö are currently developing short movies that can be used by the municipal departments.

The method consists of three main steps which are conducted every year, i.e. the RVA is successively made both broader and deeper over time. The method has been described in detail (although a previous version) in Hassel and Cedergren (2017) but will be summarized below.

3.1.1 *Step 1 – Mapping the municipal department*

First, a mapping of the functions performed by the municipal department is conducted. This mapping includes determining how critical each function is for safety and functionality of the municipality of Malmö as well as mapping of dependencies that each function has. When mapping dependencies the strength of the dependencies is judged based on to what extent the function needs the dependency in order to be maintained as well as to what extent there are alternative back-up solutions in place.

3.1.2 *Step 2 – Analysing undesired events*

Secondly, an identification of what undesired events can happen is performed. Then a selection of events are analysed. The focus is primarily how the department's functions are affected in the event and to what extent they can still perform their critical functions. The extent of the negative consequences for the municipality of Malmö is judged, based on their capability to continue to perform their functions during the event. Finally, potential causes are identified, indications/trends coupled to the event is identified and the likelihood of the event is judged.

3.1.3 *Step 3 – Create a decision basis for implementing improvements*

Thirdly, the information from the two first steps are visualised in order to identify functions, dependencies and events that are particularly critical. This visualisation is then used as a basis for suggestions risk reduction measures where risk reduction measures are first identified and then evaluated. The evaluation is made based on the measures' effects on the risk level, their costs and potential side-effects.

3.2 *Risk information from sub-system RVA to system level*

The sub-system RVAs produce several outputs that can be used as input to the system-level RVAs.

All this information is stored in a computer software that can be accessed by the RVA coordinating unit at the system level. A selection of the most relevant risk information that is uploaded includes the following:

- Time until reaching severe consequences when the function cannot be performed.
- Dependencies that each function have.
- Importance of each dependency for the functions.
- Extent of back-up solutions available for each dependency.
- List of undesired events.
- Description of each undesired events.
- Functions that are prioritized in each event.
- Capability to continue performing each prioritized function in each event.
- Description and estimation of negative consequences in each event due to reduced capability in the municipal departments functions.
- Causes, trends and indications of events.
- Estimation of likelihood of the event.
- Strength of knowledge judgements for each estimation.
- Suggested risk reduction measures.

3.3 *Aggregation mechanism and complementing analytical activities*

The purpose of the aggregation process is primarily that it should provide information and insights that can be used a basis for system-level improvements. A secondary purpose is that some information should be fed back to the municipal departments, and this will be further explored in section 3.4.

As mentioned previously, according to the regulations in Sweden the municipal RVA has two perspectives—the municipality as an organisation and the municipality as a geographical region. Note that this paper takes the former perspective and leave the latter for future research as it is a much more complex task requiring extensive input from non-municipal actors.

The aggregation in the municipality of Malmö will be performed by the central RVA coordinating unit. Below a number of outputs from this aggregation is described—this list of outputs is likely to be expanded later on in the process.

3.3.1 *Identification and ranking of joint dependencies*

A municipal department may identify that a particular dependency is very critical for being able to perform their functions. In some cases they might be able to justify taking actions to reduce the criticality of this dependency by e.g. investing in buffers, making the dependency more robust (if internal) or ensuring that external actors will be able to deliver the resources or services they are dependent on (e.g. by improving contracts). However, if many municipal departments are critically dependent on the exact same resources or services, then measures might be warranted *even though* each municipal department cannot justify it on an individual basis. Additionally, in some cases risk reduction taken at system/municipal level might become much more effective than each municipal department taking separate actions concerning the same dependency.

In order to create this output the dependencies need to be described in a standardised, consistent way. Previous experiences from dependency assessments where no standardisation was used proved aggregation was very challenging since the level of detail with which dependencies was described varied greatly (Johansson et al. 2016). Therefore, the method of RVA for the municipal departments include a list of dependencies from which the municipal departments can choose from must be established. Since it is difficult to foresee all possible dependencies at the outset of the assessments, this list must be dynamic and continuously updated.

Furthermore, the ranking of the joint dependencies should logically be a combination between how many municipal departments that have the dependency and how critically dependent they are. Hence, the rating of dependency criticality must be comparable across the municipal departments which is accomplished by used common scales.

3.3.2 *Identify dependency chains*

In the assessments at municipal department level only first-order dependencies are identified since higher-order dependencies both can be difficult to realise/have knowledge about and be time consuming to identify. However, by aggregating information about first-order dependencies, dependency chains can be constructed that provide insights about second—and higher-order dependencies.

Information about dependency chains may be important for the municipality as a whole since it could provide insights on where resources should be directed to break potential cascading effects at an early response stage (both considering preventive measures and measures as an event is unfolding).

Of course, only a partial view of the municipality can be obtained based on dependency information from municipal departments. In order to obtain a more holistic view, additional actors need to be consulted and included, such as private and other public actors.

3.3.3 *Identify undesired events*

Each municipal department performs an identification of potential undesired events. Since

event identification is also a requirement for the municipal RVA the list of events from the municipal departments can be used as an input, although some type of categorisation is likely to be needed.

3.3.4 *Inform estimations of event consequences for the municipality as a whole*

An essential part of a risk and vulnerability assessment at a municipal level is the estimation of the negative consequences of undesired events for the municipality as a whole. In the assessments performed by the municipal departments no consequence estimation is carried out considering the total consequences for the municipality. However, the dependency information, assessment of capabilities and consequence estimations collected from each municipal department can provide a necessary partial input to the estimation of the consequences of undesired events at the municipal level.

Note that it is unlikely that a mechanical procedure can be used for determining consequences at municipal level based on information from municipal departments. The reason is that the functions of each municipal department are typically tightly coupled together so there is no additivity in the consequences estimated by each municipal department. Notwithstanding, the information from the municipal departments is crucial in order to understand how the municipality as a whole will be affected. But additional information from other relevant actors must also be included and the consequence estimations should be made in a deliberative process where the information from the municipal departments is used as an input.

Furthermore, in order for the municipality to be able to make consequence estimations concerning some specific events it must be ensured that all municipal departments have included these events in their assessments. Therefore, the method for RVA used by the departments will both address events that each department has identified on their own as well as a number of events that the RVA coordinating unit has selected to be common for all departments.

3.3.5 *Create a prioritized list of critical functions, dependencies and events*

In order to get the most effect from the resources available for risk reduction, it makes sense to identify the most critical functions, dependencies and events for the municipality as a whole. By doing so the RVA coordinating unit could assist the municipal departments with a responsibility for the most critical functions, dependencies and events. Either this could be to provide financial support but it could also be to assist in getting political attention.

Again, in order to establish these system-wide criticality lists, the scales for judging criticality are designed so they are common for all departments.

3.3.6 *Improve estimations of the likelihood of events*

Estimating the likelihood of undesirable events is part of the municipal RVA. Since each municipal department make likelihood estimations, although they have a possibility to waiver the estimation if they judge their expertise to be insufficient, these estimations could be used as a basis for the likelihood estimations at the municipal level. Especially those departments that have rated their knowledge concerning the likelihood as being high could be consulted. In order to enable this the strength of knowledge, see e.g. Flage & Aven (2009), concerning the various estimations is included the RVA method used by the municipal departments.

3.4 *Provision of information and knowledge bases from system level to sub-system RVA*

In addition to providing risk information that can be used for improvements at the municipal level, the aggregation process can lead to insights that can be fed back to the RVAs in the municipal departments in order to improve the quality of these assessments. Below, a number of such possibilities will be described.

Feeding back information about *dependency chains* to the municipal departments can lead to insights regarding how reduced capability to perform their functions can give rise to downstream effects. In that way their estimations of how critical their functions are can be improved. Due to the same reasons, this can also improve the estimations of consequences of events. In addition, the municipal department's understanding of how they may become affected by undesired events may also be improved since they get more knowledge about their upstream higher-order dependencies.

Aggregating and categorizing the undesired events that have been identified in each municipal department can also give rise to insights to other municipal departments that can also occur there, i.e. creating an "event repository".

In a similar way the estimation of the likelihood of events can be improved. As mentioned previously the departments can opt not to provide a likelihood estimation—if they judge their expertise concerning this to be very limited. Then information from the aggregation process can be fed back to the municipal departments so that likelihood estimations from departments that judge their level of expertise as being high can be provided to the those departments that lack expertise.

Of course, it may be the case that basically no municipal departments have the required knowledge concerning some event likelihoods. In that case the RVA coordinating unit could conduct some additional analytic activities where appropriate expertise

is identified and involved to obtain a well-founded estimation that may be fed back to the municipal departments. This would be a much more efficient approach than forcing each department to do it on an individual basis.

In order for the municipal departments to be able to accurately assess how their functions will be affected in various events and the negative consequences this would entail, it is important that the departments have good knowledge about the potential *direct impacts* of those events (i.e. what areas of the city would be flooded? How many persons would be incapacitated in a major epidemic? etc.). Such knowledge is not necessarily something that each department would have; however, as in the case of likelihood estimations, the RVA coordinating unit could make an effort to find the appropriate expertise and provide the departments with appropriate information.

3.5 *Provide support and guidance to sub-system level*

In order to both improve the quality of the RVAs in the municipal departments and to increase the possibility of accomplishing aggregation, the RVA coordinating unit should provide continuous support and guidance to the municipal departments. In the case of Malmö this will be done through at least three annual workshops where each of the three steps of the method is addressed. In relation to each step, support needs should be indicated by the each department. In addition, support needs should be identified through assessments of the quality of the department's analyses performed by the coordinating unit. Key here is to identify whether significant mistakes or misinterpretations have been made in the analysis.

4 DISCUSSION & CONCLUSIONS

In the present paper a general model for performing aggregation of Risk and Vulnerability Analyses has been suggested. This model has then been applied to sketch out some initial ideas on how to accomplish aggregation in the context of municipal RVA, with a special focus on how analyses for municipal departments can be aggregated to the municipality of Malmö.

The model and the specific application in Malmö need to be concretized and evaluated. This will be done in the coming year as the Risk and Vulnerability Analyses performed in the municipal departments are being conducted which of course is a precondition for the aggregation process.

It is argued that this aggregation is facilitated by ensuring that key aspects of the risk assessments in the municipal departments are harmonized. At the same time, too much standardisation may also reduce the utility of the assessments for the municipal departments.

Another key point of the proposed aggregation model is that sharing of risk information between system levels must be two-way rather than only from bottom-up. As information is fed back to the subsystems it is critical that information is presented in a compact and easy-to-use way otherwise it is likely that it will not be used as time and resource constraints for municipal departments is a significant challenge (Cedergren et al., forthcoming).

A subsequent aim of the aggregation process is also to be able to include relevant risk information from other actors, such as private actors and other public actors. Therefore, a study regarding what information these other actors are willing to share should be initiated. It is of utmost importance that trust relationships with these actors are built so they feel they are able to share potentially sensitive information with the municipality.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the municipality of Malmö that has provided partial funding for the present project as well as contributed with critical insights regarding municipal RVAs. We are also grateful to partial funding from the Swedish Civil Contingencies Agency.

REFERENCES

Cedergren, A. & Hassel, H. 2017. An action research approach to developing, implementing and evaluating methods for risk and vulnerability assessment *Proceedings of the ESREL2017 conference*, 18–22 June, Portoroz, Slovenia.

Cedergren, A., Hedtjärn Swaling, V., Hassel, H., Denward, C., Mossberg Sonnek, K., Albinsson, P.-A., Bengtsson, J. & Sparf, A., *Forthcoming*. Understanding practical challenges related to risk and vulnerability assessments—The case of Swedish municipalities. Accepted to Journal of Risk Research.

David, S.R. (2009). Safety Risk Aggregation: The Bigger Picture. *Journal of Safety and Reliability,* 29 (2): 34–52.

Flage, R. & Aven, T., 2009. Expressing and communicating uncertainty in relation to quantitative risk analysis. *Reliability and Risk Analysis: Theory & Applications* 2(13): 9–18.

Hassel, H. & Cedergren, A. 2017. A method for combined risk and continuity management in a municipal context. *Proceedings of the ESREL2017 conference*, 18–22 June, Portoroz, Slovenia.

Johansson, J., Hassel, H. & Svegrup, L. 2016. Capturing Societal Interdependencies from a Flow perspective—Part I: Method and Model. *Proceedings of the*

ESREL2016 Conference, 25–29 September, Glasgow, Scotland.

Klaver, M.H.A., Luiijf, H.A.M., Nieuwenhuijs, A.H., Cavenne, F., Ulisse, A. and Bridegeman, G. 2008. European risk assessment methodology for critical infrastructures. *First International Conference on Infrastructure Systems and Services: Building Networks for a Brighter Future (INFRA)*.

Kramer, R.M. 2005. A failure to communicate: 9/11 and the tragedy of the informational commons. *International Public Management Journal*, 8(3): 397–416.

Månsson, P., Abrahamsson, M. & Tehler, H. 2017. Aggregated risk: an experimental study on combining different ways of presenting risk information. *Journal of Risk Research*. https://doi.org/10.1080/13669877.2017.1391315.

Månsson, P., Abrahamsson, M., Hassel, H. & Tehler, H. 2015. On common terms with shared risks—Studying the communication of risk between local, regional and national authorities in Sweden. *International Journal of Disaster Risk Reduction* 13: 441–553.

MSB. 2011. *Vägledning för risk- och sårbarhetsanalyser*. Stockholm: Myndigheten för Samhällsskydd och Beredskap.

MSB. 2013. *Handlingsplan för skydd av samhällsviktig verksamhet.* Stockholm: Myndigheten för Samhällsskydd och Beredskap.

MSB. 2014. *Vägledning för samhällsviktig verksamhet: Att identifiera samhällsviktig verksamhet och kritiska beroenden samt bedöma acceptabel avbrottstid.* Stockholm: Myndigheten för Samhällsskydd och Beredskap.

MSBFS 2015:5. *Myndigheten för samhällsskydd och beredskaps föreskrifter om kommuners risk- och sårbarhetsanalyser.* Stockholm: Myndigheten för Samhällsskydd och Beredskap.

SFS 2006:544. Lag om kommuners och landstings åtgärder inför och vid extraordinära händelser i fredstid och höjd beredskap.

UK Cabinet Office. 2005. Emergency Preparedness: Guidance on Part 1 of the Civil Contingencies Act 2004, its associated Regulations and non-statutory arrangements.

Vaughan, D. 1997. *The Challenger launch decision: Risky technology, culture, and deviance at NASA.* University of Chicago Press.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Toward the integration of uncertainty and probabilities in spatial multi-criteria risk analysis: An application to tanker oil spills

M. Spada
Laboratory for Energy Systems Analysis, Paul Scherrer Institute (PSI), Villigen PSI, Switzerland

V. Ferretti
London School of Economics and Political Science, London, UK

ABSTRACT: Quantitative risk assessment supports decision-making processes in an increasing variety of contexts. Within the domain of environmental decision-making, the spatial distribution of impacts, vulnerabilities and consequences associated to different risk-mitigation alternatives calls for a different framework, i.e. spatial multi-criteria risk analysis. In this paper, we propose the combined use of a hierarchical Bayesian modelling approach and Geographic Information Systems to integrate uncertainty and model probabilities into an overall risk map. The research aims at testing the operability of the integration between Bayesian modelling and spatial analysis in the context of spatial risk processes for resource allocation. To this end, we applied the model on a case study dealing with tanker oil spill risk in the Mediterranean Sea. The innovative contribution of the study stems from both the context of application point of view and the use of Bayesian modelling to calculate not only probabilities but an overall spatial risk measure.

1 INTRODUCTION

A rapidly emerging field of research within both the discipline of quantitative risk analysis (e.g. Fischhoff, 2015) and the now broad and consolidated literature on spatial decision support systems (e.g. Malczewski and Rinner, 2015) is spatial risk analysis, as demonstrated by the recent special issue on the topic in the Risk Analysis Journal (http://www.sra.org/sites/default/files/pdf/SpecialissueproposalforRiskAnalysis_Jun20.pdf). This growing interest may be explained by the following factors: (i) impacts of both positive and adverse events have heterogeneous spatial distributions across the territory, (ii) land vulnerability is characterized by spatial heterogeneity, (iii) risk-mitigation alternatives also lead to different spatial consequences, and (iv) recent advances in both decision modelling and Geographic Information Systems (GIS) technologies make it possible to bring the two fields together towards the formalization of an analytical framework for spatial risk analysis (e.g. Ferretti and Montibeller, 2017).

Despite several applications conducted in different domains (e.g. natural hazards management, health issues, etc.), the criteria they employ are often risk factors with deterministic preference modelling replacing probabilistic information. Furthermore, evaluations of spatial risks have often neglected the multi-dimensional nature of spatial impacts, such as infrastructure damage, lost lives, lost crops, etc., which typically occur in such decision problems (e.g. Ferretti and Montibeller, 2017).

There are several reasons that can explain the above weaknesses. First, the lack of publicly accessible spatial data about vulnerability and probability of occurrence spatial distributions; second, the need to involve risk analysts to facilitate probability elicitation using appropriate protocols (e.g. structured expert judgment elicitation protocols, Dias et al. (2018)); third, the need to allocate enough time to the necessary preliminary phase of expert training in probabilities elicitation (e.g. Keeney and Winterfeldt, 1991).

However, preliminary ideas and suggestions on how to formalize a framework for spatial risk analysis and how to spatially elicit preference information have recently been proposed (e.g. Keller and Simon, 2017, Ferretti and Montibeller, 2017).

In this paper, we propose the combined use of a hierarchical Bayesian modelling approach (e.g. Kalinina et al., 2016, Spada et al., 2014) and GIS to update prior probability maps and obtain an overall risk map. The research aims at testing the feasibility and operability of the integration between hierarchical Bayesian modelling and spatial analysis in the context of spatial risk processes for resource allocation. To this end, we applied the model on a case study dealing with tanker oil spill risk in the Mediterranean Sea. The main reasons for the selection of this context of application can be summarized as follows: (i) international

relevance of the phenomenon, (ii) availability of historical data collected in the PSI's Energy-related Severe Accident Database (ENSAD), and (iii) the strong spatial dimension of the phenomenon (e.g. the geographical heterogeneity of both vulnerabilities and impacts in the areas affected by oil spills).

The innovative contribution of the present study stems from two aspects. First, the context of application, i.e. tanker oil spill risk distribution, is an innovative one for the use of a Bayesian modelling approach. Indeed, Bayesian modelling in the spatial domain has been mostly used for ecological studies (e.g. He et al., 2006, Tucker et al., 1997), health concerns (e.g. Saravana Kumar et al., 2017), archeological analysis (e.g. Ford et al., 2009), and natural hazards (e.g. Liu et al., 2017) to name the most recurrent ones. In the area of maritime transportation, few applications of the Bayesian approach do exist, but they are mainly non-spatial (e.g. Goerlandt and Montewka, 2015, Bouejla et al., 2014). The second reason why this study is innovative is linked to the Bayesian modelling approach. Indeed, while many applications use a Bayesian network approach (e.g. Landuyt et al., 2015), we test a hierarchical Bayesian model not only to update an initial estimate of the probability of occurrence of tanker oil spills (e.g. Burgherr et al., 2015), using information concerning the distribution of past accidents in the geographical region under analysis, but also to generate an overall risk map.

The reminder of the paper is organized as follows. Section 2 provides an overview of the proposed methodological approach. Section 3 explains how the method has been applied to model tanker oil spill risk in the geographical region under analysis and, finally, section 4 discusses conclusions and future developments arising from this study.

2 METHOD

This study is part of a larger research project funded by the Swiss National Science Foundation aiming at the development of a methodology able to integrate Bayesian modelling, Geographic Information Systems (GIS), multi criteria decision analysis and structured expert judgement elicitation protocols to comprehensively assess spatial risks associated to adverse events. This paper focuses on the integration between the Bayesian inference and GIS and represents a preliminary test of the operability of the proposed approach.

Bayesian inference is an alternative to the classical statistical inference. In the latter, also known as frequentist inference, only repeatable events have probabilities, while in Bayesian inference probability describes both epistemic and aleatory uncertainty (e.g. O'Hagan, 2003). Indeed, Bayesian analysis combines data representing the entire likelihood function with prior knowledge about the parameters, which may come from other data sets or the modeler's experience and physical intuition (e.g. Reis Jr and Stedinger, 2005). The *a priori* distribution describes what is known before observing any data, while the likelihood reflects the information about the parameters contained in the data. Parameters estimation is made through the posterior distribution, which is computed using Bayes' Theorem (e.g. O'Hagan, 2003):

$$p(y|\theta) \propto L(y; \theta)p(\theta) \quad (1)$$

where $p(y \mid \theta)$ is the posterior distribution for the parameter θ given the observed data y, $L(y; \theta)$ is the likelihood function, and $p(\theta)$ is the *a priori* distribution of parameter θ. The proportionality ($\propto$) refers to the direct sample of the parameter values from the posterior distribution, which is commonly assessed using a Markov Chain Monte Carlo (MCMC) method (e.g. Andrieu et al., 2003).

To combine the Bayesian inference and GIS, different issues could arise due to the spatial dimension of the problem: (i) lack of local data (i.e. spatial heterogeneity in the distribution and density of the data), and (ii) potential spatial correlation among locations, to name the two most relevant ones. To cope with these challenges, a spatial hierarchical Bayesian model is employed in this framework as will be explained in the following paragraphs (e.g. Cooley et al., 2007, DiMaggio, 2012, Eastwood et al., 2014, Juan et al., 2016).

A spatial Bayesian hierarchical method produces parameter (θ) estimates for each individual analysis unit (e.g. location) by borrowing information from all analysis units (e.g. Eckle and Burgherr, 2013). This procedure is known as Bayesian "borrow of strength" effect (e.g. Zhu et al., 2006). In this way, the approach compensates lack of data for individual analysis units (e.g. Kalinina et al., 2016).

Based on this premise, equation (1) could be rewritten as follows:

$$p(y|\theta) \propto L(y; \theta)p(\theta|\varphi)p(\varphi) \quad (2)$$

An extra level is added to the standard Bayesian theorem for the hierarchical Bayesian approach. In this level, the parameter (θ) can be described with a distribution, which is conditional on the hyperparameter (φ). The distribution of this hyperparameter ($p(\varphi)$) is a hyperprior distribution. Therefore, when estimating the posterior for the parameter (θ), information from the hyperprior is used in addition to the information from the prior and likelihood.

Finally, to overcome the spatial heterogeneity in the distribution and density of the data and the potential spatial correlation among locations, in the spatial Bayesian hierarchical model they are

included as prior knowledge about the parameter (θ). This is possible, since they are commonly seen as "unexplained variance" in a model (e.g. DiMaggio, 2012, Eastwood et al., 2014). Based on this premise, by considering a set of *i* locations, the parameter for each location of interest (θ_i) is transformed into a log scale (making relationships additive rather than multiplicative) and is set equal to an intercept term (a_i) and two random effects, one non-spatial, i.e. the spatially unstructured latent covariates, (ρ_i) and the other spatial (λ_i), i.e. spatially correlated latent covariates in the model (e.g. Rodrigues and Assunção, 2012):

$$\log \theta_i = a_i + \rho_i + \lambda_i \quad (3)$$

The spatially structured component is described as a conditional autoregressive (CAR) Gaussian process ($\lambda \sim CAR.normal(W, \tau_\lambda)$) where the conditional distribution of each λ_i, given all the other λ_i 's, is normal with μ = the average λ of its neighbors and a precision (τ_λ) proportional to the number of neighbors. W represents the matrix of neighbors that defines the neighborhood structure. The non-spatial component of the model (ρ_i) is defined as normally distributed with $\mu = 0$ and precision (τ_ρ). The model is completed by assigning additional (hyperprior) distributions to the precision terms τ_λ and τ_ρ (e.g. Clements et al., 2006, Cooley et al., 2007). More details about each step are going to be discussed in the case study section.

3 CASE STUDY

The proposed methodology has been applied to tanker oil spills due to the strong spatial nature of the phenomenon (e.g. Vieites et al., 2004, Burgherr, 2007).

Crude oil and its refined products are a driving factor of many economic activities. Furthermore, the oil demand is expected to increase worldwide in the coming decades and it will continue to have the highest share in the global primary energy mix (e.g. International Energy Agency (IEA), 2015). However, oil spills are one of the major causes of ocean pollution, producing ecological disasters of wide public concern. Furthermore, linked to the damage caused to the environment are the high costs to fisheries, related industries, and tourism in the affected areas. This is especially true along the major oil tanker transport routes (e.g. Burgherr, 2007, Psarros et al., 2011).

In this study, only accidental spills from tankers were taken into account, whereas spills from acts of war and operational spillages allowed by international or national regulations, such as the International Convention for the Prevention of Pollution (MARPOL) were excluded (e.g. Burgherr, 2007).

The geographical region under analysis is the Mediterranean Sea including the Bosporus strait. This area has been selected since (i) it is still the shortest route from Asia to Europe; (ii) about 16% of the global maritime traffic and 33% of the global seaborne oil (almost 8 million oil barrels per day) is carried through the Mediterranean Sea, which represents only 0.8% of the ocean surface; (iii) is one of the most affected areas with a major number of spill events (REMPEC, 2011).

3.1 *Data*

The data used for the case study analysis come from ENSAD, which is the most authoritative database for energy-related accidents worldwide (e.g. Burgherr et al., 2017). ENSAD comprehensively collects information about accidents in the energy sector and assigns them to energy chains and activities within those chains. Accident data go back to 1970 and cover fossil, nuclear, hydropower and new renewables technologies. In contrast to databases that rely on a single or few information sources, the multitude of sources considered by ENSAD is thoroughly verified, harmonized, and merged to ensure (1) a worldwide coverage, (2) consistently high data quality across regions and over time, and (3) a high degree of completeness. ENSAD focuses on severe accidents, which are distinguished from small accidents based on seven criteria, i.e. ≥5 fatalities, ≥10 injured persons, or ≥10,000 metric tons (t) of hydrocarbons released, etc. (e.g. Burgherr et al., 2017).

With reference to oil spills in the Mediterranean Sea including the Bosporus strait, the ENSAD database registered 106 events, including small accidents (e.g. <10'000 t spilled oil), in the period 1970–2012. In this study, we updated the information about oil spills in the Mediterranean Sea until the end of 2016 by using the following sources:

- Analysis, Research and Information on Accidents (ARIA) database
- Failure and Accidents Technical information System (FACTS)
- Hazards Intelligence (HINT)
- Centre of Documentation, Research and Experimentation on Accidental Water Pollution (Cedre)
- European Maritime Safety Agency (EMSA)
- Regional Marine Pollution Emergency Response Centre for the Mediterranean Sea (REMPEC)
- Center for Tankship Excellence (CTX)
- igma by Swiss Re
- Other sources, such as newspaper, national and local publications, technical reports, etc.

In addition to severe accidents, also accidents with smaller consequences (<10,000 t and ≥0.1 t) have been included. All the accidents collected from the aforementioned sources were homogenized

prior to analysis, in order to avoid possible double counting. The final dataset comprises a total of 271 fully geo-referenced tanker oil spill accidents in the Mediterranean Sea including the Bosporus straight for the period 1970–2016.

3.2 *Overview of tanker oil spills in the Mediterranean Sea*

Figure 1 displays maritime accidental tanker spills by year and severity for the period 1970–2016. Most of the accidents (161) result in a release <7 t, followed by 59 and 57 accidents with a release between 7–700 t and >700 t, respectively. According to the severe accident definition in ENSAD (section 3.1), only 16 accidents out of 271 could be considered as severe with 3 out of them even extreme since they account for a release bigger than 100,000 t. Although the annual maxima of historically observed spills exhibit some variability, the data suggest a decreasing trend until 2000 followed by a generally constant spill rate in the period 2000–2016, which is also confirmed by decadal averages and in accordance with previous studies (e.g. Psarros et al., 2011, Burgherr, 2007). Furthermore, the relatively constant numbers of annual spills in the period 2000–2016 could be explained by the introduction of the regulation 13 F of Annex 1 of MARPOL (The Marine Environment Protection Committee (MEPC), 2003), which effectively mandated double hulls for new built oil tankers of 5000 dead weight tonnage and above starting from year 2000. However, the general trend in Figure 1 should not be taken as evidence that a catastrophic spill accident (e.g. 100.000 t or greater) can be excluded in the future, but the frequency of such an event may be expected at a lower frequency (e.g. Burgherr et al., 2012).

Figure 2 depicts the spatial distribution of the oil spills in the Mediterranean Sea including the Bosporus strait for the period 1970–2016. Historical observations show different "hot spots", such as the Gibraltar and Bosporus straights, the Peloponnese, the North Tyrrhenian Sea, the Malta and Lebanon areas. This is not always correlated with the average number of tankers per year (in 2012, http://medgismar.rempec.org/) navigating in a location shown as a 0.5° × 0.5° grid. Indeed, the Gibraltar and Bosporus strait as well as the Malta and Peloponnese areas show rather large average numbers of tankers per year, while the corresponding numbers are relatively low for the Northern Tyrrhenian Sea and Lebanon. The "hot spots" in the latter cases could be explained by the fact that most of the accidents happened in ports.

The final updated dataset is subdivided into two periods: 1970–2011 and 2012–2016. This has been done in order to use the former as input for the model presented in section 3.3 and the latter as a validation of the model.

3.3 *Model setup*

Risk is the potential for realization of unwanted, negative consequences of an event. In other words, risk can be defined as the product of probability/frequency and magnitude/severity of consequences (e.g. Committee on Foundations of Risk Analysis, 2015).

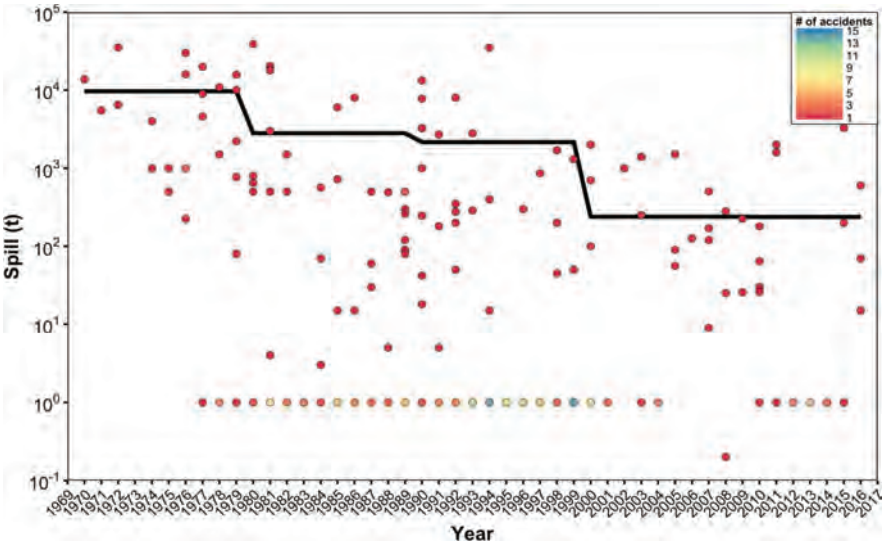


Figure 1. Distribution of tanker oil spills by year and severity in the Mediterranean Sea including the Bosporus strait. The bold black line indicates decadal averages. The colors indicates the number of accidents with the same spill size (t) in the same year.

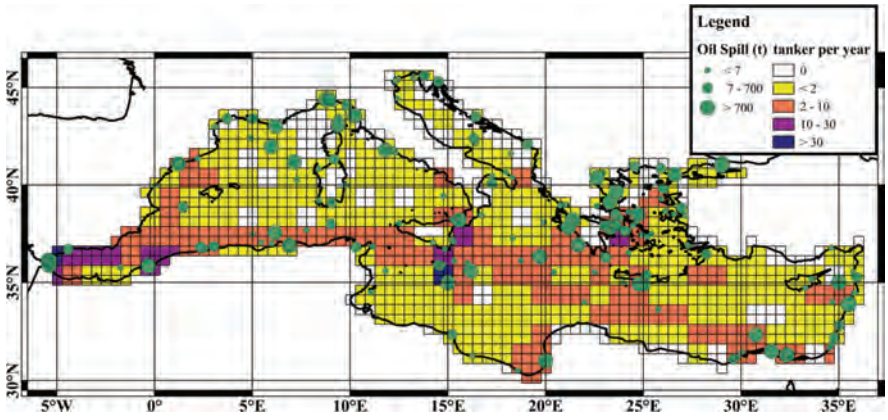


Figure 2. Spatial distribution of tanker oil spills by severity in the Mediterranean Sea including the Bosporus strait in the period 1970–2016. The average number of tankers per year (2012) navigating in the area is shown in a $0.5° \times 0.5°$ grid size (modified from AIS data from REMPEC (http://medgismar.rempec.org/)). White cells indicate areas where no tanker per year data are available.

In this study, the number of accidents per year gives the frequency, while severity measures the extent of the consequences of each accident. Furthermore, the frequency of accidents was normalized by the unit of tanker-year to allow for comparative evaluation of the modeled parameters among different areas in terms of accidents per tanker (e.g. Burgherr, 2007). It is also important to note that at a first approximation, the frequency of spills is not considered as a function of the number of tankers.

Essentially, accidents can be considered rare, independent events so that the frequency can be modeled as a Poisson distribution (e.g. Spada et al., 2014). Therefore, the frequency is modeled applying the Bayesian procedure described in section 2. In equation (2), the likelihood is described by the Poisson model. In a standard Poisson model, the variance is required to be equal to the mean. However, when dealing with spatial statistics, the Poisson models have more variances, and these are called over-dispersed Poisson models (e.g. DiMaggio, 2012). These variances are identified as either spatially-correlated effects or heterogeneity effects (section 2). The essential idea is that the probability of values estimated at any given location (e.g. the frequency at one area) are conditional on the level of neighboring values (frequencies on the neighboring areas).

Based on these premises, the parameter of interest, the frequency rate λ, is modeled as the combination of a normal distribution and the spatial random effects (section 2). For each area, the normal distribution is described by a mean and a standard deviation. This is the step where information between areas is exchanged in the hierarchical model. Both parameters of the hyper distribution are modelled with non-informative priors, so that they have no influence before the data is introduced. The parameters are modelled with normal and gamma distributions (e.g. Eckle and Burgherr, 2013), respectively.

The spatial random effects are modeled using a combination of the intrinsic conditional autoregressive (CAR) prior, which is smoothed based on the weight related to the number of the neighboring areas and their tanker-year normalization values, and a standard normal prior as described in section 2.

Once the posterior distribution for the mean frequency in each area is estimated, it is normalized by the corresponding tanker-year resulting in the expected accidents per tanker.

The severity is modeled as absolute numbers of oil spilled (tons) per accident. The expected oil spill in each area is modeled using a Lognormal (LOGNO) distribution (e.g. Burgherr et al., 2015). In this case, at a first approximation, the spill is considered independent from the area; therefore, no spatial random effects are included. The LOGNO distribution is commonly described by two parameters, namely the scale σ_i and location μ_i, which are used to define the mean value and variance of the distribution (e.g. Spada et al., 2014). In this study, the LOGNO distribution employed for the Bayesian Hierarchical modeling is described by mean and precision parameters, where the latter is the inverse of the variance. Hierarchical models are defined for the mean and standard deviation of the posterior and are modeled using a normal and gamma distributions, respectively. All the hyperparameters of both the mean and standard deviation are thus modeled with non-informative distributions.

The MCMC algorithm is run for 30,000 iterations, following a burn-in of 3,000 updates, which is also used to train the model, for both frequency

and severity cases. According to the Gelman-Rubin diagnostic (e.g. Gelman and Rubin, 1992), the simulated chains converged adequately in the MCMC practice implemented in this study for both cases. Furthermore, the models for both frequency and severity distributions have been validated using the Deviance Information Criterion (DIC), which tests how good the proposed model predicts a replicate dataset which has the same structure as the observed one, e.g. the historical observations (e.g. Gelman, 2003).

Finally, the risk in each area is estimated as the product of the expected accidents per tanker and the expected amount of oil spilled, giving the expected spilled oil in tons/tanker for each of the area in Figure 3.

3.4 *Results and discussion*

Figure 3 shows the risk of tanker oil spills in terms of expected tons of spilled oil per tanker. Areas of relatively high risk (1E-8–1E-9 tons/tanker) could be identified in the Tyrrhenian, Adriatic and Ionic Sea, in the north and west side of Sardinia, in the area of Mallorca, in the northern part of Libya and in north and east sides of Cyprus. In most of the cases, these results are driven by the comparatively large historical spills in areas with relatively low tanker traffic expressed in tanker-year (Figure 2).

Areas with a high spill risk (>1E-8 tons/tanker) could be identified in the Bosporus strait and different port areas along the Mediterranean coasts, such as in the Northern Tyrrhenian Sea, in Southern France, in the northern coast of Algeria, in the eastern side of the Peloponnese area and in Crete. This reflects the risk during port operations like loading and unloading of the tanker in ports and potential accidents due to grounding or collisions.

Relatively low risk areas (<1E-10 tons/tanker) could be found in the southern part of Spain, in the northern part of Algeria, in the Sicilian Sea, around Malta and in the Peloponnese region excluding its eastern side. These results are driven by generally low (or even absent) accidents recorded in the area combined with relatively high tanker-year (Figure 2). This result could be explained by the fact that these are open sea areas, and thus areas with limited possibilities of, for example groundings and collisions. Furthermore, in the relatively low risk areas, the result should not be taken as evidence that a spill accident can be excluded in the future, but that a spill event may be less expected than in areas of higher risk (e.g. the one with >1E-8 tons/tanker).

Finally, the remaining areas in the Mediterranean Sea show a risk level of the same order of magnitude (1E-10–1E-9 tons/tanker).

To validate our model, the oil spill accidents in the period 2012–2016 have been considered. In this period, a total number of 19 oil spills have been recorded in the Mediterranean Sea. Most of them in the Greek area and in particular in the Peloponnese and on the eastern Greek coasts.

Largest spills (>700 t) happened in areas of relatively extreme risk in the model (Figure 3), while smaller spills (<7 t) in areas of average risk in the sea, i.e. 1E-10–1E-9 tons/tanker, with the exception of an event in Cyprus (1 t released), which falls into the relative high-risk category (1E-9–1E-8 tons/tanker). Furthermore, one spill accident of 15 t in the Peloponnesus area falls into the relative low risk category (<1E-10 tons/tanker).

In general, historical observations in the period 2012–2016 are generally in good agreement with the model in terms of spill sizes/tanker, thus validating the proposed approach.

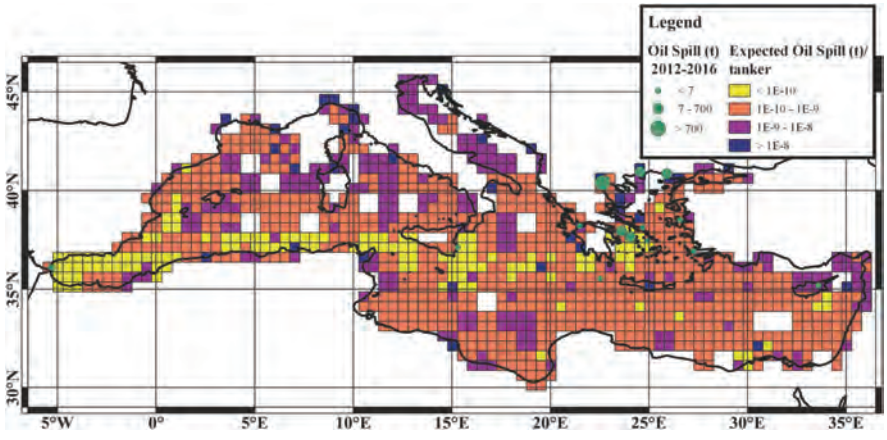


Figure 3. Spatial distribution of the Expected oil spill (t)/tanker in the Mediterranean Sea including the Bosporus strait. The oil spills (t) recorded in the period 2012–2016 used to validate the model are also shown.

4 CONCLUSIONS

In this study, we tested the operability of a spatial hierarchical Bayesian model by applying it to study the oil spill risk of tankers in the Mediterranean Sea.

A comprehensive and up to date database of oil spill accidents was used both as input to the model and for validation purposes. In particular, the results of the validation step confirm that the proposed methodological approach represents a promising tool for dealing with quantitative spatial risk assessments. Indeed, the proper consideration of the spatial component inherent in accidents that can have consequences spread over a geographic region, plays a vital role in supporting spatial monitoring procedures and risk mitigation measures.

The main benefits associated with the proposed methodological approach can be summarised as follows: (i) ability to deliver quantitative results at high spatial resolution, for example in comparison to previous studies (e.g. Burgherr et al., 2015), (ii) possibility to combine expert knowledge with empirical data (e.g. Choy et al., 2009), (iii) accountability for uncertainties and lack of data, and (iv) possibility of integrating sensitivity analysis to improve the interpretability of the parameters (e.g. Roos et al., 2015).

On the other hand, the following limitations should be acknowledged. First, we are using a simplified approach, which does not model the causality between influencing parameters (e.g. size of the tanker, marine currents, wind, spill cause, type of location such as ports, open sea, narrow strait, etc.). Second, the model is assessing an average number of accidents per year, rather than taking into account the actual yearly trend (i.e. traffic patterns).

Finally, we envisage several future developments of this research. The first one refers to the study of how structured expert judgement elicitation protocols could be included in the development of the spatial Bayesian model. Indeed, the facilitation of the quantitative expression of subjective judgement plays a crucial role in the context of probability and risk assessments. However, the presence of the spatial dimension, where basically every pixel of the map becomes an alternative to be assessed, makes this task a particularly challenging one (e.g. Ferretti and Montibeller, 2016), thus opening interesting avenues for future research. The second direction for future developments concerns the combination of the present methodological approach with multi criteria decision analysis to account for multiple impacts associated with the adverse event (e.g. Ferretti and Montibeller, 2017). The third direction of further research concerns the inclusion modelling of data about spatial vulnerabilities within the spatial multi impact Bayesian approach.

ACKNOWLEDGEMENTS

The authors acknowledge financial support from the Swiss National Science Foundation for a scientific exchange, grant ID: IZSEZ0_177357.

REFERENCES

Andrieu, C., De Freitas, N., Doucet, A. & Jordan, M. I. (2003) An introduction to MCMC for machine learning. *Machine Learning,* 50, 5–43.

Bouejla, A., Chaze, X., Guarnieri, F. & Napoli, A. (2014) A Bayesian network to manage risks of maritime piracy against offshore oil fields. *Safety Science,* 68, 222–230.

Burgherr, P. (2007) In-depth analysis of accidental oil spills from tankers in the context of global spill trends from all sources. *J Hazard Mater,* 140, 245–56.

Burgherr, P., Eckle, P. & Michaux, E. (2012) Oil tanker transportation risk: driving factors and consequence assessment. *11th International Probabilistic Safety Assessment and Management Conference (PSAM) & European Safety and Reliability Conference (ESREL)*. Helsinki (Finland).

Burgherr, P., Spada, M. & Kalinina, A. (2015) Regionalized Risk Assessment of Accidental Tanker Spills using Worldwide Data. *The 5th International Symposium on Ship Operations, Management and Economics*. Athens, Greece.

Burgherr, P., Spada, M., Kalinina, A., Hirschberg, S., Kim, W., Gasser, P. & Lustenberger, P. (2017) The Energy-related Severe Accident Database (ENSAD) for comparative risk assessment of accidents in the energy sector. *Safety and Reliability? Theory and Applications*. CRC Press.

Choy, S. L., O'leary, R. & Mengersen, K. (2009) Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology,* 90, 265–277.

Clements, A. C. A., Lwambo, N. J. S., Blair, L., Nyandindi, U., Kaatano, G., Kinung'hi, S., Webster, J. P., Fenwick, A. & Brooker, S. (2006) Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Tropical Medicine & International Health,* 11, 490–503.

Committee on Foundations of Risk Analysis (2015) Society for Risk Analysis Glossary. Society for Risk Analysis.

Cooley, D., Nychka, D. & Naveau, P. (2007) Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association,* 102, 824–840.

Dias, L. C., Morton, A. & Quigley, J. (Eds.) (2018) *Elicitation. The Science and Art of Structuring Judgement*, Springer International Publishing.

Dimaggio, C. (2012) Bayesian Hierarchical Approaches to Spatial Analysis of Injury and Disaster Data. New York, USA, Departments of Anesthesiology and Epidemiology, Columbia University.

Eastwood, J., Jalaludin, B., Kemp, L. & Phung, H. (2014) Bayesian hierarchical spatial regression of

maternal depressive symptoms in South Western Sydney, Australia. *SpringerPlus,* 3, 55.
Eckle, P. & Burgherr, P. (2013) Bayesian Data Analysis of Severe Fatal Accident Risk in the Oil Chain. *Risk Analysis,* 33, 146–160.
Ferretti, V. & Montibeller, G. (2016) Key challenges and meta-choices in designing and applying multi-criteria spatial decision support systems. *Decision Support Systems,* 84, 41–52.
Ferretti, V. & Montibeller, G. (2017) An integrated framework for environmental multi-impact spatial risk analysis. *Risk Analysis*.
Fischhoff, B. (2015) The realities of risk-cost-benefit analysis. *Science,* 350.
Ford, A., Clarke, K. C. & Raines, G. (2009) Modeling Settlement Patterns of the Late Classic Maya Civilization with Bayesian Methods and Geographic Information Systems. *Annals of the Association of American Geographers,* 99, 496–520.
Gelman, A. & Rubin, D. B. (1992) Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science,* 7, 457–511.
Gelman, A. (2003) A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *International Statistical Review,* 71, 369–382.
Goerlandt, F. & Montewka, J. (2015) A framework for risk analysis of maritime transportation systems: A case study for oil spill from tankers in a ship–ship collision. *Safety Science,* 76, 42–66.
He, H. S., Dey, D. C., Fan, X., Hooten, M. B., Kabrick, J. M., Wikle, C. K. & Fan, Z. (2006) Mapping pre-European settlement vegetation at fine resolutions using a hierarchical Bayesian model and GIS. *Plant Ecology,* 191, 85–94.
International Energy Agency (Iea) (2015) *World Energy Outlook,* Paris, France, IEA.
Juan, P., Díaz-Avalos, C., Mejía-Domínguez, N. R. & Mateu, J. (2016) Hierarchical spatial modeling of the presence of Chagas disease insect vectors in Argentina. A comparative approach. *Stochastic Environmental Research and Risk Assessment*.
Kalinina, A., Spada, M., Burgherr, P., Marelli, S. & Sudret, B. (2016) A Bayesian hierarchical modelling for hydropower risk assessment. In Walls, L., Revie, M. & Bedford, T. (Eds.) *Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016.* London, UK, CRC Press.
Keeney, R. L. & Winterfeldt, D. V. (1991) Eliciting probabilities from experts in complex technical problems. *IEEE Transactions on Engineering Management,* 38, 191–201.
Keller, L. R. & Simon, J. (2017) Preference Functions for Spatial Risk Analysis. *Risk Anal*.
Landuyt, D., Van Der Biest, K., Broekx, S., Staes, J., Meire, P. & Goethals, P. L. M. (2015) A GIS plug-in for Bayesian belief networks: Towards a transparent software framework to assess and visualise uncertainties in ecosystem service mapping. *Environmental Modelling & Software,* 71, 30–38.
Liu, R., Chen, Y., Wu, J., Gao, L., Barrett, D., Xu, T., Li, X., Li, L., Huang, C. & Yu, J. (2017) Integrating Entropy-Based Naive Bayes and GIS for Spatial Evaluation of Flood Hazard. *Risk Anal,* 37, 756–773.
Malczewski, J. & Rinner, C. (2015) *Multicriteria Decision Analysis in Geographic Information Science,* New York, USA, Springer Berlin Heidelberg.
O'hagan, A. (2003) Bayesian statistics: principles and benefits. In van Boekel, M., Stein, A. & van Bruggen, A. H. C. (Eds.) *Bayesian Statistics and Quality Modelling in Agro-Food Production Chain.* Dordrecht, Kluwer Academic Publishers.
Psarros, G., Skjong, R. & Vanem, E. (2011) Risk acceptance criterion for tanker oil spill risk reduction measures. *Mar Pollut Bull,* 62, 116–27.
Reis Jr, D. S. & Stedinger, J. R. (2005) Bayesian MCMC flood frequency analysis with historical information. *Journal of hydrology,* 313, 97–116.
Rempec (2011) Imo/unep: Regional information system; part c2, statistical analysis—alerts and accidents database. Regional Marine Pollution Emergency Response Centre for the Mediterranean Sea.
Rodrigues, E. C. & Assunção, R. (2012) Bayesian spatial models with a mixture neighborhood structure. *Journal of Multivariate Analysis,* 109, 88–102.
Roos, M., Martins, T. G., Held, L. & Rue, H. (2015) Sensitivity Analysis for Bayesian Hierarchical Models. *Bayesian Analysis,* 10, 321–349.
Saravana Kumar, V., Devika, S., George, S. & Jeyaseelan, L. (2017) Spatial mapping of acute diarrheal disease using GIS and estimation of relative risk using empirical Bayes approach. *Clinical Epidemiology and Global Health,* 5, 87–96.
Spada, M., Burgherr, P. & Hirschberg, S. (2014) Comparative Assessment of Severe Accidents Risk in the Energy Sector: Uncertainty Estimation Using a Combination of Weighting Tree and Bayesian Hierarchical Models. *12th Probabilistic Safety Assessment and Management (PSAM12)*. Honolulu, HI, USA.
The Marine Environment Protection Committee (Mepc) (2003) Resolution MEPC.111(50) – Amendments to the Annex of the protocol of 1978 relating to the International Convention for the Prevention of Pollution from Ships, 1973.
Tucker, K., Rushton, S. P., Sanderson, R. A., Martin, E. B. & Blaiklock, J. (1997) Modelling bird distributions—a combined GIS and bayesian rule-based approach. *Landscape Ecology,* 12, 77–93.
Vieites, D. R., Nieto-Roman, S., Palanca, A., Ferrer, X. & Vences, M. (2004) European Atlantic: the hottest oil spill hotspot worldwide. *Naturwissenschaften,* 91, 535–8.
Zhu, L., Gorman, D. M. & Horel, S. (2006) Hierarchical Bayesian spatial models for alcohol availability, drug "hot spots" and violent crime. *Int J Health Geogr,* 5, 54.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Risk assessment of worldwide refinery accidents using advanced classification methods: Effects of refinery configuration and geographic location on outcome risk levels

Peter Burgherr & Matteo Spada
Laboratory for Energy Systems Analysis, Paul Scherrer Institute (PSI), Villigen PSI, Switzerland

Marco Cinelli
Future Resilient Systems (FRS), Swiss Federal Institute of Technology (ETH), Zurich, Switzerland
Singapore-ETH Centre (SEC), Singapore

Jurek Blaszczynski
Institute of Computing Science, Poznań University of Technology, Poznań, Poland

Roman Słowiński
Institute of Computing Science, Poznań University of Technology, Poznań, Poland
Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Yvan Pannatier
Swiss Re Corporate Solutions Ltd., Zurich, Switzerland

ABSTRACT: A global dataset of refinery accidents for the years 1990–2016 was analyzed to evaluate the capacity of 16 attributes to differentiate between accidents that cause or not fatalities. For this purpose a Dominance-based Rough Set Approach (DRSA) analysis was carried out. The quality of approximation and accuracy measures confirmed that the established information table is able to distinguish outcome levels in terms of fatalities. Furthermore, the suitability of the extracted rules to describe hidden relationships in the accident dataset was demonstrated. Although, the predictive capacity of the decision rules was not satisfactory, the rules still proved to be useful to identify the attributes that contribute most to assign an accident to the correct outcome class. In summary, this study provided a number of new and substantial insights on worldwide refinery accidents, which complement and extend previous findings for accident frequencies and associated trends as well as different types of consequences.

1 INTRODUCTION

The field of risk assessment and management is a rather young scientific discipline, with most of its fundamental ideas, principles, concepts and applications going back to the 1970s and 1980s (Aven, 2016). One of the first articles addressing risk as a contemporary societal problem was published in Science in 1969 (Starr, 1969). The following publications provide a broad discussion of the concept of risk (Aven, 2012) as well as foundational issues (Aven and Zio, 2014) and important developments in the field (Greenberg et al., 2012, Thompson et al., 2005).

Early developments and applications of probabilistic risk assessment can be traced back to NASA's space program in the 1960s (Cooke, 2009), the aerospace industry (Keller and Modarres, 2005), and the development of nuclear energy in the same period (Otway and Pahner, 1976). In 1975, the US Reactor Safety Study (WASH-1400) marked a major milestone in probabilistic risk assessment (US NRC, 1975), and subsequently spread to other disciplines and countries (Rasmussen, 1981).

The need for systematic and consistent, comparative risk assessment of energy technologies has been recognized since the 1980s (Fritzsche, 1989, Inhaber, 1979). Since then, it became a central element both in the comprehensive evaluation of the risk performance of energy technologies (Burgherr and Hirschberg, 2014), and in the broader context of sustainability assessment (Cinelli et al., 2014, Hirschberg and Burgherr, 2015, Santoyo-Castelazo and Azapagic, 2014). Recently, an increased interest can be observed to assess the frequencies and consequences of accidents over-time, and to compare them among energy chains, chain stages, activities and

infrastructure types as well as for different configurations and locations. The systematic exploitation of existing datasets to extract such useful information is a very important and promising research avenue (Burgherr et al., 2015, Burgherr et al., 2017, Cinelli et al., 2017, Spada and Burgherr, 2016).

In the industrial realm, risk assessment and management is often carried out according to ISO 31000 (ISO, 2009a, Luko, 2013), and ISO 31010 provides a list of 31 risk assessment techniques (ISO, 2009b, Luko, 2014). Other studies present overviews of risk analysis methodologies for specific facilities such as industrial plants (Tixier et al., 2002) or distinguish between formal and informal risk handling strategies for utility companies (Mascini and Bacharias, 2012). In contrast, Jonkman et al. (2003) and Johansen and Rausand (2014) address risk metrics from a more scientific perspective, but their discussion is also relevant for industry, authorities and other stakeholders. Lastly, loss modeling of industrial and insurance companies is often based on historic incident and accident data and the use of empirical and actuarial models and other approaches (Klugman et al., 2008, Daniell et al., 2018, Mannan, 2012).

In this study, the risk of accidents in refineries is analyzed. Within the oil chain, accidents during transports of crude oil and refined products clearly dominate with a share of about 75%, whereas exploration & production (E&P) and refinery activities follow distantly with roughly 10% each (Burgherr and Hirschberg, 2014). Furthermore, E&P and refinery facilities usually represent multi-billion dollar assets, property damage losses in these two sectors are the dominant contributors in the hydrocarbon industry, and refineries exhibit an increasing trend in frequencies and extent of losses (Marsh, 2016).

Recent publications address a variety of topics, including accidents at specific facilities (e.g., Chettouh et al., 2016, Mishra et al., 2014, Saleh et al., 2014), learning from past accidents (e.g., Moura et al., 2017a, Moura et al., 2017b, Russell Vastveit et al., 2015), occurrence of major accidents (Amyotte et al., 2016), and a sustainability metric for petroleum refinery projects (Hasheminasab et al., 2018).

In a previous study, two of the authors analyzed the frequencies and consequences (i.e., fatalities, injured persons) of refinery accidents among four country clusters (Burgherr et al., 2016). This provided interesting insights and conclusion on how refinery configuration and regional differences reflecting the mode of operation are important factors potentially affecting overall refinery risk.

In the current investigation, this retrospective analysis is extended to determine the combined influence and relevance of selected attributes (e.g., country group, accident type, event chain steps, etc.) of refinery accidents on the outcome level (e.g., fatalities). Such knowledge on the potential impact of an energy accident can be used as one piece of information for the selection of response mode and resource allocation. According to the class of risk provided by the model, the Decision Makers (DMs) can decide how many resources to allocate to respond. Retrospective analysis for patterns recognition and decision support model development have been recently applied in different studies, e.g., investigation of antimicrobial activity (Pałkowski et al., 2014), selection of sustainable project portfolios (Zaras et al., 2012), and natural gas accidents (Cinelli et al., 2017), among others.

In the current research, a dataset of refinery accidents extracted from PSI's Energy-related Severe Accident Database (ENSAD) was analyzed to address three main objectives:

1. Assess the information structure of accidents from ENSAD to distinguish the events that did cause or did not cause fatalities.
2. Investigate the relationships between a set of descriptors (attributes) for the energy accidents and the outcome level (fatalities or no fatalities).
3. Study the capacity of the attributes to distinguish between accidents that cause or not fatalities.

2 METHODS

2.1 *Accident data*

The refinery accident data used in this study were extracted from PSI's ENSAD. The ENSAD has first been released in 1998 (Hirschberg et al., 1998), and since then regularly updated and extended (Burgherr et al., 2015, Burgherr and Hirschberg, 2014). Currently, the database is migrated from a standalone MS-Access to an new, interactive, web-based GIS database named ENSAD v2.0 (Burgherr et al., 2017). In ENSAD, complete energy chains are considered because accidents can occur during all stages and activities, and not just the actual power and/or heat generation (Burgherr and Hirschberg, 2008). In general, the focus of ENSAD is on so-called severe accidents because accidents with larger consequences are a major concern for industry and authorities, but also receive most attention by the general public (Burgherr and Hirschberg, 2014). To be classified as severe an accident has to fulfil at least one of seven threshold criteria, for example ≥ 5 fatalities, ≥ 10 injured persons, etc. (Burgherr and Hirschberg, 2008).

For the current study on global refinery accidents all accidents in ENSAD with at least one fatality or one injured person during the years 1990–2016 were taken into account. The lower

fatality and injury thresholds for refinery accidents were possible because a dedicated search for smaller accidents was carried out to ensure a sufficiently complete coverage. This larger dataset allows analyzing a broader set of accident attributes compared to the previous study conducted in 2016 (Burgherr et al., 2016).

In total, the refinery accident dataset comprised 698 accidents, of which 277 resulted in at least one fatality, 597 in at least one injured person, and 176 had both consequences.

2.2 *Refinery data*

Information on refinery configuration and characteristics was retrieved from the World Wide Refinery Survey (WWRS) of the Oil & Gas Journal (Penn Energy Research, 2010). The coupling between the ENSAD and WWRS databases was based on georeferenced refinery locations (i.e., geographic coordinates). WWRS data include details on refinery units and processes as well as corresponding capacities expressed in barrels per calendar day (bbl/cd), the Nelson Complexity Index (NCI) and the Equivalent Distillation Capacity (EDC). In some cases, an accident could not be assigned to a specific refinery in the WWRS database: (1) the accident description did not allow identifying the correct refinery if several are located in the same area (10 cases); (2) the refinery can be located, but is not included in the WWRS list, which is mostly the case for small, private refineries (32); or (3) the accident description is so incomplete that the refinery can only be assigned to a country, but no specific location (8).

Additionally, five more refinery attributes were based on data from Swiss Re, providing additional information on regional differences in refinery hazards. First, each refinery accident was assigned to one of four regional country clusters, representing different plant operation styles (Pannatier, pers. comm.). Generally, loss burden differs among clusters, which is attributable to operational hazard factors such as mode of operation, attitude towards safety, turnaround period, maintenance, etc. The four clusters were given the following names:

- "USA": USA, Canada, UK, Australia
- "Europe": Europe, Singapore, South Korea, Japan, Saudi Arabia, Gulf States, Egypt
- "Russia": Russia, Former Soviet Union, Eastern Europe
- "Other": South America, Africa, Maghreb, other Middle East, rest of Asia

Second, all refinery units were allocated to three complexity classes G1, G2 and G3 to reflect increasing fire and explosion hazard (see Table 1), based on a Swiss Re expert categorization (Pannatier, pers. comm.). The class G0 includes all units that do not belong to one of the three other classes. An overview of refinery units considered and their assignment to a complexity class is shown in Table 1.

Three refinery unit attributes were defined, namely (1) the unit in which the accident started; (2) the complexity class of the unit where the accident started; and (3) the complexity class of the most hazardous unit(s) in a refinery.

Third, a Swiss Re Hazard Index (SR HI) was calculated for all refineries, combining the capacity, complexity and toxic hazard of a refinery with weight factors three, two and one, respectively.

2.3 *Overview of accident attributes*

Table 2 shows an overview of the 16 attributes that were used to analyze the previously defined dataset of refinery accidents. Ten attributes concern specific accident characteristics, whereas the other six provide information about refinery configuration and operational hazard. Out of these, three were directly taken from the WWRS database, and three were established using information from Swiss Re and WWRS.

2.4 *Dominance-based Rough Set Approach (DRSA)*

The dataset of refinery accidents developed in this study can be seen as an information table, where characteristics of the accidents are condition attributes (independent variables) and the

Table 1. Assignment of refinery units to complexity classes (CC) G0 to G3.

G0	G1	G2	G3
Crude Unit	Vacuum Distillation	Catalytic Cracking	Catalytic Hydro-cracking
Coking	Catalytic Reforming	Isomerization	Alkylation
Thermal Cracking	Catalytic Hydrotreating		Polymerization
Hydrogen Production			Lubes
Hydrogen Recovery			
Coke			
Sulphur			
Asphalt			

Table 2. Overview and description of refinery attributes. O-U: ordered, but direction unknown; O-K: ordered and direction is known; N-O: not ordered.

Attribute name	Description and values	Unit	Ordering
Year	Year in which an accident occurred: 1990–2016.	yr	O-U
Country Cluster	Country clusters represent different plant operation styles (see section 2.2): USA, Europe, Russia, Other.	–	N-O
Start Unit	Name of refinery unit in which the accident started (see Table 1).	–	N-O
Start CC	Complexity class (CC) of refinery unit where accident started: G0, G1, G2, G3.	–	O-K
High CC	Unit(s) with highest complexity class (CC) in a refinery.	–	O-K
No Acc	Number of accidents that occurred in an individual refinery during the period of observation.	–	O-U
Nelson Complexity Index (NCI)	A measure for refinery complexity, replacement costs, and values (Johnston, 1996 and references therein).	–	O-U
HI	Swiss Re Hazard Index (HI) (see section 2.2).	–	O-U
CRUDE	Crude capacity of refinery	bbl/cd	O-U
EDC	Equivalent Distillation Capacity is another measure to compare refineries. is another means of comparing refinery costs, for which a refinery's atmospheric distillation capacity is multiplied by its overall complexity rating, resulting in complexity barrels (Johnston, 1996).	cbbl/cd	O-U
Acc Type	Type of accident: Technical Failure (TF), Human Factor (HF), Human-Technical (HT), Natural Hazard (NH), Intentional Attack (IA; not included in current analysis).	–	N-O
EC1 to EC5	Event chain steps 1 to 5; sequence of events as given in available information sources. In total, 27 values are used, e.g., explosion, fire, release, etc.	–	N-O
Fatalities (i.e., outcome of the accident)	Number of fatalities that an accident caused. Two outcome levels are distinguished: no fatalities vs. fatalities.	–	O-K

presence or not of fatalities is the decision attribute or outcome level (dependent variable). To achieve the research objectives, Dominance-based Rough Set Approach (DRSA) analysis (Greco et al., 2001, Słowiński et al., 2015) was applied to the refinery information table using the jRS library and jMAF software package (Błaszczyński et al., 2013). The method is well suited for this case study as it can handle quantitative and qualitative information without the need of transforming them into numerical or binary values, and it accepts inconsistencies in the dataset.

Another important characteristic of the present dataset is that it is not always known a priori whether the condition attributes are of the gain—or cost—type in relation to the fatalities. Gain-type means that the greater the value of an attribute, the less likely the accident causes fatalities; conversely, cost-type means that the lower the value of the attribute, the less likely the accident causes fatalities. DRSA is capable of discovering this type of information, named global monotonicity (Blaszczyński et al., 2012). Furthermore, local monotonicity can also be discovered, meaning the interval of attribute values in which an attribute shows a gain—or cost-type behavior in the discovered pattern (see Table 3). The applied transformation is non-invasive, i.e., it does not bias the matter of discovered relationships. With reference to the set of refinery attributes, those without an indication of their preference order for which global and local monotonicity were studied are "Year", "No Acc", "NCI", "HI", "CRUDE", and "EDC" (O-U type in column Ordering, Table 2).

The relationships between the values of the attributes of the energy accidents and the cause or not of fatalities were discovered by means of decision rules, which are objective cause-effect patterns hidden in the refinery dataset. A decision rule is in the form of EàH, namely "if E, then H". E is the condition part (also called premise or evidence) and H is the conclusion part (also called decision part). Sets of decision rules, which are essential for this analysis, were induced using the Variable Consistency Dominance-based Learning from ExaMples (VC-DomLEM) algorithm (Błaszczyński et al., 2011b). Accidents that consistently cause either fatalities or not are assigned to what are called lower approximations. VC-DomLEM uses accidents that belong to such lower approximations as training examples to induce sets of certain decision rules, which may be later used to classify new accidents.

Sets of rules, induced by VC-DomLEM, may be used to construct more accurate ensemble classifiers in variable consistency bagging (Błaszczyński et al., 2009, Błaszczyński et al., 2010). Another

Table 3. Selected examples of decision rules obtained from the DRSA of the refinery accident dataset for accidents causing fatalities or not. Mbbl: million barrels; Supp.: Support; Conf.: Confidence. Other abbreviations see Table 2.

Year	Country cluster	Start Unit	Start CC	No Acc	NCI	HI	CRUDE Mbbl/cd	EDC	Acc Type	Supp.	Conf.
Rules for accidents that **did not cause fatalities**											
> 2002 ↑	Europe				≤ 5.4 ↓					11	0.92
> 2010 ↑	Other				≤ 6.3 ↓		≤ 110 ↓			13	0.93
	USA			≤ 4 ↓			≤ 96 ↓		HT	11	0.92
	USA		≤ 0			≤ 18.57 ↓	≤ 149.5 ↓		TF	15	0.82
> 2011 ↑	Russia				≤ 7.4 ↓					13	0.93
Rules for accidents that **caused fatalities**											
[2002;2004]		Alkylation							HF	10	0.91
[1996;2004]		Crude						≥ 1521 ↓	HF	11	0.85
≤ 2007 ↑						≥ 18.42 ↓	[126.9;189]		HF	11	0.79
≤ 2011 ↑			≥ 2			≥ 21.28 ↓			TF	12	0.80
	USA		≥ 2		≥ 10.65 ↓	≥ 20.0 ↓			TF	15	0.75

extension of the bagging approach is applied when the analyzed dataset suffers from class imbalance. In our case, Neighbourhood balanced bagging (NBBag) was used to increase the predictive capabilities of constructed rule classifiers (Błaszczyński and Stefanowski, 2015). Moreover, the strategies, which were used to evaluate our proposition, consisted in learning the decision rules from a subset of the original accident dataset and testing them on the remaining accidents to evaluate whether they assign the correct fatality class or not.

The capacity of the attributes to distinguish between accidents causing or not causing fatalities was assessed through a relevance measure called confirmation (Błaszczyński et al., 2011a, Greco et al., 2001). It assesses the degree to which presence of an attribute in the condition part of a rule confirms correct prediction (i.e., fatalities or no fatalities) on a subset of accidents the rules were not extracted from. The higher the value of the relevance measure (i.e., confirmation measure) the more important for correct prediction the attribute is.

3 RESULTS

3.1 *Quality of classification and accuracy*

The ENSAD-based refinery accident dataset is close to unitary classification quality (0.99 on a 0–1 scale), which means that there are very few inconsistent accidents that have the same values for the attributes and either caused fatalities or did not. In other words, the 16 attributes selected to develop the information table are relevant to differentiate between accidents causing or not causing fatalities. This is confirmed by the accuracy for not causing or causing fatalities with values of 0.98 and 0.97 (also on a 0–1 scale), respectively.

3.2 *Decision rules*

Decision rules represent unique patterns hidden within the dataset and they unveil the cause-effect relationships between the values of the attributes and the outcome level, i.e., the fatalities class in this case. Decision rules do not only include important condition attributes, but they also contain a minimal number of elementary conditions that are necessary for representation of the cause–effect relationships existing in the information table, from which all inessential and redundant information is removed.

Decision rules are induced by DRSA using the VC-DomLEM algorithm. Table 3 shows ten representative rules, five for the description of accidents causing fatalities and five for not causing fatalities. These decision rules can be viewed as interesting patterns among all patterns that can be found to describe features of the refinery accident dataset. The rules are characterized by useful parameters, including the support (i.e., the number of accidents that support the rule) and the confidence level. The selected rules have all a strong confidence level, close or above 0.75. This indicates that at least 75% of the accidents in the dataset with the stated values for the condition attributes lead to fatalities or not.

In the following, some rather simple explanatory examples for the interpretation of the rules given in Table 3 are briefly discussed. Accidents after 2002 in the cluster Europe with a NCI smaller 5.4 caused no fatalities. For Russia the same was only true after 2011 and for a higher threshold of NCI, indicating a general better performance of Europe. The same combination of attributes for the cluster Other ranks between Europe and Russia, but additionally crude capacity has to be lower than 110,000 bbl/cd. In contrast, accidents with fatalities tend to have higher values for NCI, HI, crude capacity and EDC. Additionally, fatal

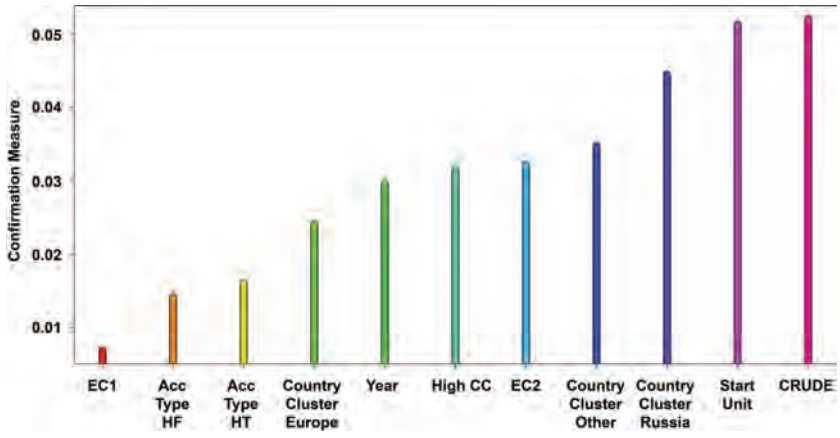


Figure 1. Confirmation measure for attributes most relevant to distinguish between accident outcomes in terms of fatalities.

accidents for combinations of these attributes generally appeared to occur more likely in the earlier years of the observation period. Finally, accidents starting in higher hazardous units (i.e., higher complexity class) cause more often fatalities, and even still in more recent years.

3.3 *Relevance of attributes*

The predictive capacity of the rules was tested by means of 10-fold cross-validation, but unfortunately the predictive accuracy of DRSA decision rules was not satisfactory. In fact, only 48% of the accidents causing fatalities and 53% of those not causing fatalities were assigned the correct class. Nonetheless, these rules can be used to extract the attribute relevance, which indicates to what extent attributes are relevant (useful) to distinguish between accidents causing fatalities or not.

The confirmation measure used for the computation of the attributes relevance quantifies the extent to which the presence of an attribute in the rule suggests a correct class (Figure 1). Attributes CRUDE, Start Unit and the Country cluster are the most relevant for discerning between the accidents in terms of the outcome. Other attributes which are still relevant, though at a lower extent, for correct prediction are the first and second event chain (EC1, EC2), the highest complexity class (High CC) of the unit(s) present in a refinery, the year when the accident took place, and the type of accident.

4 CONCLUSIONS

For the current analysis refinery accident data from ENSAD for the period 1990–2016 were combined with selected refinery characteristics using data from the WWRS database and from Swiss Re. The results demonstrated that the application of the Dominance-based Rough Set Approach (DRSA) analysis to such a complex dataset is feasible and useful to gain detailed insights about the structure and to reveal hidden relationships between accident attributes and their outcome in terms of fatalities.

The quality of approximation and accuracy measures confirmed that the information table based on 16 attributes is useful to differentiate between accidents causing or not fatalities. Furthermore, the extracted decision rules helped to evaluate hidden relationships in the accident dataset that go beyond traditional measures such as aggregated risk indicators or frequency-consequence curves. In particular, aspects of the monotonicity of attributes could be addressed as well as how higher (gain-type) or lower (cost-type) attribute values contribute to a reduced likelihood that an accident results in fatalities.

Finally, the decision rules were tested with regard to their predictive capacity to use the set of rules to assign potential future accidents to the correct outcome level. Although, the predictive accuracy was not satisfactory, the rules still proved to be useful to identify the attributes that contribute most to assign an accident to the correct outcome class.

In summary, the current study provided a number of new and substantial insights on global refinery accidents; thus extending and complementing a previous study by Burgherr et al. (2016).

Potential avenues of future research include options to enhance the predictive capacity of the

rules, which could be achieved either by adjusting the present attributes or by considering additional attributes. This is an essential aspect for decision makers and other stakeholders. They are, for example, concerned with the development of low-probability high-impact scenarios, pre- and post-event strategies to mitigate potential consequences of future accidents or the improvement of general safety culture.

ACKNOWLEDGEMENTS

This study was partially carried out under a collaboration agreement between PSI's Technology Assessment group and the Risk Engineering department of Swiss Re Corporate Solutions Ltd.

This research was also partly conducted at the Future Resilient Systems at the Singapore-ETH Centre, which was established collaboratively between ETH Zurich and Singapore's National Research Foundation (FI 370074011) under its Campus for Research Excellence and Technological Enterprise programme.

REFERENCES

Amyotte, P.R., Berger, S., Edwards, D.W., Gupta, J.P., Hendershot, D.C., Khan, F.I., Mannan, M.S. & Willey, R.J. (2016) Why major accidents are still occurring. *Current Opinion in Chemical Engineering,* 14, 1–8.

Aven, T. & Zio, E. (2014) Foundational Issues in Risk Assessment and Risk Management. *Risk Analysis,* 34, 1164–1172.

Aven, T. (2012) The risk concept—historical and recent development trends. *Reliability Engineering & System Safety,* 99, 33–44.

Aven, T. (2016) Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research,* 253, 1–13.

Blaszczyński, J., Greco, S. & SŁowiński, R. (2012) Inductive discovery of laws using monotonic rules. *Engineering Applications of Artificial Intelligence,* 25, 284–294.

Burgherr, P. & Hirschberg, S. (2008) A comparative analysis of accident risks in fossil, hydro and nuclear energy chains. *Human and Ecological Risk Assessment,* 14, 947–973.

Burgherr, P. & Hirschberg, S. (2014) Comparative risk assessment of severe accidents in the energy sector. *Energy Policy,* 74, S45–S56.

Burgherr, P., Giroux, J. & Spada, M. (2015) Accidents in the Energy Sector and Energy Infrastructure Attacks in the Context of Energy Security. *European Journal of Risk Regulation,* 6, 271–283.

Burgherr, P., Spada, M., Kalinina, A., Eckle, P. & Pannatier, Y. (2016) Accident risk assessment of refineries depending on configuration and geographic location. IN Walls, L., Revie, M. & Bedford, T. (Eds.) *Risk, Reliability and Safety—Innovating Theory and Practice.* London, UK, CRC Press, Taylor & Francis Group.

Burgherr, P., Spada, M., Kalinina, A., Hirschberg, S., Kim, W., Gasser, P. & Lustenberger, P. (2017) The Energy-related Severe Accident Database (ENSAD) for comparative risk assessment of accidents in the energy sector. IN Cepin, M. & Bris, R. (Eds.) *Safety and Reliability—Theory and Applications.* London, UK, CRC Press, Taylor & Francis Group.

Błaszczyński, J. & Stefanowski, J. (2015) Neighbourhood sampling in bagging for imbalanced data. *Neurocomputing,* 150, 529–542.

Błaszczyński, J., Greco, S., Matarazzo, B., Słowiński, R. & Szeląg, M. (2013) jMAF—Dominance-Based Rough Set Data Analysis Framework. IN Skowron, A. & Suraj, Z. (Eds.) *Rough Sets and Intelligent Systems—Professor Zdzisław Pawlak in Memoriam: Volume 1.* Berlin, Heidelberg, Springer Berlin Heidelberg.

Błaszczyński, J., Słowiński, R. & Stefanowski, J. (2009) Feature Set-based Consistency Sampling in Bagging Ensembles. *From Local Patterns To Global Models (LEGO), ECML/PKDD Workshop 2009.* Bled Slovenia.

Błaszczyński, J., Słowiński, R. & Stefanowski, J. (2010) Variable Consistency Bagging Ensembles. IN Peters, J.F. & Skowron, A. (Eds.) *Transactions on Rough Sets XI.* Berlin, Heidelberg, Germany, Springer.

Błaszczyński, J., Słowiński, R. & Susmaga, R. (2011a) Rule-Based Estimation of Attribute Relevance. IN Yao, J., Ramanna, S., Wang, G. & Suraj, Z. (Eds.) *Rough Sets and Knowledge Technology.* Berlin Heidelberg, Germany, Springer.

Błaszczyński, J., Słowiński, R. & Szeląg, M. (2011b) Sequential covering rule induction algorithm for variable consistency rough set approaches. *Information Sciences,* 181, 987–1002.

Chettouh, S., Hamzi, R. & Benaroua, K. (2016) Examination of fire and related accidents in Skikda Oil Refinery for the period 2002–2013. *Journal of Loss Prevention in the Process Industries,* 41, 186–193.

Cinelli, M., Coles, S.R. & Kirwan, K. (2014) Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment. *Ecological Indicators,* 46, 138–148.

Cinelli, M., Spada, M., Miebs, G., Kadziński, M. & Burgherr, P. (2017) Classification models for the risk assessment of energy accidents in the natural gas sector. *The 2nd International Workshop on Modelling of Physical, Economic and Social Systems for Resilience Assessment, 14–16 December 2017.* Ispra, Italy.

Cooke, R.M. (2009) *A Brief History of Quantitative Risk Assessment. Resources 172,* Washington DC, USA, Resources for the Future.

Daniell, J.E., Wenzel, F. & Schaefer, A.M. (2018) Chapter 5 - The Use of Historic Loss Data for Insurance and Total Loss Modeling. IN Michel, G. (Ed.) *Risk Modeling for Hazards and Disasters.* Amsterdam, The Netherlands, Elsevier.

Fritzsche, A.F. (1989) The health risks of energy production. *Risk Analysis,* 9, 565–577.

Greco, S., Matarazzo, B. & Slowinski, R. (2001) Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research,* 129, 1–47.

Greenberg, M., Haas, C., Cox, A., Lowrie, K., McComas, K. & North, W. (2012) Ten Most Important Accom-

plishments in Risk Analysis, 1980–2010. *Risk Analysis,* 32, 771–781.
Hasheminasab, H., Gholipour, Y., Kharrazi, M. & Streimikiene, D. (2018) A novel Metric of Sustainability for petroleum refinery projects. *Journal of Cleaner Production,* 171, 1215–1224.
Hirschberg, S. & Burgherr, P. (2015) Sustainability Assessment for Energy Technologies. *Handbook of Clean Energy Systems.* John Wiley & Sons, Ltd.
Hirschberg, S., Spiekerman, G. & Dones, R. (1998) *Severe accidents in the energy sector—first edition. PSI Report No. 98–16,* Villigen PSI, Switzerland, Paul Scherrer Institut.
Inhaber, H. (1979) Risk with energy from conventional and nonconventional sources. *Science,* 203, 718–723.
ISO (2009a) *ISO 31000:2009, Risk management—Principles and guidelines,* Geneva, Switzerland, International Organization for Standardization (ISO).
ISO (2009b) *ISO 31010:2009, Risk management—Risk assessment techniques,* Geneva, Switzerland, International Organization for Standardization (ISO).
Johansen, I.L. & Rausand, M. (2014) Foundations and choice of risk metrics. *Safety Science,* 62, 386–399.
Johnston, D. (1996) Complexity index indicates refinery capability, value. *Oil and Gas Journal,* 94, 74–80.
Jonkman, S.N., van Gelder, P.H.A.J.M. & Vrijling, J.K. (2003) An overview of quantitative risk measures for loss of life and economic damage. *Journal of Hazardous Materials,* 99, 1–30.
Keller, W. & Modarres, M. (2005) A historical overview of probabilistic risk assessment development and its use in the nuclear power industry: a tribute to the late Professor Norman Carl Rasmussen. *Reliability Engineering & System Safety,* 89, 271–285.
Klugman, S.A., Panjer, H.H. & Willmot, G.E. (2008) *Loss Models: From Data to Decisions, Third Edition,* Hoboken, NJ, John Wiley & Sons, Inc.
Luko, S.N. (2013) Risk Management Terminology. *Quality Engineering,* 25, 292–297.
Luko, S.N. (2014) Risk Assessment Techniques. *Quality Engineering,* 26, 379–382.
Mannan, S. (2012) Chapter 2 – Incidents and Loss Statistics. IN Lees, F. (Ed.) *Lees' Loss Prevention in the Process Industries (Fourth Edition)*. Oxford, UK, Butterworth-Heinemann.
Marsh (2016) *The 100 Largest Losses 1974–2015. Large property damage losses in the hydrocarbon industry. 24th edition*, London, UK, Marsh Ltd.
Mascini, P. & Bacharias, Y. (2012) Integrating a Top-Down and a Bottom-Up Approach: Formal and Informal Risk-Handling Strategies in a Utility Company. *Risk Analysis,* 32, 1547–1560.
Mishra, K.B., Wehrstedt, K.-D. & Krebs, H. (2014) Amuay refinery disaster: The aftermaths and challenges ahead. *Fuel Processing Technology,* 119, 198–203.
Moura, R., Beer, M., Patelli, E. & Lewis, J. (2017a) Learning from major accidents: Graphical representation and analysis of multi-attribute events to enhance risk communication. *Safety Science,* 99, 58–70.
Moura, R., Beer, M., Patelli, E., Lewis, J. & Knoll, F. (2017b) Learning from accidents: Interactions between human factors, technology and organisations as a central element to validate risk studies. *Safety Science,* 99, 196–214.
Otway, H.J. & Pahner, P.D. (1976) Risk assessment. *Futures,* 9, 122–134.
Pałkowski, Ł., Błaszczyński, J., Skrzypczak, A., Błaszczak, J., Kozakowska, K., Wróblewska, J., Kożuszko, S., Gospodarek, E., Krysiński, J. & Słowiński, R. (2014) Antimicrobial Activity and SAR Study of New Gemini Imidazolium-Based Chlorides. *Chemical Biology & Drug Design,* 83, 278–288.
Penn Energy Research (2010) *2010 Oil & Gas Journal Worldwide Refining Survey*.
Rasmussen, N.C. (1981) The application of probabilistic risk assessment techniques to energy technologies. *Annual Review of Energy,* 6, 123–138.
Russell Vastveit, K., Boin, A. & Njå, O. (2015) Learning from incidents: Practices at a Scandinavian refinery. *Safety Science,* 79, 80–87.
Saleh, J.H., Haga, R.A., Favarò, F.M. & Bakolas, E. (2014) Texas City refinery accident: Case study in breakdown of defense-in-depth and violation of the safety–diagnosability principle in design. *Engineering Failure Analysis,* 36, 121–133.
Santoyo-Castelazo, E. & Azapagic, A. (2014) Sustainability assessment of energy systems: integrating environmental, economic and social aspects. *Journal of Cleaner Production,* 80, 119–138.
Spada, M. & Burgherr, P. (2016) An aftermath analysis of the 2014 coal mine accident in Soma, Turkey: Use of risk performance indicators based on historical experience. *Accident Analysis & Prevention,* 87, 134–140.
Starr, C. (1969) Social Benefit versus Technological Risk. *Science,* 165, 1232–1238.
Słowiński, R., Greco, S. & Matarazzo, B. (2015) Rough Set Methodology for Decision Aiding. IN Kacprzyk, J. & Pedrycz, W. (Eds.) *Springer Handbook of Computational Intelligence.* Berlin, Heidelberg, Springer Berlin Heidelberg.
Thompson, K.M., Deisler, P.F. & Schwing, R.C. (2005) Interdisciplinary Vision: The First 25 Years of the Society for Risk Analysis (SRA), 1980–2005. *Risk Analysis,* 25, 1333–1386.
Tixier, J., Dusserre, G., Salvi, O. & Gaston, D. (2002) Review of 62 risk analysis methodologies of industrial plants. *Journal of Loss Prevention in the Process Industries,* 15, 291–303.
US NRC (1975) *Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants, WASH-1400 (NUREG-75/014)*, Washington, DC, USA, US Nuclear Regulatory Commission.
Zaras, K., Marin, J.-C. & Boudreau-Trude, B. (2012) Dominance-Based Rough Set Approach in Selection of Portfolio of Sustainable Development Projects. *American Journal of Operations Research,* 2, 502–508.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A novel navigational risk analysis method using interval type-2 fuzzy sets

C.L. Fan, D. Zhang, J.F. Zhang & H.J. Yao
Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan, P.R. China
National Engineering Research Center for Water Transport Safety, Wuhan, P.R. China

ABSTRACT: Navigational risk is critical to the shipping industry, particularly to the inland water due to its sensitive environment. To this end, this paper proposes a novel method for navigational risk analysis, incorporating Analytic Hierarchy Process (AHP), interval type-2 fuzzy sets (IT2FSs), and a similarity measure. In light of literature review, a hierarchical model for navigational risk analysis is developed including three levels: objective level, criteria level, and factor level. Then, weights of criteria and associated factors are obtained using AHP, and factors are evaluated by experts using IT2FSs. Then, results of the degree of similarity between aggregated experts' evaluation and 7-member IT2FSs representing 7 grades of risk are obtained and used to determine the degree of navigational risk of the objective level. The application of the proposed method is illustrated by assessing navigational risk of 5 vessels cruised in the Yangtze River. The proposed method provides room for more flexibility in modeling and handling subjective uncertainties in navigational risk analysis. Also, the proposed method identifies navigational risk of multiple vessels, which is useful in enhancing situation awareness of safety.

1 INTRODUCTION

Risk assessment is a critical issue in the maritime transport system. Due to market demand and development of high technology, vessels in this system have been more specific in their functions, much larger in their sizes, much faster in their cruise velocities, and much more in their quantity. This transformation, however, produces an increasing threat to navigational safety by ship collisions, grounding accidents, etc., and also to ecosystem by emissions, oil spill, ship's ballast water and sediments' discharge, etc. It has been estimated that potential losses of a 19,000 TEU containership sinking is up to US$ 1 billion and it would take two years to remove all the containers from the foundered mega containership (Allianz 2015). Therefore, it is necessary to control and manage such risk. To this end, Formal Safety Assessment (FSA), proposed by the UK in 1993 (Wang 2000), was approved by International Maritime Organization (IMO) in 2002 as principles of risk management and a systematic process, which has been broadly used and developed in maritime industry around the world (Görçün & Burak 2015, Hu et al. 2007, Montewka et al. 2014, Zhang et al. 2011, Zhang et al. 2013). The last update of the FSA guidelines was in 2015 (IMO 2015), which illustrates the framework of FSA includes 5 modules: Identification of hazards; risk analysis; Risk control options; Cost–benefit assessment; Recommendations for decision-making. The first two modules of FSA are the interest of this paper.

Traditionally, navigational risk has been assessed by statistical method (Chen et al. 2015, Jin 2014, Kujala et al. 2009, Meng et al. 2014, Wang et al. 2013), simulation (Huang et al. 2017, Lušić & Čorić 2015, Montewka et al. 2010, Qu et al. 2011, Wang et al. 2009), or evaluation of risk factors (Kum & Sahin 2015, Thieme et al. 2017, Tian et al. 2013, Xu et al. 2016, Øien 2001). However, maritime transportation is a large-scale physical and socio-technological system. Such complicated characteristics lead to that navigational risk assessment entails more specific methods which can effectively deal with epistemic uncertainties, e.g. discord in ship-ship or ship-shore communication, imprecision in ship identification, temporal and spatial differences, etc. These uncertainties have been handled by many methods, e.g. Bayesian Networks (Akhtar & Utne 2014, Brito & Griffiths 2016, Hänninen & Kujala 2012, Martins & Maturana 2013, Zhang et al. 2013), Dempster-Shafer theory of evidence (Li & Pang 2013, Talavera et al. 2013, Zhang et al. 2016), etc.

Navigational risk assessment also faces with data uncertainty (Hänninen 2014), i.e. the lack of data in detail and the difficulty in evaluating the credibility and validation of the data, and with ambiguity in seafarers' decision making in some cases.

To overcome these problems, experts' judgements have been an appropriate alternative to objective data. Experts' subjective opinions have been modeled by many studies using fuzzy set theory (Karahalios 2014, Zhao et al. 2009) or integrated fuzzy methods, e.g. Fuzzy Evidence Reasoning (Yang et al. 2014), Fuzzy Analytic Hierarchy Process (AHP) (Andrew et al. 2014, Beşikçi et al. 2016, Celik et al. 2009, Pak et al. 2015), etc. However, most of these studies are based on Type-1 fuzzy sets (T1FSs) introduced by Zadeh (1965), whose membership grade, or the height is a crisp number within [0, 1]; however, determining an exact membership function for a fuzzy set is not always possible without a loss of information (Gorzalczany 1987), which often causes biased conclusion (Liu et al. 2017). To overcome this, the concept of type-2 fuzzy sets (T2FSs) is also introduced by Zadeh (1975) as an extension of T1FS, characterized by primary and secondary membership. Nevertheless, T2FSs have not been widely applied due to complicated computation. To reduce such heavy computation effort, interval type-2 fuzzy sets (IT2FSs) is proposed by Gorzalczany (1987) with that all the values of secondary membership of T2FSs are equal to 1. IT2FSs can not only represent uncertainty better than that of T1FSs and simplify the computation compared with T2FSs (Hu et al., 2013), but also produce more accurate and robust results (Dereli & Altun 2013). Therefore, IT2FSs have been widely used in decision making, risk analysis, etc.

Regarding risk analysis using IT2FSs, there are mainly three methods which are using IF-THEN rules (Rahib et al. 2016), ranking IT2FSs (Bozdag et al. 2015), and measuring the degree of similarity (Chen & Chen 2008, Chen & Chen 2009, Chen & Sanguansat 2011, Sen et al. 2016, Wei & Chen 2009). Among them, the last one is the most popular. However, selecting a reasonable similarity measure is an open subject, depending on the real application environments (Deng et al. 2011). The similarity measure proposed by Chen & Sanguansat (2011) is adopted in this paper.

This paper aims to assess navigational risk. A hierarchical structure for identifying navigational risk factors is constructed with three levels: objective level, criteria level, and factor level. Values of relative importance of certain criteria with respect to the objective, and of relative importance of certain factor with respect to corresponding criteria are quantified using Analytic Hierarchy Process (AHP) (Saaty 1980). To evaluate risk factors, linguistic terms and corresponding IT2FSs are utilized. Then, measuring the degree of similarity between the aggregated evaluation of factors and seven grades of risk represented by IT2FSs is used to transform the aggregated result into corresponding linguistic term.

The rest of this paper is organized as follows. Section 2 presents a model for navigational risk analysis incorporating three levels—objective level, criteria level, and factor level, and corresponding grades for describing the objective level and evaluating the factor level. Based on this model, Section 3 illustrates the proposed method. Using the proposed method, navigational risk of 5 vessels is introduced in Section 4. Section 5 concludes this paper.

2 NAVIGATIONAL RISK ANALYSIS MODEL

2.1 *Evaluation model for navigational risk*

For navigational risk analysis, a model with three levels is established based on Tian et al. 2013, Zhang et al. 2011, Zhang et al. 2013, and Zhang et al. 2016, and shown in Figure 1. The objective level aims to assess the navigational risk; the criteria level comprises four kinds of criteria—static information of vessel, dynamic information of vessel, environment, and management; the factor level includes 18 factors without directly taking into account human factor which is an important parameter for navigation safety (Fan et al. 2017, Kujala et al. 2009). The reason is that it is difficult to directly determine the risk of human because human is the source of success as well as failure (Hollnagel 2014). In contrast, human behavior is partially represented, i.e. dynamic information of vessel partially represents watch officers' decision; shipowner stands for organization factor; the maintenance situation of navigational aids par-

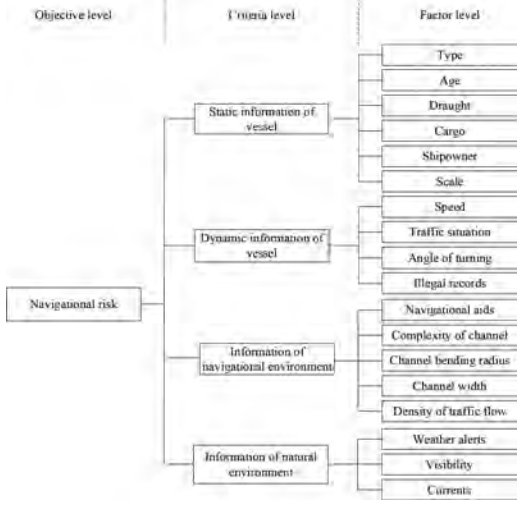


Figure 1. A hierarchical structure for navigational risk.

tially indicates the administration level. Factors in the factor level are as follows:

1. Static information of vessel includes 6 factors: shipowner, type, age, scale, cargo, and draught. For example, the attribute of cargo may lead to different navigational risk. The ratio of channel depth to vessel draught generally is larger than one, otherwise, the accident of grounding may happen.
2. Dynamic information of vessel contains 4 factors: speed, angle of turning, illegal records, and traffic situation. For example, if a vessel has many ship detention times per year by Port State Control officer or Flag State Control officer, this vessel may be critical from the view of department of administration.
3. Information of navigational environment covers 5 factors: density of traffic flow, aid facility, channel width, channel bending radius, and complexity of channel. For example, vessels benefit a lot from navigational aids which are performed and maintained well, while little from a complicated channel with a lot of bridges over it.
4. Information of natural environment incorporates 3 factors: visibility, weather alerts, and currents. For example, good visibility normally contributes to broader horizon of watching. Higher weather alert information poses a significant threat to the navigation safety.

2.2 *Evaluation grades for risk of objective level*

Navigational risk is divided into 7 grades, ranging from very low to very high, described in Table 1 from the perspective of frequency and consequence of accidents. These 7 linguistic terms are represented by 7-member trapezoidal IT2FSs, LT_k, $k = \{1, 2, \ldots, 7\}$, shown in Table 2, which are adapted from Chen & Chen 2008, Liu 2011, and Wei & Chen 2009 by deleting absolutely low and absolutely high. The reason of ignoring these two absolute terms is due to that exactly defining the absolute situations in the navigational risk analysis is not only difficult from a systematic view, but also impractical from the reality.

2.3 *Evaluation grades for risk of factor level*

Factor risk is divided into 3 grades depending on factor attribute, ad hoc condition, and considering experts' views. 3 grades for factors with respect to 4 kinds of criteria are described in Table 3, Table 4, Table 5, and Table 6, respectively. These 3 grades are represented by 3-member trapezoidal IT2FSs—L, M, and H listed in Table 2. The reason of using this 3-member trapezoidal IT2FSs is for simplification in expert's evaluating the risk of factor for a specific vessel.

Table 1. Grades of navigational risk.

Linguistic term	Explanation of linguistic term
Very low (VL)	The risk is very low due to the frequency of accident or the consequence of accident is very low.
Low (L)	The risk is low due to the frequency of accident or the consequence of accident is low.
Fairly low (FL)	The risk is fairly low due to the frequency of accident or the consequence of accident is fairly low.
Medium (M)	The risk is medium due to the frequency of accident or the consequence of accident is medium.
Fairly high (FH)	The risk is fairly high due to the frequency of accident or the consequence of accident is fairly high.
High (H)	The risk is high due to the frequency of accident or the consequence of accident is high.
Very high (VH)	The risk is very high due to the frequency of accident or the consequence of accident is very high.

Table 2. 7-member linguistic terms and their corresponding trapezoidal IT2FSs.

k	Linguistic terms	LT_k, Trapezoidal IT2FSs
1	Very-low (VL)	[(0, 0, 0.02, 0.07; 1.0), (0, 0, 0.02, 0.07; 0.8)]
2	Low (L)	[(0.04, 0.1, 0.18, 0.23; 1.0), (0.04, 0.1, 0.18, 0.23; 0.8)]
3	Fairly low (FL)	[(0.17, 0.22, 0.36, 0.42; 1.0), (0.17, 0.22, 0.36, 0.42; 0.8)]
4	Medium (M)	[(0.32, 0.41, 0.58, 0.65; 1.0), (0.32, 0.41, 0.58, 0.65; 0.8)]
5	Fairly high (FH)	[(0.58, 0.63, 0.80, 0.86; 1.0), (0.58, 0.63, 0.80, 0.86; 0.8)]
6	High (H)	[(0.72, 0.78, 0.92, 0.97; 1.0), (0.72, 0.78, 0.92, 0.97; 0.8)]
7	Very high (VH)	[(0.93, 0.98, 1.0, 1.0; 1.0), (0.93, 0.98, 1.0, 1.0; 0.8)]

3 PROPOSED METHOD

The motivation behind the development of the proposed method is to improve the flexibility in the navigational risk analysis. Flexibility is reinforced due to IT2FSs better modeling and handling uncertainty of experts' judgements when the exact membership function of T1FSs is unknown. Four steps of the proposed method are as follows:

Table 3. Grades of risk of factor with respect to static information of vessel.

	Grade of risk		
Factor	L	M	H
Type		general cargo	dangerous cargo
Age	between 10 years and 15 years	other	larger than 25 years or unknown
Draught	largely less than channel depth	normally less than channel epth	slightly less than channel depth
Cargo	container	general cargo	dangerous cargo, unloaded, or overloaded
Shipowner	government	corporation	individual
Scale	largely matching the channel	normally matching the channel	slightly matching the channel

Table 4. Grades of risk of factor with respect to dynamic information of vessel.

	Grade of risk		
Factor	L	M	H
Speed	safety speed, or obey the local regulation	other	largely less or larger than safety speed, or seriously violate the local regulation
Traffic situation	no foreign vessels	one foreign vessels around own ships	multiple foreign vessels around own ships
Angle of turning	equal to or less than 15 degree per second	less than 30 degree but larger than 15 degree per second	equal to or larger than 30 degree per second
Illegal records	less than 2 times per year	2 times per year	larger than 2 times per year or unknown

Table 5. Grades of risk of factor with respect to information of navigational environment.

	Grade of risk		
Factor	L	M	H
Navigational Aids	perfect in quality and enough in quantity	other	deficiency in quality and quantity
Complexity of channel	other	one obstructions in the channel	multiple obstructions in the channel
Channel bending radius	largely bigger than ship length	slightly larger than ship length	slightly larger than ship length
Channel width	largely bigger than ship length	slightly larger than ship length	slightly larger than ship width
Density of traffic flow	sparse	normal	dense

Table 6. Grades of risk of factor with respect to information of natural environment.

	Grade of risk		
Factor	L	M	H
Weather alerts	fourth level weather alert or no weather alert	second or third level weather alert	first level weather alert
Visibility	good	normal	bad
Currents	advection or no current	other	turbulent current

Step 1: Determine the evaluation model
Step 2: Calculate weights of criteria and factors
Step 3: Aggregate the degree of risk
Step 4: Measure the degree of similarity

Step 1: Determine the evaluation model
A hierarchical structure model including objective, several criteria, and some factors, is established to analyze navigational risk in an area during a specified period, e.g. a three-level model shown in Figure 1.

Step 2: Calculate weights of criteria and factors
In the established model shown in Figure 1, weights of criteria with respect to the objective and weights of factor with respect to corresponding criteria are obtained by AHP due to its simplicity of implementation, although some limitations exist in AHP (Ivanco et al 2017). To this end, experts' judgements are elicited. For detailed information of steps of AHP, please refer to Bian et al. 2017, Deng et al. 2014, Zhang 2011, Zhou et al. 2017. Finally, the weight of a factor with respect to the objective, the global weight of such factor, is the product of the weight of such factor with respect to corresponding criteria and the weight of corresponding criteria with respect to the objective.

Step 3: Aggregate the degree of risk
Regarding a vessel i, $i \in$ N+, the degree of risk of factor j, $\tilde{Z}_{ij}$, $j = \{1, 2, \ldots, 18\}$, is evaluated using L, M, and H shown in Table 2 and according to the grades of risk defined in Table 3, Table 4, Table 5, and Table 6. For example, given that an unloaded 1000GT oil tanker is cruising in Wuhan section of the Yangtze River during dry season with the speed of 10kn and the visibility is over 1.5km, grade of some factors' risk are as follows: "Type" and "Cargo" are H according to Table 3; "Speed" is L according to Table 4; "Navigational Aids" is L, "Channel width" is M, "Channel bending radius" is L, and "Complexity of channel" is H according to Table 5; "Visibility" is M according to Table 6.

After 18 factors are evaluated, the degree of risk of specified vessel i, $\tilde{R}_i$, is obtained by aggregating global weights of factors determined in **Step 2** and evaluation of factors $\tilde{Z}_{ij}$ based on a trapezoidal interval type-2 weighted averaging (TIT2-WAA) operator (Hu et al. 2013). Note that, the aggregation result is also a trapezoidal IT2FS. For detailed information about this proof, please refer to Hu et al (2013). Therefore, it is necessary to further analyze that what the aggregation result stands for.

Step 4: Measure the degree of similarity
To transform $\tilde{R}_i$ in **Step 3** into corresponding linguistic term represented by LT_k, $k = \{1, 2, \ldots, 7\}$, shown in Table 2, the degree of similarity between $\tilde{R}_i$ and LT_k, $S(\tilde{R}_i, LT_k)$, $k = \{1, 2, \ldots, 7\}$, is calculated based on measure proposed by Chen & Sanguansat (2011).

According to Chen & Sanguansat (2011), the lager the value of $S(\tilde{R}_i, LT_k)$, the more the similarity between $\tilde{R}_i$ and LT_k.

4 CASE STUDY

The verification of the proposed method was done through retrospective analysis of 5 vessels cruised in the Yangtze River in 2010, 5 accidents among them are from CJMSA (2010). Table 7 shows that 5 accidents with different grades and types happened in question.

[Step 1]: A three-level model is created in Figure 1. In which, the objective is to evaluate the navigational risk of these 5 vessels cruised in the Yangtze River shown in Table 7. Four kinds of criteria and eighteen factors are derived from Tian et al. 2013, Zhang et al. 2011, Zhang et al. 2013, and Zhang et al. 2016.

[Step 2]: Using AHP, a series of pairwise comparisons with respect to the criteria and factors are performed by three-expert committee to calculate weights of criteria with respect to the objective, and weights of factors with respect to corresponding criteria.

Taking the calculation of weights of criteria with respect to the objective as an example, the pairwise comparison matrix A is as follows:

$$A = \begin{array}{c} \\ \text{Static information of vessel} \\ \text{Dynamic information of vessel} \\ \text{Information of navigational environment} \\ \text{Information of natural environment} \end{array} \begin{array}{c} \begin{array}{cccc} \text{Static information of vessel} & \text{Dynamic information of vessel} & \text{Information of navigational environment} & \text{Information of natural environment} \end{array} \\ \begin{pmatrix} 1 & 1/2 & 2/3 & 1/3 \\ 2 & 1 & 4/3 & 2/3 \\ 3/2 & 3/4 & 1 & 2 \\ 3 & 3/2 & 1/2 & 1 \end{pmatrix} \end{array}$$

The largest eigenvalue of matrix A is 4.2492. The normalized eigenvector belonging to the largest eigenvalue of matrix A is: w = (0.1338, 0.2677,

Table 7. Basic information of 5 vessels.

i	Vessel name	Time	Type of accident	Grade of accident
1	Laoxiahe 828	2010.01.01	Grounding	FH
2	Fufa 888	2010.01.04	Collision	H
3	Jinzhou 656	2010.01.14	Collision	VH
4	Yuhan 805	2010.02.10	Grounding	FH
5	Xinpingjiang 1013	2010.03.05	Grounding	M

Table 8. Weights of criteria and factors.

Criteria evel	j	Factor level	Local weight	Global weight
Static information of vessel (0.1338)	1	Type (VT)	0.1596	0.0214
	2	Age (VA)	0.0641	0.0086
	3	Draught (VD)	0.2504	0.0335
	4	Cargo (VC)	0.3825	0.0512
	5	Shipowner (SO)	0.0428	0.0057
	6	Scale (VSP)	0.1006	0.0135
Dynamic information of vessel (0.2677)	7	Speed (VSE)	0.2643	0.0708
	8	Traffic situation (TS)	0.5693	0.1524
	9	Angle of turning (AT)	0.0609	0.0163
	10	Illegal records (IR)	0.1055	0.0282
	11	Navigational aids (NA)	0.4263	0.1298
Information of navigational environment (0.3045)	12	Complexity of channel (CC)	0.1591	0.0484
	13	Channel bending radius (CB)	0.0823	0.0251
	14	Channel width (CW)	0.1732	0.0527
	15	Density of traffic flow (DT)	0.1591	0.0484
Information of natural environment (0.294)	16	Weather alerts (WA)	0.5556	0.1633
	17	Visibility (NV)	0.1111	0.0327
	18	Currents (NC)	0.3333	0.0980

0.3043, 0.2940)^T. Since the ratio of consistency index to random consistency index, 0.0923, is less than 0.1, the constructed pairwise comparison matrix A is considered acceptable (Zhou et al. 2017), and the results of normalized eigenvector are values of weights of criteria with respect to the objective. Similarity, the weights of factors with respect to certain criteria are obtained by AHP. After that, the weight of a factor with respect to objective is obtained by multiplying the weight of the factor with respect to certain criteria and the weight of the criteria with respect to the objective. The weights of criteria and factor are shown in Table 8.

[Step 3]: According to the retrospective analysis of accident reports (CJMSA 2010), factors in **[Step 2]** are evaluated by expert committee using 3-member linguistic terms–L, M, and H–shown in Table 4. And evaluations of factors with respect to 5 vessels are shown in Table 9.

Regarding to vessel i, the degree of risk, $\tilde{\tilde{R}}_i$, is obtained by aggregating global weights of factors tabulated in Table 8 and evaluation of factors $\tilde{\tilde{Z}}_{ij}$ shown in Table 9 using TIT2-WAA aggregation operator, which is shown in Table 10.

[Step 4]: Based on similarity measure proposed by Chen & Sanguansat (2011), results of the degree of similarity $S(\tilde{\tilde{R}}_i, LT_k)$ with respect to each vessel are shown in Figure 2. According to Chen & Sanguansat (2011), $\tilde{\tilde{R}}_i$ can be transformed in a linguistic term represented by LT_k which is more similar to it. Therefore, the risk of these 5 vessels are as follows: Laoxiahe 828, Medium; Fufa 888, Medium; Jinzhou 656, Medium; Yuhan 805, Medium; Xinpingjiang 1013, Medium.

Table 9. Evaluation of factors for each vessel.

Factor level	Lao-xiahe 828	Fu-fa 888	Jinzhou 656	Yu-han 805	Xin-ping-jiang 1013
VT	M	M	M	M	H
VA	H	M	M	M	L
VD	M	L	L	M	H
VC	H	H	H	M	H
SO	M	M	M	M	M
VSC	M	M	H	H	M
VSP	H	L	L	H	L
TS	L	H	H	L	L
AT	H	L	L	H	H
IR	H	H	H	H	H
NA	H	M	M	H	H
CC	H	H	H	H	H
CB	L	M	H	L	L
CW	H	L	M	H	M
DT	M	M	L	L	L
WA	M	M	L	L	L
NC	M	M	H	M	M
NV	M	M	M	M	L

Table 10. Aggregation of experts' judgement for each vessel.

Vessel name	$\tilde{\tilde{R}}_i$
Laoxiahe 828	[(0.4327, 0.5052, 0.6471, 0.7054; 1.0), (0.4327, 0.5052, 0.6471, 0.7054; 0.8)]
Fufa 888	[(0.3836, 0.4600, 0.6060, 0.6669; 1.0), (0.3836, 0.4600, 0.6060, 0.6669; 0.8)]
Jinzhou 656	[(0.3937, 0.4612, 0.5888, 0.6438; 1.0), (0.3937, 0.4612, 0.5888, 0.6438; 0.8)]
Yuhan 805	[(0.3549, 0.4224, 0.5466, 0.6016; 1.0), (0.3549, 0.4224, 0.5466, 0.6016; 0.8)]
Xinpingjiang 1013	[(0.3112, 0.3763, 0.4913, 0.5447; 1.0), (0.3112, 0.3763, 0.4913, 0.5447; 0.8)]

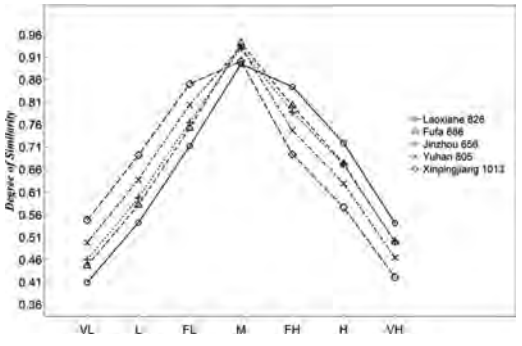


Figure 2. Results of similarity between aggregated experts' evaluation and 7 grades of navigational risk.

5 CONCLUSION

A novel method using IT2FSs for navigational risk analysis is presented in this paper. The main features of the proposed method are as follow: firstly, weights of risk factors are obtained using AHP; secondly, grades of the objective and factors are represented by IT2FSs; thirdly, degree of risk is determined by similarity measure.

Although the proposed model does not directly consider human factor, factors considered are evaluated only by 3-member linguistic terms for simplification, weights of criteria with respect to objective and weights of factor with respect to criteria are not constant but case by case, and the grades of risk are not absolutely equal to that happened in reality, 5 vessels as case study are used to illustrate the application of the proposed method.

In the future, we would specify measures to timely alleviate or effectively mitigate the identified risk, which is the third module of FSA.

ACKNOWLEDGEMENTS

The paper is financially supported by Nation-al Key Technologies Research & Development Program (2017YFC0804900, 2017YFC0804904), also partially supported by the National Science Foundation of China (NSFC) under grant No. 51579203, and No. 51711530033.

REFEREENCES

Akhtar, M. J. & I. B. Utne (2014). Human fatigue's effect on the risk of maritime groundings--A Bayesian Network modeling approach. *Safety Science. 62*, 427–440.

Allianz (2015). Safety and shipping review 2015, Allianz Global Corporate & Specialty, Munich, Germany.

Andrew, J., D. Paraskevadakis, A. Bury, Z. L. Yang, R. Riahi, & J. Wang (2014). An integrated fuzzy risk assessment for seaport operations. *Safety Science. 68*, 180–194.

Beşikçi, E. B., T. Kececi, O. Arslan, & O. Turan (2016). An application of fuzzy-AHP to ship operational energy efficiency measures. *Ocean Engineering. 121*, 392–402.

Bian, T., J. T., Hu, & Y. Deng (2017). Identifying influential nodes in complex networks based on AHP. *Physica A. 479*, 422–436.

Bozdag, E., U. Asan, A. Soyer, & S. Serdarasan (2015). Risk prioritization in Failure Mode and Effects Analysis using interval type-2 fuzzy sets. *Expert Systems with Applications. 42 (8)*, 4000–4015.

Brito, M. & G. Griffiths (2016). A Bayesian approach for predicting risk of autonomous underwater vehicle loss during their missions. *Reliability Engineering & System Safety. 146*, 55–67.

Celik, M., I. D. Er, & A. F. Ozok (2009). Application of fuzzy extended AHP methodology on shipping registry selection: the case of Turkish maritime industry. *Expert Systems with Applications. 36*, 190–198.

Chen, J. H., F. Lu, & G. J. Peng (2015). A quantitative approach for delineating principal fairways of ship passages through a strait. *Ocean Engineering. 103,* 188-197.

Chen, S. J. & S. M. Chen (2008). Fuzzy risk analysis based on measures of similarity between interval-valued fuzzy numbers. *Computers and Mathematics with Applications. 55*, 1670–1685.

Chen, S. M. & J. H. Chen (2009). Fuzzy risk analysis based on similarity measures between interval-valued fuzzy numbers and interval-valued fuzzy number arithmetic operators. *Expert Systems with Applications. 36*, 6309–6317.

Chen, S. M. & K. Sanguansat (2011). Analyzing fuzzy risk based on similarity measures between interval-valued fuzzy numbers. *Expert Systems with Applications. 38*, 8612–8621.

Chen, S. M. & L. W. Lee (2010). Fuzzy multiple attributes group decision-making based on the ranking values and the arithmetic operations of interval type-2 fuzzy sets. *Expert Systems with Applications. 37 (1)*, 824–833.

CJMSA (2010) Accident statistics of Changjiang Maritime Safety Administration.

Deng, X. Y., Y. Hu, Y. Deng, & S. Mahadevan (2014). Supplier selection using AHP methodology extended by D numbers. *Expert Systems with Applications. 41*, 156–167.

Deng, Y., F. T. S. Chan, Y. Wu, & D. Wang (2011). A new linguistic MCDM method based on multiple-criterion data fusion. *Expert Systems with Applications. 38*, 6985–6993.

Dereli, T. & K. Altun (2013). Technology evaluation through the use of interval type-2 fuzzy sets and systems. *Computers & Industrial Engineering. 65 (4)*, 624–633.

Fan, S. Q., X. P. Yan, J. F. Zhang, & D. Zhang (2017). A review on human factors in maritime accidents. *Journal of Transport Information and Safety. 2 (35)*, 1–8.

Gorzalczany, M. B. (1987). A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy Sets and Systems. 21 (1)*, 1–17.

Görçün, Ö. F. & S. Z. Burak (2015). Formal safety assessment for ship traffic in the Istanbul Straits. *Procedia-Social and Behavioral Sciences. 207*, 252–261.

Hollnagel, E. (2014). Is Safety a Subject for Science? *Safety Science. 67*, 21–24.

Hu, J., Y. Zhang, X. H. Chen, & Y. M. Liu (2013). Multi-criteria decision making method based on possibility degree of interval type-2 fuzzy number. *Knowledge-Based Systems. 43*, 21–29.

Hu, S. P., Q. G. Fang, H. B. Xia, & Y. T. Xia (2007). Formal safety assessment based on relative risks model in ship navigation. *Reliability Engineering & System Safety. 92 (3)*, 369–377.

Huang, Y. M., P.H.A.J.M. van Gelder, & M. B. Mendel (2017) Imminent ships collision risk assessment based on velocity obstacle. ESREL 2016: Risk, Reliability and Safety: Innovating Theory and Practice. Lesley Walls, Matthew Revie and T. Bedford. Glasgow (UK), Taylor & Francis Group: 693–700.

Hänninen, M. & P. Kujala (2012). Influences of variables on ship collision probability in a Bayesian belief network model. *Reliability Engineering & System Safety. 102*, 27–40.

Hänninen, M. (2014). Bayesian networks for maritime traffic accident prevention: benefits and challenges. *Accident Analysis and Prevention. 73,* 305–312.

IMO. 2015. Revised guidelines for formal safety assessment (FSA) for use in the IMO rule-making processes, MSC-MEPC.2/Circ.12/Rev.1, London.

Ivanco, M., G. Hou, & J. Michaeli (2017). Sensitivity analysis method to address user disparities in the analytic hierarchy process. *Expert Systems With Applications. 90*, 111–126.

Jin, D. (2014). The determinants of fishing vessel accident severity. *Accident Analysis and Prevention. 66*, 1–7.

Karahalios, H. (2014). The contribution of risk management in ship management: the case of ship collision. *Safety Science. 63*, 104–114.

Kujala, P., M. Hänninen, T. Arola, & J. Ylitalo (2009). Analysis of the marine traffic safety in the Gulf of Finland. *Reliability Engineering & System Safety. 94 (8)*, 1349–1357.

Kum, S. & B. Sahin (2015). A root cause analysis for Arctic marine accidents from 1993 to 2011. *Safety Science. 74*, 206–220.

Li, B. & F. W. Pang (2013). An approach of vessel collision risk assessment based on the D–S evidence theory. *Ocean Engineering. 74*, 16–21.

Liu, H. L., Z. H. Tian, A. Q. Huang, & Z. L. Yang (2017). Analysis of vulnerabilities in maritime supply chains. *Reliability Engineering & System Safety. 169*, 475–484.

Lušić, Z. & M. Čorić (2015). Models for estimating the potential number of ship collisions. *Journal of Navigation. 68 (4)*, 735–749.

Martins, M. R. & M. C. Maturana (2013). Application of Bayesian belief networks to the human reliability analysis of an oil tanker operation focusing on collision accidents. *Reliability Engineering & System Safety. 110*, 89–109.

Mendel, J. M. (2001). Uncertain rule-based fuzzy logic systems: introduction and new directions. Prentice-Hall, Upper Saddle River, NJ.

Mendel, J. M. (2007). Advances in type-2 fuzzy sets and systems. *Information Sciences. 177 (1)*, 84–110.

Meng, Q., J. X. Weng, & S. Y. Li (2014). Analysis with Automatic Identification System data of vessel traffic characteristics in the Singapore Strait. *Transportation Research Record: Journal of the Transportation Research Board, 2426*, 33–43.

Montewka, J., F. Goerlandt, & P. Kujala (2014). On a systematic perspective on risk for formal safety assessment (FSA). *Reliability Engineering and System Safety. 127*, 77–85.

Montewka, J., T. Hinz, P. Kujala, & J. Matusiak (2010). Probability modelling of vessel collisions. *Reliability Engineering & System Safety, 95 (5)*, 573–589.

Pak, J. Y., G. T. Yeo, S. W. Oh, & Z. L. Yang (2015). Port safety evaluation from a captain's perspective: the Korean experience. *Safety Science. 72*, 172–181.

Qu, X., Q. Meng, & S. Y. Li (2011). Ship collision risk assessment for the Singapore Strait. *Accident Analysis and Prevention. 43 (6)*, 2030–2036.

Rahib, H. A., K. Uyar, U. Ilhan, & E. Imanov (2016). Assessment of food security risk level using type 2 fuzzy system. *Procedia Computer Science. 102*, 547–554.

Saaty, T. L. (1980). The Analytic Hierarchy Process: planning, priority setting, resources allocation. New York: McGraw-Hill Inc.

Sen, S., K. Patra, & S. K. Mondal (2016). Fuzzy risk analysis in familial breast cancer using a similarity measure of interval-valued fuzzy numbers. *Pacific Science Review A: Natural Science and Engineering. 18*, 203–221.

Talavera, A., R. Aguasca, B. Galván, & A. Cacereño (2013). Application of Dempster-Shafer theory for the quantification and propagation of the uncertainty caused by the use of AIS data. *Reliability Engineering & System Safety. 111*, 95–105.

Thieme, C. A. & I. B. Utne (2017). Safety performance monitoring of autonomous marine systems. *Reliability Engineering and System Safety. 159*, 264–275.

Tian, L. J., S. Y. Zhang, & N. Li (2013). Real-time early-warning index system for water transportation safety in Yangtze River. *Journal of Transport Information and Safety. 2 (31)*, 69–73.

Türkşen, I. B. (2002). Type 2 representation and reasoning for CWW. *Fuzzy Sets and Systems*, *127 (1)*, 17–36.

Wang, J. (2000). A subjective modelling tool applied to formal ship safety assessment. *Ocean Engineering. 27*, 1019–1035.

Wang, N., X. Y. Meng, Q. Y. Xu, & Z. W. Wang (2009). A unified analytical framework for ship domains. *Journal of Navigation. 62*, 643–655.

Wang, Y., J. F. Zhang, X. Q. Chen, X. M. Chu, & X. P. Yan (2013). A spatial-temporal forensic analysis for inland-water ship collisions using AIS data. *Safety Science. 57*, 187–202.

Wei, S. H. & S. M. Chen (2009). Fuzzy risk analysis based on interval-valued fuzzy numbers. *Expert Systems with Applications. 36 (2)*, 2285–2299.

Xu, X., X. Geng, & Y. Wen (2016). Modeling of ship collision risk index based on complex plane and its realization. TransNav, the International *Journal on Marine Navigation and Safety of Sea Transportation. 10 (2)*, 251–256.

Yang, Z. L., A. K. Y. Ng, & J. Wang (2014). A new risk quantification approach in port facility security assessment. *Transportation Research Part A: Policy and Practice. 59*, 72–90.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control. 8*, 338–356.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning—I. *Information Science. 8 (3)*, 199–249.

Zhang, D., X. P. Yan, J. F. Zhang, Z. L. Yang, & J. Wang (2016). Use of fuzzy rule-based evidential reasoning approach in the navigational risk assessment of inland waterway transportation systems. *Safety Science. 82*, 352–360.

Zhang, D., X. P. Yan, Z. L. Yang, & J. Wang (2011). Application of formal safety assessment to navigational risk evaluation of Yangtze River. In Proceedings of the ASME 2011, 30th International Conference on Ocean, Offshore and Arctic Engineering OMAE2011-50186, Rotterdam, the Netherlands.

Zhang, D., X. P. Yan, Z. L. Yang, A. Wall, & J. Wang (2013). Incorporation of formal safety assessment and Bayesian network in navigational risk estimation of the Yangtze River. *Reliability Engineering & System Safety. 118*, 93–105.

Zhao. J. S., Z. L. Wu, F. C. Wang (2009). A basic study on synthetic judgement of shipping safety. *Journal of Navigation, 45 (02)*, 300–303.

Zhou, X. Y., X. Y. Deng, Y. Deng, & S. Mahadevan (2017). Dependence assessment in human reliability analysis based on D numbers and AHP. *Nuclear Engineering and Design*. *313*, 243–252.

Øien, K (2001). Risk indicators as a tool for risk control. *Reliability Engineering and System Safety. 74*, 129–45.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Alternative life-loss rates for failures of large concrete and masonry dams in mountain regions of OECD countries

A. Kalinina, M. Spada & P. Burgherr
Laboratory for Energy Systems Analysis, Paul Scherrer Institute, Switzerland

ABSTRACT: High safety standards for dams in Switzerland require continuous assessment of their structural stability and effectiveness of warning system. Therefore, risk assessment needs to be performed for estimation of consequences, e.g. Life Loss (LL), in case of a dam accident. With the life-loss estimation methods available nowadays, e.g. the LIFESim system, physical processes within the specific dam-failure event can be simulated. This study demonstrated the importance of adjusting the LL rates in LIFESim to reflect study-specific characteristics of the dam type and failure mode. In particular, for application to Swiss dams, alternative LL rate distributions were built based on historical events of concrete and masonry dams in the mountain regions of OECD countries. The alternative LL rates distributions had different shapes and frequency ranges than those recommended in LIFESim. A simulation example of a hypothetical dam failure showed that the alternative and recommended LL rates lead to different LL estimates.

1 INTRODUCTION

Switzerland is the country with the highest density of dams in the world. Most of the dams were built to generate hydro-electricity (90% of all dams), but they also play an important role as flood control facilities. Swiss dams are constructed and operated under high safety standards. Historically no failures of Swiss dams have occurred; where a dam failure can be defined as a collapse or movement of part of a dam or its foundation leading to a disability of the dam to retain water (ICOLD, 2016). However, the ageing of many facilities and the increasingly stricter safety standards require that dam engineers and operators need to regularly assess and update potential risks associated with dams. An example is the assessment of consequences in terms of Life Loss (LL) due to a hypothetical dam failure. The results of such a risk assessment can help to justify, for example, costly facility upgrades or investments in the warning system required for risk mitigation.

To provide a transparent and quantifiable way to determine consequences, e.g. LL, associated with dam failures and subsequent floods, various methods and models are available nowadays.

1.1 *Methods for life-loss estimation*

Methods for LL estimation differ in complexity and modeling principles. Most of the approaches are purely empirical and LL estimates are based on regressions of Population At Risk (PAR) as a function of the whole downstream population and heterogeneous Warning time (Wt) (e.g. Lee et al., 1986; Brown and Graham, 1988). Another approach by Graham (1999) is more sophisticated and provides LL rates for a mix of subgroups of PAR based on Wt, flood severity, and warning effectiveness.

However, empirical approaches have limitations, which can be summarized as following (McClelland and Bowles, 2002). Firstly, empirical methods do not differentiate between characteristics of the dam failure (e.g. breach propagation or instantaneous failure) and flood severity, which might lead to the underestimation of the impact of the flood. Secondly, information on PAR, building structures, flow quantities, etc. represent averages and is not site specific. This strongly affects the LL results, because they depend on the gender and age distribution of PAR (Salvati et al., 2018). Thirdly, evacuation is not modelled and Wt is considered as a single value, which affects the number of people that are exposed to the flood and in turn the LL estimates.

To overcome these limitations in LL estimation, it is necessary to simulate physical processes and interactions, which cannot be achieved with empirical methods only. For this purpose, complex Geographical Information Systems (GIS)-based models have been developed in recent years allowing the dynamic simulation of flood consequences, including estimation of LL due to a dam failure. In contrast to empirical methods, these models estimate LL using modules with databases about evacuation, warning time, and loss of shelter to consider site-specific conditions. The most known

models are the Life Safety Model (LSM) (British Columbia, 2006) and the HEC-LifeSim model (USACE, 2017a). LSM is an agent-based model requiring detailed information for simulation, and thus it is more suitable for studies that model the behavior of the individual receptor (i.e. microscale simulation of the impact on a person or vehicle). In contrast, LIFESim scales up the simulation from the microscale to the mesoscale, i.e. simulation for the zone; therefore, it is more suitable for LL estimation for the specific area downstream of the dam.

1.2 *Study goals*

As discussed, the application of physically-based models in dam risk assessments could lead to better LL estimation. However, available models were developed by U.S. institutions that used LL rates empirically derived from historical data on flood events that mostly happened in the USA (see Section 2.2). The direct application of these LL rates to Switzerland would provide a potentially questionable approximation, since it has been shown that accident frequency and severity strongly depend on dam characteristics and location (Kalinina et al., 2017).

Therefore, this study aims to develop LL-rate distributions that can be considered representative for the topographical conditions and characteristics of dams in Switzerland. These LL rate distributions are used in modular LL estimation models to indicate proportion of life loss, P, for different flood zones and the corresponding relative frequency of exceeding this rate. To build these distributions historical failures of concrete and masonry dams in mountain regions of OECD countries were used, which can also be considered representative for Switzerland. Furthermore, the calculated LL rate distributions were used as input to the HEC-LIFESim model to simulate the LL resulting from an instantaneous dam failure. In this way it was possible to demonstrate the robustness of the physically-based model and the sensitivity of LIFESim simulation results to different LL rates. Therefore, this study provides quantitative insights on the recommendation of McClelland and Bowles (2002) that the historical observations underlying the method for LL estimation should be adjusted according to the type of event that is likely for a particular study setting.

2 METHOD

In this section, the spatial modular software HEC-LIFESim (USACE, 2017a) is introduced. Then, the motivation for developing alternative LL rates to be applied for failures of large concrete and masonry dams in mountain areas is explained and the methodology for constructing the alternative LL-rate distributions is given. Finally, a simulation example, that will be built to demonstrate the effect of different LL rates on LL estimates, is described.

2.1 *HEC-LIFESim software*

HEC-LIFESim (or LIFESim throughout this article) is a software developed by the Hydrologic Engineering Center (HEC) of the U.S. Army Corps of Engineers (USACE, 2017a). LIFESim is a spatial dynamic system for modeling life loss or economic consequences of a natural, dam or dike flood event. LIFESim is a modular system consisting of four modules built around databases. The modules exchange data through a geo-database with various information layers and tables, as shown in Figure 1.

Within LIFESim the aforementioned data can be combined with other GIS layers such as ESRI maps (ESRI, 2017), which allows for simulations that can represent real world conditions in a more realistic and accurate manner. In other words, LIFESim can overcome some of the limitations of the purely empirical approaches for modeling LL of a dam failure.

The four modules of the LIFESim system are represented as blocks in Figure 1: 1) flood routing module, 2) loss of shelter module, 3) warning and evacuation module, and 4) loss of life module. The flood routing module of LIFESim interfaces with an existing flood routing model (e.g. HEC-RAS 5.0.3 (USACE, 2017c)) and using hydraulic and

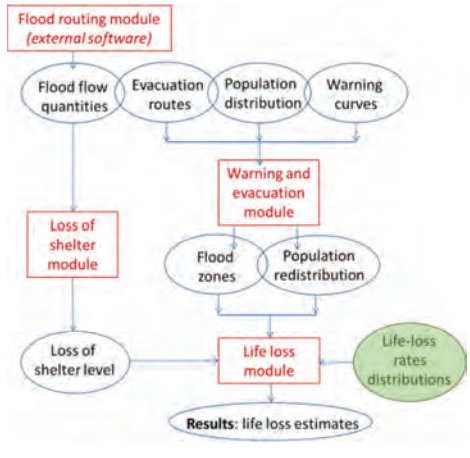


Figure 1. Simplified representation of the LIFESim approach for LL estimation (modified from Bowles, 2007).

timing editor it imports time series of flow quantities from hydraulic data source, e.g. depth time series at different points of the inundated area.

The loss-of-shelter module simulates the exposure of people in buildings during the flood event. For this, different flood zones are assigned to buildings and levels of buildings in the inundated area. In each flood zone, the physical flood environment is different, which is reflected in different historical rates of life loss. The three flood zones are defined by McClelland and Bowles (2000) by the interplay between available shelter and local flood depths and velocities, and can be summarized as follows:

- Chance zones in which flood victims are typically swept downstream or trapped underwater, and survival depends largely on chance;
- Compromised zones in which the available shelter has been severely damaged by the flood, increasing the exposure of flood victims to violent floodwaters.
- Safe zones are typically dry, exposed to relatively quiescent floodwaters, or exposed to shallow flooding unlikely to sweep people off their feet.

As input for the loss of shelter module, LIFESim utilizes the datasets of flow quantities in the simulation domain and the structure inventories obtained, for example, from HAZUS MH data (Federal Emergency Management Agency, 2003). Stability criteria for structures are set by default in LIFESim and can be changed for the specific study.

The warning and evacuation module simulates the spatial distribution of the population at risk from its initial distribution at the time when the warning is issued, to a new distribution with assigned flood zones when the flood arrives. For this module, the following information is required: GIS information on road layout, for example, from Highway Capacity Manual (TRB, 2000), information about population, for example, from HAZUS MH data (Federal Emergency Management Agency, 2003); evacuation destinations and emergency planning zones, which are location-specific and available as shape-files. Other evacuation parameters are set by default in LIFESim and can be changed for the specific study.

Finally, the loss-of-life module determines LL using the results of the aforementioned three modules. Based on the assigned flood zone categories (the loss-of-shelter module) and the value of PAR in this category (defined by the interplay between the flood map and the building inventory data), life-loss estimates are assessed using LL-rate distributions (McClelland and Bowles, 2002). For the simulations in this study, the recommended distributions were changed to alternative ones; this is indicated in Figure 1 in green and explained further in the text.

2.2 *LL rates distributions recommended in LIFESim*

For the estimation of LL, LIFESim recommends LL rates distributions developed by McClelland and Bowles (2002). To calculate the rates, the total PAR was determined and further divided into subgroups (subPAR), which help to customize the model to local conditions and to have homogeneous data for distinct areas. Three flood zones were then identified for each subPAR using the information about warning and flood severity. Finally, by estimating the ratio between the number of fatalities and the number of people in the particular flood zone of the particular subPAR, the P value was calculated for each case. Using the calculated P, the LL rates distributions were built for each flood zone (recommended distributions in Figure 3).

To construct the LL rates distributions, 38 unique flood events with 179 associated subPARs were used (Table 1). These events can be classified in three types: natural floods, floods due to a dam failure and floods due to a dike failure. Flood events due to a dam failure can be further classified in subgroups based on the dam type, among which the subgroup of floods resulting from failures of embankment dams is the largest. A similar prevalence of embankment dam failures was also found in datasets used to empirically estimate the dam-failure outflow in previous studies (Froehlich, 1995; Costa, 1985). In both cases, this can be explained by the fact that these dams commonly failed gradually; thus, data on characteristics (flow quantities, people) could be recorded. In contrast, instantaneous failures common for concrete dams give no chance to record detailed data, resulting in 8 events with 26 subPAR in Table 1.

Table 1. Flood events used by McClelland and Bowles (2002).

	Number of		
Type of event	Event	subPAR	Topography
Flood (river, flash, alluvial fan)	10	25	–
Dike	1	1	–
Dam:	27	153	–
– Embankment	16	121	–
– Buttress	1	1	mountain
– Gravity	2	8	mountain
	4	11	open/relatively flat area
– Arch	1	6	mountain
– Tailing	2	3	–
– Mill	1	3	–
Total	38	179	–

Furthermore, since the failure mode is different between embankment and concrete dams, it affects the nature of the floods and the subsequent impact on people. Thus, application of the LL rates derived on the data, which is highly dominated by embankment dam failures, can potentially bias LL estimates in studies of concrete and masonry dams (e.g. large dams in Switzerland). Therefore, the LL rates distributions need to be adjusted to reflect study-specific characteristics such as the dam type and failure mode.

2.3 *Alternative LL rates distributions*

To construct alternative LL rates distributions that are representative for large dams in Switzerland, a specific dam failure data set of large concrete and masonry dams in mountain regions of OECD countries was compiled. For this purpose, the historical experience contained in PSI's Energy-related Severe Accidents Database (ENSAD) was searched for relevant events.

The ENSAD database was developed at the Paul Scherrer Institute (PSI) in the 1990s (Hirschberg et al., 1998). Its goal is to enhance the comparative evaluation of different energy systems covering human health, environmental and economic impacts. ENSAD covers a broad range of full energy chains and continuous data collection ensures up-to-date information. Data from ENSAD are used for comparative risk assessment of energy technologies, to detect weak points in the energy infrastructure, and ultimately to support decision-making processes concerning energy supply options. For a detailed overview on ENSAD and its applications see Burgherr and Hirschberg (2014) or Burgherr et al. (2017).

The ENSAD comprises a worldwide dataset of more than 1,000 historical dam accidents in the period 1798–2017, 70% of which were in OECD countries. Each accident record has a set of characteristics, which together provide an exhaustive description of the event. Categories for some characteristics, e.g. dam type, dimensions, are adopted from other databases (ICOLD, 1995), while for others they are created to meet specific needs of ENSAD.

For this study, the ENSAD hydropower dam section was queried for dam failures that meet the following criteria: dams made of concrete and masonry; dams located in mountain areas (to be representative for the Swiss topography); dams in OECD countries (to ensure similar levels of safety as in Switzerland, see Hirschberg et al. (1998)). Consequently, calculated LL rates based on dam failure data fulfilling the above criteria can be considered a reasonable approximation for Switzerland.

For each dam failure, data about the total number of fatalities (i.e. life loss) and the population in the downstream area were searched. The latter is indicated in ENSAD only as the name of the town nearest to the dam. However, for this study, relatively homogeneous areas, subPAR, had to be defined, i.e. the total PAR had to be subdivided in areas that are different in terms of the flood severity, warning time, or flood severity understanding. Therefore, additional information on downstream population was collected from local reports, interviews, newspapers, etc. to be able to define homogenous subPAR. For example, for the Gleno dam failure, a total of 356–500 fatalities were reported by different sources; however, the total number of people affected, and the number of fatalities was certainly known only for the Bueggio Village. The availability of the information for this village defined the decision to treat this village as one homogeneous subPAR. Finally, the life-loss rate, P, was determined as the ration of LL to the population in the subPAR. For some subPAR, P was defined based on key words. For example, if the town was "washed out", than P of 0.99 was assumed with the confidence interval between 0.9 (people can survive even in a washout of the support surface) and 1 (resulted substantial destruction can lead to the complete population loss).

Furthermore, for each subPAR, the flood zone was defined based on warning time and flood severity. For this study, only one flood zone was assigned to each subPAR.

2.4 *Simulation example*

The simulation example was created to show that different LL rates defined in the model lead to different LL estimates. Therefore, to demonstrate the relevance of using LL rates that reflect dam and failure characteristic specifics for the study, e.g. studies for Swiss dams.

For the simulation, the failure of a hypothetical dam was assumed. The simulation example was built using the well-documented example project provided for the LIFESim program (USACE, 2017b). This project is not fully representative for the Swiss-dam study; however, certain data reflect the Swiss conditions. In particular, the dam-failure outflow hydrograph can be considered representative, since it reflects the instantaneous failure mode common for failures of concrete and masonry dams. For the instantaneous failure, no advance warning was initiated, i.e. dam failure warning was not issued prior to the dam failure. The area downstream of the dam is flat and open (see Figure 2), which corresponds to a hypothetical town at the end of a Swiss valley. Data for the structural inventories, emergency planning zones, road networks

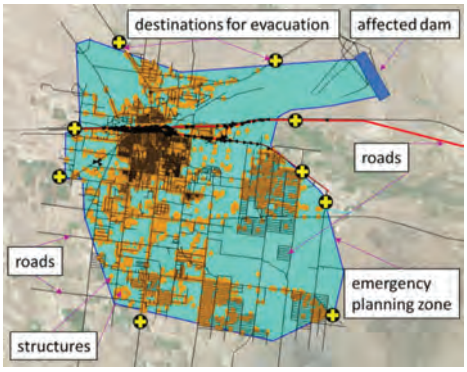


Figure 2. A map of the area flooded in the dam-failure event and symbols used for components of the LIFESim simulation (modified from USACE, 2017b).

and evacuation destinations were provided in the example project and originally taken from the sources mentioned in Section 2.1. The data for flow quantities were simulated in the HEC-RAS program (USACE, 2017d) and provided by USACE (2017b) as two-dimensional map.

Three simulations were created for this study. The first simulation corresponds to the application of the recommended LL rates by McClelland and Bowles (2002) (Section 2.2). The second simulation was run with the same model set-up except that the alternative LL rates were used (Section 2.3). For the third simulation, all downstream subPAR was assumed as a chance zone. This assumption was based on the methodology for determining PAR by the Swiss Federal Office of Energy (2017), whose procedure is based on PAR comprising the entire area affected by a dam-failure flood wave of at least 2 m height and intensity of at least 2 m2/s in a period of 2 h after a complete failure. Furthermore, this assumption was supported by McClelland and Bowles (2002) suggesting that when the buildings are about 6 meters high (i.e. one or two-story dwellings that are common for Swiss towns), then the entire PAR (alternatively, subPAR) is considered as a chance zone.

For all three simulations the simulation were done for several times of the day, namely at 12 p.m. (midnight), 6 a.m., 12 a.m. (noon), 6 p.m., and using 100 model runs for each combination to ensure convergence of results.

3 RESULTS

3.1 *Alternative LL rates distributions*

For the LL rates analysis, a dataset of 14 failures of concrete and masonry dams (buttress, gravity, and arch) located in mountain regions of OECD countries was established (Table 2). Four events from Table 1 were also considered, namely the Zerbino, St. Francis, Vajont, and Vega de Terra dams.

In the established dataset, dam name, country, and the year of the accident are given. The fatalities are indicated as the total number of fatalities resulting from the dam failure and as the number of fatalities in the defined subPAR downstream of the dam. The population at risk is also given as the total number for the entire downstream area and as the number of people for the defined subPAR. All information sources are given in Table 2. Finally, the flood zones were assigned to the calculated P values using information about warning time and flood severity (Table 2). Assigned flood zones are specified for each subPAR as letter indices (Table 2).

For each flood zone, the alternative LL rates distribution was constructed using the corresponding calculated P values (Figure 3). In addition, the confidence intervals were calculated for both alternative distributions using all the P values calculated based on different numbers of fatalities and people at risk found in the literature. Due to the lack of data for the safe zone among the failures in the new dataset, no fatalities were assumed in the safe zone.

For the alternative distribution of the chance zone, the range of P values is similar to the recommended P, i.e. between 1 and 0.4, with values of the alternative distribution potentially reaching 0.28 within its confidence interval (Figure 3). The recommended curve was built on a dataset with more realizations of high P with respect to the alternative distribution proposed in this study. This is shown for high P values (e.g. 0.8), which have a 0.9 and 0.7 frequency of exceedance for the recommended and alternative distributions, respectively. Furthermore, by considering the confidence intervals built for the alternative curve, the variability of possible P values is large for some P due to the limited data; thus, for the alternative distributions of the chance zone high P values can also increase to a frequency of 0.8.

For the compromised zone, the range for the alternative LL rates is smaller than for the recommended rates and the highest P does not exceed 0.13. However, taking into account the confidence intervals, realizations of 0.5 for P are also possible, which goes in line with the rates provided by McClelland and Bowles (2002).

3.2 *Simulation results*

Results for all three simulations are presented in Figure 4 and expressed in life loss as percentage of PAR (alternatively subPAR). The results are summarized using box plots showing the median

Table 2. Extended dataset for failures of concrete and masonry dams located in mountain regions of OECD countries.

N	Year	Dam name (country) & subPAR name	Fatalities in subPAR	Population in subPAR	Proportion of Life Loss, P	Warning	Flood Severity
1	1923	Gleno (Italy)	356–500	12,631	–		
		· Bueggio Village (Bureau of Reclamation, 2015)	209	500	0.42[a]	no	high
2	1934	Granadillar (Spain) (Gonzalez and Santamarta, 2012)	8	–	0.99[a]		
3	1928	Komoro (Japan)	7	–	–		
4	2012	Kopru (Turkey)	10 (Boston.com, 2012)	300 (Haberturk, 2012)	0.0333[b]		
5	1891	Lynx Creek (USA)	0	–	0		
6	1959	Malpasset (France) (Graham, 1999)	400–550	6000	–	no	
		· 100 ft flood depth	30	30	1[a]		high
		· 10 ft high entering Frejus	391	6000	0.0652[b]		medium
7	1925	Moyie River (USA)	0	–	0		
		Puentes (Spain)	680	–			
8	1802	· city of Lorca	608 (Saxena and Sharma, 2004)	4590 (Smedley et al., 1845; Murcia Today)	0.132[b]	some	
9	1928	St Francis (USA)	300–684	2250	–	some	
		· powerhouse N2 (McClelland and Bowles, 2002)	81		0.99[a]	no	high
		· Castaic Junction	washed away (Wikipedia, 2017b)		0.99[a]		
		· Edison tent camp at Kemp (Rogers and James, –)	89	140	0.636[a]	no	high
10	1965	Torrejon-Tajo (Spain)	39 (Wikipedia, 2017a)	50 (Extremadura, 2016)	0.6[a]		
11	1963	Vajont (Italy) (McClelland and Bowles, 2002)	1600–2600	3000	–	no	
		· Longarone town	1269	1348	0.941[a]		
		· lakeside communities	158		1[a]		
12	1959	Vega de Tera (Spain) (Graham, 1999)	140–153	415	0.347[a]	no	high
13	1944	Xuriguera (Spain)	7	–	–		
		· farm house (La Vanguardia, 1944)	6	6	1[a]		
14	1935	Zerbino (Italy)	130		–		

[a] – chance zone; [b] – compromised zone.

of the modeled distributions and the bottom and top edges of the box indicating the 25th and 75th percentiles, respectively. The whiskers extend to the maximum and minimum of the data not considering outliers, and the outliers (i.e. points distant by twice or three times the standard deviation (Ruan et al., 2005)) are plotted individually.

For simulations 2 and 3, the results were calculated using the alternative distributions shown in Figure 3 (solid lines). The confidence intervals of these distributions could not be taken into account due to specifics of the software settings.

The relative patterns of the results between different times of the day are similar across all three simulations. In particular, the highest values of P were calculated at 12 p.m. and the lowest at 12 a.m. For the former, this could be explained by the fact that most people are asleep and not aware

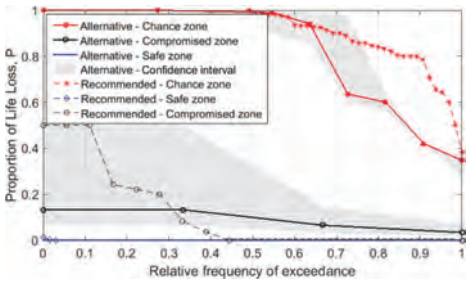


Figure 3. Historical LL rates distributions developed by McClelland and Bowles (2002) and LL rates distributions developed specifically for concrete & masonry dams in mountain regions of OECD countries with the confidence intervals (minimal and maximal values).

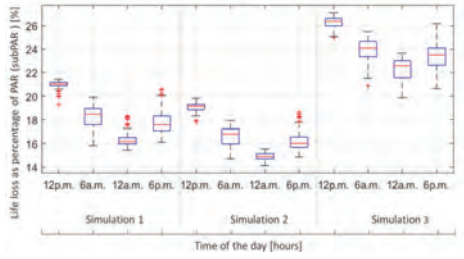


Figure 4. Simulation results for Life Loss as percentage of PAR: Simulation 1) using recommended life-loss rates by McClelland and Bowles (2002); Simulation 2) life-loss rates developed for dams in mountain areas; Simulation 3) test case with merged chance and compromised zone and safe zone without LL.

of a possible warning. On the other hand, at 12 a.m. most people are awake and at their duties; they can react faster to possible warning. The uncertainty range is higher for 6 a.m. and 6 p.m. results, because potentially at that time of the day people are traveling from home or back home; and it is quite uncertain how many people in traffic are exposed to the flood.

Comparing results between the first two simulations, the P values calculated in the second simulation are in general 10% lower than those calculated in the first simulation. This is due to the fact that, in the chance zone, the same P values have lower probabilities in the alternative distribution than in the recommended one, and in the compromised zone, the possible range of P values is in general lower (see Figure 3). The median values of P in the second simulation are shifted to lower values of the results range (i.e. distributions are left skewed) with respect to the first simulation. This can be explained by the fact that more severe P values in the chance and compromised zones in Figure 3 have lower frequency of exceedance (e.g. for a frequency of exceedance of 0.68, P is equal to 0.8 and 0.9 in the chance zone for the alternative and recommended distributions, respectively). On the other hand, probabilities for lower values of P (i.e. closer to 0.4) do not differentiate to such a high extent.

Furthermore, uncertainty in the results of the second simulation is generally lower than in the first simulation, unless in the case of the simulation at 12 p.m. Generally lower uncertainty ranges can be explained by the fact that the alternative distributions are flatter (especially the one of the compromised zone); then, sampling from the P values in the alternative distribution could potentially result in the sample with smaller range of P values.

Finally, comparing the results of the third simulation, it can be concluded that in general the defined P values are the highest among all simulations. Furthermore, uncertainty of the calculated values is significantly higher than in the results of the first two simulations. Moreover, the results show larger left skewness with respect to the other simulations results, because compromised zones became now chance zones and for all ranges of probabilities LL rates were higher. Thus, neglecting existence of compromised zones with their higher potential of people to survive, leads to potential overestimation of the overall life loss.

In general, lower rates for life losses reflected in the alternative distributions and results of the simulations are supported by the following circumstances. On the one hand, the alternative dataset of concrete and masonry dams is built exclusively for OECD countries, whereas the recommended LL rates include also events in non-OECD countries (e.g., China). Generally, higher population density downstream of non-OECD dams and differences in safety culture and awareness of people living downstream of non-OECD dams, could potentially result in higher life loss. On the other hand, mountain topography could leave higher survival chances for people located on higher altitudes in the downstream area. In contrast, in embankment dam failures, a flat area is commonly affected downstream of the dam; which makes the available shelter to be very remote.

4 CONCLUSIONS

To maintain or further improve the high safety levels of dams in Switzerland it is important to evaluate the performance of existing or planned risk mitigation measures at dams, and to estimate potential LL consequences of selected dam failure scenarios.

Generally, the LL estimation in dam risk assessment is a complex process, depending on many

parameters and circumstances. Available dynamic modular systems for LL estimation (e.g. LIFESim) can overcome known limitations of the purely empirical methods, because they allow modeling of physical processes within the dam-failure event, e.g. evacuation, population distribution. These models can also go beyond LL estimates and, for example, define times of the day with the highest risk for PAR, as it was demonstrated in the current study.

Furthermore, this study demonstrated the importance of adjusting the LL rates distributions to reflect study-specific characteristics such as the dam type, failure mode, etc. The LL rates derived from historical failures of concrete and masonry dams in mountain regions of OECD countries had different shapes and frequency ranges than the generic ones in LIFESim. The LL estimates calculated using the recommended and alternative LL rates gave different LL estimates in the simulation example of the hypothetical dam failure carried out in this study.

In summary, the importance of defining study-specific alternative LL rates distributions for dam risk assessment was demonstrated. Potential future extensions rely on the reduction of the width of the confidence intervals for the alternative LL-rate distributions considered representative for large concrete dams in Switzerland. To reduce the rather large uncertainty ranges, continuous update of information is suggested. In particular, more dam accidents need to be included and better information on subPAR, flood severity and warning availability to be provided for the events in the existing list of historical failures of concrete and masonry dams. Finally, the concept will be improved and implemented in more details to a real case study reflecting Swiss conditions.

ACKNOWLEDGMENT

This research project is part of the National Research Programme "Energy Turnaround" (NRP 70) of the Swiss National Science Foundation (SNSF). Further information on the National Research Programme can be found at www.nrp70.ch. It is also integrated within the activities of the Swiss Competence Center on Energy Research—Supply of Electricity (SCCER SoE). The authors express their sincere thanks to Dr. Christopher Robinson, who provided valuable feedback to this work.

REFERENCES

Boston.com. 2012. *Official: 10 missing in dam flooding in Turkey*. http://articles.boston.com/2012–02–24/news/31096589_1_dam-burst-hatch-rains.

Bowles, D.S. 2007. Life Loss Estimation for RAMCAP, Appendix D,. *Conventional Dams and Navigation Locks, Sector-Specific Guidance (SSG), Risk Analysis and Management for Critical Asset Protection (RAMCAP) Phase III for Dams, Locks and Levees.*

British Columbia, H. 2006. *Life Safety Model System V1.0, Guidelines, Procedures, Calibration and Support Manual.,* Rep. No. Technical Report Engineering Report E310: British Columbia Hydro.

Brown, C.A. & Graham, W.J. 1988. Assessing the threat to life from dam failure. *JAWRA Journal of the American Water Resources Association,* 24: 1303–1309.

Bureau of Reclamation 2015. *Reclamation Consequence Estimating Methodology: Dam Failure and Flood Event Case History Compilation* U.S. Department of the Interior.

Burgherr, P. & Hirschberg, S. 2014. Comparative risk assessment of severe accidents in the energy sector. *Energy Policy,* 74: S45-S56.

Burgherr, P., Spada, M., Kalinina, A., Hirschberg, S., Kim., W., Gasser, P. & Lustenberger, P. 2017. The Energy-related Severe Accident Database (ENSAD) for comparative risk assessment of accidents in the energy sector. *In:* Cepin, M.B., R. (ed.) *Safety and Reliability —Theory and Applications.* UK: CRC Press, Taylor & Francis Group.

Costa, J.E. 1985. *Floods from dam failures,* Denver, Colorado: United States Department of the Interior, Geological Survey.

ESRI. 2017. *Environmental Systems Research Institute, ESRI products, software, and services*. ESRI, Redlands, CA.

Extremadura, e. P. 2016. *51 años de la tragedia de la presa de Torrejón.*

Federal Emergency Management Agency, F. 2003. *HAZUS. Comprehensive Data Management System (CDMS) Data Dictionary, For Use with Hazus-MH Version 2.1*: FEMA, Mitigation Division, Washington, D.C.

Froehlich, D.C. 1995. Peak Outflow from Breached Embankment Dam. *Journal of water Resources Planning and management,* 121.

Gonzalez, J.G. & Santamarta, J.C. 2012. Technical development and characteristics of dam engineering in the Canary Islands. *Revista de Obras Públicas*: 33–50.

Graham, W.J. 1999. *A Procedure for Estimating Loss of Life Caused by Dam Failure,* Rep. No. DSO-99-06, Denver, Colorado: Sedimentation & River Hydraulics.

Haberturk. 2012. *Adana dam cover burst. Dam failure in Adana.* http://www.haberturk.com/yasam/haber/719281-adanada-baraj-faciasi-galeri.

Hirschberg, S., Spiekerman, G. & Dones, R. 1998. *Severe accidents in the energy sector—first edition. PSI Report 98-16*: Paul Scherrer Institut.

ICOLD 1995. *Dam Failures Statistical Analysis. Bulletin 99*, Paris, France.

ICOLD 2016. Dictionary. International Commission on Large Dams. http://www.icold-cigb.net/GB/Dictionary/dictionary.asp.

Kalinina, A., Sacco, T., Spada, M. & Burgherr, P. 2017. Risk assessment for dams of different types and purposes in OECD and non-OECD countries with a focus on time trend analysis. e-Proceedings of Hydro 2017 Conference "Shaping the future of Hydropower", 13.01, Spain.

La Vanguardia. 1944. Ha desaparecido el pantano de Xuriguera. *La Vanguardia, Espanola, Sebado, 26 de febrero de 1944*.

Lee, R., Hu, P.S., Neal, D.M., Ogles, M.R., Sorensen, J.H. & Trumble, D.A. 1986. *Predicting loss of life from floods,* Institute for Water Resources, US Army Corps of Engineers.

McClelland, D.M. & Bowles, D. 2000. Estimating life loss for dam safety and risk assessment: Lessons from case histories. Proc. 2000 Annual USCOLD Conference, U.S. Society on Dams, 2000 Denver, CO.

McClelland, D.M. & Bowles, D.S. 2002. *Estimating life loss for dam safety risk assessment—a review and new approach* Logan, Utah: Institute for Dam Safety Risk Management Utah State University.

Murcia Today. *Collapse of the Puentes dam in Lorca (1802)*. https://murciatoday.com/lorca-commemorates-the-collapse-of-the-puentes-dam_11502-a.html.

Rogers, J.D. & James, K. -. *Mapping the St. Francis dam outburst flood with geographic information systems:* University of Missouri-Rolla.

Ruan, D., Chen, G., Kerre, E.E. & Wets, G. 2005. *Intelligent Data Mining: Techniques and Applications.*

Salvati, P., Petrucci, O., Rossi, M., Bianchi, C., Pasqua, A.A. & Guzzetti, F. 2018. Gender, age and circumstances analysis of flood and landslide fatalities in Italy. *Science of The Total Environment,* 610–611: 867–879.

Saxena, K.R. & Sharma, V.M. 2004. *Dams: Incidents and Accidents*: CRC Press.

Smedley, E., Rose, H.J. & Rose, H.J. 1845. *Encyclopaedia Metropolitana, Or, Universal Dictionary of Knowledge.*

Swiss Federal Office of the Energy SFOE 2017. Methodik zur Bestimmung der Anzahl gefährdeter Personen (people at risk PAR) zur Abschätzung der hohen Gefahr (Version 1.0).

TRB 2000. *Highway capacity manual. Technical report,* Washington, D.C.: Transportation Research Board, National Research Council.

USACE 2017a. HEC-LifeSim 1.0. U.S. Army Corps of Engineers, Hydrologic Engineering Center.

USACE. 2017b. *HEC-LifeSim 1.0 Example Data*. U.S. Army Corps of Engineers, Hydrologic Engineering Center. http://www.hec.usace.army.mil/software/hec-lifesim/downloads.aspx.

USACE 2017c. HEC-RAS 5.0.3. U.S. Army Corps of Engineers, Hydrologic Engineering Center.

USACE 2017d. The Hydrologic Engineering Center's (CEIWR-HEC) River Analysis System (HEC-RAS). U.S. Army Corps of Engineers, Hydrologic Engineering Center.

Wikipedia. 2017a. *Desastre de Torrejón*. https://es.wikipedia.org/wiki/Desastre_de_Torrej%C3%B3n.

Wikipedia. 2017b. *St. Francis Dam*. Available: https://en.wikipedia.org/wiki/St._Francis_Dam.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

An experimental assessment of the MCS BDD algorithm in RiskSpectrum

O. Bäckström, R. Gamble, P. Krcal & W. Wang
Lloyd's Register, Stockholm, Sweden

ABSTRACT: The fault tree linking method for building Probabilistic Safety Assessment (PSA) models of nuclear power plants models accident sequences—combinations of safety system failures following an initiating event—by relatively small event trees. Failures of individual safety systems are modelled by fault trees. Such a model allows us to analyze selected accident scenarios or scenarios leading to a specific consequence. The analysis algorithm implemented in RiskSpectrum decomposes function events which fail along a sequence into minimal cutsets and (optionally) summarizes successful function events in an aggregate event, a so called *success module*. First order algorithms for quantification of a list of such minimal cutsets yield an approximate result. The new MCS BDD algorithm implemented in RiskSpectrum aims at improving this approximation. When computing resources suffice, it has the capability to quantify the minimal cutset list exactly. We evaluate the performance of this algorithm on real life models.

1 INTRODUCTION

A Binary Decision Diagram (BDD) is a data structure for encoding Boolean functions (Bryant 1986). It has been applied to fault tree analysis by Rauzy (1993) and Coudert & Madre (1993). Since then, it has been successfully used in many domains for a complete solution of fault trees (Rauzy 2006). However, the size of event/fault tree models emerging from Probabilistic Safety Assessment of nuclear power plants is prohibitive for an exact analysis by the means of BDDs.

The current standard practice for solving large fault tree models from the nuclear Probabilistic Safety Assessment is to decompose the tree structure into a list of Minimal c utsets (MCS). Further, first order algorithms, such as rare event approximation or Min Cut Upper Bound (MCUB), are used to quantify this minimal cutset list.

The MCS BDD algorithm implemented in RiskSpectrum presents a new method of minimal cutset list quantification. A key to its efficiency and accuracy is a heuristic procedure which assigns certain nodes for the exact treatment and certain nodes for a treatment similar to a ZBDD (Minato 1993, Jung et al. 2004). This provides us with an improved quantification for minimal cutset lists with high probability events and with a small set of high importance events. Also, it allows for event tree success quantification dependent on the failed events. In the optimal case, if the complete success information is generated, the algorithm has the capability to quantify success exactly.

In this paper, we evaluate the current implementation of the MCS BDD algorithm on real life models. The focus of the assessment lies in the following aspects:

Improved accuracy—for which types of analyses do we obtain a considerable decrease in conservatism of the results?

Success quantification—we demonstrate the capability of the algorithm to quantitatively assess the event tree success. We also discuss sensitivity of the algorithm to various model parameters.

Importance factors—we study effects of the new quantification method on the Risk Increase Factor. To what extent is this importance factor affected and where is the biggest gain?

Efficiency—the algorithm allows for user control of the trade-off between the result accuracy and the calculation resources. We investigate the effect of different settings on the calculation time and the MCS list value. Especially, we ask when one can rely on the built-in automatic parameter adjustments and in which situations one needs to steer the algorithm by fine-tuning the parameters manually.

2 BACKGROUND

A Binary Decision Diagram is a rooted binary directed acyclic graph with nodes labeled by decision variables and leaves representing one of the two values: True or False. The MCS BDD algorithm applies pivotal decomposition to the MCS

list in order to build a BDD. It selects a basic event from the MCS list as the node decision variable and builds its child BDDs from smaller MCS lists with possibly smaller cutsets. A leaf is built when the MCS list to be processed is empty or it contains an empty cutset. Figure 1 schematically depicts one step in BDD building.

One of the approximations in the analysis of large fault trees by the minimal cutset decomposition is the quantification of the resulting minimal cutset list. MCS BDD is an algorithm that has the potential to quantify the MCS list exactly. This would, however, lead to inacceptable calculation times (and possibly exceed computer memory capacities) for large real-life models like those from nuclear PSA. A complete BDD quantification of a MCS list with success modules can unfortunately be achieved only for very small cases. Therefore, we need to trade absolute accuracy for reasonable calculation times.

The MCS BDD algorithm adopts a highly pragmatic approach. It searches for events where a precise BDD quantification has the greatest effect and treats other events in an approximate way. First, we describe the most important heuristics in the MCS BDD algorithm and then we motivate them by describing types of models where this brings the greatest advantage.

For presentation purposes, we have the following assumptions on MCS lists (for the full description of the algorithm see Bäckström et al. (2014), Bäckström et al. (2016). We assume that an input MCS list contains independent basic events and possibly also success modules. A success module is a summary characterization of function events which succeed along an event tree sequence. Technically, it is also a MCS list containing cutsets which fail one of these function events. A success module is then interpreted as 'negated' in cutsets. For a more detailed description of success modules, see RiskSpectrum (2013). Cutsets are grouped according to the initiator (a frequency basic event) and a success module and then treated separately. This means that we can without loss of generality assume that there is only one initiator and only one success module in the MCS list.

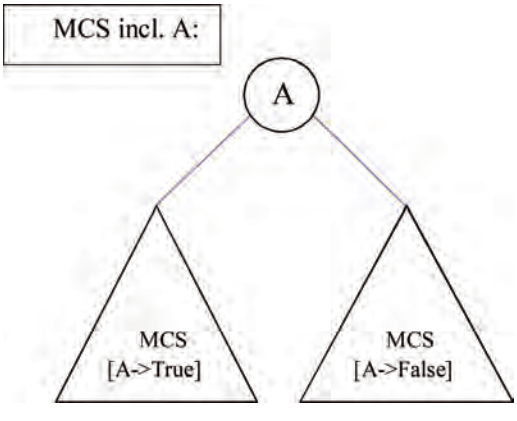


Figure 1. The minimal cutset list with cutsets containing the basic event A is decomposed into two smaller MCS lists which are then processed recursively.

2.1 *Main ingredients in MCS BDD*

The algorithm first defines the MCS list which shall be transformed into a BDD. Then, it builds the BDD. Finally, it computes the probability/frequency of the MCS list from the BDD. In the following we detail the most important parts of the process and by this also explain the main heuristic ingredients:

- Selection of the most important cutsets
- Exact and approximate nodes
- Success module quantification

2.1.1 *Selection of important cutsets*

The greatest contribution to the MCS list probability/frequency typically comes from a small portion of the minimal cutsets. The first heuristics splits the MCS list into two parts, based on the cutset values. The part which represents almost the complete MCS list value is treated by the MCS BDD algorithm. The remaining part is quantified by the Min Cut Upper Bound algorithm.

This heuristics works as a variant of a probabilistic cutoff, but its effect is purely conservative. The greater the part that is treated by the MCUB, the greater the over-approximation of the MCS list value is and also the easier is it to build the BDD.

2.1.2 *Exact and approximate nodes*

The algorithm produces BDD nodes and recursively creates their inputs from smaller MCS lists. Each node that is added is either included as an exact or approximate node. The process of determining the node type (exact or approximate) is equally important as the selection of pivotal element in the RiskSpectrum implementation.

The exact method directly follows from Shannon non-intersect decomposition. We split the MCS list (which is a probabilistic Boolean function) by a pivot basic event A (a decision variable in the formula) into two parts:

- Cutsets which are consistent with A. We remove A from these cutsets, if it is there. This corresponds to evaluating the Boolean formula with A = True.
- Cutsets which are consistent with ¬A. This removes all cutsets containing A. This corre-

sponds to evaluating the Boolean formula with A = False.

Formula 1 shows how to calculate the probability of a Boolean function, where A is one of the decision variables.

$$P(f) = P(A) \cdot P(f[A]) + P(\neg A) \cdot P(f[\neg A]) \quad (1)$$

By f[A], f[¬A] we denote the Boolean function f where A = True, A = False, respectively.

The approximate treatment splits the MCS list by a pivot basic event A into two parts:

- Cutsets which contain A. We remove A from all of these cutsets.
- Cutsets which do not contain A.

This is following the same principle as ZBDD quantification. The usual quantification yields the same result as the rare event approximation—a direct sum of the cutset probabilities. The algorithm implemented in RiskSpectrum makes use of the independence assumption of basic events. It uses the Min Cut Upper Bound approximation on independent or positively correlated cutsets. The quantification in the approximate method is performed according to Formula 2.

$$P(f) = P(A) \cdot (P(f_A) + P(f_{\neg A}) - P(f_A) \cdot P(f_{\neg A})) + (1 - P(A)) \cdot P(f_{\neg A}) \quad (2)$$

By f_A and $f_{\neg A}$ we denote cutsets that contain A and that do not contain A, respectively. Moreover, the event A is removed from cutsets in f_A. This formula quantifies both parts in the same recursive way and then it removes product of their probabilities. Note that this way of calculating the MCS list probability gives a lower value than the MCUB approximation on the whole MCS list (and by this also a lower value than the rare event approximation). Even if the whole MCS BDD was built just from approximate nodes, the result would be more precise than the MCUB quantification.

There are conditions that have to be satisfied to guarantee that the approximate method does not yield an under-approximation of the exact value. A pivot basic event A cannot be treated by the approximate method if f_A and $f_{\neg A}$ are negatively correlated due to the fact that the same event occurs in f (non-negated) and at the same time it appears in a success module.

This has shown to be a significant drawback of the approximate method, as it either adds a restriction on using approximate nodes or it limits the exact quantification of success modules.

Exact nodes and approximate nodes can be mixed. Approximate nodes make the analysis very efficient, exact nodes give a better result and allow for more precise success quantification. The MCS BDD algorithm has to balance these desirable properties. It searches for the best candidates for the exact treatment and treats the remaining basic events approximately.

2.1.3 *Success module quantification*

Success modules are quantified by the same technique. We build an MCS BDD for the success module and append it to the MCS BDD for the minimal cutset list. The success module BDD uses only exact nodes. Therefore, we can quantify its (independent) value exactly.

Disregarding dependencies between the cutsets and the success module contributes to the conservatism of the result. If we want to reduce this conservatism, we need to take these dependencies into account. However, this limits the possibilities of applying the efficient approximate treatment of nodes in the whole BDD. Therefore, we select only the most important events in the success module and disregard dependencies for the remaining ones.

2.2 *MCS BDD parameters*

Users have a possibility to steer the search for events to be treated exactly/approximately as it affects the calculation speed and the accuracy of the result. This can be done by setting algorithm parameters manually before building the BDD. The algorithm has the capability of an automatic adjustment of these parameters. This allows users to use default values for most (if not all) analysis cases in their models. The following parameters can be specified by a user:

- *MCS limit* ranges from 0 to 1 and expresses the percentage of the MCS list value above which cutsets are treated by MCS BDD.
- *Q limit* and *FV limit* parameters are used to determine the node treatment method—exact or approximate and the treatment of dependencies between events in the success module and the failure part. These parameters range from 0 (only exact) to 1 (only approximate).
- *BDD nodes limit* bounds the number of nodes that a BDD can use and by this it limits the complexity of the BDD generation.

If the algorithm cannot succeed in building a BDD with the given combination of parameters then it automatically adjusts the parameters MCS limit, Q limit and FV limit and attempts generating a BDD again. This is repeated until the algorithm successfully returns a BDD structure.

2.3 *Purpose of MCS BDD*

MCS list quantification by the new algorithm improves the accuracy of the MCS list value. Situ-

ations in which first order approximations suffer from a significant over-approximation are:

- Events with high probabilities
- Success quantification

The goal of the experimental evaluation is to assess to what extent the implementation fulfils its purpose. Additionally, we evaluate the possibilities that the algorithm parameters give users in steering the quantification.

3 EVENT TREE QUANTIFICATION

Quantification of event tree success necessarily requires handling of non-coherent fault trees (Nusbaumer & Rauzy 2013, Bäckström et al. 2012). Function event success brings negated 'events' into minimal cutsets. RiskSpectrum summarizes successful function events by a new type of event—a success module, which is simply a minimal cutset list containing basic event combinations which cannot occur in quantified sequences. Moreover, not-logic in fault trees might introduce negated basic events into minimal cutsets.

We have performed two types of assessment. The first one compares the sequence or consequence value calculated by the MCS BDD algorithm to the value computed by the Min Cut Upper Bound algorithm. By this, we analyze the accuracy increase obtained from the new algorithm. We investigate cases with significant differences closer in order to identify factors driving the accuracy increase. Large real-life models are used.

The second type of assessment compares quantification of sequences and consequences with success modules to reference values calculated by an analysis of structures which contain only function event failures (Nusbaumer & Rauzy 2013). Models with newly built event trees based on industrial PSA models are used.

3.1 *Comparison to min cut upper Bound*

We have evaluated the accuracy of MCS BDD on 15 real-life models. We split them into two categories—models which quantify event tree success (by means of success modules) and models which use event tree success only to remove basic event combinations which do not belong to the analyzed sequence. The latter group of models disregards the success probability both in the first order quantification algorithm as well as in the MCS BDD algorithm. We have analyzed individual sequences and also sequences grouped according to the accident consequence (consequence analysis cases).

3.1.1 *Analyses without ET success quantification*

First, we present results from the group of models which do not quantify Event Tree (ET) success. The results reflect only how efficiently the MCS BDD algorithm deals with dependencies between minimal cutsets. Table 1 shows decrease of the MCS list frequency from the value calculated by MCUB in percent of the new value calculated by MCS BDD. It lists the number of cases analyzed in the model, the maximal decrease, number of analysis cases with the decrease above 25%, number of analysis cases with the decrease above 10% and the minimal increase.

All of these results have been calculated with the default settings: the MCS limit has been set to 1E-3, Q limit and FV limit have been set to 1E-2. The vast majority of the MCS BDD results end up between the second and the third order approximation, which is an expected result. Exceptions rather indicate an improvement potential in the calculation of the second/third order approximation than in the MCS BDD algorithm.

Obtaining an increase of the MCS list value, even though very small, is unexpected. It could be traced to a less efficient quantification of modules in the MCS BDD algorithm and should be resolved in future versions. Most of the increases can be removed when we increase the MCS BDD accuracy by the algorithm settings.

Analysis cases with the greatest value decrease can be often identified by one property: they contain high probability events with high Fussel-Vesely importance.

3.1.2 *Analyses with ET success quantification*

As the next step, we present results from analyses where function event success in event trees is quantified by the means of success modules. Apart from dependencies between cutsets, the MCUB algorithm does not take into account dependencies between failed basic events in a cutset and basic events from the success module in this cutset.

Table 1. A summary of the MCS list value decrease with the MCS BDD algorithm compared to the MCUB algorithm. Event tree success is not quantified.

Test model	Number of cases	Max (%)	# > 25%	# > 10%	Min (%)
M-01	1200	28.8	1	15	–1.8
M-02	375	103	48	112	0
M-09	2700	89.8	160	320	0
M-10	24000	17.3	0	19	–0.83
M-11	640	309	223	323	–0.01
M-12	6000	42.9	318	712	–0.09
M-13	3400	40.8	47	188	–0.99

This might result in additional conservatism of the result.

Figure 2 illustrates this issue. Consider the second sequence. The function event "A AND B" succeeds and its corresponding success module contains the cutset {A, B} = SM. The second function event fails and produces the cutset {IE, A, SM}. The quantification method used in the MCUB algorithm calculates values of events and modules independently. The cutset value is then P(IE)*P(A)*(1-(P(A)*P(B)). This is an overly conservative quantification. We know that A has failed in this cutset. Therefore, the probability that the success module has not failed is only 1-P(B). This gives us the cutset value P(IE)*P(A)*(1-P(B)) which the MCS BDD algorithm returns (provided that the event A is considered as important by the algorithm heuristics). Moreover, the whole success module will be calculated by the means of MCS BDD.

Table 2 shows decrease of the MCS list frequency from the value calculated by MCUB in percent of the new value calculated by MCS BDD. It lists the number of cases analyzed in the model, the maximal decrease, number of analysis cases with the decrease above 25%, number of analysis cases with the decrease above 10% and the minimal increase.

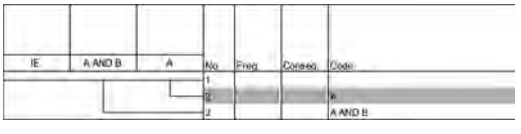


Figure 2. An event tree where the success quantification of the first function event depends on the second function event.

Table 2. A summary of the MCS list value decrease with the MCS BDD algorithm compared to the MCUB algorithm. Event tree success is quantified.

Test model	Number of cases	Max (%)	# > 25%	# > 10%	Min (%)
M-02	800	27.1	2	17	–65.6
M-03	3	–2.4	0	0	–21.4
M-04	4200	469.3	360	625	–3.4
M-05	4000	867	120	331	–2.3
M-06	7000	157	80	530	–0.7
M-07	9700	107.8	100	350	–3.0
M-08	660	0.48	0	0	–5.6
M-11	1100	1132	350	500	–64.4
M-14	3700	5.1	0	0	–0.01
M-15	2100	300.6	53	377	–22.7
M-16	910	72.1	2	27	–0.1

In some cases the MCUB results may be significantly lower than the value calculated by the MCS BDD algorithm. The cases where this has been found underestimate the success module value in the MCUB calculations. This is because the MCS list in the success module is itself quantified by MCUB which gives a conservative estimate of a cutset list. This means that we overestimate its value and then calculate its 'negation' (1-P(SM)). This in its turn means that the success module value becomes underestimated. Normally, the conservatism from the MCUB is rather limited and other approximations applied during success module creation have more significant effect. The extensive evaluation revealed cases where the difference is greater than 50% of the MCS list value. For these cases, we have investigated the success module and verified that the MCUB algorithm underestimates the success module. Note that RiskSpectrum offers a possibility to identify candidates for this effect using the second order quantification of success modules.

Mostly the MCS BDD yields a lower result—due to the increased accuracy and the treatment of dependencies between failed MCS and the success MCS. The lower results generated by the MCS BDD are mainly an effect of the treatment of dependencies between failed events and the success module. The MCS list value in some of the analysis cases decreased several times. This is typically the case when the success module contains one of the important events in the failed part of cutsets and this event occurs in combination with high probability events in the success module. This means that in all combinations where the event is failed the success module will have a very low probability.

The default settings (1E-3, 1E-2, 1E-2) seem appropriate also for analyses with quantified event tree success.

All results have been studied, and a few analysis cases per model have been reviewed to ensure that the differences in results can be explained, and that the MCS BDD produces a better estimate of the MCS list. We summarize comments on a selection of analysis cases.

Model 7 – Several sequences with MCS BDD results significantly lower than MCUB estimates have been studied in detail. The result is overestimated due to a number of high probability events. The results without success modules show the same behavior. Calculating the MCS list value up to the 6th order confirms this hypothesis.

Model 7 – two sequences yield a very low result (zero) with the MCS BDD and non-zero in MCUB. The reason is that the success module

contains an event which is a part of all cutsets in the failed part of the MCS list. This event is combined with another event in the success module which has probability one. Since the event is failed the success module probability will be zero and the value of all cutsets becomes zero as well.

Model 11 – Eight analyses with high differences were studied in detail. Four sequences with significantly different results contain many high probability events re-occurring in most important cutsets. This causes a very big overestimate from the MCUB algorithm. One analysis case has a difference in >1000% of the top value. We have shown by a reduced MCS list that the MCUB analysis indeed yields a very big overestimate in this case due to high probability events. Four analysis cases that have results more than 50% higher with the MCS BDD have big dependencies between failed basic events and basic events in the success module. Dependent quantification of success modules can take these dependencies into account and thus reduce the conservatism in the MCS list value.

3.2 *Comparison to exactly calculated ET success*

A collection of models with non-coherent structures have been set up for the assessment purpose. Most of them are based on real-life models. In each test model, the evaluated event tree has an initiating event (IE) and three function events (F1, F2, F3). Figure 3 depicts such an event tree.

Success of each function event contributes to the success module, which means that cutsets for all sequences but the last one (failure of all function events) contain success modules. Each function event takes a fault tree as input. Table 3 lists relevant features affecting the evaluated test cases for each model.

Results produced by the MCS BDD algorithm are compared to values obtained by the method which avoids function event success and quantifies only combinations of function event failures (Nusbaumer & Rauzy 2013). Function event failure combinations are also decomposed to MCS lists which are then quantified by the means of

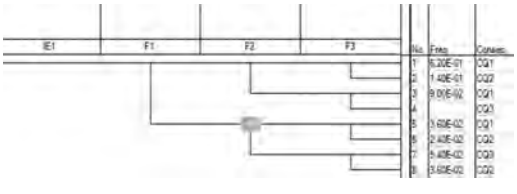


Figure 3. An example event tree used in assessment of non-coherent structure quantification.

Table 3. Model characteristics for event tree quantification.

Test model	Number of gates	Negations	High prob. events	Others
EXACT1	7			
EXACT2	69			
EXACT3	22442	Y	Y	
EXACT4	7771	Y		
EXACT5	5272	Y	Y	
EXACT6	7719	Y		IE and FE have dependencies.
EXACT7	5464	Y	Y	
EXACT8	7228			
EXACT9	13125			
EXACT10	14021	Y	Y	

Table 4. Comparison of the values computed by MCS BDD and reference values calculated by the method from Nusbaumer & Rauzy (2013).

Test model	Maximum difference in Sequence analysis case results (%)	Maximum difference in consequence analysis case results (%)
EXACT1	0.0	0.0
EXACT2	1.1	0.1
EXACT3	62	1.0
EXACT4	0.0	0.0
EXACT5	0.0	0.0
EXACT6	1.0	0.3
EXACT7	0.1	0.0
EXACT8	125.1	0.0
EXACT9	2.0	0.1
EXACT10	0.1	0.1

MCS BDD. This allows us to assess accuracy of function event success quantification in MCS BDD. Table 4 contains the comparison of these two methods.

For the test model EXACT3, three sequences differ by 14%, 19% and 62%. This is due to the accuracy of success modules. It shall be noticed that the accuracy of the success module is possible to adjust, but this has not been done in this evaluation (only when results are investigated). Sequences with low frequency in the model EXACT4 get a lower value than the reference value. This is due to cutoff application during MCS generation. The value gets slightly above the reference with a lower cutoff. One low frequency sequence in the model EXACT7 is below the reference value. This is the same cutoff issue.

In the model EXACT8, two low probability sequences differ by 99% and 125%. This is due to

success module accuracy. Increasing the success module accuracy gives differences 0.7% and 2.7%, respectively. The model EXACT10 contains four sequences with values below the reference value by 0.1%. One consequence is below the reference value by 0.1%. This is a cutoff issue during generation of minimal cutsets

Parameters used to generate the MCS BDD were set to zero for small models (EXACT1, EXACT2), ensuring exact treatment of all nodes. For other models, the MCS limit has been set to 1E-5, Q limit and FV limit have been set to 1E-3.

4 IMPORTANCE—RISK INCREASE FACTOR

The new quantification procedure does not only affect MCS list values, but also other measures such as importance factors. One of these measures most affected by inaccuracies not handled by the first order calculation is the Risk Increase Factor (RIF), also called Risk Achievement Worth (RAW).

A RIF of a component is defined as the factor of power plant risk increase when the component is unavailable (Vesely et al. 1983). Whenever any failure mode of this component occurs in a fault tree, it needs to be considered as failed (Bäckström et al. 2016). The method implemented in RiskSpectrum (RiskSpectrum 2016) calculates RIF for an object by setting the failure probability of all events in this object to one, recalculating the MCS list of the analysis case and dividing the obtained value by the nominal MCS list value.

We have evaluated RIF calculations with the MCS BDD algorithm on three models (a sample model and two real-life models). For selected basic events and event groups, we have calculated the RIF value by re-running the analysis with the analyzed basic events marked as failed in the model.

The results for basic events are summarized in Table 5. We report the maximal decrease of RIF in percent of the new RIF value (Max) and the minimal increase of RIF in percent of the new RIF value (Min).

Table 5. A summary of risk increase factor values for basic events in three evaluation models. Two models have been analyzed with and without event tree success quantification.

Test model	Number of basic events	Max (%)	# > 10%	# < 0%	Min (%)
S	126	3.0	0	2	–0.2
S(Succ)	126	3.3	0	4	–1.3
R1	320	12	37	66	–4.8
R1(Succ)	320	12	37	52	–4.8
R2	470	19	2	93	–1.9

The changes in RIF values are rather moderate. The decrease of RIF values for many basic events is relatively surprising. This decrease is small, less than 5% in all cases. The explanation of this phenomenon lies in the increased accuracy also when calculating the nominal MCS list value—the value of the MCS list generated by the original analysis.

Table 6 contains manually calculated RIF values for sample basic events from the model S and R1. It shows the RIF value calculated by the MCUB quantification, by the MCS BDD algorithm and manually with newly generated results, quantified by MCS BDD.

One can see that the MCS BDD quantification improves RIF values in the sense that it brings them closer to the actual RIF. This is even the case if the RIF value increases for a basic event, exemplified by the basic event BE-01896.

Apart from individual basic events, RiskSpectrum analyzes RIF also for groups of basic events (which could be defined in multiple ways as components, according to attributes, or directly as groups of basic events). Table 7 shows differences in RIF

Table 6. RIF values for sample basic events calculated by the MCUB quantification, MCS BDD and a manual calculation with a newly generated MCS list.

Test model	Basic event	RIF—MCUB	RIF—MCS BDD	RIF—Manual
S (Succ)	ACP-DG02-M	1.62	1.60	1.60
S (Succ)	ACP-DG01-M	1.62	1.60	1.60
S (Succ)	EFW-TR01-M	1.49	1.47	1.47
S	ACP-GT01-A	2.06	2.04	2.04
S	FEED&BLEED	3.03	3.03	3.03
R1 (Succ)	BE-00924	4.54	4.07	4.02
R1 (Succ)	BE-00882	4.54	4.07	3.46
R1	BE-01896	2020	2110	2110
R1	BE-09400	3.86	3.46	2.20

Table 7. RIF values for sample event groups calculated by the MCUB quantification, MCS BDD and a manual calculation with a newly generated MCS list.

Test model	Basic event	RIF—MCUB	RIF—MCS BDD	RIF—Manual
S (Succ)	SYSTEM:RHR	367	23.3	23.3
S (Succ)	SYSTEM:EFW	414	19.8	19.8
S (Succ)	SYSTEM:ECC	43.5	6.56	6.56
R3	EG-2	1.51E+6	1.15E+6	4.05E+5
R3	EG-3	17914	3613	N/A
R3	EG-4	2.47E+5	1.05E+4	N/A
R3	EG-5	127	121	N/A

values for groups of basic events when calculated by the MCUB algorithm, MCS BDD or manually.

The evaluation of the RIF values calculated by the MCS BDD algorithm shows that it can achieve lower (more exact) values for groups of basic events, in some cases one order of magnitude. For smaller and simpler models, it achieves the values one gets from a calculation re-generating the MCS list without the event(s) under study.

5 MCS BDD SETTINGS

The heuristics balancing scalability and accuracy of the MCS BDD algorithm can be steered by the following settings:

MCS limit—specifying the part of the MCS list which should be converted into a BDD, while the rest of the MCS list is quantified by the MCUB algorithm.

Q limit—specifying from which probability are events treated exactly when building the BDD.

FV limit—specifying from which importance are events treated exactly when building the BDD.

Node limit—absolute bound on the number of BDD nodes. If more nodes are needed, the algorithm increases other limits and restarts.

Exact usage of these limits is described in RiskSpectrum (2016). In general, the lower the MCS, Q and FV limits are, the bigger and more precise the generated BDD is, within the size defined by the Node limit. At the same time, we can expect longer generation times for larger and more precise BDDs.

Table 8. Differences in calculation times and MCS list values with varying calculation settings.

Settings	M-02 Time (s) /avg. diff. to default (%) /Max or min diff. to default (%)	M-10 Time (s) /avg. diff. to default (%) /Max or min diff. to default (%)	M-05 Time (s) /avg. diff. to default (%) /Max or min diff. to default (%)
Default	0:21 / N/A	1:19 / N/A	1:55 / N/A
1E-1, 1E-2, 1E-2	0:09 / 2.2 / 42.4	0:28 / 0.5 / 11.1	0:45 / 0.6 / 79.9
1E-5, 1E-2, 1E-2	0:46 / 7E-3 / –3.2	4:04 / 0.0 / –1.7	1:48 / 0.0 / –0.9
1E-3, **1E-1**, 1E-2	0:12 / 0.02 / 0.4	0:30 / 0.0 / 1.9	2:44 / 0.05 / 3.4
1E-3, **1E-4**, 1E-2	0:56 /–0.02 / –5.7	7:43 / 0.0 / –1.2	9:40 /–0.02 / –3.9
1E-3, 1E-2, **1E-1**	0:09 / 0.02 / 0.4	0:32 / 0.0 / 1.9	2:47 / 0.2 / 59.1
1E-3, 1E-2, **1E-4**	0:26 /–0.05 / –5.7	7:17 / 0.0 / –1.2	9:37 /–0.03 / –12.4
1E-3, **0, 0**	5:26 / 3E-3 / –5.8	N/A	N/A

Table 8 shows the MCS list values and BDD generation times with different settings. Default settings are 1E-3, 1E-2, 1E-2 for the MCS limit, Q limit and FV limit, respectively. We always write setting values in this order.

The evaluation shows surprisingly small differences in the resulting MCS list values. This can be explained by the automatic parameter adjustment and it confirms the choice of default setting values. Adjusting the settings manually makes sense in individual analyses when one suspects that the accuracy could be substantially increased by a larger BDD. In this case, one needs to accept longer calculation times, as it might require adjustment of the Node limit setting.

6 CONCLUSIONS

Experimental assessment of the MCS BDD algorithm implemented in RiskSpectrum leads to the following conclusions.

Heuristics used in the quantification algorithm have proven to be very efficient. It is possible to derive results with relatively high accuracy. The frequency estimate of a minimal cutset list might be several times smaller for cases where first order approximations give overly conservative results. This is the case in presence of high probability events which occur multiple times in cutsets with a high contribution.

More importantly, the MCS BDD algorithm allows for very accurate quantification of event tree success, both for individual sequences and (even more) for groups of sequences defined by a consequence to which they lead. Limitations in accuracy of several cases reported in this paper stem from the bounds on the success module accuracy with default settings. Improving the success quantification by lifting these bounds will be investigated in future work.

The new quantification also improves estimates of importance factors. We have evaluated the effect on the Risk Increase Factor, where especially the values for basic event groups might improve dramatically, in some cases by an order of magnitude.

Finally, the combination of default setting values and the automatic setting adjustment result in acceptable calculations times in commercial applications.

REFERENCES

Bryant, R. 1986. Graph Based Algorithms for Boolean Function Manipulation. *IEEE Transactions on Computers*, 35(8): 677–691.

Bäckström, O. & Gamble, R. & Krcal, P. & Wang, W. 2014. MCS BDD—Description and verification of the method implemented in RiskSpectrum. *PSAM 13, 2016*.

Bäckström, O. & Gamble, R. & Krcal, P. & Wang, W. 2014. Two Interpretations of the Risk Increase Factor Definition. In *European Safety and Reliability Conference (ESREL)*, 2016.

Bäckström, O. & Krcal, P. & Wang, W. 2014. Quantification of MCS with BDD, accuracy and inclusion of success in the calculation—the RiskSpectrum MCS BDD algorithm. *PSAM 12, 2014*.

Bäckström, O. & Krcal, P. 2012. A Treatment of Not logic in Fault Tree and Event Tree Analysis. *PSAM 11, 2012*.

Coudert, O. & Madre, J.-C. 1993. Fault Tree Analysis: 1020 Prime Implicants and Beyond. In *Annual Reliability and Maintainability Symposium, 1993*.

Jung W.S. & Han S.H. & Ha J. 2004, A Fast BDD Algorithm for Large Coherent Fault Trees Analysis. In *Reliability Engineering and System Safety*, Vol. 83, pp. 369–374.

Minato, S.-I. 1993. Zero-Suppressed BDDs for Set Manipulation in Combinatorial Problems. In *Proc. of DAC'93*.

Nusbaumer, O. & Rauzy, A. 2013. Fault Tree Linking versus Event Tree Linking Approaches: a Reasoned Comparison. *Risk and Reliability* 277(3): 315–326.

Rauzy, A. 1993. New Algorithms for Fault Trees Analysis. *Reliability Engineering & System Safety* 59(2): 203–211.

Rauzy, A. 2008. Binary Decision Diagrams for Reliability Studies. *Handbook of Performability Engineering*, 381–396.

RiskSpectrum. MCS BDD User's Manual. Version 3.4.0. *Lloyd's Register Consulting*, 2016.

RiskSpectrum. Theory Manual. Version 3.3.0. *Lloyd's Register Consulting*, 2013.

Vesely W.E. & Davis T.C. & Denning R.S. & Saltos N. 1983. Measures of Risk Importance and Their Applications. *NUREG/CR-3385.*

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A new approach for social vulnerability in mainland Portugal area for risk mitigation

A.O. Tavares
Earth Sciences Department, Center for Social Studies, University of Coimbra, Portugal

J.L. Barros, P.P. Santos & J.M. Mendes
Center for Social Studies, University of Coimbra, Portugal

ABSTRACT: The concept of Social Vulnerability (SV) is characterized by its multidimensionality. In the present study, Social Vulnerability was analyzed and evaluated according to the methodology developed by the Center for Social Studies of the University of Coimbra, which presents as innovative feature the incorporation of the Criticality and Support Capability components. Social Vulnerability was calculated for the 278 municipalities of mainland Portugal using factor analysis. The evaluation and calculation of the Criticality was carried out using 22 variables, selected from an initial number of 90, and the calculation of Support Capability was performed using 12 variables, from an initial number of 145 variables. The obtained outputs should be a working basis for the managers and stakeholders, authorities at different levels, and all the community with the objective of adopting adaptation and mitigation measures to natural and technological risks.

1 INTRODUCTION

The concept of Social Vulnerability (SV) is characterized by its multidimensionality, adding not only the social characteristics of the individual, but also their social and economic relations, as well as the physical and social environment where the individual is inserted (Tapsell et al., 2010). The differentiating characteristics of SV make it imperative not only in the characterization and understanding of the degree of exposure of the communities, but also in their capacity for resisting and recovering in face of hazardous events.

Historically, the concept of Social Vulnerability has emerged as an explicit critique of the dominant and conventional paradigms of analysis of disasters, with Hewitt (1983). The Sendai Framework for Disaster Risk Reduction resumes the concept of vulnerability as the conditions determined by the physical, social, economic and environmental factors or processes that increase the susceptibility of a community to the impact of hazards (UNISDR, 2015). Thus, the scientific community as recognized the need of considering social vulnerability as a particular dimension of vulnerability, developing distinct approaches for its measurement (e.g., Angeon and Bates, 2015; Rufat et al., 2015; Fatemi et al., 2017).

As noted by Wisner et al. (2004) the vulnerability to hazards is a multidimensional process that consists in a multiplicity of components related with historical, political, economic, environmental and demographic factors, which produce inequalities, dynamic pressures such as rapid urbanization and social pressures and unsafe living conditions that originates unequal exposure to risk.

There are multiple and distinct methods of measuring vulnerability (Birkmann, 2006; Fuchs et al., 2012; Birkmann, 2013; Birkmann et al., 2013;). In the present work, Social Vulnerability to natural and technological risks was analyzed and evaluated according to the methodology developed by the Center for Social Studies of the University of Coimbra (CES) and its Risk Observatory (OSIRIS) (Mendes et al., 2011). According to Mendes et al. (2011) the concept of SV is associated with the degree of exposure to natural and technological hazards and extreme events, depending closely on the resilience of individuals and communities.

Social Vulnerability must be a planning tool, supporting the implementation of a territorial model in which decision-making on risk management would be more efficiently applied.

The study is divided into 5 sections: a) Presentation of the area of study; b) methodology for the calculation of Social Vulnerability and its components; c) results at municipal level; d) discussion of the results; e) conclusions of study.

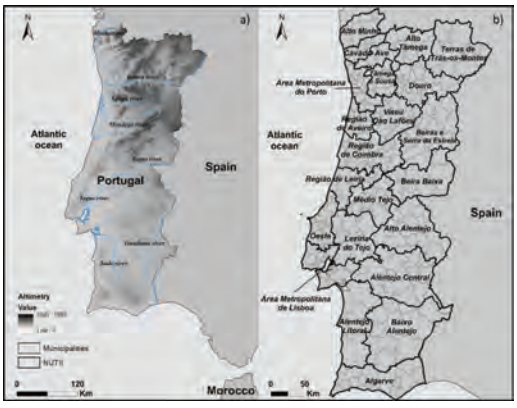


Figure 1. Location of the studied area: a) Continental Portugal (NUT I); b) Territorial organization in NUT II and LAU I (municipalities).

2 STUDY AREA

The present study was based on the calculation of SV for the 278 municipalities of mainland Portugal, with a total area of 89,089 Km^2 and a resident population of 10,044,484 inhabitants according to the 2011 Census (INE, 2012). In administrative terms Portugal is divided into three NUTS (Nomenclature of Territorial Units for Statistics) which is subdivided into three levels, defined according to population, administrative and geographical criteria and in two LAU (Local Administrative Unit), in accordance with Decree-Law 244/2002, changed in 2015 by regulation nº868 / 2014. The work presented here supports its analysis at the level of NUT III, which is composed of 23 territorial units and LAU I, which is composed of 278 municipalities (Figure 1b).

3 METHODOLOGY

The principal objective of this work is to evaluate the Social Vulnerability at municipal level in mainland Portugal. This evaluation will be assessed using principal component analysis (PCA), a technic also used by different authors like Cutter et al. (2003), Schmidtlein et al. (2008), Mendes (2009), Barros et al. (2015), with adaptations according to regional and local specificities, expressed in the type of variables and unit of analysis to be selected. For PCA was used the software SPSS®, version 23. The data that supports this evaluation were obtained using information from the Census 2011 (INE, 2012) and PORDATA database (PORDATA, 2017). In this study the conceptual understating of Social Vulnerability defined by Mendes et al. (2011) was adopted, where SV is composed by two components: Criticality and Support Capability. The evaluation of Social Vulnerability was based on PCA where redundant variables are eliminated and the remaining are normalized and grouped into factors. The PCA was carried out based on a set of premises where it stands out: a) the calculation of the Pearson correlation matrix analysis; b) the variance rate parameters (should be greater than 60%) and the Kaiser-Meyer-Olkin (KMO) sample measurement (should be greater than 0.6) with the purpose of eliminating redundant data (Comrey et al., 2009) and select the more PCA-robust dataset; c) the use of Varimax rotation to better identify the principal components. This process is done for both Criticality and Support Capability. After obtaining the respective scores in each municipality, Social Vulnerability is calculated by combining the two components mentioned above using the following equation:

$$\textit{Social Vulnerability} = \textit{Criticality} \times (1 - \textit{Support Capability}) \quad (1)$$

The results obtained are grouped into different classes that vary from very low to very high in accordance with the standard deviation (SD) and the following categories: "very low," <1 SD; "low," [–1, –0.5 SD]; "moderate," [–0.5, +0.5 SD]; "high," [0.5, 1 SD]; "very high," ≥ 1 SD (Cutter et al. 2003).

3.1 *Criticality*

The calculation of Criticality for all municipalities of mainland Portugal was carried out using 22 variables grouped into seven groups (Table 1). PCA identified 6 factors (FAC) based in the 22 explicative variables. These factors present a variance rate of 73% for the 278 municipalities under study, with a KMO of 0.726 and all communalities above 0.6.

3.2 *Support capability*

The Support Capability was performed using 12 variables grouped into four groups (Table 2).

Table 1. Groups of variables used in the calculation of municipal criticality.

Groups	Number of variables
Social support	3
Housing conditions	2
Demography	2
Economy	9
Education	2
Housing	3
Health	1

Table 2. Groups of variables used in the calculation of municipal support capability.

Groups	Number of variables
Economy	4
Civil protection resources	4
Building characteristics	2
Health facilities	2

Table 3. Criticality components.

FAC	Name	Explained variance (%)
1	Risk groups	30
2	Economic conditions	13
3	Disadvantaged population	12
4	Level of income	7
5	Employment	6
6	Dependent population	5

Based on the variables presented in Tables 2, 3 FAC's were retained, presenting a variance rate of 65% for the 278 municipalities under study, with a KMO of 0.705 and all communalities above 0.6.

4 RESULTS

4.1 *Factors of criticality*

As mentioned above the Criticality assessment identified 6 factors with different percentages in the explained variance (Table 3).

4.1.1 *Factor 1 – Risk groups*

The factor named "Risk Groups" explains 30% of the model variance where the proportion of the population under 5 years old is the dominant variable. This factor describes the most vulnerable population through the variable mentioned above and this FAC is also explained by the following variables: proportion of population with difficulties; proportion of students by secondary educational establishment and students by pre-school educational establishments. The FAC 1 is also composed by variables that are related with housing, namely: the proportion of rented accommodation and the proportion of seasonal housing. These characteristics are important because according with Cutter et al. (2003) and Mendes et al. (2011) the type of accommodation in which an individual resides reflects, in most cases, their personal, social and economic characteristics. The last variable in this factor is the average value of social security pensions which allows identifying economically and financially fragilized populations.

4.1.2 *Factor 2 – Economic conditions*

The factor 2 explains 13% of the variance where the proportion of employees on behalf of others is the dominant variant. This FAC is also constituted by the following variables: proportion of self-employed workers as an isolated employer; proportion of self-employed workers; persons employed in the primary sector; average value of social protection pensions; proportion of seasonal households. In this FAC it is considered that the better the economic condition, the greater the capacity to face and recover from hazardous events.

4.1.3 *Factor 3 – Disadvantaged population*

Factor 3 is related with the disadvantaged people and contributes with 12% of the model variance. The variable dominant is beneficiaries of the Social Integration Income (RSI) and Minimum Guaranteed Income (RMG). The proportion of housing units with renting below 100 euros, the proportion of buildings built before 1919 and the proportion of employed population in the primary sector are the other variables present in this FAC. This FAC represent, in the most cases, the population with low-income, low socio-professional and highly economic and social dependent on institutional aid.

4.1.4 *Factor 4 – Level of income*

Factor 4 explains 7% of the variance and is composed by the following variables: customer deposits in banks, savings banks and mutual agricultural credit, which is the dominant variable, and purchasing power ratio. This factor is related with the economic capacity of the population.

4.1.5 *Factor 5 – Employment*

This factor explains 6% of total variance and is composed by two variables: the proportion of employed population in the tertiary sector (dominant variable) and proportion of population employed in the secondary sector.

4.1.6 *Factor 6 – Employment*

Factor 6 explains 5% of the variance and is composed only by the variable proportion of social housing supported by social and supported income, being directly related with economic power of the population.

4.2 *Criticality factors' cartography*

The analysis of the factor 1 (Figure 2) shows that the highest values related with risk groups are located mainly in the municipalities of the central and inland areas of Portugal. This fact is directly related, in the most cases, with the areas where high percentages of elderly population and low percentages of young population are observed. In factor 2, a clear distinc-

tion is observed between the north and south areas (highest values) with the center region. These highest values of criticality related with low economic conditions are located, essentially, in the municipalities belonging to NUT III of Alto Tâmega, Terras de Trás-os-Montes and Douro (in the north) and in the Baixo Alentejo, Alentejo Litoral and Algarve southern areas, where the primary sector still plays a very important role in the regional economy.

There are also areas where there is an important proportion of self-employed workers as an isolated employer and the proportion of self-employed workers, mostly related with the primary sector. Factor 3 is related with disadvantaged population, and we can observe in the Figure 2 that the highest values of this factor emerge along the valley of Douro river and south of the Tagus river in municipalities with high percentages of population beneficiary of the RSI and RMG, living in low-rent housing and old buildings and work in the primary sector.

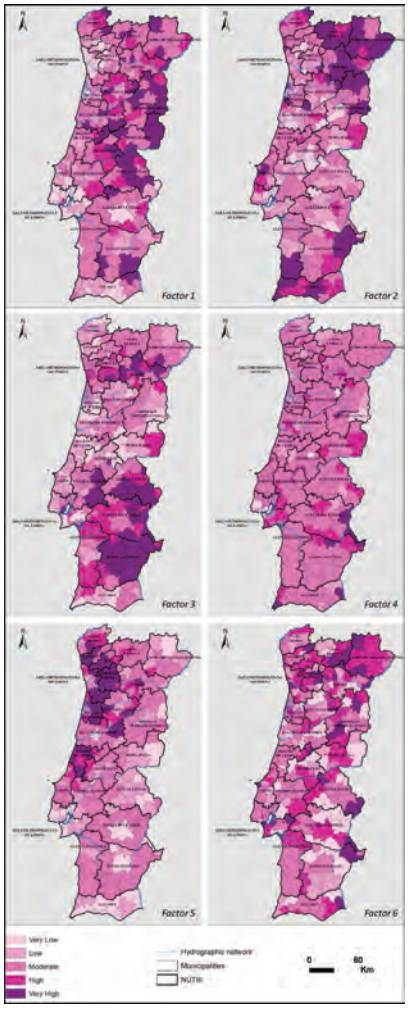


Figure 2. Cartography of the three factors that compose the criticality.

The analysis of cartography of factor 4, named level of income, allows concluding that a great territorial homogeneity exists in the different variables that compose this factor. In factor 5, related with employment, namely the population employed in secondary and tertiary sector. In this analysis we considered that employment in the secondary sector are more vulnerable. This fact is related with the predomination of small and medium enterprises, with value added (VAB) lower than tertiary sector and with greater fluctuation in productivity and employment in time of crises. The cartography identifies the highest values in the coastal northern zone of Tagus river highlighting NUT III Região de Leiria, Região de Aveiro, Área Metropolitana do Porto, Tâmega e Sousa, Cávado e Ave, which stand out as areas with strong industrial and commercial dynamism. The factor dependent population (factor 6) presents highest values north of the Tagus river, and mainly those municipalities on the right margin of the Douro river.

4.3 *Factors of support capability*

The Support Capability assessment identifies 3 factors that resulted from PCA with different percentages in the explained variance (Table 4).

4.3.1 *Factor 1 – Civil protection resources*

The factor 1 explains 30% of the total variance and is related with the municipal civil protection capability. The dominant variable is the number of fire-fighter corporations per 1000 inhabitants. The other variables of the model are: firefighters per 1000 inhabitants, average number of inhabitants per covered spaces (which represents shelter facilities), pharmacies per 10 000 inhabitants and density of road network.

4.3.2 *Factor 2 – Economic an environmental dynamic*

This factor explains 22% of the variance and is composed by the following variables: urban waste

Table 4. Support capability components.

FAC	Name	Explained variance (%)
1	Civil protection resources	30
2	Economic and environmental dynamic	22
3	Logistics and services capacity	12

collected, in kg per inhabitant, proportion of collective households, ATMs per 1000/inhabitants and accommodation capacity in hotel establishments per 1000 inhabitants, which is the dominant variable.

4.3.3 *Factor 3 – Logistics and services capability*
This factor is related with the economic dynamism and explains 12% of the variance. The dominant variable is ATMs per 1000 inhabitants. The other variables that compose the factor 3 are hospitals per 1000 inhabitants and insurance agencies per 1000 inhabitants.

4.4 *Support capability factors cartography*

Figure 3 shows the cartographic representation of each FAC expressing Support Capability.

Factor 1 is related with civil protection resources and with the analysis of the Figure 3 we can observe that the lowest values are located in the metropolitan Lisboa and Porto areas, as well as in adjacent municipalities. This factor is directly related with population density where a relatively reduced number of resources serves a greater number of inhabitants, when compared with less urbanized areas. The factor economic an environmental dynamic (factor 2) express mainly the urban character of the different municipalities. We can observe that the majority of the municipalities analyzed presents moderate values, with the lowest values principally concentrated in the northern margin of the Tagus river.

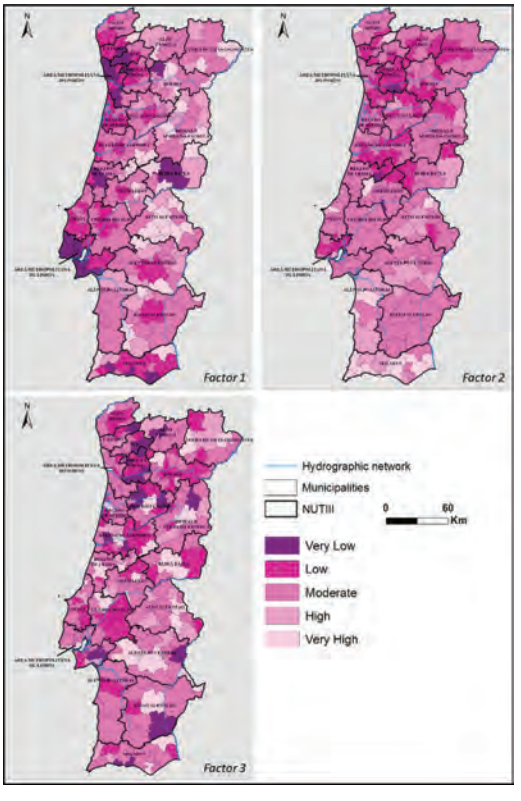


Figure 3. Cartography of the three factors that compose the suport capability.

4.5 *Criticality at municipal level*

Figure 4 presents the cartographic representation of Criticality for mainland Portugal.

We can observe that the lowest values or Criticality are mainly concentrated in the coastal area, especially in the Algarve region, and in the main regional capitals of Lisboa, Leiria, Coimbra and Porto, and their neighboring municipalities. The highest values arise predominantly at northern

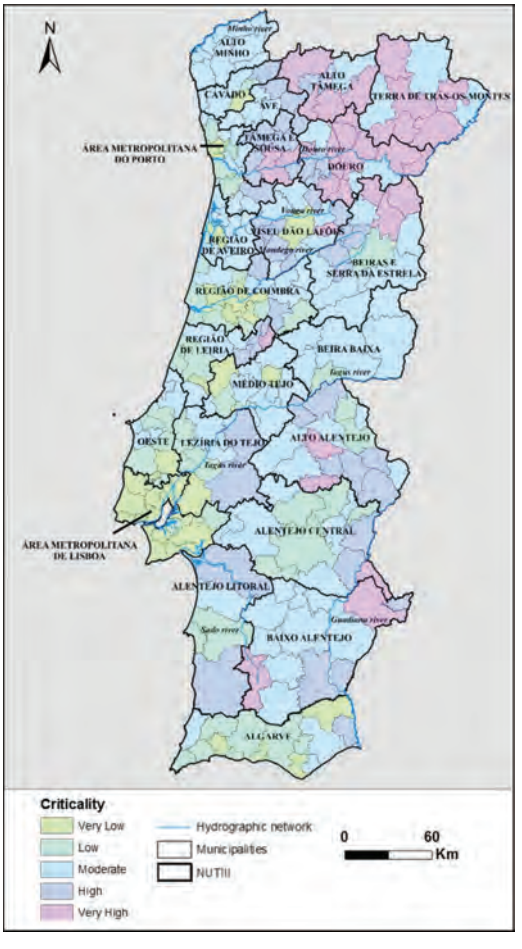


Figure 4. Criticality in mainland Portugal.

municipalities, namely in Alto Tâmega, Trás-os-Montes and along the Douro river valley. We also observe high Criticality values in the central region of Portugal, where stands out the surrounding municipalities of Viseu (located in NUTT III Viseu Dão Lafões), and along the border with Spain in the municipalities belonging to NUT III of Alto Alentejo, Alentejo Central and Baixo Alentejo.

4.6 *Support capability at municipal level*

Figure 5 shows the cartographic representation of each FAC belonging to Support Capability.

The analysis of Figure 5 allows observing that the metropolitan area of Lisboa (with the exception of the municipality of Lisboa and Oeiras) and Porto (with the exception of the municipality of Porto) presents very low and low values of Support Capability. This fact is also noted in the majority of municipalities and NUT III surrounding these areas. This fact permits to conclude that, in most cases, such low values are directly related with high population density. On the other hand, we observe that the highest values are predominantly located in the inland municipalities, especially in areas south of the Tagus river, in municipalities characterized by the availability of the resources for a small number of inhabitants.

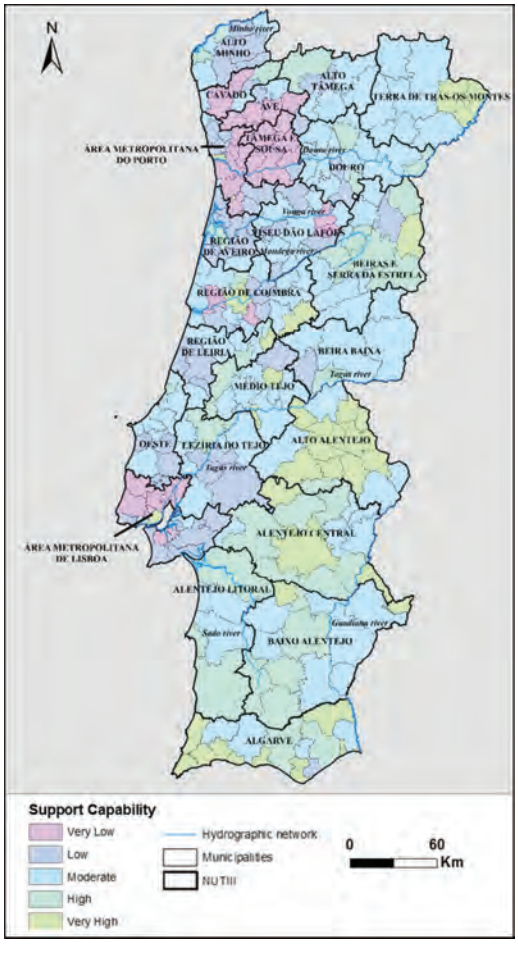


Figure 5. Support capability in mainland Portugal.

4.7 *Social vulnerability at municipal level*

The application of equation 1 that combines the Criticality and Support Capability results in the calculation of Social Vulnerability for the 278 municipalities of mainland Portugal. The analysis allows observe that the highest values of Social Vulnerability are concentrated in the northern areas, namely in municipalities located along the Douro river valley, in the region of Tâmega and Sousa, Ave, southern area of the Porto metropolitan area, Alto Tâmega, Terras de Trás-os-Montes and Viseu Dão Lafões.

In terms of lowest values we can observe that they are concentrated in areas in southern part of the country where stands outs the region of Baixo Alentejo and Algarve where the majority of municipalities has values of Social Vulnerability ranging from low to very low.

5 DISCUSSION

The analysis and evaluation of Social Vulnerability allows to conclude that we can divide, in general terms, the mainland Portugal in two areas: the area at north and the area at south of the Tagus river where the high and very high values are mainly located in the northern part. The reasons for this spatial distribution depends on several factors.

In terms of Criticality the most important factors at the municipal level are those related with the risk groups, the economic conditions and the disadvantaged population. In the total of 278 municipalities we observe that 40% of them present moderate Criticality, 30% values that varies from very low to low and 30% varying from high to very high.

About the Support Capability we can observe a relation between the highest values and the high density of population. The most important factors are associated with variables related to the civil protection resources (factor 1) and variables related to economic and environmental dynamics (factor 2). We also conclude that 39% of analyzed municipalities presents moderate Support Capability, 34% values that varies from very low to low and 27% varying from high to very high.

In terms of Social Vulnerability it possible to conclude that the final values are strongly influenced by factors related with the weak economic power of the resident population, the fragility of its economic fabric and the presence of significant percentages of dependent and disadvantaged population.

The present methodology allows compare and differentiate regions and municipalities in terms of whose characteristics of criticality, capacity of support and social vulnerability would not be evidenced in another way. The spatialization of each component and associated variables are important for the definition, application and promotion of measures related with social policies, housing, distribution and reinforcement of collective equipment, the implementation of a model of economic development more balanced in terms of employment in the inland areas and urban planning policies. The implementation and the success of this measures are important to reduce asymmetries between regions and municipalities. For the success of this measures are important promote and encourage the inter-municipal resource sharing in the sense of corresponding to the character multidimensional and multidisciplinary of Social Vulnerability and associated components.

6 CONCLUSIONS

The present work presents the calculation of Social Vulnerability for the total of 278 municipalities of mainland Portugal in accordance with the methodology presented by Mendes et al. (2011). The character multidimensional of this methodology that combine the Criticality and Support Capability allows not only the calculation of Social Vulnerability as also because of its strong territorial component, defining the Territorial Vulnerability of the analyzed areas.

The multidimensionality of this study, that is based in an extended set of variables from various dimensions like social support, housing, demography, economy, education and health allows the applicability in several risk governance dimensions. The cross-referencing of these data with existing regional or local information may result in programs that promote capacity and social cohesion. The outputs resulting from the present study allow the observation and comparison, among different places. This fact can and should be a work tool for analysis and application by different stakeholders, from multiple sectors and authorities at national, regional and local level.

The knowledge and the consciousness of the territorial distribution of Social Vulnerability and its components (Criticality and Support Capability) as well as their consideration in risk management—where spatial planning instruments are a central part of the process, is a key tool for the definition and application of multidisciplinary and multi-scale risk management strategies that not only consider the physical aspects of the territory, but all its social and institutional dimensions. In fact, the implementation of municipal and local measures that address high SV contexts would first require the existence of an adequate institutional building, drawn upon the best risk governance practices.

ACKNOWLEDGMENTS

This work was financed by national funds through FCT—Portuguese Foundation for Science and Technology, I.P., under the framework of the project FORLAND—Hydro-geomorphologic risk in Portugal: driving forces and application for land use planning (PTDC/ATPGEO/1660/2014).

REFERENCES

Angeon, V. & Bates, S. 2015. Reviewing Composite Vulnerability and Resilience Indexes: A Sustainable Approach and Application. World Development, 72: 140–162.

Barros, J.L., Tavares, A.O., Santos, A., Fonte, A. 2015. Territorial vulnerability assessment supporting risk managing coastal areas due to tsunami impact. *Water*. 7:4971–4998.

Birkmann, J. 2006. Measuring vulnerability to promote disaster-resilient societies: Conceptual frameworks and definitions, Measuring Vulnerability to Natural Hazards; Towards Disaster Resilient Societies,1:9–54.

Birkmann, J. 2013 Measuring vulnerability to natural hazards. Tóquio: United Nations University Press.

Birkmann, J., Cardona, O. D., Carreño, M. L., Barbat, A. H., Pelling, M., Schneiderbauer, S., Kienberger, S., Keiler, M., Alexander, D., Zeil, P. & Welle, T. (2013). Framing vulnerability, risk and societal responses: The MOVE framework. *Natural Hazards*, 67(2) 193–211.

Comrey, A., & Lee, H. 2009. A first course in factor analysis. New York: Physiology Press.

Cutter, S. L., Boruff, B. J., & Shirley, W. L. 2003. Social vulnerability to environmental hazards. *Social science quarterly.* 84(2): 242–261.

Fatemi, F., Ardalan, A., Aguirre, B., Mansouri, N. & Mohammadfam, I. 2017. Social vulnerability indicators in disasters: Findings from a systematic review. *International Journal of Disaster Risk Reduction,* 22: 219–227.

Fuchs, S., Birkmann, J., & Glade, T. 2012. Vulnerability assessment in natural hazard and risk analysis: current approaches and future challenges. *Natural Hazards,* 64(3): 1969–1975.

Hewitt, K. 1983. Interpretations of Calamity from the Viewpoint of Human Ecology. London: Allen and Unwin.

INE. 2012. Censos 2011 Resultados Definitivos e Portugal. Lisbon: National Institute of Statistics.

Mendes, J. M., Tavares, A. O., Cunha, L., & Freiria, S. 2011. A vulnerabilidade social aos perigos naturais e tecnológicos em Portugal. Revista Crítica de Ciências Sociais. 93: 95–128.

Mendes, J.M. 2009. Social Vulnerability Indexes as Planning Tools: Beyond the preparedness paradigm. *Journal of Risk Research,* 12:43–58.

PORDATA. 2017. Base de dados Portugal Comtemporâneo. Available at:http://pordata.pt. Acessed August, 2017.

Rufat, S., Tate, E., Burton, C. G., & Maroof, A. S. 2015. Social vulnerability to floods: Review of case studies and implications for measurement. *International Journal of Disaster Risk Reduction*, 14:470–486

Schmidtlein, M., Deutsch, R., Piegorsch, W., & Cutter, S. 2008. A sensitivity analysis of the Social Vulnerability Index. *Risk Analysis*, 28: 1099–1114.

Tapsell, S., McCarthy, S., Faulkner, H., Alexander, M. 2010. Social Vulnerability to Natural Hazards. CapHaz-Net WP4 Report, Flood Hazard Research Centre—FHRC, Middlesex University, Londres. Acessível em http://caphaz-net.org/outcomes-results/CapHaz—Net_WP4_Social-Vulnerability.pdf

UNISDR 2015. Sendai Framework for Disaster Risk Reduction 2015–2030. United Nations Office for Disaster Risk Reduction, Sendai, Japan.

Wisner, B., P. Blaikie, T. Cannon, and I. Davis 2004. At Risk: Natural Hazards, People's Vulnerability and Disasters. 2nd ed. London: Routledge.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Criticality analysis of wind turbine energy system using fuzzy digraph models and matrix method

M.K. Loganathan
School of Engineering and Technology, Kaziranga University, Jorhat, Assam, India

Indrani Bezbaurah
Mechanical Engineering, Kaziranga University, Jorhat, Assam, India

O.P. Gandhi
ITMMEC, IIT Delhi, New Delhi, India

R.C. Borah
Mechanical Engineering, Kaziranga University, Jorhat, Assam, India

ABSTRACT: Complex systems are prone to catastrophic failure as the complexity causes the system to collapse by itself. Wind energy system, which is fast growing source of electricity, is rapidly evolving into complexity and size, leading to inherently and unavoidably hazardous by their own nature. The wind turbine failures have significant impact on public health and safety risk, productivity and economy. Although design plays a major role in developing safer and reliable system, yet achieving desired operational safety and reliability remains a difficult task for the wind turbine manufacturers. In order to ensure better safety and reliability during operation, effective design and maintenance measures need to be taken. Criticality analysis of the wind turbine components or its subsystems is one way to achieve these objectives. Criticality analysis helps to identify critical failure modes or items, which in turn, assists in formulating optimal design and maintenance procedures so that better operational safety and reliability of the wind turbines can be obtained. The conventional FMECA, which is used for criticality analysis, takes care of the effect of failure on components, but does not consider the causal relations or interdependencies among failures. This paper presents an effective method of criticality analysis of wind turbine energy system using fuzzy based digraph models and matrix method by taking into account the causal relations/ interdependencies among failures. This will help to identify the critical failure modes /units of wind turbine energy system. The proposed method is useful in criticality assessment of wind turbines in design as well as in operation stages.

1 INTRODUCTION

Wind energy is a clean and renewable that offers several advantages. In order to capitalize on it, the economically leading countries have been harvesting wind energy over past many decades [1]. However, the efficiency of the wind turbine system is limited by the failures of its elements [2–5]. The critical issues related to reliability and maintenance of wind energy systems have not been addressed fully and these still remain big challenges in operating and maintaining the wind power system [6]. The large wind energy system has numerous elements at its various hierarchical levels, and each elements follow different failure patterns [7–8]. This makes the system more complex and thus leading to hard to access faulty units, difficulty in maintenance and poor reliability. The operations and maintenance cost, which are directly affected by unreliability of the elements of the wind turbine, can be reduced when the whole system continue to function without failure, or with reduced failure rate [9]. It is, therefore, crucial to achieve and sustain high operational reliability. Although, several researchers have studied reliability aspects of wind turbine systems [10–13], yet there are limited findings on how to evaluate the criticality of the elements of wind turbine with respect the failure characteristics such as failure patterns, failure interdependencies, etc. Tools like FMEA may be good in assessing criticality, but lacks in expressing failure dependencies [14]. In reality, the failures, which are the loss of functions, are interrelated, i.e., the failure (i.e., deteriorated function) of one unit does have causal relation with the failure of other unit, which is connected through physical structure [15]. The studies, which ignore this, will provide inaccurate results. It is, therefore,

essential to consider the failure interdependencies through structure while evaluating criticality. The well known structure models; Graph/digraph models and matrix methods, which are effective to model failure dependencies [15–16], are extensively employed for various engineering applications. Application of these tools in criticality analysis of wind turbine energy systems is not yet explored. It is found that FMEA based models have been used in recent years for criticality assessment of wind turbine systems [14], but these are limited to modelling of failure modes, and have not considered the failure interdependencies. Several studies have been carried out on solving engineering problems in a fuzzy environment using fuzzy theory [17]. There are not many literatures found in the application of fuzzy theory in the failure studies of wind turbine systems. This paper presents a methodology of criticality analysis of wind turbine energy system based on fuzzy decision making and digraph model and matrix methods. The method is effective in indentifying critical units in wind turbine energy system taking into account the failure modes and their interdependencies.

2 METHODOLOGY

The proposed method is based on structural models i.e., digraph models and matrix, and fuzzy decision making methods. The selected wind turbine energy system is first described for obtaining structural knowledge, which helps to understand various units, and their interconnections. The common, but important failure modes and their causal relations/interdependencies for the units are then identified and represented using digraph model, and the resultant digraph, which is known as causality digraph, will be converted into causality matrix using matrix method. The matrix will be further processed for causality analysis and evaluation of criticality index for each unit of wind turbine system. The unit, which is having high criticality index, will be considered to be critical unit that will contribute to major breakdown and safety issues of the wind turbine energy system, whereas the unit with low criticality index, will be considered as low critical unit. In this way, the criticality of the wind turbine energy system is evaluated and critical units are identified. Subsequently, appropriate preventive/corrective action will be taken.

3 SYSTEM DESCRIPTION

A typical horizontal wind turbine energy system, which is shown in Figure 1, is considered for the analysis.

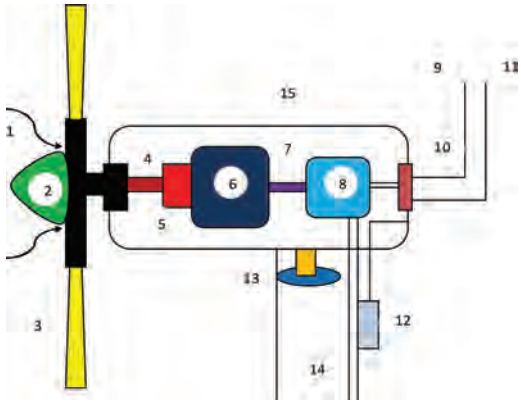


Figure 1. A typical horizontal wind turbine energy system.

The system consists of the following units, i.e., components and subsystems;

1. **Wind (fluid component)** – This sweeps over the turbine blades at high velocity and impart force on the blades to rotate.
2. **Nose and rotor hub** – The aerodynamic design coupled with rotor hub, streamline and distribute the wind over blades.
3. **Rotor blades** – These are attached to the nose and rotor, and spins at ample wind velocity.
4. **Drive train** – This is the combination of main turbine shaft and support bearing mechanism, which connects the blade with gear box, and transfer the rotational energy to the gear box, and then to the generator.
5. **Mechanical brake** – This is used to stop the turbine in order to prevent mechanical failures of the turbine components from high wind speed.
6. **Gear box** – This is used to increase the rotational speed of the turbine shaft with varied torque.
7. **High-speed turbine shaft** – This is the part of the drive train, which connects gearbox and generator.
8. **Generator** – This converts mechanical energy from gear box into electrical energy.
9. **Wind speed sensor** – It measures wind speed.
10. **Electronic control system** – When the wind speed is undesirably very high, the controller gets command from the wind sensor to stop the rotating drive elements through mechanical brakes. It also helps to re-start the rotation at low wind speed.
11. **Wind direction sensor (wind vane)** – This measures the direction of the wind, and sends the command signal to yaw drives to adjust the facing of the turbine with respect to wind direction.

There are many other types of sensor used for measuring position, and speed of all rotating elements. These are not shown in Figure 1.

12. **Yaw drive** – It receives the command signal from wind vane and rotates the turbine.
13. **Yaw motor (Hydraulically operated)** – It physically rotates the turbine based on the instructions from the yaw drive.
14. **Supporting parts (tower)** – It supports the entire wind turbine system high in the air. This also encloses the complete electrical wiring systems (Not shown in Fig. 1).
15. **Nacelle (Housing)** – This encloses entire turbine units.
16. **Electrical system (**Not shown in Fig. 1) – This consists of inverter to steady the output variables; current, and voltage, and transformer to raise or lower the voltage in AC transmission line.
17. **Hydraulic system (**Not shown in Fig. 1**)** – This consists of hydraulic pump, control valves, hydraulic motor and actuators. This is used to actuate yaw mechanism.

In order to develop causality digraph, most common failure modes of the units and their interdependencies are identified. This is discussed in the following line.

4 FAILURE DATA COLLECTION AND FAILURE MODE IDENTIFICATION

From the literature and extensive interaction with wind turbine manufacturing/installation company personnel, failure data were collected for wind turbine energy system. The percentage of failures for each unit of the system is given in Figure 2 and the down time per failure in terms of days for all units, is represented in Figure 3.

The collected failures were categorized into four major failure modes, which are listed below;

1. Surface crack, rupture and overloading (F_s)
2. Component looseness (F_l)
3. Circuitry failure (F_c)
4. Fuse blown (F_b)

The above failure modes contribute to the majority of the failures in the wind turbine energy system. This is shown in Figure 4.

These failure modes are causally interrelated/ interdependent. This means that one failure mode may be the result of the occurrence of the other failure mode, and vise versa. The failure modes and their interdependencies are presented in graphical representation by developing digraph models. This is described below.

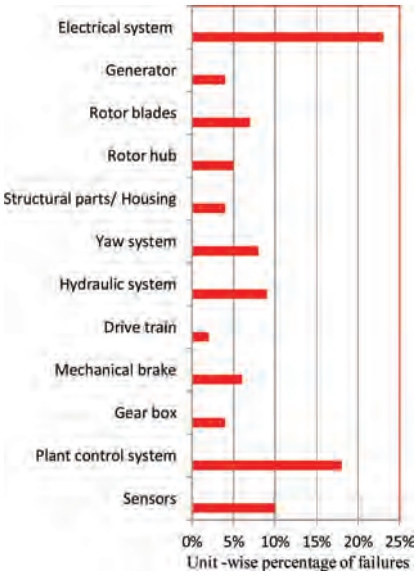


Figure 2. Percentage of failures of wind turbine energy system.

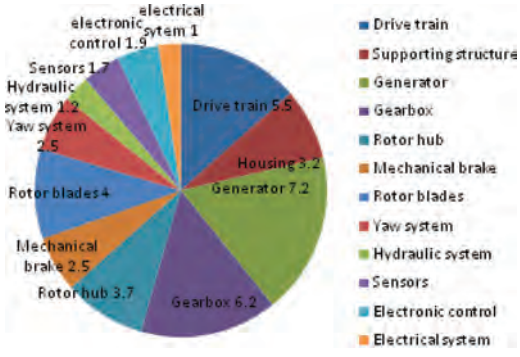


Figure 3. Down time per failure (Days).

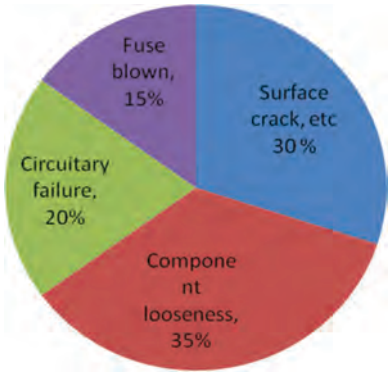


Figure 4. Percentage of important failure modes of wind turbine energy system.

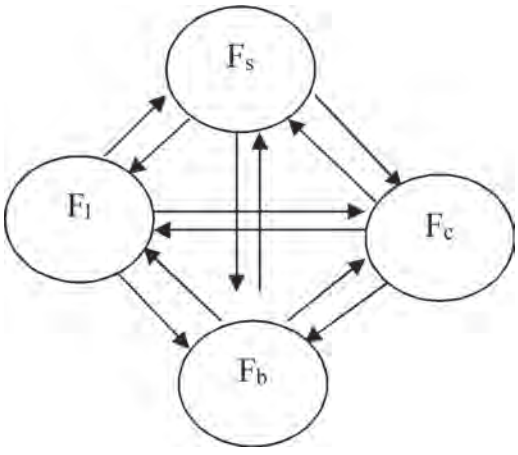


Figure 5. Causality digraph for wind turbine energy system.

5 CAUSALITY DIGRAPH

The causality digraph, $G_{CD} = (V_F, E_F)$ for wind turbine energy is developed with, nodes, 'V_F' representing the failure modes and the edges 'E_F' representing the causal relation/interdependencies among the four major failure modes indentified under Section 4. The developed causality digraph is shown in Figure 5.

The causality digraph helps to perform visual analysis. In order to carry out criticality analysis of failure modes and criticality evaluation of the units of wind turbine energy system, the causality digraph will be converted into an equivalent matrix by using matrix method. This is described in the subsequent section, i.e. Criticality Evaluation.

6 CRITICALITY EVALUATION

In this section, the criticality analysis and evaluation is performed by defining equivalent matrix for the developed causality digraph of the wind turbine energy system. As discussed under Section 5, the causality digraph represents the four major failure modes and their causal relations. If there are a large number of failure modes, then the digraph will become complex. It will be difficult to perform the visual analysis of the digraph. Moreover, the quantification of severity of failure modes and their causal relation is necessary to identify the critical element of the wind turbine system. It is, therefore, essential to convert the digraph into equivalent matrix for further processing. The equivalent matrix, which is known as 'Wind Turbine Criticality Matrix' (WTCM), E_{CM}, is written as;

$$E_{CM} = \begin{array}{c} \begin{matrix} \text{s} & \text{l} & \text{c} & \text{b} & \text{Failure mode} \end{matrix} \\ \begin{bmatrix} \text{F}_\text{s} & f_{\text{sl}} & f_{\text{sc}} & f_{\text{sb}} \\ f_{\text{ls}} & \text{F}_\text{l} & f_{\text{lc}} & f_{\text{lb}} \\ f_{\text{cs}} & f_{\text{cl}} & \text{F}_\text{c} & f_{\text{cb}} \\ f_{\text{bs}} & f_{\text{bl}} & f_{\text{bc}} & \text{F}_\text{b} \end{bmatrix} \begin{matrix} \text{s} \\ \text{l} \\ \text{c} \\ \text{b} \end{matrix} \end{array} \quad (1)$$

where the diagonal elements represent the failure modes (F_s, F_l, F_c, F_b), while the off-diagonal elements represent the causal relation/interdependency between the failure modes, *i* and *j* (fij, *i*, *j* = s, l, c, b; fij≠fji). The WTCM is analogous to permanent matrix in the graph theory. It is mentioned that the permanent is a standard function that is used in combinatorial mathematics [16]. Permanent of WTCM is criticality expression for the wind turbine energy system. The 'Wind Turbine Criticality Expression' (WTCE), which represents severity of failure modes and their causal relations/interdependencies from combinatorial consideration, is obtained from its WTCM (eqn. (1)), as;

$$\begin{aligned} &P(E_{CM}) \\ &= F_s F_l F_c F_b + f_{ls} f_{sl} F_c F_b + f_{cs} f_{sc} F_l F_b + f_{bs} f_{sb} F_l F_c \\ &\quad + f_{lc} f_{cl} F_s F_b + f_{cb} f_{bc} F_s F_l + f_{lb} f_{bl} F_s F_c + f_{lc} f_{bl} f_{cb} F_S \\ &\quad + f_{lb} f_{cl} f_{bc} F_s + f_{sl} f_{lc} f_{cs} F_b + f_{sl} f_{lb} f_{bs} F_c + f_{sc} f_{ls} f_{cl} F_b \\ &\quad + f_{cb} f_{bs} f_{sc} F_l + f_{sb} f_{ls} f_{bl} F_c + f_{bc} f_{cs} f_{sb} F_l + f_{sl} f_{ls} f_{bc} f_{cb} \\ &\quad + f_{sl} f_{lc} f_{bs} f_{cb} + f_{sl} f_{lb} f_{bc} f_{cs} + f_{sc} f_{ls} f_{bl} f_{cb} + f_{sc} f_{lb} f_{cs} f_{bl} \\ &\quad + f_{sc} f_{lb} f_{bs} f_{cl} + f_{sb} f_{ls} f_{cl} f_{bc} + f_{sb} f_{lc} f_{cs} f_{bl} + f_{sb} f_{lc} f_{bs} f_{cl} \end{aligned} \quad (2)$$

The WTCE (eqn. (2)) helps to carry out criticality analysis from combinatorial considerations as it takes care of the severity of failure modes and all possible causality relations among failure modes. The WTCE, which is the characteristic of the causal relations between failure modes of wind turbine energy system, contains number of terms. Each term in the expression has physical meaning. The first term in the expression represents the severity of the four major failure modes. Each term, from second to seventh, represents a two-failure mode causality loop and the severity of two failure modes, and each term from eighth to fifteenth represents three-failure mode causality loop and the severity of three-failure modes. Each of the last nine terms represents severity of causal relations between all major failure modes. By substituting the severity value of failure modes and their causal relations, one can carry out not only the criticality analysis of wind turbine, but also the evaluation of its criticality index.

For performing criticality analysis, each term is examined for severity. The first term in the expression contains the combination all failure modes; $F_s/F_l/F_c/F_b$, which represents the severity of

all major failure modes. By examining this term, the design or maintenance engineer will be able to identify critical failure modes or critical causal relations between failure modes. This will prompt the engineers to take appropriate corrective action. For example, if the severity of the failure mode, F_l, component looseness is higher in electrical subsystem, then an additional care must be taken to minimise its severity or remove the failure mode. In the similar way, other terms are used for criticality analysis, and appropriate preventive or corrective actions can be taken. In order to evaluate criticality index of the wind turbine energy system, the severity values of failure modes and that of failure interdependencies are substituted in WTCE, which will then be solved to obtained numerical index that represents the severity of the failure modes of the wind turbine energy system. The main objective of the work is to evaluate criticality index at subsystem level, so that the critical subsystem can be selected and ranked based on the criticality index. It is recommended that the severity value should be obtained from wind turbine shop-floor data base or experienced wind turbine service/ operational personnel. However, the severity of the failure modes and their causal relations can also be proposed based on the field experience as there is no data source that provides the severity data for the wind turbine components and no commonly accepted method available as well. The failure modes indentified in Section 4, interact with each other, making causal relations at varying degrees. Depending upon the severity of the failure modes and the degree of their interaction or influence of one failure mode on the other, appropriate severity rating may be selected for each failure mode and their causal relations. In the present work, the data of severity rating are obtained by making extensive interaction with service engineers and operational personnel of wind turbine manufacturing industries. Some of the technical report and literatures [1–10] are also useful in this regard. It is to be mentioned that obtained data is the severity ratings, which are in the form of qualitative terms, e.g. Low, Average, or High. These terms will be converted into quantitative values, i.e., severity value, for evaluating criticality index.

In order to convert the severity rating into severity values, an appropriate and accurate method needs to be selected. In this work, a fuzzy theory based approach has been chosen for converting severity rating into severity values. The severity values of the identified failure modes and their causal relation are substituted in eqn. (2), i.e., WTCE, to obtain criticality index of the wind turbine energy system, at its various subsystem level. Based on the criticality index, the critical subsystem can be selected. The following section will describe the conversion of severity rating into severity values using fuzzy theory, and the evaluation of criticality index of some selected units of wind turbine energy system.

7 FUZZY THEORY IN CRITICALITY EVALUATION

Fuzzy theory provides a tool for directly manipulating the linguistic terms that an analyst employs in making a criticality assessment for a failure modes, effects and criticality analysis (FMECA). In the proposed approach, these parameters, i.e. severity rating (linguistic term) are represented as members of a fuzzy set, fuzzified by using appropriate membership functions and are evaluated in fuzzy inference engine, which makes use of well-defined rule base and fuzzy logic operations to determine the crisp score, which represents the criticality/riskiness level of the failure. This means that the method will convert linguistic term into fuzzy numbers and the fuzzy numbers into crisp scores. The higher the value of crisp score, the greater will be the risk and lower the value of crisp score, and the lesser will be the risk.

To demonstrate the method, a 5-point scale having the linguistic terms low, below average, average, above average, and high is considered (Refer Fig. 6). The linguistic terms are converted into fuzzy number, and then to crisp score.

The crisp score of fuzzy number 'M' is obtained as follows:

$$\mu_{\max}(x) = \begin{cases} x, 0 \le x \le 1 \\ 0, otherwise \end{cases} \tag{3}$$

$$\mu_{\min}(x) = \begin{cases} 1-x, 0 \le x \le 1 \\ 0, otherwise \end{cases} \tag{4}$$

The linguistic terms with their corresponding crisp scores are chosen from Figure 6 and given in Table 1. The crisp score represent the severity value of failure modes, and that of causal relations.

The severity value is selected from Table 1, based on the degree of severity for each of the failure modes and their causal relations. By substituting these values in eqn. (2), the criticality analysis is carried out and subsequently criticality index for the wind turbine energy system is obtained at various subsystem levels. For example, let us examine the severity of the major failure modes for drive train unit for criticality analysis. From Table 1, the severity values for all major failure modes, i.e., $F_v/F_l/F_c/F_b$ of the drive train unit are taken as; 0.695/0.895/0.495/0.295 respectively, and substituted in eqn. (2) in place of the first term.

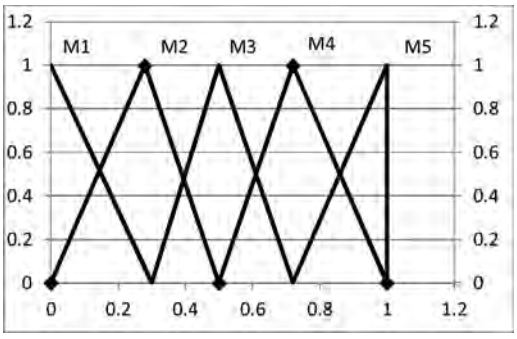


Figure 6. Linguistic terms to fuzzy numbers conversion (5-point scale).

Table 1. Conversion of linguistic terms (Severity rating) into crisp score (Severity value).

Linguistic term (Severity rating)	Fuzzy number	Crisp number (Severity value)
Low	M1	0.115
Below average	M2	0.295
Average	M3	0.495
Above average	M4	0.695
High	M5	0.895

It is observed that the failure mode, component looseness, F_l is more severe, and suitable corrective action needs to be identified and implemented to maintain or redesign the drive train keeping in mind the severity of component looseness. On the similar line, the severity values of the failure modes and their causal relations are substituted in place of remaining terms and the criticality analysis is performed for the drive train unit. Subsequently, the eqn. (2) will be solved to obtain criticality index for the drive train unit, and obtained as ***0.6741***. Similarly the criticality analysis is carried out to identify critical failure modes/critical causal relations between failure modes for the remaining units by choosing appropriate severity values from Table 1. In the example discussed here, seven units (i.e., subsystems) such as; *drive train, generator, electronic control system, yaw drive, electrical system, hydraulic system and gear box*, are considered for evaluating criticality index. The criticality index for the all subsystems are evaluated and presented in Table 2. The electrical system is identified as critical unit as it has highest value of criticality index, which is ***2.2314***; it is, therefore, obvious that the electrical system is assigned criticality rank as 1. The electrical system of wind turbine energy system requires more attention in terms of, for example, improved design, reliability, maintainability, safety, etc. The

Table 2. Subsystem-wise severity values, criticality index, and criticality rank.

	Severity value of the failure modes				Severity value of causal relations among failure modes													
Sub-system	F_s	F_L	F_C	F_B	f_{1b}	f_{b1}	f_{1s}	f_{s1}	f_{sc}	f_{cs}	f_{cb}	f_{bc}	f_{1c}	f_{c1}	f_{sb}	f_{bs}	CI	CR
Drive trains	0.695	0.895	0.495	0.295	0.695	0.115	0.895	0.895	0.295	0.115	0.295	0.295	0.695	0.115	0.695	0.115	0.6741	5
Generator	0.695	0.695	0.695	0.695	0.895	0.115	0.695	0.295	0.695	0.295	0.695	0.695	0.695	0.115	0.895	0.495	1.9953	2
Electronic control system	0.295	0.495	0.895	0.695	0.495	0.295	0.295	0.295	0.495	0.295	0.695	0.895	0.895	0.495	0.695	0.495	1.6772	3
Yaw drive	0.115	0.295	0.695	0.495	0.695	0.115	0.895	0.695	0.295	0.115	0.295	0.295	0.695	0.115	0.695	0.115	0.5422	6
Electrical system	0.695	0.895	0.295	0.115	0.895	0.295	0.295	0.295	0.695	0.295	0.695	0.895	0.895	0.495	0.695	0.495	2.2314	1
Hydraulic system	0.695	0.895	0.495	0.295	0.495	0.115	0.695	0.495	0.495	0.495	0.495	0.495	0.495	0.115	0.295	0.295	0.9383	4
Gear box	0.695	0.895	0.115	0.115	0.695	0.115	0.895	0.895	0.295	0.115	0.295	0.115	0.695	0.115	0.695	0.115	0.2244	7

CI – Criticality Index; CR – Criticality Rank.

design activity should be oriented towards removing the failure modes, reducing the probability of occurrence, and minising the severity of the failure. Similarly, other units of the wind turbine energy system are prioritized based on the criticality index, and suitable preventive or corrective actions are taken, be it in design or operating stage.

8 CONCLUSION

In this paper, a methodology for criticality analysis and evaluation of criticality index for wind turbine energy system at various subsystem level has been proposed. Criticality index for various units of the wind turbine energy system has been evaluated using fuzzy diagraph models. Criticality index is useful in ranking, and identifying the critical wind turbine units, which require immediate maintenance action, redesign or modification.

This method will be helpful for the system safety and reliability analysts in taking corrective action during design and operational stage of complex wind turbine energy system.

REFERNCES

[1] *http://www.gwec.net* [Accessed 27 June 2017].

[2] Reder, M.D., Gonzalez, N. and Melero, J.J. 2016. Wind turbine failures—tackling current problemsin failure data analysis. *Journal of Physics*: Conference Series [DOI:10.1088/1742–6596/753/7/072027].

[3] Gowdar, R.D. and Mallikarjune Gowda, M.C. 2016. Reasons for wind turbine generator failures: a multicriteria approach for sustainable power production. *Renewables: Wind, Water, and Solar* [DOI: 10.1186/s40807–016–0029–1]

[4] Alewine, K. and Chen, W. 2011. A review of electrical winding failures in wind turbine generators. *IEEE Electrical Insulation Conference*, 5–8 June, Annapolis, MD, USA.

[5] Nivedh, B.S. Major failures in the wind turbine components and the importance of periodic inspections. *www.windinsider.com* [Accessed 29 June 2017]

[6] Pardalos, P.M., Rebennack, S., Pereira, M.V.F., Iliadis, N.A. and Pappu, V. (eds.). 2013. *Handbook of Wind Power Systems*. Springer-Verlag Berlin Heidelberg.

[7] Jain, P. 2011. *Wind energy engineering*, McGraw-Hill, Inc. United States.

[8] Johnson, G.L. 2006. *Wind energy systems*. Kansas State University. Manhattan, USA.

[9] Hill, R.R., Stinebaugh, J.A., Briand, D., Benjamin, A.S. and Linsday, J. 2008. Wind turbine reliability: A database and analysis approach. *Sandia Report*.

[10] Hahn, B., Durstwitz, M. and Rohrig, K. 2006. Reliability of wind turbines—Experiences of 15 years with 1500 WTs. *German Wind Energy Report*, ISET, Kassel.

[11] Guo, H., Watson, S., Tavner, P. J. and Xiang, J. 2009. Reliability analysis for wind turbines with incomplete failure data collected from after the date of initial installation. *Reliability Engineering and System Safety* 94 (6): 1057–1063.

[12] Tavner, P. J., Xiang, J. and Spinato, F. 2007. Reliability analysis for wind turbines. *Wind Energy* 10 (1): 1–18.

[13] Arabian-Hoseynabadi, H., Oraeea, H. and Tavner, P. J. 2010. Failure modes and effects analysis (FMEA) for wind turbines. *Int J Electr Power Energy Syst.,* 32: 817–824.

[14] Loganathan, M.K., Minu Shikha Gandhi, and Gandhi, O. P. 2015. Functional cause analysis of complex manufacturing systems using structure. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 229(3): 533–545.

[15] Loganathan, M.K and Gandhi, O.P. 2015. Reliability evaluation and analysis of CNC cam shaft grinding machine. *Journal of Engineering, Design and Technology* 13 (1): 37–73.

[16] Venkata Rao, R. 2007. *Decision making in the manufacturing environment*, Springer-Verlag London Limited.

[17] Chaturvedi, S. K., and Gargama, H. 2011. Criticality assessment models for failure mode effects and criticality analysis using fuzzy logic. *IEEE Transactions on Reliability* 60 (1):102–110.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Hazard identification for a dynamic positioning and mooring system in Arctic condition: Complementary use of hazard identification study (HAZID) and Systems Theoretic Process Analysis (STPA)

T. Joung
Korea Research Institute of Ship and Ocean Engineering (KRISO), Daejeon, South Korea

H. Kim
Norwegian University of Science and Technology, Trondheim, Norway

Y. Kim, S. Cho & K. Kang
Korea Research Institute of Ship and Ocean Engineering (KRISO), Daejeon, South Korea

Y. Liu & M.A. Lundteigen
Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: Dynamic Positioning (DP) control system is currently utilized for maintaining the position of Floating Production Storage and Offloading (FPSO) in a variety of the ocean environment conditions. However, FPSO in the Arctic region, which operates under diverse ice conditions, should be designed with an advanced mooring system together with the DP system as the FPSO needs safer system compared with the normal DP systems. A novel DP and mooring system for the FPSO operating in the Arctic region is now being developed at KRISO in Korea. A platform shape of the FPSO that has minimum resistance and maximum operation efficiency is also under development. Therefore, analyses and assessments of potential risks are required, considering that the developing system has many novelties compared to the conventional DP or mooring system. Hazard identification study (HAZID) is one of the most widely used traditional methods to identify hazards of a system, and the method is simple to use and requires limited training. System Theoretic Process Analysis (STPA), on the other hand, was recently developed for modern complex control systems and provides systematic analysis procedure based on systems theory. The main objectives of this study are to suggest an approach for hazards identification associated with the design and operation of DP and mooring system, by utilizing complementary hazard identification methods (HAZID and STPA). HAZID has been applied to the structural part of the system, while STPA has been applied to the control system. The paper also includes a comparison of the strengths and weaknesses of the selected methods and a discussion of complementary use of HAZID and STPA.

1 INTRODUCTION

1.1 *Background*

It takes almost a decade to make the first oil production since finding the oil in the Northern Alaska in 1969. Various R&D for the effective and safe operation of offshore plants were conducted under the extremely difficult environment conditions in the Arctic region such as low temperature, iceberg etc. To overcome the risk factors in Arctic sea, a variety of fixed type offshore platforms were considered so far. However, the exploitation of natural resources is taking place at deeper waters in the Arctic region, so floating type offshore plant platforms are considered.

The most important factors for design and operation of floating type offshore plant in the Arctic area is that various techniques are required to keep the stable operation and production under the condition of a verity of ice conditions, e.g. icebergs, ice floe (drift ice) etc. It is well known that the Arctic conditions are much harsher/tougher environment conditions for the safety of the operating vessel, compared to the normal sea ones when considering polar lows, sea ice, low temperature, uncertainty of metocean data etc.

The improved Dynamic Positioning (DP) system and more stable mooring system is required to keep the operation positioning in the Arctic region. The DP system will help ice management during the operations when the ice floes are approaching, and

the mooring lines make the offshore plants more stable.

The equipment with the advanced technologies are installed on the offshore plant but there are still hazards (risks) during the operation. For example, if ice management is impossible due to an extremely large iceberg, the offshore plant should be removed from the operation place to the safer place based on a prepared manual. Therefore, a variety of scenarios should be prepared considering many different risk circumstances.

To identify hazards and hazardous situations that may occur in the operation of DP and mooring systems, it is important to carry out a systematic hazards identification. Several *hazard identification methods* have been developed, see e.g. Rausand (2011). One of the most widely used hazard identification methods is HAZard IDentification (HAZID). The Preliminary Hazard Analysis (PHA), a variant of HAZID, was developed by the U.S. Army (MIL-STD-882D) to evaluate hazards early in the life of a process, and has been successfully used for safety analysis of machinery and process plants (Rausand, 2011, CCPS, 2011). HAZID is a rather simple and versatile technique that requires limited training and can cover a range of safety problems (Rausand, 2011). However, this method is a brainstorming-oriented technique, with support of checklists of known hazards, and the results are strongly dependent on knowledge and expertise of the analysts (Molland, 2011). This is a particular concern when the systems increase in complexity, and have unidentified behavior due to extensive use of ICT technologies. These are typical attributes of e.g. the DP system. Dedicated hazards identification methods have therefore been developed to handle more complex, software-intensive, sociotechnical system (Leveson, 2012). One such example is the Systems-Theoretic Process Analysis (STPA).

STPA is a further development within the framework of Systems-Theoretic Accident Model and Processes (STAMP) causality model (Leveson, 2012). STPA can identify the hazards covered by traditional hazard identification methods, but it also can identify additional hazards that are not included or poorly handled in the traditional methods, like software errors, component interactions, complex human errors, and so on (Leveson and Thomas, 2013). STPA identifies unsafe control actions, using a systematic analysis of a functional control structure of the system (Leveson and Thomas, 2013). However, STPA may not necessarily be suited for analysis of systems comprising only passive component with no control action.

1.2 *Literature study*

DP and mooring system in the Arctic Ocean are seldom applied. Therefore, the DP and mooring system which have been applied and operated in normal sea areas to keep the operating position are investigated first. The DP system is closely related with the control system, so the DP system is researched for STPA and mooring system is researched for HAZID in this study.

The most frequent accidents come from mooring line failure during the operating offshore plant (FPSO) between 2001 and 2013. That is, the number of accident cases is thirteen (13) times, and other cases are ten (10) times from the chain failure and five (5) times from wire rope disconnection. In particular, accidents from the mooring line connection was the main cause of the catastrophic disaster (Offshore Magazine, 2013).

Ma et al. (2013) and Kvitrud (2014) have researched about the accident causes of the mooring system and DP system in offshore structure and found that the top chain, wire rope terminations and connectors of mooring system are the major reasons. Several incidents have been reported for floating structure in North Sea, and the main causes of accidents are the mooring lines failures which are overloaded during extreme weathers (Kvitrud, 2014). There are several QRA techniques on DP systems during drilling operations in the Arctic or normal sea condition (Pedersen, 2015, Team Energy Resources Limited, 2002), but HAZID work applied for DP and mooring system are extremely seldom.

Only limited number of studies have applied STPA to DP systems. Abrecht (2016) conducted STPA analysis for a DP system of an offshore supply vessel. After identifying unsafe control actions (UCAs) and casual scenarios of identified UCAs, the study also conducted traditional hazard identification analyses, such as Fault Tree Analysis (FTA) and Failure Modes and Effect Analysis (FMECA), for the DP system and compared the results to discuss advantages of STPA. Rokseth et al. (2017) also analyzed a typical DP system using STPA and discussed applicability of STPA compared with FMECA. However, there is no previous study that applied STPA to DP and mooring system.

1.3 *Objectives*

The main objectives of this study are to suggest an approach for hazards identification associated with the design and operation of DP and mooring system, by utilizing complementary hazard identification methods (HAZID and STPA). HAZID has been applied to the structural part of the system (e.g. hull structure, mooring lines, turret system etc.), while STPA has been applied to the control system (e.g. DP systems etc.). The paper also includes a comparison of the strengths and weaknesses of the selected methods and a discussion of complementary use of HAZID and STPA.

1.4 *Structure of the paper*

The remainder of this paper is organized as follows: DP and mooring system is introduced in Section 2. Section 3 and 4 analyze hazards of DP and mooring system using HAZID and STPA respectively. Finally, results and discussion are presented in Section 5.

2 INTRODUCTION TO ARC7 PROJECT AND DP AND MOORING SYSTEM

Korea Research Institute of Ships and Ocean Engineering (KRISO) has initiated a five years long term research project (ARC7 project) to develop a hull form design for a year-round floating type offshore structure in the Arctic condition with DP and mooring system. In order to design an offshore structure hull form for the given operating condition in Arctic region (ARC7 condition[1]), an ice performance evaluation methodology which uses KRISO's ice tank and numerical analysis methods have been developed. In this research, Ice load estimation, hull form design, configuration design of mooring & DP systems are considered as core technologies.

The aim of the research project, ARC7 project, is design hull form for the offshore plant structure in the condition of the ice sea condition and over 200m depth. The safety of the designed system, DP and mooring system for station keeping in ice condition should be proven, which is the proper one for the requirement of the aim of the project.

The designed FPSO systems are applied for the offshore plant structure, which is a ship-type planform as shown in Figure 1. The developing systems and optimized hull form are designed for the minimized ice drag force and maximized operational efficiency.

Thrusters for the DP system and mooring lines for the mooring systems are developed for the different ice conditions based on the ice management scenarios. Concept of DP and mooring system with ice management in the Arctic ice sea is shown in Figure 2.

Design for the DP and mooring system should be considered together with the hull form of the offshore plant. The design process follows design spiral to find best compromise with in the given boundaries such as environment conditions, cost, rules (regulations) etc. Developed design spiral for the offshore plant structure with the DP and mooring system in Arctic Ocean is shown in Figure 3.

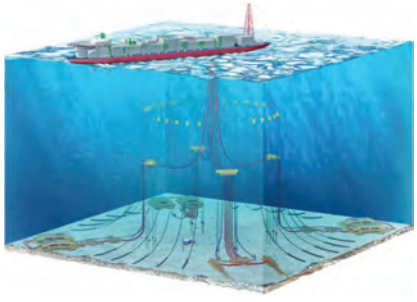


Figure 1. Configuration of offshore plant system in the Arctic region (Kim et al., 2017).

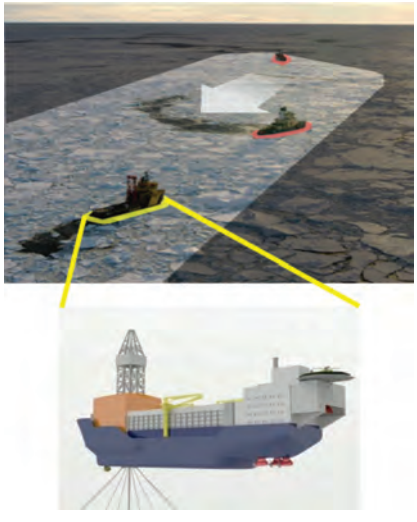


Figure 2. Concept of DP and mooring system with ice management in Arctic sea (Keinonen et al., 2006).

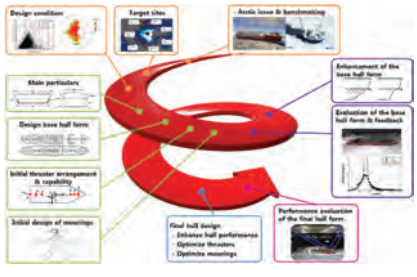


Figure 3. Design procedure (design spiral) of the offshore plant structure with the DP and mooring system in Arctic sea.

[1]ARC7: One of the ice class rule of the Russian Maritime Register of Shipping (RMRS). The ice classes are divided to non-Arctic, Arctic and icebreaker classes. The ice class notation is followed by a number which denotes the level of ice strengthening: Ice1 to Ice3 for non-Arctic ships, Arc4 to Arc9 for Arctic ships, and Icebreaker6 to Icebreaker9 for icebreakers. These ice classes can be assigned in parallel with the Finnish-Swedish ice class and/or the IACS Polar Class, provided the vessel complies with all applicable rules. The selection of ice class is based on the operating area in the Russian Arctic, time of year, ice conditions, operating tactics, and whether the vessel operates under icebreaker escort or independently (RMRS, 2017).

3 HAZID FOR THE DP AND MOORING SYSTEM

3.1 *HAZID workshop for the DP and mooring system in the arctic region*

HAZID work has been carried out for "the year-round floating offshore structure hull form development which has DP system and mooring system as the mean of the station-keeping in the condition of ARC7". The HAZID workshop was performed for the FPSO's concept design (hull form, DP system, mooring system, turret system etc.) for operating in the Barents Sea.

Hazards and risks during the installation, operation, and maintenance in the Arctic sea are identified, and the causes and consequences of the systems are discussed in the HAZID workshop.

Preventive safeguard and mitigating safeguard to prevent or minimize for the identified hazards (risks) are suggested in the workshop, and HAZID report/worksheet were prepared

3.2 *HAZID results*

The hazards (risks) in the three nodes (hull, mooring and turret systems) for the failure of structure and systems were identified in the HAZID workshop.

Main causes and consequences for the hull structural failure were identified, and the preventive/mitigating safeguard were suggested during the HAZID workshop.

Main causes of the hull structure are collision, green water, wind, wave, current, sea ice and low temperature. The mooring system failure came from improper connection between mooring lines, excessive tension, green water, wind, wave, current, sea ice, low temperature, system malfunction etc. Main causes of the turret system failure were, on the other hand, identified as collision, green water, environment (wind, wave, current), low temperature etc.

3.3 *Recommendation and guideline*

The identified hazards or risks from the HAZID workshop were classified as shown in Table 1. The hazards (risks) are mostly related to ice, temperature, and ocean environment in the Arctic region, while personal risk and downtime occurrence were the common consequences for the accident cases. Working procedure document, safety fence, rapid recovery were suggested as preventive/mitigating safeguards for the currently designed system.

Table 1. Summary of HAZID.

			Existing safeguards	
Cause		Consequence	Preventive safeguard	Mitigating safeguards
Node (1) Hull Structure	Collision	Capsizing & sinking	– Robust design – Increase S.F. – Improve damaged stability	Auto-ballast control system
	Sea ice	Ice impact	– Ice management – Ice avoidance	– Auto DP system
	Low temp.	Brittle fracture	– Winterization – Heat line – Improve material property – Mooring line upgrade	– Deicing
Node (2) Mooring/DP system	Green water	Impact on deck	– Green water protector – Proper flare angle – High freeboard – Deflector	– Safety plan
	Sea ice	Large offset	– Ice management – Ice avoidance – High DP capacity – Heading control	– Proper winch operation
	Current	DP overload	– High DP capacity – Heading control	– Proper winch operation
	Wave	Large offset	– High DP capacity – Heading control	– Proper winch operation
Node (3) Turret	Collision	Mooring disconnection	– Redundancy design – Heading control	– Heading control
	Sea ice	Fail to disconnect	– Redundancy design	– Rapid OSV support – Auto system
	Current/Wave	Risers failure and hydrocarbon release	– Redundancy design – Shut down valve	– Oil spill recovery plan

Major cause for the systems and recommendations (safeguards) from the HAZID workshop are shown in Table 1.

4 STPA FOR THE DP AND MOORING SYSTEM

The STPA focused primarily on the control system, and its interaction with ship, ship engine and sensors, environment, DP-operator, and bridge officer. The STPA was carried out in three main steps, following the STPA procedure: (1) Establish system engineering foundation, (2) Identify Unsafe Control Actions (UCAs), and (3) Identify scenarios and safety constraints.

4.1 *Establishing system engineering foundation*

The first task of STPA was to establish *system engineering foundation*, which includes defining system-level accidents, hazard and safety constraints, and establishing functional control structure.

System-level accidents of an offshore vessel with DP and mooring system are *Rupture of the riser* and *Structural damage of the vessel*. The former is caused due to unintended drift of the vessel beyond the riser disconnect limit, and collision with other vessels or icebergs can be the cause of the latter. System-level accidents, hazards, and safety constraints are summarized in Table 2.

The second task of the system engineering foundation was to establish functional control structure. A high-level control structure was developed first, as shown in Figure 4, and a detailed structure was then developed as shown in Figure 5.

The DP and mooring system is controlled by DP Operator who receives commands from Bridge Officer. DP Control System calculates vessel motion and provides control command to Power System and Thrusters for position keeping of the vessel. Responsibilities and process models of each controller in the functional control structure are defined in Table 3.

4.2 *Identifying unsafe control actions*

STPA considers that safety can be treated as a dynamic control problem, rather than a component failure problem (Leveson and Thomas, 2013), and therefore the main focus of STPA is to identify control actions that lead to unsafe situations of a system. These control actions are called Unsafe Control Actions (UCAs).

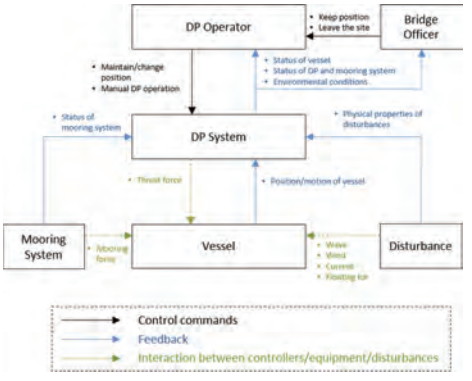


Figure 4. High-level control structure of DP and mooring system.

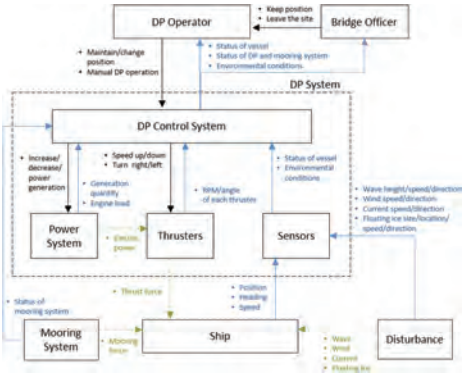


Figure 5. Detailed control structure of DP and mooring system.

Table 2. System-level accidents, hazards, safety constraints.

System-level accident	System-level hazard	System-level safety constraints
SLA1: Rupture of the riser	SLH 1: The vessel fails to maintain its position and drifts outside riser disconnect limit	SLSC 1: The vessel must never drift outside riser disconnect limit
SLA 2: Structural damage of the vessel	SLH2: The vessel fails to maintain its position and collides with other vessels	SLSC2: The vessel must never drift toward other vessels
	SLH3: The vessel fails to avoid icebergs that are beyond the structural strength	SLSC3: The vessel must avoid icebergs that are beyond the structural strength

The UCAs can be identified by examining combinations of control commands and process models that were identified from Section 4.1. Some examples of UCAs that were identified in our study are shown in Table 4.

4.3 *Identifying scenarios and safety constraint of each UCA*

The last step of STPA was to identify scenarios, casual factors, and safety constraints for each UCA. The scenarios and casual factors describe

Table 3. Responsibilities and process models.

Controller	Responsibilities	Process models
DP Operator	– Maintain or change position of the vessel depending on commands from Bridge Operator	– Command from Bridge Officer (keep/change position) – DP mode (position keeping/ moving to other site) – Status of automatic DP control system (operating/malfunction) – Status of mooring system (moored/disconnected) – Environmental conditions (thrust force required/ not required)
DP Control System	– Calculate vessel motion based on position, environmental conditions and mooring forces – Provide control commands to Power System to generate required electric power to Thrusters – Provide control commands to Thrusters to generate required thrust force to the vessel	– Command from DP Operator (maintain/ change position, manual operation) – Position of vessel – Environmental conditions – Mooring forces – Power generation quantity – Engine load – RPM of each Thruster – Angle of each Thruster

Table 4. UCAs of DP control system.

UCA.DPC01: DP Control System does not provide Speed up command to Thruster when vessel needs more thrust force		
Scenario	Associated causal factors	Safety constraints
DP Control System receives wrong measurement from reference sensors	Low accuracy of reference sensors	SC.DPC.01.01 Accuracy of reference sensors must be tested periodically SC.DPC.01.02 Reference sensors must have 2oo3 configuration
DP Control System receives no measurement from reference sensors	No power supply to reference sensors	SC.DPC.01.03 DP Control System must generate an alarm when no signal is received from reference sensors SC.DPC.01.04 Reference sensors must be connected to UPS
	Broken signal wires from reference sensors	SC.DPC.01.03 DP Control System must generate an alarm when no signal is received from reference sensors SC.DPC.01.05 Signal wires must be inspected periodically
DP Control System receives correct measurement, but DP Control System does not provide *Speed up* command	Wrong logic inside PCS	SC.DPC.01.06 Logic of DP Control System to generate *Speed up* command must be fully demonstrated during sea trial

Table 5. Scenarios, causal factors and safety constraints of UCA.DPC01.

No	Control Action	Not provided	Provided	Too early	Too late	Too short	Too long
1	Speed up Thruster no.1	UCA.DPC01 DP Control System does not provide *Speed up* command to Thruster when vessel needs more thrust force	UCA.DPC02 DP Control System provides *Speed up* command to Thruster when vessel needs to reduce or maintain thrust force	UCA.DPC03 DP Control System provides *Speed up* command to Thruster too early, before vessel needs more thrust force	UCA.DPC04 DP Control System provides *Speed up* command to Thruster too late, when vessel needs more thrust force immediately	UCA.DPC05 DP Control System provides *Speed up* command to Thruster too short, so Thruster cannot generate enough thrust force	UCA.DPC06 DP Control System provides *Speed up* command to Thruster too long, so Thruster generates excessive thrust force
2	Speed down Thruster no.1	UCA.DPC07 DP Control System does not provide *Speed down* command to Thruster when vessel needs less thrust force	UCA.DPC08 DP Control System provides *Speed down* command to Thruster when vessel needs to increase or maintain thrust force	UCA.DPC09 DP Control System provides *Speed down* command to Thruster too early, before vessel needs less thrust force	UCA.DPC10 DP Control System provides *Speed down* command to Thruster too late, when vessel needs less thrust force immediately	UCA.DPC11 DP Control System provides *Speed down* command to Thruster too short, so Thruster generates excessive thrust force	UCA.DPC12 DP Control System provides *Speed down* command to Thruster too long, so Thruster cannot generate enough thrust force

why and how UCAs occur, and safety constraints suggest requirements or guidelines to prevent scenarios from occurring, and ultimately, to prevent occurrence of UCAs.

Contrary to previous steps to identify UCAs, STPA does not provide structured guidance for this step. Identification of scenarios, causal factors, and safety constraints should therefore rely on brainstorming of analysts. Some of the results of this step is shown in Table 5.

5 RESULTS AND DISCUSSION

In this study, hazards of DP and mooring system have been analyzed using two hazard identification methods: HAZID and STPA. The scope of HAZID was limited to the hazards related to mooring system, while STPA focused on the control hazards of the DP system.

The advantages and limitations of each method, introduced in Section 1, were confirmed in the analyses of this paper. HAZID covered extensive safety problems, like hazards during installation, operation, maintenance, and so on. This method can be applied without limitation for static (passive) systems that requires no control actions, like mooring system, as well as dynamic (active) systems that requires control actions, like DP systems. However, HAZID provides less structured approach to identify hazards related with control problems compared with STPA. The results of HAZID are therefore highly affected by the knowledge and expertise of analysts. On the contrary, STPA provides well-structured systematic approach to identify hazards of a control system, and this systematic approach reduces reliance on analysts' knowledge or expertise and supports thorough hazard analysis of a system. However, STPA may encounter some problems when the method is applied to systems comprising only passive component with no control action, because STPA is a specialized analysis method for control problems. For instance, multiple mooring line failure due to heavy wind and/or wave is a critical hazardous event that can lead to rupture of the riser

or collision with other vessels, but this might not be directly identified by STPA because no control action is related with this hazardous event. This hazardous event can only be included indirectly into the analysis as a process model or as a scenario of some UCAs of the DP system.

Previous studies on STPA indicates that STPA can find wider range of safety problems than traditional hazard identification methods. In 2003, STPA was applied to U.S. Missile Defense System that had already been analyzed using traditional hazard identification methods, and STPA found so many additional flaws that the project was delayed for six months to fix the problems (Pereira et al., 2006). For an unmanned spacecraft of the Japanese Aerospace Exploration Agency (JAXA), STPA found every hazard identified by fault tree analysis and additional hazardous scenarios that were related to system design flaws, software, and so on (Ishimatsu et al., 2010). After analyzing a typical DP system, Rokseth et al. (2017) concluded that STPA can be considered as complementary to FMEA, providing a better risk picture of the DP system, and the same applies to this case study. At least for the specific case of this paper, STPA may cover a narrower scope of safety problem than a traditional hazard identification method, HAZID. The well-structured and systematic approach of STPA consequently resulted in limited applicability. This is not the only problem of STPA. HAZard and OPerability (HAZOP) study, for instance, provides specific guidewords and process parameters to identify hazards of a process system, and consequently, the application of this method is limited to process hazards (Rausand, 2011). To use HAZOP for other kinds of hazards, the guidewords and parameters should be modified for the hazards. The more a method provides guidance to identify hazards, the more restricted the scope may become.

For thorough analysis of hazards for DP and mooring system, complementary use of several hazard identification methods may be required, because each method has its own advantages and limitations, as confirmed by this study. A combined approach based on multiple hazard identification methods, to strengthen the strength and make up for the weakness of each method, would be an important further work.

REFERENCES

Abrecht, B. R. (2016) Systems Theoretic Process Analysis Applied to an Offshore Supply Vessel Dynamic Positioning System. Massachusetts Institute of Technology.

CCPS (2011) *Guidelines for Hazard Evaluation Procedures*, Wiley.

Ishimatsu, T., Leveson, N. G., Thomas, J., Katahira, M., Miyamoto, Y. & Nakao, H. (2010) Modeling and hazard analysis using STPA.

Keinonen, A., Shirley, M., Liljestrom, G. & Pilkington, R. (2006) Transit and Stationary Coring Operations in the Central Polar Pack. *International Conference and Exhibition of Performance of Ships and Structures in Ices* 16–19 July 2006, Banff, Alberta, Canada.

Kim, Y., Song, H., Won, Y. & Kang, K. (2017) Activities related with the Development of Arctic Station-keeping Methodology in KRISO. *The 32nd International Symposium on the Okhotsk Sea & Polar Oceans.* Feb. 2017, Mombetsu, Japan. pp. 58–62.

Kvitrud, A. (2014) Lessons Learned From the Norwegian Mooring line Failures 2010–2013. *Conference on Ocean Offshore & Arctic Engi-neering 2014.* San Francisco, California, USA, June 8–13 2014.

Leveson, N. & Thomas, J. (2013) An STPA primer. Cambridge, MA.

Leveson, N. (2012) *Engineering a safer world: Systems thinking applied to safety*, MIT press.

Ma, K., Shu, H., Smedley, P., L'hostis, D. & Duggal, A. (2013) A historical review on integrity issues of permanent mooring systems. *Offshore Technology Conference.* 6–9 May, Houston, Texas, USA, Offshore Technology Conference.

Molland, A. F. (2011) *The Maritime Engineering Reference Book: A Guide to Ship Design, Construction and Operation*, Elsevier Science.

Offshore Magazine (2013) Mooring Systems for Offshore Floating Installations Trends & Technology.

Pedersen, R. N. (2015) QRA techniques on dynamic positioning systems during drilling operations in the Arctic: With emphasis on the dynamic positioning operator. UiT The Arctic University of Norway.

Pereira, S. J., Lee, G. & Howard, J. (2006) A system-theoretic hazard analysis methodology for a non-advocate safety assessment of the ballistic missile defense system. Missile Defense Agency Washington DC.

Rausand, M. (2011) *Risk assessment: theory, methods, and applications*, John Wiley & Sons.

RMRS (2017) Rules for the Classification and Construction of Sea-Going Ships Part I: Classification. Russian Maritime Register of Shipping.

Rokseth, B., Utne, I. B. & Vinnem, J. E. (2017) A systems approach to risk analysis of maritime operations. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability,* 231, 53–68.

Team Energy Resources Limited (2002) Hazard assessment of well operations from vessels, Health and Safety Information, HSE InfoLine.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Risk analysis of high enthalpy fluid storage in geothermal power systems

Z. Nivolianitou
NCSR DEMOKRITOS, Aghia Paraskevi Attica, Greece

E. Kondili & G. Piperidis
Department of Mechanical Engineering, TEI of Piraeus, Egaleo Attica, Greece

ABSTRACT: The aim of this paper is to present the study elaborated through a specific risk analysis approach that highlights the potential risks and accidents in a geothermal installation in order to make the latter more reliable and environmentally friendly. The study also aims at offering a different perspective to the public benefiting from geothermal energy by influencing positively peoples 'awareness on this kind of facilities. The geothermal fluid may contain several Non-Condensable Gases (NCG), such as carbon dioxide (CO_2) and hydrogen sulfide (H_2S). Moreover, the presence of silica and boron in the geothermal brine can be hazardous both to people and to the surface environment. The safety analysis presented determines potential accidents in the storage phase and appropriate remediation actions, in the case that the geothermal fluid reinjection methodology is used.

1 INTRODUCTION

1.1 *General*

During the last years, the continuously rising energy demand worldwide is very evident. This can be explained by the sharp growth of the developing countries, as well as the improvement of living standards in the developed ones. Fossil fuels are the main energy source for most of the countries for several decades. However, the limited inventory of the former and the need for energy independence made most countries to invest gradually into renewable energy sources in order to reduce the share of technologies based on fossil fuels. Renewable resources have an unlimited availability, are usually equitably distributed around the world and are characterized as clean technologies because they produce very little waste and also have a minimal environmental impact. Moreover, they contribute not only to the reduction of CO_2 emissions but also to other pollutant gas emissions, such as sulfur, nitrogen oxides, VOCs (Volatile Organic Compounds) fostering both environmental protection, and growth sustainability. This is in line with the future socioeconomic and environmental needs of global economy according the Kyoto treaty objectives (Dincer, 2000 & deLIano-Paz et al., 2015).

Geothermal energy offers an alternative to this respect. Its utilization for electricity generation has been commercially used since 1913. As a renewable energy source, it can enhance a low carbon economy and strengthen independency from imported fuels. Geothermal power can be used as baseload renewable energy 24/7 in order to generate electricity regardless of the weather variations. Moreover, geothermal energy can fluctuate depending on the needs and can be flexible to support the intermittent renewable energy resources demands from wind and solar parks. In this context, it can be used to provide the stability of the power grid enhancing the efficiency of the entire system and increasing the security of energy supply against disruptions for geopolitical reasons and fossil fuel's price high volatility (deLIano-Paz et al., 2015).

The geothermal fluid itself may contain several Non-Condensable Gases (NCG), such as carbon dioxide (CO_2) and hydrogen sulfide (H_2S). Moreover, the presence of silica and boron in the geothermal brine can be hazardous both to people and to the surface environment. For these reasons geothermal fluid reinjection is used, a vital part of any geothermal development. The reinjection plan should be developed as early as possible in any geothermal development taking into account that the field characteristics are likely to change with time.

1.2 *Geothermal injection plants*

Few types of geothermal installations have been used so far. In 1970 on Achuapan field in El Salvador, the first injection effort has been implemented for environmental reasons, where a high Boron

content (~50 ppm) was identified and surface disposal was not admissible (Einarsson et al., 1975). Nowadays, injection seems to be the most favorable solution both environmentally and economically. Geothermal fluid injection is important to a geothermal project for a number of reasons, such as (a) to avoid surface disposal which can cause environmental impact, (b) to support the reservoir pressure, (c) to avoid any ground subsidence, and (d) to benefit from rock matrix heat.

Depending on the type of the geothermal system, reinjection can be infield, outfield or a mix of them. For vapor-dominated systems, where the water can run out, reinjection should be infield, while for hot water and liquid-dominated system a mix of infield and outfield injections is recommended. Through infield reinjection pressure support is provided and, consequently, drawdown and the potential for subsidence will be reduced. However, outfield reinjection protects the production area from the risk of cold water returns (Kaya et al., 2011).

In the present study the infield reinjection of thermal fluid is used as the technology method.

1.3 *Types of hazards*

It has been noted that the lack of planning for injection early in the development phase usually caused delays in putting power on line and reaching the planned generation level as well (Arnorsson, 2004). Thus, as it was mentioned above, the injection process is a vital part of any geothermal development, affecting directly the success or failure of any geothermal field development. Chemical pollution occurs both from gaseous components in steam that are discharged into the atmosphere and from aqueous components in spent water that may mix surface and ground waters, characterized as the most adverse environmental effect of geothermal energy utilization. Geothermal fluid may be including CH_4, CO_2, B and H_2S. The last is a noxious gas that has an unpleasant smell, when present in low and harmless concentrations and can be fatal, if inhaled in high concentrations for a longer period of time. In order to reduce chemical pollution both waste water and steam condensate should be injected into drill holes (Arnorsson, 2004).

2 TECHNOLOGY USED

2.1 *General principles*

The single-flash steam technology is adopted when the geothermal production wells indicate a mixture of steam and liquid, in order to convert the geothermal energy into electricity in a simple way. Initially, after its extraction the geothermal fluid mixture passes through a cylindrical cyclonic pressure vessel and is separated into distinct steam and liquid phases, with a minimum loss of pressure. The sitting of the separators is part of the general design of the plant and there are several possible arrangements, as shown in Figure 1 (Dipippo, 1998).

A typical 30 MW single-flash power plant includes 5–6 production and 2–3 injection wells relatively. These wells can be drilled at sites across the field or from a single pad through a directional drilling in order to intercept wider zone of the reservoir. In either case, through a piping system the geothermal fluids are transported from the production wells to the powerhouse and then to the disposal wells. Of course, the initial piping system can be modified if new power units are added later on.

2.2 *Description of problems actually connected with the injection of waste geothermal fluid*

Geothermal fluid does not necessarily require to be injected into the production geothermal reservoir; it could be injected into a different aquifer simply to avoid any environmental impact owing to surface disposal. However, in that case many problems can occur, such as ground subsidence, seismicity or leakage of the injection fluid to the surface due to the injection pressure, despite the fact that injection in a shallower aquifer than the producing reservoir saves drilling cost. On the other hand, injection into the production reservoir could be both beneficial as it was mentioned above but also risky owing to the potential cooling of the production well and the possible adverse impact on the chemistry of the extracted geothermal fluid (Sanyal et al., 1995).

The injection of waste geothermal fluid is most of the times connected with a number of problems. The suitability of injection sites is a critical choice for plant operation, production, while the injection within the same fault zone can cause serious cooling. Thus, the developer in such systems has the choice to inject in shallow ground water aquifers, if the geothermal fluid is environmentally benign and could also inject deeper within the fault zone,

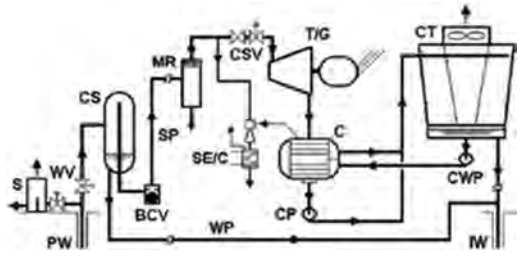


Figure 1. Simplified single-flash power plant design.

in order to be heated up before mixed again with the production line. Finally, in the case of environmentally benign fluid, the latter can be discharged on the surface.

Cooling provided by injection, seems to be the most common problem in the geothermal industry. According to Sanyal et al. (1995), there are two causes for injection—induced cooling: a) Very close distance between production and injection wells, b) "Short—circuiting" of the injected fluid to the production wells caused by a fault or fractured zone. However, the cooling problem can be identified through tracer test program conduction and could give alert to the developer.

A potential groundwater contamination can be caused by the injection of the geothermal waste on the geothermal reservoir; the main factors are the up flow of the injected water to the groundwater aquifer through a fault and the potential leakage of the injected fluid behind the casing caused by poor cement bond or probable damage due to corrosion or mechanical causes. However, through a careful geologic modeling the first cause can be avoided by locating injection wells in alternative sites.

Leakage of injection waste water to the surface can be identified in very shallow geothermal reservoirs (a few hundred meters). In order to avoid this kind of problem, injection should be deeper than the production level. Moreover, the occurrence of micro earthquakes near injection sites could be induced due to high pressure injection. More specifically, if the fluid pressure is increased beyond the original pore pressure and subsurface zones of weakness or active faults exist near the injection area, seismic activity may be induced. Thus, in order to avoid seismic activity, injections wells should be located away from known active faults, and the injection pressure should be lower than the original pore pressure of the system.

3 USE OF RISK ANALYSIS

3.1 *General*

Risk Analysis stemmed from the Major industrial accidents of the last decades involving dangerous chemicals that pose a significant threat to humans and the environment. It involves hazard identification, hazard evaluation, the development of potential risk reducing measures, and the communication of risk information to decision makers (Papazoglou et al., 1992) Risk analysis typically involves the following key steps:

- Hazard identification (HAZID)
- Frequency analysis
- Consequence analysis
- Quantification of risks using output from frequency and consequence analysis
- Investigation of potential risk reducing measures
- Development of recommendations

The common methods used in Risk Analysis are a) the hazard and Operability Analysis (HAZOP), b) the Failure Modes and Effects Analysis (FMEA), c) the "What if" scenarios, d)the Fault Tree (FT) Analysis, e) the Event Tree (ET) Analysis, f) the Risk Matrix and many other methodologies, which range from purely Qualitative to totally Quantitative and a mix of the two that are applied according to the needs and resources of the plant to be analyzed. Making assumptions about the detailed engineering of the reinjection method used, as series of accident scenarios including hydrogen sulfide (H_2S) release have been identified and are presented below.

3.2 *Accident scenarios in the geothermal plant*

The gases that exist in the geothermal re-injection fluid, namely the carbon dioxide (CO_2) and hydrogen sulfide (H_2S), also exist with the natural steam and do not condense at the condenser temperatures; these gases pass through a steam jet ejectors and, after condensers (SE/C in Figure 1) and the vacuum pumps, can removed increasing the overall pressure in the condenser and lowering the turbine power output. If a series of misfortunes happen and all protection systems (PSV, pressure control) fail to operate, the pressure will increase consequently above the normal limits and a possible break in the tank will occur. Two accidental scenarios have been analyzed together with their consequences:

a. Instantaneous failure of the tank full of H_2S
b. Failure of a pipeline carrying H_2S.

The SOCRATES toolkit has been used for these analyses; this is an in-house development of NCSR "DEMOKRITOS with gas outflow, dis-

Table 1. Outflow data for plant damage states (a) and (b).

Outflow data	Plant damage state (a)	Plant damage state (b)
Type of Installation	Tank	Tank
Storage conditions	Gas Pressurized	Gas Pressurized
Pressure in tank	$4 * 10^5$ (Pa)	$4 * 10^5$ (Pa)
Temperature	290K	290K
Diameter of tank	1.65 (m)	10 (m)
Height of tank	10 (m)	10 (m)
Height of orifice	5 (m)	5 (m)
Diameter of orifice	1.65 (m)	0.013 (m)
Duration	–	1200 (sec)

persion and consequences assessment codes based on TNO's "Yellow", "Green" and "Purple" books respectively (TNO, 2017).

The initial conditions to this software for outflow calculation are presented in Table 1 below:

4 RESULTS

SOCRATES toolkit calculated the concentration of Hydrogen Sulfide over a specified time and distance on the appropriate mesh for the case studies. In both cases scenarios, the instantaneous and the continuous release, Hydrogen Sulfide had enough buoyancy owing to its temperature and dispersed in the atmosphere in accordance with the Gauss model. All the parameters considering the meteorological phenomenon such as, the velocity, the direction and the stability class of the wind as well as the environmental temperatures are the mandatory data/ input for the Gaussian model when dealing with such dispersions and uncertainty in their values has been taken into consideration. More specifically, in the SOCRATES toolkit, sixteen (16) cases regarding the various meteorological parameters were simulated and correlated with site prevailing meteorological conditions.

Individual and Group Risk are calculated for plant damage states (a) and (b) for each mesh point in the area. The isorisk curves in Figures 2 and 3 present the individual risk for plant damage states (a) and (b), respectively.

Figure 2. Isorisk curves and consequences zone for plant damage state (a), Instantaneous release.

Figure 3. Isorisk curves and consequences zone for plant damage state (b), Continuous release.

5 DISCUSSION OF RESULTS

The Risk Analysis results presented in the previous section demonstrate the minor risk of the geothermal plant activity in the surroundings of the installation, as the 10^{-6} isorisk curve remains within the battery limits of the installation. Only plant operators are likely to suffer some discomfort owing to Hydrogen Sulfide, should an accidental release of the latter happen. However, personnel members are normally well trained and have sufficient Personal Protective Equipment (PPE) to deal with this hazard. The consequences findings though the SOCRATES toolkit running and interpretation of the results give promising results towards the low-risk operation of such plants in the proximity of inhabited areas.

6 CONCLUSIONS

In this paper, the results of a Risk Analysis of a geothermal power system caused by the injection of geothermal fluid back to the origin reservoir were presented. Additionally, environmental threats specific to geothermal fluid injection sites have been also mentioned.

Specific accidental scenarios of Hydrogen Sulfide release have been studied through the SOCRATES toolkit and the consequences owing to its dispersion are considered as non-significant to nearby communities. Plant personnel are aware of possible threats and can take specific mitigation measures.

Geothermal fluid injection condensate is a solution to environmental impacts but also helps maintain reservoirs pressure, increasing both the operational lifetime of the production well and the reservoir lifetime. Waste fluid injection is favorable against surface disposal of waste water due to its constituents which may cause adverse effect on the environment and the people.

In Greece geothermal energy has been used already in the early '90s only for direct utilization. The Greek reservoirs are not in use for electricity production in the country. One of the reasons was severe technical problems. These problems seem to be faced nowadays with the new prevention and mitigation techniques.

The final conclusion is that a reinjection plan should be developed as early as possible in any conceptual study of a geothermal development taking into account the field characteristics and also preferably the Risk Analysis results.

REFERENCES

Arnorsson, S. 2004. Environmental impact of geothermal energy utilization. Ageovhemical perspective: *Energy, Waste and the Environment;* 236, pp. 297–336.

deLIano-Paz, F. Silvosa, A.C. Antelo, S.A. Soares, I. 2015. The European low-carbon mix for 2030:The role of renewable energy sources in an environmentally and social efficient approach: *Renewable and Sustainable Energy Review*; 48, pp. 49–61.

Dincer, I. 2000. Renewable energy and sustainable development: a crucial review: *Renewable and Sustainable Energy Reviews*; 4(2), pp. 157–75.

DiPippo, R. 1998. "Geothermal Power Systems," Sect. 8.2 in *Standard Handbook of Powerplant Engineering,* 2nd ed., T.C. Elliott, K. Chen and R.C. Swanekamp, Eds., McGraw-Hill, Inc., New York, pp. 8.27–8.60.

Einarsson, S.S., Vides, R.a., Cuellar, G. 1975. Disposal of geothermal waste water by reinjection: *Proceeding Second United Nations Symposium on the Development and Use of Geothermal Resources*, San Francisco, 20–29 May, 1975, pp. 1349–1363.

Harvey, C.C. White, B.R. Lawless, J.V. Dunstall, M.G. 2010. 2005–2010 New Zealand country update, world geothermal congress. Indonesia: pp. 25–30.

Kaya, E. Zarrouk, S.J. O'sullivan, M.J. 2011. Reinjection in geothermal fields: A review of worldwide experience: *Renewable and Sustainable Energy Review*; 15, pp. 47–68.

Papazoglou I.A., Nivolianitou Z., Aneziris O., Christou M. 1992. Probabilistic safety analysis in chemical installations, *Journal of Loss Prevention in the Process Industries,* Volume 5, No 3, pp. 181–191.

Sanyal, S.K., Granados, E.E., Menzies, A.J. 1995. Injection—related problems encountered in geothermal projects and teir mitigation: the United State experience: *Proceeding of the World Geothermal Congress*; Florence, Italy, 3, pp. 2019–2022.

TNO. 2017. The Coloured Books—Yellow, Green, Purple, Red, Available at: https://www.tno.nl/en/focus-areas/urbanisation/environment-sustainability/public-safety/the-coloured-books-yellow-green-purple-red/, reached on Dec 13, 2017.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Branching rules and quantification based on human behavior in the ADS-IDAC dynamic PRA platform

M.A. Diaconeasa
Department of Mechanical Engineering, University of California, Los Angeles, USA

A. Mosleh
Department of Materials Science and Engineering, University of California, Los Angeles, USA
The B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles, USA

ABSTRACT: ADS-IDAC is a discrete dynamic PRA simulation platform in which the time-dependent changes in the functional state and parameters associated with the system elements are traced to generate scenarios by branching to new sequences at various time steps following a small set of general branching rules. These model-based branching rules have been developed to obtain a more realistic and complete solution space than the traditional static PRA methods, and avoid the sequence explosion phenomenon as the number of system states increases. This paper describes a new version of the ADS-IDAC simulation platform that includes: branching based on important human operator events – e.g., information processing, decision-making, procedure-following, or action-taking type, and full implementation of Human Error Probability (HEP) quantification rules that explicitly account for HEP dependencies based on shared performance shaping factors modeled using a dynamic Bayesian network.

1 INTRODUCTION

The dynamic Probabilistic Risk Assessment (PRA) methodologies are model-based simulations used to generate risk scenarios and their associated probabilities. This is achieved through general rules of stochastic and deterministic behaviors and interactions of the system and its elements – e.g., process variables, hardware, human operators, and environmental conditions. The simulation engine of a dynamic PRA platform tracks possible changes in the functional state and parameters associated with the elements of the system as a function of time. The nature and impact of the interactions and interdependencies among the system elements are processed by the simulation engine to generate risk scenarios. Ultimately, depending on the selected method chosen for scenario generation, probabilities of individual or clusters of scenarios are calculated for the system end states of interest. Dynamic PRA methodologies are especially important when the system includes time-dependent and complex interactions between the process variables, hardware, human, and environmental conditions. They provide a natural framework to include physical models, such as thermal-hydraulic codes for Nuclear Power Plants (NPPs), mechanistic models of hardware failure, cognitive models of human behavior, and those of natural hazards. Such a dynamic PRA simulation platform is ADS-IDAC: the Accident Dynamics Simulator coupled with the Information, Decision and Action in a Crew context cognitive model, and a realistic nuclear power plant thermal-hydraulic model. It is one of the most mature discrete dynamic platforms with an evolution that spans more than 25 years (Fig. 1).

In most of the Human Reliability Analysis (HRA) methods, the Human Error Probability (HEP), defined as the probability of an operator not completing a specific task, is quantified as a function of the Performance Shaping Factors (PSFs). In ADS-IDAC, the PSFs are quantified in terms of their contextual parameters (i.e. surrogates) and their impact on the cognitive processes is implemented through manifestation nodes as is illustrated in Figure 3 (Li, 2013).

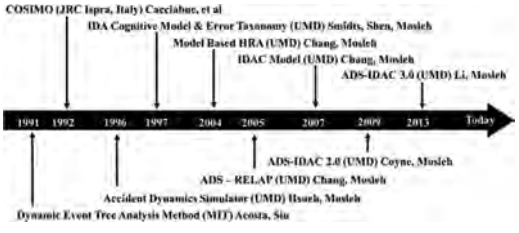


Figure 1. ADS-IDAC development history.

Although it greatly improved the explicit impact of the PSFs on the human performance, ADS-IDAC still lacks a full implementation for explicitly quantifying the HEPs based on the dynamic nature of the PSFs. For each individual or team activity, the behavioral effects of the PSFs can be accounted for through an influence diagram. Like its application in the Phoenix method (Ekanem, 2013), the Bayesian Belief Network (BBN) approach can be used to estimate the probability that a specific cognitive behavior occurs given certain conditions.

The main objective of the research reported in this paper was to introduce a set of comprehensive quantification rules to enable dynamic calculation of branch probabilities and complete risk scenario probabilities. The HFE dependencies were explicitly accounted for through the shared PSFs using a newly developed dynamic Bayesian network starting from a BBN model of PSFs developed in the Phoenix method.

2 IDAC HUMAN BEHAVIOR ADJUSTED BY CONTEXT AND OPERATOR VARIABILITY

During the simulation, the human operator behavior in IDAC is adjusted based on the context through a mechanism of surrogates—Performance Shaping Factors (PSFs) – manifestation nodes (Fig. 2). At each time step, the NPP state parameters are used to adjust the surrogate node values, the surrogates (yellow nodes) affect dependent PSFs (blue nodes) and in turn the PSFs affect manifestation (green nodes). The relationships between these nodes are based on empirical correlations found through extensive literature reviews corresponding to the appropriate human behavior mechanisms (Li, 2013).

Like all information processed by the operator model, all the dynamic PSF values are based on information perceived by the operator rather than data obtained directly from the thermal-hydraulic model or control panel. Perceived data may differ from the actual parameter value in thermal-hydraulic model or control panel due to time lags in updating perceived data and any distortions introduced by perception filtering and biasing.

The PSFs modeled in ADS-IDAC are: parameter criticality, system criticality, information load, time constraint load, cognitive task load, passive alarm load, expertise, task complexity, stress, fatigue, and problem-solving style. All of them are dynamic PSFs, except expertise and problem-solving style.

The criticality of system condition dynamic PSF represents the operator's perception of the level of degradation of key safety functions compared to normal operation. The value of the system criticality PSF corresponds to the aggregate deviation of key safety parameters from a nominal value. Each operator profile has its own parameters used to calculate this PSF: the threshold limits associated with each parameter, and the weighting factors used to aggregate the parameter contributions. The contribution from each identified parameter to the overall criticality of system condition PSF value is denoted as the parameter criticality. Given a set of high and low threshold limits, the parameter criticality corresponds to the magnitude of the parameter's deviation from a nominal safe condition.

The information loading dynamic PSF represents the operator's mental workload associated with the perception, processing, and communication of information. All information available from the NPP hydraulic model and crew communications must first pass through the operator's perception filter before it can be memorized and used. Consequently, the information flow rate through the perception filter provides an appropriate measure of each operator's information processing workload.

The time constraint load dynamic PSF represents the time available until a monitored NPP parameter exceeds a critical threshold. Because operators will normally monitor more than one important parameter, the overall PSF value is based on the most time critical parameter. The knowledge base profile for each operator includes data

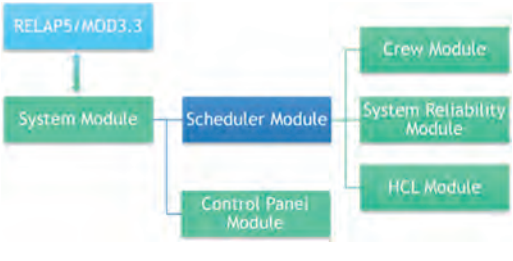


Figure 2. ADS-IDAC architecture.

Figure 3. Surrogates (yellow) – PSFs (blue) – Manifestation Nodes (green) IDAC model.

defining how the time constraint load PSF value is calculated, including a listing of NPP parameters used to calculate the time constraint PSF value along with the associated critical threshold values.

The task load dynamic PSF is indicative of the actual task demand assigned to a person quantified in terms of the number and type of tasks in a time unit. NPP control room operations do not normally involve heavy physical work, so in ADS-IDAC only the cognitive task load is of interest. Simulation HRA models possess a unique advantage of tracking each activity performed by the operator, which allows the code to count and to assess the workload specifically; therefore, the cognitive task load is also a dynamic PSF evaluated at each time step.

The passive alarm load dynamic performance factor embodies the number of salient stimuli that catch the operator's attention automatically like the alarms in the control room. Most often passive information is intrusive and grabs the operators' attention while interrupting their ongoing cognitive processes. Thus, too much passive information could be overwhelming. In addition to causing mental stress, it shifts one's attention and impedes the ability to refocus.

Operator expertise facilitates operator's coping with fast system dynamics in several ways: structuring and sorting the observations systematically, speeding the retrieval of knowledge for explaining the observation, and making connections between different pieces of information.

The task complexity dynamic PSF represents a measure of interaction among system dynamics, diagnosis confusion, and operator expertise. Amongst the system dynamics tracked at each time step are parameter trend changes, component state changes, and alarm state changes. Diagnosis confusion represents the complexity induced by inconsistent information and indicates the operator's level of understanding of the current NPP status.

The stress dynamic PSF combines various stress inducing PSFs into one factor: time constraint load, passive information load, cognitive task load, and task complexity. Each of the stressors has an equal weight on the stress value.

As the NPP control room operators' tasks do not involve heavy physical work. Thus, only the following three dimensions are considered in calculating this dynamic PSF: mental fatigue, sleepiness, lack of motivation/activity. It is evaluated based on an initial fatigue level at the beginning of their shift, a prolonged effort component due to performing tasks over a long period of time, and a sustained effort component representing the accumulation of fatigue by performing tasks. Moreover, the sustained effort component of fatigue is accelerated by the stress level.

The problem-solving style static PSF is reflected into the following of model parameters and information processing functions. In ADS-IDAC three problem solving styles have been implemented: Vagabond, Hamlet, and Garden-Path styles. They affect various parameters used to model the variation in diagnosis of operators in the reasoning module: routine monitoring time interval, maximal alarm stack length, prioritization of investigation items, investigation termination criteria, and accident awareness thresholds.

This framework performs well for adjusting the behavior of human operators based on the context. However, it is incomplete as it does not include any HFE, Crew Failure Mode (CFM), or HEP quantification that must be included in the generated Discrete Dynamic Event Tree (DDET) events for its full quantification.

3 NEW QUANTIFICATION MODEL FOR HUMAN ERROR

3.1 *Overview of HCL*

The HCL methodology (Wang, 2007) was developed for risk scenario analysis in PRAs of technological systems that considers not only the risks associated with hardware components (also called 'hard' causes), but also the risks generated by human activities, physical environment or socio-economic environment (also called 'soft' causes). This methodology offers a multi-layered modeling approach so that each individual domain of the system is modeled with the most appropriate technique. The three layers modeled in HCL are: Event Sequence Diagram (ESD) layer – it is used to model the risk context, Fault Tree (FT) layer – it is used to model the physical systems' behavior and quantify their impact on their corresponding linked events in the ESD, Bayesian Belief Network (BBN) layer – it is used to model the causal relations between events that have 'soft' root causes (Groth, Wang, Mosleh, 2010).

The HCL library was coupled with ADS-IDAC in a previous research effort for its FT-BBN quantification capabilities. The FTs were dynamically linked to the DDETs for modeling support systems and their impact on the frontline systems (Diaconeasa, 2017).

In this research only the BBN layer of the HCL architecture was necessary. Its capabilities had to be expanded to include leaky noisy OR gates that are used to reduce the conditional probability table size.

3.2 *Overview of the phoenix method*

The Phoenix method is a static HRA method that was developed out of the IDAC model. In the Phoenix method, the quantification of HFE is performed using a BBN for modeling the effect of the

PSFs on the CFMs. The construction of the BBN was made using the CFMs and PSFs as nodes and the arcs to show the relationships of influence between them through a conditional probability table. BBNs provide numerous benefits such as the ability to incorporate both qualitative and quantitative information from different sources for analysis, a causal structure for modeling interdependencies among its elements, the flexibility of updating the present state of knowledge of the model to incorporate new evidence as it becomes available, the capability of reasoning under uncertainty, and its ability to interface with existing conventional PRA models.

3.3 *HFE quantification through a dynamic Bayesian network*

The starting point for developing a quantification framework of the ADS-IDAC HFEs was the Phoenix method briefly described in the previous section. It is a natural step to include the additional elements of the Phoenix method into the IDAC model as part of the full dynamic ADS-IDAC simulation environment.

A BBN is valuable for problem domains or systems where the variables do not change over time. This assumption cannot always be assumed. For example, NPP system parameters and human operators' reasoning are clearly changing over time. In these cases, a Dynamic Bayesian Network (DBN) is necessary. A DBN is a BBN that is extended to incorporate a temporal dimension to enable the modeling of dynamic systems. The temporal extension of a BBN does not necessarily mean that the network structure or parameters changes dynamically, but it means that a dynamic system is being modeled. Hence, a DBN is a directed, acyclic graphical model of a stochastic process. It consists of time steps, with each time step containing its own variable values. The basic idea in a DBN is to specify how variables at time t influence variables at time t + 1 and replicating the structure of a model for each time step (Fig. 4). This concept of the dynamic Bayesian network was used to model the dependencies between the HFEs by replicating the network structure to represent the dynamic system and ultimately estimate the conditional HEP at each time step. This structured, causal model integrated into ADS-IDAC also helps improve the reproducibility and transparency of results produced by different HRA analysts for the same scenario.

Construction of the DBN involves building the structure of the network and defining the data describing the causal relationships between the network's nodes, and the nodes that will change in time. Unfortunately, the Phoenix method BBN cannot be adopted without modifications as some of its nodes

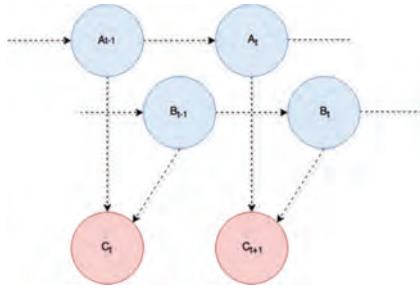


Figure 4. Simplified dynamic Bayesian network with developed on one time step with two dynamic nodes A and B influencing node C at every time step.

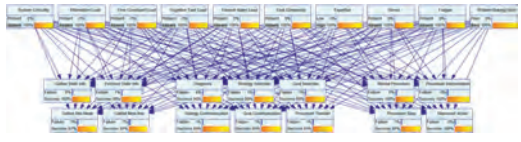


Figure 5. DBN of PSFs and HFEs.

Table 1. Mapping between the PSFs of ADS-IDAC and Phoenix.

ADS-IDAC	Phoenix
System criticality	Time constraint
Information load	Resources
Time constraint load	Time constraint
Cognitive task load	Task load
Passive alarm load	HSI
Expertise	Knowledge/Abilities
Task complexity	Procedures
Stress	Stress
Fatigue	Stress
Problem-solving style	Team effectiveness

do not have an ADS-IDAC equivalent and they do not cover all the HFEs modeled by ADS-IDAC.

The structure of the dynamic Bayesian network contains two layers in which all the top layer nodes influence all the bottom layer nodes (Fig. 5). Since the primary purpose of this dynamic Bayesian network is to model the effect of the PSFs on the crew failure modes, the top layer contains the PSF and the bottom layer contains crew failure modes. The top layer contains the PSFs described in the previous section: system criticality, information load, time constraint load, cognitive task load, passive alarm load, expertise, task complexity, stress, fatigue, and problem-solving style. As the Phoenix method does not have the same PSFs, based on their definition and purpose an equivalence relationships table was created to match the PSF used in ADS-IDAC and Phoenix (Table 1). All the PSFs except the expertise

and problem-solving style are dynamic; therefore, their value will change as the simulation progresses depending on the context. The PSFs have also been normalized to have values between 0 and 1.

The bottom layer of the dynamic Bayesian network in Figure 5 is made of crew failure modes. As in the Phoenix method, the crew failures modes specify the possible forms of human error in each of the information pre-processing, decision-making, and action execution phases.

The crew failure modes are also the generic functional modes of failure of the crew in its interactions with the NPP and represent the manifestation of the crew failure mechanisms and proximate causes of failure. They are selected to cover the various modes of crew response including procedure driven, knowledge driven, or a hybrid of both. To avoid double counting crew failure scenarios during the estimation of HEPs, the crew failure modes are defined as being mutually exclusive.

The crew failure modes within the information phase assume that the crew has failed in detecting, noticing and understanding the plant function they are supposed to be handling. Human failure in this phase can be divided into two major groups namely: failure to perceive passive information and failure to actively gather information. The crew failure mode that would occur during the perceiving of passive information is "Perceive State Info" – the crew fails to perceive the plant parameters or states from the control panel. The crew failure modes that would occur during the active gathering of information are: "Gather State Info" – the crew unintentionally try to collect the information from the wrong source, "Gather Info Mode" – the crew decide to use the old memorized information instead of collecting updated information, and "Gather New Info" – the crew failure in gathering new information. The equivalence table between the ADS-IDAC and Phoenix crew failure modes in the information phase is given in Table 2.

The crew failure modes within the decision-making phase assume that there is failure in situation assessment, problem solving and decision-making given correct information pre-processing. Therefore, the assumption is made that the crew has detected, noticed and understood the plant functions they are supposed to be handling. However, they have failed to make a correct assessment of the plant condition, diagnose, decide and plan the adequate response needed to solve the problem at hand. Moreover, the decision-making operator has the responsibility to communicate the action-taking operators the appropriate strategy. Ultimately, failures in this phase result in implementing an incorrect recovery strategy, hence failing the required function. Therefore, the following crew failure modes have been included: "Diagnosis" – decision-maker reaches the wrong assessment of the plant, "Strategy Selection" – decision-maker takes the wrong strategy given the correct situational assessment, "Strategy Communication" – decision-maker fails to communicate the correct strategy selected to the action-taker, "Goal Selection" – decision-maker selects the wrong immediate goal given the correct situational assessment, "Goal Communication" – decision-maker fails to communicate the correct goal selected to the action-taker, and "Procedure Transfer" – decision-maker switches to the wrong procedure. The equivalence table between the ADS-IDAC and Phoenix crew failure modes in the decision-making phase is given in Table 3.

The crew failure modes within the action execution phase involve failure in action execution given correct information pre-processing, situational assessment, and decision-making. It is assumed that the crew has detected, noticed and understood the NPP function they are supposed to be handling. Also, it is assumed they have made a correct assessment of the NPP condition, diagnosed, decided and planned the adequate response needed to solve the problem. However, they fail in executing the response or required action. It is assumed that the crew failure modes in the action execution phase are unintentional errors, that is the operators are always acting in the interest of recovering the NPP. The following crew failure modes have been included: "Mental Procedure" where the crew fails to adapt the instinctive response procedure to the current situation, "Procedure Step" where the crew skip or pause a procedure step in order to rely of their knowledge,

Table 2. Mapping between the CFMs in the information pre-processing phase of ADS-IDAC and Phoenix.

ADS-IDAC activity	Phoenix CFM
Perceive state info	Reading error
Gather state info	Wrong data source attended to
Gather info mode	Decision to stop gathering data
Gather new info	Team effectiveness

Table 3. Mapping between the CFMs in the decision-making phase of ADS-IDAC and Phoenix.

ADS-IDAC activity	Phoenix CFM
Diagnosis	Plant/system state misdiagnosed
Strategy selection	Inappropriate strategy chosen
Strategy communication	Information miscommunicated
Goal selection	Inappropriate strategy chosen
Goal communication	Information miscommunicated
Procedure Transfer	Inappropriate transfer to a procedure

"Procedure Interpretation" where the crew misinterpret the procedure step expectation, and "Maneuver Action" where the action-taker does not perform the requested action. The equivalence table between the ADS-IDAC and Phoenix crew failure modes in the action execution phase is given in Table 4.

Some of the crew failure modes fall into the category of errors of commission, that is they are the result of their intent given the wrong situational assessment of the NPP conditions.

The DBN structure defined by the PSFs and CFMs given above was integrated into ADS-IDAC by linking the all the human events types simulated into ADS-IDAC to the appropriate CFMs.

In the information pre-processing phase, the CFMs are linked to the human event types as follows. The "Perceive State Info" crew failure mode is used to estimate the probability of the action-taker to correctly register the perceived information for an alarm state, a frontline system state, a support system state, a parameter value, and a parameter trend value from the control panel. The "Gather State Info" crew failure mode is used to estimate the probability of any of the human operators to collect the information from the correct source on the control panel: alarm state, frontline system state, support system state, or parameter value. The "Gather Info Mode" crew failure mode is used to estimate the probability any of the crew members succeeds in collecting updated information instead of using old memorized information. The "Gather New Info" crew failure mode is used to estimate the action-taker's or decision-maker's probability of adding a parameter to the scan queue for gathering updated information.

In the decision-making phase, the CFMs are linked to the human event types as follows. The "Diagnosis" crew failure mode is used to estimate the decision-maker's probability of reaching the correct assessment of the NPP given their understanding of the NPP conditions. Note that if the operators do not correctly understand the NPP conditions, they will still reach a diagnosis, even if it's the wrong one. The "Strategy Selection" crew failure mode informs the decision-maker's probability of selecting the appropriate strategy given the correct situational assessment. The supported strategy selections in ADS-IDAC are wait and monitor, procedure following, hardwired diagnosis, and knowledge-based reasoning. The "Strategy Communication" crew failure mode is used to estimate the decision-maker's probability of communicating to the action-taker the selected strategy. Moreover, if the action-taker is in the follow instruction strategy mode, the same crew failure mode is used to estimate the decision-maker's probability of communicating to the action-taker the appropriate instruction. The type of instruction can be to obtain information about an alarm state, a frontline system state, a support system state, a parameter value, and a parameter trend value from the control panel or change their values. The "Goal Selection" crew failure mode is used to estimate the decision-maker's probability of selects the appropriate immediate goal given the correct situational assessment. Selecting the inappropriate goals can lead a delay in the appropriate recovery actions. The "Goal Communication" crew failure mode is used to estimate the decision-maker's probability to communicate the correct goal selected to the action-taker and consultant. The "Procedure Transfer" crew failure mode is used to estimate the crew's probability to switch to the correct written or mental procedure. The "Mental Procedure" crew failure mode is used to estimate the crew's ability to perform the appropriate instinctive response procedure based on the correct situational assessment. The "Procedure Step" crew failure mode is used to estimate the crew's probability of correctly skipping or pausing a procedure step in order to rely of their knowledge. The "Procedure Interpretation" crew failure mode informs the crew's probability to correctly interpret the procedure step expectation. As in the case of perceiving information, the expectations can be related to an alarm state, a frontline system state, a support system state, a parameter value, or a parameter trend value from the control panel.

In the action-taking phase, only one CFM is modeled as follows. The "Maneuver Action" crew failure mode is used to estimate the action-taker's probability to complete an action communicated by the decision-maker or from procedures.

After defining the PSFs, the CFMs with their mapping to the existing human events in ADS-IDAC, the next step was to obtain the data necessary to quantify the DBN. In order to achieve this, the estimated and calibrated parameters from the Phoenix method were adopted. The data sources used in the Phoenix method include German NPP operating experience data, other HRA methods (e.g. SPAR-H), and expert judgement. The advantage of the Bayesian network is that when new human performance data becomes available, be

Table 4. Mapping between the CFMs in the action execution phase of ADS-IDAC and Phoenix.

ADS-IDAC activity	Phoenix CFM
Mental procedure	Failure to adapt procedures
Procedure step	Procedure step omitted (intentional)
Procedure interpretation	Procedure misinterpreted
Maneuver action	Incorrect operation on component

it qualitative or quantitative, it can be easily integrated into the model parameter estimation process using Bayesian inference. Common sources of information that can be used are experimental data (e.g. control room simulator data), operating experience (e.g., licensee event reports), HRA databases, like the US NRC sponsored database project called the Scenario Authoring, Characterization, and Debriefing Application (SACADA), in addition to expert judgement.

The conditional probability table for each crew failure mode node in the Bayesian network is used to capture the strength of influence between each crew failure mode and its parent PSF nodes. This implies that the probability of the crew failure mode given all its possible combinations of the PSFs needs to be defined. This is challenging problem as the number of conditional probabilities in the conditional probability table grows exponentially with the number of nodes and states.

To reduce the conditional probability table size, the noisy OR gates can be used to specify the DBN and to build the conditional probability table for the crew failure modes nodes. In relation to the DBN quantification model included in ADS-IDAC, the leaky noisy OR gate also give the advantage of representing the probability that a crew failure can occur even when there is no influence from any of the PSFs. In other words, the leak factor provides a way to include other PSFs that are not explicitly represented in the DBN model as individual PSF nodes.

Using the HAMMLAB empirical data and the HRA results from the international empirical study (Lois, 2009), the conditional probability table has been normalized. The normalization procedure was the following: HFE 1A, failure to isolate the steam generator in the simple SGTR scenario, has been quantified using the original conditional probability table. The probability obtained has been scaled down two orders of magnitude such that the simulated HFE 1A falls inside the band of results obtained in the international empirical study.

The quantification of the successful human events captured by ADS-IDAC is a very critical aspect of the full DDET quantification. Now all the successful human events have a probability covering the full unit interval instead of an assumed fixed probability of one.

Compared to the conventional HRA methods, ADS-IDAC is able to quantify each individual activity that can lead to a particular HFE. Therefore, it is not only able to transparently predict the system and crew behavior, but quantify the probability of succeeding or failing at each time step for each activity the crew is undertaking based on the actual context through the linked DBN as is illustrated in Figure 6.

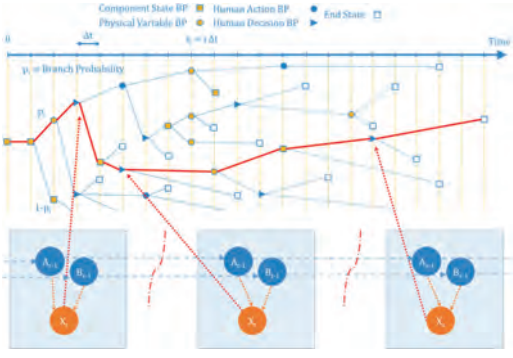


Figure 6. Graphical representation of an ADS-IDAC generated DDET where human events quantified with a simplified DBN are highlighted.

4 EXTENDED BRANCHING OF EVENTS IN GENERATION OF DDETS

The ADS-IDAC simulation engine generates a DDET by activating success, failure or partial failure branching points when certain conditions are met. The construction of the DDET is driven by a rich contextual environment simulated by ADS-IDAC and guided by the branching rules, which allow the modeling of variability in system and human operator.

The DBN covering the effect of PSFs on the CFMs at each time step was integrated into ADS-IDAC by linking all the human events types simulated in ADS-IDAC to the appropriate CFMs. That is, all the human events are assigned a success probability at each time step based on the current state of the DBN. One consequence of this framework can be seen with the following simple example. During a diagnosis, the crew may need to check the status of a component multiple times. Depending on the context, and implicitly on the value of the PSFs, the probability of the crew to correctly perceiving the status of that same component may be different. Therefore, new branching rules have been added to capture human performance variability, and to quantify the human failure events.

The branching points that were linked to be quantified with the DBN are described below:

- When a strategy is changed, based on the "Strategy Selection" crew failure mode two branches are generated: one in which the crew continue the current strategy, and another in which they switch to the new strategy.
- When a procedure step indicates a transfer to another procedure, two branches are generated: one in which the crew switches to the new procedure, and another in which they continue the current procedure.

- When an accident diagnosis threshold is exceeded based on the knowledge-based reasoning, two branches are generated: one in which the crew take recovery actions based on their reasoning, and another in which they transfer to the appropriate procedure.
- When a mental belief activation threshold is exceeded based on the heuristic reasoning two branches are generated: one in which the crew transfer to the mental belief, and another in which the crew bypass the mental belief and continue their activity.

These rules together with the newly developed quantification models help to keep the simulation space expansion under control, yet it also allows sufficient degrees of freedom for the system and crew to evolve into unexpected behaviors. For example, given the procedure step skipping probability is quantified at each time step containing written procedure steps or mental procedures either deterministically or stochastically, the skipping of procedure steps could be simulated and their impact analyzed in a consistent and transparent way.

5 CONCLUSIONS

A set of comprehensive quantification rules to enable dynamic calculation of branch probabilities and complete risk scenario probabilities was developed and implemented using a DBN. The HFE dependencies were explicitly accounted for through the shared PSFs by adapting the BBN model of PSFs developed in the Phoenix method to the dynamic environment of ADS-IDAC.

ACKNOWLEDGEMENTS

This work was funded through a Research Grant (NRC Grant HQ-60-14-G-0013) by the U.S. Nuclear Regulatory Commission (USNRC).

REFERENCES

Cacciabue, P.C. 1992. Cognitive modeling: a fundamental issue for human reliability assessment methodology? *Reliability Engineering & System Safety*, vol. 38, pp. 91–97.

Chang, Y.-H. 2007. Cognitive Modeling and Dynamic Probabilistic Simulation of Operating Crew Response to Complex System Accidents (ADS-IDA Crew). University of Maryland.

Chang, Y.H., Bley, D., Criscione, L. Kirwan, B. Mosleh, A. Madary, T. Nowell, R. Richards, R. Roth, E.M. Sieben, S. Zoulis, A.. 2014. The SACADA database for human reliability and human performance, *Reliability Engineering & System Safety,* Volume 125, Pages 117–133, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2013.07.014.

Coyne, K.A. 2009. A Predictive Model of Nuclear Power Plant Crew Decision-Making And Performance In A Dynamic Simulation Environment. University of Maryland.

Ekanem, N.J. 2013. A Model-Based Human Reliability Analysis Methodology (Phoenix Method). University of Maryland.

Groth, K., Wang, C.D. & Mosleh, A. 2010. Hybrid causal methodology and software platform for probabilistic risk assessment and safety monitoring of socio-technical systems. *Reliability Engineering & System Safety*, 95, 1276–1285.

Henrion, M. 1989. Some practical issues in constructing belief networks. Uncertainty in Artificial Intelligence, Elsevier Science Publishers, pp. 161–173.

Hsueh, K.S., Mosleh, A. 1996. The development and application of the accident dynamic simulator for dynamic probabilistic risk assessment of nuclear power plants, *Reliability Engineering & System Safety,* vol. 52, pp. 297–314

Li, Y. 2013. Modeling and Simulation of Operator Knowledge-Based Behavior. University of Maryland.

Lois, E., Dang, V.N., Forester, J., Broberg, H., Massaiu, S., Hildebrandt, M., Braarud, P.Ø., Parry, G.W., Julius, J., Boring, R., Männistö, I., Bye, A. 2009. International HRA empirical study—description of overall approach and first pilot results from comparing HRA methods to simulator data (HWR-844/NUREG/IA-0216 vol 1). OECD Halden Reactor Project/US Nuclear Regulatory Commission, Halden, Norway/Washington, DC.

Pearl, J. 1988. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann.

Siu, N. 1990. Dynamic accident sequence analysis in PRA: A comment on "Human reliability analysis-Where shoudst thou turn?" *Reliability Engineering and System Safety,* 29(3), 359–364. http://doi.org/10.1016/0951-8320(90)90019-J.

Smidts, C., Shen, S.H. & Mosleh, A. 1997. The IDA cognitive model for the analysis of nuclear power plant operator response under accident conditions. Part I: Problem solving and decision making model," *Reliability Engineering & System Safety*, vol. 55, pp. 51–71.

Wang, C. 2007. Hybrid Causal Logic Methodology for Risk Assessment. University of Maryland.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

HYPRA: A hybrid static-dynamic PRA software platform

M.A. Diaconeasa
Department of Mechanical Engineering, University of California, Los Angeles, USA

A. Mosleh
Department of Materials Science and Engineering, University of California, Los Angeles, USA
The B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles, USA

ABSTRACT: Conventional Probabilistic Risk Assessment (PRA) methodologies and software tools developed for a variety of risk-informed applications are also characterized as 'static', referring to the fact that temporal and phenomenological aspects of risk scenarios are at best implicit in the models and results. For instance, typical core melt cut sets are essentially logical combinations of contributing events, without consideration of possible effects of different time ordering of the constituent events, and timing of event initiation or termination. Over the past three decades a small community of researchers have directed toward developing and exploring possible benefits of dynamic PRA methods and tools. Dynamic methodologies provide a natural framework to include physical models, mechanistic models of hardware failure or human operator behavior models. In this paper, the capabilities to link conventional PRA platforms with dynamic PRA tools are described and the concept of cut set diffraction is introduced. This facilitates the use of newer risk analysis methods while still using the existing probabilistic information that, at least in the U.S., is available for every Nuclear Power Plant (NPP).

1 INTRODUCTION

Conventional Probabilistic Risk Assessment (PRA) methodologies and software tools developed for a variety of risk-informed applications are also characterized as 'static', referring to the fact that temporal and phenomenological aspects of risk scenarios are at best implicit in the models and results. For instance, typical core melt cut sets are essentially logical combinations of contributing events, without consideration of possible effects of different time ordering of the constituent events, and timing of event initiation or termination. Additionally, the basic events in classical PRAs are typically binary, a situation that can mask the impact of such thing as degraded component states. Furthermore, conventional PRAs have come under attack in terms of adequacy in providing important contextual information for proper modeling and analysis of operator errors.

These are some of the reasons stated in support of the need for simulation based PRA, generally known a dynamic PRA. Over the past three decades a small community of researchers have directed toward developing and exploring possible benefits of dynamic PRA methods and tools. These efforts show a significant diversity of objectives and methodology. Among the stated objectives are: study of impact of timing and sequencing of events, understanding the effects of variations in the underlying physical processes, and study of operator error, particularly errors of commission. Methodological and computational works cover the computational efficiency of generating dynamic event trees (discrete and continuous versions), scalability and convergence, post-processing of generated results, software platform generality, and user interface features.

With significant advancement of dynamic methodologies and computer power and storage capacity, the prospects of using dynamic PRA tools to answer risk questions is more real now than ever before. Despite the progress in the development of dynamic PRA methods and tools, the conventional PRA method and software platforms are expected to remain the method of choice for the current generation of NPPs. However, a strong case can be made that conventional PRAs can be augmented by dynamic approaches, especially for scenarios where dynamic characteristics are anticipated to be important.

An efficient path to enable such augmentation is to develop protocols and capabilities to link conventional PRA platforms with dynamic PRA tools. This will facilitate the use of newer risk analysis methods while still using the existing probabilistic information that, at least in the U.S., is available for every NPP.

2 CONVENTIONAL AND DYNAMIC PRA

2.1 *Conventional PRA*

In the application of the defense in depth concept, Probabilistic Risk Assessment (PRA) is used to determine the probability for breaching each barrier. In general, PRA methods are employed to identify failure scenarios and to estimate their associated risk by answering to three fundamental questions (Kaplan and Garrick 1981): 1) "What can happen?", 2) "How likely it is that it will happen?", and 3) "If it does happen, what are the consequences?" System performance reliability, uncertainty analysis, and human reliability analysis methods are explicitly integrated into the PRA framework, thereby enabling the application of defense in depth, which would be impossible to do using deterministic analysis methods alone.

PRA is a matter of scenarios and their likelihoods as illustrated in Figure 1 showing an event tree with the system actuation failure event represented through a fault tree (Garrick, 2008). The theory of structuring scenarios is rooted in field of reliability engineering and systems analysis. Reliability engineering produced graphical representations that were very useful in describing how systems perform. Fault tree analysis is a deductive reasoning process for building failure models. Its roots are switching algebra and circuit theory coming out of the Bell Labs. The inductive reasoning process of the event tree concept has its roots in decision theory and was developed into a systems performance tool in the reactor safety study WASH 1400 (Rasmussen, 1975). The theory of probability as a measure of likelihood evolved in the fields of mathematics over hundreds of years. For events that occur frequently, a frequency can be quantified to reflect the number of occurrences per unit of time. However, for events that rarely occur, or may not occur at all, the concept of probability is quantified. Probability is synonymous with 'credibility' as in the credibility of a hypothesis based on all of the available evidence. Ultimately, for any defined undesirable event frequencies, such as fatalities, in the form on probability of frequency are the main measure of safety risk for engineering systems.

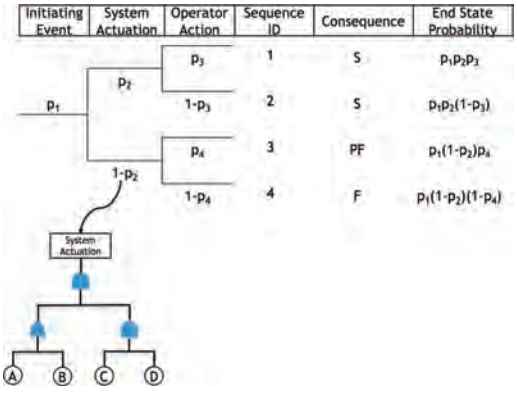


Figure 1. Conventional PRA.

A full score PRA involve the following well established steps (Garrick, 2008):

- define the system and its success state—usually involves linearizing the system into logically progressive operating states, that is fixing in time the top events based on the analyst's understanding of the system operation,
- identify and characterize the hazards also called threat assessment,
- develop and structure the 'what can go wrong' scenarios that lead to undesired outcomes or vulnerability assessment—creative exercise analyst dependent
- quantify the scenarios—the uncertainties associated with the scenarios must be part of the answer
- assemble results into measures of risk, and importance measures,
- interpret the result for meaningful risk management.

Each sequence is a path to either type of consequence: success (S), failure (F), or partial failure (PF) state. Each probability is a conditional probability. The sequences are quantified by propagating the probability density functions representing the split fractions of the top events through their corresponding paths up to their end states. One of the end results is a probability of frequency curve for each sequence end state consequence or in the form of a mean and median with its confidence interval. However, the quantification results are as important as the scenarios generated by the analysts and their relative importance to the NPP safety.

Multiple conventional PRA software platform are readily available in academia, industry, or regulatory agencies. The most common ones are IRIS (Fig. 2), RISKMAN, CAFTA, RiskSpectrum Suite, SAPHIRE. In this work, the IRIS tool was selected as the conventional PRA platform of choice, however the generic architecture used is applicable to the other software platforms.

2.2 *Dynamic PRA*

Dynamic PRA methodologies are generally those that use a time-dependent phenomenological model of system evolution and take into account its stochastic behavior to estimate the risk associated with the system response to an initiating event (Aldemir, 2013). The system evolution model

Figure 2. IRIS platform for modeling ESDs, FTs, and BBNs.

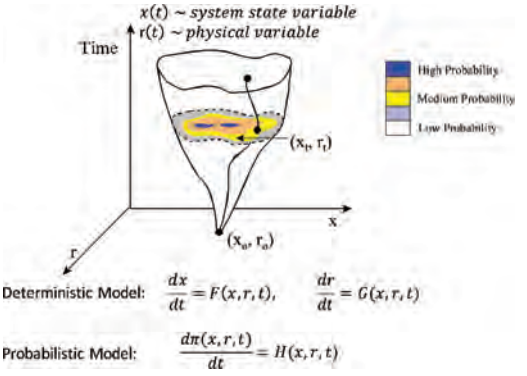


Figure 3. Dynamic PRA (Mosleh, 2015).

keeps track of the current hardware status, current level of processes variables, current operator assessment, scenario history, and time (Siu, 1990). A graphical representation of the system evolution space with its probabilities of occurrence is shown in Figure 3.

Dynamic PRA has been grouped in two main categories: continuous-time (e.g. Continuous Event Tree (CET)) and discrete-time (e.g. Dynamic Event Tree Analysis Method (DETAM), Accident Dynamic Simulator (ADS), Analysis of Dynamic Accident Progression Trees (ADAPT), and Risk Analysis Virtual Environment (RAVEN)).

The discrete dynamic PRA methodologies use Discrete Dynamic Event Trees (DDETs) that are computationally generated based on a time-dependent model of system evolution and various branching conditions.

Essentially, all discrete dynamic PRA methodologies employ a simulation engine that generates branches at each user-specified time step or conditions with their associated probabilities and computes the probability of each scenario. As can be seen in Figure 4, branching points can include system hardware states, physical variable changes, human actions, software failures or an end state if one of the stopping criteria is met.

ADS-IDAC, the Accident Dynamics Simulator with its operating crew simulation model (Identification, Decision and Action in a Crew cognitive context) and thermal-hydraulic code RELAP5/Mod 3.3 was selected as the dynamic PRA platform for scenarios that would require more extensive operator response modeling. ADS-IDAC is a simulation engine that includes a scheduler module, a hardware reliability model, an indicator module (the control panel), and the IDAC operator response model coupled with the RELAP5/MOD3.3 thermal-hydraulic code (the system model) to generate DDETs containing contextually rich scenarios that could occur given an initiating

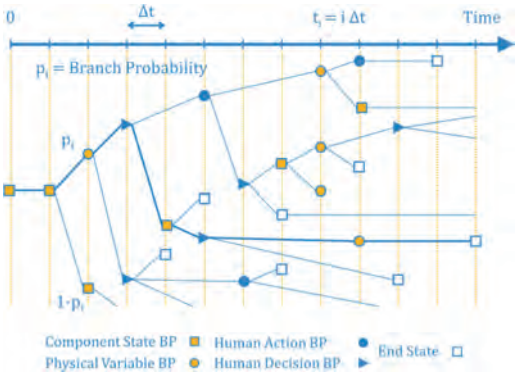


Figure 4. Discrete dynamic PRA methodology.

event. Its modular structure and the flow of information between modules are shown in Figure 5.

A scheduler module coordinates the interactions between all the other modules and generates the DDETs. As is the case for traditional ETs, the probability of each scenario or sequence is the product of conditional probabilities of its constituent branches. The indicator module simulates the control panel indicators' states driven by information from the system module. The HCL module models and quantifies the probability of Human Failure Events (HFEs) based on a Dynamic Bayesian Network (DBN) of a range of Crew Failure Modes (CFMs) and Performance Shaping Factors (PSFs) that reflect the context conditions in which the operators create situational assessments and devise recovery strategies (Diaconeasa, 2017a). The hardware reliability module simulates the failure probabilities of the system's and control panel's components, but coupled with the HCL module can model the impact of support systems on the frontline systems through dynamically linked FTs (Diaconeasa, 2017b).

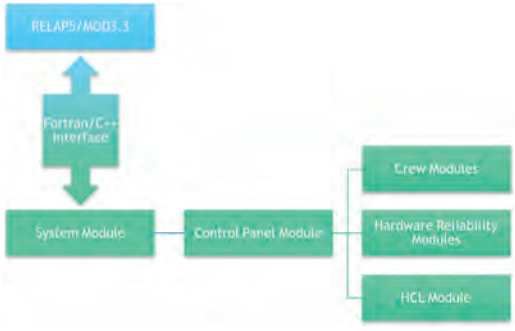


Figure 5. ADS-IDAC architecture.

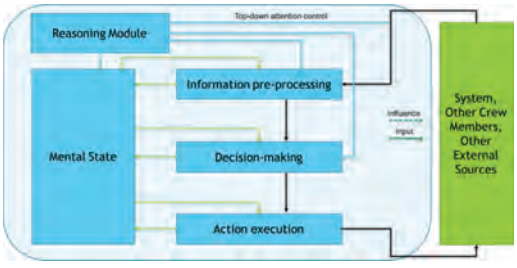


Figure 6. IDAC cognitive model for one human operator.

The IDAC model serves as underlying framework for operator behavior. IDAC decomposes the operator's cognitive flow into three main process: information processing, decision-making, and action execution (Fig. 6). The domain of applicability of IDAC is constrained to environments characterized by high levels of training and explicit requirements to follow procedures. These constraints simplify the modeling by limiting the degrees of freedom from the broader human response spectrum. In IDAC, the crew is modeled as a team of individuals working on different assigned tasks and communicating with one another. The individuals differ by the content of their memory, by their mental state, and by the goals and strategies they employ. IDAC can simulate several decision-making and problem-solving strategies, including passive and active information gathering, diagnosis, skill-, rule-, and knowledge-based actions, and procedure-following. The model includes several dynamic and static PSFs as part of the set of factors and rules that simulate the Senior Reactor Operator (SRO) and Reactor Operator (RO) responses. Each operator also has a unique knowledge base that defines his or her knowledge about nuclear plant systems and operations.

Previous research efforts in developing ADS-IDAC have shown that a small set of generic branching rules are sufficient to capture complex variations in system and crew-to-crew performance (Coyne, 2009).

3 HYBRID PLATFORM ARCHITECTURE

The hybrid static-dynamic PRA platform was designed to selectively feed conventional PRA results (e.g. cut-sets) from IRIS into the ADS-IDAC dynamic PRA platform for dynamic analysis.

The typical ET/FT/BBN analysis of conventional PRAs is used to generate a list of cut sets. Each cut set is used in the dynamic PRA analysis to constrain the branching at every time step in creating the DDET driven by the system and crew evolutions. This is graphically illustrated in Figure 7.

The optimum strategy that provides the broadest range of applications and highest compatibility with present conventional and dynamic platforms (not only IRIS and ADS-IDAC) is by creating an interface between the platforms for selectively passing information that complies with the Open-PSA model exchange format. Given that the Open-PSA model exchange format covers only ETs and FTs, the standard will be extended to cover BBNs in order to provide compatibility to all the layers of the IRIS platform. In Figure 8, the hybrid static-dynamic PRA platform is shown where the typical inputs and outputs are listed for each module.

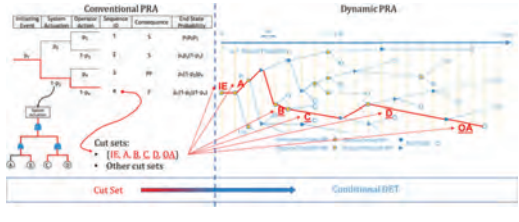


Figure 7. Hybrid static-dynamic framework.

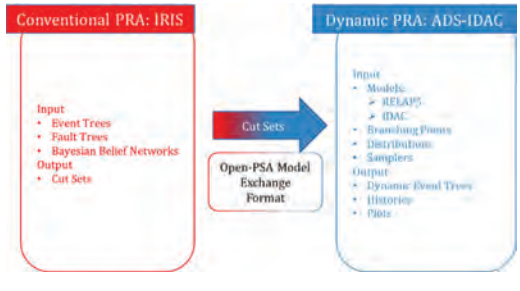


Figure 8. Hybrid static-dynamic PRA platform.

The interface acts as a gate for passing information from cut set to branching conditions in ADS-IDAC. Thus, in the ADS-IDAC model boundary conditions need to be set based on both the cut set information and the available components available in the system model either captured in the RELAP5 thermal-hydraulic model or through dynamically linked FTs.

Overall, all the branching points and the intermediate events associated with an initiating event make up the DDET during accident scenarios. The set of generic branching rules cannot create the DDET without modeling in parallel the dynamic system and human operator behaviors. Therefore, the construction of the DDET is driven by a rich contextual environment simulated by ADS-IDAC and guided by the branching rules, which allow the modeling of variability in system and human operator.

The necessity for branching rules in a dynamic PRA simulation platform like ADS-IDAC is the sequence explosion phenomenon. If the simulation engine would allow branching at every time step the number of sequences needed to be explored would grow exponentially and the simulation time would become unrealistically long with the current computational models and resources. For the same computational reasons, sequence termination conditions have been implemented to stop the engine from exploring sequences after a time period of interest, when certain physical limits have been exceeded, or when the operators enter certain procedures. For example, if the interest of the simulation is the exploration of crew variability in diagnosing a SGTR, two sequence termination conditions could be set to stop the simulation. One of them could be placed when the operators transfer to procedure E-2 "Isolation of steam generator with secondary break" or E-3 "Tube rupture in one or several steam generators." Another could be set when the simulation time exceeds a certain time period set based on previous crew performance.

At the same time, sequence termination conditions can be set to calculate an overall failure probability for an event of interest. For example, a sequence termination condition for the fuel element cladding temperature exceeding the acceptance criteria for emergency core cooling systems for light water nuclear power reactors (10 CFR Part 50.46) of 2200°F. The summation of the end state probability for all sequences that were terminated by this condition would essentially estimate an overall measure of core damage probability.

Overall, the branching rules and the sequence termination conditions do not cover all the models included in ADS-IDAC their variability is not included in the generated DDET and, ultimately, the solution space is not complete.

Branching rules that cover the failure of either frontline or support system components have been implemented. Nonetheless, by implementing branching rules alone does not mean the sequence end state probability can be quantified. Each branching point requires either a success or failure probability. The ADS-IDAC hardware reliability module covers modeling of both frontline and support system failures during operation by considering the failure rate, the number of failures desired and the time interval between them. DDET branching is modeled such that failures during operation generate two branches: success and failure branch. For a specific equipment, if more than one failure during operation is modeled, only the subsequent success branches will further allow more failures as on the failure branches this equipment had already failed. This feature further extends ADS-IDAC's capability to dynamically predict the timing importance of component failures during operation for the overall safety of the design in question is. For example, small-break LOCA scenarios in a PWR could be set up to simulate the timing of the pressurizer power operated relief valves (PORVs) stuck-open failure events and their impact on the available time for recovery actions.

The crew's recovery of frontline and support system's component failures can also be modeled (Diaconeasa, 2017b). When the crew attempt to recover a component, two additional branches are generated: a recovery branch, in which the component is successfully recovered, and a permanent failure branch, in which the component remains failed.

The branching points that were quantified with the dynamically linked FTs (Fig. 9), out of which a success and a failure branch are created, are given below:

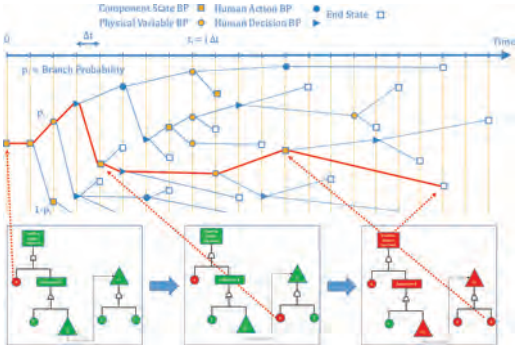


Figure 9. DDET with quantified frontline and support system branching points based on dynamically linked FTs.

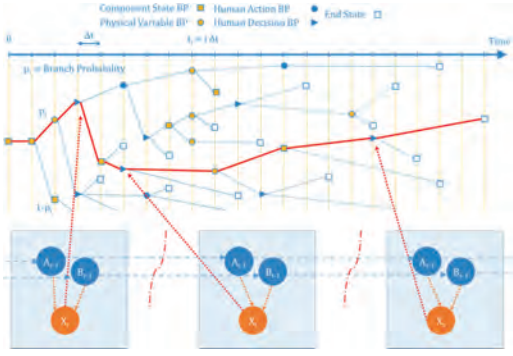


Figure 10. DDET with quantified human events branching points based on a DBN of PSFs and CFMs.

- Components of frontline systems failures at fixed time.
- Components of frontline systems failures on demand.
- Components of frontline systems failures during operation.
- Components of support systems failures at fixed time.
- Components of support systems failures on demand.
- Components of support systems failures during operation.

Using the DBN of PSFs and CFMs, a range of branching points for HFEs are implemented to model the crew's variability (Diaconeasa, 2017a). The HFEs covered in the branching rules (Fig. 10) can be of type strategy selection, procedure transfers, diagnosis of accident conditions based on either the knowledge-based reasoning or procedure following.

4 CUT SET DIFFRACTION PHENOMENON

The hardware and human operator behavior variations can act as a diffraction grating to create new scenarios starting from a single cut set, a process analogous to light diffraction in optics. This phenomenon is illustrated in Figure 11.

The number of sequences obtained from the single cut set would depend on the number of events in the cut set, but also by considering the following variations: timing and order of failures, degree of component degradation, time and degree of recovery, human decisions and actions, human dependencies, or physical variable thresholds.

The timing and order of failures are important in Fukushima-like scenarios, or hurricane

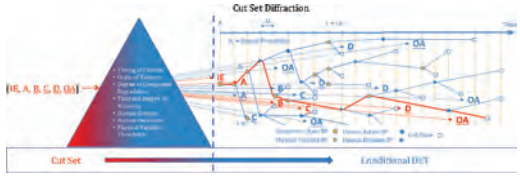


Figure 11. Graphical representation of the cut set diffraction where the input cut set is highlighted in the DDET.

scenarios, where the reactor is tripped at some time prior to the more severe plant response. How can initially smaller decay heat levels complicate the sequence of events operators are trained on? How can the order of events be shifted and how does this impact the crew response? There are scenarios where it is necessary to consider the timing of failures. For example, the failure of containment spray at some time after they successfully start will prolong the capacity of the Refueling Water Storage Tank (RWST) to supply suction before recirculation. On the other hand, the failure to secure running trains of spray per procedure would lessen the time to suction switchover.

The degree of component degradation is important in cases similar to the Beaver Valley PRA sequences initiated by a loss of an instrument bus. When can an initially benign sequence degrade and become more of a challenge? Loss of this bus fails makeup to the volume control tank, but normal charging continues. If not recovered (by switching to an alternative instrument bus), suction is shifted to the RWST. Failure to recover makeup prior to switchover to RWST could make successful termination dependent on a single check valve (in a borated environment).

The time and degree of recovery was seen to be important in the TMI-3 Loss Of Coolant Accident (LOCA). Auxiliary Feedwater (AFW) was failed, and then subsequently recovered. Once recovered, High Pressure Injection (HPI) was secured. What was not appreciated was that AFW recovery came too late, as a LOCA had already occurred.

The importance of decisions and actions of the crew can be illustrated by looking at scenarios involving the failure of Main Feedwater (MFW) and AFW. Procedures instruct the crew to recover MFW or AFW and after failure to do so, go to feed-and-bleed (F&B). Observation of different crews at simulators suggest different crews will devote different amounts of time focusing on the recovery of MFW or AFW. When would the delay result in loss of a potential success path? Also, recovery that requires steam generator depressurization (e.g., using condensate) may influence recovery characteristics. This assessment will likely

be plant-specific due to differing shutoff head capacities of the HPI pumps.

The hardware dependencies can have risk implications as is the case for shared equipment during a multi-unit initiator. Examples might be Loss of Off-Site Power (LOSP) in a NPP with shared diesels and shared equipment. Initial plant states and equipment failure conditions may be underappreciated as they could complicate plant recovery.

A consequence of the cut set diffraction phenomenon in the hybrid static-dynamic PRA platform is the consideration of non-core damage end states. PRAs are typically developed to trace the sequence of events that lead to core damage or fission products release. The frequency of such scenarios is one metric of interest in regulatory risk, as it is one common surrogate measure for public health risk. However, restricting the logic models to only determine core damage frequency, or large early release frequency, limits the breadth of information potentially available to decision makers.

5 CONCLUSIONS

A hybrid static-dynamic PRA platform was developed to leverage the newer risk analysis methods available while still using the existing probabilistic information that, at least in the U.S., is available for every NPP. The IRIS and ADS-IDAC tools make up the main elements of this platform, nevertheless similar PRA tools could be used that conform to the Open-PSA model exchange format. Finally, the cut set diffraction phenomenon was introduced and illustrated with short examples.

REFERENCES

Aldemir, T. 2013. A survey of dynamic methodologies for probabilistic safety assessment of nuclear power plants. *Annals of Nuclear Energy,* 52, 113–124. http://doi.org/10.1016/j.anucene.2012.08.001.

Coyne, K.A. 2009. A Predictive Model of Nuclear Power Plant Crew Decision-Making and performance in a Dynamic Simulation Environment. University of Maryland.

Diaconeasa, M.A., Mosleh, A. 2018. Branching rules and quantification based on human behavior in the ADS-IDAC dynamic PRA platform, *ESREL*, Trondheim, Norway.

Diaconeasa, M.A., Mosleh, A. 2017. The ADS-IDAC dynamic platform with dynamically linked system fault trees. American Nuclear Society International Topical Meeting on Probabilistic Safety Assessment and Analysis, Pittsburgh, PA, United States.

Garrick, B.J. 2008. Quantifying and Controlling Catastrophic Risks. Quantifying and Controlling Catastrophic Risks, 1–351.

Kaplan, S. and Garrick, B.J. 1981. On the Quantitative Definition of Risk. *Risk Analysis,* 1(1), 11–27. http://doi.org/10.1007/BF00134104.

Rasmussen, N. 1975. Reactor Safety Study: An assessment of Accident Risks in U.S. Commercial Nuclear Power Plants WASH 1400 (NUREG-75/014).

Siu, N. 1990. Dynamic accident sequence analysis in PRA: A comment on "Human reliability analysis-Where shoudst thou turn?" *Reliability Engineering and System Safety,* 29(3), 359–364. http://doi.org/10.1016/0951–8320(90)90019-J.

Risk management

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A framework for assessment of Technological Readiness Level (TRL) and Commercial Readiness Index (CRI) of asset end-of-life strategies

I. Animah & M. Shafiee
Cranfield University, Bedford, Bedfordshire, UK

ABSTRACT: A substantial number of industrial assets within the manufacturing, power generation, transportation, oil and gas, petrochemical processing, mining and construction sectors are facing operation beyond their anticipated design life and will be in need of intensive maintenance services in the coming years. At the end of an asset's design lifetime, the operators must make a decision on either rejuvenating the components through life-extension solutions or decommissioning the asset. This means that life extension policies (e.g. remanufacturing, reconditioning, repurpose, retrofitting) and decommissioning strategies (e.g. recycling and disposal) will continue to play a crucial role in the future management of industrial assets. However, some of the End-of-Life Management Strategies (ELMS) or their emerging technologies may not be mature yet, and therefore application of such strategies can cause extensive uncertainties. A well-documented Technological Readiness Level (TRL) and Commercial Readiness Index (CRI) for these strategies and related technologies will be a key in reducing the uncertainties involved in implementing ELMS in various industries. This paper aims to propose a systematic framework consisting of six different processes to help asset managers evaluate the TRL and CRI of different ELMS and their corresponding technologies. An essential part of developing this framework is the strong collaboration among academics and industrial experts with several years of experience in undertaking life extension and decommissioning projects. For purpose of illustrating the model, a case study involving end-of-life strategies of wind turbines is provided and the results are further discussed. The data required for this study is collected from various sources, including the published literature and industrial reports as well as by surveying academic and industrial experts. The results of this study indicate that TRL and CRI assessments are not only an effective means of evaluating the technological status of different ELMS but also a means for risk management decision making.

1 INTRODUCTION

Over the past several decades, asset owners in the manufacturing, power generation, transportation, oil and gas, petrochemical processing, mining and construction industries have focused on optimizing the design of their systems, enhancing installation techniques and improving production. However, in recent years, many of the assets operating in the above-mentioned industries are entering a new phase of development where assets are expected to reach their anticipated design lifetime. Hence, the attention of industries is now shifting towards how these ageing assets can be managed, to ensure that they continue to deliver high level of service beyond their original design lifetime.

The two most popular end-of-life management strategies (ELMS) include asset life extension and decommissioning. Asset life extension involves the application of technical and administrative procedures to extend the useful life of engineering structures, systems and components at the end of their design lives, provided they are technically and economically qualified. The benefits of extending the service life of ageing assets are enormous. For instance, extending the service life of a multi-million pound system could result in substantial return on investment. It also has the tendency to increase production volume and reduce CO_2 emissions due to slow down in the manufacturing of new products.

On the other side, decommissioning represents the last stage of the asset life cycle. It involves the total or partial removal of assets, which ensures the restoration of a site (land or seabed) to suitable condition for other uses and also maximizes material recovery from removed assets through waste management. Despite the huge potential of life extension and decommissioning to asset owners, the life extension policies and decommissioning strategies as well as their emerging technologies are not yet matured. Therefore, the application of these ELMS in many sectors often result in huge technological and commercial uncertainties.

In order to minimize the technological and commercial uncertainties involved in implement-

ing ELMS in various industries, it is important for asset owners to explore, assess and evaluate the maturity level of life extension policies and decommissioning strategies for different systems and components, taking into account technological readiness level (TRL) and commercial readiness index (CRI) of related technologies.

The TRL was first developed by NASA in 1974 as a benchmarking tool to assess and communicate the maturity levels of new technologies (Mankins, 2009). Since then, it has been applied in various industries to provide a measurement of technology maturity. Although the TRL concept is appropriate to help minimize technological uncertainties, there are often commercial or financial uncertainties characterising new programmes and technologies entering the market. Another accepted process that can be used for benchmarking the commercial maturity of new technologies is the commercial readiness index (CRI). The CRI was developed by the Australian Renewable Energy Agency (ARENA) to evaluate the commercial readiness level of renewable energy technologies.

Understanding and communicating the TRL and CRI for life extension policies and decommissioning strategies as well as their related technologies will not only enable asset owners to reduce the uncertainties involved in implementing ELMS in various industries but also can help them to determine how key assets could be competitive beyond their original design life.

The aim of this paper is to propose a framework to evaluate the maturity level of different life extension policies and decommissioning strategies by taking into account the TRL and CRI of related technologies. The proposed framework is tested with a case study involving offshore wind turbine blades that have reached the end of their original design lives. Our results indicate that the proposed framework provides a powerful decision-making tool to support asset owners to efficiently manage engineering structures, systems and components when they reach the end of their original design life.

The rest of the paper is organized as follows. Section 2 presents the conceptual framework and the steps required to aid in assessing the maturity levels of life extension policies, decommissioning strategies and the related technologies. In Section 3, the model is applied to a case study involving wind turbine blades and the results are analyzed in Section 4. Finally, the conclusions and future research directions are outlined in Section 5.

2 THE PROPOSED FRAMEWORK

The proposed framework for evaluating the maturity level of life extension policies, decommission strategies and emerging technologies is shown in Figure 1, which includes five steps. These steps are explained in details as follows:

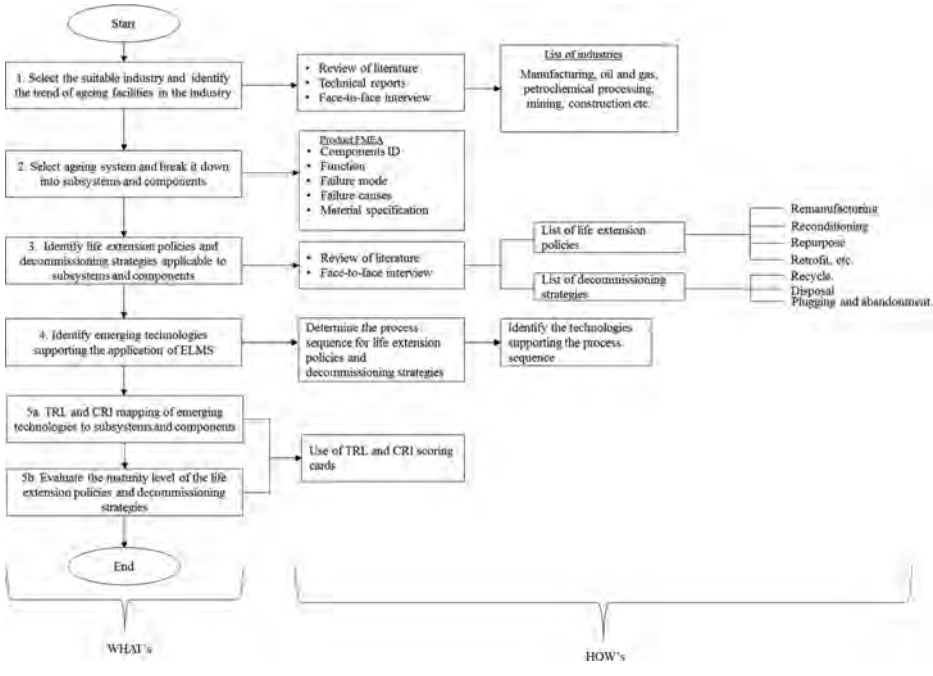


Figure 1. Steps for applying the proposed framework.

Step 1. Select the suitable industry and identify the trend of ageing facilities: As the purpose of this study is to evaluate the maturity level of different life extension and decommissioning policies as well as their related technologies to support end-of-life management of structures, systems and components within different industries, it is key to investigate the trends in the number of facilities approaching, have reached or exceeded the end of their original design life within a specific industry or company. This task can be achieved by reviewing journal articles, conference papers, technical reports, company's dossier and face-to-face interview with industrial experts. The industries/companies that can apply the proposed framework to support end of life management of ageing assets include (but not limited to) the following: power generating (nuclear energy, renewable energy and power transmission and distributions), manufacturing, transportation, oil and gas, petrochemical processing, mining, construction, etc.

Step 2. Select an ageing facility and break it down into subsystems and components: In this stage, an ageing facility is selected and decomposed into manageable units. Decomposition of ageing facility into manageable units is to facilitate the mapping of applicable life extension policies or decommissioning strategies for high-risk subsystems and components of the facility. The task of breaking down a facility into manageable units can be achieved through the use of product failure mode and effect analysis (FMEA).

Step 3. Map applicable life extension policies and decommissioning strategies to subsystems and components: At this stage, all applicable life extension policies and decommissioning strategies are identified and mapped to critical subsystems and components of the facility. The life extension policies and decommissioning strategies can be identified though literature review as well as consultations with industry experts. Examples of life extension policies include remanufacturing, reconditioning, repurpose, retrofitting, etc. while recycling, disposal and plugging and abandonment are examples of decommissioning strategies. For more comprehensive description of different life extension policies, readers can refer to Shafiee and Animah (2017).

Step 4. Identify emerging technologies for implementing life extension policies and decommissioning strategies: The goal of this stage is to identify the technologies related to the implementation of the applicable life extension policies and decommissioning strategies for the critical subsystems and components. In order to identify the related technologies, one must understand the process sequence for each life extension policy or decommissioning strategy. Understanding the process sequence for different life extension policies and decommissioning strategies will help in identifying the appropriate technologies and their features and provide specific information that allows for the allocation of TRL and CRI respectively.

The process sequences involved in extending the useful life of an engineering structure, system or component include cleaning, disassembling and inspection, repair, reassembling and testing. Whereas cutting, lifting, removal and material recovery/disposal may constitute decommissioning process sequences for systems and components, especially within the offshore oil and gas and offshore wind power industries.

Step 5. Evaluate the maturity level of the life extension policies and decommissioning strategies: This task of the proposed framework is performed by utilizing the TRL and CRI assessment scales, shown in Figure 2. The TRL scale helps to assess the technical maturity of the life extension policies and decommissioning strategies as well as their related technologies whereas the CRI scale assists in evaluating the commercial readiness of the strategies and the related technologies.

There are two rounds of activities at this stage. In the first activity, the TRL and CRI of the emerging technologies supporting the implementation of life extension policies and decommissioning strategies for subsystems and components are determined using Eq. (1) and (3) below:

$$TRL_i = \sum_{j=1}^{n} w_{ij} TRL_{ij} \tag{1}$$

$$\sum_{j=1}^{n} w_{ij} = 1; for \forall_i \tag{2}$$

where w_{ij} is the relative importance (weight) of technology j for policy/strategy i, and

$$CRI_i = \sum_{j=1}^{n} v_{ij} CRI_{ij} \tag{3}$$

$$\sum_{j=1}^{n} v_{ij} = 1; for \forall_i \tag{4}$$

where v_{ij} is the relative importance (weight) of technology j for the policy/strategy i. The weights w_{ij} and v_{ij} can be allocated by experts or estimated

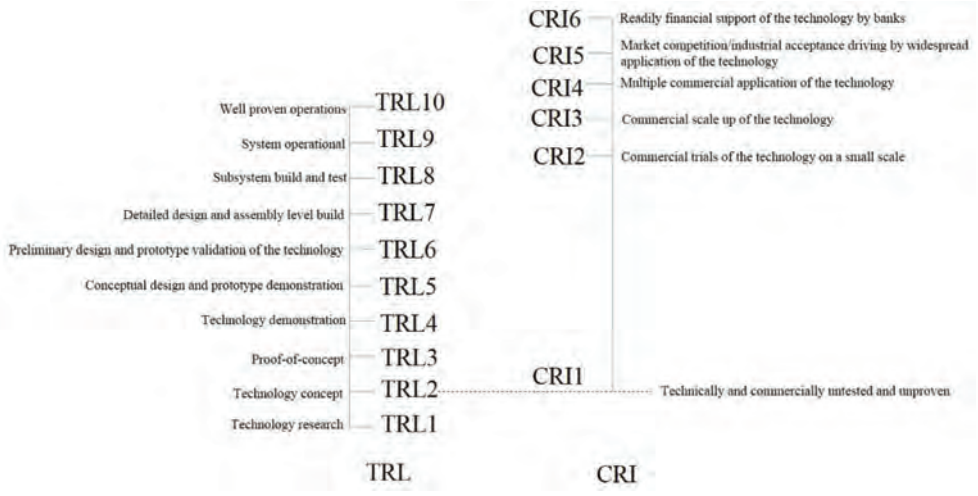


Figure 2. Technology readiness level (TRL) and commercial readiness index (CRI) scale (ARENA, 2014 and Straub, 2015).

using the Delphi-analytical hierarchy process. The second activity involves evaluating the maturity level of each life extension policy and decommissioning strategy (M) using Eq. (5):

$$M = \sum \alpha TRL_i + (1-\alpha)CRI_i, 0 < \alpha < 1 \quad (5)$$

where α and $1-\alpha$ represent the relative importance (weight) of TRL and CRI in relation to each other.

3 APPLICATION TO WIND TURBINE BLADES

The number of wind turbines reaching the end of their original design life of 20–25 years has been increasing in recent years, hence repowering, life extension and decommissioning activities are attracting the attention of both practitioners and scholars in the wind power industry. In order to illustrate the efficacy of the proposed framework, it is applied to offshore wind turbine blades which have reached the end of their original design lifetime. The data used in this study was collected from literature, academics with expertise in wind energy industry as well as blade manufacturers. In this Section, the results of the application case are presented and discussed.

Step 1. The wind power industry is the focus of this case study.

Step 2. From the product FMEA of the wind turbine provided by the operator, the proposed framework is applied to the blades, which are made of composites.

Step 3. The life extension policies and decommissioning strategies considered for the blades are briefly explained below.

- Remanufacturing: It involves the use of modern technologies and procedures to break assets down into core parts, and then engineering changes are made to the core parts in order to meet Original Equipment Manufacturers (OEM) specifications and performance. In most cases, the cost of a remanufactured system is less than that of a brand new system but with an equivalent warranty to that of a brand new system (Animah *et al.*, 2017).
- Reconditioning: According to Shafiee and Animah (2017) reconditioning involves taking appropriate actions to restore a defective system to between "as good as new (AGAN)" and "as bad as old (ABAO)" condition, hence the output is less than that of OEM's stated output.
- Repurpose: It involves taking the necessary steps to create a new use or purpose for an existing system which was originally designed for a different purpose (Aguirre, 2010; Coughlan *et al.*, 2015). With this life extension policy, engineering actions and processes are applied to make transformation to key parts of a system. The transformation is based on the technical feasibility of the engineering processes, environmental performance and economic viability of the system for future operations (Bauer *et al.*, 2017).
- Retrofitting: It involves the process of replacing old components of a system with modern

Table 1. Overview of technologies used to support remanufacturing, reconditioning, repurpose and retrofitting process for blades.

Process sequence	Technologies	Technology features
Cleaning	Jet cleaning (water jet, abrasive (sand) blasting cleaning, dry ice cleaning etc.).	With this technology, rust, grease, oil and other contaminants are removed from surfaces of components through physical interaction of the accelerated medium (sand, water, dry ice etc.) by compressed air or high pressured water (Liu et al., 2013).
	Organic solvent cleaning	This method of cleaning is performed by drenching/soaking components in organic solvent or spraying the organic solvent on component surfaces while cleaning takes place as a result of dissolution and chemical reaction (Kikuchi et al., 2011).
	Ultrasonic cleaning	Ultrasonic cleaning involves the use of high frequency (20–400 KHz) sound waves to generate agitation in a liquid (Niemczewski, 2007). The cavitation bubbles produced as a result of the agitation acts on contaminants on the surfaces for cleaning.
Disassembly	Disassembly embedded design	Disassembly embedded design integrates a disassembly mechanism into a system during design, e.g. Snap fits to dislodge locked out of position (Soh et al., 2014).
	Active disassembly	This technique makes use of external triggers such as temperature, magnetic force or pressure, for the release of fasteners. It include utilizing smart materials, freezing elements, soluble elements, pneumo-elements and hydrogen storage alloy elements as a fastening technique (Duflou et al., 2006).
Repair/ Modification/ Retrofitting process	Laser repair technology (LRT)	LRT is suitable for the repair/rebuilding of worn out metallic parts, which are considered non-repairable using traditional welding or plating techniques. The process involves injection of metal powder into a focused beam of a high powered laser in a tightly controlled atmospheric condition.
	Advanced mechanical machining processes	This technology is needed when the repair of existing components require CNC milling, turning, drilling, tapping, cutting and other machining operations.
	Industrial 3D printing	This is an additive manufacturing process which uses high powered laser fusion to produce components by building layer upon layer from a predefined 3D digital design.
	Tubercle technology	This technology mimics the bumps on humpback-whale fins to develop a more efficient wind turbine blades.
	Plastic surgery	This the use of plastic surgery to make old, smaller and less efficient wind turbine blades into bigger and efficient ones without replacing them at the end of life.
Testing	Shearography system	This is the use of optimal measurement technology or shearorgraphy sensor to detect defects such as wrinkles, delaminations, debondings and kissing bondings in wind turbine blades.

Table 2. Overview of technologies used to support recycling and disposal of wind turbine blades.

Process sequence	Technology	Technology features
Material recovery	Pyrolysis	This is the use of pyrolysis recycling technology (i.e. burning of the resin matrix with limited oxygen).
	Mechanical grinding	This is the reduction of composite materials into suitable sizes and grounding them into different grades using mechanical processes such as hammer mill.
	Solvolysis	This technology makes use of physico-chemical separation process for full material recovery of thermosets. The process uses water at sub-supercritical temperature to breakdown thermoset resin, in order to remove it from the fibre.
	Incineration	This is the burning of end of life product for energy generation.
Disposal	Landfill	Burying the materials used in the manufacturing of the system in the ground.

equivalent in order to achieve higher functionality and availability. An example of retrofitting is the blade extension program for wind turbines, in order to increase the swept area of the rotor to increase power generation at low speed.

– Recycle: Complex industrial equipment are made of different materials. In post-decommissioning, scraps from these infrastructure in the form of metals and non-metals are generated, recycling processes are then used to recover high grade materials from these scraps for other uses, thereby limiting the amount of waste generated at the end of life (Yang *et al.*, 2012). However, recycling of the products can reduce the exploitation of natural resources and protect the environment.
– Disposal: Open burning and landfills are examples of the popular disposal options for components and materials from industrial systems. However, the economic value of reusing the system or recovering materials through recycling is lost when this strategy is implemented at the end of original design life.

Step 4. A number of technologies have been proposed by both researchers and practitioners to support the process sequence for implementing life extension policies and decommissioning strategies across different industries. The technologies that can support the process sequence of remanufacturing, reconditioning, repurpose and retrofitting of wind turbine blades are shown in Table 1. On the other hand, Table 2 shows the overview of the related technologies that can support the material recovery/disposal process sequence of recycle and disposal of wind turbine blades.

Step 5. The maturity level of the life extension policies and decommissioning strategies applicable to wind turbine blades using the TRL and CRI scales in Figure 2 is evaluated. The TRL and CRI of each technology was evaluated through experts' elicitation. First, the experts were asked to allocate TRL and CRI to the emerging technologies supporting the implementation of the life extension policies and decommissioning strategies obtained from the literature. Table 3 shows the TRL and CRI of the emerging technologies supporting the implementation of life extension policies and decommissioning strategies for wind turbine blades using Eqs. (1) and (3). The weights assigned to evaluate *TRL* and *CRI* for each technology were achieved through a consensus reached by a panel of experts. Second, the maturity level of each life extension policy and decommissioning strategy has been evaluated using Eq. (5) and are ranked in Table 4.

As shown in Table 4, disposal was chosen as the most appropriate ELMS for wind turbine blades. This is because the maturity level of landfill as a technology for disposing the blades made of composites are technically well proven, commercially available and relatively cheap in many parts of the world. The second and third most suitable ELMS for the wind turbine blades are remanufacturing and reconditioning. Retrofitting, repurposing and recycling were identified as the fourth, fifth and sixth preferred ELMS for the blades. This is because the key technologies needed to support the implementation of these ELMS for products made of composite materials are still in the experimental phase. This means that the technologies are not technically matured for large scale commercial applications. For instance, pyrolysis, mechanical

Table 3. TRL and CRI of emerging technologies supporting the life extension and decommissioning strategies for wind blades.

Emerging technologies	TRL	CRI
Jet cleaning	10	5
Organic solvent cleaning	–	–
Ultrasonic cleaning	–	–
Disassembly embedded design	10	5
Active disassembly	7	1
LRT	5	1
Advanced mechanical machining processes for composites	2	1
Tubercle	7	1
Plastic surgery	7	1
Shearograph	9	2
Industrial 3D printing	2	1
Pyrolysis	3	1
Mechanical machines (grinding, stripping, crushing etc.)	3	1
Solvolysis	3	1
Incineration	9	2
Landfill	10	3

Table 4. Ranking of maturity level of life extension policies and decommissioning strategies for wind turbine blades.

Life extension policies/Decommissioning strategies	Ranking
Disposal	1
Remanufacturing	2
Reconditioning	3
Retrofitting	4
Repurpose	5
Recycle	6

grinding and Solvolysis which are considered as the most suitable recycling technologies for composite products such as wind turbine blades are at the laboratory scale in terms of technological development (Rybicka *et al.*, 2016). Thus, making landfill and incineration the most widely used technologies to support end of life management of wind turbine blades when they reached the end of their original design life.

5 CONCLUSION

In this study, a methodology was proposed to assist in understanding and communicating the maturity level of life extension policies and decommissioning strategies for industrial assets reaching the end of their original design lives. The proposed methodology considered TRL and CRI of emerging technologies supporting the implementation of different life extension policies and decommissioning strategies, in order to rank the best ELMS. To the best of our knowledge, this was the first time the TRL and CRI scales were integrated to communicate the maturity level of different life extension policies and decommissioning strategies for products reaching their end of life. For the purpose of clearly illustrating the efficacy of the proposed framework, it was applied to determine the maturity level of different life extension polices and decommissioning strategies for wind turbine blades. The findings from the case study indicated that disposal of wind turbine blades through landfilling was considered as the most appropriate strategy for managing wind turbine blades when they reached the end of their original design life. This was followed by remanufacturing, reconditioning, retrofitting, repurpose and recycle.

REFERENCES

Aguirre, D. (2010). Design for repurposing: A sustainable design strategy for product life and beyond. In: *Industrial Designers Society of America*. pp. 1–25.

Animah, I., Shafiee, M., Simms, N. and Tiwari, A. (2017). A multi-stage remanufacturing approach for life extension of safety critical systems. *Procedia CIRP*, 59, 133–138.

ARENA. (2014). Commercial readiness index for renewable energy sectors.

Bauer, T., Brissaud, D. and Zwolinski, P. (2017). Design for high added-value end-of-life strategies. In: *Sustainable Manufacturing.* (pp. 113–128). Springer International Publishing.

Coughlan, D., Fitzpatrick, C. and McMahon, M. (2015). Repurposing E-waste as a driver for desource efficiency. In: *Proceedings of the 29th EnviroInfo and 3rd ICT4S,* 7–9 September, Copenhagen, Denmark, p. 238.

Duflou, J.R., Willems, B. and Dewulf, W. (2006). Towards self-disassembling products design solutions for economically feasible large-scale disassembly. In: *Innovation in Life cycle Engineering and Sustainable development*, pp. 87–110.

https://invrecovery.org/remanufacturing-the-future-of-sustainability/. (accessed 24.10.2018).

Kikuchi, E., Kikuchi, Y. and Hirao, M. (2011). Analysis of risk trade-off relationships between organic solvents and aqueous agents: case study of metal cleaning processes. *Journal of Cleaner Production*, 19(5), 414–423.

Liu, W., Zhang, B., Li, M.Z., Li, Y. and Zhang, H. (2013). Study on remanufacturing cleaning technology in mechanical equipment remanufacturing process. In: *20th CIRP International Conference on Life Cycle Engineering,* 17–19 April, Singapore, pp. 644–648.

Mankins, J.C. (2009). Technology readiness and risk assessments: A new approach. *Acta Astronautica*, 65(9–10), 1208–1215.

Niemczewski, B. (2007). Observations of water cavitation intensity under practical ultrasonic cleaning conditions. *Ultrasonics Sonochemistry*, 14(1), 13–18.

Rybicka, J., Tiwari, A. and Leeke, G.A. (2016). Technology readiness level assessment of composites recycling technologies. *Journal of Cleaner Production*, 112, 1001–1012.

Shafiee, M. and Animah, I. (2017). Life extension decision making of safety critical systems: An overview. *Journal of Loss Prevention in the Process Industries*, 47, 174–188.

Soh, S.L., Ong, S.K. and Nee, A.Y.C. (2014). Design for disassembly for remanufacturing: Methodology and technology. *Procedia CIRP*, 15, 407–412.

Stirling, W. (2016). The value of remanufacturing. https://www.theengineer.co.uk/the-value-of-remanufacturing/. (accessed 24.10.2018).

Straub, J. (2015). In search of technology readiness level (TRL) 10. *Aerospace Science and Technology*, 46, 312–320.

Yang, Y., Boom, R., Irion, B., Heerden, D. Van, Kuiper, P. and Wit, H. De. (2012). Chemical engineering and processing : Process intensification recycling of composite materials. *Chemical Engineering & Processing: Process Intensification*, 51, 53–68.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Engineering safety recommendations: Results from a survey in aviation

N. Karanikas
Aviation Academy, Faculty of Technology, Amsterdam University of Applied Sciences, The Netherlands

ABSTRACT: Taking into account the lack of uniform guidelines for the design and classification of safety recommendations, a relevant framework was developed according to academic and professional literature. The framework includes nine design criteria for recommendations, it incorporates classifications of their scope and expected effectiveness, and it was used to perform a questionnaire survey across aviation professionals involved in the generation of safety recommendations. The goal of the survey was to capture (1) whether practitioners are knowledgeable about the design criteria, (2) the degree to which they apply those criteria along with corresponding reasons, (3) perceptions of the expected effectiveness of types of controls introduced through recommendations, (4) the frequency of generating each control type and respective explanations, and (5) the extent to which practitioners focus on each of the categories of recommendations' scope and the relevant reasons. Overall, the results showed: an adequate level of knowledge of the design criteria; a strong positive association of the knowledge on a particular criterion with the degree of its implementation; a variety of frequencies the recommendations are addressed to each of the scope areas; a reverse order of perception of the expected effectiveness of control types compared to the literature suggestions. A thematic analysis revealed a broad spectrum of reasons about the degree to which the design criteria are applied, and the extent to which the various types of recommendations are generated. The results of the survey can be exploited by the aviation sector to steer its relevant education and training efforts and assess the need for influencing the direction safety recommendations are addressed. Similar research is suggested to be conducted by organizations and regional and international agencies of any industry sector by ensuring a larger sample.

1 INTRODUCTION

Safety investigations play a crucial role in safety improvements especially because they lead to the formulation of recommendations to eliminate or mitigate identified problems with the scope to prevent similar occurrences in the future. To date, although in aviation there are established guidelines for the conduction of a safety investigation (ICAO, 2003, 2008, 2011, 2015), there is yet little guidance about the design of safety recommendations (Pooley, 2013).

To fill this gap, researchers and students from the Aviation Academy of the Amsterdam University of Applied Sciences (Zonneveld, 2016, De Vos, 2016, Kiefer, 2016) reviewed academic and professional literature and proposed a relevant framework that includes design criteria (Table 1), their scope (Table 2) and expected effectiveness (Table 3).

The particular framework was used to perform a questionnaire survey across safety practitioners to explore the extent to which the framework aspects are known and applied, and reveal any underlying reasons.

Table 1. Design criteria for safety recommendations.

Criterion	Brief explanation	Literature*
Specific	Addresses a particular problem	(Haughey, 2014, Gregson, 2017)
Measurable	Allows monitoring of its implementation	(Haughey, 2014)
Assigned	Addressed to specific responsible agent(s)	(Haughey, 2014, Gregson, 2017)
Realistic	Achievable within current boundaries	(Haughey, 2014, Gregson, 2017)
Time-bound	End dates defined	(Haughey, 2014)
Review	Review dates defined	(Haughey, 2014)
Objectives	What and not how to achieve	(Johnson, 2003, Gregson, 2017)
Action-oriented	Actionable items are preferred over studies	(Johnson, 2003)
Non-blaming	Focus on problems, not individuals	(Johnson, 2003, Dekker, 2016)

*Indicative literature references.

Table 2. Scope of safety recommendations*.

Dimension	Category	Brief explanation
Aspect of operations	Process structure	Oriented to low-level tasks (Re)design of system's architecture and functionality
	Culture context	Change of norms and behaviours
		Focus on politics and the society
Stakeholders affected	Macro-level	Governments, associations etc.
	Meso-level	Industry sector(s)
	Micro-level	Organizations and individuals
Degree of renewal	Repair	Short fixes of local problems
	Adaptation	Improvement of larger systems
	Innovation	Creation of new solutions

*Adapted from ESREDA (2015).

Table 3. Expected effectiveness of safety recommendations.

Type of control introduced*	Brief explanation
Physical	Prevent completely actions or access
Functional	Use of technology to limit actions
Symbolic	Means to alert for hazards or remind/train rules, procedures etc.
Incorporeal	Strategies, general policies, legislation

*In descending order of robustness (Hollnagel, 1999).

2 METHODOLOGY

Following the establishment of the theoretical framework about the design and classification of safety recommendations as presented above, the researcher aimed at assessing the degree to which the aspects of the framework are known and/or applied by safety investigators and professionals involved in the formulation of safety recommendations in general (e.g., safety and risk managers). The aspects of the framework comprised the topics of a questionnaire that was administered to safety managers, investigators and professionals. The data collected from the analysis with the tool and the questionnaire responses were statistically processed to obtain an overall picture and examine differences across various variables.

2.1 *Survey questionnaire*

The survey instrument was designed with the goal to capture the following information from practitioners involved in the generation of safety recommendations: whether they are knowledgeable of the design criteria and the degree to which they apply those in daily practice along with possible reasons; perceptions about the order of effectiveness of the control types referred in literature, frequency of proposing each control type as part of their role, and explanations about the latter; extent to which they focus on each of the categories included in the three dimensions of recommendations' scope and respective reasons.

The questionnaire included an introductory section where the background and aims of the study were stated along with the voluntary character and anonymity of participation. Also, the particular section referred to the estimated time investment (i.e. up to 15 minutes) and the contact details in the case that the respondents wanted to provide feedback on the questionnaire, get informed about the results of the study or raise any other inquiry.

To examine possible variations of the responses against characteristics of the sample, the subjects were asked to fill in their main job role at the time of participation (i.e. safety manager/officer, safety investigator, or other), the year they started getting involved actively in the generation of safety recommendations, the country they were practicing their vocation at the time of participation, and their highest level of education (i.e. High School, Associate Degree, Bachelor degree, Master degree, Doctoral level, and other).

For each of the design criteria for safety recommendations, the respondents were asked to state whether they know the criterion (possible choices: YES or NO), the extent to which they apply the criterion when creating safety recommendations (possible choices: 0–20%, 21–40%, 41–60%, 61–80%, and 81–100%), and explanations about their latter. Regarding the expected effectiveness of recommendations, the corresponding section provided a brief description and a few examples for each of the control types (i.e. physical, functional, symbolic and incorporeal) and asked the participants to rank the controls in the order of their effectiveness, state which control type they most frequently introduce in their safety recommendations, and justify their last answer. It is noted, that the control types were presented to the participants in a random order of effectiveness outlined in the literature (Hollnagel, 1999). The last section of the instrument referred to the scope of recommendations and included a short description for each of the dimensions (i.e. aspects of operations, stakeholders affected and degree of renewal) and their values (see Table 2). The subjects were prompted to choose the frequency to which their recommendations focus on each of the categories of the three dimensions (possible choices: 0–20%, 21–40%, 41–60%, 61–80%, and 81–100%) and state respective reasons.

It is clarified that there were no obligatory questions to be answered. The respondents could omit any demographic or safety recommendation related question. The draft version of the survey instrument was sent to four persons with relevant academic and professional background for their review. Following the revision of the questionnaire according to the remarks collected, its final version was designed online with the use of the Qualtrics platform. The functionality of the online questionnaire was tested with the participation of the same four reviewers. The survey instrument was administered through two main channels: (1) personal emails to contact persons of the network of the Aviation Academy of the Amsterdam University of Applied Sciences that covers various aviation organizations worldwide; (2) online messages to practitioners found on the LinkedIn platform and holding relevant positions (e.g., safety managers, safety investigators). Due to time constraints, three working days were devoted to the administration of the questionnaire and a three weeks period was set for the collection of responses.

2.2 *Sample and analysis of questionnaire responses*

In total, 42 questionnaires were filled. Because of the snowball sampling strategy, the author could not have any information about the number of persons whom the instrument finally reached (e.g., unmonitored or unread emails and messages). Therefore, the response rate could not be estimated. Nevertheless, since the scope of the whole study was an initial assessment of the situation around safety recommendations, the number of responses was deemed as sufficient. Table 4 presents the distribution of the sample across its demographic characteristics. It is clarified that the apart from the main job role, the rest of the demographics were grouped due to the small number of responses in some of the categories.

Table 4. Distribution of the sample.

Demographic variable	Values	Sample size	Valid percentage*
Main job role	Safety staff	18	42.9
	Safety investigator	13	31.0
	Other	11	26.1
Years involved in generation of recommendations	< = 4	10	25.0
	5–11	11	27.5
	12–18	10	25.0
	> = 19	9	22.5
Geographical region	Europe	30	71.4
	Other	12	28.6
Highest level of education received	< = Bachelor	18	43.9
	> = Master	23	56.1

*Some demographic questions were not answered.

Regarding the analysis of data, the frequencies for each the closed questions were calculated to offer an overall view of the responses. Fisher's exact tests were performed to reveal any associations of the knowledge of design criteria with the variables of Table 4. The same variables were also used to conduct Kruskal-Wallis or Mann-Whitney tests (i.e. depending on the number of categories of each variable) for the closed questions which corresponded to ordinal data (i.e. frequency of application of design criteria, frequency of focus of recommendations on the areas defined, and level of effectiveness of controls the recommendations introduce).

It is noted that, to allow the execution of statistical tests, the frequency choices were translated to ordinal figures as follows: 1: 0–20%, 2: 21–40%, 3: 41–60%, 4: 61–80%, and 5: 81–100%. The tests were run with the SPSS software version 22 (IBM, 2013). The function of Monte Carlo Exact Test under the settings Confidence Level: 99% and Number of Samples: 10.0000 was chosen to strengthen the validity of the results. The level of statistical significance was set to 0.05.

The open-ended questions concerned, a thematic analysis was performed. The researcher individually performed a coding of the answers, which was afterwards tested for reliability with two other colleagues who were not involved in the study. The comments of the raters indicated areas of disagreement as well as cases that the content of the answers had not been captured by the initial codes. Based on these remarks, the coding was revised and retested, resulting in agreement levels ranging from 77% to 92% between the researcher and each of the participants as calculated with Cronbach Alpha tests. The finalization of the list of coding themes was followed by the calculation of frequencies per code for each of the questionnaire topics.

It is clarified that in several qualitative responses the subjects did not provide explanations about their choices in the closed questions, but they restated the latter or made general comments not applicable to the particular question. These cases were excluded from the analysis. Due to the small number of valid responses, no statistical tests were conducted between the responses and the demographic characteristics of the participants.

3 RESULTS

3.1 *Results from analysis of closed questions*

The frequencies that the design criteria of safety recommendations are known and the extent to which are applied by the survey participants are shown in Table 5. The Non-blaming, Assigned and Realistic criteria were the ones most known, whereas the Review date and Actions criteria were

Table 5. Design criteria for safety recommendations.

Criterion	Median rank (application)	Frequency (%) the criterion is known
Specific	4.5	90.5
Measurable	4.0	87.8
Assigned	5.0	97.6
Realistic	5.0	95.2
Time-bound	4.0	90.5
Review	4.0	73.8
Objectives	4.0	85.4
Actions-oriented	4.0	75.6
Non-blaming	5.0	100.0

the ones that were least known by the respondents. Also, Spearman's bivariate correlations were performed between the figures of knowledge percentage and the medians. The results of the particular statistical test showed a significant and strong association ($N = 9$, $rs = 0.870$, $p = 0.002$), meaning that the higher the knowledge on a specific criterion, the higher the degree of its implementation.

The Fisher Exact tests between the design criteria and the demographics of the population resulted in significant results only for the association of years of experience in the generation of safety recommendations with the knowledge of the criteria Realistic ($N = 40$, $p = 0.046$) and Time-bound ($N = 40$, $p = 0.008$). The participants with 19 or more years of experience in safety recommendations declared less frequently that they knew about both criteria compared to participants having fewer years of experience. Regarding the frequency of application, the Objectives criterion was applied less frequently by safety managers/officers than the rest of the job roles ($p = 0.039$) and it was utilized more by the participants with increased years of involvement in safety recommendations generation ($p = 0.025$).

Regarding the scope of recommendations, the respective medians are reported in Table 6. The statistics revealed that culture-focused recommendations were applied more frequently by participants with roles other than safety staff and investigators ($p = 0.046$). Recommendations addressed to industry sectors (i.e. meso level) were generated by safety investigators more than other job holders ($p = 0.002$). Meso—and macro-level types of recommendations were made more frequently by subjects working in Europe ($p = 0.044$ and $p = 0.021$ respectively) or having a high educational background ($p = 0.008$ and $p = 0.004$ correspondingly).

The expected effectiveness of safety recommendations concerned, Table 7 presents the survey findings with regard to the perceived degree of effectiveness for each type of control introduced through recommendations. The statistics showed

Table 6. Distribution of scope areas of generated recommendations.

Dimension	Category	Median rank (application)
Aspect of operations	Process	4.0
	Structure	3.0
	Culture	2.0
	Context	2.0
Stakeholders affected	Macro level	2.0
	Meso level	3.0
	Micro level	4.5
Degree of renewal	Repair	4.0
	Adaptation	4.0
	Innovation	2.0

Table 7. Perceived effectiveness of safety recommendations.

Type of control introduced	Median rank
Physical	2.0
Functional	2.0
Symbolic	3.0
Incorporeal	3.0

no associations of the perceived effectiveness with the demographic characteristics of the sample.

3.2 *Results from analysis of open-ended questions*

The codes derived from the thematic analysis and their frequencies showed that in many cases the subjects applied the design criteria because they were mentioned in internal or external documentation or had been seen as best practice. It was widely recognized that the Specific criterion minimizes ambiguity in the implementation of safety recommendations, which according to a few respondents is expected at some degree. However, 3 out of the 20 participants declared that some flexibility is required and recommendations should not be always too specific.

The Measurable criterion was seen by most of the subjects as often unfeasible and a few respondents stated that the effect of changes is more important than their measurement, monitoring does not apply to simple recommendations, and a customization of relevant metrics to each organizational level/function is necessary. About the Assigned criterion, 3 out of the 18 participants argued that recommendations might require the engagement of more than one responsible persons, departments, agencies etc. The comments made about the Realistic criterion showed that this helps in increasing the credibility and feasibility of the recommendation and demonstrating an achieve-

ment of a balance between the resources required for its realization and the anticipated benefits. However, 5 out of the 23 answers pointed that the Realistic criterion might be difficult to meet due to the diversity of perspectives and interests of the stakeholders involved.

The Time-bound criterion concerned, the participants expressed a variety of views. Four persons recognized that the specific criterion would ensure the implementation of a recommendation, whereas three persons stated their reservations about the feasibility of the criterion. Also, two persons did not know the particular criterion and one person did not contemplate it as important. A similar picture was observed in the answers regarding the knowledge and feasibility of the Review date and Objectives criteria. The latter criterion was appreciated by three respondents because it allows flexibility in the operationalization of the recommendation.

A combination of actions and studies depending on the type of the problem to be solved was proposed by most of the participants. Four out of the 17 subjects argued that the Action criterion increases the feasibility of a recommendation. The application of the Non-blaming criterion was seen as positive by most of the participants regarding the effects on the overall culture and increase of the effectiveness of safety recommendations. Only 1 out of the 23 answers suggested that a focus on individuals/teams might be appropriate in cases of repeated unsafe behaviors.

The answers regarding the type of control most frequently introduced by the survey respondents showed a preference in symbolic controls, followed by incorporeal, functional and physical ones. The explanations given emphasize on the decreased feasibility of controls of technical nature (i.e. physical and functional) due to the demands for more resources for their realization. The views about the effectiveness of technical or non-technical controls (i.e. symbolic and incorporeal) were almost evenly distributed. In three cases, the respondents argued that symbolic and incorporeal controls are more vulnerable and require more frequent repairs. One respondent claimed that the focus on symbolic controls sources from the pressure of authorities who ask for more and better procedures.

Regarding the focus of recommendations on specific aspects of operations, 4 out of the 17 answers addressed that the decision depends on the context of the problem identified and four of the respondents answered that is a matter of organizational focus. The views about the process and culture aspects of operations were divided into two equal parts. Four subjects claimed that those aspects are difficult to change whereas other four subjects stated that the specific aspects are easier and faster to change. Concerning the degree of renewal, repairs and adaptations were seen as equally sufficient to deal with problems, partly because of the lower associated costs. Innovations were contemplated by 4 out of the 14 participants as either expensive or not appropriate to be introduced through recommendations, and two participants suggested that the required degree of renewal depends on the problem under examination.

Lastly, regarding the level of affected stakeholders, 8 out of the 17 subjects stated that their focus on the micro level is dictated by the organizational priority to deal with local problems and the difficulty to affect the meso and macro levels. One respondent recognized that micro level interventions are cheap and another respondent stated that interventions at the lowest level could lead to changes to the rest of the levels. Once more, many subjects declared that the type of each problem is as a parameter to decide which stakeholders will be addressed in a recommendation.

4 DISCUSSION

The overall results regarding the design of safety recommendations suggest that most of the participants are knowledgeable about the criteria identified in the academic and professional literature and apply these criteria to at least 60% of the recommendations but with different extents. Rather expectedly, the knowledge of a specific criterion was positively associated with the degree of its application when generating safety recommendations. However, such knowledge was obtained more through experience, best practice or organizational documentation rather than international and regional standards. The lack of guidelines in industry standards might explain partially the fact that some of the criteria were not known and consistently applied by a fraction of the survey participants. The statistical tests showed a few differences across some criteria, mainly linked to the years of experience in generating recommendations. Two of the criteria were less known and another criterion was more frequently applied by participants with higher experience.

The Non-blaming criterion was known and applied by all participants, thus indicating that the necessity for a just culture in aviation (Humpreys, 2014, Michaelides-Mateou and Mateou, 2016, Quinn, 2007) and the role that recommendations can play in realizing a culture of fairness have been well communicated across the particular industry sector. The criteria of Assigned and Realistic also scored very high regarding the level of familiarity of the respondents and the degree of their application, this partially attributed by the participants to the contribution of these criteria to increased credibility of recommendations. On the other hand, the Review and Actions criteria were not

known by about one-quarter of the survey participants. These criteria were seen as important but sometimes not feasible or binary regarding their application. In general, the respondents expressed concerns about the strict satisfaction of the whole set of criteria for every single recommendation and argued that each case must be handled differently.

Concerning the type of controls introduced through safety recommendations, interestingly, their effectiveness as perceived by the participants is not aligned with the literature suggestions, the results showing a reverse order. Whereas the work of Hollnagel (1999) implies that physical and functional controls are more robust and effective, the respondents viewed the symbolic and incorporeal controls as such. The comments collected showed that the viewpoints about the effectiveness of technology and non-technology based controls were evenly divided. According to the participants, the main factors driving the recommendation of a specific control type are the resources available and the expectations of the authorities for improved procedures. Hence, the author contemplates that the responses about the effectiveness of controls were affected by current practice and not a consideration of their technical characteristics or potential to deal with a hazard or risk more successfully.

Nevertheless, the positions of the participants regarding the types of controls preferred are in tandem with literature suggesting that the amendment of rules or introduction of new ones are favored due to the low costs involved and the timeliness and easiness of implementation (Bourrier and Bieder, 2013). Moreover, the viewpoints of the respondents seem to confirm their efforts to generate recommendations that are realistic by taking into account the existing boundaries (e.g., resources, operational needs).

The findings regarding the scope of recommendations showed that the higher and wider the level of operations, stakeholder and renewal, the less frequency corresponding changes are suggested. Recommendations that focus on the physical processes and the work floor, specific organizations and repair of problems had been more frequently introduced by the participants. On the other hand, interventions at the culture and context aspects of the operational area, suggestions to the regulatory and standardization levels and recommendations for innovative solutions were least frequently generated. The comments of the respondents seemed to agree about the effects of cost and time limitations along with political factors that can discourage the formulation of recommendations addressing wider and deeper systemic flaws. A few conflicting statements were observed regarding the easiness to make process or cultural changes, this indicating somehow misaligned perceptions.

The statistics regarding differences of the recommendations' scope across the demographics included in the study showed that meso-levels of stakeholders (i.e., whole industry sectors) were addressed more frequently by safety investigators. This finding can be explained by the fact that aviation safety investigation standards (e.g., ICAO, 2003, 2011) prompt the examination of latent factors which in turn allows the formulation of recommendations targeted beyond organizational boundaries. Furthermore, recommendations for stakeholders other than specific organizations and individuals were generated more frequently by respondents working mainly in Europe and having a higher level of education. These findings might reflect the effects of different regional and national cultures (e.g., ICAO, 2013, Stolzer et al. 2008) as well as an influence of educational breadth and depth on the tendency to adopt systemic views and address deficiencies at higher system levels.

5 CONCLUSIONS

The current study was a first initiative to map the situation in the aviation industry around the engineering of safety recommendations. Regarding the degree to which professionals are knowledgeable about the design criteria that can increase the quality and effectiveness of recommendations and the extent to which apply these criteria in practice, the results suggested an adequate level of knowledge and a satisfactory frequency of application of the nine design criteria included in the research. However, the findings were attributed more to the employment of best practice and the reference of such criteria in organizational documentation, rather than to guidelines from regional or international guidelines or topics of training. Hence, the inclusion of respective material in standards and safety management/investigation courses is highly recommended to achieve consistency in the generation of safety recommendations and establish a commonly referred framework for their design. Nonetheless, such a framework must be viewed as a guide and not a compliance-check reference. As the study participants stated, each case has a different context and the satisfaction of all design criteria might not be feasible or proper for all recommendations.

The findings concerning the classification of recommendations based on the types of controls they introduce showed a gap between perceptions of the practitioners and the suggestions of the literature. The former viewed "soft" control types as more effective than "hard" ones, whereas literature implies the opposite. It is noted that the categorization of controls does not aim at stating a preference of any type over another but rendering the industry

aware of the weaknesses and strengths of each control type. Nonetheless, the conduction of further studies is suggested as a means to provide empirical evidence about the actual level of effectiveness of each type of control and indicate any optimum combinations of those.

When considering the focus areas of recommendations, the findings suggested an emphasis on the repair of lowest activity levels rather than systemic interventions and innovative solutions. A common characteristic of all topics examined in this study was the influence of resource and political boundaries on the generation of safety recommendations, which usually lead practitioners in suggesting cheap and easy to implement fixes. From a pragmatic viewpoint, the situation mentioned above is somewhat unavoidable, but this should not exclude the visible and explicit justification of choices when engineering safety recommendations. The recognition and documentation of the boundaries imposed on each case can support the aggregation of data and their monitoring to inform stakeholders accordingly and possibly lift existing limitations when situations allow.

The researcher would like to point out that the small sample of the current study does not allow to claim generalization of the results. However, the finding of this research can trigger the execution of similar studies at larger scales at regional or international levels regardless of industry sector.

REFERENCES

Bourrier, M. & Bieder, C. 2013. *Trapping Safety Into Rules: How Desirable or Avoidable is Proceduralization?* Farnham: Ashgate.

De Vos, B. 2016. *Evaluation the Quality of Recommendations Formulated in Accident Reports*. BSc graduation thesis (unpublished). Amsterdam: Aviation Academy, Amsterdam University of Applied Sciences.

Dekker, S. 2016. *Just culture: restoring trust and accountability in your organization.* (3rd ed.). Boca Raton: CRC Press.

ESReDA. 2015. *Case study analysis on dynamic learning from accidents*. Retrieved from European Safety, Reliability & Data Association: http://www.esreda.org

Gregson, M. 2017. *What Makes a Good Air Safety Recommendation?* ISASI Forum, October-December 2017, pp. 15–17

Haughey, D. 2014. *A Brief History of SMART Goals*. Retrieved from Project Smart: https://www.projectsmart.co.uk/brief-history-of-smart-goals.php.

Hollnagel, E. 1999. *Accidents and Barriers*. Retrieved from University of Linköping, Sweden: http://www.hhs.iup.edu/CJANICAK/SAFE541CJ/Barrier%20 Analysis%20 Paper.pdf.

Humpreys, K. 2014. *Just Culture: Can there be a Just Culture in Aviation Safety Occurrence Reporting Systems.* Technical paper, Adelaide: ISASI 2014 Seminar.

IBM. 2013. *SPSS Statistics for Windows*. Version 22.0. New York: IBM Corp.

ICAO. 2003. *Training Guidelines for Aircraft Accident Investigators.* Cir 298 AN/172. Canada: International Civil Aviation Organization.

ICAO. 2008. *Hazards at aircraft accident sites*. Circular 315. Montreal: International Civil Aviation Organization.

ICAO. 2011. *Manual on Accident and Incident Investigation Policies and Procedures*. Doc 9962. Montreal: International Civil Aviation Organization.

ICAO. 2013. *Safety Management Manual*. Doc. 9859. Montreal, Canada: International Civil Aviation Organization.

ICAO. 2015. *Manual of Aircraft Accident and Incident Investigation*. Doc 9756 AN/965. Part I Organization and Planning, 2nd. Montréal: International Civil Aviation Organization.

Johnson, C. 2003. *Failure in Safety-Critical Systems: A Handbook of Accident and Incident Reporting.* Glasgow: University of Glasgow Press.

Kiefer, K. 2016. *Quality Evaluation of Safety Recommendations Formulated in Aircraft Accident Reports*, BSc graduation thesis (unpublished). Amsterdam: Aviation Academy, Amsterdam University of Applied Sciences.

Michaelides-Mateou, S. & Mateou, A. 2016. *Flying in the Face of Criminalization: The Safety Implications of Prosecuting Aviation Professionals for Accidents.* Oxon: Routledge.

Pooley, E. 2013. *Harmonisation of safety recommendations*. Technical paper. Madrid: ESASI 2013.

Quinn, K.P. 2007, December. Battling Accident Criminilization. *Aerosafety World*: 11–12.

Stolzer, A.J., Halford, C.D. & Goglia, J. 2008. *Safety Management Systems in Aviation*. Aldershot: Ashgate.

Zonneveld, J. 2016. *Assessment of Safety Recommendations*, BSc graduation thesis (unpublished). Amsterdam: Aviation Academy, Amsterdam University of Applied Sciences.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Problems of mobile risks in territory

J. Prochazka & D. Procházková
Faculty of Transportation Sciences, Czech Technical University in Prague, Czech Republic

ABSTRACT: The issue of transportation of dangerous goods is addressed in international law since 1957, when international agreements on transport of dangerous goods by road ADR was created in Geneva. Directive SEVESO was put into practice in 1982, because occurrence of severe accidents involving dangerous substances in 70 s. Directive SEVESO was step by step more particularized according to practice and regulation REACH was added in 2007. Both legislation lays down requirements for handling with hazardous substances from the manufacture, transportation and storage through to their use. Series of agreements on the transport of dangerous goods to another means of transport followed. All agreements dealing with carriage of hazardous substance have one thing in common. Dangerous substances are treated as goods and all responsibility is bear by carrier. Critical evaluation of the existing rules and traffic accidents with hazardous substances shows that in practice there are missing tools, which would: effectively reduce the distortions of security measures that arise on the part of carriers; to ensure rapid response and the protection of other road users on the roads and railways; protecting people and environment around the place of traffic accident; and the recovery of the territory afflicted by impact of road traffic accidents with hazardous substance. Analysis of a database of accidents involving dangerous substances shows number of examples where the recovery of territory after accident with hazardous goods carried more than 10 years. Therefore, we must create system tools for coping with those risks. Basic strategy in response has to be prepared at national level for major accident during the transport of dangerous substances. Local governments need to determine critical points, places with protected assets and dangerous goods transportation. Risk management and response plans need to be prepared for critical points. The article shows tools, which is being tested in practice.

1 INTRODUCTION

The issue of manipulation of dangerous substances is studied from 50 s of the 20th century. The development and wider support, however, received up to the beginning of the 80 s after a series of accidents in the industry. In practice, it is addressed separately the risks management associated with the production, storage, and processing of large quantities of dangerous substances and risk management during transport by different type of transportation.

Prevention of accidents involved dangerous substances and their impacts is addressed in the case of industrial objects by technical, human and procedural measures in accordance with the relevant standards laid down in international treaties. Emergency plans are prepared for case of measures failure by industrial plant operators for the factory side and near surroundings (on-site emergency plan) and by public administrative to the broad surroundings (off-site emergency plan).

Prevention of accidents in case of dangerous substances transportation or dangerous goods, how carriers company call such trade commodity, also include technical, human and procedural measures based on international standards. The methodology of emergency planning is, however, mostly missing. Failure of measures, similar to industrial plans, occurs because human factor, a failure of technology or natural disaster, and it can occur also in areas with high population density. The article, therefore, will be deal with the phenomena.

The second chapter describes the nature of the accidents occurrence with presence of dangerous substances on basis of an analysis of the database of accidents involving dangerous substances and statistical approaches. The third chapter will be devoted to the assessment of the criticality of infrastructure from the perspective of their technical parameters and common accidents on one side and the amount and type of protected interests in the surrounding area on other side.

2 NATURE OF MOBILE RISK OCCURENCE

The logarithmic dependence of the number of realizations of certain phenomena N on the phenomena intensity I is observed in the case of statistics of occurrence of all known disasters, $N(I) \sim a^{bI}$,

Figure 1 in general. The prediction of the threat is based on the distribution of extreme values (Gumbel 1941) in the case of natural disasters.

The logarithmic character of ***F-N*** curves (fatalities-number) is observed even in the case of technological accidents, when the intensity of phenomenon is expressed in the number of human victims. Observed dependencies lose the logarithmic character just at the range of low and high intensities, where the differences between incidents reported and occurring are (Hirschberg 1998) due to the data set non-homogeneity in low intensities range and too short observation period in high intensity range. Results of mathematical models such as the distribution of extreme phenomena and so on can only be taken as a guide.

The causes and circumstances of accidents in the shipping domain are similar to those in technological domain (insufficient technical standards, faults in vehicle construction, bad maintenance, human factor, etc.). Therefore, we can also see the logarithmic character of the F-N curve (Boot 2013). However, we observe two differences. First, our control over the entire transport network is less than our control over the surrounding of industrial area. From that reason, the nature of this risk changes dynamically, e.g. the market with dangerous goods may steeply increase or decrease, which influences the transport of this goods; shipping the new dangerous goods may also start. The other problem is the risk mobility. An accident that occurs at one point on the transport route may happen at any other point of transport road. As example, we show three similar accidents in North America on rail route:

1. June 19, 2009 Cherry Valley (Illinois, USA) Canadian train with more than 2 million gallons of ethanol derailed to cross with the road, 1 dead (NTSB 2012).
2. July 6., 2013 Lac-Mégantic (Quebec, Canada) A train with 72 oil tanker cars derailed in a small town with about 6,000 inhabitants, about 50 dead, a large part of the city destroyed (NTSB 2014).
3. December 30, 2013 Casselton (North Dakota, USA) A train carrying a large amount of crude oil derailed, the need to evacuate over 2,000 people, but fortunately no losses to life (NTSB 2017).

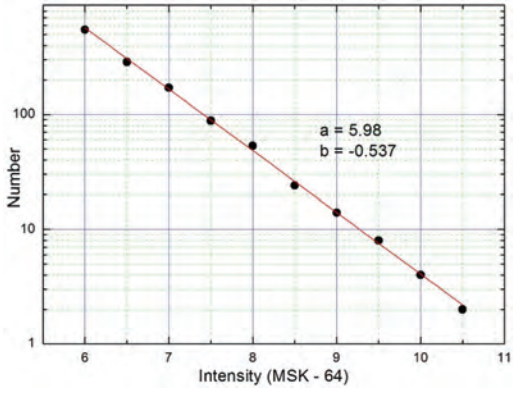


Figure 1. Logarithmic dependence of earthquake occurrences on intensity in Central Europe, circuit with a radius of 400 km (Prochazkova 2017).

The three above-mentioned accidents have common features, in addition to the region, the large-scale fires of large amounts of flammable substances. However, the impacts of individual accidents were diametrically different due to the varying population density and different protected interest concentration in the immediate vicinity of accident place.

The nature of mobile risks does not allow us to determine the real threats based only on site specific events. We need to extrapolate a wide range of information on accidents with the dangerous substances presence across the entire transport route where the accident occurred or whitherward the followed substance is transported under similar conditions. Procuration and lessons learned only from the accident site are not sufficient.

In the case of mobile risks, it needs to be taken into account that it is only a matter of time when a major accident occurs, namely even in densely populated area or even containing the high concentration of other protected interests, such as the sources of drinking water, power stations, and the like. Under certain circumstances, the transport of dangerous substances can be excluded in places with very high criticality as they present the places with big concentration of public assets; the mark "Forbidden for vehicles carrying the dangerous goods" is possible to use only exceptionally from economic reasons. In the Czech Republic, the transport of dangerous goods is, for example, excluded on the highway between Praha and Brno, between 49 and 90 km due to a drinking water source for the Praha.

However, a similar solution cannot always be used or it is connected with a lot of problems, and therefore, its application is justified only in places with especially high criticality. For less but still critical sites, it is still necessary to carry forward the lessons learned from all accidents with the presence of transported dangerous substances and to establish the plans for possible response at least on the basis of the requirements of crisis management, population protection and critical infrastructure protection.

It is necessary to introduce the role of local government in the transport of dangerous substances on the emergency management level or at least on the level of crisis management in addition to the already established obligations of dangerous goods carriers. The first step of local government is to set up the team of experts which proposes the professional solution of problem under account.

Head of the team need to be the safety management expert. He/she can be from municipality or regional Security Council, who is responsible for crisis management, Act No. 240/2010 Coll., on crisis management in the Czech Republic, or some external professional from Technical Support organization. Members from police force and firefighter force are also important. The nature of the problem also requires expert from the chemical industry. Czech chemical industry provides consulting services (TRINS 2015) for accidents with dangerous substances. Team can also be extended on transport expert and health care expert.

The team then needs to know the nature of accidents that can occur in the followed territory and it needs to determine the distribution of risks that can realized here and to characterize the possible emergency situations. Special attention should be paid to dangerous substances that are transported in large quantities through the roads at territory of given municipality. This especially involves the vicinity of chemical plants and long-distance transit corridors important for operation of specific industries.

Larger quantities of hazardous substances are transported by rail (Becherova 2017), so it is necessary to consider in this case the larger affected area than at the road transport. Responses plans need to be also prepared for pipeline transitions, especially at locations, where pipeline are placed above the surface (Hansler 2012). Pipelines surrounding can be better monitored and it is not so associated with mobile risk issues.

The municipality, in cooperation with the regional industry, can identify the specific substances that are transported on regional roads. It is obvious that fuel transports have a special position, because the transport of fuels is presented always and everywhere. The accidents with presence of fuels dominate to all transport accident statistics for dangerous substances (Procházková 2015a).

Local government needs to take into account also the transit traffic, at which the type of hazardous substances is heavily to predict. From the professional viewpoint, the response plan types would be prepared with regard to properties and hazards of substances associated with the individual hazard classes, fire hazards, explosion hazards, hazards of leakage to the environment, and also with combination of these three factors mentioned above, Figure 2.

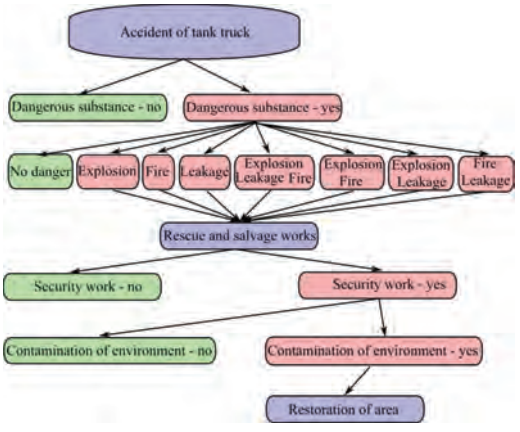


Figure 2. Process model of a crash accident response for the transport of dangerous goods.

It is necessary to determine the nature and extent of the threat, for both, the fuels and the hazardous substances that have been transported in followed location. The character of accident scenario is given by chemical, physical and other properties of transported substances, and also by site and meteorological conditions in time of accident origin. The extent of threat is based not only on the properties of dangerous substance, but especially by its amount. The amount is usually limited by relevant legal and technical standards for every shipping type.

The properties of substance, which predetermine the extent and nature of threat, are given in the safety data sheet. Everybody, who works with the hazardous substances during the shipping, has responsibility to respect the instructions in the safety data sheet. The issue of shipping the below-limit amounts of hazardous substances is solved by the REACH order in separate chapter. The Safety Data Sheet needs to be always available at substance handling site. In the case of transport, however, this is a common problem, e.g. this document is not in all languages of countries through them the shipment goes over.

The experiences from daily practice show that people often underestimate the hazards that are given in the safety data sheet. They do not consider that properties of each hazardous substance vary in dependence on local conditions that are in site of transportation. For example, the ignition temperature of unleaded petrol in the liquid state can be more than 250°C, the ignition temperature of the vapour/air mixture is substantially lower (Česká rafinérská 2012). The findings from the safety data sheet, therefore, need to be supplemented by lessons learned from studies of real-world accidents or by the results of practical experiments.

At present, many databases contain the accidents with hazardous substances. For example, it can be mentioned the energy accident database ENSAD, which include also fuel transportation. This database itself comes out from many other databases (Burgherr 2017). Compilation of databases is very expensive, and therefore, the availability of some databases is very limited. For many organizations, it is then necessary to carry out their own investigations across relevant case studies (Prochazkova 2014).

Complete statistics of traffic accidents enable to determine the most common causes of accidents. Detail study of major accidents shows the great number of combinations of phenomena that can occur at accidents, i.e. the prediction of accident scenario is not easy. Lessons learned from both, the statistics and the case studies allow us to evaluate only the general data for probability of accident and expected impacts at the sites under consideration; the real scenario will be always determined by momentary local conditions in the given site.

3 MOBILE RISK AND EXPERT METHODS

For compilation of groundwork for determination and mastering the risks, it is necessary to put together: the team of experts who well know the behaviour of hazardous substances under different conditions that can be expected during the transport; and to collect data on dangerous substances and common amounts that are transported. Then, it is necessary to use suitable tools for risk analysis, namely the Checklist and What, if analysis. In practice, a lot of software tools are used. Therefore, we take a short comment on the widely used computing software for the dispersion of dangerous substances.

Determination of the affected area is a key step for subsequently identification of impacts and compilation of response plans. The form and size of affected area are calculated using the dispersion models. These models need to take into account the properties of substances (lighter or heavier than air), atmospheric conditions (wind speed, temperature, humidity, inversion) or local relief topography (buildings, forest, terrain slope). The factors that affect the dispersion are many, so mathematical models have to use certain approximations in many cases, which reduce the accuracy of outputs. Therefore, it is suitable to apply experts´ experiences before their use in practice.

At present, the models convert from analytical calculations to calculations using the computer technology. Computer technology can combine analytical and numerical calculations, but essentially is based on the same formulas (Leksin 2015). The advantage of computer models is the accuracy, the model can take into account a larger number of input parameters, and the accessibility, the model can be used by broad public; the knowledge of model mathematical background is not necessary. The disadvantage is non-transparency of computation, because the user does not often know what omissions at calculation are performed (the omissions are hidden in software model concept).

Although computer computations allow us to use more complex models, it is necessary as it was given above to use the expert judgement at transfer to practice. It is fact, that each result, in addition, is greatly affected by the interface at the input, where we cannot include a number of geographic data in the calculation. In some cases, geographic conditions are deciding (Pontiggia 2010). The dispersion models can provide us with a lot of information still, despite all the uncertainties and unsureness. Figure 3 shows the dispersion behaviour for the chlorine tank in normal atmospheric conditions.

The example at Figure 3 suits for preliminary identification of the belt width around the considered ground road where we follow the impacts on protected interests. Common atmospheric conditions are sufficient to response at normal and may be emergency situations. For very extreme meteorological conditions it is necessary to use the principals of crisis management and for it success it is necessary to prepare both, the scenarios for unfavourable up to extremely conditions and the response plans. For real accident site it is necessary to consider the properties of affected area and local conditions. Detail investigation of critical sites, e.g. the sites on roads with frequent traffic accident, very improves the model results.

The densely populated areas or areas with high density of other public protected interests are fac-

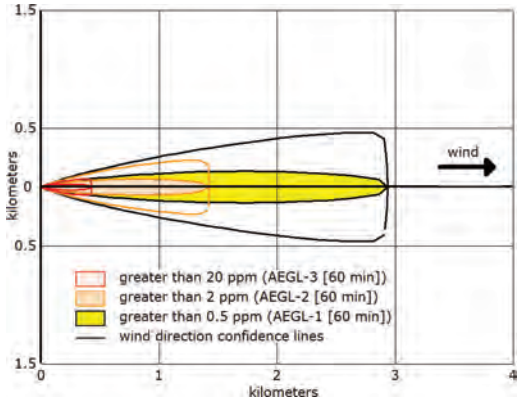


Figure 3. Dispersion of chlorine from the tank according to the ALOHA dispersion model under normal atmospheric conditions.

tors that would be taken into account when we select the critical sites, for which from safety reasons there are necessary the preventive actions and the preparation of response plans. The occurrence probability of traffic accident in critical site is further important. The local territory investigation is very important road parameter. Such investigation can be executed using the special Checklist.

The Checklist needs to take into account the physical parameters of the road (horizontal profile, vertical profile, surface properties), environmental characteristics (weather, climate), utilization of road (traffic density, purpose of traffic) and other complications (tunnel, bridge, canyon) (CSA 2002). It is appropriate to add accident statistics to the overall criticality assessment, if it is monitored for a given section. Assessment of criticality can be done in the selected sites of road only if it is performed cumulatively for road sections. The example of such a Checklist is in Table 1, where it is judged the section of highway in Central Bohemia, where transport Praha-Vienna (south-east) and Praha-Linz (south), E50, E55, E65 is located.

For identification of critical highway spots, it is necessary to determine the size of risk in a given place for possible traffic accidents with presence of hazardous substances. Assessment of risk size at real location for the fixed event is influenced by density of protected interest and by likelihood of accident. Methods for impact assessments are several, but most of them have a relatively narrow focus on the situation (input data) or the protected assets. For the public administration, as the legal guardian of territory, it is the most appropriate the method "What-If analysis" performed by brainstorming with a team of experts, and the following judgement of losses at protected interests following from the determined impacts.

The team of experts needs in the first to collect information on possible events, i.e. the characteristics of hazard connected with leaked substances. Then, it needs to determine the scenarios of affected areas at different meteorological conditions for individual hazardous substances. Next, it is necessary to determine the acceptability of impacts in individual scenarios at different meteorological conditions.

The experts can execute assessment by brainstorming. The identified impacts are best structured by the area of protected assets:

- human lives, health and security,
- property and welfare,
- environment,
- critical infrastructure and technologies.

It is also suitable to consider the development of impacts with time, how they occur apart from the structure by type of protected interests. Such data enables to build more accurate response plan. It is also necessary to identify the primary accident

Table 1. Checklist for critical assessment of the first 21 km of the D1 motorway in Central Bohemia.

D1-km	Horizontal profile	Vertical profile	Complication (tunnel, bridge, canyon)	Technical condition of communication	Traffic intensity 50000/hour	Average of accident/ 1 month	Climate condition	Total
1	1.0	1.0	1.0	0.0	1.9	1.0	0.0	5.9
2	0.0	0.0	1.5	0.0	1.9	2.0	0.0	5.4
3	0.0	1.0	0.0	0.5	1.7	1.0	0.0	4.2
4	1.0	1.0	0.0	0.5	1.7	1.0	0.0	5.2
5	1.0	1.0	1.0	0.5	1.7	1.0	1.0	7.2
6	0.0	2.0	1.0	0.0	1.7	1.0	0.0	5.7
7	1.0	1.0	2.0	0.5	1.5	1.0	0.0	7.0
8	0.0	1.0	0.0	0.5	1.5	1.0	0.0	4.0
9	0.0	1.0	1.5	0.5	1.4	1.0	0.0	5.4
10	1.0	1.0	2.0	0.5	1.4	1.0	0.0	6.9
11	0.0	2.0	0.5	0.0	1.6	2.0	0.0	6.1
12	0.0	1.0	1.5	0.0	1.6	2.0	0.0	6.1
13	0.0	1.0	0.0	0.5	1.6	1.0	0.0	4.1
14	0.0	1.0	1.0	0.5	1.6	1.0	0.0	5.1
15	1.0	2.0	0.0	0.0	1.6	1.0	0.0	5.6
16	2.0	2.0	1.0	0.0	1.3	1.0	1.0	8.3
17	0.0	0.0	0.0	0.0	1.3	0.0	0.0	1.3
18	1.0	2.0	0.0	0.0	1.3	1.0	0.0	5.3
19	1.0	1.0	0.5	0.0	1.3	0.0	0.0	3.8
20	1.0	0.0	0.0	0.0	1.3	0.0	0.0	2.3
21	0.0	0.0	1.0	0.0	1.3	1.0	0.0	3.3

impacts, the secondary impacts to which it belongs e.g. the loss of supply of affected area, damage of drinking water sources etc. An example of such analysis can be found in article (Prochazkova 2015b). According to performed tests such analysis takes about 3 hours. In the case of team composed from 5 experts, it is 15 man hours. As a rule, such analysis is necessary to perform for several different places. The result of such work is good preparation for mitigation of impacts of traffic accidents and precise response plan that reduces losses on protected interests.

4 RISK ASSESSMENT

Under the risk assessment the losses need to be quantified in adequate values. For public administration and managers, the most comprehensible formulation of losses is in finances. In the case of property and infrastructure, it is not the problem to determine financial value of losses and damages. In the case of the environment, however, there are a number of methodologies, most of which only focuses on the production effect of environment and ignores the whole area of its importance for humans.

Financial evaluation of losses on human lives and damages of health in EURs includes a certain point of cynicism; in some countries it misses the appropriate legal rule for its determination. Therefore, it is important, so the public administration may determine the clear instructions and methodologies for this case. These procedures are mostly based on local culture, legislation, and the society tolerance for different risks.

The appropriate risk is possible to assess as soon as we know the extent of the damage and the probability of its realization. The risk can be expressed in the form of value of damage per unit of time, for example, by the equation

$$R = D \cdot P = D \cdot A_r T_{ds} \quad (1)$$

where R is the risk, D is the impact (damage), P is the probability, A_r stands for accident and T_{ds} the intensity of transport of dangerous substances.

The critical matrix can be also used to express the risk. It shows much more information and allows sophisticated access to risk; especially, in the ALARP (As Low As Reasonable) approach, for hazardous substances. Figure 4 shows the combined form of the Critical Matrix (Impacts versus Probability) and the F-N Curve (Number of Death vs. Frequency), (Det Norske Veritas 2014). It determines which situations are acceptable, conditionally acceptable (ALARP), or unacceptable. Decisions on preventive measures and on response

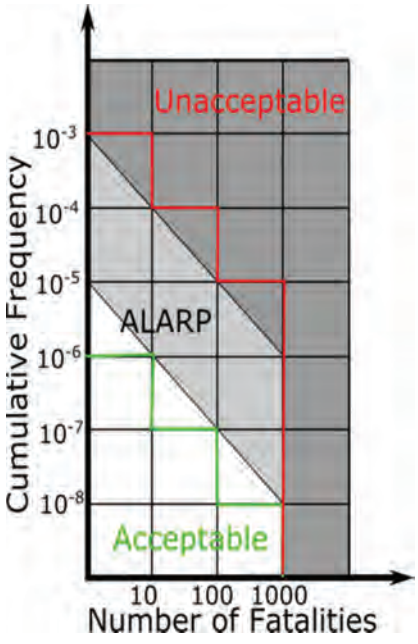


Figure 4. Criticality matrix and F-N curve for accident with dangerous substances involved (Det Norske Veritas).

Figure 5. Risk Management in public space, expert do Risk Assessment, Risk Decisions is executed by elected representatives, and Risk Mitigation is applied by technical experts.

plans are subsequently made according to position of our judgement result in the criticality matrix. The matrix given in Figure 4 only counts the human lives, but it can be transformed to other protected interests.

Establishing the position of analysed situation in the criticality matrix in Figure 4, is the final part of the risk assessment. The team of experts passes the results of risk assessment to responsible office for the territory governance that is responsible for territory safety management and for correct decision-making. The risk management issue moves to the next stage, Risk Decision, Figure 5. The decision on further approach to revealed risks and on measures for coping with identified risks belongs to the responsibility of elected representatives, who are not experts, which often influence the quality of decisions.

The roles of experts return at the final stage of Risk Management, i.e. at Risk Mitigation, but real measures are influenced by limits given by public administration decision-making (e.g., it very influences amount of finances that can be used for risk mitigation).

5 CONCLUSION

Critical judgement of present situation shows that in practice, there are missing the tools that: effectively reduce breaking the security measures on the carrier side; enable fast reaction to accident; enable protection of all road / railroad users including the humans and environment in traffic accident vicinity; and prop up the renovation of affected area so it can be used for human needs.

The article demonstrates a methodology for the risks assessment associated with the transport of dangerous goods by road or rail. The main problem is the nature of mobile risks associated with transport. Another problem is the risk dynamics. This methodology is primarily targeted to public administration, and it well supplemented the measures of legal rules that are primarily targeted on the carriers of dangerous goods. This methodology consists of three steps: risk analysis, risk judgement; and risk mitigating.

Identification and awareness with the nature of mobile risk phenomena in the reporting area is the first step. The main problem during the risk identification is neglecting the mobility and dynamics of the risk. The often error in practice is that it is not considered the experiences from accident sites in other regions. The analysis and judgement of emergency situations, which may arise as consequence of accident with dangerous substances, is the second step.

The recommended methodology seeks to combine practices using the computational models with investigation of experts based on knowledge of real territory. The final step is the risk mitigation and early response; the combination of criticality matrix with the F-N curve is used. The methodology calculates the significance only with life-threatening impacts and needs to be adjusted to include other fatal impacts, such as groundwater contamination, or the destroying of other critical infrastructure elements.

The above described methodology includes procedures and tools for risk assessment by team of experts. In co-operation with real local experts we step by step determine real results for individual real public administration (selection of appropriate preventive measures, processing the response plans for very hazardous substances at the level of emergency and crisis management).

ACKNOWLEDGEMENT

Authors thanks to the Czech Technical University in Prague for support (grant SGS2015-17).

REFERENCES

Becherova O. & Hošková-Mayerová Š. 2017. Rail infrastructure as a part of critical infrastructure: London: Balkema ISBN: 978-1-138-62937-0, p. 1615.

Boot, H. 2013. *The Use of Risk Criteria in Comparing Transportation Alternatives*. Chemical Engineering Transactions, **31**, p. 199, DOI: 10.3303/CET1331034.

Burgherr, P., Spada, M., Kalinina, A., Hirschberg, S., Kim, W., Gasser, P. & Lustenberger, P. 2017. *The Energy-related Severe Accident Database (ENSAD) for comparative risk assessment of accidents in the energy sector*. London: Balkema, ISBN 978–1-138–62937–0, p. 1417.

Česká Rafinérská. 2012. *Material Safety Data Sheet UNLEADED PETROL*. Litvínov: Česká Rafinérská a.s.

CSA. 2002. *Risk Management: Guideline for Decision-Makers*. Canada: Canadian Standards Association, CAN/CSA-Q850-97.

Det Norske Veritas Ltd. 2014. *Harmonised Risk Acceptance Criteria for Transport of Dangerous Goods*. European Commission DG-MOVE, PP070679/4.

Gumbel, E.J. 1941. *The return period of flood flows.* The Annals of Mathematical Statistics, 12, p. 163.

Hansler, R. & Laheij, G. 2012. *Failure causes for pipelines transporting hazardous substances*. New York: IAP-SAM & ESRA, ISBN: 978-1-62276-436-5.

Hirschberg, S., Spiekerman, G. & Dones, R. 1998. *Severe Accidents in the Energy Sector, project GaBE: Comprehensive Assessment of Energy Systems.* PSI Bericht Nr. 98 ISSN-1019–0643.

Leksin, A., Barth, U., Adeulov, D. & Mock, R. 2015. *Comparison of Dutch and Russian standards for calculating physical effects of hazardous substances*. London: Balkema, ISBN 978-1-138-02879-1, p. 69.

NTSB. 2012. *Derailment of CN Freight Train U70691-18 With Subsequent Hazardous Materials Release and Fire Cherry Valley, Illinois, Railroad Accident Brief.* USA: National Transportation Safety Board, RAR1201.

NTSB. 2017. *BNSF Railway Train Derailment and Subsequent Train Collision, Release of Hazardous Materials, and Fire Casselton, North Dakota, Railroad Accident Brief.* USA: National Transportation Safety Board, RAB1701.

Pontiggia, M., Busini, V., Derudi, M., Alba, M., Scaioni, M., Rota, R., Landucci, G., Molag, M., Tugnoli, A., & Cozzani, V. 2010. *Safety of LPG rail transportation in the perspective of the Viareggio accident*. London: Balkema, ISBN 978-0-415-60427-7, p. 1872.

Procházková, D. & Procházka, J. 2017. *Problems connected with determination of size of maximum expected disaster in selected site.* London: Balkema ISBN: 978-1-315-37498-7, p. 1443.

Prochazkova, D. 2015b. *Safety of complex technological facilities.* ISBN: 978-3-659-74632-1. Saarbruecken: Lambert Academic Publishing 2015, 244p. ID:200303845

Prochazkova, D., Prochazka, J. & Patakova, H. 2015a. *The results of a systematic study of the risks associated with the transportation of hazardous substances.* Lonon: Balkema, ISBN 978-1-138-02681-0, p. 1663.

Procházková, D., Procházka, J., Patáková, H., Procházka, Z. & Strymplová, V. 2014. *Kritické vyhodnocení přepravy nebezpečných látek po pozrmních komunikacích v ČR*. Praha: ČVUT, ISBN 978-80-01-05599-1.

TRINS, (2015), *Transport Information and Accident System*, Unipetrol, Czech Republic.

TSB. (2014). *Lac-Mégantic runaway train and derailment investigation summary*. Canada: Transportation Safety Board of Canada, r13d0054.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Risk-based regulation and certification of autonomous transport systems

S.O. Johnsen, Å. Hoem & T. Stålhane
Faculty of Information Technology, NTNU, Trondheim, Norway

G. Jenssen & T. Moen
SINTEF Technology and Society, Trondheim, Norway

ABSTRACT: Autonomous transport systems in all modes—road (i.e. autonomous cars), aviation (i.e. drones), shipping and rail are coming. Regulation and testing are on-going in Norway. Risks of autonomous systems are uncertain due to missing data, emerging technology and variation in framework conditions. However, accidents of autonomous cars seem to be 1/3 or 1/2 of current levels. Incidents are different, needing outside interventions sometimes. Based on review of experiences across the modes and regulations, we suggest agile and transparent learning in the whole autonomous ecosystem, between all modes. System certification are needed, and system responsibilities must be clarified. Structures for orchestrating transport (i.e. control of many autonomous vehicles with possible common failures) and marking autonomous transport, should be established. In the interfaces between humans and systems there are differences in autonomy as imagined vs. performed, leading to new incidents and accidents. Emerging safety/security issues must be explored.

1 INTRODUCTION

This paper discusses experiences of autonomous transport systems, to establish a framework for risk based governance. Risk and risk governance are based on the process described by Renn (2005), starting with problem framing; risk appraisal (hazards and vulnerabilities); risk judgment; risk communication and risk management. The implementation of autonomy can reduce transport risks but it can also introduce new risks in the interfaces between the autonomous system and the environment (such as humans). As discussed in Lund and Aarø (2004), risk reduction must be based on a broad set of actions such as regulation, technical design, training and awareness.

Based on involvement in the regulatory process in Norway and experiences of autonomous transport systems we have discussed new emerging risks and threats. We see the need for establishing framework such as regulatory actions and clarification of responsibilities as autonomy is being implemented.

In the following we have defined autonomous systems and concepts such as Levels of Automation (LOA) used to specify degree of automation.

1.1 *Definitions and terminology*

Safety is related to accidental harm, while security is related to intentional harm. Safety is defined as: "*the degree to which accidental harm is prevented, reduced and properly reacted to*", Firesmith (2003). Security: "*the degree to which malicious harm is prevented, reduced and properly reacted to*".

In Parasuman and Riley (1997) automation and autonomy is described as "*The execution by a machine agent (usually a computer) of a function that was previously carried out by a human*". Automation can be done by various means i.e. 1: Remote controlled (Surveyed and/or externally controlled); 2: Autonomous (based on own sensors and systems); 3: Cooperative and connected (based on own sensors and other traffic information) or 4: A combination of 1–3. The terms autonomous and automated has been used interchangeably in some papers. We have made a distinction. By *autonomy* we mean a system that is non-deterministic in that it has a freedom to make choices, and by *automated* we mean a system that is more deterministic in that it will do exactly what it is programmed to do. This is based on the taxonomy and discussion of autonomy from Vagia et al. (2016).

When trying to scope risks of autonomous systems we must include the regulation, risk governance, organizational framework, interfaces to humans and the autonomous system (a combination of software components and cyber physical systems). The system is often a collection of systems being developed by different stakeholders. Thus, we have used the concept of autonomous

ecosystem, AEC. This is inspired by the concept Software Ecosystems (SEC). SEC consists of components developed by actors both internally and externally of the company, i.e. outside the traditional borders to a group of private persons and actors. Manikas et al. (2013) defined a software ecosystem as: "*the interaction of a set of actors on top of a common technological platform that results in a number of software solutions or services. Each actor is motivated by a set of interests or business models and connected to the rest of the actors and the ecosystem as a whole with symbiotic relationships, while, the technological platform is structured in a way that allows the involvement and contribution of the different actors...*". Arguments for using such a concept is the realization that development increasingly is taking place outside of organisational silos due to the need for speed of development, need for supporting applications, reduction of development costs, competition. This is creating the need to address governance challenges in an ecosystem framework.

An example of an autonomous ecosystem is Intelligent Transport Systems (ITS) consisting of autonomous vehicles, integrated with traffic control, electronic payments and other systems. Autonomous ecosystems handle information, but also actual critical processes such as transport (via automobiles, boats, drones and trams). These ecosystems must be safe and secure. The systems must be able to handle unanticipated events, breakdowns and be able to go to a safe and secure (end-)state.

To explore the main risks of autonomous systems, we need to clarify responsibilities i.e. LOA in task execution. LOA is described by steps going from no automation where the humans are fully in control to a fully automated system with no human interaction. Sheridan and Verplank (1978) introduced 10 steps of automation, going from LOA1: Fully Manual Control to LOA10: Fully Autonomous Control. The LOA has been adapted to the car industry by the Society of Automotive Engineers (SAE), describing six levels of autonomy in driving, SAE (2016). Going from no autonomy (level 0), through driver assistance, partial automation, conditional automation, high automation, to full automation (level 5). The design of the autonomous transport system must ensure that the system maintains an accepted level of performance despite disturbances, including threats of an unexpected and malicious nature. Our approach is to speed up learning and knowledge sharing between modes, since the autonomous systems have different maturity and experiences in aviation, rail, road and sea.

The concept of resilience engineering is an important strategy to handle unanticipated incidents. Hollnagel, Woods and Leveson (2006) define resilience as "*the intrinsic ability of a system to adjust its functioning prior to or following changes and disturbances, so that it can sustain operations even after a major mishap or in the presence of continuous stress*". Handling of unanticipated incidents and continue to operate safe is a key ability of autonomous transport systems.

Based on the preceding introduction, the research questions (RQ) we want to explore are:

- RQ1: What are the major risks introduced by autonomous transport systems?
- RQ2: What regulatory issues should be prioritized to handle these risks?
- RQ3: What are the way forward, i.e. main approaches and issues needed to mitigate major risks of autonomous transport systems?

2 SCOPE, CHALLENGES AND METHODS

When discussing autonomous ecosystems, we include the organisational framework, regulation, human interactions and understanding in addition to the actual systems in autonomous systems and the infrastructure. This is described in Figure 1.

2.1 *Challenges and problems*

When introducing new technology such as autonomous systems, one of the basic challenges is to understand emerging risks. Safety, security and resilience have often been identified late when vulnerabilities have been exploited and unwanted incidents have been published. There has been a tradition in the software industry that vendors seldom have to pay for these unwanted incidents even if they are due to poor quality, poor focus on safety, security or resilience. The consequences and costs have been given to users, organisations and society. In autonomous transport, the consequences can be loss of lives and/or environmental damage. In addition, when discussing vulnerabilities in autonomous ecosystem, one challenge is that there is not one single supplier, but a set of suppliers involved. It can be difficult to identify responsibilities and manage competencies, if framework conditions (regulation/responsibilities) are missing.

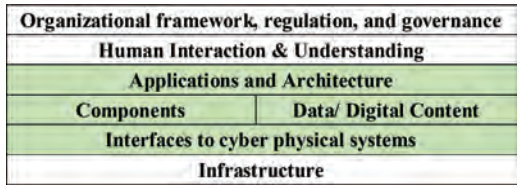


Figure 1. Scope of autonomous ecosystems—AEC.

2.2 *Methodology and approach*

We have based this paper on empirical data from users of autonomous transport systems, a targeted literature review of autonomy and safety in addition to discussion of suggested regulation of autonomous road transport in Norway.

We have explored experiences of autonomous transport systems from St. Olav Hospital in Norway, where autonomous systems have been used from 2006 to 2017. St. Olav has 10,500 employees, and covers an area of 200,000 M^2. We are involved in pilot projects with self—driving shuttle busses in three Norwegian cities. Trials addressing feasibility of Mobility as a service (MAAS) linking up to public transport (first and last mile). We are involved in trials with eco-friendly autonomous ships/vessels for cargo and passenger travel along the Norwegian coastline.

We have performed a literature review based on a keyword search of autonomy, safety, security and resilience using SCOPUS, ACM Digital Library, IEEE Explore, Springer Link and Science Direct.

We have been involved in a hearing of regulation related to testing of autonomous vehicles in Norway from the Ministry of Transport and Communications—MTC (2016). The suggested regulation was distributed in December 2016, comments to be given within March 2017 and the regulation were proposed to be approved as law in December 2017. Our comments were based on the literature review, experiences from St. Olav's and other public comments.

The taxonomy used to register incidents has been based on Blanco et al. (2016). They collected a broad set of naturalistic accident data from autonomous driving, using a taxonomy of crash seriousness going from most serious at C1 to negligent at C4.

C1: Crashes with airbag deployment, injury (needing doctor visit), rollover, more damage than $1,500, require towing, police reportable.
C2: Minimum of $1,500 worth of damage, crashes such as large animal strikes and sign strikes.
C3: Crashes involving physical conflict with another object, but with minimal damage. Includes most road departures, small animal strikes, all curb and tire strikes potentially in conflict with oncoming traffic and with higher risk potential if no curb.
C4: Tire strike only with little or no risk element (e.g., clipping a curb during a tight turn), considered to be of such minimal risk that most drivers would not consider these incidents to be crashes.

3 RESULTS AND DISCUSSIONS

In the following section, we have documented experiences from autonomous systems at St. Olav Hospital; some selected findings from our literature review; and key issues discussed during regulation of testing of autonomous transport systems.

3.1 *Findings from autonomous systems at St. Olav*

St. Olav Hospital has installed an automated transport system called Transcar LTC2 Automated Guided Vehicle System (AGV) from Swisslog. They installed seven AGVs in 2006, and additional 14 AGVs at the end of 2009. From 2010 to 2017 they have had 21 AGVs in operations. Each week the 21 AGVs transport medicine, food, clothes and garbage, in total 70–80 tonnes. (Each AGV can transport a load of 500 kg, and is transporting 3.6 tonnes each week). The speed is slow, moving at approximately 2 km/hour (maximum speed is 5 km/h). The AGVs can send signals, open doors, and reserve elevators to deliver goods. There are different suppliers of door and elevator automation. When there are conflicts that cannot be resolved, a signal is given to the operational centre. The centre is manned by an operator that can intervene through the system, or go to the place where there is a conflict.

The AGVs can communicate (i.e. deliver pre-programmed messages) such as "Please move—you are in my way", or "Elevator is reserved—please move out of elevator". A key issue related to the awareness building between automated transport systems and humans are the above-mentioned communication from the AGVs, supporting the understanding that the automated system need to inform the bystanders about their perceptions and what they are going to do next, that helps staff, patients and visitors to learn to interact with the AGV's and to anticipate their behaviour.

In the Transcar LTC2 Operations and Maintenance manual it is written "*Always maintain a distance of 1.5 meters between the vehicles and people or objects.*" This safety guideline is not possible to implement at St. Olav due to space limitations.

There are traces on the floor indicating that the AGVs are always following the same pathway, thus (new) common failures may happen.

There has been a total of 100–130 minor incidents per year (5–6 per AGV) categorised as C4 by us. Minor repairs are done on the AGVs, changing around 50 components per year. There are around 15 emergency stops each year, categorized as C3, where components must be changed. We do not have data indicating that there has been any incidents of category C2 or C1. Reported incidents are minor crashes due to faulty navigation, for example due to objects placed in the route travelled that is not detected.

When interviewing the users some incidents that can be generalized were reported:

- The AGVs have problems with pallets close to the walls. The AGV uses the wall as reference in steering. A misplaced pallet results in a lateral shift of the AGV position and may sometime end up with a collision. Initially the operators used a great deal of time to clear the transport road area (in the basement) from clutter (i.e. parked bicycles, pallets with supplies); this work has been reduced now—but maintenance and design should take into accord these limits of AVGs.
- The AGV collided several times with the forklifts, since the LiDAR sensor (light detection and ranging) had a limited vertical field of view and was seeing a free zone (space) under the forklift. This was mitigated by placing a black rubber skirt under the forklift. The same kind of collisions happened when using stepladders on the floor in the AGVs pathway, since the LiDAR did not detect the object. Thus, one issue has been the ability to see and identify objects in relation to the AGVs sense of its own size and position. This may be a general challenge with autonomous transport systems. The death accident of Joshua Brown, described by NTSB (2017) and NHTSA (2017), was between a Tesla and a trailer crossing the road—a white trailer giving poor contrast and with substantial height above the ground. Some similarity with the forklift problems at St. Olav. A rubber skirt under the trailer may have increased visibility/visual signal of the trailer.
- The AGVs can open doors, reserve and use elevators. Sometimes there has been conflicts between the AGVs and the users, needing human intervention through a central control.
- Software updates of AVGs, elevators and doors has led to interface problems, thus there is a need to look at the AVGs as a part of an ecosystem.

During the 11 years' operating the AGVs there has been no reporting of human injuries at St. Olav. However, at the AHUS hospital (AHUS, 2009), with the same system—one incident happened in 2009, where a nurse sustained a minor injury when colliding with the AGV (i.e. category C3 or C4).

In summary, the AGV system has had an impressive safety record at St. Olav's Hospital. Key issues of safe operations are related to an ecosystem approach planning the interaction between technology, organisation and humans. Based on preparation trough pilots; low speed; communication between automated systems and humans to inform surrounding people of the AGV's intended behaviour. The unexpected may happen, thus there was a need to establish a manned control centre that can intervene during operations.

3.2 *Key findings from literature review*

In Axelrod (2014) the focus is on software assurance of safety-critical and security-critical systems. The perception is that use of the current methods has not achieved the wished-for level of protection, and that there are missing security principles and standards. There seems a need for incentives or regulations to implement protective and immunizing measures in software. A requirement could be that these measures are included in a certification process. On governance, it is suggested to establish software assurance standards at the United Nation (UN) level; to have a risk based approach; to share best of breed methods; and the need to discuss liabilities for damages occurring because of an attack or security-related errors.

International governance of security of the infrastructure is addressed through several channels such as standard bodies (i.e. ISO, IEC) and international bodies such as OECD, EU, NATO and UN. Autonomous systems are international—involving many actors with different agendas. In GCIG (2016) there is a discussion of governance of emerging technology as it is integrated into critical infrastructure, such as transport systems. It is suggested that manufacturers should follow the principle of privacy and security by design, when developing new products. They must be prepared to accept legal liability for the quality of the technology they produce. Buyers should collectively demand that manufacturers respond effectively to concerns about privacy and security. Governments can play a positive role by incorporating minimum security standards in their procurement. It is suggested that government regulations should require routine, transparent reporting of technological problems to provide the data required for a transparent market-based cyber-insurance industry. It is suggested to establish an agreement (a compact) based on collaboration between government, industry and private society supporting this evidence based decision making.

In Koscher et al. (2010) vulnerabilities in cars are pointed out, such as the possibility to control a wide range of automotive functions and completely ignore driver input from dashboard, including disabling the brakes, selectively braking individual wheels on demand, stopping the engine, and so on. Attacks were easy to perform and the effects were significant. It is possible to bypass rudimentary network security protections within the car, and perform attack that embeds malicious code in the car that will completely erase any evidence of its presence (after a crash). There is a discussion of the challenges in addressing these vulnerabilities in the existing ecosystem.

In Lima, et al. (2016) semi-autonomous and fully autonomous cars are described as coming

from the development stage to operations. The autonomous systems are creating safety and security challenges. These challenges require a holistic analysis, under the perspective of ecosystems of autonomous vehicles. These systems will become important critical information infrastructures, simultaneously featuring connectivity, autonomy and cooperation. Threat analyses and safety cases should include both (random) faults and (purposeful) attacks.

In DHS (2015), there is a discussion of Cyber-Physical infrastructure risks in the future smart cities. Several examples of unwanted incidents are described in transportation systems (i.e. autonomous vehicles; trains) in electricity distribution and management and in water and wastewater systems sector. It is suggested to the regulator to work with standards and regulations in addition to communication and increased engagement by giving direct assistance. Challenges mentioned are the need to establish goal based standards and regulations as new technology is implemented and to focus on dissemination of best practices and systematic education.

In (Cerrudo, 2015) there is an empirical evaluation of "smart cities" looking at a broad set of technologies of traffic control, management of energy/water/waste and security. Known vulnerabilities are in traffic control systems, mobile applications used by citizens, smart grids/smart meters and video cameras. The issues are lack of cyber security testing and approval, lack of encryption, lack of City Computer Emergency Response Teams (CERT), and lack of cyber-attack emergency plans. There are reasons to anticipate that we establish potential for serious incidents, if these issues are not addressed and mitigated.

In Frei (2010) there is a discussion of the security dynamics of general software ecosystem (SEC), applicable to autonomous ecosystems. They examine 27000 vulnerabilities in the decade (1996–2008). The paper explores several policies such as security through obscurity, responsible disclosure of vulnerabilities (a suggested policy) or security through transparency. One key insight is that secrecy prevents people from assessing their own risks, which contributes to a false sense of security. Responsible disclosure means that the researcher discloses full information to the vendor, expecting that mitigation is developed within a reasonable timeframe. An increasing number of organizations has adopted some form of responsible disclosure. A risk based regulatory regime are dependent on such an open discussion of the risks.

In summary, if we want systems that are safe, secure and reliable, both safety, security and reliability must be built together. There has been documented several vulnerabilities and responsible disclosure of vulnerabilities to the vendors, seems to be a beneficial policy. Some sort of communities of practices, and a CERT of autonomous systems should be established. There is missing international regulation or compacts based on private public partnerships to ensure privacy, safety, security and resilience. Vendors must ensure this quality by design, and must be prepared to accept legal liability of the technology they produce. Regulations should require routine, transparent reporting of technological problems to provide data for a transparent market-based cyber-insurance industry, and a risk based regulatory regime.

3.3 *Key issues when discussing regulation*

3.3.1 *Selected issues from all forms of transport*

Risks of autonomous transport are not well known at present. To increase knowledge and learning, experiences, taxonomies, regulations and relevant incidents should be gathered and disseminated from all modes—autonomous road systems (vehicles), air transport (i.e. drones), rail (unmanned metro and rail systems) and shipping. Accident investigators and rule-makers (such as "The Accident Investigation Board in Norway") should develop methods for investigation of accidents of autonomy and report their findings.

Shipping: Completely unmanned ships seem to give large benefits and enables new transport systems, some of these issues are documented in Rødseth (2017). There is a need for onshore control centres to manage autonomous shipping operations. Norway has focused on autonomy in sea transport. A network, Norwegian Forum for Autonomous Ships (NFAS) at nfas.autonomous-ship.org, has been established. A more general research program called Centre for Autonomous Marine Operations and Systems (AMOS) has been initiated at the Norwegian University of Technology and Science, ref www.ntnu.edu/amos. The Trondheimsfjord has been selected as a national testing area in collaboration with The Norwegian Maritime Authority and The Norwegian Coastal Administration. At the end of 2017 three testing areas has been established in Norway (Trondheimsfjord, Storfjord and Horten). Test areas has also been established in Finland and China. Risk levels of autonomous ships are influenced by existing incidents and new incidents (i.e. caused by new automation, and former incidents mitigated by crew now being removed). Work is ongoing to explore safety of autonomous sea transport, and to explore a taxonomy of LOA for shipping, Rødseth et al. (2017). In Trondheimsfjord an autonomous passenger ferry is going to be tested in 2018–2019. The authorities need to set rules and requirements based on acceptable risk levels. There

is an increased need for Human Factors knowledge to improve the quality of interfaces (i.e. "human in the loop" control when needed) between humans and the autonomous systems.

Aviation: To govern the use of Remotely Piloted Aircraft Systems (i.e., drones) in Norway, regulation has been established, Civil Aviation Authority—CAA (2016). The operator must be certified through an exam, CAA (2017). Experiences of remotely piloted aircraft Systems, Waraich et al. (2013), documents that mishaps may happen (i.e. 50 mishaps occur every 100,000 flight hours' vs human-operated aircraft where there is one mishap per 100,000 flight hours). The high mishap rate is related to poor attention to human factors science and design in ground control centres, Waraich et al. (2013). Several pilot projects with drones are planned, transporting goods/persons.

Rail/Metro systems: By automated metros (rail systems) we mean systems where there is no driver in the front cabin, nor accompanying staff, also called Unattended Train Operation (UTO). UTO have been in operations from 1980. In UITP (2013), there are listed 674 km of automated metros consisting of 48 lines in 32 cities. UTO´s are found in Barcelona, Copenhagen, Dubai, Kobe, Lille, Nuremberg, Paris, Singapore, Taipei, Tokyo, Toulouse and Vancouver. Wang et al. (2016), list the arguments for UTO as increased reliability, lower operation costs, increased capacity, energy efficiency and an impressive safety record. There is substantial infrastructure cost to ensure safe on and offloading of passengers and that the track is safe and isolated from other traffic. Four distinct Levels of automation are defined: GoA1: *Non-automated train operation, with a driver in the cabin*. GoA2: *Automatic train operation system controls train movements, but a driver in the cabin observes and stops the train in case of a hazardous situation.* GoA3: *No driver in the cabin but an operation staff on board.* GoA4: *Unattended train operation, with no operation staff on board*. We have at present not found normalized accident data for UTO (incidents based on person km), but no accidents have been reported. It seems that the UTO has exceptionally high safety. However more systematic analysis and normalization of all international UTO transport incidents are needed.

Road Transportation: Google's self-driving cars, where the vehicle systems control all aspects of the driving, have been on public roads in the US since 2009. The safety record has been impressive. However Google engineers are supervising and re-taking vehicle control if necessary. The death accident in 2016 (Joshua Brown) by Tesla in Autonomous driving condition was caused by a tractor-trailer that made a left turn in front of the Tesla, and the car failed to apply the brakes. The Tesla did not "see" the trailer—it was all white and had poor contrast with the surrounding bright white sky. In addition, there was a high gap between the road and the trailer. The National Transportation Safety Board (NTSB, 2017) found that the system's "operational design" was a contributing factor to the crash because it allows drivers to avoid steering/watching the road for periods of time that were "inconsistent" with warnings. Tesla could have taken further steps to prevent the system's misuse. In addition, NTSB faulted the driver for not paying attention and "over-reliance on vehicle automation". It also seems there is a need for better training of drivers related to autonomous systems—a part of driver education and driver license requirements.

There are scarce safety data so far, but data from the period 2009 to end of 2015 has been collected from Googles cars, in Teoh et al. (2017). There were three police reportable accidents (denoted as level C1) in California while driving 2,208,199 km, giving an accident rate of 1,36 police reportable incident pr. million km. This is 1/3 of reportable accidents of human-driven passenger vehicles in the same area. Car accidents involving autonomous cars are different from human driven. Google cars get more rear-ended by other vehicles while stopped or barely moving. There is an element of risk negligence in that the human driver does not fully anticipate the action of the self-driving car. There are also challenges of sustained human attention during lengthy period of autonomous driving, making it difficult for the human operator to intervene i.e. "Human in the loop" challenges. Huffington (2017) documented that Waymo's human drivers had to take control from the automated system (i.e. "disengagement") once for every 5,000 miles in 2016. "Backup" human drivers in Uber's self-driving cars had to take over about once every mile as of March 8, ref Recode (2017). It is a challenge to get situational awareness after having been out of the driving control loop for 5,000 miles. The takeover time of the human driver varies from 2 to 26 seconds, ref Eriksson et al. (2017), challenging the design of autonomous systems to enable human intervention.

Analysing all car accidents, it is suggested that 80–90% of accidents are due to "human errors", thus autonomous cars could reduce the level of accidents substantially. However, autonomy could introduce new types of accidents, due to automation itself or due to human drivers not predicting action from the automation. In Blanco, et al. (2016) it is suggested that accident rates are reduced to ¾ of present, while Teoh et al. (2017) documents accident levels of autonomous systems as 1/3 of human driver systems. The National Highway Traffic Safety Administration (NHTSA, 2017) reported a reduction in vehicle crash-rate

by almost 40% with Autosteer activated in Tesla Model S and Model X, compared to before. In Cummings et al. (2014) it is suggested that the level of accidents could be reduced by 50%. More experiences must be gathered, but significant reduction of accidents is expected.

3.3.2 *Need for systematic open data reporting*
At present there are missing data of incidents (accidents and successful recoveries) related to autonomous systems. Open reporting must be established covering systematic safety records and security stories, being available to researchers and industry actors, such as insurance. The scope must cover actions from the autonomous system but also document perceptions and understanding from the involved human actors. The differences between espoused values (rule based actions/work as programmed in autonomous systems) and actual values (actions/work as being done by humans in interaction with autonomy) can create the basis for errors and accidents. It should be a key area of research to explore accidents because of poor design vs. blaming the human actors. Use of video recording could help, based on regulation protecting personal data; (EU 2016:679). There must be a combination of data gathering in combination with in-depth accident investigation. Accident investigation boards should explore accidents of autonomy, to support rapid learning and changes in addition to improve their methods to analyse autonomy incidents.

3.3.3 *System perspective and human factors*
Safety of autonomous systems are dependent on new designed technology, human factors and organisational issues as discussed by Cummings et al. (2014). The perception should be that most accidents in autonomous systems are a consequence of poor design and poor testing, and that "human errors" are a consequence and not a cause as described by Dekker (2002). Moving trivial functions (that can be programmed) to an autonomous system, means that tough decisions and deviations must be handled by humans. Thus, the science of Human Factors, knowing strengths and weaknesses in cognition and ergonomics, must get a significant position when automation is designed and implemented.

3.3.4 *Responsibilities and certification*
The autonomous system decides based on design approved by the manufacturer. Thus, product responsibilities of accidents and incidents must be placed at the manufacturer (OEM). This is in line with the view of the car OEMs Volvo, Google and Mercedes-Benz (Iozzio, 2016). This is also in line with the supervisory responsibility demanded in the Oil and Gas industry (i.e. where the operator is responsible for the chain of suppliers employed). This supervisor responsibility must be placed on the car OEMs, including the continued updating and adaptation of software in use. Certification is needed, such as the ISA/IEC-62443 scheme of industrial control systems used since 2010. However, certification is still being developed, a survey documenting key issues are found in Martin, et al. (2015).

3.3.5 *Security and risk-based regulation*
Security (for safety) must be included in the development of autonomous systems, and systematic testing (including penetration testing) must be done as a part of certification prior to product release. The precautionary principle must be established as a condition for autonomous transport systems, COMEST (2005).

4 CONCLUSIONS

Related to the research question RQ1 (major risks): The sensors and systems used in autonomous systems, does not have a perfect view of the surroundings, and may also act uncoordinated with their surroundings, thus new type of accidents may happen. There is a need to speed up learning from these incidents and to be aware of communication and information challenges in operations.

Human control and assistance through control centres and via human machine interactions must be designed based on the science of Human Factors in order to avoid higher levels of accidents as documented by Waraich et al. (2013).

We continue to see vulnerabilities and exploitation of software in the public and private sectors. Different perspectives are used in security and safety, due to different adversity models. The security community are addressing threats (directed, deliberate, hostile acts) and the safety community are addressing hazards (undirected events). AEC are so pervasive across all sectors that a silo approach can no longer be acceptable. To ensure that all actors in the value-chain understands this, a silo-based "need to know" principle must be replaced by transparent and open reporting. This can also support a market based cyber-insurance industry.

Related to the research question RQ2 (regulation): There is a need for regulatory action from government to set minimum standards, establish responsibility, and follow up of incidents/accidents. Prescriptive and detailed rulemaking on a national level is wanting, but should be replaced by functional approach demanding the same level of risk in automated systems as in existing systems.

Vendors must have responsibility to ensure safety, security and resilience by design, and must be prepared to accept legal liability for the quality of the technology they produce. Ideally, a formal process of product acceptance and certification (i.e. safety case) should be established before a product can be sold. The manufacturers should establish a proactive focus on (best practice) safety/security standards. There is a need to ensure that there is some sort of a structured learning process (among all relevant actors) when incidents happen.

Related to the research question RQ3 (way forward): Innovative approaches, such as the perspective of Autonomous Ecosystems (AEC) are needed to handle the challenges of autonomous transport systems. The science of Human Factors need to be prioritized to ensure that human intervention can be designed in the system and can be performed in actual operations based on actual human limitations and human strengths to improvise and handle unanticipated events.

Safety has been dependent on publicised accidents and a systematic learning loop between users, the regulator and industry. One component in the learning loop of complex software systems has been reporting and analysis of incidents through computer incident response teams (CERTS). There is a need to establish CERTS of AEC to help coordinate actions.

Rules and mechanisms for updating software in autonomous systems will become more urgent as failures can lead to accidents, thus handling of updates must be addressed in a systematic manner.

Communication between autonomous transport systems and drivers and bystanders must be improved. Autonomous systems are rule based while humans are not, thus there may be misunderstandings and common failures, creating need for interventions through transport centres controlling the flow of transport.

These AEC will be exposed to new strains—thus there must be a focus on how to handle surprises by resilience, to ensure that new demands/ stress/ failures are not impacting transportation in a catastrophic way.

REFERENCES

Axelrod, C.W. (2014, May). Reducing software assurance risks for security-critical and safety-critical systems. In *Systems, Applications and Technology Conference (LISAT), 2014 IEEE Long Island* (pp. 1–6). IEEE.

AHUS (2009) www.rb.no/lokale-nyheter/pakjort-av-robot-pa-jobben/s/1-95-4309194

Blanco, M., Atwood, J., Russell, S., Trimble, T., McClafferty, J., & Perez, M. (2016). Automated vehicle crash rate comparison using naturalistic data. Virginia Tech TI.

Cerrudo, C. (2015). An emerging US (and world) threat: Cities wide open to cyber attacks. *Securing Smart Cities—White paper - IOActive.* www.ioactive.com/pdfs/IOActive_HackingCitiesPaper_cyber-security_CesarCerrudo.pdf

CAA-Civil Aviation Authority (2016) Regulations for Remotely Piloted Aircraft Systems (Forskrift om luftfartøy som ikke har fører om bord mv) retrieved from lovdata.no/forskrift/2015-11-30-1404

CAA-Civil Aviation Authority (2017) Training requirements—retrieved from luftfartstilsynet.no/selvbetjening/allmennfly/Droner/

COMEST (2005) The Precautionary Principle from UNESCOs World Commission on the Ethics of Scientific Knowledge and Technology.

Cummings, M.L., & Ryan, J. (2014). Who is in charge? The promises, pitfalls of driverless cars. TR News, 292, 25–30.

Dekker, S.W.A. (2002). Reconstructing the human contribution to accidents: The new view of human error and performance. *Journal of Safety Research*, 33(3), 371–385.

DHS (2015) Department of Homeland Security, Office of Cyber and Infrastructure Analysis: The Future of Smart Cities: Cyber-Physical Infrastructure Risk

EU (2016:679) On the protection of natural persons with regard to the processing of personal data and on the free movement of such data; Regulation of the European Parliament and of the Council of 27 April 2016

Eriksson, A., & Stanton, N.A. (2017). Takeover time in highly automated vehicles: noncritical transitions to and from manual control. *Human factors*, *59*(4), 689–705.

Firesmith, D.G. (2003). "Common concepts underlying safety, security, and survivability engineering", *Technical note CMU/SEI-2003-TN-033*, Carnegie Mellon University.

Frei, S., Schatzmann, D., Plattner, B., & Trammell, B. (2010). Modeling the security ecosystem-the dynamics of (in) security. *In Economics of Information Security and Privacy* (pp. 79–106). Springer US.

GCIG (2016) Global Commission on Internet Governance, "One Internet" www.ourinternet.org

Hollnagel, E. Woods D. and Leveson N. (2006). "*Resilience Engineering*", Ashgate.

Iozzio, C. (2016). Who's Responsible When a Car Controls the Wheel?. *Scientific American*, *314*(5), 12–13.

Koscher, K., Czeskis, A., Roesner, F., Patel, S., Kohno, T., Checkoway, S.,... & Savage, S. (2010, May). Experimental security analysis of a modern automobile. In *2010 IEEE Symposium on Security and Privacy* (pp. 447–462). IEEE.

Lima, A., Rocha, F., Völp, M., & Esteves-Veríssimo, P. (2016). Towards Safe and Secure Autonomous and Cooperative Vehicle Ecosystems. In Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy (pp. 59–70). ACM.

Lund, J., & Aarø, L.E. (2004). Accident prevention. Presentation of a model placing emphasis on human, structural and cultural factors. *Safety Science*, *42*(4), 271–324

Martin, J., Kim, N., Mittal, D., & Chisholm, M. (2015). Certification for autonomous vehicles. *Automative Cyber-physical Systems course paper, University of North Carolina, Chapel Hill, NC, USA.*

Manikas, K., & Hansen, K.M. (2013). Software ecosystems–a systematic literature review. *Journal of Systems and Software*, *86*(5), 1294–1306.

MTC (2016) Ministry of Transport and Communications "Testing of autonomous road transport systems" from www.regjeringen.no/no/dokumenter/horing---forslag-til-ny-lov-om-utproving-av-selvkjorende-kjoretoy-paveg/id2523663/

NTSB(2017) www.ntsb.gov/investigations/AccidentReports/Pages/HWY16FH018-preliminary.aspx

NHTSA (2017) The National Highway Traffic Safety Administration, Office of Defects Investigation resume PE 16-007. Automatic vehicle control systems, Tesla Model S accident in Florida May 7 2016.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, *39*(2), 230–253.

Renn, O. (2005). "Risk Governance—Towards an Integrative Approach" *White paper no.1 – international risk governance council.*

Recode (2017) www.recode.net/2017/3/16/14938116/uber-travis-kalanick-self-driving-internal-metrics-slow-progress

Rødseth, Ø.J. From concept to reality: Unmanned merchant ship research in Norway. I: Proceedings of Underwater Technology (UT), 2017 IEEE. IEEE 2017 ISBN 978-1-5090-5266-0. OCEAN

Rødseth, Ø.J. & Nordahl H. Ed. (2017). Definition for autonomous merchant ships. Version 1.0, October 10. 2017. Norwegian Forum for Autonomous Ships. nfas.autonomous-ship.org/resources-en.html Accessed December 2017.

SAE (2016). SAE International standard "J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems." Revised: 2016-09-30

Sheridan, T.B., & Verplank, W.L. (1978). *Human and computer control of undersea teleoperators*. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

Teoh, E.R., & Kidd, D.G. (2017). Rage against the machine? Google's self-driving cars versus human drivers. *Journal of Safety Research*, *63*, 57–60

UITP (2013) Observatory of Automated Metros World atlas report. International Association of Public Transport (UITP), Brussels

Vagia, M., Transeth, A.A., & Fjerdingen, S.A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed?. *Applied ergonomics*, *53*, 190–202.

Waraich, Q.R., Mazzuchi, T.A., Sarkani, S., & Rico, D.F. (2013). Minimizing human factors mishaps in unmanned aircraft systems. *ergonomics in design*, *21*(1), 25–32

Wang, Y., Zhang, M., Ma, J., & Zhou, X. (2016). Survey on driverless train operation for urban rail transit systems. *Urban Rail Transit*, 1–8.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

How systems engineering may be useful in preparing FMECA—lesson learnt from a practical case

M. Bucelli
Department of Civil, Chemical, Environmental and Material Engineering, Alma Mater Studiorum—University of Bologna, Bologna, Italy
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology NTNU, Trondheim, Norway
Safetec, Trondheim, Norway

J. Zhang & A. Rauzy
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology NTNU, Trondheim, Norway

S. Sultana
Department of Marine Technology, Norwegian University of Science and Technology NTNU, Trondheim, Norway

ABSTRACT: Risk communication and information exchange have become more challenging in today's complex projects, which involves different stakeholders in different roles. Unstructured data communication and flaws in knowledge exchange may lead to erroneous risk perception and evaluation. The role of risk analysis has become more critical in this perspective. This paper proposes a new way to enrich risk analysis with the help of knowledge of systems engineering. Focus is given to establish a Failure Mode, Effect and Criticality Analysis (FMECA) model using models from systems engineering which are more systematic compared to models used in tradition approach. A reference case study is analyzed, which is inspired from the subsea laboratory at the Department of Mechanical and Industrial Engineering (MTP) at the Norwegian University of Science and Technology (NTNU). The investigation focuses on the series of relevant risks for the subsea gas boosting section in offshore Oil & Gas installations. Results from the present study are discussed with emphasize on three aspects: (1) cross-system effects, (2) reasonable and reachable risk reduction measures and (3) multiple system dimension. We acknowledge that the proposed method based on system thinking can be used to construct system behavior model, which in turn could be used in FMECA development to gain better understanding of risks and to improve the overall performance of system.

1 INTRODUCTION

In the last decades, the industry, from manufacturing to chemical and Oil & Gas (O&G) sector, has become more and more complex, notably by incorporating innovative technologies requiring new stakeholders. Risk analysts are responsible for the identification and the analysis of potential risks arising from the performed activities and for treating these risks, according to established acceptance criteria (ISO 31000, 2009). Many factors affect the actual risk level in different ways. For instance, a new human machine interface may cause stress and therefore influence operator performance. The risk posed by these factors raises the importance of both the risk analyst and the communication network with the other professionals, consequently, the focus moved to risk management.

Risk management involves different disciplines at different levels across the entire enterprise. However, nowadays in the industry, the coordination between the different parties is not stressed enough when instead it should be properly maintained (Rice & Spence, 2016). Bringing different areas of expertise together in analyzing the potential risks and identify threshold criteria is still challenging. Moreover, the methods adopted for information collection and analysis are often unstructured (Kirsch, Hineb, & Maybury, 2015). The inconsistence in jargon used and the difference in the theoretical background of the different stakeholders may lead to erroneous decision-making and costly correction for inadequate and inappropriate actions.

Failure Mode, Effect and Criticality Analysis (FMECA) has been widely accepted as the risk identification and the characterization tool for

hardware components since the late 1960s. However, today's industry involves fast-moving technology innovations and faces challenging operating conditions. There is no formal procedure in constructing FMECA for such complex system and then the quality and content of FMECA depend on the competence and experience of risk analysts. In some practices, analysts start FMECA without establishing a baseline system concept. Based on this scenario, the solution is to structure coordinated and distributed system information about what is taking place and what is needed for proper risk mitigation.

The existing frameworks for risk management in Oil and Gas industry (ISO 31000, 2009; NORSOK Z-013, 2010) also highlight the needs of establishing the system concept before starting risk analysis, and maintaining and updating the system concept based on given indications, however, no detailed methods and approaches are prompted. This paper suggests the use of often-cited approaches in Systems Engineering (SE) to fulfil such needs, to prevent excessive resource for reaching the agreement on the system concept.

The main objective of this contribution is to investigate how FMECA can get advantages from SE. Many companies have adopted SE as the systematic approach for the design of complex systems (Asbjørnsen, 1992; Haskins, 2008). In fact, SE may help in mediating information exchange among professionals in a simple and concrete way by means of different analyses and models at different detail levels. In this sense, it allows the right person to access the right information at the right time and use it. The adaptation and exploitation of SE approaches and models can assist in maintaining the unified context of the system and ensuring that the interests of different stakeholders are properly understood and considered.

This paper provides a notion through a practical case on to what extent and in which ways the risk analysts think SE methods as effective for supporting FMECA. The following of this paper is organized as follows: Section 2 describes the basis of the subsea gas boosting laboratory that still demands further improvements from a risk perspective. Section 3 discusses the key features of linking and coupling systems engineering and risk analysis models. In Section 4, the proposed model is executed with the practical case and the results are presented in section 5, discussed in section 6 consecutively. Sections 7 presents conclusions.

2 A PRACTICAL CASE

The paper tactically selects an accessible subsea laboratory located at the Department of Mechanical and Industrial Engineering (MTP) at the Norwegian University of Science and Technology (NTNU) to investigate a larger spectrum of risks relevant to subsea gas boosting, as shown in Figure 1. As of today, one challenge for subsea boosting is the compression of wet gas (within

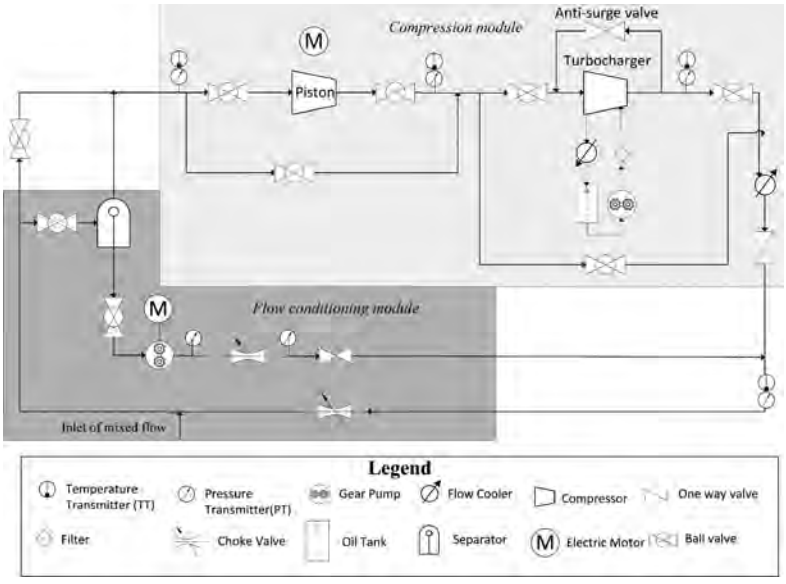


Figure 1. Subsea compression laboratory.

water fraction of 2–20%) since the stratified gas-liquid flow may run the high risk of damaging the traditional compressor. The laboratory is constructed as the pre-testing facility to emulate the different existing solutions for subsea wet gas compression in the Norwegian Continent Shelf, i.e. Ormen Lange and Åsgard (within pre-separation) and Gullfaks (without pre-separation).

This laboratory includes two major modules to test different characteristics of wet gas: the separation module (i.e. the left-bottom) and the compression module (i.e. the right-top). The mixed flow within water fraction (ranged from 0–20%) is emulated by controlling the inlet flow. The implementation of a separation module can separate the water from mixed flow before entering the compression module, and allow studying how the working efficiency and robustness of a separator can influence the whole compression process. The compression module involves a compressor driven by a piston and a compressor driven by the turbocharger, where the piston compressor is very vulnerable by particles and water. Three test scenarios are therefore formulated by the manual close/open of ball valves:

- Wet gas compression with the separation module, where the piston is bypassed
- Wet gas compression without the separation module, where the piston is bypassed
- Dry gas compression, where the piston and turbocharger can compress in the series

The laboratory is almost completed in late 2016, but still demands many improvements in respect to different aspects. This paper exclusively focuses on managing emerged risks of the current structure of the laboratory and devotes to present the obtained results and knowledge for the new development in the industry-size gas boosting system.

3 METHOD

This paper aims to suggest a structured method to enrich the scope of validity of risk assessment by taking advantage from SE. The proposed method propagates SE activities toward risk assessment activities such as FMECA, as shown in Figure 2. SE workshop and FMECA workshop that focus on very different objectives are as the heart of this method. This collaborative method of knowledge transfer enables effective risk management with an objective of dispersing expert knowledge into available tacit knowledge (Alavi & Leidner, 2001).

The starting part of the model is preliminary analysis, which includes defining process goal, defining system along with its boundary, environment and interactions. This facilitates having an overview of the process and to be informed about what is included and what isn't included. SE workshop involves experts (e.g. Designers, operators, managers) to create the static vision of the structure of the system from operational, functional and physical perspectives. SE workshop covers the

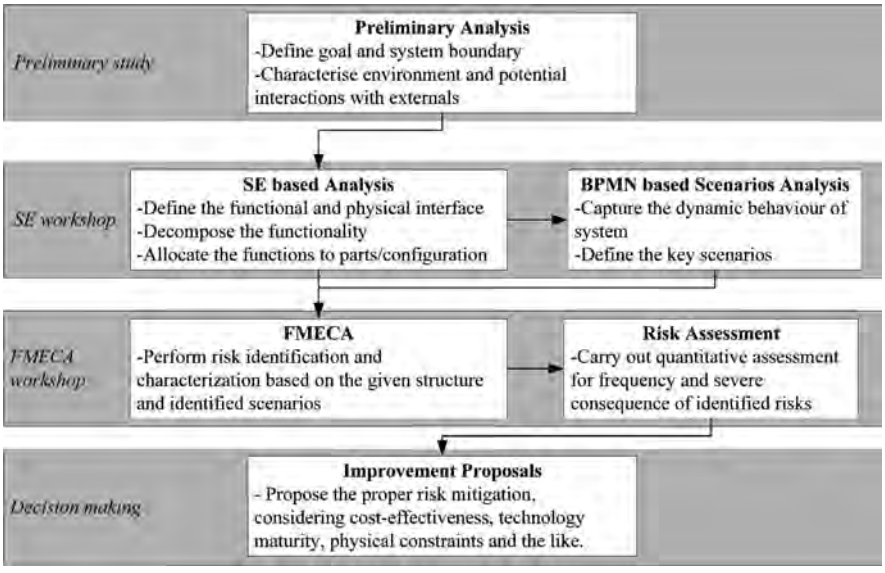


Figure 2. The conceptual map for the proposed method.

interests and the expertise of each contributing stakeholders.SE workshop consists of two concurrent analyses: SE based analysis that covers the operational, functional and physical aspects and the scenario approach that borrows Business Process Model and Notation (BPMN) (White & Miers, 2008). Operational analysis introduces behavior of the system that helps performance based assessment. We can introduce conflict among different component behavior. For example, component C will go off if the component A shuts down or one specific pump may trigger electric supply failure. It solves the conflicts by considering the system function in an ordered sequence of declarations. Functional decomposition specifies how the component functions realize the module functions and how the module functions realize the overall function. Functions that make up another function are grouped together in sets based on ways of achievement. Material, energy and signal flows are viewed as the attributes of these interactions. The functionality is allocated to the preferred parts and layouts. Physical decomposition is made based on the layout of components provided by supplier to show connections between components. This enables us to check interdependency among system components. The availability of the system, including reparable elements can be determined. If one component is out of order, the effect of that on the system or other components can be realized by following the connections. Management personnel can set their work order easily by following the connection.

Designing the scenarios is always an effective way of correctly abstracting the concerns in design. The scenario analysis of a given system is completed on the basis of static vision of its structural decomposition. BPMN can be considered as procedural knowledge representation which represents a set of interconnected procedures (Ligęza & Potempa, 2012). BPMN is therefore considered as a feasible approach to study some scenarios generated from the combination of critical failure, near-miss and even safe states on each interconnected procedure. Both business experts and process experts can easily understand the semantics of BPMN, so this tool is considered feasible to graphically represent the interested scenarios.

The result of SE work is used to conduct FMECA. To carry out the FMECA, multidisciplinary experts are invited to form the team. The team, analyses system components for failure modes. Then potential causes and effects are determined, but forgetting the internal-related or external-related failure modes (DNV-RP-D102, 2012). Involving the scenario-based approach (e.g. BPMN) offers the opportunity to enrich the traditional FMECA. SE workshop combines various discipline knowledge in one common platform and captures multidimensional knowledge into one frame. Different discipline experts analyses the same issue from different angle and may find a different solution (Su & Dou, 2013). It assures universal agreement on a conflicting issue, eliminates bias toward severity rankings, and carry out a detailed analysis of the system structure as well as its process. In addition, generic FMECA contains no dynamic features of the system being analyzed. Using scenario-based analysis as the baseline can assist the risk engineers to clarify the context of each failure modes. The similar approach, called as the scenario-based FMEA was discussed in (Issad, Kloul, & Rauzy, 2017).

After completing FMECA, one can carry out the well-round risk assessment on a basis of FMECA to provide indications for further improvements regarding risk mitigation, recourses and modifications on its structure.

The proposed method is vividly illustrated by the following analysis of the presented case.

4 ANALYSIS

4.1 *Preliminary analysis*

The preliminary analysis defines the preliminary system concept to describe what the system should do, without specifying any functionality and embodies. The analysis covers all the elements that are unmodifiable from engineering perspective, like the external environment of the system, users, legal and regulatory framework and the like. The analysis paves the ground for all the possible technical solutions for the stated problem.

Figure 3 illustrates the operational context within the subsea compressor laboratory, which only indicates what the system do without specifying how to achieve the goal. Different models can be developed on basis of defining operational context of subsea compressor laboratory. For instance, state diagram can be made to check the operational constraints by combining the states of laboratory and those of its external systems, like the energy supply system. The complete preliminary analysis can lay a solid foundation towards the SE workshop that complete the system concept.

4.2 *SE workshop*

The SE workshop executes the analysis stated and discussed above in Section 3, including the functional decomposition and physical decomposition. Functional decomposition is to define the different levels of functionality based on the system mission. Physical decomposition is based on the Process & Instrumentation Diagrams (P&ID) and layout of each component to check physical interactions among subcomponents and allocations. Figure 4 illustrates the physical decomposition of

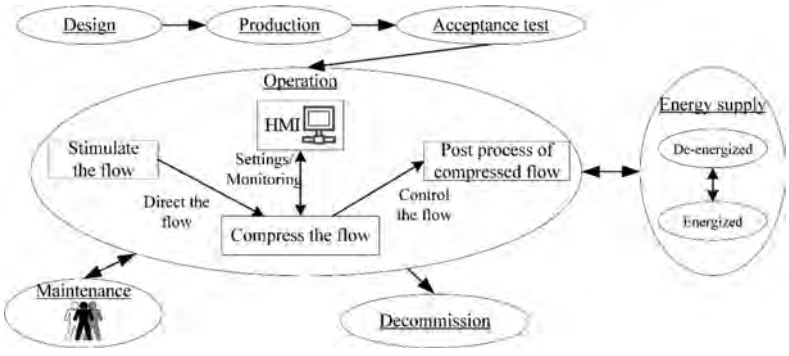


Figure 3. Operational contexts of subsea compressor laboratory.

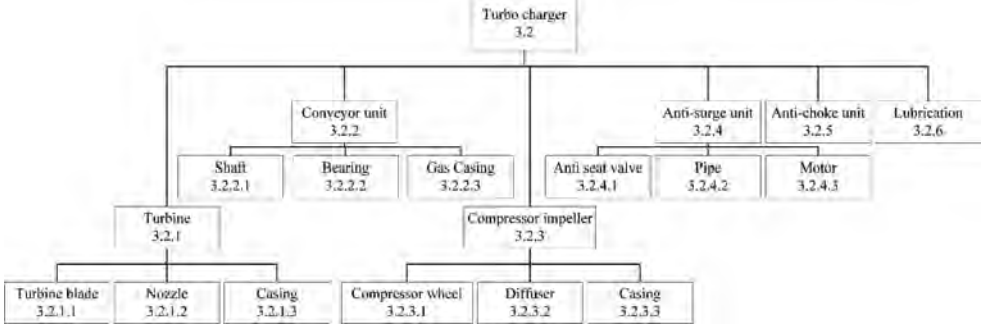


Figure 4. Physical decomposition.

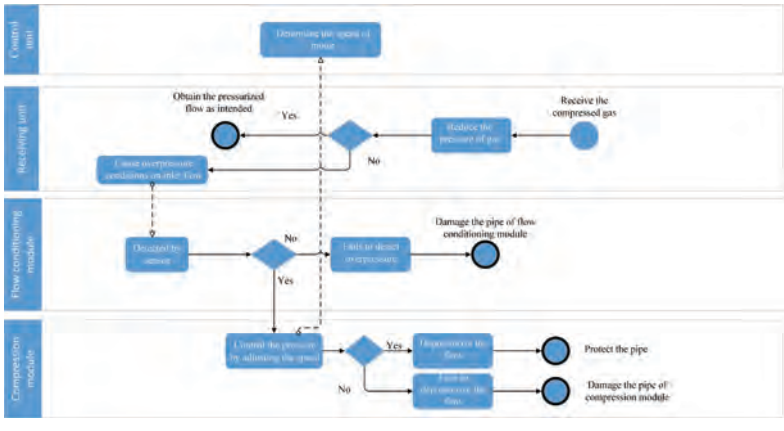


Figure 5. Simplified BPMN for overpressure scenario.

the selected system, the turbo charger to exemplify some key activities in SE workshop.

The decomposition is rather easy to apply for analyzing the functionality and physical structure of the system. One can refine all the operational contexts with sufficient details and trace the changes of functionality and structure through such method. However, this method only describes the system concept statically. Indeed, to complete the well-round risk analysis, we have to carry out the scenario analysis that dynamically describe the events that trigger the transitions between each operational context.

As discussed before, BPMN models are convenient to visualize practical scenarios. Figure 5 presents one selected accident scenario. In Figure 5, the laboratory is tactically divided into

four lanes, including a control unit that is not illustrated in Figure 1. The interfaces (information exchange) between control unit and other lane are indicated by the message flow, i.e. the dashed line. Three decision points are presented to accommodate different missions. The identified operational contexts are also reflected in BPMN model. One can explicitly observe how the control unit influences the whole process of subsea gas compression, and connects each object within distinct activities through BPMN. Here, focus is given to the overall process, system interactions, interface and sequence to work flow, where no time/temporal aspects are considered. Different consequences are specified upon the success of activities, i.e. the availability of corresponding objects. Through generating the accidental scenarios, designers are able to identify the additional needs and carry out complementary analysis to improve the design proposal, see also the summarized result in Section 5.

4.3 *FMECA workshop*

The results from SE workshop are integrated in the FMECA. One advantage is the explicit identification and evaluation of the potential failure modes across physical boundaries. The effect of failure on the subsystems can be studied by analyzing the sequence flow. For instance, the turbocharger compressor receives the flow from the gas intake under the test scenario 2. Once there is a failure or malfunction of the gas intake component, (e.g. Leakage in pipeline or contaminated with liquid from the surroundings), the compressor can be damaged or experience the temporary loss of efficiency. Such risk can be immediately registered when developing FMECA. Another advantage is that BPMN highlights major tasks within the block instead of a corpus of components. In some practices, analysts produce FMECA based on the checklists of physical components (for example Figure 4) as they did not analyze the entirety of the given system concept. By adopting the BPMN model in SE workshop, risk analysts are able to create the FMECA that covers the most significant accident scenarios, which saves a large amount of repetitive and unbalanced works when coordinating the contributions from different design teams.

5 RESULT

Table 1 summarizes some key implications raised after conducting two workshops. The differences between laboratory environment and subsea environment are considered and discussed in the last column of Table 1.

Table 1. Improvement proposal for subsea gas boosting laboratory.

Key issues	Description	Decisions	Relevance for industry-size case
Installation of additional valves, sensors or transmitter	Flow sensor in the water inlet	Must do. The humidity must be controlled to map the characteristics of flow	Not relevant for the flow from the real gas field.
	Pressure sensor in the oil loop	Should do. High pressure may blow the pipe.	Relevant.
	Level transmitter in the oil reservoir	Should do. The implementation can assist in the maintenance (oil refill). Especially when the laboratory is continuously run or stop using for a long time (volatilization of oil)	Highly relevant.
	Flow meter before compressor	Can consider for smooth operation Compressor efficiency decreases with increased mass flow.	Highly relevant.
	Flow mixer	Can consider. The stratified flow runs a higher risk of damaging the compressor than a dispersed homogenous flow.	Not relevant, stratified flows are fairly common even in the real gas field.
	Pressure relief valve in the separator	Should do. The implementation will reduce the risk of separator blow.	The relevance depends on the type and size of separator.

(*Continued*)

Table 1. (*Continued*).

Key issues	Description	Decisions	Relevance for industry-size case
Installation of additional components	Logic control unit connected to all sensors and controllers	Must do, to control the process easily and from a remote location.	Highly relevant.
	Additional filters in the inlet water line	Not necessary if have confidence that supply water is clean	The relevance depends on the needs of removing sands or other particles, requiring expertise from reservoir management.
	A Protective wall around the whole lab	Should do, which will prevent smoke dispersion and reduce fire spread	Highly relevant.
	Protective housing for piston compressor, turbocharger, water pump after the separator	Can consider, as it will prevent from damage in case of water flooding. The cost analysis is needed.	Highly relevant.
Maintenance strategy	Leakage test before the operation	Must do, as it is the most cost-efficient mean to check the integrity of the system.	Only relevant for the site-acceptance test. The leakage sensors are implemented for this purpose.
	Periodic dust cleaning	Must do, as it will reduce the blockage of valves and pipe network.	Not relevant.
	Documentation of operation strategy of system and valves	Must do, as it will reduce human error in operation.	Mostly important.

The results are obtained by considering the major risks within the existing design. One limitation of the current analysis is that the engineering efforts behind each decision are not included. The remaining works are the complementary analysis such as life cycle cost analysis to support the decision-making in the real practice.

6 DISCUSSION

This collaborative model of SE workshop and FMECA workshop makes access to a comprehensive knowledge network of practical experience and expert understanding. This enables identification of potential hazards and implementation of appropriate measures for prevention of accidents. The proposed method enables expert to capture and document the experience they have gained throughout different projects. This can be implemented both in the design phase and in modification phase.

In the traditional approach of FMECA, a failure analysis is mainly carried out on the component level, functional interactions between observed components are not included (Bertsche, 2008). Failure analysis is carried out for the individual process steps. The entire production process is not thoroughly analyzed, for example, the layout of individual component is not considered (Bertsche, 2008). The benefits of developing the FMECA through the identification of key scenarios are listed in the following, based on experiences from practical cases:

- Cross-system effects

It is possible to achieve a holistic and systematic view of the system through SE workshop. The functional analysis assists in comprehending

the effect of failure on subsystem functions and system functions.

- Reasonable and reachable risk reduction measures
Risk reduction measures cover several factors, e.g. maintenance scheduling, decision-making support and barrier management. The architectural analysis can clarify the main constraints that limit in choosing these factors. The coordination between FMECA workshop and SE workshop is therefore concerned with whether the selected structure offers the best balance of these factors.

- More than one single dimension
SE workshop involves experts (e.g. designers, operators, managers) in performing BPMN; operational, functional and physical analysis. This allows to include the interests, expertise and the needs of each stakeholder in the very beginning phase of risk analysis and reflects in the development of FMECA.

In the proposed method, important aspects like interaction between components and environmental effects are considered which gives more confidence in risk analysis and in decision making. An abstruse idea of a system introduces uncertainty in the risk assessment process. The consistent representation of knowledge and system assures that results of risk estimates can be integrated in decision making without less doubt. Only system related uncertainties are being dealt here, which can be mitigated with system knowledge based on experience and expertise.

To present the whole system on like only like P&ID will create a blur on communicating the message. This paper proposes structural breakdown type diagram to represent the physical and functional behavior of the system. The analyses include state in the behavior attribute of a system in the model that describes interdependency and helps performance based assessment.

When the system is complex enough, and one component is serving different functions of the system, it can never be edited as a whole or only through physical or functional dissection. Application of both functional and physical analysis gives a hierarchical breakdown with branches to show connections between components. When one component is out of order, the effect on other components or on the whole system can be observed easily by following the connections, so management personnel can set their work order easily.

If any new component is added to or reduced from the system, editing decomposition diagram is easy which helps to modify FMECA easily. It is often found that in traditional approaches for a change of system, risk assessors have to go through FMECA fully. Prior establishment of physical and functional analysis allows users to modify knowledge easily in case of a change of system.

Effective knowledge transfer and integrated knowledge management help to make a more resilient and reliable system by reducing vulnerability. It helps to develop a better plan for proactive measure to cope with the emergency. These workshops include models for performance and reliability, previous experience, realism of assumptions, representativeness of scenarios along with most critical issues, thoroughness of analysis and first class deliverable.

Effective maintenance also requires integrated information and knowledge system from which maintenance team can get output from other disciplines to make proper maintenance record and work order. Capturing system knowledge effectively facilitates full retrieval of information to implement preventive maintenance on a contrary to corrective maintenance. Failing to do so, increases cost significantly (Motawa & Almarshad, 2013).

Risk engineers often do the risk analysis to compare the risk level to check whether the specified activity is to be complied with the standard. In this prospect, a greater chance for improvement remains out of scope and behind the paperwork. By the arrangement of thinking around systematic activities described earlier, risk can be communicated effectively and efficiently. The quality of risk assessment can be improved by capturing and identification of all possible issues systematically. By sharing with one another's information, it is possible to get a better risk picture in more than one single dimension (Su & Dou, 2013).

However, going through all details is time consuming as there remains a lot of overlaps and repetitions among functions. It is difficult to define the scope of subsystem when one single component performs two functions. It is also questionable whether to assume the previous work as reliable enough or not. Finding necessary expert and shareholders' opinion on a timely manner need proper planning. The execution of the proposed method needs a high management capability of the organization. An organization should have a commitment to provide resources, freedom and time needed to acquire information. Implementing the analysis in development phase makes it possible to identify weak spots and comparative tests can be carried out.

7 CONCLUSION

Modern society deals with a larger spectrum of risk and a larger spectrum of stakeholders, of interest, value and knowledge. Recognizing the wider scope of risk leads to a positive evolution in managing risk. A structured communication model can help in this respect. In this paper, we

presented how SE may help in capturing different nodes of a system for effective risk evaluation, through the basic analysis technique like FMECA. The proposed method is checked with a subsea gas boosting laboratory where the case study assures more confidence in making development proposal and risk mitigation. By doing this type of analyses in the development phase, it is possible to identify weak spots and comparative tests can be carried out. Modifications in the design phase saves cost and preventive measures can be taken.

As a future improvement, a consequence study can be included in detail and should be checked with other applications. The paper also suggests AltaRica 3.0 to encode the FMECA for quantitative risk assessment. This recent achievement has been brought to the forefront in the risk analysis community, see also more details about this modelling language in (Prosvirnova, 2014). This modelling formalism suggests taking the advantages from a structuring paradigm, i.e. S2ML (Batteux, Prosvirnova, & Rauzy, 2015), and a sufficient mathematical framework, i.e. GTS (Rauzy, 2008). With the support of this modelling language, the analysts can provide indications of the system structure as well as the operational process.

ACKNOWLEDGEMENTS

The authors would gratefully thank Christian Holden, for kindly sharing his knowledge and design experience about the subsea laboratory at the Department of Mechanical and Industrial Engineering at the Norwegian University of Science and Technology (NTNU) and for patiently answering technical questions.

REFERENCES

Alavi, M., & Leidner, D.E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS quarterly*, 107–136.

Asbjørnsen, O. (1992). *Systems engineering principles and practices*. Maryland, USA: Skarpodd.

Batteux, M., Prosvirnova, T., & Rauzy, A. (2015). *System Structure Modeling Language (S2ML)*.

Bertsche, B. (2008). *Reliability in automotive and mechanical engineering: determination of component and system reliability*: Springer Science & Business Media.

DNV-RP-D102. (2012). *Failure Mode and Effect Analysis (FMEA) of Redundant Systems*. Høvik, Norway: DNV.

Haskins, C. (2008). *Systems engineering analyzed, synthesized, and applied to sustainable industrial park development*. PhD thesis: NTNU.

ISO 31000. (2009). *Risk management-Priciples and guidelines*: International Organization for Standardization.

Issad, M., Kloul, L., & Rauzy, A. (2017). *A scenario-based FMEA method and its evaluation in a railway context*. Paper presented at the Reliability and Maintainability Symposium (RAMS).

Kirsch, P., Hineb, A., & Maybury, T. (2015). A model for the implementation of industry-wide knowledge sharing to improve risk management practice. *Safety Science, 80*, 66–76.

Ligęza, A., & Potempa, T. (2012). *Artificial intelligence for knowledge management with bpmn and rules.* Paper presented at the IFIP International Workshop on Artificial Intelligence for Knowledge Management.

Motawa, I., & Almarshad, A. (2013). A knowledge-based BIM system for building maintenance. *Automation in Construction, 29*, 173–182.

NORSOK Z-013. (2010). *Risk and emergency preparedness assessment*.

Prosvirnova, T. (2014). *AltaRica 3.0: a Model-Based approach for Safety Analyses.* Computational Engineering, Finance, and Science [cs.CE] Ecole Polytechnique.

Rauzy, A. (2008). Guarded transition systems: A new states/events formalism for reliability studies. *Proceedings of the Institution of Mechanical Engineers. Part O, Journal of risk and reliability, 222*(4).

Rice, R.G., & Spence, P.R. (2016). Thor visits Lexington: Exploration of the knowledge-sharing gap and risk management learning in social media during multiple winter storms. *Computers in Human Behavior, 65*(612–618).

Su, E., & Dou, J. (2013). How Does Knowledge Sharing Among Advisors From Different Disciplines Affect the Quality of the Services Provided to the Family Business Client? An Investigation From the Family Business Advisor's Perspective. *Family Business Review, 26*(3), 256–270. doi:10.1177/0894486513491978.

White, S.A., & Miers, D. (2008). *BPMN Modelling and Reference Guide: Understanding and Using BPMN*. usa: Future Strategies.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Study on the flight landing quality evaluation model with analytical network process and matter element analysis method

Jingsong Lei, Wenbing Chang, Lei Li, Shenghan Zhou & Yiyong Xiao
School of Reliability and Systems Engineering, Beihang University, Beijing, China

ABSTRACT: This study firstly proposes an evaluation model for evaluating flight landing quality based on Analytical Network Process (ANP) and Matter Element Analysis Method (MEAM). Focus on the complexity of flight landing quality synthetic evaluation and incompatibility between single factor identification and hierarchical classification, building element matrics, using ANP to calculate the weight of each parameter and calculating correlation degree to evaluate the flight landing quality. The results of case study show that the model with ANP and matter element analysis model can be applied to evaluate the quality of landing, and compared with the classic Comprehensive Weighting Method (CWM), this evaluation result is more objective, accurate and reasonable.

1 INTRODUCTION

Aircraft landing is one of the most important stages in the whole flight phase, which is seriously affected by the environment and human factors, and the potential failure rate is the highest (Feng Yachang et al.1993). Data of fatal accidents makes it clear that about 23 percent of all fatal accidents take place in the landing phase from 2003 to 2012, though this phase just accounts for about 1% proportion in whole flight phase (X. H. Wang et al. 2009). So it is important to analyze the landing quality so as to find out the way to improve landing quality and make landing safer.

Nowadays, the Quick Access Recorder (QAR) is widely used for aircraft health status monitoring. The QAR data contains variety of flight information. It is used in many methods to analyze landing status and aid decision-making in landing process.

There are few studies studying landing quality, Zhang Rongjia analyze and discuss heavy landing of A321 flight based on QAR data (2012), Feng Yachang et al (1993) used loop separation parameter method to evaluate landing quality. Although there are few studies on landing quality, there are a number of literatures on the study of landing. Landing quality is a part of the landing. In condition of lacking studies on landing quality, it is necessary to study the literatures of landing. Julia A. Bennell et al (2017) considered the scheduling of aircraft landings on a single runway, built algorithms for creating landing schedules and verified the algorithms by random test data and real data from London Heathrow airport. B.S. Girish (2016) proposed a hybrid particle swarm optimization algorithm in a rolling horizon framework to solve the Aircraft Landing Problem (ALP).

Both Analytical Network Process (ANP) and Matter Element Analysis Method (MEAM) are common methods. ANP describes the relationship of elements in the system by network structure rather than hierarchy structure. The indexes in system may have mutual influence on each other. ANP is able to find out the correlation among indexes. Zhou Lisha (2009) used ANP to evaluate power customer satisfaction. Wang Haolun et al (2014) applied ANP to analyze SWOT for strategy decision of enterprise. Xiang Yong & Ren Hong (2014) evaluated the smart city based on the ANP-TOPSIS method. MEAM is widely used in evaluation. Huang Huiling et al (2010) used MEAM to evaluate land eco-security, Huang Jian et al (2007) used MEAM to evaluate power quality. ANP and MEAM are widely applied, yet they have not been applied together. This paper will firstly integrate ANP and MEAM. Specifically, this paper will apply ANP to calculate the weight of each index and apply MEAM to evaluate landing quality. 5 flights with their respectively flight data are studied for getting reasonable analysis results and complete analysis method.

The rest of the paper is organized as follows. In Section 2, we introduce applying ANP to calculate index weight. In Section3, we introduce the quality evaluating model of landing using MEAM. In section 4, case is studied. Finally, in Section 5, conclusions are made.

2 APPLYING ANP TO CALCULATE INDEX WEIGHT

In 1996, Saaty proposed network analysis method (ANP) on the basis of Analytic Hierarchy Process (AHP). ANP allows multiple indexes that can be quantified or difficult to quantify, and considers the association or feedback relationships between elements at different levels and elements within a set of elements. Therefore, compared with AHP, ANP reflect and describe the decision problem more realistically. This study adopts the method proposed in document by Sun Hongcai (2011), and has the following steps:

2.1 *Determining element hyper matrix*

According to the scale defined in Table 1, taking one element $P_s(s=1,2,\ldots,m)$ in the control layer as a criterion, taking the element $e_{jl}\left(l=1,2,\ldots,n_j\right)$ in $C_j=\left(j=1,2,\ldots,N\right)$ as a sub criterion, other elements in the element group $C_j=\left(j=1,2,\ldots,N\right)$ will be compared according to their impact on $e_{jl}\left(l=1,2,\ldots,n_j\right)$, and $N\times\sum n_j=\left(j=1,2,\ldots,N\right)$ judgement matrices can be obtained; calculate the eigenvectors corresponding to the maximum eigenvalue of each matrix, and then the consistency test is performed. If checked, these eigenvectors are normalized; if not checked, a comparison matrix is constructed. These normalized eigenvectors form a $\sum n_j\left(j=1,2,\ldots,N\right)$ order super matrix, and the matrix consists of $N\cdot N$ block matrices.

Table 1. Judgment scale definition.

Scale	Definition
1	Factor *i* is as important as factor *j*
3	Factor *i* is slightly more important than factor *j*
5	Factor *i* is obvious more important than factor *j*
7	Factor *i* is strongly more important than factor *j*
9	Factor *i* is extremely more important than factor *j*
2,4,6,8	The median of the adjacent judgments above
Reciprocal	If the ratio of factor *i* to factor j is a_{ij}, the ratio of factor j to factor i is $a_{ji}=1/a_{ij}$

2.2 *Determining the weight matrix of elements*

Taking one element $P_s\left(s=1,2,\ldots,m\right)$ in the control layer and one element group as criteria. Comparing other element groups with the criteria and *N* judgement matrices can be constructed. It is also necessary to check the consistency of these judgment matrices and find the eigenvectors, and these normalized eigenvectors are constructed into a weighted matrix of order *N*.

2.3 *Determining weighted super matrices*

Multiplying each element of the weighted matrix in step 2 with a block of matrices in step 1, a weighted super matrix is formed. The weighted super matrix reflects the control action of the element group on the element and the feedback of the element to the element group.

2.4 *Calculating index weight*

After obtaining the weighted super matrix, the relative sorting vector of the elements is determined by the corresponding calculation method according to the subordinate matrix type, that is, the weight of the *N* elements.

3 QUALITY EVALUATING MODEL OF LANDING USING MEAM

MEAM is the rule and method of studying the problem of contradiction, it is the edge subject of systematic science, mathematics and thinking science. It is a transverse subject which is applied widely in natural science and social science. It has two theories: one is the matter element and the matter element theory, the other is the mathematical tool based on the extension set (Cai Wen 1999).

The ordered triple consists of matter, character and quantity are regarded as the basis element to describe matter in matter element analysis, denoted as: *R = (N, C, V)*, where *N* refers to matters, *C* refers to the characteristics which can show nature, function of matters and relationship between behavior and matters. *V* is the quantity of *C*, which determine scope of one characteristic. MEAM studies the extension of matter elements by means of extension sets which is determined by correlation functions and the specific conditions of the matter element.

3.1 *Index system and the level limit of each index*

Among the data recorded, there are 6 indexes, which are ground velocity, distance to be flown, roll angle, angle of pitch, lateral acceleration, normal acceleration. Both ground velocity and distance to be flown

are indexes that reflect the condition related to ground. Similarly, roll angle and angle of pitch are angle indexes, lateral acceleration and normal acceleration belong to acceleration category. Considering these characteristic and the usage of Super Decisions Software which is used to implement ANP, this paper build the index system (see Figure 1).

Each landing quality index will be rated as "excellent quality—level 1", "quality between excellent and good—level 2", "good quality—level 3", "quality between good and qualified—level 4", "qualified—level 5", recorded as „„. All 5 levels can not only avoid excessive deviation because of few levels, but also decrease the work of calculation to a certain range (Xu Yonghai & Xiao Xiangning 2004). Level limit of each index is determined by analyzing raw flight data (see Table 2).

3.2 *The establishment of matter element model*

For one flight, V is the quantification of Index C, $R=(N, C, V)$ is the basic element (Cai Wen et al. 1997). Angle of pitch, distance to be flown, roll angle, normal acceleration, lateral acceleration and ground velocity will be used to describe the landing quality. They are recorded as c_1,c_2,c_3,c_4,c_5,c_6, and the corresponding quantity are v_1,v_2,v_3,v_4,v_5,v_6, so it can be expressed as

$$\mathrm{R}(\mathrm{N},\mathrm{C},\mathrm{V})=\begin{vmatrix} N & c_1 & v_1 \\ & c_2 & v_2 \\ & \vdots & \vdots \\ & c_6 & v_6 \end{vmatrix} \tag{1}$$

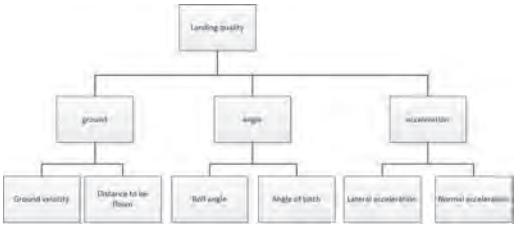


Figure 1. Evaluation index system of landing quality.

The classical field matter element of evaluation of landing quality is

$$R_{0j}\left(N_{0j},C_i,V_{0j}\right)=\begin{vmatrix} N_{0j} & c_1 & v_{01j} \\ & c_2 & v_{02j} \\ & \vdots & \vdots \\ & c_6 & v_{06j} \end{vmatrix} = \begin{vmatrix} N_{0j} & c_1 & \langle a_{01j},b_{01j}\rangle \\ & c_2 & \langle a_{02j},b_{02j}\rangle \\ & \vdots & \vdots \\ & c_6 & \langle a_{06j},b_{06j}\rangle \end{vmatrix} \tag{2}$$

where N_{0j} is the level j of landing quality, $j = 1$ indicates that the quality of the landing is excellent, $j = 2$ indicates that the quality is between excellent and good, $j = 3$ indicates that the quality is good, $j = 4$ indicates that the quality is between good and qualified, $j = 5$ indicates that the quality is qualified. C_i are six characteristics, $i = 1$–6 respectively represents angle of pitch, distance to be flown, roll angle, normal acceleration, lateral acceleration and ground velocity. v_{0ij} is the range of c_i, namely classical field. Classical field is an interval range, which indicates the basic interval of the variation of landing quality, and the range of v_{0ij} is interval $< a_{0ij},b_{0ij} >$, which can be recorded as: $v_{0ij} = < a_{0ij},b_{0ij} > (i = 1, 2, \ldots 6)$.

Segment field matter element is

$$R_p\left(\mathrm{N},C,V_P\right)=\begin{vmatrix} N & c_1 & v_{p1} \\ & c_2 & v_{p2} \\ & \vdots & \vdots \\ & c_6 & v_{p6} \end{vmatrix} = \begin{vmatrix} N & c_1 & \langle a_{p1},b_{p1}\rangle \\ & c_2 & \langle a_{p2},b_{p2}\rangle \\ & \vdots & \vdots \\ & c_6 & \langle a_{p6},b_{p6}\rangle \end{vmatrix} \tag{3}$$

where N is all quality levels, c_i are six characteristics, v_{pi} is the range of c_i, namely segment field. $v_{pi} = < a_{pi},b_{pi} > (i = 1,2,\ldots,6)$, obviously, $v_{0ij} \in v_{pi}$ $(i = 1,2,\ldots,6)$.

Table 2. Level limit of landing quality index.

indexes	Q1	Q2	Q3	Q4	Q5
Angle of pitch/°	[0,0.800)	[0.800,1.600)	[1.600,2.400)	[2.400,3.200)	[3.200,4.000]
Distance to be flown/m	[0.50.00)	[50.00,100.00)	[100.00,150.00)	[150.00,200.00)	[200.00,250.00]
Roll angle/°	[0,1.000)	[1.000,2.000)	[2.000,3.000)	[3.000,4.000)	[4.000,5.000]
Normal acceleration/m·s^{-2}	[13.000,17.400)	[17.400,21.800)	[21.800,26.200)	[26.200,30.600)	[30.600,35.000)
Lateral acceleration/m·s^{-2}	[0,1.200)	[1.200,2.400)	[2.400,3.600)	[3.600,4.800)	[4.800,6.000]
Ground velocity/m/s	[23.000,25.000)	[25.000,27.000)	[27.000,29.000)	[29.000,31.000)	[31.000,33.000]

For an evaluation flight, the measurement data and analysis results are represented by matter element R_0 called landing quality to evaluate matter-element. N_0 stands for quality level, v_i is the quantity of c_i.

$$R_0(N_0, C, \mathrm{V}) = \begin{vmatrix} N & c_1 & v_1 \\ & c_2 & v_2 \\ & \vdots & \vdots \\ & c_6 & v_6 \end{vmatrix} \tag{4}$$

Extension set is characterized with the correlation function, and the range of correlation function is the whole real axis, manage to express the correlation function of extension set by algebraic formula, so qualitative question will be quantitative. The correlation function values is calculated by formula (5):

$$K_j(v_i) = \begin{cases} -\dfrac{\rho(v_i, v_{0ij})}{|v_{0ij}|}, v_i \in v_{0ij} \\ \dfrac{\rho(v_i, v_{0ij})}{\rho(v_i, v_{pi}) - \rho(v_i, v_{0ij})}, v_i \notin v_{0ij} \end{cases} \tag{5}$$

And

$$\begin{aligned} \rho(v_i, v_{0ij}) &= \left| v_i - \frac{1}{2}(a_{0ij} + b_{0ij}) \right| - \frac{1}{2}(b_{0ij} - a_{0ij}) \\ \rho(v_i, v_{pi}) &= \left| v_i - \frac{1}{2}(a_{pi} + b_{pi}) \right| - \frac{1}{2}(b_{pi} - a_{pi}) \\ (i &= 1,2,\cdots,6; j = 1,2,\cdots,6) \end{aligned} \tag{6}$$

Comprehensive correlation degree is calculated by formula (7):

$$K_j(P_0) = \sum_{i=1}^{6} w_{ij} K_j(v_i) \tag{7}$$

where w_{ij} represents the weight of each evaluation index, $K_j(P_0)$ represents the correlation degree between the landing quality and the level j. If $K_j = max\{K_j(P_0)\}(j = 1,2,\cdots,6)$, the quality of the landing P_0 is to be evaluated is level *j*.

4 CASE STUDY

4.1 *Landing quality indexes of the subject to be assessed*

Considering the use of MEAM, this case filter out 5 flight with their data at the time of landing (see Table 3).

Table 3. Flight data.

Evaluation index	No. 1	No. 2	No. 3	No. 4	No. 5
Angle of pitch/°	1.387	1.200	3.041	3.159	1.939
Distance to be flown/m°	39.67	210.57	230.44	88.50	27.46
Roll angle/	0.879	4.136	1.538	0.088	0.857
Normal acceleration/ $m{\cdot}s^{-2}$	16.976	14.661	34.724	13.118	19.291
Latern acceleration/ $m{\cdot}s^{-2}$	1.102	0.157	0.315	0.472	3.307
Ground velocity/m/s	31.513	26.783	28.496	27.446	28.564

4.2 *Matter element model for landing quality evaluation*

The classical matrix and the segment field matrix are

$$R_{01} = \begin{vmatrix} Q_1 & c_1 & < 0{,}800 > \\ & c_2 & < 0{,}50.00 > \\ & c_3 & < 0{,}1.000 > \\ & c_4 & < 13.000{,}17.400 > \\ & c_5 & < 0{,}1.200 > \\ & c_6 & < 23.000{,}25.000 > \end{vmatrix}$$

$$R_{02} = \begin{vmatrix} Q_2 & c_1 & < 0.800{,}1.600 > \\ & c_2 & < 50.00{,}100.00 > \\ & c_3 & < 1.000{,}2.000 > \\ & c_4 & < 17.400{,}21.800 > \\ & c_5 & < 1.200{,}2.400 > \\ & c_6 & < 25.000{,}27.000 > \end{vmatrix}$$

$$R_{03} = \begin{vmatrix} Q_3 & c_1 & < 1.600{,}2.400 > \\ & c_2 & < 100.00{,}150.00 > \\ & c_3 & < 2.000{,}3.000 > \\ & c_4 & < 21.800{,}26.200 > \\ & c_5 & < 2.400{,}3.600 > \\ & c_6 & < 27.000{,}29.000 > \end{vmatrix}$$

$$R_{04} = \begin{vmatrix} Q_4 & c_1 & < 2.400{,}3.200 > \\ & c_2 & < 150.00{,}200.00 > \\ & c_3 & < 3.000{,}4.000 > \\ & c_4 & < 26.200{,}30.600 > \\ & c_5 & < 3.600{,}4.800 > \\ & c_6 & < 29.000{,}31.000 > \end{vmatrix}$$

$$R_{05} = \begin{vmatrix} Q_5 & c_1 & <3.200,4.000> \\ & c_2 & <200.00,250.00> \\ & c_3 & <4.000,5.000> \\ & c_4 & <30.600,35.000> \\ & c_5 & <4.800,6.000> \\ & c_6 & <31.000,33.000> \end{vmatrix}$$

$$R_p = \begin{vmatrix} Q_p & c_1 & <0,4.000> \\ & c_2 & <0,250.00> \\ & c_3 & <0,5.000> \\ & c_4 & <13.000,35.000> \\ & c_5 & <0,6.000> \\ & c_6 & <23.000,33.000> \end{vmatrix}$$

Matter element matrices are

$$R_1 = \begin{vmatrix} N_1 & c_1 & 1.387 \\ & c_2 & 39.67 \\ & c_3 & 0.879 \\ & c_4 & 16.976 \\ & c_5 & 1.102 \\ & c_6 & 31.513 \end{vmatrix} \quad R_2 = \begin{vmatrix} N_2 & c_1 & 1.200 \\ & c_2 & 210.57 \\ & c_3 & 4.136 \\ & c_4 & 14.661 \\ & c_5 & 0.157 \\ & c_6 & 26.783 \end{vmatrix}$$

$$R_3 = \begin{vmatrix} N_3 & c_1 & 3.041 \\ & c_2 & 230.44 \\ & c_3 & 1.538 \\ & c_4 & 34.724 \\ & c_5 & 0.315 \\ & c_6 & 28.496 \end{vmatrix} \quad R_4 = \begin{vmatrix} N_4 & c_1 & 3.159 \\ & c_2 & 88.50 \\ & c_3 & 0.088 \\ & c_4 & 13.118 \\ & c_5 & 0.472 \\ & c_6 & 27.446 \end{vmatrix}$$

Table 4. Correlation value of each flight and its respective indexes.

Correlation degree	No. 1	No. 2	No. 3	No. 4	No. 5
$k_1(v_1)$	$k_1(v_1) = -0.297$	$k_1(v_1) = -0.25$	$k_1(v_1) = -0.700$	$k_1(v_1) = -0.737$	$k_1(v_1) = -0.370$
$k_2(v_1)$	$k_2(v_1) = 0.266$	$k_2(v_1) = 0.5$	$k_2(v_1) = -0.600$	$k_2(v_1) = -0.647$	$k_2(v_1) = -0.149$
$k_3(v_1)$	$k_3(v_1) = -0.133$	$k_3(v_1) = -0.25$	$k_3(v_1) = -0.401$	$k_3(v_1) = -0.468$	$k_3(v_1) = 0.424$
$k_4(v_1)$	$k_4(v_1) = -0.422$	$k_4(v_1) = -0.5$	$k_4(v_1) = 0.199$	$k_4(v_1) = 0.051$	$k_4(v_1) = -0.192$
$k_5(v_1)$	$k_5(v_1) = -0.567$	$k_5(v_1) = -0.625$	$k_5(v_1) = -0.142$	$k_5(v_1) = -0.046$	$k_5(v_1) = -0.394$
$k_1(v_2)$	$k_1(v_2) = 0.207$	$k_1(v_2) = -0.803$	$k_1(v_2) = -0.902$	$k_1(v_2) = -0.303$	$k_1(v_2) = 0.451$
$k_2(v_2)$	$k_2(v_2) = -0.207$	$k_2(v_2) = -0.737$	$k_2(v_2) = -0.870$	$k_2(v_2) = 0.23$	$k_2(v_2) = -0.451$
$k_3(v_2)$	$k_3(v_2) = -0.603$	$k_3(v_2) = -0.606$	$k_3(v_2) = -0.804$	$k_3(v_2) = -0.115$	$k_3(v_2) = -0.725$
$k_4(v_2)$	$k_4(v_2) = -0.736$	$k_4(v_2) = -0.211$	$k_4(v_2) = -0.721$	$k_4(v_2) = -0.41$	$k_4(v_2) = -0.817$
$k_5(v_2)$	$k_5(v_2) = -0.802$	$k_5(v_2) = 0.211$	$k_5(v_2) = 0.391$	$k_5(v_2) = -0.558$	$k_5(v_2) = -0.863$
$k_1(v_3)$	$k_1(v_3) = 0.121$	$k_1(v_3) = -0.784$	$k_1(v_3) = -0.259$	$k_1(v_3) = 0.088$	$k_1(v_3) = 0.143$
$k_2(v_3)$	$k_2(v_3) = -0.121$	$k_2(v_3) = -0.712$	$k_2(v_3) = 0.462$	$k_2(v_3) = -0.912$	$k_2(v_3) = -0.143$
$k_3(v_3)$	$k_3(v_3) = -0.561$	$k_3(v_3) = -0.568$	$k_3(v_3) = -0.231$	$k_3(v_3) = -0.956$	$k_3(v_3) = -0.572$
$k_4(v_3)$	$k_4(v_3) = -0.707$	$k_4(v_3) = -0.136$	$k_4(v_3) = -0.487$	$k_4(v_3) = -0.971$	$k_4(v_3) = -0.714$
$k_5(v_3)$	$k_5(v_3) = -0.780$	$k_5(v_3) = 0.136$	$k_5(v_3) = -0.616$	$k_5(v_3) = -0.978$	$k_5(v_3) = -0.786$
$k_1(v_4)$	$k_1(v_4) = 0.096$	$k_1(v_4) = -0.378$	$k_1(v_4) = -0.984$	$k_1(v_4) = 0.027$	$k_1(v_4) = -0.231$
$k_2(v_4)$	$k_2(v_4) = -0.096$	$k_2(v_4) = -0.623$	$k_2(v_4) = -0.979$	$k_2(v_4) = -0.973$	$k_2(v_4) = 0.430$
$k_3(v_4)$	$k_3(v_4) = -0.548$	$k_3(v_4) = -0.811$	$k_3(v_4) = -0.969$	$k_3(v_4) = -0.987$	$k_3(v_4) = -0.285$
$k_4(v_4)$	$k_4(v_4) = -0.699$	$k_4(v_4) = -0.874$	$k_4(v_4) = -0.937$	$k_4(v_4) = -0.991$	$k_4(v_4) = -0.523$
$k_5(v_4)$	$k_5(v_4) = -0.774$	$k_5(v_4) = -0.906$	$k_5(v_4) = 0.063$	$k_5(v_4) = -0.993$	$k_5(v_4) = -0.643$
$k_1(v_5)$	$k_1(v_5) = 0.082$	$k_1(v_5) = 0.131$	$k_1(v_5) = 0.263$	$k_1(v_5) = 0.393$	$k_1(v_5) = -0.439$
$k_2(v_5)$	$k_2(v_5) = -0.812$	$k_2(v_5) = -0.869$	$k_2(v_5) = -0.738$	$k_2(v_5) = -0.607$	$k_2(v_5) = -0.252$
$k_3(v_5)$	$k_3(v_5) = -0.541$	$k_3(v_5) = -0.935$	$k_3(v_5) = -0.869$	$k_3(v_5) = -0.803$	$k_3(v_5) = 0.244$
$k_4(v_5)$	$k_4(v_5) = -0.694$	$k_4(v_5) = -0.956$	$k_4(v_5) = -0.913$	$k_4(v_5) = -0.869$	$k_4(v_5) = -0.100$
$k_5(v_5)$	$k_5(v_5) = -0.770$	$k_5(v_5) = -0.967$	$k_5(v_5) = -0.934$	$k_5(v_5) = -0.902$	$k_5(v_5) = -0.357$
$k_1(v_6)$	$k_1(v_6) = -0.814$	$k_1(v_6) = -0.320$	$k_1(v_6) = -0.437$	$k_1(v_6) = -0.355$	$k_1(v_6) = -0.446$
$k_2(v_6)$	$k_2(v_6) = -0.752$	$k_2(v_6) = -0.109$	$k_2(v_6) = -0.249$	$k_2(v_6) = -0.091$	$k_2(v_6) = -0.261$
$k_3(v_6)$	$k_3(v_6) = -0.628$	$k_3(v_6) = -0.054$	$k_3(v_6) = -0.252$	$k_3(v_6) = 0.233$	$k_3(v_6) = 0.218$
$k_4(v_6)$	$k_4(v_6) = -0.257$	$k_4(v_6) = -0.370$	$k_4(v_6) = -0.101$	$k_4(v_6) = -0.259$	$k_4(v_6) = -0.089$
$k_5(v_6)$	$k_5(v_6) = 0.257$	$k_5(v_6) = -0.527$	$k_5(v_6) = -0.357$	$k_5(v_6) = -0.444$	$k_5(v_6) = -0.354$

$$R_5 = \begin{vmatrix} N_5 & c_1 & 1.939 \\ & c_2 & 27.46 \\ & c_3 & 0.857 \\ & c_4 & 19.291 \\ & c_5 & 3.307 \\ & c_6 & 28.564 \end{vmatrix}$$

After inputting the matter element to be evaluated to the matter element model, the results can be output (see Table 4).

4.3 *Applying ANP to determine index weight*

As shown in chapter 2, the calculation of ANP is very complicated, this study uses Super Decisions software (SD) to calculate the weights of indexes. 3 first level indexes shown in Figure 1 are input to SD as 3 clusters, 6 second level indexes shown in Figure 1 are input to SD as 6 nodes, then establish connections between groups and groups, nodes and nodes, and the ANP model is established (see Figure 2).

The direction of the arrow indicates the dominance of one group over the other. For example, ground velocity has effect on index distance to be flown, so there is a loop on the cluster ground.

This study uses the standards shown in Table 1 to compare different indexes, according to the expert's judgment, the specific data is input into the SD and the result is calculated (see Figure 3)

The results will be calculated by inputting the data of Table 3 and Table 4 to the formula (7) (see Table 5).

4.4 *Analysis and comparison of the results*

As can be seen from the Table 5, the quality level of No. 1 flight is 1, namely excellent quality, No. 2 flight and No. 4 flight are the same as No.1. The quality level of No. 3 flight is 5, namely the quality is qualified. The quality level of No. 5 flight is 2, namely quality between excellent and good. The results calculated by CWM are compared with the results obtained from MEAM (see Table 6).

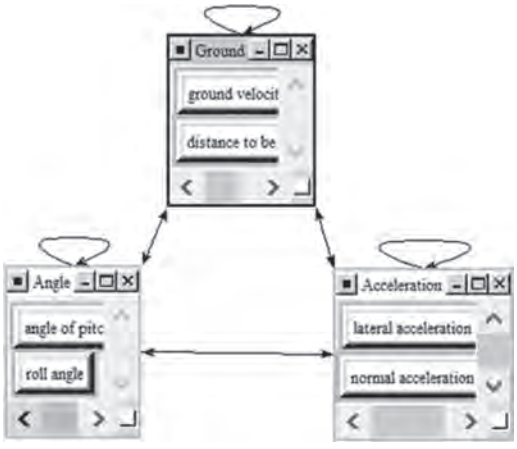


Figure 2. ANP model.

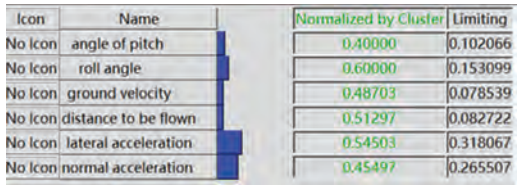


Figure 3. Weight of each index.

Table 5. Comprehensive correlation degree.

Correlation degree	Q_1	Q_2	Q_3	Q_4	Q_5
$K_j(N_1)$	–0.008	–0.354	–0.517	–0.639	–0.673
$K_j(N_2)$	–0.088	–0.581	–0.692	–0.686	–0.584
$K_j(N_3)$	–0.397	–0.581	–0.659	–0.664	–0.386
$K_j(N_4)$	0.020	–0.647	–0.704	–0.740	–0.788
$K_j(N_5)$	–0.218	–0.058	–0.083	–0.436	–0.543

Table 6. Evaluation results comparison.

Evaluation method	No. 1	No. 2	No. 3	No. 4	No. 5
MEAM	Q_1	Q_1	Q_5	Q_1	Q_2
CWM	Q_2	Q_2	Q_4	Q_2	Q_2

As shown in Table 6, the results of two methods are almost different. Only the result of No. 5 flight is same. But it is obvious that their results are similar, and there is not much discrepancy.

The reason why this happened is that the numerical values they calculated are different. The ANP calculates the weight of each index more accurately. The CWM only takes into account the level of index, but the MEAM calculates the correlation degree between the index and each level. The MEAM considers more information, the comprehensive correlation degree of MEAM is further deepened on the basis of correlation degree.

5 CONCLUSIONS

Firstly, based on extension mathematics, reasonably using correlation function to extend the interval to (–∞, +∞), MEAM considers more information, so the result can be more objective.

Secondly, ANP and MEAM is suitable to evaluate the landing quality and we can get satisfactory result. The process of analysis and the result is helpful to improve the landing quality and guide pilots. Thirdly, although CWM is the classical method, integrating ANP with MEAM is better in evaluating landing quality. Finally, ANP and MEAM is seldom used in the field of landing quality level evaluation, so the selection of evaluation index, the determination of the range of the value of evaluation can be further discussed and revised. In this study, the division of levels is based on linear data. There are lots of nonlinear data in real life, so it can be further studied to evaluate objects with nonlinear data.

The reason why ANP is chosen, rather than AHP, is that ANP can eliminate the interaction between indicators and more scientifically and objectively calculate the weight of indexes. However, in ANP, the expert scoring method is used to construct a comparison matrix between indexes. This method is very subjective, and in the future research, we will try to find a more objective and reasonable method for this problem.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (Grant No.71271009 & 71501007). The study is also sponsored by the Aviation Science Foundation of China (Grant No.2014ZG51075) and the Technical Research Foundation.

REFERENCES

Cai Wen (1999). The extension set and non-compati ble problems. Beijing International Academic Publisher.

Cai Wen, Yang Chuanyan & Lin Jinchu (1997). Method of extension engineering. Beijing Science Publishers.

Feng Yachang & Wang Yong (1993). A prediction for landing qualities of aircraft using loop separation parameter method. Journal of Beijing University of Aeronautics and Astronautics. No. 2. 36–42.

Huang Huiling, Luo Wenbing, Wu Cifang & Li Dongmei (2010). Evaluation of land eco-security based on matter element analysis. Transactions of the CSAE, vol. 26, no.3, pp.316–322.

Huang Jia, Zhou Lin, Li Qiuhua, Zhang Feng & Liu Huayong (2007). Evaluation of power quality based on the method of Matter-element. Journal of Chongqing University (Natural Science Edition), vol.30, no.6, pp.25–29.

Men Baohui & Liang Chuan (2003). Matter element method for evaluating water quality. Journal of Harbin Institute of Technology, pp. 358–361.

Sun Hongcai (2011). Network analytic hierarchy process and Decision Science [M], Beijing, National Defense Industry Press.

Wang Haolun, Gan Weihua & He Deshun (2014). A SWOT Analysis Method for Strategy Decision of Enterprise Based on ANP and TOPSIS, Science and Technology Management Research. pp.141–145.

Wang, X. H. P. Shu, X. Rong & L. Nei (2009). A Decision Support System Based on Support Vector Machine for Hard Landing of Civil Aircraft. Computer Science-Technology and Applications, 2009. IFCSTA '09. International Forum on, vol.2. 66–70.

Xiang Yong & Ren Hong (2014). The Study of Smart City Evaluation Based on the ANP-TOPSIS Method, Journal of Industrial Technological Economics.pp. 131–136.

Xu Yonghai & Xiao Xiangning (2004). Power Qualitty Problems in Deregulated Power Systems. Power System Technology, pp.48–52.

Zhang Rongjia & Zhang Lisheng (2012). Analysis and discussion on overload landing of A321. Journal of Civil Aviation Flight University of China, vol. 23, no. 1, pp. 49–53.

Zhou Lisa & Yu Xinhua (2009). Fuzzy Comprehensive Evalu ation of Power Customer Satisfaction Based on Analytic Network Process, Power System Technology. pp. 191–197.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Approach to a Bayesian decision model for cost-benefit analysis in security risk

D. Lichte & K.-D. Wolf
Institute for Security Systems, University of Wuppertal, Velbert, Germany

ABSTRACT: Security risk analysis and management for infrastructures is a challenging task as the uncertainties regarding both, the capabilities of security systems and various threat scenarios are high. Especially cost-benefit analysis regarding the investment in physical security systems to reduce the overall vulnerability of infrastructures is a complex problem. This paper presents an approach that is based on a quantitative model for vulnerability analysis previously introduced by the authors. Based on the model a Bayesian Decision Network (DN) is derived. The result of the DN is a Return on Security Investment (ROSI) based on the principle of the weakest path. The ROSI can be used to find the best outcome resulting from different configurations considering mitigation of security risks and required investments in security measures. In a last step the application of the developed approach to a simplified infrastructure is presented. Finally, the results are summarized and discussed.

1 INTRODUCTION

Security risk analysis and management of infrastructures are important issues because of a rising number of attacks with different targets, e.g. financial interests, sabotage or terrorism. At the same time they represent a challenging task, as the uncertainties regarding both, the capabilities of security systems and various threat scenarios, are high.

Especially cost-benefit analysis regarding the investment in security systems to mitigate the overall vulnerability of infrastructures subject to different types of threats is a complex problem. Although numerous approaches for (physical) security risk assessment and analysis exist, there are only few that consider uncertainty for threats and abilities of security systems. As existing methods for cost-benefit analysis are based on these approaches, they mostly do not consider uncertainties either. Additionally, most methods focus on a single attack scenario lacking a holistic view of the different feasible attack paths in an infrastructure equipped with a security system.

This paper presents an approach to a risk model that is based on a quantitative model for vulnerability analysis previously introduced by the authors. The quantitative vulnerability model is enhanced using Bayesian Decision Networks. Bayesian Networks (BN) allow a graphical representation of the security system and its function in a considered infrastructure. Additionally, a decision network (DN) based on an influence diagram is developed. The DN enables the analysis of outcomes resulting from different configurations considering mitigation of security risks and required investments in security measures. Based on a return on security invest (ROSI) analysis, the most valuable configuration can be chosen.

Therefore, the structure of the quantitative vulnerability model is transferred into BN structures and the probability density distributions of the used model are discretized to consider uncertainties. In a second step the paper explains the implementation of the BN into the DN based on security risks and security investments resulting from different configurations of the security system. The last step describes the application of the developed approach to a simplified infrastructure. Finally, the results are summarized and discussed.

2 STATE OF THE ART

2.1 *Security risk assessment*

Security comprises a number of issues covering different fields of expertise; therefore a comprehensive view is needed to conduct a holistic security assessment (Harnser Group 2010). Physical security as one part deals with the protection of infrastructures from intentional physical attacks (Beyerer et al. 2010). The aim of physical security measures is to prevent an attacker from reaching his objective by different means of protection, detection and intervention and also set up resilient structures to mitigate the consequences of successful attacks (Garcia 2008).

The corresponding security risk definition can be defined as (Contini et al. 2012, Mc Gill et al. 2007):

$$Risk = Threat \cdot Vulnerability \cdot Consequence \quad (1)$$

This definition combines—based on a quantitative analysis—consequences of attacks and probabilities of threat scenarios with the risk of individual attacks being successful, defined as vulnerability. The above quantitative definition of risk may help to deduce acceptable risks and necessary measures to mitigate risks (Broder & Tucker 2012). Inherent uncertainties regarding the three risk factors should be cautiously considered (Campbell & Stamp 2004).

Various approaches for security risk assessment have been developed that may be divided into qualitative, quantitative and hybrid methods (Meritt 2008). Qualitative methods are mostly based on expert knowledge, while existing quantitative methods use discrete probabilities. The former are more widespread because of their ease of use, while at the same time the application of expert knowledge can lead to inaccurate or even wrong results (Landoll 2011). Additionally, some quantitative methods aiming at cost-benefit analysis have been developed. Typically, cost-benefit analyses of security measures would account for potential financial losses as a result of an attack, the probability of occurrence of various attack scenarios and the vulnerability of the security system (Flammini et al. 2009). This analysis yields accurate results but raises the complexity compared to qualitative methods (Landoll 2011).

2.2 *Cost-benefit analysis in security risk assessment*

Generally, a cost-benefit analysis for security measures is difficult, as the benefits for the measures are hard to evaulate (Butler 2002). Thus, the support of decision-making concerning security investments is a developing area (Abrahamsen et al. 2015).

Different approaches were proposed in an IT-security context that are based on methods and models of vulnerability analysis to include the effectiveness of countermeasures into the general risk assessment. For example, (Bistarelli et al. 2006) propose defense trees as an enhancement of attack trees to calculate the effectiveness of possible security measures. The SAEM method proposed by (Butler 2002) is based on expert knowledge and aims at including the estimated effectiveness into the risk assessment. An interesting approach is the definition of a return on security investment (ROSI). The ROSI is a ratio of mitigated consequences of attacks and the investment for the therefore needed security measures (Sonnenreich et al. 2005). Although proposed in an IT-context, a more general use in security risk assessment is conceivable.

In the area of physical security risk assessment only a few approaches exist that provide strategies for decision-making with respect to security measures. Wyss et al. propose a security risk metric that similar to the IT—related methods is based on the vulnerability analysis. The basic idea of the method is a rather general approach that every applied measure should make the easiest path towards a successful attack as difficult as possible considering the constraints of costs, operational and programmatic restrictions (Wyss et al. 2010). Another approach analyzes costs and benefits of different measures of aviation security (Stewart & Mueller 2013). Here, a current state is compared to different security measures with the same goals by a break-even analysis to find an estimated minimal probability of successful attacks to equal investment costs.

In conclusion, cost benefit analysis in security risk assessment is mostly based on the sub-part vulnerability assessment. A precise estimation of the inherent benefits of security measures is difficult as even the overall abilities of security measures are often uncertain. Therefore, existing methods lack a detailed coupling to the basic vulnerability assessment and rather supply general decision strategies. Additionally there are no methods of cost-benefit analysis based on quantitative methods, which are able to consider the above mentioned uncertainties to tackle this problem.

2.3 *Vulnerability assessment in security risk assessment*

Quantitative vulnerability analysis as part of the quantitative risk analysis is mostly based on methods adapted from reliability and general risk analysis. Here, the considered model is dependent on given attack scenarios (French & Gootzit 2011). This dependency is detrimental to a comprehensive analysis as knowledge about the behavior of a potential attacker may be insufficient (Cox Jr. 2009). The different modeling approaches can be further split up into mainly analytical but also formal methods. An overview of approaches is given by Nicol et al. (Nicol et al. 2004). Analytical methods are often based on attack trees, which can be seen as a derivative of fault trees already introduced by reliability analysis. Attack trees were first used by Schneier (Schneier 1999) for IT-security analysis and have been further developed by different authors since then, summarized e.g. by Vintr et al. (Vintr et al. 2012).

Contini et al. have introduced incoherent attack trees to characterize the dynamic behavior of the considered system (Contini et al. 2008). Additionally, they integrated simple probability distributions for protection into attack tree models to investigate the chronologic sequence of attacks. Hence, it is possible to analyze the security system's ability for an attack intervention by comparing the probabilities of residual protection and system's response (Contini et al. 2012).

Garcia describes this relation (Garcia, 2008), and thus uses feasible attack paths as part of different attack scenarios and corresponding barriers. The model is time-based and introduces the term of the critical detection point that is the latest possible point of detection that ensures a successful intervention.

Summarizing, the different existing approaches to analytical modeling and analysis of vulnerability are lacking the consideration of uncertainties in the system parameters and overall behavior. Additionally, these approaches do not allow a scenario spanning analysis of the whole security system as the analysis depends on specific scenarios.

2.4 *Bayesian (Decision) networks and influence diagrams*

Bayesian Networks (BN) are based on Bayesian probabilistics interpreting probability as a degree of belief. BN represent a combination of probability and graph theory. A BN therefore quantifies dependencies between various data, information or knowledge considering uncertainties (Jensen & Nielsen 2007). BN consist of nodes and connecting edges in directed acyclic graphs (DAG) linking parent and children nodes. BN therefore consist of (Gribaudo et al. 2015):

- Variables (nodes) with a finite set of states
- Directed edges between the nodes
- A conditional probability table describing the result of each node

The joint probability distribution for a set of nodes $U = \{A_1, \ldots, A_n\}$, is defined as:

$$P(U) = \prod_{i=1}^{n} P(A_i \mid parents(A_i)) \tag{2}$$

Influence diagrams extend the definition of BN to enable the consideration of decision problems by using the BN DAGs (Howard & Matheson 2005). The goal of these diagrams is to find the decision alternative that delivers the highest expected utility (Shachter 1986).

As the structure is similar to BN, influence diagrams also include chance nodes. Additionally three more types of nodes are added (Howard 1988):

- Decision nodes allow decision alternatives that may be controlled by the decision maker.
- Deterministic nodes represent constant values that only depend on the states of their parent nodes.
- Value nodes represent the utility function implemented into the decision process.

Influence diagrams are used in different fields, where DN are set up to visualize the structure and to support decision-making processes. Despite a widespread use of influence diagrams and DN for a wide range of decision problems, e.g. cost-benefit investment decisions in financial disciplines their use in security risk assessments is not very common. An approach to use BN to describe the security vulnerability of gas pipelines is described in (Fakhravar et al. 2017).

3 APPROACH

This paper presents an approach for cost-benefit analysis based on a quantitative model for vulnerability assessment that uses probability density functions (pdfs) introduced already by the authors (Lichte & Wolf 2017). The approach discretizes the pdfs and sets up a model that reflects the described functional relations in a BN based on conditional probabilities. Thus, the characteristics of single barriers are defined as submodels that include the barrier vulnerability. These submodels are then assembled to attack paths resulting from the topology of a considered infrastructure at a higher level model. Within the higher level model it is possible to derive the strength of the infrastructure to block possible attacks by combining the barrier vulnerability in the BN. The strength of a single attack path is considered as the complement of its vulnerability in accordance with the underlying model. Subsequently, the cost benefit DN based on an influence diagram is added. Therefore threat and consequence as the decision utility are implemented to establish a complete risk model. The cost benefit network calculates a return on security investment (ROSI) of different configurations based on an approach introduced in (Sonnenreich et al. 2005).

Finally the proposed approach is applied to a simple exemplary infrastructure composed of four barriers that allow four partly overlapping attack paths.

3.1 *Basic assumptions*

The underlying model is based on four basic assumptions, which characterize the most relevant behavior of a security system in an infrastructure (Lichte & Wolf 2016). These assumptions are used

in the probabilistic description of the systems' relations.

1. The weakest path of the security system determines the system's vulnerability as the chosen path of the attacker is uncertain.
2. The combination of protection and observation at barriers is necessary as an attacker is always able to break through a barrier given infinite time without being detected.
3. The detection of an attack is possible only if the protection is sufficient to prevent a break-through under observation until detection.
4. After detection, an attack can be stopped only if the residual protection along the remaining attack path lasts long enough to prevent the attacker from reaching the asset until intervention is completed.

3.2 *Discretization of the input PDFs*

The model comprises the three main input parameters protection P, observation O and intervention I, which are described as pdfs. To use these input parameters in a convenient BN with limited complexity, the describing pdfs have to be discretized.

Figure 1 uses the example of a normal pdf: the values for the intervals are derived by integrating the pdf within the upper and lower boundaries l and u of the intervals.

$$\int_{l}^{u} p_i(t)\,dt \tag{3}$$

The size of the intervals used in the BN should be determined with respect to the considered infrastructure and security measures as well as the intended accuracy. The observation period for P and O is equal and related to the single barriers, while the period for intervention is usually longer as it is related to the behavior of the whole system. For calculation reasons, the discretization intervals should be of the same size, so the number of intervals n_I in the description of I grows in relation to the number of barriers.

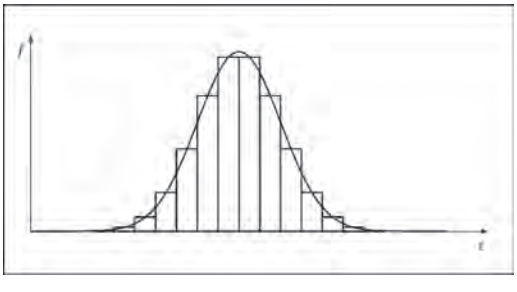


Figure 1. Schematic discretization.

$$n_I = n_B * n_{O,P} \tag{4}$$

Hence, the nodes for the input parameters protection, observation and intervention are built as prior distributions.

With the distribution nodes built at an earlier stage it is possible to derive the BN for barrier vulnerability based on the four assumptions about the behavior of a security system.

3.3 *Bayesian network for barrier vulnerability*

Basically, a combination of three different relations of the input parameters is needed to describe the characteristics of the security measures at a barrier of a security system: Detection D, residual protection R and timely intervention T.

A detection D of an attacker is triggered with the probability that the protection measure at a barrier prevents an attacker from a break-through until an observation is completed with detection. This allocates the conditional probability

$$D = P(t_p > t_o) \tag{5}$$

In the context of BN this can be interpreted as a chance node. The distribution table represents the following probability condition:

$$\forall k,l : P_D = P\left(D \mid t_{Pl}, t_{Ok}\right) = \begin{cases} 1 & \text{if } t_O\left(t_{Ok}\right) < t_P\left(t_{Pl}\right) \\ 0 & \text{else} \end{cases} \tag{6}$$

wherein t_P and t_O denote the time for protection and observation, k and l denote the running index of the corresponding time interval.

The second key relation in the vulnerability model is the ability for a timely intervention T. This parameter is based on the pdf I of the time needed for intervention t_I and the residual protection R and is therefore defined by the conditional probability given by:

$$T = P(t_{RP} > t_i) \tag{7}$$

The residual protection is the sum of all protection measures at the residual barriers of the system on the attack path. Figure 2 shows the BN for the vulnerability assessment of a barrier containing the nodes for detection and timely intervention.

As depicted, the node for timely intervention has the parent nodes "D", "I" and "P" of barrier i. Additionally, the protection nodes of *n* residual barriers are connected as barrier nodes. The resulting general distribution table reflects the following conditional probability for timely intervention:

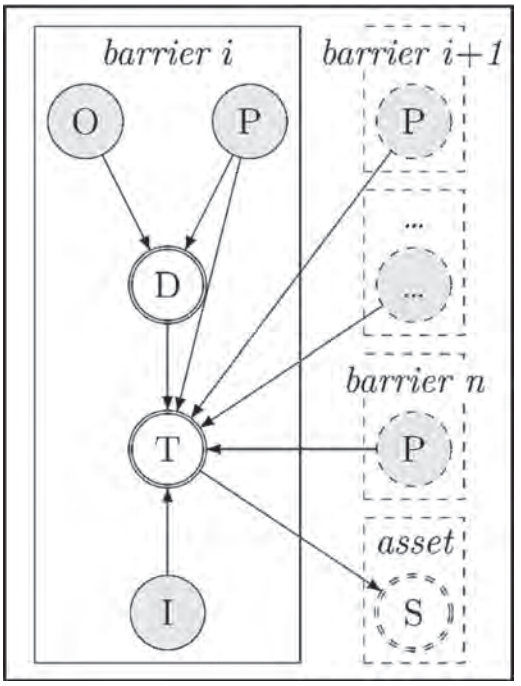


Figure 2. BN for barrier vulnerability.

$$\forall l^i,\ldots,l^n \quad \left(P_T = T \mid D, t_{\mathrm{Im}}, t^{(i)}_{Pl^{(i)}},\ldots,t^{(n)}_{Pl^{(n)}}\right) = \begin{cases} 1 & \text{if } t_I\left(t_{\mathrm{Im}}\right) < t^{(i)}_P\left(t^{(i)}_{Pl^{(i)}} + \ldots + t^{(n)}_P\left(t^{(n)}_{Pl^{(n)}}\right)\right) \\ 0 & \text{else} \end{cases} \tag{8}$$

In (8) $t^{(i)}_P$ denotes the protection time of the i-th barrier so that $t^{(i)}_{Pl^{(i)}}$ describes the state of the protection time of the i-th barrier falling into the l-th time interval.

According to the basic assumptions, the strength of a barrier is determined by the possibility to timely intervene an attack given its detection. As vulnerability is the complement of a barrier's strength, the BN for vulnerability is fully derived. In the next step the network for the system's vulnerability and risk assessment is set up.

3.4 *System vulnerability and risk assessment network decision network*

The risk assessment network consists of barriers of an attack path, an asset as target of an attack, a threat node and cost utility describing possible consequences of a considered attack. Both, the barrier nodes as well as the asset nodes comprise subsystems. The subsystem in the barriers contains the vulnerability model, while the subsystem of the asset comprises the computation of the ability of the attack path to block an attacker. Figure 3 depicts the DN for risk assessment.

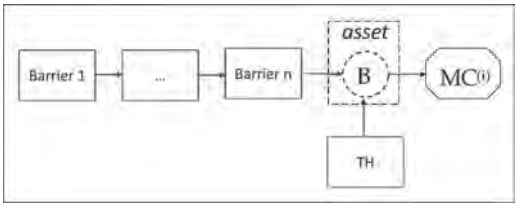


Figure 3. DN for risk assessment.

The deterministic node "B" within the submodel of the asset describes the ability of a blocked attack. The node depends on the parent nodes "T" for the timely intervention of all *n* barriers on the attack path. Additionally, the threat node "TH" serves as a parent node. An attack is blocked when the attacker is detected at a barrier and if a timely intervention—starting at the same barrier—is possible. The conditional probability for the evaluation of the deterministic function of the node "B" is then given by:

$$P_B = P\left(B \mid TH, T, \ldots, T^{(n)}\right) = \begin{cases} 1 & \text{if } P_D = 1, P_T = 1 \\ 0 & \text{else} \end{cases} \tag{9}$$

The parent node "TH" describes the probability of the occurrence of a threat and serves as a factor for the calculation of the general probability of a blocked attack (node "B").

Finally, a value node "MC" is added with the node "B" as a parent (see Fig. 4) to represent the consequences in the risk function in (1). Herein, the possible consequence is inserted as a constant monetary value. The calculated result of the utility function is the value of the consequences mitigated by the installed security measures on the attack path of the infrastructure. The linear utility function yields to:

$$MC = P_B \cdot Value \tag{10}$$

Thus, the derived DN enables a computation of the mitigated consequences and the remaining risk on the considered attack path. Similarly, other feasible attack paths of a system can be set up using the same values for the different threat nodes "TH" and the utility nodes "MC" respectively. Following the principle of the weakest path, the overall risk of a system with n paths is obtained by:

$$R_{System} = \max[\left(1 - MC^{(i)}\right),\ldots,\left(1 - MC^{(m)}\right)] \tag{11}$$

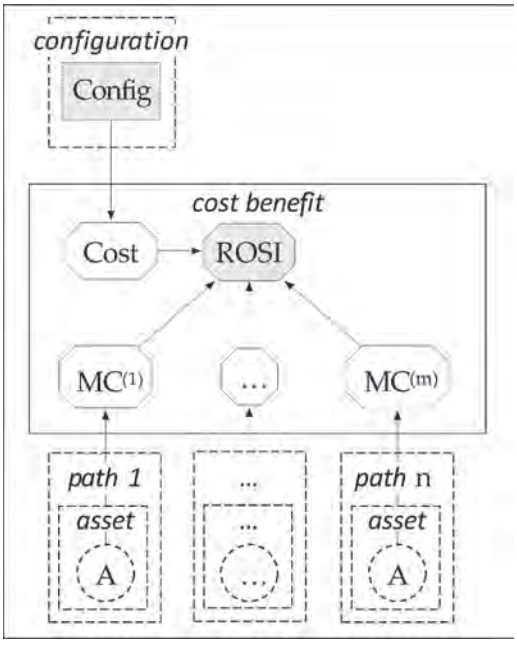


Figure 4. DN for cost benefit analysis.

3.5 *Cost benefit decision network*

Proceeding from the mitigated consequences, a cost benefit DN is constructed by adding a decision node to enhance the DN with the capabilities of an influence diagram. The decisions in the node symbolize the different configurations of the security system that are part of the analysis as shown in Figure 5. The decision node varies the values of the parameters for protection and observation measures at all barriers as estimated for the different configurations. These values need to be inserted into the "P" and "O" nodes in the subsystem of the vulnerability analysis, where the same barriers on different attack path are assigned the same values (see Fig. 4).

As depicted in Figure 4, a value node is additionally added to reflect the costs of the different configurations introduced by the decision node. Finally, the value node for the return on security invest (ROSI) is introduced. It combines the value of the mitigated consequences depending on its configuration and calculated in the risk assessment network with the costs of the configurations. The calculation of the ROSI is based on the principle of the weakest path, as the weakest path is the decisive path in case of an occurring attack. Hence we yield the following utility function for the number of *n* attack paths:

$$ROSI = \frac{\min\left(MC^{(i)},\ldots,MC^{(m)}\right) - Cost}{Cost} \qquad (12)$$

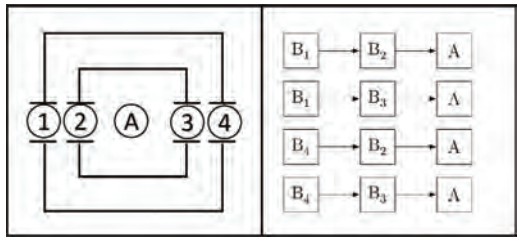


Figure 5. Structure and attack paths of the exemplary infrastructure.

This utility function computes a ratio between the mitigated consequences of the weakest path and the needed investment to realize the required security measures (Sonnenreich et al. 2005). In order to decide whether to invest or not the following basic rules do apply:

- Invest for $ROSI^{(i)} > ROSI^{(i+1)},\ldots,ROSI^{(n)} > 0$
- Do not invest for $ROSI \leq 0$

Thus, the basic decision whether to invest or not as well as the decision which configuration should be realized are possible.

In the next step the derived DN is applied to an exemplary simple infrastructure.

3.6 *Application of the DN to an exemplary infrastructure*

The DN is applied to a notional simplified infrastructure. Its structure is shown on the left part of Figure 5. The security system of the infrastructure consists of four barriers enabling four feasible attack paths.

The four attack paths within the infrastructure consist of two barriers (see Figure 6, right). Three different configurations are analyzed for a decision whether to invest and which configuration to choose. The values of the parameters for the barriers dependent on configuration and the costs for the configurations are listed in Table 1.

The intervention time does not depend on the configurations and is set to the values listed in Table 2.

The DN based on the exemplary infrastructure is depicted in Figure 6.

Finally, the vulnerability as well as the ROSI of the different configurations are calculated. The results are shown in Table 3.

Based on the results, the medium configuration is the optimum decision based on the DN and the ROSI formulation as it shows the best ratio between mitigated consequences and the required investment. It is also visible that the weakest attack path of the security system changes depending on

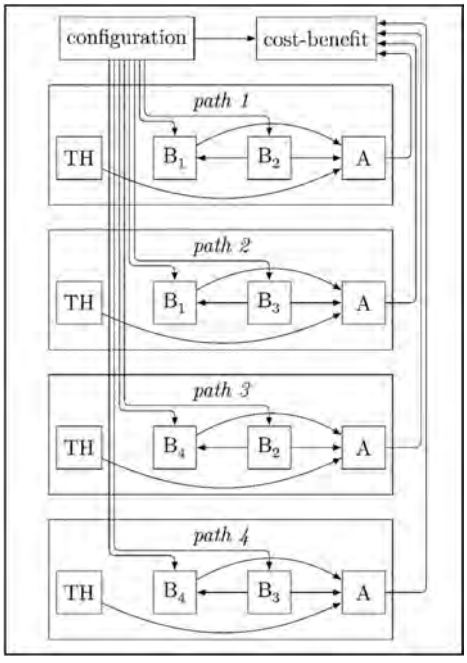


Figure 6. Overall DN for the exemplary infrastructure.

Table 1. Input parameters for P and O and needed investment for different configurations.

Config	1 (50 k$)		2 (150 k$)		3 (300 k$)	
Param. [μ, σ]	P	O	P	O	P	O
Barrier 1	180,30	240,60	180,30	150,60	240,40	240,60
Barrier 2	120,20	150,20	150,20	150,20	120,20	90,20
Barrier 3	60,20	120,30	180,20	120,30	60,20	30,10
Barrier 4	60,20	180,60	60,20	90,30	240,30	180,60

Table 2. Input parameters for I.

	I [μ,σ]
Barrier 1	360,60
Barrier 2	30,5
Barrier 3	30,5
Barrier 4	180,30

Table 3. Results for ROSI and Vulnerability.

Config	ROSI	V	Path
1	0.5692	0.4614	4
2	0.5909	0.4034	3
3	0.1950	0.1037	1

the configuration (see Table 3). The definition of the ROSI utility function considers this behavior of the vulnerability assessment model.

4 CONCLUSION

This paper proposes a new approach regarding cost-benefit analysis in physical security analysis using DN based on BN as well as influence diagrams. The approach enables the assessment of the vulnerability of the considered infrastructure and introduces a decision network that uses the ROSI as a utility function to support decision-making regarding a security investment. The vulnerability assessment is based on a model introduced by the authors using pdf-based parameter descriptions.

In a first step those pdfs which were used to describe the input parameters now are discretized for a further use in BN. Subsequently the BN submodel describing the characteristics of a security system regarding its vulnerability is derived and a risk assessment network based on attack paths is deducted introducing threat probability and a consequence utility function. Following, a decision between different configurations and related investment costs is implemented to establish a DN. In a last step the ROSI utility function is derived based on mitigated consequences of the weakest path and the costs of the different configurations. The ROSI function supports decision-making between different configurations introduced by the decision node. The application to an exemplary infrastructure shows the use of this approach and the consideration of the weakest path in the ROSI function.

Further research is needed to refine and further develop the presented approach. A software implementation would be useful to automate the discretization of the pdfs and the model building. Additionally, the definition of the threat probability as well as the consequence utility function need to be addressed to in more depth. The investment costs for security measures also should be analyzed in greater detail. Especially the implementation of a measure to relate the costs to a possible barrier related effort, e.g. the length of perimeter barriers or observation distances, would be useful.

REFERENCES

Abrahmansen, E.B. & Pettersen, K. & Aven, T. & Kaufmann, M. & Rosqvist, T. 2015. A framework for selection of strategy for management of security measures. *Journal of Risk Research 20* (3): 404–417.

Beyerer, J. & Geisler, J & Dahlem, A. & Winzer, P. 2010. Sicherheit: Systemanalyse und Design. In: *Sicherheitsforschung—Chancen und Perspektiven*. Berlin: Springer.

Bistarelli, S. & Fioravanti, F. & Peretti, P. 2006. Defense trees or economic evaluation of security investments. In: *First International Conference on Availability, Reliability and Security ARES'06*, Proc. intern. conf., Vienna, Austria.

Broder, J.F. & Tucker, E. 2012. *Risk Analysis and the Security Survey, 4th ed*. Waltham: Butterworth-Heinemann.

Butler, S.A. 2002. Security Attribute Evaluation Method: A Cost-Benefit Approach. In: *24th International Conference on Software Engineering ICSE 2002*, Proc. intern. conf., Orlando, USA.

Campbell, P.L. & Stamp J.E. 2004. *A Classification Scheme for Risk Assessment Methods.* Albuquerque: Sandia National Laboratories.

Contini, S. & Cojazzi, G.G.M. & Renda, G. 2008. On the use of non-coherent fault trees in safety and security studies. *Reliability Engineering and System Safety* 93 (12): 1886–1895.

Contini, S. & Fabbri, L. & Matuzas, V. & Cojazzi, G. 2012. Protection of Multiple Assets to Intentional Attacks. A Methodological Framework. In: *11th Probalistic Safety Assessment 2012*, Proc. intern. conf., Helsinki.

Cox Jr., L.A. 2009. *Risk Analysis of Complex and Uncertain Systems.* New York: Springer.

Fakhravar, D. & Cozzani, V. & Khakzad, N. & Reniers, G. 2017. Security vulnerability assessment of gas pipelines using Bayesian network. In: *27th European Safety and Reliability Conference ESREL 2017*, Proc. intern. conf., Portoroz, Slovenia.

Flammini, F. & Gaglione, A. & Mazzocca, N. & Pragliola, C. 2009. Quantitative security risk assessment and management for railway transportation infrastructures. In: *Critical Information Infrastructure Security.* Berlin: Springer.

French, G.S. & Gootzit, D. 2011. Defining and Assessing Vulnerability of Infrastructure to Terrorist Attack. In: *Vulnerability, Unvertainty and Risk: Analysis, Modeling and Management*, Proc. conf., Hyattsville, USA.

Garcia, M.L. 2008. *The Design and Evaluation of Physical Protection Systems. 2nd ed.* Burlington: Butterworth-Heinemann.

Giannopoulos, G. & Filippini, R. & Schimmer, M. 2012. Risk assessment methodologies for Critical Infrastructure Protection. Part I: A state of the art. In: JRC Technical Notes. Luxembourg: Publications Office of the European Union.

Gribaudo, M. & Lacono, M. & Marrone S. 2015. Exploiting Bayesian Networks for the Analysis of Combined Attack Trees. *Electronic Notes in Theoretical Computer Science*, 310: 91–111.

Harnser Group (Ed.) 2010. A Reference Security Management Plan for Energy Infrastructure. Brussels: European Commission.

Howard, R. 1988. Decision Analysis: Practoce and Promise. *Management Science* 34 (6): 679–695.

Howard, R. & Matheson, J. 2005. Influence Diagrams. *Decision Analysis* 2 (3): 127–143.

Jensen, F. & Nielsen, T. 2007. *Bayesian Networks and Decision Graphs. 2nd Ed.* Springer: Berlin.

Landoll, D.J. 2011. *The Security Risk Assessment Handbook: A Complete Guide for Performing Security Risk Assessments, 2nd ed.* Boca Raton: CRC Press.

Lichte, D. & Marchlewitz, S. & Wolf, K.-D. 2016. A Quantitative Approach to Vulnerability Assessment of Critical Infrastructures with Respect to Multiple Physical Attack Scenarios. In: *Future Security 2016*, Proc. intern. conf., Berlin, Germany.

Lichte, D. & Wolf, K.-D. 2017. Quantitative Multiple-Scenario Vulnerability Assessment Applied to a Civil Airport Infrastructure. In: *27th European Safety and Reliability Conference ESREL 2017*, Proc. intern. conf., Portoroz, Slovenia.

McGill, W.L. & Ayyub, B.M. & Kaminskiy, M. 2007. Risk Analysis for Critical Asset Protection. *Risk Analysis*, 27 (5), 1265–1281.

Meritt, J.W. 2008. A Method for Quantitative Risk Analysis. In: *22nd National Information Systems Security Conference*, Proc. nat. conf., Arlington, USA.

Morgeson, J.D. & Brooks, P.S. & Disraelly, D.S. & Erb, J.L. & Neiman, M.L. & Picard, W.C. 2011: Doctrinal Guidelines for Quantitative Vulnerability Assessments of Infrastructure-Related Risks. Volume I. Alexandria: Institute for Defense Analyses.

Nicol, D.M. & Sanders, W.H. & Trivedi, K.S. 2004. Model-based evaluation: from dependability to security. *IEEE Transactions on Dependable and Secure Computing* 1 (1): 48–65.

Schneier, B. 1999. Attack Trees. In: *Dr. Dobbs Journal* 24 (12): 21–29.

Shachter, R. 1986. Evaluation Influence Diagrams. Operations Research 34 (6): 871–882.

Solano, E.: Methods for Assessing Vulnerability of Critical Infrastructure, Institute for Homeland Security Solutions.

Sonnenreich, W. & Abanese, J. & Stout, B. 2005. Return on Security Investment (ROSI): A practical quantitative model. 3rd International Workshop on Security in Information Systems WOSIS 2005, Proc. intern. conf., Miami Beach, USA.

Stewart, M. & Mueller, J. 2013. Terrorism Risks and Cost-Benefit Analysis of Aviation Security. *Risk Analysis* 33 (5): 893–908.

Vintr, Z. & Valis, D. & Malach, J. 2012. Attack tree-based evaluation of physical protection systems vulnerability. In: *2012 IEEE International Carnahan Conference on Security Technology (ICCST)*, Proc. intern. conf., Carnahan, USA.

White, R. & Boult, T. & Chow, E.: A Computational Asset Vulnerability Model for the Strategic Protection of the Critical Infrastructure. In: *International Journal of Critical Infrastructure Protection* 7 (3): 167–177.

Wyss, G D. & Clem, J. & Darby, J. & Dunphy-Guzman, K. & Hinton, J., & Mitchiner, K. 2010. Risk-Based Cost-Benefit Analysis for Security Assessment Problems. In: *2010 IEEE International Carnahan Conference on Security Technology* (ICCST), Proc. intern. conf., San Jose, USA.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Risk prediction method of aircraft hard landing based on flight data

Liping Zheng, Jinsong Xie & Silin Qian
School of Reliability and Systems Engineering, Beihang University, Beijing, China

ABSTRACT: The paper aims to develop the risk prediction model for aircraft hard landing with flight data analysis. The statistic data shows nearly half of the incidents occurred in the aircraft landing stage and hard landing is one of the main contributions. The paper firstly analyzes the possible factors of the hard landing and the data feature of flight data. Secondly, flight data are preprocessed by height slice and Principal Component Analysis (PCA). It is to solve the prediction accuracy and data redundancy problems. Then the study builds the mathematical model with the objective function of maximize the probability of hard landing accidents through historical samples. An algorithm based on golden section was provided, and threshold values of each index were found. Finally, the proposed method is validated by empirical research. The result suggests that the proposed method is feasible in hard landing risk prediction problem.

1 INTRODUCTION

Previous aviation safety management data show that although the average time of landing phase accounts for only 1% of the total flight time, the accident rate at this stage was the highest of all phases of the flight. Therefore, the flight risk control in the landing stage plays a very important role in the flight safety assurance of the aircraft. Moreover, analysis of aircraft hard landing event is a very important work in practice.

Hard landing, also known as hard landing, is an extremely important safety hazard for the impact of flight safety during the landing stage. Hard landing may cause damage to the aircraft structure, result in direct or indirect financial loss, damage to comfort and other adverse consequences. Hard landing can cause damage to aircraft components or systems under heavy loads (eg, landing gear, wings, etc.) and, in severe cases, damage to the aircraft and casualties. Boeing points out that the acceleration of the plane's vertical to the ground exceeds the specified limit when it touchdown, and it can be judged to be a hard landing. The Airbus defines hard landing as a phenomenon that the acceleration or speed of an aircraft vertical to the ground exceeds the specified threshold. Thus, the landing load (that is, the vertical acceleration when the aircraft landed) is to determine the landing of the aircraft or not the key indicator. Accurate landing loads are predicted prior to the aircraft landing can identify the risk of hard landing events in time, and take appropriate measures in time (such as go around), which can reduce the frequency of hard landing events to a certain extent, and improve the safety of aircraft landing.

To ensure flight safety, real-time monitoring of aircraft flight status is required. The aircraft usually contains a Quick Access Record (QAR) for recording flight parameters. The flight parameters reflect real-time status information of the entire flight phase of the aircraft and have high application value in performance testing, accident investigation, flight training and assessment, equipment maintenance and safety monitoring. However, due to the complexity of the flight data and the limitations of the data analysis methods, there is still a more in-depth data development value due to the low utilization of flight data in flight safety early warning.

This article aims to establish a model based on the flight parameters to predict the risk of a landing. In a certain environment, taking the hard landing event as an object, we identify the high-risk areas that can easily trigger QAR overrun events in the "space" formed by the key factors closely related to such overrun events. The remaining part of this paper is organized as follows. In section 2, the process of data preprocessing by flight data slicing and data reduction analysis is introduced. In section 3, the risk region optimization model and determination of region division point is presented. In section 4, the model is demonstrated with real flight data.

2 DATA PREPROCESSING

2.1 *Flight data slicing*

The original flight data are time series data, and each flight parameter varies with flight time. However, it was found in the study that there was great uncertainty in predicting the landing time before the landing of the aircraft, and the prediction deviations could be as high as several minutes.

In fact, the normal landing time of aircraft is only a few seconds. In the prediction deviations of several minutes' landing time, the aircraft is likely to have completed the landing or far from completing the landing. It is highly unreliable to predict the hard landing according to the time trend. On the other hand, the height of the plane from the ground can be measured in real time by radio signals and is truly reflected in the flight data. Therefore, in order to avoid the inconsistency in the flight height data interval of each flight frame, we need to carry out the high slicing processing of flight data before establishing the risk prediction model of hard landing.

Slicing processing of flight data is a method to intercept a part of the flight data according to the altitude of the flight, the intercepted data will be used as the basis for data analysis, and the rest of the data will be removed, we set the altitude of the flight data in the range of $[h_1, h_m]$ ($h_1 < h_m$), $\boldsymbol{N}$ is the number of slices, the condition of the model is:

$$\boldsymbol{\omega} = \boldsymbol{h} = (h_1, h_2, \ldots h_N)$$

Satisfy the following formula:

$$h_i = h_1 + (i-1) \times d \tag{1}$$

$$d = \frac{h_m - h_1}{N-1} (N \geq 2) \tag{2}$$

where d stands for high interval of slice, $1 \leq i \leq$ N.

2.2 *Dimension reduction analysis*

There are many flight parameter variables in flight data. Too many variables can cause serious correlation. On the one hand, it will cause redundancy of models and reduce the efficiency of model operation. On the other hand, useless information will also cause bias in model prediction. As a method of statistical analysis, factor analysis had a good performance in data dimension reduction, the basic principle is to integrate multiple variables into a few indicators under the premise of losing less original information, so as to study all aspects of the information. After synthesizing a few indicators, the information contained are not repeated each other, i.e. variables are not related. For that reason, this paper is proposed to use factor analysis to reduce the dimension of flight data in order to extract key and valuable data from a large number of flight parameter variables.

The parameter analysis of flight parameters must first standardize the original flight parameters, and set up the k flight data with standardized treatment.

$X = (X_1, X_2 \ldots X_m)$ is a variable, m common factor of flight data is $F = (F_1, F_2 \ldots F_m)$, and $m < k$. The model is as follows:

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \ldots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \ldots + a_{2m}F_m + \varepsilon_2 \\ \ldots \\ X_k = a_{k1}F_1 + a_{k2}F_2 + \ldots + a_{km}F_m + \varepsilon_k \end{cases} \tag{3}$$

where a_{ij} is factor loading, ε_1, ε_2, …, ε_k is residual term.

Common factor of flight data in the model, "Flight data factor" for short, that is, a few key flight data information obtained after dimensionality reduction of the original flight data. This information will serve as the modeling data basis for the risk prediction model of hard landing.

3 MODEL AND ALGORITHM DESIGN

Hard landing is an important hidden danger for aircraft landing safety. There are many factors that affect the landing safety of aircraft. It can be divided into three main categories: human factors, body factors and meteorological factors. Human factors include unskilled operation, misjudgments, and operation errors caused by psychological factors, and so on. Body factors include the maintenance support of the aircraft, the level of reliability, and so on. Meteorological factors include the reduction of visibility caused by rain and snow and the deviation of the flight attitude caused by the side wind. The impact of these three kinds of factors on aircraft landing safety can be concentrated in the flight data. A large number of flight data are recorded in flight data. These flight data reflect the flight status and index parameters in real time, and provide important monitoring basis for aircraft flight safety. In this paper, we take the landing load overrun event in the landing stage as the object, and find out the high-risk area which is easy to trigger hard landing events in the "space" closely related to the event. The so-called high-risk area refers to the combination of certain factors in the span range of influencing factors for hard landing events. In the subspace represented by these combinations, the probability of a landing load over the limit event tends to be 1.

3.1 *Risk region optimization model*

After reducing the dimension of flight data, at the height of 9 *m*-2 *m* before landing, the value of the m flight data factor and Its Variation Track play a major role in the occurrence of hard landing events. Therefore, the high risk areas are divided according to the space of m flight data factor and its change rate. In the process of division, the total number of samples to meet the space is not less than a certain proportion. That is, the nature of the space is established in a certain probability. Then, the partition point is optimized and the location of the partition point is adjusted to maximize the probability of the hard landing event in the high-risk area.

The symbols are as follows:

A, B, C indicate the state value of 3 factors at the radio height 9m; ΔA, ΔB, ΔC respectively indicate the change rate of the flight data factor *A, B, C* from the radio height of 9 to 2m. *N* represents the number of all the samples; $0 < P < 1$, indicating that the sample size of the region should be kept at a certain level after the division; L^* represents the "high-risk area" threshold; *L* indicates that the number of overrun samples of triggered landing load in a certain region accounts for the proportion of the total number of samples in the region; *R* represents "high risk areas", when and only when the proportion of the flight overrun samples in the region is higher than that of L^* in the region. R_1 represents the number of flight overrun samples in "high risk areas"; R_2 represents the number of normal samples in the "high risk area"; H_1^i, H_2^i respectively represents the division point on the *i* factor, which is a decision variable.

Find the optimal model for high-risk area:

$$\max(L(H_1^i, H_2^i)) \ldots \mathrm{i} = A, B, C, \Delta A, \Delta B, \Delta C \quad s.t. \tag{4}$$

$$(R_1 + R_2)/N \geq P \tag{5}$$

$$R_1/(R_1 + R_2) \geq L(H_1^i, H_2^i) \tag{6}$$

H_1^i, $H_2^i \in$ [Minimum value of *i* factor and maximum value of *i* factor] (7)

Among them, formula (4) is the objective function, which indicates that the threshold of "high-risk area" is maximized under the value of different division points. Formula (5) indicates that the proportion of samples in the division area is not less than that of a given parameter *P*. Formula (6) indicates that the region should satisfy the definition of high-risk area; formula (7) represents the range of decision variables.

3.2 *Search of partition points*

The search for the division points adopts the golden section method with one dimension optimization. The golden section method also known as extreme and mean ratio, refers to a line segment is divided into two parts; the part with the length ratio is equal to the other part and this part of the ratio. The ratio is an irrational number, and the approximate value of the first three digits is 0.618, so it is also called the 0.618 method.

The order of Iterative Refinement will affect the result, therefore, combining the experience of the expert, the iteration sequence is ordered according to the size of the impact on the event, and the algorithm step is shown in Figure 1.

Sort $i = A, \Delta A, B, \Delta B, C, \Delta C$

$$H_2^i = i^- + 0.618(i^+ - i^-) \tag{8}$$

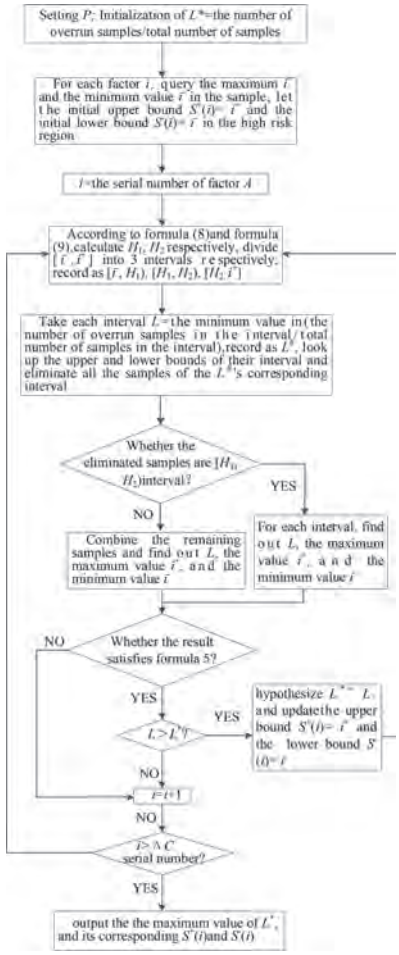


Figure 1. The diagram of algorithm step.

The point of this point on median symmetry is

$$H_1^i = i^- + 0.382(i^+ - i^-) \quad (9)$$

4 EXPERIMENT ANALYSIS

This research takes a certain type of UAV as the test object. After more than 10 years' research and test time, the UAV has completed dozens of scientific research flight tests and batch production. At present, dozens of Unmanned Aerial Vehicles (UAVs) have been delivered to many batches of the troops, with a cumulative flight time of more than 200 hours. The paper obtains the flight data of the aircraft during the landing and decline stage including 19 flight data variables in 45 sorties. 19 variables of flight data are respectively radio altitude, elevation angle, pitch rate, roll angle, roll angle rate, course angle, yaw angular velocity, aileron displacement, rudder displacement, altitude rate, elevator displacement, forward acceleration, normal acceleration, lateral acceleration, engine speed, atmospheric height, airspeed, the lateral offset and ground speed. The critical value of landing load to judge whether a hard landing is or not is 18.0 m/s^2, When the aircraft landing load is more than 18.0 m/s^2, the landing is considered to be a hard landing. During the 45 landing sorties, a hard landing occurred during the landing of fourth, fifth, sixth, 10, 14, 16, 17, 20, 24, 25, 26, 29, 30, 34, 35, 39, 40, 42, 45, respectively. Removing the flight data after landing, the unified limited flight height is 9 *m*-2 *m*, and the flight data of each landing gear are sliced at each 0.5 m by the flight height, the missing data were filled by averages, and 15 data slices were obtained and the slice data were integrated. A total of 675 data of 15 altitude values of 45 flight sorties were finally obtained, as shown in Table 1.

18 input variables of flight data were tested with KMO test. The results were 0.579 and more than 0.5, which showed that the variable of flight data was suitable for factor analysis. The results of the dimensionality reduction analysis of the parameters of the flight data are shown in Table 2.

If the extracted eigenvalue is greater than 1, the eigenvalue is considered as a flight data factor, the analysis results show that, when 18 fight data variables are reduced to 3 flight data factors, 87.278% of the raw information of the flight data can still be retained. According to the order of the number 1–3 in the table, the three flight data factor variables are defined as *A, B, C,* respectively, flight data factor after dimensionality reduction makes the model simplified, in this way, the original information of the flight data is preserved as much as possible, and the expected effect is achieved.

Table 2. The results of the dimensionality reduction analysis of the parameters of the flight data.

	Initial eigenvalue			Extraction of square sum load		
Component*	Total	Variance ratio	Cumulative ratio	Total	Variance ratio	Cumulative ratio
1	5.194	57.711	57.710	5.194	57.711	57.710
2	1.646	18.293	76.004	1.646	18.293	76.004
3	1.015	11.274	87.278	1.015	11.274	87.278
4	0.441	4.901	92.180			
5	0.348	3.863	96.042			
6	0.256	2.848	98.890			
7	0.070	0.782	99.672			
8	0.026	0.258	99.957			
9	0.004	0.043	100			

*The method of extraction is principal component analysis.

Table 1. Data slicing processing results of flight data.

Landing sortie	Flight altitude	Pitch angle	Roll angle	…	Ground speed	…	Landing load	Engine speed
1	9	–0.95859	2.90957	…	34.13035	…	27.76744	6544
1	8.5	–0.97507	1.49968	…	34.17583	…	27.54345	6544
…	…	…	…	…	…	…	…	…
1	2	2.48848	–2.87851	…	31.07449	…	10.53457	6211
2	9	–1.16459	0.79653	…	33.85126	…	31.67865	5433
2	8.5	–1.05747	0.7416	…	33.86937	…	31.12323	5431
…	…	…	…	…	…	…	…	…
2	2	2.74392	–0.13733	…	30.37497	…	13.76453	4322
…	…	…	…	…	…	…	…	…
45	2	2.16987	4.05408	…	27.26024	…	21.45645	4854

Table 3. High-risk area*.

Parameter	A	ΔA	B	ΔB	C	ΔC
Upper boundary	0.62	–0.05	0.85	0.30	0.74	–0.64
Lower boundary	0.19	–0.41	0.92	–0.70	0.34	–0.23

*$P = 0.3$.

Using the algorithm in the third section to solve the model, the result is $L^* = 0.6$, the results of the high risk area after standardization are shown as shown in Table 3, and the standardization process is as follows:

$$\overline{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

where, x_i represents the flight data.

We randomly selected 10 samples for testing and analysis, and 4 samples were found to fall into the area, of which three were QAR landing load overrun samples.

According to the calculation results, the average probability of the occurrence about landing load overrun in the analysis sample is 0.412. From the analysis results, the possibility of the occurrence about landing load overrun in the area is higher than the average occurrence probability, which indicates that the area marked in Table 3 belongs to the high-risk area of the hard landing event.

5 CONCLUSION

This paper establishes the risk prediction model for aircraft hard landing by determining the high risk area with flight data analysis. Through the high slicing processing of the original flight data, the unified flight data can be obtained at a certain height. In addition, factor analysis, as a tool for data reduction processing, can simplify more parameters and maximize the information of data. We selected flight data (including human, aircraft and environmental factors) in 9-2 *m* to predict the risk of a hard landing event. In the sense of certain probability, the division point is optimized. By adjusting the location of the division points, the probability of a hard landing event in a high-risk area is maximized. The results show that the identified high risk area can identify the exceeding limit of landing load. For future research, we will pay more attention to how to improve the accuracy of recognition. At the same time, further demonstration is needed for the selection of the predicted height.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (Grant No.71501007 & 71672006). The study is also sponsored by the Technical Research Foundation and The Graduate Student Education & Development Foundation of Beihang University.

REFERENCES

Azar, A.T. & Hassanien, A.E. 2015. Dimensionality reduction of medical big data using neural-fuzzy classifier. *Soft Computing,* 19, 1115–1127.

Brinton, M.A. 2001. Host Factors Involved in West Nile Virus Replication. *Annals of the New York Academy of Sciences,* 951, 207–219.

Cao, H., Shu, P. & Huang, S. 2008. Study of Aircraft Hard Landing Diagnosis Based on Nerual Network. *Computer Measurement & Control,* 16, 906–908.

Chen, J. 2003. About 0.618 and a Kind of the Application. *Electro-optics & Passive Countermeasures*.

Chiesa, S., Aleina, S.C., Meo, G.A.D., Fusaro, R. & Viola, N. 2014. Autonomous take-off and landing for unmanned aircraft system: Risk and safety analysis.

Chong, J.F., Bo-Chun, L.I. & Zhuo, Z.P. 2007. Comparing on Function of Moral Education in Mathematics Teaching Outline and Mathematics Curriculum Standard. *Journal of Huaibei Coal Industry Teachers College*.

Dai, Y., Tian, J., Rong, H. & Zhao, T. 2015. Hybrid safety analysis method based on SVM and RST: An application to carrier landing of aircraft. *Safety Science,* 80, 56–65.

Feng, Y., Zhu, Z., Yao, X., Ma, W., Xue, X. & Aeronautics, S.O. 2016. An Effective Safety Analysis Method of Civil Aircraft Landing Gear. *Journal of Northwestern Polytechnical University*.

Lang, G.P. 2009. Analysis of Accident Resulted from Human Factors in Safety Management. *Journal of Civil Aviation University of China*.

Luo, F., Ping, Y., Guo, S.P. & Liu, H. 2002. Civil aviation flight character surveillance and forewarning management. *Journal of Traffic & Transportation Engineering*.

Qi, M., Shao, X. & Chi, H. 2011. Flight operations risk diagnosis method on quick-access-record exceedance. *Journal of Beijing University of Aeronautics & Astronautics,* 37, 1207–1210.

Song, B.X. 2001. Application of the system of Hazard Analysis of Critical Control Points in hygienic quality surveillance of lunch box for students. *Shanghai Journal of Preventive Medicine*.

Taher, A. & Hassanien, A.E. 2014. Dimensionality reduction of medical big data using neural-fuzzy classifier. *Soft Computing*.

Wang, J.L. 2015. A New Exploration for the Language Problems of Mathematical Research in Book of Changes. *Journal of Shanxi University*.

Wang, L., Wu, C. & Sun, R. 2013. Pilot operating characteristics analysis of long landing based on flight QAR data. 8020, 157–166.

Wang, L., Wu, C. & Sun, R. 2014. An analysis of flight Quick Access Recorder (QAR) data and its applications

in preventing landing incidents. *Reliability Engineering & System Safety,* 127, 86–96.
Wang, Y., Zhang, R., Shen, S. & Wu, Z. A Modeling Technology of Aircraft Landing Safety Prediction under the Extreme Weather Conditions in the Future Based on Cloud Theory. Meeting of Risk Analysis Council of China Association for Disaster Prevention, 2016.
Xie, L., Chen, Y. & Kumar, P.R. 2014. Dimensionality Reduction of Synchrophasor Data for Early Event Detection: Linearized Analysis. *IEEE Transactions on Power Systems,* 29, 2784–2794.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

EU risk governance of migrants and refugees' influxes: A realistic foundation for crisis governance?

B.I. Kruke & C. Morsut
Centre for Risk Management and Societal Safety, University of Stavanger, Stavanger, Norway

ABSTRACT: Wars, political instability, poverty and ecosystem's alterations force several people to look for better living conditions and security. Europe has become a safe haven for migrants and refugees, particularly since 2015, when thousands of people crossed the European borders on daily basis. In this paper, we aim at studying the development and the management of the 2015 migrant and refugee influx into Europe at the European Union (EU) level in terms of risk and crisis governance, mainly through the lens of the International Risk Governance Council (IRGC) Risk Governance Framework. The 2015 mass influx into Europe showed the EU's inability to cope with such an event, with a subsequent fragmented response consisting of mainly national security initiatives. A main reason behind the inadequate overall joint crisis governance at EU level has been a weak supranational risk governance, mainly due to national political, economic, security and cultural differences.

1 INTRODUCTION

Migration is an old phenomenon. People have always been on the move. Push and pull factors will endure and people will continue to leave their home because they have to or because they want. Wars, permanent conflict, political instability, ecosystems' negative alteration due to the climate change are all push factors that force people to leave and to seek refuge or better life conditions. This kind of migration has given rise to a series of challenges to Europe: the so-called 2015 migrant and refugee crisis, characterised by a high influx of people crossing the Mediterranean Sea, is an example in this sense. In 2015, more than a million people reached Europe (UNHCR 2017) in perilous ways, putting under pressure the aid and reception mechanisms existing at national and European Union (EU) levels. Rescuing people from the sea, giving them first assistance, providing shelters and distributing food were activities performed be a variety of non-governmental, governmental and supragovernmental actors in a complex frame. National authorities, the EU, the United Nations High Commissioner for Refugees (UNHCR), the International and national Red Cross, several NGOS, but also private citizens offered their assistance in 2015. They continue these tasks since the so-called crisis has become more a structural occurrence rather than a one-off event.

A short terminological clarification of the terms migrant and refugee is needed at this point, before proceeding with the goal of our paper. The 1951 Refugee Convention defines a refugee a person crossing a national border seeking protection from political or other forms of persecution (UNHCR 2010), while a migrant is a person who chooses to move mainly to improve his/her living conditions (a better job, education or family reunion are reasons behind the choice) (UNHCR 2016). Unlike refugees, who cannot safely return home, migrants face no such impediment. However, in describing the 2015 massive influx of people into Europe, the UNHCR employed both terms, since it was difficult to differentiate between the two groups, those escaping from various political instabilities and those impelled by economic reasons (economic migrants). We will follow the UNHCR's terminology in this paper.

We aim at studying the development and management of the 2015 migrant and refugee crisis at the EU level in terms of risk and crisis governance, mainly through the lens of the International Risk Governance Council (IRGC) Risk Governance Framework. Our analysis rests upon previous research we conducted on resilient crisis management (Morsut and Kruke 2014, Kruke and Morsut 2015), on reliable crisis governance (Morsut and Kruke 2016), and on the relationship between risk and crisis governance (Morsut and Kruke 2017). This paper first presents the conceptual framework applied in our case. Secondly, by drawing from document analysis of EU policy and legal documents, it outlines the EU risk and crisis governance towards migrants and refugees according to the IRGC Risk Governance Framework, to answer the

following question: to what extent did the EU crisis governance follow a thorough risk governance process in coping with the influx of refugees and migrants?

2 CONCEPTUAL FRAMEWORK

2.1 *Risk*

Risk is related to a possible future state of affairs. Risk may be defined as "an uncertain consequence of an event or an activity with respect to something that humans value" (IRGC 2005: 19), or a situation or event where something of human value (including humans themselves) is at stake and where the outcome is uncertain (Rosa 1998, 2003). The uncertain consequences—understood in terms of likelihood and severity—can be positive or negative. The degree of 'positiveness' or 'negativeness' depends very much on peoples' perceptions. However, most people relate risk to something negative. Kates et al. (1985: 21) describe risk as the possibility that an undesired state of reality (adverse effects) may occur as a result of natural events or human activities. This implies that there is a possibility for a negative result of a natural phenomenon or human activity. However, the impact of this phenomenon or activity is unknown, as the probability for the occurrence. Uncertainty characterises the phenomenon or activity in question. Uncertainty, together with values, is central to Aven and Renn's understanding of risk, which refers to uncertainty about and severity of the consequences (or outcomes) of an activity with respect to something that humans value (Aven and Renn 2009: 1). Thus, risk is inherently a subjective phenomenon, a social construction. Expert judgements is necessary, but not adequate, to understand and manage risk. In addition, risk perception, public values and stakeholders' understanding may also be important to consider. Thus, differences in approaching the risk between the public and the so-called experts are likely to be seen (Sjöberg 1999).

Risk perception, in particular, has its roots in cognitive psychology and has given rise to a vast literature on the relationship between the perceived risk and the real foundation of risk (Slovic 2000). Culture, gender and knowledge availability influence the perceived risk and the actions taken accordingly, both by individuals and by decision makers. Risk perception, as a social and cultural construct, reflects values, symbols, history and ideology (Weinstein 1980). Differences in risk perception between the public and the so-called experts may hamper risk management approaches and rational decision-making. This conflict between expert and public risk perception is at the basis of the social dilemmas of risk management (ibidem) and of risk governance.

2.2 *Risk and crisis governance*

Governance lacks a univocal meaning since it rests in several disciplines that discuss the governance from their standpoint (see Kjær 2004; Peters 2000; Pierre 2000; Stoker 1998). Thus, our understanding of governance differs according to who exerts governance. Governance is often seen in relation to government and governability (Renn et al. 2011). Government may be understood as setting and administering the public policy, while governability is understood as the overall capacity for governance of any societal entity or system (Kooiman et al. 2008). Some scholars consider the state as the main actor exerting governance (Bevir et al. 2003). In this understanding, governance is closely related to government. Governance, on the other hand, is understood in terms of socio-political interaction patterns, where management and decision-making are conducted within a framework of institutional diversity (Morsut and Kruke 2017). This diversity is formed by several stakeholders—state authorities, trade associations, NGOs, civil society, private actors (Kooiman 2003; Krahmann 2003; Bogason 1996), networks (Sørensen and Torfing 2007) and supranational organisations, such as the EU (Marcussen and Torfing 2007). Governance, therefore, entails stakeholders' involvement (Renn 2008) to a much greater degree than government.

Scholars define and characterise governance at various levels. At a national level, Nye and Donahue (2000) describe governance as structures and processes for collective decision-making involving governmental and non-governmental actors. The joint approaches between public and private actors are prominent in this understanding of governance. The same is the case with Rosenau's understanding of governance at a global level (1992). Here, governance embodies a horizontally organized structure of functional self-regulation, encompassing state and non-state actors, bringing about collectively binding decisions without superior authority (ibidem).

The term risk governance involves the translation of the substance and core principles of governance to the context of risk-related decision-making (van Asselt and Renn 2011). A successful risk governance approach leads to a so-called dynamic non-event (Weick 2011). However, if the non-event becomes an event, in the form of a crisis, this crisis needs to be managed by a crisis governance, defined as "to what extent the relationships among economic and political, formal and infor-

mal institutions are able to manage crises" (Kruke and Morsut 2015: 187). In this respect, there is a clear relation between the quality of risk governance and an effective crisis governance.

2.3 *The IRGC risk governance framework*

The IRGC has elaborated a Risk Governance Framework for a systemic risk governance approach. The Framework goes beyond a naïve understanding of risk as an objective category and a relativistic perspective where all risk judgements are subjective reflections of power and interests (Renn 2008). The Framework is characterised by two interlinked spheres: the assessment and the management sphere.

The former deals with the generation of knowledge, whereas decisions and implementation of actions are conducted in the latter. Each sphere is divided into phases (IRGC 2005) briefly described here:

Pre-assessment:
The purpose is to capture both the variety of issues that stakeholders and society may associate with a certain risk as well as existing indicators, routines, and conventions that may prematurely narrow down, or act as a filter for, what is going to be addressed as risk. Typical activities in the pre-assessment are:

- Problem Framing: different perspectives of how to conceptualize the issue;
- Early warning: systematic search for new hazards;
- Screening: establishment of procedures for screening hazards and risks;
- Scientific conventions for risk assessment and concern assessment: assumptions and parameters of scientific modelling and evaluating methods and procedures for assessing risks and concerns.

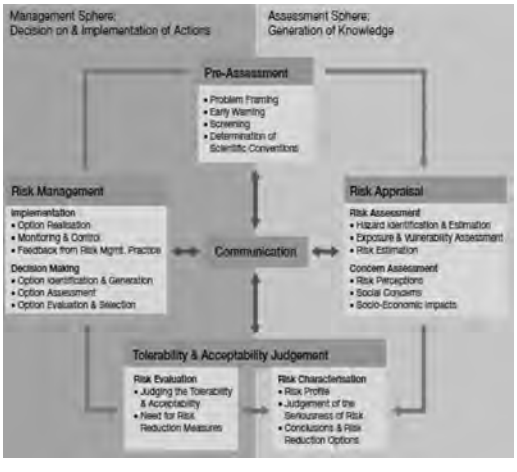


Figure 1. IRGC framework (IRGC 2005: 13).

Risk Appraisal:
Main activities in risk appraisal are the development and the synthesis of the knowledge base as a foundation for a decision on whether or not a risk should be taken. If the decision is to take the risk, then a follow-up activity is to map available options for avoiding, mitigating, reducing or handling the risk. Risk appraisal comprises both a scientific risk assessment and a concern assessment. The scientific risk assessment deals with the risk's factual, physical and measurable characteristics, including the probability of occurrence (or a probability distribution over a range of negative consequences) (IRGC 2005: 14). The concern assessment is a systematic analysis of the associations and perceived consequences (benefits and risks) that stakeholders, individuals, groups or different cultures may associate with a hazard or a cause of hazard (ibidem).

Tolerability and acceptability judgement:

- Risk characterisation: Collecting and summarizing all relevant evidence necessary for making informed choice of tolerability and acceptability of the risk in question and suggesting potential options for dealing with the risk from a scientific perspective.
- Risk evaluation: Applying societal values and norms to judge tolerability and acceptability and, consequently, to determine the need for risk reduction measures.

Risk Management:

Decision-making:

- Option identification and generation: Identification of potential risk-handling options, particularly risk reduction (i.e. prevention, adaptation and mitigation, as well as risk avoidance, transfer and retention);
- Option assessment: Investigations of the impacts of each option (economic, technical, social, political and cultural);
- Option evaluation and selection: Evaluation of options (multi-criteria analysis).

Implementation:

- Realization of the most preferred option;
- Monitoring and feedback: Observation of the effects of implementation (link to early warning). Ex-post evaluation.

In all phases risk communication is essential to enlighten the risk process for all not involved relevant stakeholders, including civil society, and to maintain trust among these (IRGC 2005).

3 THE CASE: THE 2015 INFLUX OF MIGRANTS AND REFUGEES INTO EUROPE

The UNHCR started to collect data on migrants and refugees' Mediterranean Sea crossing since 2007.

As Figure 2 indicates, the statistics show a peak in 2015 with more than one million people. Despite of a decrease in 2016, Europe still faces a substantial flow of migrants and refugees across the Mediterranean. Syria, Afghanistan, and Iraq are the top three countries of escape due to a prolonged warfare (UNHCR 2016a). Challenges of reception and management of refugees and migrants remain the same, since permanent solutions are still not in place.

Migration (by force or by choice) is a global phenomenon as much as is an old one. Refugees and migrants moving to Europe are much less than the quantity of people moving to other regions of the world. The vast majority of migrants continues to be hosted by developing countries, particularly those that are proximate to the migrants and refugees' countries of origin: for instance, the bulk of the Syrian refugees is hosted by Turkey (2.2 million), Lebanon (1.2 million) and Jordan (almost 630,000), according to figures recorded in December 2015 (IOM 2017a). In the case of Europe, the events of 2015 were particularly challenging for the frontline EU member states like Italy and Greece. They received the highest number of refugees and migrants (respectively 153,842 and 856,723 - UNHCR 2017a). Transit countries in Central and Eastern Europe and major destination states, such as Austria, Belgium, Finland, Germany, Sweden and the Netherlands, experienced challenges in handling the flow of migrants and refugees, as well.

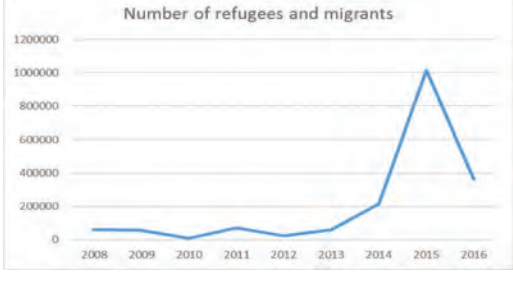


Figure 2. Overview of refugees and migrants crossing the Mediterranean Sea 2008–2016 (UNHCR 2017).

3.1 *The EU risk governance*

The EU's legal and operational asylum structure, the Common European Asylum System (CEAS), aims at harmonising the national asylum policies of the EU member states. The CEAS was decided in 1999, at the European Council of Tampere, and completed in 2005, with revisions between 2011 and 2013. The CEAS contains a series of legislative measures that cover all the phases of asylum seeking, from the entrance into the EU territory to the application process, from the rights to the duties an asylum seeker has once he/she is deemed qualified to stay. The rationale behind the CEAS is twofold: in the short term, the EU aims at achieving "common standards for fair and efficient asylum procedures in the Member States" (Council 2013: np). In the longer term, the EU wants "Union rules leading to a common asylum procedure in the Union" (ibidem).

The EU has only partially succeeded in these goals, since the reception of migrants and refugees remains largely a national affair. Despite of the rather positive experience from the refugees' flows following the Balkan wars in the 1990s, the 2015 influx showed a unilateral response among the member states, which exacerbated the distance between the EU's legal architecture and the national approaches to immigration.

The CEAS rests on two pillars: Directives and Regulations. The main Directives are the Asylum Procedures Directive (Council 2013), the Qualification Directive (Council 2011) and the Reception Conditions Directive (Council 2013a).

The Asylum Procedures Directive contains principles and criteria on how to apply for asylum in the EU territory. The member states have the main responsibility for the asylum procedure, while the EU mainly supports them through the European Asylum Support Office (EASO) (see below) and the European Refugee Fund. The Qualification Directive describes which kind of issues member states should take into account when processing an application (legislation from the country of origin, statement from the applicant and so on). In this case, as well, the member state has the primary task to deal with the application by verifying the validity of the information provided. The Reception Condition Directive specifies a minimum and common standard of access to housing, food, clothing, health care, education and employment.

Two Regulations (the EURODAC Regulation and the Dublin Regulation) complete the architecture. The former (Regulation 2013) established the EU asylum fingerprint database in 2003 and supports the latter (Regulation 2013a), which sets the

criteria for the examination of an asylum application. The Dublin Regulation aims at preventing the so-called asylum shopping (an applicant looks for the member state with the most indulgent asylum procedures) and the indiscriminate circulation of people that move inside the EU territory (the Schengen area) while waiting for the conclusion of the application process. These two criteria foresee that an asylum seeker should apply in the first country of entrance. In 2015, the high influx of migrants and refugees put under pressure two countries of entrance, namely Italy and Greece. The combination of high numbers and poor reception capacities showed that the Dublin Regulation could not work in such conditions.

Two agencies assist the member states in the implementation of the Directives and Regulations' legal measures: (1) the European Asylum Support Office (EASO 2017), created in 2011; (2) FRONTEX, which, since 2004, is the European Border and Coast Guard responsible for border monitoring and management (FRONTEX 2017). At the end of 2014, FRONTEX started Operations Triton in the Central Mediterranean and Poseidon in the Greek sea to control the EU maritime borders and to conduct search and rescue operations for people crossing the Mediterranean Sea (EEAS 2017).

It is worth mentioning the 2001 Directive on temporary protection (Council 2001), which is not included in the CEAS. This Directive foresees a temporary protection status in all EU countries, promoting a balance of efforts between member states receiving migrants and refugees, following a Council Decision that confirms a mass influx of displaced people and states the groups in need for protection.

3.2 *The EU crisis governance*

The EU crisis governance of the 2015 influx of migrants and refugees did not develop according to the Common European Asylum System. The EU response was undermined by national fragmented responses consisting of different emergency measures such as the enforcement of border controls, containment of the number of people crossing national borders, also with the use of force, construction of fences, and the rejection by force or the arrest of people who entered the national territory illegally (Morsut and Kruke 2017). These measures were taken unilaterally, with no consultation among states, which, for example, share the same border, causing tensions between neighbours. Both the European Commission and the Council launched a series of new initiatives (contained in three implementation packages—see Morsut and Kruke 2017) calling for unity and solidarity among the member states, seeking to coordinate the response at supranational level. The main attempt was to balance the burden and commitment between the frontline member states Italy and Greece and the rest of the EU member states.

The first package included a strengthening of operation Triton to handle search and rescue operations and combat smuggling in the Mediterranean Sea, and a temporary relocation scheme for asylum-seekers from Italy and Greece to the other Member States (European Commission 2015).

The second package included an extended emergency relocation proposal and a permanent crisis relocation mechanism to be activated when the Commission determined that a national asylum system was under pressure due to a large and disproportionate influx of third-country nationals (European Commission 2015a).

The third package, highly influenced by the terrorist attacks in Paris in November 2015, contained new measures regarding the protection of EU external borders (Morsut and Kruke 2017).

The relocation scheme in package two was officially concluded in September 2017, but less than a fifth of the original target was relocated (IOM 2017). The relocation mechanism recalls the temporary protection of the 2001 Directive (Council 2001), which was not used to cope with the influx of migrants and refugees in 2015.

In Italy and Greece migrants and refugees were taken to reception centres, to apply for asylum (as the Dublin Regulation foresees). However, both the Italian and the Greek reception systems were not able to absorb all the migrants and refugees giving them adequate shelter and the possibility to start the asylum process (UNHCR 2015). The EU intervened financially to implement hotspots in Greece and Italy, to identify, register and fingerprint migrants and refugees and to provide assistance. However, their implementation took time, leaving people literally left to themselves. In other parts of Europe, as well, assistance to migrants and refugees was limited: the camp called the Jungle on the outskirts of Calais was a striking example. In 2015, between 6,000 and 10,000 refugees, asylum seekers, and migrants, including many unaccompanied children, lived there under very poor conditions (Human Rights Watch 2017).

4 DISCUSSION

The application of the Risk Governance Framework to our case raises some interesting issues, especially related to the assessment sphere, but also in the management.

In the pre-assessment phase, framing and early warning "provide a structured definition of the problem and how it may be handled" (IRGC

2005a: 8). The EU legal and operational architecture, briefly described above, seems very promising on paper. It captures the main issues related to an influx of people into Europe and appears able to secure the ways to cope with this risk. It calls for a common and shared responsible immigration management; clearly defines the obligations of the member states and which kind of support they receive from the EU; provides a fair balance of efforts between member states; follows all the asylum seeking circle from the application to the granting of the protection, including the criteria to guarantee the person a full integration into the European society. The two maritime operations, Triton and Poseidon, seem relevant to decrease the risk of receiving a high number of migrants, but also for search and rescue of migrants.

However, there are signs of an inability to capture what stakeholders and society may associate with the risk (IRGC 2005) related to a massive influx of migrants and refugees. We argue that there may be a paradox between the Directives and the Dublin Regulation. The main goal of the Directives is to harmonise the national asylum systems according to the principles of solidarity and fair burden sharing. The Dublin Regulation aims at preventing asylum shopping and indiscriminate circulation of people inside the Schengen area, while waiting for the conclusion of the application process. Nonetheless, as for the asylum shopping, this is an admission that there are differences in the national asylum systems, which the Directives clearly have not solved. As for the indiscriminate circulation, the Dublin Regulation does not ensure a sustainable sharing of responsibility across the member states, which is exactly the opposite of the principles of the CEAS (solidarity and fair burden sharing) as such. The EU responded with new measures (the relocation mechanisms and the hotspot implementation in Italy and Greece) contained in the implementation packages. This showed that it was difficult to sustain the CEAS.

As for the FRONTEX operations Triton and Poseidon, their main contribution in 2015 (which continued in 2016 and 2017) became the rescue of migrants and refugees crossing the Mediterranean Sea with less focus on the control of the EU maritime external borders (FRONTEX 2017). This is a reactive approach and an indication of incapacity to frame the problem, in other words to see the pre-assessment, risk appraisal and risk management of the situation in Europe in relation to the root causes of migration. In this respect, the EU risk governance did not consider the wider political, security and economic instabilities making people leave their homes for Europe. Thus, monitoring and feedback according to the Framework (IRGC 2005) and, more specifically, the observing and feedback of the implementation of responses in 2015 need to address the reasons behind the EU reactive approaches and to develop proactive approaches and awareness of root causes, since the 2015 influx was not a one-off event.

In general, the CEAS degree of compliance in the events in 2015 was very poor, since the CEAS seems inadequate to account for a dynamic societal and political context, which is an important factor for framing and early warning, according to the IRGC. This context is made of member states, which reacted in different ways (the national fragmented responses) and undermined the principles of the CEAS, namely solidarity and fair burden sharing (the relocation mechanisms in the implementation packages). This shows the extent to which the CEAS is not implemented at national level and the member states retain their sovereignty on immigration policies. The views on the issue were clearly conflicting between the EU and the member states.

In addition, the unclear divide between people in need of protection and economic migrants was a challenge. The CEAS is designed to manage a tidy, small-scale and easy recognisable group of refugees, not the 2015 high number of people with different backgrounds crossing the Mediterranean Sea. Furthermore, the EU and its member states did not recognise and detect the risk of a massive influx in time. At supranational level, the EU crisis governance of the high influx was more based on an incremental approach rather than an implementation of the CEAS. The CEAS was the result of a risk governance approach based on an evaluation of the risk which had not taken into account certain features (high numbers, lack of cooperation, arrivals in few countries, poor receptions mechanisms and so on). At national level, the crisis governance seemed to be directed by pure nationalistic choices.

These reflections lead to the risk appraisal in our case. The EU did not consider the risk in all its characteristics and effects in the years prior to 2015. Migration is not a risk per se, but it became a risk due to inadequate risk appraisal, by not considering all issues related to migration according to a holistic view of the challenges. Concerns and perceptions at national level raised controversial issues in terms of solidarity, sustainability and capacities at EU level, which the EU was unable to address. For example, the reception facilities of Greece and Italy were not able to satisfy the needs of all the people coming. Unacceptable living conditions and slow bureaucratic national asylum processes may very easily lead people to become part of that substrate of the society, surviving through illegal expedients. This nourishes mistrust and fear towards immigrants establishing a vicious circle, where immigrants are increasingly isolated by the host

society and the host society intensifies its intolerance towards immigrants because they live at the margins of the same society. In addition, mistrust is directed also towards those segments of the society who assists the irregular migrants, also risking legal prosecution (Human Rights Watch 2017).

A scientific risk assessment, of factual, physical and measurable characteristics of the risk (IRGC 2005), using statistics at global level, could have scaled back societal negative perceptions, since developing countries are still the main recipient countries. A proper concern assessment of the perceived consequences (IRGC 2005) and a following risk communication strategy may have laid the foundation for a relationship of trust among the different stakeholders, in particular the EU and member states. The EU seemed incapable of perceiving the concerns of the various stakeholders and the public. This led to opposition and protest against the EU risk governance, through national measures of crisis governance.

Many stakeholders, and the public at large, are not formal parts of the process of addressing and handling risk. They, nevertheless, need to perform their own informed risk-related decision-making and choices about the risk in question, through a balance of factual knowledge based on risk perception and personal interests. Reliable risk communication may therefore bridge conflicting viewpoints and, as well, increase the likelihood of shared risk acceptance among different stakeholders: risk evaluators and managers, mostly involved in the risk process, researchers and policy makers, across academic disciplines and institutional barriers, and the people affected by the process (Renn 2008).

Terror was another factor not taken into consideration in the pre-assessment phase of the EU risk governance, nor was it a part of the concern assessment in the risk appraisal phase prior to 2015. The threat of terror attacks from IS fighters and European citizens returning to Europe after fighting for IS in Iraq and Syria influenced the last implementation package in 2015 and turned many European citizens against refugees coming from those areas. This threat was a major social mobilisation factor influencing national policies and approaches.

5 CONCLUSIONS

The IRGC Risk Governance Framework has been a useful tool for studying the features of a supranational system of risk governance as the CEAS and the way it was applied in our case.

We can conclude that the EU crisis governance of the 2015 influx did not adhere to the EU risk governance approach put previously in place. The responses in 2015 seemed to be more directed to preserve the Dublin Regulation at all costs, when it was clear that the system shaped by this Regulation is not made for a mass influx situation. In addition, the fair burden sharing presupposes a supranational approach, not followed by the member states which decided for a narrow national approach.

We argue that the main reason behind the inadequate overall joint crisis governance at EU level was a weak EU risk governance of the pre-crisis activities prior to 2015, which did not take into consideration the impact of national political, economic, security and cultural differences among European countries. It seems clear that the asylum shopping is an admission that the differences in the national asylum systems have not been harmonised by the Directives put in place before 2015. The EU did also not adequately consider the social mobilisation potential of a high number of migrants and refugees, including a possible number of IS fighters.

The analysis of the 2015 events through the pre-assessment and the risk appraisal phases has pointed out that the EU risk governance of the 2015 events was based on fragile premises, since both the pre-assessment and the risk appraisal were elaborated upon inadequate and not complete pieces of information. The use of big data and alternative data (Facebook, Twitter, Instagram and mobile phone data) could have contributed for a better understanding of migration-related phenomena.

REFERENCES

Aven, T. & O. Renn (2009). On risk defined as an event where the outcome is uncertain. *Journal of Risk Research* 12(1): 1–11.

Bevir, M., R.A.W. Rhodes & P. Weller (2003). Traditions of Governance: Interpreting the Changing Role of the Public Sector. *Public Administration 81(1)*: 1–17.

Bogason, P. (1996). *New Modes of Local Political Organizing: Local Government Fragmentation in Scandinavia*. New York: Nova Science Publishers.

Council (2013). *Directive 2013/32/EU of the European Parliament and of the Council of 26 June 2013 on common procedures for granting and withdrawing international protection (recast)*.

Council (2013a). *Directive 2013/33/EU of the European Parliament and of the Council of 26 June 2013 laying down standards for the reception of applicants for international protection (recast)*.

Council (2011). *Directive 2011/95/EU of the European Parliament and of the Council of 13 December 2011 on standards for the qualification of third-country nationals or stateless persons as beneficiaries of international protection*.

Council (2001). *Council Directive 2001/55/EC of 20 July 2001 on minimum standards for giving temporary protection in the event of a mass influx of displaced persons and on measures promoting a balance of efforts between*

Member States in receiving such persons and bearing the consequences thereof.
EASO (2017). *Agencies*. https://ec.europa.eu/home-affairs/what-we-do/agencies_en#6 accessed 13.11.17.
EEAS (2017). *EU operations in the Mediterranean Sea*. https://eeas.europa.eu/sites/eeas/files/5_euoperationsinmed_2pg.pdf accessed 13.11.17.
European Commission (2015). *European Commission makes progress on Agenda on Migration*. http://europa.eu/rapid/press-release_IP-15-5039_en.htm accessed 12.11.17.
European Commission (2015a). *Refugee Crisis: European Commission takes decisive action*. http://europa.eu/rapid/press-release_MEMO-15-5597_en.htm accessed 12.11.17.
FRONTEX (2017). *European Border and coast guard Agency*. http://frontex.europa.eu/ accessed 13.11.17.
Human Rights Watch (2017). *Like Living in Hell: Police Abuses Against Child and Adult Migrants in Calais*. New York: Human Rights Watch.
IOM (2017). *The EU Relocation Programme at IOM*. http://eea.iom.int/index.php/what-we-do/eu-relocation accessed 12.11.17.
IRGC (2005). *White paper on Risk Governance: Towards an Integrative Approach*. Geneva: the International Risk Governance Council.
IRGC (2005a). *An introduction to the IRGC Risk Governance Framework*. Geneva: the International Risk Governance Council.
Kates, R.W. et al. (1985) (eds.). *Perilous Progress: Managing the Hazards of Technology*. Boulder, CO: Westview Press.
Kjær, A.M. (2004). *Governance*. Cambridge: Polity Press.
Kooiman, J. et al. (2008). Interactive governance and governability: an introduction. *Journal of Transdisciplinary Environmental Studies 7(1)*: 1–11.
Kooiman, J. (2003). *Governing as Governance*. London: Sage.
Krahmann, E. (2003). Conceptualizing Security Governance. *Cooperation and Conflict 28(1)*: 5–26.
Kruke, B.I. & C. Morsut (2015). Resilience in a multilevel crisis governance context: A tale of joint implementation of community, regional, national and EU response capabilities. In Podofillini L., B. Sudret, B. Stojadinović, E. Zio, W. Kröger (eds.). *Safety and Reliability of Complex Engineered Systems*. London: Taylor and Francis: 187–194.
Marcussen, M. & J. Torfing (2007). *Democratic network governance in Europe*. London, UK: Palgrave.
Morsut, C. & B.I. Kruke (2014). *Crisis response planning in a post-Westphalian Europe: How is resilient crisis management being shaped by multilevel crisis governance?* Paper presented at 2014 NEON Conference Stavanger.
Morsut, C. & B.I. Kruke (2016). The (European) Union civil protection mechanism: A reliable crisis governance tool? In Walls L., M. Revie and T. Bedford (eds.). *Risk, Reliability and Safety: Innovating Theory and Practice*. London: Taylor and Francis: 494–501.
Morsut, C. & B.I. Kruke (2017). Crisis governance of the refugee and migrant influx into Europe in 2015: a tale of disintegration. *Journal of European Integration* https://doi.org/10.1080/07036337.2017.1404055.
Nye, J.S. & J.D. Donahue (2000). *Governance in a Globalising World*. Washington DC: Brookings Institution.
Peters, G.B. (2000). Governance and Comparative Politics. In J. Pierre (ed.). *Debating Governance. Authority, Steering, and Democracy*. Oxford: Oxford University Press: 36–53.
Pierre, J. (2000). *Debating governance. Authority, Steering, and Democracy*. Oxford: Oxford University Press.
Regulation (2013). *On the establishment of Eurodac*.
Regulation (2013a). *Criteria and mechanisms for determining the Member State responsible for examining an application for international protection lodged in one of the Member States by a third-country national or a stateless person (recast)*.
Renn, O. et al. (2011). Coping with Complexity, Uncertainty and Ambiguity in Risk Governance: A Synthesis. *Ambio 40(2)*: 231–246.
Renn, O. (2008). *Risk Governance: Coping with Uncertainty in a Complex World*. London: Earthscan.
Rosa, E.A. (2003). The logical structure of the social amplification of risk framework (SARF): Metatheoretical foundation and policy implications. In Pidgeon, N.F., R.E. Kasperson and P. Slovic (eds.). *The social amplification of risk*. Cambridge: Cambridge University Press: 47–76.
Rosa, E.A. (1998). Metatheoretical foundations for post-normal risk. *Journal of Risk Research 1*: 15–44.
Rosenau, J.N. (1992). Governance, order, and change in world politics. In Rosenau, J.N. and E.O. Czempiel (eds.). *Governance without Government: Order and Change in World Politics*. Cambridge, UK: Cambridge University Press: 1–29.
Sjöberg, L. (1999). Risk Perception by the Public and by Experts: A Dilemma in Risk Management. *Human Ecology Review 6(2)*: 1–9.
Slovic, P. (2000). *The Perception of Risk*. London: Earthscan Publications.
Stoker, G. (1998). Governance as theory: five Propositions. *International Social Science Journal 50(155)*: 7–28.
Sørensen, E. and J. Torfing (2007). *Theories of democratic network governance*. London, UK: Palgrave.
UNHCR (2017). *Operational portal: Mediterranean refugee situation*. http://data2.unhcr.org/en/situations/mediterranean accessed 01.11.17.
UNHCR (2017a). *Refugees & migrants sea arrivals in Europe*. https://data2.unhcr.org/ar/documents/download/53447 accessed 13.11.17.
UNHCR (2016a). *Refugees/Migrants Emergency Response—Mediterranean*. http://data.unhcr.org/mediterranean/regional.php accessed 13.11.17.
UNHCR (2016). *UNHCR viewpoint: 'Refugee' or 'migrant' – Which is right?* http://www.unhcr.org/55df0e556.html accessed 11.11.17.
UNHCR (2015). *The sea route to Europe. The Mediterranean passage on the age of refugees*. http://www.unhcr.org/5592bd059.pdf accessed 10.11.17.
UNHCR (2010). *Convention and Protocol relating to the status of refugees*.
Van Asselt, M. & O. Renn (2011). Risk Governance. *Journal of Risk Research 14*: 431–449.
Weick, K.E. (2011). Organizing for Transient Reliability: The Production of Dynamic Non-Events. *Journal of Contingencies and Crisis Management 19(1)*: 21–27.
Weinstein, N.D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology 39(5)*: 806–820.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Unforeseen events with a major accident potential—a study of some examples from the Norwegian oil and gas industry

W. Røed
University of Stavanger, Stavanger, Norway

ABSTRACT: Many severe accidents such as Piper Alpha and Deepwater Horizon are mainly a result of long event sequences, which have developed gradually for a significant period, before it comes to a point of no return where control is lost, and emergency preparedness has to take over. During the significant build-up period, sometimes referred to as a 'spiral to disaster', there are often several opportunities where control could have been regained, if the awareness and understanding of the sequence of events had been sufficiently understood. However, since it was not, the opportunity to prevent the major accident failed. This paper presents and discusses some selected historical accidents and near-misses in the Norwegian oil and gas industry, the aim being to improve the understanding of accident propagation and in particular signals and warnings that could have been seen, but was not. The ambition is to achieve insight that can be used to improve future risk assessments and the ability to detect unforeseen events in general.

1 INTRODUCTION

As emphasized by Vinnem and Røed (2014), many accidents in the industry such as Piper Alpha (Cullen, 1990), Longford (Hopkins, 2000), Texas City (CSB, 2007) and Deepwater Horizon (Presidential commission, 2011) are mainly a result of long event sequences, which have developed gradually for a significant period, before it comes to a 'point of no return' where control is lost, and emergency preparedness has to take over. During the significant 'build-up' period, sometimes referred to as 'spiral to disaster', there are usually several opportunities where control might have been regained, if the awareness and understanding of the sequence of events had been sufficiently understood. But since it was not, the opportunity to prevent the major accident hazard failed. The present paper studies in detail some historical events and describes what contributed to the 'spiral'.

In the study, selected historical near-accident events in the Norwegian oil and gas industry have been reviewed. They were selected based on having a potential to escalate and result in a major accident, often referred to as accidents with more than two fatalities (Vinnem and Røed, 2015) or with other extensive consequences, for example to the environment or assets (PSA, 2016).

The goal of the analysis is not to obtain more accurate probability estimation of major accident events, but to improve the overall risk management linked to such events. In particular we are motivated by the fact that involved personnel are often not able to see what is coming, although in hindsight, it may be concluded they should have—the signals and warnings were there, but the system and risk understanding was poor. And then, after the accident has occurred, the accident propagation becomes "obvious", and it is questioned "How could this occur without being noticed?" The aim of the study is to give increased understanding of how some accidental events have propagated—knowledge that can contribute to prevention of such events in the future.

The paper is organized as follows: In chapter 2 we introduce the selected events and the criteria they were selected based upon. In Chapter 3, the results and implications of the results are presented and discussed. Chapter 4 provides a discussion, and in Chapter 5 we draw some concluding remarks.

2 THE SELECTED EVENTS

The study includes 16 selected events in the Norwegian oil and gas industry in the period 2008–2015, as demonstrated in Table 1. This includes all publically known events with an accident investigation report available in the public domain and with a major accident potential, as explained in the previous section. For two events, we have used company reports, and for the remaining 14 events, the accident investigation was performed by the petroleum safety authorities in Norway. The above comprises 12 unignited hydrocarbon leaks, one non-process fire, one oil spill to the sea, one loss of well control and one situation with loss of buoyancy/stability

on a floating production unit. Events before 2008 were excluded since they are considered less relevant for today's safety regime and since the quality of the accident investigation reports has increased over time and in particular after 2008.

Table 1. Events included in the study.

Charact.	Description
10.01.2008 Draugen* Unign. HC	A hydraulic hose broke. Due to poor design, pressure surge occurred. This caused automatic release of an offloading hose, resulting in 6 m^3 oil spill.
24.05.2008 Statfjord A* Unign. HC	Oil leak during hot tapping. Release of 156 m^3 oil in the utility shaft and 70 m^3 to sea. Flashing equal to 0.9 kg/s gas.
12.09.2008 Oseberg C* Unign. HC	A solenoid valve was replaced with a wrong spare. Due to lack of flushing, the process valve opened very quickly initiating a hammer effect, causing a 26 kg/s gas release with total amount 1500 kg.
19.05.2009 Kollsnes** Unign. HC	Flange bolts were installed with the wrong torque. This resulted in a 12 ton condensate leak with initial leak rate 22 kg/s.
05.11.2009 Veslefrikk* Well event	Release of 3450 m^3 oil based cuttings and 93000 m^3 oil containing slop from an injection well to the sea bed.
08.02.2010 Mongstad** Unign. HC	During installation of insolation, it was drilled through a line filled with LNG, resulting in a 0.08 kg/s gas leak with total release of 300 kg.
12.09.2010 Mongstad** Unign. HC	A wrench was used to stop an observed leak. Technical equipment loosened and was blown 30 meters by the pressure.
04.12.2010 Gullfaks B* Unign. HC	During leak testing and maintenance work on a choke valve, internal leaks resulted in release of 1.3 kg/s and total amount 800 kg.
13.07.2011 Valhall* Non-process fire	A fire occurred in a crane engine resulting in burning/glowing particles from the exhaust igniting a vent stack.
26.05.2012 Heimdal* Unign. HC	Valves were opened in the wrong sequence resulting in a pressure build-up. Due to poor design, a pipeline burst.
12.09.2012 Ula* Unign. HC	Repair of a seepage of produced water was postponed to the upcoming revision stop. Before being repaired, the bolts failed due to corrosion, resulting in release of 20 m^3 oil and 1600 kg of gas.
07.11.2012 Floatel Superior* Loss of stab.	Anchor bolsters were not constructed to withstand expected weather conditions. This resulted in loose anchors hitting and penetrating the vessel, causing severe listing.
17.06.2013 Oseberg A* Unign. HC	Gas leak due to rupture in the blow-down line from the test separator. The line was not designed for sand in the well flow, and the segment was not sufficiently segregated from another segment.
05.01.2014 Hammerfest LNG** Unign. HC	Loss of seal liquid in a packing box resulting in a 0.1 to 0.3 kg/s HC gas leak with total amount 250–750 kg.
26.01.2014 Statfjord C* Unign. HC	Oil leak with initial leak rate 20.8 kg/s from sump tank to the cellar deck and to sea. 40 m^3 oil was spilled to sea and 2 m^3 oil was spilled on the installation.
18.01.2015 Gudrun* Unign. HC	The design included an insufficiently dimensioned control valve. This caused high vibrations resulting in a burst pipe. The result was a 8 kg/s release of 2.8 tons condensate.

In order to analyse each event at a sufficient level of detail, we have identified barrier elements that did not perform as intended during the accident propagation, and then studied each of these failures more in-depth. During the categorization process, the lists of root causes presented in the investigation reports were used as a basis, although for some of the events, some of the root causes mentioned in the reports were combined and presented in a simplified manner. To some extent, interpretation of the information in the accident investigation reports was needed during the identification process. The number of barrier elements varied from 1 to 6 for each event with a total of 53 barrier elements for all the 16 events together.

3 SAFETY BARRIER PERFORMANCE

3.1 *Safety barrier performance categorisation*

The performance of safety barriers can be measured in several dimensions such as capacity, reliability, availability, accessibility, efficiency, ability to withstand loads, integrity and robustness (PSA, 2015). In the present paper we have considered integrity/availability (was the barrier ready to be used?), functionality (did it work as intended?) and robustness (did it "survive" the accidental conditions?). For example, lack of a mandatory risk analysis is considered loss of availability, since the activity was not carried out at all. Incorrect planning is considered loss of functionality, since the

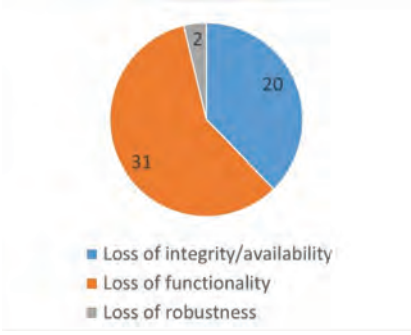


Figure 1. Barrier element performance categories.

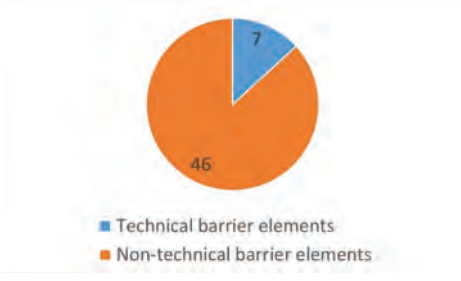


Figure 2. Technical and non-technical barrier elements.

planning was carried out (available) but failed and was inefficient. An example in the robustness category is technical equipment that failed due to vibration, i.e. the barrier element was vulnerable—not able to withstand the accidental loads (vibration). Out of the 53 barrier elements studied, 20 barrier elements had loss of the integrity/availability, 31 had loss of functionality and two had loss of robustness as shown in Figure 1.

Figure 2 presents the number of technical and non-technical barrier element failures. There are numerous barrier definitions, see for example the discussions by Lauridsen et al. (2016) and Øien et al. (2015). In the Norwegian petroleum industry it is often distinguished between operational, organizational and technical barrier elements, sometimes also simplified to technical and non-technical barrier elements as in Figure 2. Out of the 53 barrier elements studied, only 7 were technical barrier elements while 46 were non-technical, i.e. operational or organizational. This is in line with other research showing that a high fraction of root causes in accident investigation reports are non-technical, see for example Vinnem and Røed (2015) and Mostue et al. (2014).

3.2 *Barrier element failures—did anyone suspect the failures*

It is sometimes distinguished between known knowns, unknown knowns and unknown unknowns in the research literature. Secretary of Defense Donald Rumsfeld explained these terms in a press briefing February 12, 2002 (later elaborated on in Rumsfeld, 2011):

> "*…there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things [we know] we do not know. But there are also unknown unknowns—the ones we don't know we don't know*."

Inspired by the above terminology, Aven and Krohn (2014) introduced three categories of failures: a) Events that were completely unknown to the scientific environment (unknown unknowns), b) Events that were not on the list of known events from the perspective of those who carried out a risk analysis (or another stakeholder), and c) Events on the list of known events in the risk analysis but found to represent a negligible risk.

These categories correspond to unknown unknowns, unknown knowns and known knowns, correspondingly. As emphasized in Figure 3, out of the 53 barrier elements studied, none were in the a) category, 14 were in the b) category, and 39 were in the c) category. Examples in the b) category are a missing orifice that the personnel were not aware of and lack of identification of the major accident potential during a risk assessment. In the c) category, examples are overruled inspection intervals, known violations, work practice different than prescribed and required risk assessments not being performed. For the majority of the barrier element failures in category c), someone in the organization had knowledge about weaknesses indicating that the barrier element was, or could be, deteriorated. However, the probability of this resulting in a major accident hazard was considered sufficiently low to accept the risk, although in most cases the risk was not formally assessed and accepted according to company procedures.

The above classification was to some extent subjective, and in particular it was difficult to distinguish between the b) and c) categories for some events. If the b) criterion "or another stakeholder"

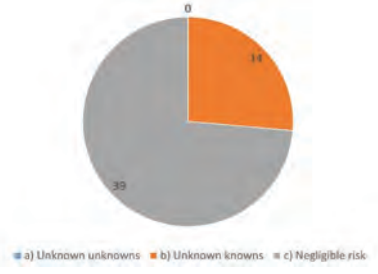


Figure 3. Event categorisation.

had been interpreted as anyone in the organization, there would be more c's and fewer b's giving even stronger evidence that known knowns was the most frequent challenge for the studied events. From a risk management perspective, it is good news that there were no events in category a) and few events in category b), since the knowledge and understanding of the situation, and thus the ability to manage the risk in a proper way, is stronger for the barrier failures in category c) than for the ones in category a) and b).

3.3 *Barrier element failures—who knew about it*

So far, we have seen that for many of the barrier element failures, someone knew there were deteriorated or weakened barrier elements, and since the activity was not stopped, this means the risk was formally or informally accepted. Now we will discuss who had this knowledge. We have distinguished between i) personnel in the sharp end of the organization, such as process operators and maintenance personnel offshore, and ii) personnel in the blunt end, such as project planners and managers working onshore. The latter was only studied when criterion i) was not met.

As illustrated in Figure 4, For 28 out of the 53 barrier element failures, the weaknesses were known by personnel in the sharp end. Examples are lack of compliance, insufficient knowledge about procedures, lack of risk assessment, insufficient maintenance and known weaknesses in design. For 12 of the barrier element failures, weaknesses were known to someone else in the organization. Examples are insufficient design or fabrication, unclear responsibilities, poor quality of a risk assessment and insufficient maintenance. For the remaining 13 barrier element failures, there is insufficient information in the accident investigation report to make a proper categorization; it is not explained in detail who knew what. This is not a surprise as it is not common to include such a level of detail in studies of near-misses and non-catastrophical events, as the ones being studies in this paper. The above implies that to some extent we have needed to read between the lines, and sometimes we have made assumptions based on the information available in the reports.

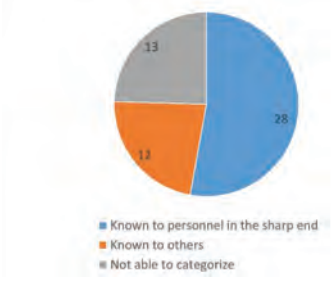


Figure 4. Who knew about the barrier element failures?.

3.4 *Surprises*

For 36 of the 53 barrier elements there were surprises. We have distinguished between two situations; i) it was a surprise that the barrier element actually failed (28 occasions) and ii) it was a surprise that the barrier element actually played a role as a safety barrier and was safety critical (8 occasions). The last category implies a lack of system understanding; the personnel involved did not realize that something being safety critical actual was. Examples in the first category are unknown built-in weaknesses, unclear responsibilities, insufficient planning or assessments ahead and technical equipment that was believed to be similar but was in fact not. Examples in the second category are safety-critical technical equipment considered non-critical for safety before the accident occurred. The majority of the surprises are of the first kind. This implies that for the majority of the situations with surprises, it was realized that weaknesses in the barrier element could potentially result in an accident, but it was a surprise that the barrier element actually failed and that the even actually occurred.

The above surprises are discussed from a viewpoint of before the accident occurred. During our work, we also wanted to find out if it, after the accident occurred, still was a surprise that some barrier elements contributed to the accident occurring. No such surprises were found. At first glance this may seem surprising in itself, but in fact, it is not: Since we used accident investigation reports as our source of information, and the barrier elements in our study were chosen because they contributed to the event occurring, it is not surprising that each of the barrier elements' contribution to the event occurring easily can be explained in hindsight. The above emphasizes the importance of realizing that the present study relies purely on information in accident investigation reports, and that this must be kept in mind when the results are interpreted. As elaborated on by Damnjanovic and Røed (2016), investigation reports by definition provide information about the propagation of the events and work processes only when an accident occurs. It is a general industry challenge that we do not have a full understanding of the events and work flow in situations with success and no accident occurs. This challenge has inspired scientists working in the field of resilience, as elaborated on by Nemeth and Hollnagel (2014), Hollnagel et al. (2013) and others.

3.5 *Early warnings*

For 29 of the barrier element failures there were early warnings indicating that something abnormal

was occurring. For 20 occasions, there were no early warnings and for the remaining four cases, we do not have sufficient information to make a proper categorization. We have divided the warnings into five groups based on their characteristics:

i. A physical warning (phenomenon) occurred in advance to the incident, with sufficient time to recover if it had been recognized (6 occasions)
ii. A similar or near to similar situation was experienced earlier on the same site (2 occasions)
iii. Uncommon solution, different from conventional solutions, but implications of this was not sufficiently recognized (5 occasions)
iv. Degradation was seen or could (should) have been seen by required maintenance/inspection (6 occasions)
v. Insufficient governing documents or poor adherence to governing documents (10 occasions)

The first category includes a physical warning in advance to the event, with sufficient time to recover the situation. Examples are an increase or drop in pressure, anchors that slammed into the vessel in harsh weather and severe vibrations in process equipment. These are 'strong' warnings since they clearly indicate that something abnormal is going on. Also category iv) includes rather strong warnings with situations that could, and should, have been detected by maintenance and inspection programs. Examples are visible degrading of a hose, valves with observed internal leakage, insufficient maintenance of a crane engine and lack of testing according to the maintenance program. Category ii) includes occasions where a similar or near to similar situation had been experienced at the same site, for example a known challenge that technical equipment had been fabricated without a safety critical item installed. Category iii) includes situations where a nonconventional solution is used, and where related implications were not fully understood or recognized. Examples are a change in the commissioning phase not being validated, and an emergency shutdown system for which parts of the system could be put out of operation. Category v) includes warnings in terms of an insufficient management system or a culture with poor adherence to procedures or requirements specified in the management system. Examples are errors and inconsistencies on a work permit, lack of sufficient resources and competence related to a work task being performed and insufficient barrier strategy and barrier design. With 10 occasions, this is the most common early warning category. Unfortunately, this category also includes the 'weakest' warnings, since there may be many such breaches without an accident or near miss occurring. This means that it was difficult to realize the 'spiral to disaster' had started.

Nine of the barrier element failures included poor quality of risk assessments or required risk assessments not being performed at all, although none of these deficiencies had early warnings. One important objective of a risk assessment is to identify potential hazards, and the above emphasizes the importance of risk assessment actually being performed—with sufficient quality, and the importance of having mechanisms in the organization ensuring that this is the case. If potential hazards are not identified at all, there may be no or few early warnings, and existing warnings may be overlooked. This may potentially result in poor risk management and acceptance of risks based on an insufficient basis.

4 DISCUSSION

A traditional view on barrier management is to establish a sufficient number of barriers and ensure they perform as required. With reference to James Reasons's Swiss Cheese model (Reason, 1990), this means installing a sufficient number of cheese slices with no (or few) holes. However, as stated by the National Commission investigating the Macondo accident (Presidential commission, 2011), 'Complex systems almost always fail in complex ways'. This statement emphasizes that barrier management is not simple at all: For the Macondo accident lack of sufficient safety barriers were easily pinpointed in retrospect, but it was not evident until the accident occurred.

For the majority of the events studied in the present paper, more than one barrier element failed and contributed to the event occurring. The number of barrier element failures varies from 1 to 6 for each of the accidents within the classification used in the paper. As emphasized by Vinnem and Røed (2015), if root causes had been studied in more detail, more factors would most likely have been identified.

All events studied are near-misses; situations that potentially could have resulted in a major accident, but due to various circumstances did not. A question is then if the situations studied are relevant for learning about major accident prevention. We believe the study is relevant since only events with a major accident potential have been included in the study, as explained in the introduction to the paper. Thus, all events included can be considered lagging major accident precursors as discussed by for example Hopkins (2009), Kjellén (2009) and Vinnem (2010). This implies that the events study may bring relevant information to the table when it comes to causes and contributing factors to major accidents, but limited information about potential consequences.

It is important to be aware that the study considers a sample of selected events based upon the criteria explained earlier in the paper. It is not an empirical study with random events. This should be kept in mind when the results are interpreted. For example, it is a relevant challenge that the

number of barrier element failures identified in the accident investigation reports may depend on the severity of the accident, since it is likely that severe accidents are investigated to a higher level of detail, and thus, more barrier element failures are highlighted in the reports. This may partly explain why the number of barrier failures varied between 1 and 6 for the events studied.

5 CONCLUSION

In this paper, 53 barrier element failures related to 16 historical events with a major accident potential in the Norwegian petroleum industry have been studied. The majority were non-technical barrier element failures associated with loss of functionality. In many cases, someone in the organization, typically in the sharp end, knew about weaknesses indicating that the barrier element was, or could be, deteriorated. For many of the events there were surprises. For the majority of the cases, it was a surprise that the barrier element(s) actually failed, but for some cases, it was also a surprise that the barrier element that failed actually was safety critical. For more than half of the barrier element failures, there were early warnings. Approximately half of the warnings were 'strong' warnings, such as a physical phenomenon that occurred in advance to the incident, with sufficient time to recover if it had been recognized and physical degradation that was seen or could have been seen by required maintenance and inspection. The resulting half were weak warnings, such as unconventional design or similar events being experienced previously on the same site.

Since only 16 historical events have been studied, care should be taken when interpreting the results. However, some general reflections have been made:

As demonstrated in previous studies, the present paper confirms that it is important to pay attention to non-technical barrier elements when risk is managed in organizations. The study also indicates that the main challenge is not events coming totally out of the blue. More often, someone in the organization, typically in the sharp end, are aware of barrier deteriorations and that these deteriorations may result in an accident. However, it is considered unlikely that an accident will actually occur. Since the activity is continued with deteriorated barrier elements, the increased risk is formally or informally accepted by someone in the organization.

ACKNOWLEDGEMENTS

The research is funded by ARCEx partners and the Research Council of Norway (grant number 228107).

REFERENCES

Aven, T. and Krohn BS. (2014) A new perspective on how to understand, assess and manage risk and the unforeseen, *Reliability Engineering and System Safety*, Vol. 121, January 2014, pages 1–10.

Chemical Safety Board (2007). Investigation report refinery explosion and fire.

Cullen WD (1990). The public enquiry into the Piper Alpha disaster. London. Department of energy.

Damnjanovic I and Røed W. (2016) Risk management in operations of petrochemical plants: Can better planning prevent major accidents and save money at the same time? *Journal of loss prevention in the process industries.* Vol. 44, November 2016, pp. 223–231.

Hollnagel E., Leonhardts J., Licu T. and Shorrock S. (2013). From Safety-I to Safety-II: A White Paper. Eurocontrol. 2013.

Hopkins A. (2000). Lessons from Longford: The Esso gas plant explosion, Sydney: CCH Australia Ltd.

Hopkins A. (2009). Thinking about process safety indicators. *Safety Science* 47 (2009), 460–465.

Kjellén U. (2009). The safety measurement problem revisited. *Safety Science* 47 (2009), 486–489.

Lauridsen O, Lootz E, Husebø T and Ersdal G. (2016). Barrier management and the interaction between technical, operational and organisational barrier elements. SPE Conference in Stavanger, April 2016.

Mostue BA, Sandvik PC, Steen-Hansen A and Storesund K., (2014). *Proceedings of ESREL 2013, Amsterdam, Netherlands* 29 September/2 October 2013.

Nemeth CP. and Hollnagel E. (2014). *Becoming Resilient—Resilience Engineering in Practice, Volume 2*. Ashgate.

Presidential Commission. (2011). Deepwater, the Gulf oil disaster and the future of offshore drilling. Report to the President, National Commission on the DP Deepwater Horizon oil spill and offshore drilling, January 2011. ISBN: 978-0-16-087371-3.

PSA (2015) Guidelines regarding the management regulations.

PSA (2016). Major accident risk website. Available at: http://www.psa.no/major-accident-risk/category1030.html. Accessed February 23, 2016.

Reason, J. (1990) The contribution of latent human failures to the breakdown of complex systems. Philosophical Transactions of the Royal Society (London), series B. 327: 475–484.

Rumsfeld, D (2011). Known and unknown—a memoir.

Vinnem JE (2010). Risk indicators for major hazards on offshore installations. *Safety Science* 48 (2010) 770–787.

Vinnem JE. and Røed W. (2014). Norwegian oil and gas industry project to reduce the number of hydrocarbon leaks. *SPE Economics and Management* (6) 2, pp. 88–99.

Vinnem, J.E. and Røed, W. (2015). Root causes of hydrocarbon leaks on offshore petroleum installations. *Journal of Loss Prevention in the Process Industries*, 36, 54–62.

Øien, K. Hauge, S. Tinmannsvik, R.K. (2015) Towards a holistic approach for barrier management in the petroleum industry, SINTEF Technology and Society.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Analysis of 985 fire incidents related to oil- and gas production on the Norwegian continental shelf

C. Sesseng, K. Storesund & A. Steen-Hansen
RISE Fire Research AS, Trondheim, Norway

ABSTRACT: Fire is a major threat in the petroleum industry. However, little has been published about the fire related incidents that have occurred in the Norwegian petroleum sector. To gain more knowledge, data from 985 incidents in the 1997–2014 period has been analysed. Examples of factors studied are type of facility involved, involved area or system, consequences and severity level. The analysis of the fire incidents reveals that even though many incidents are reported, the large majority of these have not imposed risks for severe fire accidents. It has also provided valuable information regarding possible dangerous situations, commonly involved areas, types of equipment as well as types of activity that were involved. Twenty-nine percent of the incidents were false alarms, which must be regarded as a high number in an industry where any production stop could be extremely costly.

1 INTRODUCTION

In Petroleum Safety Authority Norway's (Ptil) assessment of the risk level in the petroleum industry on the Norwegian continental shelf, defined situations of hazard and accidents (DFUs) are defined and utilised. A DFU is an unplanned event which has led, or may lead, to loss of life and other values. Also, a DFU must be an observable event which it is feasible to measure accurately (Vinnem et al., 2006). Different DFUs are associated with different risk areas. According to Ptil's annual report on trends in risk level, there has been a declining trend in DFUs occurrences associated with major accident risk from 2004 to 2014 (Petroleum Safety Authority Norway, 2016). In the period there were no hydrocarbon fires, but there were hydrocarbon leakages with ignition potential. However, a fraction of the observed DFUs were fires and explosion not involving hydrocarbons. Even though fire occurrences are few compared with other DFUs, fires have disastrous potential, and a fire preventive focus should be maintained.

The current study has a quantitative approach and look in depth into the fire incident statistics, with false alarms included, to gain more knowledge about the incidents, where they occurred and what their outcomes were, in order to found a basis for future work aiming to improve fire safety, improve detection reliability and prevent false alarms in the petroleum industry.

2 DATASET

The dataset which found the basis for this study is an extract of Ptil's fire incident statistics. RISE Fire Research received authorization from Ptil to analyse this set of data, and gained access to all relevant incidents over the defined period of time. The database contains fire incidents, small and large, which are systematically reported to the authority. The sample comprises 985 reported incidents from all facilities and operators in the areas within Ptil's jurisdiction in the 1997–2014 period. The incidents included in our selection were reported as one of the following incident types: ignited hydrocarbon leakage, fire/explosion in other areas, fire/explosion in other areas (not hydrocarbon fire), and fire/explosion in other areas (not hydrocarbon explosion).

The dataset comprise information about time of the event, at which facility and in which area or system it occurred. Furthermore, the severity together with actual and potential consequence of the incidents are classified and registered. Also, there is a free text field where each incident is described in short.

The severity of the incidents is assessed and reported according to a 5-point scale, where the different values are defined as follows: 1–not notifiable, 2–simpler follow-up, 3–potential under minor changes, 4–severe or 5–large potential/serious accident/death.

3 METHOD

In the current study, the analyses are based on fire statistics from Ptil. The study has a quantitative approach, which implies that the individual cases have not been studied in detail, except for information extracted from free text fields in the statistics database. The results from the analyses are presented as descriptive statistics.

4 RESULTS

4.1 *Sample description*

During the 1997–2014 period there were 985 reported fire incidents on the Norwegian continental shelf. From 14 incidents the first year of the period, there has been an increase over the period, ending on 66 incidents in 2014. In 2006, there was a peak with 84 reported incidents, see Figure 1.

The figure shows the development in number of incidents, distributed over severity degree over the period. Most incidents (91.2%) are classified with severity degree 2, which are incidents which require minor follow-up, whereas the remaining incidents are classified as one of the other severity degrees. For readability, the incidents in these categories are presented as one bulk in Figure 1. Most of the incidents in this bulk were incidents with severity degree 4 (n = 63). In addition there were 6 serious accidents (degree 5).

There is a leap in number of incidents from 2005 to 2006. The average number of incidents for the years before the leap, i.e. 1997–2005, is 35, whereas the corresponding number for the 2006–2014 period is 75. This corresponds to a 216% increase. As will be demonstrated in section 4.5, this is related to an increase in the number of reported false alarms.

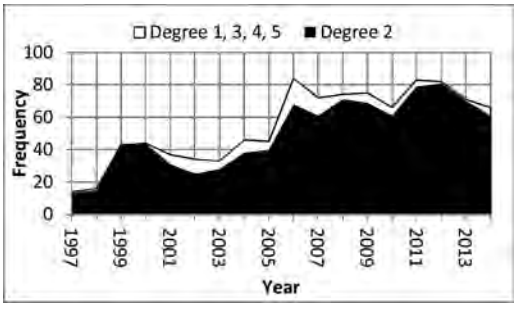


Figure 1. Number of reported incidents, distributed over severity degrees. Degree 2 (n = 898) is presented as one category, whereas degrees 1 (n = 15), 3 (n = 3), 4 (n = 63) and 5 (n = 6) are collapsed into one category for readability, N = 985.

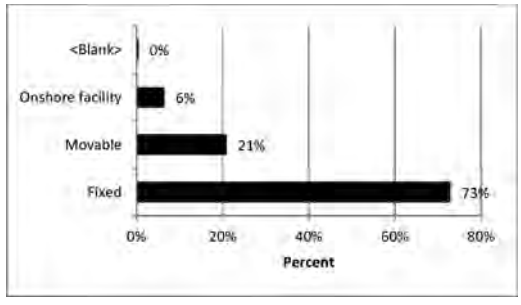


Figure 2. Distribution of incidents between different facility types, N = 985.

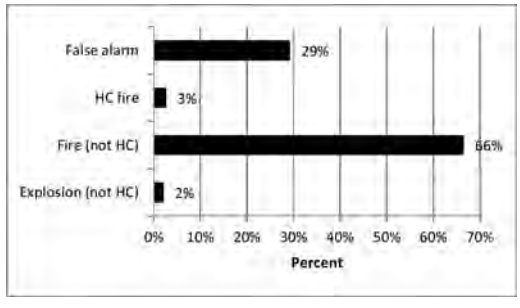


Figure 3. Actual outcome of reported alarms, N = 985.

Further, 2006 also stands out because the number of incidents with severity degree 4 is over twice as high as any other year in the period (15 incidents, compared with the year with the second highest number of degree 4 incidents: 7). The median over the period is 4.

A vast majority of the reported incidents occurred on fixed installations (73%), whereas one out of five incidents was related to movable installations. A minor proportion (6%) was incidents occurring in onshore facilities.

Over 70% of the incidents were real fire and explosion incidents. However, as Figure 3 demonstrates, a vast majority of the incidents were non-hydrocarbon fires, but rather fires in electrical systems, overheated machinery etc. Almost one third of the incidents were classified as false alarms.

4.2 *Type of area/system involved*

The incident reporting system on which this analysis is based upon is designed so that there is one variable to register both the area and system involved, i.e. one cannot discriminate between different systems in one specific area. E.g. a fire may start in an electric installation, but it is not specified if the electric installation is in the living quarter, main process or other areas of the facil-

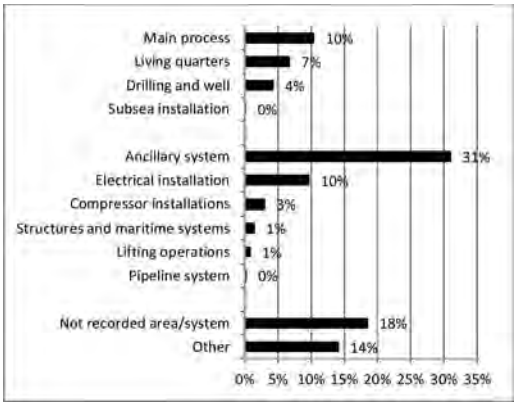


Figure 4. Distribution of incidents between different areas (top group) and systems (middle group). Incidents without recorded area or system is shown in the bottom group, N = 985.

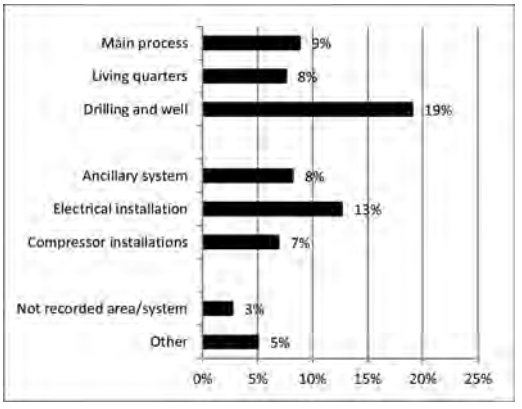


Figure 5. Incidents causing drilling downtime, distributed over the area or system the incident occurred in. Areas and systems with fewer than 20 reported incidents are excluded from the figure, n = 76.

ity. In Figure 4, and the figures following, areas and systems have been split into two separate groups, where the first group represent areas and the second group represent systems. The areas and systems are sorted with descending number of reported incidents within each group.

Further, incidents with no area or system recorded, or incidents in other areas or systems constitute a third group. The categories in the latter group are relatively large, with 18% and 14% of the incidents, respectively. The majority of the incidents within these categories (88%) were reported between 1997 and 2002. Also, it is seen that there is a distinction between the categories. While most of the incidents tagged with "Not recorded area/system" were reported before 2002, the main part of the incidents tagged with "Others" was reported after 2004.

With reference to Figure 4, one sees that of all reported incidents, nearly one third occurred in ancillary systems. These systems comprise, among others, communication systems, electrical power supply systems and water treatment facilities. In short, systems not related to separation, production and transport of hydrocarbons. An equal proportion incidents occurred in areas and systems not specified. The remaining incidents were distributed over main process (10%), electrical installations (10%), living quarters (7%) and others (9%).

4.3 *Consequences*

Of the 985 reported incidents in the period, only 8% resulted in drilling downtime whereas 23% caused production stoppage.

Not surprisingly, one fifth of all incidents causing drilling downtime occurred in the drilling and well area, see Figure 5. Of all incidents causing

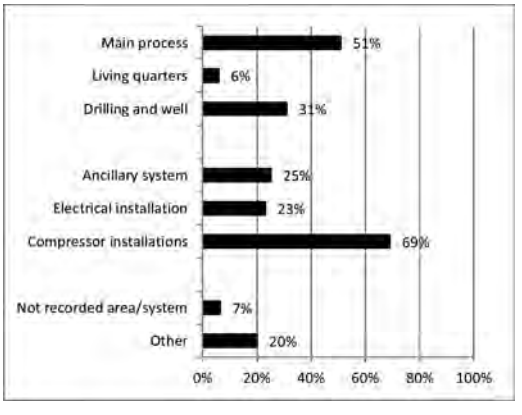


Figure 6. Incidents causing production stoppage distributed over the area or system the incident occurred in. Areas and systems with fewer than 20 reported incidents are excluded from the figure, n = 231.

production stoppage, half of them occurred in the main process and one third occurred in the drilling and well area, see Figure 6.

4.4 *Severity level*

The vast majority (91%) of the reported incidents were classified as incidents requiring simpler follow-up (severity degree 2), whereas only 3 incidents (≈0%) had the potential of becoming a severe situation under minor circumstantial changes (degree 3). Six percent of the incidents were regarded as severe (degree 4) and <1% (6 incidents) had a large potential or were large accidents, but none resulted in fatalities. Three of these occurred at onshore facilities. In addition, there

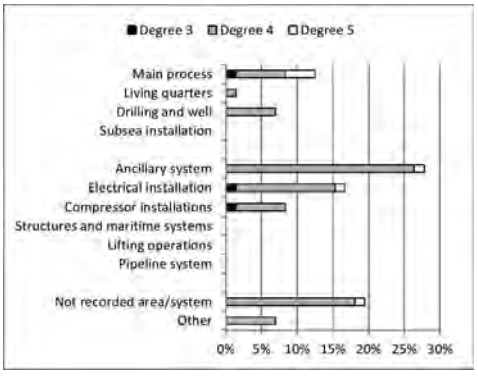


Figure 7. Incidents with severity degree 3 (n = 3), 4 (n = 63) or 5 (n = 6), distributed over different areas and systems, total n = 72.

were 15 incidents (2%) which had the lowest severity degree (degree 1), even though these incidents are not notifiable.

Examples of incidents on severity level 3 are fires that were extinguished after a short period of time, either by automatic extinguishing systems or by manual effort. A fire in an HVAC module with smoke spread to the living quarter, on the other hand, was classified as severity degree 5.

It is not straight-forward to analyse trends in severity degree over time, since there are few incidents with severity degree other than 2. However, it is seen qualitatively that the proportion of incidents with severity degree 4 is lower in the period 2008–2014 (average 3,4%) than it was between 2001–2007 (average 12%). The same trend is seen when adjusting for the increase in false alarms after 2006 which yields a decrease in degree 4 incidents from an average of 19.5% in the first period to 5.3% in the second period. Also, there has not been an incident with severity degree 5, since 2008. It therefore seems that there is a decline in the degree of severity of the incidents in the sample over time.

When studying the distribution of the incidents with severity degree 3 or higher over the different areas and systems, it is seen that most incidents occur in the main process and drilling and well areas, see Figure 7. Correspondingly, one fourth of the incidents with this severity took place in ancillary systems and 15% in electrical systems.

Even for these degrees of severity, there are around one fifth of the incidents, whereof one incident was classified as *large potential/serious accident/death* (degree 5), where area or system has not been recorded.

4.5 *False alarms*

False alarms are alarms caused by other circumstances than fire and explosion. According to ISO/DIS 17755–2, a false alarm is an alarm for which no fire occurred or […] due to accidental operation of fire alarm devices (ISO, 2010). The most frequent causes observed in the sample were detectors malfunctioning, misinterpretation of the situation by the detection system, technical and human errors. Examples of misinterpretations are sandblasting dust being detected as smoke, heat from sauna detected as heat from fire, heated leakage of lubricating oil detected as smoke and steam from cleaning detected as smoke.

The number of false alarms was quite low in the first half of the focus period, see Figure 8. Up until 2005, there were only a few cases, whereas in 2006 there is a leap, and in the years following the average number of false alarms per year is 29. This is probably not a real increase in false alarms, but rather an effect of a new reporting scheme, where more incidents are included than before.

From 2006 towards the end of the period, there is a seemingly decrease in the number of false alarms. However, the trend has not been checked statistically or adjusted for changes in the petroleum activity on the continental shelf.

One third of all false alarms occurs in relation to ancillary systems, and one fifth occurs in the main process, see Figure 9. In fact, when taking

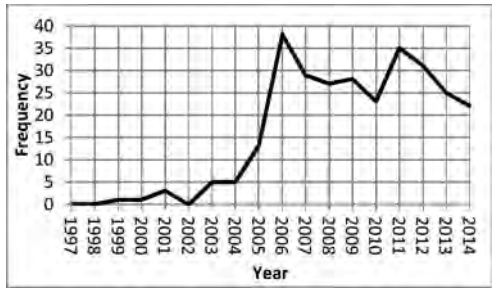


Figure 8. Number of reported false alarms each year, n = 286.

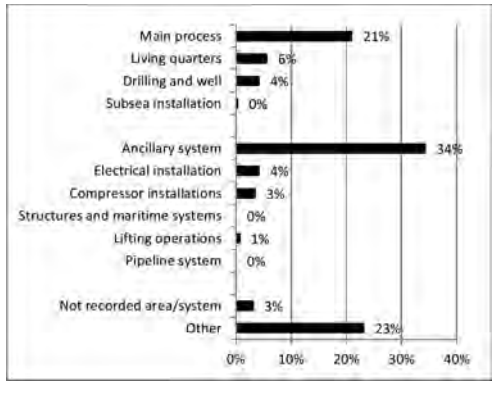


Figure 9. False alarms reported, distributed over areas and systems, n = 286.

into account the number of incidents in each area or system, it is seen that 60% of all reported incidents in the main process are false (Figure 10). This is almost twice as large proportion of false alarms than any other area or system (disregarded the incidents categorised as "other", as this category most likely constitutes numerous sub-categories).

Table 1 presents an overview of how many of the false alarm incidents, in each area or system, which caused either drilling downtime or production stoppage. The table should be read with caution, as some of the areas or systems have very few incidents, which may yield large percentages.

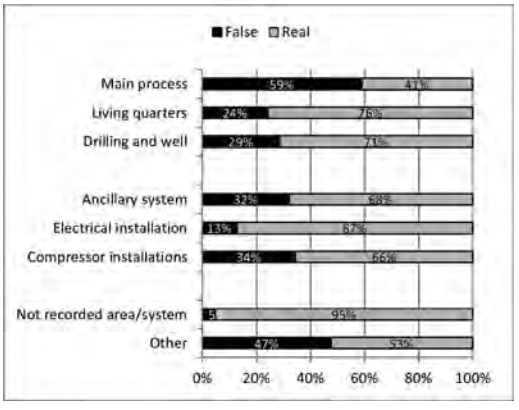


Figure 10. Proportion of real and false alarms for each area and system. Areas and systems with fewer than 20 reported incidents are excluded from the figure. Main process n = 102, living quarters n = 66, drilling and well n = 42, ancillary system n = 305, electrical installations n = 95, compressor installations n = 29, not recorded n = 182, other n = 139, total n = 960.

Table 1. The number of false alarms which caused either drilling downtime or production stoppage.

	Caused drilling downtime		Caused production stoppage		
Area/system	[freq.]	[%]	[freq.]	[%]	n
Main process	6	10%	39	65%	60
Living quarters	2	13%	1	6%	16
Drilling and well	1	8%	4	33%	12
Subsea installation	0	0%	1	100%	1
Ancillary system	6	6%	30	31%	98
Electrical installation	0	0%	1	8%	12
Compressor installations	0	0%	7	70%	10
Structures and maritime systems	–	–	–	–	0
Lifting operations	0	0%	0	0%	2
Pipeline systems	–	–	–	–	0
Not recorded area/system	0	0%	3	33%	9
Other	5	8%	23	35%	66
All	20	7%	109	38%	286

Nonetheless, 7% of the false alarms resulted in drilling downtime and 38% caused production stoppage. In addition, a total of 129 false alarm incidents (45%) resulted in personnel mustering to life boats.

Furthermore, 65% of the false alarms occurring in the main process caused production stoppage. Similarly, 31% of the false alarms caused by ancillary systems had the same consequence.

Fewer false alarm incidents caused drilling downtime. Again, false alarms occurring in the main process or ancillary systems are the main cause for drilling downtime.

5 DISCUSSION

5.1 *Incidents*

Over the 18 year period there were almost 1000 reported fire incidents on the Norwegian continental shelf, and it is seen that there was almost twice as many incidents in the second half of the period compared to the first half. The increase is probably an effect of a shift in reporting regime, where more incidents (mostly false alarms) than before were included. There is therefore reason to believe that there were even more incidents in the first half of the period than what has been reported, but that most of the unreported incidents were false alarms incidents.

The analysis of the fire incidents reveals that even though many incidents are reported, the large majority of these have not imposed risks for severe fire accidents. There also seems to be a positive trend regarding the severity degree of the fire incidents, as there is a decline in the number of severe incidents and major accidents. However, the current analysis has not adjusted for the activity level in the Norwegian sector or other possible covariates. Also, in general one should be careful to draw any conclusions based upon the trend of a single indicator, as there may be other indicators not investigated in the current study which may affect the fire safety level negatively (Vinnem, 2010; Vinnem et al., 2006).

Also, since the current study is retrospective in nature, it is important to emphasize that a possible change in the conditions which may affect the fire safety level will not be observed in the incident statistics until later, and that the apparent trend is only valid for the focus period (Vinnem et al., 2006).

The current study does not conclude on the underlying causes of fires offshore. Future studies should therefore focus on revealing such causes in addition to triggering factors for the fire incidents. This can be done by examining investigation

reports from the different incidents and by performing interviews with key personnel with the operators. A study with such a design, investigating the underlying causes for 35 fires in electrical equipment on offshore platforms in the Norwegian sector, is reported in (Storesund et al., 2012). The study categorised the causes according to the Human–Technology–Organisation perspective, which may be helpful when trying to sort and reveal patterns in causes and find suitable and targeted measures.

5.2 *False alarms*

A great proportion of the reported incidents in the period were false alarms, and a relatively large fraction of these have been shown to cause production downtime and consequently economical losses. The classic ever-returning dilemma of smoke and fire detection is that increasing the detectors' sensitivity to detect fires as early as possible also causes an increase in number of false alarms. Correspondingly, decreasing the sensitivity to eliminate false alarms affects fire detection time negatively.

One of the main reasons for false alarms is that the detection system misinterprets the situation, and for instance takes steam or dust as smoke. However, detection systems have become smarter and there are several technologies available that contribute to reduce the number of false alarms caused by misinterpretation of the situation.

E.g. studies have shown that multi-sensor detectors with CO sensor can both decrease detection time for certain types of fires in addition to reducing the number of false alarms (Cestari et al., 2005; Sesseng et al., 2016; Sesseng and Reitan, 2016). The mentioned studies have only investigated the residential case, and there may be areas where such detectors are not suitable. Still, there is reason to believe that many areas may take advantage of this technology, e.g. living quarter, workshops etc.

The next main causes for false alarms are technical errors and malfunctioning detection systems. At the same time, compared to other barrier elements, fire detection systems have the lowest failure rate when tested. Each year, some 50,000 tests of fire detectors are performed on offshore facilities in the Norwegian sector, and since the beginning of the reporting of these tests in 2002 the mean fail rate has been declining. In 2002, around 0.9% of the tested detectors failed the tests, whereas only 0.1% failed in 2015 (Petroleum Safety Authority Norway, 2016, 2015, 2010; Vinnem, 2010). The trend is positive, and it should therefore be a continued focus on maintenance and testing.

The last main cause for false alarms is human errors. This could be due to work made in the proximity of a sensor, work during service and testing of the system or failure to comply with procedures. This is most likely best managed by improved routines for work and risk assessment as well as focusing on procedural compliance.

Obviously, if the number of false alarms can be reduced there will be great economic benefits. The majority of installations on the Norwegian shelf is ageing, and anecdotal evidence suggests that some having old or outdated detection systems. It should be investigated whether an upgrade of the fire detection systems could reduce the number of false alarms and, consequently the number of false alarms resulting in production stoppage and mustering.

5.3 *Reporting*

The current reporting scheme has certain shortcomings. By registering information concerning system and area in the same variable, information is lost. The obvious consequence is that one would have to choose to register either area or system, which would be at the reporter's discretion. Besides, a specific type of system, e.g. ancillary systems, could be found in several areas, but may constitute different risks in different areas, which makes the available information of limited value. The reporting scheme ought therefore to be changed such that more details regarding involved area and system are recorded.

A large number of incidents categorised as "Not recorded area/system" was reported before 2002. Almost half of the incidents categorised as "Other" were false alarms and could, through the free text description, be derived to specific areas/systems. This shows that in some ways the procedure and culture of reporting seems to have improved over the years. At the same time it appears that it may be difficult to categorise false alarms.

6 CONCLUSIONS

The numbers show that there is room for improvement regarding fire safety in the petroleum production on the Norwegian shelf. There are many incidents, although with low degree of severity. Future work should focus on investigating the underlying causes and triggering factors of the fire incidents, to be able to find focused fire preventive measures.

There is also a large number of false alarms, which may be quite costly if they cause production downtime. A more thorough investigation would be informative concerning what types of equipment are causing false alarms. This could found the basis for targeted measures decreasing the occurrence of false alarms and downtime and thus increasing the economic profit of the installations.

The numbers also show that severe incidents do not occur often, something that may be explained by good control of barriers. However, there are still some incidents that occur that have the potential of developing into a severe incident. Hence, there must still be a focus on barriers preventing the consequences of an escalating incident.

REFERENCES

Cestari, L.A., Worrell, C., Milke, J.A., 2005. Advanced fire detection algorithms using data from the home smoke detector project. *Fire Safety Journal* 40, 1–28. https://doi.org/10.1016/j.firesaf.2004.07.004.

ISO, 2010. ISO/DIS 17755-2: Fire safety—Statistical data collection—Part 2: Definition of terms.

Petroleum Safety Authority Norway, 2016. Trends in risk level in the petroleum activity. Summary report 2015. Norwegian continental shelf. Petroleum Safety Authority Norway, Stavanger, Norway.

Petroleum Safety Authority Norway, 2015. Risikonivå i norsk petroleumsvirksomhet. Hovedrapport, utviklingstrekk 2014, norsk sokkel. Petroleum Safety Authority Norway, Stavanger, Norway.

Petroleum Safety Authority Norway, 2010. Risikonivå i norsk petroleumsvirksomhet. Hovedrapport, utviklingstrekk 2009, norsk sokkel. Petroleum Safety Authority Norway, Stavanger, Norway.

Sesseng, C., Reitan, N.K., 2016. Experimental investigation of using CO sensors to detect smouldering fires in dwellings. Presented at the Suppression, Detection and Signaling Research and Applications Symposium (SupDet), San Antonio, Texas, USA.

Sesseng, C., Reitan, N.K., Fjær, S., 2016. Mapping of gas concentrations, effect of dead-air space and effect of alternative detection technology in smouldering fires (Project number: 20053:2). SP Fire Research AS.

Storesund, K., Mostue, B.A., Christian, S., Steen-Hansen, A.E., 2012. Hendelser med brann i elektriske anlegg—Årsaksforhold og tiltak (No. NBL A12137). SINTEF NBL as, Trondheim, Norway.

Vinnem, J.E., 2010. Risk indicators for major hazards on offshore installations. *Safety Science* 48, 770–787.

Vinnem, J.E., Aven, T., Seljelid, J., Tveit, O.J., 2006. Major hazard risk indicators for monitoring of trends in the Norwegian offshore petroleum sector. *Reliability Engineering & System Safety* 91, 778–791.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Implications from major accident causation theories to activity-related risk analysis – an application to the Norwegian Atlantic salmon farming industry

X. Yang, I.B. Utne & S.M. Holen
Department of Marine Technology, NTNU, Trondheim, Norway

I.M. Holmen
SINTEF Ocean, Trondheim, Norway

ABSTRACT: The Norwegian Atlantic salmon farming industry is exploring the possibility to run fish farms in more exposed locations. The severe wave and current conditions, irregular wind, sheer remoteness, and limited weather window challenge the operational planning to avoid accidents. The objective of this paper is to present results from applying a generic list of safety-critical parameters to a net cleaning operation to assess their usefulness and relevance in aquaculture. The list was proposed based on implications from major accident causation theories in safety research. The case study demonstrated that the list is a useful operational planning tool to identify activity failure mechanisms that have potential to cause accidents. The results also have implications to how to use barrier principles in aquaculture in general.

1 INTRODUCTION

The economic value created from Norwegian aquaculture is expected to reach 25.3 billion Euro by the year 2050, which means that the production will increase fivefold compared to the year 2010 (Olafsen et al., 2012). Despite the positive prediction, the fish farming industry is facing the challenge of the fewer available locations in the sheltered coastal environment and increasing negative ecological consequences due to sea lice, fish escapes and farm waste on the seabed (Holmer, 2010). One attempt to solve these challenges is to move fish farms to more exposed locations. This means that the fish farms have to deal with the amplified risk to both fish and human due to severe wave and current conditions, irregular wind and sheer remoteness. Many technical solution concepts are initiated based on extensive offshore experience from Norwegian oil and gas industry. The Norwegian Directorate of Fisheries, which is responsible for sustainable management of marine resources and marine environment, has approved several concepts that are based on offshore technology (Norwegian Directorate of Fisheries, 2017). These include Ocean Farm 1 from Ocean Farming AS, Havfarm NSK 3417 from Nordlaks Oppdrett AS, and Aquatraz from MNH production. Ocean Farm 1 has been put into pilot operation in Frohavet, Norway (KYST, 2017). Some fish farmers have even started running test facilities by expanding existing systems towards exposed areas with few significant technological and operational changes.

Despite the amplified risk due to exposed locations, the motivation for performing risk assessments is relatively low in parts of the aquaculture industry (Holmen et al., 2017). Aquaculture is different from an offshore industry where both authorities and companies have put significant efforts into systematic risk management. The similarity of the offshore locations and operating environment opens the discussion if aquaculture can or should learn safety practices from the offshore oil and gas industry. One distinction is that the offshore industry has been implementing a system-based risk analysis, while aquaculture considers more day-to-day practical routines (Pettersen, 2017). The system-based risk analysis is represented by Quantitative Risk Analysis (QRA). Take the hazardous event hydrocarbon leak as an example. The piping items (e.g., flanges, valves, instruments and process equipment, such as pumps, vessels) and technical safety barrier systems (e.g., gas detector, fire extinguishing system, and firewall) are modelled to derive a basic risk level from the design of the facility. The results are documented, and the technical conditions of these systems are followed-up during operation.

In a typical commercial fish farm today, the major operations include transferring the fish, delivery of feed to the fish farm, fish feeding, daily inspection, health and biomass control, oxygen measurement, net cleaning, removal of dead fish, delousing, and IMR (Inspection, Maintenance and Repair) (Bjelland et al., 2016). These routine operations in the fish farms dominate the variable

risk on a daily basis. This means that not only the technical condition of the fish cage, but also the operational and organizational factors, external impacts, such as the operating environment, the involved working units (e.g., service vessel, workboat) should all considered to keep the risk at an acceptable level (Yang and Haugen, 2016).

Yang and Haugen (2018) have looked into implications from major accident causation theories to activity-related risk analysis and proposed a generic list of safety-critical parameters to reveal possible activity failure mechanisms that may lead to major accidents. The list was generated from the principles and critical factors addressed in various accident causation perspectives, including the Energy-barrier perspective (Gibson, 1961), Man-made disasters theory (Turner, 1978), Conflicting objectives perspective (Rasmussen, 1997), Normal accident theory (Perrow, 1984), System-Theoretic Accident Model and Processes (STAMP) (Leveson, 2012), High-Reliability Organization (HRO) (Roberts, 1990) and Resilience Engineering (RE) (Hollnagel et al., 2006).

The objective of this paper is to present and discuss the results of applying these generic safety-critical parameters to daily fish farming operations to assess their practical usefulness and relevance in aquaculture. Net cleaning operation, which takes place approximately every two weeks, is chosen for the case study.

2 METHODOLOGY

Net cleaning, a critical and frequent operation in the marine fish farm to remove biofoulings (e.g., seaweed and mussels) on the net, was selected as a case in this paper. A professional service company provided a net cleaning procedure and related operational risk analysis document to the authors. However, the issues covered in that analysis are somewhat coarse. The Norwegian aquaculture industry is, in general, lacking full accidents/incidents statistics and in-depth analysis. Moreover, the risk factors during operations are not systematically identified and understood by the operators (Holmen et al., 2017b), which is challenging when attempting to collect safety-critical parameters (SCP) in the case study.

A preliminary list of SCPs was identified and collected from several reports (Føre and Lien, 2014, Føre and Thorvaldsen, 2017, Holmen et al., 2017a). The list serves as an input to interviews with one expert in the net cleaning operation, one captain and one TQM (Total Quality management) manager from a marine operation service company. A joint workshop was planned but could not be arranged due to schedule conflicts of the above participates. The interviews were structured focusing on the following questions:

- Are safety-critical parameters identified in the preliminary list important? If not, what else has the direct and significant influences to the different dimensions of risk (cf. Section 3.2)?
- What are the major hazardous events in each task, and what could be proactive and reactive barriers?

As a result, the preliminary list has been verified by field experts and revision has been made based on the feedback.

3 CASE DESCRIPTION

3.1 *Net cleaning operation*

The biofoulings can be a suitable habitat for sea lice larvae and may stop the oxygen supply to the farmed salmon due to reduced flow and water exchange inside and outside the net.

The cleaning operation is performed in various ways among service companies. In this case, a Remote Operated Net Cleaner (RONC) is used, and two operators are assigned to the job (Figure 1). The cleaner is remotely controlled by monitoring its movement from a screen in the wheelhouse onboard a service vessel.

The operation can be decomposed into five tasks: service vessel arrives at the cage, launch of RONC, a cleaning operation, recovery of RONC, and departure to the next cage or return to the port. The detailed procedures are described in Figure 2.

3.2 *Operational planning from risk perspective*

The risk dimensions that need to be considered in fish farming operations are risk to personnel, risk to material assets, risk to the environment (e.g., fish escape and pollution), risk to fish welfare, and food safety (Yang et al., 2017). Prevention of fish escape, which is considered as a threat to the wild fish population and environment, is an important focus of authorities. Personal safety is gradually gaining attention, ever since the aquaculture industry is realized to be one of the most dangerous professions in Norway (Aasjord and Geving, 2009). Fish health and fish welfare are other rising concerns, as sea lice have become a significant disease problem (IMR, 2017). The risk to marine assets has received

Figure 1. Net cleaning operation (from AQS website for illustration purpose).

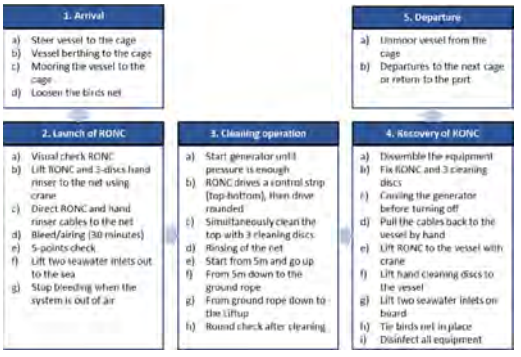


Figure 2. Stepwise procedure for net cleaning operation using RONC.

little attention. However, loss of service vessels or in a worst-case scenario, the fish farm, will give companies significant financial and reputation damage. The last dimension, food safety, is strictly controlled by the Norwegian Food Safety Authority and this aspect falls outside the consideration of this paper for short-term operational planning.

The exposed locations of the fish farms reduce the weather window and amplify the operational risk during net cleaning, especially for the personnel and the fish. The narrow weather window requires the operators to be well prepared, to foresee, and reduce the associated risks to be as low as reasonably possible while doing short-term planning.

4 APPLICATION OF THE SAFETY-CRITICAL PARAMETERS

4.1 *Safety-critical parameters*

Safety-Critical Parameters (SCPs) are the factors that have direct and significant influences on the risk involved in performing one activity (Yang and Haugen, 2016). They address possible failure mechanisms, and these are used to avoid active failures and latent errors that have major accident potential on an activity level (Yang and Haugen, 2018). The proposed generic parameters cover four aspects: input, control, process, and interaction. These SCPs are described shortly in Table 1. The readers are referred to Yang and Haugen (2018) for a full description.

Table 1. Safety-critical parameters categories descriptions adapted from (Yang and Haugen, 2018).

Aspect	SCP category	Short description
Input	Work instruction	Stepwise description of a job
	Operational limitations	e.g., weather, restricted area, restricted time, to ensure acceptable risk
	Plant drawings	Documents that demonstrate the technical systems
	Process input	Materials, tools, and spares that are consumed/used in the activity
Control	Proactive barrier systems	Barrier systems to prevent hazardous events from happening
	Recovery barrier systems	Barrier systems to reveal and recover from latent errors
	Reactive barrier systems	Barrier systems to mitigate the consequences of hazardous events
	Temporary barrier systems	Replacements for temporarily unavailable barrier systems
	Competence of operator—process model of the task	The competence of operators related to how to perform a specific task
	Competence of operator—process model of the system	The competence of the operators to interpret the current state of the controlled process
	Personnel exposure	The necessary number of workers.
Process	Physical *process* system	The physical system that the operator interacts
Interaction	Other concurrent control actions	Other on-going activities, which may have hazardous interactions with the activity.
	Environmental disturbance	Undefined or out-of-range environmental disturbance while performing the activity.

4.2 *Adaptation to net cleaning operation*

The original SCP framework focuses mainly on personal safety in the context of a hydrocarbon leak in oil and gas industry. When the framework is applied to aquaculture, it is essential to also consider fish welfare, risk to the environment (e.g., fish escape), and risk to marine assets. Further identification of special hazardous events that give rise to the above risk concerns (in addition to personal safety) is necessary. The SCP categories of barrier systems (i.e., proactive barrier systems, recovery barrier systems and reactive barriers systems) are removed from Table 1. Instead, the barrier systems are illustrated in connection with each hazardous event as shown in Figure 3 to better show their functions in preventing the hazardous events and mitigate the consequence.

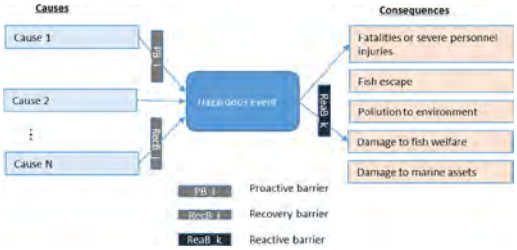


Figure 3. Illustration of a hazardous event, causes, consequences and barrier systems.

Another adaptation can be found under the category "process". The fish farming industry is dealing with live animals. During net cleaning, the salmon, and cleaner fish stay in the cage. So besides the "physical process system," another SCP; namely "fish", is added to emphasize the risk dimension of fish welfare.

5 RESULTS

The five tasks are combined into three because of their similarities: service vessel arrival and departure, RONC launch and recovery, and cleaning operation. This section describes the SCPs for these tasks in detail following the SCP categories proposed in Table 1 and Figure 3.

The possible hazardous events and their relevance to each task are listed in Table 2, with the primary concern indicated with *. These hazardous events are identified based on the findings from Holen et al. (2017a), Holen et al. (2017b), Føre and Thorvaldsen (2017), Føre and Lien (2014) and risk analysis document provided by the service company. Some hazardous events are relevant to several tasks. Therefore, the reactive barriers upon the occurrence of the hazardous event are illustrated in Section 5.4, instead of discussing them separately under each task.

5.1 *SCPs for service vessel arrival and departure*

The major hazardous event that may happen during service vessel arrival and departure is "service vessel collides/contacts with the fish farm". Figure 4 summarizes the possible causes, existing proactive (PB_i), recovery barriers (RecB_j) which are derived from discussions with the company and results in Yang et al. (2017).

Another hazardous event is "man overboard" while mooring the vessel to the cage. The anti-skid surface of the floating collar and shoes are the only existing barriers. To prevent oil spill (e.g., diesel oil, hydraulic oil) from the service vessel, the proactive barrier is also the 5-point check, which includes motor oil, hydraulic oil, coolant, seawater filter, and visual inspection.

Table 2. Possible hazardous events in net cleaning operation and their relevance to each task.

Hazardous event	Arrival and departure	RONC launch and recovery	Cleaning operation
Service vessel collides/ with fish farm	X*	X	
Holes in the net			X*
Oil spill from the service vessel	X	X	X
Dropping/swinging objects		X*	
Loss of growth of fish			X
Personal injury		X	X
Man overboard	X	X	X

*Primary concerned hazardous event.

5.2 *SCPs for RONC launch and recovery*

The primary hazardous event that may happen during launch and recovery of RONC using crane is "dropping/swinging objects." Figure 5 shows the possible causes, existing proactive (PB_i), and recovery barriers (RecB_j), which are derived from Holen et al. (2017b), and discussions with the company.

Another hazardous event could be the "holes in the net" due to entanglement between the hook and the net. However, this is relatively rare. "Man-over board" is also a possible hazardous event. Another concern is the "personal injury" while using the crane. The injuries include entanglement, crush, and back injuries (Holen et al., 2017a). In this case, no proactive barriers are set to prevent such injuries.

5.3 *SCPs for cleaning operation*

Figure 6 shows the primary hazardous event during the cleaning operation – "holes in the net" and possible causes and existing barriers, which are derived from Føre and Lien (2014), Føre and Thorvaldsen (2017), Jensen et al. (2010) and discussion with net cleaning operators.

"Personal injury" and "loss of growth of fish" are also relevant to this task. The various causes and barriers are described in Figure 7 and Figure 8 separately.

5.4 *Reactive barriers*

The hazardous events that are discussed above are further analysed to identify the barrier systems that aim to mitigate or reduce the possible consequences to the fish, the personnel, and the environment (Figure 9). The existing barrier systems are collected by interviewing the service company.

Table 3. SCPs for service vessel arrival and departure.

SCP category	SCPs
Work instruction	See Figure 2
Operational limitations *(limitations to arrival)*	Generic vessel limitations: stability, loading, maximum passenger, minimum crew
	Weather limitation: wind, wind direction, current speed, wave height, visibility
	Restricted mooring points
	Restricted time of operation
Plant drawings *(layout of the cage)*	Actual positions of mooring lines of the cage
	Actual positions of bridles of the cage and the net (due to deformation)
Process input *(Tools, materials, spare parts that are used)*	Condition of navigation system
	Condition of telecommunication system
	Condition of mooring line
Competence of operator—process model of the task *(Competence of completing arrival and departure)*	Competence of maneuvering the vessel
	Knowledge of influences of the weather condition to maneuvering: season, rain, wind, wave, current, fog, icing condition, darkness.
	Competence of safe mooring
Competence of operator—process model of the system *(knowing the vessel)*	Knowledge of the condition of the vessel; how the vessel reacts upon maneuvering
	Knowledge of the safety systems: emergency stop, rescue equipment, firefight system
Personnel exposure	Two operators in the wheelhouse
Physical process system *(The vessel itself)*	Technical condition of the vessel: propulsion system, power system, side propellers, etc.
	The disinfection status of the vessel
	The size of the vessel: length, width
Fish	Non-applicable
Other concurrent control actions	No other ongoing activities
Environmental disturbance	Unexpected big waves and strong current from, e.g., passing vessels
	Changed wind speed, wind direction

Table 4. SCPs for RONC launch and recovery.

SCP category	SCPs
Work instruction	See Figure 2
Operational limitations *(limitations in launch and recovery operation)*	Weather limitation: wind, wind direction, current speed, wave height, visibility
	Generic crane limitations: max. Lifting moment, capacity, slewing torque, etc.
Plant drawings *(layout of the cage)*	Bird net, the height of the railing
Process input *(Tools, materials, spare parts that are used)*	Technical condition of the crane
	Interface with the RONC: crane fastening, crane lifting straps
	Velocity of the crane tip
	Condition of the technical communication system
Competence of operator—process model of the task *(Competence of completing launch and recovery)*	Knowledge of the stability of the vessel: ballast tank, weight distribution, wind-exposed surface
	Knowledge of influences of the weather condition to launch and recovery: rain, wind, wave, current, fog, icing condition, darkness. Knowledge of operating crane: relative movements between the vessel and the cage
Competence of operator—process model of the system *(knowing the RONC)*	Knowledge of the condition of the RONC
Personnel exposure	One operator on the deck, one operator on the wheelhouse
Physical process system *(The RONC)*	The condition of the RONC: weight, size, disinfection status
Fish	Non-applicable
Other concurrent control actions	No other ongoing activities
Environmental disturbance	Unexpected big waves from, e.g., passing vessels
	Changed wind speed, wind direction

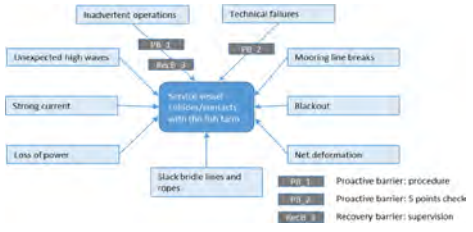


Figure 4. Causes of service vessel collision/contacts with the fish farm and existing barriers to prevent.

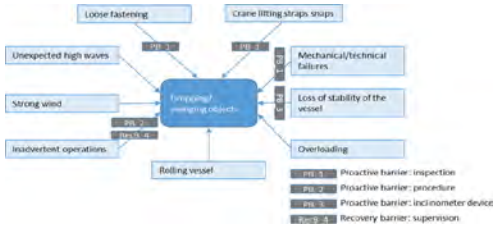


Figure 5. Causes of dropping/swinging objects and existing barriers to prevent.

Table 5. SCPs for cleaning operation.

SCP category	SCPs
Work instruction	See Figure 2
Operational limitations *(limitations in cleaning)*	Weather limitation: current speed, current direction, visibility underwater Generic RONC limitations: max. working water pressure, speed Strength of the netting Restricted time for operation
Plant drawings *(layout of the cage)*	Remaining systems in the cage Actual position of the net (due to deformation)
Process input *(Tools, materials, spare parts that are used)*	Technical condition of the equipment: RONC, high-pressure unit, inlets of the sea water, cables and hoses, power supply, sharp edges, cracks, loose screws RONC spare parts: discs, nozzles
Competence of operator—process model of the task *(Skill of cleaning)*	Knowledge of cleaning techniques Knowledge of RONC untangling and retrieve techniques Knowledge of disinfection techniques
Competence of operator—process model of the system *(know the RONC and the net)*	Knowledge of remotely maneuvering RONC and RONC behavior in different current speed and direction Knowledge of the status and actual position of the net Knowing the actual position of RONC during cleaning
Personnel exposure	One operator on the deck, one operator in the wheelhouse
Physical process system *(The net itself)*	The actual position of the net The condition of the net: type, slackness, age, degree of wear, existing holes, vulnerable parts (e.g., side rope). Possible items on the net: lost knives, fish hooks, mussels
Fish	Oxygen level, type of fouling, possible disease, possible hydroid
Other concurrent control actions	No other ongoing activities
Environmental disturbance	Unexpected big waves from e.g., passing vessels Changed wind speed, wind direction

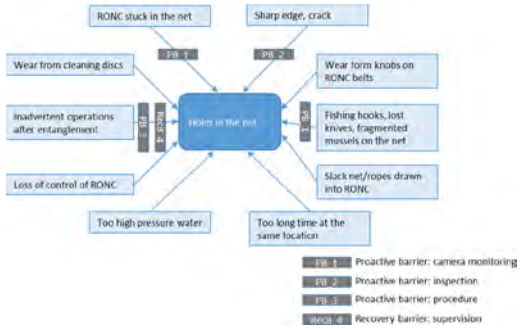


Figure 6. Causes of holes in the net and barriers to prevent.

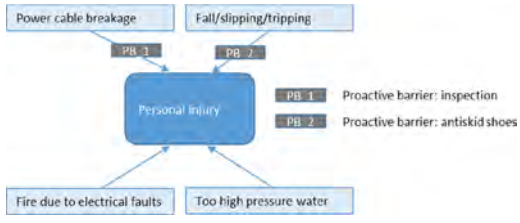


Figure 7. Causes of personal injury during cleaning and barriers to prevent.

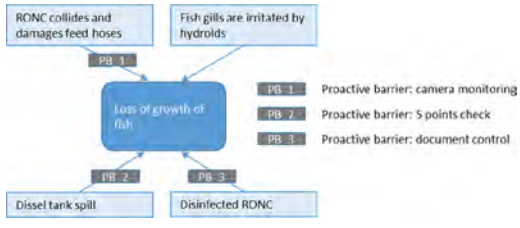


Figure 8. Causes of loss of growth of fish and existing barriers to prevent.

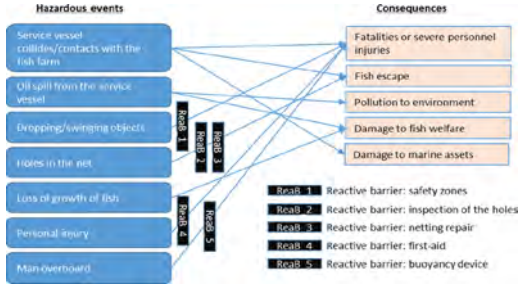


Figure 9. Reactive barrier systems of various hazardous events.

6 DISCUSSION

6.1 *Applicability of the SCP list*

In today's fish farm, it is a common practice to hire marine operation service companies to carry out various fish farming operations, such as net cleaning, delivery of feed, delousing and so on. Safe operations are heavily dependent on the experience of the operators. The rapid development of the industry follows a fast employment growth in these service companies. Concerns about employee inexperience arise from both the industry and the authorities. It is especially challenging to get skilled staff at exposed locations (Johannesen and Sæther, 2017). The industry is also experiencing technological innovations (Bjelland et al., 2015). The rapid emergence of new systems demands operators to learn about the system and operational techniques quickly to cope with unforeseen challenges. "Planning in details is the keyword," as reflected by Marine Harvest who has been running test facilities in exposed locations (Johannesen and Sæther, 2017).

The generic list of safety-critical parameters provides systematic guidance to identify the vital information that needs to be collected for a detailed operational planning. The detailed information can assist the decision-maker to approve the operation by considering the risk to personnel, environment and the fish. The information can also give the operators a reasonable anticipation of what may happen and what are or could be the proper reactions. The results from the case study show a significant number of critical operational risk factors that origin from experiences and understanding of the systems. These factors will also facilitate the training and learning process of junior operators to operate more safely.

6.2 *Challenges regarding using the SCP list*

The main challenge with using the SCP list lies in acquiring the information under each parameter. The service companies have specialized equipment, service vessel and expertise for performing the operations. They may not have enough knowledge about the specific site they are working on, such as the design of the cages, locations of mooring lines, systems equipped inside the net, the health status of the fish, and the possible presence of knives, fishing hooks and so forth. The above information could be available upon request, but a close cooperation and effective communication between the service companies and fish farmers are challenging, as revealed by Føre and Lien (2014). The listed SCPs can improve the communication by indicating clearly what information the service company is seeking.

It is also challenging to collect operational limitations, which is identified as one of the critical underlying causes of accidents by Norwegian Maritime Authority (NMA, 2017). It is not common for the service company to have written operational limits regarding weather conditions. It is subject to the captain's judgment and experience to decide whether the operation is feasible. Moreover, on some sites, the weather changes quickly, which makes it challenging to derive standard operational limits.

Another challenge is to understand the influence of the planned operation to fish welfare and fish health, which has not been paid much attention to yet. For instance, hydroid in the washed off waste damages the salmon gills (Saue, 2017). Gill diseases have been associated with reduced growth and large-scale mortality (Mitchell and Rodger, 2011). This may be out of the knowledge of the service company until symptoms become relatively severe. Acquiring such information requires a close follow-up of fish health after the operation and knowledge about the latest research findings.

6.3 *Applying barrier principle in aquaculture*

Using a barrier perspective as a strategy to reduce risk, especially fish escapes, gains more and more attention both from the industry and the government (Ministry of Trade Industry and Fisheries, 2017). The barrier perspective is one of the major accident causation theories behind the generic SCP list. Identification of barriers is an essential part in the case study, and the results show that in current fish farms, barrier systems have been partially placed and functioning, even though no explicit "barrier" label is given yet. For example, supervision is somewhat prevalent in the industry, as a response to the increasing use of junior operators. The senior operators on the site can correct the possible latent errors made by juniors during the operation. This correction is a typical recovery barrier. In fact, compared to supervision, certifications and procedures are not widely adopted practices in the service companies yet. The safe operations are heavily dependent on operator experience, from which the SCPs are derived. The results have several implications for implementing a barrier perspective in the industry in general.

First, clearly defined hazardous events are helpful to identify which barriers are in place, and which could be missing. Definition of hazardous events can be subjective, and this will influence what could be the proactive barriers or reactive barriers. For example, both holes in the net and fish escape can be defined as a hazardous event, but the reactive barriers will differ. ROV or diver

inspection and repair of the holes can be reactive barriers to "holes in the net", while escaped fish retrieve is the reactive barrier. It is beneficial to have standard definitions of hazardous events in the industry to facilitate a common understanding of the reactive barriers.

Second, the performance of the barrier systems should be evaluated in different operations. For instance, holes in the net are the most direct cause of fish escape in period 2010–2016 (Føre and Thorvaldsen, 2017). The reactions (reactive barriers) to the holes vary from operation to operation. Immediately after and during the net cleaning operation, the camera equipped with RONC can be used to inspect whether holes are present. In some cases, the RONC itself has been used as a temporary barrier in front of the hole to prevent fish escape. For other operations, such as delousing, ROVs or divers have to be sent out separately to do the inspection. Even though the inspection can be a reactive barrier in both operations to prevent fish escape, the response time in different operations is different. In turn, this influences the performance of the barrier systems, which affects potential consequences (Sklet, 2006).

Third, it is necessary to be aware that all kinds of functions, elements, and systems that are associated with safety can be given the label "barrier" (Rollenhagen, 2011). On one hand, there is a potential for setting up more effective proactive and reactive barrier systems in the case study. For example, absorbent boom/sheets can be equipped on board the vessel in case of oil spill. On the other hand, purely depending on barrier systems may not be enough for preventing accidents, even though operator's knowledge about all the SCPs (including operational limits) identified in the case study can be labeled as "barriers." Precautions should be taken when defining the barrier systems.

7 CONCLUSION

The case study in this paper illustrates that the proposed safety-critical parameters provide a systematic overview of the critical risk factors in specific fish farming operations for operational planning, considering the risk to personnel, fish escape, fish welfare and marine assets. The results also have implications to how to use barrier principles in aquaculture in general. The collected information by using SCP for fish farming operations can be used to improve the learning process of junior operators and to facilitate communication between the fish farmer and service companies.

The case study opens up the possibility of applying the SCP framework to other more complex operations in aquaculture. Furthermore, the priority of the SCP should be evaluated based on experiences from the senior operators, and relevant accidents and near misses. Prioritization of the risk factors is considered as further work in the following research.

ACKNOWLEDGEMENTS

This work has been carried out as part of the Reducing Risk in Aquaculture project. The Norwegian Research Council is acknowledged as the main sponsor of project number 254913. Heidi Moe Føre at SINTEF Ocean is acknowledged for their valuable input to this case study.

REFERENCES

Aasjord, H. & Geving, I.H. 2009. Accidents in norwegian fisheries and some other comparable norwegian industries. *IFISH 4 – The 4th International Fishing Industry Safety and Health Conference.* Iceland.

Bjelland, H.V., Føre, M., Lader, P., Kristiansen, D., Holmen, I.M., Fredheim, A., Grøtli, E.I., Fathi, D.E., Oppedal, F. & Utne, I.B. 2015. Exposed aquaculture in norway. *OCEANS'15 MTS/IEEE Washington.* IEEE.

Føre, H.M. & Lien, A.M. 2014. Report ivenstigation and measures against damage to the net during cleaning operation. Trondheim, SINTEF Ocean.

Føre, H.M. & Thorvaldsen, T. 2017. Causes for farmed salmon and trout escape during period 2010–2016. Trondheim, SINTEF Ocean.

Gibson, J.J. 1961. The contribution of experimental psychology to the formulation of the problem of safety—a brief for basic research. *Behavioral approaches to accident research.* London, Association for the Aid of Crippled Children.

Holen, S.M., Utne, I.B., Holmen, I.M. & Aasjord, H. 2017a. Occupational safety in aquaculture—part 1: Injuries in norway. *Marine Policy,* In press.

Holen, S.M., Utne, I.B., Holmen, I.M. & Aasjord, H. 2017b. Occupational safety in aquaculture—part 2: Fatalities in norway 1982–2015. *Marine Policy*.

Hollnagel, E., Woods, D.D. & Leveson, N. 2006. *Resilience engineering concepts and precepts,* Farnham, Ashgate.

Holmen, I.M., Lien, A.M., Fathi, D. & Ratvik, I. 2017a. Project note exposed: Hazards in aquaculture operations. Trondheim, SINTEF Ocean.

Holmen, I.M., Utne, I.B., Haugen, S. & Ratvik, I. 2017b. The status of risk assessments in norwegian fish farming. *Accepted for ESREL2017.*

Imr 2017. Risk assessment in norwegian fish farming 2017. Bergen, Norway., Institute of marine research.

Jensen, Ø., Dempster, T., Thorstad, E., Uglem, I. & Fredheim, A. 2010. Escapes of fishes from norwegian sea-cage aquaculture: Causes, consequences and prevention. *Aquaculture Environment Interactions,* 1, 71–83.

Johannesen, F. & Sæther, A. 2017. Marine harvest faroes, sandsvág experience and challenges. SFI Exposed, Trondheim 30–31. May 2017.

Leveson, N. 2012. *Engineering a safer world: Systems thinking applied to safety,* The MIT Press.
Ministry of Trade Industry and Fisheries 2017. Stragegy against escape in aquaculture. Ministry of Trade, Industry and Fisheries.
Mitchell, S.O. & Rodger, H.D. 2011. A review of infectious gill disease in marine salmonid fish. *J Fish Dis,* 34, 411–32.
Nma 2017. Focus on risks 2018.
Norwegian Directorate of Fisheries 2017. Overview of applications for development permits.
Perrow, C. 1984. *Normal accidents: Living with high risk technologies,* Princeton, Princeton University Press.
Pettersen, S. 2017. Risk management—learning from the petroleum industry. *Internal project.* DNV GL.
Rasmussen, J. 1997. Risk management in a dynamic society: A modelling problem. *Safety Science,* 27, 183–213.
Roberts, K.H. 1990. Some characteristics of one type of high reliability organization. *Organization Science,* 1, 160–176.
Rollenhagen, C. 2011. Event investigations at nuclear power plants in sweden: Reflections about a method and some associated practices. *Safety Science,* 49, 21–26.
Saue, O.A. 2017. Seeing the relationship between wash and gill disease. *iLaks.*
Sklet, S. 2006. Safety barriers: Definition, classification, and performance. *Journal of Loss Prevention in the Process Industries,* 19, 494–506.
Turner, B.A. 1978. *Man-made disasters: The failure of foresight,* London, Wykeham.
Yang, X. & Haugen, S. 2016. Risk information for operational decision-making in the offshore oil and gas industry. *Safety Science,* 86, 98–109.
Yang, X. & Haugen, S. 2018. Implications from major accident causation theories to activity-related risk analysis. *Safety Science,* 101, 121–134.
Yang, X., Utne, I.B. & Holmen, I.M. Submitted. MIMACHE: A methodology for the identification of major accident hazards and hazardous events in norwegian aquaculture. *Submitted to Safety Science*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Using microworlds to study critical infrastructure protection—the effect of incentives on risk management

H. Tehler
Centre for Critical Infrastructure Protection Research (CenCIP), Lund, Sweden
Lund University, Lund, Sweden

J. Lindström & H. Lindbom
Lund University, Lund, Sweden

ABSTRACT: A computer simulated microworld was used to study risk management in a critical infrastructure context. 36 students assumed the role of an electric distribution company Chief Executive Officer (CEO) making decision on how to spend resources between investments in risk reduction and other investments. We studied the effect of differences in terms of the incentives to invest in risk reduction and the extent that the participants had access to a risk assessment. We found that both independent variables influenced the total resources spent on risk reduction and total losses due to storms. If the company had to bear a larger part of the total losses due to a storm their propensity to invest in risk reduction increased. In addition, if the participants were provided with a simple risk assessment to support their decisions they invested more resources in risk reduction, and they were also more successful in limiting the total losses due to storms.

1 INTRODUCTION

The functioning of modern societies is dependent on the services provided by an interconnected web of critical infrastructures (CI:s). Although what counts as a CI might differ between countries, telecommunication, electric power, transportation and water supply systems are usually considered to be CI:s. Such systems can have a considerable geographic extent, sometimes covering entire nations, or even continents.

Two trends in the development of CI:s are especially important for the ability to manage risks associated with them. First of all, they are becoming increasingly interconnected growing into so called "system of systems" (Kröger, 2008, OECD, 2011) and thereby increasing the risk of transboundary crises, i.e. crises that "affect multiple jurisdictions, undermine the functioning of various policy sectors and critical infrastructures, escalate rapidly and morph along the way" (Ansell et al., 2010). Thus, what might previously have resulted in a local disruption of some services might nowadays spread quickly and grow into a full-blown crisis.

Secondly, the planning, operation and management of these systems have become increasingly fragmented (Almklov & Antonsen, 2010, de Bruijne & van Eeten, 2007). A system that used to be run by one organization can nowadays be managed by a multitude of different actors. Thus, at the same time as the systems are growing more interconnected, the management of them are becoming more fragmented (de Bruijne & van Eeten, 2007).

When CI:s and their associated risks are becoming increasingly connected it is no longer viable to approach risk management in a "silo-based" fashion, focusing on one actor and one system at a time. This conclusion has influence the corporate world (Gordon et al., 2009) as well as in governmental efforts to manage risk through so-called 'all-hazards' and 'whole-of-society' approaches (OECD, 2009).

Notwithstanding the fact that there are efforts to manage risk from a more holistic perspective, it is very difficult to know if approaches such as the ones exemplified above actually leads to more effective management of risk (Rivera et al., 2017, Hoyt & Liebenberg, 2011). Therefore, there is need to better understand risk management in general, and specifically in an interconnected world.

There are many different methods one can use to try to increase our knowledge of risk management. Here we present the initial results from a study making use of a method that has not been used much in the present context; a computer simulated

microworld. The focus of the study is on two aspect that can influence risk management: differences in incentives for investments in risk reduction and the extent to which risk assessments are available when making such investment decisions.

2 USING MICROWORLDS TO STUDY RISK MANAGEMENT

There exists a multitude of definitions of risk (see e.g. Aven and Renn, 2009), and there are also several definitions of risk management. For example, the SRA Glossary defines it as "Activities to handle risk such as prevention, mitigation, adaptation or sharing" (Society for Risk Analysis, 2015), and it is also noted that risk management "…often includes trade-offs between costs and benefits of risk reduction…". Other definitions, like the ones suggested by International Organization for Standardization and United Nations International Strategy for Disaster Reduction (ISO, 2009, UNISDR, 2009) also emphasize the fact that risk management is about doing something, e.g. prevent, mitigate, etc., in order to achieve something, i.e. the handling of risk.

To study risk management is thus about studying a purposeful activity, or a number of activities. Such studies can be conducted in many different contexts where the focus might be on one person, a group, an entire organization, or even multiple organization. And they can also be performed within different scientific disciplines, have different ambitions, and be based on different theoretical frameworks. It might, for example, involve field studies focusing on organizations (e.g. research on High Reliability Organizations, as described by Roberts (1990)), or laboratory research focusing on individual decision making (e.g. Tversky and Kahneman, 1974).

Here the focus is on using a computer simulation, a so called microworld to study risk management. Microworlds have been used for some time as a means to capture some of the complexity of real life problem contexts. In a micorworld-study the participants interact with a computer simulation and receive continuous feedback on the effect of their actions. Brehmer & Dörner (1993) describes the use of microworlds as a way to escape "…both the narrow straits of the laboratory and the deep blue sea of the field study". They refer to the fact that laboratory experiments often get criticizes by field researchers for lacking relevance, and vice versa that laboratory researchers criticize field researchers for lack of control of confounding variables. A micorworld is something in between a laboratory experiment and a field study. It retains the possibility to control the variables under study, but at the same time offer more realistic conditions in terms of task complexity.

In terms of realism, there are some clear characteristics of real risk management problems that microworlds are especially suitable to capture. Risk management usually require *a series* of *dependent* decisions. For example, a decision to invest in risk reduction will be influenced by previous investment decisions. The previous decisions determine, to a certain extent, the present level of risk. And, a lower level of risk will, in general, make it more difficult to find good investments to decrease the level of risk more (decreasing marginal utility of investments). Thus, the decisions that led up to the present situation (risk level) all have an influence of the current decisions.

Moreover, the risk management problem will *continuously change*. Both as a consequence of the actions taken by the risk manager, but also due to other factors. These characteristics are identical to those described for dynamic decision problems (Edwards, 1962, Brehmer, 1992).

In terms of problem complexity, microworlds allow us to extend the study of risk management from using lottery-like situations where the focus is on the participant's *judgements*, and how they make them, to more complex game-like situations where the focus is on *control*. To frame risk management as a control problem is not new but has been proposed as a prerequisite for understanding it in a dynamic society (Rasmussen, 1997). In the microworld context, control refers to "…an attempt to achieve some desired state of affairs" (Brehmer, 1992). To try to achieve control in the present context of risk management means to achieve a suitable balance between the costs and benefits of risk reduction. Clearly, judgements, for example about the likelihood of various events, are an important part of trying to achieve control. But, there is much more to the control problem than merely judgements.

Yet another feature of microworlds that makes them suitable for studying risk management is the fact that one can study how successful participants are in trading off the costs of risk reduction investments with their associated benefits. This is usually difficult to achieve in a field study, especially if one focus on low-probability high-impact events (Rivera et al., 2017). Neither the simple experiments lacking the dynamic nature of microworlds can adequately capture it.

In a critical infrastructure context, negative consequences can be described in terms of failure to maintain vital societal functions (see for example the definition of CI in the EU council directive 2008/114/EC). Thus, a relevant risk management problem is how successful the operators of CI:s are at trading off the costs of investments in risk

reduction versus the benefits the investments provide in terms of reduced occurrence and/or consequences of failures of vital societal functions. However, due to the interconnected nature of CI:s, and their potential devastating societal consequences, an equally interesting question is how their tradeoffs affect the consequences suffered by the *users of the services*. Part of the study presented here is aimed at investigating such consequences.

3 A TWO-LOOP MODEL OF RISK MANAGEMENT

One of the most straightforward control-loop models of risk management one can use is a model focusing on the relationships between risk management decision and losses/consequences (measured in a suitable way) as illustrated in the left part of Figure 1. Such a model ignores many important aspects of risk management, for example, how people make decisions (as for example studied in the Naturalistic Decision Making or the Heuristics & Biases traditions, see Kahneman & Klein (2009) for a review of the two traditions), or why and how people identify something as a risk (see for example Boholm & Corvellec (2011)). Nevertheless, it captures some of the important dynamics of risk management, i.e. past decisions may influence future ones, and it can be used as a basis for studying differences in risk management arising from different incentives to invest in risk reduction. For example, it is reasonable to assume that if the negative consequences an actor might suffer due to a specific activity are reduced, the actor's interest in investing in risk reduction will also be reduced.

The one-loop model focuses on an actor's *concrete experiences* in terms of investments in risk reduction, which might lead to a change in the system of interest, which then might lead to less negative consequences. If the likelihood of experiencing losses is relatively high one would expect the control problem to be relatively easy for a risk manager. He/she continuously gets feedback on the effect of various investment in risk reduction and can therefore find a balance between the cost of investments and the benefits they bring. However, if the focus is on low likelihood/high consequence events the problem is not so easy. Often there will be no losses at all, and therefore one might question the effectiveness of investments. Similarly, when severe losses eventually occur, it is very difficult to determine if the investments made so far had any effect on the outcome. In general, less frequent events generating losses will mean that the control loop is weaker.

Figure 1. Illustration of the two-loop model of risk management.

The one-loop model relies on what *has previously happened* in a system, but from a risk management perspective *what can happen* in the system is also very important. The question of what can happen together with judgements of how likely it is and the consequences thereof is an important part of risk assessment (Kaplan & Garrick, 1981, Aven, 2010). Therefore, we can complement the one-loop model with another loop that connects the risk management decision to a *description of risk* (Aven, 2010). The implication of this second loop is that a decision to invest in risk reduction can be influenced by a description of risk from the system of interest (e.g. in the form of a risk analysis). The decision can, in turn, influence future state of the system, which can then give rise to a new description of risk, and so on.

Thus, this second loop focuses on *what might happen* in the system of interest. The description of risk will most likely also be influenced by previous consequences in the system (and perhaps similar systems), but that relationship is not explicitly illustrated in Figure 1. Moreover, the two loops represents two types of analytical processes previously termed experiential and analytic processing (Marx et al., 2007).

The two-loop model is the basis for the experiments presented below. In a CI context, it describes the decisions that a CI operator might make to invest in risk reduction, and how the investment might influence the infrastructure (state of the system), which might then influence the occurrence and/or magnitude of negative consequences due to unwanted events (depending on the CI it might for example be storms, technical failures, floods, etc.). It also describes how the same decisions and following change of the CI might influence future descriptions of risk, which might in turn influence future decisions.

The "other factors" are included in the model as a reminder that risk management is not a closed system and that there are many other factors than the ones included here that influence risk management in practice. For example, there are many other factors that influence the state of a CI system apart from investment in risk reduction, such as various performance objectives of the CI operator.

4 THE EXPERIMENTS

4.1 *Overview*

The aim of the experiment was to investigate the effect of differences in incentives to invest in risk reduction, and if such effects could be influenced by the extent that risk assessments are available to support the investment decisions. The importance of incentives for risk management in CI:s is salient in many sectors where CI:s are operated. A specific form of incentives is related to the two-loop model presented above. It concerns the magnitude of the negative consequences affecting a CI operator in case of a failure to maintain the vital societal function in question (e.g. electric power supply). One can imagine two extreme situations in this respect. In practice the situation for most CI:s are probably somewhere in between the extremes. On the one hand, one can consider a situation where an operator of a CI does not suffer any negative consequences at all due to a failure to uphold the function of interest. Clearly, there are negative consequences of such a failure, but they will solely affect the users of the function in question. On the other hand, one might have a situation where the CI operator will have to compensate the users of the function to cover for the full consequences of a failure. The net consequences suffered by the users would then be zero and the CI operator will have to bear the full consequences of the failure.

In practice, it is difficult to imagine how these two extreme situations could occur. It would, for example, be very difficult to determine the compensation to the user of a function to compensate for "all" their losses. Nevertheless, the focus here is not to simulate real cases but rather to try to capture some of the important mechanisms influencing risk management in practice and investigate if they can be influenced in different ways.

4.2 *The microworld*

A microworld called MicroRisk was designed as part of several master thesis projects at Lund University. MicroRisk is a computer simulation that allows the participants in the experiment to interact with it using a regular computer and a web-browser. To implement the two-loop model of risk management in a CI context the particular study reported here used MicroRisk to put the participants in charge of a fictitious electric power company. A detailed description of the study, including the user interface, can be found in (Lindström, 2017). MicroRisk was run for a number of turns and after each turn the participants were asked to prioritize between investments to reduce the risk associated with the supply of electricity (e.g. strengthen grid, invest in repair capability, spare parts, etc.), and other investments. The choice was made by dividing 100 resource units between "other investments" and "risk reduction investments". Resource units (or simply units) was the way consequences (losses) and investments were measured in the simulation. Each turn there might be a serious storm affecting the power grid causing negative consequences (measured in units). The ultimate goal of the participants was to maximize the number of units spent on other investments minus the units lost due to storms.

The consequences of a storm were determined by the amount of resources the participants spent on risk reduction. The more they spent, the less the consequences of storms. This function was implemented in the MikroRisk by having the "state of the system" (Figure 1) being represented by one variable called *Level of risk*. The *Level of risk* was a number between 1 and 1000. At the start of each game the level of risk was set to 500. Each unit the participant would spend on risk reduction would then reduce one unit from the *Level of risk*. However, each turn 50 units were added to the *Level of risk* independent of the decisions made by the participants. It represents the fact that a system that is not maintained will degrade over time. Thus, a participant that chose not to spend anything on risk reduction would quickly degrade the system increasing the consequences due to storms significantly. It was only the potential consequences that were affected when changing the *Level of risk*, not the likelihood of a storm occurring. The maximum negative consequences that could occur, was a loss of 1000 units and the minimum was 1 unit. 1000 units would be lost if a storm occurred when *Level of risk* was at its highest value, and 1 units would be lost if it was at its lowest value. However, the relationship between *Level of risk* and the negative consequences was not linear. Investments when the *Level of risk* was close to the starting value of 500 was more effective than investments made when the Level of risk approached low values (diminishing marginal utility of investments).

Games were played during 30 turns. After the 30th turn of a game there was a 20% probability that it would end after each turn. Most games ended in between 30 and 40 turns, but in the analysis below we only use the 30 turns that were completed in all games. The participants did not know how many turns they were going to play. Neither did they know the probability of a severe storm each turn. Instead they had to infer the likelihood based on previous experience (left side of the loop in Figure 1). The number of storms during the thirty turns was set to four. In each game, they were randomly assigned to occur in four of the thirty turns.

There were four versions of the game played by the participants. The four versions differed with respect to two factors each having two possible states: *Incentives* (*Limit*/*No limit*) and *Risk assessment* (*RA*/*No RA*). The incentive to invest in risk reduction was varied by having two possible game-types. One type (*No Limit*) where the power-company would bear all losses due to storms and the users none. The other type (*Limit*) meant that there was an upper limit to the losses (200 units) that the power-company had to bear. Above 200 units, the company would only have to bear 15% of all exceeding losses (the remaining 85% of the losses affected the customers of the power company). The *Total losses* is the sum of the losses (in terms of units lost) affecting the power company *and* the customers. The *Total losses* is not dependent on whether incentives are limited or not, it is only dependent on the state of the system (*Level of risk*), which is only affected by the decisions made by the participants.

The *Risk assessment* factor was varied by either presenting a risk assessment to the participants to support their decisions (*RA*) or not (*No RA*). Thus, when no risk assessment was present, the loop on the right in Figure 1 was absent. The risk assessment was a simple text of the form "Given the present level of preparedness experts estimate the consequences of a severe storm to be a loss of somewhere between *X* and *Y* units". The interval between *X* and *Y* was arrived at by first calculating the true consequences, C_{true} (the consequences that would occur during a turn *if* a storm occurred) given the present *Level of risk*. Then a new value C_{est} was calculated by random in the interval [0.8 * C_{true}, 1.2 * C_{true}]. Then the interval between *X* and *Y* was calculated by multiplying C_{est} by a factor *k* ($X = C_{est} - (C_{est} * k)$, $Y = C_{est} + (C_{est} * k)$). The factor *k* was a random number (drawn every turn) between 0.1 and 0.5. Thus, the risk assessment would show the participants a relatively wide interval in the vicinity of C_{true}.

4.3 *The participants*

Participants consisted of students at Lund University that were recruited to the experiment by posting information about the experiment at different parts of the university. The 30 first to enroll in the experiment was given a movie ticket. There were 36 students that participated in the experiment. 16 (44,4%) females and 19 (52,8%) men (one person did not want his/her sex to be recorded). The students were predominately studying (or had just graduated) from engineering programs (22), but there were also students from biology (4), medicine (3), psychology (1), audiology (1), speech therapy (1), nursing (1), systems science (1), graphical design (1), and criminology (1).

4.4 *Procedure*

We used a fully crossed, 2 *Risk assessment* (*No RA*/*RA*) × 2 *Incentive* (*No Limit*/*Limit*), within-subject design. Thus, each participant played all four versions of the game. The order in which the games were played was randomized to minimize learning effects. The participants played the game using one of three computers in a closed off room at the Division of risk management and societal safety at Lund University.

In each version of the game we measured how much resources each participant spent on investments in risk reduction (*Total investments in risk reduction*), how much losses the company and the customers suffered (*Total losses*).

5 RESULTS

Three of the students did not complete all versions of the game. Their data was removed from the results. A two-way analysis of variance was conducted on the influence of the two independent variables (*Incentive*, *Risk assessment*) on the each of the dependent variables (*Total losses*, *Total investments in risk reduction*).

5.1 *Total losses*

The mean value for *Total losses* in game condition 1 (No RA, No Limit) was 1448 (*SD* = 1181), in condition 2 (RA, No Limit) it was 1183 (*SD* = 856), in condition 3 (No RA, Limit) it was 2004 (*SD* = 1324), and in condition 4 it was 1583 (*SD* = 1125). The results are illustrated in Figure 2 where a box-plot shows the median value and the 25th and 75th percentiles respectively (whiskers show the most extreme data points). The main

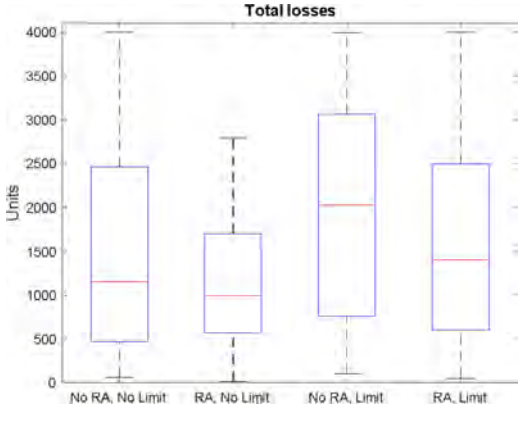


Figure 2. Box plot showing *Total losses* for all four game types.

effect for *Incentive* was significant $F(1, 32) = 7.47$, $p < 0.05$, as was the main effect for *Risk assessement* $F(1, 32) = 5.58$, $p < 0.05$. The interaction effect was not significant, $F(1,32) = .25$, $p = 0.62$.

5.2 *Total investments in risk reduction*

The mean value for *Total investment in risk reduction* in condition 1 (*No RA*, *No Limit*) was 1583 (*SD* = 354), in condition 2 (*RA*, *No Limit*) it was 1596 (*SD* = 254), in condition 3 (*No RA*, *Limit*) it was 1320 (*SD* = 475), and in condition 4 (*RA*, *Limit*) it was 1535 (*SD* = 382). The results are illustrated in Figure 3.

The main effect for *Incentive* was significant $F(1, 32) = 5.24$, $p < 0.05$, as was the main effect for *Risk assessement* $F(1, 32) = 7.42$, $p < 0.05$. The interaction effect was also significant, $F(1,32) = 6.45$, $p < 0.05$.

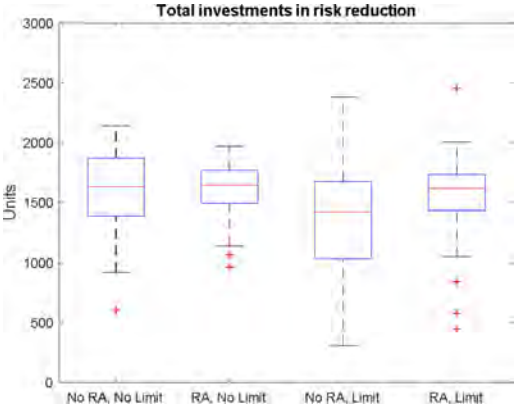


Figure 3. Box plot showing *Total investment in risk reduction* for all experimental conditions.

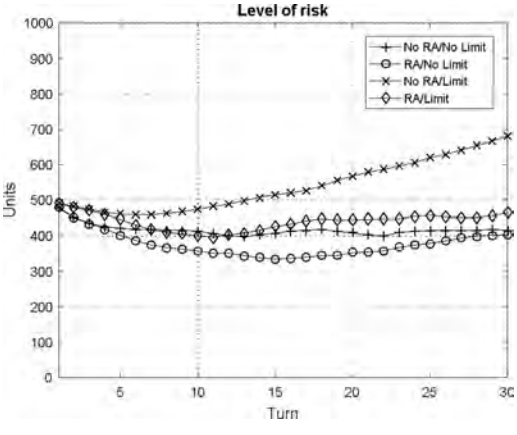


Figure 4. Mean value of the *Level of risk* for all participants.

5.3 *Level of risk*

During the games, we measured how the *Level of risk* varied. In Figure 4 the mean value of all participants *Level of risk* at different turns and during different versions of the game is shown.

6 ANALYSIS AND DISCUSSION

The results from the study show that both independent variables (*Incentive* & *Risk assessment*) have a significant effect on both *Investments in risk reduction* and on *Total consequences*. Taking the perspective of the *users* of vital societal functions, i.e. all actors dependent on the service in question and the public, it is especially interesting to note that the introduction of a simple risk assessment increased the investments in risk reduction. The increase was so strong that it essentially cancelled out the negative effect of reducing the participants' incentive to invest in risk reduction. The mean value of *Total investments in risk reduction* is essentially the same for the cases (*No RA*/*No Limit*), (*RA*/*No Limit*) & (*RA, Limit*) in Figure 3, whereas the mean value for (*No RA, Limit*) is significantly lower.

The result suggest that one can influence CI operators (and others) to invest more in robust services without using coercive measures, such as detailed regulations. Instead, one can influence their behavior in a positive way (for the users of the services) by introducing a risk assessment. This effect might be especially important in CI:s where the consequences due to severe interruptions of services are small for the operator compared to those suffered by the users of the services.

Moreover, while the results presented in Figure 2 illustrate the rise in *Total consequences* due to a reduction of incentives to invest in risk reduction (an expected effect) it also shows something interesting from a user of service-perspective. It shows that the introduction of a risk assessment not only led to more investments in risk reduction, it also had a significant effect on the *Total consequences*. Thus, the participant not only spent more on risk reduction, the investments were also effective in reducing *Total consequences*. One might assume that the reduction of *Total consequences* stem only from the fact that the participant spends more resources on risk reduction. In part, that is true, but comparing Figure 2 (*No RA*/*No Limit*, and *RA*/*No Limit*) to Figure 3 (*No RA*/*No Limit*, and *RA*/*No Limit*) also suggest that even without an increase in resources spent on risk reduction

the introduction of a risk assessment leads to less *Total consequences*. At least in contexts where the CI operator have to bear a large part of the negative consequences due to a service interruption.

Microworld experiments allow researchers focusing on risk management to study how the state of a system, and the risk associated with it, change over time (turns). Thus, using microworlds one can study risk management phenomena where longitudinal studies of the state of a system is crucial. One example is phenomena associated with a gradual increase in risk that goes unrecognized resulting in a system drifting into failure (Dekker and Pruchnicki, 2013). Figure 4 shows the mean value of the *Level of risk* in the four versions of the microworld. It is clear that one of the versions stands out from the rest in terms of a gradual increase in risk (*No RA*/*Limit*). A rising *Level of risk* due to economic pressure (in this case the incentive to select "other investments") can be counteracted by one or both of the two loops in Figure 1. The left side loop counteracts this tendency by concretely making the participants experience that they pay a price for allowing the *Level of risk* to become high in that the losses when a storm occurs also becomes high. However, when there is a limit to the losses suffered by the power-company, the concrete consequences felt by the participants are smaller compared to when there are no such limits. Thus, the left side loop is *weaker* in countering the economic pressure (that is the same in all cases) than when there are no limits.

The right-side loop counteracts the economic pressure by making clear for the participants what the consequences *can* be if a storm occurs. Although, this might not be as important as concretely experienced losses, it has the advantage of being present all the time (during all turns) as opposed to the concrete losses due to storms that is only experienced in four of the thirty turns. This makes it easier to use the risk description as a basis for the risk management tasks. An increase in investments in risk reduction would, in this case, immediately be visible in the risk description presented during the next turn. Thus, the participant would get a confirmation, although somewhat uncertain, that the investments have an effect.

However, in the (*No RA*/*Limit*) condition of the experiment this feedback is lacking and the right-loop is thus in its weakest form to counter the economic pressure. Moreover, in that condition the left-loop is also in its weakest form and that is an explanation offered by the two-loop model as to why the *Level of risk* is almost constantly increasing in one of the conditions in Figure 4, whereas in the other conditions it seems to be stabilized on much lower levels.

The relevance of the results to real problems of CI risk management is unclear and needs to be further studied. Nevertheless, based on the results it seems reasonable to focus on finding practical situations of CI risk management problems that might differ with respect to the conditions investigated here. Thus, we should look for situations where there are differences in terms of the consequences suffered by the *provider* of a vital societal function in case of a failure to supply it (left loop). And we should also look for situations that differ with respect to how often failure of functions occur (left loop). Moreover, situations differing with respect to how (if) risk analyses are used as a basis for decision making should also be identified. Based on further detailed study of such practical situations one might be able to improve the microworld studies linking them better to the practical cases and investigating whether the effects seen in the microworlds are also present in practice.

7 CONCLUSIONS

We have investigated the effect of different incentives for investing in risk reduction, and the effect of risk assessments, on risk management in a computer simulated microworld. The focus was on a situation where one actor supplies a service or function that is used by other actors. The risk management problem is related to how much the supplying actor will spend on risk reduction, focusing on protecting the service in question, given that the actor will suffer a certain amount of negative consequences in the event of a service failure.

Both factors had a significant effect on the dependent variables of the study. Manipulating the incentives to invest in risk reduction by changing the negative consequences suffered in case of a service failure affected both the amount of resources invested in risk reduction, and the total losses due to service interruptions. Moreover, a similar effect was observed when the extent to which the actor supplying the service in question had access to a simple risk assessment changed. Having access to a risk assessment (compared to not having such an access) resulted in more resources spent on risk reduction, and also in less total losses due to service interruptions.

Thus, the results indicate that risk management actions can be influenced in a predictable way just by providing access to simple risk assessments. Supplying a decision maker with a risk assessment in the experiments lead not only to an increased propensity to invest in risk reduction it also resulted in more well-balanced decisions that ultimately led to less total losses. The addition of a risk assessment can counter the effect of reducing the incentives to

invest in risk reduction. This effect is especially clear when the incentives to invest in risk reduction is low.

Although the practical implications of the results presented here are not clear, the results indicate that incentives and the extent to which risk assessments are used (and how they are designed) are two important factors for understanding risk management.

ACKNOWLEDGEMENTS

This research has been financed by the Swedish Civil Contingencies Agency (MSB).

REFERENCES

Almklov, P. & Antonsen, S. 2010. The Commoditization of Societal Safety. *Journal of Contingencies and Crisis Management* 18 (3): 132–144.

Ansell, C., Boin, A. & Keller, A. 2010. Managing Transboundary Crises: Identifying the Building Blocks of an Effective Response System. *Journal of Contingencies and Crisis Management* 18 (4): 195–207.

Aven, T. & Renn, O. 2009. On risk defined as an event where the outcome is uncertain. *Journal of Risk Research* 12 (1): 1–11.

Aven, T. 2010. On how to define, understand and describe risk. *Reliability Engineering & System Safety* 95 (6): 623–631.

Boholm, Å. & Corvellec, H. 2011. A relational theory of risk. *Journal of Risk Research* 14 (2): 175–190.

Brehmer, B. & Dörner, D. 1993. Experiments With Computer-Simulated Microworlds: Escaping Both the Narrow Straits of the Laboratory and the Deep Blue Sea of the Field Study. *Computers in Human Behavior* 9: 171–184.

Brehmer, B. 1992. Dynamic decision making: Human control of complex systems. *Acta Psychologia* 81: 211–241.

De Bruijne, M. & Van Eeten, M. 2007. Systems that Should Have Failed: Critical Infrastructure Protection in an Institutionally Fragmented Environment. *Journal of Contingencies and Crisis Management* 15 (1): 18–29.

Dekker, S. & Pruchnicki, S. 2013. Drifting into failure: theorising the dynamics of disaster incubation. *Theoretical Issues in Ergonomics Science* 15 (6): 534–544.

Edwards, W. 1962. Dynamic Decision Theory and Probabilistic Information Processing. *Human Factors* 4: 59–73.

Gordon, L.A., Loeb, M.P. & Tseng, C.-Y. 2009. Enterprise risk management and firm performance: A contingency perspective. *Journal of Accounting and Public Policy* 28 (4): 301–327.

Hoyt, R.E. & Liebenberg, A.P. 2011. The Value of Enterprise Risk Management. *Journal of Risk and Insurance* 78 (4): 795–822.

ISO 2009. *Risk management—principles and guidelines* (ISO 31000:2009). Geneva: International Organization for Standardization.

Kahneman, D. & Klein, G. 2009. Conditions for intuitive expertise: a failure to disagree. *Am Psychol* 64 (6): 515–26.

Kaplan, S. & Garrick, B.J. 1981. On The Quantitative Definition of Risk. *Risk Analysis* 1 (1): 11–27.

Kröger, W. 2008. Critical infrastructures at risk: A need for a new conceptual approach and extended analytical tools. *Reliability Engineering & System Safety* 93 (12): 1781–1787.

Lindström, J. 2017. *Incitament för investering i riskreduktion*, Report 5041. Lund University, Division of risk management and societal safety (in Swedish).

Marx, S.M., Weber, E.U., Orlove, B.S., Leiserowitz, A., Krantz, D.H., Roncoli, C. & Phillips, J. 2007. Communication and mental processes: Experiential and analytic processing of uncertain climate information. *Global Environmental Change* 17 (1): 47–58.

OECD 2009. *Innovation in country risk management*. Paris: Organisation for Economic Co-operation and Development.

OECD 2011. *Future Global Shocks: Improving Risk Governance*. Paris: Organisation for Economic Co-operation and Development.

Rasmussen, J. 1997. Risk management in a dynamic society: a modelling problem. *Safety Science* 27 (2): 183–213.

Rivera, C., Wamsler, C. & Tehler, H. 2017. Evaluating the Performance of Disaster Risk Management Systems—Is It Possible? In: Madu, C.N. & Kuei, C.-H. (eds.) *Handbook of Disaster Risk Reduction & Management*. Singapore: World Scientific Publishing.

Roberts, K. 1990. Some Characteristics of One Type of High Reliability Organization. *Organization Science* 1 (2): 160–176.

Society for Risk Analysis (2015) *SRA Glossary*. Society for Risk Analysis, (retrieved from http://sra.org/resources on the 28th of November 2017).

Tversky, A. & Kahneman, D. 1974. Judgement under Uncertainty: Heuristics and Biases. *Science* 185 (4157): 1124–1131.

UNISDR 2009. *UNISDR Terminology on Disaster Risk Reduction*. Geneva: United Nations International Strategy for Disaster Reduction.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Lessons learned from an unexpected uranium accumulation event

D.G. Harrison & A. Smith
U.S. Nuclear Regulatory Commission, Rockville, Maryland, USA

ABSTRACT: On July 14, 2016, a nuclear fuel fabrication facility licensee notified the United States Nuclear Regulatory Commission (NRC) that significant amounts of uranium were discovered, potentially exceeding their Criticality Safety Evaluation (CSE) mass limits, during an annual inspection of a scrubber ventilation system. The licensee subsequently confirmed not only significant mass several times higher than the CSE mass limits in the scrubber and associated ventilation ductwork, but also significant concentrations of uranium. As part of the NRC's platform of continuous improvement, a lessons-learned activity was initiated to explore opportunities for improving the NRC's regulatory processes for early identification of facility operational issues and preventing such events in the future. This paper describes the event, some of the licensee's root causes that led to this event, some of the reasons why the NRC did not identify this condition (and similar conditions at this and other facilities) through its regulatory processes prior to the event, and the improvements being considered to enhance these NRC regulatory processes.

1 THE EVENT

1.1 *What happened?*

On May 28–29, 2016, a licensee conducted an annual inspection and cleaning of their scrubber ventilation system. The scrubber is one of the main air scrubbers for the nuclear fuel conversion process and is connected to numerous processes. When the scrubber ventilation system was inspected and cleaned, a large mass of material was found inside the large attached ventilation ducting and subsequently within the scrubber body itself. At the time, the licensee believed that the uranium concentration of the material removed from the scrubber was low. The licensee removed the material and sent samples for analysis of the composition. The licensee received the results of the initial lab analysis on May 30, 2016, which indicated a significant concentration of uranium. The licensee did not fully consider the results and restarted operation of the system. Over a month later, on July 13, 2016, the licensee received the results of additional lab analyses that confirmed the earlier results indicating that the concentration of uranium was almost fifty percent (50%) and significantly exceeded the Criticality Safety Evaluation (CSE) mass limit for the process. The licensee reported the event to the US Nuclear Regulatory Commission (NRC) on July 14, 2016.

1.2 *Why is this important?*

This event did not result in a criticality. However, because there were no physical controls or measures available to prevent a criticality (i.e., all controls and measures failed to prevent the accumulation of uranium significantly above the CSE mass limit), this event represented a significant safety concern. The subsequent discovery of similar conditions at this and other fuel fabrication facilities has reinforced the need to address the concerns and weaknesses raised by this event in both the licensees' and regulator's processes.

2 ROOT CAUSES FOR THE EVENT

2.1 *What led up to this event?*

Throughout a period of more than a decade before the event, a combination of process changes, analysis assumptions, and operational approaches created the environment for this uranium accumulation event, including the licensee's slow response, poor decision making, and delayed reporting.

In 2002 this scrubber replaced another scrubber and over a number of years ventilation discharges from other processes were rerouted to this scrubber. The scrubber was originally designed to scrub acidic off-gas; however, many of the current feed streams contain ammoniated (basic) off-gas.

The feed streams all tied together through a network of ventilation ductwork of various diameters to a large diameter section before entering the transition section of this scrubber, reducing the linear velocity of the flow and allowing greater reaction time between the scrubber solution and the incoming feed streams.

In June 2009 the licensee implemented a new safety basis for the scrubber ventilation system that lowered the CSE mass limit by more than a factor of 60 and installed expansion plenums on a vent line, which the licensee assumed would reduce the amount of particulates that would travel to the scrubber. However, the licensee never considered the potential for the uranium to accumulate in a chronic fashion within the scrubber ventilation system. Further, the licensee incorrectly assumed that only minor amounts of uranium powder were expected to accumulate in the scrubber ventilation system. In December 2009 the licensee identified significant accumulation and performed additional modifications to remove an ammonia line. In 2010 the licensee instituted periodic cleaning of various processes. In April 2015 the licensee revised a procedure and included a note that based on "past experience the [percentage of uranium] of the trapped powder is approximately 45–48%."

Material buildup was still periodically observed and in April through May of 2016 large slabs of material would become dislodged during pressure washing and fall into the scrubber ventilation transition section. The operators were directed to continue to pressure wash the material so it would dissolve. Though not the desired result, it was fortuitous that the material did not dissolve, because the insoluble ammonium-uranyl-fluoride mixture prevented the formation of a critical mass configuration.

2.2 *Why did the licensee choose not to report the event immediately?*

In accordance with the regulations, licensees should report an event to the NRC within one hour in which there are no Items Relied On For Safety (IROFS) available and reliable to perform their function that results in the failure to meet specified regulatory performance criteria. On May 30, 2016, the licensee received the results of a sample taken from the material removed from sections of the scrubber ventilation system that indicated high uranium concentrations. However, on May 31, 2016, the nuclear criticality safety engineer, unaware of the sample results and assuming low uranium concentration, declared that the accumulated material did not challenge the CSE mass limit. As a result, the licensee did not immediately perform a detailed evaluation to determine whether the material discovered could have exceeded the safety basis.

On June 1, 2016, after completion of the cleaning activities, the nuclear criticality safety engineer communicated to the process engineer that there were no issues from the NCS group with restarting the scrubber ventilation system. Even though the process engineer was aware of the sample result that clearly indicated the CSE mass limit had been exceeded and a detailed evaluation of the credited controls (i.e., IROFS) was needed (because it had failed to prevent the accumulation), the licensee restarted the system. Only after receiving additional lab results confirming the high uranium concentration did the licensee stop the process and report the event to the NRC on July 14, 2016.

2.3 *What were some of the root causes for the event?*

Fundamentally, because the licensee's configuration management program did not ensure that design and physical changes to the scrubber ventilation system and associated controls (i.e., the designated IROFS) were properly designed and implemented to prevent adverse impact to the scrubber ventilation system safety basis, material accumulated that reduced scrubber efficiency by increasing the amount of uranium carryover to the system and generating insoluble uranium bearing compounds. Complex chemical interactions from various input streams created ammonium uranyl fluoride, which is mostly insoluble in water and plated out on the scrubber ventilation surfaces and within the scrubber body. Over time, the assumptions in the licensee's safety basis became invalid.

Furthermore, although the licensee conducted periodic inspections of the ventilation ductwork and was detecting material accumulation, they did not effectively use procedures to weigh and sample the uranium concentration in the material collected, undermining their ability to properly evaluate scrubber performance. Since scrubber ventilation system visual inspections did not effectively detect and remove significantly concentrated uranium from the system, eventually the established CSE mass limit was exceeded.

The licensee completed its own root cause evaluation in October 2016 and identified two root causes and two contributing causes for the event.

Root Cause 1: Programmatic controls for configuration management did not have the rigor to mitigate increased uranium accumulation in the scrubber ventilation system when design changes were made to the system and when operational requirements for the scrubber spray system were changed in the procedure.

Root Cause 2: Management did not scrutinize the content of the CSE and as-found conditions in the scrubber ventilation system with

the questioning attitude and conservative bias required for a healthy nuclear safety culture. Further, management did not ensure the organization had sufficient procedures and training to recognize and respond to deviations from the safety basis described in the CSE.

Contributing Cause 1: Operating experience and the corrective action processes were not effectively used to pursue the actions needed to detect, estimate, and mitigate deposited uranium in the scrubber ventilation system.

Contributing Cause 2: The scope of licensee audits and assessments did not provide a comprehensive review of the nuclear criticality safety program with an appropriate level of intrusiveness as is applied to higher risk activities.

The licensee's root cause analysis team also concluded that the event occurred due to long-standing weaknesses in the safety culture at the facility. The organization did not exhibit the behaviors expected to recognize that nuclear work is unique and that complex technologies can fail in unpredictable ways, resulting in adverse latent conditions not being recognized. Weaknesses in this pattern of thinking contributed to invalid assumptions and non-conservative decisions not being challenged. As a result, CSE mass limits were not well communicated and instructions for verifying the effectiveness of criticality controls were not well established. The licensee's root cause analysis team also identified a number of corrective actions to prevent either recurrence or significant consequences.

3 THE POTENTIAL FOR THIS EVENT WAS NOT FLAGGED BY THE REGULATORY PROCESSES

While the facility conditions and the licensee's initial responses to the conditions indicate a breakdown in their processes and programs, the NRC's overall response to the event was appropriate and as to be expected. An Augmented Inspection Team (AIT) was chartered on July 28, 2016, to: 1) review the facts surrounding the failure to maintain the CSE mass limits and controls in the scrubber ventilation system and the potential for similar failures in other production areas using the same control protocols, 2) assess the licensee's response to the failures, and 3) evaluate the licensee's immediate and planned long-term corrective actions to prevent recurrence. Performance issues identified by the AIT were submitted for additional NRC inspection follow-up and further review and enforcement activities followed normal regulatory processes.

In addition, as part of the NRC's overall platform of continuous improvement, NRC management initiated a lessons-learned activity to explore opportunities for improving NRC regulatory processes in identifying facility operational issues and preventing such events in the future. The team was chartered on October 28, 2016, to evaluate five areas: the licensing process, the inspection program, the operating experience program, roles and responsibilities, and knowledge management. The first two areas (licensing and inspection) are specific programmatic areas that periodically interface with the licensee and their analyses and programs. The other three areas (operating experience, roles and responsibilities, and knowledge management) support improving the capability, efficiency, and effectiveness of the regulatory staff in performing their responsibilities in the first two areas.

The team reviewed numerous documents related to each of the evaluated areas, including licensing review staff guidance, inspection procedures, and management directives, and also reviewed documents directly associated with the event, including the AIT report, an information notice, and a confirmatory action letter. The team also conducted individual and group interviews of nearly all project managers, technical reviewers, inspectors, and managers within the NRC's fuel fabrication arena, including the highest level of management within the region responsible for inspection of these facilities.

Through this effort, the team made a number of specific observations and recommendations associated with each evaluation area. The team issued its report on January 31, 2017, and many of the observations are summarized in the following subsections.

3.1 *The first opportunity comes during facility licensing*

These facilities are typically large process facilities with numerous individual processes and associated analyses. There is significant review effort expended during licensing and license renewal. Much of this effort ties to fully understanding the facility and its processes and the review of the licensee's identification and control of the multitude of hazards associated with these processes. A significant focus of the review is on the potential for criticality events, but detailed reviews are not performed for all areas. Instead, consistent with the licensing review staff guidance, reviewers primarily review the overarching facility safety program (a "horizontal review") and sample specific areas for more detailed ("vertical slice") review. This prioritization of the scope, focus, and detail of review is based on many aspects, including operating experience and reviewer experience, but also relies heavily on the perceived risk associated with the process as conveyed by the

licensee's Integrated Safety Analysis (ISA). In fact, the current licensing review staff guidance specifically states that the reviewers should more closely review processes and systems with a relatively high unmitigated risk than processes and systems with low risk. In the context of this event, the scrubber ventilation system was considered low risk by the licensee based on the assumptions: 1) that only minor amounts of uranium powder were expected to accumulate in the scrubber ventilation system, 2) low uranium concentration would be present within the scrubber ventilation system, 3) minimal amounts of small uranium particles were entrained within the intake ventilation ductwork, and 4) the scrubber constantly diluted the uranium concentration with the addition of makeup water during normal operation and anticipated upsets. These assumptions by the licensee are reflected in their ISA and established controls (i.e., IROFS).

The NRC licensing review staff guidance does not establish the level of review for processes and systems determined by the licensee to be low risk. Further, there is no specific guidance for reviewing processes and systems determined to be low risk that rely heavily on licensee assumptions. This lack of guidance resulted in the reviewers not reviewing this system in any depth during the prior facility license renewal. As a result, during the prior license renewal and amendment reviews, the reviewers did not challenge the overall performance of the system and related controls, including the assumption of low accumulation.

3.2 *The inspection program complements licensing*

One of the main purposes of the inspection program is to confirm continued compliance with the regulations and conformance with the approved license. Similar to the license review process, it is not practical to perform entire facility inspections, but rather, inspectors use a sampling approach. This approach is particularly relevant for facilities that do not have resident inspectors, which is the case for the subject facility (i.e., NRC inspectors are not located at the facility on a daily basis). For these types of facilities, over the year, inspectors visit the facility periodically to inspect specific programmatic aspects of the license, such as plant modifications, fire protection, operational safety, etc.

Similar to the license review process, the current inspection focus is on perceived high risk areas of the facility, which is based on the licensee's ISA. Because the licensee considered the scrubber ventilation system to be low risk, as stated above, the NRC did not consider this system for detailed inspection. Several inspectors noted that had the system been part of a detailed inspection, the licensee's deficiencies in the CSE and implementation of associated management measures and controls would likely have been identified.

Various inspection procedures appear to recognize that inspectors should examine presumably low risk processes and systems, but again, very limited guidance is provided on how to select samples from such processes and systems or the focus of such inspections.

3.3 *Operating experience could have provided insight and focus on this system*

Operating experience can be a valuable tool to help provide additional input to determining the appropriate focus and scope of facility areas to review and inspect. However, most license reviewers and facility inspectors did not rely upon the fuel fabrication facility operating experience program, which had previously been identified as needing to be improved. In fact, most inspectors and many reviewers were not aware of the fuel fabrication facility operating experience database or did not know how to access it. For those that were aware of the database, they observed that the database contained only relatively recent, publically available, US data and were unsure if it could trend events to support use in inspection planning. Furthermore, while a criticality inspection procedure had recently been revised to include the consideration of operating experience in inspection planning, other inspection procedures did not give any formal, structured guidance on considering operating experience. All of these conditions were considered to limit the usefulness of the operating experience database to the license reviewers and facility inspectors.

3.4 *Understanding roles and responsibilities*

Understanding individual and organizational roles and responsibilities is key to efficient and effective regulatory reviews and inspections. At the NRC, the licensing reviews are performed within one organization located near Washington, DC, while the inspections are performed within another organization located in Atlanta, Georgia. Communication and collaboration is essential in ensuring full understanding of licensing reviews and their implications for the inspection regime, especially when the organizations are physically separated by such a great distance.

The licensed facilities are required to provide annual summaries that describe the prior year's facility and process modifications and separately updates to their ISAs. These summaries can be, and are expected to be, used to inform inspection planning for the subsequent year. In the past, the

NRC licensing organization primarily performed the review of these summaries and provided its input to the NRC inspection organization, but in 2016 the NRC changed the lead role for the ISA summary reviews to the inspection organization to avoid overlapping efforts. However, the expectation of obtaining insights from the facility licensing review project manager and technical staff in these annual submittal reviews was not clearly established. Likewise, it was recognized through the lessons learned effort that the licensing review staff guidance did not clearly establish an expectation for obtaining insights from the inspection organization. In both cases, the potential for missing valuable insights was identified since the regulatory guidance did not establish a formal expectation for the various regulatory staff to collaborate in these areas.

3.5 *Knowledge management*

It is recognized that knowledge management is inextricably linked to all the other areas evaluated by the lessons learned team. It is an element critical to performing technical evaluations of licensee submittals, selecting relevant inspection samples, administering a successful operating experience program, clearly understanding respective roles and responsibilities, assessing the significance of an event, etc. Most of the lessons learned team recommendations involve some aspect of knowledge management. However, the lessons learned team did identify some fundamental knowledge management issues.

The current licensing and inspection qualification programs rely heavily on documentation reviews supported with some coursework and site visits. Certain skills that are important to regulatory staff success, however, are mostly left for the staff to pursue outside the qualification program, such as critical thinking, effective communication, and conflict resolution. All of these aspects require continuous practice and reinforcement and are invaluable when performing license reviews, conducting inspections, and interacting at all levels of the organization.

In addition, ensuring all regulatory staff are kept informed of current (and periodically reminded of past) licensing, inspection, operational, and technical issues improves the understanding and ultimately, performance of the regulatory staff and organization as a whole. While the inspection organization held periodic knowledge management seminars of selected topics, such a program was not being fully implemented within the licensing organization. As a result, lessons learned by some regulatory staff were not being effectively shared among all the other regulatory staff.

4 LESSONS LEARNED TEAM RECOMMENDED IMPROVEMENTS TO THE REGULATORY PROCESSES

The lessons learn team recommended improvements in all five regulatory areas. Most recommended improvements are associated with the verification of the technical bases and assumptions in the licensee's ISA and improving the knowledge bases and resources used by the reviewers and inspectors.

For the license review process, the lessons learned team identified for further evaluation the need to clarify the licensing review staff guidance to include guidance on the examination of the technical justification for processes and systems designated as low risk, especially those justifications related to key analysis assumptions.

For the inspection program, the team identified for further evaluation the need to modify the scope and focus of inspections so that all facility processes and systems with the potential for intermediate and high consequences are inspected within some periodicity, regardless of perceived risk significance. The team also suggested the development of additional guidance associated with reviewing and using the summaries of facility modifications and licensee ISA updates in support of inspection planning. Such additional guidance could also focus specific inspections on these analyses, with the intent of verifying the continuing validity of the technical bases and assumptions of the analyses.

For the operating experience program, the team identified for further evaluation the need to improve the framework and guidance for the flow of information from this program to the licensing and inspection programs. Related to the fuel fabrication facility operating experience database, the team suggested enhancing access to the database so that the information is more readily available to the licensing review staff and inspectors and to include legacy and international operating experience so that the database is more complete.

For the area of roles and responsibilities, the team suggested for further evaluation improving the guidance related to using the licensee's annual submittal of summary descriptions of facility modifications and ISA update summaries in inspection planning, setting the expectation to gain inspector facility knowledge and experiences within the licensing process, and providing rotational opportunities between the licensing review staff and inspectors to foster a better understanding of the diverse roles and responsibilities.

Finally, the need for improving knowledge management within these regulatory organizations is pertinent to all the above aspects. The team specifically identified for further evaluation the need to improve the qualification programs for the

licensing review staff and inspectors, to implement continuous knowledge management activities, such as regularly scheduled seminars and debriefings on topics of interest, and to periodically perform systematic reviews of the licensing and inspection programs to identify gaps and support continuous improvement.

5 CONCLUDING COMMENTS

The NRC created an action plan to guide and track the evaluations of the recommended improvements identified by the lessons learned team and their subsequent implementation, as appropriate. Some activities, such as the operating experience database, had previously been identified as needing to be improved and were already in the early implementation stages. Other activities involve additional considerations (e.g., priority, schedule, budget, and potential benefit) and are being evaluated and implemented, as appropriate. Through these efforts, the regulatory programs should improve, be more effective and efficient, and enhance the assurance of safety of the facilities.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Risk management for a particle therapy accelerator: The MedAustron experience

R. Filippini & P. Urschütz
EBG MedAustron GmbH, Wiener Neustadt, Austria

ABSTRACT: The Austrian facility MedAustron is one of the most advanced centers for ion beam therapy and research in the world. The MedAustron Particle Therapy Accelerator (MAPTA) is a CE-certified Class IIb medical device. Safety is the core attribute to be achieved, assessed and maintained during the entire system lifecycle. This paper tells about the experience of certifying a large particle accelerator as a medical device, from the risk management point of view. Examples of the different risk management activities are given, in order to explain the way the risks have been analyzed, controlled and evaluated.

1 INTRODUCTION

The MedAustron Particle Therapy Accelerator (MAPTA) is a CE-certified Class IIb medical device in compliance with the Medical Device Directive MDD 93/42/EEC [1]. The facility consists of four irradiation rooms, three for clinical operations (two horizontal and one vertical beam lines, one Gantry) and one for non-clinical research (one horizontal beamline). The layout of the accelerator complex is shown in Figure 1. The beam particles are generated in the ion source(s). They are pre-accelerated in the Linear Accelerator (Linac) and accelerated to the requested energies in the 80 meter-circumference Synchrotron ring, before being extracted and delivered into the irradiation rooms.

The maximum beam energies for patient treatment are 250 MeV/u and 400 MeV/u for proton and carbon ions respectively. For non-clinical research, beam energies up to 800 MeV/u can be generated. Compared to conventional radiotherapy with photons (gamma rays, X-rays) or electrons, the treatment with protons and carbon ions has a more precise dose distribution (i.e. the Bragg peak) and reduces damage to the adjacent healthy tissues significantly. This makes it possible to deliver much higher doses, with greater benefits for the patient, but also presents safety concerns, which have to be addressed by risk management.

Figure 1. MAPTA layout.

In general, the risk management for particle therapy accelerators and particle accelerators has relatively little literature, for several reasons. First, this depends on the size of the accelerator; only at high energies the consequence of a malfunction becomes relevant for costs and human lives for non-medical applications. Secondly, the particle accelerators are often research facilities, and this allows a larger degree of freedom of addressing safety without a risk management process. As a third reason, even if risk management is in place, there are restrictions that limit the dissemination of results. Because of that, most of the existing studies deal with reliability and availability, which are related to uptime and performance, and only a few of them included safety and risk in scope. The CERN Large Hadron Collider is within these exceptions. The LHC operates at very high beam energies (7TeV) and it is designed to deal with potentially catastrophic scenarios, e.g. beam losses and quench of the superconducting magnets with damage to equipment and interruption of operations for several months [1]. Several reliability and safety studies have accompanied its design and realization [3, 4, 5, 6]. Within the domain of particle therapy accelerators, a few studies exist, which are of great interest. A Probabilistic Risk

Assessment (PRA) has been performed for the PROSCAN proton therapy accelerator of the Paul Scherrer Institute [7] and similar analyses for radiation sources can be found in the IAEA report [8].

This paper presents the risk management for the particle accelerator MAPTA of MedAustron. The content of the paper is organized into five sections. Following this introduction, Section 2 contains the functioning principles and the safety architecture of MAPTA. The risk management framework is in Section 3, including the regulations and standards, the hazards of concern, the risk metrics and the acceptability criteria. Section 4 contains the description of the risk assessment activities, as well as several examples. A few concluding remarks are contained in Section 5.

2 MAPTA SAFETY ARCHITECTURE

2.1 *Functioning principles*

MAPTA produces ions beams of a given energy and intensity for the irradiation of the tumor, as specified in the patient treatment plan. A complete patient treatment consists of between 20 to 40 sessions, each with a total duration of about 30–40 minutes including a few minutes of irradiation time, depending on the tumor indication.

A treatment irradiation session starts after MAPTA has been handed over to the user (medical physicists and doctors), who has the responsibility of supervising the operations from the control room of the irradiation room. The entire irradiation process is automated. The MAPTA Medical Frontend (MF) system administers the irradiation according to a predetermined timing sequence. The scanning of the tumor target volume and the dose delivery is conceptually simple: the MF system switches on the chopper kicker magnet and actuates the switching dipoles to let the beam travel from the synchrotron into the irradiation room and then through the nozzle into the target volume. The beam scanning is done in "iso-energy slices", in order to deposit the required radiation dose at different positions and depths (higher energy = deeper penetration). During the irradiation, the MF system supervises the functioning of all components and monitors the beam parameters (intensity, position, energy, dose, etc.). Any possible malfunction, deviation out of tolerable limits or unauthorized action, triggers the interruption of treatment for safety reasons.

2.2 *Safety architecture*

The safety architecture in MAPTA covers three different domains: 1) access protection, 2) interoperability of the irradiation rooms and 3) patient treatment.

The MAPTA access protection system guarantees the "safe access and stay" in the accelerator areas and irradiation rooms for technical and medical personnel. Two states, "Green" and "Red", are defined. When the state is Green, the respective area/room is accessible. All radiation sources are turned off and the beam stoppers (or equivalent devices) prevent the beam from reaching the accelerator area or the irradiation room. When the state is Red, the beam can be sent into the respective area/room, and the accesses are locked. The violation of the Red state (e.g. unauthorized access) triggers an interlock which activates the Green state for the respective area or irradiation room, immediately stopping the beam and shutting down all radiation sources.

The second safety domain covers the risks that are caused by the simultaneous use of the irradiation rooms. Two operation modes per room are defined: 1) physics and accelerator (e.g. for commissioning, maintenance, and test) and 2) medical (for clinical operation). For safety reasons, certain combinations of the operation modes are not allowed and they are vetoed. Moreover, certain activities cannot be executed when in clinical operation. For example, if one irradiation room is in medical mode, none of the others irradiation rooms may be in the physics and accelerator modes. In case of violation of the veto, an interlock is triggered with subsequent beam stop and interruption of the treatment.

The third safety domain covers the risks during patient treatment. The goals are to guarantee that 1) MAPTA is correctly configured for the treatment plan, 2) the beam physics characteristics (energy, intensity, position, width) are within acceptable limits and 3) the dose is correctly delivered (i.e. the correct dose at the right time, into the right irradiation room). Most of the risk control measures are implemented in the MF system or other systems that assure an equivalent safety level. In case of detected errors, the beam is sent into the beam dump and the treatment is interrupted.

2.3 *Safety principles*

The architecture of MAPTA meets the "single fault safe" principle, which states that "a combination of two independent errors must not lead to a life-threatening situation" for the patient [9]. This principle requires a number of design features to be implemented in MAPTA, e.g.:

- Fail-safe behavior,
- Independency of risk control measures,
- Acknowledgement of a safety action,

- Alternative means or independent back-ups,
- Integrity of risk control measures.

The fail-safe behavior makes it impossible for a failure to develop in an unsafe way. For example, the selection of an incorrect source is fail-safe because it is impossible for the synchrotron to accelerate an ion species different from the one for which it is configured (e.g. protons instead of carbon ions and vice versa). A power supply failure is fail-safe for all MF components, e.g. it causes the demagnetization of the chopper magnet and the resulting beam dump. A pressure leakage in the beam stopper cylinders drops the shutters into the beam line by gravity, which again is fail-safe. The logic levels of the electronics are fail-safe in the case of loss of signal, and so forth.

Another important design feature is the independence of the risk control measure and the supervised system. This is guaranteed by hardware separation and/or software segregation, to avoid possible common causes of failure.

The result of the execution of a risk control measure is acknowledged and, in case of a fault, other risk control measures are triggered to complete the safety action, e.g. if the beam is still detected after the actuation of the chopper magnet, or if the position of the beam stoppers is still out of the beam path. The functioning of risk control measures is also verified before the start of every patient treatment session. This is guaranteed by functional tests and internal self-diagnostics. In addition, periodic checks are performed for the integrity of the risk control measure, e.g. by discovering dormant faults in redundant components.

2.4 *Hazardous situations and reaction times*

The majority of the hazardous situations in MAPTA develop in the order of milliseconds to a few seconds, depending on the causes and the circumstances. As a consequence, every risk control measure must have a suitably fast reaction time. All risk control measures are implemented in electronics and electro-mechanical components, including the back-up measures, which intervene in case of failure. The operators and the users are seldom called upon to respond to hazardous situations, unless they have a longer time of development for which the human reaction time is compatible and effective.

3 THE RISK MANAGEMENT FRAMEWORK

3.1 *Regulations and standards*

Medical devices have to comply with a complex and articulated sets of regulations, norms and standards. The Medical Device Directive MDD 93/42/EEC stipulates the manufacturers' responsibilities to consider all safety aspects and to be able to address them.

The Directive states that: *"the (medical) devices must be designed and manufactured in such a way that, when used under the conditions and for the purposes intended, they will not compromise the clinical condition or the safety of patients, or the safety and health of users or, where applicable, other persons".*

It continues: *"the solutions adopted by the manufacturer for the design and construction of the devices must conform to safety principles, taking account of the generally acknowledged state of the art. In selecting the most appropriate solutions, the manufacturer must apply the following principles in the following order:*

- *eliminate or reduce risks as far as possible (inherently safe design and construction);*
- *where appropriate take adequate protection measures including alarms if necessary, in relation to risks that cannot be eliminated;*
- *Inform users of the residual risks due to any shortcomings of the protection measures adopted."*

The above stipulations are extensively explained in the EN ISO 14971:2012 [10] regarding what concerns risk management. The safety requirements are defined in medical device standards, such as the general standard IEC 60601–1 [9], the safety standard IEC 60601–2-64 [11], the medical software standard IEC 62304 [12], the standard for medical IT networks IEC 80001–1 [13] and the IEC 62366 [14] for usability engineering. Industrial safety standards that are applied for MAPTA are the IEC 61508 with the derived IEC 62061 and ISO 13489 [15, 16 and 17]. The list is not complete and many other standards apply, which focus on a particular technical domain, e.g. the ÖVE/ÖNORM E-8001 [18] for electrical hazards.

3.2 *Scope of the risk management*

The risk management is performed throughout the entire life cycle of MAPTA, from the design to the decommissioning. While safety is the main goal, the manufacturer shall also take into account reliability and uptime, which eventually affects the performance, i.e. uptime and the patients' throughput. This is a well-known trade-off in safety engineering. Risk management also takes into account the interdependencies of MAPTA with the systems at the interface, such as the patient positioning system, the technical infrastructure and the IT network. The user activities related to the preparation of the treatment plan, accommodation of the patient, imaging, and supervision of treatment are out of scope.

The risk manager is responsible for the risk management process. The outcomes of the risk management process are reviewed by a team of experts, which includes representatives of the top management, accelerator technologists, medical physicists, and radiation oncologists.

3.3 *Hazards and categories of risk at MedAustron*

The EN ISO 14971 defines the applicable hazards and the categories at risk for MAPTA. The categories of risk are the patient, the user of MAPTA, the technical and medical personnel, the equipment, the environment and third persons (i.e. anybody not involved in the patient treatment, operation and/or service of MAPTA). The goal is to assure that all applicable combinations "hazard versus category of risk" are in scope of risk management. Table 1 provides an excerpt of the list of hazards (radiation, electrical, etc.) that apply to the categories of risk patient, personnel/user and environment for MAPTA, including the applicable standards.

3.4 *Risk metric and acceptance criteria*

The risk metric for MAPTA is the Risk Priority Number (RPN) [10]. The RPN is the product of the probability P and the severity S of a hazardous situation, i.e. $RPN = P \times S$. Six frequency intervals for P are defined in MAPTA:

- P = 1: Incredible;
- P = 2: Unlikely;
- P = 3: Seldom;
- P = 4: Occasional;
- P = 5: Often;
- P = 6: Frequent.

and five severity levels for S:

- S = 1: negligible;
- S = 2; minor;
- S = 3: moderate;
- S = 4: severe;
- S = 5: catastrophic.

Table 1. Hazards and categories of risk.

Hazards	Patient	Personnel/user	Environment
Radiation	ISO 14971	IEC 61508 IEC 62061 ISO 13489	UVP (RP)
Electrical	n.a.	ÖVE/ÖNORM E-8001	n.a.
Software	IEC 62304	n.a.	n.a.
IT	ISO 80001–1	n.a.	n.a.
Mech.	ISO 14971	ISO 14971	n.a.
Use	IEC 62366	IEC 62366	n.a.
…			

The frequency intervals for P and the levels for S are chosen in agreement with the risk management standard and the current state-of-the-art [10, 19, 20].

The acceptability criteria are applied by defining two risk thresholds. A risk is acceptable if $RPN < 10$, while it is not acceptable if $RPN > 12$. Risks with RPN between 10 and 12 (inclusive), and that cannot be reduced further, are evaluated by risk-benefit analysis. These risks are acceptable if the benefit for the patient outweighs the residual risk.

4 RISK ASSESSMENT FOR MAPTA

4.1 *Scientific rationale*

The risk assessment includes the activities of risk analysis, risk control and the risk evaluation. These activities are performed in sequence, one after the other, as shown in Figure 2. The Failure Mode, Effects and Criticality Analysis, FMECA, is the methodology for risk assessment [21] for all systems and components in MAPTA. The only exception is the access protection system, for which a

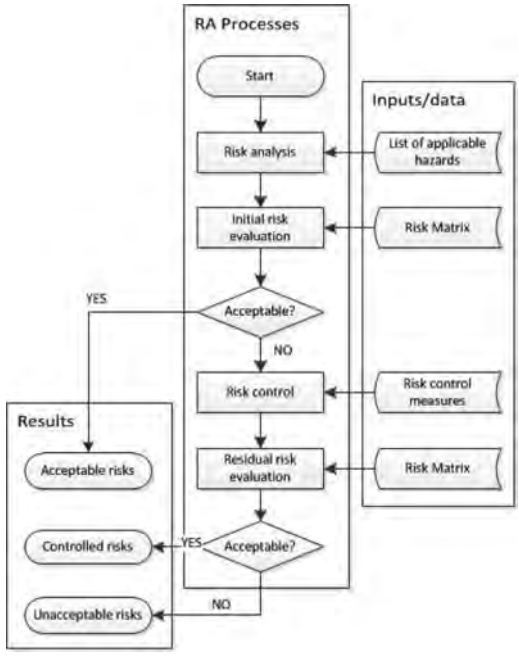


Figure 2. The risk assessment process workflow.

different method applies, as well as a different risk metric, the Safety Integrity Level (SIL) [15] instead of the RPN.

The FMECA is an inductive system analysis method. It uses the natural language to describe the hazardous situations, which has the advantage of facilitating the communication between the risk analyst and the system experts. Another advantage is that the analysis with FMECA is semi-quantitative, i.e. instead of precise failure statistics, probabilities are chosen within predefined intervals. This feature makes the risk assessment relatively easy, but also turns out to be a limitation in comparison with other methods of system/risk analysis, e.g. fault tree, which is more accurate. Another limitation exists for the analysis of common causes of failures. These drawbacks are known and have been addressed in MAPTA by ad-hoc risk assessment guidelines that assist the analyst during the compilation of the FMECA. The guidelines provide a comprehensive methodological approach and guarantee the correctness and consistency of the results.

4.2 *Risk analysis*

The risk analysis consists of three activities: definition of the hazardous situations, risk estimate and risk evaluation.

The first activity describes the failure dynamics as a causal chain of events, "initial cause → error → fault → failure → malfunction" that, together with the circumstances, lead to the hazardous situation. The more accurate the description of the causal chain is, the more effective the apportionment of the risk control measures.

The second activity makes it possible to estimate of the initial RPN. The estimate is done without considering the risk control measures. The MAPTA risk assessment guidelines provide empirical look-up tables with intervals of probabilities for the different classes of faults or errors (e.g. HW/SW faults, human errors, etc.). Analogous tables are also available for the severity. The analyst chooses the initial RPN within the suggested probability and severity intervals, for the worst case scenario. The last activity of the risk analysis is the evaluation of risks against the acceptability criteria. Those risks that are non-acceptable, i.e. RPN > 10 shall be mitigated by risk control measures.

4.3 *Risk control*

The EN ISO 14971 defines three types of risk control options: 1) inherent safety, 2) preventive/protective and 3) information for safety. Each of these risk control options has a different effectiveness, e.g. by avoiding, preventing or stopping a hazardous situation, see Table 2. Inherent safety is the most effective among the control options. It prevents the hazardous situation from developing. Preventive measures intervene while the hazardous situation is developing, while protective measures intervene when the hazardous situation has already developed and for example, it represents a real harm for the patient. Information for safety is the least effective of the risk control options. It includes organizational measures such as instructions for operators and users of MAPTA.

Table 2. Risk control options in MAPTA.

Risk control measure	*Effect to the hazardous situation*
Inherent safety (physical)	Make it physically impossible
Inherent safety (logical)	Make it logically impossible
Inherent safety (failsafe)	Turn it into failsafe
Preventive	Detect and stop development
Protective	Detect and stop at occurrence
Information for safety	Contribute to prevent/avoid

The MAPTA guidelines contain the rationale for the application of the risk control measures and the estimate of the risk reduction. This rationale follows a few general recommendations from the risk management standard and best practice. The risk control measures have to be applied altogether and not in isolation (EN ISO 14971). Secondly, a risk control measure has to be associated with the respective failure event (as identified in the causal chain). In addition, they have to be applied in the right sequence, i.e. inherent safety first, then organizational measures, preventive and protective measures. Finally, failure dynamics and the risk control measures have to be considered (and analysed) as interrelated processes. The model that supports this description is an event sequence diagram. An illustrative example is shown in Figure 3 for a hazardous situation with three risk control measures.

A risk control measure is "successful" if it prevents the hazard from developing and leads to the "end state", while it is "unsuccessful" if the hazard can develop further. In terms of P and S this means:

- Successful: P does not change, S is lower:
- Unsuccessful: P is lower, S does not change.

The amount of reduction of P depends on the type of risk control measure. The reduction of the severity S depends on the instant at which the risk control measure intervenes. If this is before the hazardous situation becomes a real harm, then the severity drops down to S = 1 (no harm), see End

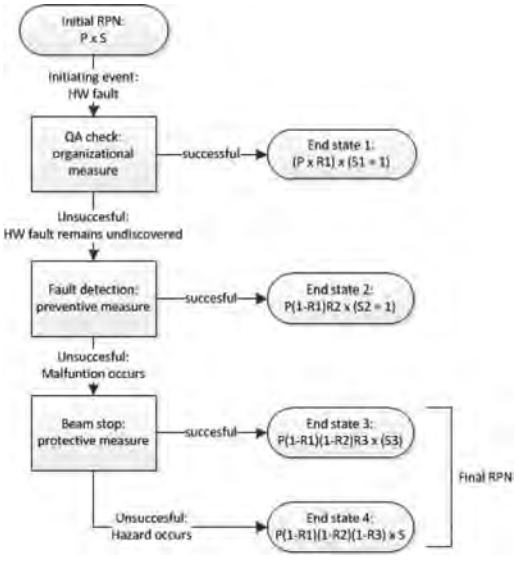


Figure 3. The risk reduction model.

states 1 and 2 in Figure 3. On the contrary, if it occurs at a later stage, then the residual severity could be higher, which is the case of End state 3.

The residual risk (i.e. the final RPN) is the risk after all risk control measures have been applied. Because the risk is reduced by a certain amount after the execution of a risk control measure, the residual risk is the highest RPN associated with the end states of the last risk control measure, e.g. End states 3 and 4 in Figure 3. In the example, End state 3 has a residual severity related to the extra dose deposited, before the beam is stopped by the protective measure. The estimate of the residual severity shall take into account: 1) the type of particles (e.g. carbon ions are heavier than protons), 2) the detection threshold of the protective measure, and 3) the reaction time, up to the complete beam stop. The limits for the maximum extra dose are defined in the standard IEC 60601-2-64, both under normal and fault conditions [11].

The amount of risk reduction is calculated in the risk assessment guidelines of MAPTA by look-up tables for every risk control measure. Figure 4 shows the risk reduction table for a preventive risk control measure. The calculation of the risk reduction is straightforward. The initial RPN = P × S identifies the cell in the table with the residual risk after the application of the risk control measure, e.g. RPN with P = 4 and S = 5 becomes RPN = 2 × 5. If more risk control measures apply, then the output of one risk reduction table becomes the input of the next risk reduction table and so forth.

Probability	Severity 1	2	3	4	5
6	3x1	3x2	3x3	3x4	3x5
5	3x1	3x2	3x3	3x4	3x5
4	2x1	2x2	2x3	2x4	2x5
3	2x1	2x2	2x3	2x4	2x5
2	1x1	1x2	1x3	1x4	1x5
1	1x1	1x2	1x3	1x4	1x5

Figure 4. The risk reduction table for preventive measures.

4.4 *Risk evaluation*

The risk evaluation is the final activity of the risk assessment. All individual risks are evaluated on the basis of the acceptability criteria. After this is done, the same individual risks are re-evaluated together in order to account for statistical cumulative effects. According to the risk assessment guidelines of MAPTA, the cumulative statistical effects are estimated by counting how many individual risks are in the same cell P × S. For every N individual risks in the same cell P × S, one cumulative risk is added in the cell (P+1) × S. The threshold N depends on the order of magnitude of the interval P. The P intervals in the risk matrix correspond to orders of magnitude 1 (P = 3, 4 and 5) and 2 (P = 1, 2), which is N = 10 and N = 100. As an example, 12 individual risks with P = 3 and S = 3 would be statistically equivalent to two individual risks in 3 × 3 and one cumulative risk in 4 × 3.

4.5 *A few results*

Figure 5 shows the results of the FMECA based on the functional description of MAPTA. The scope of this FMECA is the patient treatment session. A patient treatment session consists of four phases: 1) request and allocation of the accelerator components and the irradiation room, 2) activation, 3) irradiation, and 4) termination of treatment. In total, 132 individual hazardous situations have been identified and the respective risks have been analyzed, controlled and evaluated. The analysis also includes 26 risks related to common causes of failure. All residual individual risks (and the cumulative statistical effects) have been evaluated and they are acceptable i.e. all RPNs < 10. This result is obtained by applying 90 different risk control measures in MAPTA including:

- 22 inherent safety measures,
- 39 preventive risk control measures,
- 9 protective risk control measures,
- 20 organizational measures.

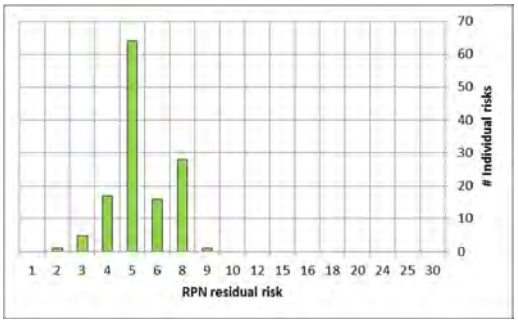


Figure 5. Distribution of the residual risks.

Another seven risk mitigations have been identified and transferred outside of MAPTA for their implementation.

The analysis of the risk control measures provides interesting insights on the MAPTA architecture for the patient safety. A risk control measure is often called for in more than one hazardous situation. Therefore, by counting the times this is required, it is possible to deduce its importance. The following statistics are obtained: the preventive measures are called for 143 times, the inherent safety measures 111 times, the organizational measures 49 times and the protective measures are called for 39 times. The preventive measures are the most required, followed by inherent safety, while the protective measures are the least required. This result is in good agreement with the general recommendation of the medical device standard, which states that hazardous situations shall be either avoided by design or prevented. Indeed, only a smaller percentage requires the intervention of protective measures.

The FMECA, based on the functional description of MAPTA, is completed by other FMECAs, which deal with risks at the system and component level. In total, more than 1700 individual risks have been analyzed and controlled in MAPTA. The overall residual risks have been evaluated and they are acceptable.

4.6 *Final reports and post-production*

The outcomes of the risk management activities and the verification by tests of the risk control measures are included in the MAPTA risk management report, which accompanies the declaration of conformity for the CE certification as a medical device. All documents produced by the risk management are organized in the respective file, which is periodically inspected by external auditors.

The risk management also covers the post-production activities, such as product changes, maintenance, commissioning and project development. These activities often require the update of the existing documents or new risk analyses, and because of that, they are constantly monitored. Reporting adverse events, errors and near misses is also within the scope of risk management. This includes the validation of the risk estimates and the estimate of MAPTA uptime, based on the operational data, as it was done, for example, in [22].

5 CONCLUSIONS

This paper presented an overview of the risk management for the particle accelerator MAPTA of MedAustron. All activities in scope of the risk management have been discussed, with several examples regarding the methodologies for the risk analysis, the risk control techniques and the outcomes of risk assessment. In total, more than 1700 individual risks have been identified, analyzed and controlled in MAPTA. The residual risks have been evaluated and they are acceptable.

The depth of this work is such that it can be barely summarized within these pages. Nonetheless, it is worthwhile sharing a few lessons learned. The first lesson learned is related to the complexity of the particle therapy accelerator on one hand, and the demanding requirements of risk management for medical devices, based on an all hazards approach, on the other hand. Complexity has been managed by the definition of different frameworks, focused on a specific domain e.g. intended use and patient safety, access protection and radiation hazards, industrial safety, etc. The second lesson learned concerns the development of the know-how. Standards for medical devices are general encompassing different applications. It is under the responsibility of the manufacturer to interpret the standard clauses, and organize risk management accordingly. A lot of groundwork has been done in this respect, with the preparation of scientific rationales, guidelines and safety concepts. Another lesson learned regards the safety culture. Risk management is a discipline that cannot be performed in isolation, and requires competencies of specialists from various fields. An essential requisite is to build adequate safety responsiveness within the different groups (engineers, accelerator and medical physicists) to be able to anticipate rather than react to potential adverse events. In order to attain these goals, risk management promotes regular exchange and dissemination of information, including training and team work. The fourth lesson learned is about the CE certification process itself. The MedAustron particle accelerator stands out among the majority of the existing particle therapy accelerators, and one of the reasons is that

risk management was part of certification process. Besides the illustration of the methodologies and the results, there is the unique experience of having dealt with a particle therapy accelerator of this size, which has challenged and possibly improved the state-of-the-art in this subject. This is one of the added values of the "MedAustron experience" in the domain of particle therapy accelerators.

Risk management was executed successfully during the phases that accompanied the certification of the particle accelerator as CE medical device, and it is presently looking after post-production activities for clinical operations with irradiation rooms IR2 and IR3 (horizontal beamlines, protons), and non-clinical operation in IR1. The vertical beamline of IR2 will be active in Spring 2018, while carbons ions are still under commissioning for a future use. The proton Gantry is planned as the last step in the commissioning sequence.

ACKNOWLEDGEMENTS

The authors are grateful to the colleagues Peter Gruebling, Gregor Kowarik, Maarten Schokker, Andreas Weinfurter and Marte Fjelland for their contribution in the risk management activities of MAPTA.

REFERENCES

[1] European Council, Medical Device Directive MDD 93/42/EEC, 1993.
[2] R. Schmidt and al., "Protection of the CERN Large Hadron Collider", New J. Phys. 8 290 (2006).
[3] R. Filippini and others, Reliability Assessment of the LHC Machine Protection System, Particle Accelerators Conference PAC 2005, pp. 1257–1259, Knoxville, USA, 16–20 May 2005.
[4] R. Filippini, "Dependability Analysis of a Safety Critical System: the LHC Beam Dumping System at CERN", CERN-THESIS-2006–054 - 2006.
[5] R. Filippini, "Safety Analysis of the Movable Absorber TCDQ in the LHC Beam Dumping System," CERN-ATS-2009–004.
[6] R. Filippini, J. Uythoven, "Reliability Analysis of the Trigger Synchronisation and Distribution System of the LHC Beam Dumping System", CERN-ATSNote-2013–043-TECH.
[7] B. Reer, V. Dang, L. Podofillini and D. Coray (2006). First Results from a Probabilistic Risk Assessment for PSI's Spot-Scanning Proton Therapy Facility, Probabilistic Safety Assessment and Management (PSAM8), New Orleans, LA, USA.
[8] Case studies in the application of probabilistic safety assessment techniques to radiation sources, IAEA-TECDOC-1494, Vienna, Austria, IAEA, 2006.
[9] International Electrotechnical Commission, Medical electrical equipment—Part 1: General requirements for basic safety and essential performance, chapter 4, edition 3.1 60601:2012–08.
[10] International Organization for Standardization, Medical devices: Application of risk management to medical devices, 14971:2007.
[11] International Electrotechnical Commission, Medical electrical equipment—Part 2–64: Particular requirements for the basic safety and essential performance of light ion beam medical electrical equipment, 60601–2-64: 2014.
[12] International Electrotechnical Commission (2006). "Medical device software—Software life cycle processes, IEC 63304:2006.
[13] International Organization for Standardization, Application of risk management for IT-networks incorporating medical devices -- Part 1: Roles, responsibilities and activities, ISO 80001–1:2010.
[14] International Electrotechnical Commission, Medical devices: Application of usability engineering to medical devices, IEC 62366:2007.
[15] International Electrotechnical Commission, Functional safety of electrical, electronic, programmable electronic safety-related systems; IEC 61508–1:2010.
[16] International Electrotechnical Commission, Safety of machinery: functional safety of safety related electrical, electronic and programmable electronic control systems IEC 62061 1.1. 2012–11.
[17] International Organization for Standardization, Safety of machinery—Safety-related parts of control systems—Part 1: General principles for design, IEC 13849–1:2008–12.
[18] Österreichisches Normungsinstitut, Errichtung von elektrischen Anlagen mit Nennspannungen bis 1000V und 1500V, ÖVE/ÖNORM E 8001–1:2003.
[19] International Commission on Radiological Protection (ICRP), Prevention of Accidental Exposures to Patients Undergoing Radiation Therapy, Publication 86, Pergamon Press, Oxford and New York (2000).
[20] International Atomic Energy Agency, Lesson learned from accidents in radiotherapy, safety report series 17, Vienna, Austria, IAEA, 2000.
[21] International Electrotechnical Commission, Analysis techniques for system reliability—Procedure for failure mode and effects analysis (FMEA), IEC 60812:2006.
[22] R. Filippini, E. Carlier, N. Magnin, J. Uythoven, "Reliability analysis of the LHC beam dumping system taking into account the operational experience during LHC run 1", ICALEPCS2013, San Francisco, October 2013.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Rescue Emergency Drone (RED) network for assessment of traffic accidents in Denmark

A.S. Kristensen, S. Mehmood & S. Ahmed
Aalborg University, Esbjerg, Denmark
Danish Centre for Risk and Safety Management, Esbjerg, Denmark

D. Ahsan
Southern Denmark University, Esbjerg, Denmark

R. Zamora
Danish Center for Risk and Safety Management, Esbjerg, Denmark

ABSTRACT: Onsite real-time video streaming of traffic accidents covering condition of inflicted person can help overcome problem of under and over-triage by emergency services. A network of Rescue Emergency Drones (REDs) that could transmit live video to emergency services is proposed to be mounted at the sites prone to frequent accidents in Denmark. A risk mapping for placement of RED docking stations at suitable places of southern Danish city, Esbjerg and its outskirts has been designed using Geographical Information System (GIS) tools ArcMap, and ArcGIS 10.5.1. The result demonstrates the robustness of RED into emergency services by providing high quality footage that helps to assess the scene of crash faster than the standard existing procedure.

1 INTRODUCTION

Lack of clarity of the condition of patients or injured persons can lead to wrong decision taken by Emergency Medical Dispatcher (EMD) and Emergency Medical Service (EMS). For instance, in Denmark 'Unclear Problem' of level B-E emergency was the chief complaint involving 66 deaths in period between 2011–2012. Because in many cases appropriate resources were not dispatched to handle the emergency which leads to degree of over-and under triage. (Andersen et al 2014)

Under and over-triage is a major problem that costs not only financial losses but also human lives. For instance, in Denmark 18 deaths could have been prevented if EMD had dispatched a targeted response. (Andersen et al 2014)

Drones have many applications in emergency services. For instance, drones can reach at the scene of the road accidents faster than conventional means of transportation. The aim of this study is to explore the potential benefit of a drone system to transmit live video footage covering the condition of the inflicted/patient that may improve the decision making of EMD. For application of drones in this regard, this study considered traffic accidents, because traffic accidents top the list of human casualties' statistics of non-natural cause of fatalities. Around 1.25 million people lost their lives and 20–50 million people suffered injuries due to traffic accidents. (WHO 2017)

Although Denmark is relatively safe country for commuters, there were 211 people killed, 1,796 suffered serious injuries and 1,432 suffered slight injuries in traffic accidents in 2016. (Statistics Denmark 2017)

Emergency medical dispatcher finds it difficult sometimes to comprehend the situation and condition of the emergency. First responders usually rush to the emergency sites with limited information that can sometime jeopardize the rescue operation. Therefore, if EMD and EMS can see and assess the severity of injuries of an inflicted person in traffic accident, it will facilitate a targeted response via live video footage.

The Danish Emergency Medical Communication Center (EMCC) receives medical emergency calls to respond and rescue patients and injured persons. EMCC staff responds the calls according to the Danish index care into five categories. Category "A" represents a life-threatening or potentially life-threatening condition; therefore, it requires immediate response. Category "B" means that a patient or injured person requires urgent help, but his/her condition is not life threatening, whereas category "C" requires an ambulance in a

non-urgent condition. Under "D" category EMD needs to send a patient transport while under "E" category no ambulance is dispatched instead taxi or other transportation is advised.

Category "A" has a pre-hospital time of 08:12 minutes, but sometime EMD can make a wrong decision in dispatching targeted response. In one of the case of an audit study, it was found, that EMD categorized an emergency as category B, however, when the ambulance arrived at the scene, a Mobile Emergency Care Unit (MECU) was summoned due to the severe condition of the patient. The life of patient would have been saved if the ambulance along with MECU could have been dispatched. (Andersen et al 2013) Therefore, fast response with right resources dispatched in saving human lives is crucial. The Danish pre-hospital median time for all emergencies is 10:27 minutes. (Andersen et al 2013) The average minimum response time of fire and rescue services (FRS) is 10 minutes and it could be 15–20 minutes depending on the location of accident sites. (Sydvestjysk Branvæsen 2015)

Rescue Emergency Drones (REDs) can reduce time of onsite assessment of the condition of inflicted person by reaching to the patient/injured person faster than the conventional means of transport and transmitting 'live video' that can help to cope with the problem of under, and over-triage.

2 RED NETWORK IN DENMARK

Providing visual aids by a drone will improve the prehospital process in case of a traffic accident this aim of the project will meet the need to reduce fatalities.

There are many potential benefits of incorporation of RED into emergency services and their improving the decision making as follows,

- Real time visual feed from the scene of crash will assist in better assessment of severity of the emergency by dispatcher.
- The sufficient of amount of resources will be saved by emergency services by overcoming problem of over and under-triage.
- A targeted and quick response will increase survival rate of inflicted persons.
- A targeted response will improve the quality of life by decreasing the severity of injuries of causalities and thereby saving them to live without physical impairments.
- Dispatcher would be able to better guide caller to handle emergency properly while the ambulance is on its way.
- Dispatcher can calm down the panicked caller.

Denmark map is developed in GIS based on the data of traffic accidents in Denmark between 2012–2016.

In Figure 1 the black dots represent the accidents, which are more frequent in populated areas of the country. There were 87,787 total accidents in Denmark recorded between 2012 to 2016 (Danish Road Directorate 2017). Majority of the accidents are reported in bigger cities for instance, Copenhagen, Odense, Aarhus and Aalborg.

Based on the audit study it is assumed that the onsite live video streaming aid would help to mitigate the consequences of accidents. For visual aid, network of REDs is proposed for Denmark. For RED network Esbjerg municipality is considered as a case study with the broader application for the rest of the country.

2.1 *Esbjerg municipality*

This study is carried out in Esbjerg municipality, which covers a total area of 794.5 km² (Sydvestjysk Brandvæsen 2015). The total population of Esbjerg municipality is 115,905. (Statistics Denmark 2017) The density of the population is 116 (individual/km²). The municipality consists of both rural and urban areas. Esbjerg municipality observed 2,515 total number of traffic accidents between 2012 and 2016. The accident data is extracted from the Danish Road Directorate and accidents coordinates are shown in the following Esbjerg municipality map.

Most of the accidents were recorded in the residential area. A total *of n* = 2515 cases of traffic accidents were reported in Esbjerg municipality between 2012 to 2016. The traffic accident casualties during this period are given in Table 1.

Table 1 shows that in 2016 total 63 casualties were recorded, among them six persons were killed, 38 injured seriously and 19 were injured slightly. One death due to traffic accidents costs Danish society up to 17.3 million DKK. (Transportministeriet 2010). To avoid such a huge loss, it is necessary to improve the pre-hospital response

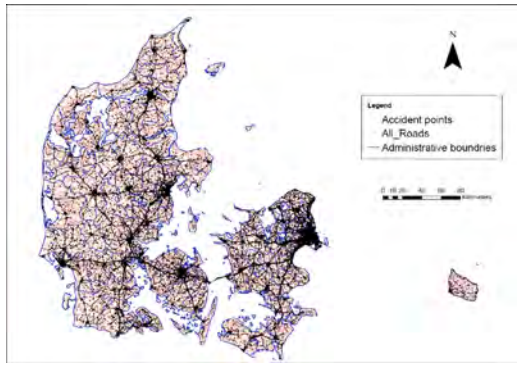


Figure 1. GIS mapping of traffic accidents coordinates on Danish roads.

Table 1. Traffic accidents casualties in Esbjerg municipality.

Years	Casualties, total	Killed	Seriously injured	Slightly injured
2012	102	7	49	46
2013	61	1	32	28
2014	83	2	47	34
2015	86	7	42	37
2016	63	6	38	19

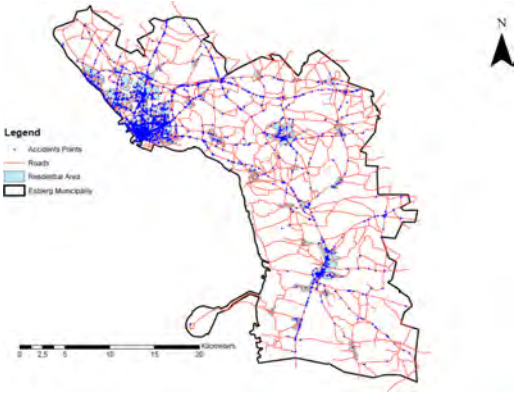


Figure 2. GIS mapping of traffic accidents coordinates on Esbjerg municipality roads.

time. There is a robust evidence for an association between short response time and survival rate for traffic accidents. (Sánchez-Mangas 2010)

Currently in Denmark, total pre-hospital median time for category "A" is 08:12 minutes whereas for B, it is 13:27 minutes. For "C" category, it is 16 minutes and 5 seconds. Similarly, for "D" category the time is 19:46 minutes. The Danish pre-hospital time median time for all emergencies is 10:27 minutes. (Anderson et al 2014). The detail of prehospital time is given in Figure 3.

Moreover, fire and rescue service (FRS) has also a crucial role in saving human lives as first responder along with EMS. As far as fire and rescue services are concerned in case of Esbjerg municipality, their time to reach at the site of accident in the municipality is depicted in the Figure 4.

Fire and rescue station of Vibevej 18, 6705 Esbjerg Ø is mainly responsible for urban area of the Esbjerg municipality. The emergency team is comprising of 7 rescues workers on 3 vehicles that is incident commander vehicle, fire truck and rescue truck. (Sydvestjysk Branvæsen 2015)

The green area of the map shows a response time of 10 minutes while yellow area represents 15 minutes of response time and rest of the area represents 20 minutes of response time. (Sydvestjysk Brandvasen 2015)

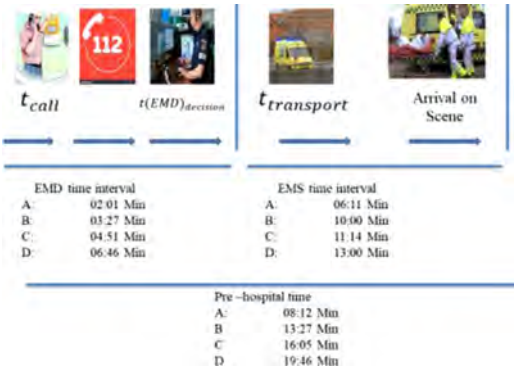


Figure 3. Current pre-hospital emergency time.

Figure 4. Esbjerg fire station response time to emergency calls.

RED can assist in emergency operations by reducing the FRS onsite assessment via video streaming by reaching faster than their time of 10 to 20 minutes depending on the location of accidents from FRS station.

2.2 *Identification of RED placement*

For optimal placement of RED networks, a spatial analysis was performed using geographic information system (GIS) tool ArcMap, and ArcGIS 10.5.1 to analyze and visualize the results. (Law & Collins 2015)

For application of RED network, DJI M210 Matrice drone is considered, which is one of the most advanced drone to date with broader industrial applications. The cruise speed of this drone with A mode is 82.8 kph. During vertical ascent, it has a speed of 5 m/s and vertical descent the speed is 3 m/s. (DJI 2017)

This drone can transmit footages with camera such a Zenmuse X4 and Zenmuse X5 s along with

Zenmuse Z30. The range of the drone is 7 km. Because of its agility, water proof and along with other specifications, it fits best to be considered as RED for building network to quickly assess the site of accident crashes and support EMD and EMS to make correct and quick decisions. RED can also assist fire and rescue team via video streaming.

Esbjerg municipality is considered for this explorative study. Esbjerg municipality is divided into urban and rural areas.

In Figures 5 and 6 optimal locations for placement of RED network are shown. To cover the urban area of the Esbjerg municipality five placements are identified to mount RED network. Similarly, five placements are also identified for rural area (Fig. 6). Each placement is the center point of the circle shown on the map. The drone range is 7 km; therefore, each circle represents 7 km of radius. The origin of the circle is for the docking station of the drone. As each UAV location covers a radius of 7 km, several traffic accident cases in the analysis are overlapping. A total *of* $n = 2515$ cases of traffic accidents were reported in Esbjerg municipality between 2012 and 2016. Out of these 2,029 were in the urban area and 486 were reported in the rural area of the municipality. For each location's (both rural and urban) median time, maximum time and minimum time is depicted in the following Tables 2 and 3.

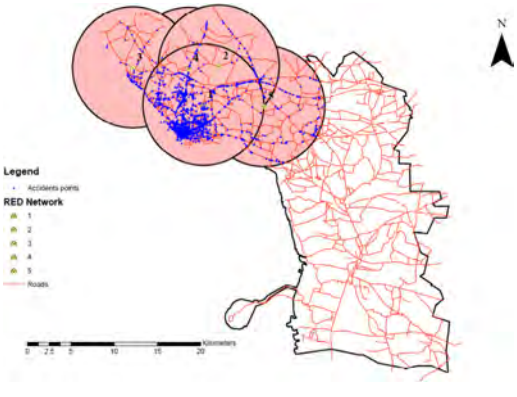


Figure 5. RED Network placement across Esbjerg municipality urban area.

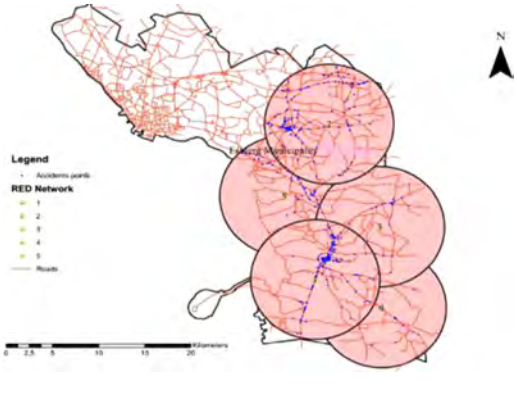


Figure 6. RED network placement across Esbjerg municipality rural area.

Each location is identified based on the number of accidents in the radius of 7 km of circle. The origin of the circle is the location for the drone placement. From drone placement to wherever accidents occurred in the circle, the distance is measured. The following formula is used to measure the distance between longitudinal and latitudinal coordinates.

$$d = \sqrt{(x_1 - x_2)^2 + (y_2 - y_1)} \tag{1}$$

Considering the speed of the drone total time between two locations is calculated and subsequently that total time is used to calculate the median time, maximum time and minimum time for each location. The preparation time for launching the drone is not considered, as this time is approximately 3 seconds. Claesson et al 2017) whereas airborne time of the drone is considered for median time calculation.

Maximum time for both urban and rural location is approximately 5 minutes 57 seconds, whereas minimum time and median time is varied across the locations.

Table 2. RED network median time to reach at the scene of accident in urban area.

Location urban	Median time	Maximum time	Minimum time
1	03:25	05:57	00:59
2	04:48	05:57	01:59
3	05:01	05:57	01:07
4	04:57	05:57	01:03
5	04:54	05:57	01:31

Table 3. RED network median time to reach at the scene of accident in rural area.

Location rural	Median time	Max time	Min time
1	03:21	05:55	01:05
2	04:22	05:56	00:53
3	04:53	05:57	01:39
4	04:03	05:56	01:01
5	04:05	05:57	01:47

3 DISCUSSION

Real time video from the scene of crash is powerful tool in supporting quick and right decisions. (Fig. 7) Due to many reasons bystander cannot clearly define the health status of the patient/injured persons. Danish medical staff supports the concept of live video streaming to deal with problem of over and under-triage. (Gerdstrøm 2017)

To have a safe operation of the live video streaming via RED, precautionary measures needs to be considered. There should not be any safety concerns for bystanders or any harm to the surroundings environment. DJI M210 has collision avoidance sensors, however, bystanders onsite must be informed of RED approaching to them. Moreover, rotors of RED should be shut down once the EMS or FRS reach at the site of accident.

Building and integrating RED Network in Denmark may bring new challenges for the emergency services to get training and implement the system as well as the interaction among EMD, bystander or inflicted persons at the site of crash.

There are some risks associated with this novel idea of RED network such as public perception of the drone technology, differentiation (colour/appearance) between the emergency drones and other drones for the public, risk of falling of a drone, risk of drone docking unit stolen or damaged, risk of data/information stolen, charging issues with the drone, bad weather and environmental effects of the drone technology etc.

Nevertheless, the pre-hospital phase of emergency services would benefit from RED due to live mutual visual inspection of the emergency. The real time video feed will help cut down the costs of the resources that are not needed at the site of crash. For instance, a procedural protocol to respond to a traffic accident alert involves dispatching of fire truck, incident commander vehicle and a rescue truck that may or may not be needed. Inter departmental and intra departmental communication of emergency services is expected to be improved. Another worth mentioning benefit of RED networks is that the implantation of the system to traffic accidents will pave the way to scalability and application of it to other emergencies.

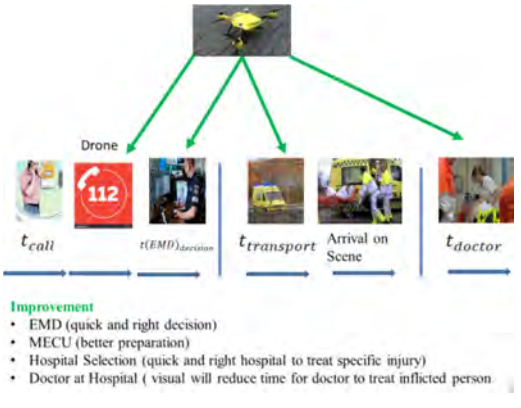


Figure 7. RED network assessment of emergency scene via live video.

4 LIMITATIONS

Beyond Visual Line of Sight (BVLOS), flight operations of drones are not allowed in Denmark.

It is important to know the acceptability of drone technology in local population, for which a there is a need to have comprehensive risk perception study.

DJI M210 drone was considered having a range of 7 km with area of network coverage of 14 km; many accidents in the analysis are therefore overlapping. The configuration of the drone along with range would have resulted different results if we would had considered another drone.

The real test flights are yet to be performed.

5 CONCLUSION

The application of GIS model results in the identification of appropriate placement of RED networks across Denmark. The real-time video transmission via RED networks can enable emergency services to take immediately right decision and dispatch a targeted response to treat injured persons. Therefore, RED could be the key to overcome the problem of under-triage and over-triage in saving lives besides cutting budgets.

ABBREVIATIONS

BVLOS: Beyond Visual Line of Sight; EMCC: Emergency medical communication center; EMD: Emergency Medical Dispatcher; EMS: Emergency medical services; FRS: Fire and Rescue Services; GIS: Geographical information systems; GPS: Global positioning systems; MECU: Mobile emergency care unit; RED: Rescue emergency drone.

REFERENCES

Aeryon skyranger (2017) Aeryon SkyRanger the benchmark for VTOL sUAS Aeryon skyranger" [Online] Available at: https://www.aeryon.com/aeryon-skyranger Accessed on: 14.04.2017.

Andersen, M.S. Johnsen, S.P., Hansen, A. E., Skjaerseth, E., Hansen, C. M., Sørensen, J.N., Jepsen, S.B., Hansen, J.B. & Christensen, E.F 2014. Preventable deaths follow-

ing emergency medical dispatch—an audit study. *Scand J Trauma Resusc Emerg* Medv.22; 2014PMC4293002

Andersen, M.S., Johnsen, S.P., Sørensen, J.N., Jepsen, S.B., Hansen, J.B. & Christensen, E.F. 2013. Implementing a nationwide criteria-based emergency medical dispatch system: A register-based follow-up study. *Scand J Trauma Resusc Emerg* Medicine 2013; 21: 53.

Claesson, A. Fredman, D. Svensson, L. Ringh, M. Hollenberg, J. Nordberg, P. Rosenqvist, M Djarv, T. Österberg, S. J. Lennartsson J. & Ban, Y. 2016. Unmanned aerial vehicles (drones) in out of—hospital-cardiac-arrest. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* (2016) 24:124 DOI 10.1186/s13049–016–0313–5

Claesson, A., Bäckman, A., Ringh, M., Svensson, L., Nordberg, P., Djärv, T., Hollenberg, J. 2017. Time to Delivery of an Automated External Defibrillator Using a Drone for Simulated Out-of-Hospital Cardiac Arrests vs Emergency Medical Services, JAMA. 2017;317(22):2332–2334. doi:10.1001/jama.2017.3957

Danish road directorate 2017. available at http://www.vejman.dk/da/Sider/default.aspx accessed on 15.10.2017

DJI 2017. Matrices 200 series. Available at: https://www.dji.com/matrice-200-series/payloads Accessed on 17.07.2017

Gerstrøm. G. 2017. Speciallæge. AnGus—Sundhed og PsykologiCVR:21841846

Kristensen, A.K., Mehmood. S., Ahmed, S., Ahsan, D. 2017. Rescue Emergency Drone (RED) for Fast Response to Medical Emergencies Due to Traffic Accidents. World academy of science, engineering and technology. *International journal of health and medical engineering* vol: 11, 2017

Law, M. & Collins, M. (fourth edition) 2015. *Getting to know ArcGIS*, 380 New York street Redland California 92373–8100: ESRI press.

Mehmood, S. & Ahmed S. 2017. Incorporation of Drones into Fire and Rescue Service of Esbjerg Muncipality for a Robout Response. A MSc Risk and Safety Management Thesis. Aalborg University, Esbjerg Denmark

Sánchez-Mangas, R., García-Ferrrer, A., Juan, A.D., & Arroyo. A.M.2010. The probability of death in road traffic accidents. How important is a quick medical response? *Accident Analysis and Prevention* 42 (2010) 1048–1056

Statistics Denmark 2017. Injured and killed in road traffic accidents by region, casualty, motor vehicles involved, age and sex. [Online] available at: http://www.statistikbanken.dk/statbank5a/default.asp?w = 1600 Accessed on: 10–12–2017

Statistics Denmark 2017. Population 1. January by municipality and time. Available at http://www.statistikbanken.dk/statbank5a/default.asp?w=1600. Accessed on: 02.12.2017

Sydvestjysk Brandvæsen 2015. Risikobasret dimensionering sydvestjysk brandvæsen. Available at: http://www.fanoe.dk/Files/Files/Dagsordner/committee_78303/agenda_237522/documents/364f243a-8d19-4f01-8134-d7d418a5c41d.pdf, Accessed on 20.05.2017

Transportministeriet (2010) værdisætning af transportens eksterne omkostninger, Rapport Juni 2010

Wissenberg M, Lippert FK, Folke F, Weeke P, Hansen CM, Christensen EF, Jans H, Hansen PA, Lang-Jensen T, Olesen JB, Lindhardsen J, Fosbol EL, Nielsen SL, Gislason GH, Kober L, Torp-Pedersen C. 2013. Association of National Initiatives to Improve Cardiac Arrest Management with Rates of Bystander Intervention and Patient Survival After Out-of-Hospital Cardiac Arrest JAMA. 2013;310(13):1377–1384. doi:10.1001/jama.2013.278483

World Health Organization 2017. Media center, fact sheet Road Traffic Injuries. [Online] Available at: http://www.who.int/mediacentre/factsheets/fs358/en/, Accessed on: 10–10–2017

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Swedish multi-level planning system for critical infrastructure protection: The regional core

C. Groβe
Mid-Sweden University, Sundsvall, Sweden

P.M. Olausson
Mid-Sweden University, Östersund, Sweden

ABSTRACT: With its growing dependence on electricity, modern society faces the risk of cascading failure of interconnected societal functions. To protect societal functions during an event of power shortage, Sweden has implemented a multi-level planning process called *Styrel*, which involves national-, regional—and local-level actors. As part of the Swedish crisis management system, the regional body operates as a co-ordinator that organises co-operation and interaction between private and public actors. This study examines the role of the regional hub in *Styrel* and the collaboration and co-operation between planning levels. It focuses on the co-ordinator's perspective and presents evidence from interviews and a survey among planners at County Administrative Boards, entrusted with the supervision and execution of *Styrel* within their regional area of responsibility. This paper indicates that the regional co-ordinator lacks the awareness, knowledge and resources to fulfil its core function in the national planning for critical infrastructure protection.

1 ELECTRICITY AND THE SWEDISH CRISIS MANAGEMENT SYSTEM

1.1 *Background*

Electricity is a vital resource in today's society, which largely depends on electricity for maintaining critical social functions. It can be argued that the reliable distribution of electricity is crucial for private households, businesses, and public operations to function and survive (Cohen 2010, Ghanem et al. 2016, Rinaldi et al. 2001). This dependency is likely to increase over time due to the continuous developments in important infrastructure such as railways and electric cars (Cedergren et al. 2015).

The power grid is vulnerable to various types of events, such as extreme weather conditions (e.g. storms and floods), technical failures due to outdated infrastructure and aging components, cyber-attacks and destruction. Disturbances in the grid can have severe consequences for society (Gheorghe et al. 2006, Pescaroli & Alexander 2016). For example, in Sweden, the storms Gudrun, Per, Dagmar and Ivar caused major problems that in some cases lasted for more than a month (EA 2006, 2007a, 2007b).

In the future, there is a risk that such extreme conditions will increase in number and magnitude due to the changing climate (Birkmann et al. 2016). Given the serious effects of such events on society, creating the necessary conditions for sustainable power supply during a crisis is an important function of the Swedish Energy Agency (EA). In order to ensure undisturbed power supply to important users in society, i.e. critical infrastructure (CI), the EA has developed a planning process called *Styrel* (an acronym for *control of power supply to prioritized electricity users*), to provide critical infrastructure protection (CIP) against short-term power shortages.

1.2 *Aim of the study*

The County Administrative Board (CAB) plays a central role in the Swedish *Styrel* process as co-ordinator (EA 2014). The aim of this paper is to examine the role of the regional hub of *Styrel* and the collaboration and interaction between planning levels that are included in the process. The focus is on the differences between CABs regarding their performance as co-ordinators in *Styrel*.

1.3 *The Swedish crisis management system*

The Swedish crisis management system depart from three principles: The first one is *the principle of responsibility*, which implies that actors who are responsible for an activity or a process in everyday life are also responsible for it during a crisis. Next, *the principle of parity* implies that societal functions

during a crisis should as far as possible be carried out in the same way as they are during normal conditions. The third *principle of proximity* states that actors closest to the event handles the crisis when it occurs; this means that a municipality or county/region should primarily handle a crisis. If local resources are insufficient, the state can act through the CAB (MSB 2014, Pramanik et al. 2015, Tehler et al. 2012). In practice, this means that the CAB is responsible for co-ordinating between relevant actors in their county (MSB 2014). The co-ordinating role may involve some problems, as there is no explicit process for resolving possible conflicts within the Swedish crisis management system.

A study of the Swedish defence directors at the 21 CABs in Sweden has revealed what the problems are (Wimelius & Engberg 2015). According to the study, clearer governance, improvement in network management and increase in resources are measures that can help to improve co-operation among the various players in the county. Several defence directors expressed the view that the Swedish crisis management system is characterised by weak governance and lack of continuity (Wimelius & Engberg 2015). A study of the river groups in Northern Sweden further substantiated this view. The river groups exchange information in events such as floods and high flows through co-operation via networks. However, vague instructions from the Swedish Rescue Services Agency have resulted in the different river groups working differently, having different objectives, and involving different actors (Olausson & Nyhlén 2017). All these reports point to the need for a more integrated and standardised system when it comes to crisis management in Sweden.

2 THEORETICAL FRAMEWORK

This study focuses on the planning process for power shortages, *Styrel*, which involves both public and private actors (Große 2017). Pierre & Peters (2000) consider the management of society as a continuum that extends from traditional top-down control, at the one end, to self-organisation (autopoesis) and networks at the other end. The concept of governance is the common element of the entire continuum. In social sciences, the concept of governance has no clear definition, in which regard Pierre & Peters (2000) note:

'…Sufficiently vague and inclusive that it can be thought to embrace a variety of different approaches and theories, some of which are even contradictory' (Pierre & Peters 2000: 37).

Governance can be regarded as a policy instrument in the context of institutionalism, rational choice, and network and policy communities, or it can be analysed based on neo-Marxist and critical theories. The concept of governance describes how a society is organized, governed and who is involved in dialogue, participation, and networking. According to both governance and public policy theories, networks are an important phenomenon (e.g. Christopoulos & Ingold 2011, Henry 2011, McGinnis 2011, Petridou 2014). In this study of *Styrel*, we use the definition of governance as a policy instrument and subsequently as a network for steering. *Styrel* can also relate to the concept of risk governance, which considers legal, institutional, social and economic contexts as well as the actors involved in each of these contexts (Renn 1998).

Governance or policy networks can be either self-organized or created and co-ordinated by the state (Sørensen & Torfing 2005). Individual organizations often use networks to achieve their strategic and operative objectives, to maximize their influence over outcomes or to avoid dependence on other actors in the system. From this perspective, governance involves managing networks (Rhodes 1996).

This study examines material from interviews and a survey of planners at CABs to portray the CABs' central role in the Swedish planning system. The analysis was based on the concept of complex systems governance, the aim of which is to ensure control, communication, co-ordination and integration of a complex system by several metasystem functions (Keating et al. 2014). In particular, the focus is on two functions of complex systems governance:

- *Policy and Identity*
- *Information and Communications.*

The aim of focusing on these two functions is to inform other functions of complex systems governance, such as learning and transformation and the operational performance of the Swedish crisis management system and its governance, i.e. the metasystem (Keating et al. 2015, Keating et al. 2017, Keating & Bradley 2015).

- Policy and Identity

The role of policies is to provide direction and identity to the system components, e.g. the planners in the Swedish *Styrel* process, and to represent the system to external constituents, e.g. the Swedish crisis management system and the wider public.

- Information and Communications

Secure and reliable information paths are particularly important in national planning for CIP. However, access to relevant information for decision-making is similarly vital for the performance

of the planning system, as is the consistent interpretation of available information throughout multi-level planning, such as in the case of *Styrel*.

This study examines the available evidence in light of these governance functions and highlights problems in the design, execution and evolvement of the Swedish multi-level planning system for CIP, in order to inform further development of this complex system and its governance.

3 METHOD AND SELECTION OF CASES

In this study, we use interviews with co-ordinators at the CABs in three counties in Sweden: one in the rural north, one including one of the three major cities in Sweden, and one including some heavy industry close to the capitol of Sweden.

This study further includes a survey with all the co-ordinators at the 21 CABs in Sweden, carried out in October 2017. Until today, 15 of these co-ordinators have responded to the survey, which means that the participation rate is 71.4%. These 15 participants provided answers to 34 questions on their perceptions of the effectiveness and efficiency of the planning in general and on the proceedings during the last planning process iteration within their area of responsibility in particular. The survey has an overall response rate of 62.2%; the answers to the remaining questions were *do not know (N/A)*.

A document study complemented the interviews and the survey and provided important background information, which allowed for data triangulation (Gerring 2007). The documents for study included a handbook for the planning process (EA 2014), evaluations of the pilot study in 2008 (Länsstyrelsen Blekinge 2009, Dalarna 2009) and evaluations of the first round of planning in 2010 at the national level (EA 2012) and in Stockholm County (Länsstyrelsen Stockholm 2012). Moreover, a report on the grid operator's plans for manual load shedding (MFK) completed the document study (Veibäck et al. 2013). We conducted the interviews after the document study, which deepened the information gained from the documents and allowed for verification of the evidence from the documents in the interviews.

4 SWEDISH PLANNING FOR CRITICAL INFRASTRUCTURE PROTECTION–*STYREL*

In Sweden, different actors are responsible for energy supply at the national level. The EA is responsible for creating the conditions for efficient, resilient, and sustainable energy use and cost-effective distribution of Swedish energy (EA 2012). The Swedish Energy Markets Inspectorate (EI) is responsible for supervision, regulation and licensing in the energy market. The Swedish Civil Contingencies Agency (MSB) bears the overall responsibility of the crisis management system and the measures taken before, during and after an emergency or crisis. Finally, the Svenska Kraftnät (SvK) is nationally responsible for the power grid. When a power shortage occurs, the SvK is responsible for MFK in as informed and socially efficient a way as possible. The mission is to ensure that local and regional power grid operators can perform such MFK within 15 min (EA 2012, Veibäck et al. 2013).

In order to enable the national, regional and local grid operators to run an MFK without affecting critical social functions, the four national-level actors (the EA, the EI, the MSB and the SvK) have developed the planning process *Styrel*. The planning and prioritisation process for power shortages has been used 2010 and then repeated in 2014. The next planning process iteration will take place in 2019. In *Styrel*, the CAB acts as co-ordinator between governmental agencies and municipalities, on the one hand, and the municipalities and power grid companies on the other, as Figure 1 depicts.

During the recent planning in 2014, the following multi-level process was agreed upon (EA 2014):

With the aid of an eight-digit scale for prioritisation of CI (see Table 1), national agencies identify and prioritise the CI that each of them operate.

In step (1) (see Fig. 1), each agency sends a portion of these ranked objects to the CAB of the regional area of responsibility in which the CI object is located. Each CAB merges the received

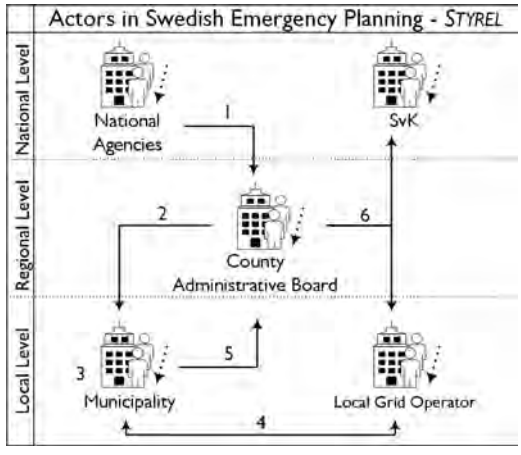


Figure 1. Actors and information paths in the Swedish multi-level planning process for CIP against power shortages.

Table 1. Priority classes of critical infrastructure.

Class	Description
Electricity consumers that have/represent:	
1	significant impact on life and health—short-term (hours)
2	significant impact on society's functionality—short-term (hours)
3	significant impact on life and health—long-term (days)
4	significant impact on society's functionality—long-term (days)
5	significant economic value
6	significant importance for the environment
7	significant importance for social and cultural values
8	others

lists of prioritised objects and divides them into portions that correspond with each municipality's area of responsibility. In step (2), the CAB forwards these portioned lists to each municipality. In step (3), the municipalities make an inventory of locally important infrastructure and prioritise the objects in accordance with the list in Table 1.

In step (4), the municipalities exchange information on the prioritised consumers with each locally operating power grid provider, which provides information on the technical feasibility of control. The CI objects merges into controllable power lines. Thereby, the used spreadsheet performs additive aggregation of the objects' ranking scores, which yields another list that contains the ranking of the power lines. After a final evaluation, the municipalities send this latter list back to the CAB in step (5). Each of the CABs merge these lists from the municipalities in their jurisdiction, resolve conflicts between lines that cross municipal or regional borders and make the final decision about the ranking of power lines. In step (6), the CABs send the final document to the SvK and dedicate portions of it to each power grid provider that operates in the region.

5 RESULTS OF THE STUDY

5.1 *Analysis of the reference process model*

The Swedish planning process for CIP involves actors from a large number of national agencies—all the CABs and municipalities and locally, regionally and nationally operating power grid providers. In Figure 1, the CAB makes two appearances as co-ordinator of the proceedings. The *Styrel* process can therefore be considered as a multi-agency planning process (Alexander 2015, Bharosa et al. 2010) that occurs at multiple hierarchical levels (Allouche & Berger 2011). This Swedish multi-level planning system consists of three hierarchical levels—the local, the regional and the national level.

The *Styrel* process can be decomposed into single problems at each level, at which responsible planners act on behalf of public or private organisations in sequential order, while the CABs play a central role as co-ordinators of the planning decisions. This role is directed top down and bottom up, but the latter role is incomplete because the procedure lacks co-ordination at the national level.

In the top-down part of the sequence (step (1) & (2)), a CAB receives information on an electricity-dependent CI that national agencies operate in the CAB's regional area of responsibility. The survey results in Tables 2 and 3 indicate that although the CABs perceived the collaboration with national agencies as good, 84.6% of the CABs stated the need for a more structured process for this activity, particularly for the consistent interpretation of priority classes. Further, on in the process, the CAB portions the information on national objects and sends them to each of the municipalities in its juris-

Table 2. Co-ordinators' perceptions of *Styrel*.

	Median	0	1	2	3	4	5
Population: 21, Response rate: 71.4%, n = 15							
Participation *Styrel:* never: 58.3%, once: 25%, twice: 16.7%							
Proceedings of *Styrel*: Knowledge: 42.5%, Perception: 78.9%							
Importance of *Styrel*	4.64	0	0	0	1	3	10
Usefulness in crisis mngt	3.80	0	2	0	2	6	5
CIP in power shortages	3.00	1	0	0	3	3	0
Collaboration with							
• National agencies	3.11	0	1	1	3	4	0
• Municipalities	4.00	0	0	1	3	4	5
Trust in							
• National agencies	2.90	0	1	2	4	3	0
• Municipalities	3.55	0	1	0	3	6	1
• Energy Agency	3.90	0	1	2	4	3	0
Impact of CABs' work	2.45	0	3	2	4	1	0
Knowledge of *Styrel*	2.36	3	2	1	4	3	1
Level of system control	3.23	1	1	2	2	4	3
Good information access	3.22	0	1	2	2	2	2
Good information security	2.75	2	1	2	1	5	1
Clear information paths	3.10	1	0	1	3	5	0
Good resource access	2.67	1	2	1	6	0	2

Note: Scale running from 0 (don't agree) to 5 (totally agree).

Table 3. Co-ordinators' experiences with *Styrel*.

	Yes	Mostly	No	N/A
Request for clearer processes				
• with national agencies	84.6%		15.4%	
• with municipalities	69.2%		30.8%	
• with power grid providers	53.9%		38.5%	
Followed the handbook	16.7%	41.7%	0.0%	41.7%
Regular meetings	15.4%	42.9%	0.0%	38.5%
… was handled of	Municip.	Collaboration	CAB	N/A
National/ regional CI	7.7%	30.7%	15.4%	46.2%
Final compilation	0.0%	25%	38.6%	33.4%
Cross-local lines	0.0%	25%	25%	50%
Cross-county lines	0.0%	30.7%	0.0%	58.3%

diction that host such assets. In addition, according to the reference process model, the CABs should provide training and guidance to their municipalities during subsequent planning at the local level. Since the questions on concrete proceedings had a low response rate of 57.5% during the survey, it is possible that the knowledge within the planning system is stunted. In addition, 58.3% of the CABs have not participated in the planning process before. This may influence their ability to co-ordinate the proceedings and to provide guidance to the municipalities. Nevertheless, the CABs' responses with regard to collaboration with the municipalities were slightly positive and indicated that they rather trusted the municipalities. However, the reference process did not provide any measures to evaluate the correctness of the planning decisions, so aside from communication, the CABs have no means of assessing the information they receive—neither in the top-down nor in the bottom-up phase.

In the second part of the sequence (steps (5) & (6)), the information flow is in the bottom-up direction. Information about local prioritisation also comprises the national CI assets, but they are masked. During the recent iteration of the planning process, information exchange was limited to power lines and the number of objects per priority class. Even though this reduction in information may ensure a certain level of information security, it makes regional—or national-level co-ordination impossible. The study results indicate that the CABs used meetings as a means to gain more information and to align the prioritisations in their area of responsibility. Nevertheless, 69.2% of the CABs stated that they require a more structured process for collaboration with municipalities. Moreover, each CAB must also merge the lists from the municipalities and then decide upon the regional ranking of power lines. Half of the CABs that answered this question decided to merge the lists entirely on their own. Further, one CAB performed the merging by itself and announced the changes to the concerned municipalities. The remaining respondents stated that they co-operated with the municipalities to align the results and to compile the final ranking list. Finally, the CAB divides this list into bundles of power lines that correspond to each local power grid operator in the region and sends this list to each of them. In addition, each CAB delivers a complete list to the national power grid provider. Interestingly, even though the process does not necessitate intensive collaboration of the CAB with the grid providers, 53.9% of the CABs stated that they required a more structured process anyway. This is probably because the CABs do not receive any feedback from the power grid providers about next-level planning for MFK because of national information security concerns.

Due to the immense information processing and process management involved, CABs bear a double burden—as participants in the process and as regional co-ordinators. Hence, the CAB represents the central hub in the current multi-level *Styrel* planning process in Sweden.

5.2 *Organisation and execution of Styrel*

Each CAB is responsible for co-ordinating work related to crisis management in its own county in Sweden. Therefore, the CAB is also responsible for co-ordinating the execution of *Styrel*, in which the CAB plays the central role in the planning approach, but with little influence on the quality of the process outcome. The evaluation after the pilot and the first round of planning showed overall, the CABs perceive *Styrel* as an important planning process for identifying CI. The survey substantiates this perception, as 92% voted on *agree*/*strongly agree*. However, due to the limited influence of the CABs on the outcomes of the *Styrel* planning and the subsequent MFK planning, the CABs expressed some doubt about the usefulness of *Styrel*'s outcomes for crisis management. Further, they expressed considerable doubt about whether *Styrel* can provide the intended protection for society during a power shortage.

The interviews show that the three CABs organised their work according to the reference model. All three CABs emphasise the importance of working within existing networks. In particular, the CABs used already existing networks, used

in ordinary work with crisis management and emergency response. No new networks emerged in the three regions. Evaluation of the first run in 2010 indicates that the CABs acknowledged the *Styrel* process' contribution to improved cooperation within existing networks (EA 2012). However, the organisation of these networks differs between the three counties, and the counties' size seems to be the main reason for the differences. The two smaller counties worked more closely together, e.g. meetings included representatives from all municipalities. The larger county also used existing networks meetings. In this case, the county divided into four or five different groups; northeast, north-west, southeast, south-west, and the large city. This division ensured a smoother planning process in the region, but it also made it difficult for the municipalities to have an understanding of the region as a whole. Instead, individual municipalities had only experienced the discussion in their part of the region, which could lead to differences in principles and priorities among the four parts. According to the evaluation of the first run in 2010, the major challenge was to find a common view on the prioritisations among municipalities in a region. Thereby, how to deal with the dependence chains and to which extent an analysis of these chains is appropriate seem unclear (EA 2012). In *the rural north county* and *the county close to Stockholm*, all municipalities participated in the discussions on principles and priorities.In the latter one, the municipalities made notes in the planning document, which made it easier for the CAB to identify the objects along the line. This could also have impact on the result: '*For the result then ... because we have a bundle of power lines, without knowing what is on them, it is extremely difficult. Because you could, in theory, cut off the hospital using a few ICA stores, or some water pumps*'. The notes made it possible for the CAB to identify such effects.

This study evinces that the CAB in general has followed the planning model as stated in the handbook for *Styrel*. However, there were some deviations from the model due to lack of time. Since some actors did not follow the predetermined schedule, the CABs ran out of time for their part of the process. Although the other actors in the process caused this delay, the three CABs perceive the initial plan as too optimistic. The co-ordinators argued that there was a risk that such a compressed schedule, which speeds up the work of municipalities and CABs, led to a widespread copy-and-paste behaviour in the municipalities: '*It may be necessary to give more time because it became very stressful when it became so delayed in the first line from government agencies*'.

Evaluation of the *Styrel* planning process in 2010 revealed that there were only a few, if any, contacts between CABs and private actors, except for contacts with the power grid providers (EA 2012). Due to time constraints, the interviews with the three CABs indicated that no other private actors or actors representing civil society have been involved in the current *Styrel* planning.

Between the two rounds of planning, the CABs' role in the process changed. In the first one, the CAB participated more actively in assessing and balancing the priorities of the CI objects at the county level, whereas in the second planning, the CAB only complied the results from the municipalities. One of the counties did not fully apply this change; instead, the municipalities, the region and the CAB made the final ranking list together. The participating municipalities were, according to the co-ordinator, unanimous about this departure from the official planning process: '*Yes, in what other way would we do? It's just like a damn long list*'.

5.3 *Integration and governance*

Styrel is an integrated part of the Swedish crisis management system. As stated, the three principles of the system are *responsibility*, *parity*, and *proximity*. The CAB is responsible for co-ordinating work with the system at the regional level. Critique from CABs against the *Styrel* process mainly includes problems with the process itself and the lack of feedback during the process in the multi-level system.

In the interviews, the co-ordinators at the CABs all agreed that it is important to identify CI objects, i.e. societally important objects, in advance in order to ensure that there is as much power supply as possible to these CI objects in the event of a power shortage. Therefore, there are certain elements of *Styrel* that are important for the functioning of society. However, all three CAB co-ordinators interviewed are critical about the design of the reference process model and process execution in the two rounds. They are also critical about, the limits of the usefulness of the planning process. Today, the process stands to some extend for itself; therefore, the co-ordinators regret the absence of a holistic, integrated view on *Styrel*. One co-ordinator envisioned that integration and transition of the planning process of *Styrel* would be an important pay-off to the Swedish crisis management system at subsequent planning levels, such as preparedness and contingency planning.

In two of the counties, the co-ordinator at the CAB described the process as smooth without any major conflicts between the included parties. The problem was primarily that the CAB, according to changes in the process in the second round of planning, could not access information about the objects themselves, but only the lines. All three co-ordinators at the CAB emphasised on the problems of this change. Since the co-ordinators did not get information about individual objects along high-

priority lines, there is a risk that important objects is down prioritised due to the process design. In all three cases, the co-ordinators preferred to have more information about the objects in order to ensure the quality of the process outcome.

There were solutions to deal with the problems. In one case, the CAB and the municipalities first discussed how to grade a certain objects. Then the municipalities made notes in the planning document indicating which objects are located along the line. In another case, the CAB, the municipalities, and the region made the final ranking together. In the third case, they only used initial discussions, but there were no discussions on individual objects. In theory, this could imply a down prioritization of the line for the major hospital in favour for other lines. Finally, before submitting the final ranking list, the CAB ensured that they were along one of the highest prioritised lines.

6 DISCUSSION AND CONCLUDING REMARKS

6.1 *Policy and identity*

From this study, it appears that *Styrel* is part of a reliable energy supply plan, even in the event of power shortages. However, the findings also indicate that the execution of *Styrel* does not follow on the process created by the EA. The risk of 'copy-and-paste' behaviour can particularly affect the prerequisites for reliable power supply during a power shortage. In accordance with the concept of resilience, this study on the co-ordinating function of the CAB highlights that there is a risk that society cannot maintain important social functions. The implementation of the process does not provide any guarantee for a resilient power supply. Further, any form of systematic co-operation between the system components, such as private and public actors, seems to be absent in the current *Styrel* process. Systematic co-operation, if any that occurs at the municipal level remains to study.

The importance of private-public co-operation in networks for enabling actors (i.e. the municipalities, regions, CABs and power grid providers) to identify and prioritise CI objects is further emphasised by the evaluation of the three pilot studies in 2009. The results from our current study reveal that none of the three CAB formed new networks for the process.

The findings signify the underrepresentation of private actors and actors representing civil society in the planning process and its reference model, developed by the EA. However, the deliberately vague definition of the reference model allows municipalities and CABs to include private actors to obtain as much information as possible for the ranking of CI objects. Thus, the system permits components to adapt to local regional commitment to improve the process. This means that the regional outcomes of a process instantiation can vary distinctly, which questions the national character of the planning. Particularly, since *Styrel* prescribes neither an over-regional nor a national alignment of CI and the power lines, it remains uncertain how local and regional proceedings during the planning affect CI objects of over-regional and national importance.

Although the policy is accepting alternative proceedings, the CABs used already existing networks, which only include public actors, mainly at the municipal level making the proceeding more effective. However, such an approach carries the risk that important information from private actors, such as private care providers, is lost ignoring proper risk communication to society.

6.2 *Information and communications*

This study implies that it is important that specific public actors, such as persons responsible for crisis management, have authorised access to crucial information on power lines that ensure power supply to CI objects, such as different care providers. It seems that there is no guarantee that the actors update available information, due to the earlier mentioned 'copy-and-paste' behaviour. Moreover, due the limited information content in the received lists, the CAB cannot control the correctness and completeness of the CI objects; instead, it has to rely on the performance and commitment of other actors.

Styrel can contribute to the maintenance of CI during power shortages, but there is no proof of its success in this role due to the absence of any assessable success factors. This presupposes that actors at the municipal level execute the planning in accordance with the national strategic objectives. However, the interviews in this study reveal that in some cases, individual interpretations of these objectives resulted in an adapted, time-saving behaviour, i.e. 'copy-and-paste' of local results from the first planning round. Since there is no way of ensuring that the available data on CI objects from the previous planning also applies four years later, there emerges a risk that the results of *Styrel* do not properly reflect the intentions and priorities of municipalities and agencies.

In addition, the absence of specific feedback from power grid providers on the planned proceedings during a power shortage hampers further reliable integration of *Styrel* in regional crisis management. These preconditions illustrate that the regional co-ordinator cannot rely on the results of *Styrel* planning for CIP in subsequent planning processes, such as preparedness and continuity planning.

7 CONCLUSIONS

Styrel as planning process does not necessarily contribute to the creation of a reliable energy supply as stated by the governmental guidelines of the EA. *Styrel* can contribute to the maintenance of CI and societally vital services, but it is difficult to gauge this in the absence of assessable success factors.

It appears that there are no integration of the *Styrel* process into the Swedish crisis management system. Such integration may further improve the effectiveness and efficiency of this complex multi-level planning system. In particular, such integration could facilitate the further development of co-ordinated information paths and directed communication. In turn, such development can assist with ensuring adequate national and international information security with regard to sensitive information about CI. Therefore, it is necessary for authorised persons to designate and monitor the necessary information with confidentiality, integrity and availability to fulfil strategic and operative objectives in the context of national CIP.

The present analysis shows that the results of the *Styrel* process, implemented in a Swedish multi-level planning system, rely on the commitment of the CABs as the co-ordinator for achieving a common understanding of the criticality of infrastructure and for mediating regional collaboration. The level of trust between the different levels of the planning system seems likely to further influence the resulting emergency response plan. Moreover, the planner's perceptions regarding the significance of the planning task, the likelihood of a power shortage situation and the crisis management capability of a county can have an impact on the effectiveness of the complex multi-level planning system in a crisis.

This paper also indicates that there is a lack of awareness at the regional level about the function of core players in the Swedish *Styrel* approach. In addition, the regional hub lacks the knowledge and resources to fulfil adequately its dedicated function in the national planning process for protecting CI objects from the consequences of a power outage.

With insights from the Swedish case, this paper highlights the regional core of *Styrel* and contributes thereby to international discussions on the identification, prioritisation and protection of CI objects.

ACKNOWLEDGEMENTS

The study is supported by the Swedish Energy Agency alongside the project 'Från myndighet till medborgare och tillbaka', which is gratefully acknowledged.

REFERENCES

Alexander, D. 2015. *Disaster and Emergency Planning for Preparedness, Response, and Recovery.* Oxford University Press.

Allouche, M. K. & Berger, J. 2011. Collaborative Multi-Level Plan Monitoring. *Journal of Defense Resources Management*.

Bharosa, N., Lee, J. & Janssen, M. 2010. Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises. *Information Systems Frontiers* 12(1): 49–65.

Birkmann, J., Wenzel, F. & Greiving, S. et al. 2016. Extreme Events, Critical Infrastructures, Human Vulnerability and Strategic Planning: Emerging Research Issues. *Journal of Extreme Events* 03(04): 1650017-(1–25).

Cedergren, A., Johansson, J., Svegrup, L. & Hassel, H. 2015. Local Success, Global Failure: Challenges Facing the Recovery Operations of Critical Infrastructure Breakdowns. In: Podofillini, L., Sudret, B., Stojadinović, B., Zio, E. & Kröger, W. (eds.) *Safety and reliability of complex engineered systems: Proc 25th Europ Safety and Reliability Conf, ESREL 2015, Zürich, Switzerland*. Taylor & Francis, London: 4343–4348.

Christopoulos, D. & Ingold, K. 2011. Distinguishing between political brokerage & political entrepreneurship. *Procedia—Social and Behavioral Sciences* 10: 36–42.

Cohen, F. 2010. What makes critical infrastructures Critical? *International Journal of Critical Infrastructure Protection* 3(2): 53–54.

Gerring, J. 2007. Is There a (Viable) Crucial-Case Method? *Comparative Political Studies* 40(3): 231–253.

Ghanem, D. A., Mander, S. & Gough, C. 2016. "I think we need to get a better generator": Household resilience to disruption to power supply during storm events. *Energy Policy* 92: 171–180.

A. V. Gheorghe, M. Masera, D. L. Vries & M. Weijnen (eds.) 2006. *Critical Infrastructures at Risk: Securing the European Electric Power System*. Springer, Dordrecht.

Große, C. 2017. Applying Systems Thinking onto Emergency Response Planning: Using Soft Systems Methodology to Structure a National Act in Sweden. In: *Proc 6th Int Conf on Operations Research and Enterprise Systems ICORES*. SCITEPRESS: 288–297.

Henry, A. D. 2011. Ideology, Power, and the Structure of Policy Networks. *Policy Studies Journal* 39(3): 361–383.

Keating, C. B. & Bradley, J. M. 2015. Complex system governance reference model. *International Journal of System of Systems Engineering* 6(1/2): 33–52.

Keating, C. B., Katina, P. F. & Bradley, J. M. 2014. Complex system governance: Concept, challenges, and emerging research. *International Journal of System of Systems Engineering* 5(3): 263–288.

Keating, C. B., Katina, P. F. & Bradley, J. M. 2015. Challenges for Developing Complex System Governance. In: Cetinkaya, S. & Ryan, J. K. (eds.) *Proceedings of the 2015 Industrial and Systems Engineering Research Conference*: I1401.

Keating, C. B., Katina, P. F., Jaradat, R.'e., Bradley, J. M. & Gheorghe, A. V. 2017. Acquisition System Development: A Complex System Governance Perspective.

Länsstyrelsen Blekinge 2009. *Styrel. Slutrapport: Länsförsök Blekinge 2009*. 20090924, Karlskrona.

Länsstyrelsen Dalarna 2009. *Styrel: Länsförsök Dalarna 09 – Slutrapport,* Borlänge.

Länsstyrelsen Stockholm 2012. *Styrel i Stockholms län: – planeringsprocessen 2011*. Rapport 2012:12, Stockholm.

McGinnis, M. D. 2011. Networks of Adjacent Action Situations in Polycentric Governance. *Policy Studies Journal* 39(1): 51–78.

MSB—Swedish Civil Contingencies Agency (MSB) 2014. *Gemensamma grunder för samverkan och ledning vid samhällsstörningar: MSB777,* Karlstad.

Olausson, P. M. & Nyhlén, J. 2017. Organization and Decision-Making in Enforced Networks: The River Groups in Northern Sweden. *Journal of Contingencies and Crisis Management* 25(4): 313–325.

Pescaroli, G. & Alexander, D. 2016. Critical infrastructure, panarchies and the vulnerability paths of cascading disasters. *Natural Hazards* 82(1): 175–192.

Petridou, E. 2014. Theories of the Policy Process: Contemporary Scholarship and Future Directions. *Policy Studies Journal* 42(1): S12-S32.

Pierre, J. & Peters, B. G. 2000. *Governance, politics and the state*. Macmillan, Basingstoke.

Pramanik, R., Ekman, O., Hassel, H. & Tehler, H. 2015. Organizational Adaptation in Multi-Stakeholder Crisis Response: An Experimental Study. *Journal of Contingencies and Crisis Management* 23(4): 234–245.

Renn, O. 1998. Three decades of risk research: Accomplishments and new challenges. *Journal of Risk Research* 1(1): 49–71.

Rhodes, R. A. W. 1996. The New Governance: Governing without Government. *Political Studies* 44(4): 652–667.

Rinaldi, S. M., Peerenboom, J. P. & Kelly, T. K. 2001. Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems Magazine* 21(6): 11–25.

Swedish Energy Agency (EA) 2006. *Stormen Gudrun: Vad kan vi lära av naturkatastrofen 2005?* ET 2006:2.

Swedish Energy Agency (EA) 2007a. *Utvärdering av stormen Per: Konsekvenser och lärdomar för en tryggare energiförsörjning.* ER 2007:37.

Swedish Energy Agency (EA) 2007b. Utvärdering av stormen Per—aktörsvisa sammanställningar av intervjuer och analyser. Dnr 17–07–2831.

Swedish Energy Agency (EA) 2012. *Slutrapport från Energimyndighetens styrel-projekt: ER 2012:04.*

Swedish Energy Agency (EA) 2014. *Styrel: Handbok för styrels planeringsomgång 2014–2015.* ET2013:23.

Sørensen, E. & Torfing, J. 2005. The Democratic Anchorage of Governance Networks. *Scandinavian Political Studies* 28(3): 195–218.

Tehler, H., Brehmer, B. & Jensen, E. 2012. Designing societal safety: A study of the Swedish crisis management system. In: *Proc PSAM 11 / ESREL 2012, Helsinki, Finland*: 4239–4248.

Veibäck, E., Stenérus Dover, A.-S., Fischer, G. & Lindgren, J. 2013. *Elnätsföretagens MFK-planering: En studie av elnätsföretagens möjligheter att genomföra manuell förbrukningsfrånkoppling baserad på Styrel.* FOI-R--3797--SE.

Wimelius, M. E. & Engberg, J. 2015. Crisis Management through Network Coordination: Experiences of Swedish Civil Defence Directors. *Journal of Contingencies and Crisis Management* 23(3): 129–137.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Identifying hazards to include in risk analyses

M. Leonhardsen & O.E. Olsen
University of Stavanger, Stavanger, Norway

A.S. Nilsen
UiT Arctic University of Tromsø, Tromsø, Norway

ABSTRACT: A risk analysis should provide decision makers with information regarding relevant hazards. The initiating phase, where the risk analysts identify hazards to be included in the risk analysis, lays the foundation for the rest of the analysis. This phase is, therefore, of great importance. In this paper, we examine how risk analysts in a municipal setting identified potential adverse events and how they chose which ones to analyse in the risk analysis. The municipalities under study had important similarities with respect to exposure to hazards and government regulation. With these similarities as a starting point and studying how the initiating phase took place, the paper focuses on impact regarding the uniformity of adverse events. Looking at events included in the Comprehensive Risk and vulnerability Analyses (CRAs), seems to reveal a predominance of uniformity. This is reasonable given the previously mentioned similarities. It is arguably also a result of many risk analysts using the same sources to retrieve ideas of potential hazards. The latter is alarming when considering risks not listed in these sources, like emergent or local risks.

1 INTRODUCTION

Communities at different levels of society strive to attain safety and security. To do so, they try to identify hazards and threats that pose a risk. This is a starting point in preparedness for emergency response and risk and vulnerability reduction (Perry & Lindell 2003). Risk Analyses (RAs) are the prominent methods in which risks and vulnerabilities are identified and assessed. They are formal and analytical (Renn 1998; Rausand & Utne 2009) and used by organizations to prepare for misfortune. They do so by providing decision makers with information about relevant hazards and threats, the likelihood of these adverse events and their potential consequences. RAs enable decision makers to make informed decisions regarding reduction of risks and vulnerabilities (Aven 2011).

The mission of RAs is in other words (Rausand & Utne 2009):

- To figure out what kind of adverse events might happen
- To figure out the likelihood of the events
- To figure out the consequences of the events
- To describe the risks (Aven 2015)

The bulleted list clearly illustrates that beneficial outcomes of RAs depend on the initiating phase, which is identifying the adverse events of relevance to the RA.

"This is one of the most important steps in the risk analysis. If a hazard source or an adverse event is not detected, it will not be included in the analysis" (Rausand & Utne 2009, p. 86).

Likewise, Cameron et al. (2017, p. 53) describe the identification of hazards as "the first and most crucial step in any risk assessment". According to Renn (2008), stages of risk assessments vary depending on risk domains and risk sources. Regardless of that, hazard identification is one of three core elements in risk assessment (Renn 2008). Not being able to identify hazards properly can result in accidents or adverse events (Cameron et. al 2017).

Risk analysis has received a lot of academic attention. A December 2017search for "risk analysis" in the Academic Search Premier database, resulted in approximately 58800 academic articles. Comparatively, searching for "hazard identification" or "identification of hazards" resulted in 1100 and 1250 hits. This paper is a supplement to the studies of this highly important element of risk analysis.

This paper focuses on the initiating phase of RAs. It presents how risk analysts in 12 municipalities identified and chose hazards to be analysed in greater detail in their RAs. These municipalities had several similarities (they are presented in section 3). With these similarities as a starting point, can we categorise the risk analysts who carried out

the work as either copycats at one end of the scale, or explorative analysts at the other?

The main focus, though, is the impact of the approaches used in the initiating phase. This impact is identified and discussed, restricted to uniformity of risks included in the RAs. To be more precise: we have studied the processes when conducting so-called Comprehensive Risk and vulnerability Analyses (CRAs).

Municipalities are exposed to both hazards and threats. The nuances between the two terms are not of importance in our study. So, for convenience, we use "hazard" as a common term for both. Further, we use the terms hazards and adverse events in an interchangeable manner, even though hazards do not necessarily lead to adverse events.

1.1 *The CRAs*

The objective of the Civil Protection Act 2011 and secondary law is to ensure that municipalities safeguard the safety and security of the population (Directorate for Civil Protection 2017). According to these legal requirements, the Norwegian municipalities must have CRAs.

The objective of the Civil Protection Act 2011 and secondary law is to ensure that municipalities safe-guard the safety and security of the population (Di-rectorate for Civil Protection 2017). According to these legal requirements, the Norwegian municipali-ties must have CRAs.

The secondary law lists a few minimum requirements for the CRA. Two of them are of importance for the initiating phase of CRAs (Directorate for Civil Protection 2017):

- First, a CRA must address both existing and future risks in the municipality, as well as external risks of relevance to the municipality.
- Secondly, critical functions in society and critical infrastructure must be addressed. Loss of electricity or water can be examples.

Beyond that, the legal requirements do not specify what kind of adverse events be included in CRAs. Risk analysts in the municipalities must identify the potentially adverse events based on idiosyncratic risks in their communities.

There is a variety of methods for risk analysis (Rausand & Utne 2009). Analysts in the municipalities are free to choose, but preliminary RAs are the common method in the municipal domain. In preliminary RAs, potential adverse events are identified, then the identified events are analysed separately regarding causes, likelihood and consequences (Aven 2006).

In addition to the requirements in the Civil Protection Act focusing on the risks from a holistic perspective, the municipalities face regulation at the sector-level.

2 THEORETICAL APPROACH

Several elements are of importance in the initiating phase of RAs. Based on our point of interest, we focus on some theoretical considerations related to the method of risk/hazard analysis, supplemented with some perspectives when suited.

Due to the framework for the paper, elements of importance are excluded, though. For instance, risk perception, i.e. peoples' judgement of hazards (Renn 2008), is not explicitly addressed. Neither is safety culture addressed, even though culture can contribute to focus the attention to some specific hazards, while other hazards are not taken notice of (Pidgeon & O'Leary, 2000, Pidgeon 1998; Douglas & Wildavsky 1983).

2.1 *Method and regulation*

A preliminary risk analysis is suited for both major and minor hazards. However, risk analysts might be restricted by a decision that the process of identifying hazards should be limited to regulatory requirements (Baybutt 2014). Such restrictions could, in extreme cases, result in a CRA of rhetorical value, symbolizing control (Clarke 1999), risking that hazards of importance or interest are omitted from the analysis.

2.2 *Imagination*

Cole (2012, p. 12) uses the phrase "broaden the mind-set of responders" as an argument for surprise scenarios in exercises. It is also requisite to broaden the mind-set of risk analysts when identifying potentially adverse events.

Imagination and creativity contribute to the identification of scenarios that would otherwise not necessarily have been identified. Hence, imagination and creativity are required, but analysts might lack these characteristics (Camerona et al. 2017). Besides, even if risk analysts are imaginative, it is not a guarantee for identifying all hazards (Baybutt 2014).

A boundary for imagination might be the ontological status of hazards and risks. They are not fixed. Risks can be viewed in different ways; as objective properties or as socially constructed (Aven & Renn 2010). Risks pre-exist in the former view, and risks can in principle be identified and measured (Lupton 2013, p. 13). Socially constructed risks, on the other hand, are the product of rhetorical processes (Lupton 2013, p. 46). Potentially this induces discussions or interpretations among risk analysts about which hazards to consider in the initiating phase of CRAs.

2.3 *Cognitive biases*

Thinking can be divided into two systems; fast and slow (Kahneman 2011). The fast mode is instinc-

tive and the slow is deliberate. The risk analysis method presupposes deliberate thinking. However, risk analysts are humans. Therefor they are not necessarily as rational as could be expected (Aakvaag 2008).

Cognitive biases are results of heuristics (Kahneman & Tversky 1982). The biases stem from the unconscious influence on human judgements and decisions (Baybutt 2016). They are deviations from the rationality of thinking (Meissner & Wulf 2013, p. 802). Researchers have found many cognitive biases, e.g. the availability bias, group thinking or the framing bias, to mention a few. We will not go into details in this paper. The point here is that cognitive biases among risk analysts can result in missed hazard scenarios (Baybutt 2016). Therefore the negative effects of cognitive biases need to be addressed. This is very difficult due to the unconscious processes involved (Baybutt 2016). However, knowledge, information and awareness can reduce biases. Another strategy is to use a devil's advocate in the risk analyst team. An appointed devil's advocate can initiate debates that might challenge the mind-set of others (Baybutt 2016). Additionally, scenario planning can alter biases (Meissner & Wulf 2011).

2.4 *Filtering risks*

There must be a limit to the number of adverse events to analyse in the CRA. It is simply a matter of resources. This implies that the number of identified adverse events in the brainstorming process must be reduced. Rausand & Utne (2009) argue that hazards where the risks are small, due to low likelihood and/or insignificant consequences, could be filtered here.

Power and interest are also important. Interests can be invested in which adverse events should be emphasised and de-emphasised (Aven 2011; Dekker & Nyce 2014). This also applies to the brainstorming phase. Being able to handle interests requires the capacity to exercise power. There are several sources of power, e.g. information, expertise, control over agenda and resources (Antonsen 2009).

2.5 *Standardization and uniformity*

Recipes and checklists can be beneficial. They provide advice and save time for risk analysts (Hale & Swuste 1998). Checklists can also mitigate a lack of imagination among risk analysts (Baybutt 2014).

The purpose of recipes of how to do things is to do the same. Ergo, uniformity is reasonable (Brunsson 2000). However, standardization can cause blindness to possible adverse events unsuited to the recipes (Hale & Swuste 1998). For instance, Baybutt (2014) holds that elements unlisted in checklists might be left out.

3 METHODS

Data were gathered in twelve of nineteen municipalities in a county in Arctic Norway. The main criterion for including municipalities in the study was location. They are all located in the same geographical region, and therefore to a certain extent exposed to the same hazards. Another criterion was time. The Civil Protection Act came into force in 2011. Requirements in the law set a new framework for CRAs. The CRAs and CRA-processes included in the study are from the timespan 2011–2017, ensuring that the municipalities had been subject to the same legal requirements. Their geographical location also meant that they had been subject to the same supervision by the same County Governor. A third criterion was the availability of the informants during the data collection period.

The data was collected via interviews and analyses of the CRAs, a qualitative approach. Twelve semi-structured interviews were conducted; one informant per municipality. A question guide with open ended questions was used. The informants all played pivotal roles in the CRA process in their respective municipalities. All of them had participated actively in the process of making the CRAs which this paper focuses on. Hence, they had first-hand knowledge of the process and the choices that were made.

The contents of interviews and CRAs were analysed and compared, so data coherence could be checked.

In a Norwegian context, the municipalities spanned from small to medium population size.

Next, we will present findings from the processes of brainstorming and filtering.

4 THE BRAINSTORMING PROCESS

The identification of potential hazards in the municipalities is called the brainstorming process in this paper. The term here refers to a process of creativity, imagination, structure and mapping. Next, we will present the "who's" and the "how's" in this process.

The municipalities had their own unique brainstorming processes. However, there were similarities. Aggregated, the processes either involved

- municipal representatives (M)
- a combination of M and external representatives (E)
- consultants who involved either M or M+E

In eight of the municipalities, both internal and external representatives participated.

It is hard to conclude unambiguously in what way the legal requirements and other regulative

attempts to influence affected the process of identifying hazards. For some it seems as if regulative involvement broadened the scope of hazards to consider. The internal focus in one of the municipalities was not motivated by legal compliance. The intention was to heighten organizational competence in this area of municipal responsibility. A contrast is the municipality with the lowest involvement in this study. Here the primary objective was a "good enough" CRA.

A devil's advocate formally appointed to challenge assumptions or stimulate ideas was not used. Ascribing such a formal role to a participant in the processes seems to be unfamiliar to the risk analysts. However, in two of the municipalities, the risk analysts responsible for the local processes deliberately sought counter-arguments. They were self-appointed informal devil's advocates.

A common trait for the municipalities is that they based identification of hazards on a combination of information sources. In some cases, their imagination was not sufficient, and other sources provided valuable inspiration and ideas. Nobody referred to checklists etc. as means to save time or resources.

In addition to their own previous municipal CRAs, most analysts used regional or national sources to assist them when identifying hazards: typically, national and regional RAs and a government guideline for CRAs. The analyses and the guideline served as checklists. At the local level, other municipal CRAs were sources of information too, but to a lesser extent than for instance the regional RA.

Another influence was the urge from national and regional government to take specific adverse events into consideration, for instance deliberate adverse events in schools etc. (threats, use of weapons), and quite recently, arrival of refugees in large numbers.

Media-coverage was also a source of information and inspiration to some risk analysts.

The informants also referred to a recently established regional arena for risk analysts. Here the analysts could exchange ideas. For instance, the hazard related to cruise tourism had been addressed by one of the municipalities. This was also relevant for some of the other municipalities. The influence from this arena will probably be apparent in future CRAs.

Arguably, these sources facilitate uniformity if not reflected on.

Interestingly, in two of the municipalities, representatives from residents were invited to participate in the process, potentially providing a local focus. Representatives from the municipality and external actors who had been invited to contribute could, of course, also add a new perspective.

The analysts were asked about the usefulness of external information sources and potential negative effects. The majority found such sources very helpful, providing ideas and serving as some sort of quality control as to the content of CRAs. The potential negative effects seem to be eliminated by the usefulness of such sources.

5 THE FILTERING PROCESS

After having identified potential hazards, the municipalities chose which hazards to include in the CRA. In this paper, this process is called filtering. Some of them had identified many potential adverse events, others had few. Most of the municipalities structured the identified events by merging related events, thus reducing the number of events. In addition, the approach of filtering out small risks was applied in several municipalities.

One of the municipalities in fact included all of the identified hazards in the CRA as they were, without filtering.

All in all, the filtering processes passed without much controversies, according to the risk analysts. Issues for debate were the severity of hazards, not their ontological status. The participants in the process came to an agreement. External actors without representation in the CRA work, such as representatives from local industry, showed no interest in trying to influence this, or other, processes. This lack of interest is interesting per se, but beyond the scope of this paper.

6 IMPACT

Looking at the type of risks included in the CRVs, there is a high degree of uniformity regarding events that are mandatory to address; e.g. critical infrastructure.

However, the analysts have not analysed all types of critical infrastructure. They have chosen the ones relevant to them. Electricity and electronic communication are the focus of attention, followed by water supply. Transportation is also addressed in some of the CRAs. Here, local circumstances are obviously of importance. E.g. municipalities with only one main road are more vulnerable than those with several.

Pandemics, nuclear accidents and extreme weather, the transboundary risks, are also included to a high degree in the CRAs. Fires, accidents, emissions or spills of dangerous substances are also addressed to a high degree. These are events of relevance to all municipalities, regardless of location. However, the detailing and the objects at risk vary from one CRA to the other. Municipali-

ties with a coastline have analysed accidents at sea, for example.

Several of the above-mentioned hazards are regulated in sector legislation: e.g. water-supply, fires and nuclear accidents.

There is also high uniformity at an aggregated level regarding deliberate adverse events. They are included in the CRAs. The types of adverse events differ, though. Events like threats and "minor" violence are addressed by almost everybody. Terror, a disastrous event, is covered in fewer CRAs. Here, analysts have varied between the events, most likely based on their assessment of relevance to their municipality. Only two CRAs include Cyber-attack. This type of hazard has not been on the public agenda for very long, and the two CRAs have recently been revised.

In addition, the CRAs encompass a few adverse events of a strictly local character: flooding and the breaking of dikes.

Finally, a few of the CRAs contain unique adverse events; such as substance abuse among municipal employees, animal diseases, violence and sexual abuse against children and breaches in information security.

Identifying adverse events is a challenging task, taking uncertainty about what the future holds into consideration. In two of the municipalities this was addressed by including "the unknown event".

7 DISCUSSION AND CONCLUSION

7.1 *Discussion*

The preliminary risk analysis method and the legal requirements per se are neutral regarding the number of people involved. The variety among the studied municipalities is vast with respect to involvement of personnel. At one end of the continuum only the consultant and a single municipal employee took part in the process. Here the rhetorical value of the CRA was the primary objective, i.e. having a CRA in compliance with regulations (Clarke 1999). At the other end of the continuum a bottom-up approach was applied with all municipal departments being mobilized. The variety of involvement does not seem to have a bearing on the number of adverse events included in the CRAs.

If other parameters were considered, like how well founded the CRAs are in the municipality, or reduction of risk, then the verdict regarding choice of processes might shift. Important hazards might be missing in the CRA (Clarke 1999).

There is a predominance of uniformity in the events that were included in the CRAs. Obviously, this can be ascribed to legal requirements and that the municipalities face the same hazards to a relatively large degree. In addition, the formal framework for CRAs confers some uniformity. The brainstorming processes also induced uniformity, because many risk analysts in the study used the same sources to retrieve ideas of potential hazards. With recipes and a framework like this, some uniformity is reasonable (Brunsson 2000).

Still, we would argue that this is alarming with regard to hazards not listed in these sources, like emergent, novel or local hazards. They might be left out, as implied by Baybutt (2014) and Hale & Swuste (1998). For instance, fish diseases were not included in one of the often-used sources, the County-RA. This hazard might be of relevance in several municipalities because fishery is a prominent part of the industrial base in these communities. Only one CRA included this hazard. Using sources to retrieve ideas, the risk analyst might miss out hazards if the process resembles copying and has a lack of imagination (Cameron et al. 2017; Baybutt 2014).

Most of the processes took place without major disagreement. A devil's advocate was not used in most of the municipalities. Hence, the processes lacked a participant who systematically could have challenged the mind-set of others (Baybutt 2016). Given the uncertainty about future adverse events and local susceptibility, the processes could have benefited from critical voices challenging both the premises for analysing risks, i.e. the formal framework for CRAs, and the local processes.

This could perhaps have provided more diversity in novel or unique adverse events. However, diversity is not an objective per se. The CRAs were in fact diverse in having variations within types of hazards. For instance, some included car accidents, others included bus accidents.

Having argued in this paper that uniformity can be alarming, a final reflection should be added regarding crisis management. Even if all hazards have not been identified and analysed, in crisis management many of the same features occur regardless of the hazard involved (Nilsen 2017). Therefore, uniformity need not be too serious in that respect. The disadvantage of uniformity is the decreased possibility of reducing or eliminating unknown hazards in advance, making crisis management redundant.

7.2 *Conclusion*

The initiating phase of any risk analysis, where adverse events are identified and filtered, is very important. This paper has addressed how risk analysts in 12 municipalities have carried out this phase and the impact of uniformity of adverse events analysed in municipal CRAs. The main conclusion is that there is a predominance of uniformity in CRA-events, as could be expected due to the circumstances and the brainstorming-processes.

Diversity is not an objective per se. However, the chance of identifying the unknown adverse event is lessened by copycats. That is alarming, and awareness needs to be raised.

REFERENCES

Aakvaag, G. 2008. *Moderne sosiologisk teori [Modern sociological theory.* Own translation*]*. Oslo: Abstrakt forlag AS

Antonsen, S. 2009. Safety culture and the issue of power. *Safety Science,* 47: 183–191

Aven, T. & Renn, O. 2010. *Risk Management and Governance. Concepts, Guidelines and Applications.* Berlin: Springer

Aven, T. 2006. *Pålitelighets- og risikoanalyse* [*Reliability—and risk analysis*. Own translation]. 4nd ed. Oslo: Universitetsforlaget

Aven, T. 2011. Selective critique of risk assessments with recommendations for improving methods and practice. *Reliability Engineering and System Safety,* 96:509–514

Aven, T. 2015. Risikostyring [Risk management. Own translation]. 2nd ed. Oslo: Universitetsforlaget.

Baybutt, P. 2014. A critique of the Hazard and Operability (HAZOP) Study. *Journal of Loss Prevention in the Process* Industries, 33:52–58

Baybutt, P. 2016. Cognitive biases in process hazard analysis. *Journal of Loss Prevention in the Process Industries,* 43:372–377

Brunsson, N. 2000. Standardization and Uniformity. In Brunsson N. and Jacobsson, B. eds. *A world of standards*. New York: Oxford University Press, p. 138–150.

Cameron, I., Mannan, S., Németh E., Park, S, Pasman H., Rogers W. & Seligmann B. 2017. Process hazard analysis, hazard identification and scenario definition: Are the conventional tools sufficient, or should and can we do much better? *Process Safety and Environmental Protection,*10:53–70

Clarke, L.B., 1999. *Mission improbable: using fantasy documents to tame disaster.* Chicago: University of Chicago Press

Cole, L.A. 2012. Preparedness, Uncertainty, and Terror Medicine. In L.A. Cole & N.D. Connell (eds.), *Local Planning for Terror and Disaster. From Bioterrorism to Earthquakes*, p. 3–15. New Jersey: Wiley

Dekker, S.W.A & Nyce, J.M. 2014. There is safety in power, or power in safety. *Safety Science* 67: 44–49

Directorate for Civil Protection (DSB). 2017. Veileder til helhetlig risiko—og sårbarhetsanalyse i kommunen [*Guideline for comprehensive risk and vulnerability analysis in the municipality.* Own translation]. Tønsberg: Direktoratet for samfunnssikkerhet og beredskap

Douglas, M. & Wildavsky, A. 1983. *Risk and Culture. An Essay on the Selection of Technological and Environmental Dangers.* Berkley: University of California

Hale, A.R & Swuste, P. 1998. Safety rules: procedural freedom or action constraint? *Safety Science,* 29(3): 163–177

Kahneman, D. 2012. *Tenke, fort og langsomt.* [*Thinking, fast and slow*]. Oslo: Pax forlag

Lupton, D. 213. Risk. 2nd ed. London and New York: Routledge

Meissner, P. & Wulf, T. 2013. Cognitive benefits of scenario planning: Its impact on biases and decision quality. *Technological Forecasting and Social Change, 80(4): 801–814*

Nilsen, A.S. 2017. What challenges can municipalities experience in crisis management? In Cepin M. and Bris, R.(eds.). *Safety and Reliability. Theory and Applications*. London: Taylor &Francis Group, p. 1747–1754

Perry, R.W. & Lindell, M.K. 2003. Preparedness for Emergency Response: Guidelines for the Emergency Planning Process. *Disasters.* 27(4): 336–350

Pidgeon, N. & O'Leary, M. 2000. Man-made disasters: why technology and organizations (sometimes) fail. *Safety Science,* 34:15–30.

Pidgeon, N. 1998. Safety culture: Key theoretical issues. *Work & Stress,* 12(3): 202–216

Rausand, M. & Utne, U.B. 2009. Risikoanalyse—teori og metode [*Risk analysis—theory and methods*. Own translation]. Trondheim: Tapir akademiske forlag

Renn, O. 2008. *Risk Governance. Coping with uncertainty in a Complex world.* London/New York: Earthscan

Tversky, A. & Kahneman, D. 1974. Judgment under uncertainty: heuristics and biases. *Science*, 185, p. 1124–1131

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Integrated monitoring of risks for Seveso plants

G. Baldissone, L. Comberti & M. Demichela
DISAT, Politecnico di Torino, Torino, Italy

T. Marcon & E. Plot
INERIS Parc Technologique Alata, Verneuil-en-Halatte, France

M.C. Leva
Dublin Institute of Technology, Dublin, Ireland

ABSTRACT: Design documentation, safety and security analysis, environmental studies, studies on organizational factors, product characterization, etc., constitute the knowledge base each process plant, with a higher or lower detail, uses for plant management.

Most of this knowledge is often lost inside an accumulation of formal documents that are not made available for practical use, while it should be disclosed and exploited within a living model of the plant (updated in real time), to which the various actors should refer to make their decisions throughout the lifecycle of the installations.

How to give a shared representation of the factory (state, history, behavior), in order to improve the reliability and flow of decision-making, investment, prevention, protection, crisis management? A Risk monitoring systems and knowledge management to be integrated in the architectures of the company IoT has been proposed, developed and tested in French national institute for industrial environment and risks (INERIS).

The initial risk modelling embedded in the knowledge management systems, based on the bow-tie methodology to identify the barriers for critical sequences to the Major accidents and to assess their availability, to be used for decision making, has been here integrated with the Integrated Dynamic Decision Analysis in order to obtain the critical sequences of events, that include the operator contribution (in terms of errors and recovery), the barrier effectiveness and the plant behavior.

The representation of the plant in the shape of sequences allow a more user-friendly management of the information and thus a simplified control of the coherence of the risk assessment modelling with the real plant behavior, and an enhanced decision-making support in the definition of plant control measures, both technical and operational. It also allows an easier integration of the data coming from the field, with traditional or new technologies, as virtual and augmented reality.

The proposed solution is exemplified through the application to an ammonia storage plant.

1 INTRODUCTION

In this paper a risk monitoring systems and knowledge management tool proposed, developed and tested in INERIS is described, named MIRA (Monitoring Intégré des Risques Actualisés, Integrated Risk Monitoring Updated). The MIRA system was developed in the Virtualis projec, update in the TOSCA projec and test in different company. It was developed with the purpose of making the risk assessment in major hazards process plants a living analysis, making available for the different actors—operators, managers, control authorities—the information from the risk assessment and management to support the risk-based decision making along the plant life cycle.

The tool is developed to be used by two main users—the plant managers and the authorities in charge of the inspections and authorization for plant activities—in order to make available all the information about risk management and to guide the decision making with reference to major accidents prevention.

As discussed in Demichela et al. (2004) and De-michela & Piccinini (2006), the strict correlation between the risk assessment and the safety management systems allows summarizing their relations and the links with the inspections activities as in Figure 1. According to this view, the tool developed in French national institute for industrial environment and risks (INERIS) manages the following set of data:

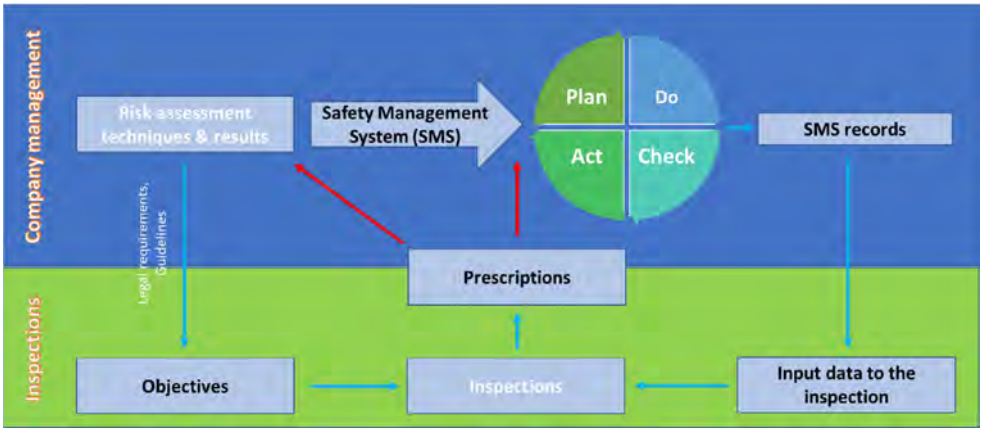


Figure 1. Processes included in the tool.

- The risk analysis data, both the probabilistic and the consequence analysis—in order to allow checking the hypotheses, the methodologies and the results;
- The prescriptions descending from the safety management system inspections, both internal and of the local authorities, that can affect the risk analysis and the related decision making;
- The operational control of the plant and its risks, designed also on the risk results that allows maintaining a tolerable the level of risk, through the safety management system;
- The records of the safety management system procedures and instructions;
- The results of the inspections and the consequent plans for improvements.

2 THE KNOWLEDGE MANAGEMENT SYSTEM

The heart of the knowledge management system is the management of the "living risk assessment", based on the updating of the risk analysis according to the modification of the processes—production processes, procedures and tasks—impacting on the possible accident sequences—probabilities of occurrence, accident scenarios, impact assessment.

The initial level of risk estimated in the "safety case" or "safety report" required by the Seveso Directive—Seveso III (Directive 2012/18/EU) –, when tolerable, should be maintained in time, maintaining the conditions that brought to the level of the risk assessed: process hazards, potential accidents sequences and barriers.

The safety management performance control allows a retrospective updating of the initial level of risk. According to Figure 2, the updating of the risk assessment will take into account the following aspects:

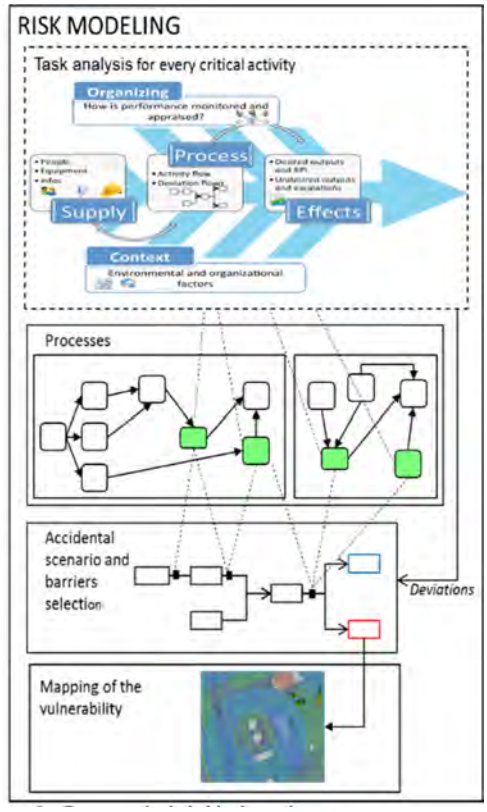


Figure 2. Risk modeling.

- The critical tasks. As an example, through task analysis, missing operational controls on criti-

cal equipment or protection systems can be traced, affecting reliability and availability of the systems.
- The analysis of technical failures and occurred accidents or near-misses. The analysis of the failures and of their treatment allows identifying the unavailability of a system or a missed critical task. Summing up the unavailability days over a given observation period allows estimating the actualized availability, e.g., of a safety barrier.

The calculation of the "level of confidence" of the system under study is carried on according to INERIS methods.

Comparing the initial risk level and the actualized one allows the inspectors to verify if the hypothesis made during risk assessment by the company management are realistic, given the real plant behavior.

In case of a difference, the plant management can revise the safety case, improve the management of the critical tasks and of the critical activities.

During inspections, the plant management has to make available the input data to the inspections, both from the risk assessment and from the operational management. The inspectors have to control the coherence of the different elements.

3 THE CASE STUDY

The case study refers to a company whose activities can be assimilated to those carried on at the AZF company, where the 21st September 2001 an explosion of ammonium nitrate killed 31 persons, together with thousands of injured (Paltrinieri et al., 2012).

In this case, the company produces and stores different types of solid fertilizers, using for their production raw materials, with different hazardous properties controlled by Seveso Directive:

- Flammable gases (natural gas and ammonia);
- Toxic gases (ammonia);
- Oxidising liquids (ammonium nitrate);
- Flammable liquids (oil);
- Toxic liquids (oil);
- Gas under pression (acétylène);
- Substances dangerous for the environment (oil, ammonia);
- Oxidising solid (fertilizers).

Within the risk assessment exercise, the company developed all the bow-ties needed to represent the possible accidental events and the barriers available for their protection (Aneziris et al., 2017). The barriers, whose availability has to be kept under control during time, being critical systems, could be technical or operational (Ahmad et al., 2015).

In Figure 3 one of the bow-ties developed for the ammonia piping system is shown as a demonstration. It refers to the release of a toxic cloud of ammonia (NH_3) and takes into account all the possible initiating events, internal and external, identifying the barriers, operational and technical, that should interrupt the possible event sequences towards an accident.

The barriers numbered in Figure 3 are:

1	Protection against lightning
2	Security guard
3	Design of the rack against earthquakes
4	Rack lay-out design
5	Site firefighting system
6	Manual decompression of the line
7a	Line section isolation for low pressure
7b	Building isolation for low pressure

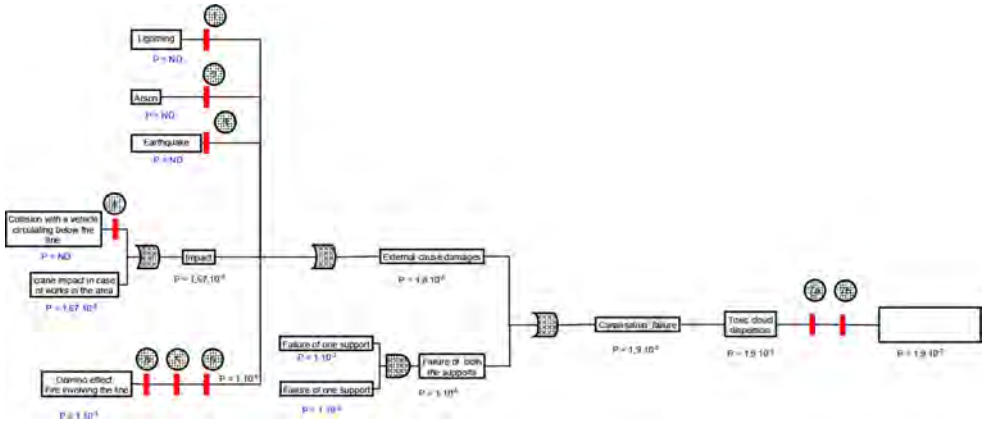


Figure 3. Sample bow-tie for the ammonia release case.

The value of barrier availability is justified in the safety case—not detailed in this paper—and can be updated according to the maintenance tasks records and the inspections' results thanks to the knowledge management system adopted.

The systematic adoption of this approach allows both the plant managers and the inspectors to easily identify the critical items and the protection systems that has to be taken under control in order to maintain in time the availabilities speculated in the risk assessment.

4 ONE STEP BEYOND

The present version of the tool, despite allowing a retrospective risk update, does not allow the risk to be updated in real time. Some experiences towards a similar purpose can be found in Paltrinieri et al. (2012), where the early warnings analysis is discussed and in Villa et al. (2016) where the dynamic risk assessment is introduced. In Raoni et al. (2018) the integration of probabilistic analysis and process simulation is also discussed and exemplified.

In general terms, to reach the goal the knowledge management tool should be interfaced with the control system of the process unit in order to collect the field data and update run time the risk figures.

In this way it could constitute an overarching real-time optimization and scheduling system, controlling and monitoring the operations of the whole plant together with the optimization of analytical tasks and the interpretation and the enhancement of the data used for risk assessment and control.

At the moment the representation of the potential accident through bow-ties does not allow this kind of integration, thus, in this seminal work, an attempt has been made of shifting toward the representation in "stories" – sequences of events—allowed by the Integrated Dynamic Decision Analysis (IDDA).

Originally developed by Galvagni, the method was tested on process case studies several times: by Demichela & Piccinini (2008) on a simple case study of a tank overflow; by Turja & Demichela (2011) and Demichela & Camuncoli (2014) for the risk-based design on an allyl-chloride production plant, where it allowed to carry on the risk analysis in a dynamic way, taking into account process time dependent occurrences; by Baldissone et al. (2016) to the analysis of competing technologies for the VOC treatment; by Gerbec et al. (2016) that compared the result of the IDDA analysis on a LPG tank cold water pressure test procedure to the results obtained with the Bayesian network.

The IDDA methodology, that can be seen as an enhanced event tree, is built on the interaction between a logical—probabilistic model and a phenomenological model of the plant under analysis.

Its use as a decision-making supporting tool in risk assessment is summarized in Figure 4, as also discussed in Baldissone et al. (2017).

The logical—probabilistic model is a logical description of the plant behavior (integrating human and technological aspects, where the case) based on the general logic theory.

In particular, when the operational aspect is relevant, a good starting point to build the logical—probabilistic model is the Task Analysis, or a

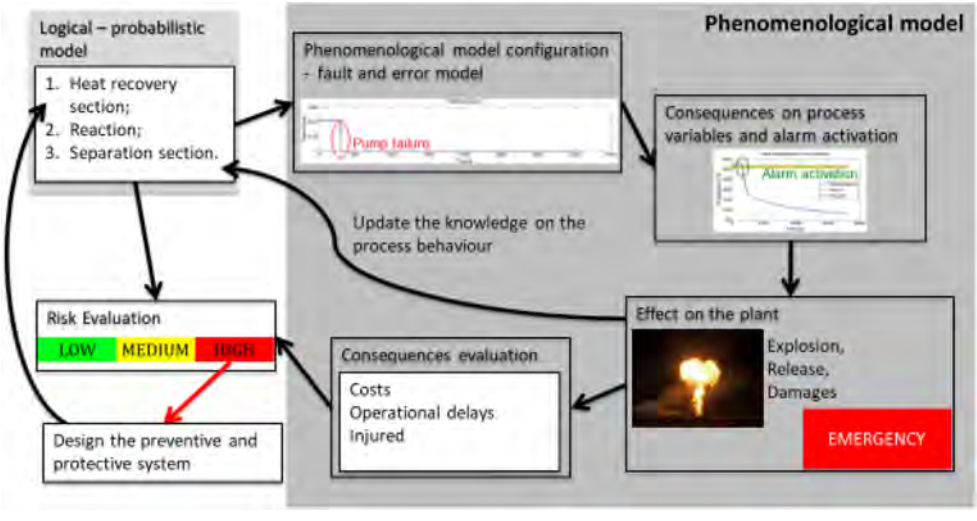


Figure 4. Conceptual scheme of the IDDA use as risk based decision-making support tool.

Table 1. Sequence n. 133 from the IDDA logical-probabilistic model elaboration.

Level	Answer	Prob.	Cumulative prob.	Description
1	1	1.00e-03	1.0000e-03	Operational error? Yes
2	1	9.00e-01	9.0000e-04	Enough space available? No
3	1	1.00e-01	9.0000e-05	Pump trip failure? Yes
10	1	1.00e-01	9.0000e-06	Level transmitter failure? Yes
20	0	1–1.00e-01	8.1000e-06	Extraction works? Yes—diluted cloud
101	0	1–1.67e-05	8.0999e-06	Crane falling down? No
102	1	1.00e-03	8.0999e-09	Fire below the pipe (Domino effect)? Yes
103	1	1.00e-02	8.0999e-11	Safety valve works? No
105	0	1–1.00e-01	7.2899e-11	Manual decompression? Yes
120	1	1.00e-03	7.2899e-14	Pipe support 1 failure? Yes
121	1	1.00e-03	7.2899e-17	Pipe support 2 failure? Yes
125	#1		7.29e-17	Both Supports fail? Yes
150	#1		7.29e-17	Pipe leaks? Yes—NH3 release
200	#2		7.29e-17	Result of the case 1? Overload—release managed
201	#1		7.29e-17	Result of the case 2? Toxic release from the pipe

Sequence probability: 7.28988e-17.

functional analysis for the technological part, as in Balfe et al. (2017).

The logical—probabilistic model elaboration brings to the development of all the possible sequences the system could undergo, with an accuracy dependent on the level of knowledge disclosed in the model itself.

Each sequence of events is coupled with its probability of occurrence, dependent on the occurrence probability of each event contained in the sequence, and on the logical and probabilistic constraints.

The phenomenological model is the mathematical description of the physical behavior of the system, assessing the status and trend of each relevant process variable with reference to the failure sequence identified and allows for a more realistic scenario both in probabilistic and in phenomenological terms; it allows the description of not only the unwanted events, but also of the resilience capacity built in the system (recover possibilities).

The full set of alternatives allows for the complete spectrum of possible probability-consequence conditions to be used as a basis for decisions in risk reduction and control with a compound knowledge encompassing system description.

The IDDA logical probabilistic model has been developed for the bow-ties related to the ammonia involving accident scenarios and returned a set of sequences the system could undergo.

The events represented in Figure 3, described through IDDA, brought to the generation of 220 alternative sequences, 90 of which bringing to a toxic release. The probability of occurrence obtained for the Top Event is comparable to the one obtained through bow-tie.

In Table 1 a sequence ("story") is represented according to the IDDA elaboration, as example of the results obtainable.

The representation in stories, coupled with the phenomenological behavior of the system allows comparing the data collected from the field—from different data sources: e.g. from the plant DCS/PLC or smart distributed sensors; from the registration of the maintenance activities, as through a Risk Register, as the one proposed in Leva et al. (2017) - with the sequences possibly occurring in the plant, thus allowing to follow up run-time the risk in the plant.

The application of the enhanced method has also identified some dependencies among events that can hardly be caught through the bow-tie approach adopted, thus bringing possibly to not conservative evaluations.

5 CONCLUSIONS

The risk assessment in process plants has often been regarded as a static document to be prepared for authorizations more than a living document for the day-to-day decision making in process plants.

The seminal work here described would like to overcome this view and include in a knowledge management system—already developed at INERIS—a way to trace the behavior of the plant towards an accident in order to guide the maintenance, inspection and even emergency preparedness in a living system.

The opportunity has been obtained integrating in the model the representation in sequences of the Integrated Dynamic Decision Analysis, that

could be used to compare the plant management and behavior and its model contained in the risk assessment.

It was demonstrated that including this approach allows some interdependencies among event to be identified—identification otherwise impossible—and that this constitutes a good support for the further developments above described.

AKNOWLEDGMENT

This knowledge management tool is one of the results of the European Project TOSCA—www.toscaproject.eu—that allowed the initial development of the Computerized Safety Barrier Management system and the company wide Risk Register.

REFERENCES

Ahmad, M., Pontiggia, M., Demichela, M., Leva, M.C. 2015. Human and organizational factors assessment and their use as potential safety barriers. *30th Center for Chemical Process Safety International Conference 2015 - Topical Conference at the 2015 AIChE Spring Meeting and 11th Global Congress on Process Safety*:729–746.

Aneziris, O.N., Nivolianitou, Z., Konstandinidou, M., Mavridis, G., Plot, E. 2017. A Total Safety Management framework in case of a major hazards plant producing pesticides. *Safety Science*, 100:183–194.

Baldissone, G., Demichela, M., Camuncoli, G., Comberti, L. 2017. Formaldehyde production plant modification: Risk based decision making. *Chemical Engineering Transactions*, 57:703–708.

Baldissone, G., Fissore, D., Demichela, M. 2016. Catalytic after-treatment of lean VOC-air streams: Process intensification vs. plant reliability. *Process Safety and Environmental Protection*, 100:208–219.

Balfe, N., Leva, M.C., Ciarapica-Alunni, C., O'Mahoney, S. 2017. Total project planning: Integration of task analysis, safety analysis and optimisation techniques. *Safety Science*, 100:216–224.

Demichela, M., Piccinini, N., Romano, A. 2004. Risk analysis as a basis for safety management system. *Journal of Loss Prevention in the Process Industries*, 17 (3):179–185.

Demichela, M., Piccinini, N. 2006. How the management aspects can affect the results of the QRA. *Journal of Loss Prevention in the Process Industries*, 19 (1):70–77.

Demichela M.; Camuncoli G. (2014) Risk Based Decision Making. Discussion On Two Methodological Milestones. *Journal of Loss Prevention in the Process Industries*, 28 (1):101–108.

Gerbec, M., Baldissone, G., Demichela, M. 2017. Design of procedures for rare, new or complex processes: Part 2 – Comparative risk assessment and CEA of the case study. *Safety Science*, 100:203–215.

Leva, M.C., Balfe, N., McAleer, B., Rocke, M. 2017. Risk registers: Structuring data collection to develop risk intelligence. *Safety Science*, 100:143–156.

Paltrinieri, N., Dechy, N., Salzano, E., Wardman, M., Cozzani, V. 2012. Lessons Learned from Toulouse and Buncefield Disasters: From Risk Analysis Failures to the Identification of Atypical Scenarios Through a Better Knowledge Management. *Risk Analysis*, 32 (8):1404–1419.

Paltrinieri, N., Oien, K., Cozzani, V. 2012. Assessment and comparison of two early warning indicator methods in the perspective of prevention of atypical accident scenarios. *Reliability Engineering and System Safety*, 108:21–31.

Piccinini, N., Demichela, M. 2008. Risk based decision-making in plant design. *Canadian Journal of Chemical Engineering*, 86 (3):316–322.

Raoni, R., Secchi, A.R., Demichela, M. 2018. Employing process simulation for hazardous process deviation identification and analysis. *Safety Science*, 101:209–219.

Turja, A., Demichela, M. 2011. Risk based design of allyl chloride production plant. *Chemical Engineering Transactions*, 24:1087–1092.

Villa, V., Paltrinieri, N., Khan, F., Cozzani, V. 2016. Towards dynamic risk analysis: A review of the risk assessment approach and its limitations in the chemical process industry. *Safety Science*, 89:77–93.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Risk and social interaction (samhandling) to meet the unforeseen

G.E. Torgersen
The Norwegian Defence College, Oslo, Norway
School of Business, Center for Emergency Preparedness, University College of Southeast Norway, Norway

T.J. Steiro
Institute for Teacher Education, The Norwegian University of Science and Technology, Trondheim, Norway

L.I. Magnussen
School of Business, Center for Emergency Preparedness, University College of Southeast Norway, Norway

ABSTRACT: Social interaction is regarded central in prevention, planning, handling and evaluation of unforeseen (UN) events. Hence, the social interaction factors may vary in different phases and conditions. This is also the case even when conditions are unpredictable compared to situations of low or without risk. We therefore apply an UN-oriented «Bow tie» model that focuses on three main stages related to the development of a serious event;

- Phase 1: Preparation/ identification of hazard signals and barrier development
- Phase 2: Occurrence of an unexpected event/accident
- Phase 3: Measures/ stabilization

The purpose of this paper is to investigate social interaction, the term and phenomenon at risk, and how social interaction factors behave under such conditions. How do social interaction factors relate to the different phases given by the «Bow tie» model? And, of what significance can this have related to education and training as well as the competence of a complex and interdependent organization? We will in this paper argue that social interaction plays a crucial role in meeting the unforeseen and discuss this accordingly.

1 INTRODUCTION

Organizations increasingly rely on their ability to adapt to and manage multifaceted, demanding situations (Herberg et al., 2018, Brozus, 2016, Roux-Dufort, 2007 Weick, 2015), particularly when facing sudden and unexpected risk events (Barnett, 2004, Bechky & Okhuysen, 2011, Cunha et al., 2006, Fornette et al., 2016). Little is known regarding how an organization can methodically identify relevant factors that influence the outcome of an event (Kaarstad & Torgersen, 2017). The current paper is based on the work of Torgersen (2018a) and Torgersen (2015). The starting point is the Norwegian/Nordic concept of *social interaction* (interaction). Traditional perceptions of the concept of *social interaction* are challenged and new models are introduced with the questions: What are the basic structures of the concept of interaction under/during risk, and further, how can interaction be created when the conditions are unpredictable? This scientific anthology "Social interaction" (Torgersen, 2018a) may be of relevance to anyone who works with warning signs, handling and stabilizing unforeseen events within most professions in society and emergency management, as well as students, teachers and researchers in education, strategy, human resources and organizational management subjects. The work was started by Torgersen & Steiro (2009). In 2015, the term unforeseen (UN) was introduced and defined as:

> "*Something that occurs relatively unexpected and relatively low probability or predictability for those who experience and must deal with it*." (Kvernbekk et al., 2015:30).

In this article we will provide an overview of the research and explaining the importance of social interaction linked to the unforeseen. We will argue that social interaction plays a crucial role in handling the unforeseen and can provide useful and applicable knowledge to the field of risk management and may add do the dynamic capabilities (Eisenhardt et al. 2000) of the organization.

2 THE DEFINITION OF SOCIAL INTERACTION

Based on a study of 15 organizations (Torgersen & Steiro, 2009), this definition of social interaction is developed:

> "*Social interaction is an open and mutual communication and development between participants, who develop skills and complement each other in terms of expertise, either directly, face-to-face, or mediated by technology or by hand power. It involves working towards common goals. The relationship between participants at any given time relies on trust, involvement, rationality and industry knowledge.*" (Torgersen & Steiro, 2009:130).

Social interaction is primarily a way to work or "act. Central to interaction is "action", first and foremost a targeted action. This action is shared or exchanged expertise—often extensive, specialized, and used in a complementary manner (Torgersen & Steiro, 2018, Steiro & Torgersen, 2013, Torgersen & Steiro, 2009). The focus on complementariness can also be seen in the work of Miles & Watkins (2007), supporting the notion that interaction is more than the sum of its parts. For *social interaction* to occur, one must also be aware that each participant contributes with their unique situational understanding ("shared situational awareness") or *shared repertoire of practice* (Wenger), based partly on their own perspective and position in the organization, and their experiences, culture, knowledge, attitudes, emotions and job satisfaction, including recommendations to the interaction process (Torgersen & Steiro, 2018, Sandeland & Boudens, 2000).

3 THE THEORETICAL FOUNDATION FOR SOCIAL INTERACTION AND RISK

In the previous paragraph, complementary elements was pointed out as important. An illustrative example of joining force a playing complementary roles is the Duomo in Firenze, Italy. Another illustrative example is the foundation of the dome of the Florence Cathedral, designed by Filippo Brunelleschi between 1417–1434. Brunelleschi had the bricks laid in a herringbone pattern to support the inner dome (Illustration 1). King (2000) explains this as an action and reaction between the bricks. We can argue that they are the same bricks but assigned different roles, and that the action and reaction creates interaction, redistributing the forces of pressure outwards and downwards. This prevents the dome from collapsing inwards (Torgersen & Steiro, 2018). It can be illustrated in the illustration below.

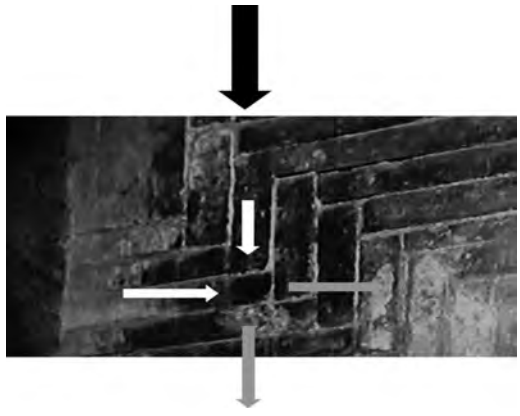


Illustration 1. The herring-bone pattern of brickwork designed by Filippo Brunelleschi (1377–1446) (Photo: Trygve Steiro, 2017).

The herring-bone pattern of brickwork designed by Filippo Brunelleschi (1377–1446), for the inner dome of Florence Cathedral, which effectively divides the pressure downwards and outwards, avoiding an inward collapse (grey arrows). The white arrows symbolize action and reaction, thereby creating an interaction, and an illustration of social interaction as something happen in the action. It also illustrate complementary aspects (Miles & Watkins, 2007).

The bow-tie diagram has been used extensively for risk management and is based on and modified after Primrose, Bentley, van der Graaf & Sykes (1996) model (Torgersen, 2018 b).

The bow—tie diagram modified after Primrose et al. (1996) in order to understand the interplay between the risk of the unforeseen through different phases and the importance of social interaction. The closer to the impact, the more demanding social interaction will be because of the actions required. However, Carlström (2018) stresses that social interaction can occur vertically, horizontally and synchronous. Scholtens (2008) emphasizes that operative staff, in most cases, make the right decisions and act in an optimal way, if they are allowed to act autonomously. Martin et al. (2016) have investigated crisis management and underlined the importance of the fours Cs; Communication, Cooperation, Coordination and Collaboration. Building on this work, we also put social interaction closest to the action.

The illustration is based on the work of Scholtens (2008), Torgersen & Steiro (2009), Martin et al. (2016) and Carlström (2018). Carlström (2018) further argue for in times of uncertainty, flexible and decentralized solutions should be sought. The same was claimed by Torgersen & Steiro (2009)

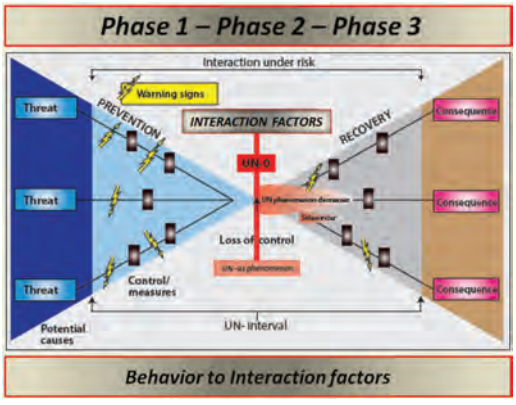


Figure 1. (Torgersen, 2018).

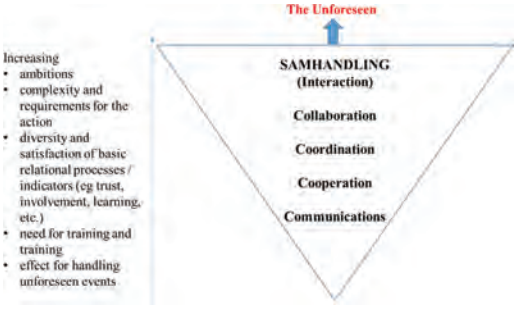


Figure 2. 4Cs plus social interaction in order to meet the unforeseen (Torgersen & Steiro, 2018).

building on the classical work by Burns & Stalker (1961). Carlstöm uses the military concept "auftragstachtik" as an example. For a further and elaborated discussion see Carlstöm (2018), Bergh & Boe (2018) Krabberød & Jacobsen (2018). Brady (2011) has looked into the Battle of Stalingrad and stressed that while the German commander Paulus sticked to the plan and doctrines too rigidly, the opponent Russian General was improvising and allowed improvising by the Russian over commando Stavka. New technology has been said will contribute to more decentralized chain of command (Zuboff, 1988; Albert & Hays, 2003). Albert & Hays introduced the term "strategical corporal". On the use of plans and military doctrines, also see Carlsten, Torgersen, Steiro and Haugdal, 2018. We are not claiming that plans are not important, since they obviously are. When foreseen interruptions occur, plans can be followed more strictly. However, when an unforeseen event appears they need to be uses as a framework and theory, but local action is apparently more important. An illustrative example on plans and social interaction are the Hudson River landing in 2009 with no fatalities (Eisen & Savel, 2009). On the 15th of January 15 in 2009, US Airways Flight 1549 hit a flock of geese shortly after takeoff from LaGuardia Airport (New York, NY), causing both of the engines to lose power. The first officer were in control of the aircraft. Captain Chesley Sullenberger took control of the airplane and the radio-communications, and then instructed the first officer to run an engine restart checklist, which was unsuccessful. Without engine power, Sullenberger decided that he was unable to reach either LaGuardia or the Teterboro airport in New Jersey. The flight crew then decided that an emergency landing in the Hudson River was a viable option. Pariés (2011) points out that a bird strike hitting and destroying both engines was thought of on beforehand. However, nobody had seem to predict or foresee the effect of losing both engines below an altitude of 800 meters. This was the case of Flight 1549. Below is excerpts of the communication of Flight 1549 and the air control of LaGuardia.

Based on National Transportation Safety Board senior official Kathryn O. Higgins's account and based on a transcript of the communication (Federal Aviation Administration, 2009). It is also worth noticing that within the cockpit, Captain Sullenberger orders *"My aircraft"*. First officer Skiles replies: *"Your aircraft"* (Eisen & Savel, 2009). Eisen and Savel (2009) write that this is according to the

Time	Event
1524:54	Tower cleared flight 1549 for takeoff
1525:51	Pilot informs control tower that they were at 700 feet
1527:01	Radar detected that 1549 hit primary targets
1527:36	Pilot, "Ah, this is Cactus 1539 hit birds, we lost thrust in both engines. We're turning back towards LaGuardia"
1527:49	Controller advised LaGuardia to stop departures
1528:05	Controller, "Cactus 1529, if we can get it to you do you want to try to land runway one three?"
1528:11	Pilot, "We're unable. We may end up in the Hudson"
1528:50	Pilot, "I am not sure if we can make any runway. Oh, what's over to our right anything in New Jersey; maybe Teterboro"
1529:02	Controller, "Do you want to try and go to Teterboro"
1529:03	Pilot, "Yes"
1529:25	Pilot, "We can't do it"
1529:28	Pilot, "We're going to be in the Hudson"
1530:30	Touchdown in Hudson River

Table 1. The terms 7T in Norwegian and the English translation.

Trust (Norwegian word *tillit*)
Assurance (Norwegian word *trygghet*)
Well-being (Norwegian word *trivsel*)
Belonging (Norwegian word *tilhørighet*)
Clarity (Norwegian word *tydelighet*)
Time (Norwegian word *tid*)
Tolerance (Norwegian word *toleranse*)

Table 2. Sources of influence and competencies for "samhandling" or "social interaction".

Competencies for samhandling	Sources of influence
Trust	Torgersen & Steiro (2009)
Assurance	Torgersen & Steiro (2009)
Well-being	Torgersen & Steiro (2009)
Belonging	Torgersen & Steiro (2009)
Clarity	Weick (1987); LaPorte & Consolini (1991); Weick & Sutcliffe (2001); Løfdali (2014); Steiro, Johansen, Andersen & Olsvik (2013); Fredriksen & Moen, (2013); Eggen & Nyrønning (1999); Simensen (2005); Leitao (2010)
Time	Weick (1993); Steiro et al. (2013)
Tolerance	Kant (1795/1991); Derida (2005a; 2005b; 2000); Torgersen & Steiro (2009); Steiro et al. (2013); Steiro & Torgersen (2018)

procedures and also accordance to Crew Resource Management Training, that is a well established concept to foster interaction within aviation crews. When Sullenberger gave the order, "*Brace for impact*," the flight attendants chanted repeatedly, "*Brace, heads down, stay down*" (Eisen & Savel, 2009:912). Pariés (2011) further reports that captain Sullenberger was very focused on the tasks. The Air Traffic Controller reported also after the accident: "*During the emergency itself, I was hyper focused, I had no choice but to think and act quickly, and remain calm. I was flexible and responsive. I listened to what the pilots said, and made sure to give him the tools he needed. I stayed calm and in control*" (Pariés, 2011:16).

All this communication can be seen as precise communication, that is so important for effective social interaction and particular when time is of such shortage (Torgersen & Steiro, 2009). The relevance of competence skills brings us on to further findings regarding competence for social interaction. Torgersen & Steiro (2018) have identified the following central competencies as important for social interaction. The first four was identified in Torgersen & Steiro (2009) and in Norwegian it was "tillit, trygghet, trivsel and tilhørighet", which was summed up in the 4Ts. In the study of Torgersen & Steiro (2018), "tydelighet" (English; clarity, both in role and communication) and "toleranse" (English; Tolerance) and "Tid" (English; time) was added. That would for Norwegian readers then be the 7Ts (Table 1). This point is not important in itself other that for educational reasons. To a not Nordic reading audience, this has no relevance.

In the following Table 2, the theoretical and empirical sources of influence to the 7T model is presented.

4 MEASURING SOCIAL INTERACTION IN RELATION TO THE UNFORESEEN

Herberg et al. (2018) found that social interaction combined with general self-efficacy and social support can account for a considerable proportion of the variance in preparedness for the unforeseen in different organizations. The study was performed using a self-completion questionnaire answered by personnel from the Norwegian Armed Forces during a three-month period in the winter of 2016 & 2017. The personnel were from all branches of the military, including commissioned and non-commissioned officers, military academy students and conscripts participated in the study. The study results are based on a survey carried out. The questionnaire was distributed to 16 units, departments and military academies throughout Norway. A total of 624 personnel participated in the study and the response rate was 77%. The sample consisted of 525 males (85%) and 92 females (15%) respondents with a mean age of 25.7 years and with the average military experience was 5.5 years (Herberg et al. 2018). Social interaction was found to be the most important predictor of preparedness for the unforeseen. The results in the study indicate that it is possible to prepare for unforeseen events by implementing measures that improve social factors in particular Herberg et al. (2018). Organizations should develop a work environment where managers and colleagues provides both moral and emotional support. Furthermore this environment where peoples communicating is recognized by their listening to each other and creating trust. The authors further suggest to create an environment of common understanding of the situation, have useful routines (and relevant) with partners and focusing on exchange and complement employee skills and knowledge (Herberg et al., 2018).

These findings lend support to the factors suggested by Torgersen & Steiro (2018) and listed in Table 1. However, Herberg et al. (2018) points to a need for more research to better understanding the relations between these factors.

5 TRAINING AND EDUCATING FOR SOCIAL INTERACTION

We have seen from paragraph 4, that social interaction plays an important role. In another study, by Nyhus, Steiro & Torgersen (2018), mentoring and coaching was studied for a joint operation course held at the Norwegian Defence College. Personnel form all weapon branches participated (Air, Army and Sea). The purpose of the course was to increase understanding of joint operations (Andersen, 2016; FHS, 2015). Since conflicts and wars rarely follow a familiar pattern, unpredictable factors are tried to be put into the education (Heier, 2015). The course is executed by more experienced officers, guiding less experienced officers and soldiers with given military cases and scenarios that are often based on experiences from actual events. The theoretical framework applied was therefore "apprenticeship learning". Apprenticeship is rooted in sociocultural learning theory (Saljø), which emphasizes that knowledge is constructed through social interaction (social interaction) in a context and not primarily through individual processes (Lave & Wenger, 1991, Maguire, 1999, Dysthe, 2001). 10 hours of observation was conducted in the sessions. In addition five group interviews were conducted with a total of 23 informants (out of a number of total 100 students). After a sufficient number of data was obtained (saturation), the interviews were stopped. A thematic analysis was adopted when interpretating the interview material. This is a suitable method for identifying, analysing, and reporting patterns within the data analysis (Braun & Clarke, 2006). The analysis reveals that some supervisors emphasize mainly on the product, and others more on how the group reached the result, i.e. the process. The product said something about the specific deliveries that the group arrived at. Furthermore, the product, among other things, referred to the supervisors who made sure that the group followed certain structured patterns of action and worked towards a goal that was embodied in doctrines and drills. Nyhus et al. (2018) found further that mentors who focused a lot on the product in the preparation phase did not take the unforeseen into account. A process-oriented supervisor with authority enough to create a shared commitment and understanding of situations. By the virtues of experience and competence, the supervisor can pave the way in preparation for meeting the unforeseen.

In the impact phase, we see how some mentors successfully combine equality with such results focus. It is seen as an advantage if the supervisor allows to focus on the result, but at the same time dares to move away from the role of an instructor who is going to make a correct or wrong answer.

In the final consequence phase, the supervisor should represent a counterweight to pure instruction, where logical conclusions about results control the conversation. Nyhus et al. (2018) found that the supervisor should not maintain a tight structure with a given focus on a blueprint product. The supervisors should investigate, challenge and test the group's knowledge and understanding, and she should ask follow-up questions where they, together with the group, assess the answers. The supervisor occupies a coaching role (Nyhus et al. (2018). The authors further suggest that guided student groups within the apprenticeship learning tradition is an alternative to traditional education, one that provides insights and valuable learning experiences.

6 CONCLUSION

We have demonstrated in this paper that social interaction offers something different than collaboration, co-operation. In a world facing the unforeseen, rather than insisting on the prediction of unexpected events, organizations should therefore investigate how to deal with unanticipated and learn from such events and how this can be developed. A change in focus and mind-set should highlight the relevance of social interaction as a basic and generic core competence. We have defined social interaction and explain underlying elements that makes it unique. At the same time trust, assurance, well-being, clarity, time and tolerance are, as we have argued, is central elements in order to establish and develop good social interaction. It does not come for free, but the potential benefits are expected to be large. We have also argued that social interactions plays key roles during risk and meeting with the unforeseen as demonstrated in paragraph 4. We have used the Bow-tie diagram to distinguish between social interaction before, during and after an incident or accident. We need further research that can contribute to the analysis of incidents and accidents within this framework. It would also been interesting to examine the developments of social interaction and to follow up on social interaction in different organizations.

REFERENCES

Alberts, D. S., & Hayes, R. E. 2003. *Power to the Edge: Command and Control in the Information Age*. Office of the Assistant Secretary of Defence, Washington DC Command and Control Research Program (CCRP).

Andersen, M. 2016. «Hva er fellesoperasjoner?» [What are joint operations? In: Andersen, M. & Ødegaard, G. (Eds.) 2016. *Militære fellesoperasjoner – en innføring* [Military Joint Operations—an Introduction]. Oslo: Abstrakt forlag.

Barnett, R. 2004. Learning for an unknown future. *Higher Education Research & Development 23*(3), 247–260.

Bechky, B. A., & Okhuysen, G. A. 2011. Expecting the Unexpected? SWAT Officers and Film Crews Handle Surprises. *Academy of Managment Journal, 54*(2), 239–261.

Bergh, J. & Boe, O. 2018. *Samhandling* [Interaction] and trust in military leadership structures. In: Torgersen, G. E. 2018. *Samhandling (Interaction) under Risk—a Step ahead of the Unforeseen.* Oslo: Cappelen Akademisk. Open Access.

Brady, M. 2011. Improvisation versus Rigid Command and Control at Stalingrad. *Journal of Management History*, 17(1), 27–49.

Braun, V. & Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology, 3* (2), p.77–101.

DOI:10.1191/1478088706qp063oa

Brozus, L. (Ed.). 2016. Unexpected, unforeseen, unplanned: scenarios of international foreign and security policy. *SWP Research Paper 1*. Berlin.

Burns, T. E., & Stalker, G. M. 1961. *The Management of Innovation.* London: Tavistock.

Carlsten, T. C., Torgersen, G. E., Steiro, T. J., & Haugdal, B. K. 2018. The Relevance of Samhandling (Interaction) in Military Doctrines. In: Torgersen, G. E. 2018. *Samhandling (Interaction) under Risk—a Step ahead of the Unforeseen*. Oslo: Cappelen Akademisk. Open Access.

Carlström, E. 2018. *Samhandling* [Interaction] during crisis work—a three-level model. In: Torgersen, G. E. 2018. *Samhandling* (Interaction) under Risk—a Step ahead of the *Unforeseen*. Oslo: Cappelen Akademisk. Open Access.

Cunha, M. P., Clegg, S. R., & Kamoche, K. 2005. Surprises in Management and Organization: Concept, Sources and A Typology. *British Journal of Management*, 17, 317–329.

Derrida, J. 2005 a. *Rogues: two Essays on Reason*. Translated by Pascale-Anne Brault and Michael Naas. Stanford: Stanford University Press.

Derrida, J. 2005 b. On Cosmopolitanism and *Forgiveness*. Translated by Mark Dooley and Richard Kearney. New York: Routledge.

Derrida, J. 2000) *Of Hospitality*. Translated by Rachel Bowlby. Stanford: Stanford University Press.

Dysthe, O. 2001. *Dialog, samspel og læring*. [Dilaogue, Interplay and Learning] Oslo: Abstrakt forlag.

Eggen, N. A., & Nyrønning, S. 1999. *Godfoten: Samhandling—veien til suksess [The Strong Leg: Samhandling—The Road to Success].* Oslo: Aschehoug.

Eisenhardt, K., Martin, J., & Helfat, Constance E. 2000. Dynamic capabilities: What are they? *Strategic Management Journal, 21*(10–11), 1105–1121

Federal Aviation Administration 2009. Preliminary accident and incident data: US Airways 1549 (AWE1549), January 15, 2009.

Available at: http://www.faa.gov/data_statistics/accident_incident/1549/. Accessed November 10, 2017.

Fornette, M-P., Bourgy, M., Jollans, J-Y., Roumes, C., & Darses, F. 2016. Enhancing Management of Complex and Unforeseen Situations Among Pilots: New Trends in Cognitive-Adaption Training. In: M. A. Vidulich, P. S. Tsang, J. M. Flach (Eds.) 2016. *Advances in Aviation Psychology: Routledge Publications, 1*, 229–247.

Fredriksen, P. & Moen. F. 2013. Ledelse [Leadership]. In: Moen, F. 2013. *Prestasjonsutvikling: Coaching og ledelse [Performance Development: Coaching And Leadership].* Trondheim: Akademika Forlag.

Heier, T. 2015. Læring i risikosamfunnet. [Learning in The Risk Society] In: Torgersen, G. E. (Eds.). (2015). *Pedagogikk for det uforutsette*. [Pedagogic for the Unforeseen] Bergen: Fagbokforlaget.

Herberg, M., Torgersen, G. E., & Rundmo, T. 2018. Competence for the Unforeseen—the Importance of Human, Social and Organizational factors.

In: Torgersen, G. E. 2018. *Samhandling (Interaction) under Risk—a Step ahead of the Unforeseen*. Oslo: Cappelen Akademisk. Open Access.

Kant, I. 1795/1991. *"Perpetual Peace": A Philosophical Sketch.* Translated by Ted Humphrey. Indianapolis: Hackett Publishing.

King, R. 2000. *Brunelleschi's Dome: The Story ofthe Great Cathedral in Florence*. New York: Penguin Books.

Krabberød, T. & Jacobsen, J. O. 2018. Military Samhandling [Interaction] in Unforeseen Situations—A Historical Perspective. In: Torgersen, G. E. 2018. *Samhandling (Interaction) under Risk- a Step ahead of the Unforeseen.* Oslo: Cappelen Akademisk. Open Access.

Kvernbekk, T., Torgersen, G. E., & Moe, I. 2015. Om begrepet det uforutsette [On the Term the Unforeseen]. In: Torgersen, G. E. 2015. *Pedagogikk for det uforutsette [Pedagogy for the Unforeseen].* Bergen: Fagbokforlaget.

Kaarstad, M., & Torgersen, G. E. 2017. Is it Possible to Assess an Organization's Preparedness for the Unforeseen? Development and Evaluation of a Methodology. Arts Social Science Journal 8, 254.

LaPorte, T. R., & Consolini, P. M. 1991. Working in Practice but not in Theory: Theoretical Challenges of "High-Reliability Organisations". *Journal of Public Administration Research and Theory, 1*, 19–47.

Lave, J., & Wenger, E. 1991. *Situated Learning:* Legitimate Peripheral Participation. Cambridge, MA: Cambridge University Press.

Leitao, C. E. 2010. I 'Godfotens' fotspor: en kvan-titativ undersøkelse om flytopplevelser og jobbutførelse hos de ansatte i RBK [In the foot step of the strong leg. A quantitative investigation regarding flow experiences and work performance in RBK]. *Unpublished master thesis, Department of Psychology, NTNU.*

Løfdali, B. (2014). *I skyggen av Eggen. Storhetstiden, fallet og veien tilbake for Rosenborg Ballklubb [In the Shadow of Eggen. The Greatness, the Fall and Way Back for Rosenborg Ballklub].* Kagge Forlag.

Martin, E., Nolte, I., & Vitolo, E. 2016. The Four Cs of Disaster Partnering: Communication, Cooperation, Coordination and Collaboration. *Disasters*, *40*(4), 621–643.

Maguire, M. 1999. Modern Apprenticeships: Just in Time, or Far too Late. In: Ainsley, P. & Rainbird, H. (Eds.). 1999. *Apprenticeship: Toward a New Paradigm of Learning*. London/New York: Routledge.

Miles, S. A., & Watkins, M. D. 2007. The Leader ship Team: Complementary Strengths or Conflicting Agendas? *Harvard Business Review, April 2007.*

Nyhus, I., Steiro, T. J., & Torgersen, G. E. 2018. Apprenticeship Learning in Preparation for Meeting the Unforeseen. In: Torgersen, G. E. 2018. *Social interaction (Interaction) under Risk—a Step ahead of the Unforeseen.* Oslo: Cappelen Akademisk. Open Access.

Paries, J. 2011. Lessons from the Hudson. In: Hollnagel, E., Pariés, J., Wood, D. & Wreathall, J. 2011. *Resilience Engineering in Practice: A Guidebook.* Boca Raton: CRC Press. Taylor & Francis Group.

Primrose, M. J., Bentley, P. D., van der Graaf, G. C.& Sykes, R. M. 1996. Shell International Exploration and Production B. V. The HSE Management System in Practice—Implementation. *Paper Presented in the Third International Conference on Health, Safety and Environment in Oil & Gas Exploration & Production. New Orleans, June 9.-12. Paper ID: SPE 35826.*

Roux-Dufort, C. 2007. Is Crisis Management (Only) a Management of Exceptions? *Journal of* Contingencies and Crisis Management, 15(2), 105–114.

Säljö, R., & Moen, S. 2001. *Læring i praksis: Et sosiokulturelt perspektiv* [Learning in Praxis—A Sociocultural Perspective]. Oslo: Cappelen akademisk.

Scholtens, A. 2008. Controlled Collaboration in Disaster and Crisis Management in the Netherlands, History and Practice of an Overestimated and Underestimated Concept. *Journal of Contingency and Management, 16*(4), 195–207.

Simensen, J. O. 2005. *Godfot—arven. Knut Torbjørn Eggen i samhandling med Nils A [Strong leg heritage. Knut Torbjørn Eggen in interaction with Nils A.]* O. Oslo: Aschehoug.

Steiro, T. J., & Torgersen, G. E. 2018. Weltbürger Perspectives and Samhandling [«Interaction»]. In: Torgersen, G. E. 2018. *Samhandling (Interaction) under Risk—a Step ahead of the Unforeseen.* Oslo: Cappelen Akademisk. Open Access.

Steiro, T. J., & Torgersen, G. E. 2013. The Terms of Interaction and Concurrent Learning in the Definition of Integrated Operations. In: Rosendahl, T. & Hepsø, V. (Eds.). 2013. *Integrated Operations in the Oil and Gas Industry: Sustainability and Capability Development*. Hersey: IGI Global.

Steiro, T. J., Johansen, P., Andersen, B., & Olsvik, L.S. 2013. Balancing Structure and Learning in an Open Prison. *International Journal of Management, Knowledge and Learning*, (1), 101–121.

Torgersen, G. E. 2018 a. *Samhandling* (Interaction) under Risk—a Step ahead of the *Unforeseen*. Oslo: Cappelen Akademisk. Open Access.

Torgersen, G. E. 2018 b. Introduction. In: Torgersen, G. E. 2018. *Samhandling (Interaction) under Risk—a Step ahead of the Unforeseen.* Oslo: Cappelen Akademisk. Open Access.

Torgersen, G. E. & Steiro, T. J. 2018. Defining the Term Samhandling [Interaction]. In: Torgersen, G. E. 2018. *Samhandling (Interaction) under Risk—a Step ahead of the Unforeseen*. Oslo: Cappelen Akademisk. Open Access.

Torgersen, G. E., & Steiro, T. J. 2009. *Ledelse, social interaction og opplæring i fleksible organisasjoner [Leadership, Social interaction and Education in Flexible Organizations]*. Stjørdal: Læringsforlaget.

Weick, K. E. 2015. Ambiguity as Grasp: The Reworking of Sense. *Journal of Contingencies and Crisis Management,* 23(2), 117–123.

Weick, K. E., & Sutcliffe, K. M. 2001. *Managing the Unexpected: Resilient Performance in an Age of Uncertainty*. Jossey-Bass.

Weick, K. A. (1993). The Collapse of Sensemaking in Organizations. The Mann Gulch disaster. *Administrative Science Quarterly, 38*(4), 628–652.

Weick, K. E. 1987. Organizational Culture as a Source of High Reliability. *California Management Review, 29* (2), 112–127.

Zuboff, S. 1988. *In the Age of the Smart Machine: The future of work and power.* Basic books.

Wenger, E. 1998. *Communities of Practice: Learning, Meaning, and Identity*. Learning in Doing: Social, Cognitive, and Computational Perspectives). Cambridge: Cambridge University Press.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Implementation guidance for resilience management of critical infrastructure

Gonçalo Cadete
Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

Bjarte Rød
UiT The Arctic University of Norway, Tromso, Norway

Miguel Mira da Silva
Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

ABSTRACT: Existing resilience management and assessment frameworks for Critical Infrastructure (CI), as well as resilience indicator indexes, define categories and methodologies that provide useful conceptual models. However, there is a lack of specific guidance on how to implement such conceptual models in practice, thus enabling stakeholders to conduct assessment, auditing, and consulting initiatives in a consistent manner. In this paper, we use an existing CI resilience management framework—based on the ISO 31000 risk management process–, and present guidance for implementing such a framework, based on generalized COBIT5 best practice. This is done by illustrative demonstration with a defined risk scenario that includes Information and Communications Technology (ICT) aspects, using a specific process from an existing CI resilience assessment framework.

1 INTRODUCTION

Assessing and managing the resilience of Critical Infrastructure (CI) present many conceptual and implementation problems, that CI operators, regulators, assessors, auditors, and consultants must address. High conceptual complexity kicks-in when hybrid threats, dependencies, and time-variable cascading effects are considered. Practical implementation issues arise, as crisis may span interdependent sectors and sovereign border, calling for frameworks and standards that promote interoperability and facilitate communication, cooperation, and collaboration among diverse stakeholders. Finally, the growing importance of the cyber dimension calls for an integrated people-cyber-physical approach to help solve the social-technical challenges of resilience management.

Significant research and development investments have been made in recent years to address such challenges (IMPROVER 2017, CIPRNet 2017, DRIVER 2017, FORTRESS 2017, PREDICT 2017, SmartResilience 2017, ERNCIP 2017).

In this paper we use the term resilience as defined by UNISDR (2017), i.e. "*The ability of a system, community or society exposed to hazards to resist, absorb, accommodate, adapt to, transform and recover from the effects of a hazard in a timely and efficient manner, including through the preservation and restoration of its essential basic structures and functions through risk management*". From this definition of resilience, three important aspects may be highlighted:

- The concept of resilience goes beyond the domains of security and safety – related to protection–, to include additional goals and functions related to crisis management – such as response, recovery, and adaptation;
- Risk management is an important function to ensure achieving the desired ability, towards a defined optimal level;
- Since UNISDR defines resilience as an *ability*, resilience management models and frameworks may take the point of view of optimizing sets of *capabilities* and *capacities*, as well as to assess resilience based on related indicators and measurement models.

Existing resilience management and assessment frameworks for Critical Infrastructure (CI), as well as resilience indicator indexes, define methodologies and categories that provide useful conceptualizations. A recent review of such resilience methodologies and frameworks can be found in Rød et al. (2017), Lange et al. (2017a, 2017b), and

IMPROVER (2016). Such conceptual models help address complexity, since they are simplified representations of the systems, intended to promote understanding within the relevant domain of discourse.

However, there's a lack of specific guidance on how to implement such conceptual models in practice, thus enabling stakeholders to conduct assessment, auditing, planning, and implementation initiatives in a consistent and effective manner.

Due to the growing importance of the cyber dimensions, it is important to note that security and resilience of CI requires consideration and integration of cyber and physical aspects, therefore a cyber-physical approach is needed (Choraś 2017, NIST 2014).

For addressing the government and management aspects of the cyber dimension, COBIT5 provides state-of-the-art conceptual models and guidance. COBIT5 is used e.g. in the NIST Cybersecurity Framework, for providing governance and management best practice (NIST 2014).

In this paper, we provide implementation guidance for resilience management and assessment of critical infrastructure, by developing, integrating, and demonstrating the two functional aspects:

- *Resilience management*: we reuse an existing CI resilience management framework from Lange et al (2017a, 2017b) –based on the ISO 31000 risk management process–, and present guidance for implementing such a framework, based on generalized COBIT5 best practice. By itself, the COBIT5 framework only addresses part of the problem, namely the enterprise governance of information and related technologies. However, as we propose and demonstrate in this paper, some of the key concepts, models, methods, and guidance of the COBIT5 framework may be generalized to encompass the broader scope of CI operator objectives, activities, and supporting technologies. Such an approach allows for a holistic solution to the resilience management problem, ensuring—by design—that the cyber dimension is integrated using a state-of-the art governance framework for information and related technologies.
- *Resilience assessment*: based on the resilience assessment framework from Cadete et al (2017), we define generic requirements for resilience assessment models, to enable seamless integration with the COBIT5 assessment and measurement models.

To illustrate the proposal, we provide a demonstration that uses a risk scenario that includes information and communications technologies (ICT) aspects, using as an example the resilience assessment framework from Cadete et al (2017).

2 RELATED WORK

In Rød et al (2017), a selection of existing resilience assessment methodologies is described and evaluated in the context of the IMPROVER project (IMPROVER 2017). The authors conclude that there is a need for a CI resilience assessment framework that is sufficiently well defined, but at the same time may be flexible to account for idiosyncrasies of the different types of CIs and their operators. Also, such a framework should remain compatible with the current guidelines for risk assessment of the European Union (EU) Member States and should integrate the paradigm of resilience into the risk assessment process according to ISO 31000.

Lange et al (2017a, 2017b) addressed those needs, proposing a framework for resilience assessment of CI, which integrates the resilience paradigm into the risk assessment (RA) process according to ISO 31000, while maintaining compatibility with the current European guidelines for national RA applied by the EU Member States (Fig. 1). Starting from definitions used in ISO 31000 for risk assessment, the authors propose a conceptual framework that maps these risk concepts to the resilience assessment domain:

- **Resilience assessment:** Resilience assessment is the overall process of resilience analysis and evaluation.
- **Resilience analysis:** Resilience analysis is the process of determining the level of resilience.
- **Resilience evaluation:** Resilience evaluation is the process of comparing the results of resilience analysis with criteria or objectives to determine whether resilience level is acceptable and identify areas for improvement.
- **Resilience treatment:** the process to modify resilience, focusing on the absorptive, adaptive or restorative capacity.
- **Resilience management:** Coordinated activities to direct and control an organization regarding its resilience, including the above processes.

According to these conceptual mappings, the authors propose that some of the expected outputs of a risk analysis could contribute to some of the indicators required to carry out a resilience analysis, i.e. the outputs may be utilized in the form of a single/multi-hazard engineering analysis to provide input to suitable resilience analysis methodology.

Dependencies (of other CI and suppliers) accounted during the risk analysis stage, may be used in the CI resilience analysis. The authors also propose that the maturity of organizational processes (see Fig. 1) may be used to provide more input to the resilience analysis methodology.

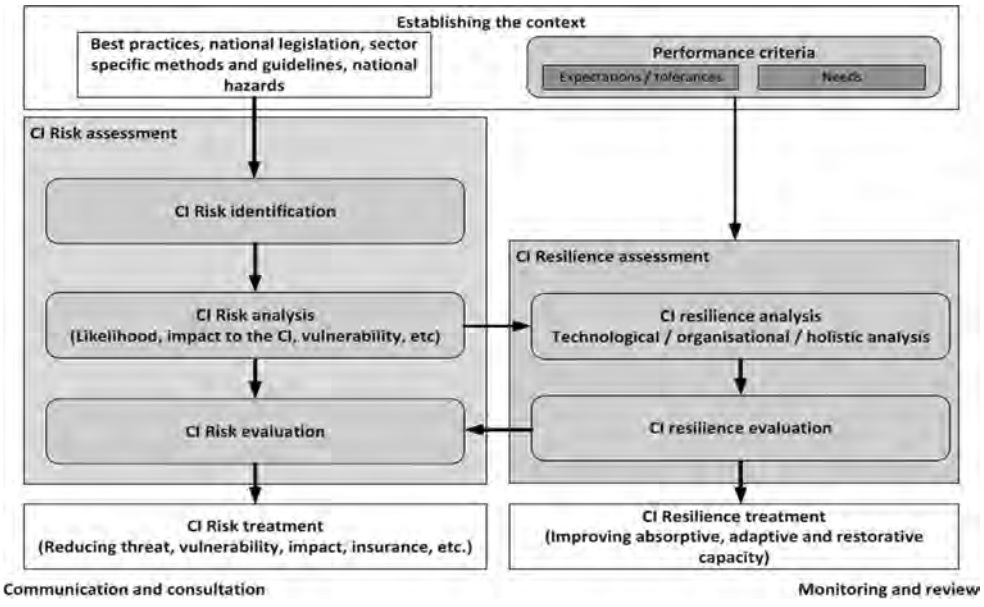


Figure 1. CI risk assessment (RA), incorporating CI resilience assessment. Taken from Lange et al (2017b).

The "*establishing the context*" activity (Fig. 1) provides input to the overall CI risk assessment process, requiring e.g. knowledge about best practices in the industry in question, national legislation, sector specific methods of RA, as well as relevant hazards identified in national RA standards.

However, the authors conclude that future work is needed to implement the proposed framework, namely to develop "*a resilience evaluation methodology which is compatible with a resilience analysis methodology and which allows a comparison between the results of a resilience analysis and real performance objectives based on the needs and tolerances of the society or community which relies on the service provided by the CI*".

Regarding resilience assessment models, Cadete et al (2017) propose a resilience assessment framework, based on rating the capability levels of organizational processes, that allows for a direct comparison between the capability levels that result from an assessment and the desired capability levels–thus enabling CI resilience evaluation. Both sets of capability levels (i.e. assessed and desired) are determined using the same goals cascade methodology, thus allowing for straightforward CI resilience evaluation, i.e. for:

- comparing the results of resilience analysis with criteria or objectives to determine whether the assessed resilience level is acceptable, and;
- identify areas for improvement.

The proposed framework consists of the following components:

- A **Goals Cascade Methodology (GCM)**, for ensuring that stakeholder needs are met, as well as for tailoring the framework usage for each CI organizational setting. This methodology is consistent with the COBIT5 goals cascade, thus enabling consideration of ICT concerns, which are increasingly important for all CI sectors. The proposed goals cascade methodology is derived from the COBIT5 goals cascade. This method allows for customizing the framework for different organizational settings, recognizing that each critical infrastructure organization has its own objectives and context. The goals cascade corresponds to the COBIT5 framework principle of meeting stakeholder needs. It provides value-based discernment for decision-making at several management levels, namely the discernment criteria for assessing risk, as well as for selecting risk indicators. Similarly, in this framework, the stakeholder needs translate to a cascade of inter-related goals at different levels, starting from high-level governance goals, cascading through all required management levels, down to relevant operational levels. Therefore, the GCM allows for tailoring the assessment rationale to any sector-specific, country-specific, as well as disaster-specific scenarios. Interestingly, the recently released *NIST Community Resilience Planning Guide for Building and Infrastructure Systems* (NIST 2015) also

approaches resilience planning activities based on understanding individual and social needs, which are then mapped to community goals and desired recovery performance goals, in recognition that "*the community's social and economic needs and functions should drive goal-setting for how the built environment performs*" (NIST 2017).

– A **Process Reference Model (PRM)**, is a set of interrelated processes that enables the governance and management of CI organizations, from a disaster risk management viewpoint. This reference model is integrated in the broader governance and management of the organization, thus embedding disaster risk management concerns in business-as-usual activities. The proposed model presents a disaster risk management viewpoint, and includes relevant concerns from the COBIT5 and the NIST Cybersecurity Framework (NCF). The NCF provides a common language for enabling communication and cooperation in cybersecurity risk management, for the information technology (IT) and industrial control systems (ICS) environments. The NCF core defines a set of relevant activities that help in the analysis, prioritization and implementation of cybersecurity countermeasures. Security functions are defined at the highest level of abstraction, helping to express activities for management of cybersecurity risk, and defined as follows:
– Identify: develop the organizational understanding to manage cybersecurity risk to systems, assets, data, and capabilities;
– Protect: develop and implement the appropriate safeguards to ensure delivery of critical infrastructure services;
– Detect: develop and implement the appropriate activities to identify the occurrence of a cybersecurity event;
– Respond: develop and implement the appropriate activities to act regarding a detected cybersecurity event; and
– Recover: develop and implement the appropriate activities to maintain plans for resilience and to restore any capabilities or services that were impaired due to a cybersecurity event.

A diagram of the PRM from Cadete et al. (2017) is shown in Figure 2. The governance and management enabling processes are aggregated in three interconnected groups:

1. *Cisis management*: a central process group for enabling specific disaster management functions, following the NCF terminology: identify, protect, detect, respond, and recover.
2. *Generic enablers*: a process group for cross-functional (generic) enablement, improvement, learning, and adaptation, for continuously enhancing disaster management capabilities, including generic governance and management capabilities.
3. *Risk*: processes for risk governance and management, conveying the idea that risk management should be continually performed on all other enabling capabilities for disaster management, thus promoting a clear disaster risk reduction perspective. Note that the risk-related processes refer to organization-wide processes (i.e. are not specific to disaster management), thus ensuring that disaster risk management concerns are embedded in the broader enterprise risk management (ERM) activities.

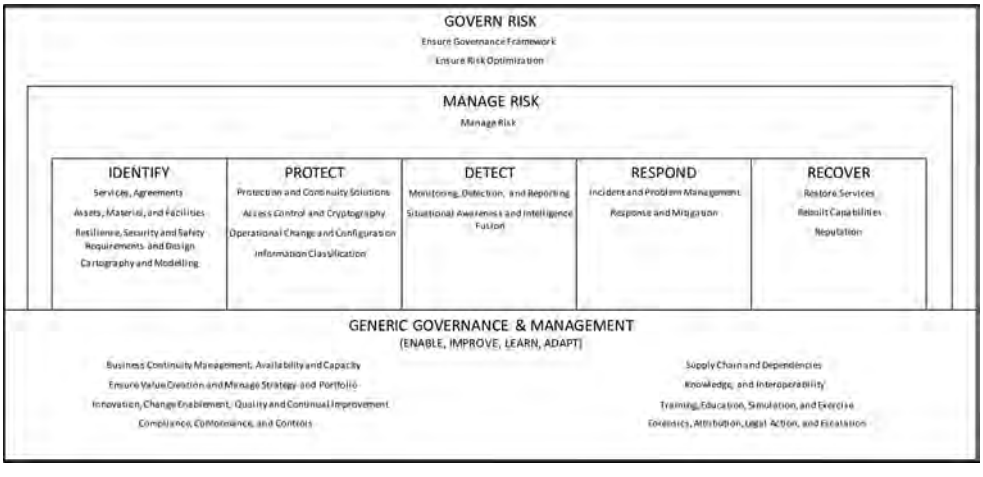


Figure 2. Process Reference Model (PRM), for the 26 disaster risk management processes. These organizational processes cover risk management (govern risk and manage risk), disaster management (identify, protect, detect, respond, and recover), as well as generic governance and management processes. Taken from Cadete et al (2017).

– A ***Process Assessment Model*** **(PAM) and a** ***Process Measurement Model*** **(PMM)**. These models are based on the ISO/IEC 33000 series guidelines. Like in COBIT5, the assessment results are expressed according to the capability level ratings achieved for each enabling process. Note that a certain capability level rating "*does not guarantee that an organization will perform its processes at any given process capability level, simply that it is capable of performing its processes at that level*" (ISO 2015). The capability levels follow the ISO/IEC 33020xx definitions, ranging from "*Incomplete*" to "*Innovating*" (the latter maps to the "*Optimizing*" COBIT5 capability level, based on ISO 15504).

The authors discuss why process assessment and measurement models are useful for conducting risk and resilience assessments, noting that risk is defined, in generic terms, as "*the effect of uncertainty on objectives*" (ISO 2009) and that processes "*have clear business reasons for existing, accountable owners, clear roles and responsibilities around the execution of the process, and the means to measure performance*" (ISACA 2012a). Note also that the first process capability level is named "*Performed*" and addresses precisely the achievement of organizational objectives in a certain process area, thus serving as a useful indicator for risk and resilience management. The achievement of higher capability levels provides stronger assurances regarding risk reduction, although generally implying higher implementation costs (Cadete et al. 2017).

3 RESEARCH PROBLEM

Although Lange et al (2017a, 2017b) proposes a CI risk assessment process model incorporating CI resilience assessment, and the resilience assessment framework from Cadete et al (2017) allows for establishing a link between resilience analysis and resilience evaluation—based on measuring and comparing capability levels of disaster risk management processes–, we currently lack a methodology for guiding the implementation of such conceptual models in practice.

Without such guidance, stakeholders will not be able to conduct assessment, auditing, and consulting initiatives in a consistent manner. Also, results from different CI assessment initiatives (e.g. different sectors, operators, auditing and consulting companies) will hardly be comparable, hindering the effectiveness of enterprise, policy, and regulatory initiatives.

4 PROPOSAL

In this paper we seek to provide a contribution to this problem, by providing guidance on how to manage, assess, and measure resilience-related capabilities which are expressed in the form of capability levels of organizational processes. As stated before, this approach is important because risk may be defined as "*the effect of uncertainty on objectives*" (ISO 2009c), and that processes "*have clear business reasons for existing, accountable owners, clear roles and responsibilities around the execution of the process, and the means to measure performance*" (ISACA 2012a). It is also important to clarify that governance and management processes have an overarching role in risk management, as stated in CISM guidance (ISACA 2016): "*most security failures can ultimately be attributed to failures of management, and it must be remembered that management problems typically do not have technical solutions*."

To this end, we propose to use generalized COBIT5 guidance for risk management, based on COBIT5 for Risk (COBIT 2013c) best practice. Note that COBIT5 provides a framework that assists enterprises in achieving their objectives for the governance and management of enterprise information technology (IT). As such, its IT scope is narrower that what we need to address the broader CI resilience assessment and management problem. However, the framework fundamentals may be easily generalized towards creating higher-order generic governance and management frameworks.

The COBIT5 extensive body of knowledge is based on the following objectives and principles (ISACA 2012b):

– *Governance objective*: enterprises exist to create value for their stakeholders. Consequently, any enterprise, commercial or not, has value creation as a governance objective. Value creation means realizing benefits at an optimal resource cost while optimizing risk. Note that, in COBIT5, IT risk is treated as business risk, therefore the COBIT5 recommendations may be generalized to other risk types.
– *Principle 1*: Meeting Stakeholder Needs: the purpose of risk governance and risk management is to help ensure that enterprise objectives are achieved throughout the goals cascade. Optimizing risk is one of the three components of the overall value creation objective for an enterprise. Note that this principle is aligned with NIST guidance, that approaches resilience planning activities based on understanding individual and social needs, which are then mapped to community goals and desired recovery performance goals, in recognition that "*the community's social and economic needs and functions should drive goal-setting for how the built environment performs*" (NIST 2017).

– *Principle 2*: Covering the Enterprise End-to-end: COBIT 5 for Risk covers all governance and management enablers in its scope and describes all required phases of risk governance and risk management. COBIT 5 does not focus on only the IT function, but treats information and related technologies as assets that need to be addressed like any other asset, by everyone in the enterprise. Also, governance of enterprise IT is integrated into enterprise governance. This principle is relevant to ensure that resilience management and assessment frameworks are able to cover all cross-sector concerns engaged in crisis management (especially due to CI interdependencies).
– *Principle 3*: Applying a Single Integrated Framework: COBIT 5 for Risk aligns with major risk management frameworks and standards, such as ISO 31000, ISO/IEC 27005, and COSO ERM (COSO 2017).
– *Principle 4*: Enabling a Holistic Approach: COBIT 5 for Risk identifies all interconnected elements of the enablers that are required to adequately provide risk governance and management, presenting a holistic and systemic approach towards risk.
– *Principle 5*: Separating Governance from Management: COBIT 5 distinguishes between risk governance and risk management activities. Good governance means that risk optimization is part of the governance arrangements that are put in place and risk information is included in the decision-making process, at the highest levels of leadership and related accountability.

A benefit of using a risk management approach that applies generic COBIT5 rationale is that the cyber and informational dimensions of CI are directly mapped to the information and related technologies content of COBIT5. Addressing the full people-cyber-physical challenge requires the seven categories of enablers of COBIT5, as per Principle 4:

– Principles, Policies and Frameworks;
– Processes;
– Organizational Structures;
– Culture, Ethics and Behaviour;
– Information;
– Services, Infrastructure and Applications;
– People, Skills and Competencies.

An adequate process reference model, such as the one presented in Figure 2, covers the "*Processes*" enabler dimension directly. Since COBIT5 is a holistic body of knowledge, the other six enablers are covered indirectly:

– The people dimension is covered by the enablers "*People, Skills, and Competencies*", "*Culture, Ethics, and Behaviour*", and "*Organizational Structures*", as well as informed and enforced by the enabler "*Principles, Policies, and Frameworks*".
– The enablers "*Information*" and "*Service, Infrastructure, and Application*" should be understood in the broad informational and technological scope of CI operator activity, i.e. including IT, operational technology (OT), material, facilities, and other infrastructural and technological artifacts.
– The enablers "*Information*", "*Service, Infrastructure, and Application*", and "*People, Skills and Competencies*" correspond to organizational resources. This entails that the concept of capability should include measures of (resource) capacity.

Considering the above COBIT5 guidance, as well as COBIT5 for Risk (ISACA 2013b), we may provide implementation guidance for CI resilience management, as described in the following sub-sections. In the remainder of this section we assume that GCM and PAM/PMM models are used—such as the one proposed in Cadete et al (2017).

4.1 *Establishing the context*

Using the GCM, we can translate stakeholder needs into specific, actionable, and customized goals cascade, from high-level enterprise goals, down to enabler goals.

For risk assessment, these enabler goals will later be associated with best practice, standards, and compliance requirements from relevant areas of concern.

During this risk management activity, the external and internal contexts should be established, and risk criteria—i.e. terms of reference against which the significance of a risk is evaluated—should be developed (ISACA 2013b).

4.2 *CI Risk Identification*

COBIT5 recommends an approach based on risk scenarios. A risk scenario is a description of a possible event that, when occurring, will have an uncertain impact on the achievement of the enterprise's objectives.

Risk scenarios can be identified and developed using two different mechanisms (ISACA 2013b):

– A top-down approach: starting from the overall enterprise objectives, identify business objectives and scenarios with highest impact on achievement of business objectives.
– A bottom-up approach: a list of generic scenarios is used to define a set of more relevant and customized scenarios, applied to the individual enterprise situation.

Note that using risk scenarios, as a specific method for identifying risk, does not exclude using other risk techniques e.g. from ISO 31010, to assist risk identification, as well as risk analysis activities.

4.3 *CI risk analysis*

In this activity, the frequency and impact of risks are estimated. Risk factors should be considered, i.e. those conditions that influence the frequency and/or business impact of risk scenarios. They can also be interpreted as causal factors of the scenario that is materializing, or as vulnerabilities or weaknesses.

Scenario analysis may be based on past experience, known current events, and also possible future circumstances (ISACA 2013b).

4.4 *CI resilience analysis*

Using a PAM and a PMM as assessment and measurement models, the analyst can relate the risk scenarios with current and forecasted organizational capability levels (including resource capacities), to calculate current risks and residual risks (equal to current risks with additional risk responses applied) (ISACA 2013b).

4.5 *CI resilience evaluation and treatment*

In this activity we compare the results of resilience analysis with criteria to determine whether resilience level is acceptable and identify areas for improvement. The purpose of defining a risk response (avoidance, acceptance, sharing/transfer, or mitigation) is to bring risk in line with the defined risk appetite for the enterprise (ISACA 2013b).

4.6 *CI risk evaluation and treatment*

The CI risk and resilience risk responses should be integrated in a common decision-making process and contribute to a single security, safety, and resilience strategy, although the implementation of specific resilience controls may be realized by specialized initiatives or projects, within a coherent portfolio.

5 DEMONSTRATION

To illustrative how to implement the proposed guidance, for conducting resilience management and assessment, we will address the ERNCIP Project Platform concerns (ERNCIP 2017) and adapt a demonstration from Cadete et al (2017).

- *Establishing the context:* we situate ourselves in the context of the ERNCIP Project Platform, addressing the concerns of the *Thematic Group (TG)* on *Chemical and Biological (CB) Risks to Drinking Water*. Adopting the framework from Cadete et al (2017) as our reference model, we use the GCM and define that the enabling process area "*Monitoring, Detection, and Reporting*" should achieve capability level "*Innovating process*" (ISO 33020), given the ERNCIP requirements for high levels of risk control and innovation.
- *Communication and Consultation:* note that the ERNCIP Project Platform itself provides a forum for communication and consultation purposes, i.e. addresses "*continual and iterative processes that an organization conducts to provide, share or obtain information and to engage in dialogue with stakeholders regarding the management of risk*" (ISO 2009).
- *CI Risk Identification*: using a top-down approach, we address the ERNCIP TG work programme goals and choose e.g. a bio-terrorism risk scenario.
- *CI Resilience Analysis:* using the PAM/PMM from Cadete et al (2017) as assessment and measurement models, we assume that the analyst (we may also assume external auditing validation) assessed the current capability level of the "*Monitoring, Detection, and Reporting*" process as "*Established process*". Note that this means that the current capability level is below the desired capability level of "*Innovating process*".
- *CI Resilience Evaluation and Treatment:* mitigation is proposed as risk response, meaning designing and executing a project to move from "*Established process*" to "*Innovating process*", to meet risk appetite criteria. Using ICT best practice from COBIT5, this project entails achieving a capability level of "*Innovating process*" for the COBIT5 processes "*DSS01 Manage Operations*" and "*DSS02 Manage service requests and incidents*". Further detailed guidance can be found in COBIT5 documentation (ISACA 2013a).
- *CI Risk Evaluation and Treatment:* finally, after designing the project and estimating related implementation costs, the plan is submitted to the approval of top management. A decision is made to include the new project in the ongoing organizational security programme, with a mandate to address all possible portfolio synergies, as well as to report to governance bodies accordingly, for *monitoring and review* purposes.
- *Monitoring and Review:* as assumed in the previous paragraph, monitoring and review should include the governance bodies, for the purposes of ensuring that controls are effective and efficient in both design and operation, obtaining further information to improve risk assessment, analyzing and learning lessons, detecting changes

in the external and internal context, and identifying emerging risks (ISO 2009).

6 CONCLUSION

The main contribution of this paper is to provide implementation guidance for resilience management and assessment of critical infrastructure, using the resilience management framework from Lange et al (2017a, 2017b). Also, we demonstrated practical implementation of this guidance, to facilitate understanding and provide validation, using the context of a real-world resilience initiative (ERNCIP 2017), a fictional risk scenario, and addressing ICT concerns using a state-of-the-art governance and management framework for information and related technologies (ISACA 2012b).

The ultimate goals of this contribution are to facilitate conducting assessment, auditing, and consulting initiatives in a consistent manner, as well to enable comparing assessment results from initiatives performed in different CI settings (e.g. different operators, sectors, or countries), using a standards-based approach.

Note that the resilience assessment framework discussed in the paper (Cadete et al. 2017) is not essential to the proposal, and may be replaced by another similar framework if the following generic requirements are met:

1. Approach: is aligned with the ISO 31000 risk management process, adopts a cyber-physical approach, and provides direct or indirect coverage for the seven COBIT5 enablers.
2. Governance: is aligned with the governance objective and principles, as described in the proposal.
3. GCM: provides a goals cascade methodology, so that risk may be address as "*the effect of uncertainty on objectives*" (ISO 2009), and we can translate stakeholder needs into a specific, actionable, and customized goals cascade.
4. PRM: provides a reference model based on organizational processes, covering all relevant resilience management and assessment areas of concern.
5. PAM/PMM: is aligned with ISO 33020 or a similar standard (such as ISO 15504).

Regarding limitations, note that this paper does not claim that the proposed guidance is the best approach for resilience management of critical infrastructure, instead assuming the framework from Lange et al (2017a, 2017b) is a reasonable approach. However, we have demonstrated that the framework from Lange et al (2017a, 2017b) can be used to derive an approach that is consistent with ISO 31000 and COBIT5 principles. Also, we have demonstrated that it can be used consistently with process-based resilience assessment frameworks.

REFERENCES

Cadete, G. et al 2017. A Conceptual Framework for Assessing the Resilience of Critical Infrastructure. *Safety and Reliability—Theory and Applications*, ISBN 978-1-138-62937-0, Taylor & Francis Group, London.

CIPRNet 2017. Official website for the CIPRNet—The Critical Infrastructure Preparedness and Resilience Research Network. Available: https://www.ciprnet.eu/.

Choraś, M. et al 2016. Cyber Threats Impacting Critical Infrastructures. *Managing the Complexity of Critical Infrastructures—A Modelling and Simulation Approach,* ISBN 978-3-319-51042-2, Studies in Systems, Decision and Control 90, Springer Open.

CIPRNet 2017. Official website for the CIPRNet—The Critical Infrastructure Preparedness and Resilience Research Network. Available: https://www.ciprnet.eu/.

COSO 2017. Enterprise Risk Management—Integrating with Strategy and Performance, Executive Summary. COSO—Committee of Sponsoring Organizations of the Treadway Commission.

DRIVER 2017. Official website for the DRIVER Project—DRIVing Innovation in Crisis Management for European Resilience. Available: http://driver-project.eu/.

EC 2010. Risk Assessment and Mapping Guidelines for Disaster Management SEC (2010) 1626 final. Brussels: European Commission.

ERNCIP 2017. *Official website for the The ERNCIP Project Platform*. Available: https://erncip-project.jrc.ec.europa.eu/.

EU 2008. Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection. Official Journal of the European Union.

EU 2016. Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union (NIS Directive). Official Journal of the European Union.

FORTRESS 2017. Official website for the FORTRESS Project—Foresight Tools for Responding to cascading effects in a crisis. Available: http://fortress-project.eu/.

IMPROVER 2016. *Report of criteria for evaluating resilience. Deliverable Number: D2.2*. Available: http://improverproject.eu/. The IMPROVER project—Improved risk evaluation and implementation of resilience concepts to Critical Infrastructure.

IMPROVER 2017. Official Website for The Improver Project—Improved risk evaluation and implementation of resilience concepts to Critical Infrastructure. Available: http://improverproject.eu/.

ISACA 2012a. *COBIT 5 Implementation*, Rolling Meadows, IL, USA, ISACA.

ISACA 2012b. COBIT 5: A Business Framework for the Governance and Management of Enterprise IT, Rolling Meadows, IL, USA, ISACA.

ISACA 2013a. COBIT Process Assessment Model (PAM): Using COBIT 5, Rolling Meadows, IL, USA, ISACA.

ISACA 2013b. *COBIT 5 for Risk*, Rolling Meadows, IL, USA, ISACA.

ISACA 2016. *CISM® Review Manual 15th Edition*, Rolling Meadows, IL, USA, ISACA.

ISO 2009. ISO 31000:2009 - Risk management—Principles and guidelines.

ISO 2012. ISO 22300:2012 Societal security—Terminology.

ISO 2015. ISO/IEC 33020 - Information technology—Process assessment—Process measurement framework for assessment of process capability.

Lange, D. et al 2017a. Incorporation of resilience assessment in critical infrastructure risk assessment frameworks. *Safety and Reliability—Theory and Applications*, ISBN 978-1-138-62937-0, Taylor & Francis Group, London.

Lange, D. et al 2017b. Framework for implementation of resilience concepts to Critical Infrastructure. Deliverable number: 5.1. Available: http://improverproject.eu/. The IMPROVER project—Improved risk evaluation and implementation of resilience concepts to Critical Infrastructure.

NIST 2014. Framework for Improving Critical Infrastructure Cybersecurity—Version 1.0. NIST—National Institute of Standards and Technology.

NIST 2015. Community Resilience Planning Guide for Buildings and Infrastructure Systems—Volume I. NIST—National Institute of Standards and Technology.

NIST 2017. *Official webpage for the Community Resilience Planning Guide*. NIST—National Institute for Standards and Technology. Available: https://www.nist.gov/el/resilience/community-resilience-planning-guide.

PREDICT 2017. Official website for the PREDICT Project—PREparing for the Domino effect in Crisis siTuations. Available: http://www.predict-project.eu/.

Rød, B. et al 2017. Evaluation of resilience assessment methodologies. *Safety and Reliability—Theory and Applications*, ISBN 978-1-138-62937-0, Taylor & Francis Group, London.

SmartResilience 2017. Official website for the SmartResilience Project. Available: http://www.smartresilience.eu-vri.eu/.

UNISDR 2009. *UNISDR Terminology on Disaster Risk Reduction*, United Nations International Strategy for Disaster Reduction (UNISDR), Geneva, Switzerland, May 2009.

UNISDR 2017. UNISDR Terminology on Disaster Risk Reduction, United Nations International Strategy for Disaster Reduction (UNISDR), Geneva, Switzerland. February 2nd, 2017 update. Available: http://www.unisdr.org/we/inform/terminology.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Impact of human factors on threats in sewage treatment plants

M. Łój-Pilch, A. Zakrzewska & E. Zielewicz
Institute of Water and Wastewater Engineering, Faculty of Energy and Environmental Engineering, Silesian University of Technology, Gliwice, Poland

ABSTRACT: Physical, biological and chemical process take place during wastewater treatment. Any, even the smallest, disruption of any process may pose a risk in operation of the treatment plant. For each type of wastewater, an appropriate method for their purification is established. It depends on the type of sewage, the population equivalent and the type of receiver The paper compares the threats that may pose a risk in municipal sewage treatment plants. Selected, for risk analysis, sewage treatment plants are located in the same city. They were designed in the same technological chain due to similar Population Equivalent (PE) and comparable conditions of discharge purified wastewater to the receiver. Previous risk analyses have shown the importance of the human factor, in the form of management, work organization, documentation and databases, technical knowledge and staff training. Employees of sewage treatment plants in their daily work can prevent negative events or cause them to occur. The location of the selected wastewater treatment plant in the same city area causes that they are managed by the same person and operating in the same way. This ensures a similar response to the same events, similar risk elimination during the operation and quality of databases. These condition allow for the unification of the human factor when comparing the treats at the municipal sewage treatment plants.

1 INTRODUCTION

Man can influence the environment in which he lives and all processes that take place in his surroundings in both positive and negative ways. His decisions and actions during exploitation of technical objects may contribute to the emergence of potentially dangerous situations and events which result in losses, both material and immaterial, or in other words may cause risks. Therefore, the human factor can be seen as one of risk factors. It affects the effectiveness of business management, work organization and keeping of records and databases. It is also worth considering in terms of risk for occupational safety, which is caused by low level or total lack of employees' knowledge in this field, their reluctance to change and improve the work process and their unfamiliarity with OSH rules [1]. Speaking about the risk in terms of the human factor the authors want to focus on the human impact on municipal sewage treatment plants operation process.

The basic stage of risk analysis is its identification and classification, which enables its management in the company [2, 3, 10]. Obtaining reliable results depends on the accuracy of conducting the process and the professionalism of the person who identifies threats. Of great importance are also exactness and diligence in maintaining a database. In other words, reliable risk identification also depends on the human factor.

2 HUMAN FACTOR IN A MUNICIPAL SEWAGE TREATMENT PLANT

Sewage treatment plants are specific technical objects in which physical, biological and chemical processes take place. Constant supervision of the technological process and introduction of monitoring and control systems are not able to exclude the emergence of failures and malfunctions of individual devices [4, 5, 11]. This happens because adverse situations are fortuitous events and it is impossible to predict when and where they may occur. Correctly conducted threat identification and competent risk management in the object allows the elimination of adverse events before they may arise, and the development of operating instructions for dealing with a given phenomenon [12]. Yet the effectiveness of these activities depends on the employees who will execute them. Therefore it can be stated that proper functioning of a treatment plant is largely dependent on the human factor.

3 CHARACTERISTICS OF EXAMINED TREATMENT PLANTS

Risk identification was performed for two municipal sewage treatment plants in the Upper Silesian Industrial District (GOP). The examined objects are characterized by similar technological lines, PE

Table 1. Characteristic of examined treatment plants.

	Plant A	Plant B
Population Equivalent—PE	52 000	62 500
Treatment technology	Activated sludge	
Elements of technological line	Expansion chamber Bar screens	Sifters
	Horizontal grit-removal tank, aerated, two-chamber	
	Biological dephosphatation chamber	
	Activated sludge chamber	
	Secondary settling tanks, radial	
Difference in the quantity of technological line elements	3 secondary settling tanks	2 secondary settling tanks
Possibility of sewage delivery with the use of sewage truck	Yes—cast station present	
Industrial objects in the catchment area	Meat processing plant, electro-plating plant	None
Receiver	Watercourse	

value and the method of sewage discharge to the receiver. The characteristics of plants in question are presented in Table 1.

The analyzed sewage treatment plants are supervised by the same manager, so they both have identical treatment plants' worksheets, in which all works carried out during the shift and any failures or adverse events are recorded. This should standardize one of the human factor components in threats identification.

4 METHODOLOGY

First step in identifying threats of analyzed sewage treatment plants was the thorough familiarization with the technology and the way of operating objects, based on submitted operating instructions. The recognition of adverse events was preceded by a thorough interview with the manager of the object and shift managers. Researchers got acquainted with current operating problems and previous methods of their elimination. Operating instructions [6,7] for the treatment plant obligate the staff to carefully keep the operating log, which consists of:

- Worksheets of the treatment plant—it is required to record all work carried out during the shift, failures and observations;
- Equipment repair and maintenance cards—contain records on repair and maintenance works and failures;
- Technological notebook—contains results of analyzes of laboratory tests and information about the amounts of waste generated in the treatment plant; and therefore further stages of work are based on these documents.

Next, after analyzing the worksheets, a list of risk factors was prepared together with the frequency of their occurrence. Any doubts and inaccuracies in the records were verified with the operator's personnel. The next stage was to get acquainted with the equipment repair and maintenance card in order to determine risk factors being eliminated during maintenance of individual objects. This allowed identification of latent risk.

The final stage of the work was to determine which of adverse events actually cause a risk in the operation of a sewage treatment plant. For this purpose, records of laboratory tests results of treated sewage at the outflow during and after the occurrence of a potential threat were checked with the use of technological notebook. In the case of incorrect values of any quality indicator, the occurrence of a given adverse event was treated as a threat causing losses—an emergence of a risk.

Identification of threats in the examined treatment plants was performed for a period of 3 years of object's exploitation. This period fell on the years 2014–2016.

5 THREATS IDENTIFIED IN MUNICIPAL SEWAGE TREATMENT PLANTS

5.1 *Identified risk factors [4]*

Risk factors in municipal sewage treatment plants may be divided into:

- internal—caused by the impact of treatment plant objects,
- external—caused by the influence of factors from outside the treatment plant,
- ordinary—easy to predict, often occurring during current operations,
- extraordinary—unpredictable fortuitous events, unusual situations,
- explicit (existing) – occurring in the past, often appearing periodically,
- latent—potentially possible, but not yet occurred (in a given object).

5.2 *Identified risk types [4]*

During the risk identification process of municipal sewage treatment plants one can distinguish the following types of risk:

- qualitative—results in lowering the level of sewage treatment or in digestibility of individual devices,

- operational—results in a decrease of the efficiency of technological line cleaning, and in extreme cases even in a breakdown of the process,
- ecological—results in a negative impact on the receiver and environment,
- financial—results in a financial outlay which must be borne by the sewage treatment plant as a consequence of the arisen threat.

6 SEWAGE TREATMENT PLANT "A"

In the analyzed three-year operational period of the sewage treatment plant, a total of 36 different threats were identified. Some of them were isolated events, others occurred several times. It was stated that a total of 111 situations could have been risk-causing events. Table 2 presents the identified threats for individual objects of the treatment plant.

The percentage share of risk-posing events (Figure 1) indicates that devices exposed to the highest risk are the bar screen and activated sludge chamber. Description of the most frequent events and specification of type of factors and type of risk they cause are presented in Table 3.

Table 2. Identified risk-posing factors for individual objects of treatment plant "A".

Device	Number of events
Cast station for delivered sewage	1
Local sewage pumping station	1
Bar screen	30
Grit chamber	18
Dephosphatation chamber	15
Activated sludge chamber	23
Clariffer	17
Sewage treatment plant	6
Total	111

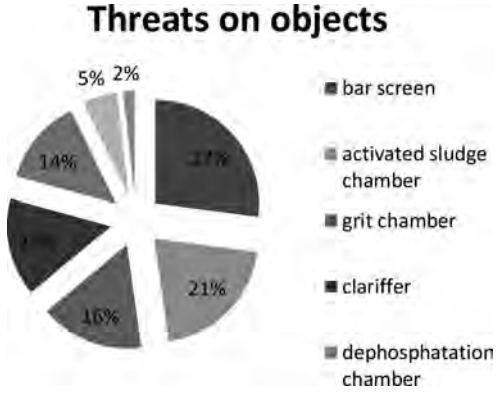


Figure 1. Percentage share of events that may cause risks in individual devices of the A plant's technological line.

Table 3. Examples of risk factors for selected elements.

Device	Event	Factor	Type of risk	Effect	Action taken/proposed
Bar screen	clogging of bars	external	qualitative	impeded sewage flow	cleaning of bars
Bar screen	large fat and meat dump	external	qualitative, operational, financial, ecological	greasing and clogging of bars, greasing of grit-removal tank	removal of excess fat, cleaning of bars
Bar screen	failure of bar screen automatics	ordinary	qualitative	closing of bar screen channel	repair of bar screen controlling automatics
Grit-removal tank	failure of grit-removal tank	ordinary	qualitative	increased amount of sludge in regulation chambers	repair of failure
Grit chamber	dump of greasy wastewater	external	qualitative	large amount of fats in the flotation area	pumping-out of fats
Dephosphatation chamber	sludge floating on the surface of dephosphatation chamber and activated sludge chamber	internal	qualitative, operational	bad sedimentation of sludge	lime dosage

(*Continued*)

Table 3. (*Continued*).

Device	Event	Factor	Type of risk	Effect	Action taken/proposed
Activated sludge chamber	freezing of activated sludge chamber surface	external	qualitative	ice formation on sludge coat	breaking the ice
Activated sludge chamber	formation of ice blockage	external	qualitative, operational	inability to draining of sewage	removal of blockage
Activated sludge chamber	problems with sludge draining, sludge putrefaction	internal	qualitative, operational	formation of scum layer	removal of scum layer and removal of putrefied sludge
Activated sludge chamber	emergence of filamentous bacteria	internal	qualitative	formation of scum layer	breaking the scum layer and bacteria removal
Secondary settling tank	freezing of the settling tank surface	external	qualitative, operational	formation of ice layer	breaking the ice layer
Secondary settling tank	auxiliary devices failure	ordinary	qualitative, operational	minor disturbance in the settling tank operation	repair of auxiliary devices
Sewage treatment	electrical power outage	external	qualitative, operational	lack of power for electrical powered devices	connection to the emergency power supply

Table 4. Identified risk-posing factors for individual objects of treatment plant "B".

Device	Number of events
Cast station for delivered sewage	2
Inflow to the plant	1
Sifters	9
Grit chamber	10
Activated sludge chamber	72
Clariffer	18
Sewage treatment plant	2
Total	114

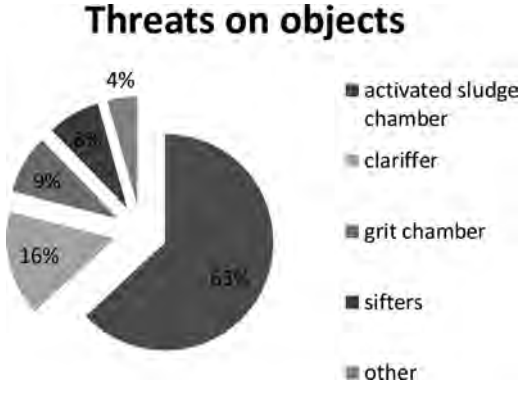


Figure 2. Percentage share of events that may cause risks in individual devices of the B plant's technological line.

7 SEWAGE TREATMENT PLANT "B"

Over the same time period, 32 different threats that could have caused risks were identified in treatment plant B. The number of risk-posing events was 114. Table 4 presents the identified threats and the number of their occurrences in individual objects of the "B" treatment plant.

Figure 2 shows that the most risk-posing device in the B plant was activated sludge chamber. The reason for this situation may be the attempt to improve the effectiveness of sewage treatment in the year 2015 and incursion in the technology used, as a result of which filamentous bacteria and scum layer appeared in activated sludge chamber. The most frequent threats are shown in Table 5.

8 COMPARISON OF THREAT IDENTIFICATION RESULTS IN THE EXAMINED TREATMENT PLANTS A AND B

Among the identified threats, 9 events occurred in both sewage treatment plants with different intensity (Table 6).

Large discrepancies in quantities of individual phenomena occurring in examined treatment plants may be caused by various circumstances, for instance:

Table 5. Examples of risk factors for selected elements.

Device	Event	Factor	Type of risk	Effect	Action taken/ proposed
Sifters	sifter scraper failure	ordinary	qualitative	clogging of sifter	repair of scrapper
Grit chamber	large dump of greasy sewage	external	qualitative	clogging of grease chamber outflow	unclogging the outflow
Activated sludge chamber	emergence of filamentous bacteria	internal	qualitative, operational	formation of scum layer	breaking the scum layer and actions aimed at stopping bacteria development
Activated sludge chamber	momentary change in quality of inflowing sewage	External ordinary	qualitative, operational	high concentration of ammonia in aeration chamber	turning on mammoth rotors
Activated sludge chamber	steering malfunction	ordinary	qualitative, operational	incorrect readings from activated sludge chamber	manual steering and repair of automatics
Activated sludge chamber	freezing of activated sludge chamber surface	external	qualitative	formation of ice layer	breaking the ice layer
Secondary settling tank	malfunction of discharge flume cleaning brush	ordinary	qualitative	cleaning of discharge flume impossible	reapair, cleaning of flumes
Secondary settling tank	auxiliary devices failure	ordinary	qualitative, operational	minor disturbance in the settling tank operation	repair of auxiliary devices
Sewage treatment	electrical power outage	external	qualitative, operational	no power for electrical powered devices	connection to emergency power supply
Sewage treatment	failure of devices visualization	ordinary internal	qualitative, operational	automatic steering of all devices impossible (aeration chambers operate in manual mode)	repair of visualization system

Table 6. Frequency of the same phenomena occurrence in sewage treatment plants A and B.

Device	Event	Treatment plant "A"	Treatment plant "B"
Cast station for delivered sewage	No flow	1	1
Grit chamber	Freezing of grit chamber elements	6	1
	Large dump of greasy sewage	8	1
Activated sludge chamber	Emergence of filamentous bacteria	1	41
	Freezing of activated sludge chamber surface	5	5
	Mammoth rotor malfunction	4	3
	Power cut of operation sensors	1	5
Secondary settling tank	Auxillary devices failure	15	7
Sewage treatment	Electrical power outage	6	1

- dump of greasy sewage was influenced by meat-processing industry in the catchment area of treatment plant A;
- the emergence of filamentous bacteria was the effect of attempts to improve the efficiency of treatment plant B, i.e. technological changes;
- momentary power outages in treatment plant A were caused by earthworks near the main power cable; treatment plant A was in five out of six situations notified beforehand.

Repetitive occurrences of only 9 adverse events in both analyzed objects does not mean that the remain-

ing events are distinctive for a given treatment plant. The same situations may be variously described in the worksheets depending on who makes the record on a given day. During the identification process, efforts were made to unify the records in operating logs, but this was not always possible.

The main operational problem of both treatment plants are fibrous substances not included in the above list. In treatment plant A, the cleaning of probes (activated sludge chamber, secondary settling tank) from fibers was called "a maintenance of individual objects". Due to the occurrence of this problem, initially assumed period between inspections was shortened from 6 to 3 months. On the other hand, in treatment plant B every 3 months on average the event called "removing of rags" has been recorded on individual objects (active sludge chamber, secondary settling tank). Only after conversation with the manager and employees of the sewage treatment plant it was found out, that it is all about the same operational activity.

The above examples show that the human factor is of great importance during identification of threats.

9 CONCLUSIONS

In both treatment plants appeared phenomena that under unfavorable conditions may be risk factors. In the analyzed period there were no exceedances of indicators examined at the outflow, so it can be concluded that potentially risk-posing situations did not adversely affect the process of sewage treatment. Employees' behavior, their decisions and actions taken when a risk-posing factor emerges have an impact on the level of sewage treatment attained by each treatment plant.

The human factor plays a decisive role in risk identification. Actions aimed at standardization of this factor (comparison of threats in two municipal sewage treatment plants with similar PE and a similar technological lines), have emphasized its role not only in the identification process but also in risk management. The methods of dealing with a given threat depend on the training and expertise of the staff.

ACKNOWLEDGEMENT

This paper was financed with funds for the directional studies of young researchers granted to the Institute of Water and Wastewater Engineering of the Silesian University of Technology in 2017.

REFERENCES

[1] Gajdzik, B., 2013. Components of risk connected with workers in the improvement of occupational safety and hygiene management system, Zeszyty Naukowe Wyższej Szkoły Zarządzania Ochroną Pracy w Katowicach, Nr 1(9)/2013, s. 44–45, ISSN-1895–3794.
[2] Kempa, E.S., 1995 Risk in Environmental Engineering Processes and Devices (Plants), Ochrona Środowiska 2(57),.
[3] Kempa, E.S., 2008 Risk analysis in water treatment systems, Ochrona Środowiska 3(50), 1993.
[4] Iwanejko, R., Rybicki, S.M., 2008. Zarządzanie ryzykiem dla oczyszczalni ścieków, Gaz, woda i technika sanitarna, February 2008.
[5] Rak, J., 2003. Methods of Risk Estimates for Water Supply Systems, Ochrona środowiska, Vol. 25, Nr 2/2003.
[6] Operating instruction of municipal sewage treatment plant "A"*.
[7] Operating instruction of municipal sewage treatment plant "B"*.
[8] Operating log of treatment plant "A"*.
[9] Operating log of treatment plant "B"*.
[10] Tchankova, L., 2002. Risk identification—basic stage in risk management, *Environmental Management and Health*, Vol. 13 Issue: 3, 2002, pp.290–297.
[11] Kirwan, B., 1998. Human error identification techniques for risk assessment of high risk systems, *Applied Ergonomics*, Vol. 29 Issue 3, June 1998, pp. 157–177.
[12] Kasap, D., Kaymak, M., 2007. Risk Identification. Step of the Project Risk Management, *International Conference on Management of Engineering & Technology*, Portland 2007.

* Internal documents of the sewage treatment plant.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Field operations in the high arctic—experienced feedback and tacit knowledge as key tools for safety management

M. Indreiten
The University Center in Svalbard

E. Albrechtsen
The University Center in Svalbard
The Norwegian University of Science and Technology, Trondheim, Norway

S.M. Cohen
The University Center in Svalbard

ABSTRACT: The paper demonstrates the importance of tacit knowledge to cope with various situations during field work in the high arctic. Two cases from field work at the University Center in Svalbard (UNIS) are shown to exemplify this: one boat trip and one snow mobile trip with researchers and students. Successful field operations depend heavily on technicians from the UNIS Field Safety Section that have the responsibility to assist in the planning and execution of every type of field work. Due to rapidly changing conditions, local variations, extreme weather conditions, lack of access to infrastructure and communication, successful safety performance is accomplished by individual's ability to adapt to situations. The paper demonstrates that this ability to a large extent is a function of the tacit knowledge of the technicians. To improve the tacit knowledge of each technicians, systems and practices of experience feedback must be run to ensure individual and organizational learning from both failures as well as successes. This is in particular important in systems with great variability in climatic conditions and systems with organizational changes.

1 INTRODUCTION

The University Centre in Svalbard (UNIS) has been operating in Longyearbyen, Svalbard since 1993. Longyearbyen is a small town located on the west side of Spitsbergen, a part of the Svalbard archipelago in the high arctic at 78 degrees north. UNIS educates more than 800 students and supports close to one hundred research projects on an annual basis. In total UNIS has close to 12 000 field days per year. The education and research is field based and the season lasts from January to December.

Operations in the high arctic prove to be challenging. Challenges encountered include, but are not limited to: lack of infrastructure, harsh and variable weather, darkness, and rapidly changing natural hazards. These are conditions that have to be handled from day to day to ensure safe operations for students and scientists.

In the last five years natural hazards have been changing at such an increased rate (e.g. avalanche danger, melting sea ice, high levels of precipitation, rapid fluctuations in air temperature etc.). For UNIS this implies that established operational procedures based from many years of experience are no longer valid. The practitioners responsible for planning and guiding students and scientists in the field often experience that the plan deviates from the performance. They have to choose deviation "constantly" to maintain a safe performance of the activity. A challenge is thus to have a system or process that secures feedback to the decision makers.

To keep up with the changing risk picture, experienced feedback and tacit knowledge are getting more and more important to maintain safety management and safe operations.

The purpose of the paper is to exemplify how experienced feedback and tactic knowledge are used to manage safety for operations in the high arctic and discuss how experience feedback can ensure organizational learning for field operations.

2 EXPERIENCE FEEDBACK

Safety management is based on the principle of experience feedback, i.e. the process by which information about the results of an activity is fed

back to decision makers as new input to modify and improve subsequent activities (Kjellén and Albrechtsen, 2017). Kamsu Foguem et al. (2008) have a similar interpretation: experience feedback is a process whereby experience at an operational, tactical or strategic level is disseminated in such a way that the knowledge is used to improve the organization's performance.

The purpose is to use information about experienced or expected safety performance as a basis for decisions that prevent accidents and reduce accident risk.

Experience feedback is based on principles from quality management such as Juran (1989) persistent feedback control and Deming's (1993) cycle. Kjellén and Albrechtsen (2017) present a safety information system based on principles of experience feedback that consist of collection of data about experienced and expected safety performance; analysis and storage of the data; distribution of analyzed data to decision-makers; and decision-making and implantation of safety measures. This system facilitates systematically improvement of safety based on experiences (incidents, non-conformities, observations, etc.); identification of current performance (inspections, audits); and excepted safety performance and challenges (risk assessments).

Another important principle of experience feedback is organizational learning and knowledge sharing. The process of organizational learning involves an organizational unit changing itself or its knowledge base as a result of experience (Cyert & March, 1963). The unit can learn directly from its own experiences, or from the experiences of other units (Levitt & March, 1988). Argyris and Schön (1996) claim that organizations learn only if the product is a change in behavior and governing variables in the organization.

Although the literature differs between individual and organizational learning, there is a clear relationship between the two (Crossan et. al, 1999). Nonaka and Takeuchi (1995) demonstrates how transitions of tacit and explicit knowledge lead to organizational knowledge (Figure 1). Through these processes, knowledge is converted from individual knowledge to shared knowledge that can be utilised by the whole organisation. The transitions are continuous processes that lead to a learning spiral. Nonaka & Takeuchi (1995) propose four basic processes whereby knowledge is converted:

- Externalisation takes place when tacit knowledge is made explicit, for example when an unwanted occurrence is observed and reported by a worker at the sharp end.
- Combination takes place when explicit knowledge is combined with other explicit knowledge, for example when a reported unwanted occurrence is compared with other reported unwanted occurrences in an effort to identify similarities.
- Internalisation takes place when explicit knowledge becomes tacit knowledge. The point is to see the importance of making practical use of knowledge through converting the explicit to practical, effective and correct actions.
- Socialisation takes place when tacit knowledge is spread as tacit knowledge to other members of the organisation, who learn over time through seeing what others do.

Principles of organizational learning implies that safety management based by experience feedback is dependent on both formal and informal processes of knowledge sharing among practitioners, safety staff and managers.

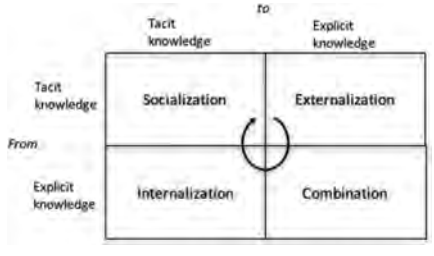


Figure 1. Knowledge creation in organizations (Nonaka and Takeuchi (1995).

3 SAFETY CHALLENGES IN FIELD OPERATIONS AT THE UNIVERSITY CENTRE IN SVALBARD

Every person going through the UNIS system expects access to teaching, learning or research in the field. Safety technicians at UNIS have the responsibility to assist in the planning and execution of every type of field work. This includes the safe transport of groups to their desired field locations, and then further technical assistance and general safety at the field site. Safety technicians at UNIS benefit from the experience of years of living and working in the Arctic. This includes hundreds of hours of work in the field on an annual basis.

At UNIS there are two distinct field seasons. Winter field season normally runs from January-May, while summer field season extends from June-October. November and December are typically slower, due to lack of snow and light.

Winter field season is characterized by snow mobile travel. Other forms of transportation may include travel by beltwagon or larger sea going vessels. Typical hazards encountered during winter season include: snow mobile driving, avalanche

terrain, sea ice, glaciers, harsh weather conditions, and polar bear encounters.

Field operations in the summer season is mainly done by boat travel. A variety of vessels are used including: zodiacs, polar circles, tourist boats and large cruise vessels. Typical hazards encountered during summer season include: harsh weather, rough ocean conditions, camp challenges, polar bear encounters, and hazards encountered when travelling in steep mountainous terrain.

The top priority in both winter and summer field seasons for the technicians is to ensure the group has safe transportation from Longyearbyen to their field location. In winter season this is controlling a group of 10–30 students and teachers on snow mobiles as they drive through variable terrain to their destination. In summer this is transporting up to 12 students and teachers at a time by polar circle to their destination.

Ideally the technicians should only operate under favorable conditions, where there are few hazards and thus low risk. Due to the environment and related hazards and challenges, all activities involve some kind of risk, and the plan often deviates from what is expected.

Two examples will be presented which illustrate the challenges associated with field travel in both winter and summer

3.1 *Case 1: Boat travel, summer season*

The objective of the field travel was to drop off one group to Colesbukta, drop off camp equipment and perform water sampling in Dicksonfjorden, see map in Figure 2. The two trips were originally scheduled to take place on different days, but this had to be changed when an attempt to drop off the first group in Colesbukta the day prior was unsuccessful. Technician A had taken three scientists, plus a boat full of field equipment towards Colesbukta the previous afternoon. The group had traveled for approximately one hour in bad conditions, heading straight into the oncoming waves which were crashing over the bow of the boat at a height of over two meters. The technician decided to turn around for several reasons:

- If the technician maintained the same speed, an hour long round trip would have taken two to four hours in conditions which were not supposed to improve
- The wind direction was not favourable for landing in the desired location
- The scientists had a lot of heavy equipment which would take a long time to unload
- The desired drop off point was extremely shallow, with lots of known objects in the water. With heavy waves and wind, chances of either grounding the boat, or hitting the propeller/engine on an object were increased
- The boat was heavy and had a broken anchor winch, making it not ideal for shore landings, especially with the given conditions

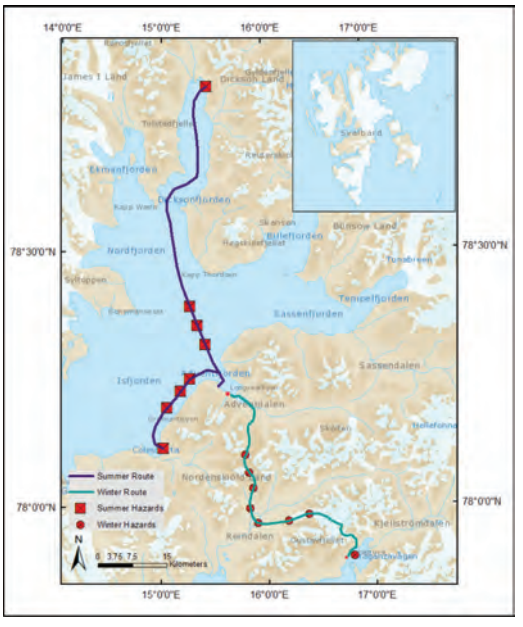


Figure 2. Map showing the routes described in the following tasks. Summer hazards are where shallows and rough seas are normally encountered. Winter hazards are where avalanche, glacier, sea ice and open water hazards are normally encountered. Basemaps © Norwegian Polar Institute.

If the technician only encountered one of the before-mentioned factors, the trip would probably have been completed. A combination of all of the factors, which the technician was able to identify during the trip, forced the decision to turn around. There was no protocol for this, but due to past experiences the technician was able to determine that in this particular situation, the risk was not worth the potential consequences.

The next day the winds and sea had calmed making it realistic to complete both trips. Due to a combination of factors it was decided that two technicians and two boats were needed to make this a successful operation. A request was made by Technician A to Technician B for assistance. Factors which were considered for this trip included:

- Colesbukta and Dicksonfjorden are in opposite directions, so if only one boat went, the trip would take a significantly longer amount of time
- The drop-off spot for the camp in Dicksonfjorden is dependent on tides. A lot of gear needed

to be unloaded which is time sensitive in order to not ground the boat
– The boat going first to Dicksonfjorden needed to pick up scientists from the camp for sampling
– Wind was coming from east, making the travel to Dicksonfjorden fine, but more challenging on the way back to Longyearbyen when heading directly into the wind

The plan was then executed safely and successfully. The two boats left at the same time, with Boat B heading first to Colesbukta to drop off scientists and gear, and then bring the rest of the gear to Dicksonfjorden to meet Boat A in time to drop off equipment before the tide started falling. Boat A went straight to Dicksonfjorden and was able to complete the sampling and drop off the other equipment. Boat A and Boat B were then able to drive back to Longyearbyen from Dicksonfjorden together, which was optimal because the winds began to pick up creating unfavorable conditions for driving alone. Boat B has a covered cabin, making it more favorable for driving in big waves. Boat A could then drive in the wake of Boat B, as to not get as many waves into the boat on the way back to Longyearbyen.

This task ended successfully for several reasons. The two technicians had combined experience from driving boats in Isfjorden. They both knew the challenges with landing in the two areas, and were able to understand the implications that the weather conditions and tides would have on the locations they needed to get to and tasks they needed to complete. Flexibility played a huge part in that both technicians were willing to change plans when it proved necessary in order to complete a safe and successful trip. The challenges and potential hazards encountered during this scenario are not uncommon or unknown. Shallows, tides, waves, wind, boat problems, etc., are all encountered by boat drivers in Svalbard. The challenges can be anticipated, but only the knowledge one needs to be equipped to deal with them, and to operate around them are only gained through experience.

Figure 3. Example of vessel used during field trips, UNIS Polaris beached.

3.2 *Case 2: Snow mobile travel, winter season*

The objective of the field travel was to transport a group of students and professors from Longyearbyen to Svea (a small mining settlement located 60 km away from Longyearbyen) in early February by snow mobiles. Travel between Longyearbyen and Svea is common during the winter and spring months by snow mobile. Transporting a group of up to 25 students and professors to do work on the sea ice close to Svea is normally not challenging. However, in recent years, Svalbard has experienced more precipitation, higher temperatures and less sea ice. This leads to more challenging conditions for winter field work. The technicians anticipated the following risk factors for this particular task:

– 25 people with little to no snow mobile experience must drive over 60 km through challenging conditions
– Driving through avalanche terrain
– Driving in places with open water
– Driving over terrain which is icy and rocky
– Driving over glaciers with crevasses
– Working on sea ice

Therefore, special precautions had to be taken, and here it is the job of the technicians to use their knowledge and expertise to ensure the group can travel and work safely in Svea. One should always expect and be prepared to encounter hazards doing this kind of work, but in this particular situation there were several risk factors which needed special attention from the technicians to ensure safe and successful travel:

– The season up to this point had brought unusual conditions which included: a lot of snow, followed by heavy rain and temperatures up to +7°C. Temperatures then quickly fell to well below 0°C.
– The normal route to Svea includes traveling through narrow valleys surrounded by steep mountains which leads to both avalanche hazards and water hazards
– The route also includes travel across a wide open valley which is vulnerable to wind which can blow all of the snow away leading to icy and rocky conditions
– The route includes a glacier crossing which creates a potential crevasse hazard
– Due to the decreased activity in Svea, the route which is normally well maintained is much more unknown and unpredictable

– Sea ice which is normally forms close to Svea is affected by warmer air and sea temperatures. Extra precaution must be taken when working on this ice.

Instead of the normal procedure of sending out one technician to follow the group as would be necessary, several scouting trips were undertaken in order to identify all of the possible hazards the group might encounter, and to confirm that safe travel was possible. Three separate scouting trips were completed until the technicians were satisfied that they were comfortable sending the group through the terrain and had identified all possible hazards and deemed them manageable for the group. The technicians identified open water, avalanche conditions, blue ice and rocks. They were able to deviate slightly from the normal route, in order to find a route which was as safe as possible for the students and staff.

On the day when the students and staff were supposed to travel to Svea, two technicians joined to ensure safe travel. The trip was completed successfully and the students and staff were able to do their work in Svea.

4 DISCUSSION

4.1 *Adaptation and flexibility*

Both field trip examples described in the previous section were successful in terms of safety because the technicians were able to anticipate the risks of the trips and thus adapt to the situation of the trips. Rankin et al. (2014) present a framework for understanding coping mechanism (adaptions) to respond to variations in a dynamic environment, see Figure 4. Adaptions are a function of 1) objectives, i.e. the outcome that the adaption aims at achieving and is related to identifying demands, pressures and conflicting goals; 2) the context in which the adaption is carried out; and 3) necessary resources and conditions for successful implementations of the coping mechanism, including both "hard" and "soft" conditions such as availability of knowledge. The adaption in itself consists of 1) the four cornerstones of resilience (Hollnagel et al., 2007) anticipating, monitoring, responding and learning; and 2) interactions between sharp-end and blunt-end.

Figure 4. Snow mobile travel to Svea.

The successful adaptions among the technicians at the field trips can be described in such a framework. Their adaption is a function of the context of the action that consist of their ability to monitor and anticipate the situation. The technicians' tacit knowledge is a key contributor to their ability to adapt to the situation, not least because the success of the actions depend on the decisions made at sharp end due to lack of communication infrastructure with the blunt end.

The tacit knowledge among the technicians and their ability to adapt to the situation are essential in both cases for maintaining safe travels in continuously changing variable conditions. Scenario 1 has several elements which made it more difficult than a normal to drop off or pick up in the field. Tacit knowledge from the technicians which is acquired through multiple seasons and hundreds of hours of driving boats around the Isfjorden area was vital to complete this task in a safe and effective way. The snow mobile trip in scenario 2 was successful for many of the same reasons that the boat trip was successful. The technicians were able to use their past experience and knowledge to anticipate the hazards and then act accordingly. Flexibility plays an important role. The technicians were able to put in much more work than is normal for this kind of trip. When many different hazards exist and combinations of hazards are not always predictable, tacit knowledge and experience are essential.

The two scenarios also show the connection between the sharp end—the practitioner, and the blunt end—the management. For both scenarios a key word is flexibility. Based on the experience in the sharp end from similar operations, the management can see the need for flexibility and use of extra time to adapt to the situation. Experience in the sharp end is building situational understanding in the blunt end.

4.2 *Experience feedback at an organizational level to improve coping mechanisms*

The available knowledge among the technicians at sharp-end during field trips could be improved by principles of experience feedback as illustrated in Figure 5. The framework by Rankin et al. (2014) is rooted in principles of resilience engineering.

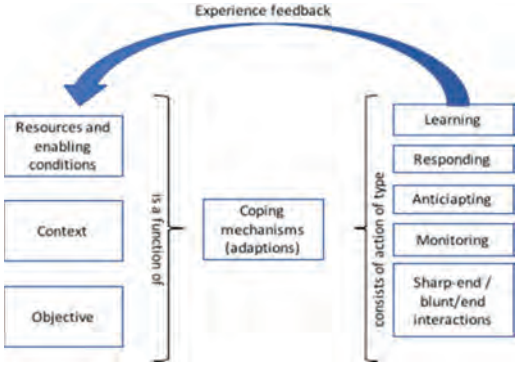


Figure 5. Framework for analysis of coping mechanisms (based on Rankin et al. 2014) and the importance of experience feedback.

Resilience engineering is centred around four main abilities: to respond, to monitor, to anticipate and to learn. The ability to learn from both failures and success is a key to improving the knowledge among sharp-end practitioners. By learning from experiences at field trips both individually and among colleagues, resources that contribute to adaptions will improve.

The technicians who participate in fieldwork with students and employees acquire new experience every day and thus maintain and develop their tacit knowledge. In addition, their tacit knowledge is generated when the they continuously close nonconformities during field trips as shown in the two cases in the prior chapter.

The individual learning is important learning for the organisation. Feedback from experiences related to everyday tasks should be shared in the organization; what works and what does not work in order.

Nonaka and Takeuchi (1995) emphasize how different transitions of tacit and explicit knowledge create shared knowledge in organizations. Socialization (from tacit to tacit knowledge) is one of the transitions that contributes to organisational learning. The technicians at UNIS start each morning with a meeting to go through the tasks/duties to be performed on that day. Ekman (2012) highlights the importance of informal conversations in making tacit knowledge visible in the organisation, and facilitating arenas that encourage small talk. The morning meetings involve a set agenda where events from the day before is discussed and the technicians inform each other of changes related to snow conditions, weather, etc. The technicians generally meet up for a cup of coffee together before this meeting. A lot of information is shared during this five-minute period that should be raised during the formal meeting. As a result, important information is not raised at the meeting with the management because it has already been shared during the small talk over coffee before the meeting. Tacit knowledge among those talking is improved, but there is potential for organizational learning in addition if this knowledge is share among more people.

Transition from tacit knowledge to explicit knowledge (externalization) will also contribute to organizational learning (Nonaka and Takeuchi, 1995). Within safety management, systems for reporting of unwanted occurrences is an important contribute to externalization of tacit knowledge, but also to combination of explicit knowledge as well as internalization (from explicit knowledge to tacit knowledge) (Kjellén and Albrechtsen, 2017).

Such learning among technicians wil happen in communities of practise (Wenger, 1998). Communities of practice is a group of humans that has a mutual engagement, common goals and activities and a common repertoire of actions and resources (Wenger, 1998). Among the primary parts of learning in communities of practice we find social participation, sharing stories, apprenticeship learning, and that learning is a complex social phenomenon dependent on context.

How can one use and systemise the informal "coffee break" to strengthen learning in the organisation? Ekman (2012) refers to the importance of horizontal meeting places for the tacit knowledge where every day experiences can be shared. It is also possible to learn from conversations about a completely normal day. It provides an opportunity to test out the prevailing knowledge and create new learning. A traditional view in the field of safety is that one learns from mistakes and incidents. However, in more recent times, it has become more common to focus on learning from successful tasks (Hollnagel, 2014), which after all is most of the tasks one performs during a working day. By learning from everyday events, it is possible to test the prevailing knowledge and, in doing so, uncover practices that are unsafe, even though no accidents have occurred. "Learning from successful operations is not only about identifying and promoting good practice, it is also about detecting the instances where no accident occurred in split of unsafe practices or unsafe systems" (Rosness et al., 2016).

4.3 *Contextual change that affect experience feedback*

UNIS experience increased student production and rapid changes in the natural environment. Ashby´s (1961) law of requite variety states that control of a system is achieved only when the variety of countermeasures matches the variety and changes of

the system. This implies that the field technicians must acquire new knowledge and put this into effect to deal with the contextual changes of their field activities. Systems and practises for experience feedback would enable improved knowledge to handle new situations.

Growth is not only a matter of increasing staffing to deal with the increased activity. One must also make structural changes to ensure that the environment for learning from tacit knowledge and making this visible is as favourable as possible. If one looks at organisations that manage to exploit tacit knowledge, facilitating communication is a key factor. Structurally, one can facilitate the rapid spreading of knowledge and spend time on systematic training and review the composition of the work group to attain a mentor effect.

5 CONCLUSION

Successful field operations at the University Centre in Svalbard depend heavily on safety technicians that have the responsibility to assist in the planning and execution of every type of field work. Due to changing conditions, local variations, extreme weather conditions, lack of access to infrastructure and communication successful safety performance is created by individual's ability to adapt to situations. This paper has demonstrated that this ability to a large extent is a function of the tacit knowledge of the technicians. To improve the tacit knowledge of each technicians, systems and practices of experience feedback must be run to ensure individual and organizational learning from both failures as well as successes. This is in particular important in systems with great variability in climatic conditions and systems with organizational changes.

REFERENCES

Argyris, C., & Schon, D. A. (1996). Organiational learning II. Addison Wesley.

Ashby, W. R. (1961). An introduction to cybernetics. Chapman & Hall Ltd.

Crossan, M. M., Lane, H. W., & White, R. E. (1999). An organizational learning framework: From intuition to institution. Academy of management review, 24(3), 522–537.

Cyert, R. M., & March, J. G. (1963). A behavioral theory of the firm. Englewood Cliffs, NJ, 2.

Deming, W.E. (1993). The new economics for industry, government and education. MIT Press, Boston, MA.

Ekman, G (2012), Fra prat til resultat—om lederskap i hverdagen, Abstrakt forlag [in Norwegian].

Foguem, B. K., Coudert, T., Béler, C., & Geneste, L. (2008). Knowledge formalization in experience feedback processes: An ontology-based approach. *Computers in Industry*, 59(7), 694–710.

Hollnagel, E. 2014. Safety-I and Safety-II: The past and future of safety management. Ashgate, Aldershot, UK.

Hollnagel, E., Woods, D.D. and Leveson, N.C. (eds.). (2007). Resilience engineering: Concepts and precepts. Ashgate, Aldershot, UK.

Jacobsen, Thorsvik (2016) Hvordan organisasjoner fungerer, 4 utgave, Fagbokforlaget. [In Norwegian].

Juran, J.M. (1989). Juran on leadership for quality—An executive handbook. The Free Press, New York.

Kjellén, U. & Albrechtsen, E., (2017). Prevention of Accidents and Unwanted Occurences. CRC Press.

Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company: How Japanese companies create the dynamics of innovation. Oxford university press.

Rankin, A., Lundberg, J., Woltjer, R, Rollenhagen, C. & Hollnagel, E. (2014) Resilience in Everydat Operations: A Framework for Analyzing Adaptions in High-Risk Work. *Journal of Cognitive Eningeering and Decision-Making*. Vol. 8, No1, pp.78–97.

Rosness, R. Haavik, T. Tinmannsvik, R.K. (2016), What do you do when you build safety? Practitioners' guide to learning from successful operations. Trykkpartner AS.

Wenger, E. (1998). Communities of practice: Learning, meaning, and identity. Cambridge university press.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Automation of the rail—removing the human factor?

T.M. Stene
SINTEF, Trondheim, Norway

ABSTRACT: Automated vehicles will be increasingly used as transport in the future. However, it is unclear if this imply full autonomy or different levels of automation. A unified definition of autonomy in transport is missing. The SAREPTA project (Safety, autonomy, remote control and operations of industrial transport systems) is established in 2017, and cover safety challenges of future intelligent transport systems that are autonomous, remotely controlled and normally not manned. The project covers both road, sea, aviation and rail. This paper focuses on issues related to rail transport, including both metros and railway. The purpose of the paper is to describe current rail accidents as a basis for questioning whether future digitalisation will improve safety. The paper will discuss the autonomy concept in relation to grades of automation. Relevant questions are: What is automation and which accidents may be prevented by automation? To what degree do automation and remote control imply removal of the Human Factor? And from a safety perspective—What is the safety potential of future automation, and how can humans contribute to safety in future intelligent transport systems?

1 INTRODUCTION

Digitalization is a global change affecting a variety of social conditions and businesses. In addition to changing products and services in businesses and the labour market, digitalization will also create radically new business models in many industries (Stene et al 2017).

1.1 *Safety and automation of transport systems*

Safety and environmental challenges of future intelligent transport systems are addressed in a newly established project founded by the Norwegian Research Council for 2017–2021. The SAREPTA (Safety, autonomy, remote control and operations of industrial transport systems) project focuses on systems that are autonomous, remotely controlled and/or periodically not manned.

In the project, four thematic areas of autonomous systems are central: (1) Risk identification and risk levels, (2) Infrastructure vulnerabilities and threats, (3) Technical, human and operational barriers to mitigate system risks, and (4) Organizational and human factors, and regulatory measures. The project includes road, sea, aviation and rail. This paper focuses on the rail. The purpose of the paper is to describe current rail accidents as a basis for questioning whether future digitalisation will improve safety. Relevant questions are: What is automation and which accidents may be prevented by automation? To what degree do automation and remote control imply removal of the Human Factor? And from a safety perspective—What is the safety potential of future automation, and how can humans contribute to safety in future intelligent transport systems?

1.2 *Current rail transport safety—fatal and frequent accidents*

European railways are the safest mode of land transport and the safety level has improved over the last decades (EU ERA (European Railways Agency) 2016). However, accidents have heavy impact on confidence in the system. Further, every accident represents a significant business cost in a highly competitive environment. It is argued that emphasis needs to be on human factors as well as on new technology which can be both an opportunity and a threat.

Compared to other transport modes, the fatality risk for an average train passenger (0.12 per billion km) is at least twice as high as commercial aircraft passengers (EU ERA 2017). However, the risk is higher for passengers traveling by bus/coach (one third of the risk) and sea vessels (nearly three times as high). Further, using individual transport means on the road is most risky. Car occupants have at least 20 times higher likelihood of dying compared to train passengers.

Even if rail transport statistically is safer than road transport, some large rail accidents have occurred. The rates of fatal train accident (five or more killed: totally 362) have fallen substantially from 1980 to 2009 on Europe's main line railways (Evans 2011). Fatality risks per million train-km (system risk) in the period 2010–2014, based on persons involved,

was 0.28 killed per billion train-km at the EU level (EU ERA 2016). For rail passengers, this was 0.14 killed passengers per billion train-km.

Although rail transport safety has steadily enhanced over the years, the number of accidents started increasing in 2014 and 2015 (Eurostat 2017). Still, the number of victims (killed or injured persons) continues to decline. Table 1 shows the number and persons killed and injured in rail transport accidents in Europe 2016. Two types of accidents are dominant – (1) Rolling stock in motion and (2) Level-crossings—followed by (3) Train collisions and (4) Derailments.

The majority are accidents to persons caused by rolling stock in motion. These are either hit by a railway vehicle or an object attached to it. Persons that fall from railway vehicles are included, as well as persons that fall or are hit by loose objects when travelling on-board vehicles.

Fatal level crossing accidents are more numerous and account for more fatalities than fatal train collisions and derailments (EU ERA 2016). Further, in contrast to collisions and derailments, the rate per train-kilometre remained unchanged in 1990–2009. Thus, level crossing accidents represent an increasing proportion of serious accidents.

The estimated accident rate in 2016 is 1.07 fatal collisions or derailments per billion train-kilometres, which represents a fall of 73% since 1990 (Evans 2011). This gives an estimated mean number of fatal accidents in Europe in 2016 of 4.7. In contrast to fatal train collisions and derailments, the rate per train-kilometre of severe accidents at level crossings fell only slowly and not statistically significantly in 1990–2016. There are statistically significant differences in the fatal train accident rates and trends between the different European countries.

Totally, the most common cause of fatal accidents is signal passed at danger, followed by signalling/dispatching errors and violation of the speed limit. Further, small numbers are train fires and groups of persons struck by trains, mostly track workers.

Table 1. Number and persons killed and injured in rail transport accidents by type of accident in Europe 2016 (Eurostat 2017).

	Number of persons		
Type of accident	Killed	Seriously injured	Total
Collisions	44	77	121
Derailments	11	27	38
Accidents involving Level-crossings	256	220	476
Accidents to persons caused by rolling stock in motion	651	438	1089
Others	2	16	18
Total	964	778	1742

The causes of level crossing accidents differ from train collisions and derailments. The most frequent cause of fatal train collisions (2) and derailments (3) is signals passed at danger. The majority of level crossing (1) accidents are caused by errors or violations by road users. Most major crossings in Europe have automatic warnings (lights, barriers and bells) operated by approaching trains. Most minor crossings have fixed warning signs only, with no indication when trains are approaching. The primary responsibility for operational safety thus rests with road users, either in obeying warnings or checking that no train is approaching before they cross.

1.3 *Animals along the track—a current challenge*

Less severe accidents and incidents strongly outnumber fatal accidents (EU ERA 2016). However, these occurrences are not collected at the EU level, and great benefits could be made from reporting them to identify and manage risks.

While the number of people killed or injured in rail accidents is well-documented, little research has been done to analyse the number of animal casualties on international railways (Gray 2015). High-speed trains often cut through sensitive wildlife habitats. Accidents involving various species are detrimental to local wildlife, are costly and a danger to travellers.

In Norway, nearly 2000 collisions with animal are recorded on the railway each year, which is a doubling of the frequency over 20 years (Roaldsen et al. 2015). Reduction of crashes—even by a few percent—can contribute to significant socio-economic savings and reduced conditions for both humans and animals.

From 1991–2014, the Norwegian National Rail Administration registered nearly 26 000 events with one or more animals (near 36 000 animals) being hit by train. Over 90 percent involve moose (57%), roedeer (15%), sheep (9%) and domesticated reindeer (8%). Topography and landscape influence the existence of animals in areas near the rail, thus increasing the accident risk. Important factors are related to food, shelter, visibility and animal corridors. Further, weather conditions as snow and rain affect where the animals are.

2 TRANSPORT TECHNOLOGY INNOVATION

2.1 *Digitalization of the rail*

Digital technology may be defined as the use of ITC (computing capacity + telecommunication)

to gather, transfer and process data to provide the communication backbone for all users of the network (BearingPoint 2017).

Rail 4.0 may be considered a parallel concept to Industry 4.0 (Stene et al 2017). The concept refers to four industrial revolutions starting at the end of 18th century with the introduction of (1) mechanical manufacturing, and continues with (2) mass production, (3) computers and automation (also labelled digital revolution) and (4) Internet. Four key components in Industry 4.0 are: CPS (Cyber-Physical Systems), IoT (Internet of Things), Smart Factory (e.g. traffic management sites) and IoS (Internet of Services).

Further, Davidsson et al (2016) divide the digital period in four waves: (1) introduction of computers in the 80s, (2) Internet in the 90s made it easy to access and share information, (3) mobile Internet making this possible regardless of where you are, and (4) is represented by Internet of Things (IoT). In addition to people, different types of entities (vehicles, machinery) may also have access to and share information.

In the rail sector, ERTMS (European Railway Traffic Management System) is a common signalling system that is to be introduced in all EU countries by 2030. A standardized system will improve the interoperability between networks and systems. ERTMS includes ETCS (European Train Control System), GSM-R (Global System for Mobile Communication-Railway, which is radio communication between train and signalling), and common European traffic regulation. A common trans-border railway transport allows trains to travel in any European country which has the ERTMS system implemented both in the rail infrastructure and in the train itself.

ERTMS has many similarities with CBTC (Communication-Based Train Control), which is the preferred signalling solution for automated subways and metros. One difference is that ERTMS is standardized, while CBTC is supplier specific. CBTC is a signalling system making use of telecommunication between train and track equipment (wayside) for traffic management. By making more exact positions of each train, the system makes it possible reduce time intervals between trains. The main objective is increased capacity.

2.2 *Automatic Train Operation (ATO)*

Generally, autonomy is often related to attributes like self-government, freedom to act or function independently. For vehicles, autonomy is generally understood as the ability to make decisions about actions to take, e.g. course or speed, independent of a human operator. Levels of autonomy or automation describe the successive shifting of responsibility from the driver to the vehicle. Different concepts are used to describe vehicle automation in each transport mode/ domain.

In addition to concepts used in each domain, Ponsard et al (2017) present a comparative over-

Table 2. Comparison of automation levels at road, rail and air. Based on Ponsard et al (2017).

Railway	Road	Aircraft	Resp.	
Grades of automation	SAE levels	Levels of automation		
GoA-0 Sight train operator	**L0** No automation	**Level 1** Raw data, no automation at all	All time	Warn Protect
GoA-1 Manual train operation Automated train protection	**L1** Driver assistance Park assist/cruise control	**Level 2** Assistance Flight director Auto-throttle	Drivers	Guide Assist
GoA-2 Semi-automated train operation (STO). Autom. train op. (ATO)	**L2** Partial automation Traffic jam assist	**Level 3** Tactical use Autopilot	Monitors all time	Manage movements within limits
GoA-3 Driverless train operation (DTO) Automated control (ATC) Some control by attendant (operating doors, emergencies)	**L3** Conditional automation	**Level 4** Strategic Flight management system	Ready to take back control	Drives itself, may give back control
	L4 High automation Highway traffic jam system	Uninterrupted autopilot project (Boing) Drones (unmanned)	May not take back control	Drives itself with graceful degradation
GoA-4 Unattended train op (UTO) Automated doors Platform screen doors	**L5** Full automation (all situations)		Not required	All time

view of the responsibility between system vs human (driver/pilot) at different levels of automation (see Table 1). In rail, the concept Grades of Automation (GoA) is used. Notice the double line in the table; this marks a shift from GoA-3 in responsibility from the driver to the system.

Rail and airplanes have already achieved much higher levels (Ibid). However, this is only true for some rail line types. Several fully autonomous metros exist. The next two sections in this paper goes more into this.

2.3 *New technology on the main line railway*

The difference between signalling and control systems in European railway is significant, and until 1980 14 national standards were in practical use (Tao & Jing 2014). ETCS (European Train Control System) is designed to replace these incompatible safety systems, and the first version was published in 2000.

As mentioned above, the GoA concept describe levels of automation in rail. Figure 1 illustrates the existence of a driver at different grades. Further, the operations are described at each grade, i.e. management agents and actions to be taken.

Implementation of ERTMS at GoA-1 implies that signal information is shown on a panel inside the cabin. The driver may use the signal as a replacement of a traditional light outside at the track. The signal tells whether the driver may drive into the next block or not. At GoA-2 the train is operated by automated control based on signals from sensors along the track. In addition to be responsible for monitoring the speed and position, the driver may take control in case of any incident or emergency.

A lot of literature on transport autonomy focus on train automation, i.e. the interaction and responsibility between vehicle—driver (see Figure 2). The inner control loop is responsible for executing the production plan (Rao & Montigel 2017), and the focus is on driving performance by

Figure 1. Levels of automation (Brodeo 2016).

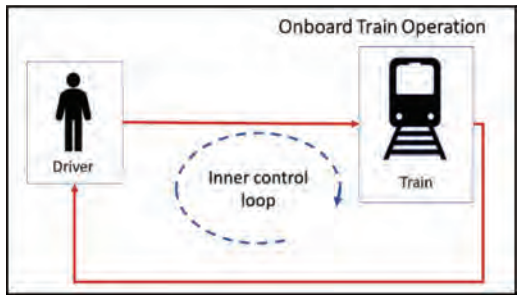


Figure 2. Train automation—Control of onboard train operation.

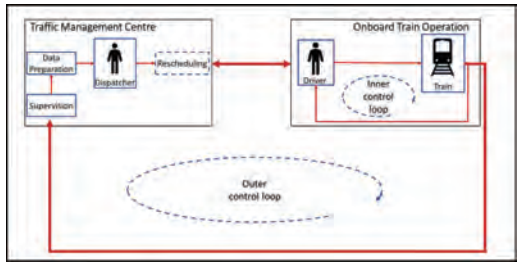


Figure 3. Traffic management—Control of traffic and infrastructure. (Based on Rao & Montigel, 2017).

providing driver assistance or introducing train automation.

Rao (2015) presents a holistic approach to the main line railway. In addition to (1) train automation, the focus is also on (2) traffic management, and the relationship between the two areas (see Figure 3). The outer control loop supervises the status of traffic and infrastructure, detects deviations and conflicts, and develops a new schedule (rescheduling) and transmits it to train operation.

Automation depends on two supports: Onboard support (as the Automatic Train Protection—ATP) system to provide train's overspeed protection and to keep a safe headway between trains, and infrastructure support (as Automatic Train Supervision—ATS) to provide dynamic traffic regulation to avoid traffic conflicts (Rao et al 2016).

Even at GoA-4 trains on are not autonomous in the sense that no control is needs. Traffic management focus both on the outer control loop (improving efficiency for the dispatcher by providing resolutions for traffic conflict) and the inner loop (improving driver performance or assisting the driver). Thus, reducing human failure are central in both control loops.

ETCS (European Train Control System) is a signalling, control and train protection system used on the main railway lines. The train detection

equipment sends the position about speed limitation, signal status etc. (Venticinque et al 2014). Three levels define the use of train control system; communication from track to train (level 1), continuous communication between the train and the Traffic Management Centre (level 2), and future implementation of a moving block technology (level 3). Several main rail tracks operate at level 2, including two main subsystems: (a) a ground system collects and transmits track data to (b) an onboard subsystem.

ETCS-2 uses digital radio transmission of signals along the trackside (Tao & Jing 2014). With its onboard positioning equipment, the train can automatically report its exact position and direction of travel at regular intervals, in addition to motion (stop/go) signals. Balises on the tack detect trains and send the position to the control centre (Venticinque et al 2014). Based on the position of all trains, the centre determines the new movement authority (MA) and sends it to the train. The onboard computer calculates its speed profile from the MA and the next braking point. This information is displayed to the driver.

2.4 *Autonomous metros*

In metro systems, automation refers to the process by which responsibility for operation management of trains is transferred from the driver to the train control system (UITP 2017).

The experience period with automated metros is over 30 years. The first was high capacity, but today we also see a trend of increase in mid-capacity trains. Between 2014 and 2015 Europe will lead in terms of growth (Hernández 2014). Asia and Europe together hold 75% of the km of fully automated metro lines.

For metros, many use the term CBTC synonym as an automated driverless system. However, at its most basic form the system provides automatic protection (ATP) only. Fully automated systems also include ATO (Automatic Train Operation) and ATS (Automatic Train Supervision).

A semi-autonomous train (GoA-2) may manage movements, but a human need to be onboard to start the train, open doors etc. (Lufkin 2015). There are also trains that can fully operate completely free of humans. Only 6% of the world's transit rails operate those trains. Several cities are aiming for automation.

There are 55 fully automated metro lines in 37 cities around the world (UITP 2016a). Fully automated metro lines, defined as those metro lines in which trains can be operated without staff onboard—a defining characteristic is the absence of a driver's cabin on the train. This type of operation is also known as Unattended Train Operation (UTO), or Grade of Automation 4 in standard IEC 62267.

2.5 *Metro automation and safety*

The positive experience of decades of automated operation highlights one of the major elements to consider in this success story: safety (UITP 2016b). There have been no significant accidents, in particular none involving casualties, in any automated metro line in the world.

Copenhagen Metro is one example of a system running fully automated, consisting of automatic train protection, operation and supervision. Although no serious accidents have occurred, incidents and accidents may point out some risk areas. The station area is strongly marked. The safety of the platform/track interface is crucial for fully automated metro lines.

The dominant safety measure is installation of platform screen doors (detection systems) preventing persons and objects from falling on the track. Currently, near 80% of stations in fully automated metro lines in operation in the world are equipped with such doors (UITP 2016).

Platform and track incidents aside, there has only been one operational incidents with UTO systems; in Osaka at the end of the 80s a train did not stop at terminus and hit a bumper stop, provoking injuries in a few dozen passengers (UITP 2017).

2.6 *Open surroundings—challenging the main railway*

Since the main railway has much more complicated infrastructure situations, currently train automation is mainly applied in metro railway (Rao et al 2016).

The open surroundings of current main rail traffic challenge safety. Rails with driverless trains are generally run on closed off networks, i.e. run underground. Thus, no one can fall onto the tracks, and there are no points where the trains cross with others.

3 DISCUSSION

3.1 *Rail 4.0 – Opportunities and challenges?*

The purpose of intelligent systems is to make the human environment more "people-friendly" technologies (Tokody & Flammini 2017). This means that infrastructural systems should be sustainable, safe, economic and easy-to-use. The development of intelligent, autonomous systems may ensure sustainability and safety.

Future IoS (Internet of Services) in a rail context will focus on offering services to the general public or specific target groups as passengers. For example, a dynamic system for Copenhagen metro, will automatically optimize trains frequency

depending on numbers passenger and changes of numbers (Razeto & Corsanego 2017). Likewise, in Switzerland, a new Trip Planner app using voice control will let customers compare, combine and book a journey with multiple modes of transport including taxi (SWI 2017b).

Integrated mobility is an example of Smart Management. According to the Federal Railways in Switzerland, integrated mobility is a central field of innovation, and thus they are developing a door-to-door service to the general public ("SBB Green Class").

One example of utilizing IoT, is goods transport in Switzerland installing various sensors in carriages. Instruments will measure temperature, vibrations and the wagon's position. Customers may get information of goods status, location and time for arrival. In Japan high-speed rail use in-ground sensors in quake-prone zones, that immediately activate emergency brakes seconds after the initial quake waves are detected.

However, one of the future challenges is related to telecommunication and traffic management. ITS includes telematics and all types of communications in vehicles, between vehicles and between vehicles and a fixed location (Brodeo, 2016). As even more transport is being digitalized, the use of radio frequencies for signalling systems may be conflicting or overloaded. Several EU countries already use radio communication systems in the same range, all on a limited duration licensing scheme.

3.2 *Scenarios – Can automation prevent future rail accidents?*

For more than three decades, rail transport safety has improved generally and presumably due to a wide range of safety measures like automatic train protection, improved signalling systems and improved operational management. The question is whether new technology may contribute to prevent the most serious and frequent accidents; (1) Rolling stock in motion, (2) Level-crossings, (3) Collisions, (4) Derailments and (5) Animals along the track.

1. The engine (rolling stock) is heavy, and as such needs a long distance to stop in case of an incident or unexpected objects on the track. A driverless train needs to have equipment that detect obstacles and stops automatically. Rail research and innovation in Europe include safety related technology development; automatic obstacle-detection systems for railway vehicles, regenerative braking, monitoring systems and satellite based positioning systems (Tokody & Flammini 2017).

   However, passenger comfort is also highly valued. An efficient and powerful breaking system may cause great discomfort and passenger injuries. This is true for passenger trains, but should be a less problem with freight trains. Even though automated trains may still include some staff onboard.

   Even though capacity is the main objective of CBTC systems used at automated metros, maintaining safety is a major requirement. In addition to distance, calculations cover speed, curves and position. Thus, controlling acceleration, retardation and stops at stations. At slower speed, the distance may be shorter. A challenge is to calculate the block length for max capacity while ensuring safety.
2. Level-crossings. Road user errors or violations contribute to most of fatal accidents, either in obeying warnings or checking that no train is approaching before they cross (EU ERA 2016). The authors point out countermeasures like those for road accidents, particularly education and enforcement. However, more autonomous vehicles may also contribute to prevent rail accidents.

   Autonomous obstacle detection systems may be beneficial for road and rail transport. The Germany SMART project focuses on rail freight and automation of railway cargo haul (Shift2rail 2016), including development of (1) a prototype of an autonomous obstacle detection system and (2) a real-time marshalling yard management system. The first system will use night vision technologies, multi stereo vision system and laser scanner to create fusion system for short (up to 20 m) and long range (up to 1000 m) obstacle detection during day and night operation, as well as during operation in impaired visibility. The second system will provide optimisation of available resources and planning of marshalling operations.
3. Collisions. Related technology development which may contribute to accident prevention are automatic obstacle-detection systems for railway vehicles, traction transformers, energy storage technologies, regenerative braking, monitoring systems, satellite based positioning systems, and smart railway technologies (Tokody & Flammini 2017).

   As mentioned in relation to rolling stock in motion, passenger comfort is highly valued, and unexpected intense breaking may contrast a safety measure. Acceleration and deceleration are essentially limited by the wellbeing and safety of the passengers (Gary 2016).
4. Derailments. One serious accident on a main line using ERTMS, was a derailment of a high-speed train in Spain in 2013. Initial reports cited driver error as the sole cause, but a deeper study of the accident says lack of a functioning onboard ETCS system was a crucial factor (Puente 2015). A high-speed train derailed trav-

elling at 180 km/h (speed limit 80 km/h) through a curve, resulting in the death of 79 people and injuring more than a hundred.

The line was equipped with ERTMS/ETCS Level 1, except for the first and the last kilometre, with a national signalling system used as a backup. However, the onboard ETCS system had been switch off in 2012 due to alleged operating problems. The train driver should manually have changed the speed, but when the train entered the low speed section the driver was speaking on the phone to staff at the train company (Johnsen 2015).

If onboard ETCS had been working, the following would have happened at the ETCS exit boundary 4km before the curve where the accident occurred (Puente 2015): (a) a text message announcing the transition would have appeared on the Driver Machine Interface (DMI) of the train, which was travelling at 200km/h, (b) the DMI would have shown a message with a yellow flashing frame and would have emitted an acoustic signal asking the driver to acknowledge the transition by tapping on the screen, and (c) if the driver failed to acknowledge the message within 5 seconds, service braking would have been applied continuously until the driver had acknowledged the transition or the train had stopped.

5. Animals along the track. Current countermeasures include building fences around the worst affected rail lines, removal of vegetation and warning systems (Roaldsen et al. 2015). The implemented strategies include installation of warning signs for train drivers, night patrols along the tracks and introducing staff to assist animal crossings. Warning signs are the most widespread accident prevention measure (Gray 2015). Most is human warnings, but acoustic signals creating fear in animals (preventing them from approaching the tracks) is also tried. As an example, Norwegian reindeer owners often warn about animals near the rail, implying that train drivers may reduce speed and the probability of incidents (Busengdal et al 2014). More general models have also been developed to predict the occurrence of animals (Gundersen & Andreassen 1998). Gray (2015) argue that manned assistance along high-speed tracks across the world is not a practical solution and better alternatives are needed. Deutsche Bahn Netz AG and OptaSense is one example of testing new warning technology. Distributed Coustic Sensing (DAS) technology uses heat and motion sensors in various areas of operation, including to detect and alert train drivers of animals approaching the tracks.

3.3 *Will automation remove the human factor?*

Automated systems are often designed to relieve humans of tasks that are repetitive. However, the more reliable the system, the more likely is it that humans in charge will "switch off" and lose their concentration, implying greater likelihood of unexpected factors and a potential catastrophe (Vedantam 2009). Technology replacing or assisting the driver can become crutches. Accidents happen when unusual events come together. No matter how clever designers of automated systems might be, they simply cannot account for every possible scenario, which is why it is so dangerous to eliminate human "interference."

The on-board personnel may be unprepared to take control and manually drive. Regular training exercises that require operators to turn off their automated systems and run everything manually are useful in retaining skills and alertness (Ibid). In addition to detect system failure, understanding how automated systems are designed to work also allows operators to recognize when it is on the brink.

As the system cannot cope with all situations, the driver must be ready to resume operations when instructed (Ponsard et al 2017). The author address issues as situational awareness (the system should make sure that driver's decisions are based on right mental pictures), human reaction capabilities (e.g. alarms may cause confusion, defect view of the entire situation, or panic), warning annoyance (trust in the system in case of e.g. frequent/inappropriate alarms) and task inversion (focus on monitoring alarm and lack of attention to real world situations). The authors claim that machine learning techniques can pay an important role for making sure the driver and the system are operating optimally together.

3.4 *How to cope with unexpected scenarios?*

The concept of black swans refers to rare and unpredictable events. Black swans are extremely rare, catastrophic, and unpredictable events that never have been encountered before (Taleb 2007). In principle, black swans cannot be anticipated. However, even though a catastrophe was not predicted, does not mean that the event could not have been prevented (Murphy 2016).

Implementing new technology and autonomous transport, black swans will occasionally occur. We have to prepare both to cope with alternative scenarios and to handle completely unexpected situations accompanied by high stress and emotions. Thus, in addition to training to identify clues of and handling anomaly situation, training should cover completely unexpected and catastrophically events with an extremely high emotional state.

Experiential training may be necessary for coping with unexpected events, especially to handle personal high stress and to communicate with others (Stene et al 2016).

Emergencies are events which happen suddenly and may destroy normal operations. Despite the presence of automated metro operation control system, the emergency management is still heavily dependent upon capabilities of dispatchers at the management centre (Wang & Fang 2014). The system may lose a part of automated safety protection function. Thus, human error behaviours during emergencies cannot be ignored. Competent humans in transport control centres may represent a safety barrier, preventing incidents and accidents (Stene et al 2017). Machines may be excellent in detecting signs and signals, but humans have to evaluate and decide action based on the context and complexity of the actual situation.

4 CONCLUSION

4.1 *Future automated trains and metros*

With more people living in urban areas than ever before, metro systems around the world will need to adapt (Lufkin 2015). The next generation of subways will develop from cities that are already at the cutting-edge, e.g. the super-fast speeds of Japan's shinkansen or the punctual, low-cost driverless trains of Copenhagen.

Self-driving trains are already being used in some countries, with varying degrees of autonomy. Autonomous driving on a complex rail system, with passenger trains and freight trains is more difficult than on a subway—but it is possible (Gary 2016). Several pilots are currently running. On a test field in Germany, trains will be fitted with cameras and other technologies to detect obstacles on the track and stop the train if necessary. The AutoHaul project in Australia, a long-distance railway system is intended to transport iron ore from 15 mines.

Switzerland will test self-driving trains on a main line without too many people, but still get a feel for how it would work in public (SWI 2017a). The trains will be fitted with sensors that should detect objects on the rails and bring the train to a stop. If rolled out, a system to automate train traffic is assumed to increase passenger and freight capacity by 30%.

4.2 *The human factor in future rail systems*

Technology can improve safety, but there may be examples where human interaction is necessary (Gary 2016). The main purpose of implementing a common European railway signalling system are: (1) Maintaining a safe distance between following trains on the same track, (2) Safeguarding the movements at junctions, and (3) Regulating the movements of trains according to the service density and the speed required (Abel, 2010).

The development relies too heavily on old inertia, meaning too much emphasize on technology. More attention should be paid to the organization, the passengers and the infrastructure (Malla 2014) and passenger evacuation procedures (Hernández 2014).

Factors contributing to the likelihood of catastrophic rail accidents are system complexity, a trend towards higher travel speed, growing infrastructure capacity constraints and the constant cost pressures on risk management activities (EU ERA 2017). Accident investigations should continue to report on both success or failure of systemic risk management methods, e.g. high-reliability organisations, redundancy, robust regulatory and enforcement regimes.

Based on experiences from operating both automated and conventional metro lines, one conclusion is that the human factor is that key for the success of an automated line. (UITP 2016b). The rail is far from being autonomous, in the sense of being independent of a human operator. Humans will still be a necessary resource to manage transport and cope with unexpected incidents.

REFERENCES

Abel, S.K. (2010). Eurepean Rail Traffic Management System—An Overview. *Iraq. J. Electrical and Electronic Engineering, 6* (2), pp 172–179.

BearingPoint (2017). *Digitalization in the rail sector*. https://www.bearingpoint.com/files/Digitalization-in-the-rail-sector.pdf&download=0.

Brodeo, G. 2016. Automation in urban public transportation. http://fsr.eui.eu/wp-content/uploads/2016/03/4-Intermodal-Forum_Brodeo.pdf.

Busengdal, A.L., Stanimirov, M. & Brynslund, T. 2014. *Handlingsplan for å redusere antall dyr påkjørt med tog 2014–2017.* Utgave nr. 1. Bane Teknikk miljø og vegetasjonskontroll.

Davidsson, P., Hajinasab, B., Holmgren, J., Jevinger, Å. & Persson, J.A. 2016. The Fourth Wave of Digitalization and Public Transport: Opportunities and Challenges. *Sustainability 2016,* 8 (12), 1248; doi:10.3390/su8121248.

EU (2017). Towards a Strategic Transport Research & Innovation Agenda (STRIA). https://ec.europa.eu/programmes/horizon2020/en/news/towards-strategic-transport-research-innovation-agenda-stria.

EU ERA 2016. *Rail Safety Performance in the European Union 2016.* European Union Agency for Railways/ERA (European Railway Agency). ISBN 978-92-9205-050-4.

Eurostat 2017. *Railway safety statistics.* Eurostat—Statistic explained (http://ec.europa.eu/eurostat/statistics-explained/index.php/Railway_safety_statistics).

Evans, A.W. 2011. Fatal train accidents on Europe's railways: 1980–2009. *Accident Analysis and Prevention 43*, 391–401.

Gary, P. 2016. *Are driverless freight trains safe?* http://www.railway-technology.com/features/featureare-driverless-freight-trains-safe-5008616/.

Gray, E. 2015. *The underdog: preventing animal casualties on railways.* Railway Technology. News views and contacts from the global Railway industry. http://www.railway-technology.com/features/featurethe-underdog-preventing-animal-casualties-on-railways-4532957/.

Gundersen, H. & Andreassen, H.P. 1998. The risk of moose Alces alces collision: A predictive logistic model for moose-train accidents. *Wildlife Biology* 4, pp 103–110.

Hernández, M. 2014. Metro automation: a proven and scalable solution. *Eurotransport* 12 (1), pp 30–32.

Johnsen, S.O. (2015). *HMI analysis of driver's cabin X60B.* SINTEF report F26781 (Restricted).

Lufkin, B. 2015. 5 ideas that could change the future of trains. https://gizmodo.com/5-ideas-that-could-change-the-future-of-trains-1720914816.

Malla, R. 2014. Automation sets a new benchmark. *Metro Report International*, March 2014.

Murphy, J.F. 2016. Surviving the black swan, strategies for process safety specialists, and companies to survive unpredicted catastrophic events. *Process Safety Progress 35* (1), pp 13–17.

Ponsard, C., Massonet, P. & Dallons, G. (2017). Cross-Domain Fertilisation in the Evolution towards Autonomous Vehicles. *ERCIM NEWS 109*, April 2017.

Puente, F (2015). ETCS: a crucial factor in Santiago accident inquiry. *International Railway Journal.* http://www.railjournal.com/index.php/signalling/etcs-a-crucial-factor-in-santiago-accident-inquiry.html.

Rao, X. (2015). *Holistic railway network operation by integration of train automation and traffic management*. ETH Zürich, Doctoral Thesis.

Rao, X. & Montigel. M. (2017). Integration of traffic management and train automation for the main line railway. *IRSE NEWS*, Issue 233.

Rao, X., Montigel. M. & Weidemann, U. (2016). A new rail optimisation model by integration of traffic management and train automation. *Transportation Research Part C71*, pp 382–405.

Rolandsen, C.M., Solberg, E.J., Van Moorter, B. & Strand, O. 2015. *Dyrepåkjørsler på jernbanen i Norge 1991–2014*. – NINA. Rapport 1145. Trondheim, Norway.

Shift2rail 2016. *Smart automation of rail transport*. http://shift2rail.org/projects/smart/.

Stene, T.M., Danielsen, B.-E. & Trevisani, D. 2016. Preparing for human spaceflight to the Moon 2020–2030. In: *Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016.* CRC Press. ISBN 9781138029972.

Stene, T.M., Wahl, R., Svarva, R. & Langlo, J.A. 2017. Digitalization of the Rail Network—Challenging the Traffic Management. In Čepin & Briš (Eds). *Safety and Reliability—Theory and Application*, pp 3113–3122. Taylor & Francis Group, London, ISBN 978-1-138-62937-0.

SWI 2017a. *Self-driving trains pilot project is on the horizon*. https://www.swissinfo.ch/eng/sci-tech/no-one-at-the-helm_self-driving-trains-pilot-is-on-the-horizon/42460870.

SWI 2017b. *Remote-controlled trains? Swiss rail bets on technology.* https://www.swissinfo.ch/eng/strategy-2020_track-to-the-future-railways-bet-on-innovation/42957374.

Taleb, N.N. 2007. The Black Swan: The Impact of the Highly Improbable. Random House, New York, 2007.

Tao, T. & Jing, X. (2014). Future of rail signalling and train control. In Mori, K. (Eds): *Concept-oriented research and development in information technology*. John Wiley & Sons Inc.

Tokody, D. & Flammini, F. 2017. The intelligent railway system theory. *International Transportation* is a special edition of Internationales Verkehrswesen, ISSN 0020-9511. 69, pp 38–40.

UITP (Union Internationale des Transports Publics). 2016a. *Statistics brief.* World report on metro automation. http://www.uitp.org/sites/default/files/cck-focus-papers-files/UITP_Statistic%20Brief_World%20Metro%20Automation%202016_Final02.pdf.

UITP 2016b. *Metro automation—a flexible and safe solution*. http://www.uitp.org/sites/default/files/Knowledge/PTI/PTI_1-2016_MetroAutomation_EN.pdf.

UITP 2017. *World report on metro automation*. http://www.uitp.org/world-report-metro-automation.

Vedantam, S. 2009. *Metro Crash May Exemplify Paradox of Human-Machine Interaction*. Washington post.

Wang, J. & Fang. W. 2014. A structured method for traffic dispatcher error behavior analysis in metro accident investigation. *Safety Science 70*, pp 339–347.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Revitalization of risk management in the Norwegian petroleum sector

B. Heide & G. Ersdal
Petroleum Safety Authority Norway, Stavanger, Norway

ABSTRACT: The PSA has launched a risk management project. The purpose is to stimulate a revitalization of the risk management in the Norwegian petroleum sector. The project is ongoing, which means that no conclusions, findings or reports are final yet. In this paper we describe the reasons for initiating this project. We also describe the process, and the preliminary focus areas of the project. Three messages to decision makers are given special attention. First, that risk management has to be an integral part of all organizational processes, and part of decision making. Second, uncertainty is a main component in the risk concept, leading to a strong focus on the need for robustness. Third, the management culture at all times absolutely has to be characterized by a sincere wish to reduce risk. Finally, we describe our intention to provide examples of practical risk management challenges, as well as some ideas for how these challenges can be handled.

1 INTRODUCTION

1.1 *Background and the reasons to initiate this project*

The PSA (Petroleum Safety Authority Norway) has launched a risk management project in order to stimulate a revitalization of the risk management in the Norwegian petroleum sector. We aim to contribute to preservation and further development of the industry's risk management. We address key issues from the viewpoint of the PSA, based on input from the stakeholders in the industry.

The primary target group for the project is managers and decision makers. Other groups that have important roles in safety management will hopefully also benefit from the content.

The need to preserve and further develop risk management is described in several sources. This has resulted in the PSA's decision to initiate a project on risk management. The purpose of the project is both process oriented and goal oriented. The process shall give new and even more appropriate attention to risk management, and a memorandum on risk management shall be issued.

Several events and reports have led to the initiation of this project. A brief presentation of the most relevant of these are:

The annual report "Risk Level Norwegian Petroleum" (RNNP) attempts to contribute to a common perception of the risk level between the stakeholders. The RNNP indicates that there have been significant improvements in several areas over a number of years, but also opportunities for further improvement (PSA, 2017).

The Norwegian government appointed an expert group to assess the current safety regime. In the "Engen report" (2013) it was concluded that the current regime is well-functioning. However, the report also indicated that there was a need for further development of risk management, especially related to major accident risk.

The Deepwater Horizon accident in 2010 showed that there is a need to review the principles and methods of risk management, and also how they are carried out in practice. Following this accident, both the supervisory authorities and the industry have taken a number of initiatives in the fields of risk management, barrier management and management follow-up (PSA 2014, Norwegian Oil and Gas Association 2012). The PSA (2014) concludes that such initiatives must be subject to continuous development in order to achieve lasting effects.

A working group at the Norwegian Oil and Gas Association (2015) has reviewed today's practice to identify improvement areas within risk informed decision making. They pointed out that in many cases the information is not available soon enough, and identified possible paths for improvement. The Norwegian Oil and Gas Association also published a report "Black Swans" (2017), which points to the need for an expanded perspective on risk, where knowledge building, experience transfer and learning become even more central.

The risk concept in the regulations was clarified in 2015. Here, attention to uncertainty was made even more explicit, and the PSA issued a memorandum in 2016 describing what we aim to achieve regarding this topic. The memo from 2016 highlighted the risk term, while the project described in this paper highlights the management aspect of risk.

1.2 *Why is risk management essential?*

Good risk management will enable the industry to find a reasonable balance between safety and economic profit. The operators on the Norwegian continental shelf have been given a great degree of freedom to find efficient ways of doing business as long as a certain safety level is achieved, through a "functional based regulatory framework". This framework both encourages and requires a certain mindset. The intention is to both encourage and require that the nature of the operations is taken into account, as well as local and operating conditions. A prerequisite for a functional based regulation is that operators and duty holders take responsibility and implement suitable risk management processes. There are therefore requirements for risk management and reduction processes in the HSE regulations, including continuous improvement.

Good risk management should also provide an opportunity to use resources in a way that has the best effect on safety and value creation. This implies that the industry must integrate risk management in the decision making processes. Also, a proper understanding of how risk management can be performed well in practice is required.

1.3 *Limitations of this paper and project*

This paper should not be read as the viewpoint of the PSA. It is merely the viewpoint of the authors of the paper.

"Risk management" is here limited to the PSA's authority area, and is based on major accident risk. It may also be relevant for working environment, natural environment, health, security and so on. The project is not intended to introduce any new requirements.

We cover only selected themes, and the project should be seen in conjunction with other areas highlighted by the PSA. See for instance the Barrier memo, the "Book about learning" and the "HSE and Culture" pamphlet.

2 THE PROCESS

Members of the PSA's Risk management group have been running the project. In the planning phase of the project, some keys to success became clear for such an ambitious project. There was a need for discussions with the stakeholders, as well as professional experts. Such discussions between the stakeholders (regulators, trade unions and industry associations) are already a formalized part of the Norwegian petroleum safety regime. See for instance the Safety Forum (PSA, 2017). Therefore we have in the early stages discussed the project with the tripartite Safety Forum and the tripartite Regulatory Forum.

Furthermore, we emphasize that risk management is not a stand-alone activity for risk management experts. Instead, the various departments need to use risk management as an integral part of their activities. Therefore, we have taken great care making sure that the risk management project is not run solely by risk management experts. We are consulting with all the PSA departments to get their input on what is the key to successful risk management. This idea of consulting with different technical disciplines has also been applied while discussing with the industry.

We have had several meetings with operating companies, contractors, labor unions and risk management experts. In these meetings we have tried to make sure that the discussions are driven by the needs of the decision makers and managers in the industry, rather than the opinions of the risk management experts.

Additionally, we have had several meetings with the other relevant safety regulators in the Norwegian petroleum sector, the Norwegian Board of Health Supervision and the Norwegian Environment Agency.

Typically, the discussions have centered on topics such as:

- What is necessary for a well-functioning risk management?
- Which tools are useful?
- What are the conditions and principles that are necessary for risk to be managed as an integral part of the activities instead of a stand-alone activity after the main decisions are made?
- How can we make sure that there is sufficient agreement between short-term and long-term objectives in various areas (HSE, profits, project progress, departments), at various levels and between various participants in the activities?
- What are the main challenges?
- What are the necessary criteria of success?

The feedback from these meetings have been very valuable to gain an understanding of challenges and good practices. However, we need to emphasize that the conclusions in the project will be our conclusions, and not merely a summary of the statements of the parties.

A well-known challenge for any organization when dealing with major accident risk, is to make sure that even if you have a strong track record on safety you still must remain vigilant. This is very relevant for our project, since the stakeholders to a large degree have presented how they act when they are performing well. Hence, the intention is that this project can help in passing on such knowledge of best practice. Further, it is our intention

that the project also serve as a reminder that one major accident is one too many.

In the span of the project, our focus area has shifted slightly based on the meetings and discussions we have had. The project is not completed, so instead of any final conclusions we can only give an overview of how the project has developed so far. When the project started, the following main themes were identified:

1. Risk management as an integral part of the management processes
2. Risk reduction processes
3. Closing the "control loop" by learning
4. Risk analysis and—acceptance criteria

We identified theme number 1 above as the main theme, while there was no "ranking order" between the three other themes. At the time of writing, theme number 1 is still the main theme. As this paper will show, the others are still included but have been somewhat restructured.

At the time of writing this paper, we have identified three areas that are of importance to our target audience, where we believe that further development is necessary. In the next section we describe these three areas. We point out that being excellent in one of these areas is not sufficient. Instead, these three areas must complement each other.

3 PRELIMINARY FOCUS AREAS IN THE PROJECT

3.1 *Risk informed management*

ISO 31000 describes risk management as "coordinated activities to direct and control an organization with regard to risk". It is further stressed that risk management should be a holistic and integrated part of all processes in the organization.

The reality that we too often see is that risk analysis is done after the decisions already are taken, and used to justify decisions that are already taken, in contrast to the "direct and guide" role described in ISO 31000. This have been recognized by many organizations, and e.g. the Norwegian Oil & Gas Producers Organization (NOROG) have produced several reports (NOROG 2015) on the importance having risk information in time for the decisions and of how to produce the risk information in time.

Integrated risk management in ISO31000 implies that risk management is a (integrated) part of the organization's processes and activities. Further, that risk management is a part of the management's responsibility.

Holistic risk management in ISO31000 implies that the risk management is able to take into account how a decision will affect and interact with other related areas of the organization's activity, and how they relate as a whole. This will imply that there is a need to collate the risk information from all related parts of the organization and balance these in the decision. It is therefore essential that risk management becomes aligned and integrated with the strategic planning processes. Further, it should enable the involvement of all stakeholders, internal and external, capturing their opinions on the companies' operation and critical issues. Risk management according to ISO 31000 should equip upper management with tools to help taking better and more informed decisions.

The Petroleum Safety Authority in Norway do experience examples of what we perceive as good risk management. In these cases, the industry describes processes where appropriate risk information are available when the decisions are being made. Further, that the decision makers takes an active role in obtaining relevant information before making decisions, and that decision makers have fruitful discussions about the acquired information and the strength of knowledge in this information. In essence that they live according to the proverb that "Doubt is the key to knowledge". Finally, that they take the strength of their knowledge into account in the decisions.

These examples of holistic and integrated risk management in accordance with ISO 31000 is what we in this project and paper call "Risk informed management".

In a risk informed management, one must start by understanding the organizations activity as a whole and the goal in this specific activity and decision (what is to be delivered). Here it is important to understand the context in which this is done and what requirements are set for the activity. Furthermore, one need to identify risk and possible reasons for why things may go wrong and the consequences of this. The decision on how to perform the activity must consider this understanding. An evaluation of the robustness of these plans is needed, in case of disturbances and changes. Obviously, the decision makers need to be well qualified to make the decision.

The execution must be carried out as planned and it is important that those who perform the activity have the necessary understanding of the context as well as the procedures for the task. It is also important that they have understood the basis for decisions, consequences and uncertainties so that they can react properly if disturbances and changes occur.

To improve, it is important that one assesses the delivery and the execution. Any learning from this evaluation should be communicated to the organization.

3.2 *Uncertainty and robustness*

Before making decisions, the responsible party shall ensure that issues related to health, safety and the environment are adequately considered. The basis for decision-making must have the necessary quality, where different alternatives and consequences have been studied and relevant experts, departments and user groups have been involved. To consider uncertainty has to be included in this decision process.

How can uncertainty be considered adequately? This is a research topic that has been given quite a bit of attention the last few years, for instance in the previous ESREL conferences. We will not go into detail of the possible methods here, but we stress the following point that has been made in the literature.

Say that a decision maker concludes that it is unlikely that a particular event will occur within a given time frame. Such statements are typically based on data, information, testing, analysis, argumentation, theory, models, assumptions, discussions with stakeholders and more. These statements may be more or less strong, and taking uncertainty into account includes clarifying what this knowledge consists of and how strong it is. If the knowledge is weak, the decision will have a weak foundation.

Obviously, if the knowledge is weak, correcting measures are necessary. In our project we stress the importance of robustness as an important measure to provide extra margins, in addition to similar concepts like resilience and the cautionary approach.

Requirements for robustness are used because deviations, unexpected changes and surprises can occur. Robustness must be emphasized especially for events with high potential.

As unforeseen events may occur, robustness is needed to ensure that the business can be operated in a safe manner. This implies that the organization are still able to operate in the event of disturbances slightly beyond the stresses they are expected to be exposed to. Surprises, unexpected changes and disturbanses can happen, and events that are considered unlikely can still happen if the assessments are based on assumptions that are incorrect.

A high degree of uncertainty must lead to a cautionary approach, for instance through barrier requirements and robust solutions or application of principles such as reducing risk as far as possible without significant disproportion between cost and effect. If there is insufficient knowledge concerning the effects of a preventive measure, further measures should be taken according to the HSE regulations.

3.3 *Management and culture*

How can risk management become a good tool for reducing risk, rather than a documentation and reporting of risk? The attention and priority of the management greatly contributes to the culture. Risk management processes and systems are important tools to achieve good risk management, but humans and organizations are preconditions for success. *The management culture always must be characterized by a sincere wish to reduce risk*.

Further, knowledge, involvement and commitment must be a core value that forms the decision-making processes in every aspect of the organization, even in times of pressure on the industry such as delays, changes and increased cost consciousness.

Based on Reason (1997) we describe a good HSE culture as just, reporting, learning and flexible. There needs to be established a belief that you will not be met by sanctions when you report an issue. This trust is easily eroded, and is difficult to reestablish. A key to earning and keeping the trust is to display consistency between HSE promises and action, especially when finding the balance between safety and other priorities. Here, communication skills are important, as well as establishing dialogue with all stakeholders instead of solely managing by commanding. This facilitates trust, commitment and valuable knowledge into the risk management process.

3.4 *Examples of risk management challenges*

In the coming memorandum from this project, the PSA also intends to highlight some practical risk management challenges. This should include some ideas for how these challenges can be handled.

However, the PSA is careful to point out that the risk owner is responsible for their risk. Thus, it is neither possible nor advisable that the authorities give detailed requirements or solutions that can be interpreted as requirements. Therefore, the PSA is careful not to spell out detailed solutions, but instead we intend to nudge the industry in a certain direction when we see such a need.

In this paper we have highlighted the importance of considering uncertainties before making decisions. It has been discussed in the literature how uncertainties can be handled in practice when using Risk Acceptance Criteria and Risk Reduction Processes. Summarized, we note that the Black Swans report describes that Risk Acceptance Criteria are fulfilled if:

- The calculations are within the criteria while the knowledge is strong, or
- The calculations are within the criteria with a large margin, while the knowledge is not weak.

Similarly, the required risk reduction processes beyond the Acceptance Criteria (ALARP) might incorporate the effect of uncertainties in a manner where:

– If the cost is low, the proposal is implemented.
– If the cost does not present a significant disproportion to the risk reduction, the proposal is implemented.
– If there are other important aspects, the proposal is considered implemented. Other aspects can be significant uncertainty, need for robustness and barriers, and more.
– Furthermore, it is not possible to set specific regulatory requirements and established minimum solutions in the industry aside based on arguments about risk informed cost benefit assessments.

We note that these bullet points are well known in the literature and in the industry, but that uncertainties are not always systematically considered.

When successful, the decision makers are actively involved in and are observant as to whether the knowledge base is strong or if there is a need to obtain more information. In good examples, statistics constitute a necessary part of the decision-making basis, in addition to other sources of information.

Other typical management challenges includes:

– To balance the different needs. For instance, regarding ventilation, where working environment and avoiding gas collection should be balanced. Other examples include keeping aware of the possibility of conflicts between short-term and long-term objectives in various areas (HSE, profits, project progress, departments), at various levels and between various participants in the activities.
– The uncertainties we have discussed above leads us to conclude that risk assessments can often be "massaged" to provide a justification of decisions that are already made. That is not the intention behind the regulation, as it does not provide any incentive to improve safety.
– Placing great emphasis on identifying all relevant risk factors. Quality is ensured by considering specific aspects, local conditions and operational conditions. Involvement, local knowledge and broad knowledge are keys to success. The opposite would be generic hazard identifications. Generic lists of hazard situations can be used as a starting point for brain storming, but will not be sufficient to provide an overall and nuanced basis for decision making.
– To execute according to the decisions, with a good understanding of the risk while vigilantly detecting possible changes and deviations. Personnel on the sharp end need sufficient knowledge of the task and the risk picture, and to understand how the activity is planned and which surrounding factors that have to be taken into account. Furthermore, major accident risk is a natural part of the various forms of risk visualization, such as Safe Job Analyses and Work permits.
– To avoid that the plans are made without involving the personnel on the sharp end, and to ensure that the reasoning behind the decisions are communicated. If not, any adjustments that need to be made in the sharp end might not be considered and handled as a change. Another challenge is if the performance is rigidly performed according to the plan, but without understanding of the risk picture and possible changes in conditions. Rigid processes and procedures can also lead to a quietly accepted practice of non-compliance.
– Transfer of experience might be the most difficult part of the management loop, especially to make sure that the experience is accessible to anyone who needs the information later. In our "Book about learning" (PSA 2013), we emphasize that organizational learning is a prerequisite for safety. The book can inspire actors to find good learning solutions themselves. The report "Black Swans" (Norwegian Oil and Gas 2017) places great emphasis on better ways to learn, and can help actors find practical ways to close the control loop.
– The regulations require that one should improve safety by continuously identifying where it is needed. A typical pitfall is a "good enough" philosophy without being conscious about when and where improvement is needed.

4 CONCLUSIONS

We have described the need for well-functioning risk management to create a reasonable balance between safety and economic profit. Additionally we have described that several reports indicate a need to revitalize the risk management of the Norwegian petroleum sector.

Our project aims to contribute to these goals. The project is ongoing, and at the time of writing this paper our draft project report is subject to a hearing at several government agencies.

We have involved the stakeholders in the project, and aim to continue this approach in the coming phases.

The preliminary message is:

– There is a plethora of risk management tools available. These may seem intricate and difficult to use for management, which can create confusion and apathy. However, we note that all good tools are based on the traditional "Control loop" (Plan-Do-Check-Act) which is the standard way to operate for managers.
– Risk management is the responsibility of the decision makers, meaning that the role of a risk

management department is to give advice to provide the best possible knowledge background for the decision makers.

- Risk management must be an integral part of decision making, as described in ISO 31000. When "risk management" is performed separately from (and after) other processes, the intentions behind the regulatory regime will not be fulfilled. Instead, we note that when the stakeholders describe successful risk management, they describe a standard control loop where the necessary information about risk is available at the right time.
- Relevant information needs to be available at the time of decision-making. This can be challenging. We observe that the industry is currently focused on this topic, and possible improved methods are examined.
- Uncertainties needs to be acknowledged when making decisions regarding major accident risk. We note that success is contingent on decision makers that take an active role in requiring relevant information, including questioning the strength of knowledge. Due to uncertainties, there is a need for a cautionary approach where the industry ensures sufficient robust solutions.
- Decision makers need to be aware of the potential for conflicting aims. Examples include project progress, safety and department aims. To ensure balance between such various goals, a holistic approach is necessary.
- Finally, the framework described above is dependent on a management culture that at all times is characterized by a sincere wish to reduce risk.

In essence, the main findings so far from the project is that:

1. risk have to be managed by risk informed management,
2. decisions need to take into account the uncertainty, taking into account the proverb "Doubt is the key to knowledge".
3. robustness is a necessity and a prerequisite for providing margins for deviations, unexpected changes and surprises,
4. safety concerned leadership and safety culture is the foundation of risk management.

ACKNOWLEDGEMENT

We gratefully acknowledge the contributions to the project from all the project members and numerous other discussion partners.

REFERENCES

ISO 31000 – Risk management.

Petroleum Safety Authority 2017; Principles for barrier management in the petroleum industry http://www.ptil.no/getfile.php/1344810/PDF/BARRIERS%20memorandum%202017%20eng.pdf.

Petroleum Safety Authority 2013; A book about learning http://www.ptil.no/getfile.php/1344467/PDF/English%20PDFs/L%C3%A6ringshefte_lavoppl%C3%B8st-engelsk.pdf.

Petroleum Safety Authority 2007; HSE and culture pamphlet http://www.ptil.no/getfile.php/13532/z%20Konvertert/Products%20and%20services/Publications/Dokumenter/hesciltureny.pdf.

Petroleum Safety Authority 2017; The Safety Forum http://www.ptil.no/safety-forum/category917.html.

Petroleum Safety Authority 2017; RNNP—Risk Level Norwegian Petroleum, http://www.ptil.no/getfile.php/1344338/PDF/RNNP%202016/ENG_summary_RNNP2016.pdf.

Petroleum Safety Authority 2014; Concluding report on our follow-up of the Deepwater Horizon accident, http://www.ptil.no/getfile.php/1326908/PDF/Deepwater/DwH_PSA%20final%20report_2014.pdf.

Norwegian Oil and Gas 2012; Deepwater Horizon—Lessons learned and follow up https://www.norskoljeoggass.no/Global/Publikasjoner/_H%C3%A5ndb%C3%B8ker%20og%20Rapporter/DWH%20rapporter/OLFs%20DWH%20rapport%20%202012.pdf.

Norwegian Oil and Gas 2015; Enhanced risk assessment and management https://www.norskoljeoggass.no/Global/L%c3%a6ring%20og%20erfaringsoverf%c3%b8ring/Form%c3%a5lstjenlige%20risikoanalyser/2015%2012%2022%20Memo%20%20Enhanced%20risk%20assessment%20and%20 management_v2.pdf.

Norwegian Standard 2008; NS5814 Requirements to risk assessments (only available in Norwegian).

Norwegian Oil and Gas 2017; Black Swans https://www.norskoljeoggass.no/Global/2017%20dokumenter/Black%20 swans.pdf.

Reason, J. (1997), *Managing the risks of organizational accidents*, Ashgate Publishing Ltd, Hampshire, England.

Simulation for safety and reliability analysis

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Effectiveness investigation of the correlation algorithms applied in a Smart ID Card system to monitor the use of PPE

M. Dźwiarek
Central Institute for Labour Protection—National Research Institute, Warsaw, Poland

T. Łempiński & M. Światowski
TENVIRK Sp. z o.o., Chorzów. Poland

ABSTRACT: A system called SMARTD ID CARD employees location and working time registration systems. Among the others, the data obtained in this system allow one to monitor whether the employees use constantly personal protective equipment. The research aimed at verification of the effectiveness in detection whether a worker used personal protection means. To this end, a robot simulated some human movements. Comparison between Pearson, Kendall's Tau, Spearman's and BiSerial correlation detection algorithms has been performed. The results obtained proved that the system was highly effective in detecting the use of personal protective equipment. The system was more effective at detecting that the personal protective equipment had not been used, what was important from the safety point of view. Additionally, the detection time of the lack of personal protection was relatively short and taking about 2 minutes. Basing on the results obtained the Authors recommend the use of SPEARMAN or BISERIAL algorithm.

1 INTRODUCTION

With the development of new technology procedures, more and more efficient machines have been designed. Especially, the development of Information Technology (IT) and telecommunication techniques has brought about the possibility of intelligent manufacturing system design. In such systems a human factor contribution to production process has been significantly reduced in view of the automation of entire process. The computational capacity of these systems in the first place, is applied to monitor the manufacturing process. However, their usage in monitoring the level of safety of system operators has been increased significantly.

In modern systems, sensors devices are increasingly used to the implementation of safety functions. Such systems detect the position of special labels, which in turn allows for localisation of the objects on which these labels have been installed (Dzwiarek 2015, Reiner et al. 2013). An example of such equipment is the Real Time Location System (RTLS), which detects the position of the label (Gomez et al. 2013, Guyoun et al. 2013). RTLS systems employ a variety of location technologies, e.g., active identification using radio waves, optical location and ultrasonic location. A sample application of the location system to improve safety consists in using it in monitoring of employee activities is showed in (Seppa 2012). That is especially useful in detection of situations requiring medical attention (Sachs et al. 2014).

At the same time, the rapid development of electronic systems applications to working time registration is observed. Currently, most common are devices containing the radiofrequency identification (RFID) (Roberts 2006), bar code or magnetic strip card readers, in which the working time is registered after the card has come closer or been inserted into the reader. The major drawback of these systems consists in the fact that they display the information only at the moment of identification (card approach or insertion). Later, the user does not have the access to working time registration data until the card is used again.

2 MATERIALS AND METHODS

2.1 *SMART ID CARD system*

Advantages of location and time recording systems combines the system developed at TENVIRK Sp z o.o. called SMARTD ID CARD. The system consists of:

- control units forming a radio Mesh Topology Network (MESH) (Huang et al. 2008) and connected with the cloud computing called "eAttendace" via Ethernet or General Packet Radio Service (GPRS) (Walke et al. 1991),

- smart cards which can be pinned to the worker's clothing, equipped with a set of sensors, please find the items, which, at most, could form the aforementioned set:
  - two independent accelerometers,
  - magnetometer,
  - moisture sensor,
  - pressure sensor,
  - ambient light sensor,
  - temperature sensor,
  - measurement of battery voltage.
- system software of class Enterprise Resource Planning/Customer Relationship Management (ERP/CRM) (Møller 2005) in the cloud, which collects real time data.

Computer control system is called "eAttendence" and is available in the cloud Software as a Service (SaaS) model (Haolong et al. 2015). The data from sensors are collected in the radio MESH network (Fig. 1) and submitted to the cloud eAttendance by internet links. Then they are processed and stored.

The smart cards are located in the MESH network based on nearest router address. The values of card parameters of Link Quality Indicator (LQI) (Qinab et al. 2013) and Received Signal Strength Indication RSSI (Yanga at all 2015), are the basis of which the system calculates theoretical distance from the nearest router. Knowing the location of the router, one can specify the card position. Additional localisation functions employ the frame time-of-flight measurement/or phase shift. Location accuracy is estimated as 1 m.

The system provides:

- recording of all inputs and outputs automatically with no the need for any actions to be taken either by employee or employer,
- automatic detection and signalling of all misuses, such as fake presence at work of the absent employee, illegitimate absence at the workstation, work, improper exchanging of cards etc.,
- monitoring of the presence at work both inside and outside the building, e.g. on the construction site, with no need for installing control devices or construct additional input gates,
- prompt, real-time information about the working time; delivered not only to employers but also to employees,
- additional information about detecting any accidents at work;i.e., fall or impact,
- detection whether PPE is used,
- automatic registration of the working time devoted to specific tasks, projects and clients,
- prompt production of corporate employee cards,
- significant reduction of costs as compared to classic systems.

2.2 *Research aims and methodology*

Among the other, the data resulting from the SMART ID CARD allow one to, monitor whether the employees use Personal Protective Equipment (PPE) constantly. To this end the correlation between signals representing movements of cards used by workers and the signals from the cards installed at the centre of PPE (e.g., helmet, glove, mask, glasses, etc.) has been examined.

The purpose of the research consisted in effectiveness verification of the detection mechanisms used in the SMART ID CARD, showing whether PPE was used or not.

Specific objectives were:

1. To check which correlation algorithms most suitable for detection of the use of PPE.
2. Finding how the correlation determining parameters affect the determination efficiency.

The research conducted in the Laboratory for Safety Techniques in Control Systems has been implemented in accordance with the methodology of the research developed within the framework of the project "Principles of the use of monitor-

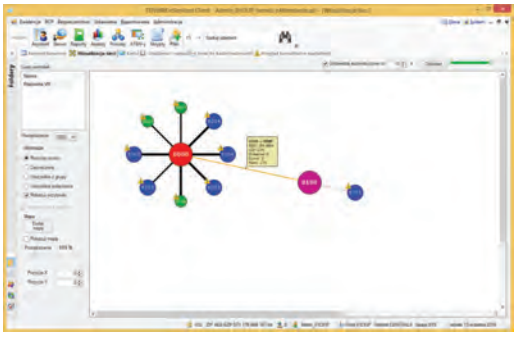


Figure 1. Sample visualisation of the network in eAttendance cloud.

Figure 2. Portal robot used in human movement simulation.

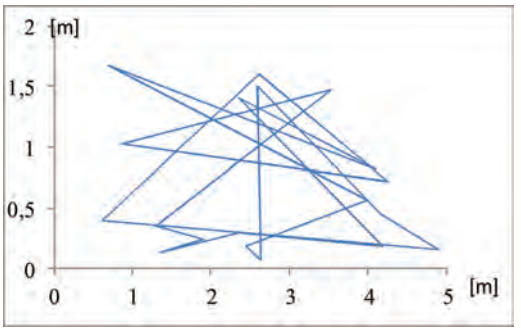


Figure 3. Robots head movement trajectory.

ing techniques for employee localisation with the use of ultra-broadband (UWB) communication to ensure machine safety" (Dzwiarek 2015). To simulate human movements the portal robot have been used (Fig. 2).

The displacements of robot working head represented human movements, as well as those of PPE. The card simulating men have been mounted on the robots head. The PPE use was simulated through a flexible connection of the second card, so that the movements of both cards were synchronised, but not identical. The robot head has been moved between consecutive points at a speed of 0.6 m/s. At each point of the trajectory the following sequence of movements was performed:

- 3 bottom-up movements,
- movement representing inactivity (a distance of 0.15 m along the Y axis, travelled at a speed of 0.01 m/s).
- 2 bottom-up movements,
- return movement to represent inactivity.

At the last point of trajectory the flexible connection of labels was broken and the smart card fell down (representing the PPE to be put away). The trajectory of robot motion that consisted of twenty randomly selected points is shown in Fig. 3. The experiment was repeated 3 times with the connection length 0.10 m, 0.50 m and 1 m, respectively.

3 RESULTS

In the course of experiment the signals from accelerometers situated in cards were registered. These were signals indicating:

- moving times,
- number of movments,
- motion sensitivity.

The data were stored in the cloud eAttendance. Sample data recorded during one experiment in the cloud eAttendance are shown in Fig. 4. The vertical line represents:

- start the experiment,
- connection fall,
- end of the experiment, respectively.

Then, the data obtained have been analysed. During the analysis effectiveness of the following correlation algorithms was compared:

- Pearson product—moment correlation coefficient (Buda &. Jarynowski 2004),
- Spearman's rank correlation coefficient, which is one of the non-parametric measures of statistical dependencies between monotone random variables (Kowalczyk 2015),
- Kendall's Tau coefficient (Kowalczyk 2015),
- BiSerial measure moment correlation coefficient (Linacre 2008).

The data were analysed basing on those registered by the Smart ID Card system, and written in the form of variables "quantity of movement" (IR), "time of movement" (CR) and "movement sensitivity" (CzR). During the tests they were compared with each other using different functions, which determined the value of the correlation coefficient:

1. The product of movement sensitivity and time of movement: CzR x CR.
2. The product of movement sensitivity and the logarithm of time of movement: CzR x log (CR).
3. The product of the logarithm of movement sensitivity and the logarithm of time of movement: log (CzR) x log (CR).
4. The product of movement sensitivity and quantity of movement: CzR x IR.
5. The product of movement sensitivity and the ratio of time of movement to quantity of movement: CzR x CR/IR.

To find the optimal values the Authors performed calculations for all control parameters:

a. width of the correlation window,
b. correlation coefficient threshold to identify the current situation,
c. type of correlation algorithm,
d. function of the measured data used as correlated variable.

During the experiment, we can distinguish two different phases:

1. Periods during which the both cards were attached to the robot head, i.e., their movements were correlated.
2. Periods during which one card remained motionless, so movements of the cards were not correlated.

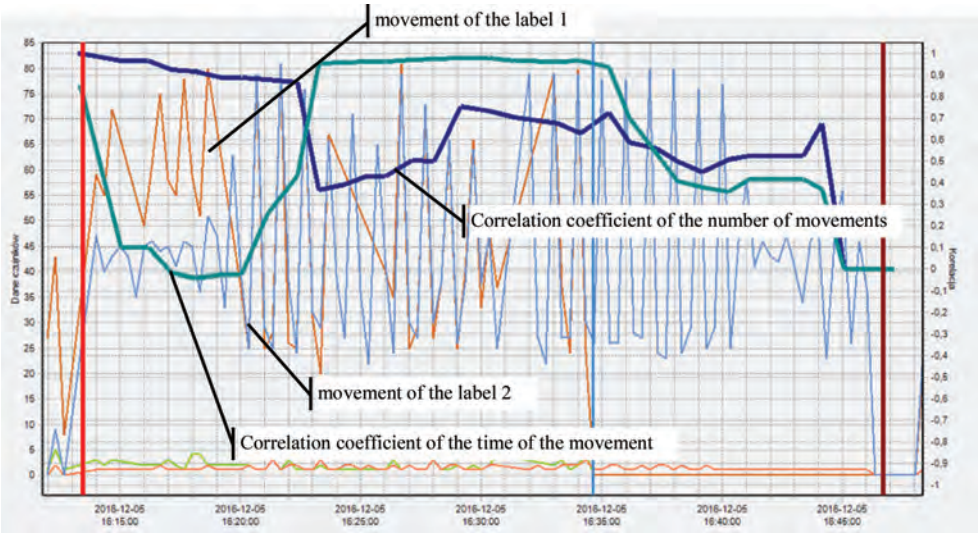


Figure 4. Example of experimental data obtained (Pearson correlation coefficient).

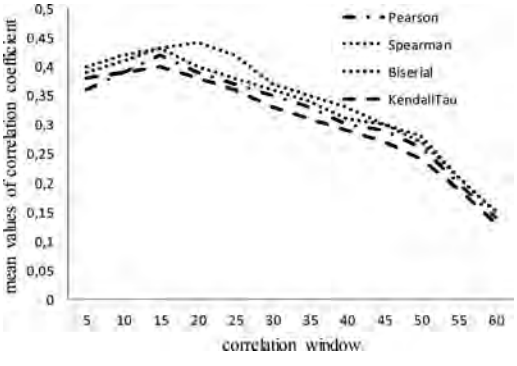


Figure 5. The averaged difference between the values for the correlate area and area without correlation depending on the correlation window width.

In the first phase, the values of correlation coefficients should be as high as possible (close to the unity), while in the second case, the correlation should not be observed, and the value of correlation coefficient should be close to zero.

To ensure that the detection of correlation loss was effective the following two functions of probability were determined:

- the probability that before the card release the correlation coefficient was not less than the assumed value 0.4,
- the probability that after the card release the correlation coefficient falls to zero after a certain period of time.

The results are presented in Table 1.

Table 1. The probability of an accurate indication of the PPE-lay-down obtained for each correlation algorithm.

Time after PPE put-away	Correlation algorithm			
	Pearson	Spearman	Biserial	Taukendall
0 (P(corrOFF > 0.4))	0.78	0.78	0.78	0.66
100 s (P(corrOFF < 0.4))	0.45	0.56	0.78	0.45
150 s (P(corrOFF < 0.4))	0.78	0.78	0.89	0.56
200 s (P(corrOFF < 0.4))	0.89	0.89	0.89	0.56
250 s (P(corrOFF < 0.4))	0.78	0.78	0.89	0.66

4 CONCLUSIONS

Test results have proved that the SMART ID CARD was highly effective in detecting the fact that PPE was worn, e.g., number of cases for which the value of the correlation coefficient has fallen below the threshold value reached 90%. System is much more effective at detecting that PPE was not used, what is important from the safety point of view. In addition, the time necessary for the detection of lack of personal protection is relatively short and reaches near 2 minutes. Taking

into account all the results obtained the course of research the use of SPEARMAN or BISERIAL algorithm for detecting lay down of the PPE is strongly recommended. Differences between these two algorithms are so small, that decision can be made depending on computational complexity the application involves.

An important innovation aspect consisted in the applied test procedure. On-line registration of all relevant test parameters in the cloud eAttendance allowed to both improve the conducted experiments, and significantly ordered their process. It is an excellent example of how the use technique of "Internet of things" can improve the research.

ACKNOWLEDGMENTS

The publication was developed on the basis of the results of the project. "The smart card to unattended time registration, access control, and monitoring of safety and productivity" conducted in accordance with the contract no POIR. 01.01.01-00-0227/15-00 in 1.1 "Projects R & d firms, Sub-measure 1.1.1" Industrial research and development work carried out by the companies' Operational programme Smart Growth 2014–2020 co-financed by the European regional development fund. Intermediary institution for this project is the National Research and Development Centre.

REFERENCES

Buda A. Jarynowski A. 2010. Life-time of correlations and its applications. ISBN 978-83-915272-9-0.

Dzwiarek, M. 2015. Real Time Location Systems for monitoring safety of the machine operators. *Proc. of the conf. "Safety of Industrial Automated Systems 2015". Königswinter 18–20 November 2015.* Berlin: Deutsche Gesetzliche Unfallversicherung.

Gomez, J. et al. 2013. Comparative Study of Localization Methods in Indoor Environments. *Wireless Personal Communications.* 72/4: 2931–2944.

Grove, A.T. 1980. Geomorphic evolution of the Sahara and the Nile. In M.A.J. Williams & H. Faure (eds), *The Sahara and the Nile*: 21–35. Rotterdam: Balkema.

Guyoun, H. et al. 2013. Design and Implementation of the Ubiquitous Sensor Network-Based Monitoring System Using RTLS (Real-Time Location System). *Sensor Letters.* 11/9: 1721–1725.

Haolong F. et al. 2015. An integrated personalization framework for SaaS-based cloud services. *Future Generation Computer Systems*. Vol 53: 157–173.

Huang, J. et al. 2008. Game Theory in Communication Systems. *IEEE Journal on Selected Areas in Communications*. 26 (7): 1042–1046. doi:10.1109/jsac.2008.080902.

Jappelli, R. & Marconi, N. 1997. Recommendations and prejudices in the realm of foundation engineering in Italy: A historical review. In Carlo Viggiani (ed.), *Geotechnical engineering for the preservation of monuments and historical sites*; *Proc. intern. symp., Napoli, 3–4 October 1996*. Rotterdam: Balkema.

Johnson, H.L. 1965. Artistic development in autistic children. *Child Development* 65(1): 13–16.

Kowalczyk, T. 2000. Link between grade measures of dependence and of separability of pairs of conditional distributions. *Statistics and Probability Letters.* Vol. 46: 371–379.

Linacre J. 2008. The Expected Value of a Point-Biserial (or Similar) Correlation. *Rasch Measurement Transactions*. Vol. 22 (1): 1154.

Møller, Ch. 2005. ERP II: a conceptual framework for next generation enterprise systems?. *Journal of Enterprise Information Management.* Vol. 18 (4): 483–497.

Polhill, R.M. 1982. *Crotalaria in Africa and Madagascar*. Rotterdam: Balkema.

Qinab, F., Daib, X., Mitchellb, J. E. 2013. Effective-SNR estimation for wireless sensor network using Kalman filter. *Ad Hoc Networks*. Vol. 11: 944–958.

Reiner, T., Reinhard, H., Sachs, J, Zwick T. Ultra-Wideband. 2013. Radio Technologies for Communications, Localization and Sensor Applications. InTech. Vienna.

Roberts C.M. 2006. Radio frequency identification (RFID). *Computers and security*. Vol. 25:18–26.

Sachs J., et al. 2014. Remote vital sign detection for rescue, security, and medical care by ultra-wideband pseudo-noise radar. *Ad Hoc Networks*. Vol. 13: 42–53.

Seppa H. 2012. The future of sensor networks. *VTT Impulse*: 20–27.

Tao Yanga, Xiaoping W. 2015. Accurate location estimation of sensor node using received signal strength measurements. *Int. J. Electron. Commun. (AEÜ)*. Vol. 69: 765–770.

Walke B., Mende W., Hatziliadis G. 1991. CELLPAC: A packet radio protocol applied to the cellular GSM mobile radio network, *Proc. of 41st IEEE Vehicular Technology Conference*, May 1991: 408–413.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A Monte Carlo method for evaluating dependability of mission repairable items

H. Cheng, J. Huang & Y. Zhang
NAA, Beijing, China

ABSTRACT: When evaluating or predicting dependability of products in reliability engineering practice, we may face one kind of problems where product is repairable and mission time is flexible, i.e. repairs are allowed during mission process only if the accumulated delayed time is less than or equal to a specified time duration. The analytic solutions to this kind of problems are difficult or impossible to be derived. To solve this kind of problems, the relevant terminologies are listed and analyzed, firstly. Then, the occurrence processes of the probabilistic events corresponding to mission success or failure are analyzed. After this, a numerical method based on Monte Carlo simulation is proposed to compute dependability of repairable items when mission time is flexible. Finally, two examples are presented. The first one illustrates the use of the proposed method to assess or to predict dependability of an exponential distributed item. The second one shows the use of the method to compare two products in the dependability when mission time is specified, but repairs are allowed in the limited accumulated time during mission. Although, for simplicity, products in the two examples are assumed to follow exponential distributions, the method is available to any items with one or more failure mechanisms that follow the common distributions in reliability field, such as exponential distribution, normal distribution, log-normal distribution and Weibull distribution.

1 INTRODUCTION

Normally, mission reliability R_m is used to identify the probability for one product to finish a specific mission successfully (Zeng *et al* 2011, Murthy & Rausand 2008). Meanwhile, availability is used instead of reliability for reparable products. However, the conventional concepts and theoretical methods are not available for some special problems in reliability engineering practice (Porry 1973, Krivtsov 2000). One of these problems is to solve the success probability, namely, dependability of repairable product during mission (Yang 2007). In such case, some of the failures occurring during the mission process can be repaired by the operators, and the mission time is somewhat flexible, namely, maintenance that does not exceed specified time is allowed.

To solve this kind of problems, a Monte Carlo (Lemieux 2008) based method is proposed by considering the procedure of the probabilistic events, failure occurrence, down for repair, successful repair, unsuccessful repair, maintenance time exceeding limit.

Examples are presented to show the availability of the proposed Monte Carlo method in dependability prediction, dependability assessment, and dependability comparison for various products or design.

2 TERMINOLOGY

Before establishing the simulation process, it is better to review and study the terminology related. The dependability is related to reliability and maintainability of a product. Here, for that the issue studied is about dependability, i.e., success probability for mission, we just pay attention to the critical failures resulting in mission stop or mission failure. The notions related is described below.

1. Critical failure: the failures causing mission break off.
2. Mean Time Between Critical Failures (MTBCF): the ratio between accumulated operating time T and critical failure times r during a specific mission profile.
3. Mean Time To Repair (MTTR): The ratio between accumulated maintenance time and the failures times r during a specific mission profile.
4. Dependability: an item's ability to successfully finish an intended mission.

3 SIMULATION MODEL

The proposed method is useful regardless of the distributions of the failure and maintenance time.

But for simplicity, the exponential distribution is used to show the procedure.

Assume that one item with n kinds of failure mechanisms, all of which are followed independent exponential distributions with distribution parameters $\{\theta_i\}, i = 1, 2, \ldots, n$. The number of repairable mechanisms is k out of n, and n-k is the number of unrepairable mechanisms.

Let T be the time duration of mission. The item can be repaired during mission, but the accumulated maintenance time must be less than or equal to a specified time T_M, where $T_M < T$. For the item can be repaired T_M hours, the smallest operating time for item will be $T - T_M$ during the mission. Therefore, the item can conduct the mission successfully if the following events happen:

Event A: the item has never failed during the mission.

Event B: the item fails at t_0, but T-$t_0 \leqslant T$.

Event C: the item has failed l times during the mission, all are repairable, and the accumulated maintenance time is less than or equal to T_m.

Based on the analysis above, the Monte Carlo simulation algorithm is established and shown in Figure 1.

The simulation steps in Figure 1 are explained as follows:

Step 1: For one item with n types of failure mechanisms, every failure mechanism may follow different probability distribution. The distribution parameters for every type of failure should be obtained before to establish the simulation program. The parameters for failure distribution can be denoted by $\boldsymbol{\theta}_i$, $i = 1, 2, \ldots, n$, regardless of one-parameter cases, say exponential distribution, or multi-parameters cases, say Weibull distribution.

Step 2: Simulate failure time t_i for the ith failure mechanism with distribution parameter $\boldsymbol{\theta}_i$ described in step 1. Conduct the simulation for all the failure mechanisms to get the failure time series $\{t_i\}$ corresponding to $\{\boldsymbol{\theta}_i\}$, $i = 1, 2, \ldots, n$.

Step 3: Find out the minimum value t_k, $1 \leq k \leq n$, in $\{t_i\}$. Denote the minimum failure time t_k as $t_{\min}$ and $t_{k,0}$, and we will use them in the subsequent steps.

Step for judgement: Judge whether the failure mechanism corresponding to $t_{\min}$ is repairable or not in mission. Go on to conduct the next step for the case that the failure mechanism is repairable, and the mission fails for that it is unrepairable.

Step for judgement: Compare the values of $t_{\min}$ and T-T_m. If the result is,

$$t_{\min} \geq T\text{-}T_m \tag{1}$$

then the continuous operating time of the item can achieve the smallest operating time $T - T_M$, and the mission will succeed. Otherwise, go on to conduct step 4.

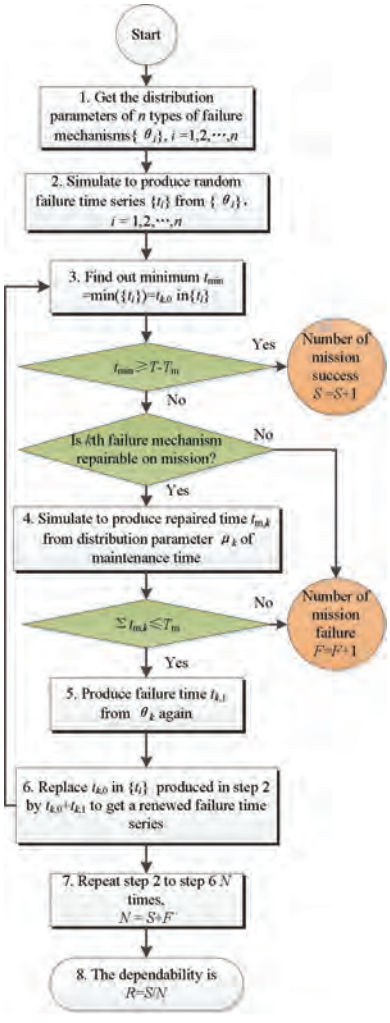


Figure 1. The Monte Carlo simulation algorithm for calculating dependability.

Step 4: Assume that the distribution parameter of maintenance time is $\boldsymbol{\mu}_k$ (the distribution of maintenance time could be any type normally used) corresponding to the failure mechanism of t_k in step 1. Simulate the maintenance time $t_{m,k}$ from distribution with parameter $\boldsymbol{\mu}_k$. The subscript "m" indicates "maintenance", and "k" indicates the kth failure mechanism**.**

Step for judgement: Compare the accumulated maintenance time $\Sigma t_{m,k}$ (Σ indicates cumulative sum) with the specified T_m, if,

$$\Sigma t_{m,k} > T_m \tag{2}$$

this means that the maintenance time used has exceeded the maximum maintenance time allowed. And mission fails.

if,

$$\Sigma t_{m,k} \leqslant T_m \tag{3}$$

then go on to step 5.

Step 5: Simulate the next failure time $t_{k,1}$ based on the distribution parameter $\boldsymbol{\theta}_k$ corresponding to the kth failure mechanism. In addition, for a prescribed problem in practice, we must know the performance of the repair that the item is repaired to new or is repaired to old first. Produce the next failure time $t_{k,1}$ based on this precondition.

Step 6: Get new failure time series by replacing t_k in $\{t_i\}$ produced in step 2 by $t_{k,0}+t_{k,1}$. And then denote the new time series with $\{t_i\}$, $i = 1, 2, \ldots, n$, as well.

Go back to step 3, continue to conduct the loop from step 3 to step 6 until that we can get a result that the mission fails or succeeds.

Step 7: Repeat the simulation procedure (from step 2 to step 6) N times, and N is determined based on the requirement of accuracy (Murthy & Rausand 2008).

Step 8: Compute dependability of the item based on the simulation result,

$$R = S/N \tag{4}$$

where N is the total number of simulated test shown in Figure 1, and S is the number of successful cases out of N times of simulation.

4 EXAMPLES

The Monte Carlo simulation method presented in section 3 could be used to predict or to assess the dependability of a specified item. Also, it is useful for comparing two items in the aspects of reliability and maintainability. Two examples are presented to show the availability of the proposed method. For simplicity, the failure time and maintenance time are both assumed to follow exponential distributions. The simulation processes are the same for the cases that the failure time and maintenance time corresponding to different failure mechanism follow different kinds of distributions.

4.1 *Example 1*

The mission time $T = 100$ h for item A. The maximum allowed accumulated maintenance time $T_m = 10$ h. There are four critical failure mechanisms which can cause mission pause for item A, and all the related parameters will be used in the simulation procedure are tabulated in Table 1. The first and second failure mechanisms are repairable during mission, and the third and fourth are not.

Calculate the dependability R of item A based on the simulation process shown in Figure 1. The change of R with simulation times N (the maximum number is 100000) is shown in Figure 2.

It can be seen from Figure 2 that the dependability R of item A converges to a somewhat stationary level with the increase of N. The maximum value of R is 1, and the minimum one is 0.5, corresponding to $N = 1$ and $N = 2$ respectively.

To show the change rule of R clearly, the abscissa scale is limited to the interval [0 10000]. The result is shown in Figure 3.

Table1. The distribution parameters for failure mechanisms of item A.

Distribution parameter of failure time θ_i	Repairable during mission	Distribution parameter of maintenance time μ_i
$\theta_1 = 400$	Yes	$\mu_1 = 2$
$\theta_2 = 500$	Yes	$\mu_2 = 3$
$\theta_3 = 1000$	No	/
$\theta_4 = 900$	No	/

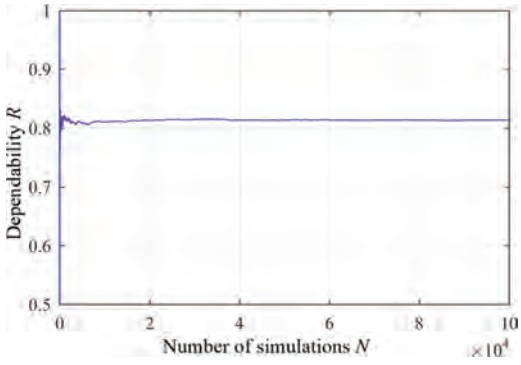


Figure 2. The change of R with simulation number N (global).

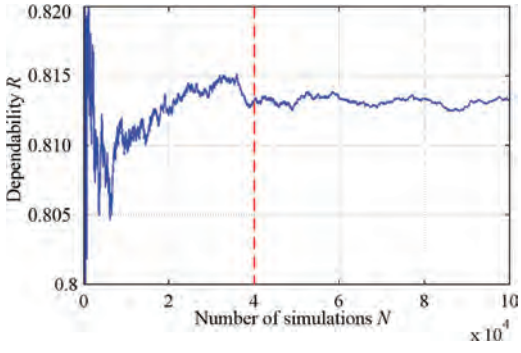


Figure 3. The change of R with simulation number N (local).

From Figure 3, it can be seen that R becomes stationary relatively when $N \geq 1107$. A step further, when N reaches 40000 and above, the result of R is very stationary, and the variation is acceptable.

To analyze the stationary property of the simulation based result, the change values of R corresponding to $N = 100000$ are divided into 1000sections with 100 continuous simulation results in each section. N' is used to denote the order of the sections. The coefficient of variation standard deviation to mean ratio, C (Ronald *et al.* 2007) of each section is calculated. The change rule of C with the section number N' is shown in Figure 4.

As shown in Figure 4, the maximum value of C is 0.1046. The coefficient of variation C converge quickly with the increase of N'. To show the result clearly, the ordinate axis is limited to the interval [0 0.0001] as shown in Figure 5.

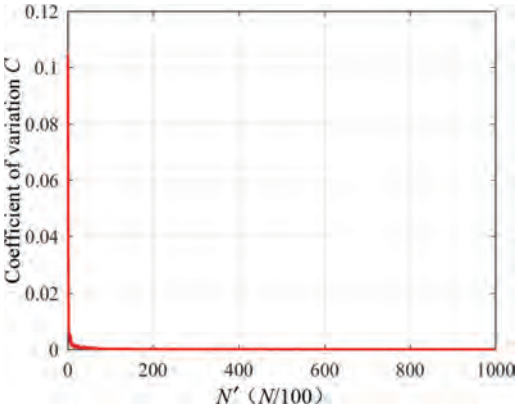


Figure 4. The change rule of coefficient of variation C of R with the divided section number N'.

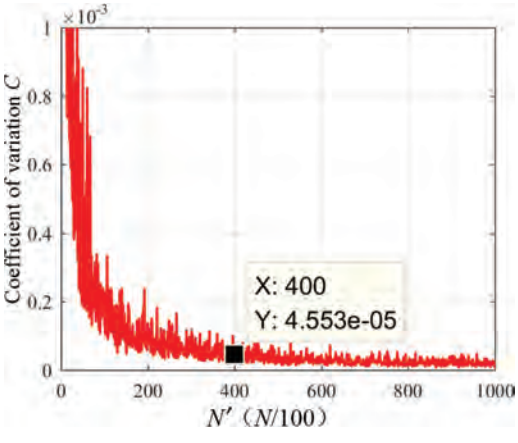


Figure 5. The change rule of coefficient of variation C of R with the divided section number N' (local).

Form Figure 5, we can see that the coefficient of variation C will be less than 10^{-5} when the simulation number N reaches $400 \times 100 = 40000$. This implies that the variation of R is extremely small already.

Based on all the analysis above, we accept the simulation results of R from $N = 40000$ to $N = 100000$ here in this example as shown in Figure 6.

The dependability of item A is shown in Figure 6, the mean value is,

$$\hat{R} = 0.8132 \tag{5}$$

4.2 *Example 2*

There are item B and item C to compare the dependability. The mission time is $T = 200$ h, and the maximum allowed accumulated maintenance time is $T_m = 12$ h.

There are four critical failure mechanisms which can cause mission pause for item B, and all the related parameters will be used in the simulation procedure are tabulated in Table 2. The first, second and third failure mechanisms are repairable during mission, and the fourth are not.

There are three critical failure mechanisms which can cause mission pause for item B, and all the related parameters will be used in the simulation procedure are tabulated in Table 3.

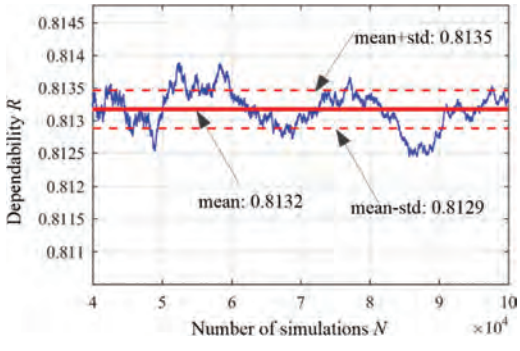


Figure 6. Simulation result of R of item A (N = 40000–100000) and its variation.

Table 2. The distribution parameters for failure mechanisms of item B.

Distribution parameter of failure time θ_i	Repairable during mission	Distribution parameter of maintenance time μ_i
$\theta_1 = 500$	Yes	$\mu_1 = 0.5$
$\theta_2 = 1000$	Yes	$\mu_2 = 1.5$
$\theta_3 = 1200$	Yes	$\mu_3 = 1.0$
$\theta_4 = 1300$	No	/

Table3. The distribution parameters for failure mechanisms of item C.

Distribution parameter of failure time θ_i	Repairable during mission	Distribution parameter of maintenance time μ_i
$\theta_1 = 800$	Yes	$\mu_1 = 3$
$\theta_2 = 1200$	Yes	$\mu_2 = 4$
$\theta_3 = 800$	No	/

Based on the data in Table 2, the MTBCF of item B through traditional method is,

$$MTBCF = \frac{1}{\frac{1}{500}+\frac{1}{1000}+\frac{1}{1200}+\frac{1}{1300}} = 217h \quad (6)$$

The MTTR of the mission repairable failure mechanisms of item B is,

$$MTTR = \frac{\frac{0.5}{500}+\frac{1.5}{1000}+\frac{1}{1200}}{\frac{1}{500}+\frac{1}{1000}+\frac{1}{1200}} = 0.8696h \quad (7)$$

Meanwhile, the MTBCF of item C is,

$$MTBCF = \frac{1}{\frac{1}{800}+\frac{1}{1200}+\frac{1}{800}} = 300h \quad (8)$$

The MTTR of the mission repairable failure mechanisms of item C is,

$$MTTR = \frac{\frac{3}{800}+\frac{4}{1200}}{\frac{1}{800}+\frac{1}{1200}} = 3.4h \quad (9)$$

Based on the analysis above, item C is more reliable than item B from the viewpoint of tradition method.

However, it is more reasonable to consider both the reliability and maintainability of a product when calculating its ability to finish a specific mission. The comparison of dependability of item B and of item C based on the proposed Monte Carlo method in Figure 1 are shown in Figure 7. The total simulation number N is 100000.

The simulation result becomes stationary when N is larger than 40000 as shown in Figure 7. A step further, we have analyzed the variation of the simulation results of item B and item C quantitively from N = 40000 to N = 100000 as shown in Figure 8 and Figure 9, respectively.

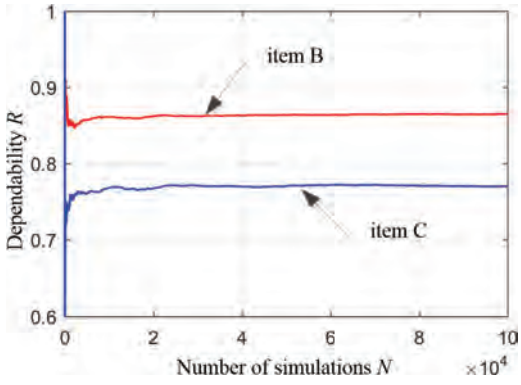


Figure 7. The comparison of simulation based dependability of item B and of item C.

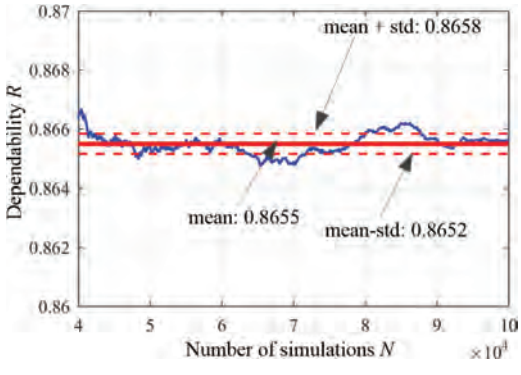


Figure 8. Simulation result of R of item B (N = 40000–100000) and its variation.

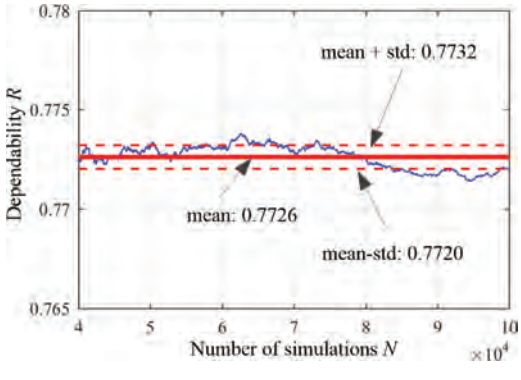


Figure 9. Simulation result of R of item C (N = 40000–100000) and its variation.

Table 4. The comparison of item B and item C in the aspects of reliability, maintainability and dependability.

Item	MTBCF (h)	Mission repairable proportion of critical failure	MTTR (h)	Dependability
B	217.27	75%	0.8697	0.8655
C	300.00	67%	3.400	0.7726

Based on the result shown in Figure 8, the dependability of item B is,

$$R_B = 0.8655 \tag{10}$$

The standard deviation of R_B is

$$S_{td}(R_B) = 3.4241 \times 10^{-4} \tag{11}$$

The coefficient of variation of R_B is

$$C(R_B) = 3.9562 \times 10^{-4} \tag{12}$$

The coefficient of variation is very small, this means that the result is stationary and believable.

Based on the result shown in Figure 9, the dependability of item C is,

$$R_C = 0.7726 \tag{13}$$

The standard deviation of R_C is

$$S_{td}(R_C) = 5.9918 \times 10^{-4} \tag{14}$$

The coefficient of variation of R_B is

$$C(R_C) = 7.7554 \times 10^{-4} \tag{15}$$

The coefficient of variation is very small; this means that the result is stationary and believable.

Based on the calculation above, the comparison of item B and item C in the aspects of reliability, maintainability and dependability is listed in Table 4.

As listed in Table 4, the inherent reliability of item C is better than item B. However, the maintainability design of item B is better comparatively. Finally, as shown in this example, the dependability (probability to finish the given mission) of item B is 12% higher than item C, although its MTBCF is 28% lower compared to the latter.

5 CONCLUSIONS

In this study, we focus on the issue calculating dependability of repairable item when the mission time is flexible, and get the following conclusions:

1. A Monte Carlo simulation algorithm is established based on the analysis of the occurrences of the random events during the mission where the item could be repaired with in limited accumulated maintenance time.
2. The use and availability of the proposed method in predicting or assessing the dependability of a product is developed and verified in the first example.
3. Based on the process and result in example 2, the dependability of one item during the mission scenario analyzed in this study is determined by both reliability and maintainability, and a better maintainability can even make up the disadvantage in the aspect of reliability.

REFERENCES

Krivtsov, V.V. 2000. *A Monte-Carlo Approach to Modeling and Estimating of the Generalized Renewal Process in Repairable System Reliability Analysis.* Maryland: University of Maryland, 20–60.

Lemieux, C. 2008. *Monte Carlo and Quasi-Monte-Carlo Sampling*. New York: Springer, 2–16.

Murthy, D.N. & Rausand, M. 2008. *Product Reliability: Specification and Performance*. London: Springer, 5–9.

Porry, K.E. 1973. *A General Monte-Carlo Simulation Model for Estimating Large Scale System Reliability and Availability.* Clemson: Clemson University, 20–40.

Ronald, W., Raymond, M. & Sharon, M. 2007. *Probability & Statistics for Engineers & Scientists*. New York: Pearson Education, 75–81.

Yang, G. 2007. *Life cycle reliability engineering*. New Jersey: John Wiley & Sons, 56–60.

Zeng, S. *et al*. 2011. *Reliability design and analysis*. National defense industry press, 1–10.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Real-time work simulations of aircraft unit fuzzy reliability evaluator

Norbert Grzesik, Robert Czapla & Aneta Krzyżak
Faculty of Aviation, Polish Air Force Academy in Dęblin, Poland

Mariusz Zieja
Air Force Institute of Technology in Warsaw, Poland

ABSTRACT: In this publication, the authors presented a proposal to use fuzzy expert inference system for evaluating the selected aircraft on-board unit reliability. Under consideration is military aircraft gun. The article is continuation of 2015 ESREL publication. There were initial analysis and now work simulations is presented. The project ensures reliable work (selected unit reliability, in addition maintenance process and safety improvements). It was verified by practical simulations. The research confirm possibility of use fuzzy expert inference systems in aircraft on-board units reliability evaluation.

1 INTRODUCTION

Prof. Lotfi Zadeh, in 1965, proposed to use multivalent logic in inference systems. This technology was and is used in many branches of science, including reliability systems research. Used by people systems reliability analysis with traditional mathematical and statistical methods is difficult because of its complexity. Research centres from all over the world started to search alternative object reliability evaluation approaches. For example in automotive (ABS system Mauer, 1995), reliability analysis (Azadeh et al. 2009), aviation (Grzesik 2004, 2012), medicine and even in music. Selected unit was M61A1 gun mounted on F-16 class aircraft. Many different factors effect gun reliability. Some of them changing slowly and some dynamically (Wrona, 2013). It can lead to gun parts damages. Proper design fuzzy expert inference system may contribute to decrease the gun maintenance costs and time. In the publication (presented on ESREL 2015 Conference, Żurek, & Grzesik 2014) fuzzy expert inference system designing process in Matlab, Fuzzy Logic Toolbox software (Mrozek & Mrozek, 1998) and initial analysis were described. Dynamic system simulation of designed models with use of Simulink software is presented as the next steps of the project development. The scope of research is to increase the system reliable work.

Extensive literature analysis and preliminary studies about very wide range of fuzzy logic applications provide the authors necessary fundamentals to continue the project in aircraft on-board units like selected one. Moreover there are several available publications in web sites describing similar problems, but not in military aircraft.

And that is why the authors conducted further system work analysis to verify proper system work practically in dynamic system work simulations which are presented in the article.

One of the most important factor determine achieving the main goal is accurate selection of the knowledge base experts during designing process of such systems. The experts were responsible for the criteria and principles of operation of these systems determination (creation of inference rules). The general overview about expert selection process is presented in Żurek & Grzesik (2014).

That kind of reliability evaluation method supports classical methods Adamski (2006), Jasztal et al. (2007), Tomaszek et al. (2011) Jasztal et al. (2008), Zio (2009) and can be used when there is no statistical information or the information is not sufficient, prognostic parameters, describing the object, change and the changes are not well known, long term prognosis are determined with combined mathematic and heuristic methods Idziaszek & Grzesik (2014).

Typical fuzzy inference system consists of fuzzy sets (membership functions). Those sets and membership functions are characteristic for fuzzy logic and provide opportunity to use the data in fuzzy inference systems.

This publication is an integral part of the authors scientific researches.

2 FUZZY EXPERT INFERENCE SYSTEM FOR SELECTED AIRCRAFT ON-BOARD UNIT RELIABILITY EVALUATION

Project is Mamdani type non-adaptive fuzzy inference system with two inputs and one output (MISO—

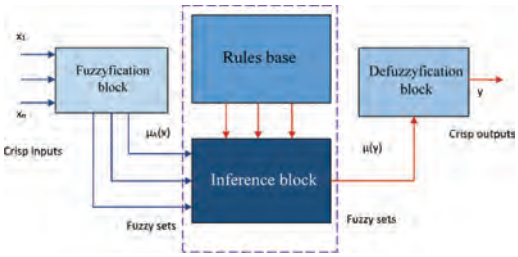


Figure 1. Fuzzy inference system (typical).

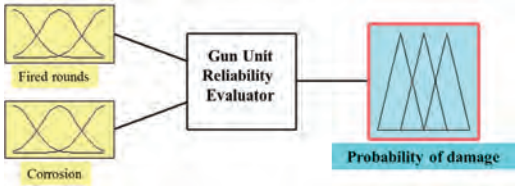


Figure 2. Fuzzy expert inference system for selected aircraft on-board unit reliability evaluation (first evaluator).

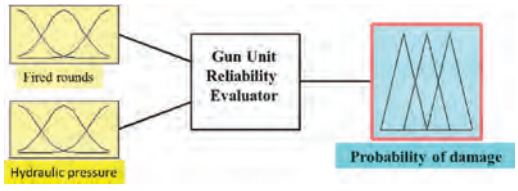


Figure 3. Fuzzy expert inference system for selected aircraft on-board unit reliability evaluation (second evaluator).

many inputs—single output), Wrona (2014). The project consists of two evaluator and is shown on Fig. 2 (first evaluator, presented on ESREL 2015 Conference) and Fig. 3 (second evaluator).

According to technical documentation, reliable gun system work depends on number of fired rounds, barrel corrosion (barrel reliability) and hydraulic pressure in the hydraulic drive. That is why authors analyzed the influence of those factors on gun probability damage.

3 REAL-TIME WORK SIMULATIONS OF THE SYSTEM

In order to carry out the simulation, the fuzzy expert inference systems designed in Matlab, Fuzzy Logic Toolbox software need to be transferred to Simulink software. The Simulink model consists of two signal blocks ("Signal Builder") and fuzzy inference system block ("Fuzzy Logic Controller", Fig. 4). In addition, each block is combined with a block showing the results of the simulation ("Scope"). Simulations were performed using the same, designed in Fuzzy Logic Toolbox models, because the structure of the test models is the same. To obtained results the simulation time is not relevant. It was selected that retrieving parameters from individual samples is easier. In the simulations, 11 samples were collected at intervals of every 10 seconds, so the simulation time is 100 seconds., Wrona (2014).

- M61A1 gun barrels evaluation damage probability simulation analysis

Three system simulations were performed in order to examine the proper project work. The input signals are: "Fired Rounds" expressed in number of total shots and "Barrel Corrosion". The output signal presents changes of gun barrels damage probability along with a change in input parameters. The input signal builders present the change of systems' input values.

Simulation I

The first simulation assumes that the input parameters are changing from the minimum to the maximum value in a linear way (Figs. 5, 6).

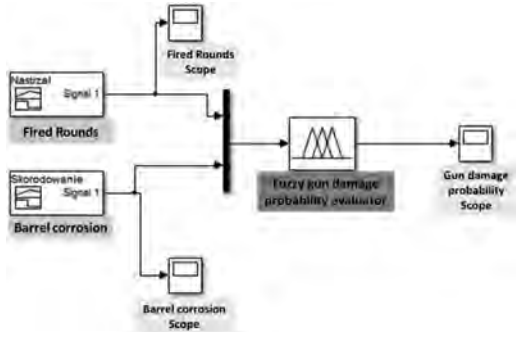


Figure 4. The Simulink model of fuzzy expert inference system.

Table 1. Signal samples parameters used in the simulation I.

Sample number	Fired rounds	Barrel corrosion [in]	Probability of damage [%]
1	0	0	3
2	5000	0,01	12,4
3	10000	0,02	21,8
4	15000	0,03	28,6
5	20000	0,04	33,3
6	25000	0,05	45
7	30000	0,06	52,4
8	35000	0,07	60
9	40000	0,08	63,4
10	45000	0,09	68,1
11	50000	0,1	97

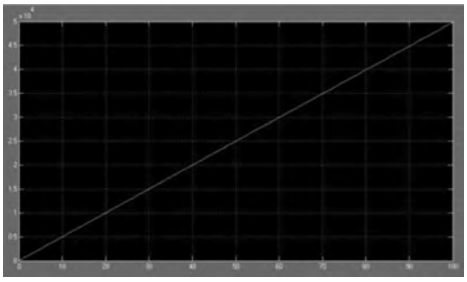


Figure 5. "Fired Rounds" linear input function in the Simulink signal builder.

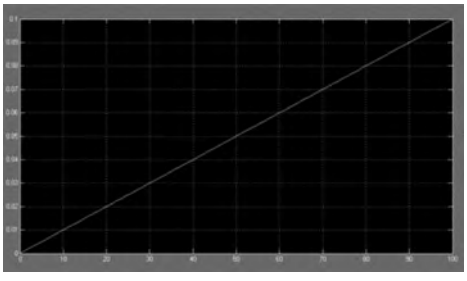


Figure 6. "Barrel Corrosion" linear input function in the Simulink signal builder.

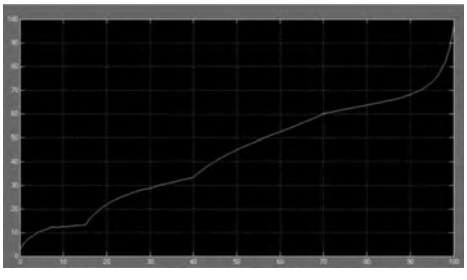


Figure 7. Gun barrels damage probability linear output function calculated in the Simulink.

The Fig. 7 shows that with simultaneous and proportionate input values increase the gun barrels damage probability increased almost linearly. From 90 seconds the damage probability increases sharply to 97%.

Simulation II

The input parameters of the "Fired Rounds" signal in the second simulation remained unchanged (Fig. 8) in relation to the parameters from the previous simulation. However, it was assumed that the corrosion parameters have changed very slightly and oscillate around a small value of 0.01 in (Fig. 9).

Taking into account the simulation II assumptions, the probability of gun barrels damage is kept

Table 2. Signal samples parameters used in the simulation II.

Sample number	Fired rounds	Barrel corrosion [in]	Probability of damage [%]
1	0	0,008	3,56
2	5000	0,0084	12,4
3	10000	0,0088	12,8
4	15000	0,0092	26,7
5	20000	0,0096	29,4
6	25000	0,01	29,7
7	30000	0,0104	30
8	35000	0,0108	33,5
9	40000	0,0112	49,6
10	45000	0,0116	61,5
11	50000	0,012	83,1

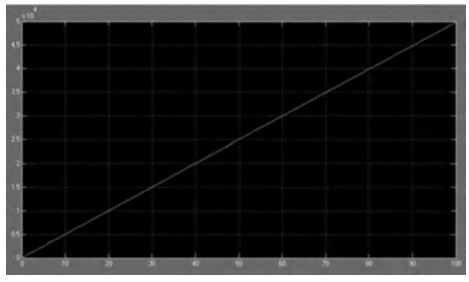


Figure 8. "Fired Rounds" linear input function in the Simulink signal builder.

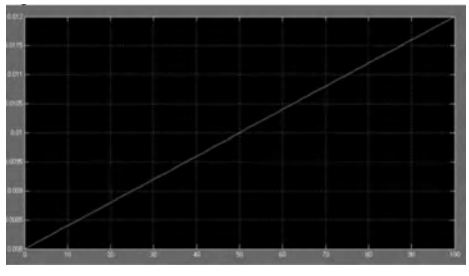


Figure 9. "Barrel Corrosion" linear input function in the Simulink signal builder.

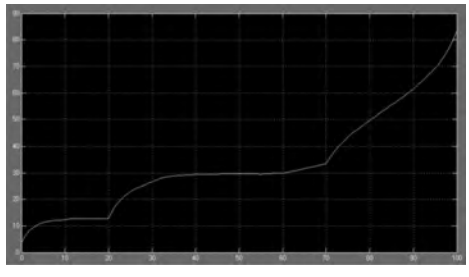


Figure 10. Gun barrels damage probability linear output function calculated in the Simulink.

at a low level for 20 seconds (Fig. 10). Then for about 70 seconds the damage probability is averaging. From 70 seconds the damage probability rapidly increases and reaches a large level due to activate antecedents with high and very high fired rounds value.

Simulation III

The third simulation imitated the situation with the intense increase in barrels corrosion with relatively small fired rounds.

Table 3. Signal samples parameters used in the simulation III.

Sample number	Fired rounds	Barrel corrosion [in]	Probability of damage [%]
1	5000	0,06	29,4
2	6000	0,064	31,1
3	7000	0,068	32,7
4	8000	0,072	38
5	9000	0,076	44,4
6	10000	0,080	49,6
7	11000	0,084	55,4
8	12000	0,088	59,3
9	13000	0,092	49,6
10	14000	0,096	70,6
11	15000	0,1	83,7

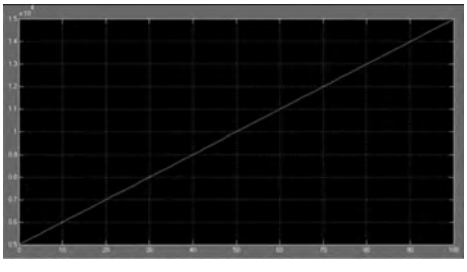


Figure 11. "Fired Rounds" linear input function in the Simulink signal builder.

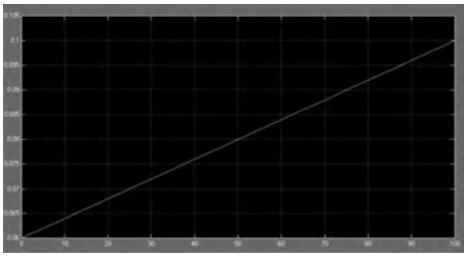


Figure 12. "Barrel Corrosion" linear input function in the Simulink signal builder.

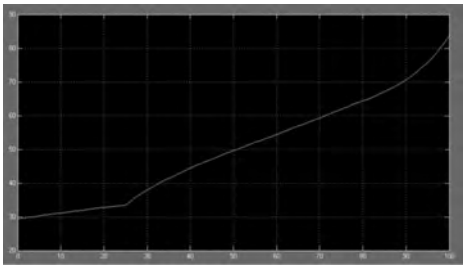


Figure 13. Gun barrels damage probability linear output function calculated in the Simulink.

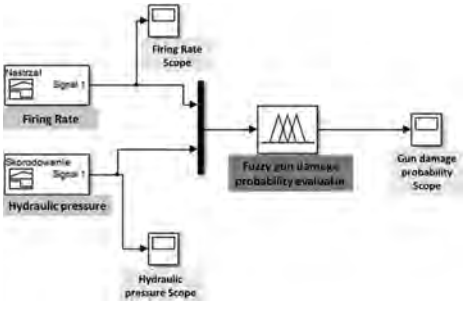


Figure 14. The Simulink model of fuzzy expert inference system.

Fig. 13 analysis of the gun barrels damage probability in the third simulation shows, that with a relatively small amount of fired rounds, rapid growth of barrels corrosion significantly increases the barrels damage probability.

- M61A1 hydraulic drive evaluation damage probability simulation analysis

Two system simulations were performed in order to examine the proper project work. The input signals are: "Firing Rate" expressed in number of shots per minute and "Hydraulic Pressure" (Fig. 14). The input signal builders present the change of systems' input values. The output signal presents changes of gun barrels damage probability along with a change in input parameters.

Simulation I

The first simulation assumes that the both input parameters are within the range of the optimal values.

According to performed simulations and taking into account project assumptions assessed that input parameters will be oscillated within the nominal system work parameters and the probability of gun hydraulic drive damage is small or very small (Fig. 17).

Table 4. Signal samples parameters used in the simulation I.

Sample number	Firing rate [rounds/min.]	Hydraulic pressure [psi]	Probability of damage [%]
1	5800	3000	3
2	5880	2900	8,3
3	5960	2800	10,6
4	6040	2700	11,8
5	6120	2600	12,5
6	6200	2500	12,9
7	6060	2560	12,7
8	5920	2620	12,4
9	5780	2680	12
10	5640	2740	11,4
11	5500	2800	10,6

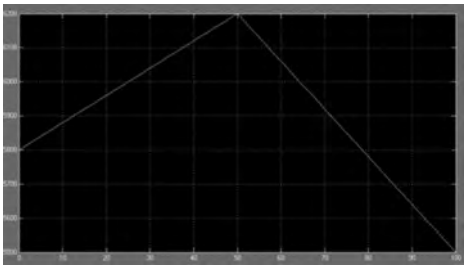


Figure 15. "Firing Rate" linear input function in the Simulink signal builder.

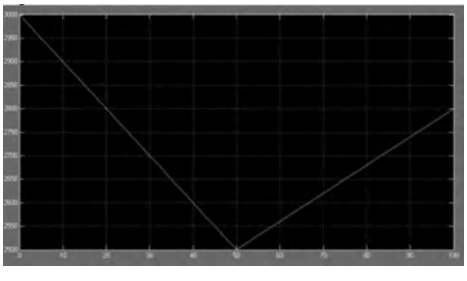


Figure 16. "Hydraulic Pressure" linear input function in the Simulink signal builder.

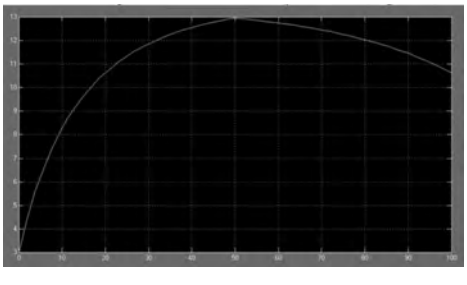


Figure 17. Gun hydraulic drive damage probability linear output function calculated in the Simulink.

Simulation II

Simulation number two assumes that gun drive supply pressure is continuously dropping to an unacceptable value. The firing rate remains the same for some time and then drops below the limit value.

Fig. 20 shows, that with the optimal gun firing rate the drop of gun hydraulic drive pressure does not affect its probability of damage. When the gun firing rate and the gun drive pressure are average, the probability of gun damage is also on average level. Dropping firing rate below 3,900 rounds/min

Table 5. Signal samples parameters used in the simulation II.

Sample number	Firing rate [rounds/min.]	Hydraulic pressure [psi]	Probability of damage [%]
1	6200	3000	3
2	6200	2830	10
3	6200	2660	12,1
4	5825	2500	13
5	5450	2280	13,3
6	5075	2070	13,5
7	4700	1860	25
8	4325	1640	29,1
9	3950	1430	37,5
10	3575	1210	63
11	3200	1000	96,4

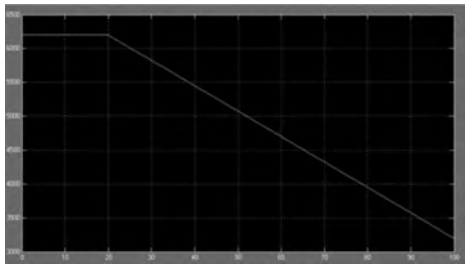


Figure 18. "Firing Rate" linear input function in the Simulink signal builder.

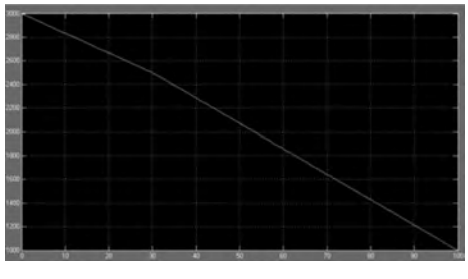


Figure 19. "Hydraulic Pressure" linear input function in the Simulink signal builder.

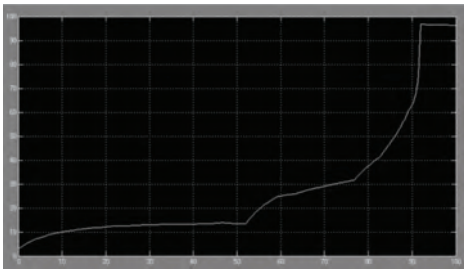


Figure 20. Gun hydraulic drive damage probability linear output function calculated in the Simulink.

leads to sharp increase the gun damage probability and when the pressure reaches a value of approximately 1000 psi, probability is at level of 97%.

4 DISCUSSIONS

Fuzzy expert systems are increasingly used as a tool to solve all sorts of scientific problems. These systems are characterized by high accuracy, while used mathematical simplicity and uncomplicated structure. In past and recent years, research are intensifying into the use of inference systems in the objects reliability and maintenance analysis and evaluation (lead to improvement, Kacprzyk & Yager 1985, Sergaki & Kalaitzakis 2002, Yager 2004). Moreover, these systems also are more and more used in the aircraft on-board systems. That is why authors decided to try to demonstrate the possibility of use fuzzy expert inference systems in aircraft on-board systems reliability evaluation (based on M61A1 gun damage probability evaluation systems). Some simplifying assumptions were used during designing process. However, it did not cover the benefits of use the technology in the complex objects reliability analysis.

The advantage of fuzzy expert inference systems is the fact that they can be used to assess the objects damage probability, that reliability depends on rapidly changing in time parameters (like: hydraulic pressure, firing rate), as well as on the parameters that changing during exploitations (like: fired rounds, barrel corrosion, Wrona 2013).

Another advantage the fuzzy systems is the ability to perform a simulation based on the real-time changing parameters of which it would be possible to determine when the individual maintenance procedures are necessary, and when it can be done later. Such action could lead to reduction in amount of unnecessary maintenance procedures and reduce operating costs.

5 CONCLUSIONS

The possibilities resulting from the use of fuzzy expert systems, are very flexible and depend on the creativity of engineers designing such systems. The results obtained in Simulink software are satisfactory at this level of designing process. The project development would require the M61A1 gun damages and malfunctions data collection.

The main contribution of the publication is that there is possibility of use the fuzzy logic (fuzzy expert inference systems) in selected aircraft on-board systems/units reliability evaluation. This approach is pioneer according to the systems reliability evaluation problems.

REFERENCES

Adamski M. 2006. Wpływ wybranych parametrów techniczno-taktycznych działek lotniczych na efektywność niszczenia celów naziemnych. IV Konferencja Naukowa "Kierowanie ogniem systemu obrony powietrznej (OPL)". Koszalin.

Azadeh A., Ebrahimipour V., Bavar P. 2009. A fuzzy inference system for pump failure diagnosis to improve maintenance process: The case of a petrochemical industry. *Expert Systems with Applications* 37 (2010) 627–639.

Grzesik N. 2012. Podstawy sterowania rozmytego. Dęblin, Wyższa Szkoła Oficerska Sił Powietrznych.

Idziaszek Z., Grzesik N. 2014. Object characteristics deterioration effect on task realizability—outline method of estimation and prognosis. *Eksploatacja i Niezawodnosc—Maintenance and Reliability* Vol.16, No. 3, 2014.

Jasztal M, Szajnar S, Ważny M. CFD-FASTRAN—software package for numerical analysis of flow around a body by the air stream. Eksploatacja i Niezawodnosc—Maintenance and Reliability 2008; 4(40): 55–62.

Jasztal M., Żurek J., Tomaszek H. 2007. A method of evaluating fatigue life of some selected structural components at a given spectrum of loads—an outline. *Eksploatacja i Niezawodnosc,* Nr 3(35)/2007, str.69–71.

Kacprzyk J., Yager R.R. 1985. Emergency-Oriented expert systems: A fuzzy approach. *Information Sciences*, Volume 37, Issues 1–3, December, Pages 143–155.

Mauer, G.F. 1995. A fuzzy logic controller for an ABS braking system. Fuzzy Systems, *IEEE Transactions on Fuzzy Systems* (Volume:3, Issue: 4), Page(s): 381–388, ISSN: 1063–6706.

Mrozek B.; Mrozek Z. 1998. Matlab 5.x Simulink 2.x—Poradnik użytkowania. Warszawa: Wydawnictwo PLJ.

Sergaki A., Kalaitzakis K. 2002. A fuzzy knowledge based method for maintenance planning in a power system. *Reliability Engineering and System Safety*, 77 (2002) 19–30.

Tomaszek H., Jasztal M., Zieja M. 2011. A simplified method to assess fatigue life of selected structural components of an aircraft for a variable load spectrum. *Eksploatacja i Niezawodnosc—Maintenance and Reliability;* 4: 29–34.

Wrona O. 2013. Analiza możliwości wykorzystania nieadaptacyjnych sterowników rozmytych podczas sytuacji awaryjnych w lotnictwie. Dęblin, Wyższa Szkoła Oficerska Sił Powietrznych.

Yager R.R. 2004. Uncertainty modeling and decision suport, *Reliability Engineering & System Safety,* Volume 85, Issues 1–3, July–September: Pages 341–354.

Zio E. 2009. Reliability engineering: Old problems and new challenges. *Reliability Engineering & System Safety.* Volume 94, Issue 2, February: Pages 125–141.

Żurek J., Grzesik N. 2014. Fuzzy expert aircraft onboard control systems assistant. Safety reliability and risk analysis: Beyond the horizon. *ESREL conference proceedings.* London: Taylor & Francis Group: Pages 250–251.

Żurek J., Grzesik N. 2015. Fuzzy expert inference system for selected aircraft on-board unit reliability evaluation. Initial project analysis. *ESREL conference proceedings*. London: Taylor & Francis Group.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Selecting correct architecture for mission critical safe control systems

E.H. Dogruguven
Aselsan Inc., Ankara, Turkey

I. Ustoglu
Department of Control and Automation Engineering, Yildiz Technical University, Istanbul, Turkey

ABSTRACT: Developing safety critical systems require long years of planned investments, broad theoretical knowledge and domain experience. Data interchange between CPUs, synchronization, computation speed and diagnostic measures shall exhaustively be evaluated along with the effects of the parameters used in the reliability and safety calculations ex tunc. This study focuses on the effects of calculation parameters for different architectures. Special attention is paid for the architecture 1oo2D regarding its model and normative definition. It has also been revealed that there are correlations between some parameters which seem independent. An advising route map is created to distill what kind of methods can be applied to decrease the hazard rates. For concretizing some concepts and sharing field experience, railway domain is selected, however the study is fully applicable to other domains due to deeming the norm IEC 61508 along the entire paper.

1 INTRODUCTION

Murthy et al. (2008) mentions that regulatory requirements, customer requirements, and technical requirements are to be fulfilled when developing a safety instrumented system (SIS). For safety related systems in railway, automotive, nuclear power etc. the domain specific norms dealing with system, hardware (HW), software (SW) and transmission are derived from the core norm IEC 61508 (Functional safety of electrical/ electronic/ programmable electronic safety-related systems) as depicted below in Figure 1. CENELEC Norms are the railway norms derived from this standard.

The norm IEC 61508 defines in Part I four discrete safety integrity levels (SIL) with regards to the low demand and high demand or continuous. In case the yearly demand rate is lower than one, then the low demand mode is applied, else continuous mode should be considered. For the low demand mode, SIL is determined according to the Probability of a dangerous failure on demand (PFD) explained as safety unavailability of an E/E/PE safety-related system to perform the specified safety function when a For the high demand mode, SIL is determined according to the average frequency of a dangerous failure per hour (PFH) explained as average frequency of a dangerous failure of an E/E/PE safety related system to perform the specified safety function over a given period of time.

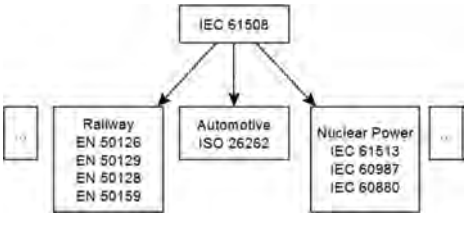


Figure 1. Derivation of standards from IEC 61508.

A similar approach is applied also in EN 50129 (Railway applications—Communication, signalling and processing systems—Safety related electronic systems for signalling). IEC 61508 and CENELEC norms describe both qualitative and quantitative requirements to be applied for these integrity levels (IL). Architecture is both relevant with qualitative and quantitative requirements. For instance, hardware fault tolerance (HFT) is defined as qualitative requirement and PFH_G is allocated as quantitative requirement for the pertinent IL. Here, the subscript "G" represents the group of voted channels which is also used in the remaining part of the paper in this manner.

Although several formulas are provided in the norms or stochastic processes can be utilized for the calculation of the PFH_G, it is not comprehensible to judge the effects of the parameters in these formu-

Table 1. SIL vs PFD_{avg} and PFH.

SIL	PFD_{avg}	PFH [h^{-1}]
1	1E-1 > PFD_{avg} ≥ 1E-2	1E-5 > PFH ≥ 1E-6
2	1E-2 > PFD_{avg} ≥ 1E-3	1E-6 > PFH ≥ 1E-7
3	1E-3 > PFD_{avg} ≥ 1E-4	1E-7 > PFH ≥ 1E-8
4	1E-4 > PFD_{avg} ≥ 1E-5	1E-8 > PFH ≥ 1E-9

las. However while developing on-board computer, it has been realized that such a study would be very useful to decide the design path. Chen et al. (2008) simulated RAMS of triple-modular-redundant system and a dual-modular-duplex-redundant system and compared using the estimates of actual hardware failure rates. King (2014) analyzed Hazardous Event Frequency per year with respect to demand rate. But, neither effect of the parameters in Table 2 nor the crucial dependent failures were discussed in these studies. Smith & Gruhn (1995) showed dependency of the safety system on some parameters at some level, however this study is not very detailed, the results were shown in bar charts with two only values. Moreover, the architectures 1oo2D and 1oo3 were not covered and high demand/ continuous mode was not evaluated, but the behavior of parameters in high demand mode are very different as the algebraic equations differ essentially from each other in low and high demand mode for the same architecture. There are some studies like the one from Liu and Rausand (2016) about proof testing effect or the one from Ilavsky et al. (2013) about ß factor effect, but no detailed work covering all parameters and all plausible architectures has been met in the literature. Therefore, it is believed that a study examining different architectures to find out the influences of parameters to reach very low PFH_G could be very beneficial and interesting. These very low challenging values are required since the result of a possible hazard at some systems like nuclear power plants, autonomous vehicles, high speed trains or air planes could be catastrophic. For instance, European rail traffic management system/European train control system (ERTMS/ETCS) defines the core hazard for ETCS on-board in Subset – 091 in clause 4.2.1.8 as exceedance of the safe speed or distance as advised to ETCS with tolerable hazard rate of $2 \times 1E\text{-}09$ (1/h) for the entire system which results in a PFH_G requirement of the vital train computer about less than 1E10–12 (1/h). A similar or even less quantitative target for the on-board vital computation unit is obtained for the metro lines considering the safety requirement in the CBTC standard IEEE 1474.1. It sets the requirement that the CBTC wayside and train-borne equipment located within any contiguous portion of a one-way route (including the maximum number of other trains that can be located in this contiguous portion of a one-way route under the specified peak operating headway) that can be traversed by a train traveling at the specified maximum authorized speed for one hour or less shall have a total calculated aggregate MTBHE (total of all critical and catastrophic hazards) of at least 1E-09 operating hours. This information shows that quantitative values that are thousand times better than SIL 4 are to be obtained for several constituents like vital train computer or vital wayside Automatic Train Protection computer. Another novelty of the paper is a route map providing structural information to develop safety critical constituent to reach very low dangerous failure rates.

This paper is organized as follows. The introductory part gives a brief information about the norms, functional safety and motivation of the study. Section II distills technical background of the architectures and the calculation methodologies. In Section III, the simulation results of PFH_G as a function of different parameters are described by highlighting the most interesting results in accordance with the pertinent architecture in the form of deductions. Furthermore, the architectures are compared in this part. Section IV provides a route map as a proposal to reach very low PFH_G ranges. Finally, some concluding remarks are conveyed.

2 SAFETY ARCHITECTURES

Karydasa & Brombacherb (1999) explain that architectural modeling deals with the development of a detailed block diagram of the programmable electronic system identifying each subsystem and the interconnections related to the safety function under consideration. IEC 61508 defines six architectures providing the formulas depending on the parameters defined in Table 2 to calculate the PFH_G. However, according to IEC 61508, part 2, Table 3, for SIL 4, the HFT must at least be one, if diagnostic coverage (DC) is over 99%, and two, in case DC is between 90% and 99%. The architectures 1oo1 and 2oo2 thereupon are eliminated for

Table 2. Parameters and their explanations.

λ	Total failure rate (1/h) of a channel in a subsystem	ß	The fraction of undetected failures that have a common cause
λ_s	Safe failure rate (1/h)	$ß_d$	Of those failures that are detected by the diagnostic tests, the fraction that have a common cause
λ_{sd}	Detected safe failure rate (1/h)	MTTR	Mean time to restoration (h)
λ_d	Dangerous failure rate (1/h)	MRT	Mean repair time (h)
λ_{dd}	Detected dangerous failure rate (1/h)	K	Fraction of the success of the autotest circuit in the 1oo2D system
λ_{du}	Undetected dangerous failure rate (1/h)	T_1	Proof test interval (y)
DC	Diagnostic coverage		

this analysis. Chen et al. (2008) points out that triple modular redundancy and dual-duplex modular redundancy are main architectures used for safety-critical computer systems. Furthermore, according to our investigations in the railway industry, 1oo2, 1oo2D and 2oo3 are the main vital computer architectures for Communication Based Train Control (CBTC) systems used in metro lines and ERTMS ETCS systems used in high speed lines. In case availability requirements are high, then redundant 1oo2 and 1oo2D architectures are selected. 1oo3 architecture is usually used in transmitters and electromechanical relays, but not preferred for computing units due to relative lower availability since the unavailability of the computing unit would result in the unavailability of the entire system.

Reliability block diagram paradigm is used for deriving the equations in IEC 61508. Another widely accepted and used method for sophisticated architectures is Markov Diagrams, a stochastic process. In this work, the formulas in IEC 61508 are utilized. Applying different paradigms will surely result in different algebraic outcomes, however as Zhang et al. (2003) shows the difference is not so crucial when applied correctly.

The calculation formulas and architecture explanations provided in IEC 61508 are distilled in following. In these algebraic formulations, ß-Factor paradigm is selected for modeling the dependant failures.

2oo3 Architecture: This architecture consists of three channels connected in parallel with a majority voting arrangement.

$$PFH_G = 6\left((1-\beta_D)\lambda_{DD} + (1-\beta)\lambda_{DU}\right)(1-\beta) \lambda_{DU} t_{CE} + \beta\lambda_{DU} \quad (1)$$

1oo3 Architecture: This architecture is similar to 2oo3. It consists of three channels connected in parallel with a voting arrangement such that one channel is enough to perform the safety function.

$$PFH_G = 6\left((1-\beta_D)\lambda_{DD} + (1-\beta)\lambda_{DU}\right)^2(1-\beta) \lambda_{DU} t_{CE} t_{GE} + \beta\lambda_{DU} \quad (2)$$

1oo2 Architecture: This architecture consists of two channels connected in parallel, such that either channel can process the safety function. It is assumed that any diagnostic testing would only report the faults found and would not change any output states or change the output voting.

Figure 2. 1oo2, 2oo3 and 1oo2D architectural descriptions.

$$PFH_G = 2\left((1-\beta_D)\lambda_{DD} + (1-\beta)\lambda_{DU}\right)(1-\beta) \lambda_{DU} t_{CE} + \beta\lambda_{DU} \quad (3)$$

$$t_{CE} = \frac{\lambda_{DU}}{\lambda_D}\left(\frac{T_1}{2} + MRT\right) + \frac{\lambda_{DD}}{\lambda_D} MTTR \quad (4)$$

1oo2D Architecture: This architecture consists of two channels connected in parallel. During normal operation, both channels need to demand the safety function before it can take place. In addition, if the diagnostic tests in either channel detect a fault then the output voting is adapted so that the overall output state then follows that given by the other channel. If the diagnostic tests find faults in both channels and a discrepancy that cannot be allocated to either channel, then the output goes to the safe state. In order to detect a discrepancy between the channels, either channel can determine the state of the other channel via a means independent of the other channel. The channel comparison/switch over mechanism may not be 100% efficient therefore K represents the efficiency of this inter-channel comparison/switch mechanism, i.e. the output may remain on the 2oo2 voting even with one channel detected as faulty. The parameter K will need to be determined by an FMEA.

$$PFH_G = 2(1-\beta)\lambda_{DU} + \left((1-\beta)\lambda_{DU} + (1-\beta_D)\right) \lambda_{DD} + \lambda_{SD}) t_{CE}' + 2(1-K)\lambda_{DD} + \beta\lambda_{DU} \quad (5)$$

$$t_{CE}' = \frac{\lambda_{DU}\left(\frac{T_1}{2} + MRT\right) + \left(\lambda_{DD} + \lambda_{SD}\right) MTTR}{\lambda_{DU} + \lambda_{DD} + \lambda_{SD}} \lambda_{SD}\frac{\lambda}{2} DC \quad (6)$$

Although this formula does not make a distinction regarding the DC between detected and undetected dependant failures, Hokstad (2005) showed how the DC is influenced when comparison of channels is applied as part of the diagnostic testing for multiple channel systems and suggests an alternative to define two betas, i.e. β for DU failures, and β_D for DD failures. Notwithstanding, this paper takes the formulation provided in IEC 61508 part 6.

3 SIMULATION RESULTS AND DEDUCTIONS

In this section, the effects of parameters are explained. At each step, one parameter and of

Table 3. Parameters used for the analysis of λ effect.

λ (PoI)	λ_s	λ_{sd}	λ_d	λ_{dd}	λ_{du}	DC	SFF	β	β_d	MTTR	MRT	K	T1
1E-07<λ<1E-05	v	v	v	v	v	0.99	0.995	0.02	0.01	8	8	0.98/0.9999	1

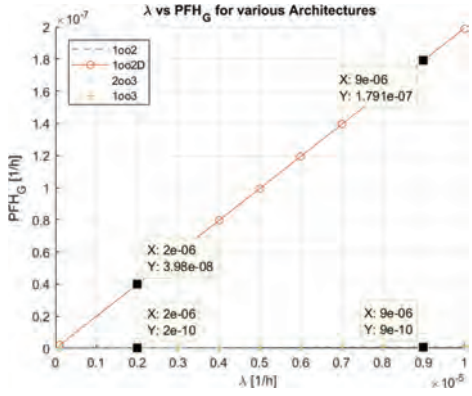


Figure 3. The influence of λ (K = 0.98 for 1oo2D).

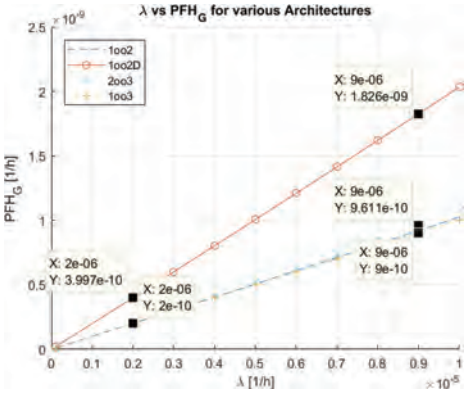


Figure 4. The influence of λ (K = 0.9999 for 1oo2D).

Table 4. Parameters used for the analysis of ß effect.

λ	λ_s	λ_{sd}	λ_d	λ_{dd}	λ_{du}	DC	SFF	β (PoI)	β_d	MTTR	MRT	K	T1
1E-07	5E-08	4.95E-08	5E-08	4.95E-08	5E-10	0.99	0.995	0.02 < β < 0.2	beta/2	8	8	0.98/0.9999	1

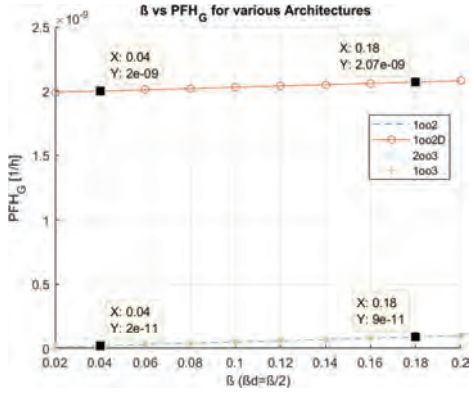


Figure 5. The influence of ß (K = 0.98 for 1oo2D).

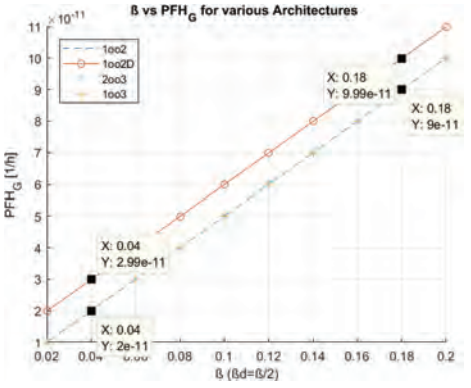


Figure 6. The influence of ß (K = 0.9999 for 1oo2D).

course the parameters derived from this parameter if there exist are varied while remaining ones are fixed. A table is provided to track the values of variables in the equation. The parameter of interests is shown as "PoI", variating value as "v". Having provided the simulation results, a deduction part is also provided to summarize the impact of the pertinent parameter. It is found that setting plausible values for fixed variables is essential. A particular importance should be given for the value K used for the calculation of the architecture 1oo2D. In IEC 61508, part 6, an example value for K is given 0.98. However, according to our field experience, it is possible to reach 0.9999 for K, and with this valuation, the results change for 1oo2D substantially. In the simulations, attention is drawn in case a change from 0.98 to 0.9999 for K has a great impact on the results.

Table 5. Parameters used for the analysis of $ß_d$ effect (for ß = 0.02).

λ	λ_s	λ_{sd}	λ_d	λ_{dd}	λ_{du}	DC	SFF	β	β_d (PoI)	MTTR	MRT	K	T1
1E-07	5E-08	4.95E-08	5E-08	4.95E-08	5E-10	0.99	0.995	0.02	0.002 < βd < 0.2	8	8	0.9999	1

Table 6. Parameters used for the analysis of $ß_d$ effect (for ß = 0.2).

λ	λ_s	λ_{sd}	λ_d	λ_{dd}	λ_{du}	DC	SFF	β	β_d (PoI)	MTTR	MRT	K	T1
1E-07	5E-08	4.95E-08	5E-08	4.95E-08	5E-10	0.99	0.995	0.2	0.02 < βd < 0.2	8	8	0.9999	1

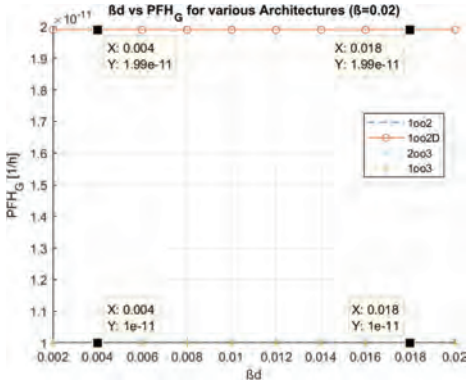


Figure 7. The influence of $ß_d$ while ß is 2%.

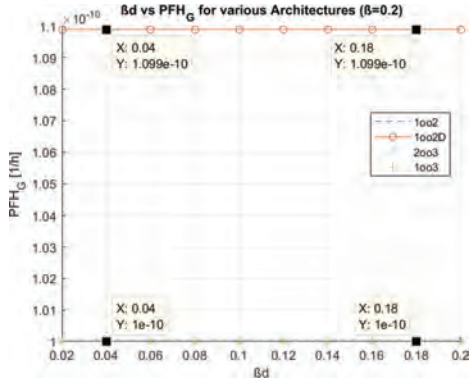


Figure 8. The influence of ßd while ß is 20%.

3.1 *The influence of λ (Table 3 and Figures 3 and 4)*

To compare 1oo2D with other architectures in a correct way, K is chosen firstly as 0.98 and then as 0.9999.

According to above parameters, the simulation re-sult is shown below.

Deduction: There is a positive correlation between λ and PFH_G for all architectures. 1oo2D is the most affected architecture. K should be selected correctly to judge the results and compare the architectures.

3.2 *The influence of ß (Table 4 and Figures 5 and 6)*

Deduction: To make an exhaustive effort for decreasing the ß ten times affects 1oo2, 2oo3 and 1oo3 almost linearly such that ten times better PFH_G can be reached. Similar is also valid for 1oo2D in case K is relative high (0.9999). If it is not the case, the effect of ß on 1oo2D is negligible.

3.3 *The influence of $ß_d$*

The influence of $ß_d$ is examined for two different ß factors to see the difference between the cases with relative high and low ß values. First, a ß value of 0.02, afterwards a ß value of 0.2 are handled.

3.3.1 *The influence of $ß_d$ for ß= 0.02 (=2%) (Table 5 and Figure 7)*

3.3.2 *The influence of $ß_d$ for ß= 0.2 (=20%) (Table 6 and Figure 8)*

Deduction: It is found that any change in the $ß_d$ fraction of ß does almost not affect the PFH_G which is very strange, because lots of effort is to be paid for increasing the diagnostics and the return of this endeavor is almost zero. Besides, when compared with DC, a measure for detecting independent failures λ, this behavior is found again too strange, since in this case, the failures are also detected which would cause a hazard if they would not be revealed. We believe this equation should be reconsidered from this perspective. According to this simulation result, we decided to scrutinize this issue in our next study in more detail.

3.4 *The influence of DC (Table 7 and Figures 9 and 10)*

Deduction: An increase of DC causes a decrease in the PFH_G for 1oo2, 2oo3 and 1oo3. In case the DC

Table 7. Parameters used for the analysis of DC.

λ	λ_s	λ_{sd}	λ_d	λ_{dd}	λ_{du}	DC (PoI)	SFF	β	β_d	MTTR	MRT	K	T1
1E-07	5E-08	4.95E-08	5E-08	v	v	0 < DC < 0.99	v	0.02	0.01	8	8	0.98/0.9999	1

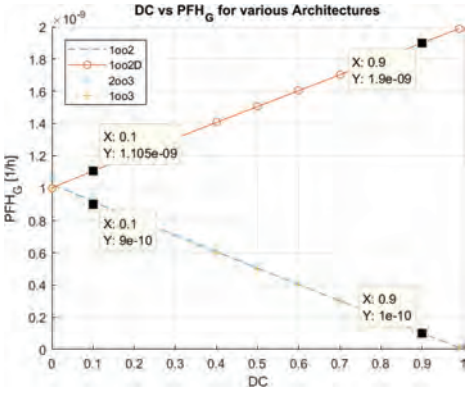


Figure 9. The influence of DC (K = 0.98 for 1oo2D).

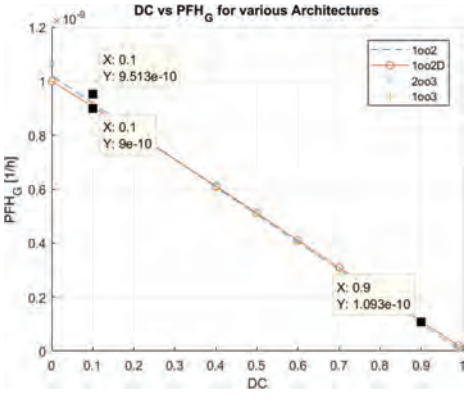


Figure 10. The influence of DC (K = 0.9999 for 1oo2D).

Table 8. Parameters used for the analysis of MTTR.

λ	λ_s	λ_{sd}	λ_d	λ_{dd}	λ_{du}	DC	SFF	β	β_d	MTTR (PoI)	MRT	K	T1
1E-07	5E-08	4.95E-08	5E-08	4.95E-08	5.00E-10	0.99	0.995	0.02	0.01	0 < MTTR < 1000	8	0.9999	1

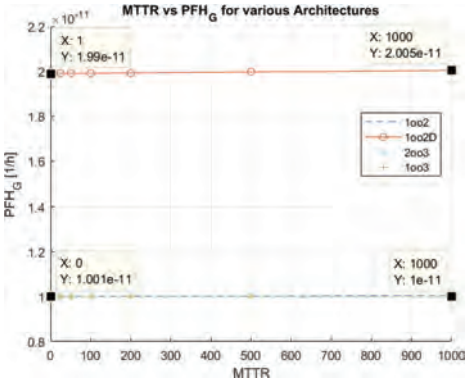


Figure 11. The influence of MTTR.

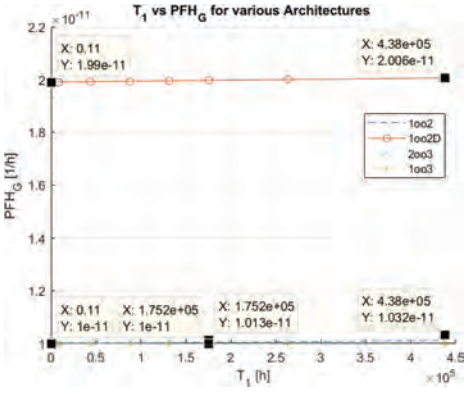


Figure 12. The influence of T_1.

Table 9. Parameters used for the analysis of T_1.

λ	λ_s	λ_{sd}	λ_d	λ_{dd}	λ_{du}	DC	SFF	β	β_d	MTTR	MRT	K	T1 (PoI)
1E-07	5E-08	4.95E-08	5E-08	4.95E-08	5.00E-10	0.99	0.995	0.02	0.01	8	8	0.9999	1/(365.24) < T1 < 50

Table 10. Parameters used for the analysis of K.

λ	λ_s	λ_{sd}	λ_d	λ_{dd}	λ_{du}	DC	SFF	β	β_d	MTTR	MRT	K (PoI)	T1
1E-07	5E-08	4.95E-08	5E-08	4.95E-08	5.00E-10	0.99	0.995	0.02	0.01	8	8	0.98 < K < 0.9999	1

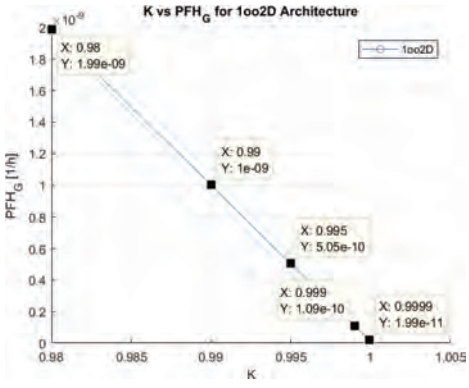


Figure 13. The influence of K.

is doubled from 0.5 to 0.99, then the PFH_G falls for these by half. For 1oo2D, the situation is observed again as strange. It is found that when K is 0.98, then there is a positive correlation between PFH_G and DC such that if DC goes up from 0.5 to 0.99, the PFH_G increases about 33%. However, the question arises here why DC would affect negatively the PFH_G. On the other hand, in case K is 0.9999, then the behavior is similar to the other architectures. It is unveiled that when K is about 0.99, then PFH_G is not influenced by DC. According to this simulation result, we decided again to scrutinize this issue in our next study in more detail.

3.5 *The influence of MTTR (Table 8 and Figure 11)*

Deduction: MTTR is usually selected not more than 8 hours. In this simulation, as the affect during the 8 hours was found as none, the time range is changed up to1000 hours. Even in this case, for all architecture, the MTTR has almost no impact on the PFH_G.

3.6 *The influence of T_1 (Table 9 and Figure 12)*

Deduction: T_1 can be selected from one day to many years according to the type of the unit. For vital computers it can be the life time such as 20 years. However, for both very short and very long time, the T_1 does almost not affect PFH_G. Besides, when doing proof tests, the effectiveness of the tests shall also be considered. Therefore, taken into account the effort needed for proof tests and also the effectiveness of such tests, the life time can be selected for proof test interval such that no proof test is required during the normal life time of the constituent.

3.7 *The influence of K (Table 10 and Figure 13)*

Deduction: The parameter K has an important impact on the PFH_G for 1oo2D. If it is increased from 0.95 to 0.99, five times better PFH_G can be obtained. And if it changes from 0.98 to 0.99, then two times, if from 0.98 to 0.999, then twenty times and if from 0.98 to 0.9999, then hundred times lower PFH_G could be obtained.

Beside the deductions mentioned above, the simulation results show surprisingly that the architectures 1oo2, 1oo3 and 2oo3 have very similar PFH_G values and similar characteristics although there are major differences for the HW/SW design and implementation of these architectures. At first glance, a hypothetical judgement would claim that 1oo3 should have much more better safety performance than 1oo2 as there exist one additional computation unit, however the results show this is not the case.

Taking into account the highlights for the architectural perspectives, we have eliminated 1oo3 architecture and decided to select 2oo3 architecture for ERTMS/ETCS on-board vital computer and 2 × 1oo2 (cold stand-by redundancy) value for CBTC on-board vital computer, since for the CBTC systems full redundancy for sensors and actors are given as requirement due to very high density traffic. We have favored 1oo2 over 1oo2D, because with this selection, the time and cost efforts needed to spend to reach for very high K values could be excluded. On the other hand, instead of developing two different systems and going through the exhaustive certification processes for two different systems, we have unified the requirements and decided to develop 2 × 1oo2D (cold stand-by redundancy) platform with relative high K value, since this effort has been evaluated as less than going through the second independent assessment process. We have chosen 1oo2D also for utilizing cross channel comparision. We have also come to conclusion for our projects that it is plausible to utilize diverse HW to reduce dependent failures and get better hazard rates while it helps to reduce systematic failures at the same time. Moreover, according to the analyses, hot stand-by design would cause additional failures in comparison to cold stand-by, its design would be much more

complex, and as cold stand-by can cover the availability requirements, too, the cold stand-by system is selected.

4 ROUTE MAP TO GET BETTER PFH_G WITH REGARDS TO THE RELIABILITY AND SAFETY PARAMETERS

As mentioned in the introduction, todays challenging technologies require a PFH_G of the computation unit about less than 1E10-12 (1/h). Regarding these though requirements, an ordinary development procedure would not be sufficient for reaching the quantitative requirements. According to the results and deductions provided in the previous section, a route map is developed for this purpose in this section. Note that ß factor is utilized for modeling dependant failures in this study, so these clauses are valid in case ß factor is used. If this is not the case, simulations are to be repeated to ensure the correctness and consistency of the results.

i. Decrease λ as much as possible; for each potential architecture. A linear relationship with a slope one is valid between λ and hazard rate. To decrease this reliability parameter, selecting higher quality parts, setting correct duty cycle, applying derating, consulting engineering experience to adjust data and using field data can be utilized. For instance, if the duty cycle is reduced from 100% to 75%, we experienced 23.47% better failure rates. Moreover, the reliability calculations can be performed using different methods such as MIL 217F, Bellcore, 217 Plus etc. Sometimes, ten times better results can be obtained as provided in the study of the RIAC (2010). At our experience, 217Plus developed by RIAC (2010) gave 5.29 times better results in comparison to MIL HDBK 217F N2 developed by USA Department of Defence (1995). If there is no special requirements, then the most effective handbook can be utilized. Besides, it has been observed for some COTS products in the industry that they are claimed to work at mobile and ground environments, however there is only one failure rate and one MTBF (Mean Time Between Failures) value provided. But, according to our calculations, if the environment changes from ground fixed to ground mobile, the failure rate increases 83.71%. As a result, if the constituent is designed to word at different environments, with different duty cycles etc., the quantitative values should be provided for these different operating conditions.
ii. Decrease common cause factor, namely ß as much as possible; for 1oo2, 1oo3, 2oo3 architectures, approximately ten times better ß results in ten times better PFH_G; however the effect for 1oo2D is very limited if K is relative low like 0.98 such that ten times better ß results in two times better PFH_G. Diversity is a very decisive character to get low dependant failures. Another approach would be trial modeling dependant failures with another method than ß factor. Fleming (1987) explains details of the dependant failure models such as binomial failure rate, multiple Greek letters, alpha factor etc. which can be alternatives to ß factor modeling.
iii. The effect of detecting the common cause failures, namely $ß_d$ has almost no effect on PFH_G for any architecture, hence do not spend much effort to detect common cause failures. On the other hand, as explained previously, we believe that some issues are overlooked for this parameter in the equations of IEC 61508 and we will investigate this issue in more detail at the next studies.
iv. Increase DC for all architectures except 1oo2D if K is less than 0.99. Similar to the previous item, the behavior of DC for 1oo2D with a K of less than 0.99 is also strange. Therefore, we recommend to bring up K at least to 0.99 to get plausible results.
v. Do not perform much effort to decrease MTTR for the safety performance as it has almost no effect on PFH_G for any architecture.
vi. Similar to the case with MTTR, do not perform much effort to decrease T_1 for the safety performance as it has almost no effect on PFH_G for any architecture.
vii. If 1oo2D is chosen, then increase K as much as possible. If K changes from 0.98 to 0.9999, then hundred times better PFH_G could be resulted.

5 CONCLUSION

In this study, the effects of calculation parameters for several architectures are simulated. Attention is drawn for surprising, interesting sides of the results while providing detailed deductions regarding the architectures for each parameter in the calculations. It has been observed that some parameters like K or DC for the architecture 1oo2D are crucial while some other parameters like MTTR or T_1 has almost no effect on the results of any architecture. The simulations can be utilized when designing safety critical systems, subsystems or equipment. The selected architectures for on-board computer to be used in ERTMS ETCS and CBTC domains are shared providing the most important judgements. An advising route map is newly created, including project

examples, to distill what kind of measures can be taken to decrease the dangerous failure rate which is at challenging low levels in today's high-tech mission critical systems like high speed trains or driverless vehicles performing vital safety functions.

REFERENCES

BS EN 50129, 2003. Railway Applications—Communications, signaling and processing systems—Safety related electronic systems for signaling.

Chen, X., G. Zhou, Y. Yang and H. Huang, 2013. A newly developed safety-critical computer system for China metro, *IEEE Trans. Intell. Transp. Syst., vol. 14, no. 2*: 709–719.

ERTMS ETCS Subset 091, 2015. Safety Requirements for the Technical Interoperability of ETCS in Levels 1 & 2.

Fleming, K.N., 1987. Parametric Models For Common Cause Failure Analysis, *Advanced Seminar on Common Cause Failure Analysis in Probabilistic Safety Assessment, 16–19 November 1987*: 159–174. Ispra, Italy.

Hokstad, P. 2005. Probability of Failure on Demand (PFD) – the formulas of IEC 61508 with focus on the 1oo2D voting ESREL, Gdansk, Polen.

IEC 61508, 2010. Functional safety of electrical/ electronic/ programmable electronic safety related systems.

IEEE 1474.1, 2004. IEEE Standard for Communications-Based Train Control (CBTC) Performance and Functional Requirements.

Ilavsky, J., Rastocny, K., Zdansky, J. 2013. Common-cause failures as major issue in safety of control systems, *Information And Safety-Related Systems, v. 11:* 86–93.

Karydasa, D.M., Brombacherb A.C., 1999. Reliability certification of programmable electronic systems, *Reliability Engineering and System Safety, v. 66*: 103–107.

King, A.G., 2014. SIL determination: Recognising and handling high demand mode scenarios, *Process Safety and Environmental Protection, v. 92*: 324–328.

Liu, Y., Rausand, M., 2016. Proof-testing strategies induced by dangerous detected failures of safety-instrumented systems, *Reliability Engineering and System Safety, v. 145:* 366–372.

Murthy, D., Rausand, M., Østeras, T., 2008. Product reliability: specification and performance. Springer, London.

Reliability Information Analysis Center (RIAC), 2010. Reliability Modeling—The RIAC Guide to Reliability Prediction, Assessment and Estimation.

Smith, S., Gruhn, P., 1995. The "primary integrity parameters" - De*sign parameters for safety systems, ISA Transactions, v. 34*: 311–318.

USA Department of Defense, 1995. MIL HDBK 217 FN2 Military Handbook Reliability Prediction of Electronic Equipment.

Zhang, T. Long, W., Sato, Y., 2003. Availability of systems with self-diagnostic components—applying Markov model to IEC 61508–6, *Reliability Engineering and System Safety, v. 80:* 133–141.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Equal load-sharing models of cascades in interdependent network infrastructures

A. Scala
Istituto Sistemi Complessi CNR, Università "Sapienza" Roma, Rome, Italy

P.G. De Sanctis Lucentini
Gubkin Russian State University of Oil and Gas, Moskow, Russia

G. D'Agostino
ENEA, CR Casaccia, Rome, Italy

ABSTRACT: We extend the equal load-sharing model of cascades to investigate the abrupt breakdown behavior of coupled distribution grids. In particular, we mimic the effects of the ever-increasing customer demand in a foreseen scenario where energy hubs interconnect the different energy vectors. In the load growth scenario we find evidence of first order transitions (i.e. abrupt breakdowns of the system) due to the long-range nature of the flows. Our results indicate that the foreseen increase in the couplings between the grids has two competing effects: on the one hand, it increases the safety region where grids can operate without withstanding systemic failures; on the other hand, it increases the possibility of total failure of all the interconnected systems at once.

1 INTRODUCTION

In this paper we will consider a simple mean-field model of cascading that applies both to single and to interdependent networks. To highlight the possibility of emergent behavior, we will first abstract *PNI*s in order to understand the basic mechanisms that could drive systemic failures; in particular, we will consider finite capacity networks where a commodity (a scalar quantity) is produced at source nodes, consumed at load nodes and distributed as a Kirchoff flow (e.g. fluxes are conserved) and introduce a simplified model that is amenable of a self-consistent analytic solution. Subsequently, we will extended such model to the case of several coupled networks and study the cascading behavior of such a model under increasing stress (i.e. flow magnitudes).

2 MODEL

Let's consider a weighted network $G=(V,E,\boldsymbol{c})$ where $V=\{1\le i\le|V|\}$ is the node set, $E\subseteq V\times V$ is the set of edges and $\boldsymbol{c}=\{c_{(i,j)}\}$ is the vector characterizing the capacities of the edges (i,j). We associate the nodes a vector $\mathbf{s}=\{s_i\}$ that characterize the production ($s_i>0$) or the consumption ($s_i<0$) of a commodity. We further assume that there are no losses in the network (i.e. $\sum_i s_i = 0$); hence, the total load on the network is

$$L=\sum_{i:s_i>0} s_i$$

The distribution of the commodity is described by the fluxes $\boldsymbol{f}=\{f_{(i,j)}\}$ on the edges $(i,j)\in\mathcal{E}$ and is supposed to respect Kirchoff equations, i.e.

$$\sum_j f_{(i,j)} = s_i \tag{1}$$

The relation among fluxes and demand/load is described by constitutive equations

$$\boldsymbol{f}=F(\boldsymbol{s},G) \tag{2}$$

where in general Eq. (2) is non-linear but satisfies Eq. (1).

More often, constitutive equation rely on a relation among fluxes on the line and the values of a physical field ϕ defined on the nodes. As an example, in networks transporting fluid commodities (e.g. gas, water) the constitutive equations take the form

$$f^{\gamma}_{(i,j)} \propto \phi_i - \phi_j \tag{3}$$

where $\gamma \sim 2$ and ϕ_i is the pressure at the *ith* node. In the case of DC currents, the relation is linear ($\gamma = 1$), $f_{(i,j)}$ is the electric current and ϕ_i the voltage; the same linear equations hold in the case of steady-state DC power flow in electric networks, where now $f_{(i,j)}$ is the power flow and ϕ_i the phase angle. For further details, see (Scala 2017) and references therein.

3 METHODS

Cascading failures represent a critical vulnerability in our world where network infrastructures grow both in complexity and interdependencies (D'Agostino and Scala 2014, DAgostino and Scala 2015). When detailed information about the topology are at hand, centrality measures can be used to identify vulnerable subsets of the infrastructural system (Scala et al. 2016). However, in this section we will concentrate on cascading due to the violation of the link capacities while disregarding the effects of shocks due to strong transients.

For the flow networks described in the previous section, the finite capacity $c_{(i,j)}$ of a link (i, j) constrains the maximum flux on such a link

$$\left| f_{(i,j)} \right| < c_{(i,j)}$$

while above such flux, the link will cease functioning. As an example, power lines are tripped (disconnected) when power flows go beyond a certain threshold. Since flows will redistribute after a link failure, it could happen that other lines get above their flow threshold and hence consequently fail, eventually leading to a cascade of failures. A typical algorithm to calculate the consequences of an initial set of line failures $\mathcal{F}^0 = \{(ij)\ failed\}$ is the alg.(1). Here $F(\boldsymbol{p}, G \,|\, \mathcal{F})$ calculates the flows subject to the constrains that flows are zero in the failure set of edges $(i, j) \in \mathcal{F}$.

Algorithm 1 Network cascading

```
Set initial failures F^0
t ← 0
repeat
  t ← t + 1
  Calculate flows f^t ← F(s, G|F^{t-1})
  Calculate new failures ΔF^t ← {(ij) : |f_ij^t| > c_ij}
  F^t ← F^{t-1} ∪ ΔF^t
until ΔF^t ≡ ∅
```

To develop a general model that helps us understanding the class of failures that can affect Kirchoff-like flow networks, let's start from rewriting Eq. (1) in matrix form

$$\mathcal{B}^T \boldsymbol{f} = \boldsymbol{s} \tag{4}$$

using the incidence matrix B that associates to each link $(i.j)$ its nodes i and j and vice-versa. B is an $|v| \cdot |\varepsilon|$ matrix where each column corresponds to an edge (i, j); itscolumns are zero-sum and the only two non-zero elements have modulus 1 and are on the *ith* and on the *jth* row.

The matrix B is related to the Laplacian B^TB of the system; in particular, it shares the same right eigenvalues and the same spectrum (up to a squaring operations); hence, it is a long-range operator since perturbation on a node of the system can be reflected on nodes far away on the network (Pahwa et al. 2014).

Due to the long range nature of Kirchoff's equations, to understand the qualitative behavior of such networks we can resort to a mean field model of flow networks where one assumes that when a link fails, its flow is re-distributed equally among all other links. Such model, introduced in (Pahwa et al. 2014), is akin to the fiber-bundle model (Peirce 1926, Daniels 1945) and has been considered in more details in (Scala and Lucentini 2016, Yagan 2015) for the case of a single system. In this model, each time lines trip, flows are recalculated, the lines above their threshold trip again, their flows would be re-distributed and so on, up to convergence; recalling that L is the total load of the system and assuming the each link (i, j) has an initial flux $f = L/|E|$, we can describe such a model by alg.(2).

Algorithm 2 Mean Field cascading

```
t ← 0
F^t ← 0 initial number of failed links
repeat
  t ← t + 1
  M ← |E| − F^{t-1} number of working links
  l ← L/M average flux on the working links
  F^t ← |{(ij) : l > c_(ij)}|
until F^t = F^{t-1}
```

Such algorithm can be cast in the form of a single equation in the case where the system is composed by a large number of elements with capacity c. In fact, in such limit we can describe the links' population by the probability function $p(c)$ of their capacities. Indicating with $M = |E|$ the initial number of links, we see that if we apply an overall load L to the system, all the links will be initially subject to a flow $l^0 = L/M$. Thus, a fraction of links $f^1 = \int_0^{L/M} p(c)dc$ would immediately fail, since their thresholds are lower than the flux l they should sustain. After the first stage of a cascade, there will be $M^1 = (1 - f^1)M$ surviving links and the new load per link is $l^1 = L/M^1$. The following cascade's stages follow analogously; we can thus write the mean field equations for the $(t + 1)^{th}$ stage of the cascade:

$$f^{t+1} = P\left(\frac{l}{1-f^t}\right) \tag{5}$$

where $l = L/M$ is the initial load per link and $P(x) = \int_0^x p(c)dc$ is the cumulative distribution function of link capacities; the initial conditions are $f^{t=0} = 0$. The fix-point f^* of Eq. (5) satisfies the equation

$$f^* = P\left(\frac{l}{1-f^*}\right) \tag{6}$$

and represents the total fraction of links broken at the end of the cascading stages (Pahwa et al. 2014).

The behavior of f^* depends on the functional form of $p(c)$. In particular, by defining $\pi(c) = 1 - P(c)$ and $x = l^{-1}(1-f)$, we have that

$$f = \int_0^{1/x} p(c)dc = 1 - \pi\left(\frac{1}{x}\right)$$

and can rewrite Eq. (5) as

$$lx^{t+1} = \pi\left(\frac{1}{x^t}\right)$$

(see Fig. (1)). This equation has a trivial fix-point $x^* = 0$ (representing a total breakdown of the system) since $\pi(\infty) = 0$. Such fix-point is unstable for $l \to 0$ and becomes stable for $l > \partial_x \pi(x^{-1})|_{x\to 0}$. We notice that if $P(c)$ does not change convexity

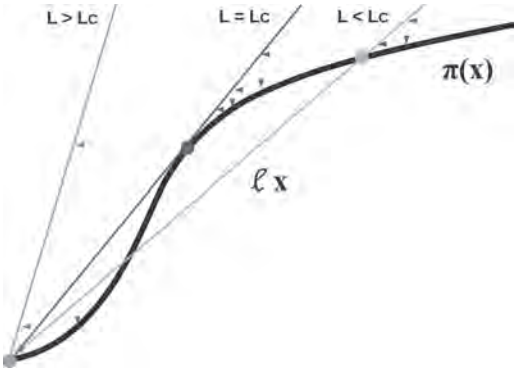


Figure 1. Graphical solution of the fixpoint equation. For $L = L_c$, there is only one solution corresponding to a critical point. For $L > L_c$, the stable fixpoint x^* corresponds to a small fraction of broken links f^*; in the limit of $l \to 0$, $f^* = 0$. On the other hand, for $L < L_c$ the only stable solution is $x^* = 0$, corresponding to the situation $f^* = 1$ where all the links are broken and hence the whole system has failed.

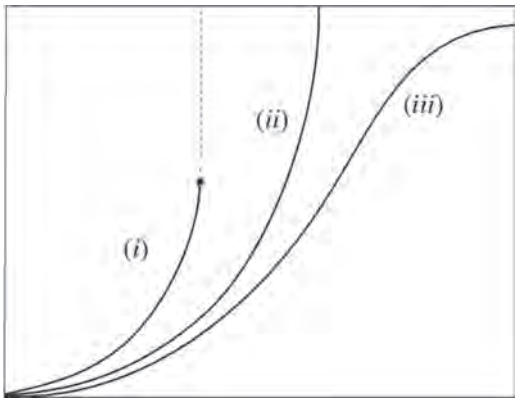


Figure 2. The behavior of the fix-point x^* depends on the tail of the distribution of link capacities $p(C)$ and is known to present a first order transition for a wide family of curves; in particular, when the cumulative distribution function $P(C)$ goes to zero faster than C^{-2}, we are in case *(iii)* where the transition is discontinuous, with a jump at a critical value L_c and a divergence of $dM/dL \sim (L - L_c)^{-1/2}$.

(i.e. has no bumps) and the transition is first order, the system will breakdown directly to the total collapsed state $f = 1$.

In general, the behavior of the fix-point x^* depends on the tail of the distribution $p(c)$ and is known to present a first order transition for a wide family of curves (da Silveira 1998) (see Fig. (2)); in particular, the fixpoint behavior is dictated by the behavior of $\frac{d\pi}{dC} = \frac{p(C)}{C^2}$ for $C \to \infty$, i.e.

- for $p(C) \sim C^{-\gamma}$ with $1 < \gamma < 2$, no transition occurs
- for $p(C) \sim C^{-2}$, the transition is continuous but there is a critical point L_c with a divergence of $dM/dL \sim (L - L_c)^{-1/2}$
- for $C^2 p(C) \to 0$ when $C \to \infty$, the transition is discontinuous, with a jump at a critical value L_c and a divergence of $dM/dL \sim (L - L_c)^{-1/2}$.

Depending on the functional form of $p(c)$, Eq. (6) could sometimes be solved analytically. Otherwise, the fix-point of Eq. (6) can be solved numerically either by iterating the Eq. (5) or by finding the zeros of Eq. (6) by Newton-Raphson iterations.

4 RESULTS

Commodities are defined substitutable when they can be used for the same aim; when commodities are substitutable, they can expressed in the same units. An example of such commodities are electricity and gas can, since be both used for domestic heating. Hence, an increase on the cost of the gas (as the one that has been recently experienced by

Ukraine) could provoke stress on the electric network of the country since most customer will possibly switch to the cheaper energy vector. To take account for such effects, we will extend the model described by Eq. (5) to the case of several coupled systems that transport substitutable commodities.

We will consider n coupled systems assuming that when a system a is subject to some failures, it sheds a fraction $T_{a\to b}$ of such the flow increase due to such failures on system b. In other words, upon failure system a decreases its load by a quantity $l_a f_a \sum_{b\neq a} T_{a\to b}$ and increases the load of all systems $b \neq a$ by $l_a f_a T_{a\to b}$. Thus, the n coupled systems are described by a set of n equations of the form of Eq. (5)

$$f_a^{t+1} = P_a\left(\frac{\tilde{l}_a^t}{1-f_a^t}\right) \tag{7}$$

$\tilde{l}_a^t$ is the load per link experimented by system a at the t^{th} stage of the cascade and $P_a(x) = \int_0^x p_a(x)\,dx$ is the cumulative of the probability distribution function $p_a(x)$ for the capacities of the a^{th} system. Equations (7) are not independent, since the systems' coupling is reflected by the dependence of $\tilde{l}_a^t$ on the fractions f_b^t of failed links in all the other systems, i.e.

$$\begin{aligned}\tilde{l}_a^t &= l_a\left(1-f_a^t\sum_b T_{a\to b}\right)+\sum_b T_{b\to a} l_b f_b^t \\ &= l_a + \sum_b \mathcal{L}_{ab} l_b f_b^t\end{aligned} \tag{8}$$

where $\mathcal{L}_{ab} = (1-\delta_{ab})T_{b\to a} + \delta_{ab}\sum_b T_{a\to b}$ has again the form of a Laplacian operator. Thus, the full equations for n coupled systems are

$$f_a^{t+1} = P_a\left(\frac{l_a + \sum_b \mathcal{L}_{ab} l_b f_b^t}{1-f_a^t}\right) \tag{9}$$

For simplicity, we will consider the case of two identical system with a uniform distribution of link capacities and solve the fix-point of Eq. (7) numerically. We show in Fig. (3) the cascading behavior of two coupled systems; we observe that—as in the single system case—transitions are in the form of abrupt jumps, i.e. are first order. Let's rewrite Eq. (9) in the case of symmetric couplings $T_{1\to 2} = T_{2\to 1} = 1$ and same probability distribution for the capacities

$$\begin{cases} f_1^{t+1} = P\left(\dfrac{l_1}{1-f_1^t}\left[1-T\left(f_1-\dfrac{l_2}{l_1}f_2\right)\right]\right) \\ f_2^{t+1} = P\left(\dfrac{l_1}{1-f_2^t}\left[1-T\left(f_2-\dfrac{l_1}{l_2}f_1\right)\right]\right) \end{cases} \tag{10}$$

If the two systems described by Eq. (10) are stressed at the same pace (i.e. $l_1 = l_2 = l/2$), we get the case

$$\begin{cases} f_1^{t+1} = P\left(\dfrac{l}{1-f_1^t}\left[1-T\Delta f_{12}\right]\right) \\ f_2^{t+1} = P\left(\dfrac{l}{1-f_2^t}\left[1+T\Delta f_{12}\right]\right) \end{cases}$$

from the symmetric solution $\Delta f_{12} = 0$ we see that the breakdown of both systems happen at the same critical load as the uncoupled systems. Such situation is shown in the left panel of Fig. (3).

In the general, only one of the systems will be the first one to break down (i.e. the fraction of broken links jumps to $f = 1$): correspondingly, also the other systems will experience a jump in the number of broken links. Let's consider the symmetric case described by equations (10) and suppose that $l_1 > l_2$, so that system 1 is the first to breakdown (i.e. $f_1 = 1$); hence, the equation for the fix-point of the second system becomes

$$f_2^* = P\left(\frac{l}{1-f_2^*}\left[1+T\left(1-f_2^*\right)\right]\right) = P\left(\frac{l^*}{1-f_2^*}\right)$$

i.e. the system behaves like a single system starting with a renormalized load $l^* > 1$. Thus, if $l^* < l_c$ the

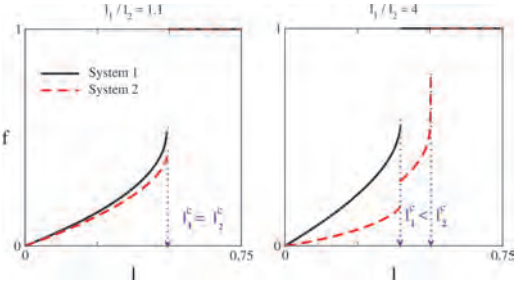


Figure 3. Behavior of the number of failed nodes respect to the total stress $l = l_1 + l_2$ of the systems. For simplicity, we present the case of two identical systems with a flat distribution of link capacities and symmetric couplings $T_{1\to 2} = T_{2\to 1} = 0.5$. We show the result of increasing the total stress l in the two systems along the lines $l_1/l_2 = const$. **Left panel:** we show the case $l_1/l_2 = 1.1$ where both systems are subject to a similar stress while increasing l. In such case both system break down together at the same critical load $l_1^c = l_2^c$; in the region $l > l_1^c = l_2^c$ both systems are failed. **Right panel:** we show the case $l_1/l_2 = 4$ where when increasing l systems 1 is more stressed than system 2. In this case, the break down of system 1 at the critical load l_1^c induces a jump in the number of failures system 2, but system 2 is still able to sustain stress and will break down only at higher values of l. Respect to the $l_1 \sim l_2$ case, there is now a region $l_1^c < l < l_2^c$ where only system 1 is failed.

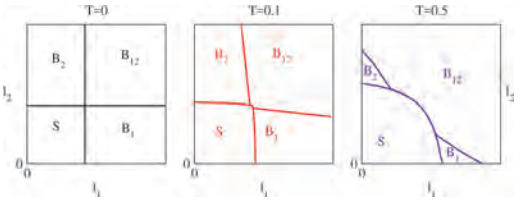


Figure 4. Phase diagrams of two identical coupled systems with symmetric interactions $\left(T_{1\to 2} = T_{2\to 1} = T\right)$. The plane of initial loads l_1 and l_2 is separated in four different regions by critical transition lines. The labels B_i (i = 1,2) mark the areas where only system i suffers systemic cascades ($f_i = 1,\ f_{j\neq i} < 1$), while the label B_{12} marks the area where both system suffer system wide cascades $\left(f_1 = f_2 = 1\right)$. The label S marks the (safe) area near the origin where no systemic cascades occur. **Left panel:** the case $T = 0$ corresponds to two uncoupled systems: thus, each system suffers systemic failure at $l_i > l^c$ (where l^c is the critical load for an isolated system); both systems are failed in the B_{12} area corresponding to the quadrant $(l_1 > l^c, l_2 > l^c)$. **Central panel, right panel:** when couplings are introduced, each system is able to discharge stress on the other one and the area S where both systems are safe increases. On the other hand, the area B_{12} where *both* systems are failed increases.

critical value of Eq. (5), system 2 will break down at higher values of the stress. Such situation is shown in the right panel of Fig. (3).

In Fig. (4) we show the full phase diagrams of two coupled systems while varying the coupling among them. As discussed before, due to the symmetry all the systems have a transition point $\left(l^c, l^c\right)$ along the $l_1 = l_2$ line, where l^c is the critical load of the single system. According to the initial loads, we can distinguish an area S near the origin where the system is safe and three separate cascade regimes: B_1, B_2 where either system 1 or 2 fails, and B_{12} where both systems fail. We notice that, by increasing the coupling among the systems, both the area S where the two systems are safe and the area B_{12} where they fail together grow; accordingly, the areas B_i where only one system fails shrink.

5 DISCUSSION

In this paper we have introduced a model for cascade failures due to the redistribution of flows upon overload of link capacities. For such a model, we have developed a mean field approximation both for the case of a single network and for the case of coupled networks. Our model is inspired to a possible configuration for future power systems where network nodes the so-called energy hubs (Geidl et al. 2007), i.e. points where several energy vectors converge and where energy demand/supply can be satisfied converting one kind of in another. Hubs condition, transform and deliver energy in order to cover consumer needs (Favre-Perrod 2005). In such configurations, one can alleviate the stress on a network by using the flows of the other energy vectors; on the other hand, transferring loads from a network to the other can trigger cascades that can eventually backfire.

By analyzing the case of two coupled systems and by varying the strength of the interactions among them, we have shown that at low stresses coupling has a beneficial effect since some of the loads are shed to the other systems, thus postponing the occurrence of cascading failures. On the other hand, with the introduction of couplings the region where not only one system fails but both systems fail together also increases. The higher the couplings, the more the two systems behave like a single one and the area where only a system is failed shrinks. Notice that our model in the present form does not apply to islanding strategies in power systems, where some sub-networks can even enhance their reliability upon failure of part of the remaining system (Mureddu et al. 2016); such subject will deserve further investigations.

6 CONCLUSIONS

It is worth noting that while fault propagation models do predict a general lowering of the threshold for coupled systems (Wang et al. 2013), in the present model a beneficial effect due to the existence of the interdependent networks is observed for small enough overloads, while the expected cascading effects take place only for large initial disturbances. This picture is consistent with the observed phenomena for interdependent Electric Systems. Moreover the existence of interlinks among different networks may increase their synchronization capabilities capabilities (Martin-Hernandez et al. 2014).

REFERENCES

DAgostino, G. & A. Scala (2015). Systemic interdependences. In W.S. Bainbridge and M.C. Roco (Eds.), *Handbook of Science and Technology Convergence*, pp. 1–11. Springer International Publishing.

Daniels, H.E. (1945). The statistical theory of the strength of bundles of threads. i. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 183*(995), 405–435.

da Silveira, R. (1998, Apr). Comment on "tricritical behavior in rupture induced by disorder". *Phys. Rev. Lett.*, 3157–3157.

D'Agostino, G. & A. Scala (Eds.) (2014). *Networks of Networks: The Last Frontier of Complexity*. Understanding Complex Systems. Springer International Publishing.

Favre-Perrod, P. (2005, July). A vision of future energy networks. In *Power Engineering Society Inaugural Conference and Exposition in Africa, 2005 IEEE*, pp. 13–17.

Geidl, M., G. Koeppel, P. Favre-Perrod, B. Klockl, G. Andersson, & K. Frohlich (2007). Energy hubs for the future. *IEEE Power & Energy Magazine 5*(1), 24–30.

Martin-Hernandez, J., H. Wang, P.V. Mieghem, & G. D'Agostino (2014). Algebraic connectivity of interdependent networks. *Physica A: Statistical Mechanics and its Applications 404*(0), 92–105.

Mureddu, M., G. Caldarelli, A. Damiano, A. Scala, & H. Meyer-Ortmanns (2016, October). Islanding the power grid on the transmission level: less connections for more security. *Scientific Reports*, 34797–.

Pahwa, S., C. Scoglio, & A. Scala (2014, January). Abruptness of cascade failures in power grids. *Sci. Rep. 4, –*.

Peirce, F. (1926). Tensile tests for cotton yarns, part "v": the weakest link theorems on strength of long and composite specimens. *Journal of Textile Institute*, T355–T368.

Scala, A. & P.G.D.S. Lucentini (2016). The equal load-sharing model of cascade failures in power grids. *Physica A: Statistical Mechanics and its Applications*, 737–742.

Scala, A. (2017). *Complex Networks and Infrastructural Grids, Volume 5 of Order, Disorder and Criticality Advanced Problems of Phase Transition Theory*, Chapter 7, pp. 341–396. WORLD SCIENTIFIC.

Scala, A., V. Zlatić, G. Caldarelli, & G. D'Agostino (2016). Mitigating cascades in sandpile models: an immunization strategy for systemic risk? *The European Physical Journal Special Topics* (10), 2017–2023.

Wang, H., Q. Li, G. D'Agostino, S. Havlin, H.E. Stanley, & P. Van Mieghem (2013, Aug). Effect of the interconnected network structure on the epidemic threshold. *Phys. Rev. E*, 022801.

Yagan, O. (2015, Jun). Robustness of power systems under a democratic-fiber-bundle-like model. *Phys. Rev. E 91*, 062811.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Optimizing terminal logistics and dimensioning

S.L. Isaksen & T. Lilleheier
Safetec, Oslo, Norway

N.J. Edwin
Safetec, Trondheim, Norway

ABSTRACT: Design and correct dimensioning of oil and gas terminal facilities is often a challenge due to a variety of operational uncertainties as well as the volatility in supply and demand. Various factors come into play to ensure smooth uninterrupted operation, as well as optimal and effective resource utilization. This paper discusses the challenges of optimization and presents an integrated approach to optimize terminal logistics and dimensioning, where parts of the objective function is solved by discrete event simulation. A case study of the new Veidnes terminal to be commissioned in the north of Norway, is also presented. This terminal shall serve as a central hub for export of oil from the already producing Goliat field and the upcoming offshore facilities in the Barents Sea, including Johan Castberg as well as the later producers Alta Gotha and Wisting. A variety of factors including weather disturbances, production profiles, shuttle tankers, jetties, etc. are included in the analysis to help provide a decision-basis on the number and size of tanks required at the terminal.

1 INTRODUCTION

Oil produced offshore is usually transported to terminals onshore either via pipelines or by oil tankers. These terminals typically have facilities for storage and export of oil, and some also have facilities for further treatment of oil, such as fractionation. Export of oil from the terminal can be done via pipeline, ship, train or trucks. In practice, an oil terminal represents a buffer in the middle of the supply chain, reducing the consequences of disturbances in the transportation.

During the design process of such terminals, important decisions are made. One such decision is related to the dimensioning, where the difference between good and bad decisions can be worth tens or even hundreds of millions of dollars. Dimensioning of these terminals is difficult as the required capacity may vary due to high volatility in supply and demand. Tank storage space is expensive in terms of CAPEX (Capital Expenditures) but too little storage can turn out to be much more expensive in terms of deferred production and delays in the supply chain. This can lead to the design of an extra buffer, which in practice is expensive overcapacity. Also, tank suppliers may have a tendency to recommend more volume than necessary.

Despite the expense of bad decisions in design, it is not uncommon to base such decisions on experience, general knowledge, and gut feeling. Experience is certainly always valuable, but is not sufficient since each terminal is unique. Some sort of simplified calculation is usually involved as well in the design or planning phase. However, a common misconception in the industry is that complexities and uncertainties make the problem impossible to fully approach mathematically.

This paper discusses how a mathematical model may be used to approach the optimization problem. It discusses an integrated approach involving an objective function and discrete event simulation. The paper also describes a specific case to demonstrate the application of the mathematical approach for a Norwegian oil terminal.

2 PROBLEM AND CASE DESCRIPTION

2.1 *The optimization problem*

Mathematical optimization constitutes a large area of applied mathematics. It includes finding the 'best' value of some objective function within a given domain. The 'best' value is depending on the context but will typically be the one which maximizes or minimizes the value of the objective function.

Terminal dimensioning can be viewed as such an optimization problem, where the aim is to minimize overall costs over a defined period. Both CAPEX (Capital Expenditures) and OPEX (Operational Expenditures) are important to include. Another

important aspect is the revenue from exported oil. Lost revenue due to deferred production as a consequence of shortage in storage capacity, can be regarded as an expense. Alternatively, the objective function can contain all revenue from exported oil, subtracting the CAPEX and OPEX, in which case the function should be maximized. In that case, the objective function would in a simplified form look something like this.

$$\max_{\bar{x}\in(0,\infty)} f(\bar{x}) - g(\bar{x}) - h(\bar{x}) \quad (1)$$

where f = revenue from exported oil; g = CAPEX; and h = OPEX. $\bar{x}$ is a vector of elements such as number and size of offshore tanks, number and size of shuttle tankers, pump capacities, etc.

CAPEX is usually the easiest function of the three, consisting of sums of unit costs multiplied by number of units. But it may be hard to express the cost as a continuous function, since e.g. tank producers offer a fixed selection of designs with a certain size. Uncertainties may also be present at an early stage, but these are quite small compared to the other parts of the objective function.

OPEX is more complex as these are running costs over a long period of time, i.e. the expected lifetime of the terminal, which in itself is a large uncertainty factor. On the other hand, OPEX may not be very largely influenced by the size of many of the elements in $\bar{x}$. I.e. maintenance of a 500,000 barrel tank is not necessarily much more expensive than maintenance of a 400,000 barrel tank. But for both, the Life Cycle Cost (LCC) is associated with high uncertainty.

The revenue from exported oil or the revenue loss from deferred export due to shortage in storage capacity, is the most difficult parameter to assess. There are many factors associated with large uncertainties. In order to understand what could cause deferral in oil production or export, we must understand the dynamics of the whole supply chain. These dynamics contain a lot of time-dependent and stochastic elements and are difficult to represent with equations.

Consider the following example: An oil tanker is on its way to an offshore installation to load oil. The approximate loading operation and sailing times are known. However, it turns out that the waves are too high for a safe loading operation and the tanker must wait. If we are unlucky, the bad weather period lasts for so long that the oil storage tank on the platform gets full and there is a forced production shutdown. By the time the weather improves, the tanker loads and heads to the terminal. This delay may cause a delay in the export, or export of off-spec oil quality because the export tanker was depending on the load from the shuttle tanker getting there in time.

One could say that the weather was the cause of two financial loss factors; the offshore disruption and the delayed or off-spec export load. However, this could have been avoided with e.g. larger tanks offshore and onshore. Weather is just one of several factors influencing the oil export, and as the example illustrates, the dynamics or timing is an important element. To express revenue as part of the objective function is thus a very challenging task. As will be discussed in the next chapter, discrete event simulation is a suitable way to solve the problem, since the dynamics are well captured and the objective function to a large part is reduced to drawing random numbers from given distributions.

2.2 *Case description*

2.2.1 *Veidnes terminal development*

Statoil are building a new terminal in Veidnes in the northern parts of Norway, which will receive oil from various offshore installations in the Barents Sea by a fleet of winterized shuttle tankers. Oil stored at Veidnes will be shipped by export tankers to various destinations in continental Europe. The sizing and number of tanks located at Veidnes is important, since overcapacity is expensive and insufficient storage may lead to expensive losses. Figure 1 contains an overview of the Veidnes terminal with associated offshore fields.

Optimal sizing and number of tanks is influenced by the sizing and number of shuttle tankers and export tankers, as well as several other parameters. In addition, exported oil is required to have a certain quality, measured in API (American Petroleum Institute) gravity value. The storage and mixing of different oil qualities, as well as requirements for a certain API value of exported oil, also influence the optimal size and number of tanks.

One of the biggest challenges is the volatile nature of supply. In the case of Veidnes, one of the offshore fields, Goliat, has just started producing. The plan for Veidnes is to start operation when also Johan Castberg, with larger oil reserves than Goliat, starts producing. Later, additional fields,

Figure 1. Location of the Veidnes terminal and associated oil fields.

Alta Gotha and Wisting, will start production and the winterized shuttle tankers will transport oil from four different fields to Veidnes.

In addition to the four fields starting production at different times, their production profiles are dynamic, and their forecasts are associated with a high degree of uncertainty. There are also other reserves in the area, which could be relevant for oil production and shipping to Veidnes at some point in the future. However, these factors are not considered in the first stage of designing the terminal.

2.2.2 *Phased development*

When the Wisting and Alta/Gotha fields begin production, the terminal needs to be expanded. For now, we consider the three phases listed in Table 1.

In a scenario like this, there are two main options. The first is to account for the highest supply somewhere in phase 3 (when the sum of all four fields reaches peak production) and dimension the terminal capacity accordingly from the start. In this case, there is excess terminal capacity in the first two phases. The second option is to build a smaller terminal at first and then expand capacity in the later phases. Which of the options is best also requires analysis, since it can be argued that building a large terminal from the start is cheaper than doing it in steps. But there is also a potential NPV (Net Present Value) gain in delaying some of the construction for later. This evaluation was already made on a high level, with the decision to construct in two phases, one for the first two phases and then add additional tanks and another jetty for the third phase. Since some decisions had been made already, the case study was both a verification, as well as an optimization study.

In other words, for larger planned increases in supply, such as the start-up of a new producing field or plateau production of fields, the existing structure can be increased. For smaller and highly uncertain increases in production profile, it is not feasible to add to existing capacity.

For predicted reductions in supply such as end of plateau production, reductions in capacity are of course not feasible either, as it hardly affects the costs, since the investment has already been made.

Table 1. Phases considered for the Veidnes terminal project.

Phase	Period	Involved fields
1	2022 Q4–2023 Q4	Goliat, Johan Castberg
2	2023 Q4–2024 Q3	Goliat, Johan Castberg, Alta/Gotha
3	2024 Q3–2030 Q4	Goliat, Johan Castberg, Alta/Gotha, Wisting

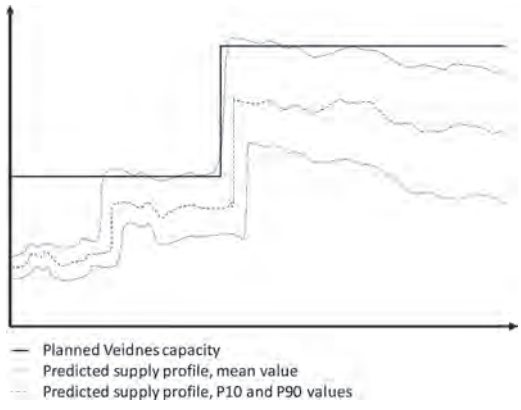


Figure 2. Illustration of supply, uncertainties and phased development.

Figure 2 illustrates the main elements as discussed. The planned capacity should account for the peak supply in all phases. But planned capacity must take uncertainties in supply into account. The optimization approach consists of finding the right balance between minimizing the risk and associated financial loss of too little capacity and the investment in sufficient capacity, even though overcapacity is inevitable in certain periods.

3 APPROACH

3.1 *Discrete event simulation*

As mentioned in Section 2.1, the dynamic nature of the terminal activities and resulting oil export, presents a significant challenge to express the $f(\bar{x})$ function from equation 1. Monte Carlo simulation, however, is suitable for dealing with dynamics. The progression of time is simulated and stochastic events can be dealt with by conditions implemented in the simulation algorithm.

Monte Carlo simulation has the advantage of simplifying a mathematical problem significantly. If a stochastic element is represented by a probability distribution function, the simulator draws just one number at a time from this distribution, making calculations much easier. But if this process is repeated a sufficient number of times, the numerical result still approaches the theoretical result. Hence, the challenge shifts from mathematical complexity to computational power in the case of Monte Carlo simulation. This challenge becomes smaller as the computational power in general keeps increasing.

Discrete event simulation is a branch of Monte Carlo simulation which is suitable for solving problems like these, because they are characterized by

events occurring at discrete points in time. Events are generated at simulation start or initialization. These are exemplified by the arrival of a shuttle tanker, the tank getting full or empty based on current loading/unloading rate, the shift from waves under the critical threshold value to above the critical threshold value, etc. These events are sorted in an event queue, according to time of occurrence. Instead of moving forward in time in pre-defined small incremental steps, which is the traditional Monte Carlo approach, the simulator skips to the next event in the queue. At the event, calculations are done and new events drawn from distributions, which might change the existing event queue.

In a case like this, there are elements of both discrete events and continuous flow. When e.g. a shuttle tanker offloads at Veidnes, there is a steady flow into the tank and the tank level increase is continuous in time. However, it is not necessary to monitor the tank level continuously. With a discrete event simulation model, the next event can be the time when the tank is full or the shuttle tanker is empty. If the flow rate varies in time, e.g. full flow at the start of the offloading operation and then slowing down when the tank approaches full, the point in time changing to a lower flow rate can be regarded as an event.

In models with both elements of continuous flow and discrete events, it is generally more efficient to create a discrete event simulation model and include the continuous flow elements, than vice versa.

The difficult part of the revenue function is twofold. It is expressed as a product of the amount of exported oil and the oil price. The problem with the oil price is simply the fact that it is highly uncertain, looking over a future 10-year period. It should be discounted as well and the discount rate adds to the uncertainty.

The difficulties about the amount of exported oil have been explained in Section 2.1. It is a time-dependent function which is difficult to express due to many parameters and high complexity. Hence, this part is covered by the simulation model.

In order to assess the amount of exported oil, it was important to consider the supply chain and evaluate all significant influencing factors. The simulation model includes the following parameters:

- Production profiles
- Capacity of pumps, tanks and tankers
- Number of tanks, jetties, and tankers
- Oil quality (API grade)
- Weather
- Criteria for critical wave heights and oil quality blend
- Travel speeds and distances
- Berthing and de-berthing times
- Export strategy (blended or segregated oil quality)

Some of these parameters are dynamic and change during the simulation. For instance, the number of jetties increase from one to two at a certain point in time, when more offshore fields start producing.

3.2 *Simulation tool and algorithms*

The simulation software, ExtendSim, was used to create and run the model of the Veidnes terminal. It belongs to the class of generic simulation tools, not specifically built for a particular industry or simulation paradigm. These are typically very flexible and can be used for almost any modelling of production, processing, logistics, resource management, etc. or combinations of these.

ExtendSim consists of different blocks, which perform various functions, e.g. drawing a random number from a distribution, creating an item, assigning an attribute to an item, performing an activity, acting as a storage tank, and even implementing a custom-made function. These existing blocks cover the necessary functionality for the basics of the model, i.e. producing oil offshore, loading shuttle tankers, transporting to Veidnes and offloading, storage at Veidnes, loading export tankers and transport it to various destinations. It also enables the specification of oil quality requirements and the inclusion of a weather model and its interaction with loading and unloading operations.

Despite much existing functionality, the custom-made function block has been used extensively for implementing algorithms for specific operational rules. For instance, rules must be implemented to determine which offshore field the next available shuttle tanker will go to and when. This requires some calculations and functions. In this case, it was decided that the next available shuttle tanker (the one that just finished offloading at Veidnes) goes to the offshore field that requires offloading first. Loading starts when the tank will be full in x hours, where x is a parameter that can be sensitized.

Another challenging issue which needed hard-coding of algorithms, is the rule-set of loading and offloading at Veidnes. In order to know if there is capacity to receive or even to know if an export tanker can be filled with the right API blend, it is necessary to keep track of the total volume of each API value and the used and potential volume for each tank at Veidnes. This is fairly straight-forward. However, with the introduction of a second jetty in phase 3, one can easily end up in a situation where e.g. an export tanker is about to load when the shuttle tanker arrives at the other jetty. Hence,

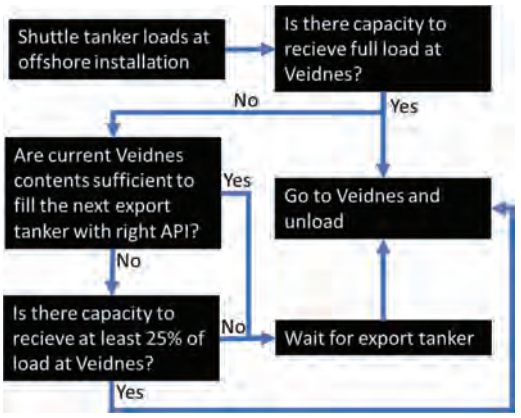


Figure 3. Algorithm for prioritization of shuttle and export tankers.

it is not enough to know the status at Veidnes, but also the remaining volume and API grade of a tanker at the jetty in the middle of a loading/offloading process. Operational rules implemented in the algorithm are illustrated in Figure 3.

Yet another operational rule must deal with the loading and offloading in case of bad weather. Critical wave height limits are typically established for given vessels. The critical limit for start of loading operation is lower than the limit for abortion, which may occur if the wave height increases during the operation. If the operation is aborted due to increased wave height during the operation, exceeding the critical limit, it must also be decided if one should take the current load to the destination or stand-by for a new weather window to complete the operation.

As for the weather itself, we obtained weather data for the last 25 years. There are several possibilities to apply these data to model the weather over the next 15 years. A common approach would be to fit the duration of the good weather and bad weather periods to probability distributions and draw a number from those distributions during the simulations. Good and bad weather would then be defined by the critical wave height limits. In the Veidnes project, another approach was used. The weather data are stored in a database linked to the simulation model. At the beginning of a new year in the simulation, a random historical weather year is drawn. Then the weather in the model replicates the weather in that particular year.

3.3 *Sensitivity cases*

In order to solve the objective function, $f(\bar{x})$ must be simulated, while $g(\bar{y})$ and $h(\bar{z})$ can be calculated. However, the vectors, $\bar{x}, \bar{y}$ and $\bar{z}$ contain many of the same elements, so they cannot be optimized independently. For instance, the amount and size of the tanks at Veidnes contribute to CAPEX, OPEX and revenue. Due to the fact that we simulate revenue rather than attempt to express $f(\bar{x})$ mathematically, it is more difficult to genuinely find an optimum, as we simulate rather than derivate the function. When simulating the oil export, a specific tank configuration must be selected. Hence, each simulation produces results for a specified vector, $\bar{x}$. In theory it can take an infinite range of values, making the task of finding an optimal value impossible by just selecting arbitrary values for $\bar{x}$ and running simulations. In practice, there are of course restrictions with regards to many of the parameters. Not only is the tank size restricted to a positive value, which cannot exceed the size of the largest tank available on the market, but the tank supplier usually has a finite set of products to offer.

Still, there are many parameters and thus a large amount of combinations of values, even if they can take finite values or at least values in a finite interval. Simulation thus impedes the calculation of an optimum value somewhat, but this is the price to pay in order to be able to realistically model the revenue from exported oil.

In the Veidnes case, sensitivity analyses are easily done, with all essential model input parameters listed in MS Excel with a link to the simulation model. Each Excel worksheet represents a unique sensitivity case. By changing parameter values and running sequential simulations for all cases, results are exported back to Excel for each case for comparison.

There are two main cases dealing with the different oil qualities. One alternative is to keep all API grades segregated. The other alternative is to blend them to a mix with a given API. There is also a question of how to blend, since it can be done either in the Veidnes tanks or in the export tanker. Only the second option has been considered in this analysis, i.e. the tanks at Veidnes never contain more than one oil quality and the blending is done when loading the export tanker from different tanks.

Except for the 'segregated' and 'blended' cases, all other sensitivity cases are defined by changes in parameter values. Thus, the model becomes a tool for Design of Experiments (DOE), to contribute to the economic optimization. That is, the model is suitable for analyzing the effect of changes in all the parameters listed in Section 3.1. The DOE includes all of these, as well as the timing for the time-dependent parameters.

With the objective function calculated for each sensitivity case rather than derivated, we obtain a

set of results. Of these we can select the one which gives the highest profit. If we are uncertain about this being the optimal value, new simulations can be run for values in the vicinity of the previous ones.

4 RESULTS

Detailed results of the case study are not presented in this section due to confidentiality. Also, only preliminary results have been produced so far, due to high uncertainties, especially in the economical parameters. Several interesting observations have been made, however. First off, it was observed that the assumed tank volume at Veidnes is sufficient and could even be somewhat reduced, thus reducing CAPEX, if one is willing to accept a few deviations. That is, in the segregated oil quality case, one must accept that a small fraction of export tankers is not fully loaded. For the blended case, one must accept that a small fraction of export tanker loads is off-spec with respect to oil quality. Due to variations in the production profiles for the different fields and the batch-like nature of the supply to Veidnes, situations like these are difficult to avoid entirely. If one has zero tolerance of these issues, the number of Veidnes tanks must increase significantly, which generally will result in over-capacity except for those few instances. In order to conclude, we need to obtain the cost associated with off-spec tanker loads.

A few offshore disruptions or so-called tank tops are also expected over a 9-year period. The simulation model verified that the planned number of shuttle tankers is sufficient and disruptions are not due to late arrivals, but solely long periods of bad weather. The only way to avoid these, is changes in design of shuttle tankers and loading facilities, enabling higher tolerance for waves. However, the number of disruptions seemed acceptable to the project.

In the base case, it was assumed that an increase to three shuttle tankers and two jetties were needed in phase three. This seems to lead to some over-capacity. For instance, average jetty occupation is right below 30%. It would seem sufficient with one jetty occupied 60% of the time. However, this average is calculated over the period, 2022 to 2030. In the peak year, 2026, the jetty occupation is close to 40%. With one jetty this would be 80% and lead to much waiting time.

5 CONCLUSIONS

This paper has demonstrated that mathematical optimization of terminal dimensioning and logistics is difficult but possible to a large extent. Drivers of production deferral in a complex supply chain are difficult to identify using analytical approaches, but discrete event simulation is a helpful tool as it is suitable for capturing the dynamics and reduce mathematical complexity. This does, however, require a flexible tool, allowing for algorithms to represent operational rules in a realistic way.

Even though results are preliminary, there are several key learning points for the project. E.g. optimum tank configuration depends on both the blending strategy and the size of the export tanker.

It has been demonstrated that results from simulation of DOE lead to very valuable insight by quantifying important parameters and reveal how they are influenced by changes in input values. Although a single optimum is difficult to conclude from these results alone, they provide valuable insight and a solid basis for cost-benefit evaluations.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

An integrated bayesian network and cost-benefit analysis model for blowout preventer configuration selection in deepwater offshore fields

E.M. Enjema, M. Shafiee & A. Kolios
Cranfield University, Bedford, Bedfordshire, UK

ABSTRACT: Due to the capital intensive nature, limited supply quantities, infeasible and unviable prospects, among several other setbacks of other emergent energy sources, huge importance continues to be placed on Blowout Preventers (BOPs), the principal defense mechanism against blowouts during any drilling/workover operation in the oil and gas sector. Particularly so after the Macondo disaster, BOPs have been the center of regulatory change and sector development. BOP availability and reliability become even more important as drilling advances into deep and ultra-deep water offshore fields. The BOP configuration choice for such variable environments will have far reaching consequences. Reliability, though hugely important and vital, is one of the several criteria that operators must use for determining the most cost-effective configuration as the cost of accidents in deeper waters increases proportionately. In the current paper, an integrated framework for the selection of the most appropriate BOP configuration in deep and ultra-deep water conditions is proposed. The framework captures all evaluation criteria such as BOP reliability, handling/deployability, overall weight and CAPEX/OPEX ratio. Appropriate mathematical and evaluation tools such as Bayesian Network (BN) and Lifecycle Cost Analysis (LCCA) are employed to evaluate different configurations. The models are applied to a commonly used CLASS VII subsea BOPs in deeper waters. The results indicate that configuration 1 (with 2 annular, 2 pipe rams, 1 blind shear ram, 1 casing shear ram) is slightly less reliable than configuration 2 (with 1 annular, 2 pipe rams, 1 blind shear ram, 2 casing shear rams), however, the operation and maintenance (O&M) costs are higher for the latter configuration. Our framework can serve as a valuable decision making tool for BOP stakeholders as varying facets of information regarding the device are obtained.

1 INTRODUCTION

The increasing demand in world's energy consumption has made the Blowout Preventer (BOP) a crucial and indispensable complex technical system. Coupled with improved safety awareness and growing stringent environmental protection policies, the safe operation of the device cannot be overemphasized. However, depletion of shallow oil reserves has forced drilling and oil exploration into deeper, erratic sea environments. Such deep and ultra-deep water developments rely on new technology, which is yet to be field proven, hence, increased uncertainty related to occurrence of unforeseen events and higher capital expenditure, costs of production interruption and subsea intervention costs (Enjema *et al.*, 2017). A combination of unpredictable and erratic environmental conditions as well as increased associated costs is a major challenge in this sector.

Several studies have been carried out on different individual aspects of the BOP system, including its reliability (see, e.g. Cai, *et al.*, 2012; Holand & Skalle, 2011; Holand, 2001), cost analyses (see, e.g. American Petroleum Institute (API), 2015), etc. More so, aspects such as maintenance, repair and inspection, or configuration and classification are well regulated and documented by regulatory bodies. However, there exist great dependence and interconnectivity among these aspects and evaluating them holistically provides a better view of the entire system functioning and operation.

Given the above-mentioned research gap, this study proposes a framework that considers many of these factors simultaneously and the interrelationship between them. The model is comprehensive and built upon modern intuitive techniques such as Bayesian Network (BN) for technical reliability analysis, Cost Benefit Analysis (CBA) for economic analysis and eventually, a comparison scheme for selection of the best solution. The proposed approach is validated with a case study of a class VII BOP system and the results are subsequently discussed and evaluated. The generic nature of the proposed framework makes it applicable to various other complex engineering systems.

The rest of the paper is organized as follows. Section 2 presents the selection framework developed in this research. In Section 3, the model is applied to a case study and the results are analysed in Section 4. Finally, the research is concluded in Section 5.

2 PROPOSED FRAMEWORK

In this Section, a conceptual framework is developed for the purpose of analysing and selecting suitable BOP configuration for varying operational conditions. Comparative analyses of the reliabilities of various system configurations under these conditions and a cost-benefit evaluation of the configurations is performed. Decision making based on economies of scale and suitability for particular fields of operation is simplified. Information used in the concept development emanate from published literature in the offshore oil and gas sector, face-to-face semi structured interviews and correspondence with BOP experts. The proposed framework, as shown in Figure 1, consists of three main phases. Phase 1 is a preparatory stage in which the premise for selection is determined. The system requirements, functionality and safety levels are investigated. This provides further information on the related parameters involved in operating the system. Evaluating and discretising the operating condition(s) is achievable at this stage. All related correlations are then analysed and data is collected. Phase 2 is the core of the framework, in which mathematical and technical determination of all correlations related to BOP selection is performed. The final phase captures the comparative analyses and selection process. The entire framework will provide a modern comparative and selection tool, given several interrelated parameters. Key tasks in each phase are described in the following subsections.

2.1 *Preparation*

2.1.1 *Setting premise*

Setting the ground rules for the technical, operational and economic aspects is the first step. BOPs

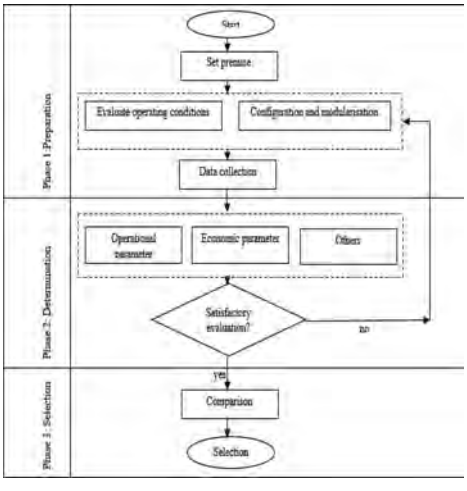


Figure 1. The proposed integrated framework.

are deemed the most safety critical component in a driller's toolbox but also the single largest agent of unproductive time (Sattler, 2013). BOPs are no doubt complex, multi-configuration and multi-phase devices, synchronising several individual components of hydraulic, electrical and mechanical nature. Aside the failsafe function of monitoring and maintaining well integrity, BOP system's primary functions are therefore:

- to confine or seal off well fluids in the well bore
- to provide means of adding or withdrawing controlled volumes of fluid to and from the well bore.
- to shut or 'kill' the well and seal the wellhead.

Typical BOPs are huge pieces of equipment with some current weighs of about 450 tons and 60 ft. in height (Tulimilli *et al.*, 2014). Principal factors affecting safety and reliability of BOPs particularly in high pressure high temperature (HPHT) operations include (Montgomery, 1995):

- Manufacturing specifications
- MIT (Maintenance, Inspection and Testing) techniques.
- Temperature and pressure (and corresponding fluid effects).
- Stack configurations.

Deep and ultra-deep water developments rely on new technology, which is yet to be field proven. The uncertainty related to occurrence of unforeseen events increases as novel technologies introduced for deep waters are not encountered in shallow waters. Deeper waters are also characterized by large capital expenditures with relatively high operational expenditures and high sustainable production rates – hence large losses for production interruption. Furthermore, subsea interventions become more expensive and are associated with longer waiting times for the required mobilization of intervention vessels. Subsea well system repairs and economic penalty for delayed/lost production also soar, characterized by long delays centered on availability, particularly in ultra-deep water environments.

Some BOP configurations have been shown to be more reliable than others (see, Cai *et al.*, 2012; Sattler & Gallander, 2010). Individual components' reliability affects overall system reliability due to variations in numbers, position and even type within stack configuration. Common-cause failures are popular within BOP systems, with a dominant impact on accidents (Cai *et al.*, 2012). Redundant components (control panels, control pods, annular preventers, ram preventer, valves and regulators etc.) may fail from a single event. Historically, multiple low frequency failures lead to blowouts (Whooley *et al.*, 2011). A failure in a BOP system with series-fashion connectivity will affect the entire system functionality. It is therefore obvious that reliability

is directly related to configuration, which intends to have a direct bearing on size/weight and hence total cost of operation and maintenance (O&M).

Safety Integrity Level (SIL) requirements for the BOP are specified in IEC 61508 and 61511standards which are widely accepted for the basis of operation and design of Safety Instrumented Systems (Pinker, 2012). The Norwegian Petroleum industry accepts a minimum SIL 2 requirement for the SIF (Safety Instrumented Function) of the subsea BOP systems (NOG 070, 2004).

2.1.2 *Evaluation conditions*

Technology progressed significantly and by the 1990s it was advanced into deeper depths. Deepwater drilling refers to water depth between 400 m – 1500 m. Depletion shallow water reserves coupled with huge advances in technology has seen many drilling companies venture into depth greater than 3000 m. Though the risks in shallow water drilling are considerably less, the economics of deep-water production are highly attractive and worth the risks (Latham, 2002). Areas such as the Atlantic Margin in Europe, Gulf of Guinea in West Africa, the Gulf of Mexico and the Compos Basin in Brazil are currently drilling and exploring deep-water environments of over 2000 m (Oyeneyin, 2009). These future oil and gas development regions are burdened with complex geological features and profiles. Conditions that affect the operation, maintenance and testing (MIT) of the BOP are also considered. These new adventure presents varying oceanographic and geological environments, flooded with high gas-oil ratios, HPHT regions, elevated tides and wave currents, difficult formations and even lack of experienced personnel (Skogdalen & Vinnem, 2012). Specific data based on some of these conditions provide some foundation in this study. Erratic conditions are limited to HPHT, where 'high' is defined based on API TR 1PER15 K-1, API 17TR8, and expert literature regarding current operating limits (Lehr & Collins, 2015):

- High Pressures ≥ 15000psi (~1000bar)
- High Temperatures ≥ 350°F (175°C)
- Water depths ≥ 5000 ft or 1500 m

Classification for HPHT wells are expected wellhead shut in pressure ≥ 10.000psi (690bar) or pore pressure gradient >1.81bar/10 m and high temperature when reservoir temperature or wellhead temperature > 150°C (Masi *et al.*, 2011).

2.1.3 *Data collection*

Data for analyses, particularly for safety critical technical systems in the oil and gas industry, is constant challenge (Khakzad *et al.*, 2014). The accuracy and integrity of the data is sometimes questioned. Expert judgements through rigorous elicitation techniques (see Clemen & Winkler, 1999; Keeney & Winterfeldt, 1991) are used and due to the dynamic nature of the proposed framework, updating is possible when new and coherent data becomes available. In a bid to overcome this problem, subject experts with extensive years of experience are selected to provide required information. More so, different types and forms of data is sourced from different systems, the trends are observed and aggregation is done where possible to eliminate and minimise ambiguity. In the second phase, different types of data about system design, component interaction and interrelationships and operational procedure are required. Primary qualitative data is obtained via questionnaires completed by subject experts. Slightly more precise and delicate qualitative assertions and quantitative information such as failure rates and conditional failure probabilities, load margins and corresponding effects are gathered. In order to control the quality and ensure specificity of the data obtained, interviews and questionnaires are employed. Specific primary data here provides greater confidence in the results obtained. Additional secondary data is also available within literature. However, greater focus is placed on primary data obtained from questionnaires.

2.2 *Determination*

2.2.1 *Operational parameter (Reliability)*

A few studies such as Holand (2001), Holand & Skalle (2011), Holand & Awan (2012) have considered the reliability of the BOP in deep-water conditions, using the fault tree analysis (FTA) and the reliability theory. Some other studies, e.g. Cai *et al.* (2012) and Cai *et al.* (2013) use a more contemporary technique called Bayesian Network (BN). The effects of inherent complex technical systems characteristics such as common-cause failures are considered, alongside important reliability attributes such as maintenance and repair can be captured with this technique. A Bayesian network (BN) is a compact representation of a multi-variate statistical distribution function (Langseth & Portinale, 2007). It has a qualitative side, portrayed by a directed acyclic graph with nodes (representing random variables) and arcs between the nodes (representing dependencies) which together define the joint probability distribution over all random variables (Boudali & Dugan, 2006). Nodes represent cause and effect in real-world situations and arcs connect the nodes. The quantitative part, made up of conditional probabilistic tables, could easily be ascertained by a domain expert (Langseth & Portinale, 2007).

BNs perform both predictive and diagnostic analyses and are now considered as a viable alternative technique (Khakzad *et al.*, 2013). Such predictive analysis is used to compute the reliability of systems in a process called marginalization. Qualitative and quantitative parameters, discrete and

continuous data, equations and data sets can be modelled into systems or processes through BNs. Erratic conditions such as extreme temperature and pressure and their immediate corresponding effects are easily incorporated into complex technical system analyses via this methodology. A simple step by step frame is presented in Figure 2.

2.2.2 *Economic parameter (cost-benefit ratio)*
Several factors influence the choice of BOP employed in any drilling or workover development and this is common with other complex technical systems which are associated with high turnovers and productivity. Cost-benefit analysis (CBA) as used in economics enables long-term decision making by comparing present value of costs with the present value of benefits. An activity is considered worthwhile when the sum of its benefits outweighs the sum of its costs, i.e. benefit/cost ratio is greater than 1. It becomes necessary to identify the associated benefits and costs. Based on experts' input, a summary of potential benefits and costs associated with BOP operation and functioning are shown in Table 1.

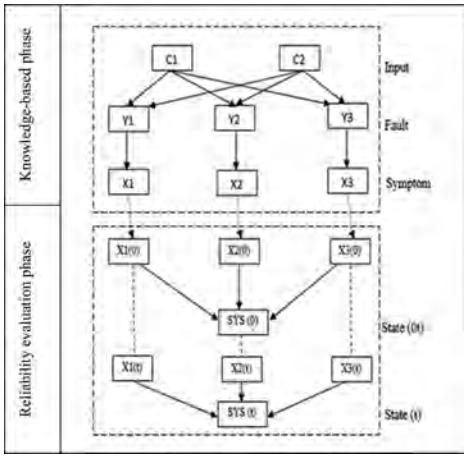


Figure 2. Bayesian network framework.

Table 1. Associated BOP costs and benefits.

Costs	Cost elements	Benefits	Benefit elements
Capital costs	Purchase Upgrades	Production	Revenue
Installation costs	Labour Logistics	Increased Safety	
Operating cost	Maintenance Royalties Taxes Logistics	Time savings	

2.2.3 *Others*
Given that the proposed framework is a living tool, stakeholder decision-making is facilitated as new data is obtained. Though reliability, of paramount importance, is the focus of the study carried out, other related factors which affect the selection process are compared. The economic benefits are determined in CBA whilst other criteria such as associated weight/size are obtained from varying available data sources and expert opinion.

2.3 *Selection*

In the case of different BOP configurations under study, the most appropriate option is chosen on the basis of a number of criteria such as reliability, cost, weight/size, technical handling/deployment issues, etc. For this purpose, a simple weighted decision matrix is used to rank options. Buying or rental and MIT expenses may be of utmost importance to an operator but reliability and availability affect several other stakeholders, the environment and beyond. Adding weights to these criteria will enable clear decision making and eventual optimisation.

3 APPLICATION AND RESULTS

In this Section, the proposed model is applied to a BOP of class VII. This class is typically used in deep water application and comprises of either a dual annular, five (5) ram arrangement or a single annular, six (6) ram configuration. The required data is collected from 4 BOP experts, two BOP vendors and supported by extensive literature review. Employing the proposed model, prior to setting the appropriate premise, the reliability for both Class VII configurations is calculated. The selected tool, incorporates a continuous discretised evaluation range for both temperature and pressure, considering one-year period, and the results are presented in Table 2.

The results indicate that the configuration 1 (with 2 annular, 4 pipe rams, 1 blind shear ram) is slightly less reliable (obtained with Bayesian Network analysis) than configuration 2 (with 1 annular, 4 pipe rams, 1 blind shear ram, 1 casing shear rams). Increase in reliability may be due to

Table 2. Reliability approximation of different configurations.

	Reliability	
Configuration	Temperature model	Pressure model
1) 2 annular, 5 rams	≈ 99.7	≈ 97.2
2) 1 annular, 6 rams	≈ 99.8	≈ 97.5

Table 3. Preventer type and overall configuration weight.

BOP Type	Weight (lbs)	Configuration 1 quantity	Total	Configuration 2 quantity	Total
Annular	40,632	2	81264	1	40632
Pipe (single)	23,300	4	93200	4	93200
Casing shear (single)	28783	–	–	1	28783
Blind shear (single)	28783	1	28783	1	28783
			Total = 203247		Total = 191398

Table 4. Final selective decision matrix.

Decision matrix			
Criteria	Weight/ Rating	Configuration 1	Configuration 2
A: Reliability	0.4	2	2
B: Weight/size footprint	0.2	–2	1
C: Technical feasibility	0.1	1	–1
D: Economic feasibility	0.3	1	–1
	Total = 1	0.8	0.6

Feasibility scale:
2 = better than average.
1 = slightly better than average.
–1 = slight worse than average.
–2 = worse than average.

the casing shear ram. Temperature models also seem more tolerance compared with their pressure counterparts as pressure has a direct bearing on the operation of the rams. Costs and benefits are also evaluated for each configuration. Approximate purchase prices for single annular, pipe, and shear preventers are $29500, $39500, and $105000 respectively. The Net Present Value (NPV) can be appropriately estimated by calculating the Net Cash Flow (NCF), depicted by Liu and Ford (2008) as:

$$NCF = Revenue - CAPEX - OPEX - Tariffs - Tax$$

The analysis captures most factors, parameters and risks involved in operating a complex technical system. Holistic analyses and suitability assessment of the individual configurations examined for the loading parameters involved is then made possible. As mentioned earlier, configurational changes affect the weight and size of the entire BOP stack. This has bearings on maintenance as number of components may have increased or reduced. More so, the larger the stack, the more complicated installation, deployment and decommissioning will be. Weight and size depend largely on the manufacturer and if the rams are studded, flanged or hub type. Table 3 provides the details on different BOP types available in the market.

Finally, based on expert opinions, the weights are assigned to the parameters and the best alternative, given the period and conditions considered is chosen.

4 CONCLUSIONS

This study developed a framework for BOP selection, particularly in deep and ultra-deep water environments. The model provided a powerful decision making tool for incorporating the effects of erratic/extreme loading conditions and selecting the set of system modules, considering several related factors. Application to a class VII BOP configuration was also shown and the results demonstrated the validity of the proposed framework. Correlating several aspects related to the selection process facilitated decision making and provided a better overview of the many intricacies involved in deep and ultra-deep water exploration and drilling. Simultaneously analyses and incorporation of technical, operational and economic selection aspects were incorporated. The entire framework, once completed efficiently resulted in varied forms of information relating to system operation, design (redundancy configuration), loading condition boundaries and even economic consideration. The proposed model can be applied to other modular complex technical systems in general as it is comprehensive, intuitive and dynamic. Its application to other operational conditions is a possible step in the future. Appropriate and more coherent data will also provide refined results in further analyses. The development of appropriate modelling tools to reduce uncertainty in cost assessment and the development of a software based selection mechanism after the evaluation of the necessary correlates are other future prospects.

REFERENCES

American Petroleum Institute (API), 2015. BSEE Proposed Well Control Rule Cost and Economic Analysis, Texas: American Petroleum Institute.

API 17TR8, 2015. HPHT Design Guidelines for Subsea Equipment, Washington: API Publishing Services.

API TR 1PER15 K-1, 2013. Protocol for Verification and Validation of High-pressure High-temperature equipment, Washington: API Publishing Services.

Boudali, H. & Dugan, J. B., 2006. A Continuous-time Bayesian network reliability modeling and analysis framework. *IEEE Transactions on Reliability*, 55(1), pp. 86–97.

Cai, B. et al., 2012. Performance Evaluation of Subsea Blowout Preventer Systems with Common-cause Failures. *Journal of Petroleum Science and Engineering*, Volume 90–91, pp. 18–25.

Cai, B. et al., 2012. Reliability analysis of subsea blowout preventer control systems subjected. *Journal of Loss Prevention in the Process Industries*, Volume 25, pp. 1044–1054.

Cai, B. et al., 2013. Performance evaluation of subsea BOP control systems using dynamic Bayesian networks with imperfect repair and preventive maintenance. *Engineering Application of Artificial Intelligence*, Volume 26, pp. 2661–2672.

Clemen, R. & Winkler, R., 1999. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2), pp. 187–203.

Duell, C., Fleming, R. & Strutt, J., 2001. Implementing Deepwater Subsea Reliability Strategy. Texas, s.n.

Enjema M, Shafiee M & Kolios A (2017) A study on the reliability of oil and gas blowout preventer (BOP) technologies under deep-water erratic conditions. In: *European Safety and Reliability Conference (ESREL)*, Portoroz, Slovenia, 18.6.2017 – 22.6.2017.

Holand, P. & Awan, H., 2012. Report:Reliability of Deepwater Subsea BOP Systems and Well, unrestricted version. [Online] Available at: www.bsee.gov [Accessed 19 October 2016].

Holand, P. & Skalle, P., 2011. Deepwater kicks and BOP Performance, Trondheim: SINTEF.

Holand, P., 2001. Reliability of Deepwater Subsea Blowout Preventers. Society of Petroleum Engineers, pp. 12–18.

Keeney, R. & Winterfeldt, D., 1991. Eliciting probabilities from experts in complex technical problems. *IEEE Transactions on Engineering Management*, 38(3), pp. 151–173.

Khakzad, N., Khan, F. & Amyotte, P., 2013. Quantitative Risk Assessment of Offshore Drilling Operations: A Bayesian Approach. *Safety Science*, Volume 57, pp. 108–117.

Khakzad, N., Khan, F. & Paltrinieri, N., 2014. On the application of near accident data to risk analysis of major accidents. *Reliability Engineering and Safety Systems*, Volume 126, pp. 116–125.

Langseth, H. & Portinale, L., 2007. Application of Bayesian networks in reliability analysis. In: M. Ankush & A. Kassim, eds. Bayesian Network Technologies: Applications and Graphical Models. s.l.:IGI Global, pp. 84–102.

Langseth, H. & Portinale, L., 2007. Bayesian networks in reliability. *Reliability Engineering and System Safety*, 92(1), pp. 92–108.

Latham, A. J., 2002. Commercial Realities in Deep and Ultradeep Water. s.l., Energy Institute.

Lehr, D. J. & Collins, S. D., 2015. The High-Pressure/High-Temperature Completions Landscape – Yesterday, Today and Tomorrow. Amsterdam, SPE Annual Technical Conference and Exhibition.

Liu, S. & Ford, J., 2008. Cost/Benefit Analysis of Petrophysical Data Acquisition. Edinburgh, Society of Petrophysicists and Well Log Analysts.

Masi, S. et al., 2011. Blowout Probability in Dangerous Wells: A Sensitivity Analysis between CHCD and HP/HT Environments. Ravenna, Offshore Mediterranean Conference.

Montgomery, M. E., 1995. Inspection and Testing procedures improve BOPs for HPHT drilling. *Oil and Gas Journal*, 93(6), pp. 49–53.

NOG 070, 2004. 070 Norwegian Oil and Gas Application of IEC 61508 and IEC 61511 in the Norwegian Petroleum Industry. [Online] Available at: www.norskoljeoggass.no [Accessed 19 September 2016].

Oyeneyin, M. B., 2009. Cost Effective Deepwater Well Construction – A Case for Managed Pressure Drilling with Casing. Aberdeen, Society of Petroleum Engineers.

Pinker, R., 2012. Improved method for reliability assessment of safety- critical systems: An application example of BOP systems, Trondheim: Norwegian University of Science and Technology.

Sattler, J. & Gallander, F., 2010. Just How Reliable is Your BOP Today? Results from JIP,US GOM 2004–2006. New Orleans, IADC/SPE 128941.

Sattler, J., 2013. BOP Performance- Developments and Consequences in a Post-Macondo World. Amsterdam, SPE/IADC.

Skogdalen, J. E. & Vinnem, J. E., 2012. Quantitative Risk Analysis of Oil and Gas Drilling using Deepwater Horizon as case study. *Reliability Engineering and System Safety* 100, pp. 58–66.

Tulimilli, B. et al., 2014. Design Study of BOP Shear Rams based on validated simulation model and sensitivity studies. San Francisco, ASME.

Whooley, A., Deegan, J., Goldsmith, R. & Botto, A., 2011. Tools and Techniques for the Selection and Design of Safer Deepwater Risers Systems for Mobile Offshore Drilling Units. Rio de Janiero, Offshore Technology Conference.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

The use of reliability simulation techniques in data-driven facility simulation

F. Reinecke & S. Bracke
Chair of Reliability and Risk Analytics, University of Wuppertal, Wuppertal, Germany

ABSTRACT: Simulation methods are widely used in different industrial sectors like economics, logistics and also in the field of reliability engineering. In general, simulation pursues the target of imitating the operation of a real-world system over time in order to gain knowledge of its behavior which is transferable to reality. It is especially applied when the testing of the real-world system is too expensive, risky or too complex. The aim of this paper is to present different simulation modelling techniques which are used in reliability engineering and other subject areas and to evaluate them regarding the specific application of Data-driven Facility Simulation (DFS). After presenting the techniques, important issues of DFS are discussed. Subsequently, the presented modelling methods are evaluated with respect to a wind park consisting of wind turbines as an illustrative example. They are discussed regarding different advantages and disadvantages as well as limitations.

1 INTRODUCTION

Simulation is widely used in different industrial sectors like economics, logistics and also in the field of reliability engineering. It pursues the target of imitating the operation of a real-world system over time in order to gain knowledge of its behaviour, which is transferable to reality (Banks 2005). It is especially applied when the testing of a real-world system is too expensive, risky or too complex (Kolonko 2008).

One of the first steps in performing a simulation study is to develop a (mathematical) simulation model, which consist of objects and their mutual relationships. It is expressed in a mathematical, logical or symbolic way and can either be static or dynamic (Banks 2005).

In reliability engineering, various simulation approaches have been used, i.e. for estimating performance characteristics of system components, the system availability and safety or the system's degradation. For example the Monte Carlo Simulation (MCS) as well as its specific application, the Discrete Event Simulation (DES), is used for the evaluation of reliability characteristics of multicomponent technical systems as well as for the estimation of the reliability of mechanical structures (VDI 1999). Gathered simulation results can be used for reengineering activities or for further improvement of the system.

The aim of this paper is to show selected and established simulation modelling methods, which are used in reliability engineering and other subject areas. The stated modelling methods are evaluated regarding the specific application of DFS. This is done by means of a wind park consisting of wind turbines as an illustrative example. They are discussed regarding different possibilities of application, advantages and disadvantages as well as limitations.

2 MONTE CARLO SIMULATION IN RELIABILITY ENGINEERING

The MCS is a well-known and established method for simulating stochastic systems. Within MCS, random numbers are used for solving or rather estimating solutions of a deterministic or stochastic problem (Law 2007; Zio 2013). It is often applied when it's too complex or impossible to evaluate a system by analytical methods. In reliability engineering, MCS is used for the determination of reliability characteristics of technical systems and can be performed by the following 3 steps (cf. VDI 1999):

1. Building a stochastic simulation model $y = \varphi(\underline{x})$, which contains random variables (i.e. component state variables x_i). The model should describe the behavior of the system under study in a sufficient way.
2. Implementation of the model in a computer program for simulating system failure processes
3. Simulate the model with a sufficient amount of simulation runs N. Within a simulation run,

input random variables $x_i, i = 1,...,n$ are generated with consideration to their determined distribution (i.e. by inverse transform sampling or rejection sampling, etc.) and the model output (the simulated state of the system y) is calculated. Lastly, statistical parameters of the generated output data $y_k, k = 1,...,N$ are calculated (i.e. mean value, variance, etc.)

According VDI (1999), the MCS can be performed at different types of models, which can be static or dynamic. In literature, the MCS is often associated with simulating only static models whereas for discrete dynamic models the terms of discrete event and discrete time simulation (DES and DTS) are used. However definitions of MCS vary in literature.

3 METHODS FOR SIMULATION MODELLING

This section deals with selected dynamic simulation modelling techniques, which are used in reliability engineering and are mostly based on MCS. The methods are presented and instances of applications in the field of reliability are shown.

3.1 *Discrete event simulation*

Discrete Event Simulation (DES) describes a technique of system modelling where the system's state variable can only change at discrete points in time (Banks 2005) when an event occurs which is defined according (Law 2007) *"as an instantaneous occurrence that may change the state of the system"*. After establishing a model, it can be run and an artificial history in form of generated data can be produced (Banks 2005).The DES can be seen as particular application of the MCS if events occur at random.

A typical application of using DES in reliability engineering is the simulation of a model (e.g. a block diagram model) over a specific operation time t. For each component, failure and repair events are generated from its time-to-failure or time-to-repair distribution. At each point of time an event occurs, the system state (success or failure) is determined. Moreover, in such simulations, different constraints can also be considered (e.g. repair strategies and capacities or operation phase-switching) (cf. VDI 1999). With a sufficient number of simulation runs N, reliability measures (e.g. MTTF) can be estimated. There are also other applications where DES is used. For instance, it is used for reliability and availability assessment of civil engineering structures (Juan 2009). Moreover, Sharda & Bury (2008) developed a DES model for understanding facility failure effects on a chemical plant's production capability.

3.2 *Discrete time models*

According Zeigler (2010), within discrete time models, the time advances stepwise in a discrete way (i.e. in a one-second interval). At each specific point in time, the model persists in a certain state. The model is executed at each point in time and the next state is determined for the subsequent time step. In this subsection, selected discrete time modelling methods are presented and discussed.

3.2.1 *Discrete-Time markov chain models*

Discrete-Time Markov Chains (DTMC) can be used for modelling and simulation studies (cf. Nfaoui, Essiarab & Sayigh 2004; Papaefthymiou & Klockl 2008; Sahin & Sen 2001; Shamshad 2005). A discrete time stochastic process $X_t, t \in \tau, \tau = \mathrm{N}_0$ can be described by a markov chain, where for all states $i_0,...,i_t \in I$ the following condition (1) is valid (Waldmann & Helm 2016).

$$\begin{aligned} &P\left(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1},..., X_0 = i_0\right) \\ &= P(X_{n+1} = j \mid X_n = i) \\ &= P_{ij} \end{aligned} \tag{1}$$

I denotes the state space and P_{ij} the one-step transition probabilities which can be represented in a one-step transition matrix **P**.

The simulation of the stochastic process is carried out for each time step and can be performed as follows (cf. Sigman 2007).

1. Select an initial value $X_n = i_0, n = 1$
2. Generate X_{n+1} by sampling from the conditional distribution $P(X_{n+1} \mid X_n)$ and set $n = n + 1$
3. Set $X_n = X_{n+1}$ and go back to step 2 until the desired number of n is reached.

Markov chain models are also used in the field of reliability. Skalny (2013) used Discrete-Time Markov Chains in combination with the MCS Method for estimating the failure probability of companies for satisfying an order to industrial partners. Markov chains were also applied in modelling reliability structures (cf. Koutras 1996).

3.2.2 *ARMA models*

Autoregressive Moving Average Models (ARMA models) (Box & Jenkins 1979) are used for time series forecasting and simulation (especially for wind speed time series (cf. Kamal & Jafri 1997; Sfetsos 2000)). In ARMA models, it is assumed that the present value of a variable is the result of a linear function of a specific number of past values and a random error term. It can be represented generally in the following form (see equation 2):

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} + e_t \quad (2)$$

y_t denotes the actual value and e_t the random error which is assumed to be i.i.d. with mean of zero and constant variance σ^2 at time t. $\phi_i (i = 1,...,p), i \in N_0$ are the autoregressive and $\theta_j (j = 1,...,q), i \in N_0$ are the moving average model parameters, p and q are referred to as the order of the model. ARMA models can be used to model only stationary time series. If a time series, which has to be modelled, isn't stationary, it has to be differenced d times or transformed. The determination of the model (model orders p, d, q and model parameters θ_l and ϕ_l) can be performed by the Box and Jenkins methodology (Box & Jenkins 1979).

The simulation process can be done by performing the following three steps:

1. Select an initial value $y_t, n = 1$
2. Generate a value of e_t by sampling from the fitted normal distribution function for e_t with MCS
3. Calculate a new value of y_t by equation 2, shift all former values one step in the past and go back to step 2.

ARMA models are also applied in reliability engineering. For instance (Ho & Xie 1998) used ARMA models for repairable system failure reliability forecasting. Another application example is given in (Karki, Hu & Billinton 2006) and (Billinton, Chen & Ghajar 1996). They used an ARMA model for simulating wind speeds in combination with a power generation model for the reliability evaluation of wind power systems.

3.2.3 *Artificial neural network models*

Another technique for simulation modelling are artificial neural networks (ANN) (Kruse 2012). ANN's are used in a wide range of applications (i.e. wind speed forecasting). It is a technique for mapping input vectors to output vectors by learning from examples without implying a specific relationship between them (Li & Shi 2010). There are different types of ANN's available (i.e. feed forward, radial basis function networks or recurrent neural networks). This section deals only with the type of feed forward ANN models (FNN) (Kruse 2012). The mapping is realized by interconnected neurons, arranged in different layers called input, hidden and output layer and no feedback loops between the layers exist. The strength of each connection from a neuron j to a neuron i is expressed by a connection weight. Hereby an activation level v_i is calculated for each neuron i in the network as follows (Mohandes 1998):

$$v_i = \sum_{j=1}^{n} w_{ij} x_{ij} - w_{i0} \quad (3)$$

w_{ij} denotes the connection weights and x_{ij} the input values to neuron i. The bias value w_{i0} denotes the shift of the function $\varphi(v_i)$. This function determines the output of a neuron from a calculated activity level and is a nonlinear function (i.e. a sigmoidal function, see equation 4).

$$sig(x) = \frac{1}{1 - e^{-x}} \quad (4)$$

In order to get adequate results by the FNN for a given input, the connection weights of the network need to be trained (Welch 2009). For training, the well-known back propagation algorithm (BP) (Kruse 2012) can be applied. Hereby, an iterative training process minimizes the mean square error of the network output in comparison to the real output by adjusting the connection weights which are assigned with random values initially (Li 2010; Mohandes 1998).

Concerning time series simulation or forecasting, an ANN describes a nonlinear function $f(.)$ which represents the relation of past recorded values of a time series to future values (see equation 5) (Zhang 2003).

$$y_t = f(y_{t-1},...,y_{t-n},\underline{w}) + \varepsilon_t \quad (5)$$

ε_t describes the error term of the model, whereas $\underline{w}$ denotes the vector of the model parameters (connection weights between the neurons). After training the ANN, the simulation can be performed by selecting initial values of $y_{t-1},...,y_{t-n}$ and calculating a new value of y_t. Then all former values are shifted one step in the past and the simulation process is performed again until a specified number of simulation runs is reached. ANN models were also applied in reliability engineering. They are widely used in structural reliability analysis (cf. Chojaczyk 2015; Hurtado 2001). Moreover (Rajpal, Shishodia 2006) used ANN's for modelling and simulating the behaviour of a complex, repairable system (helicopter transport facility) under various constraints. Results can be used for optimizing the operation of the system.

4 MODELLING METHODS FOR DATA-DRIVEN FACILITY SIMULATION

In recent years the availability and functionality of condition monitoring and diagnostic systems regarding a wide range of technical facilities has

increased. The systems offer the ability to record field data of different variables during field operation which can be used for the identification of damages and the facility state history. The data are acquired by different sensors, which are integrated within a plant and are available in form of multivariate time series.

They can be used for prediction purposes like the estimation of the Remaining Useful Life (RUL) or the future facility state. However the amount of recorded data is often small in case of a short field duration time which hinder prognoses. Additionally, the gathered data often represent just a small time frame within the usage phase concerning the whole expected facility service life. Thus long-term predictions concerning the facility state can be afflicted with high error because aspects like changing of environmental conditions; degradation behavior as well as changing of the facility usage, which also leads to a change in the facility load scenarios, cannot be included. In order to address these problems, the behavior of the facility can be simulated over a desired period of time in form of generating synthetic operation data. Thereby the simulation process can be performed under assumptions (i.e. changing usage behavior after a specific operation time) based on expert knowledge which gives the possibility to perform more accurate predictions in case of the occurrence of changing conditions during the facility life.

4.1 *Important issues of data-driven facility simulation*

For performing DFS, several issues (i.e. requirements on data structure) have to be taken into account which are stated as follows:

Data structure and recording
For performing DFS, multivariate input time series data need to be recorded which fulfill requirements of a constant sampling rate and consistent time stamps at all variables. This is needed for instance for identifying cross correlation structures between the variables. Another important issue is the amount of recorded data. For instance, if a simulation model is built in form of a markov chain, the amount of recorded data needs to be sufficient for determining state transition probabilities.

Another important issue is the occurrence of incorrect measurements as well as missing values within the recordings. Therefore, it is recommended to treat implausible and obviously incorrect data as missing values. Missing values in time series can then be replaced with the help of data imputation methods within a data preparation process. On the contrary, outliers are treated as rare events by the simulation model and thus don't need to be rejected.

Data analysis
In order to perform a DFS, a profound data analysis has to be made. Results are serving as input for the simulation model building as well as for the evaluation process. The data which exist in form of time series can be analyzed among others in the following way:

Statistical parameters: Calculating general statistical parameters for evaluation purposes (mean, variance, percentiles, min/max, etc.)
Probability density functions: Fitting probability density functions (i.e. Weibull density function) to the time series data. This can be done either for the whole recorded time series (i.e. also for evaluation purposes) or within a system state (i.e. a specified range of wind speed) for modelling transitions into other system states.
Autocorrelation: An analysis of the autocorrelation structure of the recorded time series needs to be performed for considering the series chronology within the simulation process (Feijóo & Villanueva 2016). This can be done by the following equation given in Shamshad (2005) for a specified number of lags k

$$\rho_k = \frac{\frac{1}{N-K}\sum_{i=1}^{(N-K)}(x_i-\bar{x})(x_{i-k}-\bar{x})}{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(x_i-\bar{x})} \tag{6}$$

where $x_i,...,x_N$ is the analyzed time series and $\bar{x}$ is its mean. The autocorrelation structure describes the correlation of time series values with its past values at different lags. The calculated autocorrelation structures of simulated and real data can be compared and thus the simulation model performance can be evaluated.
Event identification: Technical facilities behave in a different way in various operation phases which also results in the data recordings which must be considered during simulation. For instance wind turbine data recordings clearly differ between starting up phases and shutdown phases as well as in idle time phases (i.e. during maintenance). Therefore it is important to label the recorded time series for allocating data to different operation phases in order to acquire better simulation results. Additionally, the occurrence of events needs to be analyzed regarding frequency, sequences and occurrence probability.
Dwell times and idle times: Another important issue within data analysis is the determination of dwell times in system states because in simulation the time duration within different load states of a facility which is an important factor for the degradation estimation needs to be reproduced in an adequate way. For increasing the accuracy of simulated data

and for better RUL estimation, the determination of idle time durations has also to be taken into account.

4.2 *Evaluation of simulation modelling methods for data-driven facility simulation*

The reviewed simulation modelling methods shall be evaluated regarding DFS by means of a wind farm as an illustrative example. The evaluation is based on a data set consisting of operational time series data of wind turbines located in the western part of Germany. The data set consists of 10 minute values. Its essential structure is shown in Table 1.

Discrete event simulation

The operational time series variables 3 to 7 and 12 (see Table 1) in the data set change at almost every point in time their value or remain just a short time in a value level. This leads to an enormous number of system changes which are not seen as events in the proper sense. DES is usually not applied here because a system change takes place at almost every discrete point in time and a system change can't be assigned to a specific reason. However with DES, sequences of clear specified events concerning a wind turbine can be modelled in which the behaviour of operational time series variables change. An example for such events are maintenance activities, facility shutdowns caused by shadowing or occurring facility failures. An advantage of applying DES for simulating a wind park on event basis is that it can help for evaluating maintenance and repair strategies of the plant or for estimating the plants availability. However, a major disadvantage is that a plant specified model has to be developed. The wind turbine data set includes high numbers of different types of facility states. If DES is applied, a complex model has to be established. For DFS, a DES model can be combined with an DTS model (cf. Cha & Roh 2010). This would take potential impacts from the facility states (e.g. maintenance actions) onto the operational time series data into account as well as their interdependencies. For establishing such a model, expert knowledge of the facility is an essential condition. DES models are usually very process specific which another drawback is.

Table 1. Structure of the wind farm data set.

No.	Variables	Available data
1	Timestamp	Date, hour, minute
2	Plant number	1 to 14
3	Wind speed [m/sec]	Mean, min, max
4	Rotor speed [1/min]	Mean, min, max
5	Active power [kW]	Mean, min, max
6	Reactive power [kW]	Mean, min, max
7	Nacelle position [°]	Mean
8	Blade angle A,B,C [°]	Mean
9	Rainfall [mm/min]	Mean, min, max
10	Visual range [km]	Mean, min, max
11	Ambient brightness [lux]	Mean
12	Temperature of machine parts [°C]	Mean
13	Facility status	Nominal status values

Discrete time models—ARMA models

The operational time series variables of the wind park data set (variables 1 to 12) show a strong autoregressive behaviour which is given by a slight decrease of the ACF function. This is a strong indicator that the time series are non-stationary which means that the series doesn't have invariant characteristics. Invariant characteristics of a (weakly) stationary time series would be a constant mean, a constant and finite variance over time as well as a covariance structure which is only dependent on the lag and not on time (Davies 2014). As with ARMA models only stationary series can be modelled, they are not suited to model operational time series variables of a wind park without stationarizing the series (e.g. by applying data transformation or decomposition). Differencing would lead to a stationary series, however the effect of a huge information loss leads to problems in simulating a time series which has similar characteristics as the observed one. With decomposition techniques, the time series can be divided into a trend-cycle component and a stationary component. The former still can't be modelled by ARMA in contrary to the latter component.

The advantage of ARMA is that just a few parameters for generating a time series with a specified ACF are needed. However, generally the probability density function differs from the measured data which can lead to wrong estimations (Papaefthymiou 2008). Furthermore, concerning wind turbines, it is assumed that ARMA models are not suited for simulating idle times of a facility within a variable (e.g. active power) as well as for instance dwell times (e.g. in specific temperature value level of a machine part) because the simulation of its trajectory is driven by sampling from an error distribution (see equation 2). Another disadvantage of ARMA models is reasoned by its linear nature which make them potentially not well suited for modelling operational time series data of wind turbines (cf. Chen & Yu 2014).

Discrete time models—ANN models

ANN Models are another way for time series simulation as described in section 3.2.3. They can be

used for approximation of various nonlinearities in data (Zhang 2003) in contrary to ARMA models. Moreover, they give the possibility to model a huge number of functions with high accuracy. They were successfully applied in wind speed simulation as well as in simulation of series of solar radiation (cf. Li 2010; Mihalakakou 2000). However, there is no rule available for determining the ANN topology and the number of needed neurons for modelling a function. This has to be examined for every different time series in a parameter study.

A disadvantage of ANN lies in modelling time series in which values persists a certain duration. For example, regarding the wind park data set: During an idle time no output power is generated. That means the future value of past zero values is also zero as long as the facility is in idle mode. However, after a while the facility is re-started which leads to future values different to zero. This can't be modelled solely by an ANN as described in section 3.2.3. The ANN would persist in the idle stage for the rest of the simulation duration. A possible way to face this problem is to use hybrid models. For example, the ANN model can be trained for just simulating changes of value levels and the dwell time within a level can be simulated by a Monte Carlo approach. Another way for applying ANN models in DFS is the usage for regression. For instance, the variables rotor speed and active power are highly correlated. After simulating one of these variables, the other one can be determined through modelling the relationship between the two variables. This gives the possibility for generating multivariate correlated time series.

Discrete time models—markov chains

Instead of ARMA models, Discrete-Time Markov Chains can be used to model operational time series data of the wind park. The methodology is suited for modelling non-stationary and nonlinear time series which are existent in the wind park data set. Moreover, the methodology considers the underlying ACF as well as the probability distribution function of the control sample series. It is also able to simulate dwell times (of temperature levels of machine parts of the wind turbine) as well as idle times associated with the belonging state. However, when applying this methodology, the stochastic process needs to be discretized by defining a number of states (Papaefthymiou 2008), which leads to a loss of information. According Papaefthymiou (2008), when applying MC, a tradeoff between the model accuracy and complexity (represented by the number of states and parameters) has to be made. With a high number of states the process can be better represented. However it is difficult to assess transition probabilities in case of a low data volume. Regarding wind speed series, in literature, MC of higher order are applied, which leads to more adequate results with better ACF and a better retention of the probability density function of the simulated series compared to the real data (Feijóo 2016). However, by using a higher order MC, the model complexity as well as the needed amount of data increases enormously which limits their usability (cf. Aksoy 2004).

The MC approach is also not able to simulate a trend. As the series of the data set don't show a trend over the observation time, this restriction is negligible.

5 CONCLUSIONS

In this paper, different simulation modelling techniques were reviewed regarding the field of reliability engineering. After discussing important issues of data-driven facility simulation, the presented modelling methods were evaluated regarding the specific applicability in DFS and in general. Its advantages, disadvantages and limitations were pointed out by means of a wind park as an illustrative example. The stated models (ARMA models, DES, ANN as well as Discrete-Time Markov Chains) can be seen as tools which can be used for establishing a comprehensive simulation concept for operational time series simulation of a technical facility. In future studies, the stated modelling methods shall be applied on the wind park data under consideration of the stated important issues of DFS. Furthermore, the methods shall be applied in combination for performing the DFS.

REFERENCES

Aksoy, H. et al. 2004. Stochastic generation of hourly mean wind speed data. *Renewable Energy* (issue 14): 2111–2131.

Banks, J. 2005. *Discrete-event system simulation*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Billinton, R., Chen, H. & Ghajar, R. 1996. Time-series models for reliability evaluation of power systems including wind energy. *Microelectronics Reliability* (issue 9): 1253–1261.

Box, G.E.P. & Jenkins, G.M. 1979. *Time series analysis*. Oakland: Holden-Day.

Cha, J.-H. & Roh, M.-I. 2010. Combined discrete event and discrete time simulation framework and its application to the block erection process in shipbuilding. *Advances in Engineering Software* (issue 4): 656–665.

Chen, K. & Yu, J. 2014. Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach. *Applied Energy*: 690–705.

Chojaczyk, A.A. et al. 2015. Review and application of Artificial Neural Networks models in reliability analysis of steel structures. *Structural Safety*: 78–89.

Davies, R., Coole, T. & Osipyw, D. 2014. The Application of Time Series Modelling and Monte Carlo Simulation—Forecasting Volatile Inventory Requirements. *Applied Mathematics* (issue 08): 1152–1168.

Feijóo, A. & Villanueva, D. 2016. Assessing wind speed simulation methods. *Renewable and Sustainable Energy Reviews*: 473–483.

Ho, S.L. & Xie, M. 1998. The use of ARIMA models for reliability forecasting and analysis. *Computers & Industrial Engineering* (issue 1–2): 213–216.

Hurtado, J.E. & Alvarez, D.A. 2001. Neural-network-based reliability analysis—A comparative study. *Computer Methods in Applied Mechanics and Engineering* (issue 1–2): 113–132.

Juan, A.A. et al. 2009. Applications of discrete-event simulation to reliability and availability assessment in civil engineering structures, *Proceedings of the 2009 Winter Simulation Conference (WSC)*, *2009 Winter Simulation Conference - (WSC 2009)*, *Austin, TX, USA*, *13.12.2009 - 16.12.2009*: IEEE.

Kamal, L. & Jafri, Y.Z. 1997. Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan. *Solar Energy* (issue 1): 23–32.

Karki, R., Hu, P. & Billinton, R. 2006. A Simplified Wind Power Generation Model for Reliability Evaluation. *IEEE Transactions on Energy Conversion* (issue 2): 533–540.

Kolonko, M. 2008. *Stochastische Simulation*. Wiesbaden: Vieweg+Teubner Verlag / GWV Fachverlage GmbH Wiesbaden.

Koutras, M.V. 1996. On a Markov chain approach for the study of reliability structures. *Journal of Applied Probability* (issue 02): 357–367.

Kruse, R.J. et al. 2012. *Computational Intelligence*. Wiesbaden: Vieweg + Teubner.

Law, A.M. 2007. *Simulation modeling and analysis*. Boston, Mass.: McGraw-Hill.

Li, G. & Shi, J. 2010. On comparing three artificial neural networks for wind speed forecasting. *Applied Energy* (issue 7): 2313–2320.

Mihalakakou, G., Santamouris, M. & Asimakopoulos, D.N. 2000. The total solar radiation time series simulation in Athens, using neural networks. *Theoretical and Applied Climatology* (issue 3–4): 185–197.

Mohandes, M.A., Rehman, S. & Halawani, T.O. 1998. A neural networks approach for wind speed prediction. *Renewable Energy* (issue 3): 345–354.

Nfaoui, H., Essiarab, H. & Sayigh, A.A.M. 2004. A stochastic Markov chain model for simulating wind speed time series at Tangiers, Morocco. *Renewable Energy* (issue 8): 1407–1418.

Papaefthymiou, G. & Klockl, B. 2008. MCMC for Wind Power Simulation. *IEEE Transactions on Energy Conversion* (issue 1): 234–240.

Rajpal, P.S., Shishodia, K.S. & Sekhon, G.S. 2006. An artificial neural network for modeling reliability, availability and maintainability of a repairable system. *Reliability Engineering & System Safety* (issue 7): 809–819.

Sahin, A.D. & Sen, Z. 2001. First-order Markov chain approach to wind speed modelling. *Journal of Wind Engineering and Industrial Aerodynamics* (issue 3–4): 263–269.

Sfetsos, A. 2000. A comparison of various forecasting techniques applied to mean hourly wind speed time series. *Renewable Energy* (issue 1): 23–35.

SHAMSHAD, A. et al. 2005. First and second order Markov chain models for synthetic generation of wind speed time series. *Energy* (issue 5): 693–708.

Sharda, B. & Bury, S.J. 2008. A discrete event simulation model for reliability modeling of a chemical plant, *2008 Winter Simulation Conference*, *2008 Winter Simulation Conference (WSC)*, *Miami, FL, USA*, *07.12.2008 - 10.12.2008*: IEEE.

Sigman, K. 2007. *Simulating Markov Chains. Notes on MC simulation*. http://www.columbia.edu/~ks20/stochastic-I/stochastic-I-MCI.pdf. 27.11.2017.

Skalny, P. & Krajc, B. 2013. Discrete-Time Markov Chains in Reliability Analysis-Case Study. In Herrero, Á., Snášel, V., Abraham, A., Zelinka, I., Baruque, B., Quintián, H., Calvo, J.L., Sedano, J. & Corchado, E. (eds), *International Joint Conference CISIS'12-ICEUTE´12-SOCO´12 Special Sessions*: 421–427. Berlin, Heidelberg: Springer.

VDI-Gesellschaft Produkt—und Prozessgestaltung. 1999. *Monte-Carlo-Simulation*. Berlin: Beuth Verlag.

Waldmann, K.-H. & Helm, W. 2016. *Simulation stochastischer Systeme*. Berlin, Heidelberg: Springer Gabler.

Welch, R.L., Ruffing, S.M. & Venayagamoorthy, G.K. 2009. Comparison of feedforward and feedback neural network architectures for short term wind speed prediction, *2009 International Joint Conference on Neural Networks*, *2009 International Joint Conference on Neural Networks (IJCNN 2009 - Atlanta)*, *Atlanta, Ga, USA*, *14.06.2009 - 19.06.2009*: IEEE.

Zeigler, B.P., Praehofer, H. & Kim, T.G. 2010. *Theory of modeling and simulation*. Amsterdam: Academic Press.

Zhang, G.P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*: 159–175.

Zio, E. 2013. *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*. London: Springer London.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Safety for automated warehouse exhibiting collaborative robots

Rafia Inam & Elena Fersman
Ericsson Research, Ericsson AB, Stockholm, Sweden

Klaus Raizer, Ricardo Souza, Amadeu Nascimento Jr. & Alberto Hata
Ericsson Research, Ericsson Telecomunicações S.A., Indaiatuba, SP, Brazil

ABSTRACT: The trend of automation in industrial production has led to massive use of autonomous robots. In classical approaches, safety is usually guaranteed by isolating robots from humans. Collaborative robots, i.e., humans and robots working together, are expected to increase both productivity and performance. However, removing fences and putting the robot working in collaboration with humans causes new hazardous situations. Therefore, proper risk assessment should be performed to avoid those hazardous situations without compromising the productivity. We present an automated warehouse where autonomous robots load trucks with products while sharing the same environment with human workers. In this position paper we propose a safety strategy that is modeled based on dynamic safety fields around the robot, which is consistent with important guidelines in collaborative robotics (i.e., ISO15066). We propose three different safety levels of dynamic fields: red (critical), yellow (warning) and green (clear). Instead of completely stopping the robot in the presence of humans it can keep performing its operations with some enforced constrains for safety reasons. We also propose a risk assessment of hazardous situations based on proprioceptive and exteroceptive data. This evaluation generates different warnings or actions to be performed based on those safety levels and is responsible for changing the size of the dynamic fields.

1 INTRODUCTION

In the present phase of industrialization, automation represents an important role, that leads towards an extensive use of robots from factories and manufacturing industries to other business areas, like in large logistic systems and warehouses. The traditional safety strategy is to separate robots and humans completely by isolating robots behind fences, and stop the robot immediately if anything goes wrong, thus minimizing the contact between robots and workers, as in Kiva project (Guizzo 2008). *Collaborative robotics*, in which robots and humans collaborate to accomplish tasks, introduces new advantages in terms of productivity and efficiency, but also new hazards, like increased possibility of collision with workers.

Avoiding hazardous situations is an absolute prerequisite for autonomous robots. Due to elimination of barriers around the robot in this new collaborative situation, a robot should interact with other robots and workers at different levels. It is crucial to ensure the correct and safe operation of the robot so that it cannot cause injuries to the workers, other objects or to itself (Robla-Gómez et al. 2017). This issue is aggravated in the warehouse scenario, where mobile robots can move freely in the presence of other moving robots and workers. The regulations that incorporate robot related risks for workers include the international standard ISO 10218 (ISO 2011a, ISO 2011b). A recent technical specification ISO/TS 15066:2016 (ISO 2016) for collaborative robots introduces new concepts, such as *collaborative operation, collaborative work-space* (a shared space where worker and robot can perform tasks concurrently), and *collaborative robot*, which are a direct focus of this work. Please note that in the rest of the paper, we use human and worker synonymously.

Many safety approaches for collaborative warehouse rely only on sensor data (e.g., cameras and Li-DARs) to perform high precision detection and tracking of the workers (Krug et al. 2016, Sabattini et al. 2017). By knowing the position of the workers, robots simply deviate from them as they come closer. Although such reactive approach may be sufficient in some situations, there are other solutions in which robot actions can be determined from its proximity to the obstacle and the type and nature of this obstacle. This leads to the *safety field* concept which creates a virtual circle around the robot, where robot reaction changes according to the region of the circle that is occupied (Magnanimo et al. 2016, SafeLog 2017).

In this position paper, we present a two-fold safety strategy and a detailed architecture including all the required components to implement safety for collaborative operations within an automated

warehouse. We base our safety analysis on creating three-layered safety fields around the robot: red (critical), yellow (warning) and green (clear). It is based on *speed and distance monitoring* and *power and force limiting* approaches of the ISO/TS 15066:2016. An advantage of using different safety fields is the increased performance of the collaborative operation. Instead of completely stopping the robot in the presence of humans it can keep performing its operations with a decreased speed. Our two-fold safety strategy consists of:

1. *An offline safety analysis:* performs a simulation-based quantitative safety assessment of the scenarios. It evaluates the high-level plans (high-level task descriptions for the robots), generated by the system before sending it to the robots.
2. *An online safety analysis:* creates the safety field around the robots within the robot control loop at runtime. It uses the sensors' data and performs *risk assessment* to calculate the risks. Based on the calculated risk values, it generates 3-layered safety fields around the robot. These fields are dynamic in size, and its sizes are computed based on the risk assessments and on both environmental and operational context of the robot.

Paper Outline: Related work is presented in Section 2. Section 3 describes details of the automated warehouse, its scenarios and the proposed architecture. Our idea for three-layered safety strategy is presented in Section 4, and finally, Section 5 concludes the paper with a description of ongoing and future works.

2 RELATED WORK

The problem of robot safety in warehouses is a relatively recent issue, though the safety problem involving robots have already been discussed for decades (Robla-Gómez et al. 2017). One of the first initiatives in warehouse automation was the Kiva project (Guizzo 2008) which used omnidirectional mobile robots to load shelves and bring them close to workers. This enables workers to pick products from the inventory and accomplish the orders efficiently. In this scenario, humans are placed in a separate area from the robot and an alarm is triggered when someone enters the robot area. In comparison to this traditional approach, our approach allows collaborative operations inside the working space and hence the safety requirements are less restrictive than keeping the robot within a limited area.

The problem of automated picking and placing of objects using mobile robots is addressed in (Krug et al. 2016), in which the robots operate in the same area as the workers. The robots are equipped with a set of safety LiDARs and camera sensors for safety assessment. Workers use reflective vests (ANSI/ISEA 107-2004 standard) to facilitate the process of human detection and make it possible even in low light conditions. The robot determines the position, velocity and the body configuration (e.g., sitting, standing) of the worker through its sensors and uses this information for safe navigation. The mobile robot conforms with the EN 1525 standard (CSN 1998) which tolerates contact with a human. The robot solution for warehouse proposed by (Sabattini et al. 2017) relies on 3D object detection using the fusion of four LiDARs and two cameras. The LiDARs ensure coverage of 360° of the scenario. The object detection and tracking is performed by both camera and LiDAR. The outputs from these sensors are used by the robot to deviate from obstacles and workers during the navigation.

In both approaches of (Krug et al. 2016, Sabattini et al. 2017), robots share a collaborative workspace with humans and a safe distance is kept to avoid collisions with objects and humans. However, the distance is calculated based on sensor data only and no risk assessment or safety analysis is performed, as we propose in this paper.

SafeLog is a more recent initiative which deals directly with safety in a collaborative scenario between robots and humans inside a warehouse (SafeLog 2017). It uses three safety levels (A, B and C) which creates a three-layered virtual circle around the worker. The safety level A creates a virtual circle around the worker which repulses the robot in a similar approach as potential fields (Taquet et al. 2017). Robots should stop when entering region covered by level A, which is the closest one to the worker. In the safety level B, robots should send a notification to the worker about the risk and the travel speed can be limited. Finally, in the safety level C, the safety system re-plans the robot and human paths to avoid close encounters between them. Further, safety vests are also used to facilitate the localization and detection of humans.

Another collaborative scenario presents a two-level dynamic safety fields creation around the moving robot (Magnanimo et al. 2016). The protection field behaves like a cage; that is, the robot immediately halts when detecting a human inside this field, while detecting any object inside the warning field results in less rigorous action like reducing the robot speed. The sizes of the fields are changed dynamically based on sensor data.

Both SafeLog and the work of (Magnanimo et al. 2016) use collaborative scenarios and are more similar to our proposed approach. However, differently, we introduce a three-layered safety field which dynamically changes the sizes of each field based on 1) the environment conditions and 2) the results of the risk assessment performed on the captured data. We also present an offline safety analysis before sending plans to the robots. Moreover, in the approach of (Magnanimo et al. 2016),

their warning field is intended for human safety assessment and does not focuses on interaction with other robots, as done in our approach.

3 WAREHOUSE LOGISTICS MANAGEMENT USING COLLABORATIVE ROBOTS

The first part of this section presents details of our logistics use case within an automated warehouse and describes collaborative and non-collaborative scenarios to be performed safely inside the warehouse. The second part of this section describes the proposed architecture and its components.

3.1 *Collaborative scenarios*

Our use case is an automated warehouse where autonomous robots and humans work together in a shared environment to perform the logistics management operations. The high-level task is to load trucks with products. Multiple robots pick up products from the shelves and deliver them to conveyor belts, that in turn take the products to the trucks. Robots can move freely in the environment to perform the tasks and are equipped with a robotic arm for pickup operations. Humans interact with shelves by placing or moving products on them. Thus, the shelves and warehouse floor are shared among humans and robots, presenting a collaborative work-space and a collaborative scenario, as shown in Figure 1. Here, the human and the robot can come into close interactions with each other, thus leading to severe safety risks.

Other situations can include the human intervention when a product is dropped by the robot and human comes to clean up, or for the maintenance of a broken robot. Proper procedures must be adopted and safety must be ensured during this operation.

Figure 1. Illustration of robots in the warehouse and the dynamic safety fields. The blue cube is the recharging station, the small boxes in red, green and brown are products on the shelf, and a white square on the floor next to the conveyor belt is a way-point. The green (clear), yellow (warning) and red (critical) circles are the safety zones around the robot.

A robot can also come close to another robot in multiple situations. Encounters can happen during navigation from shelves to conveyor belts and when the robot is coming back to the shelf after delivering the product. Another encounter can occur when two or more robots are picking up products from a different row or column of the same shelf. Although the plan provided to each robot will ensure that two or more robots will not be at the same row and column at a particular time (using offline safety analysis), still a situation can arise when the products are placed closely and there is a chance of colliding with nearby robots.

3.2 *System architecture*

This section describes the basic architecture model with some core functionalities necessary to provide all the features of an automated logistics warehouse with collaborative robots. Figure 2 presents an overview of the basic architecture.

Some physical components of Figure 2 and their respective functions within the automated warehouse are described in Section 3.1, i.e., products at the *shelves*, *robots* picking up the products from the shelves and placing on the *conveyor belt*. A *human worker* is responsible for placing products on the shelves.

To fully automate the warehouse scenario and for simulation purposes, a digital representation of all these physical components/actors is required. We use *digital twins* for this purpose, which is a digital/virtual representation of an actor/component in the system. It provides a well-known communication interface and resource description, therefore, hiding all the specific complexity of heterogeneous devices/resources. A *Warehouse Controller* component performs the main control of the system. It is responsible to perform all initial configurations and discovery mechanisms setup. When a resource becomes online it will need to register itself to the system and that, in turn, will trigger the spawning of a digital twin associated with that resource, and from that point on, all interactions between the system and the actual resources are performed through the digital twin.

The system needs to perform task planning and monitor the execution of the tasks. Thus, a *planning service* is implemented, employing PDDL traditional technique (Mcdermott et al. 1998) to generate an overall plan for the whole system[1].

While dealing with the collaborative scenarios, there is a strong requirement of assuring human

[1]Please note that the details of the warehouse controller, planning service, and digital twins are presented to depict a complete picture of the architecture and are beyond the scope of this paper.

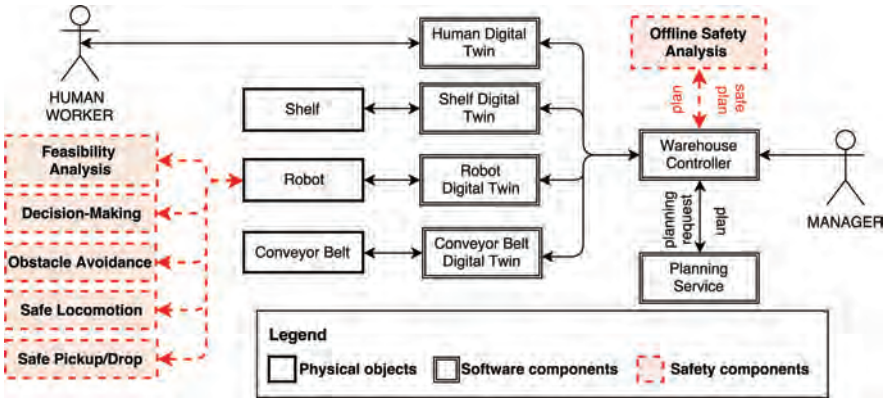


Figure 2. System architecture of the collaborative warehouse, the safety components are highlighted in red dotted boxes and arrows.

worker's safety aspects during operation. Our two-fold safety strategy is presented by dotted components and arrows in Figure 2. First, there is an initial assessment of the whole computed plan performed within the *Offline safety analysis* element. If the plan is deemed safe, the tasks are sent to the respective resources for execution. Second, during the entire operation, local *Online safety analysis* is performed within the robots to guarantee safety. It consists of multiple components namely *feasibility analysis, decision making, obstacle avoidance, safe locomotion and safe pickup*. All these safety-related functionalities are described in the next section.

4 SAFETY STRATEGY

A proper safety analysis is required for multiple collaborative scenarios in the automated warehouse (described in Section 3.1) where a robot can come close to humans, other robots or objects.

For collaborative scenarios, the ISO/TS 15066:2016 (ISO 2016) establishes a series of guiding points towards a situation where robot and human worker can share the same workspace. A first requirement on the technical specification enforces that the robots used in collaborative operations shall meet safety requirements defined in ISO/TS 10218:2011 (ISO 2011a, ISO 2011b). Secondly, the document restricts robots to have at least one of the following modes: *safety-rated monitored stop*, *hand guiding*, *speed and separation monitoring,* and *power and force limiting*. In the first method when the human is outside the collaborative workspace the robot can operate normally. For human and robots working together in the workspace, the robot must be at safety-rated monitored stop, resulting in a stop category 2 (actuators are still powered on) as defined in (IEC 2016). If any of previous conditions are violated the robot must issue a protective stop resulting in a stop category 0 (actuators are powered off). In hand guiding mode, the human must control the robot using a device located near or at the robot's end-effector. For the human to enter in the workspace, the robot must be in a stop category 2. The third method allows human and robots to be moving while both are inside the workspace, but the robot shall keep a protective distance from the human (ISO 2016). This distance varies based on some parameters (such as velocity, robot reaction time, robot distance to stop, operator velocity) and to keep that distance robot's velocity can be restricted. The power and force limiting method allows intentional and non-intentional contacts between robot and human. In this mode, the robot should keep values of force, pressure and energy transfer limited according to different parts of human body and contact situations.

Our proposed safety approach is based on both *speed and distance monitoring* and the *power and force limiting* methods. Although distance monitoring is performed throughout the process, during navigation most safety measures are based on limiting the speed of the robot. On the other hand, during the robotic arm operations, all the safety measures are done by imposing limitations on power and force.

The following sections present our main approach to achieve safety. Our key aim is to enhance safety through a dynamic system capable of adapting the robot's behavior based on its current context. To achieve this purpose, we present a *3-layered safety strategy* in Section 4.1. Sections 4.1.1 and 4.1.2 present how tactical and operational aspects of safety will be handled by different components at different moments. In the context of an automated warehouse where all devices are interconnected (i.e., share information),

make decisions and perform actions, a robust *safety analysis* should be performed. Section 4.2 briefly discusses the need of a real-time implementation of *risk assessment* approach to achieve a robust hazard management. Another important aspect to obtain effective and safe co-working in the automated warehouse is the human psychological state. This aspect of *establishing human trust* is discussed in Section 4.3.

4.1 *Three-layered safety strategy*

Our safety strategy is based on dynamic safety fields around the robot. The strategy defines three distinct levels of safety: *critical (red), warning (yellow)* and *clear (green)*. The levels classify the "degree of safety" in which the robot is currently operating in relation to outside elements such as humans and other robots while also taking into account the nature of the operation being performed by the robot itself.

The green level corresponds to a clear state, i.e., even if the robot is detecting obstacles (i.e., human, robot, warehouse infrastructure) those are at a safe distance from the observing robot. Therefore, the robot is clear to continue working with its current setup and parameters.

The yellow level is a warning state. Here, the robot has detected an obstacle closer than a certain threshold and has to *adapt its behaviour* to guarantee that it is operating in a safe state. The dimension and impact of this adjustment is dependent on its current operation and the nature of the detected obstacles. If the robot is moving inside the warehouse, it might be necessary to reduce its velocity, but if it is performing an operation with its manipulator it might be necessary to not only decrease the speed of the joints but also to change the areas within which the arm is allowed to move about.

Finally, the red level is the critical level. When the robot is at this level, a human or another robot is very close and it is under safety threat. Therefore, every action needs to ensure that neither the human, other robot nor the robot itself are injured or damaged. In most cases, the natural decision is to perform category 2 stop on the robot, i.e., stop the movement of the arm or the robot platform completely in a controlled manner.

To classify the safety state (green, yellow or red) and alter the robot's behavior accordingly, the nature of the detected object (i.e., human, robot, infrastructure) and the current context (i.e., operation being performed) should be taken into consideration. To help the human establish *trust* on the robot, it is also important to inform the close-by detected human about the robot's safety state. This can be performed through triggering a cue and/or even using augmented reality devices to get the alerts and detailed information of the robot, similar to the approach of (SafeLog 2017).

Additionally, in a truly collaborative scenario, there may arise situations where the human and robot will have to almost touch each other in order to complete some task. For instance, a scenario where a robot needs to give an item to a human or the other way around. Here, the human will most likely be very close to the robot and, based on the nature of the interaction, the decision to completely stop the robot might not be applicable. Although the robot needs to be allowed to perform some actuation to conclude the previously mentioned operation, other types of actuation should be extremely limited if not forbidden. E.g., the robot needs to be able to actuate the gripper and perhaps some other parts of the manipulator itself, but these movements should be performed under severe constraints, while the overall movement of the robotic platform should be forbidden.

We envision that an advantage of using different safety fields will be the increased performance of collaborative operations. Instead of completely stopping the robot in the presence of humans, it would be better if the robot could keep performing its operations in a constrained manner. Further, due to dynamically changing sizes of the fields based on the input data and calculated risk[2], the robot could maneuver efficiently even in small places, e.g., by reducing field sizes in the presence of no risk.

4.1.1 *Offline safety analysis*

To analyze the safety of the generated plan (generated by the planning service) from a tactical perspective, offline safety analysis is performed before actually sending those plans to robots. This analysis is implemented by running the candidate plan in a realistic simulated environment (i.e., using 3d physics-based robotics simulator such as V-REP (Rohmer, Singh, and Freese 2013) or Gazebo (2017)). The purpose of this step is to reduce the possibilities of unforeseen situations, e.g., robots moving too close to one another, which could lead to unnecessary recalculations by the planning service.

The simulation reproduces all objects, robots and other devices[3], and receives as input a starting state for the warehouse and a plan to be implemented. The offline safety analysis module returns a feasibility status (whether the plan is feasible/safe or not) and a level of risk associated to following the given plan. In order to calculate this level of risk the following measurements can be used: the number of robots crossing closely to other robots or obstacles,

[2]Risk assessment module will calculate the risk level for situations in real-time as described in Section 4.2.

[3]At this moment we are not considering modeling humans in the simulated scenario due to the complexity and unpredictability of human behavior. Future work could address this point to bring the safety evaluation closer to that of the real scenario.

the waiting time for other robots to move away in order to pick up or drop products and the number of times robots have to access the same shelf.

4.1.2 *Online safety analysis at robot*

While the offline analysis is responsible for dealing with safety at a tactical level, robots are responsible for handling most of the safety requirements at the operational level. The robots must be capable of not only detecting static and dynamic obstacles but also identifying the human workers. To detect the human workers, a variety of techniques and sensors can be applied. When a worker is detected within a minimal distance, the robot must enter a mode where its power and force are limited to a threshold to ensure no physical harm to the worker, but may still allow physical contact between them (ISO 2016). This method of operation avoids a complex tracking of all human movements and distances from the robot parts, yielding a simpler and more straightforward system.

Interactions that strictly happen among robots, have fewer safety requirements. In these situations, robots must be aware of their poses and collision avoidance techniques such as Singh and Krishna (2013) and Belkhouche (2017) can be applied. When dealing with static objects (such as shelves, conveyor belts, walls and non-interactive equipment), the robot can follow classical collision avoidance behavior like (Khatib 1986), since in those scenarios safety requirements are not so demanding.

Each robot in our scenario is modeled as an autonomous cognitive agent which is capable of highlevel decision-making[4], i.e., it has its own goals, perception, actuation and decision-making capabilities. Even though our control strategy contemplates a centralized planning service, which can perform planning at the level of the whole warehouse and send tasks to each robot, a minimum level of autonomy at each robot is necessary to deal with unforeseen events. For example, when the robot receives a high level task, its cognitive agent would then perform a local safety analysis for the received task/instructions. If the instructions are deemed unfeasible or too risky, the robot tries to find an alternative way of accomplishing the tasks. A diagram that illustrates the connection between robot, its cognitive agent, its digital twin and the environment can be seen in Figure 3.

Each robot must be able to acquire information about its current environment from a multitude of sources, of both proprio- and exteroceptive nature, and be able to fuse this information to build a complete perspective of its state and the world around it.

[4]We are implementing the high-level control of robots as cognitive agents (Laird 2012), which are capable of reactive and deliberative decision-making. We believe such an approach can produce more resilient robotic agents that are able to adapt to unforeseen scenarios.

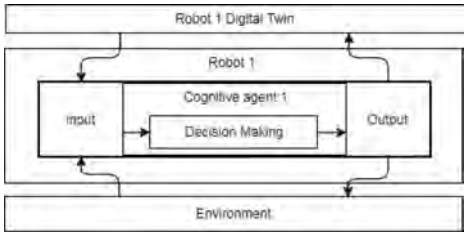


Figure 3. Cognitive control at each robot is responsible for monitoring the environment, with instructions from warehouse controller coming via digital twin. The robot then reasons about the most appropriate way to perform those instructions given what it knows about the local environment through its sensors' input.

For environment perception, we choose cameras and LiDARs to detect static and dynamic obstacles in the vicinity of the robot. Cameras can have additional application by capturing some special vests/tags on the humans (e.g., ANSI/ISEA 107-2004) to facilitate the process of human detection. Robots' set up also includes wheel encoders and IMU for self localization. As the position estimation is a key component to avoid close encounters between robots and workers, we consider the usage of radio triangulation, through technologies such as Wi-Fi, Bluetooth Low Energy, Ultra-Wideband, together with the encoders and IMU data (Jiménez and Seco 2017).

4.2 *Risk assessment*

We consider to implement a risk assessment algorithm (e.g., fuzzy logic, neural networks, or neuro-fuzzy as described in (Viharos and Kis 2015)) to calculate the risk level (i.e., high, medium, low) and then calculating the sizes of the dynamic safety fields based on the calculated risk level. The algorithm will be implemented within the cognitive agent's *decision making* module. Proprioceptive and exteroceptive data is input from the chosen sensors to the module, based on which the risk assessment algorithm will calculate the current risk level for the robot.

Depending on the level of risk calculated by the agent, module 1) changes/recalculates the sizes of the fields and 2) the robot's behavior is modified in order to increase safety accordingly, as shown in Figure 4.

For example, if no object detected within the warning or critical fields, and the area where robot has to move is narrow, then the size of the field can be reduced so that robot can move safely within the area. If a human/object is detected moving fast towards the robot then it should not only reduce the speed but also increase the size of the fields around it to avoid any hazardous situation.

For safety reasons, in our scenario the arm is not used while the robot's base is in movement. The arm

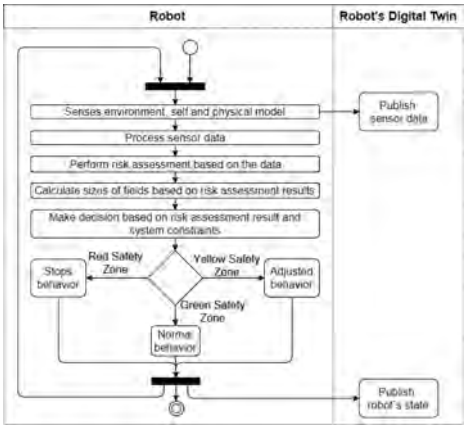


Figure 4. Workflow to execute the proposed safety strategy.

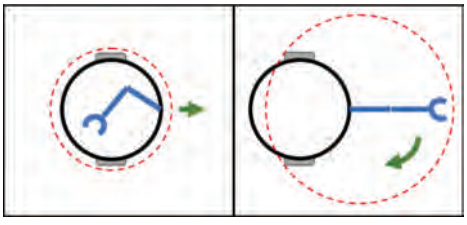


Figure 5. Difference in risk assessment for the cases when the robot is moving (to the left) and when it is standing still and moving its arm around (to the right). Green arrow indicates movement.

stands still at a position within the base's bounding box. When the robot's base is standing still, the robotic arm is then able to move for performing its pick-up and drop actions. This scenario is illustrated in Figure 5, where red dotted lines represent the areas considered to be critical for human interaction.

4.3 *Discussion on establishing human trust*

It is necessary to ensure that the worker feels comfortable and safe when cooperating with a robot, and that mental strains associated with such tasks are bearable. Three influential factors to assess the worker's mental strain (including distance, speed and warnings of motion) were varied in order to define design criteria to improve human worker comfort in (Arai, Kato, and Fujita 2010). Suitable training of the worker is reflected in the ISO/TS 15066:2016 (ISO 2016). Training clearly has an influence on his/her confidence and stress levels as well as their safety.

We propose that the robot informs the identification of human/other robot in its safety zones through a visual cue (a visual indication, for example a led panel on the robot), sound alert or augmented reality. To exemplify, if used visual cues, a red light can be displayed when human is identified inside the red zone, a yellow light for yellow zone and green light for the rest to specify that robot is performing its operation normally). In this way the human feels informed about the robot working and feels safe while performing collaborative tasks. The robot performs similar actions when another robot comes closer to it.

Most works often focus on the task of increasing trust of humans on machines. However, human trust on the automated system should not be blind. Excessive trust could be as harmful (or even more) as a lack of it. Therefore, the concept of "calibrated trust" should be explored (Lee and See 2004).

Calibrated trust tries to minimize the mismatches between trust and the capabilities of automated systems. Over-trust means poor calibration in which trust exceeds what the system is capable of delivering, and distrust means not trusting the system enough, thinking it is not capable of delivering what in fact it can. Both scenarios are undesirable because they could generate safety issues (e.g., over-trusting an autonomous car with level 3 autonomy to behave like one with level 4 autonomy *Automated driving* (...), see weblink in 'References' section) and reduced efficiency (e.g., a worker insisting on doing by hand something that is already automated).

With that in mind, we intend to increase the level of human-machine mutual understanding (Azevedo, Raizer, and Souza 2017), by making clear to the human user the intentions and motivations that led the automated system to make its decisions.

5 CONCLUSIONS AND FUTURE WORK

In this position paper, we have presented a detailed architecture to realize a safe collaborative automated warehouse scenario. We have proposed a safety strategy that combines the safety analysis performed globally by the warehouse controller (offline safety analysis) and locally by the robot itself (online safety analysis). The simulation-based offline analysis will check the safety of the high level plans for the robots at a higher level, and will ensure that collisions will not happen.

Considering the dynamism of the warehouse, unexpected changes in the robot and human plan may occur at run-time. To address unforeseen events, we have presented a three-layered safety field approach to be applied by the robots to adjust their behavior locally.

Currently, we are implementing the proposed architecture in a simulated scenario using V-REP simulator. For future works, we will extend the simulation with our proposed three-layered safety strategy. We will also implement it using physical

robots. It will consists of the implementation of the risk assessment module, and the three-layered safety strategy. Results will be evaluated using a set of Key Performance Indicators (KPIs) based on safety requirements, and the overall performance of the warehouse. We also intend to imbue each robot with learning capabilities, so it can adapt to situations that are particular to the warehouse. For instance, robot A is picking up a product at the shelf while robot B is waiting to do the same. Robot B could learn that pick up action at this particular shelf takes longer than at others, and adjust its behavior accordingly (e.g. should it keep waiting or should it move to another shelf?). Another future direction could be to extend the simulation with a human model and work on trust aspects to bring the safety evaluation closer to the real world.

ACKNOWLEDGEMENT

SCOTT (www.scott-project.eu) has received funding from the Electronic Component Systems for European Leadership Joint Undertaking under grant agreement No. 737422. This Joint Undertaking receives support from the European Unions Horizon 2020 research and innovation programme and Austria, Spain, Finland, Ireland, Sweden, Germany, Poland, Portugal, Netherlands, Belgium, Norway.

REFERENCES

Arai, T., R. Kato, and M. Fujita (2010). Assessment of operator stress induced by robot collaboration in assembly. *CIRP Annals 59*(1), 5–8.

Automated driving: levels of driving automation. https://web.archive.org/web/20170903105244/https://www.sae.org/misc/pdfs/automated_driving.pdf. Accessed: 2017-12-11.

Azevedo, C.R.B., K. Raizer, and R. Souza (2017, March). A vision for human-machine mutual understanding, trust establishment, and collaboration. In *2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, pp. 1–3.

Belkhouche, F. (2017, May). An optimal time strategy for collision avoidance between collaborative agents. In *2017 American Control Conference (ACC)*, pp. 1328–1333.

CBS (1998). Safety of industrial trucks. driverless trucks and their systems. technical report. Technical report, EN 1525.

Gazebo: robot simulation made easy. http://gazebosim.org/. Accessed: 2017-12-11.

Guizzo, E. (2008, July). Three engineers, hundreds of robots, one warehouse. *IEEE Spectrum 45*(7), 26–34.

IEC (2016). *IEC 60204-1:Safety of machinery – Electrical equipment of machines – Part 1: General requirements Robot systems and integration*. Geneva, Switzerland: International Electrotechnical Commission.

ISO (2011a, July). *ISO 10218-1 (2011): Robots and robotic devices – Safety requirements for industrial robots – Part 1: Robots*. Switzerland: International Organization for Standardization.

ISO (2011b, July). *ISO 10218-2 (2011): Robots and robotic devices – Safety requirements for industrial robots – Part 2: Robot systems and integration*. Geneva, Switzerland: International Organization for Standardization.

ISO (2016, February). *ISO/TS 15066:2016 Robots and robotic devices – Collaborative robots*. Geneva, Switzerland: International Organization for Standardization.

Jiménez, A.R. and F. Seco (2017, Sept). Finding objects using uwb or ble localization technology: A museum like use case. In *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8.

Khatib, O. (1986, April). Real-time obstacle avoidance for manipulators and mobile robots. *Int. J. Rob. Res. 5*(1), 90–98.

Krug, R., T. Stoyanov, V. Tincani, H. Andreasson, R. Mosberger, G. Fantoni, and A.J. Lilienthal (2016, Jan). The next step in robot commissioning: Autonomous picking and palletizing. *IEEE Robotics and Automation Letters 1*(1), 546–553.

Laird, J.E. (2012). *The Soar cognitive architecture*. MIT press.

Lee, J.D. and K.A. See (2004). Trust in automation: Designing for appropriate reliance. *Human factors 46*(1), 50–80.

Magnanimo, V., S. Walther, L. Tecchia, C. Natale, and T. Guhl (2016, Oct). Safeguarding a mobile manipulator using dynamic safety fields. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2972–2977.

Mcdermott, D., M. Ghallab, A. Howe, C. Knoblock, A. Ram, M. Veloso, D. Weld, and D. Wilkins (1998). PDDL – The Planning Domain Definition Language. Technical report, CVC TR-98-003/DCS TR-1165, Yale Center for Computational Vision and Control.

Robla-Gómez, S., V.M. Becerra, J.R. LLata, E. Gonzlez-Sarabia, C. Torre-Ferrero, and J. Pérez-Oria (2017). Working together: A review on safe human-robot collaboration in industrial environments. *IEEE Access PP*(99), 1–1.

Rohmer, E., S.P.N. Singh, and M. Freese (2013, Nov). V-rep: A versatile and scalable robot simulation framework. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1321–1326.

Sabattini, L., M. Aikio, P. Beinschob, M. Boehning, E. Cardarelli, V. Digani, A. Krengel, M. Magnani, S. Mandici, F. Oleari, C. Reinke, D. Ronzoni, C. Stimming, R. Varga, A. Vatavu, S.C. Lopez, C. Fantuzzi, A. Mayra, S. Nedevschi, C. Secchi, and K. Fuerstenberg (2017). The pan-robots project: Advanced automated guided vehicle systems for industrial logistics. *IEEE Robotics Automation Magazine PP*(99), 1–1.

SafeLog: safe human-robot interaction in logistic applications for highly flexible warehouses. http://safelog-project.eu/. Accessed: 2017-12-11.

Singh, A.K. and K.M. Krishna (2013, Dec). Reactive collision avoidance for multiple robots by non linear time scaling. In *52nd IEEE Conference on Decision and Control*, pp. 952–958.

Taquet, J., G. Écorchard, and L. Preucil (2017). Realtime visual localisation in a tagged environment. *CoRR abs/1708.02283*.

Viharos, Z. and K. Kis (2015). Survey on neuro-fuzzy systems and their applications in technical diagnostics and measurement. *Measurement 67*(Supplement C), 126–136.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A flexible simulation model of the operation and maintenance process of a complex technical system

J. Malinowski
Systems Research Institute, Polish Academy of Sciences, Warsaw School of Information Technology, Warsaw, Poland

ABSTRACT: The paper presents a versatile simulation model of the operation process of a complex technical system. Among other features, the model assumes dependencies between the system's components, and a priority-based maintenance/repair policy implemented by a limited personnel. Over the recent decades various maintenance models were investigated with the aim of optimizing the maintenance costs. Most researchers pursued analytical approach, hence the obtained results hold true under quite restrictive assumptions imposed for the purpose of analytical tractability (mutual independence of components, exponential time-to-failure and time-to-repair, no waiting time for repair, etc.). The model described herein is free from such limitations, as it is designed to compute the system's reliability or performance indices by way of simulation rather than analytically. The following features make it more realistic than many other models from the relevant literature: 1) the components are mutually dependent, i.e. a component's state change affects the failure rates of some other components, 2) repairs can be perfect, imperfect, or minimal, 3) due to a limited maintenance personnel, a failed component may have to await its repair in a priority queue, where the priority level depends on the component's importance, 4) a component state may be unobservable, in which case hidden failures are revealed by inspections. Using such a model, the optimal or near optimal parameters of an assumed maintenance policy can be found by repeated simulation. Any performance index expressed as a priori known function of these parameters can be subject to optimization. The author's main result is a non-trivial simulation algorithm encompassing the model's comprehensiveness. Clearly, the simulation approach, although time-consuming, allows to avoid elaborate analytical derivations which, in the absence of simplifying assumptions, become unreasonably complicated or impossible as the system's complexity increases.

1 INTRODUCTION

Over the recent decades numerous researchers investigated various maintenance models with the aim of optimizing the maintenance policies of the modeled systems. They mainly used analytical approach, hence most of the obtained results hold true under quite limiting assumptions imposed for the purpose of analytical tractability (mutual independence of components, exponential TTF/TTR, no waiting time for repair). See Couchan et al. (2013), Nakagawa (2008a), Nakamura et al. (2017), Sarkar et al. (2011), and Wang (2002) for surveys of more or less recent works on the subject.

The model presented in the current paper is free from such limitations, as it is constructed for the purpose of simulating the system's behavior rather than analytically computing its reliability or performance indices. This model assumes that the system is composed of two-state mutually dependent components, which means that the aging rate of a component may depend on the states of other components. When failed, a component is either replaced or repaired, depending on its wear-related age measured by integrating the component's changeable aging factor over time up to the moment of failure. A component is also replaced on reaching the age limit for its operation, even if it is still in operating condition (preventive replacement). Repairs and replacements are performed by a limited maintenance personnel, hence a failed component may have to wait in a maintenance queue for being serviced. There are multiple queues corresponding to various priorities assigned to individual components. The components waiting in a higher priority queue have precedence over those in the lower priority queue, and each queue is a FIFO sequence.

Based on the above model, a simulation framework for the reliability analysis of a complex technical system is constructed. By assumption, the simulation begins at time $t_0 = 0$, when all the components are new and operable. The simulation procedure produces a sequence of time instants at which changes of components' states occur, i.e. if

t_k is the time when a component changes its state, then t_{k+1} is the time of the subsequent state change of the same or another component, $k \geq 1$. More accurately, t_1 is the time of the first component failure, and either of the following events takes place at each subsequent t_i, $i > 1$:

- a component fails
- a component's hidden failure is detected
- a component's age reaches age limit for operation
- a component's repair or replacement is completed

Depending on a component, its failures can be self-revealing or hidden. In the first case a failure and its detection are simultaneous events; in the second case a failure can only be detected at the next inspection or preventive replacement, whichever comes first, and until then the component remains in the state of undetected failure. A binary vector defined in the next section is used to distinguish between the two types of components. Upon a failure detection, either the component's repair or replacement is started (if at least one maintenance team is available), or (if all the teams are busy) the component is placed in the respective maintenance queue, following the rules given in section 4.

The proposed simulation model has a number of changeable (adjustable) parameters. One of the key parameters is the age limit on a failed component's repair; if a component's failure is detected before its operating age has reached that limit, then the component is scheduled for repair, otherwise it is scheduled for replacement. Another key parameter is the age limit for a component's operation; when this limit is reached, the component is put out of operation and scheduled for replacement regardless of its state. Clearly, the value of the former parameter should be smaller than the value of the latter one.

Each failure, repair, and replacement incurs a cost, and so does a sojourn in a failed state. The system operator's aim is to minimize the overall operating cost over a certain period of time, or the average operating cost per unit time over an indefinitely long period. This aim can be achieved by appropriate adjustment of the aforementioned changeable parameters (decision variables). This can only be done by means of the trial-and-error method combined with the repeated simulation, as the model's complexity practically excludes analytical optimization methods.

2 ACRONYMS AND NOTATION

CDF – cumulative distribution function
PDF – probability density function
TTF – time-to-failure
TTR – time-to-repair
IFR – increasing failure rate
n – number of components
$e_1,\ldots,e_n$ – the individual components
Q – number of priority levels (maintenance queues) for maintenance scheduling
R – number of maintenance teams
D(i) – set of components whose states affect the aging rate of e_i; let us note that $i \notin D(i)$
$X_i(t)$ – the reliability state of e_i at time t; $X_i(t) = 1$ if e_i is in operation, and $X_i(t) = 0$ if e_i is failed or under repair
$Y_i(t)$ – the operational state of e_i at time t; $Y_i(t) \leq -1$ when e_i is awaiting repair, $Y_i(t) = 0$ when e_i is undergoing repair, $Y_i(t) = 1$ when e_i is in operation, and $Y_i(t) = 2$ when e_i is in the state of undetected failure
$a_i(t)$ – the aging rate of e_i at time t; $a_i(t)$ can differ during e_i's lifetime and, if $Y_i(t) = 1$, it is dependent on the states of e_j, $j \in D(i)$
$A_i(t)$ – the age of e_i at time t, defined as $\int_{[0,t]} a_i(s)\,ds$, where t is counted from the last moment when e_i was first put in operation or replaced;
$T_i(t)$ – the prospective sojourn time of e_i in the state $Y_i(t)$, as counted from t, provided that the aging rate after t is constant and equal to a(t). The prospective and actual sojourn times may differ, because e_i's aging rate may change before the prospective time elapses.
V[i] – age limit for e_i's repair, i.e. if t is the time of e_i's failure (or its detection) and $A_i(t) \geq V[i]$ then e_i will be replaced rather than repaired
W[i] – age limit for e_i's operation i.e. e_i will be replaced at time t, regardless of its state, when $A_i(t)$ reaches W[i]. Clearly, W[i] > V[i]
dtc[i] – a binary variable stipulating whether failures of e_i are self-revealing; if so, then dtc[i] = 1, otherwise dtc[i] = 0
S – time between consecutive inspections that are necessary to reveal failures of e_i for which dtc[i] = 0
A_i^* – age of e_i at its failure or preventive replacement, whichever comes first; A_i^* is simulated each time when e_i enters state 1, i.e. when new, replaced, or repaired e_i is put in operation
sim(i, 1, A) – a function simulating the random age of e_i at which its failure will occur, given that e_i is in state 1 and its already accumulated age equals A
sim(i, 0, A) – a function simulating the random duration of e_i's repair (A < B[i]) or replacement ($A \geq B[i]$), given that A is the already accumulated age of e_i
prt[i] – the maintenance priority level of e_i
len[q] – the current length of the repair queue no. q
ind[q, r] – the index of the component awaiting its turn in the r-th place in the q-th queue
avl_r – the number of currently available repair teams

$C_f[i]$, $C_{rpr}[i]$, $C_{rplc}[i]$ – the costs of e_i's failure, repair, or replacement respectively

$c_{und}[i]$ – the unit cost of e_i's sojourn in state 2 (undetected failure)

$C_{idl}[i]$ – the unit cost of e_i's sojourn in idle (≤ 0) state (awaiting or under maintenance)

c_{team} – the unit cost of employing one team

c_A – average total operating cost per unit time

3 A COMPONENT'S LIFE CYCLE

The functioning of a single component can be modeled by a stochastic process with the state-space $\{\ldots, -1, 0, 1, 2\}$. The assignment of individual states is shown below:

–1 or less: a component placed in a repair queue,
0: a component under repair,
1: an operable component,
2: a component with undetected failure.

If a component is in a "negative" state, then this state's absolute value of determines the component's place in the respective repair queue. Fig. 1 illustrates all possible inter-state transitions. Let us note that all the "negative" states are grouped into one.

The above diagram can be simplified if a component's failures are self-revealing, i.e. it never enters state 2. In such a case the node representing the state 2 can be deleted along with the adjoining links.

The diagram in Fig. 2 illustrates the transitions to or from the "negative" states, i.e. placements in the repair queue, forward moves in the queue, and repairs completions. It is assumed that multiple repairs are never completed simultaneously (simultaneous completion is an event of probability 0).

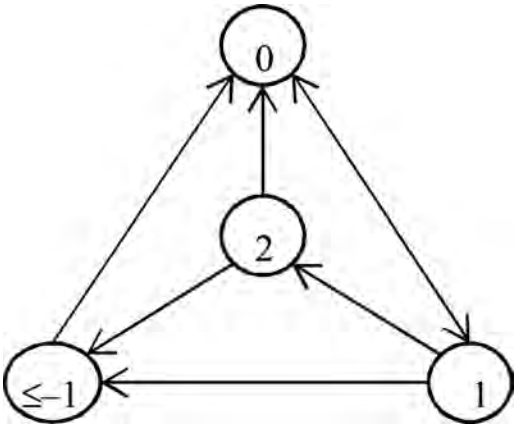


Figure 1. The diagram of a component's inter-state transitions.

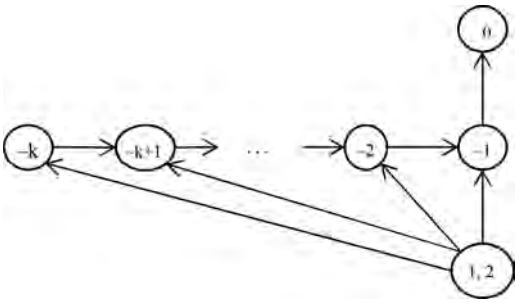


Figure 2. A diagram of transitions to or from the "negative" states.

Otherwise, transitions from – k to min(–k+s,0), where $k \geq 2$ and $2 \leq s \leq r$, would also be possible, which would increase the diagram's complexity.

4 THE MAINTENANCE POLICY RULES

A failure of e_i is revealed immediately if $dtc[i] = 1$. In turn, if $dtc[i] = 0$, then e_i's failure is revealed at the time of the next inspection. When this happens, e_i's maintenance is performed immediately provided that the maintenance personnel is available, otherwise e_i must wait for its turn. Furthermore, when the age of e_i reaches $W[i]$, it is either replaced or scheduled for replacement, no matter what its state is (if e_i is still operating, its operation is stopped). The components awaiting maintenance are placed in Q (logical) queues, corresponding to different maintenance priority levels. The levels assigned to individual components are stored in one-dimensional array $prt[\cdot]$, where $prt[i]$ is the level assigned to e_i. Components on priority level q are placed in the q-th queue, $1 \leq q \leq Q$. Each queue has the FIFO property, i.e. components with equal priority are placed in the respective queue in the order of their failure times—the recently failed one is placed at the end of the queue. The components in the q-th queue are scheduled for maintenance ahead of those in the (q+1)-st queue, $1 \leq q \leq Q–1$, i.e. the lower the component's priority level, the sooner its maintenance will start. Summing up, the component e_i awaits repair in the queue whose number is given by $prt[i]$, and the (negative) state of e_i determines its place in this queue.

If e_i is repairable and the time of its maintenance arrives, then e_i undergoes repair, provided that its age has not reached the limit $V[i]$. If e_i's age is greater or equal to $V[i]$, then it undergoes replacement rather than repair. It is necessary to be aware of the difference between $V[i]$ and $W[i]$. $V[i]$ is the age limit for e_i's repair. If e_i's age reaches or exceeds $V[i]$ and e_i fails, then e_i is only subject to

replacement. W[i] is the age limit for e_i's operation. If e_i's age reaches or exceeds W[i], then e_i will be replaced even if it is in the operating state.

5 THE SIMULATION ALGORITHM

According to the maintenance model presented in the previous sections, the following algorithm, simulating the system's operation, will now be presented. The core of this algorithm is the procedure generating the sequence $(t_k, k \geq 1)$ which, as indicated in the Introduction, is the sequence of time instants at which components change their operational states. The algorithm operates in a loop consisting of the following steps:

1. At $t_0 = 0$ all variables characterizing the maintenance process, defined in the Notation section, are given their initial values. The respective pseudocode is given below.

   k ← 0; t_0 ← 0; len[q] ← 0, q ∈ {1,…,Q}; avl_r ← R;
   for i = 1,…,n do {
   ··$Y_i(t_0)$ ← 1; $A_i(t_0) = 0$; A_i^* ← min(sim(i,1,0), C[i]); ··$T_i(t_0)$ ← $A_i^*/a_i(t_0)$
   }

   Since all components are in state 1 at t_0, their prospective sojourn times in this state are simulated with the use of sim(i, 1, 0). In order to obtain $T_i(t_0)$ the simulated age has to be divided by $a_i(t_0)$, as follows from the definitions of $A_i(t)$ and $T_i(t)$.
2. Increase k by 1
3. Compute t_k by taking the smallest $T_i(t_{k-1})$ over all e_i that are in non-negative states at t_{k-1}; the components in negative states, i.e. those awaiting repair, are irrelevant to the system behavior in the interval $[t_{k-1}, t_k)$. Also, add $a_i(t_{k-1})\cdot(t_k - t_{k-1})$ to $A_i(t_k)$ for each e_i that is not in state 0 at t_{k-1} (the aging rate of a component under maintenance equals 0).
4. Assign $Y_i(t_{k-1})$ to $Y_i(t_k)$ for each i, so that each component that does not change its state at t_k will remain in the same state in which it was at t_{k-1}.
5. Compute $Y_i(t_k)$ for each e_i that is in non-negative state at t_{k-1} and changes its state at t_k. Each such e_i is identified by checking if its sojourn time in the state $Y_i(t_{k-1})$, as counted from t_{k-1}, equals $t_k - t_{k-1}$.[1] For any e_i fulfilling this condition we have:

   1. if $Y_i(t_{k-1}) = 1$, then e_i changes its state to 2 at t_k provided that dtc[i] = 0 and inspection or preventive replacement does not coincide with t_k; otherwise (i.e. dtc[i] = 1, or dtc[i] = 0 and inspection or preventive replacement is performed at t_k) e_i is queued for maintenance, i.e. e_i enters an appropriate negative state at t_k.
   2. if $Y_i(t_{k-1}) = 2$, e_i is queued for maintenance at t_k,
   3. if $Y_i(t_{k-1}) = 0$, e_i's maintenance is completed at t_k, i.e. e_i changes its state to 1 at t_k.

   In case (3) one maintenance team is released, hence avl_r is increased by 1. Moreover, if e_i's age was greater or equal to V[i] when e_i entered state 0, e_i underwent replacement in state 0, hence its age is set to 0 at t_k. In view of the above analysis, step 5 is implemented in the following way:

   for each i such that $Y_i(t_{k-1}) \geq 0$ and $T_i(t_{k-1}) = t_k - t_{k-1}$ do {
   ··if ($Y_i(t_{k-1}) = 1$ and dtc[i] = 0 and $\lceil t_k/S \rceil \cdot S \neq t_k$
   ····and $A_i(t_k) \neq W[i]$)
   ··then $Y_i(t_k)$ ← 2
   ··else {q ← prt[i]; len[q] ← len[q] + 1;
   ····$Y_i(t_k)$ ← –len[q]}
   ··if ($Y_i(t_{k-1}) = 2$)
   ··then {q ← prt[i]; len[q] ← len[q] + 1;
   ····$Y_i(t_k)$ ← –len[q]}
   ··if ($Y_i(t_{k-1}) = 0$)
   ··then {$Y_i(t_k)$ ← 1; avl_r ← avl_r + 1;
   ····if $A_i(t_k) \geq V[i]$ then $A_i(t_k)$ ← 0}
   }
6. If avl_r > 0, set to 0 the states at t_k of at most avl_r components placed at the head of the repair queue, then update avl_r and the repair queue accordingly.
7. Compute the cost incurred during $(t_{k-1}, t_k]$ and add it to the total cost. The respective pseudocode is given below.
   C ← C + $(t_k - t_{k-1})\cdot R\cdot c_{team}$;
   for i = 1,…,n do {
   ··if ($Y_i(t_{k-1}) \leq 0$) then
   ····C ← C + $(t_k - t_{k-1})\cdot c_{idl}$[i];
   ··if ($Y_i(t_{k-1}) = 2$) then
   ····C ← C + $(t_k - t_{k-1})\cdot c_{und}$[i];
   ··if ($Y_i(t_{k-1}) \neq 0$ and $Y_i(t_k) = 0$) then
   ····if ($A_i(t_k) < V[i]$)
   ·····then C ← C + C_{rpr}[i]
   ·····else C ← C + C_{rplc}[i];
   ··if ($Y_i(t_{k-1}) = 1$ and $Y_i(t_k) \neq 1$ and fail[i] = 1) then
   ····C ← C + C_f[i];
   }

   Remark 1: The cost of repair or replacement is added when the respective action starts, because it is only then known whether ei's age has reached V[i].

   Remark 2: The last "if" adds C_f[i] to the total cost if e_i fails at t_k; fail[i] is defined in the explanation to step 8.

[1] Possible state changes at t_k of the components awaiting repair in the interval $[t_{k-1}, t_k)$ are secondary w.r.t. state changes of those which are in non-negative states in this interval. The former components' states at t_k are computed in step 6.

8. Determine $T_i(t_k)$ if e_i changes its state to a non-negative one at t_k, or e_i remains in state 1, but at least one e_j in D(i) changes its state. Otherwise assign $T_i(t_{k-1}) - (t_k - t_{k-1})$ to $T_i(t_k)$. The binary value fail[i] is set to 1 if e_i enters state 1 at t_k (e_i's repair or replacement is completed at t_k) and the age of e_i at its failure (simulated at t_k) doesn't exceed W[i] (e_i fails before or when it reaches its planned replacement age); fail[i] is used in cost calculation in step 7. Step 8 is implemented by following pseudocode:

```
for i = 1,…,n do {
··fail[i] ← 0;
··if (Yi(tk) ≠ Yi(tk–1) ) then {
····if ( Yi(tk) = 0) then
·····Ti(tk) ← sim(i, 0, Ai(tk));
····if (Yi(tk) = 1) then {
·····Ai* ← min( sim(i, 1, Ai(tk)), W[i]);
·····Ti(tk) ← (Ai* – Ai(tk))/ai(tk);
·····if (Ai* ≤ W[i]) then fail[i] ← 1
····}
····if (Yi(tk) = 2) then
·····Ti(tk) ← min( ⌈tk/S⌉·S—tk, (W[i]–Ai(tk))/ai(tk)
)
··}
}
```

Remark 1: cf. step 5 for the conditions of entering state 2.

```
··else if (Yi(tk) = Yi(tk–1) = 1
····and Yj(tk) ≠ Yj(tk–1) for at least one j∈ D[i] )
····then Ti(tk) ← (Ai* – Ai(tk))/ai(tk);
```

Remark 2: If any e_j, $j \in D(i)$, changes its state at t_k then e_i's aging factor also changes, hence $a_i(t_k) \neq a_i(t_{k-1})$ and $T_i(t_k) \neq T_i(t_{k-1}) - (t_k - t_{k-1})$, thus $T_i(t_k)$ must be calculated using the new aging factor.

```
··else
····Ti(tk) ← Ti(tk–1) – [tk – tk–1]
}
```

The function sim(i, 1, A) simulates, using a random number generator, the age of e_i at which its failure will occur, provided that e_i is in state 1 and its already accumulated age equals A. This function is called when e_i is put in operation, and then A = 0 for new or replaced e_i, or $0 < A < V[i]$ for a repaired e_i. The simulated value is used to calculate A_i^* which, divided by e_i's aging rate, yields the prospective sojourn time of e_i in state 1 from the moment when e_i enters state 1, provided that e_i's aging rate remains unchanged during e_i's sojourn in state 1. Since e_i's aging rate depends on the states of e_j, $j \in D(i)$, it is necessary to recalculate $T_i(t)$, using $a_i(t)$, at each time t when any e_j, $j \in D(i)$, changes its state during e_i's sojourn in state 1.

The function sim(i, 0, A) simulates the duration of e_i's repair or replacement depending on whether $A < V[i]$ or $A \geq V[i]$ respectively. If e_i enters state 2, i.e. $Y_i(t_{k-1}) \neq Y_i(t_k) = 2$, then $T_i(t_k)$ is calculated as $\min(\lceil t_k/S \rceil \cdot S - t_k, (W[i] - A_i(t_k))/a_i(t_k))$, because e_i must wait in state 2 until the time of the next inspection or until e_i's age reaches W[i], whichever comes sooner.

9. Go to step 2

For illustration, the algorithm's block diagram is shown in Fig. 3.

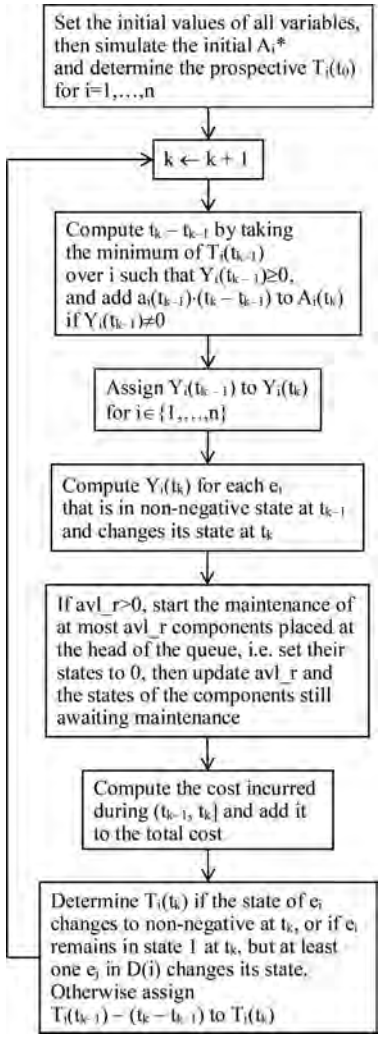


Figure 3. The simulation algorithm's block diagram.

6 THE ILLUSTRATIVE EXAMPLE

We will now present the numerical results obtained by the program based on the algorithm from the

previous section. As the complete version of the program is still under preparation, the results come from its simplified version developed according to the following assumptions:

- The components are independent, i.e. a component's aging rate does not depend on the other components' behavior and is constant during a component's lifetime ($T_i(t_k)$ needs not to be recalculated with the new aging rate when e_i remains in state 1 and another component changes its state at t_k).
- The network is built of non-repairable components; if a component fails, it is always replaced, irrespectively of its age. TTF and TTR of e_i are simulated by the functions sim(i,1) and sim(i,0) respectively. They do not variable A, because the age of a new or replaced component equals 0, and the age of an old component is irrelevant to its replacement time. In consequence, there is no age limit for e_i's repair, thus the only age limit—W[i] – is imposed on e_i's operation time.
- The failures of all components are self-revealing (dtc[i] = 1 for i = 1,…,n), thus no inspections need to be performed and the set of component states does not include state 2.

The example system has the following parameters:

Number of components: 5
Type of distribution of a component's TTF: Weibull
Scale parameter for TTF: 0.05
Shape parameter for TTF: 1.5
Type of distribution of a component's TTR: Weibull
Scale parameter for TTR: 0.5
Shape parameter for TTR: 1.5
Time unit: 1 day
Number of maintenance queues: 2
Components with priority 1: e_1, e_2, e_3
Components with priority 2: e_4, e_5
Number of maintenance teams: 1, 2, 3
C_rplc[i] – 1000
C_fail[i] – 4000
c_idl[i] – 900
c_team – 40

For the sake of this paper the following definition is adopted: a (two parameter) random variable T has Weibull distribution if the CDF of T is given by the following formula:

$$\Pr(T \le t) = 1 - \exp[-(\lambda t)^{\alpha}] \quad (1)$$

where λ and α are the scale and shape parameters respectively. It should be noted that in our example α is equal to 1.5 which is greater than 1. This is due to the fact that, as follows from the reliability theory, planned replacements are only cost-effective if the distribution of a component's TTF has the IFR property (see Barlow & Proschan (1975)). The Weibull distribution has this property only if $\alpha > 1$. Other than Weibull distributions of TTF and TTR are discussed in Barlow & Proschan (1975) and O'Connor & Kleyner (2011).

Table 1. The results of cost calculation for various W and R (* denotes the local minimum of c_A).

W	k_fin = 10.000 c_A for R = 1, 2, 3	k_fin = 100.000 c_A for R = 1, 2, 3
60	1815, 1732, 1801	1792, 1747, 1786
50	1778, 1721, 1783	1790, 1743, 1777
45	1807, 1715, 1776	1784, 1737, 1787
40	1774, 1725, 1788	1774*, 1739, 1772
35	1776, 1734, 1777	1782, 1725, 1765
34	1799, 1743, 1769	1778, 1726, 1763
33	1763, 1759, 1750	1794, 1722, 1757
32	1771, 1739, 1762	1786, 1724, 1752
31	1769, 1721, 1764	1774*, 1724, 1759
30	1771, 1717, 1753	1781, 1713*, 1753
29	1766, 1704, 1758	1779, 1720, 1748
28	1776, 1714, 1728	1782, 1718, 1746
27	1777, 1702, 1739	1784, 1720, 1759
26	1779, 1727, 1754	1788, 1719, 1742*
25	1776, 1704, 1746	1789, 1709*, 1750
24	1781, 1709, 1758	1787, 1719, 1748
23	1795, 1693, 1792	1800, 1716, 1747
22	1814, 1690, 1763	1807, 1717, 1742*
21	1790, 1681*, 1725	1808, 1714, 1754
20	1796, 1737, 1743	1813, 1717, 1759
19	1835, 1725, 1760	1829, 1729, 1756
18	1846, 1730, 1772	1844, 1732, 1759

Our aim is to find the optimal age limit for a component's operation, i.e. the optimal age at which its preventive replacement should take place. The objective function to be minimized is the total operating cost per unit time. Let W*[i] denote the optimal age for e_i. As the components are stochastically identical, W*[i] are equal for all five components, i.e. W*[1] = … = W*[5] = W*. The obtained results of cost calculation for various values of W = W[1] = … = W[5] are given below:

The values of c_A in the first column are obtained for k_fin = 10.000, where k_fin is the number of cycles of the algorithm's main loop. Based on this accuracy (k_fin can serve as a measure of the simulation's accuracy) we conclude that c_A attains its minimum for W = 21 and R = 2. However, it is advisable to perform a more accurate simulation, because we can see that c_A fluctuates in the sampled interval $18 \le W \le 60$ and for each R it has more than one local minimum there, therefore a

more precise analysis of c_A's behavior is indicated. The computations for k_fin = 100.000 show that c_A as a function of W is not unimodal (has more than one local minimum) for each R = 1, 2, 3. The respective local minima are marked with asterisks. The optimal number of maintenance teams and the optimal value of W are equal to 2 and 25 respectively. Also, let us note that the intervals $30 \le W \le 45$ and $20 \le W \le 32$ for R = 1 and R = 3 respectively are uncertain" as regards determining the optimum value of W, but these values of R are of no practical interest, as the optimum number of maintenance teams equals 2.

7 CONCLUSION AND FURTHER RESEARCH

A newly developed algorithm simulating the maintenance process of a complex technical system, and the results of a computer program implementing this algorithm, have been presented. The author aimed at developing a software tool that would enable the system operator to assess the total cost incurred over a long operating time, and to minimize it by properly adjusting the decision variables, i.e. R, S, and V[i], W[i], i = 1,…,n, as defined in Section 2. Due to the complexity of the system model the analytical computation and optimization methods cannot definitely be used, thus Monte Carlo simulation combined with heuristic optimization is the only feasible solution.

The algorithm demonstrated in the current paper bears resemblance to the one presented in George-Williams & Patelli (2015). However, our approach focuses on the simulation of time points in which any system component changes its operational state. The time axis is divided into state-invariant intervals (no component changes its state in any such interval) thus allowing for very accurate analysis of the system operation process. It is also important that the components are mutually dependent and a component's aging rate depends on other components' states. The components states distinguished in the last-mentioned paper are different than in the present one, but the model considered herein is adjustable in this respect. Also, since failed components may await repair in a priority queue, our model makes it possible to analyze the system from the queuing theory (QT) viewpoint, enabling its operator to optimize the relevant (QT related) parameters.

The current paper defines the system operation cost as the sum of individual components' operation costs. However, the state of the whole system, expressed as an a priori defined function of the components' states (the structure function) can also be included in the overall cost calculation. Apart from the cost, other performance characteristics (e.g. the system availability) can be easily calculated, in a similar way as in step 7 in section 5.

Clearly, the main drawback of stochastic simulation is its high time complexity. The simplified version of the program, based on the assumptions from Section 6, executes in approximately 12 seconds if k_fin = 10.000, and 120 seconds if k_fin = 100.000 (on a PC machine with Intel® Core™ i5 CPU). This shows that, as expected, the execution time increases linearly with k_fin. Admittedly, the above execution times are relatively short as far as Monte Carlo simulation is concerned. However, this is due to a small number of components (5). If their number is changed to 10, the execution time increases to 22 seconds (k_fin = 10.000), which means that the algorithm's time complexity increases with n somewhat slower than at a linear pace. This is explained by the fact that steps 3 through 8 of the algorithm, except step 6, are implemented as "for" loops with n cycles, while step 6 requires avl_r (i.e. at most R) compound operations, and in a real system R is much smaller than n. Obviously, if the components are not identical (as in the provided example) and their number is high then the program may have to be run a large number of times in order to find the optimum values of the decision variables. However, the components in many systems can be divided in a number of groups such that the components in one group have identical parameters, which significantly reduces the number of decision variables, and, as a consequence, decreases the time complexity of the optimization task.

Summing up, we have endeavored to develop a possibly comprehensive algorithm that would encompass a wide range of maintenance models found in the relevant literature, and would also be applicable for practical purposes. It seems that this goal has been at least partly achieved, but there still remains considerable conceptual work to be done. The further research will be mainly concentrated on properly defining the dependence relations among the system's components, i.e. how the states of components in the set D(i) affect the aging rate of e_i. The issue of components dependence is considered in the following works: Zhang & Horigome (2001), Dukhovny & Marichal (2012), Yang et al. (2013), Nakamura et al. (2017), Zhang & Wilson (2017). Also, the algorithm should support various failure modes. Then, whether a failure is self-revealing or not may depend both on its mode and the failed component's type. Last but not least, the impact of external conditions (e.g. ambient temperature and humidity) on the components' aging rates should be taken into account in constructing future versions of the presented algorithm.

REFERENCES

Barlow, R.E. & Proschan, F. 1975. Statistical Theory of Reliability and Life Testing; Probability Models. Holt, Rinehart and Winston.

Chouhan, R. & Gaur, M. & Tripathi, R. 2013. A Survey of Preventive Maintenance Planning Models, Techniques and Policies for an Ageing and Deteriorating Production Systems. In Raj Gaurav Mishra (ed.), *HCTL Open International Journal of Technology Innovations and Research* 3: 89–107.

Dukhovny, A. & Marichal, J.-L. 2012. Reliability of Systems with Dependent Components Based on Lattice Polynomial Description. *Stochastic Models* 28: 167–184.

George-Williams, H. & Patelli, E. 2015. Monte-Carlo Based Reliability/Availability Analysis Algorithm for Efficient Maintenance Planning. *Transactions, 23-rd Conference on Structural Mechanics in Reactor Technology*, Manchester, UK, Aug. 10–14 2015.

Jardine, A.K.S. & Tsang, A.H.C. 2013. *Maintenance, Replacement, and Reliability*. CRC Press, Taylor & Francis Group.

Nakagawa, T. 2008a. *Advanced Reliability Models and Maintenance Policies. Springer Series in Reliability Engineering*. London: Springer.

Nakagawa, T. 2014b. *Random Maintenance Policies. Springer Series in Reliability Engineering*. London: Springer.

Nakamura, S. & Qian, C.H. & Nakagawa, T. (eds.) 2017. *Reliability Modeling with Computer And Maintenance Applications*. Singapore: World Scientific.

O'Connor, P. & Kleyner, A. 2011. *Practical Reliability Engineering*. John Wiley & Sons.

Rubinstein, R.Y. & Kroese, D.P. 2008. *Simulation and the Monte Carlo Method, 2-nd Edition*. John Wiley & Sons

Sarkar, A. & Panja, S.C. & Sarkar, B. 2011. Survey of Maintenance Policies for the Last 50 Years. *International Journal of Software Engineering & Applications* (IJSEA) 2(3): 130–148.

Wang, H. 2002. A survey of Maintenance Policies of Deteriorating Systems. *European Journal of Operations Research* 139(3): 469–489.

Yang, Q. & Zhang, N. & Hong, Y. 2013. Reliability Analysis of Repairable Systems With Dependent Component Failures Under Partially Perfect Repair. *IEEE Transactions on Reliability* 62(2): 490–498.

Zhang, T. & Horigome, M. 2001. Availability and Reliability of System With Dependent Components and Time-Varying Failure and Repair Rates. *IEEE Transactions on Reliability* 50(2): 151–158.

Zhang, X. & Wilson, A. 2017. System Reliability and Component Importance Under Dependence: A Copula Approach. *Technometrics* 59(2): 215–224.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Feasibility study of a simulation driven approach for estimating reliability of wind turbine fluid power pitch systems

Jesper Liniger & Mohsen Soltani
Department of Energy Technology, Aalborg University, Esbjerg, Denmark

Henrik C. Pedersen
Department of Energy Technology, Aalborg University, Aalborg East, Denmark

Nariman Sepehri
Department of Mechanical Engineering, University of Manitoba, Winnipeg, Canada

ABSTRACT: Recent field data indicates that pitch systems account for a substantial part of a wind turbines down time. Reducing downtime means increasing the total amount of energy produced during its lifetime. Both electrical and fluid power pitch systems are employed with a roughly 50/50 distribution. Fluid power pitch systems generally show higher reliability and have been favored on larger offshore wind turbines. Still general issues such as leakage, contamination and electrical faults make current systems work sub-optimal. Current field data for wind turbines present overall pitch system reliability and the reliability of component groups (valves, accumulators, pumps etc.). However, the failure modes of the components and more importantly the root causes are not evident. The root causes and failure mode probabilities are central for changing current pitch system designs and operational concepts to increase reliability. This paper presents a feasibility study of estimating pitch system reliability based on a failure rate prediction method for generic fluid power components. Special attention is given to the use of computer simulations for assessing working conditions such as flow, pressure, work cycle, fluid contamination concentration etc. The fluid power pitch system is co-simulated with the 5MW NREL wind turbine implemented in the FAST software. The estimated failure rates is compared to field data and comments are given to the correlation and discrepancies based on the uncertainties of the simulated conditions.

1 INTRODUCTION

Pitch systems are today employed on all modern multi-megawatt turbines and enable the turbine blades to rotate along their longitudinal axis in order to facilitate aerodynamic braking. This is used at wind speeds above rated and also for enabling safe emergency stopping of hub rotation. Multiple studies on turbine reliability and downtime have indicated the pitch system to be the most unreliable sub-system of the turbine (Wilkinson and Hendriks 2010, Carroll et al. 2015). Contributing to over 20% of total downtime, pitch systems not only introduce high risk, they also cause a significant loss of power production during the lifetime of turbines.

Typically modern turbines use either electrical or fluid power pitch systems, where this paper focus on reliability estimation of the latter. Currently, the most detailed publicly available field data on pitch system failures show the failure rate distribution among system components such as valves, pump, accumulators, cylinders, etc. (Carroll et al. 2015). Yet, the failure modes and root causes are not evident. Such information is crucial in order to identify critical areas of the system and to enable development of more reliable and safe concepts. Also, precise reliability estimates of system components allow for strategic maintenance planning which potentially reduces maintenance time and costs. In an attempt to reveal the highrisk areas of the pitch system, Liniger et al. conducted a systematic qualitative study on identifying critical components (Liniger et al. 2017). While the study showed promising results, the occurrence of failure modes was qualitatively determined using expert knowledge of generic fluid power systems. Thus, actual failure rates and operational dependent failure mechanisms of the pitch system in a wind turbine was not directly considered. Two quantitative studies have been conducted with the purpose of modeling fluid power pitch system reliability (Yang et al. 2011, Han et al. 2012). While these studies have aimed at creating the model basis for calculating

reliability, both the origin of failure modes and failure rates has not been fully covered. The only publicly available source of failure rate estimation for fluid power components known to the authors is the *Handbook of Reliability Prediction Procedures for Mechanical Equipment* (Jones 2011). This source presents failure rate models as functions of component dimensions, material properties and operating conditions.

The main contribution of this paper is a feasibility study of estimating pitch system reliability using the empirical failure rate models of (Jones 2011). Estimation of failure rates has the potential to close the gap in knowledge between the qualitative and quantitative studies while also incorporating turbine operating conditions into the framework. Operating conditions are generated using a simulation model of a fluid power pitch system and the 5MW National Renewable Energy Laboratory (NREL) wind turbine implemented in the Fatigue, Aerodynamics, Structures, and Turbulence (FAST) software. The estimated failure rates are compared to the most detailed and recent field failure rates available.

2 PITCH SYSTEM DESCRIPTION

The pitch system configuration used in this analysis is depicted in Figure 1 and component label description is found in Table 1. The system consists of a supply located in the nacelle of a turbine and the actuation located in the rotating hub. The supply consists of a fixed displacement fixed speed pump where pressure and flow are conditioned by dump valve V2 and relief valve V1. The supply connects to the rotating hub through rotary union R1. The actuation is a conventional fluid power cylinder drive, where cylinder position is controlled closed-loop and flow is metered by the proportional valve V6. Accumulators A1 and A2 stores energy which is used for extending the cylinder C1 in the event of an emergency shutdown.

The fluid power pitch system presented in Figure 1 is very similar to the system analyzed in the previous qualitative study (Liniger et al. 2017). The results of the previous study showed to correlate well to field failure data which indicates that the system presented here is similar to the real-life systems which are confidential. Note that conventional fluid power pitch systems also employ a locking circuit for keeping the blade pitch angle fixed when the turbine is shut down. The locking circuit and all system transducers are omitted in this analysis as their failure rates are negligibly small compared to that of the other system components (Carroll et al. 2015).

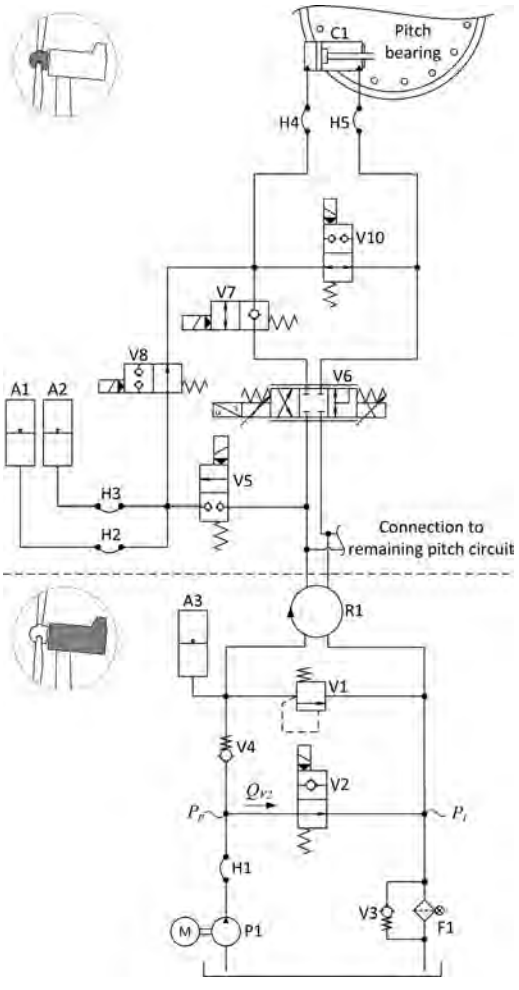


Figure 1. Fluid power pitch system diagram with indication of supply and actuation circuit locations in the wind turbine.

Table 1. Description of component labels.

Label	Description	Type
V1	Relief valve	Cartridge, poppet
V2	Solenoid dump valve	Cartridge, poppet
V3,V4	Check valves	Cartridge, poppet
V5,V7-V10	Solenoid valves	Cartridge, poppet
V6	Proportional solenoid valve	Module, spool
C1	Differential cylinder	
H1-H5	Flexible hoses	
A1, A2	Emergency accumulators	Gas charge, piston
A3	Pump accumulator	Gas charge, piston
P1	Fixed displacement pump	Internal gear

Table 2. Main data for the wind turbine and pitch system simulation model.

Turbine data	Value
Nominal power	5 [MW]
Nominal hub speed	12 [RPM]
Tower height	90 [m]
Blade length	63 [m]
Wind speed (Rated)	11.4 [m/s]
Turbulence model	Normal Turbulence Model (NTM)
Pitch system data	
Pitch cylinder (Rod/Piston/ Stroke)	Ø90/Ø140/1350 [mm]
Pump flow (Rated)	20 [l/min]
System pressure (Rated)	250 [bar]

3 WIND TURBINE SIMULATION MODEL

A simulation model of the fluid power pitch system operating in a wind turbine is utilized for generating operating conditions subsequently used for reliability estimation. The wind turbine model is based on the open-source data for a 5MW NREL turbine implemented in the FAST software (Jonkman and Buhl 2005). The main specification are given in Table 2.

It is noted that power capacity of the simulated turbine is above the 2–4MW range covered by the field data. The simulated operating loads may, therefore, be larger than those found in the real-life systems and possibly yielding higher estimated failure rates.

The dynamical model of fluid power pitch system is based on the layout described in the previous section and is developed in a previous study by the authors (Pedersen et al. 2015). The dynamical model is implemented in Matlab/Simulink and co-simulated with FAST. The model incorporates the compressibility of fluid in the cylinder chambers, proportional valve dynamics and kinematics of the cylinder-blade coupling. The pitch angle is controlled closed-loop using a gainscheduled PI-compensator. The main specification for the pitch system is given in Table 2.

4 OPERATING CONDITIONS

The operating conditions for the real-life turbines used in this feasibility study are unknown. Thus, the simulated system is operated under a wide range of conditions which are considered to fully cover the conditions of the real-life turbines. The considered range of operating conditions is mean wind speed, turbulence intensity, ambient temperature and fluid temperature. The full-field wind is generated using Turb-Sim (Jonkman 2009) and based on the IEC61400-1 wind turbine design standard Design Load Case (DLC) 1.2 (IEC 2006). This load case is used for evaluating fatigue loads of wind turbines during normal operation. While the pitch system is used for both stopping and starting the turbine, normal operation constitutes the majority of the system lifetime. Based on the availability of the real-life turbines (Carroll et al. 2015) and considering out-of-range wind speeds, the utilization percentage can be assumed to be 90%.

The mean wind speed is normally described using the Weibull probability density distribution which can be described by a shape and scale parameter (Hansen 2008). A 20-year baseline distribution shown by blue bars in Figure 2 from the Østerild location near the Danish shore is selected for the study. The black bars show the range of wind speed distributions considered by selecting shape and scale parameters ±20% from the baseline values. The wind distribution is discretized in twelve wind bins from cut-in to cut-out wind speed of the 5MW NREL turbine. To simplify the analysis, the wind direction is assumed to be ideal, that is, orthogonal to the turbine. Estimated failure rates for each wind bin is multiplied by the probability density and summarized to yield values comparable to the field data.

According to the DLC 1.2, turbulence intensity is categorized in either high (class A, 16%), medium (class B, 14%) and low (class C, 12%) for the Normal Turbulence Model (NTM). All three turbulence intensity classes are considered in the feasibility study.

The pitch systems are located in the hub and nacelle of the turbine which in most modern turbines is conditioned to be within a desirable operating temperature interval. Operating temperatures considered are T_{amb} = [0 20 60]°C.

Lastly, the fluid temperature is controlled during normal operation of the pitch system. However, it is not uncommon for the fluid temperature to be locally different than the desired value. T_{fluid} = [30 50 70]°C is selected to cover the expected range of temperatures.

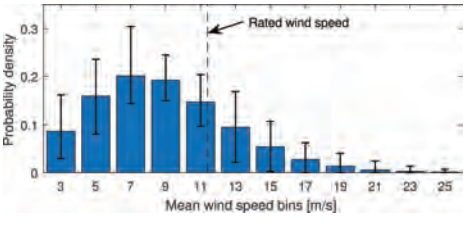


Figure 2. Mean wind speed distribution used in simulations. The black bars indicate the considered range.

5 FAILURE RATE ESTIMATION

Failure rates are estimated using the *Handbook of Reliability Prediction Procedures for Mechanical Equipment* (Jones 2011). The failure rate models are constructed such that an empirically determined base failure rate, λ_B, for a generic component is multiplied with several non-dimensional factors, C_1, C_2, C_3…, describing both material, dimensions and operating conditions for estimating the failure rate.

Some failure rates estimations are dependent on operating temperature. The operating temperature for components with direct fluid contact is set to the fluid temperature. Solenoids are if they are operated, set to the ambient temperature plus a constant offset accounting for joule heating. The offset is 100°C for valves that are continuously on during normal operation. Valve V2 is operated intermittently, thus reducing the temperature offset to 70°C. The temperature offsets are confirmed from measurements (Liniger et al. 2018).

Failure rate estimates are influenced by the amount of allowable leakage for seals and valves in the system. For pitch systems, external leakage is in most cases much more critical than internal leakage due to environmental contamination hazard of the turbine surroundings. External leakage is generally set to a very low value of $2 \cdot 10^{-7}$ l/min corresponding to a few drops a month. Allowable internal leakage is set to 10^{-4} l/min for seat valves and seals. Proportional spool type valves are normally associated with higher internal leakage, and the allowable limit for valve V6 is therefore set to 2 l/min.

Operating cycles of the valves, cylinders and accumulators are used for assessing the failure rates. Operating cycles for on/off valves are simply determined from the number of activations during normal operation. For proportional valve V6, cylinder C1 and the accumulators, the operating cycles are determined using rain-flow counting and a minimum travel threshold. The minimum threshold for C1 and the accumulators is selected to 2 mm. For valve V6, the minimum threshold is 0.2 mm. Due to the uncertainty of the threshold values, a sensitivity study is conducted in the results Section 5.2.

Contamination concentration in the fluid is also considered in the failure rate estimation. The contamination concentration at each component N_{10} is determined by the particles generated from upstream components according to the rates specified in (Jones 2011). Additionally, a particle filtration size of $C_n = 3$ μm for filter F1 is utilized in the calculations.

To simplify the description, the details for failure rate estimation of one component are given in the following section. The cartridge poppet type valve V2 is selected since valves are the most used component type in the system. Also, the failure rate estimation procedure for the parts in valve V2 is similar to most of the remaining components. All component specifications are similar to those found in actual pitch systems working in turbines with similar power capacity as for the real-life systems.

5.1 *Cartridge poppet valve V2*

Cartridge Valve V2 is shown in Figure 3. Valve V2 consists of several parts where dimensions, material and specifications are given in Table 3. All notations follow (Jones 2011) and pressures and flows are denoted according to the diagram in Figure 1. The operating time t_H is given in hours and N_{V2} denote the operating cycles of valve V2. Note that both the factors for surface finish (roughness) for the seat valves and Young's modulus for O-rings increase with time. The estimated failure rates for these parts are therefore increasing with time.

As an example, the multiplication factors for the main poppet are given in Table 4 and values are determined from a nominal operating scenario. The nominal operating scenario covers one calendar year of operation at ambient temperature T_{amb} = 20°C, fluid temperature T_{fluid} = 50°C, rated wind speed and turbulence class B. The pressure and flow rate multiplication factors depend on simulated pressure and flow time series. Multiplication factor values C_p and C_w represented in Table 4 are mean values. Due to length limitations, the time series are not shown but can be found in the work by Pedersen et al. (Pedersen et al. 2015).

The multiplication factors C_p, C_q, C_v, C_n, C_s, C_w are seen to cover operating conditions for pressure and flow. C_f, C_{dt}, C_{sw} depend on the dimensions and manufacturing of the valve. The main poppet failure rate for the nominal operating scenario is determined according to:

$$\lambda_{V2,m} = \lambda_{B,SV} C_p C_q C_f C_v C_n C_s C_{dt} C_{sw} C_w \frac{N V_2}{t_H} \qquad (1)$$

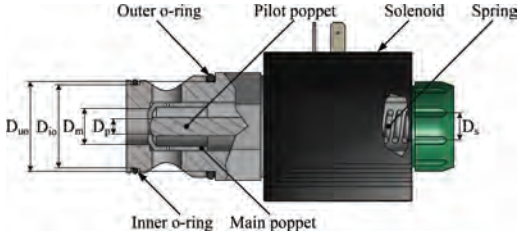


Figure 3. Cartridge poppet type valve V2 with part dimensions.

Table 3. Cartridge poppet type valve V2 data related to failure rate estimation.

Part	Dimension		Material	Specification	
Main	Seat diameter	$D_m = 15$ mm	Steel	Seat valve base failure rate	$\lambda_{B,SV} = 1.4 \cdot 10^{-6} \frac{\text{failure}}{\text{cycle}}$
Poppet	Seat width	$D_{mw} = 2$ mm		Seat pressure drop	$\Delta P_m = P_p - P_t$[bar]
				Rated flow	$FR_m = 140$ l/min
				Allowable leakage	$Q_{mf} = 10^{-4}$ l/min
				Surface finish	$F_m = \begin{cases} 15.10^{-5} \cdot N_{V2} + 0.2[\mu m] \\ \quad \text{for} \quad N_{V2} <= 4000 \text{ cycles} \\ 1.5 \mu m \\ \quad \text{for} \quad N_{V2} > 4000 \text{ cycles} \end{cases}$
Pilot Poppet	Seat diameter Seat width	$D_p = 5$ mm $D_{pw} = 0.5$ mm	Steel	Rated flow Allowable leakage	$FR_p = 1$ l/min $Q_{pf} = 10^{-4}$ l/min
Spring	Coil diameter	$D_s = 15$ mm	Steel	Spring base failure rate	$\lambda_{B,S} = 23.8 \cdot 10^{-6} \frac{\text{failure}}{\text{hour}}$
	Wire diameter	$D_{sw} = 2$ mm		Operating cycle rate	$\frac{N_{V2}}{t_H} \left[\frac{\text{cycle}}{hour} \right]$
				Active coils	$N_a = 7$
				Compression length	$C_L = 5$ mm
Outer	Inner diameter	$D_{io} = 30$ mm	NBR-70	Seal base failure rate (static)	$\lambda_{B,SS} = 2.4 \cdot 10^{-6} \frac{\text{failure}}{\text{hour}}$
O-ring	O-ring diameter	$D_{uoø} = 2.62$ mm		Allowable leakage	$Q_{uof} = 2 \cdot 10^{-7}$ l/min
				Seal pressure drop	$\Delta P_{uo} = P_p - P_{atm}$ [bar]
				Youngs modulus	$E_{NBR70} = 59 \cdot 10^{-6} \cdot t_H + 6.2$ [MPa]
				Mating surface finish	$F_{uo} = 0.8$ μm
Inner O-ring	Inner diameter O-ring diameter	$D_{uo} = 28$ mm $D_{uoø} = 1.78$ mm	NBR-70	Allowable leakage Seal pressure drop	$Q_{iof} = 10^{-4}$ l/min $\Delta P_{io} = P_p - P_t$ [bar]
Solenoid			Insulation	Solenoid base failure rate	$\lambda_{B,S} = 2.77 \cdot 10^{-6} \frac{\text{failure}}{\text{cycle}}$
			class-H	Coil temperature	$T_{coil} = T_{amb} + 70$ [°C]
				Operating cycle rate	$\frac{N_{V2}}{t_H} \left[\frac{\text{cycle}}{hour} \right]$

The main poppet and remaining part failure rates of valve V2 are given in Table 5.

Clearly, the solenoid contributes with the highest failure rate of the valve. The lowest failure rate exists for the inner o-ring. The main and pilot poppet and the spring are seen to yield similar failure rates.

5.2 *Results*

The feasibility study of the described estimation method is performed by comparing estimated failure rates to field failure rates. The field failure rates are divided into six component groups, namely accumulators, valves, pump, cylinder, rotary union, and hoses. The system presented in Figure 1 is likewise divided in the six component groups. The accumulator group consists of A1-A3. The valve group cover V1-V10 and the hose group is constructed from hoses H1-H5. Each estimation case utilize all combinations of wind speed profiles as given in Figure 2 and turbulence intensity classes A, B and C.

Figure 4 shows the comparison between the estimated and field failure rates for a range of ambient temperatures. The black error bars indicate the

Table 4. Failure rate estimation and multiplication factors for main seat valve in V2 under nominal operating conditions.

Factor name	Description	Value
Pressure	$C_p = (4.8 \cdot 10^{-3} \cdot \Delta P_m)^2$	0.44
Allowable leakage	$C_q = \begin{cases} 9.10^{-4} \cdot Q_{mf} & \text{for } Q_{mf} > 4.9 \cdot 10^{-4} \text{ l/min} \\ 4.2 - 4.8 \cdot 10^3 \cdot Q_{mf} \end{cases}$	3.72
Surface finish	$C_f = \frac{(39.4 \cdot F_m)^{1.65}}{353}$	0.46
Fluid viscosity	$C_v(T_{fluid})$ SAE 10 fluid look-up table	0.47
Fluid contamination	$C_n = \left(\frac{C_O}{10} \right)^3 F_{Rm} N_{m10} \cdot 3.79$	0.024
Contact pressure	$C_s = 0.26 \cdot \left(\frac{9000}{3 \cdot \Delta P_m \cdot 1.5 \cdot 10^{-4}} \right)^{1.5}$	0.31
Seat diameter	$C_{dt} = 1.1 \cdot D_m \cdot 0.04 + 0.32$	0.97
Land width	$C_{sw} = \begin{cases} 3.55 - 0.97 \cdot D_{mw} + 73 \cdot D_{mw}^2 - 86 \cdot D_{mw}^3 \\ 0.25 \text{ for } D_{mw} > 1.34 \cdot 10^{-2} \end{cases}$	2.0
Flow rate	$C_w = 1 + \left(\frac{Q_{V2}}{FR_m} \right)^2$	1.0
Operating cycle rate	$\frac{N_{V2}}{t_H}$	$33 \frac{\text{cycle}}{hour}$

Table 5. Part failure rate for valve V2 under nominal operating.

Part	Failure rate $\left[\frac{\text{failure}}{\text{cycle}} \right]$
Main poppet	$\lambda_{V2,m} = 2.4 \cdot 10^{-7}$
Pilot poppet	$\lambda_{V2,p} = 1.9 \cdot 10^{-7}$
Spring	$\lambda_{V2,sp} = 1.5 \cdot 10^{-7}$
Outer o-ring	$\lambda_{V2,uo} = 1.0 \cdot 10^{-8}$
Inner o-ring	$\lambda_{V2,io} = 2.7 \cdot 10^{-9}$
Solenoid	$\lambda_{V2,so} = 1.2 \cdot 10^{-6}$
Valve V2	$\lambda_{V2} = 1.8 \cdot 10^{-6}$

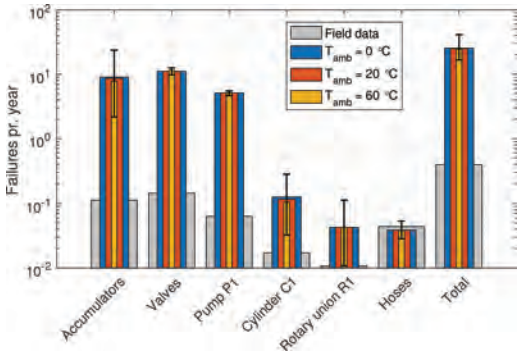


Figure 4. Component group failure rates for varying ambient temperatures. Other conditions are fluid temperature $T_{fluid} = 50°C$ and low travel threshold.

variation due to the considered wind speeds, turbulence intensities and wear models. The colored bars are mean values for each estimation case. Feasible failure rates estimation means that the field data must fall within the error bars. From Figure 4, the estimated failure rates are seen to be at least an order of magnitude larger than the field failure rates for component groups other than hoses. Increasing the ambient temperature slightly reduces the estimated failure rates. The only component group to fall within the estimated range is hoses.

The estimated failure rates for varying fluid temperatures are seen in Figure 5. Generally, the estimated failure rates increase with increasing fluid temperature. A significant change is seen for the failure rates of valves. At lowest value, the estimated failure rate of the rotary union covers the field data.

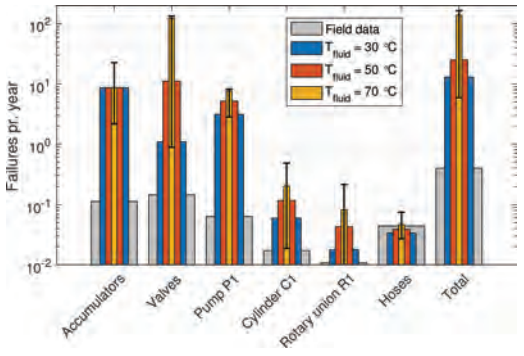


Figure 5. Component group failure rates for varying fluid temperatures. Other conditions are ambient temperature T_{amb} = 20°C and low travel threshold.

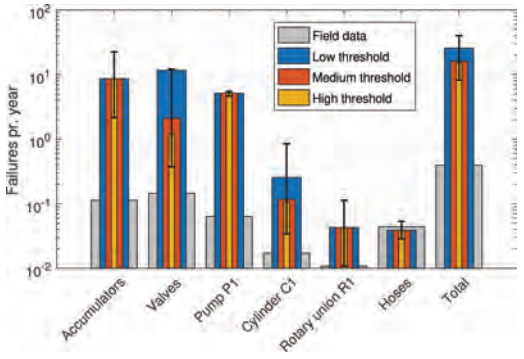


Figure 6. Component group failure rates for varying travel thresholds. Other conditions are T_{amb} = 20°C and T_{fluid} = 50°C.

The travel threshold for counting operating cycles is associated with a high degree of uncertainty. Thus, the effect of three threshold levels are analysed. The low levels are described in Section 5. Medium and high thresholds are double and triple w.r.t. the low values. Increasing the threshold lowers the operating cycles which in Figure 6 is seen to have a minor decreasing effect to the failure rates. At a low threshold value both the valves and cylinder C1 failure rates are increased significantly.

The results are generally not satisfying, and the large discrepancy between estimated and field values is either related to wrongful modeling assumptions or non-describing field data. Reasons for the estimated failure rates being larger than the field data could be a consequence of the simulated wind turbine being 5 MW rather than the 3–4 MW range of the real-life systems and non-modeled repair or replacements of components. Also, over 30% of field failures are known to be insufficiently documented and therefore not considered in this comparison (Liniger et al. 2017). On the other hand, the real-life systems are known to contain more components than presented in Figure 1 which potentially could further increase the estimated failure rates.

In spite of these facts, the large discrepancies are most likely caused by the estimation procedure being over-conservative. This is indicated by the tendency seen in Figure 6 for low threshold, where the failure rates for all component groups except hoses follow the field data with an offset. As evident from the references in the *Handbook of Reliability Prediction Procedures for Mechanical Equipment* (Jones 2011), the empirical base failure rates date back to the late 1960's. Latest developments in manufacturing processes, component design, and fluid properties are therefore not considered in the estimation procedure. A suggestion for increasing the precision of the estimation procedure could, therefore, be to adjust the base failure rates to more current data and evaluate if the multiplication factor can be used as is.

6 CONCLUSION

A feasibility study of estimating fluid power pitch system reliability has been conducted using an estimation procedure for generic fluid power components. The estimated values were determined based on operating conditions obtained from a simulation model of a pitch system in normal operation in a turbine. The estimated failure rates were compared to failure rates of real-life turbines. Large parameter variations related to wind speed, turbulence intensity, and system temperatures were performed since the operating conditions of the real-life turbines were unknown.

The estimated failure rates were over-conservative in relation to the field failure rates for accumulators, valves, pump, cylinder and rotary union. The method yielded feasible results only for failure rates of hoses.

While being over-conservative, the estimated failure rates followed the same tendency of the field data for accumulators, valves, pump, cylinder and rotary union. The similar tendency indicated that the base failure rates are incorrect for components in modern pitch systems and should be updated using more recent test data.

REFERENCES

Carroll, J., A. McDonald, & D. McMillan (2015). Failure rate, repair time and unscheduled O&M cost analysis of offshore wind turbines. *Wind Energy*, 1107–1119.

Carroll, J., I. Dinwoodie, A. McDonald, & D. McMillan (2015). Quantifying O&M savings and availability improvements from wind turbine design for maintenance techniques. In *Europe's Premier Wind Energy Event*, Volume 9. European Wind Energy Association (EWEA).

Han, X., H. Zhang, Y. Chen, X. Zhang, & C. Wang (2012). Fault diagnosis of hydraulic variable pitch for wind turbine based on qualitative and quantitative analysis. *World Congress on Intelligent Control and Automation (WCICA), 2012 10th*, 3181–3185.

Hansen, M.O. (2008). *Aerodynamics of wind turbines* (2. ed.). earthscan London.

IEC (2006). Wind turbines part 1: Design requirements (IEC 61400-1:2005).

Jones, T.L. (2011, May). *Handbook of Reliability Prediction Procedures for Mechanical Equipment*. West Bethesda, Maryland 20817-5700: Naval Surface Warfare Center NSWC.

Jonkman, J. & M.L.J. Buhl (2005, August). FAST user's guide. Technical Report NREL/EL-500-29798, National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, Colorado 80401-3393.

Jonkman, J. (2009, August). TurbSim user's guide: Version 1.50. Technical Report NREL/TP-500-46198, National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, Colorado 80401-3393.

Liniger, J., M. Soltani, H.C. Pedersen, J. Carroll, & N. Sepehri (2017, June). Reliability based design of fluid power pitch systems for wind turbines. *Wind Energy 20*(6), 1097–1110.

Liniger, J., S. Stubkier, H.C. Pedersen, & M. Soltani (2018). Early detection of coil failure in solenoid valves. *Submitted to: IEEE/ASME Transactions on Mechatronics*.

Pedersen, H.C., T.O. Andersen, & J. Liniger (2015, October). Investigation of load reduction possibilities in wind turbines using a fluid power pitch system. In *Proceedings of the ASME/BATH 2015 Symposium on Fluid Power & Motion Control*. American Society of Mechanical Engineers.

Wilkinson, M. & B. Hendriks (2010). Reliability-focused research on optimizing wind energy system design, operation and maintenance: Tools, proof of concepts, guidelines & methodologies for a new generation. *Collaborative Project: Large Scale Integrated Project, FP7-ENERGY-2007-1-RTD*.

Yang, X., J. Li,W. Liu, & P. Guo (2011). Petri net model and reliability evaluation for wind turbine hydraulic variable pitch systems. *Energies 4*(6), 978–997.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Simulator training in driver education—potential gains and challenges

G.B. Sætren, P.A. Pedersen, R. Robertsen & P. Haukeberg
Nord University, Business School, Norway

M. Rasmussen & C. Lindheim
NTNU Social Research, Studio Apertura, Norway

ABSTRACT: Norway is currently ranked as one of the top nations in regard to road safety. However, continued efforts are applied as we stretch towards a goal of zero deaths and serious injuries in road traffic accidents. In this paper we explore if Norwegian driver education could benefit from simulator training. Possible advantages are cost effectiveness, environmentally friendly training, repeatability, accessibility to different scenarios (accident scenarios and dangerous situations, darkness and snow outside of winter, difficult weather conditions and extreme road traffic density), the possibility to make errors in a safe environment, and interaction with new technology such as advanced driver assistant systems. However, there are challenges such as how to increase the number of simulators in Norway, and legal obstacles as current legislations require all mandatory parts of the Norwegian driver education to be conducted on the road. Our overall impression is that the driver education in Norway could have advantages in applying a more systematic approach to simulator training.

1 INTRODUCTION

The purpose of this paper is to investigate the use of simulator training in driver education in Norway, discuss the potential gains and challenges and look at the possibility of increasing the availability and use of driving simulators. In Norway, like in many other countries, the public authorities have established a formal theoretical and practical driver education (NPRA 2017), based on scientific and policy factors, where professional driving teachers employed by approved driving schools are the main responsible bodies to conduct the education. The driver learner program is an extensive and systematic module based program with a comprehensive syllabus. The program is to a large degree based on the Goals of Driver Education-matrix (GDE-matrix; Keskinen 1996 in Hatakka et al. 2002; Keskinen et al. 2010). In this program it is estimated that the average learning period, from novice to the issuing of the driver's license, is two years. The authorities recommend that training starts at the age of 16 in order to get the driver's license at the age of 18 – which is the lower limit for receiving a car driver's license in Norway. To reduce accident risks in novice drivers, elements of the driver training are carried out in real life situations where driver learners are accompanied by skilled driver instructors. Additionally, in Norway it is legal and recommended for experienced drivers (normally parents) to provide driver learners with extra practice. The only premise is that the driver learner has completed an introductory course and that the experienced driver must have held their driver's license for a minimum of five years without receiving any penalties or driver's license endorsements (FOR 2017). Such additional training is meant to increase the driver learners' experience behind the wheel prior to their exams and the license issuance. Our question is to whether driving simulators could be a training platform in Norway to increase driver learners' driving experience, and if they can complement or even substitute some of the more traditional learning methods used in today's education.

In many industries where human errors are likely to have critical outcomes, such as aviation, hospital medicine and commercial nuclear power, simulator training is frequently used as part of training. Simulator training can be cost efficient and can provide training in situations that are rarely seen (e.g. accident scenarios; Bye et al. 2011; McGaghie et al. 2010; Salas, Bowers & Rhodenizer 1998). Currently driving simulators are not the standard way of learning how to drive, however, in some European countries, such as in The Netherlands and the United Kingdom, simulator training has gained some acceptance as part of the driver education (Baten & Bekiaris 2003), and there are reports showing an increased use of simulators in Germany (Stiegler & Vennefrohne 2017) and France (Goepp 2017). There are several

factors explaining why simulator training is more common in other industries where human errors are likely to have critical outcomes than in driving. In medical surgery for instance, the risk of letting unskilled personnel practice on people is considered too high, so simulation has become a natural way of acquiring skills. Doing simulator training means that there is room to learn from mistakes. The same is seen in aviation. Additionally, the costs and emissions of flying a large aircraft are so substantial that doing all the training necessary to obtain a commercial pilot license is not considered economically or environmentally beneficial. Even though developing, building and handling a simulator also result in costs, it is far less expensive than training in airplanes. In aviation, as well as industries such as commercial nuclear power, simulators can be used to train personnel to avoid serious accidents and to minimize the overall consequences if unwanted events occur.

It is our impression that all of the reasons mentioned above, concerning reduced risk through extra training, prevention of fatalities and injuries, handling accident scenarios, and reduced cost and emissions, can be used as reasons to introduce car driver training in simulators. In this paper we will attempt to clarify current usage and potential gains of simulators in driving education in Norway (section 2). This is discussed in the light of the rapid technological development in today's automobile industry, and how new technology can be included in simulators. This is followed by a discussion on the structural and practical obstacles in implementing an increase in simulator training in driver education (section 3).

2 POTENTIAL GAINS IN SIMULATOR TRAINING

It has never been common to use driving simulators as part of the driver education in Norway. Currently, only 5–10 out of 1033 driving schools, offer simulator training for driving-license category B driver training (vehicle weight less than 3500 kg), and the simulators are mainly used for learning the basic introductory elements of handling and maneuvering a car. These schools seem to lack a systematic pedagogical or educational plan in their simulator use. Additionally, The Norwegian Public Road Administrations are rather strict on what is allowed to be taught in a simulator only. Any topic that is mandatory in the education will not be approved using only a simulator (NPRA 2017), despite research indicating that for instance that the mandatory dark driving demonstrations have the same learning outcome taught in real life and in a simulator (Mikkonen 2007; Robertsen et al. 2017). A different approach is taken in Finland where dark driving sessions are approved using a simulator, so these aspects are not internationally agreed upon.

There have not been many empirical studies measuring and discussing the learning outcomes from using simulators in driving education. We only found one published study on use of simulator training in driver education in Norway (Robertsen et al. 2017). This study was regarding theoretical learning outcome when comparing traditional training and simulator based training on dark driving demonstration. Dark driving is a part of first module (basic handling of the car) in the Norwegian driving education program. This study showed no significant differences in the outcome between these two groups on theoretical knowledge of dark driving. According to some of the international empirical studies concerning driving simulators, it seems like simulator training could be useful in driving education. In a study carried out in in The Netherlands by de Winter et al. (2009), they found that better driving simulator performance increased the actual driving skills on the roads and the chance for passing the final driving test. Additionally, Crundall et al. (2010) found that commentary training in a driving simulator has beneficial effects on driving behavior in the UK. For instance, it was found to improve responsiveness to hazards on the roads. Wang et al. (2010) have pointed out that road hazard performance was significantly higher for a simulator trained group of novice drivers than others. Divekar et al. (2016) also report that novice drivers' outcome in PC-based simulator training increases the awareness and driving skills in real life operations. Additionally, a German study showed that the training period could be reduced by 21 days when using a simulator instead of traditional training with a driver instructor (Reindl, Gunther, & Wottge 2016). However, in all these empirical studies there are methodological challenges in isolating and measuring the learning outcome from simulator training, and determining the transferability from improvements in the simulator to improved driving on real roads. Another common challenge in these studies is the difficulty to measure the long term effects on the drivers' skills and behavior. Nevertheless, it seems to be an agreement in these empirical studies that especially novice drivers have a significant short term positive learning outcome from training specific elements in simulators.

Based on this earlier research, a systematic offer of simulator training in the official driving education might make it easier to learn the basic skills in handling a car and making the soon-to-be drivers trained in adjusting their road traffic behavior to the circumstances on the road. Hence,

it seems likely that driving in a simulator, accompanied by other training methods, could be used to improve the various driving- and safety skills during the phase of learning to maneuver a car. Particularly, in order to reduce risks of young drivers, training with professional driving instructors combined with simulator training, seems to gradually become accepted as a useful tool in developing driving skills.

The gains of supplying adequate simulator training are also related to the possibility to train in a secure environment where the negative consequences of making mistakes are eliminated. It is also environmentally friendly, flexible, could train driver leaners in different road traffic environments any time of the year. For instance, Norway has a large road traffic density variance. This means that different mandatory training scenarios, such as urban and rural driving, are fairly easy to obtain in the cities and other densely populated areas. However, in some rural areas access to urban driving might entail a long journey. A widespread access to driving simulators could potentially reduce the number of those long journeys. Due to the long dark winter in Norway, it is particularly important to learn how to drive in the dark and handling the challenges of darkness. However due to the lack of darkness in summer, the mandatory dark driving demonstration can only be conducted from the end of October till mid-March. With a simulator of sufficient quality, dark driving demonstrations can be given all year around. Other environmental challenges are seen in other countries, such as southern France (Goepp 2017), where it is not unlikely to go through the entire driver education without facing rain. The possibility for all driver learners to experience different weather conditions and road traffic densities is a good argument for using a simulator. Additionally, simulator training has the potential to be cheaper for the driver learner, if sufficient instructions are given virtually and thereby removing the necessity of one driving instructor per driver learner.

Simulator training should also be seen in connection to the rapid technological development seen in the automobile industry. When a driver learner has finished the driving education, should he/she only be able to handle the basic technology found in every car, or should he/she have learned how to use and interact with new technology introduced in new cars to assist the driver? Staying in touch with the technological race while using a traditional training approach would entail a very frequent replacement of vehicles. Using a simulator approach might be easier as a software update could potentially provide the new technology or features to be used without replacing the simulator.

From other industries, research has shown that training for new and more automated technology is of importance in order to avoid unwanted incidents (Salas et al. 2006; Sætren and Laumann 2015). It would be beneficial to have a system for training drivers who buy new cars with new Advanced Driver Assistant Systems (ADAS) technology. Research shows that buyers have very limited knowledge of the new technology in their new cars and on how it can be used. One of the reasons could be that only 24% reported that they received instructions regarding the ADAS technology from the manufacturer when buying the car (Harms & Dekker 2017). Training in simulators could provide an important alternative for learning how to drive with ADAS technology for instance for drivers who already hold their driver license but need training for new technology.

3 CHALLENGES IN SIMULATOR TRAINING

In order to implement a broader use of simulator training in driver education there are a lot of challenges and obstacles that have to be discussed and solved. First, there are technological challenges in developing adequate hardware and software making the simulators relevant for learning to drive on the road. For instance, to what degree could and should simulators be designed to give the driver learners a "car-like" experience when training, and should the simulators be designed to make it possible to adjust for different equipped cars? A software-challenge would be to design adequate road traffic situations training the driver learners to handle and control the vehicle on the road in different road traffic settings.

One of the main challenges in simulator based training, if thought of being used for more advanced driver education, is to adjust the training to the GDE-matrix. As mentioned, the hierarchical GDE-matrix is an important base for the Norwegian driver education. The GDE-matrix originally consisted of four levels, where the first level is vehicle maneuvering, second level is mastering road traffic situation, the third level is goals and context of driving, and fourth level is goals for life and skills for living (Keskinen 1996 in Hatakka et al. 2002) and later a fifth level, social skills, was added (Keskinen et al. 2010). The skills are learned through theoretical and practical teaching in addition to individual and group work. To reach level four and five it takes time to mature, thus, the authorities recommend starting at age 16 in order to have a driver license at age 18. The main reason for this is that adequate psycho-motoric skills and physiological functions are found not to be sufficient

for good and safe driver performance. For instance, when the lowest levels of the hierarchy are learned, they are applied under guidance of higher level objectives. Hence, the training of basic skills is important but the driver learner should also be able to deal with goals higher in the hierarchy such as dealing with social pressure (Hatakka et al. 2002). There is little doubt that simulator training could be of help for the lower levels in the GDE-matrix, but in order to deal with the higher levels including self-evaluation, simulator training might not be optimal. Thus, this argues that simulator training cannot completely replace the traditional driver's education, but be a supplement.

Increasing the use of simulators in Norway has specific legal challenges. The mandatory driver's education is regulated such that it must be given by professional driver instructors while the driver learners are sitting behind the wheel of an actual car. Hence, training in simulators can only be seen as an additional part of an education program and not a part of the mandatory education. That being said, learning how to drive entails a large amount of training outside of the mandatory elements allowing driving schools to use simulators a substantial amount if they want to. The main obstacle in Norway seems to be the lack of simulators. There might be several reasons for this, but it seems that in general, the driving schools do not consider it economically beneficially to offer simulator training. In addition to the investment cost of simulator, the driving schools have to handle the cost of software updates, maintenance, and training staff in simulator handling. Without simulators, the main income of a driving school is hours spent on the road with driver learners. Introducing a simulator (particularly one that is cheaper than traditional training) the school has to change its business model for selling man-hours to include selling simulator-hours. For the driving instructors this would undermine their occupation. This is further underlined with the argument for having simulators in countries such as France, where it was emphasized that a school with 5–6 instructors, one driver instructor can be replaced with a simulator (Goepp 2017). There is no shortage of driving instructors in Norway, like for instance in Germany (Stiegler & Vennefrohne 2017), thus, the Norwegian market does not provide that need for a simulator for better efficiency.

Another challenge is simulator sickness. Simulator sickness is a subset of motion sickness, leading to many experiencing nausea after only a short while in a driving simulators, influencing the usefulness of simulator training (de Winter et al. 2012). Simulator sickness is due to the perceived discrepancies between the motion expected by the participant and the motion displayed in the simulator. This has been a problem in a wide range of simulators and virtual reality applications, but it is gradually decreased as the simulations improve in terms of both responsiveness and reduced delays. Research shows that younger individuals are less prone to simulator sickness than older (Brooks et al. 2010). This could be beneficial for driver learners, but might be a hindrance for using simulators for upholding driving skills for those who have had a driving license for some time.

4 CONCLUDING REMARKS

Our impression is that the car driving education in Norway could have advantages in using simulators more systematically than what has been done until now. There are many possible advantages from introducing simulators in driving schools such as cost effectiveness, environmentally friendly training, repeatability, accessibility to different scenarios (accident scenarios and dangerous situations, darkness and snow outside of winter, difficult weather conditions and extreme road traffic density), the possibility to make errors in a safe environment, and interaction with new technology such as advanced driver assistant systems. However, there are challenges such as how to increase the number of simulators in Norway, and legal obstacles as current legislations require all mandatory parts of the Norwegian driver education to be conducted on the road.

The experiences from other European countries and the few empirical studies that exist provide some insight into the potential for simulators to be used in driving education in Norway. However, more research should be done to find out which parts of the driving education that could be performed in a simulator, and how the simulator could set up to optimize the learning outcomes.

REFERENCES

Baten, G. & Bekiaris, E. 2003. *System for driver training and assessment using interactive evaluation tools and reliable methodologies* TRAINER. Final report GRD1-1999.10024.

Brooks et al., 2010. Simulator sickness during driver simulation studies. *Accident Analysis and Prevention, 42,* 788–796.

Bye, A. et al. 2011. International HRA Empirical study—Phase 2 report. Results from comparing HRA method predictions to simulator data from SGTR Scenarios. *US Nuclear Regulatory Commission.*

Crundall, D. at al. 2010. Commentary training improves responsiveness to hazards in a driving simulator. Accident Analysis and Prevention 42, 2117–2124.

de Winter, J.C.F. at al. 2012. Advantages and disadvantages of driving simulators: A discussion. *Proceedings of measuring behavior.*

de Winter, J.C.F at al. 2009. Relationships between driving simulator performance and driving test results. *Ergonomics 52*, 2, 137–153.

Divekar, G. at al. 2016. Effects of a PC-Based Attention Maintenance Training Program on Driver Behavior Can Lasy Up to Four Months. Simulator Study. Transportation Research Record: Journal of the Transportation Research Board, No. 2602, 121–128.

EC European Commission 2017. *2016 road safety statistics: What is behind the figures?* Downloaded December 12th 2017 from http://europa.eu/rapid/press-release_MEMO-17–675_en.htm.

Forskrift om trafikkopplæring (FOR) 2017. https://lovdata.no/dokument/SF/forskrift/2004-10-01-1339?q=trafikkopplæringsforskrift [Norwegian Regulations for traffic education] The Norwegian Ministry of Transport.

Goepp, M. 2017. How to develop further professional driving education and examination using simulators and/or VR in France. *Proceedings CIECA The International Commission for Driver Training, Munchen November 8th.*

Harms, I.M. & Dekker, G.-M. 2017. ADAS: from owner to user. Insight in the conditions for a breakthrough of Advanced Driver Assistance Systems. *Connecting Mobility NL.*

Hatakka, M. et al. 2002. From control of the vehicle to personal self-control; broadening the perspectives to driver education. *Transport Research Part F, 5,* 201–215.

Keskinen, E. et al. 2010 *GDE-5PRO and GDE-5SOC: goals for driver education in a wider context–professional and private drivers in their environment* Unveröffentlichtes Manuskript, Universität Turku, Finland.

McGaghie et al. 2010. A critical review of simulation-based medical education research: 2003–2009. *Medical Education*, 44, 50–63.

Mikkonen, V. 2007. Using simulators to teach driving in the dark as part of driver training. Report: Finnish Vehicle Administration.

Norwegian Public Road Administration (NPRA) 2017. *Læreplan for førerkort klasse B, BE og Kode 96* (Curricullum for driver training category B, BE and code 96). www.vegvesen.no.

Reindl, S. et al. 2016. *Einsatz von fahrsimualtoren in fahrschulen*. Report from: Institut fur automobilwirtschaft.

Robertsen, R. et al. 2017. Theoretical learning outcome of night driving. A comparison study of traditional real life training and simulator training. *Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016 (Glasgow, Scotland, 25–29 September 2016)*. CRC Press 2017 ISBN 9781138029972. p. 1018–1022.

Salas, E. at al. 1998. It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology,* 8(3): 19–208.

Salas, E. at al. 2006. *Design, delivery, and evaluation of training systems.* In G. Salvendy (Ed.) Handbook of human factors and engineering (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Stiegler, J., & Vennefrohne, R. 2017. The current situation of the use of simulator and VR in professional driver training and testing in Germany. *Proceedings CIECA The International Commision for Driver Training, Munchen November 8th.*

Sætren, G.B. & Laumann, K. 2015. Effects of trust in high-risk organizations during technological changes. *Cognition, Technology & Work, 17*, 131–144.

Wang, Y. at al. 2010. Effects of Simulation-Based Training Intervention on Novice Drivers' Hazard Handling Performance. Traffic injury Prevention, 11:16–24.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

A flow-based method for identifying critical pipelines in complex natural gas supply systems

Huai Su
National Engineering Laboratory for Pipeline Safety/MOE Key Laboratory of Petroleum Engineering/Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum-Beijing, Beijing, China

Enrico Zio
Dipartimento di Energia, Politecnico di Milano, Milano, Italy
Chair System Science and the Energy Challenge, Fondation Electricité de France (EDF), CentraleSupélec, Université Paris Saclay, Chatenay-Malabry, France

Jinjun Zhang & Xueyi Li
National Engineering Laboratory for Pipeline Safety/MOE Key Laboratory of Petroleum Engineering/Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum-Beijing, Beijing, China

ABSTRACT: Natural gas supply systems are complex pipeline networks, whose features can cause severe consequences under unexpected scenarios. In this work, we present the research being performed to develop a method for efficiently measuring and locating pipelines which are critical to gas distribution. Graph theory, thermal-hydraulic simulation, optimization and network flow method are integrated to calculate the pipeline criticality with respect to the supply of gas. Graph theory is applied to model the supply-transmission-demand systems. A capacity model, superposed on the graph model, and a combination of thermal-hydraulic simulation and optimization is used to simulate the system operation under different scenarios and estimate the different supply performances of the pipelines. This allows to assess the criticality of the pipelines by the network flow method. For demonstration, the method is applied to a relatively complex gas pipeline network. The results of the application show that the method can provide analytical information useful to improve the system robustness and perform a more efficient and effective protection of the gas supply system.

1 INTRODUCTION

Reliable service of supply from natural gas pipeline networks is important for economy development and society stability. Although many efforts have been done, potential vulnerabilities still exist because of uncertain environment, system structure complexity, demand fluctuation and resource limitation. When unexpected events occur, e.g., political crisis, terrorist attacks, third-party activities, extreme weather events, etc., severe consequences may follow (Su et al., 2017; Enrico Zio, 2016).

In many application areas, the vulnerability analysis of complex transmission networks has been given increasing attention, e.g., transportation systems (Hong, Ouyang, Peeta, He, & Yan, 2015; Mattsson & Jenelius, 2015), supply chain (Thekdi & Santos, 2016), power grids (Cadini, Agliardi, & Zio, 2017; E. Zio & Golea, 2012; Enrico Zio, 2014; Enrico Zio & Sansavini, 2013), etc. However, the importance of vulnerability of natural gas transmission networks has not been given enough efforts to.

Vulnerability is a term which is applied with several different definitions in the literature (Kröger & Zio, 2011). In this paper, vulnerability is defined as inherent system defects to absorb the effects of failures or to restore the system to normal condition. Vulnerability analysis focuses on the inherent properties rather than environment and probabilities, as reliability and risk do. Applying vulnerability analysis to evaluate criticalities of components of gas pipeline network systems can help to find the "bottlenecks" of a natural gas pipeline network system, improve the ability to withstand unexpected damages and reduce potential risk of loss.

There are several common methods to quantify the element criticality with respect to vulnerability. Generally, in a transmission network system, element criticality evaluation is usually performed based on the consequences of failures or the system structure. In the consequence-based methods, criticality of element is evaluated according to the direct, and sometimes indirect, effect of the failure (Johansson & Hassel, 2010). From the system struc-

ture perspective, some identification approaches have been proposed based on different considerations, e.g. flow-based performance (Fang, Pedroni, & Zio, 2015; Nicholson, Barker, & Ramirez-Marquez, 2016a; Enrico Zio & Piccinelli, 2010), network efficiency (Deng, Li, & Lu, 2015; Han. F; E. Zio, 2014; E Zio, 2007), accessibility (Demirel, Kompil, & Nemry, 2015), etc. In this paper, we use a measure of network performance (network flow) rather than the graph-theory-based measure to evaluate the criticality. Several works have been carried out recently to compare the topological model and the flow-based model for power grids (Ouyang, Zhao, Pan, & Hong, 2014) and the results show similarities.

The flow-based element criticality analysis requires a model that is capable of capturing the system behaviours, the amount of gas requested at demand sites and the supply capacity. System dynamic methods and thermal-hydraulic methods can analyse the consequences in detail; but, they are not suitable to the element criticality evaluation because of the huge cost of computation due to rather detailed numerical simulations. Hence, it is important to develop a computation-efficient model which can reflect system behaviours and calculate the consequences of element failures.

In this paper, a method of element criticality analysis of natural gas supply service pipeline networks is developed considering the system vulnerability, based on the performance of network flow. In Section 1, a consequence analysis model is developed based on network flow theory, combined with an optimization part. The model can simulate the system responses of element failures and estimate the capacities of supply of both the demand sites and the global system, with the consideration of constraints of physical limitation. Section 2 provides three criticality measures emphasizing the transmission performance based on different considerations. Finally, in Sections 3 and 4, the method is performed based on a complex natural gas pipeline network and the results are analysed in detail.

2 NATURAL GAS PIPELINE NETWORK MODEL FOR CONSEQUENCE CALCULATION

For the development of the consequence assessment model, only the components relating to gas supply capacity in the pipeline networks are considered, i.e. pipelines, compressor stations, gas sources and consumers. This is because the objective of the assessment is supply service. The development of the model of capacity/consequence analysis has followed the steps in Figure 1.

The inputs of the model include:

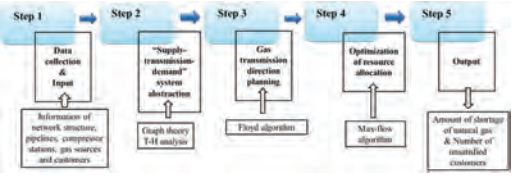


Figure 1. Process of modeling for consequence analysis.

A. Pipeline parameters: diameter, length, roughness etc.
B. Parameters of compressor stations;
C. Properties of natural gas;
D. Information of gas sources: locations, types, capacities;
E. Information of consumers: locations, demands;
F. Topology information of the pipeline network structure.

2.1 *Natural gas supply-transmission-demand system abstraction*

The pipeline network is abstracted in the form of a directed weighted network. The pipeline junctions, consumers, gas sources are represented as nodes and pipelines are represented as arcs. The weights on the arcs denote the capacities of the pipelines. The capacities are calculated by thermal-hydraulic analysis, based on the above inputs A, B and C. When the pipeline capacities change, the weights in the weighted network change accordingly.

In general, if an unexpected event occurs, operators will take actions to reduce the negative impacts on the system. Generally, the actions include adjustment of supply of gas sources and re-distribution of gas. In this model, all these actions can be swiftly performed by changing the weights on the arcs. The optimization process will be introduced in the next Sections.

2.2 *Modeling of gas transmission optimization*

In the optimization process, supply distance and economic efficiency are considered as the most relevant attributes to gas transmission planning. Firstly, a "standard cost" matrix C is developed based on network topology and the network element cost is calculated by equation 1. This so-called "standard cost", which is determined by the cost of transmission and pipeline length, represents a factor for the optimization of the transmission path, not a real cost:

$$C_{ij} = \alpha L_{ij} \cdot \beta\left(Q_{ij} \cdot c\right) \tag{1}$$

where C_{ij} represents the optimization factor combining distance and cost ($). Although the unit

of the "standard cost" is $, it is different from the actual cost. Q_{ij} represents the designed quantity of natural gas transported from i to j (MCM); L_{ij} represents the length of the pipeline from node i to node j (km); α and β, ranging from 0 to 1, represent the importance weights of distance and cost of transmission, respectively; c represents the cost of gas transportation ($/(km·MCM)).

Then, "standard cost" vectors are calculated based on matrix C. Each of the vector contains the "shortest paths" based on the "standard cost" between a gas source and all remaining nodes (except the other gas sources), found by the Floyd algorithm (Ahuja, Magnanti, & Orlin, 1993). In order to find out the sequence of the nodes that the gas flow from different sources should follow in its transmission path, the "standard cost" vectors are sorted in ascending order. For a specific gas source, priority is given to the nodes with lower "standard cost", which means lower cost and relatively higher efficiency.

The algorithm will also check whether the supply capacity of the source is exhausted: if there are still residual capacity and unsatisfied customers, the algorithm will search for the next unsatisfied demand node in the node sequence. This process continues until all demands are satisfied or the supply capacity of the source is exhausted.

2.3 *Supply capacity calculation*

On the basis of the transmission directions determined in Section 2.3, we proceed to optimize the plan of natural gas distribution in the pipeline network (volume of gas supplied by different sources and gas flow in the pipelines). This problem is converted to a max-flow problem in graph theory and solved by Ford-Fulkerson algorithm.

In this process, two constraints are to be respected:

- The sum of the flows exiting a node is equal to the sum of those entering the node (except for the sink nodes and the source).
- The flow in the arcs is within the capacity limitations.

Ford-Fulkerson algorithm is carried out by the following steps (Ahuja et al., 1993):

- Searching paths which connect the sinks and the sources with available capacities on the arcs.
- Repeating the search process until no additional flow can be added to the path.

In general, gas pipeline networks usually have more than one source or sink, however, Ford-Fulkerson algorithm is used to solve the "single source and sink" problem. To convert the "multiple sources and sinks" problem to a "single source and sink" problem, we assume a "super sink" and a "super source" connecting with all the customers and sources by arcs with unlimited capacities.

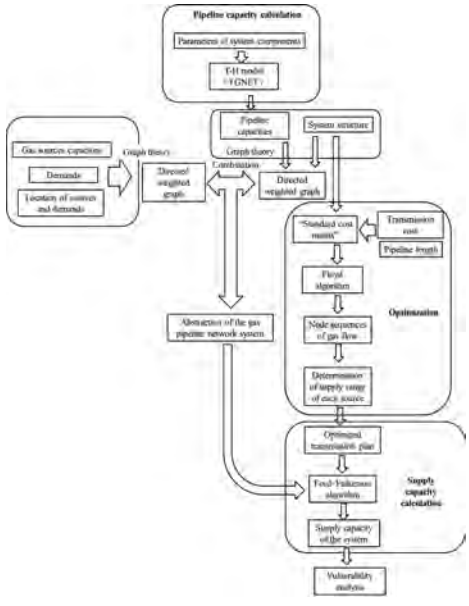


Figure 2. Flowchart of the capacity/consequence analysis model.

Finally, the model will output the supply capacities of the overall system and the customers in the gas pipeline network. By comparing with the demands requested, the amount of unsatisfied customers and the gaps between supply and demand are calculated as inputs of the element criticality analysis. The flowchart of this process is shown in Figure 2.

3 CRITICAL PIPELINE ANALYSIS OF GAS SUPPLY SERVICE

Because of its structure complexity, there may be some unknown, perhaps previously unimagined system weaknesses in a large natural gas pipeline network. Critical element analysis focuses on identifying this kind of weakness that contributes to the supply vulnerability of the gas pipeline network. A pipeline, or a combination of pipelines, is a critical component depending on how essential it is for the supply service of the pipeline network system. Compressor stations, which can also affect the supply capacity, but the impacts of their failures will eventually amount to degradation of pipeline transmission capacities.

Generally, the critical component analysis is performed by estimating the consequences of failures

of single component or multiples (Fang & Zio, 2013; Nicholson, Barker, & Ramirez-Marquez, 2016b; E. Zio & Golea, 2012). This method is performed exhaustively up to a given number (from 1 to P) of simultaneous failures. The value of P is usually from 3 to 5. In this range, every possible failure or failure combinations should be simulated, with the consequence analysis model developed in Section 1. This kind of analysis gives a comprehensive picture of the system vulnerability within the range of simultaneous failures considered.

However, in a complex natural gas network, the attack-consequence analysis method sometimes can be impossible because of the huge burden of computation. To overcome this problem, a flow-based method is applied in our research, which combines topology properties and flow properties of the gas pipeline network, to measure the criticalities of the pipelines.

Several works have explored the measure of arc criticality in a network by graph theory and network flow theory. On account that the importance of a pipeline depends on its role in the network topology and its flow capacity, an index, named weighted flow capacity rate (*WFCR*) (Nicholson et al., 2016a), is chosen to evaluate the criticalities of the pipelines. *WFCR* is the combination of flow capacity rate (*FCR*) (Nicholson et al., 2016a) and flow centrality (*FC*) (Freeman, Borgatti, & White, 1991). *FCR* is the sum of the percentages of arc flows to arc capacity for all *s-d* max flow problems: $c_{i,j}$ represents the transmission capacity of pipeline (i, j); N is the number of *s-d* pairs. *FC* represents the sum of flow in pipeline (i, j) for all possible s-d pair max-flow problems divided by the sum of all pairs max flows: $MF_{sd}(i, j)$ denotes the flow on pipeline (i, j) when the max flow MF_{sd} is from s to d (s represents gas sources and d represents demand sites).

The *FC* provides the contribution of pipeline (i, j) to the transmission capacity of the pipeline network and *FCR* accounts for its potential criticality due to constraints of capacity. Therefore, by weighting each term in equation 3 by the *FC* value, *WFCR* represents the expected impact of pipeline (i, j) to the system performance of supply:

$$FC_{i,j} = \frac{\sum_{s,d \in V} MF_{sd}(i,j)}{\sum_{s,d \in V} MF_{sd}} \tag{2}$$

$$FCR_{i,j} = \frac{1}{N} \sum_{s,d \in V} \frac{MF_{sd}(i,j)}{c_{i,j}} \tag{3}$$

$$WFCR_{i,j} = \frac{1}{N} \sum_{s,d \in V} FC_{i,j} \frac{MF_{s,d}(i,j)}{c_{i,j}} \tag{4}$$

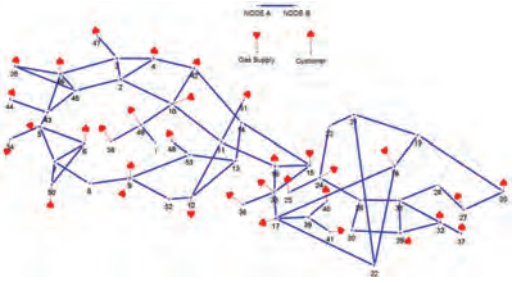


Figure 3. Topology of the gas network.

Table 1. Properties of the gas sources of the gas network.

Location	Type	Limit (MCM/d)
9	LNG terminal	4
10	Pipeline	31
15	LNG terminal	10
18	Pipeline	25
50	LNG terminal	7.1

Table 2. Demands of customers of the gas network.

Location	Demand (MCM/d)	Location	Demand (MCM/d)
4	1.43	34	1.00
5	1.57	35	1.00
6	1.66	36	1.74
9	1.46	37	1.30
12	4.40	38	1.00
16	1.54	40	2.00
17	0.50	41	1.40
20	1.50	42	0.50
24	1.60	44	1.06
25	1.80	46	1.82
27	2.50	47	0.68
29	2.00	48	1.17
32	0.80	49	2.00
33	0.80	51	0.98

4 CASE STUDY

We consider a complex natural gas pipeline network, assuming coherent and reasonable data and information of the pipeline network, including customer demands, pipeline parameters, compressor station parameters and gas source capacity. The pipeline network is shown in Fig. 3. Assumptions about locations, capacities and types of gas sources are reported in Table 1, whereas the demands are listed in Table 2.

5 RESULTS OF CRITICAL COMPONENT ANALYSIS

Firstly, to estimate the criticalities of pipelines, an overview of consequences of the "directly attack" method have been analysed. The exhaustive analyses were performed for *N*-1 to *N*-3 simultaneous failures. All the possible consequences are sorted from high to low and presented in Figure 4. The pipelines or the combinations with high consequences have relative high criticalities.

However, for a complex gas pipeline network with hundreds, or even thousands of pipelines, the direct attack method requires a quite large number of simulations and it is impossible to perform exhaustive analyses under high-order scenarios. Hence, in Section 2, the weighted flow capacity rate (*WFCR*) in equation 2, is proposed to measure the criticalities of pipelines. The pipelines were sorted from highest to lowest according to *WFCR* in Table 3. For a further comparison analysis, *FCR*, the measurement of the pipelines potential criticality due to capacity limitation, and *FC*, the measurement of contributions of pipelines to system transmission capacity, were also calculated. The sequences of pipeline criticalities based on *FC* and *FCR* are shown in Table 3.

In Table 3, the pipelines which are high-ranked by *FCR* are more prone to become bottlenecks because of the low margins between their capacities and loads. The pipelines with higher *FC* values have heavier burdens during the process of gas transmission in the gas pipeline network. The *WFCR* index combines the concepts of both *FCR* and *FC*, and the pipelines with higher *WFCR* values are more critical for the gas supply capacity of the gas transmission network.

The effectiveness of the proposed method was verified by a "random attack & preparedness policy" method. In the "random attack & preparedness policy" method, random attacks were firstly carried out on the pipeline network. The numbers of random attacks of *N*-1, *N*-2, *N*-3 simultaneous failures are respectively 100, 10000 and 500000. When the pipelines are sampled, their capacities will be reduced to zero. Secondly, six preparedness policies were performed. The first policy selects no pipeline to harden. The second policy randomly selects 15% of the pipelines to harden. The 3rd – 5th policies select the top 15% critical pipelines based on FC, FCR and WFCR, to harden, respectively. The 6th policy selects the top 15% critical

Table 3. Pipeline ranking according to importance calculation based on different methods.

FC		FCR		WFCR	
From	To	From	To	From	To
10	2	10	2	10	2
16	11	18	19	10	11
18	19	17	39	10	49
10	11	43	5	18	19
10	49	16	33	16	11
11	12	23	24	16	33
19	20	31	29	11	51
16	33	10	11	11	12
2	3	20	27	20	27
20	27	18	22	17	39
2	45	9	52	23	24
18	22	9	53	18	22
11	51	11	51	10	42
10	42	14	13	18	17
18	17	31	32	33	17
33	17	39	40	43	5
19	21	46	43	19	21
21	23	33	17	21	23
23	24	10	42	2	4
17	39	18	17	3	47
3	46	22	21	50	6
27	28	19	21	31	29
28	31	21	23	22	21
45	43	15	33	9	53
13	12	29	32	14	13
43	5	2	45	39	40
31	29	16	11	3	46
21	20	21	20	2	45
22	21	10	49	29	32
2	4	2	4	2	3
3	47	3	47	21	20
22	17	5	34	15	33
50	6	30	29	46	35
9	53	32	37	19	20
14	13	33	36	24	25

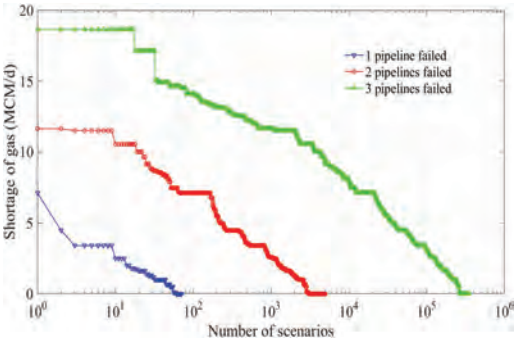


Figure 4. Distribution of shortage for N-1, N-2, N-3 simultaneous failures.

pipelines or pipeline combinations, according to the consequences of direct attacks. In the preparedness policies, the selected pipelines will maintain 70% of their normal capacities after attack. The consequences were represented by "100% × (shortage/normal demand)". The results of the "random attack & preparedness policy" simulations are shown in Figures 5–7 in the form of box plots. Supplementary statistic information is listed in Table 4.

According to the information in Table 4 and Figures 5–7, we can conclude that all the criticality-based policies can reduce the loss significantly compared with "random policy" and "do nothing". The combination index, *WFCR*, is more effective than *FCR* and *FC*, and is a little worse than the criticalities based on direct consequence calculation. However, the burden of computation of the flow-based method is far less than the direct attack method.

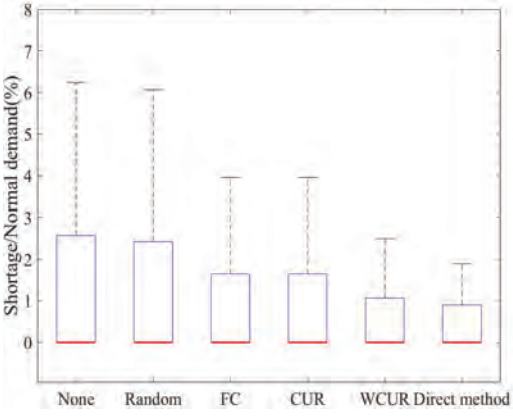


Figure 5. Supply vulnerability by different preparedness policies (1 pipeline failure).

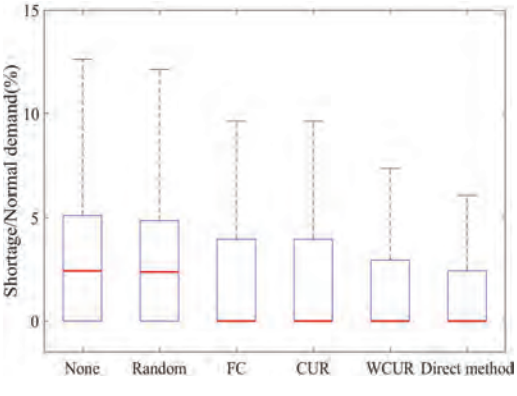


Figure 6. Supply vulnerability by different preparedness policies (2 pipelines failures).

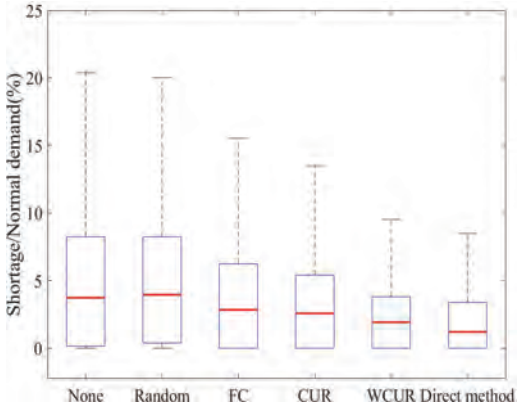


Figure 7. Supply vulnerability by different preparedness policies (3 pipelines failures).

Table 4. Supplementary information of vulnerability of the gas pipeline network by preparedness policy.

	1 pipeline failure		2 pipelines failures		3 pipelines failures	
Policy	Mean (%)	Var	Mean (%)	Var	Mean (%)	Var
None	1.77	10.68	3.56	21.44	5.43	31.17
Random	1.64	9.93	3.32	19.73	5.19	31.06
FC	1.18	7.54	2.63	15.94	4.13	24.29
FCR	1.22	7.16	2.49	15.54	3.85	23.49
WFCR	1.01	4.41	1.89	10.23	2.70	12.89
Direct attack	0.93	4.61	1.37	4.10	2.19	8.30

6 CONCLUSIONS

In this work, we develop a flow-based method to identify the critical pipelines in natural gas pipeline network from the vulnerability perspective. Comparing with the large computational cost of the direct attack method when it is used for large complex pipeline networks, the flow-based criticality measurement used in this work is much more efficient with a little loss of effectiveness. The flow-based method combines the topology and the capacity contribution of a pipeline to represent its criticality to the supply service of the gas transmission system. In the case study, both approaches have been performed and their effectiveness has been compared by the "random attack & preparedness policy" simulation. The results show that accuracy of the flow-based method is slightly lower than that of the direct "attack-consequence" method; but, the computation burden of the latter is far more higher than that of the former.

ACKNOWLEDGMENTS

This work is supported by the fund [grant number 51134006].

REFERENCES

Ahuja, R. et al. (1993). Network flows/: theory, algorithms, and applications (1st ed.). Prentice Hall.

Cadini, F. et al. (2017). A modeling and simulation framework for the reliability/availability assessment of a power transmission grid subject to cascading failures under extreme weather conditions. *Applied Energy*, *185*, 267–279. https://doi.org/10.1016/j.apenergy.2016.10.086.

Demirel, H. et al. (2015). A framework to analyze the vulnerability of European road networks due to Sea-Level Rise (SLR) and sea storm surges. *Transportation Research Part A: Policy and Practice*, *81*, 62–76. https://doi.org/10.1016/j.tra.2015.05.002.

Deng, Y. et al. (2015). A research on subway physical vulnerability based on network theory and FMECA. *Safety Science*, *80*, 127–134. https://doi.org/10.1016/j.ssci.2015.07.019.

Fang, Y. & Zio, E. (2013). Hierarchical Modeling by Recursive Unsupervised Spectral Clustering and Network Extended Importance Measures to Analyze the Reliability Characteristics of Complex Network Systems. *American Journal of Operations Research*, *3*(1), 101–112. https://doi.org/10.4236/ajor.2013.31A010.

Fang, Y. et al. (2015). Optimization of Cascade-Resilient Electrical Infrastructures and its Validation by Power Flow Modeling. *Risk Analysis*, *35*(4), 594–607. https://doi.org/10.1111/risa.12396.

Freeman, L. et al. (1991). Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks*, *13*(2), 141–154. https://doi.org/10.1016/0378-8733(91)90017-N.

Han.F;E.Zio. (2014). Analyzing controllability, efficiency and reliability of network systems by dynamic simulation.

Hong, L. et al. (2015). Vulnerability assessment and mitigation for the Chinese railway system under floods. *Reliability Engineering and System Safety*, *137*, 58–68. https://doi.org/10.1016/j.ress.2014.12.013.

Johansson, J., & Hassel, H. (2010). An approach for modelling interdependent infrastructures in the context of vulnerability analysis. *Reliability Engineering and System Safety*, *95*(12), 1335–1344. https://doi.org/10.1016/j.ress.2010.06.010.

Kröger, W., & Zio, E. (2011). Chap1 Introduction and Definition of Key Terms. In *Vulnerable systems* (pp. 1–7). London: Springer London. https://doi.org/10.1007/978-0-85729-655-9.

Mattsson, L.G., & Jenelius, E. (2015). Vulnerability and resilience of transport systems—A discussion of recent research. *Transportation Research Part A: Policy and Practice*, *81*, 16–34. https://doi.org/10.1016/j.tra.2015.06.002.

Nicholson, C.D. et al. (2016a). Flow-based vulnerability measures for network component importance: Experimentation with preparedness planning. *Reliability Engineering and System Safety*, *145*, 62–73. https://doi.org/10.1016/j.ress.2015.08.014.

Nicholson, C.D. et al. (2016b). Flow-based vulnerability measures for network component importance: Experimentation with preparedness planning. *Reliability Engineering and System Safety*, *145*, 62–73. https://doi.org/10.1016/j.ress.2015.08.014.

Ouyang, M. et al. (2014). Comparisons of complex network based models and direct current power flow model to analyze power grid vulnerability under intentional attacks. *Physica A: Statistical Mechanics and Its Applications*, *403*, 45–53. https://doi.org/10.1016/J.PHYSA.2014.01.070.

Su, H. et al. (2017). An integrated systemic method for supply reliability assessment of natural gas pipeline networks. *Applied Energy*, *209*(October 2017), 489–501. https://doi.org/10.1016/j.apenergy.2017.10.108.

Thekdi, S.A., & Santos, J.R. (2016). Supply Chain Vulnerability Analysis Using Scenario-Based Input-Output Modeling: Application to Port Operations. *Risk Analysis*, *36*(5), 1025–1039. https://doi.org/10.1111/risa.12473.

Zio, E. (2007). From complexity science to reliability efficiency: A new way of looking at complex network systems and critical infrastructures. *International Journal of Critical Infrastructures*, *3*(3–4), 488–508. https://doi.org/10.1504/IJCIS.2007.014122.

Zio, E. (2014). Vulnerability and Risk Analysis of Critical Infrastructures. *Vulnerability, Uncertainty, and Risk*, *1*(2), 23–30. https://doi.org/10.1061/9780784413609.003.

Zio, E. (2016). Challenges in the vulnerability and risk analysis of critical infrastructures. *Reliability Engineering and System Safety*, *152*, 137–150. https://doi.org/10.1016/j.ress.2016.02.009.

Zio, E., & Golea, L.R. (2012). Analyzing the topological, electrical and reliability characteristics of a power transmission system for identifying its critical elements. *Reliability Engineering and System Safety*, *101*, 67–74. https://doi.org/10.1016/j.ress.2011.11.009.

Zio, E., & Piccinelli, R. (2010). Randomized flow model and centrality measure for electrical power transmission network analysis. *Reliability Engineering & System Safety*, *95*(4), 379–385. https://doi.org/10.1016/j.ress.2009.11.008.

Zio, E., & Sansavini, G. (2013). Vulnerability of smart grids with variable generation and consumption: A system of systems perspective. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, *43*(3), 477–487. https://doi.org/10.1109/TSMCA.2012.2207106.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Evaluation of a community pharmacy dispensing process using a coloured Petri Net

M. Naybour & R. Remenyte-Prescott
Resilience Engineering Research Group, Faculty of Engineering, University of Nottingham, University Park Nottingham, Nottingham, UK

M. Boyd
Pharmacy Policy and Practice Research Group, Faculty of Science, University of Nottingham, University Park Nottingham, Nottingham, UK

ABSTRACT: UK customers visited community pharmacies to receive NHS prescriptions 1.104 billion times in 2016. One study of dispensing errors found an error rate of 3.3%. Severe dispensing inaccuracies often receive a high level of media attention, however, lower level errors could also be causing significant inefficiencies in the delivery of primary healthcare. This paper presents a modelling approach for analysing the reliability and efficiency of community pharmacies performance using a Coloured Petri Net (CPN) methodology. The model considers how single prescriptions are processed, the use of staff resources, and the occurrence of errors. The CPN evaluates performance over a set of key performance indicators. Results are validated, where possible, against published studies of community pharmacies.

1 INTRODUCTION

1.1 *Background*

Over the past 50 years there has been a growing awareness that healthcare systems are capable of inflicting harm to patients, and this harm should be reduced (Health Foundation, 2011). Two key reports by the US Institute of Medicine (Mullan et al, 2001) and the UK Department of Health (DoH, 2000) helped to spread the message that iatrogenic patient harm within healthcare systems is an important issue. Notably, if the community pharmacy dispensing error rate of 3.3% (Franklin & O'Grady, 2007) is considered, this could mean that around 36 million UK prescriptions per year contain errors.

As well as safety concerns, studies have shown that patient satisfaction with pharmacy services is linked to waiting times (Afolabi & Erhun, 2003). Extended waiting times have been given as a reason why patients will not return to a particular pharmacy (Somani & Daniels, 1982), and content customers are increasingly likely to return to their specific healthcare provider (Dansky & Miles, 1997).

1.2 *Reliability engineering*

Reliability engineering techniques are used by many industries and it has become common for complex systems to be subjected to risk assessment processes (Andrews, 2009). These assessments have historically been carried out in conventional high risk industries, such as the aviation (Netjasov & Janic, 2008), nuclear (Hsueh & Mosleh, 1996) and space sectors (Garrik, 1988), where effects of failure can be catastrophic.

Fault trees and event trees are an example of a widely used reliability engineering techniques. They use combinatorial logic to combine events to produce both qualitative and quantitative analysis of failures (Vesely et al, 2002). Fault tree analysis requires that the occurrence of events is independent.

Markov models are memoryless processes capable of modelling more complex systems, which might typically contain repair strategies and dynamic behavior (Boyd, 1998). A key limitation to implementing a Markov model for a given system, arises from the fact that the number of system states to consider grows exponentially with the number of components in the system.

Petri Nets are an effective tool for modelling processes or systems exhibiting concurrency (Schneeweiss, 1999). Since the publication of Carl Adam Petri's thesis in 1961, a number of extensions of the basic technique have been developed. Two important examples of Petri Net extensions are timed and Coloured Petri Nets (Jensen, 1996). Timed nets use either deterministic or stochastic delay timings, to control the timing of transitions. This gives the opportunity to model temporal processes. Meanwhile, incorporating token colour sets into Petri Net modelling enables token specific information to be propagated around the net. This can then be used to

control and manipulate the nets behavior. Coloured Petri Nets have been utilized to model complex systems in a wide range of areas (Liu, 2017).

The healthcare sector, primary care especially, represents a relatively new area for reliability modelling. Previous healthcare modelling studies have been centred in secondary healthcare settings. In this field, Petri Nets have been used to model hospital departments (Dotoli et al, 2010), hospital information systems (Darabi & Galanter, 2009), and mental health care services (Dammasch & Horton, 2007). Michael R. Cohen et al utilized fault trees to conduct a risk assessment of dispensing in community pharmacies (Cohen et al, 2012), and their error probabilities are also used in this paper.

The novelty of the proposed approach in this paper is the ability to perform safety and efficiency evaluation within the framework of a single modelling technique. Therefore, a timed CPN model is developed and a wide range of performance indicators is obtained, using simulations. Model outputs can be used to support resource management and safety improvement decisions. The community pharmacy dispensing process is presented in section 2, section 3 outlines how the model is built, section 4 presents results and analysis and section 5 concludes the paper.

2 COMMUNITY PHARMACY DISPENSING PROCESS

2.1 *The main stages of dispensing*

A standard community pharmacy dispensing process is described in this section. The six key stages of the community pharmacy dispensing process are given in Figure 1 (Langley & Belcher, 2009 & NPSA, 2007 & Waterfield 2008).

To begin with, prescriptions must be received by a member of staff as and when patients bring them into the pharmacy. Prescriptions are then legally and clinically checked, to ensure that the prescription is clinically appropriate before continuing. After being received, the prescriptions' labels are generated. The labels include key information about the medicine. The next stage of the process is bringing the constituent parts of the prescription together to create the final product. First, the set of items included on the prescription is gathered together from the pharmacy stock. After this, an intermediate accuracy check is recommended, before applying the labels to medicines. After the prescription is fully assembled, it is passed onto either a pharmacist or an ACT (Accredited Checking Technician) to perform a final accuracy check on the prescription. The final accuracy check is the final opportunity for a pharmacy to intervene if a prescription has been dispensed incorrectly at some point in the process. The accuracy check involves making sure that the prescription being provided by the pharmacy exactly matches what has been written on the prescription form. This includes checking that the labels, items, doses, quantities and form of medication are all correct before handing the prescription out. Any mistakes that go unnoticed at the final accuracy check are likely to reach patients. Each stage of the process can be completed by a single member of staff, although, only pharmacists or ACTs are qualified to final accuracy check prescriptions.

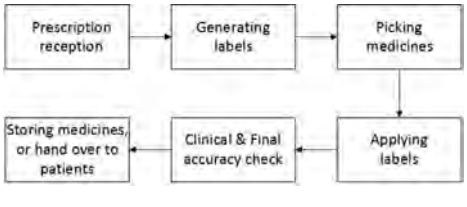


Figure 1. Dispensing process flow chart.

2.2 *Resources*

A typical community pharmacy staff team consists of a group of pharmacists, ACTs and dispensers, but the number of staff varies between pharmacies. Larger stores can have teams of up to 12 people, while the smallest independent store may be run by a single pharmacist. However, for a pharmacy to be allowed to dispense prescriptions, there must be a responsible pharmacist present during all hours of operation.

The full list of resources used in the dispensing process is as follows: prescriptions, dispensers, pharmacists, medicines, labels, labelling stations and a private room.

2.3 *Non-dispensing tasks*

As well as completing dispensing tasks, there are a number of non-dispensing tasks in pharmacies that members of staff are required to complete (Davies et al, 2014). These non-dispensing tasks include, stock management, patient counselling, advanced pharmacy services, non-prescription services, staff training, and general housekeeping. Advanced services are a set of 6 services offered in pharmacies, one example of which is the smoking cessation service.

In this study, the set of non-dispensing tasks requiring to be completed by staff is limited to stock management, advanced services and patient counselling. Although not strictly a task, lunch hours for dispensers are also included in the model.

2.4 *Failure modes*

Dispensing correct prescriptions reliably and in a time that is convenient for customers are the two main goals of community pharmacies. Therefore, the dispensing process can be considered to fail if either:

1. A prescription is incorrect when handed/delivered to a patient.
2. A prescription takes an extended amount of time to be dispensed, causing the patient to decide not to return in the future.

Prescriptions can be incorrect in a number of different ways, for example, the labels may indicate to take too much or too little of the medicine. This would be classified as a labelling error. Other examples include, items being included which are different to those prescribed. This would be classified as a contents error, and it can be due to wrong dose, wrong volume, or being a completely different medicine. Additionally, it may be the case that the labels and items were generated and picked correctly, but they are mixed up when applying the labels, this is classified as a label application error.

If one of the above errors makes it through the final accuracy check and is handed out to a patient, this is then classified as a dispensing error. If however, the error is spotted and rectified at the final accuracy check, this is classified as a near miss (Chua et al, 2003).

2.5 *Definitions: Process reliability and efficiency*

Reliability of the dispensing process, R, is defined in Equation (1) as:

$$R = \frac{p_{cc}}{p_{total}} \tag{1}$$

where p_{cc} is the number of prescriptions dispensed which are completely correct, and p_{total} is the total number of prescriptions dispensed.

Process efficiency is commonly defined as the ratio between an output gained and the level of resources needed to maintain the process. Since the cost of resources is not factored into this study, a set of efficiency indicators are used. Two examples of efficiency indicators are, the total number of prescriptions completed, and the average time to dispense walk-in prescriptions. Results for all performance indicators can be found in Table 4. The ideal outcome of the process in terms of efficiency is a high number of prescriptions completed quickly.

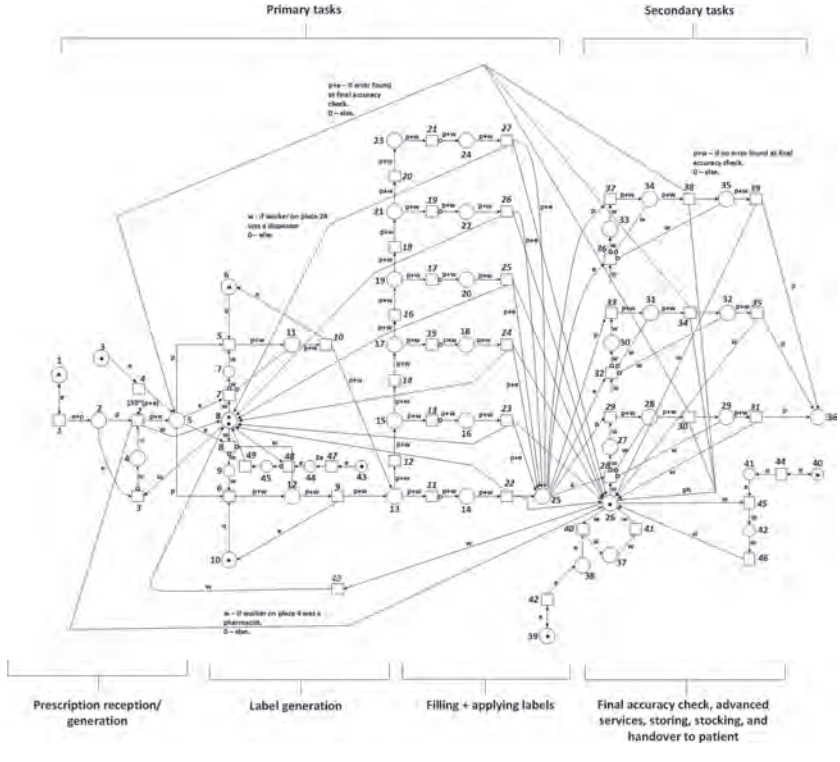


Figure 2. A CPN model for community pharmacy dispensing.

Table 1. Places.

Place	Description	Type
1	Walk-in task generator.	e
2	Customer at counter.	e
3	Delivery task generator.	e
4	Staff receiving.	w
5	Prescriptions to be dispensed.	p, e
6, 10	Labelling stations available.	e
7, 9	Staff member choosing prescription.	w
8	Staff available for primary tasks.	w
11, 12	Staff member generating labels.	w, p
13, 15, 17 19, 21, 23	These places are used to separate staff into parallel work streams.	w, p
14, 16, 18 20, 22, 24	Staff are assembling, and applying labels to prescriptions.	w, p
25	Prescriptions waiting for secondary dispensing tasks.	p, e
26	Pharmacists available to complete secondary tasks.	w
27, 30, 33	Pharmacist allocated to complete secondary tasks for a prescription.	w
28, 31, 34	Pharmacists is checking a prescription.	w, p
29, 32, 35	Pharmacist is handing out/storing for delivery.	w, p
36	All completed prescriptions.	p
37	Advanced service being completed.	w
38	Advanced service waiting.	e
39	Advanced service task generator.	e
40	Stocking task generator.	e
41	Stocking waiting.	e
42	Stocking task being completed.	w
43	Dispenser lunch break generator.	e
44	Lunch break ready to be taken.	e
45	A dispenser is on their lunch break.	w

Table 2. Transitions.

Transition	Description	(Y/N)*
1	Walk in generation: Exp(0.0033)	N
2	Receive a prescription: Uni(30, 60)	N
3	Move staff to counter: Det(0)	N
4	Delivery generation: Det(6000)	N
5, 6	Staff choose prescription: Uni(5,10)	N
7, 8	Allocate a staff member: Det(∈)	N
9, 10	Label generation: Det(15)	Y
11–21	Spreaders: Det(∈)	N
22–27	Filling & label application: N(50,10)	Y
28, 32, 36	Pharmacist allocation: Det(∈)	N
29, 33, 37	Choose prescription: U(10, 15)	N
30, 34, 38	Final accuracy check: Uni(5,10)	Y
31, 35, 39	Hand out and counsel: Exp(0.025)	N
31, 35, 39	Store for delivery: Exp(0.05)	
40	Allocate to advanced service: Det(∈)	N
41	Complete advanced service: Uni(300, 600)	N
42	Advanced service generator: Exp(0.00006)	N
43	Move pharmacist primary: Det(10)	N
44	Stocking task generator: Det(6600)	N
45	Allocate to stocking: Det(∈)	N
46	Finish stocking: Uni(300, 900)	N
47	Begin triggering of lunch break: Det(7200)	N
48	Allocate dispenser to lunch: Det(∈)	N
49	Dispenser finished lunch: Det(3600)	N

*This column designates transitions as processors.

3 MODELLING APPROACH

3.1 *Overview*

This section of the paper presents the development of a Coloured Petri Net (CPN) for modelling the dispensing process. The dispensing process being modelled in this study is that of manual dispensing pharmacy, as opposed to automated dispensing. Figure 2 shows the CPN model of a community pharmacy. Overall, the model is built according to the process flow, considering resources and errors. Model outputs are obtained after the CPN model is simulated.

3.2 *Places and transitions*

Table 1 shows the description of each place and the type of token that may occupy the places. Note that the net uses three token types: e (basic), w (staff), and p (prescriptions).

Overall, some places are used to keep track of resources, and others are used as task generators, controlling when new tasks arrive.

Table 2, shows the description and distribution of each transition. Note that Det(x) stands for a deterministic delay. Some transitions directly represent the community pharmacy dispensing tasks seen in Figure 1. Other transitions are purely used to move tokens around the net. The types of distributions and their parameter values have been assumed in this paper.

In Table 2 each transition is also designated as either a 'processor' transition, or not. A processor transition represents a task that is affected by the number of items in the prescription. For example, the transition, modelling generating labels, is a processor transition, since it will take longer to generate labels for a large prescription.

3.3 *Model assumptions*

Tasks in the model are separated into primary and secondary tasks, where primary tasks may be com-

pleted by all staff, whereas secondary tasks may only be completed by pharmacists. In addition a number of assumptions about staff behaviour and pharmacy specification are made. Below is a list of modelling assumptions about how staff behave.

- Staff complete tasks in an identical way, i.e. the same probability distributions are used to determine how long tasks take, and to generate error probabilities for different staff.
- Dispensers may only complete primary tasks, and pharmacists prioritise secondary tasks. Pharmacists are able to move to primary tasks if they are idle.
- Once primary work is begun on a prescription, the same member of staff continues working on it until the primary tasks are finished.
- Upon a customer arriving with a walk-in, the first member of staff to become available for primary tasks go to serve them.
- Dispensers have a lunch hour. It is assumed that pharmacists fit their lunch in during moments when they are not working.

Below are assumptions about the labelling stations, pharmacy opening hours, and prescriptions.

- The pharmacy is open from 9 am-5 pm.
- Walk-in prescriptions are prioritised over deliveries. Within the same type, there is a first come first served order. They arrive with increments of an Exponential distribution, as shown in Table 2.
- Delivery prescriptions arrive at the pharmacy in a single large bulk, at 10 am, 1 hour after the pharmacy opens.
- The pharmacy has 2 labelling stations capable of generating labels for prescriptions.
- Walk-ins taking longer than 15 minutes to be dispensed are classed as delayed.

3.4 *Prescription modelling*

In the CPN model, prescription tokens each have 8 colour fields which represent:

1. Delivery or walk-in
2. The number of items
3. Time taken to dispense
4. Number of iterations to compete
5. The overall outcome
6. Label error
7. Content error
8. Label application error

Table 3. Error probabilities.

Task	Error probability
Labelling	0.06
Filling	0.05
Label application	0.03
Final accuracy check	0.05

In particular the number of iterations to complete is determined by how many times a pharmacist has had to send the prescription to be corrected after a final accuracy check. The overall outcome is one of 3 outcomes: completely correct, near miss, or dispensing error. The last 3 colours, labels, contents and label application, are Boolean variables, which indicate whether an error of each type is contained within the prescription.

Upon arrival, every prescription is allocated a random number of items by sampling from a Geometric (0.35) random variable (mean = 2.86). This was chosen using two assumptions. Firstly, patients with a prescription will have at least 1 item on the prescription. Secondly, prescriptions with more items are increasingly less likely to occur than those with fewer. This number of items is then used to determine how long the processor transitions, designated in Table 2, take to fire. For example, a prescription containing 5 items will use the sum of 5 samples from the distribution that describes the duration of label generation.

3.5 *Failures*

Failures are modelled using Bernoulli random variables. At three points of the process, label generation, prescription assembly and label application, an error can occur. The error probabilities were taken from Cohen et al (Cohen et al. 2012), and are shown in Table 3.

The outcome of the final accuracy check depends on the state of the prescription being checked. It is assumed that prescriptions that are correct will always pass through the check. If there is an error present in the prescription, the pharmacist will spot it with probability 0.95, otherwise they will fail to spot it with probability 0.05.

4 PHARMACY SIMULATION SCENARIOS AND THEIR ANALYSIS

4.1 *Scenario specification*

This paper uses three pharmacy scenarios to demonstrate the ability to evaluate performance using the CPN model. These three scenarios have been chosen to demonstrate the impacts, or efficiency improvements, of adding an additional staff member.

a. Scenario 1

Staff – 1 pharmacist, 2 dispensers

Failures—Chance of failure in labelling, filling, label application and final accuracy check stages.

Advanced services—Included.

Stocking—Pharmacist must do 4 stints of stock management, each period lasting 5–15 mins.

Lunch hours – 1 hour for each dispenser, taken sequentially (only 1 dispenser may be off at the same time).

b. Scenario 2

Same as scenario 3, but with 1 pharmacist and 3 dispensers.

c. Scenario 3

Same as scenario 1, but with 2 pharmacists and 2 dispensers.

4.2 *Results and analysis*

A 9-5 day of pharmacy operation was simulated a total of 6000 times for each scenario. A test for convergence was conducted to find whether 6000 was a large enough number to reach convergence. A further 1000 simulations were carried out for each scenario, then the indicator values for the set of 7000 simulations were compared to the values calculated for 6000 simulations. Every field was the same between the two sets of data to 2 significant figures.

Results of key performance indicators for each scenario are shown in Table 4.

Since walk-in (WI) prescriptions are given priority over delivery prescriptions, walk-in prescriptions get completed first, but a smaller pharmacy which takes longer to dispense prescriptions is unable to complete all their deliveries. This can be seen in scenario 1, where 39 of the 150 delivery prescriptions are unfinished. In both scenarios 2 and 3, having an additional staff member of either type (pharmacist or dispenser) improved the efficiency of the pharmacy sufficiently so that on average almost all the deliveries were being completed. This suggests that the pharmacy may be able to complete a larger number of delivery prescriptions when employing 4 staff. The average time to dispense was also improved by more staff in scenarios 2 and 3. A large decrease (of 217s) in the average time to dispense walk-ins was seen when introducing an extra pharmacist in scenario 3. A smaller decrease (of only 75s) was gained by introducing an extra dispenser to the pharmacy team in scenario 2.

Previous studies have reported near miss rates of between 0.024% (Knudsen et al, 2007) and 1.84% (Sanchez, 2013), and dispensing error rates of between 0.014% (Knudsen et al, 2007) and 3.3% (Franklin & O'Grady, 2007). There are many more near misses occurring during the simulations than have been seen in previous studies of errors, i.e. all 3 scenarios had near misses occurring in over 10% of all prescriptions being dispensed. This may be due to underreporting of near-misses in self report based studies, or the final accuracy check failure probability is set too low in the model. The dispensing error rate produced by simulations fell within the reported range.

These simulations suggest that the simulated dispensing process has good reliability. The reliability for scenarios 1, 2 and 3 were as follows, $R_1 = 0.992$, $R_2 = 0.992$, $R_3 = 0.992$. The same reliability for tall three scenarios is due to the fact that the error rates do not depend on the type of staff and pharmacy set-up.

4.2.1 *Distribution of time to dispense*

Figure 3 shows how the distribution of the time to dispense walk-ins depends on the scenario. The duration of 600,000 walk-in prescriptions were used for comparison, i.e. around 100 walk-in prescriptions from each of the 6,000 simulations. It can be seen in Figure 3 that all scenarios have a similar underlying distribution. However the skewness decreases with each additional member of staff. A larger decrease in skew is seen when an additional pharmacist is added. Note that the dashed vertical line represents 15 min dispensing time limit.

4.2.2 *Causes of delays*

A prescription could be delayed due to one of many reasons, such as, prescriptions containing more items taking longer to dispense, delays due to a large amounts of walk-ins already being processed or waiting in the queue when a patient arrives, members of staff being busy with non-dispensing activities, or due to a near miss that has been picked up at the final accuracy check.

Table 5 shows how looking at single scenarios, for more increasingly delayed prescriptions, the average size, and number of iterations required to complete prescriptions increases. This appears to confirm the prospect that prescriptions which contain more items, or need to be dispensed multiple times are more likely to be delayed.

Table 4. Simulation results.

	Efficiency					Reliability		
Scenario	Deliveries completed	Total completed	Advanced services completed	Delayed	WI dispense time mean sec	R	Near misses	Dispensing errors
1	111.1	211	1.8	25.0	711	0.992	29.7	1.6
2	149.4	250	1.8	19.3	636	0.992	32.8	1.9
3	149.5	250	1.8	8.3	494	0.992	32.9	1.9

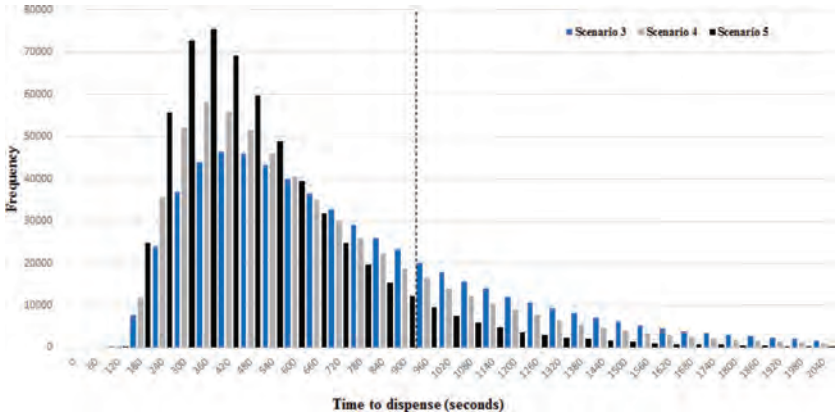


Figure 3. Distributions of the time taken to dispense walk-in prescriptions.

Table 5. Causes of delays.

Scenario		t < 15 mins	15 ≤ t < 20 mins	20 ≤ t < 25 mins	25 ≤ t < 30 mins	30 ≤ t < 35 mins	35 ≤ t < 40 mins	40 ≤ t mins
1	% of total	75.08	12.64	6.15	3.01	1.49	0.754	0.884
	Avg items	2.37	3.83	4.20	4.49	4.76	5.02	5.79
	Avg itts	0.074	0.226	0.349	0.498	0.673	0.838	1.24
2	% of total	80.66	10.40	4.72	2.17	1.03	0.495	0.533
	Avg items	2.40	4.10	4.52	4.88	5.25	5.70	6.70
	Avg itts	0.0803	0.276	0.425	0.606	0.796	0.997	1.33
3	% of total	91.74	5.27	1.77	0.66	0.295	0.141	0.122
	Avg items	2.50	5.70	6.46	6.57	7.33	8.21	9.51
	Avg itts	0.104	0.485	0.765	1.027	1.232	1.37	1.73

Comparing scenarios, it can be seen that scenarios 2 and 3 offer an improvement in the number of walk-in prescriptions being completed on time. Scenario 3 increased the percentage of prescriptions being completed on time by 16%, while scenario 2 managed an increase of only 5.5%.

5 CONCLUSION

In conclusion, this paper has demonstrated the use of CPNs as an effective tool for modelling the community pharmacy dispensing process. CPN is a suitable tool to evaluate efficiency and safety in one model. Pharmacy dispensing complexity is captured through: the inclusion of all major dispensing stages, their duration, and a variety of staff roles, errors and remedial action. Adding a pharmacist improved the pharmacy efficiency more than adding a dispenser. Dispensing errors are within the range reported in the literature, whereas near misses are overestimated.

Process reliability remained constant in all scenarios. By assigning staff wage costs to scenarios, this model could support decisions related to the cost-benefit of employing extra staff member.

Future work will focus on optimizing a pharmacy dispensing process. This would involve finding the optimal choice of how many staff should work in the pharmacy, given the working conditions and cost of staff wages. Metaheuristics such as, genetic or ant colony optimisation algorithms, are promising methodologies for this purpose. In addition, in-field data collection would be carried out, and ethical approval has been granted by the University of Nottingham. Other routes for future research could include constructing an alternative model capable of comparing the performance of automated and manual dispensing pharmacies. Future iterations of the model could be designed to include the dependency between the overall state of the pharmacy, and staff error rates. For example, if a pharmacy is busy, with many patients waiting for walk-ins to be dispensed, this could put pressure onto staff, who may be then more likely

to make errors. Another possible improvement to the model could be to consider how errors of each type, labelling, contents or label application, can actually occur in each item in a prescription.

ACKOWLEDGEMENTS

The authors would like to acknowledge and thank the ESPRC for their support and funding, without which, this research would not be possible (Grant ref EP/M50810X/1).

REFERENCES

Afolabi, M. & Erhun, W. 2003. Patients' response to waiting time in an out-patient pharmacy in Nigeria. *Tropical Journal of Pharmaceutical Research.*

Boyd, M. 1998. An Introduction to Markov Modelling: Concepts and Uses. *Reliability and Maintainability Symposium.*

Chua, S. & Wong, I. & Edmondson, H. & Allen, C. & Chow, J. & Peacham, J. & Hill, G. & Grantham, J. 2003. A Feasibility Study for Recording of Dispensing Errors and 'Near Misses' in Four UK Primary Care Pharmacies. *Drug Safety.*

Cohen, M. & Smetzer, J. & Westphal, J. & Comden, S. & Horn, D. 2012. Risk models to improve safety of dispensing high-alert medications in community pharmacies. *Journal of American Pharmacists Association.*

Dammasch, K. & Horton, G. 2007. Active Tokens for Modelling Mental Health Care with Coloured Stochastic Petri Net. *2007 Innovations in Information Technologies (IIT).*

Dansky, K. & Miles, J. 1997. Patient Satisfaction with Ambulatory Healthcare services: Waiting and Filling Time. *Hospital and Health Services Administration.*

Darabi, H. & Galanter, W. & Lin, J. & Buy, U. & Sampath, R. 2009. Modelling and integration of hospital information systems with Petri nets. *IEEE/INFORMS International Conference on Service Operations, Logistics and Informatics.*

Davies, J. & Barber, N. & Taylor, D. 2014. What do community pharmacists do?: results from a work sampling study in London. *International Journal of Pharmacy Practice.*

Department of Health Chief Medical Officer, 2000. An Organisation with a Memory: Report of an Expert Group on Learning from Adverse Events in the NHS. *Stationary Office.*

Dotoli, M. & Fanti, M. & Iacobellis & Ukovich, W. 2010. Modelling and management of a hospital department via Petri nets. *Health Care Management Conference.*

Franklin, B. & O'Grady, K. 2007. Dispensing errors in community pharmacy: frequency, clinical significance and potential impact of authentication at the point of dispensing. *International Journal of Pharmacy Practice.*

Garrick, J. 1988. The approach to risk analysis in three indus-tries: nuclear power, space systems, and chemical process. *Reliability Engineering & System Safety.*

Health Foundation, 2011. Health Foundation Research scan: Levels of Harm. *The Evidence Centre.*

Hsueh, K. & Mosleh, A. 1996. The development and application of the accident dynamic simulator for dynamic probabilistic risk assessment of nuclear power plants. *Reliability Engineering & System Safety.*

Jensen, K. 1996. Coloured Petri nets: basic concepts, analysis methods, and practical use. *Berlin: Springer.*

Knudsen, P. & Herborg, H. & Mortensen, A. & Hellebek, A. 2007. Preventing medication errors in community pharmacy; frequency and seriousness of medication errors. *Qual Saf Health Care.*

Langley, C. & Belcher, D. 2009. Applied pharmaceutical practice. *London: Pharmaceutical Press.*

Liu, X. & Zhang, J. & Zhu, P. 2017. Modeling cyber-physical attacks based on probabilistic colored Petri nets and mixed strategy game theory. *International Journal of Critical Infrastructure protection.*

Mullan, Z. & Benham, L. & Cumber, H. & Dehnel, T. & Clark, S. 2011. To err is human. *The Lancet.*

Netjasov, F. & Janic, M. 2008. A review of research on risk and safety modelling in civil aviation. *Journal of Air Transport Management.*

NPSA. 2007. Design for patient safety: A guide to the design of the dispensing environment. *National Patient Safety Agency.*

Schneeweiss, W. 1999. Petri nets for reliability modeling: (in the fields of engineering safety and dependability). *Hagen: Liole-Verlag.*

Somani, S. & Daniels, C. & Jermstad, R. 1982. Patient satisfaction with outpatient pharmaceutical services. *American Journal of Hospital Pharmacy.*

Sánchez, M. 2013. Medication errors in a Spanish community pharmacy: nature frequency and potential causes. *International journal of clinical pharmacy.*

Vesely, W. & Dugan, J. & Fragola, J. & Minarick, J. & Railsback, J. 2002. Fault Tree Handbook with Aerospace Applications. *NASA Office of Safety and Mission Assurance.*

Waterfield, J. 2008. Community pharmacy handbook. *London: Pharmaceutical Press.*

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A simulation-based safety analysis framework for autonomous vehicles—assessing impacts on road transport system's safety and efficiency

L.F. Vismari, C.B.S.T. Molina, J.B. Camargo Jr. & J.R. Almeida Jr.
Safety Analysis Group—GAS, School of Engineering of the University of São Paulo (Poli-USP), São Paulo, SP, Brazil

R. Inam & E. Fersman
Ericsson Research, Ericsson AB, Stockholm, Sweden

M.V. Marquezini
Ericsson Research, Ericsson Telecomunicações S.A., Indaiatuba, São Paulo, SP, Brazil

ABSTRACT: Advances in Information and Communication Technologies (ICT), pervasive computing and Artificial Intelligence (AI) are affecting all daily human life domains in multiple ways. Safety-critical transport systems are being massively affected by this technological shift. The Autonomous Vehicles (AV) are being considered the most promising new element in this future transport paradigm. It is expected that the AVs bring relevant benefits to the society, mainly reducing accidents and increasing accessibility. Given that any new safety-critical system, concept or technology is placed in operations only if its benefits outweigh the safety risks, assuring these future transport systems will be safe during their operation is a mandatory duty. This paper proposes a safety analysis Framework to be applied to the future intelligent Roadway Transport Systems (RTS) paradigm. Based on a combined fast and real time simulation approaches, the Framework allows analyzing, in a broad sense, the impacts of concepts, technologies and procedures on RTS safety and efficiency. A Proof-of-Concept (PoC) is provided, where representative RTS scenarios are modeled, simulated and analyzed using OpenDS, Matlab and Sumo+Veins+OMNet++. The RTS scenario elements (environment, vehicles, roadways, and driver) interact with each other. Autonomous and 'non-autonomous' vehicles share the RTS infrastructure, and AV is modeled combining a pair of RTS elements "driver+vehicle", with the control algorithms embedded in the autonomous driver element to manage AV movement. In this PoC, both the autonomous vehicle behavior and the impact of its embedded autonomous control algorithms over the RTS safety and efficiency are analyzed. Concluding, the Framework is capable to deal with the representative future RTS scenarios, mainly regarding to AV, analyzing the impact of concepts in component level (e.g. AI-based algorithms) over system level properties (e.g. safety) and the relationship among properties at various levels of RTS.

1 INTRODUCTION

Advances in Information and Communication Technologies (ICT) and pervasive use of computing and Artificial Intelligence (AI) are affecting all domains of daily human life, including those ones that could lead to deaths and health, environmental and patrimonial injuries. One of these safety-critical domains, the Transport systems, is being directly and massively affected by this new technological changing wave. Autonomous Vehicles (AV) - also called as self-driven or robot cars—could be considered the most notably new element in this future paradigm of Intelligent Transport System—ITS (PARLIAMENT & UNION 2010). Regarding to the Road Transport Systems (RTS), it is expected that the AVs will bring significant improvements in traffic safety and efficiency for all modes of transport. For example, it is expected to reduce significantly the number of traffic accidents when the need for a human driver is eliminated, or even reduced, given that the causes to more than 90% of accidents are attributed to human error (Singh 2015). Besides, AVs would increase accessibility to the individual transport by people with disabilities, elderly people, children and even vehicles without people on board (e.g. delivery or rental fleet reallocation).

As like any new safety-critical system, concept or technology, the AV will be incorporated into daily human life only if its benefits outweigh the safety risks. Therefore, it is mandatory assuring that the future safety-critical AVs, as well as the whole transport systems, are going to be safe enough during their operations on public roads in the real world environment. Given that information, communications systems and networks (i.e.

ICT) and Machine Intelligence (MI) are key enablers to the future ITS—mainly to the autonomous vehicles—it is necessary to evaluate how ICT, MI and other technologies and concepts will impact on (and contribute to) the autonomous transport safety risks and, consequently, how to apply them in an innovative, cost-effective and safe way.

U.S. National Highway Transportation Safety Administration (NHTSA) released a guidance to orient developers of Automated Driving Systems (ADS) – AV included—to "*analyze, identify, and resolve safety considerations prior to deployment using their own, industry, and other best practices*" (NHTSA 2017). It recommends the combination of simulations, test track, and on-road testing to validate the ADS safety performance. However, it is observed that the current AV safety evaluation and validation are being supported by track and on-road testing. Virtual testing approaches are a recent, in-development field interesting to manufacturers (Kim et al. 2017).

Tests are useful to improve the confidence on the system. However, it is impossible to cover all the possible system's conditions (behaviors) using tests. Kalra and Paddock (2016) show that '*fully autonomous vehicles would have to be driven hundreds of millions of miles and sometimes hundreds of billions of miles to demonstrate their reliability in terms of fatalities and injuries*', which is unfeasible in terms of costs and other available resources. In this way, they recommended that innovative, alternative methods of demonstrating safety and reliability need to be developed, as the "*virtual testing and simulation*", to supplement real-world testing in order to assess autonomous vehicle safety.

Thus, there is a need to develop new virtual testing and simulation approaches, mainly using accelerated time scales, in order to analyze the safety levels of AVs on a Roadway Transport System (RTS), as well as the safety levels of the RTS as a whole. This paper proposes a simulation-based safety analysis Framework to be applied to the RTS in this future ITS paradigm. Based on a combined fast-time and real-time computer simulation approaches, this Framework allows analyzing the impacts of concepts, technologies and procedures on the future RTS safety and efficiency, in a broad sense.

In this proposed Framework, specific RTS in predefined traffic scenarios (or 'Use Case') are modeled, simulated and analyzed. A Use Case represents a complete road traffic scenario under analysis, including the RTS architecture and specification, traffic configuration (vehicles trajectories and other traffic elements behavior) and the procedures to be applied during the analysis (simulations). By means of a Use Case, it is possible to analyze the influences of any RTS element inserted in a specific traffic configuration over the system safety metrics (e.g. collision rates) and over the way the hazardous situations evolve (system behavior). Mostly, it is possible to analyze how the AVs impact the safety and efficiency of the whole transport system they belong to, and identifying hazardous situations that emerge from the behavior of the AV on transport systems.

In addition to proposing and presenting the safety analysis Framework, this paper provides a Proof-of-Concept (PoC), where a representative Use Case is modeled, simulated and analyzed. The RTS elements interact with each other, and autonomous and 'non-autonomous' vehicles are sharing the same RTS infrastructure and running in the predefined traffic scenarios. Further, the validation of the Framework is presented using an AI-based autonomous control approach (Naufal et al. 2017) that manages risks related to the operations of AVs in a simulated RTS scenario.

This paper is structured in 5 sections. Section 2 presents details about the development of the proposed simulation-based safety analysis Framework. Section 3 provides a Framework PoC, in which a representative RTS scenario (Use Case) is implemented (modeled) and simulated by real-time and fast-time Open Source Simulation Tools. Section 4 presents the results obtained from the PoC and used to validate the proposed Framework. Finally, Section 5 presents conclusions about this work and some future directions.

2 THE SAFETY ANALYSIS FRAMEWORK FOR THE ROADWAY TRANSPORT SYSTEMS

Our proposed simulation-based safety analysis Framework is intended to be applied to the future, intelligent RTS paradigm. It is based on an Automotive (road and vehicle) Cyber Physical Systems (CPS) safety and resilience theoretical conceptual bases in the RTS domain. The Automotive CPS (ACPS) concept is a fusion among RTS, CPS and ITS concepts, as detailed in section II and III of Naufal et al (2017). These theoretical conceptual bases were used to understand the future, intelligent road transport system context in which AV are going to be inserted. Consequently, they were used to identify the elements and characteristics demanded to execute a simulation-based safety analysis of a RTS.

The developed Framework is composed of 5-modules as illustrated in Figure 1:

1. the "ACPS Elements module", a repository (class library) containing the classes of ACPS-based elements used to model an ACPS-based Roadway Transport System (RTS);
2. Framework users design a specific Use Case in the "Use Case elaboration" module using classes and attributes available in ACPS Elements module. ACPS-based RTS architecture, specification

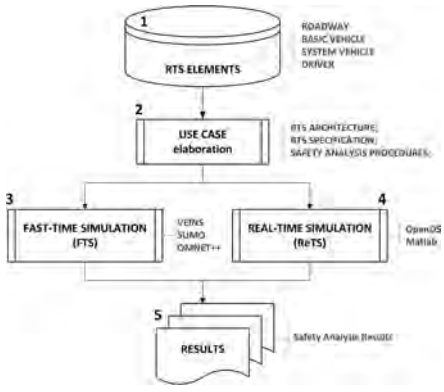


Figure 1. The simulation-based safety analysis framework.

and safety analysis procedures (safety metrics and parameters) are then submitted to the simulation-based safety analysis approach;

3. "Fast-Time Simulation (FTS) module", where the Use Case is instantiated and simulated—following the safety analysis procedures—using the open source fast-time computer-based simulation tools available in this Framework;
4. "Real-Time Simulation (ReTS) module", where the Use Case is instantiated and simulated—following the safety analysis procedures—using the open source real-time computer-based simulation tools available in this Framework;
5. The simulation reports with the simulation results are analyzed in "RESULTS module".

2.1 *ACPS elements*

Based on these ACPS conceptual bases, four classes of elements were identified: Roadway (physical place where vehicles run), Vehicle (the central traffic element, moving people and goods), Driver (manages the Vehicle movement according to the traffic rules and vehicle model) and Environment (weather conditions—e.g. rain, fog, snow—and obstacles—e.g. pedestrian, animals, roadway blockages) other than Vehicles). Table 1 presents these ACPS elements (classes) and their characteristics (attributes). These elements are both transport system safety-related (physical) elements and represent a real RTS in the ITS context (US-DoT 2017).

Regarding to the ACPS Elements, the 'DRIVER' class has no defined characteristic (attributes) because the DRIVER, in this Framework, only execute actions (methods). ROADWAY and Environment were included in the same class of elements (ROADWAY). Figure 2 illustrates the interactions among ACPS elements used to model the Use Cases in the proposed Framework. Different from a typical RTS in which Drivers get information directly from Environment and Roadway (observing them), DRIVERs monitor the Environment and Roadway using the sensors and other technologies embedded in the VEHICLE that they are driving. This enables modeling the future highly automated and fully-autonomous vehicles, given that they will take driving decisions using data obtained by themselves from the Environment.

In the proposed Framework, these ACPS elements embody a repository (a class library) containing classes of ACPS-based elements (elements, attributes, methods and interfaces). These elements represent an ACPS-based RTS and, consequently, this repository is used by Framework users to model any specific ACPS-based road traffic scenario (Use Case) and to develop a simulation-based safety analysis.

Table 1. RTS main elements and characteristics.

ACPS Elements (*Classes*)	Characteristics (*Attributes*)
ROADWAY (Environment included)	Interdicted area
	Direction of the road
	Obstacles on the road
	Physical characteristics of the road
	Signs
	Weather conditions
VEHICLE	Turn right command
	Turn left command
	Brake command
	Speed up command
	Distance from detected Vehicles/ Obstacles on the Road (DVOTR)
	Vehicle size
	Maximum braking rate
	Maximum vehicle speed
	Maximum acceleration rate
	Mechanical vehicle system status
	Steering wheel status
	Accelerator pedal status Brake pedal status
	Vehicle Speed status
	Direction of the vehicle
DRIVER	

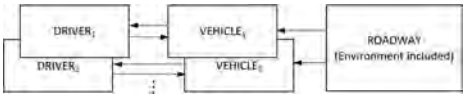


Figure 2. Relationship among ACPS elements.

2.2 *Use case elaboration*

During the Use Case elaboration, the RTS architecture, specification and safety analysis procedures

(safety metrics and parameters) are specified and modeled. The Use Case (ACPS-based RTS model and the analysis procedures) is instantiated and simulated using real-time and fast-time approaches by specific, open source simulation tools. Results (Reports) are obtained from these tools (safety analysis) and used to analyze the impacts of concepts, technologies and procedures on RTS safety and efficiency.

The Use Case shall contain the ACPS-based RTS architecture and specification (i.e. configuration of the ACPS elements characteristics—Roadway, Vehicle and Driver—which represent the specific use case scenario), as well as the safety analysis procedures—the safety metrics and parameters that will be analyzed in this scenario. For the safety analysis efficiency purposes, the Use Case should represent a road traffic scenario that maximizes the occurrence of conflicts points, hazardous situations and accidents. Besides, it is possible to analyze the influences of ACPS elements and their characteristics (Use Case Specification) over system safety metrics (e.g. collision rates) and hazardous situations evolution (behavior).

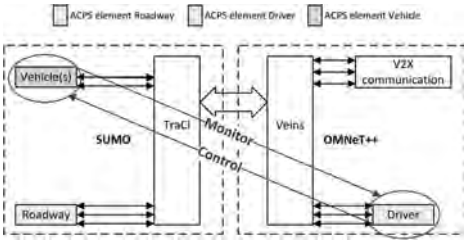


Figure 3. FTS high-level architecture.

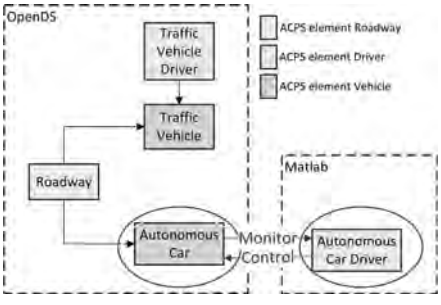


Figure 4. ReTS high-level architecture.

2.3 *Fast-Time and Real-Time simulation modules*

The Use Case is implemented (instantiated), simulated and analyzed using Fast-Time and Real-Time computer-based Simulations approaches in Framework modules 3 (FTS) and 4 (ReTS) using two different sets of open source tools. In the FTS (Framework module 3), *Veins* is used. It is an event-based network simulator that allows simulating vehicular communication, and integrates with *SUMO* (*Simulation of Urban MObility*) – a road traffic simulator—and *OMNeT++* (*Objective Modular Network Testbed in C++*). In the ReTS (Framework module 4), it is used the OpenDS(TM) – a driving simulator primarily intended for research and driving training, and Matlab(TM).

Molina et al. (2017) assessed these two open-source simulation tools sets suitability with the modeling and simulation capabilities demanded by the analysis Framework, and evaluated if they are capable to model and simulate, using fast-time and real-time simulation approaches, the Road Transport Systems (indeed, ACPS) elements characteristics required by the Framework. Besides, the tool's best features were highlighted and the process of adapting these tools for the safety analysis purposes was presented.

Figure 3 and Figure 4 illustrate the high-level architecture of FTS and ReTS modules, respectively. Roadway, Vehicles and Driver are represented in both FTS and ReTS. A functional Vehicle (AV) is modeled combining a pair "Driver + Vehicle" elements, the vehicle control algorithms are embedded in the Driver element, which manages the Vehicle's movement. In ReTS, Vehicles are represented by 'Traffic Vehicles' and 'Autonomous Cars'.

Using FTS approach, it is possible running several simulations over a Use Case scenario, allowing the estimation (statistical) of risk metrics, as well as identifying worst cases scenarios. On the other hand, using ReTS approach, system behavior during specific scenario conditions (including failure modes), especially to the worst cases scenarios can be analyzed in details, as identified by the fast-time simulation.

2.4 *Results*

Finally, '*RESULTS*' (module 5) represents the safety analysis reports generated by the simulation tools. These reports contain the safety and efficiency metrics, as well as any other relevant information, obtained by simulating the Use Case scenarios according to the Safety Analysis Procedures. These reports may contain the status of the system elements when a collision (or any safety-related event) occurs, as well as safety risk metrics and other parameters specified by the analyst in the Use Case. Therefore, using information provided in these reports, it is possible to systematically identify the potential causes of hazards or failure modes and identify how failures or defects in the system elements might contribute to hazards or accidents (e.g. collisions) – i.e. performing a Safety Analysis (MoD 2017).

3 PROOF-OF-CONCEPT (PoC) FRAMEWORK IMPLEMENTATION

This section presents a Proof-of-Concept (PoC) to the proposed Framework. The PoC is oriented by a Use Case composed by a representative number of ACPS Elements to effectively test, and validate the Framework. Besides, the Use Case presents a road traffic scenario that maximizes the occurrence of hazardous situations and accidents ('Vehicle to Anything Else' – V2X collisions). Thus, it is possible to analyze both the influences of ACPS elements (autonomous vehicles included) over RTS safety and efficiency metrics (e.g. collision rates and traffic flow and capacity) and the hazardous situations evolution (system behavior).

3.1 *AI-based Autonomous Vehicle (AV)*

The proposed Framework is used to analyze the capabilities of an AI-based autonomous vehicle (AV) control approach in managing safety risks related to the operation of an autonomous vehicle (AV) in a simulated ACPS-based RTS scenario. The analyzed AI-based AV control approach was proposed by Naufal et al (2017) and named as "A^{2}CPS Engine". It is a run-time, software-based, vehicle-centric risk management approach, which its algorithm was developed using Fuzzy Logic (an AI technique). A^{2}CPS Engine algorithms monitor the environment and the AV behavior and control AV actions, mitigating hazards and reducing accidents severity related to the AV operation. Table 2 presents the risk management rules performed by the analyzed AV Control algorithms. Depending on the calculated Risk Level and its Risk Criteria, a specific Treatment (action to be executed by the AV) is performed.

The AV Control algorithms were embedded in the autonomous vehicle DRIVER, monitoring and controlling the autonomous VEHICLE. Figure 5 illustrates the AI-based AV Control algorithms (based on A^{2}CPS Engine proposal) embedded in the autonomous DRIVER and the I/O interface with Autonomous VEHICLE. In this implementation, the AV Control algorithms receive its Speed from

Table 2. Risk management rules performed by the AV control.

Risk level	Risk criteria	Risk treatment	Action (brake intensity)
No risk	Negligible	Do noting	No brake
Low	Undesirable	Speed reduction	Small
Moderate			Medium
High	Intolerable	Stop vehicle	High
Very high			Very high

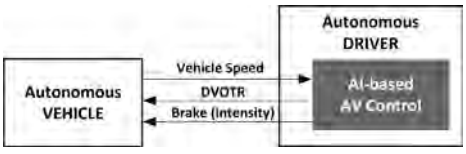


Figure 5. AI-based AV control inputs/output interface.

Autonomous Vehicle and the Distance from detected Vehicles and Obstacles on the Road (DVOTR), and it controls the brake intensity application.

The proposed safety analysis Framework is used for analyzing both (i). the autonomous vehicle behavior in the road traffic scenarios and (ii). the impact of AI-based autonomous control algorithms (autonomous vehicle behavior) over the RTS safety and efficiency levels. It is worth noting that this analysis has considered both *safety* and *efficiency*, because they are competing system characteristics in transportation systems. Traffic efficiency (e.g. traffic flow and capacity) is often impaired when system safety is prioritized. So, a tradeoff between safety and efficiency (availability included) is required. Consequently, a Use Case is designed to analyze:

1. Both *safety metrics* (#collisions) and *traffic efficiency metrics* (average speed and standard-deviation (E[v]; VAR[v]); average total distance traveled). These metrics were obtained in function of *maximum vehicle speed* (V_{MAX}) and *Incident Zone Warning* (IZW) configuration—a main characteristic of the AI-based AV Control approach (A^{2}CPS Engine), representing a dynamic zone of protection around the vehicle (Naufal et al. 2017). By these metrics, it could be possible to analyze the influence of AV Control over the autonomous vehicle behavior and how it affects the global, transport system safety versus efficiency relationship (i.e., the analysis objective 'ii.'). Given the statistical characteristic of metrics, they were obtained by the Fast-Time Simulation (FTS) approach (Framework module 3);
2. The autonomous vehicle Stopping Distance (SD) – distance the autonomous vehicle stops from a static obstacle. SD was obtained in function of the maximum vehicle speed (V_{MAX}) and the IZW configuration. By this metric, it could be possible to analyze how the AV Control algorithms affects the autonomous vehicle braking behavior (i.e., the analysis objective 'i.'). This analysis was performed by the Real-Time Simulation (ReTS) approach (Framework module 4).

3.2 *Use case details*

The ACPS-based RTS high-level architecture considered in the Use Case is illustrated in Figure 6.

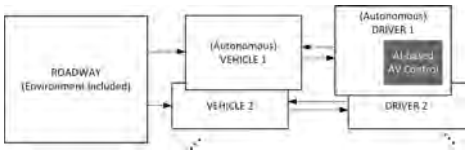


Figure 6. RTS high-level architecture USE CASE.

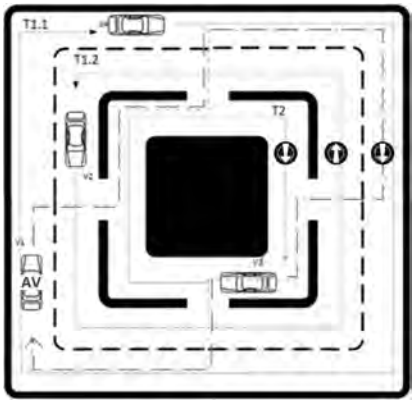


Figure 7. Use case road traffic scenario (FTS).

The AI-based AV Control (A^2CPS Engine) algorithms are embedded in the Autonomous Driver, managing the Autonomous Vehicle. The pair "Driver + Vehicle" models the Autonomous Vehicle (VEHICLE 1), which runs along the Roadway together with other Vehicles (2, 3, …).

The RTS traffic scenario considered in the FTS approach (Framework module 3) is illustrated in Figure 7. Four (4) Vehicles (V1 to V4) are travelling, in a closed traffic circuit, with pre-defined trajectories and speed profiles through a Roadway composed of 2 roads (T1 and T2) – a two-way road (T1.1 and T1.2) and a one-way road (T2). One of the four vehicles (V1) is an autonomous vehicle, and its (autonomous) Driver has the embedded AI-based AV control approach under analysis. The Roadway architecture, including each road direction, is represented by arrows. V1 to V4 trajectories are shown, where V1 (the AV) runs through all the roads (T1.1, T1.2 and T2) following the signed trajectory; V2 runs only on T1.2; V3 runs only on T2; and V4 runs only on T1.1. In this road traffic scenario, V1 trajectory exposes itself and the other vehicles (V2-V4) to the occurrence of harmful events (collisions), most of than on the crossing points. Consequently, the occurrence of hazardous situations and accidents are maximized, improving the safety analysis efficiency.

The RTS traffic scenario considered in the ReTS approach (Framework module 4) is illustrated in Figure 8. The autonomous vehicle running in a straight trajectory and with its maximum speed

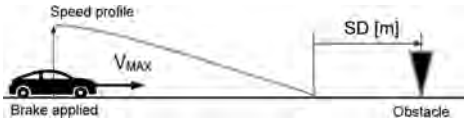


Figure 8. Use case road traffic scenario (ReTS).

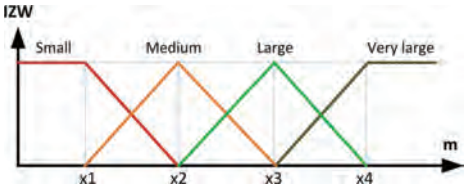


Figure 9. IZW fuzzy logic variable (x1–x4 parameters).

(V_{MAX}). In the end of this trajectory is placed a static obstacle. During the simulation it is measured the time to stop, speed and the distance from autonomous vehicle to the obstacle when vehicle stops (the stopping distance—SD) to different IZW's configurations and V_{MAX}. Using these values, it is possible to analyze if the autonomous vehicle keeps a safe and efficient distance from other vehicles.

3.3 *RTS and FTS Modules*

The Use Case traffic scenarios, including the RTS architecture and specification, and the AI-base AV control algorithms are modeled (instantiated), simulated and analyzed in the FTS and ReTS modules of Framework using its open-source tools sets (Figure 3 and Figure 4, respectively). The Use Case attributes configured to the RTS elements were: VEHICLEs = {max. acceleration = 3,3 m/s^2; maximum braking rate = 7,6 m/s^2; maximum speed = 120 km/h}; ROADWAY = {track geometry = uniform—without spines}. Drivers managed both longitudinal and lateral movement of their vehicles. Vehicles 2, 3 and 4 (non-autonomous vehicles) were driven by the native control embedded in the FTS simulation tool. In FTS, the total simulation time was 10.000 s.

In the FTS approach, it was considered 3 maximum speed (V_{MAX}) to the autonomous vehicle (V1): V_{MAX} = {15; 20; 30} m/s = {54; 72; 108} km/h; Vehicles 2, 3 and 4 run in constant speed of 5 m/s (18km/h). In the ReTS approach, it was considered 3 maximum speed (V_{MAX}) to the autonomous vehicle (V1): V_{MAX} = {30; 60; 120} km/h. Both simulation approaches (FTS and ReTS) considered the same 4 configurations to the IZW fuzzy logic variable parameters (Figure 9): {{20; 60; 95; 180}; {10; 30; 47,5; 90}; {5; 15; 23,75; 45}; {20; 60; 150; 180}}.). IZW_1 version ({20; 60; 95; 180}) was defined during the model calibration and it was the configuration-base to the other tests. IZW_2 version ({20; 60;

150; 180}) has changed only one of the parameters and used to analyze the impact of IZW shape on safety and efficiency. IZW_3 ({10; 30; 47.5; 90}) and IZW_4 versions ({5; 15; 23.75; 45}) configurations were based on the IZW_1, which was divided by 2 and 4, respectively. Parameter 'x4' physically represents the size of protection zone (45 m, 90 m and 180 m).

4 RESULTS AND DISCUSSION

The results obtained by the FTS approach are presented in Table 3. For all four IZW configurations (IZW_1 to IZW_4), three (3) autonomous vehicle maximum speeds (V_{MAX}) were considered. The number of collisions, average speed, standard-deviation and total distance travelled by AV (during the simulation length of 10,000 s) were obtained by fast-time simulation.

The results obtained by the ReTS approach are presented in Table 4. Just as it was adopted for the FTS approach, three (3) autonomous vehicle maximum speeds (V_{MAX}) were considered to each one of the 4 IZW configurations. The Stopping Distance (SD) was obtained by real-time simulation.

Table 3. FTS results (Metrics × V_{MAX} × IZW).

	V_{MAX} (AV) [m/s]		
Metrics	15	20	30
IZW_1 = {20; 60; 95; 180}			
#collisions	28	28	28
Average speed [m/s]	5.0	5.0	5.0
Standard deviation	3.1	3.1	3.1
Total distance [m]	50,050	50,049	50,049
IZW_2 = {20; 60; 150; 180}			
#collisions	30	30	30
Average speed [m/s]	5.0	5.0	5.0
Standard deviation	2.5	2.5	2.5
Total distance [m]	50,072	50,072	50,072
IZW_3 = {10; 30; 47,5; 90}			
#collisions	24	24	24
Average speed [m/s]	5.0	5.0	5.0
Standard deviation	2.3	2.3	2.3
Total distance [m]	50,115	50,115	50,115
IZW_4 = {5; 15; 23,75; 45}			
#collisions	184	195	315
Average speed [m/s]	10.0	11.5	15.9
Standard deviation	4.4	6.1	8.4
Total distance [m]	99,855	114,398	142,327

Table 4. ReTS results (SD × V_{MAX} × IZW).

V_{MAX} (AV) [km/h]	Stopping Distance (SD, in [m])			
	IZW_1	IZW_2	IZW_3	IZW_4
30	10.00	10.10	1,9	collide
60	7.47	9.92	collide	collide
120	6.67	8.80	collide	collide

Safety is measured by the number of collisions observed during 10.000 seconds (about 3h) of traffic execution. Efficiency is measured by means of average speed and standard-deviation, and total distance traveled during the simulation. The autonomous vehicle braking behavior is analyzed by means of the Stopping Distance (SD) metric (Figure 8). It is observed how these metrics change when V_{MAX} and IZW configuration are modified.

Considering the ReTS results (Table 4), it is possible to analyze the capability of four AI-based AV Control (A^2CPS Engine) algorithms versions (IZW_1 to IZW_4) in identifying a potential collision condition, assessing the related level of risk during runtime execution and stopping the autonomous vehicle *safely* (before a collision) and *efficiently* (as close as possible to the obstacle). The IZW_4 version could not stop the AV safely in any speed. Consequently, the AV has not a safe behavior using this A^2CPS Engine version. This is confirmed by the FTS results (Table 3), which IZW_4 has the largest number of collisions obtained among all IZW versions. Therefore, it is possible concluding that a smallest tested IZW (zone protection shorter than 45 m) cannot protect the AV against collision.

IZW_3 version produces an AV zone protection twice larger than IZW_4. But, IZW_3 version protects the AV against collisions only at low speeds (30km/h) with a high efficiency (AI-based AV Control algorithms stop the vehicle near to the obstacle—about 2 m). Now, considering IZW_1 and IZW_2 – with a zone protection twice larger than IZW_3 and four times larger than IZW_4 – they can protect the AV against collisions at any tested speed (from 30km/h to 120 km/h). This observation is confirmed by the FTS results (Table 3), which the speed has no influence in the traffic safety (the same number of collisions was observed in IZW_1 and IZW_2 at any speeds). The difference between these IZW_1 and IZW_2 versions is observed in the braking efficiency, where SD was between 6,67 m and 10 m (five times less efficient than the last analyzed version at 30 km/h which), and the worst braking efficiency is observed in the IZW_2 version.

The internal shape of linguistic variable IZW (Figure 9) has influence on SD (braking) efficiency, but it has neither influence on traffic safety (the number of collisions is the same in both versions) nor on traffic efficiency (same average speed and total distance).

Finally, it is possible to observe that best AI-based AV Control (A^2CPS Engine) algorithms version in terms of traffic safety (lower number of collisions) and efficiency is the IZW_3. However, regarding to the ReTS results (Table 4), this

version protects the AV against collisions only at low speeds (30 km/h). Observing the average speed and standard deviation, AV has travelled in this scenario only at low speeds. Therefore, this version is safe and efficient in terms of traffic flow/productivity.

5 CONCLUDING REMARKS

This paper proposed a simulation-based safety analysis Framework to be applied to the future, intelligent Roadway Transport Systems (RTS) paradigm. Based on a combined Fast-Time Simulation (FTS) and real-time simulation (ReTS) approaches—implemented by the open-source tools OpenDS, Matlab, Sumo, Veins and OMNet++, this Framework allows analyzing the impacts of concepts, technologies and procedures on RTS safety and efficiency. For demonstration and validation purposes a Framework Proof-of-Concept (PoC) was provided. It was applied to analyze the capabilities of a AI-based autonomous vehicle control approach—the A^{3}CPS Engine proposed by Naufal et al (2017) – in managing safety risks related to the operation of an autonomous vehicle in a simulated RTS scenario.

PoC was useful to demonstrate the Framework capability of modeling and simulating real-world, representative ACPS-based Road Transport System scenarios, in which Vehicles, Roadway (environment included) and Driver elements are interacting with each other. It enables modeling roadway traffic scenarios where both autonomous (unmanned) and manned vehicles share the same roadway infrastructure. The autonomous vehicles are modeled combining a 2-tuple of ACPS elements "Driver + Vehicle". The autonomous vehicle control algorithms, even those algorithms developed by a third-party (not by Framework user), are embedded in the autonomous Driver element and they manage the (autonomous) Vehicle movement. Using these approaches, it is possible to analyze the behavior of the autonomous vehicle's behavior in traffic scenarios and evaluate the impact of autonomous vehicles on the of RTS safety properties.

Representing an Autonomous Vehicle by the combination of a 'off-the-shelf', general purpose Vehicle element whose behavior is managed by an autonomous control algorithm embedded in the Driver element is adherent to the state-of-the-art adopted by the autonomous vehicles developers. Today, AV developers are using current manned vehicles from the consolidated automotive industry worldwide and embedding in them systems, software and machine intelligence. Therefore, this Framework can be used efficiently to evaluate the Autonomous 'Driver', the part that makes a vehicle be 'autonomous', as well as its impacts on system safety when operating in a simulated, but complete, RTS scenario.

Using the proposed Framework, it is possible observe the impact of a new concept/technology (component level) over several system parameters (in this case, traffic safety and efficiency) and, at the same time, understand and analyze the relationship among several system properties (some of them intrinsic to systems elements, others intrinsic to the system). 'Safety' and efficiency are competing metrics in transport systems, where the traffic efficiency (e.g., traffic capacity or flow rate) is often impaired when system safety is prioritized. Therefore, a trade off between safety and efficiency (availability included) is required to these safety-critical systems.

Concluding, this work has advanced on addressing the question '*how to assure that the future safety-critical autonomous transport systems are going to be safe enough during their operation?*'. Relevant ACPS-based RTS elements and characteristics to be considered in a safety analysis were identified. As the following works, authors are evolving this Framework, implementing both environment/system changes during runtime simulation and collaborative communication among system elements (V2X) capabilities. This Framework evolution will allow resilience analysis over the ACPS-based Road Transport Systems, evolving on the development of new approaches to safety assurance.

ACKNOWLEDGMENTS

This work is supported by the Research, Development and Innovation Center, Ericsson Telecomunicações S.A., Brazil.

REFERENCES

Kalra, N. & Paddock, S.M., 2016. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? Transportation Research Part A: Policy and Practice, 94, pp.182–193. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0965856416302129.

Kim, B. et al., 2017. Testing Autonomous Vehicle Software in the Virtual Prototyping Environment. IEEE Embedded Systems Letters, 9(1), pp.5–8. Available at: http://ieeexplore.ieee.org/document/7797233/.

MoD, 2017. Safety Management Requirements for Defence Systems—Requirements, UK.

Molina, C.B.S.T. et al., 2017. A comparison of two simulators to support safety analysis in autonomous vehicles. In European Safety and Reliability Conference, 2017. Portoroz, Eslovênia: Taylor & Francis Group, pp. 1903–1911.

Naufal, J.K. et al., 2017. A2CPS: A Vehicle-Centric Safety Conceptual Framework for Autonomous Transport Systems. IEEE Transactions on Intelligent Transportation Systems, pp.1–15. Available at: http://ieeexplore.ieee.org/document/8054749/.

NHTSA, 2017. AUTOMATED DRIVING SYSTEMS 2.0: A VISION FOR SAFETY, Washington DC. Available at: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf [Accessed December 7, 2017].

PARLIAMENT, E. & UNION, C.O.E., 2010. On the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport (DIRECTIVE 2010/40/EU), EU: Official Journal of the European Union. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:207:0001:0013:EN:PDF.

Singh, S., 2015. Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey, Washington, DC. Available at: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115.

US-DoT, 2017. Architecture Reference for Cooperative and Intelligent Transportation (ARC-IT). Available at: http://local.iteris.com/arc-it/html/viewpoints/physical.html [Accessed November 8, 2017].

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Robust management of distributed energy resources for frequency control in microgrids with unreliable communication

H.D. Mo & G. Sansavini
Reliability and Risk Engineering Laboratory, Department of Mechanical and Process Engineering, Institute of Energy Technology, ETH Zurich, Switzerland

ABSTRACT: Effective coordination of Distributed Energy Resources (DERs) in power systems via control strategies mitigates the frequency fluctuations stemming from stochastic renewables and uncertain demand. Recently, open communication networks are used to deploy Load Frequency Control (LFC) strategies to overcome the lack of dedicated communication infrastructures and the ubiquity of DERs system. However, open networks are exposed to communication degradation and can reduce the LFC performance. This work investigates the real-time performance and reliability of the integrated DER system and open communication networks, i.e. the cyber-physical microgrid system, with reference to LFC against communication degradation. In particular, LFC is provided by a discrete PID controller tuned via particle swarm optimization. The cyber-physical microgrid system is implemented on a real-time platform simulating various MAC protocols and open-communication-network architectures, developed in the Truetime simulator. The impact of communication degradation on LFC performance is assessed. Simulation results demonstrate that transmission delays and packet dropouts jeopardize the ability of cyber-physical microgrid systems to maintain system frequency deviations within tolerance bounds. In particular, the use of Ethernet ensures higher reliability as compared to 802.11 b/g. Moreover, the impact of interfering traffic and of the percentage of used bandwidth on the LFC performance reduction is evaluated. The optimized PID controller is able to compensate for communication degradation and uncertainty of the microgrid, and ensures robust LFC against unknown network configurations.

1 INTRODUCTION

The power sector is experiencing a structural trend towards decentralization stemming from the integration of large shares of Renewable Energy Resources (RERs) (Driesen and Katiraei, 2015). This is fostered by Distributed Energy Resources (DERs), which require the integration of power generation means located at or near the end-user side (Kumar et al., 2017). However, the stochastic nature of RERs and of the load demand induces system frequency fluctuations (Shotorbani et al., 2012). An effective control strategy is needed to keep the system frequency to its nominal value by balancing power generation and demand in real time. To this aim, Automatic Generation Control (AGC) schemes are developed for damping frequency oscillations in Distributed Generation Systems (DGS) (Shotorbani et al., 2012; Lee and Wang, 2008).

Recently, the AGC has been integrated with the open communication network, due to low cost, high speed, simple structure and flexible access. Data exchanges among PMUs, generators and control center are provided by the open communication network in the form of time stamped data packets (Kuzlu et al., 2014). Stable AGC depends heavily on the performance of the open communication network (Ahmadi and Aldeen, 2017).

However, open communication networks are exposed to various types of degradation processes, i.e. network-induced time delays (Lee and Wang, 2008) and packet dropouts (Mo et al., 2016). As a result, the measurement signals (control signals) received by the control center (ESSs or generators) degrade, effective AGC cannot be carried out and the system frequency response worsens (Pan and Das, 2016). Studying the performance of open communication networks is critical for understanding the occurrence of time delays and packet dropouts.

Time delays are variable, challenging to predict, deteriorate the AGC performance and reduce the stability region (Pan and Das, 2016). Packet dropouts refer to lost messages, which occupy network bandwidth but cannot reach destination. They affect the operations of DERs and the reduction of frequency fluctuations, particularly in uncertain network environments. Optimal feedback AGC regulators for DERs are investigated in numerous

works for perfect communication networks and the impact of transmission delays and packet dropouts on the controller cannot be captured (Ghoshal, 2004). Robust PID controllers against constant or uniformly distributed time delays are designed to cope with perturbations of the control parameters (Pan and Das, 2016). However, constant or uniformly distributed time delays cannot be generally assumed in realistic communication networks.

In this work, we investigate (a) the operations of the integrated DER system and open communication network, and (b) the design of optimal AGC strategies in the face of communication degradation. The ability of the integrated system to maintain system frequency within tolerance margins quantifies system reliability, and is evaluated by Monte Carlo simulation (MCS). Stochastic time delays are modelled by generating random congestion based on MAC protocols. Congestion of network channels depends on the activity level of the interfering traffic, which is the root cause of network-induced delays and packet dropouts (Peng and Han, 2016). The open communication network model is implemented via Truetime simulator testing different MAC protocols (Grenier and Navet, 2004; Cervin et al., 2010). Multiple activity levels of the interfering traffic are simulated via the disturbance node, which sends random interfering packets over the network. Packet dropout is described by Bernoulli-distributed variables (Minero et al., 2009). To stabilize system frequency against RER, demand variability and communication degradation, a discrete PID controller is used (Pan and Das, 2016). Particle Swarm Optimization (PSO) is adopted to minimize the stochastic objective function and achieve the optimal PID controller for various architectures and conditions of the open communication network (Ghoshal, 2004; Pan and Das, 2016). Finally, the robustness of the optimum-PID-controlled AGC against communication degradation is assessed.

The rest of the work is organized as follows. Section 2 describes the DGS model. The open network and the communication degradation models are described in Section 3. Section 4 defines the reliability of the integrated system, and introduces the PSO-based PID controller. Section 5 presents the simulation results. Section 6 concludes the work.

2 MODEL OF DGS WITH DERS

The schematic of the integrated system, which is made of a cyber layer and a physical layer, is illustrated by Fig. 1. The structure of physical layer is general, representative of DGS, and widely adopted in the literature (Pan and Das, 2016). It models a hybrid microgrid, which consists of conventional

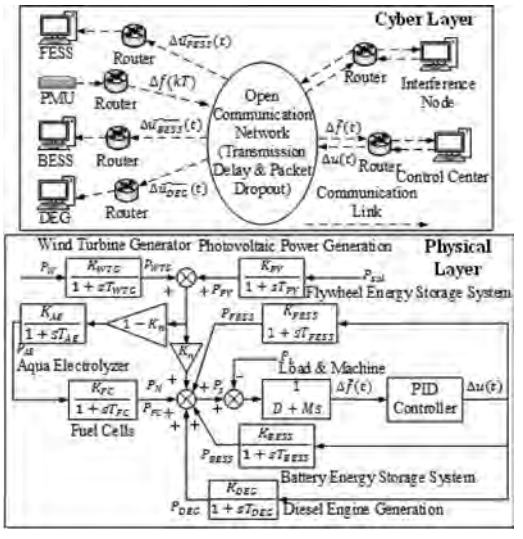


Figure 1. Schematic of the integrated system.

generators (Diesel Engine Generator, DEG), RERs (wind turbine generator, WTG, and photovoltaic generator, PV), ESSs (battery and flywheel energy storage system, BESS and FESS) and power-to-gas technology (aqua electrolyzer, AE, and fuel cell, FC). The cyber layer, i.e. the open communication network, provides data exchange between the control center and the controllable components, i.e. the ESSs and DEG, in the physical layer. PMU measurements and control signals are transmitted via the shared and open communication network. The cyber layer is detailed in Section 3. Power imbalance is the root cause of system frequency fluctuations. The control center remotely monitors the ESS and the diesel generator to reduce the imbalance and to ensure good AGC performance.

The small signal stability analysis of the hybrid microgrid in Figure 1, is based on time-domain simulations, i.e. transfer function models. In the AGC, the WTG, PV, AE, FC, DEG, FESS and BESS are described by first order transfer functions with specified gain and time constant. A centralized controller is used by Pan and Das (2016), as opposed to multiple decentralized controllers for each controllable component (Ray et al., 2011). It enables easier maintenance and reduces wiring cost, and makes the AGC design problem traceable by reducing the number of controller parameters. On the other hand, the centralized controller impacts the AGC performance, because a unique control signal is used by all the components. Nevertheless, current studies show that the centralized controller can ensure acceptable real-time AGC performance (Pan and Das, 2016). The transfer

functions $G_{WTG}(s), G_{PV}(s), G_{FC}(s)$ and $G_{DEG}(s)$ of the WTG, PV, AE and DEG, respectively, are expressed as

$$G_{WTG}(s)=\frac{K_{WTG}}{1+sT_{WTG}}=\frac{P_{WTG}}{P_W} \tag{1}$$

$$G_{PV}(s)=\frac{K_{PV}}{1+sT_{PV}}=\frac{P_{PV}}{P_{sol}} \tag{2}$$

$$G_{FC}(s)=\frac{K_{FC}}{1+sT_{FC}}=\frac{P_{FC}}{P_{AE}} \tag{3}$$

$$G_{DEG}(s)=\frac{K_{DEG}}{1+sT_{DEG}}=\frac{P_{DEG}}{\widetilde{u_{DEG}}(t)} \tag{4}$$

where K_{WTG}, K_{PV}, K_{FC} and K_{DEG} are the gain, and T_{WTG}, T_{PV}, T_{FC} and T_{DEG} are the time constant of the WTG, PV, AE and DEG. P_{WTG} and P_{PV} are the electrical power produced from the RERs, i.e. wind power P_W and solar power P_{sol}. P_{AE} and P_{FC} are the output of the AE and the power produced of the FC. The DEG is controlled by the control signal $\widetilde{u_{DEG}}(t)$ sent by the remote control center and it generates power only when the RERs cannot meet the demand.

In the DGS, the AE is used to absorb the rapidly fluctuating output power from the WTG and PV by producing hydrogen (Das et al., 2012). The hydrogen is stored and used as fuel in the FC to feed power to the grid. The dynamic property of AE is described by the transfer function $G_{AE}(s)$

$$G_{AE}(s)=\frac{K_{AE}}{1+sT_{AE}}=\frac{P_{AE}}{(P_{WTG}+P_{PV})(1-K_n)} \tag{5}$$

where K_{AE} and T_{AE} are the gain and time constant of the AE. $1-K_n$ denotes the fraction of power generated by the WTG and PV to produce hydrogen in the AE. K_n is equal to $P_N/(P_{WTG}+P_{PV})$ and is set to 0.6 (Pan and Das, 2016).

The ESSs are critical in eliminating frequency fluctuations due to their fast response to the control signal. Based on Lee and Wang (2008), the transfer functions $G_{FESS}(s)$ and $G_{BESS}(s)$ of the FESS and BESS are given as

$$G_{FESS}(s)=\frac{K_{FESS}}{1+sT_{FESS}}=\frac{P_{FESS}}{\widetilde{u_{FESS}}(t)} \tag{6}$$

$$G_{BESS}(s)=\frac{K_{BESS}}{1+sT_{BESS}}=\frac{P_{BESS}}{\widetilde{u_{BESS}}(t)} \tag{7}$$

where K_{FESS} and K_{BESS} are the gain, T_{FESS} and T_{BESS} are the time constant, P_{FESS} and P_{BESS} are the output power, $\widetilde{u_{FESS}}(t)$ and $\widetilde{u_{BESS}}(t)$ are the control signal of the FESS and BESS.

Remark 1: The DEG, FESS and BESS have rate constraint, i.e. $|P_{DEG}|<\overline{P}_{DEG}$, $|P_{FESS}|<\overline{P}_{FESS}$ and $|P_{BESS}|<\overline{P}_{BESS}$, where $\overline{P}_{DEG}$, $\overline{P}_{FESS}$ and $\overline{P}_{BESS}$ are the maximum rated output power of the DEG, FESS and BESS, respectively.

The transfer function $G_{HPS}(s)$ of the hybrid power system (HPS) models the relationship between the power imbalance, i.e. $\Delta P_S-\Delta P_L$, and the system frequency $\Delta f(t)$

$$G_{HPS}(s)=\frac{1}{D+Ms}=\frac{\Delta f(t)}{\Delta P_S-\Delta P_L} \tag{8}$$

where M and D are the inertia constant and dampi constant of the HPS (Pan and Das, 2016), P_S is the total power generated, denoted by $K_n\cdot(P_{WTG}+P_{PV})+P_{FC}+P_{DEG}+P_{FESS}+P_{BESS}$, and P_L is the power demand. The models for P_{WTG}, P_{PV} and P_L are detailed in (Lee and Wang, 2008).

3 THE OPEN COMMUNICATION NETWORK

The model of the data transmission across the open communication network accounts for the composition of network-induced delays and for the stochastic packet dropout. We assess two general architectures for the communication network used in power systems, i.e. the Ethernet (Grenier and Navet, 2004) and a mix of Ethernet and WLAN (henceforth called hybrid network) (Pan and Das, 2016), and implement them via Truetime simulator (Cervin et al., 2010).

3.1 *Model of time delay and packet dropout*

The time delay in the communication channel consists of four components, i.e. the preprocessing time T_{pre}, the waiting time T_{wait}, the time for traveling across the channel T_{tx} and the post-processing time T_{post}. (Ghoshal, 2004). Therefore, the total time delay T_d can be expressed as

$$T_d=T_{pre}+T_{wait}+T_{tx}+T_{post} \tag{9}$$

where T_{pre} and T_{post} depend on the processing speed of the device firmware. T_{tx} depends on the physical bandwidth, propagation spend and transmission distance. T_{wait} is the major source of the te delay and it is influenced by the MAC protocol.

The time delays τ_{sc} at forward channel and τ_{ca} afeedback channel are determined by (9), and their impact on the transfer function from the control signal $U(s)$ to the system output $Y(s)$ is described as

$$\frac{Y(s)}{U(s)} = G(s)\, e^{-(\tau_{sc}+\tau_{ca})s} \tag{10}$$

where $G(s)$ denotes the transfer functions of the DERs.

A second source of disturbance in the source-destination communication is packet dropout, which occurs for three major reasons, i.e. the network disconnection, time-out transmission and time-out re-transmission (Cervin et al., 2010). Binary switching sequences are usually applied, which specify the expected packet dropout probability. The stochastic parameter γ_k of the binary switching sequence is Bernoulli distributed (Liang et al., 2010), taking value of 0 or 1 with

$$\mathrm{P}\{\gamma_k = 0\} = L_d \tag{11}$$

where $0 < L_d < 1$ is the packet loss probability.

3.2 *Network architecture and implementation*

The architecture for the AGC of DERs via Ethernet and hybrid network, is illustrated by the cyber layer in Figure 1. The data exchange among the control center, DERs, interference node and PMU is wired in the Ethernet architecture. In the hybrid architecture, the data exchange between the routers and the RTU (BESS, FESS, DEG and PMU) is wireless and provided by the WLAN. The data exchange among routers is wired and provide by the Ethernet. Low product prices make WLAN more convenient and cost-effective compared to the traditional LAN. As such, WLAN has become an efficient approach to provide flexible data communication between routers and RTU, and is employed for monitoring and controlling DERs, offshore wind farms and smart home energy managemensystems (Pan and Das, 2016).

The interference node simulates the interference user in the open network, who sends disturbing traffic over the channels and cause congestion. Generally, the length of data traffic is constant (i.e. 80 bytes in this work), and the interference node sends it to the network at every time period T_i if

$$\mathrm{UN}_i < \mathrm{BWShare} \tag{12}$$

where UN_i is a uniformly distributed random number sampled at time period T_i in the interval $[0,1]$, and BWShare is the expected ratio of the network bandwidth used by the interference node.

The architectures of the real-time AGC of DERs via the Ethernet and hybrid network all consists of:

- PMU node (time-driven): the PMU takes measurements of the system frequency at every sampling interval T_s = 0.01s and sends $\Delta f(kT_s)$ to the control center over the network. The size of data packets is 80 bytes and the phase delay caused by the filter of the PMU is 0.006 s (Martin et al., 1998).
- Control center node (event-driven): when a data packet from the PMU reaches the control center, the controller takes 0.002 s to compute the control signal $u(t_n)$ and sends it to the DERs. The length of control signal is 500 bytes.
- DERs nodes (event-driven): the DEG, BESS and FESS adjust their operations based on the control signal received.
- Interference node (time-driven): it sends disturbing traffic over the network with period T_i and causes congestion, to generate different scenarios of time delay and packet dropout.

4 THE PSO-TUNED PID CONTROLLER

4.1 *Discrete-time PID controller*

Due to periodic PMU measurements and packet dropouts, the AGC process is discrete. Therefore, a discrete-time PID controller is adopted to compute the control signal

$$\begin{aligned} u(t_n) = -(&K_P \Delta\tilde{f}(t_n) \\ &+ K_I \sum_{k=0}^{n} \Delta\tilde{f}(t_k) * T_s + K_D \frac{\psi}{T_s}) \end{aligned} \tag{13}$$

where K_P, K_I and K_D are the proportional, integral and derivative gain of the controller, and $\psi = \Delta\tilde{f}(t_n) - \Delta\tilde{f}(t_{n-1})$.

For the implementation of the discrete-time PID controller, the role of a first-order pole filter on the derivative action should be considered [36]. Therefore, the transfer function of the discrete-time PID controller in Eq. (18) is

$$C(z) = -\left(K_P + K_I a(z) + K_D \frac{N_L}{1 + N_L b(z)} \right) \tag{14}$$

where N_L is the filter's coefficient indicating the location of the pole in the derivative filter and $a(z) = b(z) = T_s/(z-1)$.

4.2 *Reliability of the integrated system*

Frequency values outside the tolerance band compromise the components in the physical layer, reduce their useful lifetime and deteriorate the AGC performance. In this work, the reliability is defined as the ability of the integrated system to

maintain system frequency within a predefined tolerance band, in the face of stochastic RERs, uncertain load demand, variable time delays and random packet dropouts. The value of the reliability R measures the ability of the integrated system to provide adequate AGC performance, and is computed as

$$R = T_I / T \tag{15}$$

where T_I is the total time in which the system frequency remains within the predefined tolerance band and T is the total operating time of the AGC.

Due to the uncertainties in the DGS and the random communication degradation effects, the AGC performance evolves through stochastic trajectories. Thus, the reliability in Equation 15 and objective function in Equation 16 are stochastic and generally evaluated as the expected values of a stochastic process (Cervin et al., 2010). Thus, the reliability and t objective function are estimated via the MCS method (Pan and Das, 2016). In order to achieve a statistically acceptable estimate, the reliability is computed via 200 MCS-based samples, i.e. the integrated system is evaluated for 200 trials to compute the expected reliability (each sample simulates a time frame of 100 s and requires around 4 s on a 64 b Wdows desktop with 32 GB memory and an Iel(R) Xeon(R) E5-1650 v3 @ 3.50GHz CPU).

4.3 *Performance of the PID-controlled AGC*

The AGC performance of the integrated system depends on the discrete-time PID controller. Therefore, the PID controller is optimized to mitigate the communication network disturbances, and offer optimum AGC performance by reducing system frequency fluctuations. As a result, system reliability is also optimized. The objective function for the optimization of the PID controller is an integral performance index over the total operating time T, which quantifies frequency and control signal deviations (Pan and Das, 2016):

$$J = \int_0^T \left[\eta_1 (\Delta f)^2 + (1 - \eta_1) * \eta_2 (\Delta u)^2 \right] dt \tag{15}$$

where η_1 indicates the relative importance of the two terms and η_2 is the normalizing constant to scale both terms in a uniform range and is set to 0.002. The first term is directly related to the reliability of the integrated system and the second term measures the disturbance rejection ability of the controller, which is the total control effort to be minimized. The total process time T is set to 100 s.

Many heuristics algorithms, e.g. the genetic algorithm (Das et al., 2016) and the PSO (Pan and Das, 2016), are appropriate in such case to find the near-optimum solutions, taking into account the stochastic nature of the objective function. The fitness value of the stochastic population-based algorithms is the expectation of the stochastic objective function obtained from MCS

$$E[J] = \sum_{i=1}^{N} J_i \tag{16}$$

where J_i is defined in Equation 16 and $N = 200$ samples. We use PSO to solve the optimization problem for $\vec{x} \in \mathcal{R}^3$

$$\text{minimize } \mathrm{E}\left[J(\vec{x}) \right] \tag{17}$$

where $J(\vec{x}) : \mathcal{R}^3 \to \mathcal{R}$ is defined in Equation 16. The variable $\vec{x}$ are the controller parameters, i.e. K_P, T_I and T_D, and the 3-dimensional search space $\mathrm{G} \in \mathcal{R}^3$ is pre-specified as $\mathrm{G} = [0, 100]^3$ to widely span the optimization range of the controller design.

5 RESULTS AND DISCUSSION

This Section investigates the reliability of the integrated energy-communication system equipped with optimum PID controller to enhance the AGC performance and reduce system frequency fluctuations. We assess the impact of the two architectures, i.e. Ethernet and hybrid, and of various configurations of the open communication network on the system reliability and on the AGC performance.

5.1 *Specifications for the integrated system*

The physical system operates in nominal conditions, i.e. the stochastic wind speed, the variable sun irradiance and uncertain load, affecting, respectively, P_{WTG}, P_{PV} and P_L, given by Das et al. (2016). The coupled algebraic and ordinary differential equations for the DGS and the open communication network in Fig. 1, are numerically integrated using the Dormand-Prince method implemented in Matlab ode45 function with a fixed step size of 0.005 s. The parameters of the transfer functions for the DERs in Figure 1 are provided in Das et al. (2016). Additionally, the physical layer has the following specifications:

- The base value for the apparent power is 150 kW.
- The rated apparent power of the WTG is 150 kW, the rated electrical frequency is 60 HZ and the rated voltage for the induction machine is 440 V.

- The structure of a 150 kW PV power generation has 93 parallel strings and consists of 7 series-connected modules (SunPower SPR-230E-WHT-D).
- Based on ***Remark 1***, $0 \le \Delta P_{DEG} \le 0.7$ pu, $|\Delta P_{FESS}| \le 0.3$ pu and $|\Delta P_{BESS}| \le 0.3$pu, provide the largest rated output of the DEG, FESS and BESS (Pan and Das, 2016).

The open communication network is implemented in Truetime simulator (Cervin et al., 2010). The data rate of the communication architectures is 800 Kbits/s, and the WLAN is characterized by transmission power 20 dbm, receiver signal threshold –48 dbm, ACK timeout 0.04 ms and the retry limit 5.

5.2 *Reliability of the integrated system*

For illustrating purpose, the frequency tolerance band is set to ± 0.05. The parameters of the PID controller are tuned as $K_P = 37.78$, $K_I = 17.24$ and $K_D = 0.16$ (Pan and Das, 2016). Five combinations of BWShare, T_i and L_d are taken into account to investigate multiple communication scenarios in the Ethernet and in the hybrid network. For comparison purpose, the reliability of the integrated system with perfect communication, i.e. no time delays and no packet dropouts, is $R = 98.36\%$ and the value of objective function is $J = 16.36$.

Figure 2 illustrates the relationship between the reliability of the integrated system with Ethernet, and the communication parameters. Figure 2 shows that the AGC becomes unstable and the system reliability degrades if $L_d \ge 40\%$, the controller does not have sufficient observations of the system frequency measurement, and is unable to derive correct control signals. The case (BWShare = 0.1 & $T_i = 0.01$) has the largest reliability due to light congestion. In this case, data collisions between the PMU and the interference node are fewer than for other cases with smaller T_i. In the case (BWShare = 0.2 & $T_i = 0.005$), the interference node sends out disturbing traffic every 0.005 s and consumes 20% of network bandwidth, therefore data collisions between the PMU and the interference node increase. As a result, increased time delays degrade system reliability as compared to the other two cases. Increasing the activity level of the interference user and the used bandwidth leads to increasing data collisions, and therefore network congestion, which causes longer time delays. Thus, system reliability decreases, i.e. comparing case (BWShare = 0.1 & $L_d = 0.1$) and case (BWShare = 0.2 & $L_d = 0.1$), and comparing case ($T_i = 0.01$ & $L_d = 0.1$) and case ($T_i = 0.002$ & $L_d = 0.1$), respectively. Additionally, if $T_i \ge 0.01$, the competition for sending data packet between the interference node and the PMU reduces, network conditions improve and thus the system reliability increases.

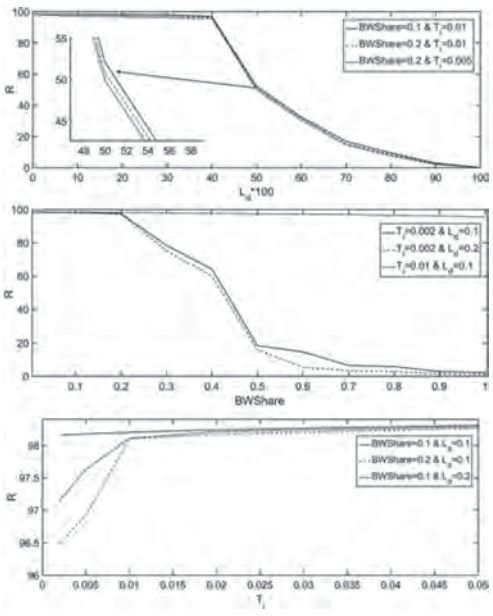


Figure 2. System reliability as a function of BWShare, T_i and L_d.

Table 1 provides the estimated reliability of the integrated system with Ethernet and with hybrid network. Such values of communication parameters are selected because they generate 10 ms to 2 s time delays, and losses of 1% to 10% of the overall packet stream, which are representative of real communication networks (Cervin et al., 2010). A negative correlation between system reliability and $E\left[J(\vec{x}) \right]$ can be inferred from Table 1. This is expected because when the system reliability is high, most of the system frequency measurements are within the tolerance band, which thereby leads

Table 1. Reliability of the integrated system for different network configurations [BWShare T_i L_d] of the two architectures.

Configurations	R_1	$\mathrm{E}\left[J_1(\vec{x}) \right]$	R_2	$\mathrm{E}\left[J_2(\vec{x}) \right]$
[0.2 0.005 5%]	98.20	17.87	96.43	19.40
[0.3 0.005 5%]	97.33	18.41	88.55	45.44
[0.1 0.005 5%]	98.24	17.25	97.14	18.73
[0.2 0.0025 5%]	96.72	19.30	91.65	42.75
[0.2 0.01 5%]	98.21	17.67	97.74	18.19
[0.2 0.005 10%]	96.92	19.17	96.27	20.03
[0.2 0.005 1%]	98.23	17.34	97.84	18.12

*R_1 and $\mathrm{E}\left[J_1(\vec{x}) \right]$ are the reliability and objective value of the integrated system with Ethernet.
*R_2 and $\mathrm{E}\left[J_2(\vec{x}) \right]$ are the reliability and objective value of the integrated system with hybrid network.

to a small value of the objective function. Table 1 shows that the Ethernet architecture ensures higher reliable to the integrated system as compared to the hybrid architecture. This is expected because the data exchange between the AP and the RTU in the hybrid architecture is provided by WLAN, which introduces additional time delays and packet dropouts, and results in lower system reliability.

5.3 *Optimal PID controller for the AGC of DERs*

PSO is employed to design the optimal PID controller. In the PSO, the number of particles and the number of iterations are set to 30 and 50. For each particle, the stochastic objective function $\mathrm{E}\left[J(\vec{x}) \right]$ is estimated from 200 MC samples. The performance of the adopted PSO is discussed by Pan and Das (2016) with respect to the convergence rate. The PSO can achieve the local optimal PID controller after 50 iterations (Pan and Das, 2016).

Table 2 lists the optimal control parameters $[K_P\ K_I\ K_D]$ for the integrated system with two network architectures, which minimize $\mathrm{E}\left[J(\vec{x}) \right]$ under various configurations of [[BWShare T_i L_d]. Different combinations of BWShare, T_i and L_d generate different levels of network traffic in different scenarios of a day, i.e. light, medium and heavy [27]. Large BWShare, small T_i and large L_d simulate intense communication flows in the open communication network. For the perfect communication, the optimal control parameters are $K_P = 36.06$, $K_I = 16.06$ and $K_D = 0.23$, the reliability is $R = 99.37$ and $\mathrm{E}\left[J(\vec{x}) \right] = 9.99$. Table 2 quantifies the amount of system reliability and control effectiveness that is lost due to the effects of communication degradation with respect to the perfect communication case. As an example, $\mathrm{E}\left[J(\vec{x}) \right] = 17.47$ is the minimum expected value of the objective function for the Ethernet work configuration [0.2 0.005 5%], which implies that the PID controller can provide the best AGC performance and highest system reliability. The increase in the BWshare and L_d, and decrease in T_i cause the reduction of R and the increase of $\mathrm{E}\left[J(\vec{x}) \right]$, which is in line with common intuition. Such changes definitely deteriorate the performance of communication network, and lead to inaccurate control of the DERs and large system frequency fluctuations.

Table 2. Optimal PID controllers $[K_P\ K_I\ K_D]$ for different network configurations [BWShare T_i L_d] of the two architectures.

Configuration	Optimal PID and performance	Network architectures: Ethernet	Network architectures: Hybrid
[0.2 0.005 5%]	$[K_P\ K_I\ K_D]$	[37.37 17.24 0.16]	[43.02 17.63 0.12]
	R	98.22	98.15
	$\mathrm{E}\left[J(\vec{x}) \right]$	17.47	17.96
[0.3 0.005 5%]	$[K_P\ K_I\ K_D]$	[39.54 23.57 0.13]	[32.99 21.68 0.09]
	R	97.64	94.43
	$\mathrm{E}\left[J(\vec{x}) \right]$	18.22	31.22
[0.1 0.005 5%]	$[K_P\ K_I\ K_D]$	[37.96 16.02 0.15]	[36.12 16.55 0.16]
	R	98.30	98.20
	$\mathrm{E}\left[J(\vec{x}) \right]$	16.75	17.71
[0.2 0.0025 5%]	$[K_P\ K_I\ K_D]$	[32.54 17.78 0.15]	[36.83 21.24 0.12]
	R	96.94	94.17
	$\mathrm{E}\left[J(\vec{x}) \right]$	19.15	30.57
[0.2 0.01 5%]	$[K_P\ K_I\ K_D]$	[37.76 18.51 0.15]	[41.00 15.85 0.12]
	R	98.27	98.22
	$\mathrm{E}\left[J(\vec{x}) \right]$	16.99	17.51
[0.2 0.005 10%]	$[K_P\ K_I\ K_D]$	[41.04 18.73 0.14]	[41.50 23.09 0.13]
	R	98.05	97.04
	$\mathrm{E}\left[J(\vec{x}) \right]$	18.01	18.79
[0.2 0.005 1%]	$[K_P\ K_I\ K_D]$	[39.71 22.65 0.16]	[43.04 17.72 0.14]
	R	98.26	98.21
	$\mathrm{E}\left[J(\vec{x}) \right]$	17.04	17.64

5.4 *Impact of communication degradation on optimum AGC operations*

The degradation of the communication network performance can affect the AGC operations even for optimally-controlled DER components. We quantify the extent of this impact for three architectures, i.e. perfect communication, Ethernet, and hybrid network, and the same WTG output, PV output and load demand. The three integrated systems are equipped with optimum PID controllers for the communication configuration [0.2 0.005 5%], given in Table 2 Line 1. Figure 3 shows the power curves of FESS, BESS and DEG, the power imbalance $\Delta P_S - \Delta P_L$, and the system frequency Δf, in the three integrated systems for $t \in [49.5\,s, 60\,s]$ of one MC simulation. The DER system experiences a sudden load increase at $t = 50$ s, and ΔP_L rises from 0.8 pu to 0.95 pu. In the perfect communication case, at $t = 52$ s, the output of the FESS, BESS and DEG rises, respectively, from 0.026 pu to 0.124 pu, from 0.008 pu to 0.038 pu, and from 0.009 pu to 0.029 pu. As a

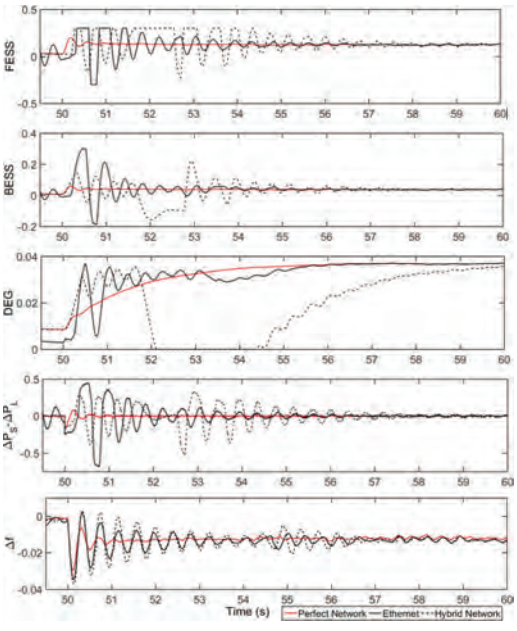


Figure 3. Evolution of FESS, BESS and DEG output, power imbalance $\Delta P_S - \Delta P_L$ and system frequency for a sudden load increase (0.15 pu) at $t = 50\ s$. Positive values of FESS and BESS power result in net injection into the grid.

result, $\Delta P_S - \Delta P_L$ is negative and very small, and, therefore, a negative and small frequency deviation $\Delta f = -0.013$ occurs.

For the integrated system with Ethernet and hybrid network, the FESS, BESS and DEG output, $\Delta P_S - \Delta P_L$ and Δf converge to the same values as the AGC with perfect communication with small convergence rates due to time delays and packet dropouts. The strong effects of communication degradation introduced by the WLAN in the hybrid network yield the smallest convergence rate in Figure 3. Additionally, Figure 4 highlights delays in the response of the controllable components to the control signal for the integrated systems with imperfect communication. For example, the output of FESS in the hybrid architecture and the output of BESS in the Ethernet architecture do not change in the time interval [50s, 50.2s], as compared to the perfect communication case. FESS and BESS keep absorbing (releasing) the same power from (to) grid even in unbalanced conditions because the data packets containing the control signals do not reach the destination. The control center lacks enough frequency measurements from the PMU, and issues inaccurate commands. Delays in component response and missing control signals ultimately cause the FESS, BESS

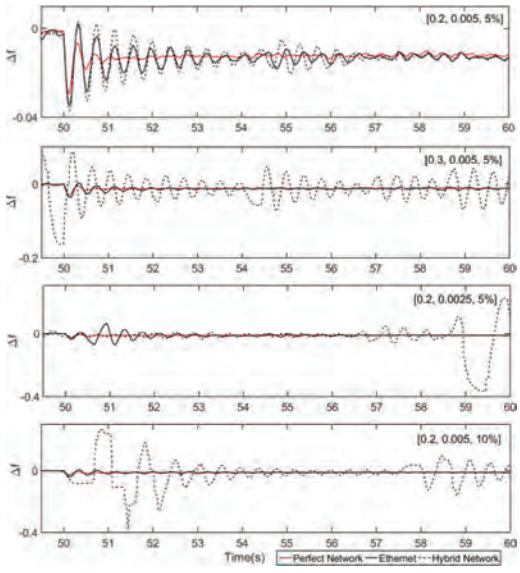


Figure 4. Evolution of system frequency for a sudden load increase (0.15 pu) at $t = 50\ s$ under increasing communication degradation.

and DEG to perform untimely and inaccurately, leading to large frequency oscillations.

Finally, we investigate the robustness of the PID-controlled AGC operations optimized for the configuration [0.2 0.005 5%] in Table 2, Line 1, subjected to increasing communication degradation. Figure 4 plots the system frequency Δf in the three integrated systems for $t \in [\,49.5\ s, 60\ s\,]$ of one MC simulation. Four configurations are shown, respectively, the optimization conditions, BWShare. increased from 0.2 to 0.3, T_i decreased from 0.005 to 0.0025, and L_d increased from 5% to 10%. Similar to Figure 4, ΔP_L rises from 0.8 pu to 0.95 pu at $t = 50$ s. Fig. 5 unveils the abilities of various controllers in providing satisfied AGC performance for unknown communication configurations. The optimized controller for Ethernet is still robust to degraded communication but small frequency fluctuations occur. The reliability in the four configurations is, respectively, 98.22%, 96.17%, 95.84% and 95.63%. Conversely, the optimized controller for hybrid network cannot provide robust AGC operations under degraded communication, and the system frequency becomes unstable. The reliability in the four WLAN network configurations is, respectively, 98.15%, 71.58%, 85.9% and 73.85%. The analysis of the PID tuning indicates that the PSO produces different optimal PID controllers for different runs of the optimization algorithm [10]. Nonetheless, even if PID controllers tuned using heuristics have different values

of R and $\mathrm{E}\left[J(\vec{x}) \right]$, they are capable of stabilizing the system frequency quickly in the analyzed network configurations.

6 CONCLUSION

This work proposes a system-of-systems framework for investigating the operations of integrated DER systems and open communication networks, and the design of optimal AGC strategies in the face of communication degradation. The results show that the activity level of interference traffic and the expected percentage of bandwidth used by the interference user have a significant impact on system reliability. Low system reliability is observed for large BWShare and small T_i. The AGC based on a discrete-time PID controller is developed to suppressed the frequency oscillations in the integrated system and enhance reliability. Optimization results show that the PSO-based PID controller is capable of reducing system frequency fluctuations and improve AGC performance. The communication degradation delays the component response to the control signal and cause their output to remain unchanged even in the presence of power imbalance until an updated packet is received. Finally, the AGC with Ethernet shows the largest reliability and robustness against unknown communication configurations, and results in small frequency oscillations.

Future work will focus on the time-delay compensation by designing a real-time Smith predictor based on the estimated communication delay. Additionally, the effect of packet dropout can be mitigated by reconstructing missing the data set to form a complete one.

REFERENCES

Ahmadi A. & Aldeen, M. 2017. Robust overlapping load frequency output feedback control of multi-area interconnected power systems. *International Journal of Electrical Power & Energy Systems* 89: 156–172.

Das, D.C., Roy, A.K. & Sinha, N. 2012. GA based frequency controller for solar thermal–diesel–wind hybrid energy generation/energy storage system. *International Journal of Electrical Power & Energy Systems* 43(1), 262–279.

Driesen J. & Katiraei, K. 2008. Design for distributed energy resources. *IEEE Power and Energy Magazine* 6(3): 30–40.

Ghoshal, S.P. 2004. Optimizations of PID gains by particle swarm optimizations in fuzzy based automatic generation control. *Electric Power Systems Research* 72(3): 203–212.

Grenier, M. & Navet, N. 2008. Fine-tuning MAC-level protocols for optimized real-time QoS. *IEEE Transactions on Industrial Informatics* 4(1): 6–15.

Kumar, L.V.S., Kumar, G.V.N. & Madichetty, S. 2017. Pattern search algorithm based automatic online parameter estimation for AGC with effects of wind power. *International Journal of Electrical Power & Energy Systems* 84: 135–142.

Kuzlu, M., Pipattanasomporn M. & Rahman, S. 2014. Communication network requirements for major smart grid applications in HAN, NAN and WAN. *Computer Networks* 67: 74–88.

Lee, D.J. & Wang, L. 2008. Small-signal stability analysis of an autonomous hybrid renewable energy power generation/energy storage system, part I: Time-domain simulations. *IEEE Transactions on Energy Conversion* 23(1), 311–320.

Liang, Y., Chen, T.W. & Pan, Q. 2010. Optimal linear state estimator with multiple packet dropout. *IEEE Transactions on Automatic Control* 55(6):1428–1433.

Martin, K.E., Benmouyal, G., Adamiak, M.G. & Begovic, M. 1998. IEEE Standard for Synchrophasors for Power Systems. *IEEE Transactions on Power Delivery* 13: 73–77.

Minero, P., Franceschetti, M., Dey, S. & Nair, G.N. 2009. Data rate theorem for stabilization over time-varying feedback channels. *IEEE Transactions on Automatic Control* 54(2): 243–255.

Mo, H.D., Li, Y.F. & Zio, E. 2016. A system-of-systems framework for the reliability analysis of distributed generation systems accounting for the impact of degraded communication networks. *Applied Energy* 183: 805–822.

Pan, I. & Das, S. 2016. Fractional order AGC for distributed energy resources using robust optimization. *IEEE Transactions on Smart Grid* 7(5): 2175–2186.

Peng, C. & Han, Q.L. 2016. On designing a novel self-triggered sampling scheme for networked control systems with data losses and communication delays. *IEEE Transactions on Industrial Informatics* 63(2): 1239–1248.

Ray, P.K., Mohanty, S.R. & Kishor, N. 2011. Proportional–integral controller based small-signal analysis of hybrid distributed generation systems. *Energy Conversion and Management* 52(4): 1943–1954.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Bayesian information fusion for non-competing relationship degradation process

Junyu Guo, Hong-Zhong Huang, Yan-Feng Li, Jie Zhou & Xiang-Yu Li
Center for System Reliability and Safety, University of Electronic Science and Technology of China, Chengdu, Sichuan, P.R. China

ABSTRACT: The degradation analysis method is widely used to the reliability analysis of the products with long life and high reliability. There is a non-competing multi-degenerate situation widely exist in some mechanical products, non-competing relationship multiple degradation model is proposed in this paper to describe it. The degradation process of products which proposed in this paper have two degradation indicators, the gamma process is used to characterize the degradation process. A Bayesian fusion framework of degradation analysis is presented to deal with the small sample size problem of the products and the MCMC method is used to simulate the samples of model parameters. A case-study of the spool valve is discussed to demonstrate the model and method we proposed.

1 INTRODUCTION

Nowadays, with the progress of the times and the development of science and technology, the complex systems such as industrial, transportation and military, are generally required to performance with a high reliability level (Ye et al., 2014). Methods such as degradation analysis have been developed to enhance the reliability analysis of these systems (Chen and Tsui, 2013, Nelson, 1981). Degradation analysis methods have been widely used in academic research and industrial productions. In general, the complex products with long life and high reliability may have multiple performance processes. The multiple degradation processes based on the competing relationship have been widely studied (Peng et al., 2016, Pan et al., 2011). In the competing relationship multiple degradation models, the products confront a failure when anyone of the performance indicator reaches a predefined threshold. However, there is a non-competing multi-degenerate situation widely exist in mechanical products (Yang et al., 2016), the competing relationship multiple degradation models may not be suitable for this situation. To address the problem, a modified multiple degradation processes model is proposed. A classic example, i.e., spool valve, is investigated in this paper to demonstrate the proposed method. The wear rate of spools and sleeves are two non-competing performance indicators of spool valve.

Earlier spool valve reliability models did not consider the small sample size problem in reliability analysis, but in some cases, small sample size problem is inevitable. A modified model that considers the small sample problem is introduced in this paper. In order to obtain more accurate analysis result under limited degradation observations, it is imperative to take full advantage of the degradation information from Original Equipment Manufacturers (OEMs) and user plant's. Degradation information from different sources have the mathematically common, so they can be fused during the reliability analysis (Feng and Zhou, 2008).

The objective of this paper is to present a hierarchical Bayesian method of gamma process model for degradation process modeling and analysis. The hierarchical Bayesian method could fusion degradation information from different sources. The Markov chain Monte Carlo (MCMC) method is used to estimate the model parameters.

The remainder of this paper are organized as follows. Section 2 introduces the gamma process model for the non-competing relationship multiple degradation model. In Section 3, a Bayesian information fusion method is presented. Section 4 presents the degradation analysis of spool valve to illustrate the proposed method. Section 5 summarizes the whole work and prospects of the future work.

2 THE DEGRADATION MODELS

2.1 *Gamma process model*

The performance degradation process of a product is usually an evolutionary process. The stochastic process model can describe this process well.

The gamma process is a stochastic process generally used to describe the performance degradation of mechanical products. Because the increment of the gamma process is non-negative, it is consistent with the performance evolvement of the mechanical products.

The gamma process is a stochastic process $\{X(t), t>0\}$ with the shape parameter α(t) and scale parameter λ is denoted as $X(t) \sim Ga(\alpha(t), \lambda)$. The gamma process has the following properties: $X(t)$ is independent incremental process with $X(0)=0$, and the degradation increments $\Delta X(t) = X(t+\Delta t) - X(t)$ follow gamma distributions as $\Delta X(t) \sim Ga(\Delta\alpha(t), \lambda)$ with $\Delta\alpha(t) = \alpha(t+\Delta t) - \alpha(t)$. The mean and variance of $X(t)$ are $\alpha(t)/\lambda$ and $\alpha(t)/\lambda^2$, respectively, where $\alpha(t)$ is a right-continuous non-decreasing function in $[0,\infty)$, with $\alpha(0)=0$ (Abdel-Hameed, 1975).

The probability density function (PDF) $g(x)$ of the gamma process is defined by

$$g_{X(t)}(x|\alpha(t),\lambda) = \frac{\lambda^{\alpha(t)}}{\Gamma(\alpha(\mathrm{t}))} x^{\alpha(t)-1}\exp(-\lambda x) I_{\mathrm{X}(0,\infty)}(x) \tag{1}$$

where

$$I_{(0,\infty)}(x) = \begin{cases} 1 & x \in (0,\infty) \\ 0 & x \notin (0,\infty) \end{cases} \tag{2}$$

and $\Gamma(\alpha) = \int_0^\infty x^{a-1}e^{-x}dx$ is gamma function (Noortwijk, 2009).

Suppose that a product's degradation process is characterized by $\{X(t), t>0\}$, it fails when the performance indicator reaches a predefined threshold C. Thus, the lifetime of the product is defined as $T = \inf\{t | X(t) \geq C\}$. The reliability $R(t)$ of product or system at time t can be expressed as:

$$R(t) = \Pr\{\sup_{s\leq t}(X(s)) \leq C\} \tag{3}$$

$$\begin{aligned} R(t) &= P_r\{T \geq t\} = P_r\{X(t) < C\} \\ &= \int_0^C g_{X(t)}(x|\alpha(t),\lambda)dx \\ &= \frac{1}{\Gamma(\alpha(\mathrm{t}))}\int_0^{C\lambda} u^{\alpha(t)-1}e^{-\lambda}du \end{aligned} \tag{4}$$

2.2 *The non-competing relationship multiple degradation model*

Suppose that the m performance indicators of the product are mutually independent, $D_p(t)$ is the degradation data of the Lth degradation processes at the time t. All degradation processes lead to one failure mode, the reliability $R(t)$ at time t can be defined as: $R(t) = \Pr\{D_1(t) + D_2(t) + ... + D_p(t) < C\}, p = 1,...,L$. If $D_p(t)$ follows a gamma process with shape parameter α_p(t) and scale parameter λ with $p = 1,2,...,L$, when the shape parameter function α_p(t)=$\alpha_p t$, then the degradation increment ΔD_p follows the gamma distribution as follows:

$$\Delta D_p \sim Gamma(\alpha_p \Delta t, \lambda) \tag{5}$$

Based on the additivity of the gamma distribution, the following relationship can be obtained as:

$$\sum_{p=1}^{L} \Delta D_p(t) \sim Gamma\left(\sum_{p=1}^{L}\alpha_p \Delta t, \lambda\right) \tag{6}$$

According to the properties of gamma process, $(D_1(t) + D_2(t) + ... + D_p(t))$ follows a Gamma process with shape parameter function α(t) = $\sum_{p=1}^{L}\alpha_p t$ and scale parameter λ. The reliability $R(t)$ at time t can be defined as:

$$R(t) = \frac{\int_0^{C\lambda} u^{\sum_{p=1}^{L}\alpha_p t - 1} e^{-\lambda}du}{\Gamma\left(\sum_{p=1}^{L}\alpha_p t\right)} \tag{7}$$

3 RELIABILITY ANALYSIS BASED ON DEGRADATION MODEL

3.1 *Basic Bayesian framework of degradation process*

Bayesian method has two advantages for degradation analysis: (1) The capability of Bayesian method for data and information fusion, making the OEMs and user plant's performance degradation data fusion analysis possible (Peng et al., 2016). (2) The handling capacity of Bayesian approach to the uncertainty, making the performance degradation analysis and reliability analysis results are guaranteed (Efron, 2013). These two points are the basic principles to construct the basic framework of Bayesian performance degradation analysis and reliability analysis.

This framework is based on Bayesian method to take the degradation data of the OEMs and the user plant as the essentials, including the acquisition of the prior distribution and the solution of the posterior distribution.

According to the different sources of degradation data, the framework divides the acquisition of the prior distribution into two methods. The method of subjective information quantization is used to obtain the prior distribution. The prior

distribution of the degradation analysis of the user plant is directly transformed by the posterior distribution of the degradation analysis of the OEMs.

In order to estimate the posterior distribution, this paper construct the posterior distribution of parameters in degradation process model based on Bayesian method, and sampling of posterior distribution from the degradation data through MCMC method.

3.2 *Degradation process analysis method*

In this paper, we analyze the non-competing relationship multiple degradation model of two degradation indicators. Suppose that the degradation data of $D_1(t)$ and $D_2(t)$ are observed for N units. Let $D_1(t_{ij})$ and $D_2(t_{ij})$ denote the jth observation for unit i at time point t_{ij} with $j=1,...,M$ and $i=1,...,N$. Let $\Delta d_{ij}=D_1(t_{ij})-D_1(t_{i,j-1})$ and $\Delta d'_{ij}=D_2(t_{ij})-D_2(t_{i,j-1})$ be the degradation increment. According to the gamma process model, the Δd_{ij}^{p} follow the gamma distribution $\Delta d_{ij}^{p}\sim Gamma(\alpha_p\Delta t_{ij},\lambda)$.

The performance degradation data of the non-competing relationship multiple degradation model includes the OEMs data D_p^O and user plant's data D_p^U. When the performance degradation data D_p^O and D_p^U are unified as D_p with $\nu=(\alpha_1,\alpha_2,\lambda)$, the likelihood function can be presented as:

$$\begin{aligned}
&L(D_1,D_2,v\mid\alpha_1,\alpha_2,\lambda)\\
&=\prod_{i=1}^{N}\prod_{j=2}^{M}g(\Delta d_{ij}\mid\alpha_1,\lambda)g(\Delta d'_{ij}\mid\alpha_2,\lambda)\\
&=\prod_{i=1}^{N}\prod_{j=2}^{M}\frac{\lambda^{\alpha_1\Delta n_{ij}}}{\Gamma(\alpha_1\Delta n_{ij})}\Delta d^{\alpha_1\Delta n_{ij}-1}\exp(-\lambda\Delta d_{ij})\\
&\quad\frac{\lambda^{\alpha_2\Delta n_{ij}}}{\Gamma(\alpha_2\Delta n_{ij})}\Delta d'^{\alpha_2\Delta n_{ij}-1}_{ij}\exp(-\lambda\Delta d'_{ij})
\end{aligned}\tag{8}$$

where $\Delta d_{ij}=D(t_{ij})-D(t_{i,j-1})$, $\Delta d'_{ij}=D_2(t_{ij})-D_2(t_{i,j-1})$ and $g(\bullet)$ is the PDF of a gamma distribution.

Suppose that the prior information about the samples is quantified and obtained as joint prior distribution for the parameters of degradation process model as $\pi(\theta)=\pi(\alpha_1,\alpha_2,\lambda)$. Following the Bayesian method, the joint posterior distribution of the parameters can be obtained as:

$$\begin{aligned}
&p(\alpha_1,\alpha_2,\lambda,\nu\mid D_1,D_2)\propto\pi(\theta)L(D_1,D_2,\nu\mid\theta)\\
&=\pi(\alpha_1,\alpha_2,\lambda)L(D_1,D_2,v\mid\alpha_1,\alpha_2,\lambda)\\
&=\pi(\alpha_1,\alpha_2,\lambda)\prod_{i=1}^{N}\prod_{j=2}^{M}\frac{\lambda^{\alpha_1\Delta n_{ij}}}{\Gamma(\alpha_1\Delta n_{ij})}\Delta d_{ij}^{\alpha_1\Delta n_{ij}-1}\\
&\exp(-\lambda\Delta d_{ij})\frac{\lambda^{\alpha_2\Delta n_{ij}}}{\Gamma(\alpha_2\Delta n_{ij})}\Delta d'^{\alpha_2\Delta n_{ij}-1}_{ij}\exp(-\lambda\Delta d'_{ij})
\end{aligned}\tag{9}$$

From the Equation (9), it is a comprehensive description of the a prior information and degradation process.

3.3 *Bayesian information fusion method*

According to the degradation process modeling and Bayesian method, the degradation data of the OEMs and the user plant are analyzed to obtain the joint posterior distribution of the model parameters. The realization process is as follows:

$$\begin{aligned}
&p_I(\alpha_1,\alpha_2,\lambda,\nu\mid D_1^O,D_2^O)=\pi(\alpha_1,\alpha_2,\lambda)\\
&\prod_{i=1}^{N}\prod_{j=2}^{M}\frac{\lambda^{\alpha_1\Delta n_{ij}}}{\Gamma(\alpha_1\Delta n_{ij})}\Delta d_{ij}^{O\,\alpha_1\Delta n_{ij}-1}\exp(-\lambda\Delta d_{ij}^{O})\\
&\frac{\lambda^{\alpha_2\Delta n_{ij}}}{\Gamma(\alpha_2\Delta n_{ij})}\Delta d_{ij}^{O'\,\alpha_2\Delta n_{ij}-1}\exp(-\lambda\Delta d_{ij}^{O'})
\end{aligned}\tag{10}$$

$$\begin{aligned}
&p_{II}(\alpha_1,\alpha_2,\lambda,\nu\mid D_1^U,D_2^U,D_1^O,D_2^O)\\
&=p_I(\alpha_1,\alpha_2,\lambda,\nu\mid D_1^O,D_2^O)\\
&\prod_{i=1}^{N}\prod_{j=2}^{M}\frac{\lambda^{\alpha_1\Delta n_{ij}}}{\Gamma(\alpha_1\Delta n_{ij})}\Delta d_{ij}^{U\,\alpha_1\Delta n_{ij}-1}\exp(-\lambda\Delta d_{ij}^{U})\\
&\frac{\lambda^{\alpha_2\Delta n_{ij}}}{\Gamma(\alpha_2\Delta n_{ij})}\Delta d_{ij}^{U'\,\alpha_2\Delta n_{ij}-1}\exp(-\lambda\Delta d_{ij}^{U'})
\end{aligned}\tag{11}$$

where D_1^O and D_2^O are the degradation data from the OEMs, $\pi(\alpha_1,\alpha_2,\lambda)$ is prior distribution quantified by the prior information for products, $p_I(\eta,\delta,\gamma,\nu\mid D^O)$ is the posterior distribution of the model parameters obtained by combination of prior information and the information contained in the degradation data of the OEMs. D_1^U and D_2^U are the degradation data from the user plants, $p_{II}(\eta,\delta,\gamma,\nu\mid D^U,D^O)$ is a model parameters posterior distribution, it is a description of the combination of prior information and information contained in the degradation data from OEMs and user plants.

From the Equations (10) and (11), the MCMC method is used to estimate the model parameters because of their computational complication. In most actual applications with Bayesian methods, it is difficult to obtain the posterior distribution. The MCMC method is used to construct a Markov chain, the invariant distributionis of the Markov chain is the posterior distribution that is needed to be accurately estimated. In this paper, we assume that the prior distributions of the model parameters are non-informative, and the OpenBUGS is used to perform the Gibbs sampling after the model parameters is estimated.

4 ILLUSTRATIVE EXAMPLE

The reliability of hydraulic systems is determined by various subsystems, and the spool valves is one

of the most important subsystems. The reliability of the spool valves directly affects the reliability of the hydraulic system. The main failure mode of the spool valves is the wear degradation of spools and sleeves. When the total wear degradation of spool and sleeve reach predetermined threshold, the spool valves is considered as failure. The Bayesian information fusion.

To obtain information about degradation of spool valves in the hydraulic system, the wear degradation of spools and sleeves of six spool valves are monitored. The wear degradation paths of OEMs are presented in Figure 1, and the wear degradation paths of user plants are shown in Figure 2.

As discussed in Section 3, the wear degradation increment of sleeves is $\Delta d_{ij} = D_{SL}(t_{ij}) - D_{SL}(t_{i,j-1})$ which is modeled as $Ga(\alpha_{SL}\Delta t_{ij}, \lambda)$ and the wear degradation increment of spools is

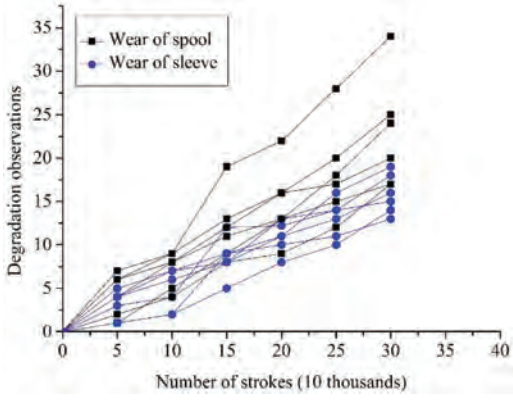


Figure 1. The wear degradation paths of OEMs.

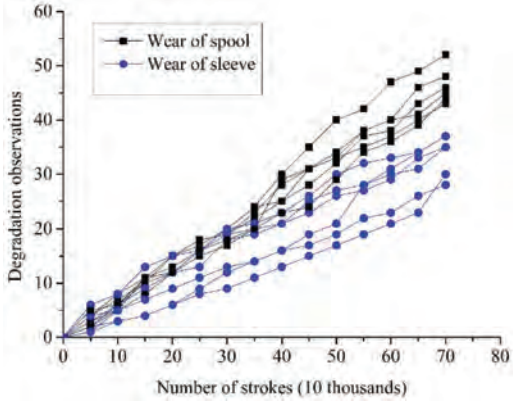


Figure 2. The wear degradation paths of user plants.

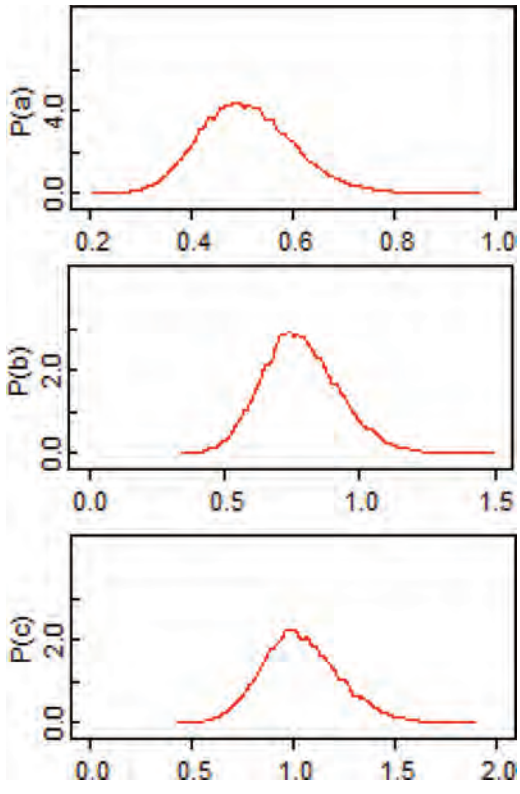


Figure 3. The posterior pdf of model parameters: (a) $\alpha_{SL}^{(OEMs)}$, (b) $\alpha_{SP}^{(OEMs)}$, (c) $\lambda^{(OEMs)}$.

$\Delta d'_{ij} = D_{SP}(t_{ij}) - D_{SP}(t_{i,j-1})$ that is modeled as $Ga(\alpha_{SP}\Delta t_{ij}, \lambda)$. Owing to the limitation on the available prior information, the non-informative prior distribution, which is the prior distributions for parameters of the degradation process model of OEMs, quantized from the subjective information. It could be given as:

$$\alpha_{SL} \sim U(0,100), \alpha_{SP} \sim U(0,100), \lambda \sim U(0,100)$$

The $U(a,b)$ is a uniform distribution.

The MCMC method is used to simulate the samples of model parameters, and then we use the OpenBUGS to generate 20000 samples.

The posterior pdf of model parameters for degradation process of the OEMs are presented in Figure 3. Based on the framework discussed in Section 3, these posterior distributions transformed into the prior distribution of degradation process model parameters of user plants.

The results for parameter estimation given in Table 1 are obtained from the generated posterior samples. The predefined degradation threshold is

Table 1. The results for parameter estimation.

	Mean	Standard deviation	Confidence interval	
			2.5%	97.5%
α_{SL}	0.6542	0.06317	0.5374	0.7827
α_{SP}	0.9189	0.0886	0.7538	1.102
λ	1.348	0.1308	1.104	1.62

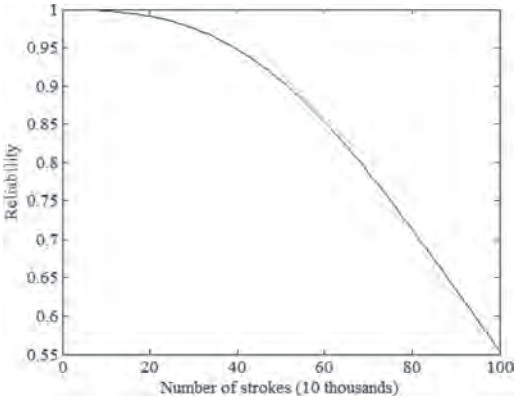


Figure 4. Reliability of the spool valves.

$C = 120$, the reliability of the spool valves can be obtained from the posterior distribution of model parameters as presented in Figure 4.

5 CONCLUSIONS

In this paper, a reliability analysis model for the non-competing relationship multiple degradation process of products with two performance indicators is presented. The characteristics of these degradation process are modeled by gamma process. Then, a hierarchical Bayesian method is presented to fusion the degradation information from different sources. Therefore, the Bayesian method is used to estimate the model parameters and the MCMC method is used to simulate the samples of model parameters. A case-study of a spool valve is provided to demonstrate the proposed model and method.

It should be noted that there are some open questions for future work. According to the individual heterogeneity of the products, the random effect could be introduced to the non-competing relationship multiple degradation process. The different working environment of the products in the OEMs and the user plant could be considered into the modeling of the degradation process.

ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China (Contract No. 51775090).

REFERENCES

Abdel-Hameed, M., 1975, A gamma wear process. IEEE transactions on Reliability, 24(2): 152–153.

Chen, N., & Tsui, K.L., 2013. Condition monitoring and remaining useful life prediction using degradation signals: Revisited. IIE Transactions, 45(9): 939–952.

Efron, B., 2013, Bayes' theorem in the 21st century. Science, 340(6137), 1177–1178.

Feng, J., & Zhou, J., 2008, Small-sample reliability information fusion approach based on bayes-fuzzy logistic operator. *Journal of Aerospace Power,* 23(9): 1633–1636.

Nelson, W., 1981, Analysis of performance-degradation data from accelerated tests. IEEE *Transactions on Reliability*, 30(2): 149–155.

Noortwijk J.M.V., 2009, A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, 94(1): 2–21.

Peng, W., Li, Y.F., Yang, Y.J., Zhu, S.P., & Huang, H.Z., 2016, Bivariate analysis of incomplete degradation observations based on inverse Gaussian processes and copulas. IEEE *Transactions on Reliability,* 65(2): 624–639.

Peng, W., Li, Y.F., Mi, J., Yu, L., & Huang, H.Z., 2016, Reliability of complex systems under dynamic conditions: A Bayesian multivariate degradation perspective. *Reliability Engineering & System Safety,* 153, 75–87.

Pan, Z., & Balakrishnan, N., 2011, Reliability modeling of degradation of products with multiple performance characteristics based on gamma processes. *Reliability Engineering & System Safety*, 96(8): 949–957.

Yang, Y.J., Peng, W., Zhu, S.P., & Huang, H.Z., 2016, A Bayesian approach for sealing failure analysis considering the non-competing relationship of multiple degradation processes. Eksploatacja i Niezawodnosc - *Maintenance and Reliability,* 18(1): 10–15.

Ye, Z.S., Xie, M., Tang, L.C., & Chen, N., 2014, Semiparametric estimation of gamma processes for deteriorating products. *Technometrics*, 56(4): 504–513.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Evaluation of the reliability of composite materials used in aviation

A. Krzyzak, G. Bemowski, R. Szczepaniak & N. Grzesik
Polish Air Force Academy, Faculty of Aeronautics, Deblin, Poland

L. Gil
University of Economy and Innovation, Lublin, Poland

ABSTRACT: Polymer composites used in engineering structures are exposed to various types of mechanical stress. The most commonly used methods of assessing their strength are the ones based on static or dynamic testing. Rare attempts are made to assess reliability of composites under a given load. The influence of the manufacturing technique on the mechanical strength of composites is also a well-known fact. The production of fibre reinforced composites using an autoclave allows for the production of high strength composites and a very small number of structural defects. However, this method is expensive and available with restriction. Much more often composites are produced by cheaper methods like hand lay-up, infusion or vacuum bag. The aim of this study is to determine the influence of the technique of making selected polymeric composites on the probability of failure before achieving a certain tensile strength threshold. Composite materials reinforced with carbon and glass fabrics have been prepared as the most commonly used composites in aviation. Composites were made by such methods as hand-lay-up, vacuum bag and pressing. Static stretching was performed, followed by a statistical analysis and a reliability analysis. The reliability analysis was performed using the Weibull model. It was found that, in terms of tensile strength, the differences between the composites made by the compression method and the manual method are negligible. Vacuum bag composites exhibit the lowest tensile strength compared to other composites. On the other hand, the analysis of reliability indicates that the highest probability of maintenance of structure continuity under load is exerted by composites with fiberglass made by the pressing technique.

1 INTRODUCTION

Reliability, which is understood quite well intuitively, is a synonym for self-assurance in operation. Since the very beginning of constructing broadly understood technical objects, this terms has been comprehended in this sense. The intention was always to maximize suitability for use, generally described by the Q quality. It needs to be stressed that the R reliability is one of the constituent elements of quality, along with e.g. efficiency, usefulness or accuracy. As a narrower concept, it describes the ability of an object to perform a specific task under certain conditions and at a given time (Godziszewski 1983, Bentley 1999).

The theory of reliability has got two main objectives (Szopa 2009):

- the formulation of theoretical grounds for the description of the laws underlying the occurrence of damage and malfunction in technical systems,
- seeking methods of design and rules of systems operation so as to minimize the possibility of damage at certain expenditures.

The reliability practice, however, is geared towards (Smalko 1972):

- exploring physical processes of the occurrence of damage,
- removing damage,
- the ability to predict the occurrence of damage and counteract them.

Thus, the science of reliability is of multidisciplinary nature. The stimulus for its development was the necessity to meet increasingly higher demands for modern technical objects. At the beginning of its existence, it served to project the expected behaviour of complex objects during their operation (e.g. aircraft, computers). Additionally, taking into account modern trends, research is conducted on the methodology of designing objects whose operation is reliable in a defined time (Słowinski 2002).

The assessment of the reliability of composite materials is based on the analysis of physical changes which affect them. The factors which influence the behaviour of the laminate can be divided into three main categories:

- mechanical load,
- material factors,
- environmental factors

One of the greatest threats in maintaining reliability of the composite strength is the separation of the matrix and the reinforcement. On the boundary of the phases there occur forces of adhesion and adsorption, which account for their overall strength (Hart-Smith 1990, Godzimirski 2008).

The foreseeable duration of composite utility will be specified by means of the likelihood of damage occurrence, which is determined on the basis of static and dynamic investigations. The most trustworthy results are those of tensile strength and bending (Krzyzak 2015, Esfandiari 2008, Wu, Cheng & Kang 2000).

Damage is defined as a transition from a fully-operational state to the state of deficiency or failure to comply with at least one major parameter (Godziszewski 1983).

2 METHODOLOGY

The main objective of this study is to examine the influence of a manufacture technique of selected polymer composites upon the probability of damage prior to reaching a specified threshold tensile strength. In aviation, the autoclave method is used for the manufacture of advanced aircraft parts. However, the hand lay-up and vacuum bag are the most common methods used for repair. In the research, the authors proposed an alternative pressing method.

The object of research was composites which are most commonly used in aviation, with a particular focus on polymer-matrix laminates. The authors examined four-layer composites reinforced with commonly available woven fabrics of glass fibre (G) and carbon fibre (C). The matrix was epoxy resin C.E.S. R70 on the basis of bisphenol A/F of density $\rho_R = 1.16$ g/cm^3 combined with C.E.S. H72 hardener whose density equalled $\rho_H = 1.02$ g/cm^3 in the ratio of 100:54.

The polymer composites were produced with three methods: hand lay-up (hand), pressing method (press) and vacuum bag (bag). During the hand lay-up lamination, particular attention was drawn to the total saturation of the reinforcement fabric with resin, as well as the reinforcement setup in such a way that the fibers ran parallel. The subsequent layers measuring 0.6×1 m were pressed together by a roller. In order to ensure complete curing of the resin, the composites were left for 72 hours.

Another batch of samples was made by means of the hydraulic press Mecamaq PDM 50S. The initial production process was identical as in the case of the hand lay-up method. Having laid and saturated the reinforcement layers with resin, the sheets were placed in a press and subjected to a compressive force of 5,000 kg or 49,033 N for 2 hours. Subsequently, they were removed and left for another 48 hours. During the pressing process, the composite was subjected to the pressure of 0.082 MPa.

The third series of the pieces was produced by the vacuum bag method. The reinforcement layers were saturated manually with resin and placed in a tape-sealed working area. Next, the delamination fabric was placed for an uninterrupted separation of the manufactured laminate from the auxiliary materials. Another layer of the perforated film was to allow removing excess resin. Next, lignin, which was responsible for the resin absorption, was laid. Finally, everything was tightly covered with a polymer film. The negative pressure was achieved by the TW-1A 1/6 HP pump.

The pump generated a negative pressure of approximately 1 bar, thus it equalled 0.1 MPa when calculated for the laminate surface which pressed the reinforcement layers.

Similarly to pressing, after 24 hours, the laminate was extracted and left for 48 hours. Due to practical limitations associated with the availability of a proper carbon fabric, it was merely possible to manufacture glass-reinforced samples by means of the vacuum bag method.

The obtained boards served for cutting test pieces of 10×100 mm. Special attention was drawn to the cutting direction. It was important to ensure the assumed direction of the reinforcement. The cutting was performed with a stream of water and garnet (for abrasive cutting) under high pressure. It was possible to obtain smooth and equal edges and high dimensional accuracy.

3 RESULTS

3.1 *Tensile strength*

The static tensile tests were performed on a fatigue testing Zwick/Roell machine, equipped with pneumatic grips, in accordance with the research norm DIN EN ISO 527-1 at a constant speed of the traverse movement equal to 2 mm/min. The measurement length equalled 50 mm and the length of the gripping hands was 25 mm. The samples were fitted in the grips, paying particular attention to

Table 1. Basic statistics results of composite tensile strength.

Composite material	Number of sizes	R_m	Standard deviation of the mean	Coefficient of variation
C (hand)	20	457.38	54.70	11.96%
C (press)	20	452.31	66.47	14.69%
G (hand)	20	275.64	14.93	5.42%
G (press)	20	291.33	22.38	7.68%
G (bag)	18	240.89	21.62	8.97%

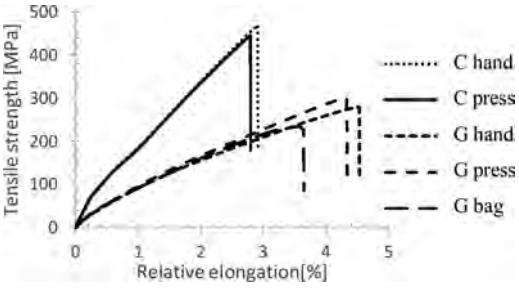


Figure 1. Tensile strength in function of relative elongation for composites made by diffirent methods (examples).

their vertical orientation. Even minor deviations in this respect would negatively affect the reliability of the results obtained.

The highest average tensile strength of materials reinforced with a carbon fabric was exercised by the laminate made with the hand method C (hand) – 457 MPa, which only slightly exceeded the pressed product C (press) – 452 MPa. In the case of the glass composites, the situation is quite different. The highest tensile strength was exercised by the composite G (press) – 291 MPa, a slightly lower by G (hand) – 276 MPa. The laminate manufactured with the vacuum bag method G (bag) reaches the minimum tensile strength at 241 MPa. Both the standard deviation, the standard error of the mean and the coefficient of variation indicate that the repeatability of research results is far greater in the case of carbon-reinforced composites, irrespective of the manufacture method, than for glass laminates.

In the case of composites, due to their two-phase structure and the occurrence of tiny matrix cracks affecting the emergence of permanent deformations, it is difficult to determine a typical elasticity boundary (Mallick 2007, Zhang, Li, Is & Yu 2013). While analysing an increase in force with respect to elongation in the process of stretching, however, it may be concluded that the test pieces underwent purely tensile deformation, only in the initial phase of the test. In the following part, until the breaking, they behaved like fragile bodies (Figure 1). Given the characteristics of composite materials and their two-phase structure, it can be deduced that the emerging matrix microcracks were to blame.

Relative extension of carbon composites during stretching was on average equal to 2.8%. The differences between C (hand), and C (press) were negligible. The situation concerning various types of samples made by hand and by pressing was similar. Both types showed similar relative deformation of 4.35% and 4.22%, respectively. Taking into account the standard deviation of results, this small discrepancy may be accidental. It was noted, however, that there is a significant maximum decline in relative elongation of the samples produced by the bag vacuum method, with regard to the samples obtained by other methods. The average value of the considered parameter was, in this case, 3.57%.

3.2 *Permissible value of mechanical strength*

The results obtained and the descriptive statistical analysis may give rise to determining permissible tensile strength to stretching of the examined composites. The safety coefficient is determined on the basis of, among others, numerous studies and a statistical analysis, however in the case of composite materials there are no rigid guidelines for its adoption. This is explained by a variety of factors determining the strength and structural characteristics of composites and, in the selection of an index, the computations as well as the constructor's experience play a large role.

Thus, the authors adopted the permissible strength value n_m as an average strength of the composites made by hand decreased by the safety coefficient on the level of 5%. In the case of composites made by the manual lamination method, with carbon reinforcement, the permissible value of the tensile strength equalled 435.51 MPa, whereas for composites with glass reinforcement it was 261.86 MPa. It was assumed that the largest structural defects occur in composites made by manual lamination, therefore the above-mentioned permissible strength will also be adequate for composites produced with methods of more advanced technology.

3.3 *Analysis of reliability*

The Weibull reliability analysis was conducted. Thus, the authors initially determined the regression functions of reliability likelihood and coefficients of determination R^2.

The formula of the function determined through approximation, using the method of least squares

errors, in a simple form $y = ax + b$, was obtained from the following formulas:

$$a = \frac{\sum (xy) - N\bar{x}\,\bar{y}}{\sum x^2 - N\bar{x}^2}$$

$$b = \bar{y} - a\bar{x}$$

where:
x – predictor value,
y – dependent variable value,
$\bar{x}, \bar{y}$ – mean values,
N – number of observations.

The coefficient of determination informs how much the accepted model explains the collected measurement data. The better model adjustment, the closer is its value to one. It is calculated using the formula:

$$R^2 = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}$$

where: $\hat{y}_i$ – prediction on the basis of the regression model of the variable value.

The figures (Figure 2, Figure 3) show the approximation of the probability distribution of maintaining strength with regard to the type of reinforcement and the methods of manufacturing the composite.

The approximated functions of probability distribution indicate slight differences between the examined carbon composites. The situation is slightly different in relation to laminates reinforced with glass. In this case, the obtained functions clearly indicate that the biggest damage in the slightest stretching stresses occur in materials manufactured by means of the vacuum bag method. On the other hand, pressed composites exercised the biggest strength, demonstrating a significant advantage over laminates produced manually.

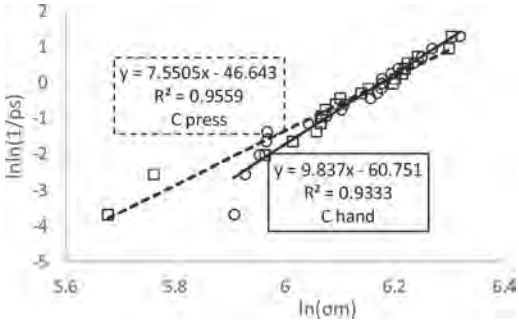


Figure 2. Approximation of the probability distribution of failure when stretching carbon-reinforced composites.

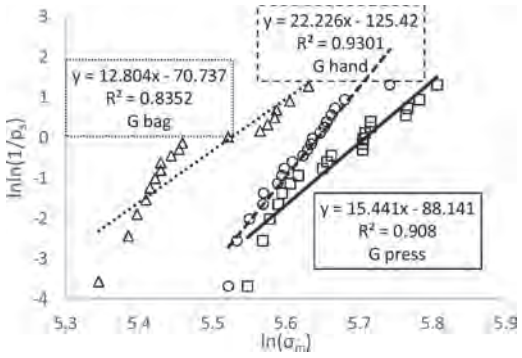


Figure 3. Approximation of the probability distribution of failure when stretching glass-reinforced composites.

The obtained values of the coefficient of determination prove that the regression function and experimental results are quite coherent. It also means that results modelling are based on proper assumptions as for the testing and adoption of distribution. It can be concluded that the obtained function can be used for the interpolation of the results within the value of variable research factors exploited in the conducted experimental studies.

When the investigated parts become damaged, mainly due to sudden wear (cracks, fractures, etc.), it is recommended to use the Weibull distribution (Saghafi, Mirhabibi & Yari 2009). It is characterized by a variable intensity of damage, which is monotone (non-decreasing) in its nature. The distribution applies to the description of fatigue service life of different materials and machinery. The reliability function, i.e. the dual parametre model has got the form:

$$R(t) = \exp\left(-\frac{t^m}{\sigma_0}\right)$$

where: t – expected operating time of objects, in strength investigations predefined as permissible strength of materials,
σ_0 – scale parametre of the Weibull model [MPa],
M – shape parametre of the Weibull model [–].

The Weibull model m is equal to the directional coefficient a of the determined approximating straight line. In other words, it corresponds to the angle of inclination of the regression line to the X-axis (Maksymiuk 2003, Yadav, Singh & Goel 2006). Alongside the increase in the Weibull module, there is a decrease in load spreading, where the critical damage is likely to occur.

The scale parameter σ_0 is a tension determined for the accumulated likelihood of failure at (Warszyński 1988):

$$p = 1 - \frac{1}{e} \approx 0.632$$

The designated parameters of the Weibull model have been listed in Table 2.

The graphs (Figure 4, Figure 5) show the functions of failure, or probability distribution functions of damage occurrence. They depict predetermined permissible tensile strength values $\sigma_m = n_m$ which constitute the adopted safe operation border (vertical dotted line). The likelihood read out at the intersection of the graph with this line indicates the percentage of the population in which there was decohesion with tension below the permissible one. The continuous horizontal line also marked probability $p_f = 0.632$ at the level. Above this value, there is accumulated occurrence of catastrophic damage for the whole population. The second quarter of the plane determined by the two restrictions is the area which excludes material out of operation. Obtaining the likelihood of damage above $p_f = 0.632$ under the influence of tension less than $\sigma_m = n_m$ is tantamount to the exclusion of the composite from use.

Table 2. Parametres of the Weibull method.

Fibre	Method	m [–]	σ_0 [MPa]
Carbon	hand	9.84	480.89
	press	7.55	484.53
Glass	hand	22.23	280.12
	press	15.44	300.23
	bag	12.80	241.18

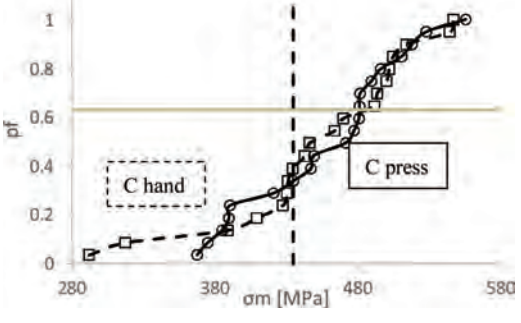


Figure 4. Empirical function of failure of carbon composites due to tensile strength.

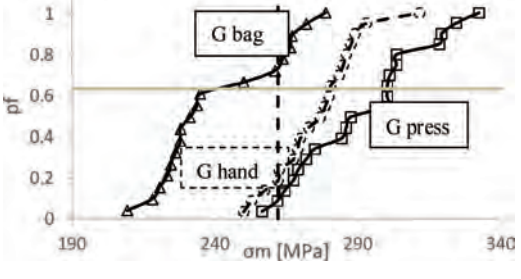


Figure 5. Empirical function of failure of glass composites due to tensile strength.

By analysing the above diagrams, it can be concluded that polymer composites manufactured by means of pressing and the manual method, both carbon and glass ones, are highly reliable and can be safely used with the adopted values of tensile stress. Laminates manufactured using the bag vacuum do not satisfy the conditions of operation at a certain level of load with assumed reliability.

It is assumed that fulfilling the condition $n_m > \sigma_m$ indicates a high probability of uninterrupted service life of a part manufactured with a given technique.

The following table (Table 3) shows the likelihood $p_f(n_m)$ of reducing the strength of the material below a predetermined permissible value. The authors adopted the following criteria of risk gradation of substantial debilitation:

$p_f(n_m) \leq 0.02$ – very low,

$0.02 < p_f(n_m) \leq 0.2$ – low,

$0.2 < p_f(n_m) \leq 0.5$ – average,

$0.5 < p_f(n_m) \leq 0.8$ – high,

$p_f(n_m) > 0.8$ – very high.

In accordance with the earlier observations, only materials produced by the vacuum method indicate a high probability of a decline in strength below the permissible safety value. The composites produced by hand and pressed are characterized by average and high reliability.

Table 3. Characteristic strength n_m and lack of reliability $p_f(n_m)$ of polymer composites manufactured with various techniques.

Fibre	n_m [MPa]	Method	σ_0 [MPa]	$p_f(n_m)$
Carbon	434.51	hand	480.89	0.34
		press	484.53	0.40
Glass	261.86	hand	280.12	0.19
		press	300.23	0.09
		bag	241.18	0.74

4 CONCLUSIONS

The technique of making polymer composites exerts an influence the likelihood of damage before achieving a certain threshold tensile strength, which can be clearly proved by analysing the results of glass laminates. The smallest probability of the emergence of such damage can be seen in the pressed samples, the reason for which might probably be more compact and free from air bubbles composite structure as compared to the samples made manually, obtained by the pressure exerted on the material in the production process. Improved matrix continuity allows an even distribution of stresses on the reinforcement in the whole volume of the element.

It was also found that the composite materials produced by the vacuum bag method (bag) showed by far the biggest likelihood of the occurrence of destructive damage prior to reaching the set threshold strength. This is surprising since the pressure exerted on the laminate in the production process should theoretically lead to a decrease in such a likelihood, similarly to the pressing method. However, these two techniques varied. Due to the pressure exerted on the laminate, the matrix may have been squeezed out of the reinforcement layers and its excess was absorbed by the lignin. This may have led to an insufficient volume share of resin in the final product. The consequence might be the failure to distribute tensions over the whole of the reinforcement, which could explain a reduction in strength. In the pressing method, resin could not find its way out of the manufactured composite. Taking into account the peculiarity of the vacuum bag technique, it can therefore be assumed that the reduction in the negative pressure generated by the pump or not using lignine could result in the reduction of the probability of the sample decohesion by reducing the outflow of the matrix from the laminate at the production stage.

Due to the fact that the approximated functions of probability distribution indicate slight differences between the examined carbon composites, and due to a relatively small number of tests performed, it was found that any conclusions so as to improve or deteriorate the properties of laminates, depending on the technique of making them, would be a too far-reaching presumption. The pressure exerted by the press appeared to be insufficient to obtain similar effects to those in glass composites, due to more compact structure of the carbon fabric used as reinforcement. The results obtained do not therefore deny the previously formulated conclusions, however they are incapable of confirming them. Presumably, it is necessary to make composites by the pressing method again, this time with more compressive load.

REFERENCES

Bentley J.P. 1999. *Introduction to Reliability and Quality Engineering*. Addison-Wesley Longman Ltd., Edinburgh Gate, Harlow.

Esfandiari A. 2008. The Statistical Investigation of Mechanical Properties of PP/Natural Fibers Composites. *Fibers and Polymers* 2008, Vol.9, No.1, 48–54.

Godzimirski, J. 2008. *Lotnicze materiały konstrukcyjne.* Warszawa: Wojskowa Akademia Techniczna.

Godziszewski, J. 1983. *Badanie niezawodności maszyn oraz ich elementów.* Zielona Góra: Wyższa Szkoła Inżynierska im. Jurija Gagarina.

Hart-Smith, L.J. 1990. Some observations about test specimens and structural analysis for fibrous composites. W.S.P. Garbo, *Composite materials, Testing and Design* (86–120). Philadelphia: ASTM STP 1059.

Krzyzak, A. 2015. *Ocena niezawodności ze względu na wytrzymałość kompozytów polimerowych ze wzmocnieniem z włókien roślinnych.* Warszawa: Polskie Naukowo-Techniczne Towarzystwo Eksploatacyjne.

Maksymiuk, J. 2003. *Niezawodność maszyn i urządzeń elektrycznych.* Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej.

Mallick P.K. 2007. *Fiber-Reinforced Composites: Materials, Manufacturing, and Design.* Third Edition. Boca Raton: CRC Press.

Saghafi A., Mirhabibi A.R., Yari G.H. 2009. Improved linear regression method for estimating Weibull parameters. *Theoretical and Applied Fracture Mechanics* 52: 180–182.

Słowiński, B. 2004. *Podstawy badań i oceny niezawodności obiektów technicznych.* Koszalin: Wydawnictwo Uczelniane Politechniki Koszalińskiej.

Smalko, Z. 1972. *Wprowadzenie do metodyki wdrażania systemów klasy CMMS.* Warszawa: Wydawnictwo PWN.

Szopa, T. 2009. *Niezawodność i bezpieczeństwo.* Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej.

Warszynski, M. 1988. *Niezawodność w obliczeniach konstrukcyjnych.* Warszawa: Państwowe Wydawnictwo Naukowe.

Wu W.F., Cheng H.C., Kang C.K. 2000. Random field formulation of composite laminates. *Composite Structures* 49: 87–93.

Yadav O.P., Singh N., Goel P.S. 2006. Reliability demonstration test planning: A three dimensional consideration. *Reliability Engineering and System Safety* 91: 882–893.

Zhang Y., Li Y., Ma H., Yu T. 2013. Tensile and interfacial properties of unidirectional flax/glass fiber reinforced hybrid composites, *Composites Science and Technology* 88: 172–177.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Research on failure mechanism and reliability of áircraft lock mechanism

Huan Pang & Ning Wang
School of Automobile, Chang'an University, Xi'an, China

Tianxiang Yu
School of Aeronautics, Northwestern Polytechnical University, Xi'an, China

ABSTRACT: Lock mechanism is one of the important and problematic components on aircraft, whose performance and reliability directly affect the mission or even safety of the aircraft. However, lock mechanism always have characteristics of complex forms, few samples and high reliability, traditional reliability assessment methods and reliability test method based on statistics is not appropriate for its reliability assessment. In addition, damage of the lock components increase with work time, the damage will make the performance of lock mechanism degradation or even cause failure. Therefore, reliability of the lock mechanism is not only function of random variables but also function of work time. In this paper, a time-variant reliability analysis method based on physics of failure was put forward, in the method, both time-invariant factors and time-variant factors are considered, and the degradation model of the components can be obtained from both practicality experiment or damage mechanism model. Taking a lock as the study object, failure modes and failure mechanism of the lock mechanism were analyzed combined with dynamical responses. Then, Response Surface Method (RSM) was used to obtain relationship between design variables and dynamic responses. Component damage was regard as a random process, after damage degradation law was obtained, considering randomicity of all the parameters, lock mechanism reliability with work time was obtained and the reliable life was predicted.

1 INTRODUCTION

Numerous mechanical systems on aircraft have lock mechanisms, such as landing gear system, cargo door system, et al. The reliability of the lock mechanisms has a great influence to the mission accomplishment and the safety of the aircraft. If the lock of landing gear door won't open in the landing time, the landing gear door cannot put down; If the cargo door lock mechanism cannot open, the airborne or airdrop mission will fail.

Several studies have been focused on the reliability of the lock mechanism. (Ouyang 1994) analyzed wear depth of the lock hook surface based on Archard's wear model, gained the changing trend of the wear reliability with the opening and closing times. (Gu, et al. 1995) used the fault tree method, from the startup and running two aspects, assessed reliability of the landing gear door's lock. (Ouyang, 1994) only focused on the partial wear of the lock mechanism, without considering the impact on the overall performance. Although lock reliability is analyzed with overall performance in consideration (Gu, et al. 1995), the degradation process and the physical processes that cause the degradation were ignored. Reliability problems of other mechanisms like lock mechanism have received extensive attention. These studies mainly focused on kinematical accuracy (Tsai et al. 2008, Wang et al. 2011), joint clearance effect (Rhee & Akay 1996, Erkaya & Yzmay 2012) and motion seizure (Ballu 2008) and other aspects. To the performance degradation caused failure problems, (Wang & Chen 2011) analyzed the time variant reliability for gear with multi-failure mode, presented the relation between reliability and work time. (Jin et al. 2013) accord to performance degradation data come from actual physics of failure test, based on failure-physics method established a model to describe MW's failure process and finally predict the work lift-time. To the component damage caused component failure or function failure, there's no too much literatures. To sum up, the present literatures don't give out an intensive study towards mechanism time-variant reliability.

The lock mechanism reliability was affected by many factors, such as the manufacturing and assembling errors, material properties, external load and other non-time-variant factors. Therefore, the lock mechanism has the possibility of

failure in the early service period, and in general, the design safety margin can make sure that the mechanism has a higher reliability in early service. However, there is a kind of factors, whose parameters varies with serving time, such as joint wear, material stress relaxation, gasket aging and corrosion. In the early service period, their influence on the lock mechanism is not significant, but with the increase of service time, these factors will degrades performance of mechanism, for example, joint wear will cause the hinge gap increases, the contact condition will become worse, causing friction increases; the stress relaxation and creep will cause the motion accuracy decreases; the sealing parts corrosion and aging may cause the performance of hydraulic system degradation, cause the driving ability decrease. Therefore, the failure probability is not only function of the variable, but also function of service time. With the service time growth, the reliability of mechanism will continue deteriorates until the failure occurs.

In this paper, failure modes and failure mechanism are analyzed, and a lock is used as an object of study. This paper is organized as follows: failure modes and failure mechanism were analyzed based on the dynamical simulation results, use failure physics method to establish damage model of the lock part, use response surface method (RSM) to obtain the deterministic function of the system, which reveals the relationship between system performances and the design variables. Then, part damage results are used to update the simulation model. Finally, considering the randomicity of variables, and using the knowledge of statistics, obtained the reliability degradation law and predicted the lifespan of mechanism.

2 RELIABILITY ESTIMATION METHOD BASED ON PHYSICS OF FAILURE

2.1 *Reliability estimation framework based on physics of failure*

Different from electronic system, mechanisms in aircraft always have properties of form complex, small sample size, and long life, high reliability requirements, so the traditional reliability assessment method based on statistics is not suitable for aircraft mechanism, and the reliability assessment

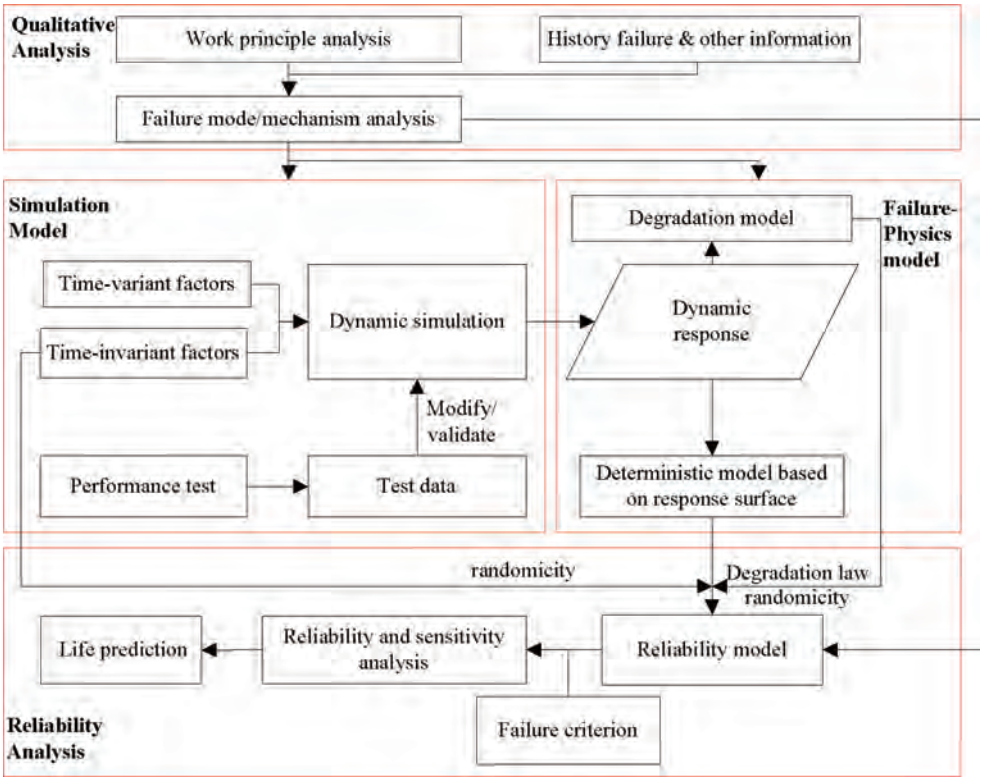


Figure 1. Time-variant reliability estimation framework based on physics of failure.

based on test method needs to spend too much time and cost. In order to solve these problems, time-variant reliability assessment framework based on failure physics was presented, as shown in Figure 1:

The main work of the time-variant reliability assessment based on failure-physics method consists of following aspects:

1. Failure modes and mechanism analysis

In this stage, failure modes and failure mechanism were obtained based on work principle analysis, history failure and other test information. Failure modes analysis is the foundation of reliability analysis, the common methods are Failure Modes Effect Analysis (FMEA) and Failure Tree Analysis (FTA). Where FMEA is a bottom-up method, which begins from components failure and follows certain logic to find out the effect to the system and then find out the crucial components and main failure modes. Its effect is very accurate at the level components but becomes progressively weaker as the fault effect propagates further away from the component and into the subsystem and system-level effects. While FTA is an up-down method, start form high level failure to find out the reasons which cause the fault.

2. Simulation model establishment

Mechanism is a mechanical system consists of several components and related joints. For some complex mechanisms, establishing and solving the dynamic models are very difficult. But the existing commercial software such as MSC ADAMS and LMS Virtural.Lab, present an easy way to model mechanism, and the dynamic results can be accurate if only the simulation model is reasonable. Therefore, this paper established the dynamics simulation model of mechanism in LMS Virtual. Lab, the model was modified and validated by the performance test, and then the model can provide accuracy results compared to real mechanism. Thus we can obtain the physical models which reflect the relationship between inputs and outputs of mechanism system.

3. Physics of failure model

Physics of failure model of mechanism contains performance function which reflects the relationship between inputs and outputs of mechanism system and the damage model which reflects the geometry and physical parameters change law with the work time. In the former step, the performance function is implicit, if use simulation model to analysis the reliability, every sample should be simulated in the model, which make it quite time consuming, for example, a single simulation cost 20 seconds, that is to say 10000 samples need 200000 seconds (about 56 hours). In order to obtain an explicit performance function, response surface method was used, this time, only a few samples need to be run in the simulation model, if bucher's sampling method is applied, the number of samples is $2n+1$, here, n is the number of random variables.

Damage in the mechanism of behaves as joint wear, plastic deformation, seal parts corrupt and aging, spring stress relaxation and so on. These damage models can be obtained by fitting the degradation data or by the damage model. Presently, many scholars have done a lot of research work in such domains (Wen Tsinghua University press, Su, Tianjin University press). In the reliability analysis process, we can directly choose the appropriate model to application.

4. Reliability assessment and lift prediction

For a certain mechanism, it is likely that two or more failure modes exists at the same time, and different failure modes may be in series, parallels or mix relations. To analysis reliability, proper reliability model needs to be established. Then bring failure criterion and variables' distribute parameters into the response surface, the reliability and sensitivity can be obtained, at last, based on the reliability result give a lift prediction of the mechanism. The operation time at which the reliability of the mechanism is lower than the required reliability is the lifespan of the mechanism.

2.2 *Time-variant reliability analysis method for mechanism*

Unlike structure's static response, dynamic responses of the mechanism changes with its configuration, so the mechanism performance can be described by a set of performance curves, namely $\boldsymbol{F} = \boldsymbol{F}(t)$. These curves can describe position, velocity, acceleration, load et al. In addition, time-variant factors makes the response quantity of different motion cycle not the same, namely $\boldsymbol{F} = \boldsymbol{F}(t,n)$, wherein t represents a runtime mechanism in single cycle, n represent the cycle number of the mechanism. In the reliability assessment process, it's not necessary to focus on the response of whole cycle time, but care about the response of the maximum value, the minimum value or a certain time value of each cycle, namely $\boldsymbol{F}(n) = \max(\boldsymbol{F}(t,n))$, $\boldsymbol{F}(n) = \min(\boldsymbol{F}(t,n))$ or $\boldsymbol{F}(n) = \boldsymbol{F}(t_c,n)|_{t = tx}$. When any $\boldsymbol{F}(n)$ exceed its allowed domain, the mechanism will failed.

In a generally way, mechanism performance can be represented by a set of response expressions, namely $F_i(\mathrm{n}) = \max(F_i(t,n))$, $i = 1,2,\ldots,s$. And the performance is mainly influenced by the m factors of x_j, $j = 1,2,\ldots,m$. Among which p factors are time-invariant and q factors are time-variant.

In this paper, linear response surface function was used to describe the relationship between the inputs and outputs of system, the form is

$$F(\bar{x}) = b_0 + \sum_{i=1}^{m} b_i x_i = B[1 \quad x], i = 1,2,...s \tag{1}$$

Assuming that time-invariant variables obey normal distribution, time-variant obey normal process and the distribution parameters of both time-invariant and time-variant factors are obtained, i.e.

$$x_j \sim N(u_j, \sigma_j^2), j = 1,2,...,p$$
$$x_j \sim N(u_j(n), \sigma_j^2(n)), j = 1,2,...,q$$
$$p + q = \boldsymbol{m}$$

Generally, $F_i(n)$ is a monotonic function, here assume it is a monotonic increasing function, namely, $F_i(n)$ increase with the mechanism performance degradation. Then the failure probability, reliability and lifespan of each failure mode can be represented as

$$P_{f_i}(n) = P\{F_i(n) > L_i\}$$

$$R_i(n) = P_r\{F_i(n) < L_i\} = \int_o^{N_i} f[F_i(n)]d[F_i(n)] \tag{2}$$

$$N_i = inf\{N | F_i(n) > L_i, \mathrm{n} \geq 0\}$$

where, L_i represents the failure threshold of each dynamic response and the inf{·} is the infimum function.

All the failure modes of the mechanism can be regard as in series, failure probability of the system can be written as

$$P_{sf} = \sum_{i=1}^{s} P_i - \sum_{1 \leq i < j}^{s} P_{i,j} + \sum_{1 \leq i < j < k}^{s} P_{ijk} - \cdots + (-1)^{s-1} P_{12\cdots s} \tag{3}$$

Previously, performances of mechanism system are treated independently, the correlations between different failure modes are been ignored (Leira et al. 2005, Franchin et al. 2003), namely, assume reg $\rho_{ij} = 0$. So the probability of the mechanism can be described as

$$P_f(n) = \sum_{i=1}^{s} P\{F_i(n) > F_{iC}\} = \sum_{i=1}^{s} P_i \tag{4}$$

However, all components in mechanism suffer a common load environment. Effects of Multi-factors coupling and multi-body coupling made failure modes in a mechanism dependent with each other, it means different failure modes have relations with each other. So the previous method which ignoring the correlations between different failure modes will lead to an overestimation of the system probability of failure and underestimate of the system reliability (Wang et al. 2007, Levitin 2001).

If correlations between different failure modes are been considered, for a mechanism have two or three failure modes, equation (4) can be written as:

For $s = 2$, $P_{2f} = P_1 + P_2 - P_{12}$ (5)

For $s = 3$, $P_{3f} = P_1 + P_2 + P_3 - P_{12} - P_{13} - P_{23} + P_{123}$ (6)

For functions of Equ. (1), the correlation coefficient ρ and covariance matrix C between different failure modes can be described as

$$\rho_{i,j} = \frac{Cov(F_i, F_j)}{\sqrt{DF_i} \cdot \sqrt{DF_i}} \tag{7}$$

$$C = \begin{bmatrix} \sigma_{F_1}^2 & \rho_{12}\sigma_{F_1}\sigma_{F_2} & \rho_{12}\sigma_{F_1}\sigma_{F_3} & \cdots \\ \rho_{12}\sigma_{F_1}\sigma_{F_2} & \sigma_{F_2}^2 & \rho_{12}\sigma_{F_2}\sigma_{F_3} & \cdots \\ \rho_{12}\sigma_{F_1}\sigma_{F_3} & \rho_{12}\sigma_{F_2}\sigma_{F_3} & \sigma_{F_3}^2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{8}$$

According to knowledge of probability theory, failure probability of k failure modes occur in the same time can be obtained through integrate joint density function of the multi-dimension random valuables in the failure domain,

For $k = 2$,

$$P_{ij} = \iint_G f(F_i, F_j) \mathrm{d}F_i \mathrm{d}F_j \tag{9}$$

For $k = 3$,

$$P_{ijk} = \iiint_G f(F_i, F_j, F_k) \mathrm{d}F_i \mathrm{d}F_j \mathrm{d}F_k \tag{10}$$

And we can also use multivariate normal cumulative distribution function in MATLAB to calculate the failure probability, the form is

$$P_{ij\ldots} = \mathrm{mvncdf}\ (\boldsymbol{L}, \boldsymbol{u}_F, \boldsymbol{C}) \tag{11}$$

Because distribute random parameters of time-variant valuable is variant with operation time, the failure probability calculated form equation (6) is the function of operation cycle. Assume that the reliability of the system is required to be high than Rc, then the lifespan N can be calculated as

$$N = inf\{T | R(n) > Rc, n \geq 0\} \tag{12}$$

3 FAILURE MECHANISM OF A LOCK MECHANISM

3.1 *Working principle of the lock mechanism*

The lock mechanism consists of eleven parts named 1 to 11, the closed state is shown in Figure 3. The opening process can be divided into three phases: (1) the hydraulic system work to push part 2, which makes part 3 rotate clockwise. At the same time, wheels assembled on part 3 shove part 4, make part 4 rotate clockwise until part 4 and part 5 apart; (2) When part 4 is open, wheels roll along the up surface of part 4 and shove part 5, at the same time, force transfer though part 6 to part 7 and part 7 rotate anticlockwise, until points A, B and C get in a line; (3) after that, with the effect of compress spring 11, part 8 open automatically, and with the spring force of 10, part 4 rotate anticlockwise and keep the lock in the open state. The closing process is in opposite to the opening process.

3.2 *Failure modes and mechanism analysis of lock mechanism*

There are too many potential failure modes in lock mechanism, in order to find out the most probable failure modes, dynamic simulation was implemented, dynamic responses of main structure are show in Figure 2.

Obviously, max value of contact force occurs in the first period and the max value of link and driving force occur in the second period.

Force analysis for the first period is shown in Figure 3, part 4 bear five forces of F_{n1}, F_{f1}, F_{n2}, F_{f2}, F_s, among which, only F_{n1} help to open the lock and all others prevent it from opening. Here F_{n2} is determined by lock hook force F_p, F_{1n} is supplied by F_L, and F_{f1} is determined by F_{n1} and friction

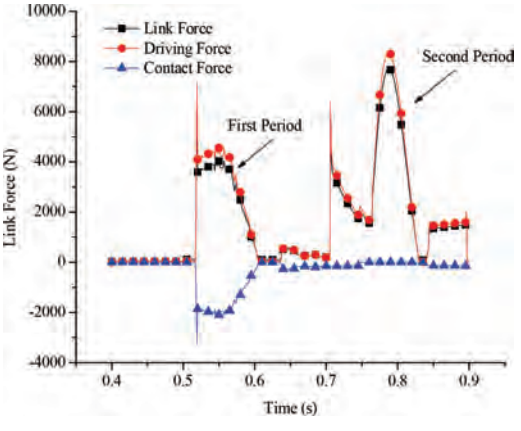


Figure 2. Dynamic simulation results.

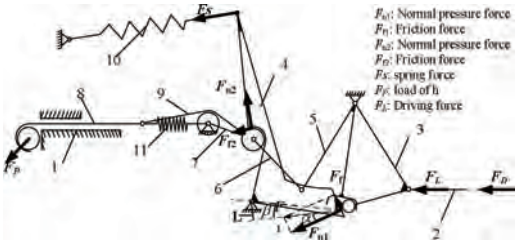


Figure 3. Force analytical of first period during opening phase.

coefficient f_1. From the direction of F_1, which is the composition force of F_{n1}, F_{f1}, it can be seen that with the increase of f_1, the arm of F_1 to the rotation axis decreased when the friction coefficient f_1 increased. And the force arm decrease to very small value or even zero if f_1 increases to a certain extent. At the same time, the wear and impact lead to slope surface of the key roughness and unevenness, which will directly affect the magnitude of the positive pressure and friction force, and affect the direction of the composition force as well.

In the performance test, we record the pressure of the hydraulic pressure in the cylinder actuator and the force in the link, after several operation cycles we found that the driving force and the link force increased compared to the beginning trial. Then the lock mechanism is carefully examined and found that there are wear and extrusion marks in the slope surface of the part 4, other parts without obvious change. Damage modality of part 4 is shown in Figure 5.

In the second period of lock opening process, part 7 contrarotate and make point C move from underside to upside of the line made of point A and point B, during this process, part 7 suffer F_1 and F_2, force analytical model is shown in Figure 4. Here F_1 supplied by hydraulic system act as the driving force and F_2 determined by Fp act as the drag force. As point C gradually close to the line form underside, distance between point A and point C increased, this result in Fp increased because much large force needed to conquer the deformation of the port. And the radius of the part 8 also influences F_P, if F_p is too large, the hydraulic pressure will not large enough to open the lock, this time the lock mechanism will get seizure.

According to what have been said, the lock mechanism have two main potential failure modes: 1) in the first period, the roller load is become larger and larger, when the maximum load exceed the terminal load, which is the maximum load the structure can bear, the structural damaged, manifested as roller plastic bending or link buckling; 2) in the second period, link buckling or lock mechanism

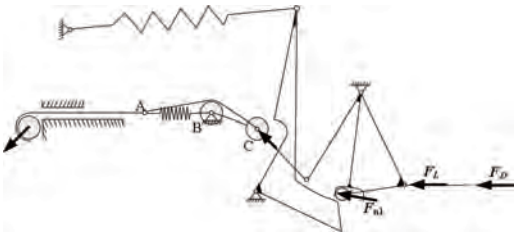


Figure 4. Force analytical of second period during opening.

Figure 5. Damage modality.

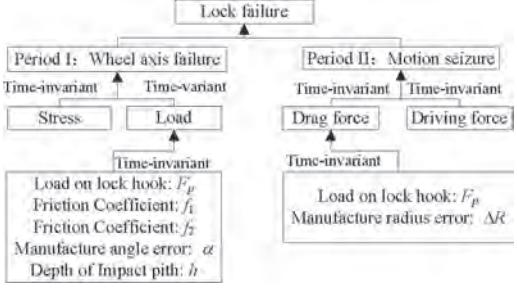


Figure 6. Failure mechanism of the lock mechanism.

get seizure. The failure mechanism is shown in Figure 6.

To sum up, the failure modes of lock mechanism is roller plastic bending in the first period and mechanism seizure in the second period.

Further, as Figure 7 shows, with the effect of degradation factors, characteristic parameters of the lock mechanism will increase with the work time, and make the reliability degradation.

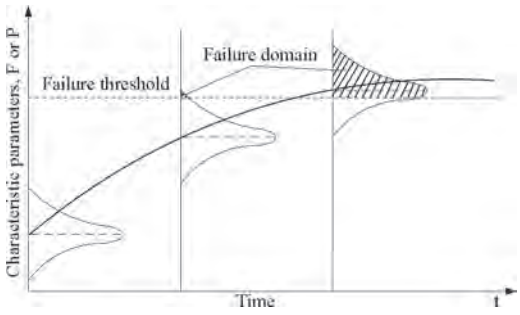


Figure 7. Sketch map for performance degradation.

4 TIME-VARIANT RELIABILITY ANALYSIS OF THE LOCK MECHANISM

4.1 *Parameter characterization of local damage*

When the lock begins to open, wheel move along X direction quickly in the hydraulic action, and then wheel assembled on the axis will impact part 4, since hardness of axis is greater than part 4, after repeated impact, surface of part 4 will generate an impact pit. The curvature of pit is approximate to radius of wheel R, the sketch map of damage modality is shown in Figure 8. Here h represents the depth of impact pit, and the dimension is equal to difference of R and D.

4.2 *Reliability model of lock mechanism and distribution parameters of variables*

4.2.1 *Reliability model of lock mechanism*

Section 3.2 shows that potential failure modes of lock mechanism is roller failure, manifested as roller plastic bending in the first period and mechanism seizure in the second period. Analysis result shows the critical force F_c lead to roller plastic bending is 8346 N and the maximum driving force is 10000 N. In order to simply the analysis process, the dispersion of the value has been ignored. Then failure criterion of the failure mode can be described as

$$g_1(\vec{x},n) = F_{C1} - F_1(\vec{x},n) < 0 \tag{13}$$

$$g_2(\vec{x},n) = F_{C2} - F_2(\vec{x},n) < 0 \tag{14}$$

According to Equ. 4, the failure probability of the lock mechanism can be written as

$$p_f(n) = P\{g_1 < 0\} + P\{g_2 < 0\} - P\{g_1 < 0 \cap g_2 < 0\} \tag{15}$$

In the above equation, $F_1(\vec{x},n)$ is the max contact force in the nth cycle and $F_2(\vec{x},n)$ is the force needed to open lock in the nth cycle.

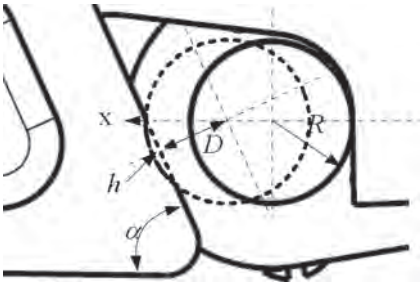


Figure 8. Depiction for the damage modality.

4.2.2 *Distribution parameters of variables*

Failure analysis of the lock in section 3 shows four factors affect the maximum roller force in the open process, they are load of hook F_p, friction coefficient between key and roller, slope of key f_1, friction coefficient between key and rocker f_2. And the key's slope α. Assuming all variables obey normal distribution.

Load of hook is a random variable affected by the flight state of the aircraft, which has no relation with the lock operation times.

The performance test shows there's almost no wear between key and rocker, so the friction coefficient between key and rocker can be regarded as a time-invariant variable.

4.3 *Performance function modeling based on response surface method*

For complex mechanism, it is unpractical to establish and solve the dynamic equations and the through test to obtain the dynamic responses and their relationships between design parameters because of the limitation of time and costs. However dynamic simulation is a good choice to solve these difficulties.

Establish the lock simulation model in LMS Virtual.Lab platform, then use test data to modify and the simulation model until the results are precise enough, and then effective simulation model is obtained.

According to the distribute parameters of valuables, adopt Bucher sampling method to get the samples, then simulation all the samples in the simulation model and obtain correlative results.

From equation (1), we get the performance functions of max contact force in the first period and the max driving force in the second period like

$$F_1(\mathbf{x},n) = B_1[1 \quad f_1 \quad f_2 \quad \alpha \quad F_p \quad h]^T \tag{16}$$

$$F_2(\mathbf{x},n) = B_2[1 \quad \Delta R \quad F_p]^T \tag{17}$$

Thereinto,

$$B_1' = \begin{bmatrix} -6295.1208 \\ 4418.1333 \\ 8495.65 \\ 78.8655 \\ 363.0785 \\ 191.5783 \end{bmatrix} \quad B_2' = \begin{bmatrix} 8291.466 \\ 132.1717 \\ 34.691 \end{bmatrix}$$

From B_1, it can be seen that all the coefficients are positive except the constant term. It means contact force increases with the increase of each parameter. B_2 shares the same conclusion. The conclusion is accord with our intuitive feelings.

Comparison is made for measured values and fitted values in Figure 9, it can be seen that the two kinds of values almost superposition, that is to say, the linear response surface is precision enough to replace the simulation model.

4.4 *Reliability analysis and life-span prediction*

Substitute the distribute parameters into limit state equations, the mean value and standard deviation can be obtained.

$$u_{F1}(n) = 2128.1 + 0.4418n + 0.0192n^{1/3}$$

$$\sigma_{F1}(n) = \sqrt{2.0189 \cdot 10^6 + 13.9868n + 0.1748n^2 + 0.00367n^{2/3}}$$

$$u_{F2}(n) = 8464.921$$

$$\sigma_{F2}(n) = 881.2249$$

The driving force and contact force distribution in different work cycles are show in Figure 10. From which it can be seen that driving force distribution is constant during lifetime, while contact force distribution changes constantly during lifetime, the mean value increase obviously with the increase of the work cycles and the deviation also increase tardily.

According to the reliability analysis method represented above, reliability of lock mechanism was obtained, as Figure 11 shows, reliability of failure mode II (motion seizure in the second period) is a constant value of 0.9592 during lifetime, while reliability of failure mode I (wheel axis damaged) decreased with the increase of the work cycles, the change is much obviously after 1000 work cycles.

Another conclusion is that in the beginning of service time, failure mode II plays as the main failure mode and after 1000 work cycles, failure mode I plays as the main failure mode.

Table 1. Summary statistics of the basic variables in the modal.

Variables	Mean, u	Standard deviation, σ	Max	Min	Average
f_1	0.04+0.0001*n	0.1*u	0.1	0.04	0.07
f_2	0.2	0.0166	0.15	0.25	0.2
α/deg	60	0.3333	61	59	60
ΔR/mm	0	0.00667	0.02	–0.02	0
F_p/kN	5	0.3333	6	4	5
h/mm	0.0001*$n^{1/3}$	0.1*u	0.3	0	0.15

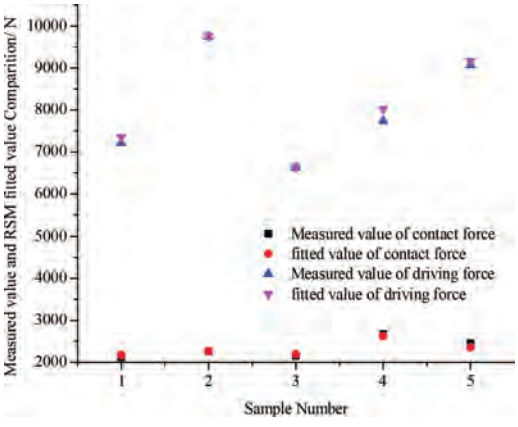


Figure 9. Comparison of measured values and fitted values.

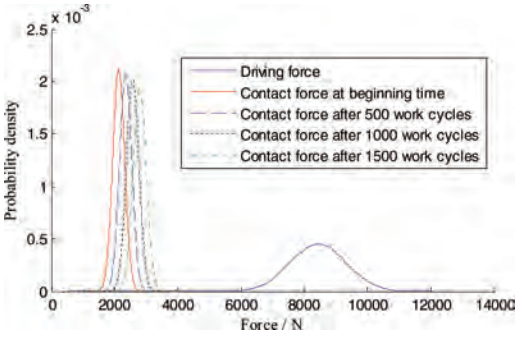


Figure 10. Diving force and contact force distribution.

From system reliability curve in Figure 11, it can be seen that lock reliability is 0.9592 at the beginning time, and decreased slowly in the following 1000 work cycles. Then, the lock reliability decreased quickly. According to the requirement that reliability much higher than 0.9, the reliable life is 5027 work cycles.

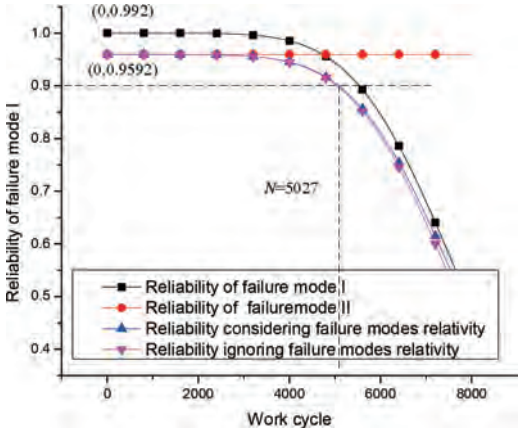


Figure 11. Degradation law of lock reliability.

5 CONCLUSIONS

1. In order to solve problems exists in the traditional statistic method and test method for lock mechanism reliability analysis, a time-variant reliability analysis method was presented, which is based on virtual simulation and with component damage into consideration.
2. Based on dynamic simulation results, failure modes and failure mechanism was analyzed, obtained that two main potential failure modes of lock mechanism are wheel axis failure and motion seizure, then analyzed the failure mode influence factors.
3. Performance functions were found through RSM, time-variant reliability of lock mechanism was evaluated with lock key damage and randomicity of variables into consideration. The results show lock reliability is 0.9592 in the early days. After 1000 operation times, its reliability decreases to 0.9451. As the reliability is required to be higher than 0.9, the reliable operation time is 5027 times.

ACKNOWLEDGEMENTS

This work was financially supported by National Natural Science Foundation of China (No. 51675428) and Central University Innovation Team Support Project (No. 310822173702).

REFERENCES

Ballu A, Plantec J.Y., Mathieu L. Geometrical reliability of overconstrained mechanics with gaps, *CIRP Annals-Manufacturing Technology*, 2008; 57: 159–162.

Erkaya S, I. Uzmay. Effects of balancing and link flexibility on dynamics of a planar mechanism having joint clearance. *Scientia Iranica* 2012; 19(3): 483–490.

Flores P, J. Amrosio, J.C.P. Claro, H.M. Lankarani, C.S. Koshy. A study on dynamics of mechanical systems including joints with clearance and lubrication. *Mechanism and Machine Theory* 2006; 41: 247–261.

Flores P. Modeling and simulation of wear in revolute clearance joints in multibody systems. *Mechanism and Machine Theory* 2009; 44: 1211–1222.

Franchin P, Lupoi A, Pinto PE. Seismic fragility of reinforced concrete structures using a response surface approach. *Journal of Earthquake Engineering*, 2003, 7: 45–77.

Gu CH, Sheng YX, Zhang SL. Analysis of mechanism reliability of landing gear uplock system. *Journal of Beijing University of Aeronautics and Astroautics* 1995; 21(4): 18–23.

Jin G, David Mattews, Fan YW, Liu Q. Physics of failure-based degradation modeling and lifetime prediction of the momentum wheel in a dynamic covariate environment. *Engineering Failure Analysis* 2013; 28: 222–240.

Jungkeun Rhee, Adnan Akay. Dynamic response of a revolute joint with clearance. *Mechanism and Machine Theory* 1996; 31(1): 121–134.

Leira BJ, Holma T, Herfjor K. Application of response surfaces for reliability analysis of marine structures. *Reliability Engineering and System Safety*, 2005, 90: 131–139.

Levitin G. Incorporating common-cause failure into no repairable multi-state series-parallel system analysis. *IEEE Transactions on reliability*, 2001, 5(4): 380–388.

Ming-June Tsai, Tien-Hsing Lai. Accuracy analysis of a multi-loop linkage with joint clearances. *Mechanism and Machine Theory* 2008; 43: 1141–1157.

OuYang ZH, Yi MD. Study on evaluation method of wear reliability for lock mechanism system of cargo-bridge. *Acta Aeronautica et Astronautica Sinica* 1994; 15(3): 324–330.

Peter Ravn. A continuous analysis method for planar multibody systems with joint clearance. *Multibody System Dynamics* 1998; 2: 1–24.

Saad Mukras, et al. Analysis of planar multibody systems with revolute joint wear. *Wear* 2010; 268: 643–652.

Su D.D. Stress relaxation and prevention of spring (Materials). *Tianjin university press*, 2002, 9.

Wang JG, Zhang JF, Du XP. Hybrid dimension reduction for mechanism reliability analysis with random joint clearance. *Mechanism and Machine Theory* 2011; 46: 1396–1410.

Wang MQ, Chen ZY. Analysis of time variant reliability for gear with multi-failure mode. *Journal of mechanical transmission*, 2011, 35(4): 50–53.

Wang Z, Xie L.Y., Zhang J.Y. Reliability model of system with common cause failure under repeated random load. *ACTA AERONAITICA ET ASTRONAUTICA SINICA*, 2007, 28(Z1): 116–120.

Wen S.Z., Huang P. Principles of tribology. *Tsinghua university press*, 2008,

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

A metal-oxide-semiconductor devices reliability assessing method based on physics of failure

Hantian Gu, Ming Zhu, Wei Zhang, Lei Zhang, Hengjing Zhu & Min Tang
China Aerospace Components Engineering Center, CASC, Beijing, P.R. China

ABSTRACT: With the development of Metal-Oxide-Semiconductor (MOS) devices, reliability is becoming a key differentiator in a competitive market. Considering the different stress types and levels during application, an algorithm is proposed to evaluate the reliability of MOS devices under complicated operation conditions. This algorithm is based on the theories of Physics of Failure (PoF). Failure modes, mechanisms and effects analysis is conducted to achieve the potential failure mechanisms and physics models. Then through modeling and simulation, thermal, mechanical and electrical parameters of MOS devices under different conditions are obtained. Using the physics models, cumulative damage theory and random sampling algorithm, the matrix of time to failure of each unit is taken. At last, competing failure model is applied to acquire the time to failure of MOS devices in working condition. The algorithm provides a new approach based on PoF for evaluating the reliability of MOS devices under complex environments.

1 INTRODUCTION

Reliability is an important requirement for almost all users of integrated circuits (ICs). Scaling for enhanced performance and cost reduction has pushed existing MOS devices materials much closer to their intrinsic physical and reliability limits. Besides, A newly developed semiconductor technology node cannot be released to user without going through a rigorous work of reliability evaluation (Saeidi et al., 2013, Hava et al., 2013). All of these demonstrate that the reliability evaluation for MOS device is necessary and it is facing a huge challenge.

For reliability evaluation, a common and realistic practice is to perform accelerating tests under conditions that are much more severe than those under operation conditions. The severe conditions in accelerating tests generally mean a much higher stressing temperature or stressing current or both than the operation conditions. After obtaining the time-to-failure and failure distribution on severe conditions, the evaluation of reliability under normal operation is realized through accelerating models by extrapolation. However, there are two disadvantages in this approach. The first one is only one failure mechanisms are considered during accelerating tests, so the time-to-failure under normal operation by extrapolation is only associated with that mechanism. The second one is that the devices are always operating in a constant stress level. It is not match with operating conditions in actual. In other words, the results are qualitative and they cannot be used to evaluate the lifetime of devices under actual environment.

Therefore, a new algorithm is put forward to evaluate the reliability of MOS devices in this paper, which consider different failure mechanisms and stressing conditions in the entire life of devices.

In the next second of the paper basic theories associated with the algorithm are introduced. In the third part the procedure of reliability evaluation for MOS devices will be given and the detailed algorithms will be told. The fourth part of this paper is a case study. In this section we will describe each step in details combined with a simple MOS device and make a discussion with the results of simulation. At last, conclude this paper and look into the future of this methodology.

2 BASIC THEORIES

This methodology that put forward in this paper is based on Physics-of-Failure (PoF). The concept of PoF, also known as Reliability Physics, involved the use of degradation algorithms that describe how physical, chemical, mechanical, thermal, or electrical mechanisms evolve over time and eventually induce failure (Borgarino et al., 2001, Pecht and Gu, 2009). In this paper, PoF is used to establishment the relationship between the operation environment and the degradation of performance

parameters of MOS devices. The used failure modes include three groups according to their origin: (i) failure modes associated with chip; (ii) failure modes resulting from leads; (iii) failure modes due to overstress. In this paper, we analyze the failure modes and mechanisms which have higher probability of occurrence and greater impact on the performance of MOS device at first. Then we calculate the stress of device under operation conditions to find the reasons that possibly lead to failure through simulation analysis. At last, *TTF* is predicted by PoF models. Hence, all of the work in this paper is about PoF.

3 ALGORITHM

3.1 *General algorithm*

As Figure 1, the input of this algorithm consists of three parts: parameters of structure, material properties, potential failure mechanisms and local stress under operation conditions.

1. Parameters of structure include dimensions of packaging and die;
2. Material properties describe the deformation of material under load;
3. Failure Mode, Mechanism and Effects analysis (FMMEA) helps to understand the potential failure mode, mechanisms and models of MOS devices during under operation conditions.
4. Local stress under operation conditions will be obtained from simulation analysis, including thermal analysis, random vibration analysis and electronic parameters analysis.

According to inputs, the main part of algorithm is developed with the work of random sampling, cumulative damage and data processing. Considering the uncertainty of size, properties of materials and operation conditions in practice, we treats the uncertain parameters as random variables described by some appropriate statistical distribution, which may be sampled using Monte Carlo methods. After sampling we get the matrix of the uncertain parameters for failure mechanisms. In practice, the MOS devices experience variable amplitude loading. And cumulative damage theory allows us to calculate the damage for each cyclic loading. Here we obtain the matrix of *TTF* of simulation unit for each critical failure mechanisms under operation conditions. At last, we need to calculate the *TTF* of MOS device according to the matrix of *TTF* of simulation unit by data processing.

In the following section, we will discuss this methodology in detail and give the algorisms in each phase.

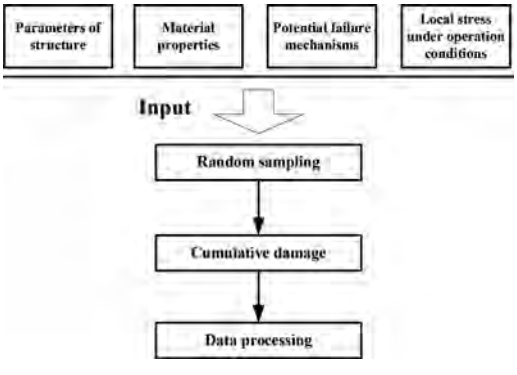


Figure 1. The workflow of reliability evaluation by simulation.

3.2 *Input information*

The input information is divided into three classes:

1. Parameters of structure include dimensions of packaging and die;
2. Material properties describe the deformation of material under load;
3. Failure Mode, Mechanism and Effects analysis (FMMEA) helps to understand the potential failure mode, mechanisms and models of MOS devices during under operation conditions.
4. Local stress under operation conditions will be obtained from simulation analysis, including thermal analysis, random vibration analysis and electronic parameters analysis.
5. Parameters of structure

For structure parameters, different approaches are selected to obtain. The following table shows the approaches to extract them:

3.3 *Material properties*

Material properties include the density, elastic modulus, poisson's ration, conductivity, coefficient of thermal expansion and so on. They are easy to get from handbook.

Table 1. Approaches of parameters extraction.

Parameters	Sources
Parameters of packaging	Datasheets Measurement
Parameters of die	Design files of layout of chip Measurement

3.4 *Potential failure mechanisms*

Considering the interaction among performance, physical characteristics of products, materials and environment stress, failure mode, mechanism and effects analysis (FMMEA) can be used to determine the potential failure mechanisms and models. Besides, the priority of failure mechanisms can be determined based on the severity and probability of occurrence.

After FMMEA, the critical failure mechanisms, which mean higher probability of occurrence and greater impact on the performance of the specific MOS device, and models are determined. They are the focus of analysis and calculation in the following phases.

3.4.1 *Local stress under operation conditions*

Since the parameters of operation conditions are the environment which the system working in and it is different from the MOS devices' working environment, the simulation modeling and analysis are necessary. Local stress of MOS devices will be given by simulation modeling and analysis. In this phase, there are three simulation are indispensable: thermal analysis, random vibration analysis and electrical parameters analysis.

- Thermal analysis

Thermal analysis is used to obtain the local temperature distribution under specific operation conditions. The solution procedures in this paper are based on computational fluid dynamics (CFD) techniques, which are concerned with the numerical simulation of fluid flow, heat transfer and related processes.

- Random vibration analysis

Random vibration analysis is used to obtain the equivalent stress, equivalent strain and model under specific operation conditions. The solution procedures in this paper are based on finite element analysis (FEA) techniques, which are concerned with the force-balance equation, deformation compatibility equation and material properties.

- Electrical parameters analysis

Electrical parameters analysis is used to obtain the related electrical parameters of MOS structure under specific bias voltage, such as the current density for electromigration and oxide field for TDDB.

3.5 *Algorithm of random sampling*

Considering the uncertainty of size, properties of materials and operation conditions in practice, we treats the uncertain parameters as random variables described by some appropriate statistical distribution, which may be sampled using Monte Carlo methods. Here, the parameters of PoF models are subject to various kinds of uncertainty, which may include: uncertainty of material properties, the effect of variations in manufacturing process and the uncertainty associated with stochastic fluctuations of operational stresses. In order to generate the samples fitting the given distribution, the algorism for random sampling is,

1. Calculate the inverse function of cumulative distribution function (CFD), $F^{-1}(x)$;
2. Generate samples of uniform distribution in the interval (0,1), $\vec{a} = (a_1, a_2, \ldots, a_n)$;
3. Calculate $\vec{x} = F^{-1}(\vec{a})$, so vector $\vec{x}$ is the samples that we wanted.

The matrix of the uncertain parameters for specific failure mechanism is,

$$\alpha = \begin{vmatrix} \alpha_{11}, \alpha_{12} \cdots \alpha_{1n} \\ \alpha_{21}, \alpha_{22} \cdots \alpha_{2n} \\ \cdots \\ \alpha_{1m}, \alpha_{m2} \cdots \alpha_{mn} \end{vmatrix} \tag{1}$$

Here, α_{mn} means the nth sample of uncertain parameter, α_m.

3.6 *Algorithm of cumulative damage*

In practice, MOS devices experience not a few of different stress levels during lifetime, but a life profile which the different stress levels are arranged in a certain order. Calculating the *TTF* under different stress levels respectively cannot descript the actual situation in application. So it is necessary to develop cumulative damage analysis. In this paper, we introduce two approaches: acceleration factor method and cumulative damage rule method.

- Acceleration Factor Algorithm (AFA)

Acceleration Factor is a multiplier that relates a product's life at an accelerated stress level to the life at the use stress level. AFA is suitable for failure mechanisms of electromigration, hot carrier injection, gate-oxide time dependent dielectric breakdown and so on. Here, in order to think about the influence of life profile with several stress levels, we convert it to a new profile with only one stress level, which we call it reference stress level using acceleration factor at first. In terms of temperature, the algorithm is as following,

1. Start;
2. Input the profile of temperature with two vectors $\vec{t} = [t_1, t_2, t_3, \cdots, t_n]$ and $\vec{T} = [T_1, T_2, T_3, \cdots, T_n]$, reference temperature T_0, $t\_trans = 0$;
3. Set $i = 1$;
4. If $T_{i+1} = T_i$, $t' = (t_{i+1}) \times \mathrm{AF}(T_i)$, otherwise, $t' \int_{t_i}^{t_{i+1}} AF(T(t)d(t))$. In the formula, $AF(T)$ is

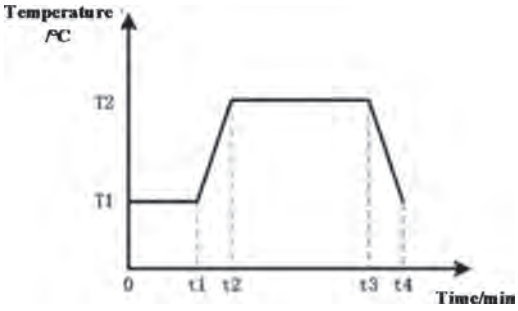


Figure 2. A profile of temperature used to help describe AFM.

the is the acceleration factor of temperature T which has been transferred to T_0;

5. Calculate $t\_trans = t\_trans + t'$;
6. $i = i + 1$;
7. Repeat the step 4) to 6) until $i = n - 1$.

Then calculate the *TTF* under operation condition of new profile and it is easy to transfer it to the *TTF* under original profile.

For example, a profile is as following,

In this profile, there are two different stress levels, *T1* and *T2*, and the dwell is *t1* and (*t3* – *t2*), respectively. The two stages with temperature change are (*t2* – *t1*) and (*t4* – *t3*). Here, assuming that we transfer them to the stress level of *T0*, the formula is,

$$t\_trans = t1 \times AF(T1) + \int_{t1}^{t2} AF(T(t))d(t) \\ (t3 - t2) \times AF(T2) + \int_{t3}^{t4} AF(T'(t))d(t) \quad (2)$$

Here, *t_trans* is the equivalent time of original profile under stress level of *T0*, *AF(Ti)* is the acceleration factor of temperature *Ti* which has been transferred to *T0*, *T(t)* and *T'(t)* are the function of *T* vs. *t*.

Calculate the *TTF* under temperature *T0*, than *TTF* under original profile is,

$$TTF_o = \left[\frac{TTF}{t\_trans}\right] \times t4 \quad (3)$$

Here, [*] mean the number that is nor more than *; *TTF* is time-to-failure of specific MOS structure under *T0* in terms of certain mechanism; TTF_o is time-to-failure of specific MOS structure under original profile.

- Cumulative Damage Rule Algorithm (CDRA)

In terms of mechanisms such as random vibration fatigue, CDRA is more appropriate to solve the problem of the cumulative damage. Here, considering the order and interaction effects of variable amplitude loading, the Corten-Dolan approach is determined to calculate the *TTF* of failure mechanisms. The formula is,

$$N_g = \frac{N_1}{\sum_{i=1}^{m} \alpha_i \left(\frac{\sigma_i}{\sigma 1}\right)^d} \quad (4)$$

Here, N_g is the cycles to failure for variable amplitude loading; N_1 is the cycles to failure for the maximum amplitude loading; σ_i is the equivalent stress of leads for the i^{th} amplitude loading, which has obtained from the random vibration analysis of phase 4; σ_1 is the equivalent stress of leads for the maximum amplitude loading; d is a constant that is decided by materials, m is the number of stress level, α_i is a percentage that can given from,

$$\alpha_i = \frac{\sigma_i}{\sigma_1 + \sigma_2 + \cdots + \sigma_m}, i = 1, 2, \cdots, m \quad (5)$$

According to the life profile of random vibration, N_g is easy to transfer to *TTF*.

Then we can calculate the *TTF* using the algorism of PoF models and cumulative damage theory (Hall and Strutt, 2003, Haggag et al., 2000).

Assuming a specific PoF model can be expressed as,

$$TTF = f(S, M, E) \quad (6)$$

Here, *TTF* is the short for the time-to-failure, S, M and E means parameters of structure, properties of materials and operation conditions, respectively. So the vector of *TTF* of simulation unit for specific failure mechanism under certain stress level is,

$$\vec{TTF} = f(\vec{S}, \vec{M}, \vec{E}) \quad (7)$$

Calculate the *TTFs* for each critical failure mechanisms obtained from phase 2, the matrix is,

$$TTF = \begin{bmatrix} TTF_{11}, TTF_{12}, \cdots, TTF_{1n} \\ TTF_{21}, TTF_{22}, \cdots, TTF_{2n} \\ \cdots \\ TTF_{k1}, TTF_{k2}, \cdots, TTF_{kn} \end{bmatrix} \quad (8)$$

Here, k is the k^{th} critical failure mechanism of MOS device, n is the number of samples.

According to matrix (8), choose different cumulative damage approaches for different failure mechanisms to calculate and the matrix of the

TTF of simulation unit for each critical failure mechanisms under life profile is,

$$TTF_o = \begin{bmatrix} TTF_{o11}, TTF_{o12}, \cdots, TTF_{o1n} \\ TTF_{o21}, TTF_{o22}, \cdots, TTF_{o2n} \\ \cdots \\ TTF_{ok1}, TTF_{ok2}, \cdots, TTF_{okn} \end{bmatrix} \quad (9)$$

3.7 *Algorithm of data processing*

Now the analysis objects are simulation units, which are parts of MOS devices, such as MOS structure. In order to evaluate the *TTF* of MOS devices, the following work is necessary.

- *TTFs* for single mechanism to *TTFs* for multiple mechanisms

At first, we need to transfer the *TTF* vectors for single mechanisms of certain simulation unit to vector for multiple mechanisms of certain simulation unit.

The algorithm for distribution fitting and testing is as follows:

1. Start;
2. Set $i = 1$;
3. Extract the vector $TTF_o(i,:) = [TTF_{oi1}, TTF_{oi2}, \ldots, TTF_{oin}]$ from matrix (9);
4. Assume that data of vector obey certain distribution;
5. Calculate distribution parameters $\theta i = [\theta i_1, \theta i_2, \theta i_3]$;
6. Hypothesis testing and repeat step 4) and 5) until the null hypothesis is accepted.
7. $i = i + 1$;
8. Repeat step 3) to 7) until $i > k$;
9. Then we get the matrix of distribution parameters of simulation unit,

$$\theta = \begin{bmatrix} \theta_{11}, \theta_{12}, \theta_{13}, \\ \theta_{21}, \theta_{22}, \theta_{23}, \\ \cdots \\ \theta_{k1}, \theta_{k2}, \theta_{k3}, \end{bmatrix} \quad (10)$$

   here if the number of distribution parameters of some failure mechanisms is p and $p < 3$, the parameters $\theta_a = 0 (q = p + 1, \cdots 3)$;
10. Set $j = 1$;
11. Random sampling according to distribution parameters $\theta(i,:) = [\theta_{i1}, \theta_{i2}, \theta_{i3},]$ and we get vector $t_{mj} = [t_{mj1}, t_{mj2}, \cdots t_{mjk}]$;
12. Competing failure model help to get the minimum value, $t'_m = [t_{mj1}, t_{mj2}, \cdots t_{mjk}]$;
13. $j = j + 1$;
14. Repeat step 11) and 12) until $j > n$;
15. Now we get the *TTF* vector for multiple mechanisms of certain simulation unit, $t_m = [t_{m1}, t_{m2}, \cdots t_{mn}]$.
16. End.

- *TTFs* for each simulation unit to *TTFs* for MOS device

In order to obtain the TTFs for MOS devices, repeat algorithm listed above. At last, we achieve the vector of *TTF* for MOS device,

$$TTF_d = (TTF_{d1}, TTF_{d2}, \cdots, TTF_{dh},) \quad (11)$$

The last work is to distribution fitting and testing for vector TTF_d. The vector is sorted into ascending order and fitted to a suitable distribution to obtain the TTF of MOS device. Assuming that the vector is fitted a distribution of f(t), the *TTF* of MOS device is,

$$TTF = \int_0^{+\infty} tf(t)dt \quad (12)$$

4 CASE STUDY

In this paper, we select a MOS device with plastic dual inline-pin (DIP) package which provide the system designer with direct implementation of the NOR function with the weight of 1.1 g and power dissipation of 200 mW. In the following, we will depict the process of reliability evaluation by simulation in detail with this device.

4.1 *Input information*

4.1.1 *Parameters of structure*

At first, it's necessary to extract the parameters of structure. According to datasheet, design files and, they are listed as Table 2.

1. Structural parameters

4.1.2 *Material properties*

This part include the parameters about weight, power dissipate and material properties and so on.

According to the results of FMMEA, the potential failure mechanisms which have a greater impact on the performance of the specific MOS device include: solder joint thermal fatigue,

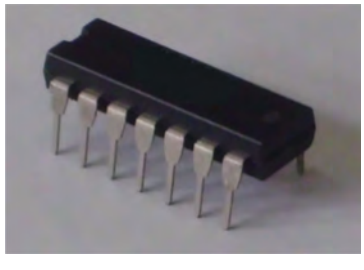


Figure 3. The MOS device which is applied in case study.

Table 2. Structural parameters for MOS device.

Packaging parameters	MATERIAL	Plastic
	LENGTH	19.3 mm
	WIDTH	6.2 mm
	THICK	3.3 mm
Lead parameters	MATERIAL	Copper alloy
	I/O	14
	PITCH	2.4 mm
	L_3	0.6 mm
	R	0.3 mm
	L_2	2.4 mm
	L_1	3.2 mm
	t	0.2 mm
	W_2	1.7 mm
	W_1	0.5 mm
Die parameters	DIE_LENGTH	4.3 mm
	DIE_WIDTH	3.5 mm
	DIE_THICKNESS	0.2 mm
	WIDTH OF INTERCONNECTS	1μm
	THICKNESS OF INTERCONNECTS	0.8μm
Solder parameters	Solder Height	1.6 mm
	Solder Joint Bond Area	1.58 mm^2

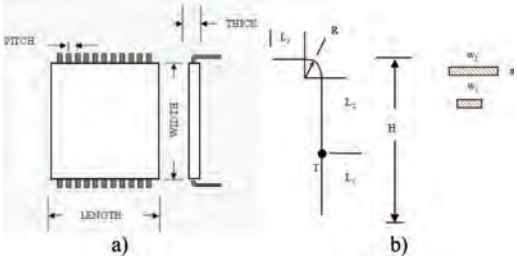


Figure 4. This figure illustrates the structural parameters: a) for packaging parameters and b) for lead parameters.

random vibration fatigue, corrosion, gate-oxide time dependent dielectric breakdown (TDDB), electromigration, hot carrier injection (HCI) and shock. So in the following steps, we focus on these seven failure mechanisms and evaluate the reliability with them. In Table 4, the failure mechanisms and PoF models are listed.

4.1.3 *Local stress under operation conditions*

At first, we need to introduce the operation conditions.

The MOS device is located in the middle of printed circuit board (PCB) with the dimensions of 120 mm × 80 mm × 2 mm. The operation conditions consist of three types of stresses, ambient environment, power spectral density of random vibration and relative humidity. The profile of

Table 3. Material properties of MOS device.

Plastic of Packaging	Density	1206 kg/m^3
	Elastic Modulus in X direction	15900 MPa
	XY Poisson's Ratio	0.25
	Conductivity	0.67 W/m*°C
	Coefficient of Linear Thermal Expansion	15e-006°C
Copper Alloy (C197)	Density	8360 kg/m^3
	Elastic Modulus in X direction	118410 MPa
	XY Poisson's Ratio	0.3
	Conductivity	0.67 W/m*°C
	Coefficient of Linear Thermal Expansion	16.8e-006/°C
	Yield Strength	405 MPa
	Tensile Strength	450 MPa
	Activation Energy	1.64 eV
FR4	Density	1938 kg/m^3
	Elastic Modulus in X direction	17200 MPa
	XY Poisson's Ratio	0.11
	Conductivity	0.2 W/m*°C
	Coefficient of Linear Thermal Expansion	17.6e-006/°C
	Tensile Strength	276 Pa
Impact Amplification Factor		1.0
Activation Energy of Interconnects		0.59 eV
Thermal Activation Energy of MOS Structure (TDDB)		0.1 eV
Activation Energy (HCI)		0.05 eV

MOS device during lifetime is as Figure 5, Table 5 and Table 6.

The relative humidity is always 20%.

According to the parameters of structure and operation conditions, the CAD, CFD and FEA models are as Figure 6.

We develop researches on thermal analysis, random vibration analysis and electrical parameters analysis and results are as Table 7.

4.2 *Random sampling*

Considering the uncertainty of parameters, monte carlo approach is widely used. Now we give the distribution types of the different parameters as Table 8.

In following section, we will introduce how to develop the monte carlo simulation to obtain the vector of *TTF* for electromigration.

At first, get the 10000 samples of uncertain parameters (W, d, T and j) by monte carlo approach as following,

$$\alpha \begin{bmatrix} 0.790114832, 0798267848, 0.79834634, 0.800647454, \ldots, 0.799123104 \\ 0.981133085, 1.007929821, 1.008361467, 1.000597051, \ldots, 1.015567806 \\ 55.84119123, 56.5393662, 55.23497947, 56.09943779, \ldots, 55.60634757 \\ 0.298663927, 0.302170728, 0.301507547, 0.299965692, \ldots, 0.299706913 \end{bmatrix}$$

Table 4. PoF models for potential failure mechanisms.

Failure mechanism	PoF model	
	Formular	Remark
Shock	$Z_{allow}\ Z_{\max}$	—
Solder Joint Thermal Fatigue	$N_f = 0.5(\Delta\gamma_T/2\varepsilon'_f)^{1/c}$	Coffin-manson
Random Vibration Fatigue	$N_f = C\left[\frac{z_1}{z_2 \sin(\pi x)\sin(\pi y)}\right]^{\frac{1}{b}}$	—
Electro-migration	$MTTF = \frac{Wd^{Tm}}{C^{jn}}\exp\left(\frac{Ea}{kT}\right)$	Black's equation
Gate-Oxide Time	$MTTF_E = \tau \cdot \exp(-\gamma E_{ox}) \cdot \exp(Ea/kT)$	E model
Dependent Dielectric Breakdown (TDDB)	$MTTF_{1/E} = \tau \cdot \exp(G/E) \cdot \exp(Ea/kT)$	1/E model
Hot Carrier Injection (HCI)	$MTTF = B \cdot (I_{sub})^{-N} \cdot \exp(Ea/kT)$	For nMOS
	$MTTF = B \cdot (I_g)^{-M} \cdot \exp(Ea/kT)$	For pMOS
	$MTTF = B \cdot (I_{sub})^{-N} \cdot \exp(Ea/kT)$	
Corrosion	$MTTF = A(RH)^{-n}\exp\left(\frac{Ea}{kT}\right)$	—

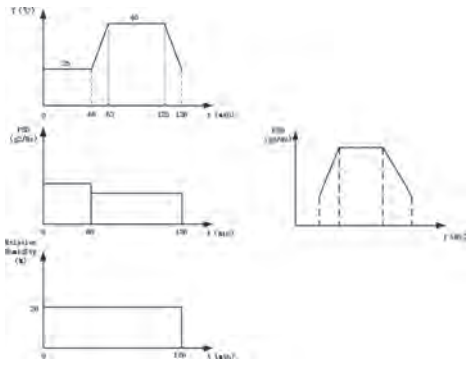


Figure 5. The profile of MOS device under operation conditions.

Table 5. Parameters of profile for temperature.

No.	Ambient temperature	Dwell in minutes
1	25	60
2	60	60

Table 6. Parameters of profile for random vibration.

No.	Frequency (Hz)	PSD (g2/Hz)
1	5	0.0266
	75	0.4
	200	0.4
	2000	0.004
2	5	0.0133
	75	0.2
	200	0.2
	2000	0.002

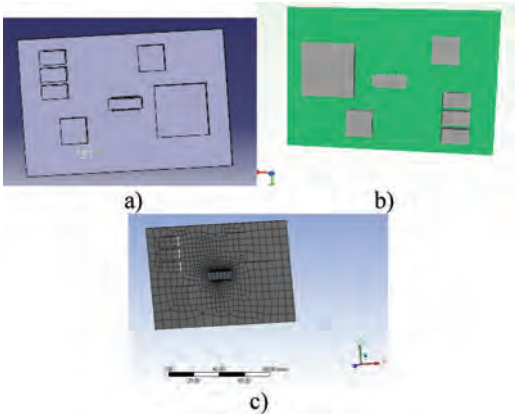


Figure 6. Three models of MOS device that is located in the middle of board and PCB: a) is CAD model, b) is CFD model and c) is FEA model.

Table 7. Results of simulation analysis.

	Parameters		Result
Thermal Analysis	Lead	Ambient Temp. 25°C	44.5°C
		Ambient Temp. 40°C	61.7°C
	Die	Ambient Temp. 25°C	55.7°C
		Ambient Temp. 40°C	72.3°C
Random Vibration Analysis	Lead	Min Natural Frequency	317 Hz
		Max PSD	8.135e-9 g^2/Hz
		Max PSD	2.683e-9 g^2/Hz
Electrical Parameters Analysis	MOS	Current Density in Interconnects	0.3 MA/cm^2
		Oxide Field	8.9 MV/cm
		Peak Substrate Current	22 µA

4.3 *Cumulative damage*

In terms of the seven potential failure mechanisms, we divide them into depletion-type and overstress-type. It is only need to confirm whether it is strong enough to stand the specific stress under

Table 8. The distribution type of each uncertain parameters.

Failure mechanism	Parameter	Distribution
Solder Joint Thermal Fatigue	L_D: *LENGTH*	Triangular Distribution
	h: Solder Height	Triangular Distribution
	ΔT: Cyclic Temperature Swing	Weibull Distribution
	Tsj: Mean Cyclic Temperature of the Solder in Degrees C	Normal Distribution
	α_c and α_s: Coefficients of Linear Thermal Expansion for Component and Substrate	Triangular Distribution
Random Vibration Fatigue	*B*: Length of the PCB Edge Parallel to the Component Located at the Center of the Board	Normal Distribution
	L: LENGTH	Triangular Distribution
	t: Thickness of PCB	Normal Distribution
	f_n: Minimum Natural Frequency	Normal Distribution
Electro-migration	*W*: Width of Interconnects	Normal Distribution
	d: Thickness of Interconnects	Normal Distribution
	T: Temperature	Normal Distribution
	j: Current Density	Triangular Distribution
TDDB	Eox: Oxide Field	Triangular Distribution
	T: Temperature	Normal Distribution
HCI	T: Temperature	Normal Distribution
Corrosion	RH: Relative Humility	Normal Distribution
	T: Temperature	Normal Distribution

operation conditions for overstress-type while it is necessary to consider the cumulative damage for depletion-type.

According to the result of FMMEA, potential failure mechanism of shock is the only overstress-type and the stress is not enough to make device failure. Besides, the change of temperature is already be considered by solder joint thermal fatigue itself, there is no need to calculate the cumulative damage. So for the other five potential failure mechanisms, the methods of cumulative damage analysis and relative formula are as Table 9.

Table 9. The algorithms of cumulative damage analysis and relative formula.

Failure mechanism	Methods of cumulative damage analysis	Formula
Electromi-gration	AFA	$AF_i = \frac{MTTF_0}{MTTF_i} = \left(\frac{T_0}{T_i}\right)^m \exp\left[\frac{Ea}{k}\left(\frac{1}{T_0}, -\frac{1}{T_i}\right)\right]$
TDDB	AFA	$AF_i = \frac{MTTF_0}{MTTF_i} = \exp\left[\frac{Ea}{k}\left(\frac{1}{T_0}, -\frac{1}{T_i}\right)\right]$
HCI	AFA	$AF_i = \frac{MTTF_0}{MTTF_i} = \exp\left[\frac{Ea}{k}\left(\frac{1}{T_0}, -\frac{1}{T_i}\right)\right]$
Corrosion	AFA	$AF_i = \frac{MTTF_0}{MTTF_i} = \exp\left[\frac{Ea}{k}\left(\frac{1}{T_0}, -\frac{1}{T_i}\right)\right]$
Random vibration fatigue	CDRA	$N_g = \frac{N_1}{\sum_{i=1}^{m} \alpha_i \left(\frac{\sigma_i}{\sigma_1}\right)^d}$

For electromigration, according to PoF model and matrix α, the vector of *TTF* is,

$$TTF = [11297.991, 11113.816, 12055.017, 11445.712, \ldots, 11954.772]$$

For each element of vector of *TTF*, calculate the time-to-failure considering the cumulative damage theory under the profile given in section 4.1.The result is,

$$t = [6312.42, 6209.51, 6735.38, 6394.95, \ldots, 6679.37]$$

Then calculate the *TTFs* of other potential failure mechanisms and form the matrix of *TTF* for one MOS structure,

$$t = \begin{bmatrix} 6312.42,\ 6209.51,\ 6735.38,\ 6394.95, \ldots,\ 6679.37 \\ 8944.57,\ 8418.34,\ 9010.22,\ 8262.84, \ldots,\ 9900.97 \\ 11396.04, 11361.27, 11360.94, 11351.17, \ldots, 11357.64 \\ 9344.20,\ 10396.72,\ 9835.80,\ 9241.89, \ldots,\ 9546.54 \end{bmatrix}$$

Since potential failure mechanisms of solder joint thermal fatigue and random vibration fatigue will never happen in MOS device, the matrix of *TTF* only includes four failure mechanisms.

4.4 *Data processing*

Calculate the matrix of *TTF* for different simulation unit and develop the data fitting and hypothesis testing. At last we transfer them to a vector of *TTF* for MOS device,

$$T = [6435.21, 5547.34, 6859.25, 6378.53, \cdots, 6678.67]$$

The data is well fitted as a weibull distribution and the *TTF* of MOS device is 6415.9h under the operation conditions given in section 4.1.

5 CONCLUSION

With the development of semiconductor technology, it is more and more obvious that the field of reliability engineering is facing the problem of reliability evaluation under complex operation conditions. So a methodology is put forward to evaluate the reliability of MOS devices in this paper. At first we need to extract large amounts of parameters of MOS device as input of following steps. Then FMMEA used to determine potential failure mechanisms and they are the focus that we will consider. The third step is simulation modeling and analysis to obtain the stress parameters under operation conditions. Then considering the cumulative damage and uncertainty of parameters, we can obtain the *TTF* of each potential failure mechanism for every simulation unit by MCS. At last, we evaluate the *TTF* of MOS device according to the data obtained before. In the end of this paper, we select a typical MOS device and it is evaluated as the case study of this methodology. The *TTF* of MOS device under certain operation conditions is 6415.9h.

REFERENCES

Borgarino, M., Menozzi, R., Dieci, D., Cattani, L. & Fantini, F. (2001) Reliability physics of compound semiconductor transistors for microwave applications. *Microelectronics Reliability,* 41, 21–30.

Haggag, A., Mcmahon, W., Hess, K., Cheng, K., Lee, J. & Lyding, J. (2000) A probabilistic-physics-of-failure/ short-time-test approach to reliability assurance for high-performance chips: models for deep-submicron transistors and optical interconnects. *Integrated Reliability Workshop Final Report, 2000 IEEE International.* IEEE.

Hall, P. & Strutt, J. (2003) Probabilistic physics-of-failure models for component reliabilities using Monte Carlo simulation and Weibull analysis: a parametric study. *Reliability Engineering & System Safety,* 80, 233–242.

Hava, A., Qin, J., Bernstein, J. B. & Bot, Y. (2013) Integrated circuit reliability prediction based on physics-of-failure models in conjunction with field study. *Reliability and Maintainability Symposium (RAMS), 2013 Proceedings-Annual.* IEEE.

Pecht, M. & Gu, J. (2009) Physics-of-failure-based prognostics for electronic products. *Transactions of the Institute of Measurement and Control,* 31, 309–322.

Saeidi, N., Schuettler, M., Demosthenous, A. & Donaldson, N. (2013) Technology for integrated circuit micropackages for neural interfaces, based on gold–silicon wafer bonding. *Journal of Micromechanics and Microengineering,* 23, 075021.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Case study of the effects of hurricanes on the coupled electricity and water systems of St Kitts

C.A. Johnson & R. Flage
University of Stavanger, Stavanger, Norway

S.D. Guikema
University of Michigan, Michigan, USA

ABSTRACT: When modelling critical infrastructure, it is important to account for the effects of interdependencies. Although there is much literature on how to account for interdependencies, there is a shortage of real-world examples. In an effort to increase the number of real-world examples, a model of the interdependent power and water systems of the Caribbean island of St Kitts has been developed. System dependencies arise due to electrically powered water pumps within the water system. Given the location of the island, it is not uncommon for it to encounter tropical storms, which may result in disruptions to the island's power system. Depending on the severity of the disruption to the power system, the effects may be able to propagate, through the dependencies, into the water system. The developed model uses the track and wind speed of past and simulated hurricanes, or up to date weather predictions, to simulate possible disruptions to the island's power system. Any propagation of these disruptions to the water system through the interdependencies are also simulated. The recent occurrence of hurricane Maria provides a useful case study to compare the output of the coupled system model with the actual effects that resulted due to the hurricane. The models are run with the known track and wind speeds of hurricane Maria. The predicted disruptions to the power system, as well as the cascading effects throughout the water system are then compared to the actual disruptions exhibited in the interdependent systems. The results can be used to validate the model and give an indication if improvements to the current model can be made.

1 INTRODUCTION

When modelling disruptions to critical infrastructure, it is agreed that any interdependencies present need to be taken into account (Buldyrev et al. 2010). There is much literature available on the various methods to model infrastructure interdependencies, however there is a lack of examples which model real interdependent infrastructure systems (Ouyang 2014). Some examples which look at real interdependent systems include Johansson and Hassel (2010) and Dueñas-Osorio et al. (2007).

To provide another real example of interdependent infrastructure modelling, a model of the coupled water and electrical power system on the Caribbean island of St Kitts has been developed by the Guikema Research Group. The model was developed with the intension to see how tropical storms affect the island's coupled power and water systems. The coupled water and power system of St Kitts provides a good opportunity to model a coupled infrastructure system given that they are relatively simple and self-contained as they service only the island of St Kitts.

In the next section, a brief overview of the island of St Kitts and the natural hazards, such as Hurricane Maria, that affect the island will be given. Section 3 will provide more detail about the model of the island's power system, water system and the dependency that exists between the two systems. The results of the model when simulating the effects of Hurricane Maria will be presented in Section 4. This will be followed by a discussion in Section 5 about the challenges associated with developing and validating this model, and more general any interdependent infrastructure model that aims to simulate real systems. The final section, Section 6 will then draw the conclusions of the paper.

2 THE ISLAND OF ST KITTS

St Kitts is the larger of the two islands of the Federation of St Christopher (St Kitts) and Nevis, which are part of the Leeward Island group in the Eastern Caribbean (The Commonwealth, The Offical Website of St Kitts and Nevis). The total

population of St Kitts and Nevis is estimated to be just over 50,000. St Kitts is of volcanic origin and thus the centre of the island is mountainous with a highest peak at Mount Liamuiga. Therefore the majority of the population live along the coastline of the island (The Offical Website of St Kitts and Nevis).

The total area of St Kitts is 69 square miles or roughly 180 square kilometers and thus is a relatively small island. The small size of St Kitts means that modelling infrastructure systems of the island is feasible, without having to overly simplify the system. These systems are also self-contained due to the geographic constraints that they service only the island of St Kitts.

Due to the location of St Kitts, the natural hazard of tropical storms is an annual threat to the island. The most notable storm to hit St Kitts in recent years was Hurricane George in September 1998, for which the estimated damage costs to St Kitts and Nevis was 445 million United States Dollars (USD) (Relief Web 2002). Other storms that have had great impacts recently on St Kitts include Hurricane Lenny in 1999 with estimated damage costs of around 41 million USD, Hurricane Hugo which hit the island in 1989 and caused damage to roughly 20 percent of the poles in the electricity power system, and Hurricane Luis in 1995 which caused great damage to the power and water systems of the island (Relief Web 1995, Relief Web 1999, US Aid 1990).

2.1 *St Kitts and Hurricane Maria*

The National Oceanic and Atmospheric Administration (NOAA) first issued a tropical storm watch on 16th September 2017, with the first advisory report being issued at 11:00 Atlantic Standard Time (AST). By 17:00 AST, NOAA were recording tropical storm force surface winds of what was then Tropical Strom Maria. At this time a hurricane watch was in effect for St Kitts and Nevis along with Antigua, Barbuda and Montserrat. At the time of Advisory 6 of Maria 17:00 AST, hurricane force winds had been recorded and the storm had been upgraded to a hurricane. The St Kitts hurricane watch had been upgraded to a hurricane warning. As of 19th September 05:00 AST, St Kitts was experiencing tropical storm force winds as Hurricane Maria passed by the south west coast of the island (NOAA). The eye of Hurricane Maria was reported as passing within 90 miles of the island (St Kitts & Nevis Observer 2017). Advisory 15, issued on 19th September at 17:00 AST was the last advisory to issue a hurricane warning to the island of St Kitts (NOAA).

Figure 1 shows the path of the centre of Hurricane Maria as it passed by St Kitts, as well as the wind speed severity category bands (The New York Times 2017). From Figure 1 it can be seen that St Kitts experienced tropical storm severity wind speeds as Hurricane Maria passed by the south-west of the island.

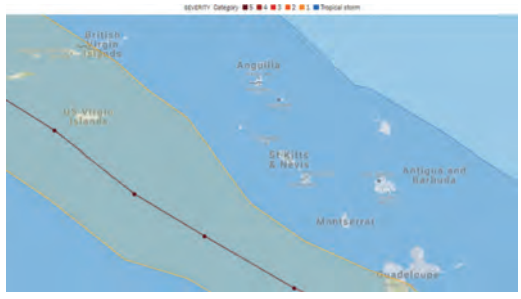


Figure 1. Track of Hurricane Maria—The New York Times, 2017.

3 COUPLED POWER AND WATER SYSTEM OF ST KITTS

The island of St Kitts is a relatively small island with an area of 69 square miles. The power and water systems which provide electricity and clean water to the island are self-contained as they do not provide utilities to other islands and can be modelled without having to be over simplified. The water system of St Kitts contains 30 wells which rely on the electricity system to pump the water from the wells into the water system. This dependence can be modelled along with the power and water systems to develop a model of the coupled power and water system for St Kitts.

The water system model was developed using information provided by St Kitts Water Department. The model encompasses the entire clean water distribution system, including the supply sources and demand nodes. Figure 2 shows the modelled water system of St Kitts, including the pipelines, reservoirs and wells. The blue lines in Figure 2 represent the water pipes, the red dots are the water nodes (wells and junctions) and the black nodes are the reservoirs and tanks within the water system.

The water system was modelled using EPANET 2.0, a program that can "perform extended period simulation of hydraulic and water quality behaviour within pressurised pipe networks" EPANET (2000).

The power system was modelled using the limited information available on St Kitts Electricity Company Limited (SKELEC) website. The main lines of the power system were modelled, however

Figure 2. Model of St Kitts water system.

Figure 3. St Kitts electricity power system model.

due to lack of available information the smaller distribution lines of the power system could not be included in the model. As stated in Section 2, most of the island's population resides on the coastal areas of St Kitts. Therefore the 3 main power lines that surround the island are modelled. The model contains 157 nodes, representing 157 wooden electricity poles that the power cables are attached to. The three main power lines can be seen in Figure 3. The northern line is represented by red squares and contains 70 poles, the southern line is shown as blue dots and contains 39 poles and the western line is shown as purple triangles and contains 48 of the 157 poles.

For each pole, a fragility curve giving the probability of failure given the on-island wind speed has been determined based on historical damage reports of hurricanes that were readily available to the public. If one pole fails in a power line, then all the poles downstream of the failed pole will also be modelled as inoperable.

The water system depends on the power system at each of the 30 wells within the system to get the water from the well into the water distribution system. To model the water system's dependency on the power system, each well is dependent on the closest electricity node, or pole, to provide power to the well. During the simulation, if a node becomes inoperable that a well depends on, the well that depends on it is then removed from the water system model. Figure 4 shows the modelled electricity lines, as black triangles, and the water wells as red squares and other water nodes as blue dots. Sixteen wells depend on nodes within the northern power line, 7 wells depend on electricity from the western line and 8 wells depend on the southern line.

The track as well as the maximum wind speed recorded every 6 hours for the centre of Hurricane Maria from 17th to 28th September 2017 was used to estimate the maximum speed of the hurricane over St Kitts using a model previously developed by Guikema et al. (2014). This provided the probability of failure for each of the 157 poles within the electricity power system. These probabilities were then used in the MATLAB model developed within the Guikema Research. To look at the effects of Hurricane Maria on St Kitts coupled power and water systems, the MATLAB model was run for 60000 iterations, where first it was modelled which poles would fail or break due to the hurricane before modelling the cascading effects of these poles down each power line and through to the water system.

Figure 4. Coupled power and water system of St Kitts.

4 RESULTS

4.1 *Electricity power system*

For each of the 60000 iterations that the model completes, each pole is recorded as failing due to the wind or not. Therefore, the model gives the frequency of initial pole failure, which can be seen in Figure 5. The frequency, given as a percentage, of each pole failing due to Hurricane Maria is shown using the colour scale as shown in Figure 5.

There are a few poles with relatively low frequency of failing in the southern line, closest to the centre of the island. However, the southern line is mostly unaffected by Hurricane Maria. On the northern power line there is one pole, around a third of the way along the line that has a high frequency of failure. Failure of this pole would cause disruptions to the rest of the line up to the north-east of the island, and thus any wells that depend on any nodes on this section of the northern line to also be disrupted. On the western power line there are more poles within the line that have initial failures due to the hurricane. This is expected as Hurricane Maria passed by the south-west of the island. The most frequent initial failure in the western line is a pole that lies around halfway along the western coast. This would cause disruptions along the rest of the line, all along the western coastline, and these disruptions would cascade to any wells that depend on this section of the western power line.

4.2 *Actual power outages*

When looking for data of what affects Hurricane Maria had on the electricity system of St Kitts, the only information found came from SKELEC website. This information was a restoration update that named residential areas that were experiencing power outages due to Hurricane Maria. The press release stated "Most of our feeders remained intact and online during the passage of the storm. Some like the Canada feeder which services Conaree, Halfmoon and Canada Estates came offline, also Basseterre North Buckley's to Trinity also is fully offline" SKELEC (2017).

Figure 6 shows the areas that SKELEC reported to have experienced outages due to Hurricane Maria. The three areas to the north of the island are Canada, Halfmoon and Conaree. Canada has been filled red rather than just outlined as the Canada feeder was the feeder identified as causing these outages to the north of the island. These three outlines were not specified by SKELEC but are the outline of the three areas named by SKELEC in the press release. Due to no further information being found in relation to power outages due to the hurricane this is best representation of possible affected areas that can be given. The larger section on the south-west of the island represents the area from Basseterre North Buckley's to Trinity. The outline encases all residential areas between, and including, Basseterre North Buckley's to Trinity. Again, the press release from SKELEC is the only information on which to base the actual areas affected and thus the areas cannot be more specific.

Figure 7 shows the results of the model for frequency of initial pole failure compared to the areas that SKELEC reported to experience power outages due to Hurricane Maria. When looking at the predictions and actual areas to exhibit outages on the northern power line, the high frequency of a

Figure 5. Frequency of initial pole failure.

Figure 6. Areas of St Kitts that experienced power outages due to Hurricane Maria.

Figure 7. Frequency of initial pole failure compared to areas to experience power outages.

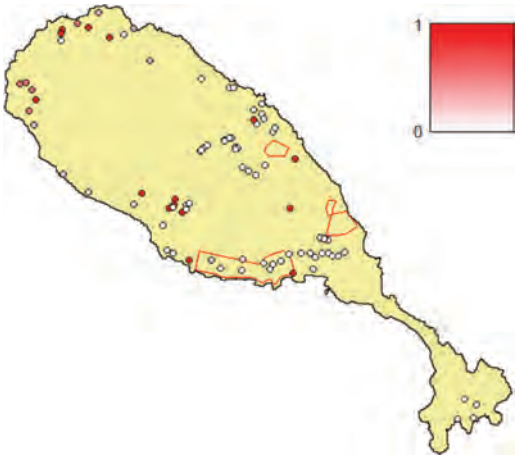


Figure 8. Frequency of water nodes with pressure lower than 20 psi.

pole break on the northern side of the island given by the model coincides with the press release that the northern side was affected such that a feeder came offline. When looking at the south-west area that experienced outages, if the main line was affected by the hurricane, it would understandable to see more outages along the line. However, these outages could have been caused to failures or disruptions within the smaller distribution lines that are not included in the model.

4.3 *Water system*

When looking at disruptions to the water system, the frequency that nodes exhibit pressure below 20 psi as well as 0 psi are both considered. The pressure of the nodes within the water system is important to measure as if the pressure becomes negative this can cause the water to flow in the wrong direction, or cause pipes to break. These both can cause the water to become contaminated. Pressures lower than 20 psi are investigated as in the USA this is the minimum pressure required to use a water system for firefighting purposes (EPA 1992).

Figure 8 shows the frequency that water nodes exhibited pressures lower than 20 psi. Comparing Figure 8 to Figure 5 the low pressures towards the lower end of the island coincide to the areas in which the initial pole failures occurred. The low pressures exhibited towards the north-western part of the island are due to the initial failures in both the northern and western power lines cascading to the north of the island.

Figure 9 shows the frequency that the water nodes exhibited pressures less than 0 psi. There are

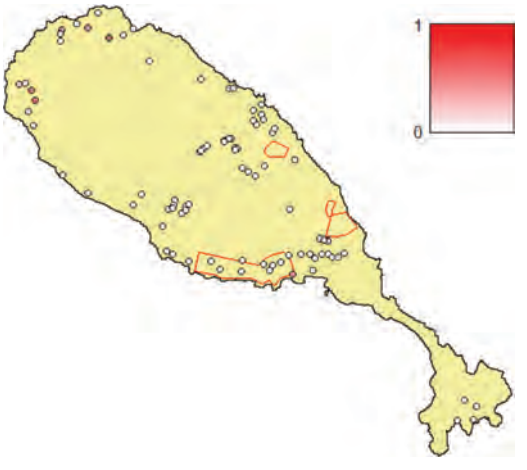


Figure 9. Frequency of water nodes with pressure lower than 0 psi.

few instances of nodes exhibiting pressure lower than 0 psi, with a few occurring on towards the north-west end of the island. Although the initial aim was to compare the results of the model to how Hurricane Maria affected the coupled infrastructure power and water system of the island, it was challenging to find any data related to disruptions to the water system.

The location of wells compared to the actual outage data was also considered to see if any wells would have been without power during the outages.

When looking at Figure 10, there are three wells that are situated in the south-west area of

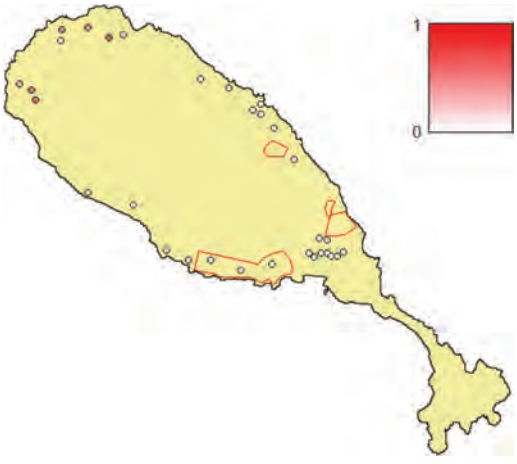


Figure 10. Frequency of water wells with pressure lower than 0 psi and reported power outages.

St Kitts that experienced power outages. To further test the water system model, these three wells were removed from the model, as is they had lost power. However, all other nodes in the water system model maintained a pressure above 0 psi, thus even if these three wells were without power, the water system would not have any disruptions.

Although no information directly related to disruptions to the water system was found, the Prime Minister of St Kitts and Nevis stated that the island's infrastructure such as the electricity and water systems "sustained extensive damage" (Relief Web 2017).

5 CHALLENGES WHEN MODELLING AND VALIDATING REAL INFRASTRUCTURE MODELS

When modelling the coupled power and water infrastructure system of St Kitts, information regarding the water system was provided by St Kitts Water Department. This, along with ongoing communication with the water department while developing the model, allowed an extensive model of the system to be developed. However, when developing the electricity model, the only information available was that in the public domain, primarily on SKELEC website. This meant that the model was an oversimplified representation of the main power lines in the island's electricity system.

Accessing and collecting data to form the basis of infrastructure models is a well-documented issue with constructing models of independent infrastructures as well as interdependent infrastructures (Ouyang 2014, Johansson and Hassel 2010, Rinaldi et al. 2001). Another problem highlighted by Ouyang (2014) is that infrastructure are large, complex systems that are constantly changing and thus the data used to produce the model may not be relevant for very long after the model has been developed. Any major changes to the infrastructure would have to be updated within the model.

When trying to validate the model, information on the disruptions to both the water and power systems due to Hurricane Maria were also hard acquire. The electricity company did have some information available, however, this is aimed at the customers of the infrastructure and is there to reassure them that the company is aware of the problems and are fixing the issues. The information relates only to general areas, and not specific parts of the system that would be useful when validating the model.

Within the USA, publicly available outage data of power companies is becoming increasingly more available. Most power utility companies are now showing daily outage data on their websites, see the website Power Outage for more information on which utilities provide outage data. However, other utilities in the USA, such as those who manage water infrastructure are still less willing to share outage data.

6 CONCLUSION

The aim of the paper was to compare the results of the coupled power and water system of St Kitts to the disruptions the systems exhibited during Hurricane Maria. Although the electricity model is a simplified representation of only the island's main power lines, the areas that experienced power outages did coincide with poles that had a high frequency of initial failure. This showed that the model is quite accurate on estimating the maximum wind speed for the different areas of the island. However, as the smaller distribution lines could not be included in the model and limited outage information is available, it is difficult to access the accuracy of the model. Improvements could be made by increasing the complexity of the model, if the necessary information can be obtained from St Kitts Electricity Company.

Although the water system model is a good representation if the island's water system, the lack of data available relating to disruptions due to Hurricane Maria again means it is difficult to validate the model. The lack of available information relating to infrastructure systems is a well-recognised challenge when developing models of infrastructure systems.

REFERENCES

Buldyrev, S.V., Parshani, R., Paul, G., Stanley, H.E. & Havlin, S. 2010. Catastrophic cascade of failures in interdependent networks. *Nature,* 464, 1025.

Dueñas-Osorio, L., Craig, J.I., Goodno, B.J. & Der Kiureghian, A. 2007. Seismic response of critical interdependent networks. *Earthquake Engineering & Structural Dynamics,* 36, 285–306.

EPA, U. 1992. Manual of Small Public Water Supply Systems.

EPANET 2000. [Online] Epanet 2 Users Manual Available at: https://nepis.epa.gov/Adobe/PDF/P1007 WWU.pdf [Accessed on 21 December 2017].

Guikema, S.D., Nateghi, R., Quiring, S.M., Staid, A., Reilly, A.C. & Gao, M. 2014. Predicting Hurricane Power Outages to Support Storm Response Planning. *Access, IEEE,* 2, 1364–1373.

Johansson, J. & Hassel, H. 2010. An approach for modelling interdependent infrastructures in the context of vulnerability analysis. *Reliability Engineering and System Safety,* 95, 1335–1344.

NOAA National Oceanic and Atmospheric Administration. [Online] Hurricane Maria Advisory Archive Available at: http://www.nhc.noaa.gov/archive/2017/MARIA.shtml? [Accessed on 21 December 2017].

Ouyang, M. 2014. Review on modeling and simulation of interdependent critical infrastructure systems. *Reliability Engineering and System Safety,* 121, 43–60.

Relief Web 1995. [Online] Caribeean—Hurricane Luis Sep 1995 UN DHA Situation Reports-10 Available at: https://reliefweb.int/report/antigua-and-barbuda/caribbean-hurricane-luis-sep-1995-un-dha-situation-reports-1-10 [Accessed on 21 December 2017].

Relief Web 1999. [Online] Hurricane Lenny OCHA Situation Report No. 7 Available at: https://reliefweb.int/report/anguilla/hurricane-lenny-ocha-situation-report-no-7 [Accessed on 21 December 2017].

Relief Web 2002. [Online] Caribben—Hurricane George Appeal No. 29/1998 final report. Available at: https://reliefweb.int/report/antigua-and-barbuda/caribbean-hurricane-georges-appeal-no-291998-final-report [Accessed on 21 December 2017].

Relief Web 2017. [Online] Prime Minister Dr. the Hon. Timothy Harris' Post-Hurricane Maria Address Available at: https://reliefweb.int/report/saint-kitts-and-nevis/prime-minister-dr-hon-timothy-harris-post-hurricane-maria-address [Accessed on 21 December 2017].

Rinaldi, S.M., Peerenboom, J.P. & Kelly, T.K. 2001. Identifying, understanding, and analyzing critical infrastructure interdependencies. *Control Systems, IEEE,* 21, 11–25.

Skelec 2017. [Online] Hurricane Maria Restoration Update Available at: http://www.skelec.kn/hurricane-maria-restoration-update-wednesday-september-20-2017/ [Accessed on 21 December 2017].

St Kitts & Nevis Observer 2017. [Online] Hurricane Maria remains powerful, dangerous hurricane Available at: http://www.thestkittsnevisobserver.com/breaking-news/hurricane-maria-remains-powerful-dangerous-hurricane-news-update/ [Accessed on 21 December 2017].

The Commonwealth [Online] St Kitts and Nevis Available at: http://thecommonwealth.org/our-member-countries/st-kitts-and-nevis [Accessed on 21 December 2017].

The New York Times 2017. [Online] Maps: Hurricane Maria's Path Across Puerto Rico Available at: [Accessed on 21 December 2017].

The Offical Website of St Kitts and Nevis [Online] About St Kitts and Nevis Available at: https://www.gov.kn/ [Accessed on 21 December 2017].

US AID 1990. [Online] After-action Report of the Hurricane Hugo OFDA Disaster Relief Team Available at: http://pdf.usaid.gov/pdf_docs/Pnabg072.pdf [Accessed on 21 December 2017].

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Availability simulation model of global navigation satellite system based on operation

A.G. Zhao, X. Sun & Y. Sun
School of Reliability and Systems Engineering, Beihang University, Beijing, China

B.D. Li
China South Industry Institute, Beijing, China

ABSTRACT: Availability is a core attribute of Global Navigation Satellite System (GNSS) and many topics are studied in this area. However most of them are about a part of GNSS, which is single satellite availability, navigation constellation availability or the availability of ground stations. Very few studies are concerned with the availability of the whole system. In view of this, the availability simulation modeling of the whole GNSS including space segment and ground segment were studied in this paper. The influence of satellite failure, interruption, coverage and ground station fault, repair etc. on system operation was considered to establish the model. Then the simulation logic was developed to simulate the operation of navigation system. Based on the STK/MATLAB hybrid simulation, the service availability of the GPS for Beijing was analyzed finally. The simulation results verify the applicability of the model and the simulation algorithm.

1 INTRODUCTION

Global Navigation Satellite System (GNSS) is used in various fields and bring great conveniences to people's lives. It has become an indispensable necessity in people's life. Availability is a core parameter of GNSS and reflects the degree to which the needs of users are met. Therefore, the level of availability of GNSS is directly related to the quality of the service it provides. At present, many scholars at home and abroad have carried out relevant research on it and achieved some results.

In the aspects of availability of ground stations, Joo built a series model of availability considering early stage and stabilization stage and obtained the availability of the ground control sections by the actually used data (Joo et al. 2007). Li took the coverage performance of the ground station as its available basis and put forward a availability model of the monitoring station (Li et al. 2010). In the aspects of availability of navigation constellation, Xiang established the constellation system reliability model and analyzed the impacts of the reliability of the satellite, the constellation configuration and the number of backup satellites on the reliability of the constellation system (Xiang et al.2007). Zhou analyzed the impacts of the single-satellite reliability change on the constellation availability aiming at the network supplement mode of terrestrial backup satellite (Zhou et al. 2014). Zhao proposed a Markov model for calculating per-slot availability considering the influence of standby satellite on slot availability (Zhao et al. 2013). Hou calculated the availability of the navigation constellation based on the failure of the subsystem component and built service availability calculation model based on constellation state probabilities (Hou et al. 2014). In the aspects of availability of navigation system, Zhao established a multidimensional parameter system of the availability of satellite navigation system (Zhao et al. 2014). Liang modeled and simulated the system structure of the ground system of the remote sensing satellite based on DoDAF (Liang et al. 2017). Yang proposed a system efficiency modeling and analysis method that comprehensively considers various kinds of factors, such as various types of interrupting randomness and lossy failure performance, and used Markov chain to build the availability model of navigation satellite (Yang et al. 2017).

In summary, the availability classification and model of GNSS present the diverse features such as the accuracy availability and integrity availability, single-point availability and service area availability, single-satellite availability and navigation constellation availability. At present, the researches on the availability of GNSS are more focused on local issues. For example, some scholars consider the availability of single satellite while others study the

availability of navigation constellations in terms of reliability and maintainability, or only consider the availability of ground stations, all of which are difficult to support the comprehensive analysis and evaluation of the overall availability of GNSS. In view of this, based on the composition, mission requirements, usage patterns and user requirements of GNSS, a systematic and comprehensive availability simulation model was built considering the operation, interruption, maintenance and guarantee of GNSS and the overall simulation logic flow was developed in order to realize the simulation of GNSS operation, networking, network supplement, interruption and other process, and analysis of its availability level. At the same time, backup satellite program can also be provide with a basis to shorten the satellite supplement waiting time and improve the navigation system availability.

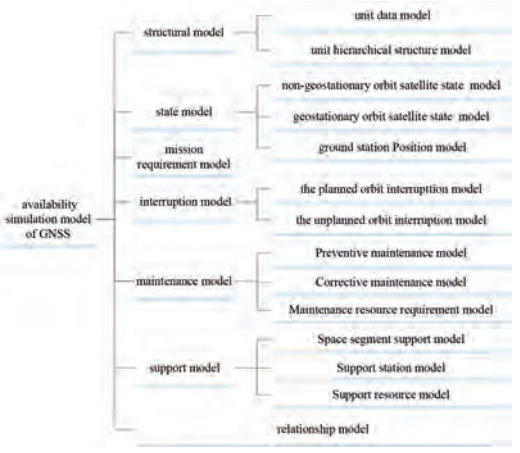


Figure 1. Framework of GNSS availability simulation model.

2 FRAMEWORK OF AVAILABILITY SIMULATION MODEL OF GNSS

Availability is the description of the working state of a system. According to the different mission requirements of GNSS, its availability can be divided into instantaneous availability, single point availability, service area availability and moving target availability. As GNSS is affected by the internal failures and refurbishment and external weather and environment, it cannot be available at all time. The availability of GNSS is mainly affected by the following factors: (1) the reliability of equipment (2) equipment maintenance (3) system structure (4) backup strategy. The above four factors were considered comprehensively and the availability simulation model of GNSS was established in this paper.

According to the overall structure and operation characteristics of the navigation system, in order to describe the whole process of composition, operation, maintenance and support of GNSS completely, the framework of the overall availability simulation model was established, as shown in Figure 1.

The model consists of seven parts, which respectively are the structural model, the state model, the mission requirement model, the interruption model, the maintenance model, the support model and the relationship model among them. The structural model is mainly used to describe the constituent elements, the hierarchical structure and the reliability characteristics of the constituent elements of the system. The state model is used to describe the state of each satellite to calculate its coverage area. The mission requirement model is used to describe different types of missions for different user groups. The orbit interruption model is used to describe the repair process of the space segment of GNSS, including the type of outage, repair time and the station information. The maintenance model is used to describe the information including the purpose of maintenance, the type of maintenance, the time required for maintenance and the resources required for maintenance. The support model is used to describe the relationship among the maintenance support factors and operational support factors including the storage and supply of satellites, the support equipment and the support facilities and personnel etc. The relationship model is used to describe the relationship among the system structure, state, missions, interruption, maintenance and support. With the above seven models, all the elements describing the operation, interruption and maintenance of GNSS are realized.

3 DISCRIPTION OF AVAILABILITY SIMULATION MODEL

3.1 *Structural model*

The structural model mainly describes the organizational relationship of the constituent units of the system at the functional structure level, and provides reference and support for the system composition and reliability data of the simulation modeling.

GNSS is a typical complex system, which has a variety of equipment types such as space segment equipment, ground segment equipment and user segment equipment, and each equipment has its own composition and characteristics. Therefore, it was divided the functional structure of GNSS by

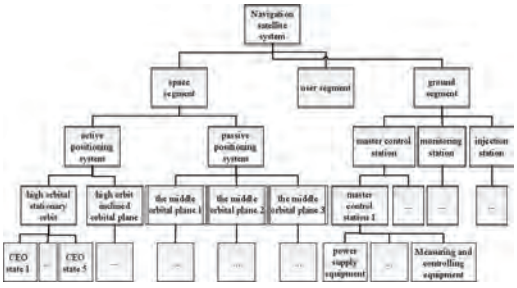


Figure 2. GNSS structural model.

layers from top to bottom in this paper, which is divided into large system layer, system layer, subsystem layer, module layer and equipment layer. Taking into account the acquisition of equipment fault information, it was intended to decompose the space segment to the satellite level, and the ground segment to the power supply equipment level, as shown in Figure 2. Data structures are shown in Table 1–2 below. Table 1 describes the unit fault mode and basic reliability data, and Table 2 describes the main relationship between units, including the relationship of structure and quantity.

3.2 *State model*

The state model is used to describe the initial state of each unit. The longitude, latitude and coverage of dynamic satellite of each satellite that change with time are calculated by the satellite operation rules, so as to complete the calculation of the coverage of the mission area. The satellite navigation system includes space satellite and ground station equipment. The space satellite can be subdivided into non-geostationary orbit satellite and geostationary orbit satellite. The operating rules and status of each unit are different. According to the different operating states, this paper divides the state model into non-geostationary orbit satellite model, geostationary orbit satellite state model and ground station state model. The data structures are shown in Table 3.

From the below three groups of state models, it can be seen that the operational state model of satellite determines the under-satellite point and coverage area at each moment of time, and the state of the ground station determines its geographical position so as to determine whether the coverage of satellite to the ground meets the mission conditions.

3.3 *Mission requirement model*

Mission requirement model is mainly used to describe the mission requirements and the logical relationship between different submissions. The main function of GNSS is to use satellite to provide users with the quick navigation and positioning, short digital message communication and timing services. According to the different functions, its main missions can be divided into the following three aspects: (1) high-precision navigation (2) high-precision positioning (3) high-precision timing. For different types of missions, and targeting different user groups, the geographical location information, movement information, and time information of the stationary users, mobile

Table 1. Unit data.

Data item	Definitions
Unit name	The unit containing the system
Fault mode	Different types of fault of the unit
Parameter name	Fault parameters, such as failure rate, MTBF, etc.
Distribution type	Parameter distribution type, including the exponential distribution and normal distribution
Parameter 1	Parameter corresponding to the distribution type
Parameter 2	Parameter corresponding to the distribution type
Initial working time of unit	The start of service of each satellite or equipment

Table 2. Unit hierarchical structure.

Data item	Definitions
Unit name	Subunit identification
Parent unit name	Parent unit identification
Unit matching number	Matching quantity of the unit that the current parent unit is affiliated with
Unit type	Determine whether the unit is a system, subsystem, module, or unit

Table 3. State model.

Non-geostationary orbit satellite state	Geostationary orbit satellite state	Position of ground station
Unit ID	Unit ID	Unit ID
Orbit semi-long axis	Longitude	Longitude
Eccentricity	Orbit height	Latitude
Orbital inclination	Field angle	Ground height
Right lifting node longitude		
Perigee argument		
True near corner point		
Field angle		

users and regional users shall be provided. In view of this, the data structure of mission requirement model was established as shown in Table 4, to describe the requirements of different missions.

The three typical missions of GNSS shall be completed in the space segment, user segment and ground segment together. It is assumed that the user segment is intact in this paper. Therefore, the navigation system is represented as a series logical relationship between the space segment and the ground segment. Reliability Block Diagram is used to express the success criterion of the mission, as shown in Figure 3, when the functions of different devices comply with the following logic condition of reliability, the navigation system mission is succeeded.

Table 4. Mission requirement model.

Data item	Definitions
Mission type	Divided into navigation, positioning and timing
Coordinate 1 Coordinate 2 Coordinate 3 Coordinate 4	For positioning and timing, the coordinates refer to the longitude and latitude of the spatial location of the mission initiation point or the latitude and longitude coordinates of the mission initiation region. For the navigation, the coordinates can express a point, which refer to the spatial location of the mission initiation point; or express a distance, which refer to that from the mission initiation point coordinate to the mission end point coordinate.
Height	The level of the ground surface where the user is
Speed	For the navigation mission, the client speed affects the satellite service area switching, so the speed parameter needs to be given
Accelerated speed	For navigation users, in addition to the user speed, the user status, acceleration, deceleration or uniform speed shall be described
Time to start work	Time to start work
Time to end work	Time to end work

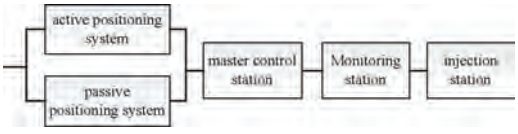


Figure 3. RBD for success criterion of the mission.

3.4 *Orbit interruption model*

The orbit interruption model is used to describe the repair process of the space segment orbit interruption of the satellite navigation system, including the type of outage, the time of interruption and the information of the orbit which it is located. Depending on different causes of the orbit interrupt, the interrupt time and repair measures are also different. Therefore, two types of space orbit interruption models are established in this paper: the planned orbit interrupt model and the unplanned orbit interrupt model. The former is used to describe the conditions that the satellite orbit is stopped and the satellite cannot provide service caused as operation and maintenance activities of satellite and the replacement of satellites, and the specific data structure is shown in Table 5. The latter is to describe the interruption repair measures and interruption repair time after the orbit is stopped caused as short-term hard fault and long-term hard fault caused as the accidental fault, and the specific data structure is shown in Table 6.

Table 5. Planned orbit interruption.

Data item	Definitions
Outage unit	The name of the unit that is the planned interruption is in, given by the structural model, as defined in Table 1
Planned outage type	The types of outages corresponding to planned interruption include operation and maintenance and end of life
Calendar time interval	Calendar time intervals of preventive maintenance
Repair time	The time consumed for the planned interrupt operation
Name of orbit	Orbit plane that the satellite is in

Table 6. Unplanned orbit interruption.

Data item	Definitions
Outage unit	The name of the unit that is the unplanned interruption is in, given by the structural model
Non-planned outage type	The types of outages corresponding to unplanned interruption include short-term hard faults and long-term hard faults
MTTI distribution type	The distribution function type of the average interrupt time of the unit
MTTI parameter1	Shape parameter 1 of interruption time distribution function
MTTI parameter2	Shape parameter 2 of interruption time distribution function
Name of orbit plane	Orbit plane that the satellite is in

3.5 *Maintenance model*

The maintenance model is primarily used to describe maintenance activities on equipment that fails on ground segments, including preventive maintenance activities and corrective maintenance activities. In order to fully describe the maintenance activities, the data structure of the preventive maintenance, the corrective maintenance and the maintenance resource requirement were established in this paper details can be seen in Tables 7–9.

Table 7. Preventive maintenance.

Data item	Definitions
Maintenance unit name	The name of the unit that the preventive maintenance operation is in, as defined in Table 1
Preventive maintenance type	Refer to the type of preventive maintenance operations, including operations and maintenance
Calendar time interval	Calendar time interval of preventive maintenance
Duration	Time consumed for this preventive maintenance operation
Maintenance station name	Maintenance site where the unit is in

Table 8. Corrective maintenance.

Data item	Definitions
Maintenance unit name	The name of the unit performing the corrective maintenance
MTTR distribution type	The distribution function type of the unit repair time
MTTR parameter 1	Shape parameter 1 of repair time distribution function
MTTR parameter 2	Shape parameter 2 of repair time distribution function
Maintenance station name	Maintenance site where the unit is in

Table 9. Maintenance resource requirement.

Data item	Definitions
Maintenance unit name	The name of the unit performing the corrective maintenance
Maintenance method	Describe the type of maintenance of the maintenance unit
Resource name	Maintenance resource for the unit to perform the corresponding level of maintenance work
Resource quantity	Maintenance resource quantity for the unit to perform the corresponding level of maintenance work

3.6 *Support model*

The support process is a complex process that contains various dynamic factors. The support model is a model that describes the support system of the navigation system, including the storage and supply of spare parts, support equipment and facilities. According to the different composite objects, the space segment support model and the ground segment support model were respectively established. The former is used to describe the information about the backup satellite configuration at all levels and the ground network supplement time, as shown in Table 10. The latter mainly describes the information about the ground support station and the support resources, as shown in Tables 11–12.

3.7 *Relationship model*

Based on the above models, if we want to establish a complete satellite navigation system availability model, we must establish the relationship between the models to lay the foundation for the subsequent establishment of simulation logic. The relationship model mainly describes the relationship of system structure, status, mission and interrupt maintenance station, as shown in Figure 4.

1. Status and missions

The operating status of the satellite determines its subastral point and coverage area and the status of the ground station determines its coverage. During the mission process, the coverage of satellite and

Table 10. Space segment support.

Data item	Definitions
Orbit plane name	The orbit plane that the repair orbit is in
Level	Describe the support level of station, usually refer to the orbit plane level
Storage quantity in orbit	Quantity of backup satellites in the orbit
Network supplement time in orbit	Refer to the time interval from that the back-up satellite in the orbit reaches the assigned orbit and debugged completely to the start of work.
Superior name	Describe the support organization relationship between the orbit plane and the ground
Ground storage quantity	Refer to the quantity of backup satellite in the ground launching base
Ground network supplement time	Describe the time required for network supplement, launch, and commissioning between the ground and the orbit plane

Table 11. Support station.

Data item	Definitions
Station name	The station used to maintain the equipment
Station level	Describe the support level of the station, including the orbit plane level, station level and launch base level
Superior station level	Represent the superior station of this station, and describe the support organization relationship between stations
Support time between stations	Describe the transit time from the superior station

Table 12. Support resource.

Data item	Definitions
Station name	The station used to maintain the equipment
Resource name	Resources required in the maintenance process, including manpower and equipment
Resource quantity	Resource quantity provided in this station

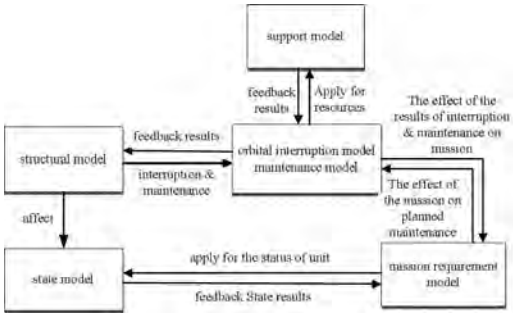


Figure 4. Relationship model.

ground station in the mission area is calculated, and then whether the mission is complete is judged by success criterion of the mission.

2. Structure and status

When the unit breaks down in the course of carrying out the mission, it is fed back to the state model. When the satellite needs to be supplemented during the repair process, it also needs to be fed back to the state model after updating the unit.

3. Maintenance interruption and mission

During the mission, the mission may be affected when the unit is interrupted or repaired. The mission process can also have some impact on the planned maintenance.

4. Interruption and support

During the repair process, whether the satellite is needed or not shall be determined. When the orbit plane does not comply with the required satellite, the interruption repair event is standby, when the quantity of the back-up satellites in the orbit plane is less than the setting value, an application is made to the ground station for satellite supplement.

5. Maintenance and support

In the maintenance process, the spare parts and other support resources are needed. When the number of resources in the maintenance station is insufficient, the maintenance event is standby. After the maintenance is completed, release the resources, and wait for the maintenance of other fault parts. After the standby maintenance event complies with the maintenance conditions, judge whether it can be maintained and determine to maintain or continue to wait.

4 AVAILABILITY SIMULATION LOGIC

4.1 *The main simulation logic*

Taking the components of the navigation system as the subject of the simulation and mission events as the traction to trigger failure events and all events in repairing model, interrupt model and support model, in the thought of discrete event simulation technology, the main process of the overall availability simulation for GNSS was established, which is shown in Figure 5. With the advancement of simulation events, determine whether the failure and maintenance events affect the implementation of the mission one by one, and carry out the corresponding maintenance activities, until the end of the mission, and then perform the availability statistics.

Based on the main simulation logic, each event has a detailed process. Eight events and their simulation algorithms were developed in this paper. The list of simulation events are shown as Table 13.

4.2 *The simulation process of typical events*

In view of the limitation of pages, only several simulation processes of typical events are presented in this paper.

4.2.1 *Simulation process of mission event*

Mission event is the key for implementation of satellite navigation system availability simulations. Driven by the mission, the equipment failure and maintenance process are simulated, and the impact on the mission is determined. In this paper, the satellite visibility is used as the evaluation criteria for

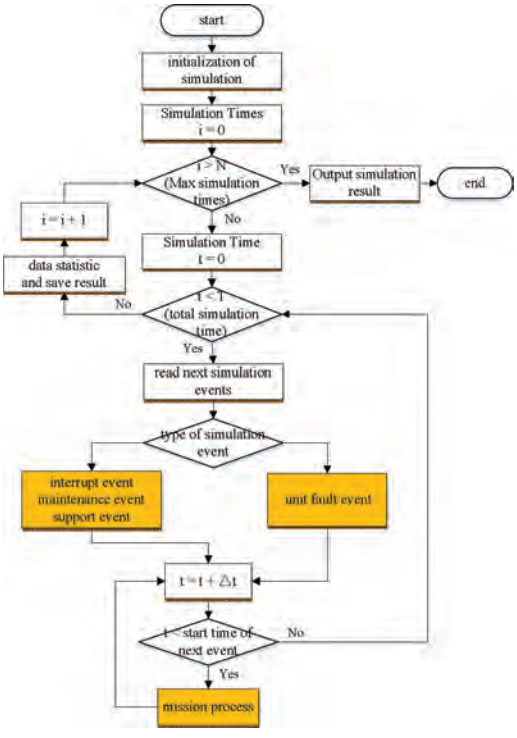


Figure 5. Main simulation logic.

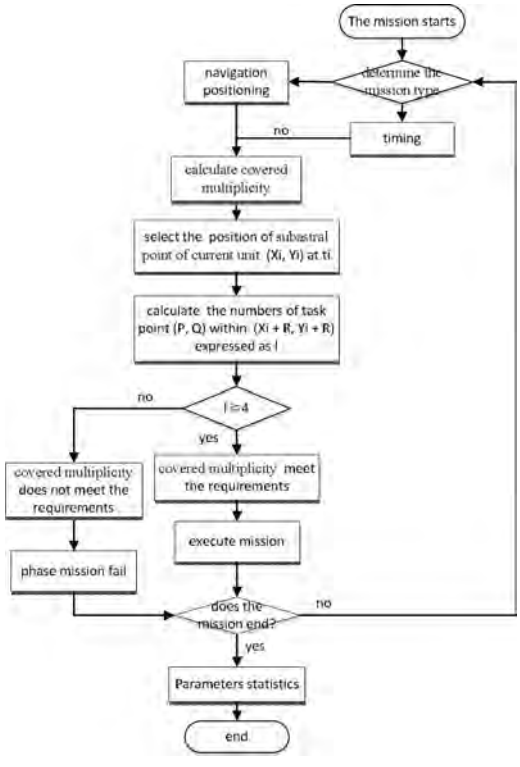


Figure 6. Simulation process of mission.

Table 13. Simulation events.

Serial number	Simulation events
1	Mission event
2	Unit fault event
3	Unplanned interrupt repair event
4	Planned interrupt repair event
5	Preventive maintenance event
6	Corrective maintenance event
7	Support event for satellite-replacing in the space
8	Ground support resource

the availability of the navigation system. The mission simulation process is shown in Figure 6.

First determine the mission type, where the navigation, positioning and timing missions are subject to different success criteria of missions, and then determine the mission implementation under each simulation step. According to the mission requirements of the navigation system, the state of each unit in the state model is read to determine whether the covered multiplicity complies with the coverage of the user in the area, that is, whether the coverage of the mission point by the space unit is met. For a given mission, the satellite has defined the location information of the mission point. So read the dynamic position information in the structural unit and determine the covered multiplicity. Since the minimum number of units required for each mission type is constant, the phase mission is judged to be successful when the number of units that can be covered can reach the minimum value specified by the mission. The phase mission fails if the number of available systems is not enough.

4.2.2 *Simulation process of unplanned interrupt repair event*

Maintenance event is also an important part of the availability simulation and is the necessary event in the simulation of the real operation of the navigation system. the simulation process of unplanned interrupt repair event for the space segment is shown in Figure 7.

If the satellite is interrupted, first determine whether the satellite can receive the ground signal. If it cannot, it is judged as a long-term hard fault, belonging to the unplanned long-term interrupt type, where the satellite needs to be replaced. It is necessary to determine whether there are backup satellites for the replacement satellites.

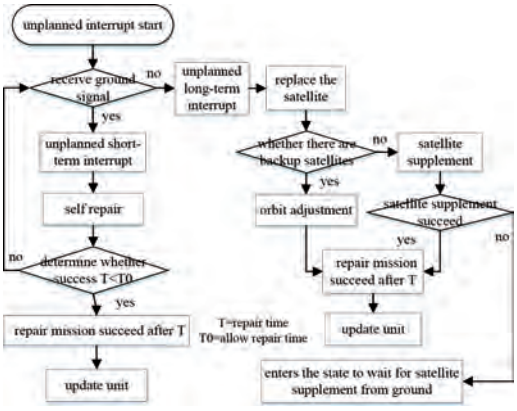


Figure 7. Simulation process of unplanned interrupt repair.

If it does not met, make a request to ground for satellite supplement. If the satellite supplement fails, it enters the state to wait for satellite supplement from ground. If the satellite supplement succeeds, the event of satellite supplement complete will be sent out after the time T, and the state unit will be updated at the same time. If there are backup satellites for the replacement satellites, then the satellites in the orbit will be subject to orbital position adjustment, and a request to ground for satellite supplement is made, and then it enters the state to wait for satellite supplement from ground. If it can receive the ground signal, the fault type is short-term interruption, belonging to short-term unplanned interruption caused by short-term hard faults. At this point, the satellite will be subject to self-repair, if the repair fails, the satellite will be redetected whether it can receive terrestrial signals. If the repair is successful, the event of repair complete will be sent out after the time T, and the state unit will be updated at the same time.

5 CASE ANALYSIS

Taking GPS as the target, Beijing area was selected to perform the single-point positioning mission and STK/MATLAB hybrid simulation method was used to simulate and analyze the service availability of GPS in Beijing. The latitude and longitude coordinates of Beijing is (40, 116), and the ground elevation is 20 m.

The data of GPS was derived from (U.S. Department of Defense. 2008). Failure data and maintenance data were gave based on experience. It is assumed that one backup satellite in B, D and F orbit and two backup satellites in ground launch base. The availability simulation model of GPS was built according the method mentioned in this paper. Setting the frequency of the simulation as 10 times, the simulation time as July 01 2007 to June 30 2015 and the simulation step as 1 minute, STK/MATLAB hybrid simulation was conducted. The trajectory of each unit is shown as below:

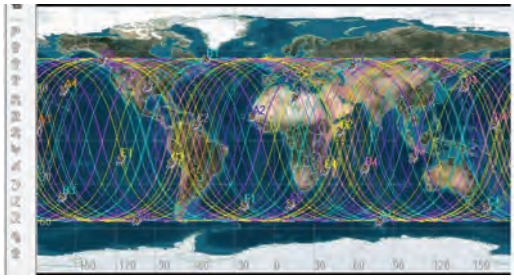


Figure 8. GPS satellite orbit ground trace in some state.

Minimum N Asset Coverage
Satisfaction: At Least 4
FOM value range check is not enabled.

N Asset Coverage for YH

Time (UTCG)	FOM Value
1 Jul 2007 12:00:00.000	9
1 Jul 2007 12:01:00.000	9
1 Jul 2007 12:02:00.000	8
1 Jul 2007 12:03:00.000	8
1 Jul 2007 12:04:00.000	8
1 Jul 2007 12:05:00.000	8
1 Jul 2007 12:06:00.000	8
1 Jul 2007 12:07:00.000	8
1 Jul 2007 12:08:00.000	8
1 Jul 2007 12:09:00.000	8
1 Jul 2007 12:10:00.000	8
1 Jul 2007 12:11:00.000	7
1 Jul 2007 12:12:00.000	7
1 Jul 2007 12:13:00.000	7
1 Jul 2007 12:14:00.000	7
1 Jul 2007 12:15:00.000	8
1 Jul 2007 12:16:00.000	8
1 Jul 2007 12:17:00.000	8
1 Jul 2007 12:18:00.000	8
1 Jul 2007 12:19:00.000	8

Figure 9. Quantity of visible GPS satellite in Beijing.

During the simulation, the available time section of each satellite to the ground mission area can be obtained. The satellite coverage multiplicity in each time section in Beijing is obtained by statistics, shown in Figure 9. Finally, we got the coverage multiplicity data list for each event. The single point service availability of the statistical system in each state is obtained by Matlab simulation.

The availability statistics model is as below:

$$\Phi_i=\begin{cases} 0, & \text{F-N} < 0 \\ 1, & \text{F-N} \geq 0 \end{cases} \tag{1}$$

$$A_o = \frac{\sum_{i=1}^{n} \Phi_i \cdot \Delta t}{T} \tag{2}$$

where F is the actual coverage multiplicity, N is the minimum coverage multiplicity required for the success of the mission, and is the available state function of the unit.

The GPS satellite navigation system availability was obtained to be 99.95% by statistics.

6 CONCLUSION

Based on the characteristics of satellite navigation system and considering the operation, interruption, maintenance and support of navigation system, the framework of availability simulation model of GNSS was established and the detailed model including space ground state, interruption, maintenance, support model and the relationship model were described. The availability simulation logic flow based on operational process were also developed, which supports the availability simulation of GNSS. Combined with the simulation analysis data, it can provide basis for improvement of operational program, and development of system dynamic interrupt plan and in-orbit backup scheme, in order to increase the system availability.

REFERENCES

Hou, Hongtao & Xie, Fei. 2014. Constellation usability analysis of navigation system based on Markov process [J]. *Systems Engineering and Electronics,* 36(4):685–690.

Joo, Inone & Lee, Jeom-Hun & Kim, Jae-Hoon. 2007. Availability Analysis of COMS Satellite Ground Control System Considering Infant Mortality[A]. *25th AIAA International Communications Satellite Systems Conference.*

Li, Zuohu & Hao, Jinming. 2010. Satellite Navigation System Monitoring Station Coverage Performance Analysis and Layout Method [A]. *CSNC first China Satellite Navigation Academic Annual Conference.*

Liang, Guilin & Zhou, Xiaoji. 2017. Modeling and Simulation of Remote Sensing Satellite Ground System Architecture Based on DoDAF [J]. *Command Control and Simulation,* 39(2):105–112.

U.S. Department of Defense. 2008. G1oba1 Positioning System Standard Positioning Service Performance Standard [R].

Xiang, Junlin & Zhang, Yulin. 2007. Design of Backup Strategy Based on Satellite Reliability and MTTR Constellation [J]. *Systems Engineering and Electronics,* 29(9):1576–1580.

Yang, Zhuopeng & Zheng, Heng. 2017. A Method of System Effectiveness Modeling and Analysis of Navigation Satellite [J]. *Journal of Aeronautics,* 38(6):647–654.

Zhao Guangyan & Sun, Yufeng & Hu, Weiwei & Qin, Tong. 2013 Study on Constellation Single-slot Availability [A]. *2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*, 286–289.

Zhao Guangyan & Sun, Yufeng. 2014. The availability parameters system of satellite navigation system [A]. *The 60th Annual Reliability and Maintainability Symposium.*

Zhou, Shanshan & Jiao, Jian. 2014. Constellation Usability Modeling and Simulation Based on Single Star Reliability [J]. *Computer Application,* 344–347.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A safe flow-management method for air traffic considering the UAS presence into the non-segregated airspace

Euclides Carlos Pinto Neto, Derick M. Baum, Marco A. Brinati, Jorge R. Almeida Jr., Paulo S. Cugnasca & João B. Camargo Jr.
School of Engineering, University of São Paulo (Poli - USP), São Paulo, Brazil

ABSTRACT: Air transportation is essential for society and it is increasing gradually. Furthermore, new technologies that improve the airspace operation in terms of safety are under development, such as the Unmanned Aircraft Systems (UAS). In the past few years, there has been a growth in UAS numbers into the segregated airspace. However, there are many challenges to be faced in order to integrate these autonomous aircraft into the non-segregated airspace. In this context, the relationship between UAS and Air Traffic Controller (ATCo) is a important aspect due to the fact that the presence of UAS into the non-segregated airspace may represent an increase in ATCo workload and, ultimately, a reduction in safety levels. This increase results from the lack of familiarity of the ATCo with the UAS, which leads these professionals to perform additional activities (e.g. cognitive activities) in order to conduct the air traffic safely. In fact, the Technology Maturity Level (TML), which is a measurement system that models the level of familiarity of ATCo with a particular aircraft, shows the impacts on workload levels of different aircraft. The main goal of this research is to propose a flow-management method for balancing the workload of sectors in order to reduce the impact on safety levels of the UAS presence in the non-segregated airspace. This workload balancing method considers the TML of Manned Aircraft (MA) and UAS, which may increase (in case of UAS) or decrease (in case of MA) the ATCo workload. The results show that this method can be employed for balancing workload effectively, even considering the UAS presence, and that integrating a small number of UAS into non-segregated airspace may present acceptable safety levels from the ATCo workload perspective.

1 INTRODUCTION

The increasing importance of air transportation for society (Marquart, Ponater, Mager, & Sausen 2003) has been considered an enabler for the development of technologies that improve the airspace operation, such as Unmanned Aircraft Systems (UAS) and Decision Support Tools (DST) for Air Traffic Controllers (ATCos) (e.g. Arrival and Departure managers) (Noskievič & Kraus 2017). These new technologies are proposed for improving airspace operation from many perspectives, such as efficiency and capacity. The DSTs, for instance, ensure the Air Traffic Controller (ATCo) the decisions made are effective, which lead to a reduction on his/her workload (i.e. the time spent on controlling aircraft) and, ultimately, on airspace complexity from ATCo perspective (Majumdar & Ochieng 2002). Although these technologies are used in many cases, they may bring uncertainties due to the fact that the personnel (e.g. ATCos) are not familiar to deal with them.

Furthermore, in the past few years, there has been a growth in UAS numbers into segregated airspace (Guerin 2015). These aircraft are composed of subsystems such as Unmanned Aircraft Vehicle (UAV), payload, control station and communication subsystems (Fasano, Accado, Moccia, & Moroney 2016) (Austin 2011) and have several militaries and civil applications (e.g. firefighting). However, the integration of these aircraft into the non-segregated airspace may lead to the creation of new ways of reaching the wellknown unsafe states. For instance, human mistakes in paloting can be made by bug in software. As the acceptance of this new technology is increasing due to its advantages compared to manned aircraft (e.g. efficiency), it stands out as a challenge for ATCos due to the lack of familiarity.

Moreover, the workload of ATCo is a result of the interaction of several factors, including the Air Traffic Control (ATC) complexity (Mogford, Guttman, Morrow, & Kopardekar 1995), and challenging situations for ATCos are related to a higher workload (Meckiff, Chone, & Nicolaon 1998), i.e., ATCo workload is related to safety (Neto, Baum, Hernandez-Simes, Almeida, Camargo, & Cugnasca 2017) as well as complexity, which is one of

the main factors that impact on ATCo workload (Majumdar & Ochieng 2002).

In this sense, the Air Traffic Flow Management (ATFM) enables the Air Traffic Management (ATM) to be effective in terms of safety, efficiency, cost-effectiveness, environmental sustainability and interoperability of ATM systems (ICAO 2014). The ATFM is conducted by a ATC cell. For instance, in European airspace, ATFM activities are carried out by Eurocontrols Network Manager Operations (Alam, Chaimatanan, Delahaye, & Féron 2017). However, there are many challenges in terms of flow management, such as the unpredictability of weather conditions and unexpected delays. Furthermore, the interaction between the UAS and the ATCo may present impacts, in terms of safety, in the airspace. Considering a high air traffic density and that a higher workload level leads to a lower airspace capacity (Majumdar & Polak 2001), the lack of adaptation of the sectors' capacity due to the presence of the UAS during the ATFM process may compromise the safety levels.

The main goal of this research is to propose a flow-management method for balancing the workload of sectors in order to reduce the impact on safety levels of the UAS integration in the non-segregated airspace. This workload balancing method considers the level of familiarity of ATCos with Manned Aircraft (MA) and UAS, which presents a lower (in case of UAS) or higher (in case of MA) impact on the ATCo workload.

This paper is organized as follows: Firstly, Section 2 presents the related works. Secondly, Section 3 presents the aspects of flow management into the airspace. Thirdly, the ATCo workload considering the relationship between ATCo and UAS is presented in Section 5. After that, Sections 4, 6 and 7 present, respectively, the aspects of UAS, the method and the case studies adopted in this research. Then, Section 8 presents a discussion on the results achieved in the experiments. Finally, Section 9 shows the conclusions of this research.

2 RELATED WORKS

In (Alam, Chaimatanan, Delahaye, & Féron 2017), the authors present a distributed air traffic flow management model for addressing the four-dimensional trajectory planning over the European Functional Airspace Blocks (FAB), which is a concept adopted in European airspace for allowing cooperation for improving the air traffic flow in an efficient and safe manner. Thus, this distributed approach of ATFM enables the information sharing between airspace blocks in strategic planning minimizing interaction between trajectories. Furthermore, this method is implemented and tested with a real air traffic data over the European airspace and interaction-free 4D trajectories are produced in short computational time (according to the time constraints of this operation). The results showed that this distributed method is viable and the interaction is reduced. In addition to this contribution, we propose a ATFM method for balancing the ATCo workload over the ATC system. Furthermore, we consider the presence of UAS into the non-segregated airspace.

In (Ivanov, Netjasov, Jovanovi, Starita, & Strauss 2017), the authors propose a two-level mixed-integer optimization model for solving the en-route demand-capacity imbalance problem in order to explore the possibility of controlling the ATFM delay distribution. Thus, the minimization of the delay propagated to subsequent flights can be achieved. Considering that this research comes from a practical background and that the model proposed is compatible with the models that are being used in the real-world, tests are conducted in realistic experiments. In order to mitigate the effects of delays in flights, aircraft operators usually embed a buffer time in their schedules. The current practice for allocating ATFM delays does not consider, though, if flights are able to absorb ATFM delay and still reduce delay propagation to subsequent flights, i.e., if the flights have any remaining schedule buffer. The results show that the proposed method reduces the delay propagated to subsequent flights and improves airport slot adherence. However, although this is an interesting contribution in terms of ATFM and delay reduction, this research does not consider the UAS presence.

The authors in (Neto, Baum, Almeida, Camargo, & Cugnasca 2017) present a simulation tool that aims evaluate safety (from workload perspective) and efficiency in aircraft sequencing in final sector considering the UAS presence. In this paper, a novel approach for measuring the impacts of the integration of UAS into the non-segregated airspace is proposed. This approach, called the Technology Maturity Level (TML), models the level of familiarity of ATCo with these aircraft. The realistic results achieved in the experiments showed that, depending on the familiarity of ATCo with the UAS, these aircraft may present a considerable impact in the workload levels and, ultimately, in the airspace capacity. However, although the authors present an approach for integrating the UAS into the non-segregated airspace, aspects of Air Traffic Flow Management (ATFM) are not taken into account.

The authors in (Clothier, Denney, & Pai 2017) propose a manner to create a Risk Informed Safety Case (RISC) applied to the context of safety assurance of UAS operation. This approach aims to

facilitate safe and cost-effective operations of sUAS by presenting the comprehensive measures considered in order to eliminate, reduce, or control the safety risk. The RISC proposed is composed by barrier models of safety, which support the development of safety measures, and structured arguments, which provide assurance of safety in operations (through, for instance, appropriate evidence). The authors also propose a model for UAS operational risk, which considers, for instance, specific hazards (e.g. collision) and operational risks (depending on the UAS). Ultimately, this paper shows key safety-related assurance concerns to be addressed and the development of a layered framework for reasoning about those concerns, which can be useful for regulators and various stakeholders in justifying confidence in operational safety in the context of the absence of the relevant aviation regulations for UAS.

3 AIR TRAFFIC FLOW MANAGEMENT (ATFM)

According to (ICAO 2014), the ATFM is "an enabler of Air Traffic Management (ATM) efficiency and effectiveness. It contributes to the safety, efficiency, cost-effectiveness, and environmental sustainability of an ATM system". ATFM enables the airspace to operate smoothly and resiliently, considering the possible difficulties that may be faced (e.g. bad weather conditions). The planning of the traffics movements (e.g. scheduling) helps the ATC units to, collaboratively, adapt the airspace for current needs in a global view.

Among the objectives of the ATFM, we can highlight (ICAO 2014): (1) improvement of the safety of the ATM system; (2) Optimization of flow in all phases of flights; (3) Simplification of collaboration between stakeholders (e.g. ATC units and airlines); (4) A better understanding of the ATM system resource constraints in terms economic and environmental priorities.

The approach of involving more stakeholders in the planning is a strategic manner of predicting future problems (e.g. unexpected delays and conflicts), mitigating risks (e.g. by using a delay buffer in each flight, regions with bad weather conditions can be avoided by the aircraft) and avoiding unsafe states. Thus, the traffic planning enables balanced definitions on the trade-off between safety and efficiency. As efficiency is, naturally, a metric to be maximized (due to, for instance, profit increase), the flow in planned in this sense. On the other hand, the impacts on the safety levels of the airspace must be taken into account in the decision-making process. Finally, safety represents the most important restriction in this context, i.e., the efficiency is desired to be maximized but the safety must remain at an acceptable level.

4 UNMANNED AIRCRAFT SYSTEM (UAS) INTEGRATION INTO THE NON-SEGREGATED AIRSPACE

UAS, in which the interest of engineering community has increased in the past few years (Fasano, Accado, Moccia, & Moroney 2016), is an autonomous system composed of subsystems (e.g. communication system and control station) (Fasano, Accado, Moccia, & Moroney 2016) (Austin 2011). There are many advantages provided by the UAS in large scale (e.g. airspace efficiency improvement) and in small-scale (e.g. reduction of risks associated with pilots in applications, such as firefighting).

Nowadays, the UAS that are being built vary in terms of size and can be classified into three categories in terms of weight (Romero & Gomez 2017): (1) Small UAS; (2) Medium UAS; (3) Large UAS. The class One (small UAS) represents the UAS that has many applications in smaller scenarios, with weight less or equal to 149 kg. Class Two (medium UAS) weights up to 600 kg. Finally, large UAS weights more than 600 kg. In this research, as we consider a futuristic scenario and the impacts of UAS are measured from the workload perspective instead of performance perspective (e.g. speed difference between large aircraft and small UAS), large UAS with size and performance similar to the commercial aircraft are adopted.

In order to control the aircraft, UAS and Manned Aircraft (MA), throughout the airspace, a set of activities must be performed by the ATCo (Neto, Baum, Hernandez-Simes, Almeida, Camargo, & Cugnasca 2017). In this research, activities are proposed by specialists of Safety Analysis Group (GAS), of the University of São Paulo. These specialists are actual Air Traffic Controllers and have more than 10 years of hands-on experience. The activities performed by the ATCo in each sector and their duration (in seconds) are: (1) First Contact (10s); (2) Instruction (15s); (3) Surveillance (10% of the mean time spent by the aircraft within the sector); (4) Communication with adjacent ATC unit (10s);

In this context, the Technology Maturity Level (TML), which is a measurement system that measures the familiarity between the ATCo and the aircraft, is employed (Neto, Baum, Almeida, Camargo, & Cugnasca 2017). Aircraft with higher TMLs are related to operations with lower workload levels, whereas aircraft with lower TMLs are related to operations with higher workload levels. For instance, nowadays, it is reasonable to consider that the UAS have a lower TML, whereas

the Manned Aircraft have a higher TML. In terms of workload, it is reasonable to consider that the familiarity of the ATCo with MA is higher than the familiarity of the ATCo with UAS. This is due to the fact that the ATCos, nowadays, are used to deal with MA but are not used to deal with UAS (which do not operate in the non-segregated airspace) (Złotowski, Yogeeswaran, & Bartneck 2017). In this context, it is reasonable to consider that the UAS has a Technology maturity Level (TML) equals to 0 and MA equals to 10. Thus, the time spent in the activities performed by the ATCo in the MA operation is multiplied by 1, whereas the time spent in the activities performed by the ATCo in UAS operation is multiplied by 2 (Neto, Baum, Almeida, Camargo, & Cugnasca 2017).

5 WORKLOAD BALANCING MODEL (WBM)

This section presents the main contribution of this research, the ATCo Workload Balancing Model (AWBM) for safe Air Traffic Flow Management (ATFM), from the workload perspective, considering the UAS presence. Firstly, the flow network employed in the AWBM is presented. Then, aspects of the allocation process are highlighted. Finally, the final considerations are shown.

5.1 *Flow network*

The airspace is composed by (but not limited to) sectors and tracks. The aircraft that operate into the non-segregated airspace use these track in order to fly to different regions. Naturally, these flights lead the aircraft to cross a set of sectors, which are controlled by ATC units. Each ATC unit is responsible for one sector, i.e., although a collaborative work is conducted, metrics of each ATC unit are independent (e.g. ATCo workload) and, in order to maintain the safety levels, the ATCo workload of a given sector must not exceed the maximum acceptable workload established. Finally, a reasonable maximum acceptable workload, i.e., a reasonable workload threshold, represents 80% of an hour (Majumdar & Polak 2001) (Majumdar, Ochieng, Bentham, & Richards 2005). The safety levels may be, then, compromised if the the current workload exceeds the workload threshold.

Figure 1 illustrates a region of the airspace and its respective set of sectors. Each sector has its own ATC unit and, consequently, its own workload threshold. Note that there are points (b_1, b_2, b_3, …, b_8) that connect the sectors according to the tracks that represent the sectors boundaries, which are logical elements that identify the region in which the aircraft changes from one sector to another. In order to abstract this scenario, a flow network is

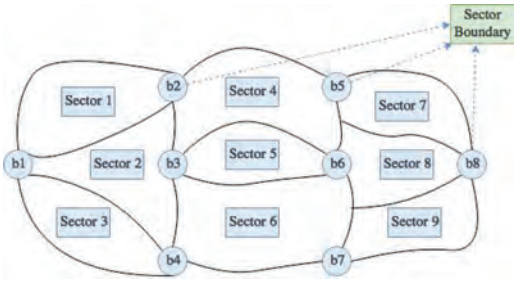


Figure 1. Region of the airspace and its respective set of sectors.

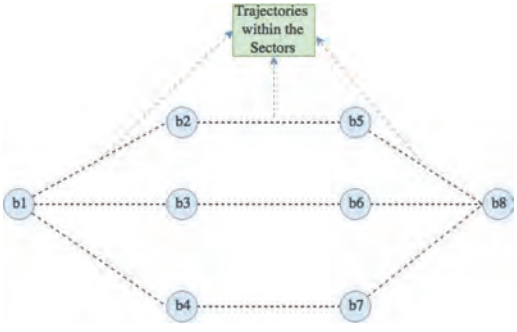


Figure 2. Flow network used for representing the scenario presented in Figure 1.

appropriate due to its suitability for dealing with problems in which goods (in this case, aircraft) must be sent from one point to another, considering specific constraints (in this case, workload and safety). In order to model the problem considered in this research, the flow network can be faced as a graph that considers a flow, a capacity, source and sink nodes, goods (i.e. aircraft) to be delivered and specific constraints (Ford & Fulkerson 1962).

Suppose a undirected graph G with nodes b_1, b_2, …, b_n and edges e_1, e_2, …, e_m. The nodes represent the sector boundaries, i.e., each node is a logical element that illustrates the point in which the aircraft changes from one sector to another. One should note that this point refers to the change of ATC unit that controls the aircraft due to the fact that each ATC unit is responsible for one sector. Furthermore, the edges represent the direct trajectory from one sector to another. Each trajectory has a particular length. Finally, a fleet composed by a set of aircraft (MA and UAS), a_1, a_2, …, a_n, must be allocated to paths that connect the source and the target of the network, i.e., origin and destination of each aircraft.

Figure 2 illustrates the flow network used by the AWBM for representing the scenario presented in

Figure 1. In this scenario, each trajectory has a specific length. One should note that as a trajectory is within a sector and a sector has a limited workload threshold[1], there is a limited number of aircraft that can operate simultaneously in each trajectory.

5.2 *Allocation process*

The scheduling problem faced in this research is constituted by allocating the set of aircraft to selected paths. This allocation process is aimed to balance the ATCo workload levels in the airspace in order to maintain a proper safety level in all sectors. For instance, if all aircraft are delivered to a single path, that path is expected to have a high ATCo workload level whereas other possible paths have a low ATCo workload level, i.e., the workload is unbalanced and some sectors may present critical scenarios in terms of safety. Firstly, in order to select a path for a given aircraft going from node b_k to the node $b_{k'}$, all possible paths are found in the graph. Secondly, an evaluation of the ATCo workload of the sectors that composed each path is conducted in order to identify which path presents the lowest workload level. Finally, the path with the lowest sum of ATCo workload is selected and the aircraft is allocated to that path.

Once the aircraft is allocated, the attributes of the path are updated. For each sector, the ATCo workload is updated with the ATCo workload related to controlling that aircraft, which depends on its Technology Maturity Level (TML). This research is focused on the planning of flights, i.e., the flights assigned to paths are expected to generate a certain level of ATCo workload and, thus, controlling the amount of traffic that fly in each path is an effective manner of balancing the workload.

Thus, this challenge can be faced as a optimization problem: Equation 1 shows the function that must be minimized in order to balance the workload in the airspace. In fact, this equation measures how different the workload of the different sectors are from each other, i.e., scenarios with higher variations of workload present a higher value in this equation. Note that the minimum value achieved is 0, which is achieved in the case in which the ATCo workloads of all sectors are the same. The W function measures the workload for a given sector, which is represented by an edge e_m. Finally, this equation sums the absolute value of the subtraction of the workload of all sectors. However, the restrictions of this problem must be respected. Equation 2 illustrates a restriction that highlights that the actual workload of a given sector, represented by the edge e_i, must be less or equals to the workload limit of its workload threshold (Wt). This restriction aims to ensure that the safety levels, from the workload perspective, are maintained at an acceptable level. Furthermore, the Equation 3 shows that each aircraft must be assigned to one path as a restriction. Solutions of this optimization problem, considering specific scenarios, represent solutions of the ATFM considering the UAS presence.

$$\min \quad f = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \left| W(e_i) - W(e_j) \right| \tag{1}$$

$$W(e_i) \leq Wt(e_i) \quad i = 1,2,3,\ldots,m \tag{2}$$

$$Paths(a_k) = 1 \quad k = 1,2,3,\ldots,n \tag{3}$$

5.3 *Final considerations*

Finally the ATCo Workload Balancing Model (AWBM) has some considerations on the scenarios modeled. Firstly, the distance between the paths is expected to be similar. Secondly, the performance of the aircraft is expected to be similar and bad weather conditions are not present in the scenarios. Finally, unsafe states are not expected, i.e., the aircraft minimum separation is respected during all flights.

6 METHOD

This section presents the steps that guide the application of our proposal in different problems. Note that this approach can be applied in scenarios of different sizes (e.g. the number of sectors) and type of aircraft (e.g. MA and UAS). This process is illustrated in Figure 3.

Firstly, the input data is provided. This input data is composed of the list of aircraft with the same source and target nodes. Furthermore, the type of each aircraft must be provided, i.e., each aircraft must be identified as Manned Aircraft (MA) or Unmanned Aircraft System (UAS). This information is used in the workload evaluation. Secondly, the scenario is described. This description is based on the transformation of a scenario into a flow network, which contains sectors (represented by edges) and sectors boundaries

Figure 3. Method adopted in this research.

1. The workload threshold is the maximum workload supported by a given sector in order to maintain the safety levels.

(represented by nodes). In fact, this phase identifies the characteristics of the situation that is used in the model.

Thirdly, The adjustment process is conducted, which is conducted in order to make sure that the description represents a viable scenario, i.e., if the flow network considered connects the source to the sink. After that, the scenario building is conducted. This process aims to build a flow network pragmatically based on the description provided. This scenario is, then, validated by checking if all aspects of the model are properly defined. If the validate is conducted successfully, the experiments can be conducted, otherwise, the adjust is conducted again. The experiments constitute the process of allocating the aircraft to the available tracks in a suitable manner. In this phase, the data the workload levels of each sector are estimated. Finally, the evaluation aims to identify the insights on the results in order to show the impact of our proposal in the workload balancing.

7 CASE STUDIES

This section presents the case studies considered in this research. Firstly, a simpler and abstract scenario is presented. After that, a more complex scenario is shown. Finally, a scenario based on a realistic scenario is presented.

7.1 *Case study I*

The main goal of this case study is to show the applicability of our proposal in a simpler scenario, in which the fleet is reduced. We consider a fleet of 5 aircraft, in which 2 are UAS (UAS_1 and UAS_2) and 3 are MA (MA_1, MA_2 and MA_3). In this scenario, illustrated in Figure 4, the aircraft must be sent from the node b_1 to the node b_8.

The characteristics of the tracks of this scenario are illustrated in Table 1. In this Table, the distances, in Nautical Miles, the mean time of the aircraft in that track (or sector) T and the ATCo workload threshold, both in seconds. The workload threshold represents the maximum ATCo workload supported by each sector, are presented for each track.

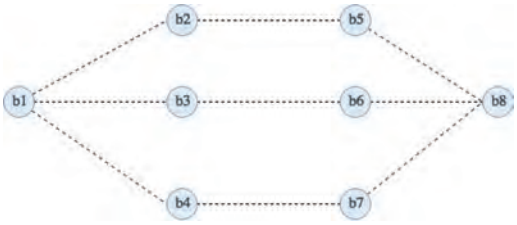


Figure 4. Scenario adopted in case study I.

Table 1. Characteristics of tracks in case study I.

Track	Distance (nm)	T (s)	Threshold (s)
b_1 - b_2	52	624	681,8
b_1 - b_3	45	540	623
b_1 - b_4	51	612	673.4
b_2 - b_5	60	720	749
b_3 - b_6	58	696	732.2
b_4 - b_7	57	684	723.8
b_5 - b_8	48	576	648.2
b_6 - b_8	43	516	606.2
b_7 - b_8	51	612	673.4

Table 2. Solution proposed in case study I.

Aircraft	Path
MA_1	b_1 - b_2 - b_5 - b_8
MA_2	b_1 - b_3 - b_6 - b_8
MA_3	b_1 - b_2 - b_5 - b_8
UAS_1	b_1 - b_4 - b_7 - b_8
UAS_2	b_1 - b_3 - b_6 - b_8

During the allocation process, the solution achieved is illustrated in Table 2. In this solution, the aircraft are allocated in a balanced manner throughout the network in order to share the ATCo workload.

7.2 *Case study II*

The main goal of this case study is to show the applicability of our proposal in a more complex scenario, in which the fleet is larger than the fleet considered in the case study I. We consider a fleet of 10 aircraft, of which 5 are UAS (UAS_1, UAS_2, …, UAS_5) and 5 are MA (MA_1, MA_2, …, MA_5). In this scenario, illustrated in Figure 5, the aircraft must be sent from the node b_1 to the node b_{10}.

Furthermore, the characteristics of the tracks of this scenario are illustrated in Table 3. Similarly as presented in the case study I, the distances, in Nautical Miles, the mean time of the aircraft in that track (or sector) T and the ATCo workload threshold (s), both in seconds. The workload threshold represents the maximum ATCo workload supported by each sector, are presented for each track.

The results achieved in the allocation process, illustrated in Table 4, showed that the aircraft are sent from the node b_1 to the node b_{10} in a balanced manner, i.e., many paths composed by different tracks are explored.

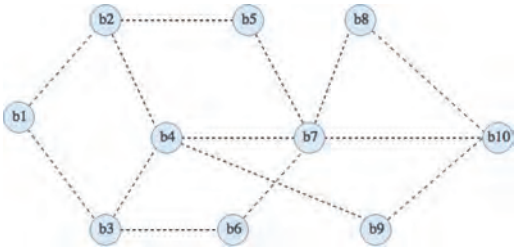


Figure 5. Scenario adopted in case study II.

Table 3. Characteristics of tracks in case study II.

Track	Distance (nm)	T (s)	Threshold (s)
b_1 - b_2	52	624	1353,8
b_1 - b_3	45	540	623
b_2 - b_4	57	684	723.8
b_2 - b_5	55	660	597.8
b_3 - b_4	42	504	707
b_3 - b_6	49	588	656.6
b_4 - b_7	49	588	656.6
b_4 - b_9	80	960	917
b_5 - b_7	57	684	723.8
b_6 - b_7	57	684	723.8
b_7 - b_8	57	684	723.8
b_7 - b_{10}	65	780	791
b_8 - b_{10}	62	744	765.8
b_9 - b_{10}	58	696	732.2

Table 4. Solution proposed in case study II.

Aircraft	Path
MA_1	b_1 - b_2 - b_4 - b_7 - b_8 - b_{10}
MA_2	b_1 - b_2 - b_5 - b_7 - b_{10}
MA_3	b_1 - b_2 - b_4 - b_7 - b_8 - b_{10}
MA_4	b_1 - b_3 - b_4 - b_9 - b_{10}
MA_5	b_1 - b_2 - b_4 - b_9 - b_{10}
UAS_1	b_1 - b_3 - b_4 - b_9 - b_{10}
UAS_2	b_1 - b_3 - b_6 - b_7 - b_{10}
UAS_3	b_1 - b_2 - b_5 - b_7 - b_{10}
UAS_4	b_1 - b_3 - b_6 - b_7 - b_8 - b_{10}
UAS_5	b_1 - b_3 - b_4 - b_7 - b_{10}

7.3 *Case study III*

The third case study presents a realistic scenario present in the Brazilian airspace. We consider a fleet of 20 aircraft, in which 10 are UAS (UAS_1, UAS_2, …, UAS_{10}) and 10 are MA (MA_1, MA_2, …, MA_{10}). In this scenario, illustrated in Figure 6, the aircraft must be sent from the node *belém* (*bl*) to the node *santarém* (*sr*). In this realistic scenario, there are 11 tracks and their characteristics are

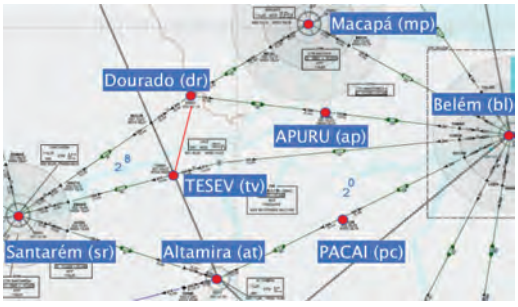


Figure 6. Scenario adopted in case study III adapted from (DECEA 2017).

Table 5. Characteristics of tracks in case study III.

Tracks	Distance (s)	T (s)	Threshold (s)
bl - mp	178	2136	1740.2
mp - dr	108	1296	1152.2
dr - sr	162	1944	1605.8
dr - tv	60	720	749
tv - sr	124	1488	1286.6
bl - ap	150	1800	1505
ap - dr	105	1260	1127
bl - tv	223	2676	2118.2
bl - pc	126	1512	1303.4
pc - at	106	1272	1135.4
at - sr	162	1944	1605.8

Table 6. Solution proposed in case study III.

Aircraft	Path
MA_1	*bl - mp - dr - sr*
MA_2	*bl - pc - at - sr*
MA_3	*bl - pc - at - sr*
MA_4	*bl - pc - at - sr*
MA_5	*bl - ap - dr - sr*
MA_6	*bl - ap - dr - sr*
MA_7	*bl - mp - dr - sr*
MA_8	*bl - pc - at - sr*
MA_9	*bl - pc - at - sr*
MA_{10}	*bl - pc - at - sr*
UAS_1	*bl - ap - dr - tv - sr*
UAS_2	*bl - tv - sr*
UAS_3	*bl - mp - dr - sr*
UAS_4	*bl - tv - sr*
UAS_5	*bl - pc - at - sr*
UAS_6	*bl - tv - sr*
UAS_7	*bl - ap - dr - sr*
UAS_8	*bl - tv - sr*
UAS_9	*bl - mp - dr - sr*
UAS_{10}	*bl - tv - sr*

illustrated in Table 5 (the distance and the mean time of the aircraft in the track).

The results of the allocation, presented in Table 6, shows that all available tracks are used, i.e., the ATCo workload is shared among the 11 sectors.

8 DISCUSSION

In the first case study, the ATCo workload level in each track is illustrated in Figure 7. Note that the workload levels of all tracks are much lower than their respective workload threshold, i.e., they support more traffics (Figure 7 and Table 1). However, this balanced manner of aircraft distribution maintains the different sectors with similar workload levels. Finally, the result of the unbalancing measure (Equation 1) is 3094.8s. Note that, for instance, if the only the path use is $b_1 - b_2 - b_5 - b_8$ and the ATCo workload threshold is respected, the unbalancing measure is 21390.6s, which is considerably higher than the value achieved by the proposed solution and, consequently, indicates a possibility of safety issues in those sectors.

The second case study presents a higher number of tracks and the workload of each track is illustrated in Figure 8. Although the path $b_7 - b_{10}$ presented a workload (756s) similar to its workload threshold (791s), the method proposed in this research ensures that the threshold is respected. Finally, the result of the unbalancing measure (Equation 1) is 30614.1s. One should note that this value may be considerably higher depending on the paths the aircraft are allocated to, i.e., if the workload levels of the tracks are very different, this measure increases.

Finally, the third case study also presented balanced results, which are illustrated in Figure 9.

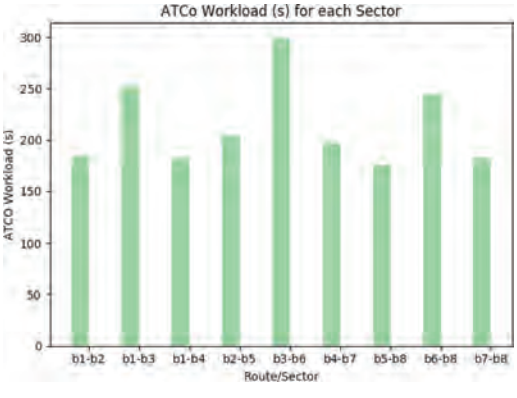


Figure 7. ATCo workload level of each track (case study I).

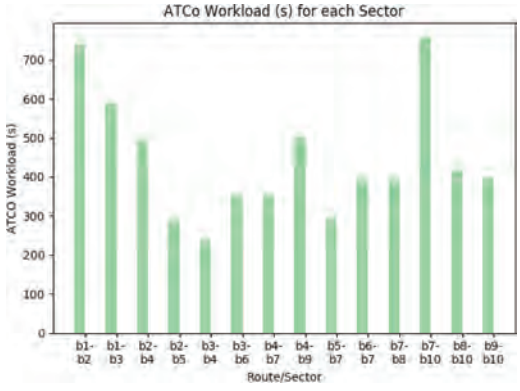


Figure 8. ATCo workload level of each track (case study II).

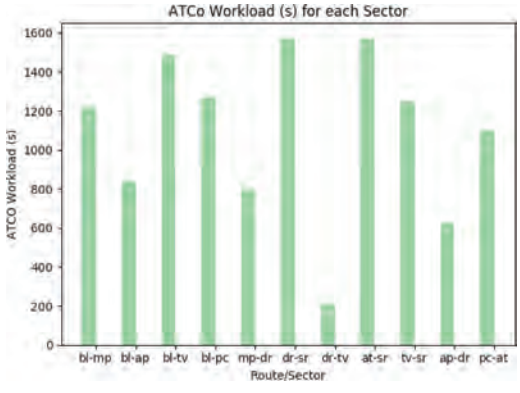


Figure 9. ATCo workload level of each track (case study III).

Note that the values vary considerably. However, this solution presents a balanced solution compared to solutions that use few of these paths. Furthermore, all sectors present a workload level below its respective workload threshold (presented in Table 5). Finally, the unbalancing measure of this case study achieved 51096s. Note that other solutions may present a considerably higher unbalancing measure, for instance, if all aircraft are sent through the path $bl - tv - sr$ respecting the ATCo workload threshold, which represents 72018.8s of unbalancing measure and may indicate problems, from the safety perspective, in those sectors.

9 CONCLUSION

In this paper, a safe flow-management method for integrating the UAS safely into the non-segregated airspace was presented. The method aims

to distributes the workload among the sectors in order to maintain the safety levels. The results achieved in the proposed experiments showed that our method distributes the workload among the sectors effectively, even considering the UAS presence. As future intentions, the authors aim to consider aspects such as delays buffer and resilience in case of failures in the UAS operation (e.g. failures in the command and control link).

ACKNOWLEDGEMENT

The authors would like to thank Boeing Research & Technology Brazil (BR&T-Brazil) for the support for this research and for its institutional support to the Safety Analysis Group (GAS) of the School of Engineering of the University of São Paulo (Poli-USP).

REFERENCES

Alam, S., S. Chaimatanan, D. Delahaye, & E. Féron (2017). A distributed air traffic flow management model for European functional airspace blocks. In *12th ATM R&D Seminar*.

Austin, R. (2011). *Unmanned aircraft systems: UAVS design, development and deployment*, Volume 54. John Wiley & Sons.

Clothier, R., E. Denney, & G. Pai (2017). Making a risk informed safety case for small unmanned aircraft system operations. *Safety (TLOS) 3*, 4.

DECEA (2017). Enroute Chart (ERC) - L2. Available in: https://www.aisweb.aer.mil.br/cartas/rotas/_enrc-l1_-enrc_20171012.pdf?CFID=2ddf8f52-fa4b-4bd4-bd7a-0dbf2f9425adCFTOKEN=0. Accessed in: December 2017.

Fasano, G., D. Accado, A. Moccia, & D. Moroney (2016, November). Sense and avoid for unmanned aircraft systems. *IEEE Aerospace and Electronic Systems Magazine 31*(11), 82–110.

Ford, L.R.J. & D.R. Fulkerson (1962). Flows in networks. In *Princeton University Press, Princeton, N.J.*

Guerin, D. (2015). Consideration of wake turbulence during the integration of remotely piloted aircraft into the air traffic management system. In *Unmanned Aircraft Systems (ICUAS), 2015 International Conference on*, pp. 926–935. IEEE.

ICAO (2014). Manual on collaborative air traffic flow management - doc 9971.

Ivanov, N., F. Netjasov, R. Jovanovi, S. Starita, & A. Strauss (2017). Air traffic flow management slot allocation to minimize propagated delay and improve airport slot adherence. *Transportation Research Part A: Policy and Practice 95*(Supplement C), 183–197.

Majumdar, A. & W. Ochieng (2002). Factors affecting air traffic controller workload: Multivariate analysis based on simulation modelling of controller workload. *Transportation Research Record: Journal of the Transportation Research Board* (1788), 58–69.

Majumdar, A., W.Y. Ochieng, J. Bentham, & M. Richards (2005). En-route sector capacity estimation methodologies: An international survey. *Journal of Air Transport Management 11*(6), 375–387.

Majumdar, A. & J. Polak (2001). Estimating capacity of europe's airspace using a simulation model of air traffic controller workload. *Transportation Research Record: Journal of the Transportation Research Board* (1744), 30–43.

Marquart, S., M. Ponater, F. Mager, & R. Sausen (2003). Future development of contrail cover, optical depth, and radiative forcing: Impacts of increasing air traffic and climate change. *Journal of Climate 16*(17), 2890–2904.

Meckiff, C., R. Chone, & J.-P. Nicolaon (1998). The tactical load smoother for multi-sector planning. In *Proceedings of the 2nd USA Europe air traffic management research and development seminar*.

Mogford, R.H., J. Guttman, S. Morrow, & P. Kopardekar (1995). The complexity construct in air traffic control: A review and synthesis of the literature. Technical report, DTIC Document.

Neto, E.C.P., D.M. Baum, J.R. Almeida, J.B. Camargo, & P.S. Cugnasca (2017). Evaluating safety and efficiency in aircraft sequencing in final approach considering the USA presence. In *XXXI Congresso de Pesquisa e Ensino em Transportes*.

Neto, E.P., D.M. Baum, C.E. Hernandez-Simes, J.R. Almeida, J.B. Camargo, & P.S. Cugnasca (2017, June). An airspace capacity-based safety assessment model considering USA integration. In *2017 International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 961–970.

Noskievič, T. & J. Kraus (2017). Air traffic control tools assessment. *MAD-Magazine of Aviation Development 5*(2), 6–10.

Romero, J. & L. Gomez (2017). Proposal for rpas integration into non-segregated airspaces. In *Integrated Communications, Navigation and Surveillance Conference (ICNS), 2017*, pp. 6C2-1. IEEE.

Złotowski, J., K. Yogeeswaran, & C. Bartneck (2017). Can we control it? autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies 100*, 48–54.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Active power dispatch strategy of wind farms under generator faults

K. Ma & J. Zhu & M. Soltani & A. Hajizadeh
Department of Energy Technology, Aalborg University, Esbjerg, Denmark

P. Hou & Z. Chen
Department of Energy Technology, Aalborg University, Aalborg, Denmark

ABSTRACT: Generator is a critical component of the wind turbine. It transforms the mechanical energy to the electric energy and transmits electricity to the grid. Based on the previous experience, the fault rate of the generator is fairly high which increases the downtime of the turbine. It has a significant effect on the performance and the economic benefit of the wind farm. The traditional wind farm controller adopts the proportional dispatch strategy and does not consider the fault severity. In this paper, a power dispatch strategy of wind farm focusing on the fault condition of the generator is proposed. The main faults are divided into three levels according to their severities. The wind farm controller dispatches power references to wind turbines onthe basis of fault level, wind condition and power demand from the grid. When the fault level is high, the strategy gives priority to protect the generator. If the fault level is not high, the wind farm should follow the power demand. In addition, the wake effect is also an important factor that must be taken into account in the wind farm control. Because it can affect the power production directly. To verify the advantage of this strategy, it is compared with the proportional dispatch strategy in the simulation. The result shows that the proposed strategy can make a good trade-off between component protection and power production.

1 INTRODUCTION

Wind energy is one of the most widely used renewable energies. With the rapid development of technology and industry application, it plays an important role in the energy market of the world. By 2016, wind energy overtook coal and became the second largest form of power generation capacity in Europe (Europe 2016). Wind Farm (WF) is seen as a mature and effective form to utilize wind energy for commercial operation. The large-scale offshore WF is the development trend of wind energy. This also implies that the wind energy penetration in the grid is larger and larger. However, wind energy has its congenital defects. One of them is the high failure rate of the Wind Turbine (WT) due to the harsh operating environment and the highly turbulent wind speed. Therefore, the stability and reliability of WF are more and more important to the grid. The fault of WT should not be neglect in the operation of WF as well.

In recent years, wind farm control draws more and more attention of the researchers. Previous Wind Turbine Control (WTC) can no longer guarantee the good performance of WF during operation, no matter from the aspects of power production, fatigue loads, etc. For the coordination, control and management of all WTs in the entire WF, it is apparent that Wind Farm Control (WFC) is more direct and effective. The research on WFC can be divided into two categories depending on whether participating in the individual turbine control. One is to achieve the control objectives of WF through adjusting WTC from the WF level (Ebrahimi et al. 2016, Tian et al. 2017, Gonzalez et al. 2013). While the other is opposite, it only coordinates and distributes the variables that are not involved in WTC such as start/stop and power reference (Soleimanzadeh and Wisniewski 2011, Soleimanzadeh et al. 2012, Zhang et al. 2015). The former method seems to be able to get better results due to more controllable variables. However, the latter method does not need to change the original controller from the manufacturer, it is easier to apply in the practice. Normally, the objectives of WFC includes power maximization, fatigue and wake reduction, frequency response, voltage control, following some specific requirements from TSO, etc.

The generator faults will influence the power output of WT directly and even result in the emergency stop. Although the failure rate of the generator is not the highest in all components, the average fault-removal time is quite long and the maintenance cost is high as well. Most research on generator faults focuses on Fault Detection and Diagnosis (FDD). The common methods need to detect some physical quantities such as voltage, current, power and then analyze the signals based on the model or signal-processing (Balasubramanian and Muthu 2017, Qiao and Lu 2015). However,

there are few articles considered generator faults in WFC. Some similar research on the fault of blade and drive-train can be found in (Badihi et al. 2015, Odgaard et al. 2009).

The analysis of generator faults shows that some faults can be mitigated by reducing load, such as turn-to-turn short circuit (Lešić et al. 2013). The down-regulation of WT can not only prevent the faulty generator from further damage but also reduce the downtime.

The contribution of this paper is to distribute power demand to the individual WTs reasonably when the generator fault occurs. According to the severity and mechanism, we divided the faults of Doubly-Fed Induction Generator (DFIG) into three levels. The proposed power dispatch strategy will take different distribution type depending on the fault level and try to fulfil the power demand under the premise of ensuring WT's safety, assuming that all faults can be detected.

This paper is organized as follows. Section 2 describes DFIG faults briefly and the fault classification. The WF model with wake effect is presented in section 3. Section 4 gives the active power dispatch strategy of WF. Section 5 provides the simulation results of different cases.

2 FAULT CLASSIFICATION

DFIG is a widely used generator in WT because of its low capital cost and good energy yield (Hansen and Hansen 2007). The components with high fault rate include slip ring/brush, bearing, cooling system, winding insulation, encoder, etc (Shipurkar 2015). Although DFIG has many different types of fault modes, the only control variable in this paper is the power reference which can adjust the electrical load of WT. So we focus more on the fault modes that can be mitigated by reducing the load. The characteristic of the fault according to Fault Tree Analysis (FTA) and Failure Modes Effect Analysis (FMEA) is used for reference. In addition, the safety of WT is always the most important. Fault severity is another factor need to be considered. The fault classification is as follows:

Fault level 1 (FL1): this fault level includes the faults that are not serious or cannot be mitigated by down-regulation. For example, some fault of the redundant sensor, minor rotor misalignment and bearing vibration.

Fault level 2 (FL2): this fault level includes the early inter-turn short-circuit faults of the stator and rotor, as well as the initial fault of the cooling system. The characteristic of this level is that the overheating caused by fault can have a more serious effect on the generator. However, down-regulation operation can reduce generator heating by decreasing the current. For protecting generator, there is a limit value of power $P_{limit,f}$. This value depends on the specific fault mode and means that the load of WT should not exceed it. The specific value of $P_{limit,f}$ can be estimated according to the fault mechanism and lifetime estimation.

Fault level 3 (FL3): this level has the highest severity, such as the phase to phase short-circuit of stator and rotor. It will result in the generator failure directly. Therefore, WT must be shut down when the fault of this level is detected for protecting WT.

3 WIND FARM MODEL

WF consists of several WTs. However, simply taking the ambient wind speed as the wind speed of all the WTs, then adding all power output of WTs as the power output of WF is not accurate due to the aerodynamic interaction. To show the results of power dispatch strategy, a WF model that taking wake effect into account is used. WF model includes three parts: WT model, wake model and WF layout.

3.1 *Wind turbine model*

A 5MW DFIG WT model is adopted here. According to the Betz theory, the mechanical power extracted by the turbine from wind energy can be calculated by the following equations:

$$P_m = \frac{1}{2}\rho\pi R^2 v^3 C_p(\lambda,\beta) \tag{1}$$

$$\lambda = \frac{\omega_r R}{v} \tag{2}$$

$$F_t = \frac{1}{2}\rho\pi R^2 v^2 C_t(\lambda,\beta) \tag{3}$$

where, P_m is the mechanical power extracted by turbine; ρ is the air density; R is the radius of rotor; v is the wind speed; C_p is the power coefficient, which depends on λ and β; λ is the tip speed radio; β is the pitch angle; ω_r is the rotor speed; F_t is the thrust force; C_t is the thrust coefficient, which depends on λ and β as well. The surfaces of C_p and C_t are shown in in Figure 1.

Pitch system adjusts the aerodynamic characteristic of blades by controlling the pitch angle. It can change the efficiency of energy capture and offer aerodynamic brake. Gearbox supports the necessary speed conversion for DFIG. Generation system includes a DFIG and a partial scale power electronic converter.

These subsystems are the major parts of DFIG WT and will be simplified as an inertial system.

The control strategies of normal operation are Maximum Power Point Tracking (MPPT) and Constant Power in low and high wind speed region respectively. The down-regulation strategy should be emphasized here because it can affect the wake.

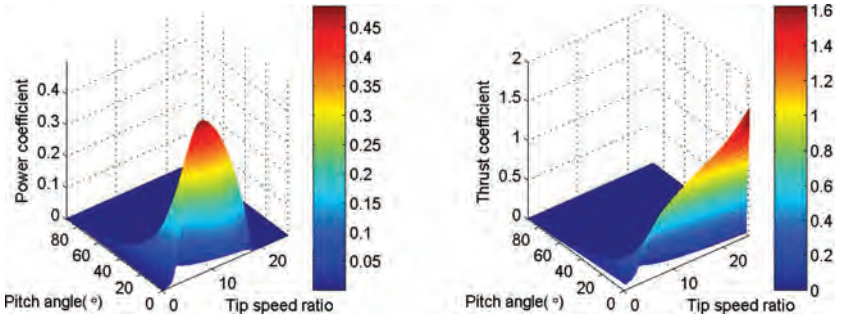


Figure 1. Surfaces of C_p and C_t.

The strategy adopted here is Max ω, which is the most widely used in the industry (Mirzaei et al. 2014).

3.2 *Wake model*

Wake effect can decrease the power production of WF and increase the fatigue load of downwind WT. Therefore, it should be considered in WFC. There are several research focus on how to describe wake effect accurately (Göçmen et al. 2016). Among engineering applications, Jensen wake model is widely used because its practicality and simplicity. In this paper, we adopt Jensen wake model for the single wake and quadratic summation method for the multiple wakes (Katic, Højstrup, & Jensen 1986). The velocity deficit can be expressed as:

$$1-\frac{v}{u}=\frac{1-\sqrt{1-C_t(\lambda,\beta)}}{(1-2\alpha X/D)^2} \tag{4}$$

where, v is the downwind wind speed at position X; u is the ambient free wind speed; X is the distance between upwind and downwind WT; D is the diameter of rotor; α is decay constant, choosing 0.05 for offshore WF.

To calculate the velocity deficit of the j^{th} WT in multiple wakes, the following equation can be used.

$$\left(1-\frac{v_j}{u}\right)=\sum_{i=1}^{n}(1-\frac{v_{ij}}{u})^2 \tag{5}$$

where, v_j is the wind speed of the j^{th} WT; n is the number of wakes that the j^{th} WT is in; v_{ij} is the wind speed of the j^{th} WT under the influence of the wake of the i^{th} WT.

3.3 *Wind farm layout*

In order to reflect the wake effect to WF and to simplify the calculation, five WTs are arranged in a straight line. The selected wind direction makes the downwind WTs are in the full wake of the upwind WTs. This layout allows the study of the worst-casewake effect. The distance between two WTs is 6.5 times the diameter of the rotor. The layout is shown in Figure 2.

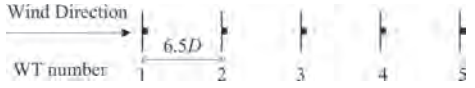


Figure 2. Wind farm layout.

4 ACTIVE POWER DISPATCH STRATEGY WITH FAULT ACCOMMODATION

The traditional active power dispatch strategy is proportional dispatch strategy (Grunnet, Soltani, Knudsen, Kragelund, & Bak 2010). Power demand is distributed to WTs proportionally to the available power of each WT by the WF controller as follows:

$$P_{a,i}=\frac{1}{2}\rho\pi R^2 v^3 C_{p,max} \tag{6}$$

$$P_{r,i}=\frac{P_{a,i}}{\sum P_{a,i}}P_{dem} \tag{7}$$

where, $P_{a,i}$ is the available power of the i^{th} WT; $C_{p,max}$ is the maximum coefficient of power of WT; $P_{r,i}$ is the power reference of the i^{th} WT; P_{dem} is the power demand of WF from TSO.

The traditional proportional strategy neither considers the impact of the wake effect on power production, nor does it consider the fault of WT. The proposed strategy is also based on the proportional dispatch strategy. But it takes generator fault classification and wake effect into consideration, and tries to follow the power demand ensuring WT's safety as the precondition.

The whole strategy according to the fault level is also divided into three parts:

1. Strategy for FL1: The fault in FL1 is not severe and cannot be remitted by down-regulating

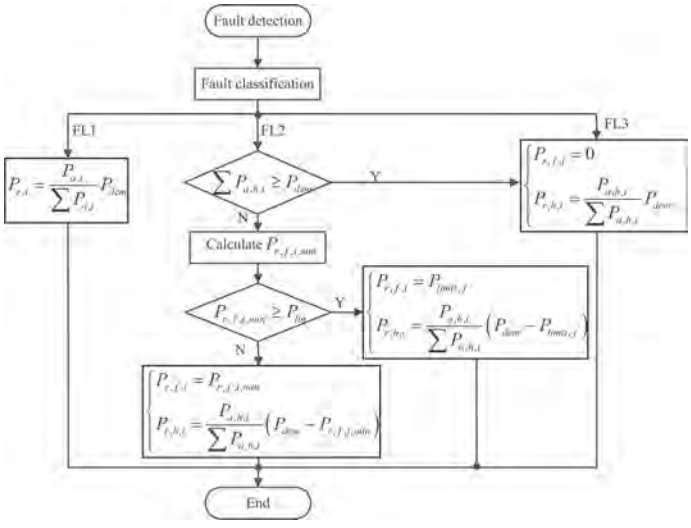


Figure 3. Flowchart of the proposed active power dispatch strategy.

WT. Therefore, the active power dispatch will maintain the original proportional strategy for both healthy WT and faulty WT.

2. Strategy for FL2: This strategy is the focus of this paper as it relates to the trade-offs between fault protection and following power demand. The general idea includes three steps.

Step 1. If the sum of the power of healthy WTs can follow the Power demand enough, it is allowed to shut down the faulty turbine. Then the power demand will be distributed as follows:

$$\begin{cases} P_{r,f,i} &= 0; \\ P_{r,h,i} &= \dfrac{P_{a,h,i}}{\sum P_{a,h,i}} P_{dem}. \end{cases} \tag{8}$$

where, $P_{r,f,i}$ is the power reference of the i^{th} faulty WT; $P_{r,h,i}$ is the power reference of the i^{th} healthy WT; $P_{a,h,i}$ is the available power of the i^{th} healthy WT.

Step 2. If the sum of the power of healthy WTs cannot follow the Power demand. The minimum power reference of the faulty WT that can fulfil the power demand should be found by the exhaustive method. There are two possible reasons why the power demand can still be followed by reducing the power of a particular WT. One is that the available power of WF is higher than power demand. The other reason is that, due to the wake effect, the down-regulation of upwind WT will increase the wind speed of the downwind WTs, and the power loss because of down-regulation will be compensated by the downwind WTs. The wind speeds of WTs are highly coupled due to wake effect and are related to the power references of all the WTs. Therefore, the exhaustive method is the simplest and quickest way.

Step 3. If $P_{r,f,i,min}$ is higher than $P_{limit,f}$, the power reference of faulty WT will be set as $P_{limit,f}$. So the power output of WF must be less than the power demand. This power deviation is ineluctable for protecting WT. The power distribution will be as follows:

$$\begin{cases} P_{r,f,i} &= P_{limit,f}; \\ P_{r,h,i} &= \dfrac{P_{a,h,i}}{\sum P_{a,h,i}} \left(P_{dem} - P_{limit,f} \right). \end{cases} \tag{9}$$

If $P_{r,f,i,min}$ is smaller than $P_{limit,f}$, the power reference of faulty WT will be set as $P_{r,f,i,min}$. In this way, the strategy can follow power demand and reduce the load of faulty WT as much as possible. The power distribution will be as follows:

$$\begin{cases} P_{r,f,i} &= P_{r,f,i,min}; \\ P_{r,h,i} &= \dfrac{P_{a,h,i}}{\sum P_{a,h,i}} \left(P_{dem} - P_{r,f,i,min} \right). \end{cases} \tag{10}$$

3. Strategy for FL3: The generator fault with high severity is a serious threat to WT's safety. Therefore, the WT with FL3 must be shut down as soon as possible. The power distribution is the same with equation 8. The whole strategy is shown in the flowchart in Figure 3.

5 SIMULATION

To verify the advantage of the proposed active power dispatch strategy, the WF model mentioned in Section 3 is used. The parameters of a 5MW DFIG WT are shown in Table 1. Figure 2 shows the

distribution of wind speed under wake effect in this WF. The ambient wind speed of WT1 is 10 m/s. All power references are set to 5MW as MPPT strategy. The distribution shows that the wind speed of the downwind WT drops due to the wake effect. The following simulation will also use this WF model.

The strategies for FL1 and FL3 will not be simulated here. Because these two parts can be understood easily and there is no difference with the existing protection measure. To illustrate the fault protection, the phase current of faulty WT, $I_{N,f}$ is used in the comparison. Because all of these faults has the relationship with the current. With respect to the simulation of strategy for FL2, $P_{limit,f}$ is set to 2 MW. Three cases will be studied here to show the simulation results in different situations.

In Case 1, wind speed is 12 m/s, and power demand is 15MW. An FL2 fault occurs on WT3. The sum of power of healthy WTs can follow the Power demand enough, so the faulty WT can be shut down. It will not affect the power demand tracking. The power output of each WT is shown in Fig. 4. The red and blue bars represent the power output in Traditional Proportional Strategy (TPS) and Proposed Proportional Strategy (PPS) respectively. The power references, Power output of WF and the phase current of faulty WT are shown in Table 2. The result shows that the faulty WT is shut down, but WF can still produce enough power to follow the power demand. The faulty WT is also protected from further deterioration.

In Case 2, wind speed is 12 m/s, and power demand is 17.5MW. An FL2 fault occurs on WT5. The healthy WTs cannot supply enough power. Through the exhaustive method, the value of $P_{r,f,i,min}$ is 2.28MW in this case. It is higher than $P_{limit,f}$. According to the proposed strategy, although WF power cannot follow the power demand, in the consideration of protecting generator, the faulty WT has to be limited to 2MW. The result is shown in Table 3. The phase current of faulty WT has been decreased from 2.70kA to 1.71kA.

In Case 3, wind speed is 10 m/s, and power demand is 10MW. An FL2 fault occurs on WT3. The healthy WTs cannot fulfill the power demand either. The value of $P_{r,f,i,min}$ in this case is 0.73MW and much lower than $P_{limit,f}$. Therefore, the faulty WT can be down-regulated to 0.73MW. Meanwhile, the power demand can also be followed. The result

Table 1. Parameters of a 5MW wind turbine.

Parameter	Value
Rated Power	5 MW
Rotor Diameter	126 m
Min. and Max. Rotor Speed	6.9 rpm, 12.1 rpm
Cut-in, Rated, Cut-out Wind Speed	3 m/s, 11.4 m/s, 25 m/s
Gearbox Ratio	97:1
Synchronous Frequency	50 Hz
Electrical Generator Efficiency	94.4%
Number of Pole-pairs	3

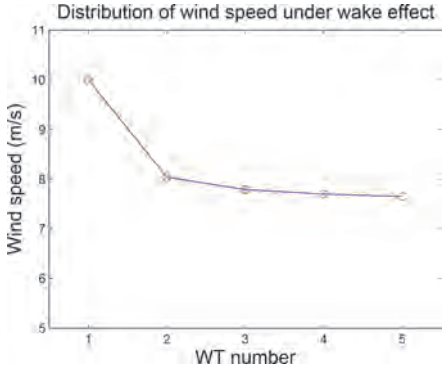


Figure 4. Distribution of wind speed in WF.

Table 2. Comparison of two strategies in Case 1.

	TPS	PPS
$P_{r,1}$(MW)	3.19	4.01
$P_{r,2}$(MW)	3.16	3.71
$P_{r,3}$(MW)	2.93	0
$P_{r,4}$(MW)	2.87	3.97
$P_{r,5}$(MW)	2.85	3.31
P_{WF}(MW)	15.06	15.06
$I_{N,3}$(kA)	2.50	0

Table 3. Comparison of two strategies in Case 2.

	TPS	PPS
$P_{r,1}$(MW)	4.08	3.49
$P_{r,2}$(MW)	3.75	1.81
$P_{r,3}$(MW)	3.31	0.73
$P_{r,4}$(MW)	3.20	2.26
$P_{r,5}$(MW)	3.16	1.71
P_{WF}(MW)	17.57	10.00
$I_{N,3}$(kA)	2.70	1.71

Table 4. Comparison of two strategies in Case 3.

	TPS	PPS
$P_{r,1}$(MW)	2.77	3.49
$P_{r,2}$(MW)	1.94	1.81
$P_{r,3}$(MW)	1.82	0.73
$P_{r,4}$(MW)	1.77	2.26
$P_{r,5}$(MW)	1.75	1.71
P_{WF}(MW)	10.04	10.00
$I_{N,3}$(kA)	1.55	0.62

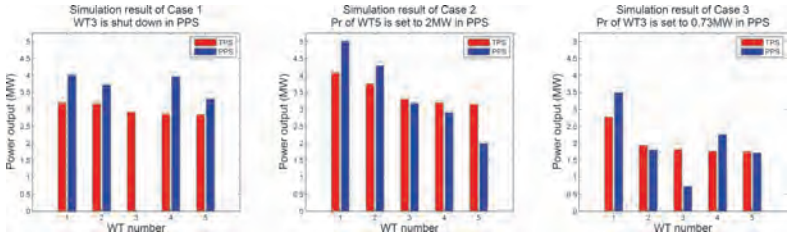


Figure 5. Comparison of power output of WTs in two strategies.

can be found in Table 4. The phase current of faulty WT can be decreased from 1.55kA to 0.62kA.

Figure 5 shows the power output of WTs in these three cases. From the simulation results, it can be seen that the proposed active power dispatch strategy can effectively ensure the safety of WT in the event of a generator fault, and follow the power demand as much as possible.

6 CONCLUSIONS

In this paper, an active power dispatch strategy of WF under generator fault is proposed. It gives the idea of implementing power distribution based on fault level. With the comparison of traditional proportional strategy, the simulation results show that the proposed strategy can make a good balance between generator protection and power production. It not only reduce the downtime caused by non-severe fault, but also follow the power demand as much as possible under the premise of ensuring WT's safety. The general idea can also be extended to some other faults in WT.

However, the shortcoming is lack of detailed research on the specific fault mode and the reliability evaluation. The fault model, lifetime estimation of generator and reliability model should be studied to combine the failure, condition, power production and reliability together in the further research.

REFERENCES

Badihi, H., Y. Zhang, & H. Hong (2015). Active fault tolerant control in a wind farm with decreased power generation due to blade erosion/debris build-up. *IFAC-PapersOnLine 48*(21), 1369–1374.

Balasubramanian, A. & R. Muthu (2017). Model based fault detection and diagnosis of doubly fed induction generators–a review. *Energy Procedia 117*, 935–942.

Ebrahimi, F., A. Khayatiyan, & E. Farjah (2016). A novel optimizing power control strategy for centralized wind farm control system. *Renewable Energy 86*, 399–408.

Europe, W. (2016). Wind in power: 2016 european statistics. Technical report.

Gonzalez, J.S., M.B. Payán, & J.R. Santos (2013). Optimal control of wind turbines for minimizing overall wake effect losses in offshore wind farms. In *EUROCON, 2013 IEEE*, pp. 1129–1134. IEEE.

Grunnet, J.D., M. Soltani, T. Knudsen, M.N. Kragelund, & T. Bak (2010). Aeolus toolbox for dynamics wind farm model, simulation and control. In *The European Wind Energy Conference & Exhibition, EWEC 2010*.

Göçmen, T., P. Van der Laan, P.-E. Réthoré, A.P. Diaz, G.C. Larsen, & S. Ott (2016). Wind turbine wake models developed at the technical university of denmark: A review. *Renewable and Sustainable Energy Reviews 60*, 752–769.

Hansen, A.D. & L.H. Hansen (2007). Wind turbine concept market penetration over 10 years (1995–2004). *Wind energy 10*(1), 81–97.

Katic, I., J. Hojstrup, & N.O. Jensen (1986). A simple model for cluster efficiency. In *European wind energy association conference and exhibition*, pp. 407–410.

Lešić, V., M. Vašak, N. Perić, G. Joksimović, & T.M. Wolbank (2013). Fault-tolerant control of a wind turbine with generator stator inter-turn faults. *automatika 54*(1), 89–102.

Mirzaei, M., M. Soltani, N.K. Poulsen, & H.H. Niemann (2014). Model based active power control of a wind turbine. In American Control Conference (ACC), 2014, pp. 5037–5042. IEEE.

Odgaard, P.F., J. Stoustrup, &M. Kinnaert (2009). Fault tolerant control of wind turbines–a benchmark model. *IFAC Proceedings Volumes 42*(8), 155–160.

Qiao, W. & D. Lu (2015). A survey on wind turbine condition monitoring and fault diagnosispart ii: Signals and signal processing methods. *IEEE Transactions on Industrial Electronics 62*(10), 6546–6557.

Shipurkar, U. (2015). Wind turbine generator systems failuresprobabilities and mechanisms. Technical report.

Soleimanzadeh, M. & R. Wisniewski (2011). Controller design for a wind farm, considering both power and load aspects. *Mechatronics 21*(4), 720–727.

Soleimanzadeh, M., R.Wisniewski, & S. Kanev (2012). An optimization framework for load and power distribution in wind farms. *Journal of Wind Engineering and Industrial Aerodynamics 107*, 256–262.

Tian, J., D. Zhou, C. Su, F. Blaabjerg, & Z. Chen (2017). Optimal control to increase energy production of wind farm considering wake effect and lifetime estimation. *Applied Sciences 7*(1), 65.

Zhang, J.-h., Y.-q. Liu, D. Tian, & J. Yan (2015). Optimal power dispatch in wind farm based on reduced blade damage and generator losses. *Renewable and Sustainable Energy Reviews 44*, 64–77.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Probabilistic assessment of the impact of connecting a new distributed generation unit to a potentially congested power system

J. Sun & P.E. Labeau
Université Libre de Bruxelles, Brussels, Belgium

A. Vergnol
Elia System Operator, Brussels, Belgium

ABSTRACT: The large-scale use of distributed electricity generation is possibly associated with increasing grid congestion; hence, Distributed Generation (DG) curtailment is envisioned to solve those congestions in affected power systems. By resorting to curtailment as part of an Active Network Management (ANM) scheme under real-time supervision/control of the DG units, more distributed generation can be accommodated in a distribution grid, while keeping the power system secure. In this paper, a method, based on an importance sampling scheme systematically targeting electrically challenging scenarios, is given to assess the curtailment risk associated to DG unit connections to the grid. It simultaneously resorts to correlated sampling in order to estimate the risk evolution due to "perturbed" situations, i.e. due to a newly connected DG unit. This significantly reduces the computation time of the risk indicators. The effectiveness of the proposed method is demonstrated on a test power grid.

1 INTRODUCTION

The European Union (EU) has set very aggressive emission reduction targets, establishing a 20% reduction in greenhouse gases with respect to 1990 levels by 2020, and a 80% reduction and 100% clean electricity by 2050 (European Climate Foundation Roadmap 2050, 2010). In this context, a lot of distributed electricity generation is expected to be part of the future power systems. This requires additional investments in the network, including expensive and time-consuming reinforcements and modernization of the power system infrastructure. In addition, the variations in power generation and the interconnection may constitute barriers to achieving an optimal system for connecting distributed sources to the grid. Therefore, the adaptation of current grid planning to the new situation in order to facilitate the integration of Distributed Generation (DG) units is more affordable (Do, M.T., Francois, B. 2017).

However, the connection of large amounts of DG units to the power system introduces many technical and economic challenges. Reverse power flows, leading to line congestions and voltage problems, are likely to take place when the non-locally consumed power is injected to the high-voltage (HV) grid (Faghihi, F., Labeau, P.E. et al. 2014). As a consequence,Transmission System Operators (TSOs) will need to operate the grid by means of an Active Network Management (ANM) scheme in (almost) real-time control (Järventaustac, P. et al. 2009), by possibly curtailing their production in case of grid congestion in order to maintain grid security and reliability.

In general, power system reliability assessment focuses on the evaluation of some indicators (such as system average interruption duration index, interruption frequency index and expected energy not supplied) to reflect the ability to supply adequate electric service on the long term, while Probabilistic Risk Assessment (PRA) aims to estimate the probability or frequency of disturbances to system operation and their consequences, both of these two elements being the constituents of the risk (Rocchetta, R. et al. 2015). In this study, indicators related to the expected curtailment are envisioned, by propagating on a grid model the uncertainties on the loads and productions. As a direct Monte Carlo Sampling (MCS) turns out to be extremely time-consuming, an alternative approach was proposed, consisting in decoupling the probabilistic assessment of possibly challenging situations from the corresponding curtailment assessment using an Optimal Power Flow (OPF) (Labeau, P.E., Faghihi, F. et al. 2014).

This methodology rests on the concept of net balance (algebraic sum of all productions minus total load in a substation). The net balance state space associated to the grid under study is discre-

tized and the cases corresponding to this mesh are first systematically analyzed in order to identify unsafe cells with respect to possible grid congestion. Then, this database of unsafe cells is updated based on possible congestion in the MV/HV transformers. The curtailment of DG units is calculated, based on a targeted (systematic) importance sampling of only those detailed variants (all individual productions and loads) likely to lead to congestion, i.e. those corresponding to net balances belonging to unsafe cells. Risk indices can therefore be calculated efficiently. However, this approach only allows calculating the risk indices of interest for a global set of q DG units already connected to the grid. The present work aims to develop an accurate way of estimating the impact of a (q + 1)th DG unit to be connected to a node of the grid, on the previous value of the risk indices for the grid with q DG units. Resorting to correlated sampling, one computation will simultaneously bring the risk estimation for a reference case (with q + 1 DG units) and for both a first perturbed case (the previous situation with q DG units) and for a second perturbed case (corresponding to a different installed capacity of the (q + 1)th unit), for each season (associated to a given rating of the elements) and each grid topology considered (N and N-1 conditions).

The paper is organized as follows. Section 2 defines the risk indices and the mathematical theory supporting their estimation. Section 3 presents the application of this methodology to the grid. A test case with realistic power grid data is used to illustrate the approach. Results and analyses are given in Section 4 and conclusion is in Section 5.

2 RISK INDEX CACULATION

The classical definition of risk is adopted as the product of the probability of occurrence of the undesired event (i.e. contingency) and the related consequence (i.e. severity). To take into account all undesired events, the definition is extended by summing all contributions as:

$$RI = \sum_i P(E_i) \times f(E_i) \tag{1}$$

where $P(E_i)$ is the probability of occurrence of the undesired event E_i and $f(E_i)$ is the severity of the related consequence. In probabilistic risk assessment, contingency frequencies are used as probabilities and severity functions as consequences.

2.1 *Risk index in a grid with DG units*

In the context of power systems, a contingency is defined as the unexpected loss of one/more electric elements (e.g. line, cable, or transformer) that the power system is made of. Overloads, influenced by seasons, related with the feeders' thermal limits, and bus voltage magnitude, related with frequency and system balance, are both indicators of power system stress and correspond to the consequences for the risk calculation (P.A. Gooding et al. 2014). Thus, the risk index associated with one contingency can be expressed as follows for the whole power network:

$$RI(C_k/\chi) = \sum_{l=1}^{n_s} q_l \times P_l(C_k/\chi) \times f_l(C_k,\chi) \tag{2}$$

where χ is the set of all seasonal conditions (e.g. summer, inter-season, winter). C_k is the kth contingency (e.g. N/N-1 configuration). q_l is the probability of the lth season. $f_l(C_k,\chi)$ is the performance function under congestion for electric element i in the conditions of the lth season. $P_l(C_k/\chi)$ is the probability of occurrence of congestion k. n_s is the total number of seasonal conditions, n_c is the total number of configurations. Let p_k be the probability of configuration k. The risk index due to all contingencies is, then, obtained as:

$$RI = \sum_{k=1}^{n_c} \sum_{l=1}^{n_s} p_k \times q_l \times P_{kl}(C_k/\chi) \times f_{kl}(C_k,\chi) \tag{3}$$

Suppose that the power grid under study comprises n substations, to which wind farms (WFs) and/or combined heat and power (CHP) units are connected. The total load in substation i is denoted L_i while the total production of DG units connected to node i is denoted by P_i. The total risk under all contingencies in a grid can be expressed as:

$$RI = \sum_{k=1}^{n_c} \sum_{l=1}^{n_s} p_k q_l \iint f_{kl}(\bar{L},\bar{P}) \varphi_{kl}(\bar{L},\bar{P}) d\bar{L} d\bar{P} \tag{4}$$

where $\bar{P}$ and $\bar{L}$ are the vectors of DG units generation and load, respectively. $\varphi_{kl}(\bar{P},\bar{L})$ is the joint probability density function(pdf) of the variants in configuration k and season l. $f_{kl}(\bar{P},\bar{L})$ represents the severity function associated to the curtailment solving a possible congestion. For a given situation kl, eq. (4) can be simplified as:

$$RI_{kl} = \iint f_{kl}(\bar{L},\bar{P}) \varphi_{kl}(\bar{L},\bar{P}) d\bar{L} d\bar{P} \tag{5}$$

In order to identify more easily possible congestions, the concept of net balance NB is used again (see section 1). Suppose that n_i is the total number of DG units connected to node i. The production of DG unit j connected to node i is denoted P_{ij}. The risk index formula now writes:

$$NB_i = \sum_{j=1}^{n_i} P_{ij} - L_i \tag{6}$$

$$RI = \iint f(\overline{L}(\overline{NB},\overline{P}),\overline{P})\theta(\overline{NB})\varphi(\overline{L}(\overline{NB},\overline{P}),\overline{P})d\overline{NB}d\overline{P} \tag{7}$$

where $\theta(\overline{NB})$ is an indicator function: if $\overline{NB}$ causes agrid congestion, it is equal to 1; otherwise it is 0. This will divide the net balance space into two regions, a safe region where there is no congestion and an Unsafe Region (UR) where congestion can occur. The UR is delimited based on a discretization of the accessible region in the net balance space. Hence the UR is bounded by n_{cel} unsafe cells (UC). Therefore, Eq. (7) can be written as:

$$\int ...d\overline{NB} = \sum_{c=1}^{n_{cel}} \int ...d\overline{NB} \tag{8}$$

$$RI = \int_{UR} d\overline{NB} \int f(\overline{L}(\overline{NB},\overline{P}),\overline{P})\varphi(\overline{L}(\overline{NB},\overline{P}),\overline{P})d\overline{P} \tag{9}$$

$$NB_{i,\min}^{UC} \le P_{w,i} + P_{CHP,i} - L_i \le NB_{i,\max}^{UC} \tag{10}$$

$$RI = \sum_{UC} \int d\overline{NB} \int f(\overline{L}(\overline{NB},\overline{P}),\overline{P})\varphi(\overline{L}(\overline{NB},\overline{P}),\overline{P})d\overline{P} \tag{11}$$

where $P_{w,i}$ is the total production of WFs connected to substation i; $P_{CHP,i}$ is the total generation of CHP units at node i; L_i is the total load at substation i. $NB_{i,min}^{UC}$ and $NB_{i,max}^{UC}$ are the extreme values of net balance for unsafe cell UC in the dimension corresponding to node i. The loads at the different nodes are correlated, and so are the wind productions, but correlations between wind productions and loads are negligible. CHP productions are independent of the other variables. Therefore, our risk index is calculated as follows:

$$\varphi(\overline{P}_w,\overline{P}_{CHP},\overline{L}) = \varphi_w(\overline{P}_w)\cdot\varphi_L(\overline{L})\cdot\prod_{i=1}^{n}\varphi(\overline{P}_{CHP,i}) \tag{12}$$

$$RI = \sum_{UC} \iiint f(\overline{P}_W,\overline{P}_{CHP},\overline{\mathrm{L}})\cdot\varphi(\overline{P}_W,\overline{P}_{CHP},\overline{\mathrm{L}})d\overline{P}_W d\overline{P}_{CHP}d\overline{\mathrm{L}} \tag{13}$$

where $\varphi(\overline{P}_w,\overline{P}_{CHP},\overline{L})$ is the total joint pdf of all variants, consisting of the product of the joint pdf $\varphi(\overline{P}_w)$ of all wind productions, the individual pdf's $\varphi(\overline{P}_{CHP,i})$ of all CHP generationsand the joint pdf $\varphi(\overline{L})$ of the loads. The detailed variant sampling procedure will be illustrated in section 3.3.

2.2 *Risk index under perturbed cases*

Considering that a new DG unit of a specific type and installed capacity is likely to be connected to a substation, its impact on the previously assessed risk indices should be estimated. As already mentioned, the risk indices presented above correspond to a global set of q DG units connected to the grid. The UR related to (q + 1) DG units is an extension of that obtained with q units, in the dimension corresponding to the node where this new unit could be connected. Additionally, a different installed capacity for this new DG unit means changing this extension range, see Eq. (14) and Eq. (15). It is thus possible to resort to correlated sampling to assess more accurately the difference in the risk indices between the three situations, considering the case with (q + 1) units under a certain capacity as the reference situation (*ref*), and the case with q units as perturbed situation no. 1 (*per1*), and the case with a newly connected DG unit with a smaller installed capacity as perturbed situation no. 2 (*per2*).

$$NB_{\min_i}{}^{(ref)} = NB_{\min_i}{}^{(per1)} + \sum_{j=1}^{n_i} P_{add}\min_{ij} \tag{14}$$

$$NB_{\max_i}{}^{(ref)} = NB_{\max_i}{}^{(per1)} + \sum_{j=1}^{n_i} P_{add}\max_{ij} \tag{15}$$

Of course, the joint pdf's of the uncertain loads and generations are affected by the addition of the new DG unit. The risk indices will thus be estimated according to the previous procedure for the reference case, and those associated to the perturbed cases are derived from a correlated sampling procedure. We can then write for the risk index definitions and Monte Carlo estimates on N sampled variants:

$$RI^{(reference)} = \iiint_{UC} f(\overline{P}_W,\overline{P}_{CHP},\overline{\mathrm{L}})\cdot\varphi^{ref}(\overline{P}_w,\overline{P}_{CHP},\overline{L}) \cdot d\overline{P}_w d\overline{P}_{CHP}d\overline{L} \tag{16}$$

$$RI^{(perturbed)} = \iiint_{UC} f(\overline{P}_W,\overline{P}_{CHP},\overline{\mathrm{L}})\cdot\frac{\varphi^{per}(\overline{P}_w,\overline{P}_{CHP},\overline{L})}{\varphi^{ref}(\overline{P}_w,\overline{P}_{CHP},\overline{L})} \cdot\varphi^{ref}(\overline{P}_w,\overline{P}_{CHP},\overline{L})d\overline{P}_w d\overline{P}_{CHP}d\overline{L} \tag{17}$$

$$\tilde{RI}^{(reference)} = \frac{1}{N}\sum_{j=1}^{N}[\sum_{UC} f^{ref}(\overline{P}_{W,i},\overline{P}_{CHP,i},\overline{\mathrm{L}}_i)\cdot\prod_{i=1}^{n}\varphi^{ref}(\overline{P}_{CHP,i})] \tag{18}$$

$$\tilde{RI}^{(perturbed)} = \frac{1}{N}\sum_{j=1}^{N}[\frac{\varphi_w^{per}(\overline{P}_w)}{\varphi_w^{ref}(\overline{P}_w)}\cdot\frac{\varphi_L^{per}(\overline{L})}{\varphi_L^{ref}(\overline{L})}\cdot\frac{\prod_{i=1}^{n}\varphi^{per}(\overline{P}_{CHP,i})}{\prod_{i=1}^{n}\varphi^{ref}(\overline{P}_{CHP,i})}\cdot \sum_{UC} f^{ref}(\overline{P}_{W,i},\overline{P}_{CHP,i},\overline{\mathrm{L}}_i)\cdot\prod_{i=1}^{n}\varphi^{ref}(\overline{P}_{CHP,i})] \tag{19}$$

Therefore, $\frac{\varphi^{per}(\bar{P}_w,\bar{P}_{CHP},\bar{L})}{\varphi^{ref}(\bar{P}_w,\bar{P}_{CHP},\bar{L})}$ is the correction factor ('statistical weight') of the various contributions tothe risk indices of the reference situation,allowing obtaining in the same simulation an estimate of the risk indices inthe perturbed situations. In fact, the correction factor calculation is specially defined for two scenarios:

a. the added DG unit is WF and the joint pdf of WFs, $\varphi_w(\bar{P}_w)$, has been changed, the correction factor is simplified by $\frac{\varphi^{per}(\bar{P}_w)}{\varphi^{ref}(\bar{P}_w)}$;
b. the added DG unit is CHP and the corresponding correction factor will be represented by $\frac{\varphi^{per}(\bar{P}_{CHP})}{\varphi^{ref}(\bar{P}_{CHP})}$.

3 ASSESSMENT METHODOLOGY

According to Eq. (13), the risk index is computed by the probability of the load-generation patterns leading to a specific congestion in the grid, curtailment power of each DG unit and the correction factor of reference variables. Therefore, the main structure of this practical methodology consists of three parts: curtailment evaluation, probabilistic assessment and correction factor estimation, shown in Figure 1. The details of the risk calculation process is introduced in the following paragraphs.

3.1 *Preprocess*

The preprocess stage, structuring the variants sampling, aims to partition the net balance space into safe and unsafe cells. SCADA system performs the network monitoring, which collects every 15 minutes multiple low-frequency signals including the existing generations and the loads. Based on this data collection, the net balance domain can be determined. Variants within the net balance domain can either lead to an acceptable solution of the load flow calculations, applied to both the N situation and all the N-1 ones, or not. The net balance variants can be grouped based on their similarity in making congestion on the grid. As shown in Figure 2, a security boundary divides the net balance space into the corresponding safe and unsafe regions. Adding a DG unit capacity to node 1 corresponds to extending the net balance domain by the yellow area, without affecting the security boundary.

In order to obtain an approximation of this security boundary between the safe and unsafe regions of state space, a mesh is defined on the NB domain, creating cells. If all the corners of a cell are

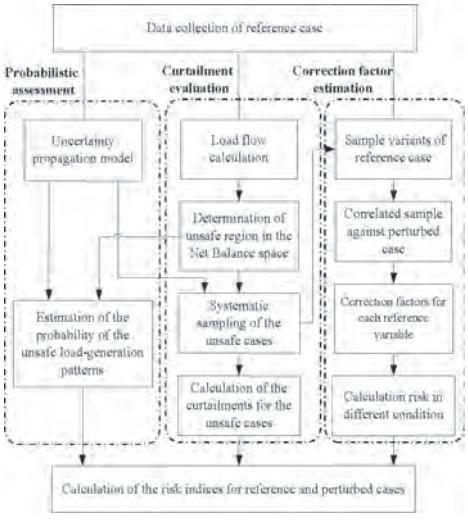


Figure 1. Framework flowchart of the proposed methodology.

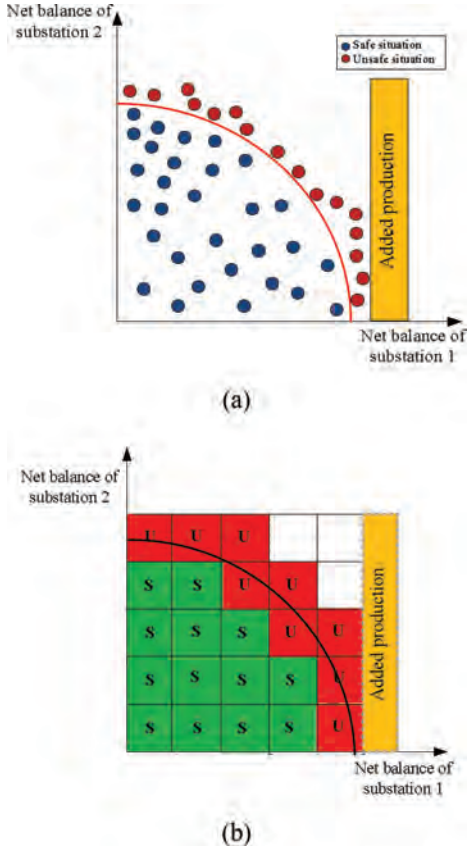


Figure 2. Preprocess:(a) safe and unsafe regions in the net balance space; (b) discretization of the net balance space in safe and unsafe cells.

safe (i.e. if they cause no congestion on the grid), then this cell is considered as a safe cell; otherwise it is considered as an unsafe cell, see Figure 2.

3.2 *The framework of curtailment*

Assessing the curtailment of the different DG units is not an easy task. For such a problem, Monte Carlo algorithm is an appropriate solution. Through Monte Carlo Sampling inside the unsafe cells, the input sample will be produced and the probabilistic modeling will be built.

At this stage, the curtailment calculation of each DG unit considers the curtailment cost of each specific DG type as well as its type of access. The Principles of Access (PoA) correspond generally to considering that the last connected producer is the first one to be curtailed. An OPF tool is applied to the sampled variants, both in N and N-1 situations, in order to calculate the (possible) curtailments of all DG units via the objective function. In this research, the objective function is the minimization of the total cost of curtailment.

In parallel with the curtailment analysis, the Capacity Factors (CF) and utilization factors (UF) are computed for each DG unit. The UF is defined as the ratio of the actual energy (after the curtailment) that can be produced in one year over the corresponding theoretical maximum at peak value, while CF is the ratio with no curtailment. Both of them are obvious indicators of the economic performance under reference and perturbed cases that a producer expect.

3.3 *Probabilistic assessment*

The risk analysis of congestion/curtailment is only associated to these unsafe cells, and the variants corresponding to the safe cells have no contribution to the risk indices. Thus, the systematic MCS based on a probabilistic modeling using historical data is performed by focusing on the unsafe cells. The stochastic input variables, i.e. distributed generation and loads, are modeled by their joint pdf's.

For WFs, the pdf that has been most used to represent the wind speed distribution is the Weibull law, which is mathematically described as:

$$f(v)=\frac{k}{c}\left(\frac{v}{c}\right)^{k-1}\times\exp\left[-\left(\frac{v}{c}\right)^{k}\right] \tag{20}$$

where v is the wind speed; k and c are the shape and scale parameters, respectively. An example is illustrated in Figure 3. Based on these marginal distributions, their joint pdf is constructed thanks to a Gaussian Copula. Sampled wind speeds are converted to electric power using the power curve.

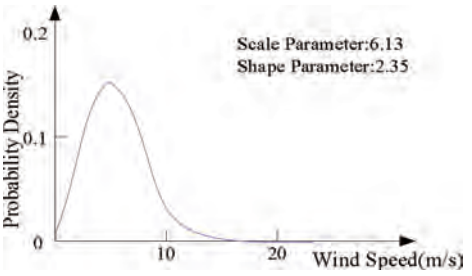


Figure 3. Wind speed marginal distribution.

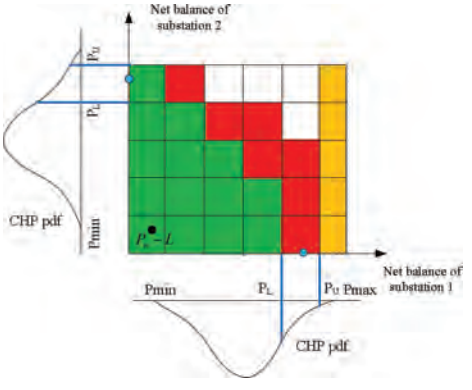


Figure 4. CHP sampling from the truncated pdf.

The load modelis obtained bynormalizing the historical load data of each year to the maximum consumption of the corresponding substation in that year. These normalized data implicitly containsa more informative consumption pattern in the different substations. A copula is used as well to define the joint pdf.

For the CHP units, their productions are independent and historical data are fit to obtain generation profiles. In the sampling procedure, after sampling wind and load from the corresponding joint pdf's, significant values of CHP productions correspond to values likely to make the net balance fall in an unsafe cell. In Figure 4, the net balance limited to wind and load, i.e. $\left(\Sigma P_{w,i}-L_i\right)$, is presented. For each unsafe cell UC, a compatible value of the CHP production at each node i is sampled from the corresponding truncated pdf in this UC, if possible, which is then given by:

$$\tilde{\varphi}_i^{UC}(P_{CHP,i})=\varphi_i(P_{CHP,i})/P_{CU,i} \tag{21}$$

where $\tilde{\varphi}_i^{UC}\left(P_{CHP,i}\right)$ is the truncated pdf of the CHP generation inside an unsafe cell UCat the ith substation,while $\varphi_i\left(P_{CHP,i}\right)$ represents the joint pdf of the CHP generation at node$i$. $P_{CU,i}$ is probability

of having the CHP production at node *i*in the right range.

The expression of the risk index for the perturbed case is then expressed as:

$$\tilde{RI} = \frac{1}{N}\sum_{j=1}^{N}\left[\frac{\varphi_w^q(\bar{P}_w)}{\varphi_w^{q+1}(\bar{P}_w)}\cdot\frac{\varphi_L^q(\bar{L})}{\varphi_L^{q+1}(\bar{L})}\cdot\frac{\prod_{i=1}^{n}\varphi^q(\bar{P}_{CHP,i})}{\prod_{i=1}^{n}\varphi^{q+1}(\bar{P}_{CHP,i})}\right. \quad (22)$$
$$\left.\sum_{UC} f(\bar{P}_{W,i},\bar{P}_{CHP,i}^{UC1},\bar{L}_i)\cdot\prod_{i=1}^{n}P_{CU,i}^{UC}(P_{wj,i},L_{ji})\right]$$

3.4 *Correction factor calculation*

In correlated sampling, the correction factor is the ratio between the probability of the perturbed variant and that of the reference case in the corresponding unsafe cells. The calculation of this factor uses the variants from the reference probabilistic assessment, see Figure 5. In the meantime, a correlated sampling, expressed in these 3 steps, is adopted so as to decrease the curtailment computation time for the perturbed cases.

a. Identify the perturbed pdf's and the type of the added DG unit (WF or CHP);
b. Sample variant from the joint pdf of the reference case;
c. Calculate the corresponding correction factors for each reference variable.

As shown in Figure 5, calculate the correction factor from the reference situation to the case of a new WF with different installed capacity connected to a grid. It will be necessary to evaluate the Weibull parameters of this perturbed case's new WF and filter the reference data which are more than this perturbed case's newly WF installed capacity. While for the case of a new CHP unit accepted to the grid, since CHP sample is to a large extent relied on the joint pdf's of wind-load pattern data and itself historical data, it is unavoidable to firstly execute reference wind-load correlated sampling so as to get perturbed CHP correlated sampling.

However, to apply efficiently the correlated sampling, the correlated sample size must be large enough.

3.5 *Risk estimation*

Finally, the risk indices for reference and perturbed cases can be estimated by averaging the product of the curtailment magnitude and the corresponding correction factor which is linked to the probability of the corresponding unsafe cells.

The weighting factor in each unsafe cell actually means the corresponding probability and is the ratio of the number of CHP samples taken from its truncated pdf inside the cell divided by the total number of random samples coming from the CHP pdf. Inevitably, the weighting factor calculation in each situation (e.g. season condition, N/N-1 contingency) needs to be taken into account.

A case study on a test power grid using this methodology will be treated in the following section.

4 TEST RESULTS AND DISCUSSION

In this section, the effectiveness of this practical methodology is illustrated on a test power grid. As

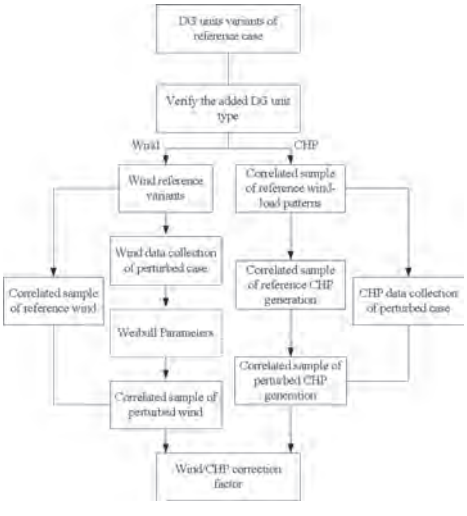


Figure 5. Flowchart of correction factor calculation.

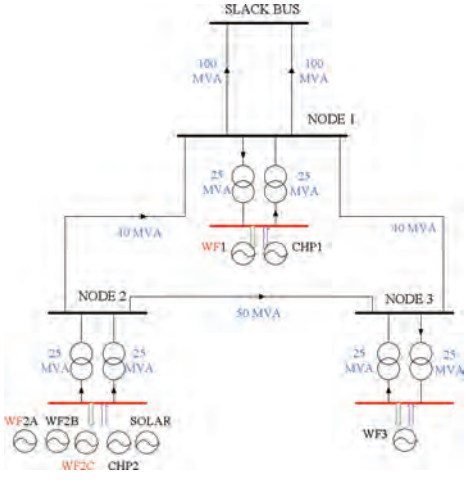


Figure 6. Test grid.

shown in Figure 6, the reference case corresponds to connecting a new WF to NODE 2. The grid ad is composed of 4 wind farms (57 MW), 2 CHP units (15 MW), 1 PV unit (1 MW) and a new connected wind farm (20 MW). The transmission network is operated at two voltage levels and comprises 4 high-voltage substations and 11 transmission lines (including transformers) connected to the MV side. The collection data has been conducted for this power grid during January 1,2015 to December 31,2015 every 15 min.

As for the perturbed case no.1, there are 7 DG units (without accounting for the added WF) connected to the substations in the power grid, while perturbed case no.2 corresponds to a new 10MW installed capacity WF accepted at NODE2. Assume that all DG units are connected to the MV side, the risk indices are estimated for the N and N-1 configurations (i.e. out of the HV/MV transformer at node 2), and for the different seasons leading to different line/transformer ratings. Due to the correlated sampling method used, the risk index for both the reference and perturbed cases are simultaneously obtained in one computation.

Figure 7 shows the Wind2A conditional risk under different cases. The probability of the seasons

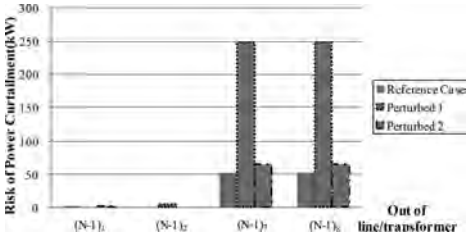


Figure 7. Wind2 A conditional risk in Summer.

and N/N-1 configurations are not considered. The solid rectangle represents the conditional risk in the reference situation, while the point rectangle and dotted line rectangle corresponds to the per1 and per2 cases, respectively. The detailed values are provided in Table 1. It obviously shows that the conditional risk of Wind2 A power curtailment in the reference case is smaller due to the addition of a new WF to node2. However, if the system lacks the transformers connected to NODE2, it will need more curtailment than for the other N-1 cases. The power curtailment of the newly connected WF in NODE2 will increase with a more adding installed capacity. And owing to the seasons' influence of wind energy, curtailment will decrease in winter.

In order to integrate the different energy curtailments under different N-0/N-1 rules, multiply the electrical elements rating and get the total risk of energy curtailment, see Figure 8. It clearly describesthe influence of different installed capacity of adding a new DG unit. The larger capacity case will produce more curtailment for this new DG unit but for the others DG units in the NODE2 will decrease. Therefore, it seems that the reference case will need more total energy curtailment.

As shown in Table 2, the utilization factors validates, there is a decrease of first and second curtailed producer under both the reference and perturbed 2 cases. Especially for the reference case, the impact is very visible, both capacity and utilization factors of Wind2A fall down from the perturbed 1 case where there is no new DG unit connected. However, the perturbed case 2 is able to have a higher value of capacity and utilization factors than reference case since the new added WF installed capacity is more suitable to NODE2.

Table 1. Conditional risk of different cases (in kW).

Seasons and $(N-1)_i$		Reference		Perturbed 1	Perturbed 2	
		Wind2A	Wind2C	Wind2A	Wind2A	Wind2C
Summer	$(N-1)_1$	0.25	176.42	0	0.31	14.18
	$(N-1)_2$	0	105.51	4.47	0	0
	$(N-1)_7$	52.35	1291.81	248.71	65.87	515.13
	$(N-1)_8$	52.18	1292.33	248.26	65.68	515.49
Inter	$(N-1)_1$	0.25	151.85	0	0.31	10.64
	$(N-1)_2$	0	34.66	4.47	0	0
	$(N-1)_7$	52.11	1283.13	247.13	65.53	515.49
	$(N-1)_8$	51.53	1282.09	245.67	64.68	511.10
Winter	$(N-1)_1$	0.25	38.28	0	0	0
	$(N-1)_2$	0	4.93	4.47	0	0
	$(N-1)_7$	52.12	1009.43	247.65	65.69	168.91
	$(N-1)_8$	51.83	1003.23	246.40	65.23	166.95

Table 2. Capacity and utilization factors.

	Reference		Perturbed 1	Perturbed 2	
	Wind2A	Wind2C	Wind2A	Wind2A	Wind2C
CF (%)	27.19	27.97	27.34	27.34	28.59
UF (%)	27.19	27.96	27.33	27.33	28.58

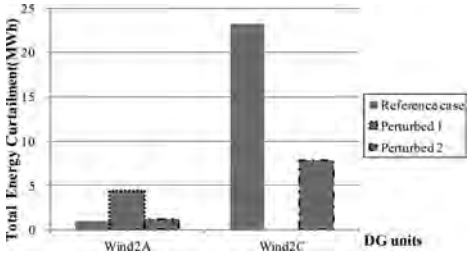


Figure 8. Total risk energy curtailment.

5 CONCLUSION

The large-scale use of distributed electricity generation is possibly associated with increasing grid congestion; hence, Distributed Generation (DG) curtailment is envisioned to solve those congestions in affected power systems. By resorting to curtailment as part of an Active Network Management (ANM) scheme under real-time supervision/control of the DG units, more distributed generation can be accommodated in a distribution grid For the purpose of retaining the power system secure and adapting the grid code, any possible new connection is able to accept a certain risk level in long-term planning.

In the work presented in this paper, an efficient methodology for probabilistic assessment of the risk of connecting a new DG unit to an existing grid, using an ANM scheme for a defined season and a configuration of the grid (N and N-1 conditions), is proposed. Furthermore, the use of correlated sampling allows significantly reducing the computation time so that one computation is able to simultaneously provide the risk estimation for both reference case and perturbed cases in order to choose a more suitable installed capacity DG unit. The concept of net balance in the preprocess stage is on one side to account for the dimensions of the problem (i.e. number of DG units and substations), and on the other side, it allows for a rather quick identification of the unsafe region by performing a set of load flow calculations on the corners of the cells in the constructed mesh. A probabilistic wind model has been implemented, resorting to a Gaussian Copula to build the joint pdf of all WFs at the different nodes of the grid; in the meantime, the compatible CHP variants inside each unsafe cell were produced. Thus, the variants belonging to an unsafe cell acted as input variables of the optimal power flow (severity function), based on the curtailment cost, for congestion/curtailment calculations.

Last but not least, the advantages of this practical methodology lie on fast estimating the impact of the possible acceptance of new DG units with different installed capacities into an existing grid, what provides foundation for the future grid planning and its reinforcements.

REFERENCES

Do, M.T., Francois, B. 2017. Probabilistic approach to evaluate the cost and constraints of the renewable production curtailment in MW network. *In Proceedings of IFAC2017, Toulouse(France).*

European Climate Foundation, Roadmap 2050: Practical guide to a prosperous, low carbon Europe 2010. http://www.roadmap2050.eu/attachments/files/Roadmap2050-AllData-MinimalSize.pdf (last consulted on 19/12/2017).

Faghihi, F., Labeau, P.E., Maun, J.C., De Wilde, V. &Vergnol, A. 2014. Towards a probabilistic risk assessment of distributed generation curtailment in saturated TSO/DSO networks. *In Proceedings of CIRED2014, Rome (Italy).*

Gooding, P.A., Makram, E. and Hadidi, R. 2014. Probability analysis of distributed generation for island scenarios utilizing Carolinas data. *Electric Power Systems Rearch* 107: 125–132.

Hagspiel, S., Papaemannouil, A., Schmid, M. and Anderson, G. 2012. Copula-based modeling of stochastic wind power in Europr and implications for Swiss power grid. *Applied Energy* 96: 33–44.

Järventaustac, P., Repo, S., Rautiainen, A. and Partanen, J. 2009. Smart grid power system control in distributed generation environment.*6th IFAC symposium on power plants and power systems control, Finland.*

Labeau, P.E., Faghihi, F., Maun, J.C., De Wilde, V. and Vergnol, A. 2014. Mathematics of PRA applied to Distributed Generation curtailment in saturated grids. *In Proceedings of Esrel2014, Wroclaw (Poland): Balkema.*

Rocchetta, R.., Li, Y.F. and Zio, E. 2015. Risk assessment and risk-cost optimization of distributed power generation systems considering extreme weather conditions. *Reliability Engineering and System Safety* 136: 47–61.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Integrated deterministic and probabilistic safety assessment of the cooling circuit of a superconducting magnet for nuclear fusion applications

R. Bellaera, R. Bonifetto, N. Pedroni, L. Savoldi & R. Zanino
NEMO group, Dipartimento Energia, Politecnico di Torino, Torino, Italy

F. Di Maio & E. Zio
Dipartimento di Energia, Politecnico di Milano, Italy

E. Zio
Chair on Systems Science and the Energetic Challenge, European Foundation for New Energy—Fondation EdF, CentraleSupelec, France

ABSTRACT: In recent years, there has been a growing interest in nuclear fusion as energy source due to its several principle advantages over fission, which include reduced radioactivity in operation and in waste, large fuel supplies, and increased safety. The most promising configuration of a nuclear fusion system is currently the tokamak, the largest of which, called the largest of which (ITER), is under construction in Cadarache, France. The safety of nuclear fusion systems has to be proved and verified by a systematic analysis of the system behavior under normal transient and accidental conditions. One challenge to the analysis is that the operation of tokamaks presents complex dynamic features as it is based on the transformer principle: in particular, they employ superconducting magnets, a subset of which operates with variable current to generate one of the components of the magnetic field needed to confine the plasma in the chamber where nuclear fusion reactions occur. In the present paper, we apply techniques of Integrated Deterministic and Probabilistic Safety Assessment (IDPSA), which combine phenomenological models of system dynamics with stochastic process models, taking for the first time as reference system the cooling circuit of a superconducting magnet for fusion applications, subject to a Loss-Of-Flow-Accident (LOFA).

1 INTRODUCTION

The need of satisfying a growing demand of energy, while protecting the environment and reducing the dependence on fossil fuels, has recently increased the interest in the use of nuclear fusion as energy source. Nuclear fusion presents several advantages over fission, which include reduced radioactivity in operation and in waste, large fuel supplies, and increased safety (EC, 2004).

The most promising configuration for electrical energy production from fusion is the tokamak, the largest of which (ITER) is now under construction at Cadarache (France), under an international collaboration between seven member entities (i.e., China, the European Union, India, Japan, Korea, Russia and the United States) (ITER, 2014).

Nuclear fusion reactors will use superconducting (SC) magnets to generate a powerful magnetic field needed to confine the plasma in the shape of a torus, where D-T fusion reactions occur at a temperature of the order of 10^8 K. On the other hand, all the SC coils need to be cooled at cryogenic temperatures in order to avoid the quench of the magnets (i.e., the loss of their superconducting state) during operation: for example, the ITER SC coils are cooled by supercritical helium (SHe) at a pressure of 0.5–0.6 MPa and temperature of about 4.5K (ITER, 2014). Dedicated cryogenic cooling loops remove the heat load from the magnets, releasing it to saturated liquid helium (LHe) pools, acting as thermal buffers in the transfer of the load to the refrigerator (Hoa et al., 2012; Zanino et al., 2013).

The safety of nuclear fusion systems has to be proved and verified by a systematic analysis of the system behavior under normal transient and accidental conditions (Taylor and Cortes, 2014; Rivas et al., 2015; Perrault, 2016; Wu et al., 2016), for two main reasons. First, the presence of radioactive sources (e.g., tritium and materials activated by the neutrons produced by the fusion reactions)

imposes a careful study to avoid the contamination of the workers, the public and the environment (Taylor, 2015; Taylor et al., 2017). Second, in view of the cost of the SC magnet system (Mitchell et al., 2008 and 2012), its protection and integrity should be guaranteed (ITER, 2014; Savoldi Richard et al., 2014; Savoldi et al., 2017 and 2018).

One challenge to the related safety analyses is that the operation of tokamaks presents *complex dynamic features*, as it is based on the transformer principle. In particular, they employ SC magnets, a subset of which operates with *variable* current to generate one of the components of the magnetic field needed to confine the plasma in the chamber where nuclear fusion reactions occur (Zohm, 2014). The *order* and *timing* of failure events occurring along an accident scenario, the *magnitude* of failures and the *values* of the process variables at the time of event occurrence are critical in determining the evolution of the system response (Aldemir, 2013; Kirschenbaum et al., 2009; Zio and Di Maio, 2009, 2010; Zio et al., 2010; Zio, 2014; Turati et al., 2017 and 2018).

To perform the safety assessments we resort to the Integrated Deterministic and Probabilistic Safety Assessment (IDPSA) framework (Di Maio et al., 2016) to combine (deterministic) phenomenological models of system dynamics with (probabilistic) stochastic process models. The IDPSA methodology is here used for the first time to analyze the response—to abnormal transient conditions—of the cooling system of a SC magnet, namely a *single* ITER Central Solenoid Module (CSM) in a reference (cold) *test facility* (Spitzer et al., 2015). In particular, a Loss-Of-Flow Accident (LOFA) is considered as reference abnormal scenario. The 4C code (Savoldi Richard et al., 2010) is employed for the (deterministic) simulation of the system behavior. Multiple Value Logic (MVL) is adopted for the description of the components failures (Garibba et al., 1985). The MVL allows describing that the components can fail at any time along the scenario, with *different* (discrete) *magnitudes* (Di Maio et al., 2015; Zio, 2013).

The paper is organized as follows. In Section 2, the single ITER CSM in the cold test facility is presented together with the corresponding simulator. In Section 3, the different regimes of system operation are described. The components failures causing the deviations from the nominal CSM conditions in the test facility are described in detail in Section 4, together with the use of MVL for generating accident scenarios. In Section 5, the results obtained are illustrated and analyzed. Finally, conclusions are drawn and summarized in Section 6.

2 DESCRIPTION OF THE SYSTEM AND PRESENTATION OF THE SIMULATOR

The ITER CS allows inducing the current in the plasma and maintaining it during long plasma pulses. It is composed by 6 CS modules (CSM) stacked in the vertical direction; each of them is being manufactured and will be independently tested. Each module is composed by 7 pancake-wound conductors, namely 6 hexa-pancakes and 1 quad-pancake. Each pancake is cooled in parallel, resulting in 40 parallel cooling channels per module, each featuring 14 turns. The He inlets are located at the coil bore, while the outlets are at the outer side of the magnet. All the pancakes of each module are electrically connected in series through suitable joints (Libeyre et al. 2015).

The main components of a reference facility for the CSMs *cold tests* (which is the subject of the present analysis) are the He refrigerator, producing the supercritical He (SHe) at 4.5 K, which is used to cool the magnets during the tests, and the test chamber where the module will be put into a cryostat. Besides these two main components, several manifolds, pipelines, heat exchangers (HXs) control valves (CV) of the cryoplant and a liquid He (LHe) thermal buffer will connect the refrigerator to the coil, and a dedicated power supply system will provide a current up to ~50 kA.

From the hydraulic point of view, the analysis reported here is focused on the loop cooling a single CSM. This closed loop, much simpler than that foreseen in ITER for the cooling of the CS, provides SHe at ~4.5 K to the inlets of the 40 hydraulic paths, collects the (warmer) SHe at their outlets and by means of a cold circulator drives it to two HXs, where the heat removed from the coil is transferred to a LHe buffer. The LHe evaporated in the buffer is, then, extracted and cooled down by the refrigerator.

The 4C thermal hydraulic code (Savoldi Richard et al. 2010) is used here to model both the coil and its SHe cooling loop. Figure 1 shows a scheme of the loop model. The SHe at the outlet of the cold circulator is cooled in HX1 to remove the heat generated by the compression process. Then, it is driven to the coil inlets through CV1 (fully open in normal operation) and a supply cryoline (cryoline 1). At the coil outlet, the SHe reaches HX2 through a return cryoline (cryoline 2) and CV2 (fully open in normal operation). The main input parameters of each component model are reported in Table 1. For the details of the model of each component, please refer to (Bonifetto et al. 2012) and (Zanino et al. 2013). A realistic characteristic of the cold circulator has been implemented in the circuit model, as described in detail in (Zanino et al. 2013) for another system.

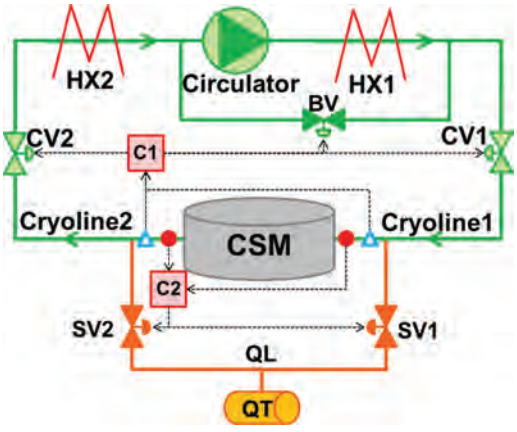


Figure 1. Scheme of the SHe cooling loop of the CSM. Solid red circles are pressure taps, open cyan triangles are flow meters (see the text for other abbreviations).

Table 1. Main input parameters of the circuit component models (L = length, D = diameter, K_v = flow coeffient).

Component	L [m]	D [mm]	# of parallel pipes
HX1,2	31	20	11
Cryoline1	28	46	
Cryoline2	24	46	
		K_v [m^3/h]	
CV1,2, BV, SVin, SVout		71	

With reference to a Loss-Of-Flow Accident (LOFA) (scenario of interest in the present paper, see the Introduction), the strategy adopted by the control system is assumed here to be similar to that adopted in ITER and described in (Savoldi et al., 2018) for the Toroidal Field (TF) coils. In particular, in order to protect the CS, a *controlled discharge* of the CS circuits is carried out in ITER, consisting of a *current ramp down* of about 30s (ITER_D_K7G8GN v2.1, 2014), driven by the plasma control system. Obviously, the current variation causes AC losses, which induce a (possibly significant) *heat deposition* in the conductor. In the reference system at hand, a similar fast controller discharge is assumed to be taken by the control system in case of LOFA. As far as the cryoplant is concerned, the basic circuit control in case of a LOFA includes the isolation of the circulator from the coil by means of the full closure of both CVs and the opening of the by-pass valve (BV) to equalize the pressure at the circulator suction and discharge (preventing any damage to the pump itself as fail-safe condition). In the present work, this action is taken (by controller C1 in Figure 1), when a SHe mass flow rate below 10% of the nominal value is measured both at the inlet and at the outlet of the CSM, after a validation time of 1s (Savoldi et al., 2018). Notice that the "nominal value" of the mass flow rate here considered corresponds to the nominal CSM conditions *during a test* in the reference facility, *not* to the "future" normal operating conditions in ITER. In case of excessive pressurization at the coil boundaries, two Safety Valves (SV) at the CSM inlet and outlet open, driven by the PID controller C2 (gain = 1 $e^{-7}Pa^{-1}$, integration time = 0.2 s, derivation time = 1 s), with set-point 1.8 MPa (Savoldi Richard et al., 2012); the controller parameters and set point have been assumed here equal to those in (Savoldi et al., 2018). When the SV opens, the He is released in a Quench Tank (QT) by means of suitable Quench Lines (QL).

The detailed CSM model solves the 1D transient mass, momentum and energy conservation equations, computing the temperature, pressure and velocity distribution, in each of the two regions (cable bundle and central, low impedance channel) of each pancake, as described in detail in (Zanino et al. 1995). Then, the inter-turn and inter-pancake thermal coupling between adjacent turns and pancakes, respectively, is computed considering the insulation as a thermal resistance to evaluate the heat transfer between neighboring conductors (Savoldi et al. 2000).

3 SYSTEM OPERATION REGIMES

During the tests, the facility and the CSM will be operated in different regimes, which can be briefly described as:

a. Cold mode standby operation (e.g. during night or weekend), when the CSM is not charged and kept at nominal ~4.5 K; no dangerous transients are expected from a LOFA in these conditions, so this regime is not analyzed here.
b. Cold mode experimental operation (i.e. during tests), when the CSM is charged at full or partial current (which is the case analyzed in the present paper). In this regime, an accidental temperature increase up to (or above) the so-called *current sharing temperature* T_{CS} may lead to a *quench*, i.e., to a loss of the superconducting state (notice that T_{CS} is defined as the temperature, above which the current starts to flow also in the Cu matrix of the SC strands and in the pure Cu strands, developing a non-zero voltage and causing Joule heat generation in the cable). The consequent fast local Joule heat deposition can induce thermal stresses that may seriously

damage the conductor, causing a degradation of its performance or, in the worst case, the loss of integrity of the conductor. The presence of a *normal* (non-SC) zone and, to some extent, variations in the conductor temperature above T_{CS} can be detected by measuring the voltage at the extremities of the each pancake.

The coil inlet temperature is here taken as the nominal one. The objective of the analysis is to study the response of the system described above (in the operating mode b.) to abnormal conditions, e.g., to accident scenarios driven by stochastic components failures (see the following Section 4). In particular, our interest is mainly devoted to LOFAs (see the previous Section 2). Actually, in absence of helium coolant flow, the heat deposition induced by AC losses (caused by the controlled discharge of the CS circuits) may a priori lead to a dangerous increase in the conductor temperature (with possible quench of the magnet), even in a cold test configuration like the one analyzed here. Two variables are monitored during each transient as "critical indicators" of the state of the system: the cooling helium pressure at the inlet and outlet of the CS magnet and the voltage measured at the coil extremities. Actually, if the pressure in the conductor exceeds 25 MPa, the conductors can be damaged (ITER_D_2NBKXY v1.2, 2009); also, if the voltage on a single pancake goes above 0.1 V for more than 1s, it means that the conductor temperature has exceeded the current sharing temperature T_{CS}, i.e., that the superconducting state of the magnet is lost.

4 FAILURE SCENARIOS GENERATION BY MULTIPLE VALUED LOGIC

The following component failures can occur at random times in the time horizon [0,600] seconds:

1. the Centrifugal Pump (CP) reduces exponentially the rotational speed, directly affecting the mass flow rate that can be reduced to i) 75%; ii) 50%; iii) 25%; iv) 0% of the nominal mass flow rate, i.e. in the last case down to a total loss of pumping capacity.
2. the two Control Valves (CVs) can fail in three different modes: i) stuck (open) at the nominal position; ii) stuck closed at 50% of the nominal position; iii) stuck totally closed.
3. the By-pass Valve (BV) can fail in three different modes: i) stuck (closed) in nominal position; ii) stuck open at 50% of the flow area; iii) stuck totally open.
4. the two Safety Valves (SVs) can fail in three different modes: i) stuck (closed) in nominal position; ii) stuck open at 50% of the flow area; iii) stuck totally open.

It can be shown that the system reaches a steady state condition in ~100 seconds, irrespectively of the failure occurred. Therefore, we set a mission time $T_M = 700$ s.

A Multiple Value Logic (MVL) scheme has been adopted to generate the accidental scenarios, considering the stochastic (discrete) time interval (t) of occurrence of component failures, their (discrete) magnitude (m) and the order (ord) of events along the sequence. The random realizations of the discretized time and magnitudes values are included into a sequence vector that represents a generated scenario to be simulated, [m_{cp}, t_{cp}, ord_{cp}, m_{cv1}, t_{cv1}, ord_{cv1}, m_{cv2}, t_{cv2}, ord_{cv2}, m_{bv}, t_{bv}, ord_{bv}, m_{sv1}, t_{sv1}, ord_{sv1}, ms_{v2}, ts_{v2}, ord_{sv2}] (Di Maio et al., 2017). The following values are considered:

- time (t) discretization: we use the label $t = 1, 2, 3, 4, 5$ and 6, for failures occurring in the intervals [0, 100] s, [101, 200] s, [201, 300] s, [301, 400] s, [401, 500] s, [501, 600] s, respectively; $t = 0$ means that the component does *not* fail within the T_M of the scenario and the value "NaN" is used to identify the respective (non-)failure order (ord) in the sequence vector of the accidental scenario. Notice that, even if the failure order (ord) may seem redundant in the MVL representation, it is actually used to discriminate between scenarios where different components fail in the *same* time *interval t*. Also, it is worth highlighting that, once a time *interval t* is identified by MVL, the *actual* time of component failure (used in the deterministic simulation of the accident scenario) is randomly sampled within the interval.
- Magnitude (m) discretization:
  - the CP magnitude is indicated with the label $m_{cp} = 1, 2, 3$ or 4 for failure states corresponding to an exponential decrease of the rotational speed down to 75%, 50%, 25% and 0% of the nominal value, respectively; if $m_{cp} = 0$, the component does not fail;
  - for each CV, the magnitude is indicated by the label $m_{cv} = 1, 2$ or 3 if the component stays stuck (open) at the nominal position, stuck closed at 50% of the nominal position and stuck closed, respectively; if $m_{cv} = 0$, the component works correctly;
  - the BV magnitude is indicated by the label $m_{bv} = 1, 2$ or 3 if the component stays stuck in (closed) nominal position, stuck open at 50% of nominal flow area and stuck totally open, respectively; if $m_{bv} = 0$, the component does not fail;
  - for each SV, the magnitude is indicated by the label $m_{sv} = 1, 2$ or 3 if the component stays stuck (closed) in nominal position, stuck open at 50% of nominal flow area and stuck totally open, respectively; if $m_{sv} = 0$, the component works in that scenario.

As an example, the accidental sequence vector [4, 6, 5, 0, 0, NaN, 1, 2, 1, 2, 4, 3, 2, 5, 4, 1, 3, 2] represents a scenario where: the CP fails completely to 0% of the nominal value at a time in [501,600] s (fifth event occurring along the sequence); the CV1 correctly works throughout T_M; the CV2 fails stuck (open) at the nominal position at a time in [101,200] s (first event occurring along the sequence); the BV fails stuck open at 50% of the nominal flow area (third event along the sequence) at a time in [301,400] s; the SV1 fails stuck open at 50% of the flow area at a time in [401,500] s (fourth event along the sequence); finally, the SV2 fails stuck (closed) in nominal position at a time in [201,300] s (second event along the sequence).

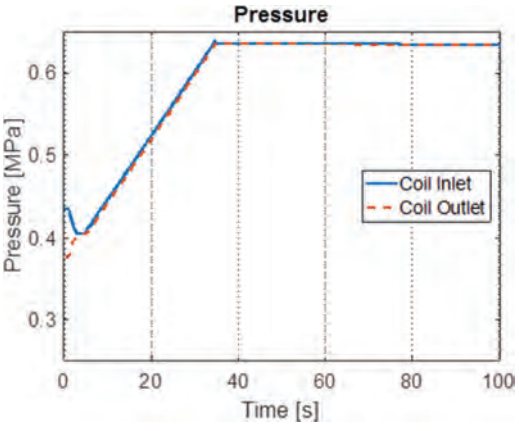


Figure 2. Pressure at the inlet and outlet of the coil (following the CP failure at 0% of the nominal mass flow rate).

5 RESULTS

In principle, the MVL accidental scenario generation procedure described above would entail the creation of 4×10^8 scenarios. We report here only the "bounding analysis" of the system response, analyzing a *reduced set* of MVL sequences, which are *expected* to be the *most challenging* for the system safety. Notice that *none* of the selected accident sequences turns out to be *critical* for the CS module integrity. In particular, the voltage at the extremities of the SC coil will remain far below the threshold of 0.1V (i.e., the temperature of the conductor does not exceed the T_{CS}): this means that the magnet maintains its superconductive properties.

Let us consider the MVL sequence [4, 1, 1, 0, 0, NaN, 0, 0, NaN, 0, 0, NaN, 0, 0, NaN, 0, 0, NaN], which entails the pump to fail according to an exponential decrease of the rotational speed till complete stop; the other components work correctly during the scenario. This sequence has been chosen because in this system no CP redundancies are designed and, therefore, a CP unavailability might lead the CSM to critical conditions. Notice that when the loss of flow is detected (Section 2), in order to protect the CS, a controlled discharge of its electrical circuits is carried out, consisting of a current ramp down to 0 in about 30s. A significant heat is, thus, generated in the conductors, due to the AC losses produced by the fast current variation. Figure 2 shows that when the LOFA occurs, the pressure of the coil both at the inlet (solid line on Figure 2) and outlet (dashed line) increases up to 50% with respect to the nominal value, without however exceeding the pressure safety threshold of 1.8 MPa.

Two additional MVL that might deserve attention are [0, 0, NaN, 0, 0, NaN, 0, 0, NaN, 0, 0, NaN, 3, 1, 1, 0, 0, NaN] and [0, 0, NaN, 0, 0, NaN, 0, 0, NaN, 0, 0, NaN, 0, 0, NaN, 3, 1, 1]. These sequences entail the failure of safety valves SV1 and SV2, located respectively at the inlet and outlet of the coil, that have the function of keeping the pressure of the system below the critical value of 1.8 MPa in accidental situations. In normal operation, the pressure at the inlet of SV1 is equal to 0.433 MPa and at the outlet is 0.42 MPa, corresponding to the assumed pressure of the quench tank. If component SV1 fails, helium starts flowing in the quench line leading to the reduction of the pressure at the inlet of the coil (solid line in Figure 3): both the inlet and the outlet pressure of the coil decrease, the mass flow rate remains at its nominal value and a new steady state situation is reached with large safety margin with respect to the pressure safety threshold 1.8 MPa.

Sequence [0, 0, NaN, 0, 0, NaN, 0, 0, NaN, 0, 0, NaN, 0, 0, NaN, 3, 1, 1], instead, entails SV2 failure. In Figure 4, it can be seen that the pressure at the inlet of the valve is equal to 0.376 MPa (solid line) whereas at the outlet (dashed line) it is equal to the assumed pressure of the quench tank (i.e., 0.42 MPa). When valve SV2 opens, the pressure at the outlet of the coil increases up to 0.42 MPa. Since the CP is working correctly, the nominal pressure drop between the inlet and the outlet of the component is maintained. As a consequence, the increase in the outlet coil pressure is followed by an increase in the inlet coil pressure. Also in this case, a new steady state condition is reached with a large safety margin with respect to the pressure safety threshold of 1.8 MPa.

The relevance of the dynamic features is witnessed by the outcome of sequence [0, 0, NaN, 0, 0, NaN, 0, 0, NaN, 0, 0, NaN, 3, 1, 1, 3, 1, 2], whose evolution is shown in Figure 5. The first failure is represented by the SV1 stuck open. This causes a slight decrease in the pressure at the inlet

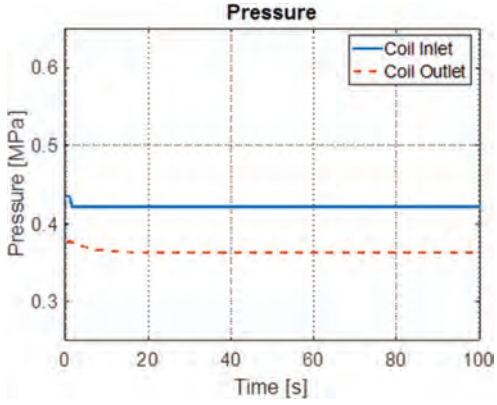


Figure 3. Pressure at the inlet and outlet of the coil following SV1 failure (stuck totally open).

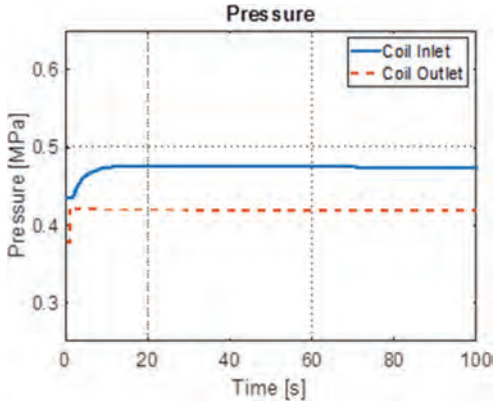


Figure 4. Pressure at the inlet and outlet of the coil following SV2 failure (stuck totally open).

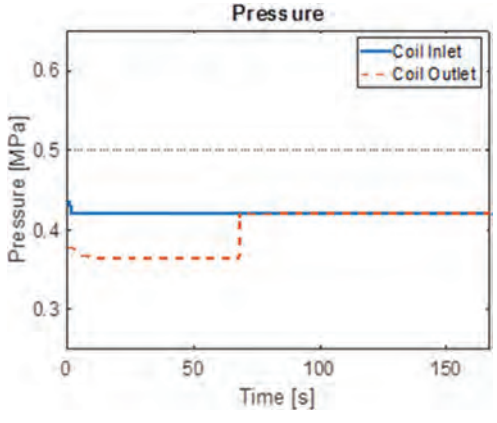


Figure 5. Pressure at the inlet and outlet of the coil along the sequence of failures: SV1 stuck totally open; SV2 stuck totally open.

(solid line) and at the outlet (dashed line) of the coil. Then, also safety valve SV2 fails stuck open, causing an increase in the pressure at the outlet of the coil up to the value of 0.42 MPa, the pressure assumed for the quench line (this is due to the fact that the pressure at the outlet of the coil is lower than that in the quench line). As a result, the pressure at the inlet and at the outlet of the coil equalizes: the LOFA happens (and is detected) and the current is reduced to zero producing AC losses, which lead to heat deposition. Throughout the rest of the transient, the value of the pressure at the boundaries of the coil stabilizes at 0.42 MPa. This response of the system justifies the necessity of a dynamic approach. In fact, if we consider the single failures of the SV1 and SV2 at the most conservative magnitude (see, e.g., Figures 3 and 4), a new steady state condition is reached *without* any *problem* from the point of view of the *availability* of the CSM in the test facility; on the other hand, the combination of failures considered just now leads to a different end state.

In all the scenarios reported here, the pressure at the boundaries of the SC coil is kept below the safety threshold of 1.8 MPa, i.e., with a positive safety margin. In other words, as for the analysis here presented, no critical situations are expected and the coil is not damaged. In addition, the voltage always stays well below the threshold of 0.1 V, which means that the current sharing temperature T_{CS} is not exceeded.

6 CONCLUSIONS

In this paper, we have considered the safety analysis of the simplified cooling system of a *single* module of the ITER Central Solenoid (CS) in a cold *test facility*, subject to a Loss-Of-Flow Accident (LOFA). For this, the deterministic 4C code has been employed to simulate the system behavior and a Multiple Value Logic (MVL) has been adopted for building a comprehensive set of combinations of times and (discrete) magnitudes of components failures to run stochastic accident scenarios. The cooling helium pressure at the inlet and outlet of the SC coil, and the voltage measured across the pancakes have been selected as "critical" safety parameters to monitor during each transient.

The application of the MVL has generated a list of more than 10^8 possible scenarios. However, for the sake of brevity, we have carried out only a "bounding analysis" of the system response by inspecting a *reduced set* of MVL sequences, in particular, those *expected* to be *most challenging* for the system safety. Results have shown that these scenarios are *not critical* for the CS module integrity: in particular, in all the cases considered, the

pressure at the boundaries of the SC coil is below the safety threshold of 1.8MPa, i.e., with a positive safety margin. Also, the voltage keeps well below the threshold of 0.1V, which means that the temperature of the conductor does not exceed the current sharing temperature T_{CS} and the magnet does *not* lose its SC properties.

However, a final remark is in order with respect to the positive results obtained. The CS magnet has been analyzed here in "cold mode experimental operation" (see Section 3), with the coil inlet temperature taken as the *nominal* one. In some cases, this temperature can be artificially increased, in order to perform specific tests (for example, the T_{CS} measurement (Savoldi et al. 2000)). This situation would reduce the temperature *margin* (namely, the difference between the operating cable temperature and the T_{CS}), shifting the operating point closer to the T_{CS} and, thus, reducing the amount of the heat needed to induce a quench. The analysis of such "more severe" operating conditions will be possibly considered in future research.

REFERENCES

Aldemir, T., 2013. A survey of dynamic methodologies for probabilistic safety assessment of nuclear power plants. *Ann. Nucl. Energy* 52: 113–124.

Bonifetto, R., Casella, F., Savoldi Richard, L., Zanino, R. 2012. *Dynamic modeling of a supercritical helium closed loop with the 4C code.* AIP Conference Proceedings (1434): 1743–1750.

Di Maio, F., Baronchelli, S., Zio, E., 2015. A computational framework for prime implicants identification in non-coherent dynamic systems. *Risk Analysis* 35(1): 142–156.

Di Maio, F., Vagnoli, M., Zio, E., 2016. Transient Identification by Clustering based on Integrated Deterministic and Probabilistic Safety Analysis Outcomes. *Annals of Nuclear Energy*, Volume 87, pp. 217–227, 2016.

Di Maio, F., Baronchelli, S., Vagnoli, M., Zio, E., 2017. Determination of Prime Implicants by Differential Evolution for the Dynamic Reliability Analysis of Non-Coherent Nuclear Components. *Annals of Nuclear Energy*, 102, pp. 91–105, 2017.

Garibba, S., Guagnini, E., Mussio, P., 1985. Multiple-Valued Logic Trees: Meaning and Prime Implicants. *IEEE Transactions On Reliability 34*: 463–472.

Hoa, C., Bon-Mardion, M., Bonnay, P., Charvin, P., Cheynel, J.N., Lagier, B., Michel, F., Monteiro, L., Poncet, J.M., Roussel, P., Rousset, B., Vallcorba-Carbonell, R., 2012. Investigations of pulsed heat loads on a forced flow supercritical helium loop—Part A: experimental set up. *Cryogenics*; 52(7–9): 340–8.

ITER, 2014. Central Interlock System Strategy for ITER Magnet Protection: Machine Protection Functions. Report *ITER_D_K7G8GN v2.1*, January 24, 2014.

ITER_D_2NBKXY v1.2, 2009. *ITER Design Description Document: Magnets—Conductors*, 09/09/2009.

ITER_D_K7G8GN v2.1, 2014. *Central Interlock System Strategy for ITER Magnet Protection: Machine Protection Functions*, January 24, 2014.

Kirschenbaum, J., Bucci, P., Stovsky, M., Mandelli, D., Aldemir, T., Yau, M., Guarro, S., Ekici, E., Arndt, S.A., 2009. A benchmark system for comparing reliability modeling approaches for digital instrumentation and control systems. *Nucl. Technol*. 165: 53–95.

Libeyre, P., Cormany, C., Dolgetta, N., Gaxiola, E., Jong, C., Lyraud, C., Reiersen, W., Everitt, D., Martovetsky, N., Rosenblad, P., Cole, M., Freudenberg, K., Sheng Liu, Smith, J., Jing Wei, Lin Wang, Xiaowu Yu, Xiaoyu Dong, Jijun Xin, Chao Li, Wangwang Zheng, Chao Fang 2015. Status of design and manufacturing of the ITER Central Solenoid and Correction Coils. *Proceedings of IEEE 26th Symposium on Fusion Engineering (SOFE)*: 1–8.

Mitchell, N., Bessette, D., Gallix, R., Jong, C., Knaster, J., Libeyre, P., Sborchia, C., Simon, F. 2008. The ITER magnet system. *IEEE Trans. Appl. Supercond*. 18: 435–440.

Mitchell, N., Devred, A., Libeyre, P., Lim, B., Savary, F. 2012. The ITER magnets: design and construction status. *IEEE Trans. Appl. Supercond*. 22: 4200809.

Perrault, D., 2016. Safety issues to be taken into account in designing future nuclear fusion facilities. *Fusion Engineering and Design* 109–111: 1733–1738.

Rivas, J.C., Dies, J., Fajarnés, X., 2015. Revisiting the analysis of passive plasma shutdown during an ex-vessel loss of coolant accident in ITER blanket. *Fusion Engineering and Design* 98–99: 2206–2209.

Savoldi, L., Bonifetto, R., Pedroni, N., Zanino, R. 2018. Analysis of a protected Loss Of Flow Accident (LOFA) in the ITER TF coil cooling circuit. *IEEE Transactions on Applied Superconductivity* 28(3): 4202009.

Savoldi Richard, L., Bonifetto, R., Bottero, U., Foussat, A., Mitchell, N., Seo, K., and Zanino, R., 2014. Analysis of the effects of the nuclear heat load on the ITER TF magnets temperature margin. *IEEE Trans. Appl. Supercond*. 24(3): Art. ID 4200104.

Savoldi Richard, L., Bessette, D., Bonifetto, R. and Zanino, R. 2012. Parametric analysis of the ITER TF fast discharge using the 4C code. *IEEE Trans. Appl. Supercond.* 22(3): Art. no. 4704104.

Savoldi Richard, L., Casella, F., Fiori, B., Zanino, R. 2010. The 4C Code for the Cryogenic Circuit Conductor and Coil modeling in ITER. *Cryogenics* (50): 167–176.

Savoldi, L., Bonifetto, R., Zanino, R., 2017. Analysis of a loss-of-flow accident (LOFA) in a tokamak superconducting Toroidal Field Coil. In: *Safety and Reliability—Theory and Applications, Proceedings of the ESREL 2017 Conference*; 18–22 June 2017; Portoroz, Slovenia; pp. 67–74; ISBN 9781138629370.

Savoldi, L., Zanino, R. 2000. Thermal-hydraulic analysis of TCS measurement in conductor 1A of the ITER central solenoid model coil using the M&M code. *Cryogenics* (40): 593–604.

Spitzer, J., Stephens, A., Schaubel, K, Smith, J., Norausky, N., Khumthong, K., Gattuso, A. 2015. ITER Central Solenoid Module fabrication program. *Proceedings of IEEE 26th Symposium on Fusion Engineering (SOFE)*: 1–6.

Taylor, N.P. 2015. Safety and licensing of nuclear facilities for fusion. *Proceedings of the 2015 IEEE 26th Symposium on Fusion Engineering (SOFE)*, 31 May-4 June 2015; Austin, TX, USA; DOI: 10.1109/SOFE.2015.7482293; ISBN: 978-1-4799-8264-6.

Taylor, N., Ciattaglia, S., Boyer, H., Coombs, D., Zhou Jin, X., Liger, K., Mora, J.C., Mazzini, G., Pinna, T., Urbonavicius, E., 2017. Resolving safety issues for a demonstration fusion power plant. *Fusion Engineering and Design* 124: 1177–1180.

Taylor, N., Cortes, P. 2014. Lessons learnt from ITER safety & licensing for DEMO and future nuclear fusion facilities. *Fusion Engineering and Design* 89: 1995–2000

Turati, P., Pedroni, N., Zio, E. 2017. Simulation-based exploration of high-dimensional system models for identifying unexpected events. *Reliability Engineering and System Safety* 165: 317–330.

Turati, P., Cammi, A., Lorenzi, S., Pedroni, N., Zio, E. 2018. Adaptive simulation for failure identification in the Advanced Lead Fast Reactor European Demonstrator. *Progress in Nuclear Energy* 103: 176–190.

Wu, Y., Chen, Z., Hu, L., Jin, M., Li, Y., Jiang, J., Yu, J., Alejaldre, C., Stevens, E., Kim, K., Maisonnier, D., Kalashnikov, A., Tobita, K., Jackson D., and Perrault, D., 2016. Identification of safety gaps for fusion demonstration reactors. *Nature Energy* 1: Article number: 16154.

Zanino, R., Bonifetto, R., Hoa, C., Savoldi Richard, L. 2013. 4C modeling of pulsed-load smoothing in the HELIOS facility using a controlled by-pass valve. *Cryogenics* (57): 31–44.

Zanino, R., De Palo, S., Bottura, L. 1995. A two-fluid code for the thermohydraulic transient analysis of CICC superconducting magnets. *Journal of Fusion Energy* (14): 25–40.

Zio, E. 2013. The Monte Carlo Simulation Method for System Reliability and Risk Analysis. London, UK: Springer.

Zio, E. 2014. Integrated deterministic and probabilistic safety assessment: Concepts, challenges, research directions. *Nuclear Engineering and Design* 280: 413–419.

Zio, E., Di Maio, F., 2009. Processing dynamic scenarios from a reliability analysis of a nuclear power plant digital instrumentation and control system. *Ann. Nucl. Energy* 36(9): 1386–1399.

Zio, E., Di Maio, F., 2010. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear power plant. *Reliab. Eng. Syst. Saf*. 95: 1: 49–57.

Zio, E., Di Maio, F., Stasi, M., 2010. A data-driven approach for predicting failure scenarios in nuclear systems. *Ann. Nucl. Energy* 37: 482–491.

Zohm, H., 2014. *Magnetohydrodynamic Stability of Tokamaks*. Wiley-VCH Verlag GmbH & Co. KGaA.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Reliability-based design optimization by using support vector machines

Niclas Strömberg
Department of Mechanical Engineering, Örebro University, Sweden

ABSTRACT: In this work we perform Reliability-Based Design Optimization (RBDO) by classifying the limit states by using soft non-linear Support Vector Machines (SVM). By adopting the kernel trick in the dual formulation, by using e.g. the Gaussian kernel,we classify non-linear states of fail or safe obtained from design of experiments. The Most Probable Point (MPP) of the SVM is established in the physical space where the distance is minimized in the metric of Hasofer-Lind. The solution to the corresponding optimality conditions is obtained by using Newton's method with an inexact Jacobian and a line-search of Armijo type. At the MPP, we perform Taylor expansions of the SVM using intermediate variables defined by the iso-probabilistic transformation. In such manner, we derive a Quadratic Programming (QP) problem which is solved in the standard normal space. This is done for several probability distributions such as e.g. lognormal, Gumbel, gamma and Weibull. The optimal solution to the QP problem is mapped back to the physical space and new Taylor expansions of the SVM are derived and a new QP problem is formulated and solved. This procedure continues in sequence until we obtain convergence of our RBDO problem. The steps presented above constitute our proposed FORM-based sequential QP approach for RBDO by using SVM. The target of reliability appearing in the FORM-based QP problem might also be adjusted using different SORM formulas such as e.g. Breitung, Hohenbichler or Tvedt, or by applying importance-based Halton or Hammersley sampling. A nice feature of the proposed SVM-based RBDO approach is that several limit state functions can be represented simultaneously by only one single SVM. Thus, the proposed SVM-based RBDO methodology might be considered to be a rational approach for the treatment of RBDO problems including system reliability. This is demonstrated by solving established RBDO benchmarks.

1 INTRODUCTION

The soft non-linear support vector machine (SVM) introduced by Cortes & Vapnik (1995) defines a paradigm shift in machine learning and the paper has been cited more than 15000 times. By adopting the kernel trick and the soft penalization, we are able to classify non-linear separable data including misclassified data points. In this work, we suggest to use a single SVM to represent several limit state functions simultaneously when performing reliability based design optimization. For readers not familiar with SVM, an excellent introduction to this machine learning discipline is found in the textbook by Hamel (2009).

Although the number of papers on SVM-based RBDO is small, the idea of using SVM in order to represent limit state functions in RBDO is not new. Basudhar et al. (2008) solved RBDO problems using SVM and a particle swarm algorithm. Song et al. (2012) performed sampling-based RBDO by using probabilistic sensitivity analysis and virtual support vector machines. Khatibinia et al. (2013) suggested a gravitational search algorithm with a weighted least square support vector machine for RBDO. Wang et al. (2015) proposed a new SORA-based RBDO method using SVM as a surrogate model for the limit state function. Most recently Liu et al. (2017) used SVM-based sampling in order to improve Kriging models for RBDO. Another recent paper is by Yang & Husada (2017), who study seven state of the art methods from data mining in order to improve accuracy and efficiency of a single-loop RBDO method.

In this work, we will adopt the SORM-based sequential quadratic programming (SQP) approach for RBDO recently suggested by Strömberg (2017), where we now represent the limit state functions by using SVM instead of analytical expressions. The main idea of the proposed method is to represent several limit state functions simultaneously with only one single SVM. In such manner, a RBDO problem with several reliability constraints boil down to a problem with only one constraint. We also think that this approach is a step towards a rational method for treating problems with system reliability. Reliability analysis without optimization using a similar idea was recently investigated by Li et al. (2016).

The proposed method might be consider to be a metamodel-based approach for RBDO even though the SVM is not representing the overall behaviour of the reliability constraint function but only is a representation of the limit state or a classification of states in fail or safe. Metamodel-based RBDO is a most powerful approach for treating the reliability constraints of complex models such as non-linear finite element models. Popular metamodels for this kind of applications are e.g. Kriging Kriging (Hu et al. 2016), artificial neural networks (Zhu et al. 2011) and radial basis function networks (Lv et al. 2015). In Strömberg (2016) FORM- and SORM-based RBDO of a underrun protection profile was performed using a new type of radial basis function networks with a priori bias suggested by Amouzgar & Strömberg (2017).

The outline of the paper is as follow: in the next section we review the FORM- and SORM-based SQP approach for reliability based design optimization that recently was suggested in Strömberg (2017), in section 3 we present the theory of the soft non-linear support vector machine by deriving the dual formulation of the maximum margin problem using the Karush-Kuhn-Tucker (KKT) conditions and introducing the soft penalization. In section 4 two RBDO examples are solved using the proposed SVM-based methodology, in particular an established benchmark with three reliability constraints are treated using a single SVM and we also suggest how to apply SVM-based adaptive sampling in order to improve the solution. Finally, some concluding remarks are presented.

2 SQP-BASED RBDO

Most recently a SORM-based SQP approach for RBDO was suggested in Strömberg (2017). A brief presentation of this approach is given in this section.

Let us consider the following RBDO problem:

$$\begin{cases} \mu\min \quad f(\mu) \\ \text{s.t.} \quad \Pr\left[g(\boldsymbol{X}) \le 0\right] \ge P_s, \end{cases} \tag{1}$$

where $\boldsymbol{X}$ is a vector of N_{VAR} uncorrelated random variables X_i. The mean value μ_i of each variable X_i is collected in μ. The functions $f = f(\mu)$ and $g = (X)$ represent the objective function and constraint, respectively. Thus, the reliability constraint reads that the probability that $g \le 0$ must be greater than the target of reliability P_s. In this work, we treat this reliability constraint by representing the limit state $g = 0$ by support vector machines, see the next section. The SVM can also be used to classify states of fail or safe, but that is not the main purpose of using SVM in this work. The main idea is to represent several limit states simultaneously by using a single SVM and then toperform RBDO following the method presented in this section.

The cumulative distribution of each variable is given by

$$F_i(x;\boldsymbol{\theta}_i) = \int_{-\infty}^{x} \rho_i \, \mathrm{d}x, \tag{2}$$

where $\rho_i = \rho_i(x;\boldsymbol{\theta}_i)$ is the probability density function for distributions parameters collected in $\boldsymbol{\theta}_i = \boldsymbol{\theta}_i(\mu_i)$. So far, the following distributions have been implemented: normal, lognormal, Gumbel, gamma and Weibull.

The problem in (1) is solved by the SQP approach as outlined below. At an iterate k with mean values collected in μ^k, we perform Taylor expansions of f and g in intermediate variables Y_i defined by the iso-probabilistic transformation, i.e.

$$Y_i = Y_i(X_i) = \Phi^{-1}\left(F_i\left(X_i;\boldsymbol{\theta}_i\left(\mu_i^k\right)\right)\right), \tag{3}$$

where $\Phi = \Phi(x)$ is the cumulative distribution of the standard normal distribution. In addition, the Taylor expansion of g is done at the most probable point x_i^{MPP} on the limit surface. The Taylor expansion of f becomes $f(\boldsymbol{\eta}) \approx f(\boldsymbol{\mu}^k) +$

$$\sum_{i=1}^{N_{\text{VAR}}} \left.\frac{\partial f}{\partial X_i}\right|_{\boldsymbol{X}=\boldsymbol{\mu}^k} \frac{\phi\left(Y_i^k\right)}{\rho_i\left(\mu_i^k;\boldsymbol{\theta}_i^k\right)} \eta_i + \frac{1}{2} \sum_{i=1}^{N_{\text{VAR}}} \sum_{j=1}^{N_{\text{VAR}}} \tilde{H}_{ij} \eta_i \eta_j, \tag{4}$$

where $\boldsymbol{\eta}$ is the mean of $\boldsymbol{Y}$ and

$$\tilde{H}_{ij} = \left.\frac{\partial^2 f}{\partial X_i \partial X_j}\right|_{\boldsymbol{X}=\boldsymbol{\mu}^k} \frac{\phi\left(Y_i^k\right)}{\rho_i\left(\mu_i^k;\boldsymbol{\theta}_i^k\right)} \frac{\phi\left(Y_j^k\right)}{\rho_j\left(\mu_j^k;\boldsymbol{\theta}_j^k\right)}. \tag{5}$$

Furthrmore, $\tilde{g} = \tilde{g}(Y) \approx$

$$\sum_{i=1}^{N_{\text{VAR}}} \left.\frac{\partial g}{\partial X_i}\right|_{\boldsymbol{X}=\boldsymbol{x}^{\text{MPP}}} \frac{\phi\left(y_i^{MPP}\right)}{\rho_i\left(x_i^{MPP};\boldsymbol{\theta}_i^k\right)} \left(Y_i - y_i^{\text{MPP}}\right), \tag{6}$$

where $y_i^{\text{MPP}} = Y_i\left(x_i^{\text{MPP}}\right)$ is the most probable point defined by

$$\begin{cases} \min\limits_{X} \dfrac{1}{2} \boldsymbol{Y}(\boldsymbol{X})^T \boldsymbol{Y}(\boldsymbol{X}) \\ \text{s.t. } g(\boldsymbol{X}) = 0. \end{cases} \tag{7}$$

In this work, we solve (7) when the limit state $g(\boldsymbol{X}) = 0$ is represented by a soft non-linear SVM as presented in the next section. This is in turn done by solving the necessary optimality conditions by using Newton's method withan inexact Jacobian.

Finally, by inserting (4) and (6) into (1), we derive the following QP-problem:

$$\begin{cases} \min\limits_{\eta_i} \quad f(\boldsymbol{\eta}) \\ \text{s.t.} \quad \begin{cases} \mu_{\tilde{g}} \le -\beta_t \sigma_{\tilde{g}}, \\ -\int \le \eta_i \le \int, \end{cases} \end{cases} \tag{8}$$

where

$$\mu_{\tilde{g}} = \sum_{i=1}^{N_{\text{VAR}}} \left.\frac{\partial g}{\partial X_i}\right|_{\boldsymbol{X}=\boldsymbol{x}^{\text{MPP}}} \frac{\phi\left(y_i^{MPP}\right)}{\rho_i\left(x_i^{MPP};\boldsymbol{\theta}_i^k\right)}\left(\eta_i - y_i^{\text{MPP}}\right),$$
$$\sigma_{\tilde{g}} = \sqrt{\sum_{i=1}^{N_{\text{VAR}}} \left(\left.\frac{\partial g}{\partial X_i}\right|_{\boldsymbol{X}=\boldsymbol{x}^{\text{MPP}}} \frac{\phi\left(y_i^{MPP}\right)}{\rho_i\left(x_i^{MPP};\boldsymbol{\theta}_i^k\right)} \right)^2}. \tag{9}$$

Here, $\beta_t = \Phi^{-1}(P_s)$ is the target reliability index which can be corrected by any SORM approach or Monte Carlo sampling. So far, four SORM approaches have been implemented, e.g. Breitung, Hohenbichler and Tvedt. In addition, Halton- and Hammersley-based importance sampling at the MPP are also implemented. The optimal solution to (8), denoted η_i^*, is mapped back from the standard normal space to the physical space using

$$\mu_i^{k+1} \approx \mu_i^k + \frac{\Phi\left(Y_i^k\right)}{\rho_i\left(\mu_i^k;\boldsymbol{\theta}_i^k\right)}\eta_i^*.$$

Then, a new QP-problem is generated around $\boldsymbol{\mu}^{k+1}$ and this procedure continues in sequence until convergence is obtained. The QP-problem in (8) solved using quadprog.m in Matlab.

3 SUPPORT VECTOR MACHINE

In this section, we present the dual formulation of the soft non-linear support vector machine. First we introduce the original linear SVM, which actually was suggested already in the 60 s by Vapnik, then we apply the kernel trick and, finally, regularize the problem.

Let us consider N sampling points x^i, which take values $y^i = 1$ (fail) or $y^i = -1$ (safe). Furthermore, we assume that it exists hyper-planes

$$w \cdot x + b = 0, \tag{10}$$

which separate these sampling points into two subsets; one that only takes values $y^i = 1$ and the other one with values $y^i = -1$. This is shown in Figure 1, where the SVM-based RBDO methodology is illustrated. We also assume that the following constraints are satisfied:

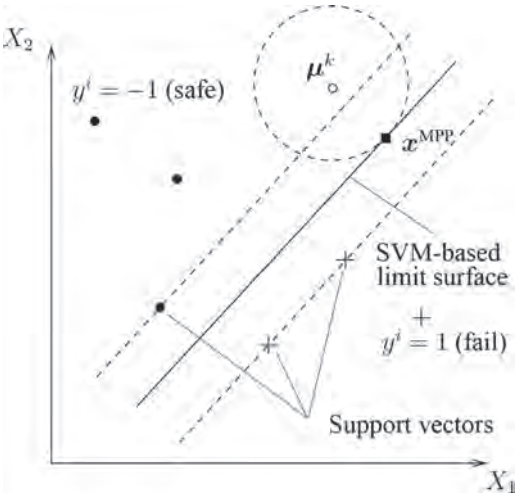


Figure 1. Illustration of the proposed SVM-based RBDO approach.

$$y^i\left(\boldsymbol{w}\cdot\boldsymbol{x}^i + b\right) \ge 1, \qquad i = 1,\ldots,N. \tag{11}$$

The shortest distances to $\boldsymbol{x}^i$ from a hyper-plane defined in (10) is given by

$$\boldsymbol{x}^i = \boldsymbol{x} + \gamma^i \frac{\boldsymbol{w}}{\| \boldsymbol{w} \|}. \tag{12}$$

(12) inserted in (11) yields

$$y^i\left(\boldsymbol{w}\cdot\boldsymbol{x} + b + \gamma^i \| \boldsymbol{w} \|\right) \ge 1. \tag{13}$$

By utilizing (10), one obtains

$$\gamma^i \ge 1/\| \boldsymbol{w} \| \text{ for } y^i = 1, \tag{14a}$$

$$\gamma^i \le -1/\| \boldsymbol{w} \| \text{ for } y^i = -1. \tag{14b}$$

Thus, the lower bound on the shortest distance $|\gamma^i|$ is maximized by minimizing $\| \boldsymbol{w} \|$. This is the key idea of the original linear support vector machine formulation, which reads

$$\begin{cases} \min\limits_{(w,b)} \dfrac{1}{2} \| \boldsymbol{w} \|^2 \\ \text{s.t.}\, 1 - y^i\left(\boldsymbol{w}\cdot\boldsymbol{x}^i + b\right) \le 0, i = 1,\ldots N. \end{cases} \tag{15}$$

Obviously, the closest sampling points to the optimal hyper-plane

$$\boldsymbol{w}^*\cdot\boldsymbol{x} + b^* = 0 \tag{16}$$

are obtained when

$$y^i\left(\boldsymbol{w}^*\cdot\boldsymbol{x}^i+b^*\right)=1. \tag{17}$$

Sampling points satisfying (17) are called support vectors, see Figure 1. It is also obvious that in the region between the optimal hyper-plane defined by (16) and the support vectors is empty of sampling points. Thus, the support vector machine formulation in (15) finds a hyper-plane that maximizes the size of this region. This region is augmented with sampling points in the SVM-based adaptive sampling approach discussed in the next section.

The Karush-Kuhn-Tucker conditions of the support vector machine in (15) are given by

$$\mathbf{0}=\boldsymbol{w}-\sum_{i=1}^{N}\lambda_i y^i\boldsymbol{x}^i, \tag{18a}$$

$$0=\sum_{i=1}^{N}\lambda_i y^i, \tag{18b}$$

$$\lambda_i\geq 0, \tag{18c}$$

$$1-y^i\left(\boldsymbol{w}\cdot\boldsymbol{x}^i+b\right)\leq 0, \tag{18d}$$

$$\lambda_i\left(1-y^i\left(\boldsymbol{w}\cdot\boldsymbol{x}^i+b\right)\right)=0. \tag{18e}$$

The corresponding Lagrangian function is $\mathcal{L}=\mathcal{L}\left(\boldsymbol{\lambda},\boldsymbol{w},b\right)=$

$$\frac{1}{2}\parallel\boldsymbol{w}\parallel^2+\sum_{i=1}^{N}\lambda_i\left(1-y^i\left(\boldsymbol{w}\cdot\boldsymbol{x}^i+b\right)\right). \tag{19}$$

Furthermore, the dual formulation of the support vector machine in (15) reads

$$\max_{\boldsymbol{\lambda}\geq\mathbf{0}}\min_{(w,b)}\mathcal{L}\left(\boldsymbol{\lambda},\boldsymbol{w},b\right). \tag{20}$$

By inserting (18a) in (19), one obtains $\mathcal{L}\left(\boldsymbol{\lambda},w(\boldsymbol{\lambda}),b\right)=$

$$-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y^i y^j\boldsymbol{x}^i\cdot\boldsymbol{x}^j+\sum_{i=1}^{N}\lambda_i-b\sum_{i=1}^{N}\lambda_i y^i. \tag{21}$$

In addition, the latter part is zero by (18b). In conclusion, the dual support vector machine formulation is given by

$$\begin{cases}\min\limits_{\lambda}\dfrac{1}{2}\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N}\lambda_i\lambda_j y^i y^j\boldsymbol{x}^i\cdot\boldsymbol{x}^j-\sum\limits_{i=1}^{N}\lambda_i\\ \text{s.t.}\begin{cases}\sum\limits_{i=1}^{N}\lambda_i y^i=0,\\ \lambda_i\geq 0,\boldsymbol{i}=1,\ldots,N.\end{cases}\end{cases} \tag{22}$$

From the optimal solution $\boldsymbol{\lambda}^*$ of the dual support vector machine in (22), we obtain the corresponding support vector machine solution from the Karush-Kuhn-Tucker conditions in (18) as

$$\boldsymbol{w}^*=\sum_{i=1}^{N}\lambda_i^* y^i\boldsymbol{x}^i \tag{23}$$

and

$$b=1/\,y^i-\boldsymbol{w}^*\cdot\boldsymbol{x}^i \tag{24}$$

for any $\lambda_i^*>0$. Notice, by using (23), the optimal hyper-plane in (16) can be written as

$$\sum_{i=1}^{N}\lambda_i^* y^i\boldsymbol{x}^i\cdot\boldsymbol{x}+b^*=0. \tag{25}$$

Notice also that you only need to do the summation over support vector indices, because otherwise λ^* equals zero by the KKT-conditions. This is utilized in the implementation in order to speed up the evaluation of the SVM.

For non-separable sets of sampling points x^i, the support vector machine approach presented above will of course not work. However, one might transform the sample set to a new space where it become separable, let say by $\boldsymbol{\xi}=\boldsymbol{\xi}(x)$. In this new space, the only difference in the derivations of the dual support vector machine in (22) and (25) is the appearance of a new scalar product $\left\langle\boldsymbol{\xi}^i,\boldsymbol{\xi}^j\right\rangle$ instead of $\boldsymbol{x}^i\cdot\boldsymbol{x}^j$. Thus, we donot have to know the explicit expression of the transformation $\boldsymbol{\xi}=\boldsymbol{\xi}(\boldsymbol{x})$, but only the expression of the scalar product of the new space. The explicit expression of this scalar product is known to be the kernel function, i.e.

$$k\left(\boldsymbol{x}^i,\boldsymbol{x}^j\right)=\left\langle\boldsymbol{\xi}^i\left(\boldsymbol{x}^i\right),\boldsymbol{\xi}^j\left(\boldsymbol{x}^j\right)\right\rangle. \tag{26}$$

Consequently, by using an appropriate kernel function in (22) instead of $\boldsymbol{x}^i\cdot\boldsymbol{x}^j$, e.g. the Gaussian kernel

$$k=k\left(\boldsymbol{x},\boldsymbol{z}\right)=\exp\left(-\frac{\parallel\boldsymbol{x}-\boldsymbol{z}\parallel^2}{2\sigma^2}\right), \tag{27}$$

the sample set can be separated by

$$\sum_{i=1}^{N}\lambda_i^* y^i k\left(\boldsymbol{x}^i,\boldsymbol{x}\right)+b^*=0. \tag{28}$$

Another kernel function is the polynomial kernel, i.e.

$$k=k\left(\boldsymbol{x},\boldsymbol{z}\right)=(1+\boldsymbol{x}\cdot\boldsymbol{z})^p. \tag{29}$$

Even if we perform a suitable kernel trick, we might have some misclassified points such that

(22) does not converge to a solution. This can be treated by applying a regularization of (22). The established soft regularization of (15) is

$$\begin{cases} \min\limits_{(w,b,v)} \dfrac{1}{2} \| \boldsymbol{w} \|^2 + C \sum\limits_{i=1}^{N} v_i \\ \text{s.t.} \begin{cases} 1 - v_i - y^i \left(\boldsymbol{w} \cdot \boldsymbol{x}^i + b \right) \le 0, i = 1, \ldots N, \\ v_i \ge 0, i = 1, \ldots N. \end{cases} \end{cases} \tag{30}$$

The Karush-Kuhn-Tucker conditions in (18) then modify by adding the following conditions:

$$C - \lambda_i \ge 0, \tag{31a}$$

$$v_i \ge 0, \tag{31b}$$

$$v_i \left(C - \lambda_i \right) = 0. \tag{31c}$$

The corresponding Lagrangian becomes $\mathcal{L} =$

$$-\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y^i y^j \boldsymbol{x}^i \cdot \boldsymbol{x}^j + \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} \lambda_i v_i + C \sum_{i=1}^{N} v_i, \tag{32}$$

where the two latter terms cancel out due to (31c). Thus, the only difference of the dual support vector machine in (22) for this regularization is the appearance of an upper bound on λ_i, i.e.

$$\begin{cases} \min\limits_{\lambda} \dfrac{1}{2} \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} \lambda_i \lambda_j y^i y^j k\left(\boldsymbol{x}^i, \boldsymbol{x}^j \right) - \sum\limits_{i=1}^{N} \lambda_i \\ \text{s.t.} \begin{cases} \sum\limits_{i=1}^{N} \lambda_i y^i = 0, \\ 0 \le \lambda_i \le C, i = 1, \ldots, N. \end{cases} \end{cases} \tag{33}$$

Finally, $0 < \lambda_i < C$ must be satisfied in order for (24) to be valid. Here, we have also introduced the kernel $k(\boldsymbol{x}, \boldsymbol{y})$ in the objective function. The soft non-linear SVM in (33) is solved using quadprog.m in Matlab.

4 EXAMPLES

In this section, we demonstrate by solving two examples that the soft non-linear SVM presented in the previous section can represent the limit states properly such that the SQP-based RBDO methodology can be applied for solving RBDO problems with SVM-basedlimit state functions. The first problem was considered in Strömberg (2016), where SLP-based RBDO was performed by adopting radial basis function networks (RBFN) as metamodels. The second example is a most well-know benchmark for evaluating new RBDO approaches, see e.g. Youn and Choi (2004).

The first example reads

$$\begin{cases} \mu_i \min \quad \sqrt{1000 \left(\dfrac{4}{\mu_1} - 2 \right)^2 + 1000 \left(\dfrac{4}{\mu_1} - 2 \right)^2} \\ \text{s.t.} \begin{cases} \Pr[(X_1 - 0.5)^4 + (X_2 - 0.5)^4 \le 2] \ge P_s, \\ 1 \le \mu_i \le 4, \end{cases} \end{cases} \tag{34}$$

where $P_s = 0.999$ and $\text{VAR}[X_i] = 0.1^2$. The deterministic solution is (1.5,1.5) and the minimum of the unconstrained objective function is found at (2,2). The constraint $g > 0$ $g > 0$ is plotted to the left in Figure 2. The solution to (34) obtained by our SQP-based RBDO approach is (1.2705,1.2705). The corresponding reliability is 99.9%.

Now, let us consider the problem in (34) as a "black-box", which we represent with metamodels for a sample set of 100 Hammersley points as depicted in Figure 4 for the next example. In particular, we let the objective function be represented by a radial basis function network as in Strömberg (2016) and the limit state is represented by a SVM

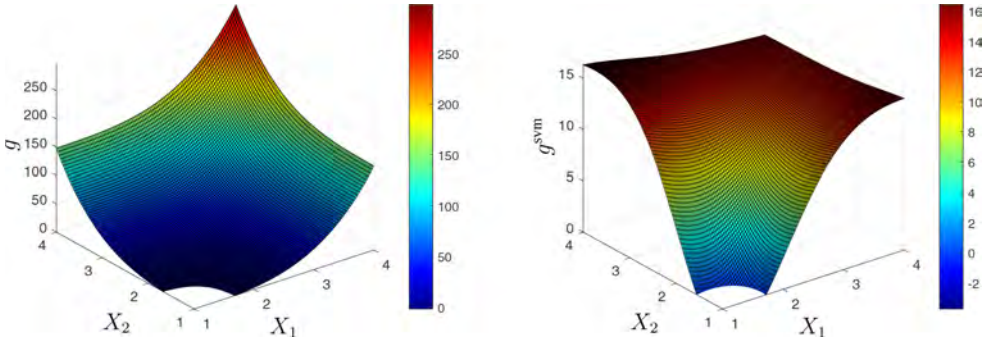


Figure 2. The left plot shows $g > 0$ and in the right plot $g^{\text{svm}} > 0$ is given. Despite the overall behaviour is not captured by the SVM, the limit state is represented properly.

g^{svm}. Figure 2, $g^{\text{svm}} > 0$ is plotted. Notice that g^{svm} only represents the limit state properly but not the overall behaviour. Of course, the classification of fail or safe is obtained by taking sign(g^{svm}). Instead of solving (34), we solve the metamodel-based representation of (34), i.e. the radial basis function taken as our objective function and g^{svm} is the limit state function. The corresponding solution is (1.3410, 1.2315) and the reliability for this solution becomes 99.7%, slightlylower than the target of 99.9%. Histograms for the objective function and the constraint are plotted in Figure 3. Notice the non-symmetric characteristic of the solution. This as well as the reliability of the solution can be improved by adopting adaptive sampling. This is a topic for future work. In fact, the SVM is most appropriate for improving the sample set with data along the limit state in the margin of the SVM. Some ideas of how this can be done is outlined for the next example.

The second example is a most well-known benchmark with three constraints and reads

$$\begin{cases} \mu_i \min \quad (\mu_1 + \mu_2)^2 \\ \text{s.t.} \begin{cases} \Pr\left[g_1 = 20 - X_1^2 X_2 \le 0\right] \ge \Phi(3), \\ \Pr\left[g_2 = 1 - \dfrac{(X_1 + X_2 - 5)^2}{30} - \dfrac{(X_1 - X_2 - 12)^2}{120} \le 0\right] \ge \Phi(3), \\ \Pr\left[g_3 = X_1^2 + 8X_2 - 75 \le 0\right] \ge \Phi(3), \\ 1 \le \mu_i \le 7, \end{cases} \end{cases} \tag{35}$$

where $\text{VAR}[X_i] = 0.3^2$. A small modification of the original problem is done by taking the square of the objective function. This problem was recently considered in Strömberg (2017) by using a SORM-based SQP approach for reliability based design optimization. The example was in that work also generalized to 50 variables and 75 constraints for five different distributions simultaneously (normal, lognormal, Gumbel, gamma and Weibull). The analytical solution for two variables with normal distribution is obtained to be (3.4525,3.2758) with our RBDO algorithm.

In this work we consider (35) to be a "blackbox" for which we setup a design of experiments using Hammersley sampling with 100 points as shown in Figure 4. From this sampling we define a training set g^{train} for our support vector machine in the following manner:

$$g^{\text{train}} = \begin{cases} 1 & \text{if any } g_i > 0, \\ -1 & \text{otherwise.} \end{cases} \tag{36}$$

For this training set, we obtain a SVM g^{svm} according to Figure 4. The plot to the right clearly shows that the the limit state functions are well represented by the global SVM $g^{\text{svm}} = 0$. The corresponding classification of fail or safe using sign(g^{svm}) is given in Figure 6. As in the previous example, we represent the sampling data for the objective function with a RBFN. Thus, our RBDO problem to be solved is this RBFN-based objective function with the reliability constraint on the SVM $\Pr\left[g^{\text{svm}} \le 0\right] \ge \Phi(3)$. In such manner, we reduce the problem size from three constraints to a problem with only one constraint. For this problem, the RBDO algorithm produces the following solution: (3.4247,3.2218). The corresponding reliability indices for the two first active constraints are 2.84 and 2.87, respectively. Histograms for these constraints are given in Figure 5.

Instead of solving (35) using a uniform sampling as presented in Figure 4, we adopt a sequential

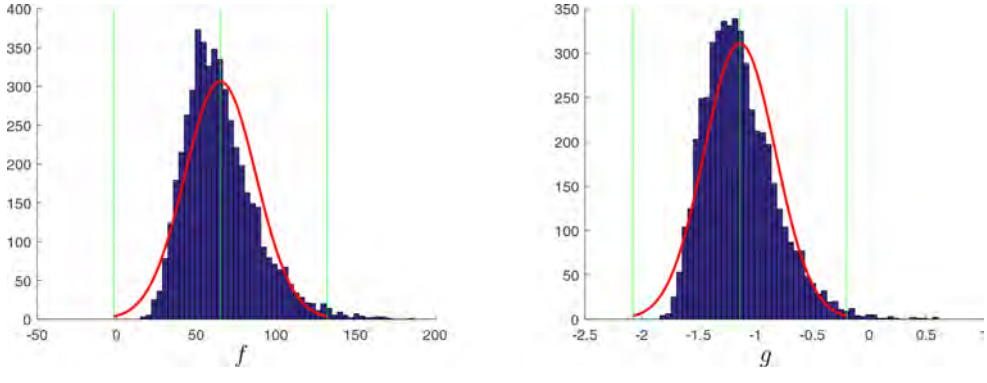


Figure 3. Histogram of the objective function f and the constraint g for the solution obtained by using the SVM. Red lines are showing the corresponding normal distribution, where green lines mark the $\mu \pm 3\sigma$ intervals.

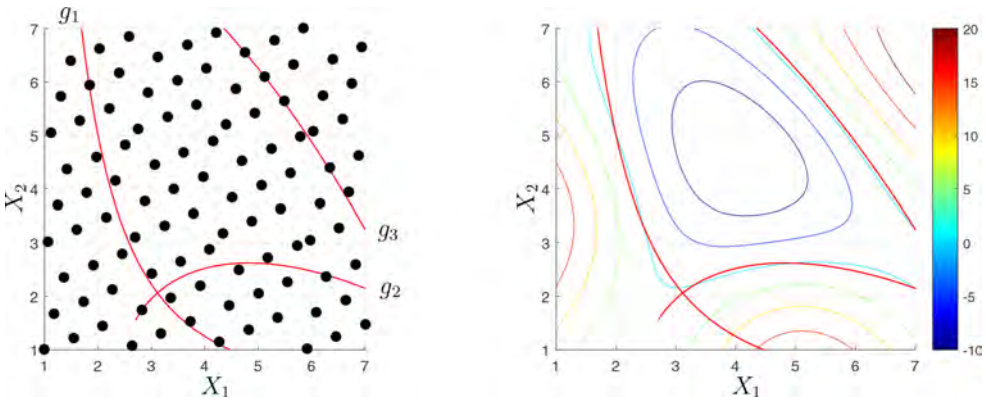


Figure 4. In the left plot the Hammersley sampling of 100 points is shown together with the explicit constraints g_i. To the right the corresponding SVM is depicted. The limit state surface in cyan corresponds well to the actual limit surfaces g_i.

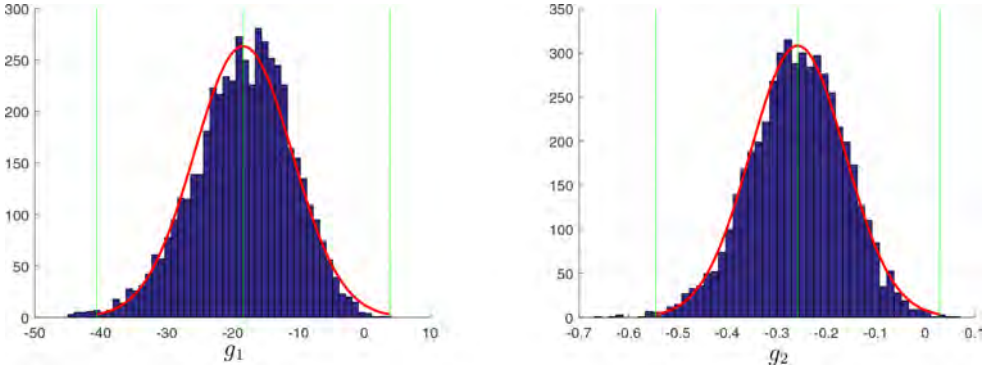


Figure 5. Histograms for the two first active constraints g_1 and g_2, respectively. Red lines are showing the corresponding normal distribution, where green lines mark the $\mu \pm 3\sigma$ intervals.

adaptive sampling approach by distributing points in the margin of the SVM along the limit states as well as adding the deterministic optimum and the most probable point obtained in each sequence. For instance, in Figure 7, we start from a set of 50 Hammersley points and then add 10 points (8 SVM-based points, 1 deterministic optimum, 1 MPP) sequentially five times following this SVM-based sampling approach in order to generate the set of 100 points shown in Figure 7. In this figure the corresponding SVM is also plotted. Comparing the SVMs in Figure 4 and Figure 7, one can see that the latter onebetter capture the local behavior of the limit surface close to the optimal solution. The optimal solution for this SVM-based formulation of the example is (3.4805,3.2585) and the corresponding reliability indices are 3.08 and 2.95, respectively, for the two active constraints. This is rather close to the target of 3. It is clear that this

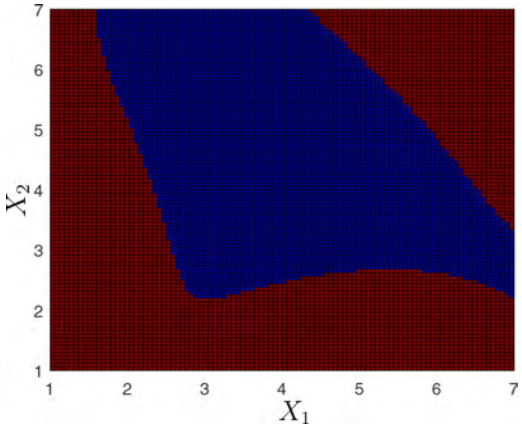


Figure 6. Classification in fail (red) or safe (blue) using the SVM.

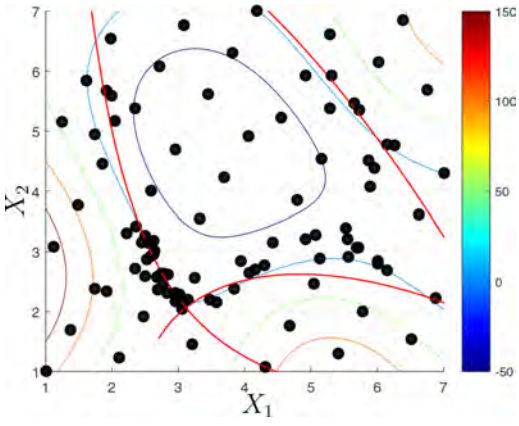


Figure 7. SVM-based adaptive sampling augmented with deterministic optimum and MPP.

is an improvement compared to the previous solution presented above for the same example.

Finally, we solve (35) for all three constraints modelled by three separate SVM models using the adaptive DoE used above. Then, we obtain the following solution (3.4304,3.3053), and the corresponding reliability indices for the active constraints: 2.97 and 3.12.

5 CONCLUDING REMARKS

In this work SVM-based RBDO is investigated by implementing soft non-linear SVM together with a SQP-based RBDO method recently developed in Strömberg (2017). The implementation is done so far for two random variables and the idea is to represent the limit state functions with a single SVM. It is demonstrated that the proposed approach works well for an established benchmark with three reliability constraints. It is also demonstrated how the SVM can be utilized in adaptive sampling. For future work it would beinteresting to investigate the approach for more than two variables with several constraints.

REFERENCES

Amouzgar, K. & N. Strömberg (2017). Radial basis functions as surrogate models with a priori bias in comparison with a posteriori bias. *Struct. Multidisc. Optim. 55*, 1453–1469.

Basudhar, A., S. Missoum, & A. Sanchez (2008). Limit state function identification using support vector machines for discontinuous repsonses and disjoint failure domains. *Probabilistic Engineering Mechanics 23*, 1–11.

Cortes, C. & V. Vapnik (1995). Support-vector networks. *Machine Learning 20*, 273–297.

Hamel, L. (2009). *Knowledge Discovery with Support Vector Machines*. Hoboken, New Jersey: Wiley-Blackwell.

Hu, W., K. Choi, & H. Cho (2016). Reliability-based design optimization of wind turbine blades for fatigue life under dynamic wind load uncertainty. *Struct. Multidisc. Optim. 54*, 953–957.

Khatibinia, M., E. Salajegheh, J. Salajegheh, & M. Fadaee (2013). Reliability-based design optimization of reinforced concrete structures including soil-structure iteraction using a discrete gravitational search algorithm and a proposed metamodel. *Engineering Optimization 45*, 1147–1165.

Li, H., A. Zhao, & K. Tee (2016). Structural reliability analysis of multiple limit state functions using multi-input multioutput support vector machine. *Advances in Mechanical Engineering 8*, 1–11.

Liu, X., W. Yizhong, B. Wang, J. Ding, & H. Jie (2017). An adaptive local range sampling method for reliability-based design optimizatio using support vector machine and kriging model. *Struct. Multidisc. Optim. 55*, 2285–2304.

Lv, X., X. Gu, L. He, Z. D., & W. Liu (2015). Reliability design optimization of vehicle front-end structure for pedestrian lower extremity protection under multiple impact cases. *Thin-Walled Structures 94*, 500–511.

Song, H., K. Choi, I. Lee, L. Zhao, & D. Lamb (2012). Sampling-based rbdo using probabilistic sensitivity analysis and virtual support vector machine. In *ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference IDETC/CIE*, Chicago, Illinois, USA.

Str¨omberg, N. (2016). Reliability based design optimization by using a slp approach and radial basis function networks. In ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference IDETC/CIE, Charlotte, North Carolina, USA.

Str¨omberg, N. (2017). Reliability-based design optimization using sorm and sqp. *Struct. Multidisc. Optim. 56*, 631–645.

Wang, Y., Y. Xiongqing, & D. Xiaoping (2015). Improved reliability-based optimization with support vector machines and its application in aircraft design. *Mathematical Problems in Engineering 2015*, 1–14.

Yang, I.-T. & W. Husada (2017). Improving classification accuracy for single-loop reliability-based design optimization. In *IMECS 2017 the InternationalMultiConference of Engineers and Computer Scientists*, Hong Kong.

Zhu, P., Y. Zhang, & G. Chen (2011). Metamodeling development for reliability-based design optimization of automotive body structure. *Computers in Industry 62*, 729–741.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Crisis management in extreme situation: The Model of Resilience in Situation (MRS) as a support to observe the organization with simulation

Q. Baudard, P. Le Bot & C. De la Garza
EDF Research and Development, Human and Organizational Factors Group

ABSTRACT: The traditional approach to safety engineering for a nuclear reactor is mainly focused on its technical dimension, while the Human Factors approach focuses on how to optimize the value added by human to the reliability of the system. Safety management seeks to organize the work as well as possible, train employees, develop their safety culture, and so on. This juxtaposition of approaches to reliability illustrates the recurring dilemma faced by at-risk organizations in choosing between the technical anticipation of pre-defined situations and the optimization of the management of the situations by people who, in real time, through their skills and understanding of the situation going on, will adapt their strategy and actions according to that particular situation. These last three years, we have performed a series of Extreme Situation Simulations on Full Scale Simulators in order to test the resilience of the Crisis Organization of EDF during accidents with characteristics similar to the accident of Fukushima. The realization of these tests was an organizational challenge which required up to 80 people for the simulation, and about 5 months for both the preparation and the analysis of the tests.

1 OBJECTIVE OF THE STUDY

Following the Fukushima accident, EDF implemented additional crisis management measures (organizational, material, etc.) to respond to an accident in an Extreme Situation (ES). An Extreme Situation is a situation in which a nuclear site is isolated and inaccessible as consequences of a large-scale external event which has an impact on all the reactors of the site, and during which the operating teams have limited means of communication. The objective of our works was to study the design of the operating teams in Extreme Situations and the National Crisis Management Organization of EDF, in order to identify the strengths and the areas of improvement for crisis management.

To do this, we observed simulations involving the entire crisis management organization of EDF, which would operate in an ES, from the operating teams on-site, to the experts from the National Technical Support Team, along with members of the National Direction Command Post. External stakeholders (Public Powers, Regulatory Authority…) have not been taken into account in the simulation.

We applied a multidisciplinary approach to these simulations, cross-referencing the analyses of experts in Cognitive Ergonomics, Human Reliability and Nuclear Safety. Our method of analysis of these simulations is based on the Model of Resilience in Situation (Le Bot & Pesme, 2010), which explains how a socio-technical system such as the one we have observed can continue to operate safely using a process of anticipation and a process of adaptation. Therefore, beyond the operational objective of studying the organizational system anticipated in an ES, we examined more closely the link between resilience and crisis management.

The purpose of this communication is to present the method we have developed for the implementation of the Extreme Situation Tests and the one used for their analysis. We will first see how the specific context of an ES is complex to simulate, and how we managed to achieve it. We will then detail the method of observation of the tests and analysis, focusing on the functional analysis of the resilience in accordance with the Model of Resilience in Situation. Finally, we will present some of the conclusions we have drawn from the analyses and the first lessons for the implementation of simulations.

2 TESTS IN EXTREME SITUATIONS

2.1 *What is an extreme situation?*

The concept of Extreme Situations emerged following the accident at Fukushima in 2011. The situation we considered as representative for the study of the crisis organization is a situation with characteristics similar to this one.

We suppose that a nuclear site is hit by a major earthquake leading to the loss of internal and external power supplies on at least two reactors, with the isolation of the site preventing the immediate arrival of the local on-call emergency response teams and the Safety Engineer.

Furthermore, nearly all internal and external communication systems of the site have been rendered unavailable by the event: in the first tests, the control room could not communicate remotely with the field operators when they are working on field manoeuvers. In later tests, an autonomous communication system was available for the teams to communicate with the field operators. In the short-term, the team in the control room can only receive support from the National Emergency Response Team by communicating with the "National Direction Command Post" via an emergency satellite telephone.

2.2 *How to simulate an extreme situation?*

During our study, we observed two series of three tests, each involving complete operating teams on full scale simulators. Four tests were carried out on two simulators in parallel, representing a beyond-design-basis "multi-unit" site accident, simultaneously affecting two reactors on one site. During one of these exercises, the two simulated units were of different technologies. The two remaining tests were focused on thermohydraulic design basis accidents in situations combined with another event (fire or flooding), in order to study the design of the team in a non-isolated site situation.

A multi-disciplinary work group, consisting of representatives of the site operations departments, training instructors on simulators and experts on ergonomics, human reliability and safety, prepared the tests. This group produced the test protocol and defined a scenario able to provide sufficient data to understand crisis organization in a beyond-design-basis extreme situation. The accident scenarios were tested and validated in technical and documentary terms, and lastly, for each test, a prior "dummy" test was used to check the entire simulation system with site operating teams.

This study meant that for the first time within the company, tests could be performed on two full scale simulators in parallel built in the training centre of the sites, a "hardware" one reproducing the control room exactly, and another "digital" one with touch screens representing the devices of the control room. During one of the tests, a third unit was simulated on paper: it was a world's first. The simulation of several units raises numerous problems, for example, certain units are paired and share common equipment. This is not the case for the full scale simulators, they are technically independent. Thus, during tests simulating two paired reactors sharing common systems, the unavailability of shared equipment needed to be simulated for one simulated reactor if the other simulated reactor used it.

Furthermore, following the Fukushima accident, it was decided to equip each reactor with complementary equipment for facing extreme situations, such as emergency unit cooldown diesel generator sets: these were not yet taken into account in the simulators, or in the training of the operating teams at the time of the tests. Since the objective of the simulations is to study the operation of the crisis organization as it would be with this equipment deployed, these devices were provisionally simulated for the tests and the operating teams participating in the tests were specially prepared in their use. The validation of the new procedures was not an objective of the tests.

Lastly, it emerged from our preliminary analyses of the Fukushima accident (Baudard, 2017) that the field actions had played a very important role in the crisis management. We therefore involved the field operators of each operating team in the simulation by asking them to simulate the completion of the actions requested by the control room in a degraded environment (poor lighting, access path blocked, etc.). This participation by the field operators helped to ensure a more realistic simulation, even though they have not simulated their actions directly in the field.

2.3 *The observation system*

To limit as far as possible any technical contingencies in the progress of the test, each scenario was played out twice with different teams before being observed during the final test. The aim of these precautions was to seek possible faults in the procedures

Table 1. Data collection methods.

In-situ observation	Ergonomics	Human reliability
Note taking	Chronologies, actions performed, decision making, communication, difficulties observed.	
Video & audio recording	Detailed subsequent analysis of sequences chosen	Used only in case of doubt
Types of instrumented collection		Process evolution Logbooks
Post-simulation debriefings	Debriefing focused on the Notable Events in the organization, noted during the observation and discussed between observers during the preparation of the debriefing	

which were being designed, but also to allow the trainers to adapt to this unusual scenario.

The tests involved up to 80 people, from its preparation through to the implementation:

- two complete operating teams and one observer per team member
- field operators for each team and one observer per group
- the experts of the national Technical Support Team and two dedicated observers
- members of the National Direction Command Post and two dedicated observers
- Scenario creators and trainers
- In-house specialists and others from outside of the company coming to observe the method of data collection
- etc.

The system observed is adapted to the crisis organization which would take place in an ES in order to represent it as accurately as possible. However, entities from outside of the company (prefecture, regulatory agency, media, etc.) were not simulated. The system is characterized by the six observation posts (two simulators, two teams of field operators, the National Technical Support Team (NTST) and National Direction Command Post) across which the ergonomics, human reliability and safety observers were spread. The simulation takes place over a period of five hours, followed by an on-the-spot debriefing and post-analysis of the simulator logbook. The data collection methods are detailed in the following table:

Five work themes guided the organization, the observation and the analysis of the Extreme Situation Simulations:

- The design of the operating team
- Field actions management
- Information Exchange with the National Crisis Organisation
- The use of the tools and resources available in an ES
- The resilience of the organisation, which we will look at in more detail later.

3 DATA ANALYSIS

3.1 *Analysis method*

The main stages in the analysis method are summarized in Table 2 below:

Following the observation of a test, the multidisciplinary analysis group identifies the favorable and unfavorable factors for the organizational resilience of the socio-technical system in Extreme Situations. The common document base for analysts is as follows:

- The exhaustive chronologies by group (operating team, Technical Support Team, etc.) is reproduced based on the notes taken by each observer, distinguishing the actions carried out by the group (application of procedures, launching field actions, etc.) and events independent of the group (equipment failure, action carried out by an independent group, etc.)
- Identification of "Notable Events" (NE). A Notable Event is an event or the repetition of an event which reveals an action, the absence of an action, a decision made, a collective or individual initiative, a fact that can strengthen the reliability of the organisation of the operating team, and/or the emergency response team, operations, its robustness, facilitating sensemaking, or on the contrary which may make the socio-technical system less reliable and make safety barriers more fragile, damaging sensemaking. The NE are observed during the in situ observation, subsequently during the group debriefing, or during the reconstruction of the overall chronology. NE are the central elements in the analysis of the resilience of the socio-technical system.

The NE are then categorized according to the groups they refer to (control room 1 or 2, emergency response team, management control unit, field operators), and a first level of analysis is used to identify:

- the Technical NE, resulting from operations on the process

Table 2. Contribution of the different disciplines to the analysis of the situations.

					Ergonomics		
					Cognitive analyses and analyses of group operations		
Observers	Human reliability + ergonomics		Human reliability				
Chronology by observer, then by Group, then overall chronology	Raw analysis from observations and debrefings Notable Events	Technical points Summary of notable events by group and overall	Monacos chronologies progress of operations	Functional analysis of resilience	Timing charts of control room activity and Site/ National interactions	Themed analyses of debriefings	Analysis by hypothesis

- the Expertise NE on the assistance that the experts provide for the operations
- the Organizational NE on the operation of the group as a whole.

Lastly, the Human Reliability analysts link together the Notable Events, the characteristics of the system observed, and the resilience functions defined in the MRS.

3.2 *Model of resilience in situation*

The Model of Resilience in Situation (Le Bot & Pesme, 2010) is an empirical model which was build out of analyses of real or simulated accidents, based on the theoretical framework of Social Regulation (Reynaud, 1997). The model serves to reconcile two seemingly opposite rationales: the anticipation of potential situations on the one hand and adaptation to the situation on the other. It is used to dynamically describe the operational management of a crisis situation, alternating between periods of constructing ad'hoc operating rules and periods of application of these rules. Should the rules being applied become obsolete, the process will repeat.

The MRS applies to crisis organization as a whole, as a dynamic network of work groups (operating team, experts, etc.) that interact, cooperate, collaborate and coordinate themselves. Each of these groups, taken within its own environment (procedures, HMI, etc.), is considered a distributed cognitive interactive system. The overall resilience of the organization therefore results from interactions within the groups and interactions between groups.

The model links together two processes: the execution process, and the adaptation process based on Figure 1:

In stable operating conditions, the system executes the operation rules (EXECUTION). The functions to be performed continuously are:

– INFORMATION: selection and sharing of information based on the surveillance criteria
– ACTION: act based on the objectives and their priorities with the corresponding resources
– CONTROL: ensure that the action complies with the operating rules

If the continuous VERIFICATION detects that the rules are not appropriate or are obsolete (objective achieved), the system initiates a Rupture phase after a RECONFIGURATION: interruption of the rules which are not relevant, mobilization of resources, then ADAPTATION in order to readjust the operating rules by carrying out DIAGNOSIS, PROGNOSIS, SELECTION of relevant procedures and parameters, PRIORITISATION of objectives, COLLABORATION to negotiate and define the operating rules, COOPERATION to distribute the tasks and the resources.

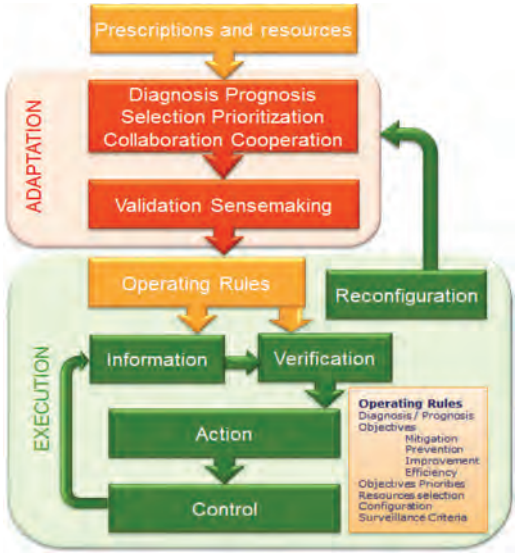


Figure 1. The model of resilience in situation.

The operating rules are validated (VALIDATION) and shared to make the situation relevant, implementing them and linking the past experience, present actions and future projections of the parties involved.

3.2.1 *Definitions of the MRS functions*

Adaptation process:
The adaptation process involves redefining operation: objectives, strategy and resources to achieve the objectives. This means:

– Anticipating the behaviour of the installation and the actions to be carried out (diagnosis and prognosis functions)
– Selecting the relevant information, procedures, instructions and means/resources (selection function)
– Collaborating to adapt them autonomously if necessary, to define the new strategy with the operational objectives and resources (collaboration and prioritisation functions)
– Validating the implementation of the rule by the authorised party (validation function).

Execution process:
The execution process involves robustly implementing the agreed strategy:

– The robustness is obtained by the execution (action function) and the control of the ongoing actions (control function).

– The group constantly checks that this strategy remains appropriate in regard of the situation (verification function)
– These functions are carried out by the acquisition and the sharing of information (information function), and guided by sensemaking (sensemaking function)

3.3 *Functional analysis of resilience*

During our analyses of the tests, we seek to relate Notable Events with these resilience functions and estimate whether they are favorable or unfavorable. For example, the contribution of a member from outside of the operating team to the production of the rules to be followed is a favorable factor for COLLABORATION between groups. On the contrary, the workload of parties involved in Extreme Situations slowed down the management of field actions, which was an unfavorable factor for the ACTION and CONTROL functions to be carried out, as well as for the PRIORITISATION of actions.

Our analysis can be summarized by the diagram on Figure 2. Taking the example of the use of a new field actions management tool, the Field Actions Monitoring Device, we would have the analysis on Figure 3.

This tool was introduced in the second series of tests, trying to resolve the difficulties observed. More specifically, its use in extreme situations allowed us to observe that the tool ensured that the field actions to be carried out were managed correctly. With minimal preparation in this new tool, the team was able to implement organization ensuring effective management of field actions, particularly important in an ES. Furthermore, this tool facilitates the prioritization of field actions and the monitoring of field operators and their optimization in a situation requiring many field actions. Therefore, the Field Actions Monitoring Device was a favorable factor in controlling the state of the reactor in a degraded situation.

4 GENERAL DISCUSSION

The results of our analyses show that the design basis of the crisis management organization in Extreme Situation was not called into question. It also appears that some observed difficulties require more consideration in the preparation of the operating teams in order to strengthen resilience.

4.1 *Management of field actions*

The main observation relates to the management of field actions and of the "field operator" resources. In a design basis accident, if electrical power sources are lost, the necessary number of safety-related equipment items for the facility are backed up by Emergency Diesel Generators and can still be controlled remotely in the control room. Field actions then involve trying to return the systems to service or checking the shutdown of the stopped systems, and therefore these actions are not generally essential for the operation of the reactor. In an Extreme Situation with a beyond-design-basis

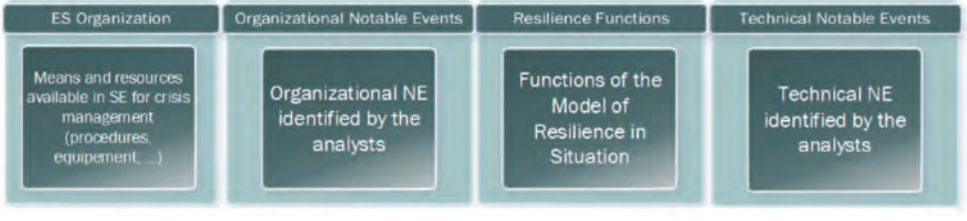


Figure 2. Functional analysis of resilience method.

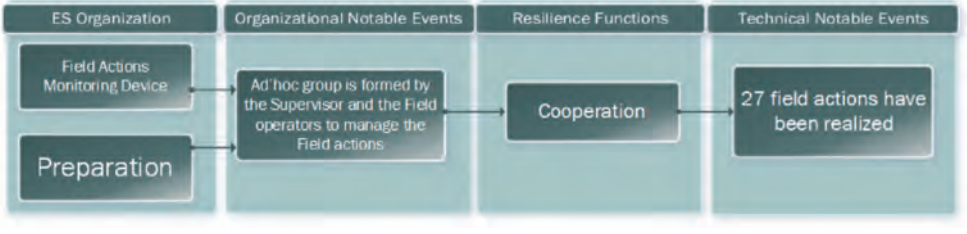


Figure 3. Example of functional analysis of resilience method: Field actions monitoring device.

loss of electrical power sources, these backups may be lost and the operation of the installation may require direct field action to control certain equipment, try to return it to service or take information. The field operators are in fact in high demand. They are sent in the field on a case-by-case basis depending on each action demanded by the team in the control room, with one or more field action sheets, and their actions are not carried out instantly since they need to travel to the premises before executing their sheet, and the action itself must often be executed manually. Each sheet therefore requires human resources, one or more field operators will need time to perform the action demanded.

Prioritizing and re-prioritizing the field action sheets becomes a condition for the success of the crisis management in the control room. This activity has considerable cognitive requirements for the person responsible of it, and an impact on the collective operation of the team. During our tests, this activity takes up a lot of the supervisor's time, therefore leaving them less available for the team and for their own supervision tasks.

To help to manage the field actions and the field operators, a prototype of a Field Actions Monitoring Device has been designed as part of an agile process (De la Garza et al., 2016). This is a triptych panel providing an overview of the actions pending, allowing them to be prioritized and then to see those which are in progress and who is carrying them out, and finally the successes and failures. This system is a support for ES operation, making it easier to monitor and ensure the safety of the field operators who leave to carry out actions. It also allows information to be shared within the operating team and becomes an area for exchanges, or even collective problem-solving.

4.2 *Reflections on the ES simulation system*

The organization and analysis of an Extreme Situation test is costly in terms of time and resources: 60 to 80 people from different entities (operators, engineering, trainers, R&D, etc.) working on the organization and progress of the tests requiring 4 to 6 months of preparation, then 4 to 6 months of analysis… We have adopted a procedure of continuous improvement in the organization of the tests, progress of which has been spread over more than three years, getting the various stakeholders in the preparation involved earlier in the process, and optimizing the analysis method.

However, we can see that there remain questions pending and areas for improvement to be followed in the organization of simulations on such a large scale, and in particular based around four points discussed above relating to the preparation, creation of scenarios and control of certain variables.

4.2.1 *Unpredictability of the scenario*

The scenario we designed for the tests was a succession of equipment failures (loss of off Site Power, Loss of on-site Power…) similar to the damages that the Fukushima Daiichi NPP faced during the crisis. In order to successfully manage the accident, the operating teams have to apply rarely used procedures.

Before the final test, we had to validate the technical aspects of de simulation, like the behavior of the Full Scale Simulator in this situation where it had to cope with a lot of equipment failures, or how well the procedures matched with the situation.

The problem is that, because of the autonomy we gave to the operating teams during the simulations, if we want to validate the adequacy of the procedures, we have to make sure that they are correctly applied by the operating team: operators might choose another procedure to apply, if they consider it more appropriate. Instructors however are trained to strictly apply the procedures, so we preferred having them for the technical validation of the scenarios.

Even though the technical validation is necessary, it does not protect from problems during the test. But since their objective was to study the resilience of the teams, we left them a lot of autonomy in the operations, and we would not have stopped the simulation should they have applied a procedure we had not expected.

4.2.2 *Simulating equipment which is scheduled but not yet installed*

How can the players be prepared to manage a simulation which will require equipment and an organization which are not yet scheduled as part of their training, and at the same time make the situation as close as possible to the target? Here we have a modification of an existing situation.

To train the teams, we offered them preparatory information meetings, during which the objective of the tests, the scheduled progress for the day and the new systems scheduled for crisis management were presented, without giving them any indication of the scenario which would be played out.

Research is in progress at EDF on rapid means of prototyping control room interfaces, which can be combined with the simulator design codes. The equipment scheduled as part of the post-Fukushima project could therefore be integrated into the simulators during extreme situation tests.

4.2.3 *Getting the players involved*

How can the players get involved in the simulation of a faulted condition, in a highly-degraded environment?

It is of course impossible to recreate in the control room or in the field the degraded condi-

tions encountered by the operators during the Fukushima accident. However, using field operators, we regularly provided the control room with information on the degraded state of the facilities. For example, on returning from a simulated field inspection, the field operator drew up for the supervisor the list of inaccessible premises, damaged equipment, etc.

We have seen tests in which the teams attached great importance to the safety of the participants in the field, avoiding sending them for field actions which are irrelevant given the situation, or sending them in pairs. These observations were interpreted as a good understanding of the situation in hand.

It has however not always been clear in the mind of the teams that the site was isolated, to the extent that some were waiting for the arrival of the on call teams in order to launch important actions. The question arises of how to simulate the consequences of the external hazard on the environment of the plant which have a major impact for the operators and their actions.

4.2.4 *Importance of multi-reactor accident simulations*

Our studies into the Fukushima accident highlighted the interactions which took place between the different damaged units on the site (Baudard, 2017): immobilization of resources, focus on one reactor at the expense of the others, transfer of experience, etc.

Thanks to the autonomy we left to the teams, they were able to perform out of the procedures actions, and we identified transfer of experience mechanisms during the simulations. For example, one operating team benefited from the experience of the team from the neighboring plant unit in restoring the power supply using the backup means scheduled in the post-Fukushima provisions.

Our work on the ES tests highlighted favorable factors allowing these beyond-design-basis crisis situations to be managed, but there are also factors which will require progress. Since these simulation of multi-units accidents are relatively recent, we think that it would be helpful, for understanding and improving resilience, to develop these simulation situations, but also simplifying them.

4.3 *Our proposition for future preparations*

Indeed, even though the organization, observation and analysis of the tests have been highly beneficial for the development of knowledge on Extreme Situations management by the parties involved within the company, these tests are too costly and too time consuming.

We have already started investigating lighter simulation methods for the teams involved in crisis management (through Serious Games or Storytelling for example), focusing on one specific group rather than the whole organization, and simulating its environment. A prototype has been tested with the National Technical Support Team (Alengry et al, 2018).

REFERENCES

Alengry, J. et al 2018. What is "training to cope with crisis management situation"? A proposal of a reflexive training device for the National Technical Support Team. *International Ergonomics Association*.

Baudard, Q. & Le Bot, P. 2017. Modelling Human Operations during a nuclear accident: The Fukushima Daiichi accident in light of the MONACOS Method. In Marko Cepin, Radim Bris (ed), *Safety and Reliability Theory and Application, ESREL Proceedings 17–21 June 2017*. CRC Press, Balkema.

De La Garza et al, 2016. D'un « document » à un « dispositif » de suivi des actions en local dans le nucléaire. *Ergo'IA 2016*.

Le Bot, P. & Pesme, H. 2010. The Model of Resilience in Situation (MRS) as an Idealistic Organization of At-risks Systems to be Ultrasafe. *PSAM10–10th International Conference on Probabilistic Safety Assessment & Management*.

Reynaud, J.D., 1997. Les Règles du jeu: L'action collective et la régulation sociale, *Armand Colin, Paris, 1997.*

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A stochastic-based evacuation model for risk assessment in road tunnel fire accidents and the importance of educating users

P. Ntzeremes
School of Mechanical Engineering, National Technical University of Athens, Greece

K. Kirytopoulos
School of Natural and Built Environments, University of South Australia, Adelaide, Australia

ABSTRACT: Road tunnel management is oriented in protecting tunnels from fire accidents that can evolve to disasters causing several human losses and extended infrastructure destruction. In order to enhance tunnels' safety, analysts usually employ Quantitative Risk Assessment (QRA) models for predicting human losses. However, human behaviour includes significant uncertainties during the evacuation process. This paper aims to address this problem proposing a stochastic-based evacuation model. Employing Monte Carlo approach, a set of simulations is conducted in which users' behaviour modelling is based on data from the existing theoretical framework, post-accident reports, conducted experiments and legislation requirements. Applying the model to a typical Greek tunnel, the outcome highlights a significant proportion of scenarios that exceed losses estimated by traditional approaches revealing also potential fallacies. The model's contribution rests on the provision of a stochastic-based simulation that is closer to describing reality than a simple deterministic model, as far as users' evacuation is concerned.

1 INTRODUCTION

Road tunnels are regarded as a key element of the road network as they connect remote regions since they create short-cuts in mountainous ranges. In addition, the growing concentration of population in large urban areas has inevitably lead to the extended use of underground road tunnels in order to relief the increasing traffic volumes (PIARC, 2016).

Due to the aforementioned implementations, road tunnels are considered critical infrastructures. A critical infrastructure can be defined as "… facilities of key importance to public interest whose failure or impairment could result in detrimental supply shortages, substantial disturbance to public order or similar dramatic impact" (Gheorghe, et al., 2006, p. 6). This criticality was confirmed by the adverse effects in terms of human losses and infrastructure destruction from various past accidents. For instance, the Mont Blanc tunnel accident in 1999, the worst accident in Europe, cost the life of 39 people and around 350 million Euros for the repair (AADT, 1999). Likewise, the Sasago tunnel accident in 2012, the worst tunnel accident in Japanese history, where the tunnel ceiling collapsed causing fire and cost the life of 9 people. Meanwhile, the road network of the region remained closed for almost two months (Maskura, et al., 2015). To this respect, road tunnels, like any complex socio-technical and critical system, must be designed to ensure an acceptable level of safety (Kirytopoulos & Kazaras, 2011). For tunnels specifically this would mean protection against fire accidents considering that can evolve to disasters.

In order to enhance tunnels' level of safety, safety analysts usually apply the risk assessment process based on various Quantitative Risk Assessment (QRA) models (Kirytopoulos, et al., 2010). The ultimate goal of a QRA model is to estimate the tunnel level of safety in case of a fire accident focusing primarily on predicting human losses amongst trapped-users. With this respect, QRA models often examine a subset of some crucial fire scenarios in a deterministic approach based on specific guidelines and regulations that each country has adopted (PIARC, 2012). However, post-accident reports, full-scale and virtual experiments towards the aspect of human behaviour under emergency situations have illustrated significant uncertainties regarding the self-evacuation process of the trapped-users. Despite that, national guidelines as well as QRA models do not take into account this information during the analysis in sufficient detail.

This paper aims to address this problem by proposing a model with a stochastic-based approach. The proposed model employs Monte Carlo

approach for conducting a set of simulations in which users' behaviour during the discrete stages of their self-evacuation process is based on probabilistic data. The probabilistic data used in the model were sourced from various findings from existing theories, post-accident reports, conducted full-scale as well as virtual experiments and legislation requirements. The outcome provides safety analysts with a prediction of the distribution of both successful and non-successful self-evacuations in order for them to estimate the overall tunnel safety and select additional measures, if required. Furthermore, by surfacing the impact of self-evacuation to the evolvement of the accident, it points out the importance that authorities and tunnel managers should give on educating tunnel users in order to react appropriate in such accidents.

2 EVACUATION PROCESS IN TUNNEL FIRE ACCIDENTS

2.1 *General principles of the evacuation models*

The safety ground of the road tunnel system encompasses all the crucial elements of the system and classifies the latter into five basic categories, namely: (a) the facilities, (b) the infrastructure, (c) the traffic, (d) the users and (d) the vehicles (Kirytopoulos, et al., 2017). The risk assessment methods deal with these categories that combine to form the tunnel level of safety as high as reasonable possible.

Users' losses is the representative parameter that provides the preparedness of the tunnel system in confronting fire accidents. Thus, each of the risk assessment methods focuses primarily on estimating the possible losses amongst trapped-users and is concerned about how to reduce them (PIARC, 2012).

Indeed, tunnel users consist the most vulnerable factor of the system. They are the first who confront with the fire consequences in a tunnel and in most of the cases without being adequately experienced in such circumstances. Moreover, they do not have appropriate equipment with them and often they do not have the education of other groups, like the members of the rescue teams, on how to react in critical situations (Kirytopoulos, et al., 2017).

Furthermore, fire in tunnel has much different behaviour than fire in the open road. In particular, it has rapid development, remains in maximum Heat Release Rate (HRR_{max}) longer and releases much more fumes (Beard & Carvel, 2012; PIARC, 2017). As a result, trapped-users have to evacuate themselves in a strictly limited time interval, which does not allow for any delay in the beginning of the self-evacuation process for the anticipation of external rescue teams.

Servicing the principles of managing risks in which users' factor must be taking into account, users' behaviour during the evacuation is in the centre of attention in order to assess and enhance road tunnels' level of safety.

The majority of the evacuation models, like engineering hand calculations and computer tools, are used in order to calculate the time it takes for trapped users to evacuate the tunnel walking away from the fire environment and heading towards a safe place as the emergency exit doors or the tunnel portals (it depends from the fire location) as soon as possible. An important prerequisite in this direction is the establishment of two basic time parameters, the Available Safe Egress Time (ASET) and the Required Safe Egress Time (RSET). ASET is defined as the time which is actually available for trapped users from fire sparking and the time point at which conditions become inadequate for human life, because of the high rates of pollutant concentration and radiation. On the other hand, RSET refers to the time that trapped users actually need for a successful outcome of their self-evacuation. The aim of the safety analysts is to achieve the ASET to be greater than RSET (Kinateder, et al., 2015). To do so, they have to forecast two main things of the evacuation process: the actions that people take and the time it needs for these actions to be performed.

However, most of the risk assessment methods as well as evacuation models implement oversimplified assumptions about the different stages of evacuation process and the behaviour of trapped-users. Furthermore, they have to follow the country-specific regulatory requirements. These approaches might increase the uncertainty of evacuation models that subsequently lead to a considerable uncertainty of the whole safety approach. Hence, it is important to design models taking into account potential information or clues from both real accidents and studies about human behaviour in fire evacuation and adapting them to the models, accordingly. If so, the uncertainties regarding the overall level of tunnel safety could be diminished.

2.2 *Human behaviour in evacuation process*

2.2.1 *Theoretical framework*

Human behaviour in fire evacuation process depends on which phase of the process the user performs. The big picture of the evacuation process in tunnels can be separated in three main phases (Kuligowski, 2013).

Initially, the first phase is the ignition phase. This point has some perturbation as not all the

users perceive the fire location promptly and also at the same time. For this reason, there is possibility for another accident to occur, such as crashing, upstream of the fire location. The second phase is the pre-evacuation phase. In this phase the user receives, collects and processes the cues in order to decide his own self-evacuation strategy. The progress of this phase depends on both the knowledge of the user and user's risk perception, as well. These two phases constitute the no-moving period and the time that elapses can prove fatal for the users. Experiments showed that under certain circumstances even a couple of seconds delay can neutralise the user (Ntzeremes, et al., 2018). The last phase of the evacuation process is the moving period. In this phase, trapped-users begin to walk away from their positions inside the tunnel and heading towards to a safe location. In tunnels, depending on the fire location, emergency exit doors and/or the tunnel portals are considered as safe locations. However, post-accidents reports indicated that not all users were willing to abandon their cars hoping that rescue teams could extinguish the fire soon or underestimating the fire behaviour (AADT, 1999).

With regard to the aforementioned phases of the evacuation process, there are some common theories in order to interpret trapped-users behaviour (Fridolph, et al., 2013). The first theory is the behaviour sequence model, which separates the human behaviour in four distinct mental and physical actions. So, a user initially receives, after interprets, subsequently prepares and in the end acts. The next theory is the role—rule model. This model makes the assumption that every user in a critical situation would behave according to the set of rules steaming from his position. A subsequent theory is the affiliation model, which assumes that a person would head to places or follow people that are familiar to him. The last theory deals with the social influence. This approach considers that the presence of other people would affect the user's evacuation process.

2.2.2 *Evidence from real and virtual cases*

Placed next to theoretical framework, a valuable source on human behaviour in fire accidents is the evidences from real and virtual accidents, both for the no-moving period and also for the moving period.

Regarding the post-accident reports, in the most disastrous tunnel accident in history, the Mont Blanc accident, one of the most important observations is that a significant number of 27 users delayed enough the evacuation process or they did not even started remaining in their vehicles. On the contrary, in Burnley accident in Sydney in 2007, the quick reaction of both the users and the emergency response units reduced the RSET significantly reducing both the causalities and injuries (Fridolph, et al., 2013).

Furthermore, studies based on virtual accidents can also provide analysts with information about users' behaviour. For instance, some studies illustrate the important factor of social influence during evacuation (Kinateder, et al., 2014). Subsequently, some other studies explored the factor of risk perception of users and the relationship between fire and awareness of successful evacuation process (Ronchi, et al., 2015). Other studies explore users' behavioural intentions and knowledge in case of fire accidents. Outcomes were far away from reflecting the expected ones, as users often had totally underestimate the criticality of the accidents (Kirytopoulos, et al., 2017). The aforementioned information illustrates the high perturbation amongst the information on the way that trapped-users might behave during fire accidents. Thus, there is a lack of available data for use by evacuation models as far as the no-moving period of users' self-evacuation process.

However, similar perturbation and lack of data exists in the estimation of the moving period, too. One of the most important factors when assessing safety in tunnels is the users walking speed through smoke, which is the basic evacuation performance characteristic. Smoke caused by fires flows stratified in the tunnel tube and large space during the initial fire, and after the stratification of the smoke is disrupted by the heat absorption by the ceiling wall. Subsequently, it starts to diffuse inside the tunnel, making evacuation activities, rescue activities and fire extinguishing activities extremely difficult. Studies have shown that walking speed does not only steams from the physical ability of the user but also affected from the speed of information exchanged, the experience and the knowledge of the user as well as from the social influence (Ronchi, et al., 2015).

Furthermore, a number of full-scale experiments estimating walking speed have been conducted. In such experiments, researchers examine different group of participants in different smoke situations (irritant and non-irritant smoke). As a result, the outcome shows a strong interrelation between the extinction coefficient, which shows how easily the air can be penetrated by a beam of light, and the estimating walking speed in normal and emergency situations (Seike, et al., 2016).

2.2.3 *Country-specific regulations*

Last but not least, the country-specific regulations, usually determine the context in which a safety analyst should operate affecting the analysis. Often, in these regulations, the walking speed of the user in relation to opacity is provided as a deterministic

number, which is something that does not always reflect reality (PIARC, 2012).

3 METHODOLOGY

In order to estimate the tunnel level of safety, a simulation-based research is used. This approach can aid a better understanding of the phenomenon, the outcome of the self-evacuation process of trapped-users in a fire accident, having the advantage that includes and integrates available data from existing theories and experiments. The structure of the process is as follows:

Initially, the fire scenario along with the fire location are selected. Trapped-users' evacuation is illustrated as a trajectory, depicted as line segments depending on the walking speed, with regard to tunnel length and time. Fig. 1 provides an example of the evacuation process.

Road tunnels have generally simple and straight geometries in which the relative distance of the users from the fire source, the fire characteristics and the time spent inside the tunnel are the main factors affecting life safety. In order to estimate the impact of fire on trapped-users, firstly, the tunnel's airflows (the air temperature and the air opacity) should be estimated. The Computational Fluid Dynamic (CFD) process is performed with the aid of Camatt 2.0 software. The outcome provides two sources of data, the air temperature and the air opacity with regard to the tunnel length and time. Subsequently, these data are used for the estimation of the effect of the heat and pollution on the evacuation trajectories of the users.

Trapped-users safety are affected by accumulating heat of both radiation and convection. The estimation of the Fractional Effective Dose of heat (FED_h) in every time step and location is estimated by the following equation:

$$FED = \Sigma(1/t_{iconv.}+1/t_{irad.})*\Delta t \quad (1)$$

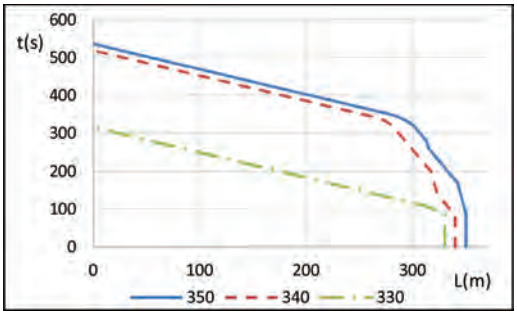


Figure 1. Evacuation process with fire location at 350 m.

where $t_{iconv.}$ is the time duration for convective heat, which is calculated as:

$$t_{iconv.} = (5*107)*T^{-3.4}\text{: for light clothing} \quad (2)$$

where {t} the time in minutes and {T} the temperature in °C, and where $t_{irad.}$ is the time duration for burning of skin by radiation, which is calculated as:

$$t_{irad.} = 4*q^{-1.36} \quad (3)$$

where {t} the time in minutes and {q} the radiant heat flux in kW/m^2 (for q > 2.5 kW/m^2; for q < 2.5 kW/m^2 this time is equal to 30 min).

Along with the effect of heat, trapped-users are also affected by accumulating toxicity by pollutants concentration. Carbon monoxide (CO) is the toxic gas on which toxicity is raised, based on the national standards. The estimation of the Fractional Effective Dose of CO (FED_{co}) in every time step and location is estimated as follows:

$$FED_{co} = \Sigma(0.00083*CO^{-1.036}*\Delta t)/D \quad (4)$$

where {Δt} is the exposure time in minutes, {CO} is the CO concentration in ppm and {D} is the concentration at incapacitation.

In order to estimate the FED of heat and pollution concentration further information is needed regarding the evacuation trajectory of the users. In the presented model, contrasting to the commonly used deterministic approach, the stochastic variables of the evacuation process are defined exhibiting the interrelation amongst current literature and national regulations.

Initially, a time that indicates the no-moving period, is selected. The no-moving period, consisting of the ignition and the pre-evacuation phases, is defined by the stochastic variable of how much time it takes before moving period begins. In Purser's study, this time regarding the Mont Blanc accident was specified between 40s and 90 s. Another study from Japan specified this time between 30 s and 120 s (Seike, et al., 2016). On the contrary, Greek regulations do not determine this time, but the most commonly time used in such analyses is approximately 90 s. Synthesising the existing data along with the Greek provisions, a uniform distribution with 60s as a lower limit and 120 s as an upper limit, is considered in order to be at the safe side.

In addition, the stochastic variable that characterises the moving period is the walking speed of the users. With a view to represent as closer to reality, the process employs existing studies results towards the connection between air opacity and walking speed (Seike, et al., 2016; Ntzeremes, et al., 2018). Synthesising these data with the

Table 1. Users' walking speed.

Speed (m/s)			
Distribution	Normal	Normal	Normal
Mean	1	0.5	0.2
St. Div.	0.067	0.067	0.05
Opacity (m^{-1})	(0, 0.50)	[0.50, 0.70]	[0.70, …)

Greek provisions, three moving intervals are defined (Table 1).

At the beginning, the process estimates human losses using the deterministic values of the Greek provisions (AAT, 2011). These values are the mean points of the aforementioned distributions. As a result, the outcome that safety analysts would estimate according to the current deterministic approach, is provided.

After building the model, the simulation is executed, with the aid of Matlab 2017b software, for 1000 iterations using the time distribution of the no-moving period and Table 1 elements for the moving period. This step forms the "60–120" scenario. Having created the distribution of losses, the results of both the deterministic and the stochastic approach are compared. Afterwards, another two repetitions of the process are conducted changing first the upper limit of the no-moving period, reducing it by 30 s creating a new uniform distribution between 60 s and 90 s ("60–90" scenario). Subsequently, reducing generally the delay time of trapped-users a second new uniform distribution between 30 s and 90 s ("30–90" scenario) is created. These two aforementioned scenarios are related to two user training scenarios, one with the goal to reduce the upper reactive limit and the other decreasing the upper and the lower one.

In the end, the outcome provides a prediction of the distribution of non-successful self-evacuations.

4 ILLUSTRATIVE CASE

4.1 *Case description*

A de-identified road tunnel in Greece fulfilling both the Greek and European safety requirements, is selected. The examined fire scenario, which is a standardised fire scenario of Greek regulations, involves a fire of HRR_{max} 100MW sparking from a HGV without carrying dangerous goods (PIARC, 2017; Ntzeremes, et al., 2018). The location of the fire, 350 m from the entrance of the tunnel, is part of the entry area, which regarded as the most common and one of the most vulnerable locations of the tunnel (PIARC, 2017, p. 12). Tunnel's attributes are shown in Table 2.

Table 2. Tunnel's attributes.

Designing features	One dimension—single sector—rural	
	Total length	3.000 m
	Slope	–1,5%
	Number of exit doors	6
	Number of traffic interruptions	7
	Starting time of traffic lights after fire ignition	5 min
Mechanical ventilation	Number of jet-fan-array	8
	Starting time	2 min
Traffic conditions	Vehicle flux	55 veh./hr.
	Proportion of HGVs	30%

4.2 *Results*

In order to simulate trapped-users trajectories, their initial locations are required. To this respect, the analysis assumes uniform traffic condition and uniform distribution of the stopped-vehicles. So, at the time fire outbreaks, the traffic simulation results in 10 trapped-vehicles shaping one row with 10 m distance in-between. Hence, the model estimates the losses corresponding to the locations that the vehicles have stopped (e.g. Fig. 1), taking into account that each vehicle has two users.

On the one hand, the analysis following the traditional deterministic approach using the mean points of the aforementioned distributions estimates eight losses (there are four fatal trajectories corresponding to the fire location and the three consecutive locations upstream the fire). However, if a more conservative approach is applied, the outcome would be with less losses. For instance, if the time when the evacuation starts is determined at 60 s and the users' speed remain the same, only two losses (only the fire location's trajectory is fatal), are estimated. This outcome highlights how the expert judgment or an incorrect threshold can alter the estimated safety level.

On the other hand, the stochastic approach illustrates a much different outcome. Fig. 2 presents a wider distribution of the scenarios.

Fig. 2 shows that regarding the simulation of "60–120" scenarios set only 18.8% of the scenarios include eight losses. The outcome provides an approximately equal probability amongst the scenarios of four, six, eight, ten and twelve losses arising significant uncertainty. This outcome raises serious concerns about the safety of the tunnel. There is a significant probability of 41.2% scenarios to have ten or more losses (up to fourteen), although there is an equal probability of 40% of

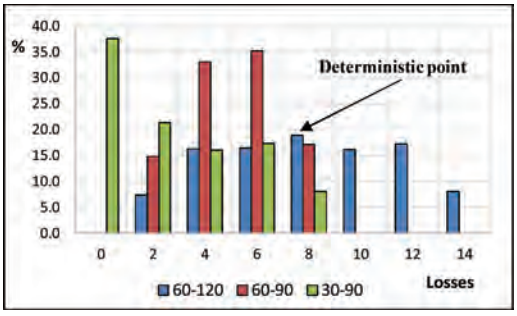


Figure 2. Distribution of evacuation scenarios.

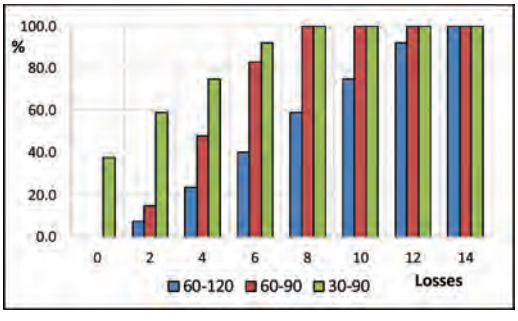


Figure 3. Cumulative function of evacuation scenarios.

scenarios to have less than eight losses, which cannot counterbalance the probability of very adverse consequences. The potential additional measures in case of fourteen losses require quite different additional measures than if there were only four, as this discrepancy raises in 30 m longer the lethal length of backlayering.

Furthermore, the "60–90" simulation shows that if the evacuation starts no longer than 90s, there is 83% probability of having favourable outcome with regard to the deterministic approach and 40% with the "60–120" scenario. In this outcome, the trapped-users corresponding to the fifth, the sixth and the seventh stopped-vehicle are considered totally safe. However, as in the "60–120" case, still all the scenarios include losses.

Only the last case ("30–90" scenario set) has 37.5% probability of including scenarios without losses. However, there is an equal probability of having up to six losses, which indicates that the need for taking additional measures is inevitable.

However, the results indicate also the importance of users' attitude in fire accidents. To this respect, reducing the upper limit of the time corresponding to the no-moving period, trapped-users before the first 40 m are successfully evacuating the tunnel. Nevertheless, the safety problem of the tunnel still remains. So, if the users are educated enough in order to start the evacuation process immediately, the outcome shows that approximately 40% of the scenarios have no losses without changing any other parameter of the system.

In order to illustrate the safety level of the tunnel the cumulative diagram of losses is employed. Based on this, safety level can be a strict threshold (e.g. 2 losses should be not exceed 20%) or can be estimated based on as low as reasonable possible (ALARP) principle.

In addition, the effect of users' education is also reflected in Fig. 3. Consequently, through the simulation one can observe the wider spread of losses in contrast to the deterministic approach. It is remarkable how a small reduction of only 30s in the range of the time of the beginning of the evacuation can benefit the overall safety level. Thus, a huge effort must be given towards the sensitisation of authorities and tunnel users towards safety.

5 CONCLUSIONS

The presented study illustrates that the current safety approach towards road tunnels, which constitute a critical infrastructure of the road network, by applying QRA models should focus further on the users' factor. The existing literature provides safety analysts with the theoretical amendments to interpret human behaviour in fire accidents. Furthermore, the cues from previous accidents as well as the full-scale experiments and virtual reality experiments enlighten better scientists in discovering crucial parameters of the evacuation process.

However, there is lack in data that would assist evacuation models to give outcomes with lower uncertainties. In the presented illustrative case, the use of the deterministic approach together with the country-specific regulatory requirements cannot estimate the tunnel's actual level of safety. The first gives four losses when the stochastic approach estimates 60% of the scenarios above this value. Nevertheless, the outcome provides safety analysts with a prediction of the distribution of non-successful self-evacuations in order to estimate the overall tunnel safety and select additional measures, if required. Hence, this approach is closer to describing reality than a simple deterministic model, as far as users' evacuation is concerned. In a nutshell, it aids risk analysts make better-informed decisions.

Furthermore, the change in the parameter of the beginning of the evacuation process shows the importance of educating tunnel users in confronting with these accidents. The outcome showed that a reduction of time by 30s could result in 40% of

the scenarios be free of losses, when in the basic case all scenarios have from two to fourteen losses.

REFERENCES

AADT, 1999. *Task Force for Technical Investigation of the March 1999 Fire in the Mont Blanc Vehicular Tunnel,* Paris: Ministry of Equipment, Transportation and Housing.

AAT, 2011. *Risk Analysis Method without considering vehicles carrying Dangerous Goods/Scenario-based approach,* Athens: Administrative Authority of Tunnels.

Beard, A. & Carvel, R., 2012. *Road Tunnel Fire Safety.* 2nd ed. London: Thomas Telford.

Fridolph, K., Nilsson, D. & Frantzich, H., 2013. Fire Evacuation in Underground Transportation Systems: A Review of Accidents and Empirical Research. *Fire Technology,* 49(2), pp. 451–475.

Gheorghe, A., Masera, M., Weijnen, M. & Vries, L., 2006. *Critical infrastructures at risk.* Dordrecht: Springer.

Kinateder, M., Kuligowski, E., Reneke, P. & Peacock, R., 2015. Risk Perception in fire evacuation behaviour revisited: definitions, related concepts and empirical evidence. *Fire Science Reviews,* Volume 4, pp. 1–26.

Kinateder, M., Ronchi, E., Gromer, D., Müller, M., Jost, M., Nehfischer, M., Mühlberger, A. & Pauli, P., 2014. Social influence on route choice in a virtual reality tunnel fire. *Transportation Research Part F,* 26(6), pp. 116–125.

Kirytopoulos, K. & Kazaras, K., 2011. *The need for a new approach to road tunnels risk analysis'.* Troyes, CRC Press.

Kirytopoulos, K., Kazaras, K., Papapavlou, P., Ntzeremes, P. & Tatsiopoulos, I., 2017. Exploring driving habits and safety critical behavioural intentions among road tunnel users: A questionnaire survey in Greece. *Tunnelling and Underground Space Technology,* 63(3), pp. 244–251.

Kirytopoulos, K., Rentizelas, A., Kazaras, K. & Tatsiopoulos, I., 2010. *Quantitative operational risk analysis for dangerous goods transportation through cut and cover road tunnels.* Rhodes, Taylor & Francis.

Kuligowski, E., 2013. Predective Human Behaviour Towards Fires. *Fire Technology,* 49(1), pp. 101–120.

Maskura, S., Takagi, R., Kobayashi, A., Chihara, T. & Takahashi, K., 2015. *Review on the configuration of outlet duct fair air supply in a tunnel—Aiming for safety improvement.* Seattle, BHR Group Limited, pp. 37–48.

Ntzeremes, P., Kirytopoulos, K. & Benekos, I., 2018. Exploring the effect of national policies on the safety level of tunnels that belong to the trans-European road network: a comparative analysis. *International Journal of Critical Infrastructures,* in press.

PIARC, 2012. *Current Practice for Risk Evaluation for Road Tunnels,* Paris: World Road Assosiation, Technical Committe 3.3, Road Tunnel Operation, ISBN 978-2-84060-290-3.

PIARC, 2016. *Road tunnels: Complex Underground Road Networks,* Paris: World Road Association, Technical Committe 3.3, Road Tunnel Operation, ISBN 978-2-84060-404-4.

PIARC, 2017. *Design Fire Characterisics for Road Tunnels,* Paris: World Road Association, ISDN 978-2-84060-471-6.

Ronchi, E. et al., 2015. Evacuation travel paths in virtual reality experiments for tunnel safety analysis. *Fire Safety Journal,* Volume 71, pp. 257–267.

Seike, M., Kawabata, N. & Hasegawa, M., 2016. Experiments of evacuation speed in smoke-filled tunnel. *Tunnelling and Underground Space Technology,* 53(1), pp. 61–67.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Incremental fatigue damage simulation for reliability assessment of steel wire ropes under fretting fatigue conditions

S. Ahmad & S. Badshah
Department of Mechanical Engineering, International Islamic University, Islamabad, Pakistan

M.F. Abdulhamid, H.S. Kang, A.S. Kader & M.N. Tamin
Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

ABSTRACT: This paper describes the newly-developed damage-based fatigue life model for the long-term reliability assessment of the drawn steel wires and the wire ropes. The methodology is based on the computed local stress field in the critical trellis contact zone of a stranded wire rope by the FE simulation and the degradation of the Young's modulus of the drawn steel wires. The fatigue damage model is based on Lemitaire's damage equations for quasi-brittle material with a damageable microinclusion embedded in an elastic mesoelement. The incremental fatigue damage calculations employing the load-cycle block is described. The routine is integrated into commercially available Finite Element Analysis (FEA) software. A case study using a single strand (1 × 7) steel wire rope with 5.43 mm-dia. drawn wires is employed to illustrate the damage-based fatigue life prediction procedures. The simulated tensile fatigue cycles consist of the load range of 145 kN and load ratio, $R = 0.1$. The peak applied load corresponds to 50% of the maximum breaking load of the steel wire rope. The FE-calculated results indicate that the von Mises stress, maximum principal stress and the contact pressure cycle in-phase, and with an identical stress ratio as the applied axial load. The trellis contact point is relatively small and experiences elastic stresses, thus the fatigue damage prediction for the fretting fatigue condition is appropriate. The damage initiation life at the trellis contact along the core wire is calculated at $N_o = 1050$ cycles. An additional and improved data set of the damage model parameters of the drawn steel wires is required to achieve an accurate and validated life prediction model of the wire ropes.

1 INTRODUCTION

Steel wire ropes have found numerous applications in marine and civil construction including the mooring system for FPSO in the deep sea, bridge suspension elements for suspended and stayed structures, and lifting cranes. In these applications, the wire ropes are subjected to complex loading, often involving tension-tension fatigue induced by both the lifted weight and self-weight of the component, bending over the sheave fatigue, free bending fatigue due to transverse load such as ocean current, and torsion fatigue. The sea operating environment further amplifies the reliability issue through stress corrosion cracking of the wire rope elements. The rated or operating load for the wire rope is often a fraction of the Maximum Breaking Load (MBL) of the wire rope and induces stresses in the elastic range. However, the combination of the relatively high stress amplitude and the positive load ratio, particularly at the trellis contact region, could initiate damage and nucleate fatigue crack at the locality. The subsequent crack propagation leads to fracture of the critical drawn wire, redistribution of stresses in the fracture locality and failure of the neighboring drawn wires thus leading to premature fatigue failure of the wire rope. Considering the associated high cost of installation, and if needed, the very costly and difficult to replace wire rope (Raoof & Davies 2008), an efficient and accurate predictive tool is required for reliability assessment of the wire ropes over their typical design life of 20 years.

The common causes of fatigue failure of the wire rope are due to the contact pressure and slip (Waterhouse 2003). The relatively high contact pressure developed at the trellis contact between the layers of the strand wires in a small contact region. Consequently, this induces partial slip with small displacement amplitude causing the fretting fatigue failure. In a wire rope design having a large lay angle, a line contact is established between the wires in the strand causing a large slip contact region. The relatively large amplitude of

displacement results in gross slip and failure occurs by fretting wear of the wire rope. In both cases, the development of a failure model should account for both the mechanics of deformation and mechanism of failure of the wire rope. The work describe in this paper is limited to wire ropes failing under the fretting fatigue mode.

Fatigue life prediction of the wire rope is a challenging task because the wire rope is a system consisting of many drawn wires arranged in different windings and simultaneously enduring the applied load. Design against failure due to the tensile load that exceeds the breaking strength of the wire rope has been established (Prawoto & Mazlan 2012). Fatigue testing of the wire rope samples to generate the fatigue-life curve seems a straight-forward way of life prediction. Most of the available fatigue-life data for the wire ropes are generated from tests at zero-depth (in-air environment). In this respect, extensive axial fatigue-life test data on sheathed steel spiral strands have been established (Raoof & Davies 2008). The work also quantifies the effects of external hydrostatic pressure, representing the deep-water loading condition, on the fatigue performance of the sheathed wire ropes. Fatigue design of the wire ropes that considers both the high-cycle fatigue of the stranded wires and the inherent fretting effects have been examined (Winkler et al. 2015, Wang et al. 2013, Cruzado et al. 2012, Sasaki et al. 2007). The fatigue-life data for the wire ropes are based on data taken from available standards such as DNV OS-E301 (2004) and open literature (Raoof & Davies 2008, Alani & Raoof 1997, Birkenmaier 1980, Suh & Chang 2000). However, the reported fatigue strength of these wire ropes is relatively low at $(S_{lim} / S_U) = 0.15–0.3$ (Kao & Byrne 1982). The continuum-based high cycle fatigue life model such as the modified Goodman and Soderberg approach could only predict the terminal life of the most critical wire. While fatigue failure of the wire rope is often defined by the area fraction of the fractured wires. Reliability model based on the local stress concentration due to the erosion of the wire material at the contact locations has also been proposed. However, the evolution of the local geometry of the area is difficult to establish.

This paper describes the framework for the development of a damage-based fatigue life model for the steel wire ropes. The methodology for damage and failure assessment of the wire rope materials under the high-cycle fretting fatigue condition is discussed. An incremental calculation routine based on cyclic degradation of the elastic modulus of the drawn wires is deliberated. A preliminary simulation for the case of a single strand (1 × 7) steel wire rope under tensile fatigue loading is demonstrated.

2 FRAMEWORK FOR THE DAMAGE-BASED FATIGUE LIFE PREDICTION

The reliability assessment of the steel wire ropes is aimed at the prediction of the fatigue life of the wire rope when subjected to the general loading conditions. Figure 1 illustrates the framework for the development of a validated damage-based fatigue life prediction model for the wire ropes. The fatigue life of the wire rope is governed by many factors including the design and material of the drawn wires, fabrication process, operating load conditions and environment. The wire rope design also indicates if fretting fatigue at trellis contact or fretting wear causing gross slip is dominating the failure scene. The fretting fatigue failure is being addressed in this paper. Fatigue tests of the wire rope establish the corresponding fatigue life, N_f, represented by *Task (A)*. *Task (B)* constitutes of the tension tests of the drawn steel wires to establish the mechanical properties of the material. Fatigue tests of the drawn wires are performed to establish the strength-life and fatigue limit of the wire material. Damage parameters are also extracted from the resulting S-N curve. The damage-based fatigue life model of the drawn wire material is developed in *Task (C)*. The properties and damage model parameter values obtained in *Task (B)* are employed. The damage model is then incorporated in the finite element analysis (FEA) through user subroutine *UMAT* for the *Abaqus FEA* software employed in this study, in *Task (D)*. The FE-computed fatigue life is compared with the measured life from *Task (A)* to establish the validity of the fatigue life prediction model.

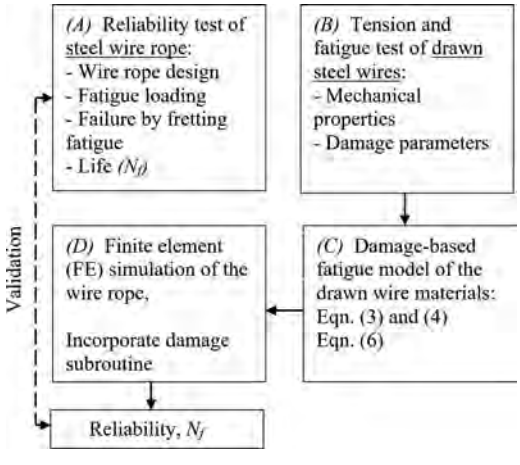


Figure 1. Framework for the damage-based fatigue life prediction of the steel wire ropes.

2.1 *Fatigue damage model*

Damage in a material can originate from the debonding of atoms or the nucleation, growth and coalescence of microcavities or microcracks. It leads to the gradual degradation of stiffness and strength of the material. Although the damage process is discontinuous in nature, it can be modelled by a continuous variable at the mircoscale (Lemaitre 1985, Kachanov 1958). In a representative volume element (RVE) of the material, the damage parameter is defined by scaling the damaged area, A_D by the total cross-sectional area, A of the RVE such that:

$$D = \frac{A_D}{A} \tag{1}$$

The damage parameter is continuous and represents the failure of microdefects over the mesoscale volume element. The value of the scalar damage variable, D is bounded by $0 \le D \le 1$, where $D = 0$ represents the undamaged state while rupture (or separation of the material point) occurs at the value of $D = 1$. Following damage initiation, the effective stress tensor, $\{\tilde{\sigma}\}$ can be represented by:

$$\hat{\sigma}_{ij} = \frac{\sigma_{ij}}{1 - D} \tag{2}$$

where σ_{ij} is the Cauchy stress tensor.

The evolution of this damage variable that depends on the expected value of the micro-defect density can then represent the history of the inelastic strain in the material. In the high-cycle fatigue where the amplitude of the loading is low, the amplitude of the plastic strain is relatively small or negligible at the mesoscale when compared to the elastic strain amplitude. For the drawn steel wires that are assumed to behave in a quasi-brittle manner, the behavior is brittle at the mesoscale but localized damage growth occurs at the microscale. Thus, the material is modeled as a damageable microinclusion embedded in an elastic mesoelement, as illustrated in Figure 2. While the mesoscale matrix is elastic, the microscale inclusion experiences elastic, plastic and damage. In addition, it is assumed that the inclusion is subjected to the same strain state (or strain rate) as the mesoscale matrix.

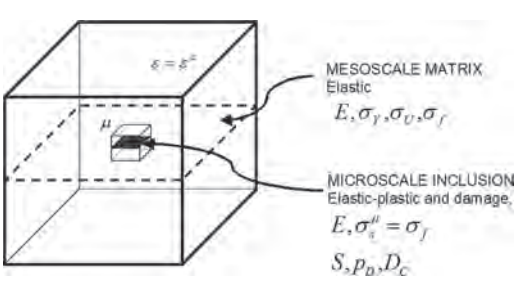


Figure 2. Representative Volume Element (RVE) for the two-scale model (Adapted from Lemaitre (1985)).

The prediction of the rupture or the separation of the material point requires the coupling between the plastic flow and damage at the material constitutive level. Detailed derivation of the damage equation is found in the work of Lemaitre (1996) and employed by the authors (Salleh et al. 2017). Damage initiation occurs when the equivalent strain reaches the damage strain threshold, p_D defined as:

$$p_D = \frac{\sigma_u - \sigma_f}{\sigma_f - \dfrac{\sigma_f^2}{\sigma_Y}} \tag{3}$$

where σ_Y, σ_u and σ_f is the yield stress, tensile strength and fatigue limit of the material, respectively. The number of cycles, N_o needed for the equivalent strain to reach the damage threshold, p_D is given by:

$$N_o = \frac{p_D}{2\Delta\varepsilon} = \frac{E\,p_D}{2\Delta\sigma} = \frac{E\,p_D}{4\left(\sigma_M - \bar{\sigma}\right)} \tag{4}$$

where E is the elastic modulus of the material, σ_M denotes the maximum stress and $\bar{\sigma}$ is the cyclic mean stress in the cycle.

Based on the kinetic damage law for the inclusion, and considering the relatively large plastic strain compared to the elastic component of a micro-volume, μ the evolution of damage is expressed as:

$$\dot{D} = \frac{Y^{\mu}}{S}\dot{p}^{\mu} \tag{5}$$

The superscript, μ refers to quantities for the microelement and S is a material damage parameter. It is desirable to express the damage strain energy release rate, Y^{μ} and the accumulated plastic strain rate, $\dot{p}^{\mu}$ terms as functions of the macroscopic quantities such as strain and stress. The plastic strain rate, $\dot{p}^{\mu}$ of the elastic-perfectly plastic microvolume is equal to the equivalent strain rate of the elastic mesoscale matrix.

We consider pure elasticity at the mesoscale and further approximate the damage rate, $\dot{D} = 0$ if the equivalent stress, $\sigma_{eq} < \sigma_f$, the fatigue limit of the material. In addition, an increase in the mean stress causes a decrease in the corresponding stress amplitude needed to induce fatigue failure at a certain number of cycles. Under these conditions, eqn. (5) can be conveniently written in incremental form as (Salleh et al. 2017):

$$\Delta D = \frac{2\sigma_f^2}{E^2 S}\left[\begin{array}{l}\frac{2}{3}(1+\nu)\left(\sigma_M - \sigma_f\right) \\ +\frac{1-2\nu}{9\sigma_f^2}\left(\sigma_M^3 - \sigma_f^3\right)\end{array}\right](\Delta N) \quad (6)$$

The term, ΔD represents the increment of the fatigue damage accumulated over the elapsed load cycles, ΔN while E and ν is the elastic modulus and Poisson's ratio of the material, respectively. It should be acknowledged that the Young's modulus of the material is likely to degrade with the accumulated load cycles. In addition, the rate of the stiffness degradation is dependent on the magnitude of the acting stress cycles and the stress ratio, as described in the next section. The separation or rupture of the critical material point would occur when the accumulated fatigue damage variable reaches a critical level (denoted by D_C in Table 1).

2.2 *Incremental fatigue damage calculations routine*

Fatigue damage is estimated based on the gradual degradation of strength and modulus of the drawn steel wire over the applied load cycles. Each simulated load cycle represents a block of specified number of cycles on the wire rope. During this load increment, the decrease in the modulus of the drawn steel wire is prescribed as observed experimentally. The calculations of the fatigue damage that occur at a material point during the given increment of the load cycles, ΔN is described by the process flow as illustrated in Figure 3. This constitutes a user-defined *SUBROUTINE UMAT* for the commercially available *Abaqus FEA* software employed in this work. The current stress tensor, σ_{ij} fatigue damage level, D the accumulated fatigue cycles, N residual stiffness, $E(N)$ properties of the drawn steel wire (S, ν, σ_y, σ_u, σ_f) and the load-cycle block size, ΔN are transferred from the main FE program.

STEP2 calculates the simulated load cycle parameter, namely the maximum, σ_M and mean stress, $\bar{\sigma}$. The number of load cycles, N_o required to initiate fatigue damage at this material point is estimated. The calculations proceed to the subsequent steps if the material point experiences damage initiation.

Table 1. Properties and damage model parameters for the drawn steel wires.

Parameter	Symbol	Value
Elastic modulus	E_o (GPa)	202
Yield strength	σ_y(MPa)	1690
Tensile strength	σ_u(MPa)	2164
Poisson's ratio	ν	0.28
Fatigue limit	σ_f(MPa)	305
Damage parameter	S(MPa)	6
Critical damage	D_C	0.8

SUBROUTINE UMAT

STEP1 Transferred from main program:
σ_{ij}, D, N, $E(N)$, ΔN
S, ν, σ_Y, σ_u, σ_f

STEP2 Increment $N = N+\Delta N$
Determine σ_M, $\bar{\sigma}$
Calculate N_o (eqn. (4))
IF $N \geq N_o$, calculate the damage
ELSE RETURN

STEP3 Determine (N_f, σ_f) from S_{max}-N curve (as corrected using Miner's rule)

STEP4 Determine residual $E(N)$ corresponding to N cycles (eqn. (7))

STEP5 Compute ΔD (eq. (6))

STEP6 Update for N:
$D = D+\Delta D$
$\sigma_{ij} = \frac{\sigma_{ij}}{1-D}$

RETURN

Figure 3. Flow chart of the incremental damage calculations over a load-cycle block.

In *STEP3,* the fatigue strength-life curve (see Figure 5) is corrected since the material point has been overstressed for finite number of cycles [Miner 1945]. The number of cycles to failure, N_f corresponding to σ_M is established. It is used in eqn. (7) to establish the residual elastic modulus, $E(N)$ of the drawn steel wire at the end of the load-cycle block in *STEP4.* Next, the increment of the damage parameter, ΔD over the increment of the load cycles is established using eqn. (6).

Finally, in *STEP6* the damage variable and the stress tensor are updated to correspond to the end the load-cycle block. These updated variables are returned to the main program for the next load-cycle block.

3 ILLUSTRATIVE CASE STUDY

A case study employing a single strand (1 × 7) steel wire rope construction and subjected to axial

fatigue loading is considered to illustrate the characteristic fatigue damage evolution and the life prediction of the wire rope. Although the 1 × 7 wire rope is commonly used as stayed cable, the validated fatigue life model could be used for other wire rope designs and loading conditions. The chemical composition of the drawn wires is (in wt.%) 0.83*C*, 0.91*Si*, 0.717*Mn*, 0.0124*P*, 0.0031*S*, 0.015*Cu*, the remaining being *Fe*. The observed preferred orientation of the grains along the drawing direction provides the superior strength of the wires in the axial direction. Since the damage model is developed for fretting fatigue failure, this failure mechanism should be reproduced by the steel wire rope under study.

3.1 *Finite element modeling*

The geometry of the wire rope model is illustrated in Figure 4, along with the design parameters. The core wire is straight. The total length of 330 mm of the wire rope model considered accounts for a full pitch length of 230 mm in the central length region, while the remaining lengths at both end regions are used to apply the load and boundary conditions. One end of the wire rope is fixed in both translational $(U_x = U_y = U_z = 0)$ and rotational $(UR_x = UR_y = UR_z = 0)$ displacements, while the other free end is subjected to an applied axial load, but without rotation $(UR_y = 0)$. An axial fatigue load consisting of the load range, ΔP = 145 kN and load ratio of minimum-to-maximum load, $R = 0.1$ is applied. The peak load cycle is at 50% of the maximum breaking load (i.e. 50% MBL) of the wire rope. The contact condition of the wires is prescribed to follow Coulomb's law with the assumed coefficient of friction, *cof* = 0.5 considering the surface roughness of the drawn wires. Based on the outcome of the mesh convergence study, the single strand (1 × 7) wire rope geometry is discretized into 203208 8-node continuum elements. *Abaqus FEA* software with user-defined *SUBROUTINE UMAT* is employed for the analysis.

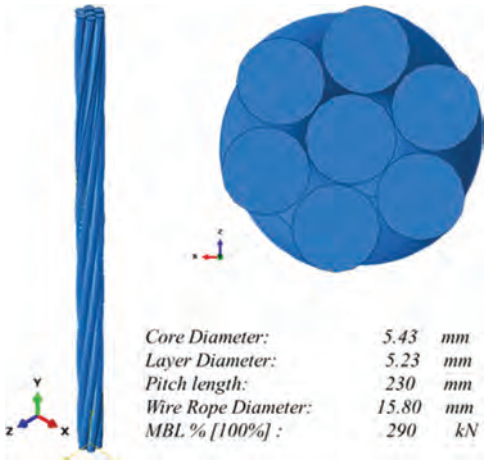


Figure 4. Geometry and the geometrical properties of the single strand (1 × 7) wire rope.

3.2 *Mechanical testing of the drawn steel wires*

The tension test is performed on the drawn steel wire specimens in the as-fabricated condition. The gage length is 100 mm. The resulting mechanical properties are listed in Table 1. The true fracture strength of the drawn steel wires, σ_R = 2.164 GPa with the corresponding true plastic strain at fracture, ε_R = 6.78% are established.

Fatigue life tests were performed on the drawn steel wires in the as-received condition at the load ratio of minimum-to-maximum load, $R = 0.1$ and loading frequency of 30 Hz. The 5.43 mm-diameter drawn wires are tested in the as-fabricated condition. The resulting fatigue strength-life (S-N) curve is shown in Figure 5 (circle symbols). The solid line compares the published fatigue life of 0.9 mm eutectoid steel wires. Effect of size (wire diameter) on fatigue life is consistent in that the larger diameter wires display a relatively shorter fatigue life. This is postulated by the greater number of manufacturing defects or microcracks inherent in a larger volume of the material within the gage section. The fatigue strength of the drawn steel wires is determined at σ_f = 305 MPa, corresponding to $N = 2.5 \times 10^5$ cycles (R = 0.1). The damage model parameter values are approximated based on the limited test data on alloy wires, as shown in Table 1 (Lemaitre 1996).

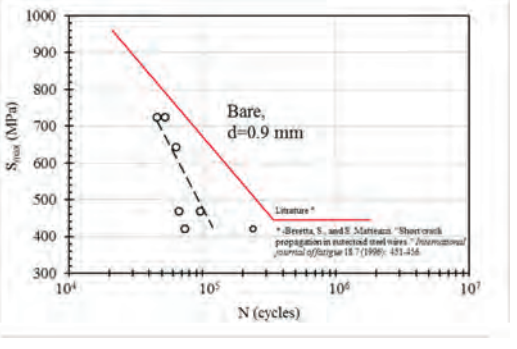


Figure 5. Fatigue strength-life (S-N) curve of the drawn steel wires, d = 5.43 mm, R = 0.1.

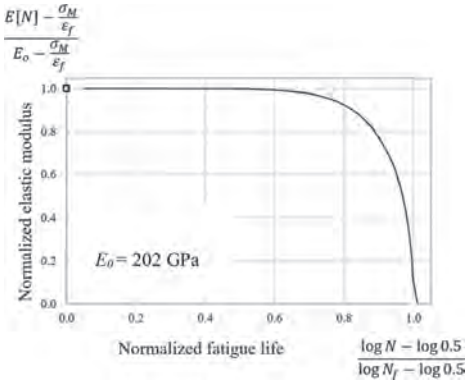


Figure 6. Normalized residual Young's modulus of the drawn steel wires.

The characteristic degradation of the Young's modulus of the drawn steel wires is established through the interrupted fatigue tests of the wire specimens. In each test, the wire specimen is subjected to a specified fatigue loading *(ΔP, R)*. The test is interrupted at every block of fatigue cycle, ΔN and the sample load-displacement response up to the maximum load cycle magnitude is recorded. The interrupted fatigue test is continued until the wire fractured. The loading and unloading residual Young's modulus as function of the accumulated load cycles are then extracted from the slope of the resulting stress-strain curves. Three different combinations of the fatigue loading is employed. The residual Young's modulus data could be presented in terms of the normalized quantities as (Adam et al. 1989, Gathercole et al. 1994):

$$E(N)=\left[1-\left(\frac{\log N-\log 0.5}{\log N_f-\log 0.5}\right)^{\alpha}\right]^{\frac{1}{\beta}}\left(E_o-\frac{\sigma_M}{\varepsilon_f}\right)+\frac{\sigma_M}{\varepsilon_f} \quad (7)$$

where α and β are curve-fitting constants. The effect of different load ratios could be incorporated through the different fatigue life, N_f of the respective specimen. The trend of the normalized residual Young's modulus curve is illustrated in Figure 6. The interrupted fatigue tests conducted by the authors for the drawn steel wires are still on-going.

4 RESULTS AND DISCUSSION

A typical distribution of the maximum principal stress in the wire rope material corresponding to the peak applied load cycle is shown in Figure 7. In the single strand wire rope under tension-tension loading cycles, the core wire experiences the greatest axial displacement under the prescribed iso-strain end condition. This is due to the core wire element having the shortest length within the pitch distance of the wire rope. The surrounding wires stretch around, inducing the contact pressure at selected locations along the core wire during the tensile stressing. However, the computed highest magnitude of 1.27 GPa is artificially induced at the localized end region of the wire rope model due to the applied boundary conditions, thus should not be considered in the fatigue analysis. The highest principal stress occurring in the relatively small trellis contact region of the core wire is about 83% of the yield strength level of the drawn wires at 1.690 GPa, as shown in the figure. Thus, the fretting fatigue model is appropriate for the analysis. Each material point in each drawn wire of the wire rope experiences different stress amplitude but at the same stress ratio of $R = 0.1$, as the applied load cycles. In addition, It is also noted that the equivalent stresses (von Mises and maximum principal stress) and the contact pressure evolve in-phase

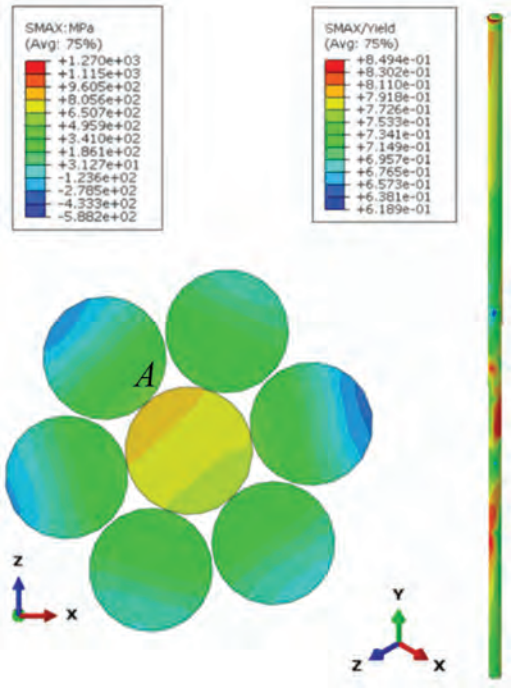


Figure 7. The maximum principal stress distribution across the critical section containing the highest contact pressure (left) and similar normalized distribution in the core wire (right).

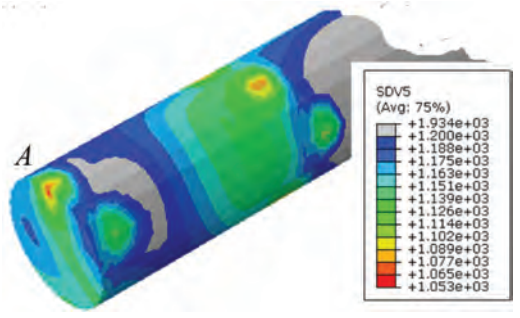


Figure 8. Distribution of the fatigue variable, N_o in part of the core wire showing the critical section (left) with the trellis contact region (marked A).

with the applied sinusoidal axial load. The peak von Mises stress in the trellis region is 1.240 GPa with the stress range of 1.116 GPa.

The calculated number of cycles required to initiate fatigue damage, N_o for the applied load-cycle is shown in Figure 8. The left cross-section represents the critical section of the core wire containing the highly stressed trellis contact region. Fatigue damage is predicted to initiate first in this localized critical region (marked A) of the core wire at $N_o = 1050$ cycles.

The grey region represents material point with $N_o > 1200$ cycles. Following damage initiation, the damage evolves with the continuous redistribution of the stresses in the wire rope to satisfy the force equilibrium and strain-displacement compatibility. The path of damage, thus fatigue crack propagation could fairly be inferred, at this early stage of the analysis, from the high-to-low gradient of N_o as illustrated in the figure.

Once the critical drawn wire fractured, the load carried by the wire is transferred to the neighboring intact wires in the strand. In a wire rope design consisting of large number of drawn wires, the fatigue life of the wire rope could be established based on the prescribed area fraction of intact-to-total load bearing area of the wires.

5 CONCLUSIONS

The framework for the development of a validated damage-based fatigue life model of the steel wire ropes has been presented and discussed. The fatigue damage model is based on the gradual degradation of strength and Young's modulus of the drawn steel wires. FE simulation of the single-strand (1×7) wire rope under axial fatigue loading ($\Delta P =$ 145 kN, $R = 0.1$) indicates that:

– the von Mises, maximum principal stress and contact pressure at the trellis contact point evolve with an identical stress ratio as the applied axial load.
– the trellis contact point is relatively small and experiences elastic stresses, thus the fatigue damage prediction for the fretting fatigue condition is appropriate.
– the critical material point at the trellis contact in the core wire experiences the first damage initiation event at $N_o = 1050$ cycles.

ACKNOWLEDGMENT

This research is supported by The EU HORIZON2020 Funding No. 703888 (UTM Grant No. 4B313) and Kiswire R&D Sdn. Bhd., Malaysia (UTM Contract Research Grant No. 4C043).

REFERENCES

Adam, T., G. Fernando, R. Dickson, H. Reiter & B. Harris (1989). Fatigue life prediction for hybrid composites. *Int. J. Fatigue*, 11, 233–237.

Alani, M. & M. Raoof (1997). Effect of mean axial load on axial fatigue life of spiral strands. *Int. J. Fatigue*. 19, 1–11.

Birkenmaier, M. (1980). Fatigue resistant tendons for cable-stayed construction. International Association for Bridge and Structural Engineering.

Cruzado, A, M. Urchegui & X. Gómez (2012). Finite element modeling and experimental validation of fretting wear scars in thin steel wires. *Wear.* 289, 26–38.

Gathercole, N., H. Reiter, T. Adam & B. Harris (1994). Life prediction for fatigue of T800/5245 carbon-fibre composites: I. Constant-amplitude loading. *Int. J. Fatigue,* 16, 523–532.

Kachanov, L.M. (1958). Rupture time under creep conditions. Izv AN SSSR, Otd Tekhn Nauk. 8, 26–31.

Kao, P.-W. & J. Byrne (1982). Fatigue initiation study of TMT eutectoid steel. *Metall. & Mat. Trans.* A. 13, 855–64.

Lemaitre, J. (1985). A continuous damage mechanics model for ductile fracture. *Trans ASME J. Engng. Mat. & Tech*. 107, 83–89.

Lemaitre, J. (1996). *A course on damage mechanics.* Springer Berlin Heidelberg.

Miner, M.A. (1945). Cumulative damage in fatigue. *J. Appl. Mech*. 12, Trans. *ASME*. 67, A159-A164.

Prawoto Y, R & B, Mazlan (2012). Wire ropes: Computational, mechanical, and metallurgical properties under tension loading. *Comput. Mat. Sci*. 56, 174–178.

Raoof, M. & T.J. Davies (2008). Axial fatigue design of sheathed spiral strands in deep water applications. *Int. J. fatigue.* 30, 2220–38.

Salleh, S., M.A. Abdullah, M.F. Abdulhamid & M.N. Tamin (2017). Methodology for reliability assessment of steel wire ropes under fretting fatigue conditions, *J. Mech. Engng. & Sci.* 11 (1), 2488–2502.

Sasaki, K., S. Iwakura, T. Takahashi, T. Moriya & I. Furukawa (2007). Estimating the fatigue life of wire rope with a stochastic approach. *J. Solid Mech. & Mat. Engng*. 1052–1062.

Suh, J.-I. & S.P. Chang (2000). Experimental study on fatigue behaviour of wire ropes. *Int. J. Fatigue*. 22, 339–347.

Wang, D, D. Zhang, S. Wang & S. Ge (2013). Finite element analysis of hoisting rope and fretting wear evolution and fatigue life estimation of steel wires. *Engng. Failure Analysis*. 27, 173–193.

Waterhouse, R.B. (2003). Fretting in Steel Ropes and Cables—A Review, Advances in Basic Understanding and Applications, STP 1425, ASTM International, PA, USA.

Winkler, J, C.T. Georgakis & G. Fischer (2015), Fretting fatigue behavior of high-strength steel monostrands under bending load. *Int. J, Fatigue*. 70, 13–23.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

An efficient computational strategy for robust maintenance scheduling: Application to corroded pipelines

E. Patelli & M. de Angelis
Institute for Risk and Uncertainty, Chadwick Building, University of Liverpool, UK

ABSTRACT: The ability to predict correctly the future remaining life time of components is of paramount importance to improve the safety and reliability of systems and networks via an effective maintenance policy. However, simplifications and assumptions are usually adopted to compensate lack of data, imprecision and vagueness, which cannot be justified completely and may, thus lead to biased results. To overcome these issues, an imprecise probabilities approach is proposed for reliability analysis and risk-based maintenance strategy. A novel efficient computational approach is proposed for identifying robust maintenance strategies. The optimal solution is obtained through only one reliability assessment based on Advanced Line Sampling and reusing the outcome of maintenance activities in a force Monte Carlo approach. The proposed methodology remove the huge computational cost of reliability-base optimization making the analysis of industrial size problem feasible. The applicability of the approach is demonstrated by identifying the optimal maintenance policy of buried pipelines and it is shown how this approach can improve the current industrial practice.

1 INTRODUCTION

One of the most important degradation/deterioration mechanisms that affect the long-term reliability and integrity of metallic pipelines is corrosion. Corrosion which leads to metal loss is the most prevailing time dependent threat to the integrity, safe operation and cause of failure for oil and gas pipelines (Caleyo et al. 2002). The corrosion process is affected by large uncertainty making the assessment of pipelines a complex and challenging task (Bazán and Beck 2013, Qian et al. 2011). For instance, uncertainties are in relation to operational data variation, associated to the randomness of the environment, form imperfect measurements of pipeline geometry, in the material strength and from the ageing processes of the pipeline.

The remaining strength of a pipeline with corrosion defects can be assessed using one or all of the international design codes viz: B31G (AMSE 1991), B31Gmod (ASME 2012), Battelle (Leis and Stephens 1997a, Leis and Stephens 1997b), DNV-101 (AS 2015) and Shell-92 (Klever et al. 1995). The associated methods use deterministic values for load and resistance variables, thereby assuming no uncertainty. In the light of the existing inherent uncertainties in the corrosion process, the obtained results are obviously quite coarse approximations, which may deviate from reality significantly. A key challenge in this regard is the probabilistic modelling, which relies on substantial information and data required to define parameter distributions. Sahraoui et al. (2013) proposed a Bayesian modelling to take into account imperfect inspection results while Li et al. (2017) suggested using Bayesian network and Markov process approach to develop an optimal maintenance strategy for corroded subsea pipelines.

However, the amount of data required to define unequivocally those distributions might not be available in practice, assumptions and simplifications are applied and often they cannot be justified completely. To solve this conflict, the use of imprecise probabilities (Beer and Ferson 2013, Beer and Patelli 2015) is proposed to realistically reflect the vagueness of the available information in the probabilistic model. In fact, since these assumptions and simplifications can be quite decisive, an imprecise probabilities approach provides a promising pathway towards a robust maintenance strategy. This paper therefore proposes the use of a novel reliability metric redefined within the framework of imprecise probabilities.

Another challenging task is the identification of optimal inspection interval time in order to reduce the overall costs of pipelines including cost of inspection, repair and failure. For instance, areas needing repairs must be accurately pinpointed as to minimise excavations for verifications. Likewise, early observations of failure mechanisms, and determination of the likelihood of failure in association with the pipeline must be handy.

The identification of optimal maintenance scheduling requires in turn the evolution of the model reliability that can be computational expensive to evaluate (Gomes et al. 2013). Approximate methods, e.g. FORM may not be sufficiently accurate or applicable for large scale problems, and we have to resort to simulation based methods.

In this paper, a novel and efficient computational technique is proposed for the identification of a robust maintenance scheduling taking into account uncertainty and imprecision. More specifically, the proposed approach allows determining the optimal inspection interval and the repair strategy that would maintain adequate reliability level throughout the service life of the pipeline obtained through only one reliability assessment. In turn, the reliability analysis is performed using Advanced Line Sampling (de Angelis et al. 2015). This allows to estimate reliability bounds with only one simulation and, in addition, it efficiency is independent of the reliability level. Hence, the proposed approach is applicable to the analysis of industrial size problem. The proposed reliability strategies are implemented in the general purpose software OpenCossan (Patelli et al. 2018, Patelli 2016, Patelli et al. 2012) and freely available.

2 MODELLING CORRODED PIPELINE

Metal losses due to corrosion affect the ultimate resistance, safety and serviceability of the structure and cause changes in its elastic and dynamic properties. These are major concerns in structural reliability assessment of existing structures and infrastructures, also in the prediction of the safe and serviceable life for both new and existing structures.

2.1 *Failure criteria*

The failure modes considered here are the loss of structural strength of pipelines through reduction of the remaining pressure strength, and pipe wall thickness caused by corrosion defects. The failure pressure are assessed using four international design codes: Shell-92, B31G, DNV-101 and Modified B31G models, respectively. The summary of all the failure pressure models is shown in Tables 1 and 2. In Table 1 W is the pipe wall thickness; L is the longitudinal length of defect; D is the outside diameter of pipe and M is the Folias' factor. In Table 2, F_p is the failure pressure and d represents the defect depth. σ_y and σ_u are the material yield stress and the ultimate tensile strength, respectively.

The assumption and limitation of these models are reflected on the individual flow stresses which

Table 1. Flow stess and Folias' factor according different international design codes.

Model	Flow stress	Folias' factor (M)
B31G	1.1 SMYS	$\sqrt{1+0.893\frac{L^2}{DW}}$
Modified B31G	SMYS + 68.95 MPa	$\sqrt{1+0.6275\frac{L^2}{DW}-0.003375\frac{L^4}{D^2W^2}}$ for $L \le \sqrt{50DW}$
		$0.032*\frac{L^2}{DW}+3.3$ for $L > \sqrt{50DW}$
DNV-101	SMTS	$\sqrt{1+0.31\left(\frac{L}{DW}\right)^2}$
Shell-92	SMTS	$\sqrt{1+0.893\frac{L^2}{DW}}$

Table 2. Failure pressure of pipelines according different international design codes.

Model	Defect (area and shape)	Failure pressure expression (F_p)
B31G	$2/3dL$ Parabolic	$1.11\frac{2\sigma_y W}{D}\left(\frac{1-\frac{2d}{3W}}{1-\frac{2d}{3W}M^{-1}}\right)$
Modified B31G	$0.85dL$ Arbitrary	$\frac{2\left(\sigma_y+68.95\right)W}{D}\left(\frac{1-0.85\frac{d}{W}}{1-0.85\frac{d}{W}\cdot M^{-1}}\right)$
DNV-101	dL Rectangle	$\frac{2\sigma_u W}{D-t}\left(\frac{1-\frac{d}{W}}{1-\frac{d}{W}\cdot M^{-1}}\right)$
Shell-92	dL Rectangle	$\frac{1.8\sigma_u W}{D}\left(\frac{1-\frac{d}{W}}{1-\frac{d}{W}\cdot M^{-1}}\right)$

are the measure of the strength of steel in presence of a defect. Folias' factors, M, is the geometry correction factor to account for the stress concentration due to radial deflection of the pipe surrounding a defect. Failure is assumed to occur as a result of the flow stress, defined by yield strength (in B31G and Modified B31G codes) or ultimate tensile strength (in DNV-101 and Shell-92) as their tensile properties. Further considerations and

assumptions on different shapes and areas of corrosion defect can also be made which might lead to different definition of failure critaria.

The failure criteria is defined as the difference between the failure pressure, F_p, of the pipeline and the maximum allowed operating pressure, ($MAOP$):

$$g = F_p - MAOP \tag{1}$$

where g is the so-called limit state function.

The easiest way to estimate the reliability of pipelines is based on safety factors (also known as level I analysis) calculated using the capacity equations or codes presented in Table 2. Such analysis do not model explicitly the uncertainties that might have occurred and increased over the years of the pipeline service. The effects of the uncertainty are considered in terms of safety margins and factors. Worst-case scenario is used for loads and capacity of the structural system and in turn, this might leads to greater safety/reliability but also to huge costs associated with the overdesign or overmaintenance of pipelines.

Level II analysis based on partial safety factors includes the first and second moment of the parameter distributions. Partial safety factors take care of uncertainties for defect depth and failure pressure (burst) capacity. For instance, DNV-101 code uses analytical expression to derive the values of standard deviation of relative corrosion defect σ_{rd}, and the failure pressure.

In modern engineering systems and critical infrastructures to assure adequate level of safety and reliability an explicit quantification of the uncertainty must be performed. A full probabilistic approach (level III analysis) requires the evaluation of multidimensional integral shown in Eq. 2. The probability of failure, P_f, is defined as:

$$P_f = P(g \le 0) = \int_{g(\boldsymbol{\theta}) \le 0} f(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{2}$$

where $f(\boldsymbol{\theta})$ represents the multivariate distribution function of the uncertainty vector $\boldsymbol{\theta}$. In realistic application a large number uncertainties need to be considered. Hence, analytical and approximate methods like FORM and SORM result to be inadequate for solving Eq. (2) (Valdebenito et al. 2010). Monte Carlo simulation methods are then required to evaluate the integral of Eq. (2). However, when dealing with rare case events, plain Monte Carlo simulation might become infeasible due to the large number of the samples required to achieve a specific level of accuracy. To overcome this limitation, advanced Monte Carlo techniques such as Line Sampling (de Angelis, Patelli, & Beer 2015) and Subset simulations (Au & Patelli 2016) can be adopted for analysing complex real world problems.

2.2 *Maintenance strategy*

In order to understand the status of pipelines, different inspection tools can be used characterised by different quality and sensitivity. The inspection activities may assess the damage incorrectly or may not even detect any damage at all based on the quality and assiciated inspection costs.

The most common tools for metal loss and crack inspection are based on the Magnetic Flux Leakage or Ultrasonic techniques (Version 2009). Pigging data is gathered through in-line inspection activities using Magnetic Flux Leakage intelligent pig, whereby the values of parameters in the model is as a result of the operations and inspection histories of the pipeline. Geometry tools are available for detecting and sizing of deformations and mapping tools for localization of a pipeline and/or pipeline features (Version 2009).

In this paper the probability of detection (PoD) associated with the non-destructive inspection techniques is modelled as (Pandey 1998):

$$PoD = 1 - \exp^{-qd} \tag{3}$$

where d represents the defect depth and q the quality of inspection.

Following an inspection, if a defect is detected, it can be repaired or not. In fact, repairing a buried pipeline is an expensive process because it requires the excavation and the replacement of part of the pipeline. For this reason, the repair is perform immediately after an inspection only if the pipe defects produce a failure pressure safety factor lower than a prescribed threshold otherwise the pipe is left unrepaired. In this case a useful remain life is estimated and a preventive maintanance can be scheduled. The threshold level of the safety factor is between 1.25 and 1.5 (Pandey 1998). These values are in agreement with the level of integrity established by actual pipeline hydro testing, and corresponds with the repair factor for a class 2 pipeline in Canadian code (Association 2007).

3 MODELLING THE UNCERTAINTIES

A full probabilistic analysis requires the proper characterisation of the uncertainties. In other words, each variable is associated with a proper probabilistic distribution function. For instance, it is practice to describe the variability of measurements as a Gaussian process characterised by its mean and standard deviation. A proper estimation

of the characteristic of the distribution (or even the shape of the distribution itself) requires the availability of data. When the amount of data is not enoughfor unambiguous characterisation of the uncertainties, expert judgement and often unjustified assumptions. This is the case in many practical situations where very limited data are available. To avoid the inclusion of subjective hypothesis, the imprecision and vagueness of the data can be treated combining probabilistic and set theoretical components in a unified construct allowing the identification of bounds on probabilities for the events of interest in order to give a different prospective to the resultss (Beer and Patelli 2015).

For the treatment of imprecise knowledge, nonconsistent information and both epistemic and aleatory uncertainty multiple mathematical concepts can be used including intervals (Augustin 2004), probability boxes (Ferson et al. 2003), normalized fuzzy sets (also known as possibility distributions) (Verma et al. 2007), Dempster-Shafer structures (Dempster 1967, Shafer 1976), Bayesian frameworks (Faber 2005) and Random Set theory. In particular, Random Set theory is a general framework suited to model uncertainty represented as cumulative distribution functions (CDFs), without making any implicit or explicit assumption at all. Explanatory examples of such flexible frameworks are provided in (Patelli et al. 2015, Rocchetta et al. 2018).

In this paper, the concept of probabilistic boxes (P-boxex) is used (Ferson, Kreinovich, Ginzburg, Myers, & Sentz 2003). P-boxes can be seen as a generalization of the Dempster-Shafer structures where the sets are represented by distributions. Hence, P-boxes are sets of Cumulative Distribution Functions (CDFs) for which lower and upper bounds are assigned $\left[\underline{F}_X, \overline{F}_X\right]$. The probability distribution associated to the random variable x can be either specified or not. The former are generally named distributional P-boxes, or parametric P-boxes, the latter are named distribution-free P-boxes, or non-parametric P-boxes. In literature the upper bound on probability is referred as plausibility and the lower bound as belief.

Distributional p-boxes appear when there is indetermination in the representations of the parameters of a given CDF. These parameters are imprecisely specified as intervals. For instance, consider a quantity that is known to be Gaussian with mean within the interval [1,2] and standard deviation somewhere in [3,4]. All CDFs that are normal and have means and standard deviations inside these respective intervals will belong to this probability box. The upper and lower CDF bounds $\underline{F}$ and $\overline{F}$ of the p-box enclose many non-normal distributions, but these would be excluded from the p-box by specifying the normal CDF as the *parental distribution family*.

The calculation of the bounds of the quantity of interest such as the probability of failure requires significant computational resources. This because it will be necessary to estimate the integral of Eq. (2) for all the possible probabilistic model considered and then identify the bounds of the response. Fortunately in many engineering applications the response of the model is monotonic with respect to the imprecision of the input parameters. In general, this allows to estimate the bounds of the probability of failure with only 2 full probabilistic analysis (Rocchetta et al. 2018). Advanced Line Sampling (de Angelis et al. 2015) method can further reduce the computational cost allowing the estimation of the bonds of the probability of failure with only 1 efficient probabilistic analysis (de Angelis et al. 2014).

4 ROBUST MAINTENANCE STRATEGY

Inspection and monitoring of pipelines is necessary in order to ensure their continued fitness for purpose, which entails protection from any time-dependent degradation processes, such as corrosion. Also, pipeline failures have significant impact on the economic, environmental and social aspects of the society. Therefore, the proper assessment and maintenance of such structures are crucial; negligence will lead to serviceability loss, failure and might lead to catastrophic environmental and financial consequences. On the other hand, maintenance is an expensive activity and the availability of robust maintenance scheduling is of paramount importance. The premise for these decisions is supplied by reliability estimation inculcating the impact of inspection scheduling and reparation activities over the pipeline's service life.

4.1 *Optimization problem*

In reliability-based optimization, the total expected costs in relation to maintenance and failure is the objective function that needs to be minimised, see e.g. (Beer et al. 2014). The time and number of the inspections represents the design variables of the optimization problem while the expected monetary costs associated with inspection, repair and failure form the objective function that can be formulated as:

$$\underset{N_I, q, t_i}{\arg\min} \quad E\left[C_T\left(N_I, q, t_i\right)\right] \tag{4}$$

where N_I, q, t_i and C_T denote the number of inspections, the qualities of inspections, the i-th time of inspection and the expected total cost, respectively. The expected total costs are defined as:

$$E\left[C_T\left(N_I,q,t_i\right)\right]=E\left[C_I\left(N_I,q,t_i\right)\right]+E\left[C_R\left(N_I,q,t_i\right)\right]+E\left[C_F\left(N_I,q,t_i\right)\right] \quad (5)$$

where C_I, C_R and C_F are the costs of inspection, repairs and failure, respectively.

In addition, the optimisation problem must satisfy some constraints. For instance, it might be necessary to guarantee a minimum level of reliability, i.e.:

$$R(t)=1-P_f(t)\geq\beta \qquad \forall t\in\left[0,T_m\right] \quad (6)$$

where $P_f(t)$ is the probability of failure at time t and T_m represents the so-called mission time. Hence, reliability based maintenance strategy requires the evaluation of the reliability over the time as summarised in Section 2.1. Constrained optimisation techniques are then adopted to identify the minimum of the objective function, Eq. (4), satisfying the constrain of Eq. (6).

4.2 *Inspection costs*

The expected inspection costs $E(C_I)$ are calculated as the product of the unit inspection cost, c_I, that depends on the quality of inspection q, corrected by the discount rate, r, and the probability that inspection will take place at time t_i: $1 - P_f(t_i)$. In other words, the pipeline has not to be failed before the i-th inspection time scheduled at t_i. This expected costs are expressed in mathematical form as:

$$E\left[C_I\right]=\sum_i \frac{c_I(q)}{(1+r)^{t_i}}\cdot\left(1-P_f\left(t_i\right)\right) \quad (7)$$

4.3 *Repair costs*

The evaluation of the expected costs associated with repair, $E\left[C_R\right]$, is quite challenging since depends on the probability of performing a repair after the i-th inspection, $P_R(t_i)$. This, in turn depends on the probability of detection (POD) (i.e. the probability to detect a defect). The expected repair costs are modelled as:

$$E\left[C_R\right]=\sum_{i=1}^{N_I} c_R P_R\left(t_i\right)\frac{1}{(1+r)^{t_i}} \quad (8)$$

where c_R, is the unit cost of a repair. The probability of repair is calculated by computing the reliability analysis of the pipeline where the repair threshold represents the limit state function weighted by the probability to detect the crack, i.e.

$$P_R(t)=\int_{1.25\leq g(\boldsymbol{\theta},t)\leq 1.5} f(\boldsymbol{\theta})\cdot PoD(d,t)\,d\boldsymbol{\theta} \quad (9)$$

4.4 *Failure costs*

The total capitalized expected costs, $E[C_F]$, due to failure are the costs associated with failure over the region of the corresponding demand functions. Hence, the calculation of the failure costs requires the estimation of the probability of failure of the pipeline over the time. Teh computational strategy proposed in the next Section allows to estimate these costs by performing a single reliability analysis and the reusing the results in the optimisation loop.

The cost of failure at a specific inspection time is calculated as the cost of the failure of i-th time t_i (that is proportional to $P_F\left(t_i\right)$ minus the cost of failure at previous time t_{i-1}: $\propto P_F\left(t_{i-1}\right)$. This allows to take into account the fact that the system has survived till the time t_{i-1}. Taking into account all the inspection times and the discount costs, the expected failure cost becomes:

$$E\left[C_F\right]=\sum_{i=1}^{N_I} c_F\frac{\left(P_f\left(t_i\right)-P_f\left(t_{i-1}\right)\right)}{(1+r)^{t_i}} \quad (10)$$

5 COMPUTATIONAL STRATEGY

5.1 *Reliability analysis*

The estimation of the probability of failure requires in general significant computation efforts. In particular for highly reliable pipelines, the number of model evaluations easily exceed the computational resources available. In addition, the presence of imprecision adds another level of complexity because the propagation of intervals and p-boxes requires the adoption of an additional optimization loop making the required computational cost quite challenge. For this reason, the Advanced Line Sampling (de Angelis et al. 2015) method is adopted to estimate the bounds of the probability of failure. One of the key feature of this approach is the ability to estimate different probabilities of failure (associated to different levels of the performance function) with only one analysis. For instance this can be used to estimate with only one reliability analysis the bounds of probability of failure due to imprecision in the inputs and the probability of repairs as well.

5.2 *Robust maintenance*

The robust maintenance is computed adopting a novel computational strategy that allows to reuse

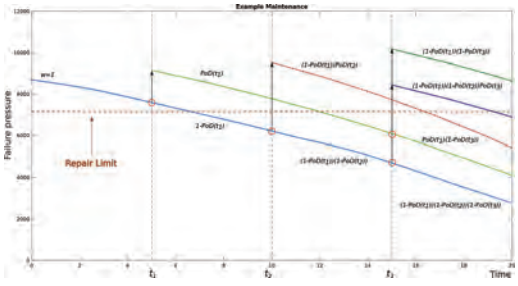


Figure 1. Effect of maintenance (repairs) on the weights associated with realisations of evolution on the time of the failure pressure.

the results of the reliability analysis in the optimisation problem.

In order to explain the simulation approach, we first consider a simplified model without imprecision and solved using plain Monte Carlo method. The idea is to first simulate the evolution of defects/cracks in pipelines without considering inspections and repairs. This is performed by sampling the parameters of the model and then solving the equations in Tables 1–2 till the time of interest. At this point we have a number of possible cracks evolution (or failure pressures) over the time. Then, we add the effect of maintenance and update the corresponding pipeline reliability as shown in Figure 1 by calculating the weights associated to each possible event outcome. The procedure is repeated for all the simulated cracks evolution. The computed weights are then used to calculate the probability of failure at time of interest. For instance, the probability of failure at time t_i is estimated by the summation of the weights associated to the failure events divide by the total numberof simulations.

Finally an optimisation tools is used to identify the number and time of inspections that minimise Eq. (4). When a new inspection time is proposed, it is necessary to recompute the weights starting from the original simulation but thisstep does not require the re-analysis of the model (i.e. evaluating the evolution of the crack/defect till the time of interest).

6 EXAMPLE APPLICATION

The optimal maintenance scheduling of a pipeline with characteristics shown in Table 4 is performed.

6.1 *Reliability analysis of pipelines*

First, the probability of failure of the pipeline as a function of time has been computed using the DNV-101, Shell-92, B31G and B31Gmod codes without considering inspection and maintenence. The uncertainties are modelled as shown in Table 5. Different level of imprecision on the parameter vaules has also been condidered.

Advanced Line Sampling (de Angelis et al. 2015) is adopted to estimate the reliability of the pipelines with 20 lines resulting in 120 model evaluations. Advance Line Sampling is able to deal with imprecision in the parameter values and it allows to compute the bounds of the reliability. In addition, the number of model evaluations are independent of the reliability level. As expected, the probability of failure of the corroded pipeline increases with time as shown in Figure 2. The Figures shows lower and upper bounds of the probability of failure when 10% of imprecision is considered on the input variables. Shell-92 and the DNV-101 are the most conservative models followed by Modified B31G and the least conservative is the B31G model. than 0.6) respectively. This is in accordance with results from literature obtained without considering imprecision (see e.g. Caleyo et al. (2002)). The results of the analysis are also summarised in Table 3.

6.2 *Robust maintenance*

Maintenance is a very effective way to improve the safety of corroded pipelines. The aim of this section is to identify the optimal number of inspections that are able to minimise the overall costs. Maintenance is only performed is a defect is detected. The typical minimal detectable depth of a high resolution Magnetic Flux Leakage tool for uniform corrosion is 0.1 W with a POD of 0.9 (Version 2009). Using these values and the pipeline wall thickness $W = 9.52\ mm$, the quality of inspection can be estimated as $q = 2.42$ (from Eq. 3). However, if the length of the defect is $l < 30mm$ we have a pitting defect. In this case the quality of inspection is reduced to $q = 1.61$.

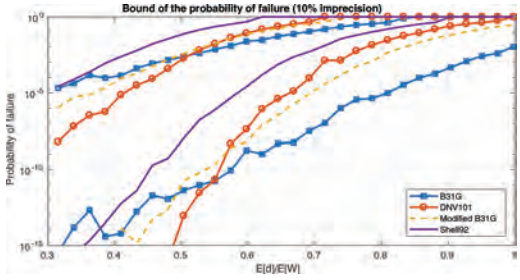


Figure 2. Lower and upper bounds of the probability of failure of a pipeline with 10% of imprecision on the variables using Shell-92, B31G, Modified B31G and DNV-101 failure pressure models.

In this example, it is assumed that the inspection time are equally spaced from the initial time till the final time of 50 years (mission time). Figure 3 shows the pipeline failure probability at mission time against the number of inspections for different models and parameter imprecision. From the results presented in the Figure 3, it can be deduced that using B31G model and 3 inspections suffice reducing the probability of failure of the pipeline below 10^{-6}. However, when the modified B31G model is used more than 6 inspections are required (using the upper bounds of the parameters). These results allows to identify the minimum number of inspections required to guarantee a prescribed level of safety. The very small probability of failure have been calculated adopting the approach presented in Section 5.

Figures 4 and 5, show the total expected cost as a function of the number of inspection obtained using DNV-101 model. Obviously, the costs of inspection increases with the number of inspections performed during the lifetime of the pipeline. Costs of failure decreased with the number of inspections. For very small number of inspections the total costs are governed by the costs associated with failure while for large number of inspections, the total maintenance costs are due to the costs associated with repairs. The optimal number of inspection is always a trade-off between costs of failure and costs of repairs. Using the DNF-

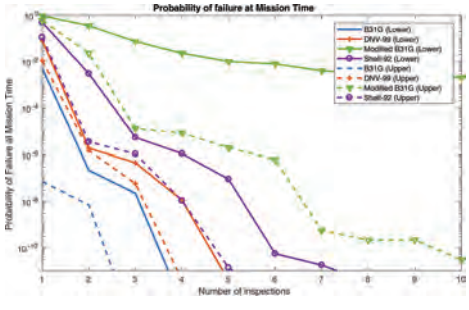


Figure 3. Probability of failure (at mission time) for a pipeline as a function of the number of inspections.

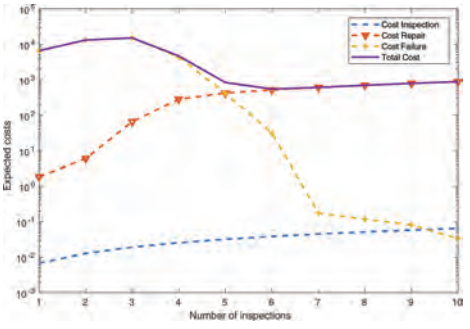


Figure 4. Expected total maintenance cost as a function of the number of inspection using the DNF-99 model, with a mission time of 50 year and upper bounds of imprecise parameters.

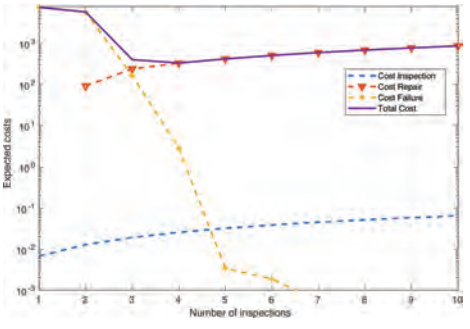


Figure 5. Expected total maintenance cost as a function of the number of inspection using the DNF-99 model, with a mission time of 50 year and lower bounds of imprecise parameters.

Table 4. Pipeline characteristics.

Parameter	
Transported substance	Crude oil
Pipe outlay	Below ground
Outside Diameter	609.6 mm
Material Class X52	SUTS 496 MPa, SMYS 358 MPa, MAOP 4.96 MPa.
Nominal wall thickness	9.52 mm

Table 3. Bounds of the relative corrosion defect for different safety levels with 10% imprecision on model parameters.

Safety level	B31G	DNV-101	Mod-B31G	Shell-92
10^{-3}	[0.4799, 0.9273]	[0.5035, 0.7389]	[0.4799, 0.7860]	[0.4093, 0.6683]
10^{-4}	[0.4093, 0.8566]	[0.4564, 0.7154]	[0.4328, 0.7389]	[0.3622, 0.6212]
10^{-5}	[0.3151, 0.8095]	[0.4328, 0.6683]	[0.3858, 0.7154]	[0.3151, 0.5977]
10^{-6}	[0.3151, 0.7625]	[0.4093, 0.6447]	[0.3151, 0.6683]	[0.3151, 0.5741]
10^{-7}	[0.3151, 0.7154]	[0.3622, 0.6212]	[0.3151, 0.6212]	[0.3151, 0.5270]

Table 5. Stochastic model used for the corroded pipeline and monetary unit cost for operation (i.e. multiplicative factors).

Parameter	Symbol & unit	Distribution	Mean	CoV	Parameter	Symbol	Value
Diameter	D [mm]	Normal	609.6	0.02	Inspection cost	c_I	0.018
Defect depth	d [mm]	Normal	3	0.10	Repairs cost	c_R	0.243
Wall thickness	W [mm]	Normal	9.52	0.02	Failure cost	c_F	36.55
Ultimate Tensile Strength	σ_i [MPa]	Log-Normal	496	0.07	Discount Rate	r	0.05
Pipe Yield Stress	σ_y [MPa]	Normal	358	0.07			
Defect length	l [mm]	Normal	200	0.1			
Operating Pressure	Op [Mpa]	Log-Normal	4.96	0.10			
Radial corrosion rate	vd [mm/yr]	Log-Normal	0.5	0.10			
Long. Corrosion rate	vl [mm/yr]	Log-Normal	0.5	0.10			

99 model the optimal number of inspections is between 4 an 6.

7 CONCLUSIONS

In this paper, an efficient numerical approach for robust optimal pipeline inspection time scheduling has been proposed. This allows to determine the optimal inspection interval and the repair strategy that would maintain adequate reliability throughout pipeline service life. The computational framework allows to take into account the uncertainties of the model and imprecisions on the knowledge of model parameters. The proposed approach is efficient since allows to perform reliability based optimisation with only one reliability analysis.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. David A. Opeyemi for the preliminary work on pipelines corrosion and Mohamed El Amine Ben Seghier for the useful comments on the paper.

This work has been supported by the UK Engineering and Physical Sciences Research Council with the project “Smart on-line monitoring for nuclear power plants (SMART)” (Grant EP/M018415/1) and by the European Unions Research and Innovation funding programme (Framework Programme) under the PLENOSE project (PIRSES-GA-2013-612581).

REFERENCES

AMSE (1991). Asme b31: Manual for determining the remaining strength of corroded pipelines. Technical report, American Society of Mechanical Engineers (ASME).

AS, D.N.V. (2015). *Recommended Practice DNV-RP-F101*.

ASME (2012). Manual for determining the remaining strength of corroded pipelines. Technical report, American Society of Mechanical Engineers (ASME).

Association, C.S. (2007). Csa z662–07: Oil and gas pipeline systems. Technical report.

Au, S.K. & E. Patelli (2016). Subset simulation in finiteinfinite dimensional space. *Reliability Engineering & System safety 148*, 66–77.

Augustin, T. (2004). Optimal decisions under complex uncertainty—basic notions and a general algorithm for data-based decision making with partial prior knowledge described by interval probability. *Special Issue of ZAMM—Zeitschrift fr Angewandte Mathematik und Mechanik 84*(10–11), 1–10.

Baz´an, F.A.V. & A.T. Beck (2013, Sep). Stochastic process corrosion growth models for pipeline reliability. *Corrosion Science 74*, 5058.

Beer, M. & S. Ferson (2013). Special issue of mechanical systems and signal processing ”imprecise probabilities-what can they add to engineering analyses?”. *Mechanical Systems and Signal Processing 37*(1–2), 1–3.

Beer, M., I.A. Kougioumtzoglou, & E. Patelli (2014). *Maintenance and Safety of Aging Infrastructure*, Chapter Emerging Concepts and Approaches for Efficient and Realistic Uncertainty Quantification, pp. 121–162. Structures & Infrastructures. Taylor & Francis Publishers.

Beer, M. & E. Patelli (2015). Editorial: Engineering analysis with vague and imprecise information. *Structural Safety Special Issue: Engineering Analyses with Vague and Imprecise Information 52, Part B*(0), 143. Engineering Analyses with Vague and Imprecise Information.

Caleyo, F., J. González, & J. Hallen (2002, Jan). A study on the reliability assessment methodology for pipelines with active corrosion defects. *International Journal of Pressure Vessels and Piping 79*(1), 7786.

de Angelis, M., E. Patelli, & M. Beer (2014, June). Line sampling for assessing structural reliability with imprecise failure probabilities. In *Vulnerability, Uncertainty, and Risk*, pp. 915–924–. American Society of Civil Engineers.

de Angelis, M., E. Patelli, & M. Beer (2015). Advanced line sampling for efficient robust reliability analysis. *Structural safety 52*, 170–182.

Dempster, A.P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics 38*, 325–339.

Faber, M.H. (2005). On the treatment of uncertainties and probabilities in engineering decision analysis. *Journal of Offshore Mechanics and Arctic Engineering 127*, 243–248.

Ferson, S., V. Kreinovich, L. Ginzburg, D.S. Myers, & K. Sentz (2003, January). Constructing probability boxes and Dempster-Shafer structures. Report SAND2002–4015, Sandia National Laboratories, Albuquerque, NM. Available at http://www.ramas.com/unabridged.zip.

Gomes, W.J., A.T. Beck, & T. Haukaas (2013). Optimal inspection planning for onshore pipelines subject to external corrosion. *Reliability Engineering & System Safety 118*, 18–27.

Klever, F., G. Stewart, & C. v. d. Valk (1995). New developments in burst strength predictions for locally corroded pipelines. In *International conference on offshore mechanics and arctic engineering, Copenhagen (Denmark), 18–22 Jun 1995*. American Society of Mechanical Engineers, New York, NY (United States).

Leis, B. & D. Stephens (1997a). An alternative approach to assess the integrity of corroded line pipe - part i: Current status. In *The Seventh International Offshore and Polar Engineering Conference, 25–30 May, Honolulu, Hawaii, USA*, pp. 624–641. International Society of Offshore and Polar Engineers.

Leis, B. & D. Stephens (1997b). An alternative approach to assess the integrity of corroded line pipe - part ii: Alternative criterion. In *The Seventh International Offshore and Polar Engineering Conference, 25–30 May, Honolulu, Hawaii, USA*. International Society of Offshore and Polar Engineers.

Li, X., H. Zhu, G. Chen, & R. Zhang (2017). Optimal maintenance strategy for corroded subsea pipelines. *Journal of Loss Prevention in the Process Industries 49*, 145–154.

Pandey, M.D. (1998). An effective approximation to evaluate multinormal integrals. *Structural Safety 20*(1), 51–67.

Patelli, E. (2016). *Handbook of Uncertainty Quantification*, Chapter COSSAN: A Multidisciplinary Software Suite for Uncertainty Quantification and Risk Management, pp. 1–69. Cham: Springer International Publishing.

Patelli, E., D.A. Alvarez, M. Broggi, & M. de Angelis (2015). Uncertainty management in multidisciplinary design of critical safety systems. *Journal of Aerospace Information Systems 12*, 140–169.

Patelli, E., H.M. Panayirci, M. Broggi, B. Goller, P.B.H.J. Pradlwarter, & G.I. Schu¨eller (2012). General purpose software for efficient uncertainty management of large finite element models. *Finite Elements in Analysis and Design 51*, 31–48.

Patelli, E., S. Tolo, H. George-Williams, J. Sadeghi, R. Rocchetta, M.D. Angelis, & M. Broggi (2018). Opencossan 2.0: an efficient computational toolbox for risk, reliability and resilience analysis. In *Proceedings of the joint ICVRAM ISUMA UNCERTAINTIES conference*.

Qian, G., M. Niffenegger, & S. Li (2011, Mar). Probabilistic analysis of pipelines with corrosion defects by using fitnet ffs procedure. *Corrosion Science 53*(3), 855861.

Rocchetta, R., E. Patelli, & M. Broggi (2018, February). Do we have enough data? robust reliability via uncertainty quantification. *Journal of Applied Mathematics 54*, 710–721.

Sahraoui, Y., R. Khelif, & A. Chateauneuf (2013). Maintenance planning under imperfect inspections of corroded pipelines. *International Journal of Pressure Vessels and Piping 104*, 76–82.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.

Valdebenito, M., H. Pradlwarter, & G. Schu¨eller (2010). The role of the design point for calculating failure probabilities in view of dimensionality and structural non linearities. *Structural Safety 32*(2), 101–111.

Verma, A., A. Srividya, & R. Gaonkar (2007). Maintenance and replacement interval optimization using possibilistic approach. *International Journal of Modelling and Simulation 27*(2), 193–199.

Version (2009). Specifications and requirements for intelligent pig inspection of pipelines. Technical report, Pipeline operators' forum.

Structural reliability

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Integrity detection of mooring chains by the approach of thermography

Wenxian Yang
School of Mechanical Engineering, Hunan Institute of Engineering, Hunan Province, Xiangtan, China
School of Engineering, Newcastle University, Newcastle upon Tyne, UK

Kexiang Wei
School of Mechanical Engineering, Hunan Institute of Engineering, Hunan Province, Xiangtan, China

Zhike Peng
State Key Laboratory of Mechanical System and Vibration, Shanghai Jiaotong University, Shanghai, China

ABSTRACT: Reliability and safety issues of mooring chains are causing concern in recent years. Accordingly, some efforts have been made for detecting the structural integrity of mooring chains. However, a fully successful mooring chain condition monitoring technique has not been achieved today. This is largely due to the fact that mooring chains are submerged in water and the currently available non-destructive testing technologies are difficult to apply in wet environment. To overcome this issue, a new mooring chain condition monitoring method is studied in this paper with the aid of thermography technique. The research is conducted based on two philosophies, i.e. (1) the mooring chain material has much higher thermal conductivity than that of water. Therefore, when the mooring chain is heated, the thermal energy will transmit mostly inside the chain, rather than dispersing in water; (2) the defects occurring in the mooring chain will disturb the transmission of thermal flow inside the mooring chain and consequently change the distribution of the temperature in the adjacent area. To demonstrate the effectiveness of the proposed method, both numerical and experimental researches are conducted in this paper. The research results have shown that thermography is indeed valid in detecting the integrity of mooring chains.

1 INTRODUCTION

There are a variety of non-destructive testing techniques that have been developed for addressing various structural integrity testing and assessment issues in different fields, such as those depicted in References (see Blitz 2012, Sanjeev et al. 2013, Amenabar 2011 and Garcia-Martin et al. 2011). However, few of them is applicable to monitoring the structural integrity of mooring chains as the mooring chains are full submerged in the water located in harsh marine and offshore environment.

In order to tackle this issue, much effort has been made by the scholars and industrialists in recent years, although a cost and technically effective mooring chain condition monitoring technique has not been successfully achieved till today. This is because so far, almost all the existing mooring chain condition monitoring techniques and systems (see AVT Reliability 2017, Seatools 2017, Lugsdin 2017), with the exception of the ultrasonic guided wave technique developed by TWI (see TWI 2013), are originally designed for detecting a broken mooring line rather than detecting and monitoring the growth of the defects occurring in it. In reality, a defective mooring chain may but is not necessary lead to broken of the mooring line when it is subject to extreme loads. For example, Remotely Operated Vehicle (ROV) inspection has been popularly adopted for inspecting typical damage and loss of integrity of marine structures. It can be equally applied to the inspection of mooring chains. However, ROV inspection only provides snapshot of the surface of mooring chains, which could be covered by thick layer of marine lives. Therefore, visual inspection via ROV is unable to provide the operator with reliable information about the actual structural health condition of mooring chains. Moreover, the application of ROV inspection is limited by weather windows and access, it is unlikely to realize the continuous monitoring of the mooring chain. In order to obtain continuous monitoring data from mooring chains, AVT Reliability attempted to use strain gauge to monitor the integrity of mooring chains and applied the devised strain gauge based measurement system to assessing the integrity of the 9 mooring chains installed on a 870,000 barrel oil storage tanker (AVT Reliability 2017). The novelty of such a system is that it does not

directly measure the chain tensions but instead, monitors the stresses in the buoy structural steelwork in reacting those same chain tensions. This has the advantage that the instrumentation can be mounted internally inside the buoy in a clean dry environment. However, the measured stresses from the buoy structural steelwork are not only dependent on the integrity of mooring chains, but also affected by the motions of the storage tanker and the external loads acting on it. Therefore, the AVT system is effective in detecting a broken mooring line, however ineffective in detecting and monitoring the incipient defects occurring in it. Apart from AVT Reliability, the other companies also develop mooring chain integrity monitoring systems using different techniques. For example, Seatools developed the mooring chain inclination measurement technique (Seatools 2017), Tritech International Ltd developed a multi-beam sonar technique (Lugsdin 2017), and so on. But they all for detecting whether the mooring lines are well connected to the floating structures, rather than for detecting the defects occurring in the chains. To tackle this issue, TWI developed an automated ultrasonic guided wave technique for monitoring mooring chains (TWI 2013). Laboratory test has shown that such a technique does work in improving the accuracy, consistency and repeatability of inspection results. But it requests to make minimal surface preparation before conducting inspection. However, this is very difficult to implement in the practical application. Additionally, the high cost of the associated robotic delivery system also limits the extensive application of such an ultrasonic guided wave technique. In view of this, a new mooring chain condition monitoring technique is studied in this paper with the aid of thermography technique. The details of the numerical and experimental research are given below.

2 HYPOTHESES

According to the fundamental theory of thermodynamics, the rate of heat flow can be described as (Borgnakke et al. 2003):

$$\frac{Q}{t} = \frac{\kappa A}{d}\Delta T \tag{1}$$

where Q = the amount of heat transferred in a time t; κ = the thermal conductivity constant for the material; A = the cross sectional area of the material transferring heat; d = the thickness of the material; and ΔT = the difference in temperature between one side of the material and the other.

From (1), there are two hypotheses can be inferred that: (1) since the thermal conductivity κ for steel is 46 Watts/meter–°C, which is much higher than the thermal conductivity of water (i.e. κ for water is only 0.58 Watts/meter–°C), the heat flow will be transferred much faster in steel than in water. Accordingly, when the steel mooring chain is heated from one end, the majority of heat flow will be transferred inside the steelwork of the chain rather than being dissipated in the water around it; (2) equation (1) shows that the cross sectional area A is inversely proportional to the temperature difference ΔT. That means when the same amount of heat is transferred inside the chain, the defect resulted change in the cross sectional area A will be indicated by the change in temperature or temperature distribution. In the meantime, the heat flow will be transferred automatically along the path which has smaller thermal resistance.

If the above two hypotheses are true, the integrity of the mooring chain then can be detected via observing the distribution of mooring chain temperature. In other words, the discontinuity in temperature distribution may indicate the presence of defect in the structure of mooring chain.

3 NUMERICAL RESEARCH

In order to demonstrate the aforementioned two hypotheses and investigate the transferring process of the heat flow insider a mooring chain when it is heated, the numerical model of mooring chains is developed in ANSYS 15. In each chain, the center to center distance is 24 mm, the width is 18 mm. The radius of the side circle is 4 mm. Then, different types of defects were artificially made on the model through changing its geometries. The mooring chains with different sizes of cracks with 0.5 mm clearance are shown in Figure 1.

In ANSYS 15, the steady state thermal analysis are conducted. In the engineering data section of the analysis, steel was chosen as the material for the mooring chain. In the calculation, the mooring chains are meshed and moreover, mesh refinement properties were used for building finer meshes particularly in the vicinity areas of the defects, such as the example given in Figure 2.

In the numerical calculations, both environmental temperature 22 Celsius and heating temperature

Figure 1. The mooring chains with different integrity conditions.

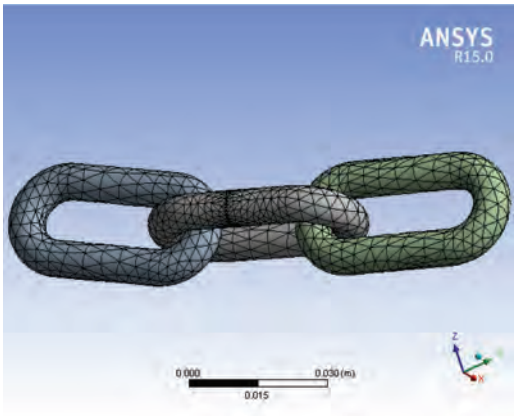


Figure 2. Mesh and mesh refinement of the mooring chains.

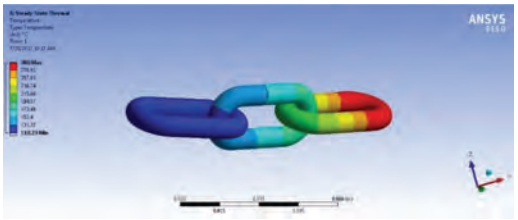


Figure 3. Numerical result obtained when a partially cracked mooring chain is heated from one end.

300 Celsius are specified in the steady state thermal column. The convection film coefficient is also defined in this section. The value is taken as 22 W/m^2°C. When the heating temperature is applied at one end of the mooring chains while the other end is fixed and no temperature is applied, the numerical calculation results for a partially cracked mooring chain are graphically shown in Figure 3.

From Figure 3, it is found that: (1) the discontinuity of the temperature observed from the upper part of the defective mooring chain proves that the defect in the mooring chain does disturb the transfer of heat flow, thus may cause visible temperature difference in the vicinity area of the defect; (2) the asymmetric distribution of the temperatures of the upper and lower parts of the defective mooring chain proves that heat flow is more easily to be transferred along the path that has smaller thermal resistance; (3) although in the numerical simulation, the mooring chain is assumed to be placed in air rather in water, the visible temperature differences prove that when the steel mooring chain is heated from one end, the majority of heat flow will be transferred inside the steelwork of the chain rather than being dissipated outside the chain as the thermal conductivities of air and water are similar and much smaller than that of steel mooring chain; and (4) moreover, the profile of the temperature distribution around the defect indicates the size of the defect. In order to further verify the findings from Figure 3, a fully cracked mooring chain is heated from both ends, as shown in Figure 4.

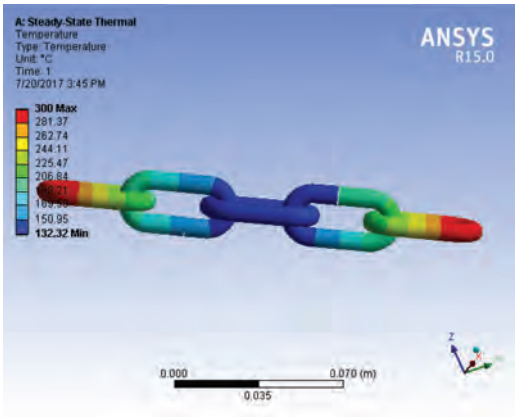


Figure 4. Numerical result obtained when a fully cracked mooring chain is heated from both ends.

From Figure 4, it is clearly seen that the aforementioned four findings are also valid when the mooring chain is completely cracked and heated from both ends. This indicates that, in the sense of theory, the thermography do have potential to be applied to detect and monitor the defects occurring in mooring chains.

4 EXPERIMENTAL RESEARCH

In order to physically demonstrate the interesting findings observed in numerical research, experimental research is organized in laboratory. The perfect and defective chains with different sizes of cracks are shown in Figure 5. Where, the chain cracks are artificially made using hacksaw.

In the experiment, the influences of the cracks on the chain temperature distributions in their vicinity areas are investigated in two scenarios in order to find a better heating method that can lead to more reliable condition monitoring result. In the first scenario, the mooring chain being investigated is heated from the both ends of it; while in the second scenario the mooring chain is headed from only one end. The experimental results obtained in the first scenario are shown in Figure 6.

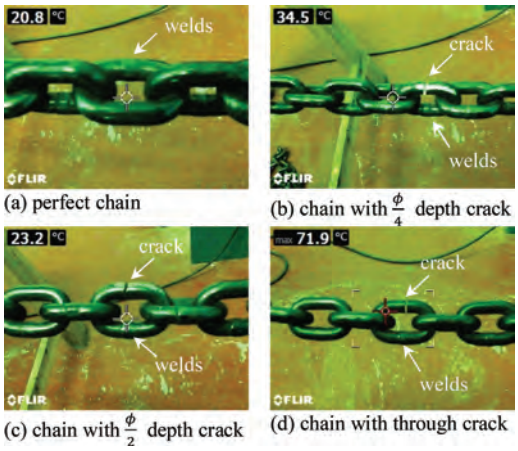


Figure 5. Mooring chains used in the experiments.

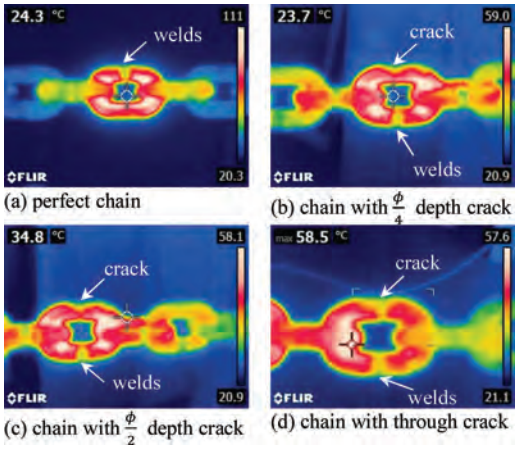


Figure 6. Experimental results obtained when the chains are heated from both ends.

From Figure 6, it is found that when the mooring chain is perfect and has no any defect inside the chain structure, the temperature is distributed evenly and smoothly over the chain except at welds, where the temperature is obviously smaller than in others due to the much higher thermal conductivity of the welding material. But when a crack is present in the chain, the even distribution of the temperature over the chain will be discontinued. Consequently, a concave profile can be observed from the cracking area in the thermal image. Moreover, it is found that the larger the size of the crack, the deeper the concave profile tends to be. This is because the air or water in the clearance of the crack has lower thermal conductivity than that of the mooring chain material. Therefore, the air or water temperature in the crack clearance is lower than that of the steelwork of the chain. From such an observation, it can be inferred that a partial through crack can partially stop the transfer of heat flow, although the heat flow is still able to be transferred through the un-cracked section. But when the crack continues to propagate and finally becomes a full through crack in the end, the transfer of heat flow will be significantly limited by the crack. In this worst case, the heat flow in the cracking area is transferred only via the air or water in the clearance of the full through crack. In general, from these experimental results obtained in the first scenario, it can be concluded that the crack occurring in the mooring chain can be readily detected using thermographic technique. Moreover, the size of the crack can be approximately understood through observing the concave depth of the temperature profile in the vicinity area of the crack. However, more accurate evaluation of the crack is difficult to achieve due to the limitation of observation.

In order to further improve the accuracy of the crack evaluation, the experiment is repeated in the second scenario. But the difference from the first scenario is the mooring chains being investigated are heated from only one end. The corresponding experimental results are shown in Figure 7.

From Figure 7, it is found that the similar phenomena observed from the first scenario (see Figure 6) also can be observed. But the assessment of the crack is not easy to achieve through observing the concave depth of the temperature profile because the concave feature caused by the crack cannot be clearly observed in the second scenario. Accordingly, a quantitative assessment method is developed in the following by using the temperature reading function of the thermographic camera. In addition, it is noticed that there is a temperature value displayed at the top left of the picture, such as $80.8°C$ in Figure 7a, $26.6°C$ in Figure 7b, $25.2°C$ in Figure 7c, and $33.8°C$ in Figure 7d. These values indicate the temperatures at the positions located by the circle with a cross. With the aid of this special temperature reading function provided by the thermal camera, the temperature at any position in the thermal image can be readily obtained. Then, the temperatures at two specific positions are measured for developing the quantitative assessment criterion. One is the temperature T_h measured at the heating source position, another is the temperature T_c measured at the other side of the crack. Since the temperature at the heating source position is the highest temperature in the thermal image, the value of T_h is usually the maximum value shown in the grey scale of the image, i.e. $T_h = 216\,^oC$ for Figure 7a, $192\,^oC$ for Figure 7b, $263\,^oC$ for Figure 7c, and $99.6\,^oC$ for

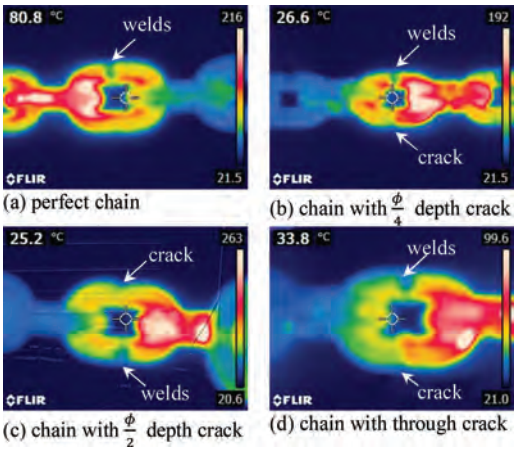


Figure 7. Experimental results obtained when the chains are heated from one end.

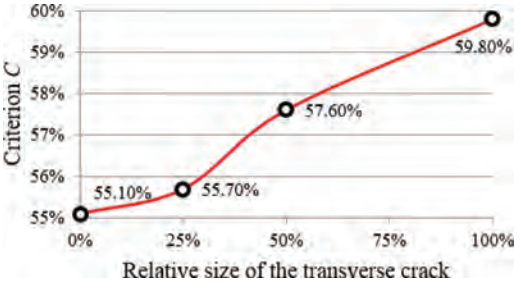


Figure 8. Quantitative assessment of the crack occurring in the mooring chain.

Figure 7d. Therefore, in fact T_c is the only value that needs to be measured by moving the circle with a cross to the other side of the crack. Once the value of T_c is obtained, the following quantitative assessment criterion can be calculated, i.e.

$$C = \frac{T_h - T_c}{T_h} \times 100\% \qquad (2)$$

In essence, the criterion C measures the significance that the crack disturbs the heat transfer along the mooring chain. When the crack increases in size, the local heat transfer ability of the chain will decrease due to the reduced contact area of the metal. This will lead to a small value of T_c and consequently a large value of criterion C when the heating temperature T_h is constant. Therefore, the criterion C can be used to quantitatively assess the size of a transverse crack occurring in the mooring chain. In the experiment, the values of T_c have been read from the images shown in Figure 7. They are $T_c = 97\,^oC$ for Figure 7a, $85\,^oC$ for Figure 7b, $111.5\,^oC$ for Figure 7c, and $40\,^oC$ for Figure 7d. Substitute the values of T_h and T_c into (2), the criterion C is calculated and the results are shown in Figure 8.

From Figure 8, it is clearly seen that as expected, the value of criterion C does increase gradually with the increasing depth of the crack. This means that the larger the size of the crack, the more significant the influence of the crack tends to be in the heat transfer process. This fully demonstrates that the proposed quantitative evaluation method does work in assessing the crack occurring in the mooring chain.

5 CONCLUSIONS

In order to explore a feasible method for monitoring the health condition of mooring chains, both numerical and experimental researches are conducted in the paper to investigate the potential of thermography technique in detecting transverse cracks occurring in mooring chains. From the work reported above, the following conclusions can be drawn:

1. When the mooring chain is perfect in structural integrity, the temperature is smoothly distributed over the chain except at welds, where the temperature is smaller than in other chain parts due to the higher thermal conductivity of welding materials in comparison of that of the steel material of mooring chain;
2. When the mooring chain is heated from its both ends, a clear concave will appear in the temperature profile in the presence of a transverse crack in the mooring chain. Moreover, the larger the size of the crack, the deeper the concave tends to be. Therefore, the concave in the temperature profile is a good indicator of the crack and its propagation when the mooring chain is heated from the both ends of it;
3. In comparison of the scenario that the mooring chain is heated from the both ends of it, more accurate assessment of the crack can be achieved when the mooring chain is heated from only one end. Experiment has shown that the proposed quantitative evaluation method can successfully predict the presence and growth of the crack when the chain is heated from only one end;
4. The aforementioned numerical and experimental researches have demonstrated that thermography technique does have potential to be applied to condition monitoring mooring chains. However, further research is still required to develop an appropriate method to heat the mooring chain in a completely wet environment.

ACKNOWLEDGEMENT

The work reported in this paper was supported by National Natural Science Foundation of China with the reference numbers of 11772126 and 11632011.

REFERENCES

Amenabar I., Mendikute, A., Lopez-Arraiza, A., Lizaranzu, M. & Aurrekoetxea, J. 2011. Comparison and analysis of non-destructive testing techniques suitable for delamination inspection in wind turbine blades, *Composites Part B: Engineering* 42(5): 1298–1305.

AVT Reliability, https://www.avtreliability.com/media-centre/case-studies/mooring-line-integrity-monitoring, latest access on 20/10/2017.

Blitz, J. 2012. *Electrical and magnetic methods of non-destructive testing*, Springer.

Borgnakke, C. & Sonntag, R.E. 2003. *Fundamentals of thermodynamics*, John Willey & Sons Inc.

Garcia-Martin, J., Gomez-Gil, J. & Vazquez-Sanchez, E. 2011. Non-destructive techniques based on eddy current testing, *Sensors* 11(3): 2525–2565.

Lugsdin, A., Real-time monitoring of FPSO mooring lines, risers. http://www.tritech.co.uk/uploadedfiles/RAMSSeaTech%20editorial.pdf, latest access on 20/10/2017.

Sanjeev, K.V., Sudhir, S.B. & Saleem A. 2013. Review of non-destructive testing methods for condition monitoring of concrete structures, *Journal of Construction Engineering* 2013:1–11.

Seatools, http://www.seatools.com/news/completion-mooring-monitoring-project/?gclid=CI2wpNnVuNYCFYqU7QodQpUEFQ, latest access on 20/10/2017.

TWI, An effective methodology for implementing structural health monitoring and in-service inspection of mooring chains, 2013.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Bayesian updating of stochastic process-based models for corroding gas pipelines based on imperfect inspection information

K. Pesinis & K.F. Tee
Department of Engineering Science, University of Greenwich, UK

ABSTRACT: This paper presents an efficient approach for Bayesian analysis of corroding gas pipelines containing metal-loss corrosion defects subjected to internal pressure. The methodology considers stochastic dependence among individual defects. The Nataf model is used to model the interdependence (correlation) among the defects. The generation of new defects and the growth of existing ones are included in the stochastic modelling, by employing a Non-Homogeneous Poisson Process (NHPP) and a Homogeneous Gamma Process (HGP), respectively. Information on a corroding pipeline obtained through multiple In-Line Inspections (ILIs) is used for the Bayesian updating of stochastic models. The Probability of Detection (PoD) and measurement error of the inspection tools are also taken into consideration. The Bayesian updating is conducted by using Structural Reliability Methods (SRM), a recently proposed method in literature that sets an analogy between Bayesian updating and a reliability problem. The SRM adopted herein is Subset Simulation (SuS) and the whole analysis is referred to as BUS-SuS. The updating is conducted in conjunction with the Data Augmentation (DA) technique which accounts for both detected and undetected defects, by treating the undetected ones as the missing data. Multiple simulated corrosion defects from different ILI inspections are employed for the implementation and validation of the methodology. A parametric study that examines the impact of correlations among defects on the posterior stochastic models is also included in the numerical example.

1 INTRODUCTION

Metal-loss corrosion is the most predominant gradual deterioration process for gas pipelines, based on historical failures. Reliability-based integrity management programs are increasingly adopted by pipeline operators to ensure the safe operation of pipelines against metal-loss corrosion (Khan & Tee 2016). In the literature, stochastic growth models are considered the most efficient modelling option for metal-loss corrosion in gas pipelines (Maes et al. 2009, Zhang & Zhou 2013, Pesinis & Tee 2017). However, a comprehensive analysis requires the modelling of the generation of new crack features too. Moreover, the dependence (or correlation) among individual defects should be considered, to express the spatial dependence among them. This is typically due to the similar corrosive environment, similar pipe properties at the defects' location and the fact that defects are subjected to the same loading conditions (Zhou et al. 2012, Khemis et al. 2016).

In Maes et al. (2009) and Zhang & Zhou (2013) gamma process was employed to characterize the growth of corrosion defects on the pipeline. In Qin et al. (2015) the generation of new metal-loss corrosion defects was taken into consideration in the analysis, using a stochastic process-based model. In all three studies, the hierarchical Bayesian analyses updated the stochastic models based on ILI data by means of Markov Chain Monte Carlo (MCMC) simulation, without considering correlations among defects. In fact, there is only limited information available on modelling correlations of deterioration in energy pipelines (Qian et al. 2013, Zhou et al. 2017) and to the best of the authors' knowledge, none when it comes to Bayesian updating. Furthermore, when it comes to MCMC simulation, this is thought to involve an underlying uncertainty around ensuring the final samples have reached the posterior distribution and has limited capacity in ultimately quantifying small failure probabilities (Straub et al 2016). An alternative method to MCMC is Bayesian Updating with Structural reliability methods (BUS), which sets an analogy between Bayesian updating and a reliability problem (Straub & Papaioannou 2014). This formulation enables the use of established SRM to conduct the Bayesian updating. The SRM adopted in this study is Subset Simulation (SuS).

Thus, hierarchical Bayesian updating is conducted herein, by using BUS-SuS in conjunction with the Data Augmentation (DA) technique. Simulated data corresponding to a gas pipeline

subjected to internal pressure loading are used to illustrate and validate the proposed methodology. The growth of multiple metal-loss corrosion defects is characterized through adopting a Homogeneous Gamma Process (HGP) model and incorporating it into the hierarchical Bayesian framework based on multiple ILI data, accounting for the associated measurement errors as well. Furthermore, the interdependence among different defects is considered using the Nataf model, also known as the Gaussian copula. The generation of new metal-loss corrosion defects is realised in the analysis by means of Non-Homogeneous Poisson Process (NHPP). At the end, the impact of different dependence scenarios is investigated.

The contributions of this paper include first the development of a robust hybrid hierarchical Bayesian framework for the updating of both generation and growth stochastic model, with respect to metal-loss corrosion. The proposed methodology, based on BUS-SuS and DA technique, eliminates the uncertainty regarding whether the final samples have reached the posterior distribution. Then, the methodology efficiently accounts for different spatial correlation scenarios among individual defects, which has not been conducted before in Bayesian updating for energy pipelines. The contents of this paper are structured as follows. Section 2 defines the uncertainties concerning the inspection data. The formulations of the NHGP and HGP models for the defect generation and growth, respectively, are presented in Section 3. The hierarchical Bayesian method for updating the model parameters by means of BUS-SuS and DA are presented in Section 4. An application based on simulated ILI data corresponding to a gas pipeline is presented in Section 5, in order to illustrate and validate the proposed methodology. Finally, conclusions are drawn in Section 6.

2 UNCERTAINTIES OF INSPECTION DATA

2.1 *Measurement error*

Inspection data for corrosion defects on gas pipelines are subject to random measurement errors because of the accuracy limitations of the ILI tool. Thus, the measured size of a defect is expected to differ from its actual size. Based on a measured size, the actual defect size can be evaluated through the following (Qin et al. 2015, Zhang and Zhou 2013):

$$d_{ki} = a_i + b_i c_{ki} + \varepsilon_{ki} \tag{1}$$

for a corrosion defect k (k = 1, 2,…, n) that is detected in the ith (i = 1, 2, …, m) inspection, where a_i and b_i denote the constant and non-constant biases, respectively, associated with the defect depth and ε_{ki} is the random scattering error associated with the measured depth. It should be noted that if $a_i = 0$ and $b_i = 1$ the tool is unbiased. The random scattering errors associated with different defects for a given ILI tool were assumed to be mutually independent and those associated with different ILI tools for a given defect were also assumed to be mutually independent (Stephens & Nessim 2006).

Given a measured defect k, the probability distribution of the random scattering error ε_{ki} in the ith inspection can be determined from tool specifications that characterize the sizing accuracy in terms of the probability that the error will fall within prescribed bounds e_{min} and e_{max}. Then, the mean value (μ) and standard deviation (σ) of the random scattering error can be determined, for any distribution type. Assuming that the mean and standard deviation of the error are from a multivariate normal distribution as follows:

$$\mu = (e_{\min} + e_{\max})/2 \tag{2}$$

$$\sigma = (e_{\max} - \mu)/\left[\Phi^{-1}\left(\frac{1+p_e}{2}\right)\right] \tag{3}$$

where Φ^{-1} is the inverse standard normal distribution function (Stephens & Nessim, 2006).

2.2 *Probability of detection*

Probability of detection (PoD) is related to the capacity of an ILI tool to detect a metal-loss corrosion defect (Zhang & Zhou 2013). It can be defined as a function of the defect size together with constants that refer to the tool's accuracy. The following exponential function can be defined (Qin et al. 2015):

$$\text{PoD} = 1 - e^{-q(c-c_{th})} \quad \text{for } c \geq c_{th} \tag{4}$$

The value of q refers to the inherent tool detection capacity and can be specified from vendor-supplied tool specifications. In addition, c denotes the actual defect depth and c_{th} is the detection threshold, i.e. the minimum detectable defect depth.

3 DEFECT GENERATION AND GROWTH

3.1 *Stochastic defect generation*

The non-homogeneous Poisson process stochastic process-based model was employed for the generation of new defects on a pipe segment over time (Qin et al. 2015, Tee & Pesinis 2017).

The total number of defects in a time interval [0, t] is assumed to follow a Poisson distribution with a probability mass function (PMF), $f_p(N(t)|\lambda(t))$:

$$f_p(N(t)\Big|\lambda(t)) = \frac{(\lambda(t))^{N(t)} e^{-\lambda(t)}}{N(t)!} \left(t \geq 0\right) \tag{5}$$

$N(t)$ is the expected number of defects generated over the time interval [0, t] and $\lambda(t)$ is the instantaneous generation rate:

$$\lambda(t) = \int_0^t \lambda_o \tau^{\delta} d\tau \tag{6}$$

with the parameters λ_0 and δ quantified from the ILI data.

In case of m ILIs that have taken place over a certain period, it was assumed that each inspection can detect new and existing defects, in terms of their spatial positions (Qin et al. 2015). The total number of defects X_i on the time of the ith inspection (i = 1, 2, ... m) t_i, can be divided into those defects that have initiated prior to the (i-1) th inspection, X_i^0 and those ones that have initiated between the (i-1)th and ith inspections X_i^s. The value of X_i^s can be evaluated by assuming that it follows a Poisson distribution with PMF given by:

$$f_p(X_i^s\Big|\lambda_0, \delta) = \frac{(\lambda_i)^{X_i^s} e^{-\lambda_i}}{X_i^s!} \tag{7}$$

The detected defects are typically less than the actual ones and that is due to the imperfect detectability of the ILI tool. If X_i^{sd} and X_i^{su} are the detected and undetected values, respectively, of the total number X_i^s. Following the Poisson splitting property (Qin et al. 2015, Kulkarni 1995), X_i^{sd} and X_i^{su} are assumed to follow the Poisson distributions with the respective PMFs:

$$f_p(X_i^{sd}\Big|\lambda_0, \delta) = \frac{(\overline{\text{PoD}}_i \lambda_i)^{X_i^{sd}} e^{-\overline{PoD}_i \lambda_i}}{X_i^{sd}!} \tag{8}$$

$$f_p(X_i^{su}\Big|\lambda_0, \delta) = \frac{[(1-\overline{\text{PoD}}_i)\lambda_i]^{X_i^{su}} e^{[-(1-\overline{\text{PoD}}_i)\lambda_i]}}{X_i^{su}!} \tag{9}$$

where PoD_i with the overline refers to the average PoD that corresponds to the X_i^s defects and can be evaluated from:

$$\overline{\text{PoD}}_i = \int \text{PoD}(x) f_{X_i^s}(x) dx \tag{10}$$

where $f_{Xi}^s(x)$ denotes the probability density function (PDF) of the depths of the X_i^s defects at time t_i.

3.2 *Stochastic defect growth*

The depth $c(t)$ of a defect at year t, (i.e. $t = 0$ corresponds to the installation year of the pipeline) was assumed to follow a homogeneous gamma process (Zhang & Zhou 2014). The PDF of $c(t)$ is gamma distributed:

$$f_{c(t)}(c(t)\Big|\gamma(t), \omega) = \frac{\omega^{\gamma(t)} c(t)^{\gamma(t)-1} e^{-c(t)\omega}}{\Gamma(\gamma(t))} \mathrm{I}_{(0,\infty)}(c(t)) \tag{11}$$

with $\gamma(t)$ representing the time-dependent shape parameter which is a linear function with time:

$$\gamma(t) = \alpha_1 (t - t_0) \tag{12}$$

where t_0 denotes the initiation time of a defect. In addition, ω ($\omega > 0$) is the time-independent rate parameter or the inverse of the scale parameter, $\Gamma(\bullet)$ denotes the gamma function and $I_{(0,\infty)}(c(t))$ is the indication function (i.e. it is equal to unity if $c(t) > 0$ and zero otherwise). The quantity α_1/ω represents the mean of the defect depth. The parameters α_1 were assumed to be common for all defect depths, whereas t_0 and ω were assumed to be defect-specific. t_{0z} and ω_z were defined with respect to the initiation time and rate parameter for the zth defect, as opposed to the index k that was used to characterise detected defects, i.e. z refers to the total number of defects (both detected and undetected).

The growth of the jth defect among the (i-1)th and ith inspections Δc_{iz} is gamma distributed with a time-dependent shape parameter $\Delta\gamma_{iz}$ as follows:

$$\Delta\gamma_{iz} = \alpha_1 (t_i - t_{0z}) \quad \left(i = y\right) \tag{13}$$

where y is a quantity that satisfies $t_{y-1} < t_{iz} < t_y$. The depth of each defect z at the time of each inspection i c_{iz}, is the sum of consecutive incremental depths between the (i-1)th and ith inspections:

$$c_{iz} = c_{i-1,z} + \Delta c_{iz} \tag{14}$$

4 BAYESIAN ANALYSIS

4.1 *Likelihood functions*

Given the number of detected defects in a total of m inspections and assuming a defect k first detected in the ith (i = 1, 2, … or m) inspection, $\mathbf{d}_k = (d_{ik}, d_{i+1,k}, \ldots, d_{i+y,k}, \ldots, d_{mk})'$ and $\mathbf{c}_k = (c_{ik}, c_{i+1,k}, \ldots, c_{i+y,k}, \ldots, c_{mk})'$ can be defined that denote the vector of the actual depths for defect k and the vector of the corresponding ILI-reported depths of defect

k, respectively. Considering the measurement error, the likelihood of $\boldsymbol{d}_k$ conditional on $\boldsymbol{c}_k$ can be defined as follows:

$$L(\mathbf{d}_k|\mathbf{c}_k) = (2\pi)^{-\frac{(m-i+1)}{2}} \left|\Sigma_{\mathbf{E}_k}\right|^{-\frac{1}{2}} \exp[-\frac{1}{2}(\mathbf{d}_k - \alpha - \mathbf{bc}_k)' \Sigma_{\mathbf{E}_k}^{-1}(\mathbf{d}_k - \alpha - \mathbf{bc}_k)] \quad (15)$$

where $\boldsymbol{\alpha} = (a_i, a_{i+1},\ldots,a_m)'$ and $\mathbf{b}$ is an $m - i + 1 \times m - i + 1$ diagonal matrix with the yth element equal to a_{i+y}.

The updating of the number of detected defects inherently contains the possibility that some of the newly detected defects in the ith inspections might have been generated prior to the previous inspection (i-1)th but remain undetected until then. However, it was assumed that the newly detected depths in the ith inspection, have initiated between the (i-1)th and ith inspections. That is a conservative approach, since it finally leads to overestimation of the instantaneous rate of the generation model (Qin et al. 2015). The likelihood function for the newly detected defects in m inspections X_i^{sd} (i = 1, 2, …, m) was defined as follows:

$$L(X_i^{sd}|\lambda_0,\delta) = \prod_{i=1}^{m} \frac{\left[\overline{PoD}_i \lambda_0 (t_i^{\delta} - t_{i-1}^{\delta})\right]^{X_i^{sd}}}{S_j^{kde}} \exp\left\{-\left[\overline{PoD}_i \lambda_0 (t_i^{\delta} - t_{i-1}^{\delta})\right]\right\} \quad (16)$$

where the PoD_i with the overline, includes both detected and undetected defects in the ith inspection. The undetected defects were treated as the missing data and the data augmentation (DA) technique was employed to incorporate these in the Bayesian analysis (Tanner & Wong 1987). This technique is described in more detail in Section 4.4. Thus, the PoD serves as a link in the Bayesian updating, between the generation and growth models.

4.2 *Prior distributions*

The gamma distribution was selected as the prior distributions of α_1, ω_k parameters of the growth model and also for the λ_0 and δ parameters of the generation model. The truncated normal distribution with an upper bound equal to the time of the inspection that each defect is detected for the first time and a lower bound equal to the time of the previous inspection was chosen as the prior distribution of t_{0j}. The prior distributions for t_{0k}, α_1 and ω_k for individual defects were assumed to be mutually independent. The shape (rate) parameters of the gamma prior distributions for α_1, ω_k and λ_0, δ were defined as $\xi_1(\xi_2)$, $\kappa_1(\kappa_2)$ and $\varphi_1(\varphi_2)$, $\zeta_1(\zeta_2)$, respectively. Dependence among defect behavior is modelled through correlation coefficients among the HGP model parameters. In this study, the model parameters ω_k were considered equi-correlated among all defects (k = 1, 2,…, n) with correlation coefficients θ. The correlation coefficient represents the statistical dependence due to common fabrication quality, common material characteristics and common loading characteristics. The joint distribution of all model parameters is subsequently modelled through a Nataf (Gaussian copula) model (Qian et al. 2013).

4.3 *Updating with BUS-SuS and DA technique*

Bayes' rule enables updating the joint prior distribution of the parameters $\boldsymbol{\eta}$ that correspond to both generation and growth models, based on the inspection data $\mathbf{C}$. The prior distribution is $f(\boldsymbol{\eta})$ and is converted into a 'posterior' distribution $f'(\boldsymbol{\eta}|\mathbf{C})$, based on the following (Straub & Papaioannou 2014):

$$f'(\boldsymbol{\eta}|\mathbf{C}) = \frac{L(\mathbf{C}|\boldsymbol{\eta}) f(\boldsymbol{\eta})}{\int_{\Omega} L(\mathbf{C}|\boldsymbol{\eta}) f(\boldsymbol{\eta})} \quad (17)$$

where $\boldsymbol{\Omega}$ denotes the domain of definition of $\boldsymbol{\eta}$.

In this study, the Bayesian updating was conducted with BUS-SuS along with the aforementioned DA technique which constitutes a robust hybrid simulation technique, in an effort to numerically evaluate the complex denominator of Eq. 17, since an analytical evaluation would not be feasible. Bayesian updating with BUS-SuS is an extension of the classical rejection sampling approach to Bayesian analysis (Straub & Papaioannou, 2014, Straub et al. 2016). It is more advantageous over MCMC, which is typically used in the energy pipeline literature, since it diminishes the uncertainty around ensuring that the final samples have reached the posterior distribution and therefore provides more accurate samples of the posterior. The simple rejection sampling algorithm however, is quite inefficient and therefore subset simulation is employed that can compute very small probabilities. Thus, the Bayesian updating becomes very efficient, while maintaining the advantages of the simple rejection sampling algorithm (Straub & Papaioannou 2014, Straub et al. 2016).

A set of random samples of the prior distribution $\boldsymbol{\eta}$ are generated and then accepted as samples of the posterior distribution. The prior samples are accepted with a probability $p = cL(\boldsymbol{\eta})$, where c is a positive constant that ensures $cL(\boldsymbol{\eta}) \leq 1$ for all $\boldsymbol{\eta}$. If c is selected too small, this might decrease the efficiency of the method, since the acceptance

rate is a linear function of c. If c is too large, the resulting samples might not follow the posterior distribution. An optimal choice of c is considered to be (Straub & Papaioannou 2014):

$$c = \frac{1}{\sup L(\boldsymbol{\eta})} \tag{18}$$

For a single ILI with error ε, $\sup L(\boldsymbol{\theta})$ is equal to maximum of the PDF of ε. For the multiple ILIs of this study, $\sup L(\boldsymbol{\eta})$ was thought to be equal to the maximum of the multivariate PDF of the error ε_{ki}. This is spatially independent, following a multivariate normal distribution with a zero mean and known covariance matrix Σε associated with the total number of inspections m.

After defining the augmented outcome space [$\boldsymbol{\eta}$; p] and the domain $\{p \le cL(\boldsymbol{\eta})\}$ a structural reliability problem results, with the posterior distribution obtained by censoring the joint distribution of p and $\boldsymbol{\eta}$ to $\{p \le cL(\boldsymbol{\eta})\}$ and marginalizing $\boldsymbol{\eta}$:

$$f'(\boldsymbol{\eta}|\mathbf{C}) \propto \int_0^1 I([\boldsymbol{\eta}, p] \in \{p \le cL(\boldsymbol{\eta})\}) f(\boldsymbol{\theta}) dp \tag{19}$$

where I is an indicator function which takes value 1 if $\{p \le cL(\boldsymbol{\eta})\}$ and 0 otherwise (p is a standard uniform random variable). In the structural reliability convention, the domain $\{p \le cL(\boldsymbol{\eta})\}$ describes an observation event (in terms of the Bayesian updating) through a limit state function r, such that it corresponds to a respective domain $\{r(p,\boldsymbol{\eta}) \le 0\}$:

$$r(p, \boldsymbol{\eta}) = p - cL(\boldsymbol{\eta}) \tag{20}$$

The probability of the observation event can be efficiently computed by a robust method such as SuS. It is typically efficient to apply SuS in the standard normal space, therefore in this study the outcome space of the original random variables p and $\boldsymbol{\eta}$ was transformed to a space with independent standard normal random variables $\mathbf{V}$ (p and $\boldsymbol{\eta}$ are independent and thus they can be transformed separately). The transformed observation domain is a function U:

$$U(\mathbf{v}) = v_0 - \Phi^{-1}(cL(\mathbf{T}(\mathbf{v}))) \tag{21}$$

where Φ is the standard normal CDF.

An algorithm based on SuS can generate samples from the transformed observation domain subsequently. SuS is a well-established method and a detailed step-by-step presentation of the method can be found in Au & Beck (2001) and Straub & Papaioannou (2014). It should be noted that the final finding of interest in this study is the samples that belong to the posterior distribution. As a result, a final step should be defined. In that extra step, J additional samples conditional on Y from the output domain $\{p \le cL(\boldsymbol{\eta})\}$ are produced, so that a total number of (at least) Z samples from the posterior distribution is obtained.

Both the detected and undetected defects were considered for the Bayesian updating in this study. The depths of the detected defects were related to the ILI-reported depths through the likelihood function given by Equation 15, while the real depths of the undetected defects were treated as the missing data and imputed using the DA technique (Tanner & Wong 1987). As a result, the joint posterior distribution of the model parameters was evaluated from the depths of the total defect population (both detected and undetected). It should be noted that it is straightforward to couple DA in BUS-SuS. DA is an iterative process that in each iteration contains two steps; the imputation and the posterior step. The imputation step generates the samples of the missing data from its corresponding probabilistic distribution conditional on the current state of model parameters, whereas the posterior step is used to generate new samples of model parameters, from their corresponding posterior distributions conditional on both the observed and missing data. More information and details of the DA process can be found in Tanner & Wong (1987), Rubin (2004) and Little & Rubin (2014).

5 APPLICATION

5.1 *Case study*

Simulated data are used to illustrate and validate the methodology. A gas pipe segment is considered, with 609 mm diameter and 7.9 mm wall thickness, from pipe material API 5 L grade X52. Five future ILIs were assumed to take place on years t = {4, 7, 9, 11, 15} after the installation of the pipeline. The growth of individual defects was assumed to be a linear random variable corrosion growth model, with the rate of defect growth for each defect selected from a uniform distribution with lower bound 0.29 mm/year and upper bound 0.50 mm/year. Metal-loss corrosion defects were assumed to initiate based on the NHHP model by deterministically setting the parameters $\lambda_0 = 0.2$ and $\delta = 1.0$ and to grow with the aforementioned individual random rate.

Furthermore, a depth detection threshold of >1 mm deep was considered and a PoD of 98.3%. The measurement error was assumed to fall within the bounds of ±1 mm with 90% probability. Therefore $q = 4.0785$ in Equation 4 and measurement error follows a multivariate normal distribution with mean zero and standard deviation $\sigma_r = 7.29\%wt$. The constant and non-constant

biases included in the measurement error given by Equation 1 were assumed to be equal to zero and unity for all inspections respectively, while the random scattering errors associated with different inspections were defined as mutually independent with the same standard deviation of unity. Table 1 summarises the simulated data.

5.2 *Validation of Bayesian formulation*

The Bayesian analysis was carried out for the defect generation and growth models, based on the ILI data. The shape and scale parameters of the gamma prior distributions for the parameters of the models α_1, ω_k and λ_0, δ were all set to unity. A total of 20,000 samples were generated to evaluate the probabilistic characteristics of the model parameters. The means, medians and standard deviations of the posterior marginal distributions of the parameters λ_0, δ of the NHPP model, together with the average PoD for the defect generated prior to the first inspection year and among the rest of inspection years respectively, are summarised in Table 2. Furthermore, the same information for the HGP parameters α_1 that are common for all defects is also presented in Table 2. When compared with the actual (i.e. simulated) values, the posterior mean and median values of δ and λ_0 are considered to be in good agreement, which validates the Bayesian formulation described in Section 4. This is further illustrated in Figure 1 and Figure 2 for the NHPP and HGP models, respectively.

In Figure 1, results for a period from when the generation of new defects initiates to the last inspection year (i.e. 15), are presented. The mean values of the number of generated defects were estimated, with the values λ_0, δ set equal to their corresponding posterior medians. For comparison, $\lambda(t)$ evaluated by the actual values of λ_0 and δ, the simulated total number (i.e. detected and undetected) of defects, along with the simulated numbers of detected defects on the five ILIs are also illustrated in Figure 3 $\lambda(t)$ agrees with the actual mean very well, which proves the validity of the Bayesian model.

Next, the depths of the detected defects at year 15 were evaluated and compared with the corresponding actual defect depths. Each predicted defect depth was set equal to the corresponding mean derived from the HGP growth model, with the parameters of the model (i.e. α_1, ω_k and t_{0k}) equal to the respective posterior medians. In the Bayesian updating, constant c of Equation 18 was evaluated on each inspection year, based on the procedure described in Section 4.4 and the respective

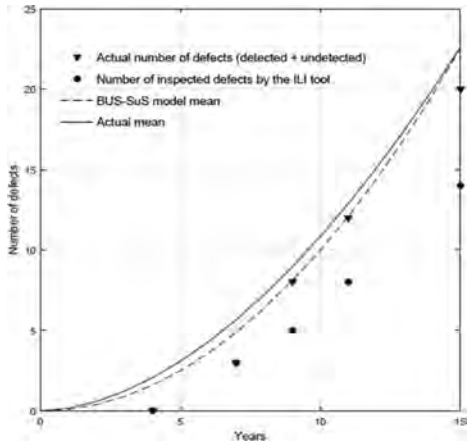


Figure 1. Comparison of predicted and actual number of detected and undetected defects.

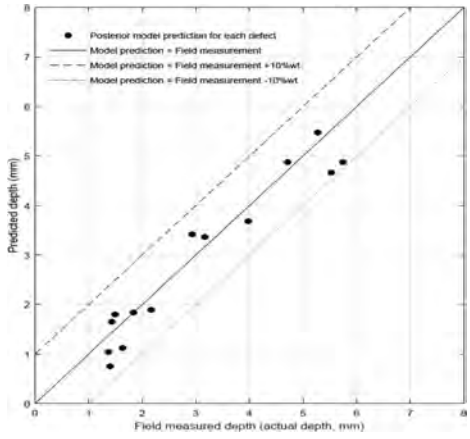


Figure 2. Comparison of the predicted depths and actual depths of defects at year 15.

Table 1. Summary of the input simulated ILI data.

Time of inspection	Year 4	Year 7	Year 9	Year 11	Year 15
Number of detected defects	0	3	5 (3)	8 (3)	14 (6)
Measured depth (mm) (Mean)	0	1.32	2.31	2.66	2.74
(Standard deviation)	0	0.07	0.86	1.14	1.48

Table 2. Posterior data of the model parameters.

Parameter	NHPP		PoD				
	λ_0 (0.2)	δ (1.0)	1	2	3	4	5
Mean	0.30	0.81	0.03	0.99	0.41	0.72	0.73
Median	0.30	0.80	0.03	0.97	0.41	0.72	0.73
Std	0.28	0.17	0.08	0.12	0.07	0.07	0.07

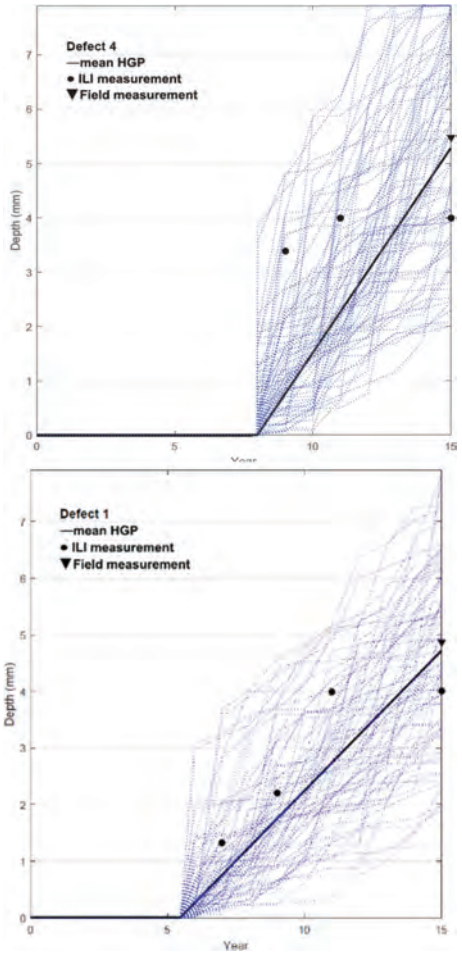


Figure 3. Predicted growth paths of defects 1 and 4 based on the HGP model.

values are illustrated in Table 3. From Figure 2 it can be observed that the model predictions are in good agreement with actual values. In fact, all defect depths fall within the range of ±10%wt of the corresponding actual depths, which is a commonly adopted confidence interval for ILI tools' accuracy in industry (Zhang & Zhou 2013). In summary, the methodology is validated against the corresponding actual data and its accuracy is verified.

For illustration purposes, two defects were chosen for analysis and their means were estimated, based on the results of BUS-SuS for the same 15-year period. The mean and standard deviation of defect depths were defined as $\alpha_1(t-t_{0k})/\omega_k$ and $(\alpha_1(t-t_{0k})/\omega_k^2)^{0.5}$ respectively for each defect, based on the HGP corrosion growth model. The parameters α_1, ω_k and t_{0k} were set equal to the deterministic median values from their corresponding marginal

Table 3. Constant c values for different ILI years.

Time of inspection	Year 4	Year 7	Year 9	Year 11	Year 15
Constant c	0	6.2796	4.3460	3.0084	2.0840

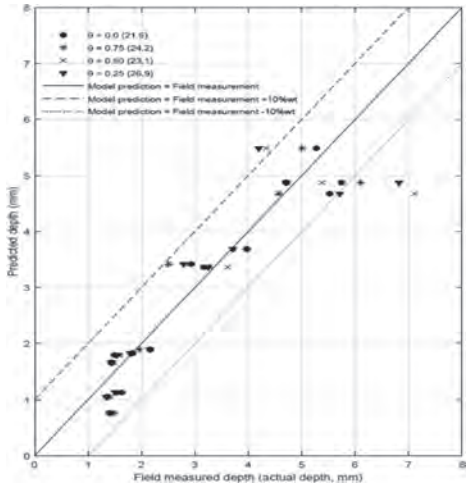


Figure 4. Comparison of the predicted depths and actual depths at year 15 for different correlation coefficients among defects.

posterior distributions derived from BUS-SuS. Also, 80 random realisations of the HGP model are illustrated for each defect. It is observed that for both defects, the BUS-SuS predicted depths and actual depths are in very good agreement, even though the ILI data is not in great proximity with the actual depths. Thus, the proposed model can account successfully for the imperfections of the inspection data and capture the actual initiation times and growth rates of the defects.

Furthermore, in order to investigate the effect of the correlations of defects on the growth model, different dependency scenarios were examined. It was assumed that c_{ik} $(i = 1, 2, ..m)$, $(k = 1, 2,\ldots, n)$ are identically distributed and equi-correlated with the corresponding coefficient θ, set equal to 0.25, 0.5, 0.75 and 0.0 (i.e. 0.0 corresponds to independent and identically distributed samples).

Figure 4 compares the predictions of the growth model (i.e. HGP) at year 15 corresponding to different correlation scenarios. The Mean Squared Error of Prediction (MSEP), defined as:

$$1/n\sum_{k=1}^{n}(c_{pr}-c_{ar})^2 \qquad (22)$$

quantitatively evaluated the predicting accuracy of the growth model, where c_{pr} and c_{ar} denote the predicted

and actual depths of the *k*th defect. The MSEP values are illustrated in brackets on Figure 4 and it is given that the higher the accuracy the lower the corresponding MSEP value is expected to be. It is observed that the most accurate correlation scenario is $\theta = 0.0$. Nevertheless, there is not a clear tendency indication with respect to different correlation scenarios, by either the depth predictions or the MSEP results. In fact, for almost half defects, the depth predictions are negligibly different among correlation scenarios. However, for all correlation scenarios the methodology proposed for updating of the growth model is validated against the actual data. It should be noted that for all correlation scenarios the posterior NHPP generation model results remained unchanged. Therefore, it is thought that correlations do not impact the updating of the methodology significantly, but the selection of a specific correlation structure is likely to affect the accuracy of the growth model upon updating.

6 CONCLUSIONS

A stochastic process-based hierarchical Bayesian methodology was proposed for corrosion management of gas pipelines. The metal-loss corrosion defect generation was characterised by a NHPP and the growth by a HGP. Furthermore, the Nataf model was used for the spatial dependence among defects. The hierarchical Bayesian framework proposed, captured the imperfect detectability of the ILI tool as defined by the PoD and also the measurement errors associated with the ILI data. The Bayesian updating was performed by employing BUS-SuS in conjunction with the data augmentation technique.

The proposed model was applied on simulated defects and multiple ILI inspections. Different defects were assumed identically distributed and equi-correlated with the coefficient θ equal to 0.25, 0.5, 0.75 and 0.0. Results from the Bayesian updating of the generation and growth models, indicate that the predicted overall defect population corresponding to the base case (i.e. $\theta = 0.0$) is in good agreement with the actual defect population, which validates the proposed methodology. The other three correlation scenarios also lead to predictions of good agreement with the actual data and therefore the Bayesian methodology is accurate irrespective of the correlation scenario. There are some discrepancies in the accuracy of predictions among the different scenarios nonetheless, identified by the MSEP. The most accurate among all four scenarios was the one with θ equal to 0.0.

REFERENCES

Au, S.-K. & Beck, J.L. 2001. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4): 263–277.

Khan, L. & Tee, K.F. 2016. Risk-Cost Optimization of Buried Pipelines Using Subset Simulation. *Journal of Infrastructure Systems, ASCE*, 22(2): 04016001.

Khemis, A. Hacene-Chaouche, A. Athmani, A. & Tee, K.F. 2016. Uncertainty Effects of Soil and Structural Properties on the Buckling of Flexible Pipes Shallowly Buried in Winkler Foundation. *Structural Engineering and Mechanics*, 59(4): 739–759.

Kulkarni, VG. 1995. *Modeling and Analysis of Stochastic Systems*. Chapman & Hall, Ltd., London, UK.

Little, R.J.A. & Rubin, D.B. 2014. *Statistical analysis with missing data*. John Wiley & Sons.

Maes, M.A. Faber, M.H. & Dann, M.R. 2009. Hierarchical modeling of pipeline defect growth subject to ILI uncertainty. In: *ASME 2009 28th International Conference on Ocean, Offshore and Arctic Engineering*. American Society of Mechanical Engineers, 375–384.

Pesinis, K. & Tee, K.F. 2017. Statistical model and structural reliability analysis for onshore gas transmission pipelines. *Engineering Failure Analysis*, 82: 1–15.

Qian, G. Niffenegger, M. Zhou, W. & Li, S. 2013. Effect of correlated input parameters on the failure probability of pipelines with corrosion defects by using FITNET FFS procedure. *International Journal of Pressure Vessels and Piping*, 105: 19–27.

Qin, H. Zhou, W. & Zhang, S. 2015. Bayesian inferences of generation and growth of corrosion defects on energy pipelines based on imperfect inspection data. *Reliability Engineering & System Safety*, 144: 334–342.

Rubin, D.B. 2004. *Multiple imputation for nonresponse in surveys.* John Wiley & Sons.

Stephens, M. & Nessim, M. 2006. A comprehensive approach to corrosion management based on structural reliability methods. In: *2006 International Pipeline Conference*. American Society of Mechanical Engineers, 695–704.

Straub, D. & Papaioannou, I. 2014. Bayesian updating with structural reliability methods. *Journal of Engineering Mechanics*, 141(3): 4014134.

Straub, D. Papaioannou, I. & Betz, W. 2016. Bayesian analysis of rare events. *Journal of Computational Physics*, 314: 538–556.

Tanner, M.A. & Wong, W.H. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398): 528–540.

Tee, K.F. & Pesinis, K. 2017. Reliability prediction for corroding natural gas pipelines. *Tunnelling and Underground Space Technology*, 65: 91–105.

Zhang, S. & Zhou, W. 2013. System reliability of corroding pipelines considering stochastic process-based models for defect growth and internal pressure. *International Journal of Pressure Vessels and Piping*, 111: 120–130.

Zhang, S. & Zhou, W. 2014. Cost-based optimal maintenance decisions for corroding natural gas pipelines based on stochastic degradation models. *Engineering Structures*, 74: 74–85.

Zhou, W. Hong, H.P. & Zhang, S. 2012. Impact of dependent stochastic defect growth on system reliability of corroding pipelines. *International Journal of Pressure Vessels and Piping*, 96: 68–77.

Zhou, W. Xiang, W. & Hong, H.P. 2017. Sensitivity of system reliability of corroding pipelines to modeling of stochastic growth of corrosion defects. *Reliability Engineering & System Safety*, 167: 428–438.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Effect of the manufacturing defects on the reliability of disposal packages for high level radioactive waste

A. Persoons, P. Beaurepaire & A. Chateauneuf
CNRS, SIGMA Clermont, Institut Pascal, Université Clermont Auvergne, Clermont-Ferrand, France

F. Bumbieler
Andra, Rue Jean Monnet, Châtenay-Malabry, France

ABSTRACT: In the deep geological disposal facility for radioactive waste to be built in France, it is planned to encapsulate high level waste in carbon-steel overpacks before inserting them into horizontal micro-tunnels. The main function of the overpack is to isolate the waste from the environment long enough for its radio-toxicity and heat to significantly decrease. Several phenomena affect the overpack during its lifetime. The prediction and modelling of these phenomena are subjected to many uncertainties due to their complexity and their extrapolation over long time periods (several centuries). This paper presents an overall framework for the estimation of the failure probability of the overpacks over time. The failure scenario considered is the fracture of the overpack. The analysis is based on a parameterized finite element model taking into account the evolution of the geometry due to the non-uniform corrosion process, the evolution of the mechanical loading and their variability. In addition the manufacturing process may induce defects which are modeled as cracks and included in the model of uncertainty. The whole process is run for several time steps until the complete failure of the overpack. The reliability estimation is based on a two-level Monte Carlo analysis and according to a fracture mechanics criterion based on a probabilistic critical stress intensity factor.

1 INTRODUCTION

Nuclear plants are one of the principal sources of electricity in many countries and are responsible for most of the electricity production in France. Nuclear industry produces 60% of the French radioactive waste, 27% are from research activities, 9% from defense, 3% from industry and 1% from medical activities (ANDRA 2015). Radioactive waste are categorized by their level of hazardousness. The European council directive 2011/70/EURATOM states that, at this time, deep geological storage is the safest option as the end point of management of high-level waste. Cigeo is the French project of deep geological disposal facility for radioactive waste. Safety and reliability are major goals of the project and it is necessary to guarantee that the radioactive material will not be in contact with water during the early time of storage. In the project, high level waste are planned to be conditioned in carbon-steel overpacks before being inserted into horizontal micro-tunnels. The main function of the overpack is to isolate the waste from the environment long enough for its radio-toxicity and heat to significantly decrease. This period is estimated at 500 years. Therefore, the reliability of overpacks is an important element in the safety assessment of the overall Cigeo project.

For this project, studies are lead on the evolution of corrosion processes (type and kinetics), the properties of the steel grades and their behavior in repository conditions, the thermos-hydro-mechanical properties and behavior of the rock, the prediction and characterization of failure mode of the steel structures (the buckling of the lining, cracking of the overpack). Reliability methods have not been explored yet in this project although they are already widely used in other fields with comparable constraints such as oil and gas industry (Dundulis et al. 2016). The interest of the study is to estimate the failure probability of the overpacks taking into account the aging process and uncertainties of the system. The failure probability is investigated considering a fracture mechanics criterion to take into account an eventual Stress Corrosion Cracking (SCC) sensitivity of the steel.

The analysis presented in this paper proceeds as follows: a realization of the random variables is generated and the finite element analysis of the degraded model is performed for all the time steps over the lifetime. The stress field is recovered and the stress intensity factor is calculated at each time

step for numerous randomly simulated cracks. The reliability of the system over time is estimated by a Monte-Carlo method using two levels of simulations. At the first level, a limited sample of the random variables affecting the finite element analysis is evaluated. For each of them (second level) the stress intensity factor of the random crack requires only post processing. This strategy allows us to analyze a large number of crack configuration with moderate numerical efforts. The reliability is then evaluated for every time step facing a random critical stress intensity factor criterion.

2 STATEMENT OF THE PROBLEM

The repository will take place 500 meters below ground in an impermeable argillaceous rock (Callovo-Oxfordian claystone) able to contain radioactivity over a very long period (ANDRA 2005). High Level Wastes (HLW) disposal cells consist of horizontal micro-tunnels cased with a steel liner (Fig. 1). A cement-based filling material that imposes corrosion-limiting environmental conditions is also injected between the rock and the lining. HLW are vitrified and sealed in stainless steel containers conditioned in low-alloy overpacks which are inserted in the lining (Figs. 2–3). The whole system may be compared to Russian nesting dolls, in the center is the waste, then the container, the overpack, the lining, the filling material and finally the rock.

The corrosion and the evolution of the mechanical loading could affect the structural integrity and the reliability of the overpack. Therefore, it is important to describe and model the in-situ conditions of the system along its lifetime, the possible variations of this scenario and the possible failure causes.

Overpacks are made of low-alloy steel to promote general corrosion in repository conditions. Water coming from the host rock will progressively fill the HLW cell implying that two corrosion rates (either under or above water) affect the overpack.

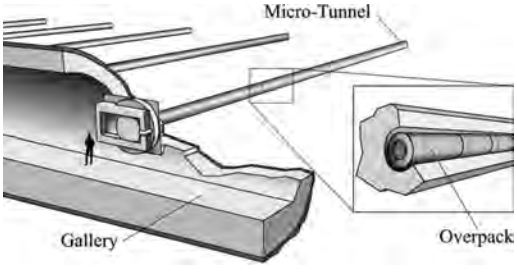


Figure 1. Overview of the storage facility.

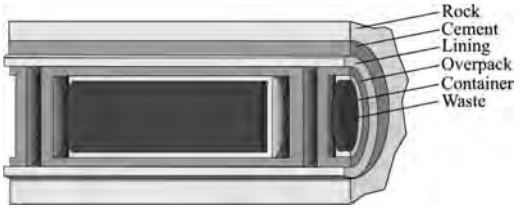


Figure 2. Schematic cross-section of the system studied.

Figure 3. R7-T7 vitrified waste disposal package.

These two corrosion rates lead to non-uniform corroded thickness around the overpack.

The evolution of the water level over time is the equilibrum result of several hardly predictible phenomena, involving the saturation and water flow in the rock and the gas production by the corrosion processes. Therefore, uncertainties remain about the long term stabilized height of the water level. A liquid water extraction system avoids the water filling of the micro-tunnel during the 100 years operating phase, so that the water level is null for this period. The equilibrium of the gas and liquid phases has been studied for intermediate level waste disposal facilities of the same project i.e. the same rock in (Croisé et al. 2011; Brommundt et al. 2014). It has been stated that once the repository is sealed, the rock saturation will increase and water will flow in. After a while, an equilibrium between water and gas pressure will be reached and the water level will stabilize. This study focuses only on the case of the water level stabilizing at an intermediate value filling half of the micro-tunnel (Fig. 4).

Corrosion also affects the steel lining which is under the pressure of the rock. The reduction of its operational thickness will eventually cause buckling, it will come into contact with the overpack

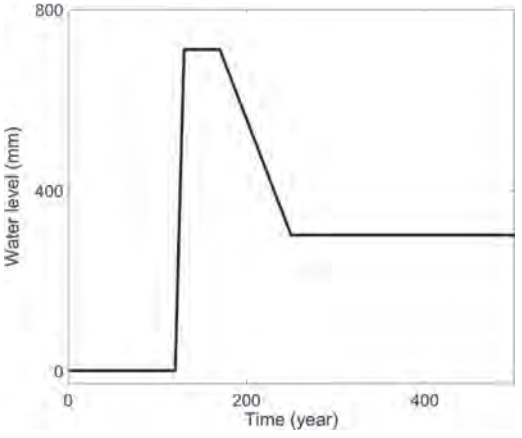


Figure 4. Water level over time.

and transmit the pressure. The contact area will progressively expand until it completely covers the overpack which will be directly subjected to the pressure of the rock. Once the micro-tunnel is sealed, it is also subjected to a fluid pressure.

The overpack is made of 3 forged parts welded by electron beam. Each overpack is inspected for defect from either the welding or the forge process, but defects such as surface cracks may remain undetected, being too small for the precision of detection. In this study, only the case of cracks with a constant depth over time is considered.

Various processes affect the geometry of the overpack and the mechanical loading. It is therefore important to model all of these processes with their variability to use them as input parameters of a finite element model. The study focuses only on the reliability facing a fracture criterion, to take into account a potential SCC sensitivity of the steel which is a conservative approach. To this end, a parameterized finite element model of the overpack has been developed. The corrosion process and its variability have been modeled as well as the mechanical loading over time. Cracks are positioned and numerically simulated in the highest tensile stress concentration area and the stress intensity factor is estimated using analytical equations.

3 MODEL OF UNCERTAINTY

Various uncertainties should be considered because of the novelty of the whole project, the complexity of the system and the long periods of time studied. These uncertainties can affect significantly the lifetime of the overpack. Therefore, it is important to quantify and to properly model them.

Two different strategies have been used to describe the variability of the system. Uncertain parameters are modelled by random variables. For corrosion rates and water level over time, the evolutions of the phenomena are uncertain because of the inherent complexity of the whole system. Sets of possible scenarios are described and each case can be studied individually.

The corrosion rates are the result of several chemical processes which are hardly predictable. The in situ corrosion process of the overpack has been studied in (Schlegel et al. 2014; Necib et al. 2017). It has been stated that the corrosion rates are decreasing as a result of the pseudo-passivation of the metal surface. However, the precise path that this decrease follows is uncertain, this study focus only on a very severe estimated scenario (Fig. 5). In this one the corrosion rates are steady for 100 years because of a potential oxygen inflow from the access gallery during operating phase of the HLW cell. Once the micro-tunnel is closed and the access drift backfilled, the corrosion rates decrease until they reach a new equilibrium. The uncertainty considered in this paper is related to the overall chemical activity of the system, where both of the corrosion rates are fully correlated. From expert opinion, the corrosion rates have a nominal value c and the interval which is the most likely to contain their actual value is $[\frac{1}{1.5}c;1.5c]$. Therefore, it has been modelled by a multiplying factor A following a lognormal distribution with the median equal to one and affecting both corrosion rates. This way, the probability for the variable to be larger than 1.5 is equal to the probability for it to be less than $\frac{1}{1.5}$. The standard deviation is then fitted so that 99.7% of the realizations are in the interval.

The contact pressure P_c transmitted by the lining after buckling is uncertain. Three different

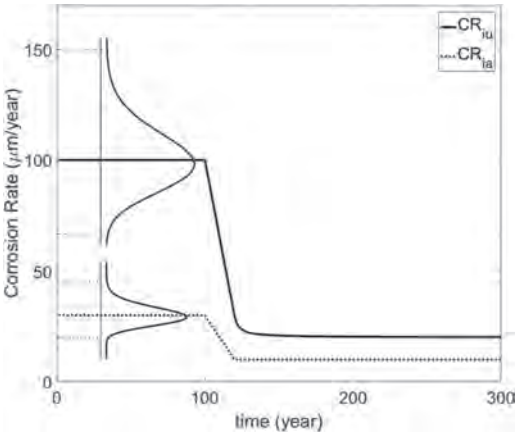


Figure 5. CR_{ia} and CR_{iu} evolutions over time.

studies have been carried out on the contact pressure transmitted by the lining after buckling. The dispersion of the results have been simplified by the confidence interval [$0.5P_{nominal}$;$2P_{nominal}$] which is used for this study. The same strategy as described for the corrosion rates have been used to model the contact pressure with lognormal distribution.

The reliability analysis is performed using a fracture mechanics criterion. An uncertain semi-elliptical crack is simulated on the external surface. The size of the crack is taken deterministic with both crack depth and length of 2 mm. These values have been chosen according to the biggest acceptable crack of the best quality of control described in the European standard EN 10228-3 about non-destructive testing for the forged parts. To reduce the calculation time, the crack is considered at a deterministic coordinate value in the axis of the cylinder (z_c). Preliminary simulations have highlighted a tensile stress concentration localized at this specific position. The position of the crack on the cross-section profile at $z = z_c$ is defined by the angle θ_1, the angle with respect to the circumferential direction. The orientation of the crack on the surface is defined by the angle θ_2. It is assumed that the crack does not have any preferential position or orientation, and that all the configurations are equally likely. Therefore, both of these parameters are taken as independent random variables following a uniform distribution defined by the interval [0; 180]. Early studies (Necib et al. 2017) have been performed to estimate the fracture toughness of the overpack under representative corrosion conditions K_{1scc}. CT specimens have been loaded to 40 $MPa\sqrt{m}$ during 4000 hours exhibiting very limited propagation (<150 µm). Further studies are in progress but no statistical data are currently available. By default, a normal distribution has been defined by the following parameters:

- mean value of $K_{1scc} = 40MPa\sqrt{m}$
- coefficient of variation of 0.1

4 FINITE ELEMENT MODEL

The reliability analysis is based on a finite element model that should be representative of the geometry and loading of the system in operational conditions for all its lifetime. To reduce calculation time, the finite element analysis is limited to the elastic behavior of the material. The goal is to retrieve the stress field in the whole area where cracks will be simulated. This model is evaluated for time steps of 100 years until complete failure. This occurs when the corroded thickness at one location reaches the initial thickness. All these evaluations correspond to one realization of the random variables affecting the finite element model. In order to perform a reliability analysis, many of these realizations are required. Therefore, the model must be as fast as possible and completely parameterized to be automatically generated.

Corrosion is taken into account as a reduction of the thickness, implying that the geometry of the overpack is changing over time. The corroded thickness is non-uniform around the overpack and depends on the time, the history of the water level, and the corrosion rates. The exact thickness of the cross-section can be evaluated at any point of the profile. However, the water level is supposed steady after a period and the corrosion process is highly simplified. These simplifications lead to strong thickness discontinuities implying sharp edges on the corroded cross-section profile. These sharp edges are not representative of the real geometry of the corroded overpack and would lead to stress concentrations in the finite element analysis. To define a continuous smooth cross-section profile representative of the thickness loss by corrosion, the exact corroded thickness is calculated for 16 points on the profile, every 22.5° as described in Equation (1).

$$CT(i,t) = \int_0^T \left(I^a(i,t) CR_{ia}(t) + I^u(t,i) CR_{iu}(t) \right) dt \quad (1)$$

with CT(i,t) the corroded thickness at the i^{th} point at time t, $CRia$(t) and $CRiu$(t) are the values of the corrosion rates above and under water at time t, I^a and I^u are defined as follow:

$$I^a(\boldsymbol{i},\boldsymbol{t}) = \begin{cases} 1 \text{ if } h_{wat}(t) < h(i) \\ 0 \text{ if } h_{wat}(t) \geq h(i) \end{cases} \quad (2)$$

$$I^u(\boldsymbol{i},\boldsymbol{t}) = 1 - I^a(\boldsymbol{i},\boldsymbol{t}) \quad (3)$$

with h_{wat}(t) the height of the water level at time t and h(i) the height of the i^{th} point. The radius of the points are reduced by the corroded thickness and a closed spline curve is defined passing through the points (Fig. 6). For each time increment, the spline is calculated and used as the corroded cross-section profile to generate the geometry of the part.

The mechanical loading is composed of a constant fluid pressure and a contact pressure. The contact pressure over time depends on the date at which the lining buckles. This date is calculated by an empirical equation determined by previous experimental and numerical studies lead at Andra (Nguyen 2017). The equation is based on the rock pressure value and the corrosion rates. The contact area where the pressure is applied is defined by a contact angle α (Fig. 7). The angle

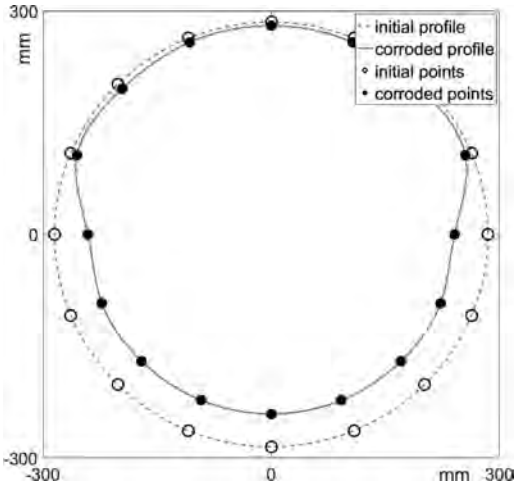


Figure 6. Sample of corroded cross-section profile.

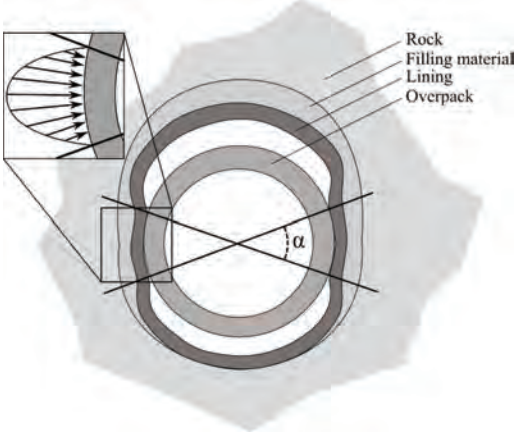


Figure 7. Schematic drawing of the contact angle.

defines two contact areas modelling the case of a two lobes buckling. The contact angle evolution is very difficult to predict due to the complexity of the system (post-buckling behavior of the lining, long term mechanical loading applied to the rock) and as a first approximation, it is assumed that it increases linearly with time and is expected to reach complete contact is about 3000 years. The pressure profile over the contact surfaces is taken as parabolic, such that pressure is equal to zero at the edges and nominal at the middle. The nominal pressure is governed by the random variable P_c as discussed in section 3. Once α reaches 180°, it is necessary to adapt the pressure profile such that its evolution is smooth. At this time, the two contact areas are joined and the pressure is null at the two edges separating them. The overall pressure profile is close to an ellipse. However, the long-term profile pressure is a circle because minor horizontal and vertical stresses in the rock are almost equal. Therefore, to ensure continuity of the loading over time, α is considered increasing even after 180° and the pressure profile is updated following the same evolution. The difference being that the profile pressure is still applied on the two 180° areas, this way the stress values at the edges increase and the pressure profile get closer to a circle (Fig. 8).

The finite element analysis is performed on the uncracked model, the stress field around the crack position is then analyzed to estimate the stress intensity factor as described in (Pommier, Sakae, and Murakami 1999). These equations are a generalization of the work of Newman & Raju (1981). In these equations the stress field around the crack is defined as a polynomial of the geometrical coordinates. The order and parameters of the polynomial are input parameters of the equations. They provide good evaluations of K_I for semi-elliptical surface cracks with a wide range of shapes given by the two input parameters, the crack depth and length. The stress field is exported from the finite-element analysis on a grid with five points in the axial direction centered on z_c, five points in the radial direction and 500 points in the circumferential direction. Once the position of the crack associated with θ_I is determined, the stress field at several points around the crack and the coordinates of these points are retrieved. The area where the data are retrieved is the slice of the scan area that is centered on the crack position (5 axial points, 5 radial points, 6 circumferential points). The components of the stress field are modelled

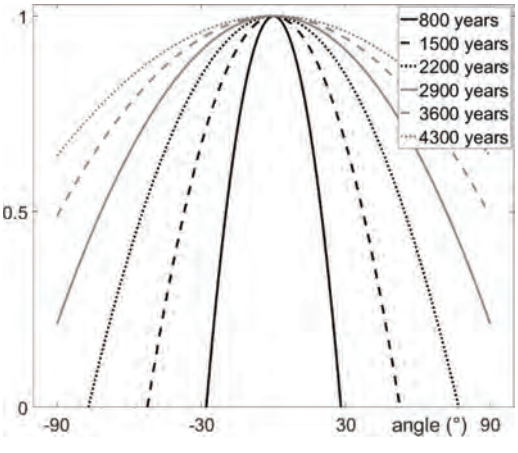


Figure 8. Pressure profile for several time steps after buckling.

with a linear regression in the global coordinate system. Then, the parameters of the fitted linear model of the stress field are calculated in the local coordinate system attached to the crack. These parameters are used to estimate the stress intensity factor associated with the opening mode along the crack edge. This way all positions and orientation of cracks can be simulated with a single finite-element analysis.

5 RELIABILITY ANALYSIS

The results of the finite element model are used to perform the reliability analysis using Monte-Carlo simulations. The failure probability P_f is expressed as follows (Lemaire 2013):

$$P_f = \int_{-\infty}^{+\infty} I(\boldsymbol{X})\, f(\boldsymbol{X})\, d\boldsymbol{X} \tag{4}$$

where f is the joint probability density function of the random variables $\boldsymbol{X}$ and I is defined as:

$$I(\boldsymbol{X}) = \begin{cases} 1 \text{ if } g(\boldsymbol{X}) \le 0 \\ 0 \text{ if } g(\boldsymbol{X}) > 0 \end{cases} \tag{5}$$

with g is the performance function which is an auxiliary function introduced to define the failure event. The failure domain is associated with the negative values of g, as discussed in Equation (5). In Monte-Carlo simulations, realizations of the random variables are generated and the integral expressed in Equation 4 is estimated by:

$$P_f \approx \frac{1}{N} \sum_{j=1}^{N} I\left(\boldsymbol{X}^{(j)}\right) \tag{6}$$

where N is the number of simulations and $\boldsymbol{X}^{(j)}$ denotes samples of the uncertain parameters obtained using a (quasi-)random number generator. This approximation is valid for N sufficiently large, usually for $N \ge 100/P_f$.

In this study, random variables are managed following two different strategies depending on how time consuming they are to evaluate. The random variables affecting the finite element model (A, P_c) are associated with considerable numerical efforts because for each realization the finite element analysis has to be performed for the whole lifetime. The random variables affecting only the crack $(\theta_1, \theta_2, K_{1c})$ are easy to evaluate because they only require post processing without further evaluation of the finite element model. Therefore, to maximize the simulations and to use the advantage of the variables associated with moderate numerical effort, the Monte-Carlo simulations are lead on two levels with two different numbers of simulations. The random variable set can be divided into two sets $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$. where $\boldsymbol{X}_1 = (\theta_1, \theta_2, K_{1c})$ and $\boldsymbol{X}_2 = (A, P_c)$. The previous equation can be written as:

$$P_f = \int \int_{-\infty}^{+\infty} I(\boldsymbol{X}_1, \boldsymbol{X}_2)\, d\boldsymbol{X}_1\, d\boldsymbol{X}_2 \tag{7}$$

The Monte-Carlo simulation is applied twice in order to transform both integrals into sums.

$$\begin{aligned} P_f &\approx \int_{-\infty}^{+\infty} \frac{1}{N_1} \sum_{j=1}^{N_1} I\left(\boldsymbol{X}_1^{(j)}, \boldsymbol{X}_2\right) d\boldsymbol{X}_2 \\ &\approx \frac{1}{N_1 N_2} \sum_{k=1}^{N_2} \sum_{j=1}^{N_1} I\left(\boldsymbol{X}_1^{(j)}, \boldsymbol{X}_2^{(k)}\right) \end{aligned} \tag{8}$$

In this study, 100 simulations of the variables related to the finite element model (N_2) have been evaluated and 100,000 cracks (N_1) have been simulated for each of them. Therefore the failure probability is expressed as follow:

$$P_f(t) \approx \frac{1}{N_1 N_2} \sum_{k=1}^{N_2} \sum_{j=1}^{N_1} I\left(\theta_1^{(j)}, \theta_2^{(j)}, K_{1c}^{(j)}, A^{(k)}, P_c^{(k)}, t\right) \tag{9}$$

and I defined with the performance function g:

$$g\left(\boldsymbol{X}^{(j,k)}, t\right) = K_{1c}^{(j)} - K_1\left(\theta_1^{(j)}, \theta_2^{(j)}, A^{(k)}, P_c^{(k)}\right) \tag{10}$$

Once complete failure is reached (i.e. corrosion of all the thickness of the overpack), the finite element model can no longer be created. Therefore, stress intensity factor and the sign of g cannot be evaluated. The system is then considered defective regardless of the sign of g and the failure probability over time is estimated with:

$$\text{if } t \ge t_{cf} \text{ then } I(\boldsymbol{X}_1, \boldsymbol{X}_2, t) = 1$$

6 RESULTS

The stress intensity factor K_1 have been evaluated every 100 years until complete failure for the $N_2 * N_1$ simulations. The results have been represented by the median curve and confident interval bounds at 95% (Fig. 9). The estimation of the stress intensity factor gives negative values as response of compressive stresses. These values have been considered as equal to zero which does not affect the median curve or the estimation of the failure probability. The results show that the dispersion

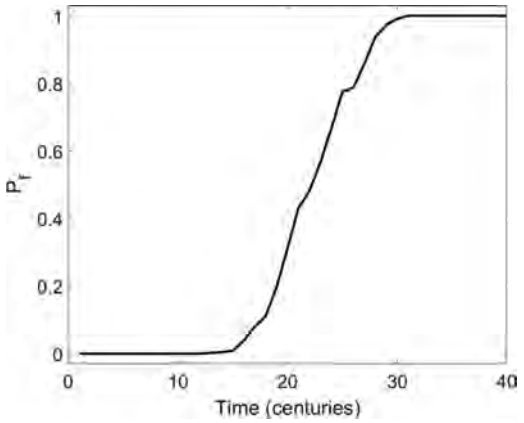


Figure 10. Estimated failure probability over time.

increases with time while the median value remains almost unchanged. This result can be explained by three phenomena. First, with time increasing, the average thickness shrinks and the overall stresses slowly increases. Second, the stress concentration area as discussed in section 3 is only local in the area where cracks are simulated but it may lead to high stress intensity factors. This area grows with time and the stress values involved increase rapidly, thus increasing the extreme values of stress intensity factor. Finally, depending on their orientation, the cracks are almost equally likely to be subjected to tension than to be subjected to compression. Therefore, even with increasing stresses involved, the median value of the stress intensity factors remain balanced. Because of the impossibility to evaluate K_1 once compete failure is reached as discussed in section 5, with time increasing the number of samples available to estimate the variability of K_1 decreases, explaining the increasing irregularities of the upper bound of the confident interval.

The failure probability over time estimated by the Monte-Carlo simulation is represented in Fig. 10. The first failures occurs after 1100 years and the failure probability at this date is 10^{-6}. Therefore, the failure probability at 500 years which is the minimum required lifetime is less than 10^{-6}.

7 CONCLUSION

The failure probability of the overpacks for HLW is estimated taking into account the evolution of the operational conditions and the aging of the system. The presented study is based on assumptions made to ensure conservationism and reduce computation time. The failure probability at 500 years is smaller than 10^{-6} which is in good agreement with the lifetime requirement of the overpack. This study gives a first estimation of the reliability of the design choices made for the Cigeo project. The reliability and safety of the infrastructures are major issues of the project and these results are relevant for the validation and improvement of the design choices.

This study could be completed by more simulations and the implementation of a metamodel to reduce numerical efforts. Further improvement would be to take into account a wider range of parameters involved in the problem. For example the other corrosion cases and water level evolution scenarios could be evaluated, as well as cracks on a wider surface area and the possibility of internal cracks in the overpack thickness.

ACKNOWLEDGMENTS

This work is founded by ANDRA, which is gratefully acknowledged for the support.

REFERENCES

ANDRA. 2005. 'Dossier 2005 Synthesis Argile: Tome Phenomenological Evolution of a Geological Repository'.

ANDRA. 2015. 'National Inventory of Radioactives Materials and Wastes'.

Brommundt, J. et al. 2014. 'Full-Scale 3D Modelling of a Nuclear Waste Repository in the Callovo-Oxfordian Clay. Part 1: Thermo-Hydraulic Two-Phase Transport of Water and Hydrogen'. Geological Society, London, Special Publications 400 (1):443–67. https://doi.org/10.1144/SP400.34.

Croisé, J. et al. 2011. 'Impact of Water Consumption and Saturation-Dependent Corrosion Rate on Hydrogen Generation and Migration from an Intermediate-Level Radioactive Waste Repository'. Transport in Porous Media 90 (1):59–75. https://doi.org/10.1007/s11242-011-9803-0.

Dundulis, G. et al. 2016. 'Integrated Failure Probability Estimation Based on Structural Integrity Analysis and Failure Data: Natural Gas Pipeline Case'. *Reliability Engineering & System Safety* 156 (December):195–202. https://doi.org/10.1016/j.ress.2016.08.003.

Lemaire, Maurice in collaboration with Alaa Chateauneuf and Jean-Claude Mitteau. 2013. *Structural Reliability.* John Wiley & Sons.

Necib, S. et al. 2017. 'Assessment of the Resistance to Environmentally Assisted Cracking (EAC) of C-Steel Casing and Overpack in the CO_x Claystone'. Corrosion Engineering, Science and Technology 52 (sup1):95–100. https://doi.org/10.1080/1478422X.2017.1336003.

Newman, J.C. & I.S. Raju. 1981. 'An Empirical Stress-Intensity Factor Equation for the Surface Crack'. *Engineering Fracture Mechanics* 15 (1–2):185–192.
Nguyen, T. 2017. 'Flambage Sous Contact d'une Coque Cylindrique Soumise à Pression Externe'. Lyon: INSA.
Pommier, S. et al. 1999. 'An Empirical Stress Intensity Factor Set of Equations for a Semi-Elliptical Crack in a Semi-Infinite Body Subjected to a Polynomial Stress Distribution.' I*nternational Journal of Fatigue* 21 (3):243–251.
Schlegel, Michel L. et al. 2014. 'Corrosion of Metal Iron in Contact with Anoxic Clay at 90°C: Characterization of the Corrosion Products after Two Years of Interaction'. Applied Geochemistry 51 (December):1–14. https://doi.org/10.1016/j.apgeochem.2014.09.002.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Probabilistic fatigue damage prediction of relative short edge crack using direct optimized probabilistic calculation

M. Krejsa & J. Brozovsky
Department of Structural Mechanics, Faculty of Civil Engineering, VSB—Technical University of Ostrava, Czech Republic

S. Seitl
High Cycle Fatigue Group, Institute of Physics of Materials, Brno, Czech Republic
Institute of Structural Mechanics, Faculty of Civil Engineering, Brno University of Technology, Czech Republic

Z. Kala
Institute of Structural Mechanics, Faculty of Civil Engineering, Brno University of Technology, Czech Republic

V. Krejsa & P. Lehner
Department of Structural Mechanics, Faculty of Civil Engineering, VSB—Technical University of Ostrava, Czech Republic

ABSTRACT: Three sizes are important for the characterization of the propagation of fatigue cracks—initial size, detectable size and acceptable size. The theoretical model of a fatigue crack progression can be based on a linear elastic fracture mechanics. Depending on location of an initial crack, the crack may propagate in structural element that could be described by calibration functions. Single edge-cracked steel element with rectangular cross-section under relative short edge fatigue damage under pure tension, pure bending, three and four point bending load have been chosen for applications of the theoretical solution suggested in the studies. When determining the required level of reliability, it is possible to specify the time of the first inspection of the construction which will focus on the fatigue damage. Using a conditional probability and Bayesian approach, times for subsequent inspections can be determined based on the results of the previous inspection. For probabilistic calculation of fatigue crack progression, the original and new probabilistic method—the Direct Optimized Probabilistic Calculation (DOProC), which uses a purely numerical approach based on optimized numerical integration without any simulation techniques or approximationapproach. This provides more accurate solutions to probabilistic tasks, and, in some cases, allows to considerably fasten completion of computations with the taking into account the statistical dependence of random input variables.

1 INTRODUCTION

Fatigue phenomenon is one of the main factors influencing the life of steel structures and bridges subjected to cyclic loading. A substantial increase in the overall weight load from vehicle axles and crossing frequencies leads to higher fatigue damage than considered during the design of bridges. Due to above mentioned reasons, it is highly relevant to develop methods for the calculation and assessment of the residual fatigue life and time-dependent analysis of the reliability of existing steel structures, e.g., Soliman et al. (2016), Maljaars & Vrouwenvelder (2014) and Partov & Kantchev (2014). Numerous numerical methods, mostly based on the Finite Element Method (FEM), e.g., Major et al. (2017), Nemec et al. (2017), Vican et al. (2015) Cajka (2013) and Kormanikova and Kotrasova (2011), have been developed to aid in the understanding of the behavior of the fatigue phenomena.

In the design of structures, information related to time-variable load and detection of cracks from measurement during the operational period of the structure should be incorporated into reliability calculations. The essential tools for these calculationsare provided by fracture mechanics and the reliability theory, e.g., Hradil et al. (2017) and Michalcova & Lausova (2017). Some of approaches used for the fatigue crack prediction are based on stochastic methods, e.g., Kralik (2016), Antucheviciene et al. (2015) and Fedorik

et al. (2015). Insight into the stochastic interactions among random factors (load, geometric and material characteristics), e.g., Krivy & Konecny (2013), affecting the reliability of steel bridges is namely essential and crucial to understanding the progress of failure probability of steel structures over time, e.g., Schneider, Thons, & Straub (2017). Moreover, due to the presence of significant uncertainties associated with crack initiation and propagation, inspection, monitoring and/or repair actions planning should be performed and applied to prevent sudden failures of damaged structural components and their associated consequences, e.g., Lotsberg, Sigurdsson, Fjeldstad, & Moan (2016).

The paper focuses on the probabilistic approach based on optimized numerical integration—the newly developed Direct Optimized Probabilistic Calculation method (DOProC), published in details, e.g., in Janas et al. (2017). The DOProC method is distinguished by higher accuracy than the other probabilistic methods. Another advantage is the easy implementation on platforms with multiple processing units or cores, enabling the parallel computing of this probabilistic procedure, e.g., on supercomputers – (Krejsa et al. 2016). This new probabilistic approach has allowed to describe and implement a very precise methodology for stochastic prediction of fatigue damage of steel structures and bridges exposed to cyclic loading. Probabilistic modeling of fatigue crack progression leads to designing a system of regular inspections of structures and is based on linear elastic fracture mechanics and Paris-Erdogan's law, e.g., Seitl et al. (2017). This article describes this approach with a particular focus on the demonstration of using the above mentioned methodology for the single edge-cracked steel element with a rectangular cross section under relatively short edge fatigue damage.

2 THEORETICAL BACKGROUND

2.1 *Fatigue crack propagation*

When investigating the propagation, the fatigue crack that deteriorates a certain area of the structure component is described with one dimension only—fatigue crack length a. In order to describe the propagation of the crack, the linear elastic fracture mechanics is typically used. This method uses Paris-Erdogan's law and defines relation between propagation rate of the crack size a, and range of the stress rate coefficient, ΔK, in the tip of the crack:

$$\frac{da}{dN} = C \cdot \Delta K^m , \tag{1}$$

where C, m are material constants, that are determined experimentally, N is the number of loading cycles and ΔK is range of the stress intensity factor in front of the crack tip and it is defined as follow:

$$\Delta K = \Delta\sigma \cdot \sqrt{\pi \cdot a} \cdot F\left(\frac{a}{h}\right), \tag{2}$$

where $\Delta\sigma$ is constant stress range (the value of $\Delta\sigma$ corresponding to each way of loading, see Fig. 1, is shown in Table 1), h is the height of the rectangular cross-section of the component and $F(a/h)$ is the calibration function which represents the course of propagation of the crack (e.g., at the edge or on the surface of the component) and various boundary conditions.

Three sizes are important for the description of the characteristics of the propagation of fatigue cracks. The fatigue crack will propagate in a stable way only if the initial crack a_0 exists in the place where the stress is concentrated. Existence of the initiation cracks during the propagation should be revealed, along the detectable length of the crack a_d, e.g., during inspections. The crack propagates in a stable way until it reaches the third important size—acceptable length of the crack a_{ac}, which is a limit for the required reliability.

2.2 *Stochastic reliability assessment*

The main assumption is that the primary design should take into account the effects of the extreme loading and the fatigue resistance should be assessed. The probabilistic methods should be used for the investigation of the propagation rate of the fatiguecrack until the acceptable size is reached because the input variables include uncertainties and reliability should be taken into account. The resistance of the structure can be evaluated using Eqs. (1) and (2) as:

$$R(a_{ac}) = \int_{a_0}^{a_{ac}} \frac{da}{\left(\sqrt{\pi \cdot a} \cdot F\left(\frac{a}{h}\right)\right)^m} \mathrm{d}a \,. \tag{3}$$

If the upper integration limit a_d is used, the resistance of the structure $R_{(a_d)}$ can be specified similarly. Similarly, it is possible to define the cumulated effect of loads that equals to:

$$E(N) = \int_{N_0}^{N} C \cdot \Delta\sigma^m \mathrm{d}N = C \cdot \Delta\sigma^m \cdot (N - N_0), \tag{4}$$

where N is the total number of oscillations $\Delta\sigma$ for the change of the length from a_0 to a_{ac}, and N_0 is the number of oscillations in the time of initialization of the fatigue crack (typically, the number of oscillations is zero). For details see, e.g., (Krejsa, Koubova, Flodr, Protivinsky, & Nguyen 2017).

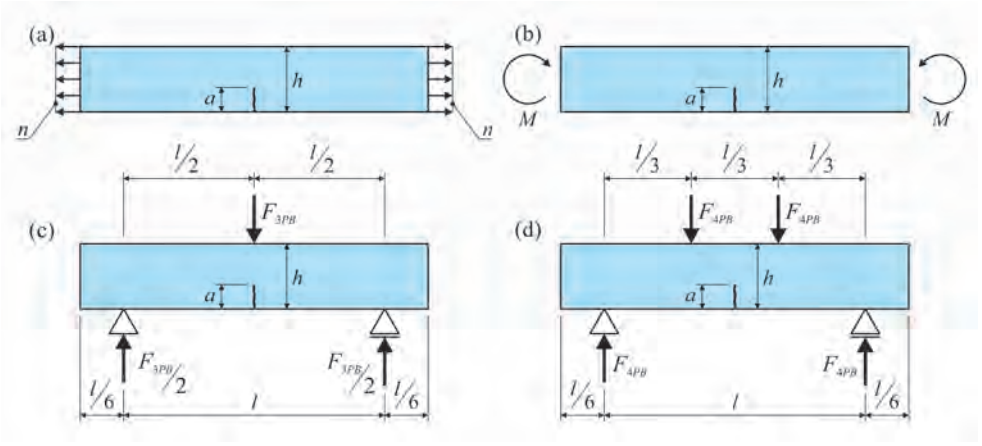


Figure 1. Static scheme for single edge-crack steel specimens: (a) loaded by tension, (b) loaded by pure bending, (c) loaded by three-point bending, (d) loaded by four-point bending.

Table 1. Stress range $\Delta\sigma$ and the acceptable length of the crack a_{ac} for single edge-crack steel specimens and various loads according Fig. 1; l is span of the element, w and h are width and height of the rectangular cross-section, n is flat load [N/m²], N is axial force and f_y is yield stress.

Type of load	Stress range $\Delta\sigma$	Acceptable length of the crack a_{ac}
Tension	$\frac{N}{w\cdot h}, N = n\cdot w\cdot h$	$h-\frac{N}{w\cdot f_y}$
Pure bending	$\frac{6\cdot M}{w\cdot h^2}$	$h-\sqrt{\frac{6\cdot M}{w\cdot f_y}}$
Three-point bending – 3PB	$\frac{3\cdot F_{3PB}\cdot l}{2\cdot w\cdot h^2}$	$h-\sqrt{\frac{3\cdot F_{3PB}\cdot l}{2\cdot w\cdot f_y}}$
Four-point bending – 4PB	$\frac{2\cdot F_{4PB}\cdot l}{w\cdot h^2}$	$h-\sqrt{\frac{2\cdot F_{4PB}\cdot l}{w\cdot f_y}}$

The probability of failure P_f equals to:

$$P_f = P\left(G_{fail}(\boldsymbol{X})<0\right) = P\left(R(a_{ac})-E(N)<0\right), \quad (5)$$

where G_{fail} is reliability function and **X** is a vector of random physical properties such as mechanical properties, geometry of the structure, load effects and dimensions of the fatigue crack.

2.3 *Inspections planning*

When the probability of failure P_f according to Eq. (5) exceeds the specified designed probability, P_d, the inspection should be performed. On the basis of the results of the first inspection, a system of following inspections can be established using conditional probability, as e.g. in Krejsa et al. (2017).

Because it is not certain in the probabilistic calculation whether the initiation crack exists and what the initiation crack size is and because other inaccuracies influence the calculation of the crack propagation, a special inspection is necessary to check the size of the measurable crack in a specific period of time. The acceptable crack size influences the time of the inspection. If no fatigue cracks are found, the analysis of inspection results give conditional probability during occurrence.

While the fatigue crack is propagating, it is possible to define following random phenomena that are related to the growth of the fatigue crack and may occur in any time, t, during the service life of the structure. Then:

- $U_{(t)}$ **phenomenon:** No fatigue crack failure has been revealed within the t-time and the fatigue crack size $a_{(t)}$ has not reached the detectable crack size a_d. This means:

$$a_{(t)} < a_d\,, \quad (6)$$

- $D_{(t)}$ **phenomenon:** A fatigue crack failure has been revealed within the t-time and the fatigue crack size $a_{(t)}$ is still below the acceptable crack size a_{ac}. This means:

$$a_d \le a_{(t)} < a_{ac}\,, \quad (7)$$

- $F_{(t)}$ **phenomenon:** A failure has been revealed within the t-time and the fatigue crack size $a_{(t)}$ has reached the acceptable crack size a_{ac}. This means:

$$a_{ac} < a_{(t)} . \tag{8}$$

If the crack is not revealed within the *t*-time, this may mean that there is not any fatigue crack in the construction element. This might be also an initiative phase of nucleation of the fatigue crack (when a crack appears in the material) and this phenomenon is not taken into account in the fracture mechanics. Even if the fatigue crack is not revealed it is likely that it exists there but the fatigue crack size is so small that it cannot be detected under existing conditions.

Using the phenomena above, it is possible to define following probabilities:

- The probability that the failure is not detected within the *t*-time, this means the probability that the fatigue crack size $a_{(t)}$ is below the measurable crack size a_d:

$$P\left(U_{(t)}\right) = P(a_{(t)} < a_d) , \tag{9}$$

- The probability that the failure detected within the *t*-time has the crack size $a_{(t)}$ that is less than the acceptable size a_{ac}:

$$P\left(D_{(t)}\right) = P(a_d \le a_{(t)} < a_{ac}) , \tag{10}$$

- The probability that the failure occurs within the *t*-time, this means the probability that the fatigue crack size $a_{(t)}$ reaches the acceptable size a_{ac}:

$$P\left(F_{(t)}\right) = P(a_{ac} < a_{(t)}) . \tag{11}$$

Those three phenomena cover the complete spectrum of phenomena that might occur in the *t*-time. This means:

$$P\left(U_{(t)}\right) + P\left(D_{(t)}\right) + P\left(F_{(t)}\right) = 1 . \tag{12}$$

In order to specify the time for the next inspection, it is necessary to determine the conditional probabilitie which can be expressed using the full probability rule:

$$P(F_{(T)} \mid U_{(t_I)}) = \frac{P\left(F_{(T)}\right) - P\left(F_{(t_I)}\right)}{P\left(U_{(t_I)}\right)} - \frac{P\left(D_{(t_I)}\right) \cdot P(F_{(T)} \mid D_{(t_I)})}{P\left(U_{(t_I)}\right)} \tag{13}$$

and

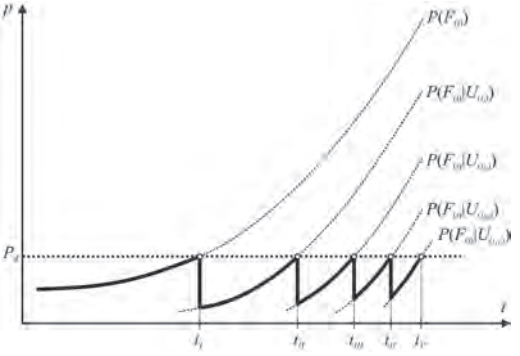


Figure 2. Probabilities of failure P_f calculated according Eqs. (5) and (13) with the times of 5 inspections.

$$P(F_{(T)} \mid D_{(t_I)}) = \frac{P\left(F_{(T)}\right) - P\left(F_{(t_I)}\right)}{P\left(D_{(t_I)}\right)} - \frac{P\left(U_{(t_I)}\right) \cdot P(F_{(T)} \mid U_{(t_I)})}{P\left(D_{(t_I)}\right)} \tag{14}$$

For details see Fig. 2 and, e.g., (Krejsa, Brozovsky, & Mikolasek 2017), (Krejsa, Kala, & Seitl 2016).

3 APPLICATION

The demonstration of using introduced methodology was made for the single edge-cracked steel element with rectangular cross-section under short edge fatigue damage defined according scheme in Fig. 1(c). Calibration curves $F(a/h)$ were experimentally derived for a specimens with relative crack length a/h: 0.01–0.3 loaded by tension, pure, three-point and four-point bending.

The resulting calibration functions, e.g., for 3PB test and ration $l/h = 2$ is according to Seitl et al. (2017):

$$F(\frac{a}{h})^{\frac{l}{h}=2}_{3PB} = +1.0259 - 1.4659 \cdot \left(\frac{a}{h}\right) + 4.9318 \cdot \left(\frac{a}{h}\right)^2 - 2.4637 \cdot \left(\frac{a}{h}\right)^3 , \tag{15}$$

for ration $l/h = 4$ is:

$$F(\frac{a}{h})^{\frac{l}{h}=4}_{3PB} = +1.0691 - 1.3496 \cdot \left(\frac{a}{h}\right) + 5.1865 \cdot \left(\frac{a}{h}\right)^2 - 3.3509 \cdot \left(\frac{a}{h}\right)^3 , \tag{16}$$

for ration $l/h=8$ is:

$$F\left(\frac{a}{h}\right)_{3PB}^{\frac{l}{h}=8}=+1.0963-1.3052\cdot\left(\frac{a}{h}\right)+5.2829\cdot\left(\frac{a}{h}\right)^2-3.5972\cdot\left(\frac{a}{h}\right)^3, \quad (17)$$

for ration $l/h=16$ is:

$$F\left(\frac{a}{h}\right)_{3PB}^{\frac{l}{h}=16}=+1.1079-1.2328\cdot\left(\frac{a}{h}\right)+5.0551\cdot\left(\frac{a}{h}\right)^2-3.2837\cdot\left(\frac{a}{h}\right)^3 \quad (18)$$

and for ration $l/h=80$ is:

$$F\left(\frac{a}{h}\right)_{3PB}^{\frac{l}{h}=80}=+1.1180-1.1964\cdot\left(\frac{a}{h}\right)+5.0176\cdot\left(\frac{a}{h}\right)^2-3.3127\cdot\left(\frac{a}{h}\right)^3. \quad (19)$$

The value of the calibration function $F(\frac{a}{h})_{3PB}$ for intermediate ratio values of *l*/*h* is determined by linear interpolation.

Eqs. (3), (4) and (5) allow for effective computation of the inspection times for each single edge-cracked steel components using DOProC method using FSCProbCalc code with exactly defined random input quantities.

The allowable crack size a_{ac} for the single edge-crack steel specimen loaded by three-point bending can be expressed by a relationship in Table 1 considering the derived weakening of the cross-sectional area of the element (with the limit length defined by ratio $a/h = 0.3$), similarly as in (Wang, Zhai, Duan, & Wang 2015):

$$a_{ac}=h-\sqrt{\frac{3\cdot F_{3PB}\cdot l}{2\cdot w\cdot f_y}}. \quad (20)$$

Deterministic and random input quantities are given in Table 2 and Table 3.

If a period of time t is specified and the time step is 1 year, it is possible to determine resistance of the construction $R_{(a_{ac})}$ and $R_{(a_d)}$ pursuant to Eq. (3) – see Figs. 3 and 4, load effects, $E(N)$, pursuant to Eq. (4) – see Fig. 5, and reliability function $G_{fail}(\boldsymbol{X})$ according Eq. (5) – see Fig. 6, as well as the probability of elemental phenomena, *U*, *D* and *F*, pursuant to Eqs. (9) through (11) for each year of the structural operation – see Fig. 7, which are the basis for specification of inspection times.

Table 2. Overview of deterministic input quantities.

Quantity	Value
Material constant m	3
Material constant C	$2.2\cdot10^{13}$ MPa^{m}m$^{(m/2)+1}$
Height of the rectangular cross-section h	0.1 m
Width of the rectangular cross-section w	0.01 m
Span of the element l	0.4 m
Target probability of failure P_d	0.02277 ($\beta=2$)

Table 3. Overview of random input quantities expressed in a bounded histograms.

Quantity	Type of parametric probability distribution	Mean value	Standard deviation
Total number of stress peaks per year N	Normal	10^6	10^5
Yield stress f_y	Lognormal	200 MPa	20 MPa
Loading force in three-point bending test F_{3PB}	Normal	6 kN	0.6 kN
Initial size of the crack a_0	Lognormal	0.2 mm	0.05 mm
Smallest detectable size of the crack a_d	Normal	2 mm	0.2 mm

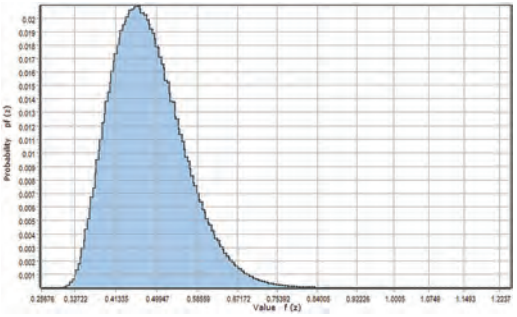


Figure 3. Resulting histograms of the structural resistance $R_{(a_d)}$.

When the probability of failure P_f according Eq. (5) exceeds the specified designed probability, P_d, the inspection should be performed. The Table 4 include numerical values for the final inspection times—for the first inspection and subsequent inspections resulting from the conditional probability pursuant to Eq. (13).

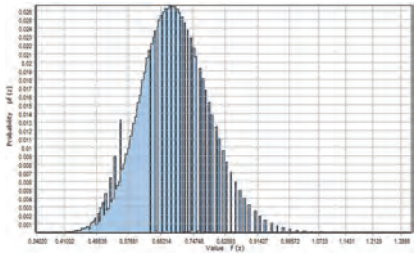


Figure 4. Resulting histograms of the structural resistance $R_{(a_{ac})}$.

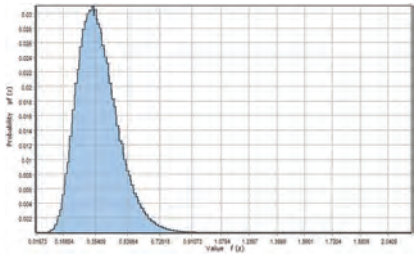


Figure 5. Resulting histogram of the load effect $E(N)$ of the calculation for $t = 35$ years of structural operation.

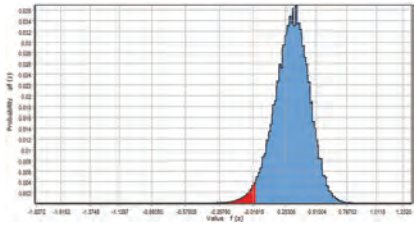


Figure 6. Resulting histogram of the reliability function $G_{fail}(\mathbf{X})$ of the calculation for $t = 35$ years of structural operation.

Table 4. Calculated times for the first five inspection of the structural element.

Inspection no.	Time of inspection [years]
#1	35
#2	46
#3	48
#4	50
#5	51

4 CONCLUSION

The article demonstrates the probabilistic calculation of the fatigue damage prediction of relatively short edge crack under various loading using the newly developed Direct Optimized Probabilistic Calculation (DOProC), which appears to be a very efficient tool to make probabilistic assessment of the structural reliability on the basis of the exact definition of the acceptable size of the fatigue crack. The theoretical model of fatigue crack progression is based on a linear fracture mechanics and Paris-Erdogan law. The computational procedure is capable to make probabilistic assessment of the structural reliability on the basis of the exact definition of the acceptable size of the fatigue crack. The probabilities were obtained for three basic phenomena, which are related to propagation of the fatigue cracks. On the basis of those data, the probability of failure can be calculated for each year of operation of the structural element. When determining the required degree of reliability, it is possible to specify the time of the first inspection of the structure, which will focus on the fatigue damage. Using a conditional probability, times for subsequent inspections can be determined. The article describes how to design the system of regular structural inspections in case of the simple demonstration example.

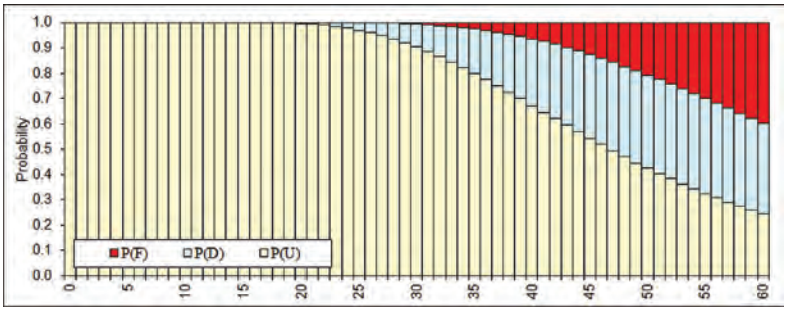


Figure 7. Resulting probabilities of random events U, D and F for first 60 years of structural operation under various load.

The DOProC method and its application in probabilistic prediction of fatigue crack damage can considerably improve estimation of maintenance costs for the structures and bridges subject to cyclic loads. This methodology is developed further. The goal of investigations seems to be, in particular, application of Bayesian networks in the computational model, such as e.g. in Mahadevan et al. (2001), which describes propagation of fatigue cracks in the system.

ACKNOWLEDGMENTS

This contribution has been developed as a part of the research project GACR 17-01589S "Advanced computational and probabilistic modelling of steel structures taking account fatigue damage" supported by the Czech Grant Agency.

REFERENCES

Antucheviciene, J., Z. Kala, M. Marzouk, & E.R. Vaidogas (2015). Solving civil engineering problems by means of fuzzy and stochastic mcdm methods: Current state and future research. *Mathematical Problems in Engineering 2015*, 1–16. Article ID 362579.

Cajka, R. (2013). Accuracy of stress analysis using numerical integration of elastic half-space. *Applied Mechanics and Materials 300–301*, 1127–1135.

Fedorik, F., J. Kala, A. Haapala, & M. Malask (2015). Use of design optimization techniques in solving typical structural engineering related design optimization problems. *Structural Engineering and Mechanics 55*, 1121–1137.

Hradil, P., V. Salajka, & J. Kala (2017). Requirements of technical standards for the dynamic analysis of the load-bearing structures of footbridges. *MATEC Web of Conferences 107*, 1–6.

Janas, P., M. Krejsa, J. Sejnoha, & V. Krejsa (2017). Doprocbased reliability analysis of structures. *Structural Engineering and Mechanics 64*, 413–426.

Kormanikova, E. & K. Kotrasova (2011). Resonant frequencies and mode shapes of rectangular sandwich plate. *Chemicke listy 105*, 535–538.

Kralik, J. (2016). Probabilistic safety assessment of the design of a tall buildings under the extreme load. In *International Conference of Numerical Analysis and Applied Mathematics 2015, Volume 1738 of AIP Conference Proceedings*, Melville, NY. American Institute of Physics. Article Number 480088.

Krejsa, M., J. Brozovsky, P. Janas, R. Cajka, & V. Krejsa (2016). Probabilistic calculation using parallel computing. In *Engineering Mechanics*, Svratka, Czech Republic. Acad Sci Czech Republic.

Krejsa, M., J. Brozovsky, & D. Mikolasek (2017). Probabilistic reliability assessment of steel elements exposed to fatigue using a bayesian approach. In *Safety and Reliability – Theory and Applications*, London, pp. 2119–2125. Taylor & Francis Group,.

Krejsa, M., J. Brozovsky, D. Mikolasek, P. Lehner, & P. Parenica (2017). Using doproc method in reliability assessment of steel elements exposed to fatigue. *MATEC Web of Conferences 107*, 1–8.

Krejsa, M., Z. Kala, & S. Seitl (2016). Inspection based probabilistic modeling of fatigue crack progression. *Procedia Engineering 142*, 145–152.

Krejsa, M., L. Koubova, J. Flodr, J. Protivinsky, & Q.T. Nguyen (2017). Probabilistic prediction of fatigue damage based on linear fracture mechanics. *Frattura ed Integrita Strutturale 11*, 143–159.

Krivy, V. & P. Konecny (2013). Real material properties of weathering steels used in bridge structures. *Procedia Engineering 57*, 624–633.

Lotsberg, I., G. Sigurdsson, A. Fjeldstad, & T. Moan (2016). Probabilistic methods for planning of inspection for fatigue cracks in offshore structures. *Marine Structures 46*, 167–192.

Mahadevan, S., R. Zhang, & N.J. Smith (2001). Bayesian networks for system reliability reassessment. *Structural Safety 23*, 231–251.

Major, M., K. Kulinski, & I. Major (2017). Dynamic analysis of an impact load applied to the composite wall structure. *MATEC Web of Conferences 107*, 1–6.

Maljaars, J. & A.C.W.M. Vrouwenvelder (2014). Probabilistic fatigue life updating accounting for inspections of multiple critical locations. *International Journal of Fatigue 68*, 24–37.

Michalcova, V. & L. Lausova (2017). Numerical approach to determination of equivalent aerodynamic roughness of industrial chimneys. *Computers and Structures in Press*, 1–7.

Nemec, I., H. Stekbauer, A. Vaneckova, & Z. Vlk (2017). Explicit and implicit method in nonlinear seismic analysis. *MATEC Web of Conferences 107*, 1–8.

Partov, D. & V. Kantchev (2014). Gardner and lockman model in creep analysis of composite steel-concrete sections. *ACI Structural Journal 111*, 59–69.

Schneider, R., S. Thons, & D. Straub (2017). Reliability analysis and updating of deteriorating systems with subset simulation. *Structural Safety 64*, 20–36.

Seitl, S., P. Miarka, L. Malikova, & M. Krejsa (2017). Fcomparison of calibration functions for short edge cracks under selected loads. *Key Engineering Materials 754*, 353–356.

Seitl, S., T. Thienpont, & W.D. Corte (2017). Fatigue crack behaviour: comparing three-point bend test and wedge splitting test data on vibrated concrete using paris' law. *Frattura ed Integrita Strutturale 11*, 110–117.

Soliman, M., D.M. Frangopol, & A. Mondoro (2016). A probabilistic approach for optimizing inspection, monitoring, and maintenance actions against fatigue of critical ship details. *Structural Safety 60*, 91–101.

Vican, J., J. Gocal, J. Odrobinak, M. Moravcik, & P. Kotes (2015). Determination of railway bridges loading capacity. *Procedia Engineering 111*, 839–844.

Wang, C.S., M.S. Zhai, L. Duan, & Q. Wang (2015). Fatigue service life evaluation of existing steel and concrete bridges. *Advanced Steel Construction 11*, 305–321.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Reliability analysis of structural health monitoring systems

E. Etebu & M. Shafiee
Cranfield University, College Road, Bedfordshire, UK

ABSTRACT: Structural Health Monitoring (SHM) systems are comprised of a grid of sensors installed at a fixed location on structures to detect the presence of defect, localize the detected defect, quantify its severity, and estimate the Remaining Useful Life (RUL). SHM system performance is currently assessed based on Probability of Detection (POD) of defects, which is a function of defect size. This performance parameter was inherited from Non-Destructive Testing (NDT), where a human operator performs inspection on a structure at a given location, with mobile sensors. For SHM systems, POD and Probability-of-False-Alarm (PFA) are a measure for only detection of defects. Furthermore, these parameters could vary over time as sensors degrade. This paper presents a methodology to characterize the performance of SHM systems with respect to damage detection, localization, and assessment. Probability theorem is used to characterize uncertainties associated with the SHM process, and Bayes theorem is employed to determine its reliability. The methodology is then tested on vibration-based modal strain energy SHM technique applied to a numerical Finite Element Analysis (FEA) study conducted on an offshore energy structure.

1 INTRODUCTION

Structural health monitoring (SHM) is a technique used to assess the integrity of in-service structures that are exposed to operational and environmental loads on a continuous basis, with the goal of improving the monitored structure's reliability, reducing inspection cost, and emergency repair expenditure (Doebling *et al.* 1998). SHM systems are typically comprised of a grid of sensors installed at a fixed location on monitored structure, where signals measured from the structures response are transferred though a communication network, processed to extract damage sensitive features, after which a damage detection algorithm is employed to determine the structure's integrity. In overall, the four functions of SHM systems include detection of defect, localizing the detected defect, quantifying its severity, and estimating the structure's remaining useful life (RUL).

SHM system performance is currently assessed based on probability of detection (POD), which describes the probability of detecting defects as a function of their size at the time of inspection (Annis, 2009). This performance parameter was inherited from non-destructive testing (NDT), where POD is a function of replicability achieved by the human operator, instruments and sensors used at a given location. Moreover, the instruments and sensors can be transported to multiple locations for inspection. This is vastly different from SHM where a human operator is absent; instruments and sensors are placed at an unchanging location after installation. For a SHM system, the replicability of POD is influenced by the environment, measurement noise, deterioration of sensors and communication network.

Statistical characterization of SHM systems performance based on POD was established in earlier works (see Thompson, 2007; Aldrin, 2010; Aldrin *et al.*, 2011; Etebu and Shafiee, 2017) to alleviate the number of experiments associated with SHM operational parameters, by implementing a validated model to assess the changes in SHM outcome at varying damage severity, location, as well as operational conditions. Outcomes from these models are used to generate POD curves using the Hit/Miss data fitted to a binary regression model, where the POD curve is generally plotted with 95% confidence interval (CI).

In addition to POD, the probability of false alarm (PFA) has also been applied to evaluate SHM performance, where PFA is a performance measure of detecting defect in the absence of defect [3]. Although these two parameters are indicators of SHM performance, POD and PFA are associated with only damage detection. Hence, these parameters do not evaluate SHM performance with respect to localization, and damage assessment.

In this paper, a methodology is presented to characterize the performance of SHM systems with respect to damage detection, localization, and assessment using the Model Assisted Probability

approach, and Vibration based SHM technique to assess varying levels of damage on multiple members of a fixed offshore platform in operational conditions.

The rest of the paper is organized as follows. Section 2 develops an assessment framework for SHM system performance. In Section 3, the proposed approach is applied to a numerical study and the results are analysed in Section 4. Finally, the research is concluded in Section 5.

2 SHM PERFORMANCE: DAMAGE DETECTION, LOCALIZATION AND ASSESSMENT

The statistical characterization of SHM systems performance based on POD and PFA has been established by previous studies (Annis, 2009; Aldrin *et al.*, 2011; Thompson, 2008), where the output O_{dd} of a SHM damage detection system either generates an event that indicates the occurrence damage D_d in a structure, or no damage detected D_{nd}. The classification of damage occurrence O_{dd} for SHM is based on a minimum detectable damage size known as the damage threshold D_{th} which separates D_d and D_{nd}. POD and PFA are then expressed as

$$POD(t) = P(O_{dd} \mid D_d - D_{th} \geq 0), \tag{1}$$

$$PFA(t) = P(O_{dd} \mid D_d - D_{th} < 0). \tag{2}$$

In order to address the performance of SHM systems with respect to localizing damage and estimating its severity, two additional performance functions are required to be developed. The performance of a SHM system with respect to locating damage on a structure—in the event of damage being detected—is characterized using the Probability of Accurate Localization (POAL), which is constructed based on the zone of true damage location (ZTDL) and the predicted location of damage occurrence (PLD). The zone of true damage location is characterized by the true damage location (TDL) and an allowable localization tolerance (ALT) corresponding to a specific TDL. Based on the given definition, the limit state function (LSF) for POAL is expressed mathematically as:

$$g(\mathrm{x,y,z,t}) = |\mathrm{TDL} - \mathrm{PLD}| - \mathrm{ALT}. \tag{3}$$

When the LSF is less than zero, then an accurate location of damage will be attained from the SHM system. An event where the LSF is greater than zero indicates the occurrence of an error in damage localization; this error either indicates a false positive or false negative location. The Probability of Accurate Localization (POAL) can be expressed in Cartesian coordinates using the localization LSF as follows:

$$\mathrm{POAL(t)} = P\Big[\, O_{dd} = Dd \,|\, \big\{ \big| TDL(x,y,z,t) - PLD(x,y,z,t) \big| \big\} - ALT(x,y,z,t) \leq 0 \Big]. \tag{4}$$

Evaluation of damage severity is conducted by either quantifying or qualifying its magnitude. The error between the predicted damage severity (PDS) and actual damage severity (ADS) is implemented to construct a damage assessment LSF. In order to construct a suitable range of error associated with damage assessment, an acceptable damage severity threshold (DST) is required. The Probability of Accurate Assessment (POAA) is constructed using conditional probability as:

$$POAA(t) = P\Big[\, O_{dd} = Dd \,|\, \big\{ \big| ADS - PDS \big| \big\} \leq DST \Big]. \tag{5}$$

3 SIMULATED CASE STUDY

The proposed approach is applied to a numerical study in order to evaluate SHM performance. Vibration based Modal strain energy (MSE) by Stubbs *et al*. (1995) is employed to assess the integrity of fixed offshore platform with varying degrees of damage at its structure members. The topology and foundation stiffness of the fixed offshore platform are given in detail by Karadeniz (2001); the operational and environmental loads applied are presented in Table 1.

Finite element model of the fixed offshore platform was constructed using beam elements, where damage was modeled by reducing the stiffness of a given structure member. Steel was assigned as the material property of all members in the structure. Baseline response of the fixed offshore platform condition was initially attained where no damage was present in the fixed offshore platform, after which cracks were simulated at 8 different locations, where 10 levels of damage severities were

Table 1. Environmental and operational loads applied in the numerical simulation.

Mass of Deck	4800 ton
Water Depth	50 m
Significant wave height	2.5 m
Drag coefficient	1.3
Inertia Coefficient	2.0
Sea spectrum	Pierson-Moskowitz

simulated at each location. In this work only the first 3 modes of vibration are implemented during the SHM process.

Damage detection is conducted by MSE based on the observed change in each structure member j at vibrational mode i using the following equation:

$$MSE_j = \frac{\sum_{i=1}^{3}\left[\phi_i^{dT} K_{j0} \phi_i^d + \sum_{j=1}^{118} \phi_i^{dT} K_{j0} \phi_i^d\right] \phi_i^T K \phi_i}{\sum_{i=1}^{3}\left[\phi_i^{T} K_{j0} \phi_i + \sum_{j=1}^{118} \phi_i^{T} K_{j0} \phi_i\right] \phi_i^{dT} K \phi_i^d}. \quad (6)$$

At each mode, λ_i represents the eigenvalue, ϕ_i represents the mode shape, K_{j0} represents the stiffness matrix of a structure member, and K is the global stiffness matrix (Stubbs *et al.*, 1995).

Damage localization in the structure is based on the normalization of the modal strain energy based on its mean σ_{MSE} and standard deviation $\overline{MSE}$ as (Li *et al.*, 2016):

$$ZMSE_j = \frac{MSE_j - \overline{MSE}}{\sigma_{MSE}}. \quad (7)$$

Assessment of damage in each element is conducted by quantifying the damage magnitude using the cross modal relationship in Equation (8), written in matrix form. Parameter α is the reduction factor in the stiffness of a given structure member. The magnitude of α varies from 0 to –1, where –1 indicates a total loss in stiffness (Li *et al.*, 2007):

$$[\alpha] = \left([\gamma]^T [\gamma]\right)^{-1} [\gamma]^T [\varphi], \quad (8)$$

where

$$\varphi = \left[\frac{\lambda_j^d}{\lambda_i} - 1\right] \phi_i^T K \phi_j^d, \gamma = \sum_{n=1}^{DE} \phi_i^T K_n \phi_j^d. \quad (9)$$

The values of O_{dd} attained from $ZMSE_j$ at different levels of damage are used to create a POD curve based on a fixed value of D_{th} which was set as 1. A hit/miss graph was plotted and fitted to the logistic binary regression model to calculate POD.

The performance function POAL was also assessed by implementing ALT, as a function of braced members sharing the same node with the damaged member. Furthermore, non-connected load sharing members on the same floor with the damaged member were also implemented to calculate ALT. The position of PDL was attained from $ZMSE_j$, which in conjunction with TDL, and ALT were used to generate hit/miss graph which was fitted to attain a POAL curve. A similar process was applied to achieve the POAA curve, where PDS was quantified using α, and DST was set at 20% of ADS.

4 RESULTS

The undamaged structures modal frequencies were 7.328 Hz, 7.328 Hz and 33.193 Hz for the first 3 modes respectively. The directions of excitement for the first two modes exhibited translation motion in the diagonal direction of the X-Y plane, while the third mode displayed torsional motion.

The 80 scenarios of damage assessed with indexes in Stubbs *et al.* (1995) resulted in at least one member element with $ZMSE$ value greater than 1, thus indicating the presence of damage for all damage scenarios. The performance of damage localization and severity were observed to be dependent on the extent of damage, type of structure member, and orientation of damaged structure member. Platform legs were most sensitive structure members to damage; a 0.5% decrease in stiffness was accurately localized. Furthermore, the estimated loss of stiffness was within 80% of the true stiffness value (see Figure 1). For diagonal structural members lying horizontal to the floor

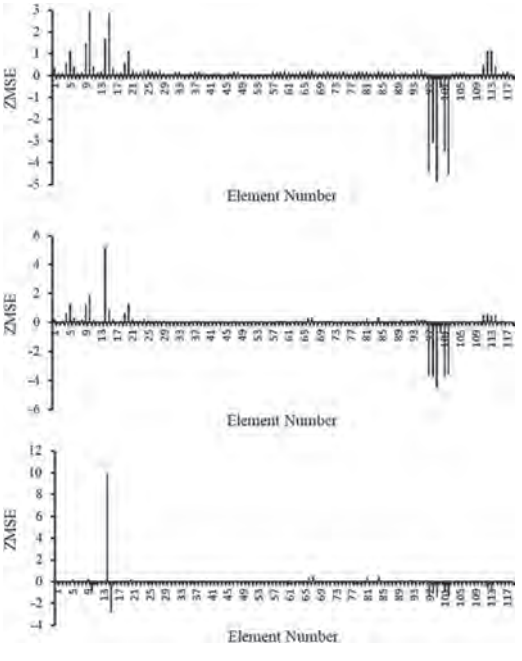


Figure 1. ZMSE outcome on structure member 5 (platform leg) at (top) 0.5% (middle) 10% (bottom) 50% reduction in member stiffness.

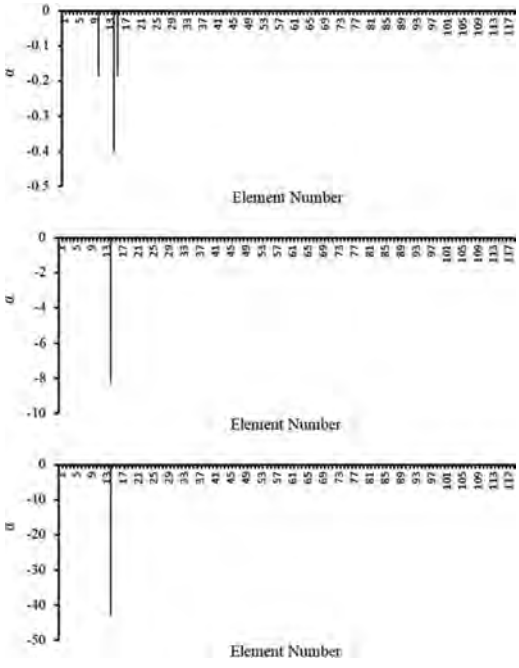


Figure 2. PDS outcome on structure member 14 (platform leg) at (Top) 0.5% (Middle) 10% (Bottom) 50% reduction in member stiffness.

plain, damage localization and severity estimation was unsuccessful across all damage scenarios. Damage in vertical bracings were accurately localized and estimated for damage scenarios where the reduction in member stiffness was 5% or greater. For horizontal bracings, a reduction in member stiffness of 15% resulted in accurate localization and estimation of damage.

Further analysis of all structure members with simulated damage indicated a correlation between the orientations of mode shape excitement, and the location of damaged structure member. Damaged structure members positioned in the diagonal planes of mode shape excitement were correctly localized with an acceptable PDS at lower magnitudes of ADS in comparison to equivalent structure members on the same floor (see Figure 2).

The data compiled from all 80 outcomes of ZME, and α were used to attain a model assisted POAL, and POAD via the hit/miss technique. A logit model was employed to generate both probability curves. The attained hit/miss data for damage localization and damage severity were equivalent in this study, when damage was not localized within the ALT, the magnitude of PDS was less than 80% of TDS.

The attained POAL and POAD curve represented in Figure 3 indicates that the vibration

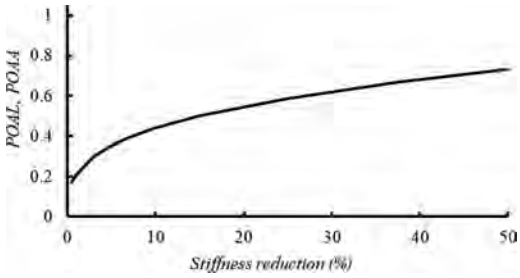


Figure 3. Probability of Accurate Localization (POAL), and Probability of Accurate Assessment (POAA) generated from hit/miss data.

modal strain energy method was able to correctly localize and assess 18% of damage that result in a 0.5% reduction in stiffness, as well as 75% of damage that generate a 50% reduction in element stiffness.

5 CONCLUSIONS

Two performance functions were presented for SHM systems with regards to its ability to accurately predict the location of damage and the damage severity. For a given damage magnitude, the Probability of Accurate Localization (POAL) quantifies the SHM systems ability to localize damage, while Probability of Accurate Assessment (POAA) is a measure of the predicted damage severity in comparison to the actual damage severity. These performance functions work in conjunction with the widely used probability of detection (POD) performance function to characterize three of the four major objectives for SHM systems.

Numerical studies were conducted to detect, localize and quantify damage on a fixed offshore platform under operational loads at varying damage magnitudes and locations using the Modal strain energy (MSE). The results indicate that MSE was able to detect damage in all simulated scenarios, thus leading to constant POD curve across all extent of damage. The POAL and POAA curves were identical, as all damage correctly localized also generated a satisfactory estimate for damage severity. The performance of MSE method with respect to POAL and POAA varied based on the type of member with an assigned damage, damage magnitude and its location with respect to the orientation of mode shape excitement.

The current numerical study did not include errors associated with Eigen value and Eigen vectors attained from the finite element model vibration data. However, noise is always present in real world measurements of vibration data. Hence,

future studies will include the addition of noise in vibration parameters.

REFERENCES

Aldrin, J.C. et al. Protocol for reliability assessment of structural health monitoring systems incorporating model-assisted probability of detection (MAPOD) approach. International Workshop on Structural Health Monitoring: From Condition-based Maintenance to Autonomous Structures. September 11–13, Stanford, California, USA.

Aldrin, J.C. et al. (2010, February). Model-assisted probabilistic reliability assessment for structural health monitoring systems. In: *AIP Conference Proceedings*, Volume 1211, No. 1, 1965–1972.

Annis, C. (2009). MIL-HDBK-1823A, Nondestructive Evaluation System Reliability Assessment. Standardization Order Desk, Philadelphia.

Doebling, S.W. et al. (1998). A summary review of vibration-based damage identification methods. *The Shock and Vibration Digest*, 30(2), 91–105.

Etebu E., & Shafiee M. (2017) Contributions of structural health monitoring to the reliability of an offshore fixed platform. In: European Safety and Reliability (ESREL) Conference, Portoroz, Slovenia, 18.6.2017–22.6.2017.

Karadeniz, H. (2001). Uncertainty modeling in the fatigue reliability calculation of offshore structures. *Reliability Engineering & System Safety,* 74(3), 323–335.

Li, H., Fang, H., & Hu, S. L. J. (2007). Damage localization and severity estimate for three-dimensional frame structures. *Journal of Sound and Vibration*, 301(3), 481–494.

Li, Y., Wang et al. (2016). An improved modal strain energy method for damage detection in offshore platform structures. *Journal of Marine Science and Application,* 15(2), 182–192.

Stubbs, N. et al. (1995). Field verification of a nondestructive damage localization and severity estimation algorithm. In: Proceedings-SPIE the international society for optical engineering, pp. 210–210.

Thompson, R.B. (2007). A Unified Approach to the Model-Assisted Determination of Probability of Detection. In: 34th Annual Review of Progress in Quantitative Nondestructive Evaluation, Volume 975; 22–27 July 2007, 1685–1692.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Serviceability criteria for structural design in prescriptive documents

J. Markova & M. Holicky
Klokner Institute, Czech Technical University in Prague, Czech Republic

L. Navarova
College in Ceske Budejovice, Czech Republic

ABSTRACT: The reliability of structures in the serviceability limit states is analysed. The Eurocode EN 1990 for basis of structural design provides general recommendations which should be further supplemented including classification of irreversible serviceability limit states with respect to assumed consequences of construction works. Further harmonisation of National Annexes to Eurocodes and other prescriptive documents is needed. The standards recommend almost identical serviceability criteria for similar environment which are further compared with differently calculated characteristic values of stresses, deflections or crack widths based on different analytical models and combinations of actions. An example of verification of the serviceability limit states of crack width for a reinforced concrete member indicates that the resulting crack width might be found in a rather broad range. The probabilistic methods are applied for the verification of the reliability of the structural member and also of crack width model. It is shown that the reliability level of the structural member designed according to Eurocodes fulfils the recommended target reliability level given in EN 1990.

1 INTRODUCTION

Construction works are designed using methods recommended in national, European or international standards. The first generation of Eurocodes allows the national selection of about 1600 Nationally Determined Parameters (NDPs) including alternative design approaches, load combinations, values of partial factors for actions and material properties and other reliability elements, and also serviceability constraints. Therefore, the reliability of a designed structure depends on applied national codes or specified parameters NDPs in the National Annexes to nationally implemented Eurocodes in CEN Member States (MS). It is expected that in the second generation of Eurocodes the number of NDPs will be reduced and most procedures harmonised.

The load-bearing capacity and serviceability of a structure designed in accordance with national codes or nationally implemented Eurocodes could be expected within a broad range. The actual structural resistance depends not only on used theoretical models and selected partial factors and other reliability elements, but also on prescriptive rules including structural detailing. Moreover, in some cases the theoretical models given in various standards for determining structural resistance provide different probability of over-crossing the specified design value.

It is shown that the reliability of structural members designed for the serviceability limit states according to current national standards, Eurocodes and also fib Model Code have a considerable scatter which should be analysed and further harmonised.

2 DESIGN PROCEDURES IN CODES

The partial factor method which is a basic method for the structural design in Eurocodes, in international standards including ISO 2394 (2015) and also in many national standards, deals with uncertainties of basic variables by means of design values assigned to the variables. The design limit state function may be expressed in terms of the set of design values of a vector $\boldsymbol{x}$ of basic variables given as

$$g(\boldsymbol{x}_\mathbf{d}) = \mathrm{g}(F_\mathrm{d}, f_\mathrm{d}, a_\mathrm{d}, \theta_\mathrm{d}, C_\mathrm{d}) \quad (1)$$

where F_d = the design value of action, f_d = the design material property, θ_d = the design model of uncertainty, a_d = the design value of geometrical quantity and C_d = the serviceability constraint.

The design condition is given as

$$g(\boldsymbol{x}_\mathbf{d}) \geq 0 \quad (2)$$

for the design vector of basic variables in the limit state function which is commonly obtained on the

basis of the characteristic values of basic variables and a set of partial factors for actions and material properties.

For the verification of the serviceability limit states, the following inequality is given

$$E_d \leq C_d \tag{3}$$

determined on the basis of relevant combination of actions following EN 1990 (2002). The limiting serviceability constraints are not defined here, however they are newly proposed in the final draft of revised prEN1990 (2017).

For the verification of various serviceability requirements, the characteristic or quasi-permanent combinations of actions are often applied for which some serviceability constraints are provided in material oriented Eurocodes. It should be noted that further guidance for serviceability constraints are commonly given in the National Annexes of nationally implemented Eurocodes of CEN MS taking into account the effects of short-term and long-term duration of actions and various design situations. However, the serviceability constraints (e.g. the limit deflection, the limit crack width) are themselves subjected to uncertainties, and should be therefore included in the probabilistic assessment.

3 RELIABILITY ANALYSIS

The knowledge of the reliability level of the structure designed according to the national codes or nationally implemented Eurocodes and also the reliability (credibility) of prescriptive analytical models can be used for optimisation of design procedures or for further harmonisation of standards.

The structural member or theoretical model may be considered as reliable, if the condition $p_{F<}$ p_t is satisfied where the failure probability p_F is given as

$$p_F = \int_{g(x)<0} \varphi_x(x)\,\mathrm{d}x \tag{4}$$

The failure probability can be expressed by the reliability index $\beta = -\Phi^{-1}(p_F)$, where Φ is the distribution function of standardised normal variable. The failure probability p_t is the target value that should not be exceeded during the intended reference period.

The reliability differentiation of structures in EN 1990 (2002) is based on three different levels of failure consequences with respect to the ultimate limit states (Consequence Classes CC1 to CC3).

However, for the serviceability limit states similar differentiation has not been provided yet. In some cases this differentiation of structures in serviceability limit states might also be useful for distinguishing potential consequences.

EN 1990 (2004) recommends for the reversible serviceability limit states the target reliability index $\beta_t = 0$ and for the irreversible limit states $\beta_t = 1{,}5$ (for the fifty year reference period). Some further recommendations for the target values of reliability indices in the serviceability limit states are given in the JCSS Probabilistic Model Code (2014) where three consequence classes are proposed. Therefore, for structures categorized to reliability class RC1 the target reliability index $\beta_t = 1{,}3$ and for structures in class RC3 the target value $\beta_t = 1{,}7$ could be considered.

The reliability analysis of structural members for the ultimate or serviceability limit states can be determined through the probability p_{F1} of the action effects $E(\boldsymbol{X})$ randomly exceeding the structural resistance $R(\boldsymbol{X})$ according to the following relationship

$$p_{F1} = \mathrm{P}\{(\xi_R\, R(\boldsymbol{X}) - \xi_E\, E(\boldsymbol{X})) < 0\} \tag{5}$$

where $\boldsymbol{X}$ = vector of basic variables, ξ_R and ξ_E = coefficients of model uncertainty of resistance and action effects.

The reliability (credibility) of theoretical models given in standards can be analysed by means of the credibility of specified design value $v_d(\boldsymbol{x}_d)$ (e.g. deformation, crack width) determined on the vector of design variables.

The probability p_{F2} of exceeding the design value $v_d(\boldsymbol{x}_d)$ specified according to relevant theoretical formulae and recommendations of prescriptive design procedure, can be analysed

$$p_{F2} = \mathrm{P}\{(v_d(\boldsymbol{x}_d) - \xi_E\, v(\boldsymbol{X})) < 0\} \tag{6}$$

The serviceability requirements in the limit states of crack width are analysed for an example of a selected reinforced concrete member as follows.

4 CRACK WIDTH VERIFICATION

Cracking in reinforced concrete members due to the load effects can be controlled by applying theoretical (analytical) crack width model recommended in Eurocodes or other codes or fulfilling appropriate practical rules. In common cases the prescriptive rules for detailing and stress control are applied according to nationally implemented Eurocodes. However, in some cases the calculation of crack width is needed e.g. for the design of water retaining structures.

Presently various theoretical models exist for predicting the characteristic crack width w_k. Some crack width model may be found nearly in any standard for the design of concrete structures. The preliminary ENV Eurocode for the design of concrete structures and also the Eurocode EN 1992-1-1 (2004) provide different models for the assessment of crack width. The Model Code 2010 (2012) currently introduces different theoretical crack width model than the previous MC 1990 (1998). Presently a new crack width model is proposed in the second generation of EN 1992-1-1 which is presently under development. Evidently the theoretical models have not been well established till now.

The design of the structural members for the serviceability limit states of crack width is based on the inequality between the analyzed crack width and the crack width limit given as

$$w_k(\boldsymbol{x}_k) \leq w_{lim} \tag{7}$$

where the characteristic value of crack width $w_k(.)$ for a vector of characteristic values $\boldsymbol{x}_k$ of basic variables is specified on the basis of a prescriptive formula and w_{lim} is the crack width limit.

For the analysis of crack width, the basic variables are commonly taken into account as deterministic ones where the geometric properties are mostly considered by nominal values, the actions and material properties by characteristic values. Therefore, the calculated values of crack width might have different statistical meaning (mean, characteristic, extreme). The variability of the basic variables may considerably influence the resulting value of crack width. It appears that the design value of crack width may be estimated only with certain reliability.

Moreover, the specification of the required limit w_{lim} for crack width is mostly based on past experience without appropriate scientific backgrounds. The crack width limits should be determined on the basis of given performance requirements, taking into account commonly adverse environment of the structure and possible consequences of excessive cracking.

5 ANALYSIS OF SLAB CRACK WIDTH

As an example, for the analysis of the theoretical models of crack width, a reinforced concrete member subjected to bending moment due to external forces is presented. A simply supported reinforced concrete slab having thickness from 0,19 to 0,29 m, span of 5 m is considered to be loaded by permanent and recommended imposed load of category B (office areas) given in EN 1991-1-1 (2002).

The slab is designed for the Ultimate Limit States (ULS) according to the Eurocodes, but in some cases considering also rules of selected national standards (in those cases the national values of partial factors are applied). The limit states of crack width of concrete slab are verified with respect to the theoretical crack width models given in Eurocodes, Model Code 2010 and also in some national standards. The common crack width limit w_{lim} = 0,3 mm is considered here. It may be noted that the requirement of various standards for limit crack width is almost identical for the similar type of environment. However, different combinations of actions are in some cases recommended in national standards for verification of serviceability limit states of crack width (e.g. quasi-static or serviceability loads combinations). Thus, different values of crack width w_k determined on the basis of a broad range of prescriptive recommendations are being compared with the same limiting value w_{lim}.

The resulting values of crack width of a reinforced concrete slab based on different models given in the Eurocode EN 1992-1-1 (2006) and its prestandard ENV, in Model Code 2010 (2012) and also in the previous document Model Code 1990, in ACI 318 (1989), BS 8110 (1989) and CSN 73 1201 (1986) are illustrated in Figure 1, partly based on the previous works Holicky 2007, Markova & Holicky (2014). It should be noted that the names of standards are shorten in all Figures (the standard in parenthesis indicate that the slab is designed for the ultimate limit states according to national recommendations with respect e.g. to thickness of concrete cover and load combination.

It is shown that the value of crack width is influenced by selection of theoretical crack width model and also previous design of structural member for the ultimate limit state. Prescriptive documents give different requirements for material properties,

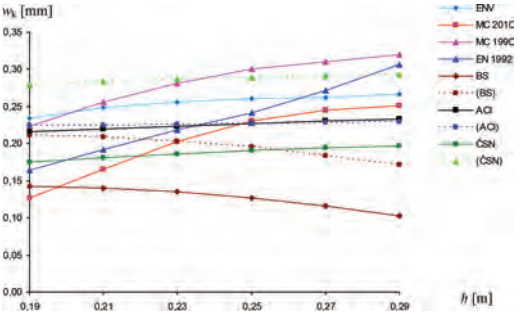


Figure 1. Characteristic values of crack width w_k of a reinforced concrete slab based on selected models in Eurocodes, Model Code documents and selected national standards.

geometrical parameters, loading, structural detailing, stress limitation, minimum area of reinforcement and its cover thickness.

6 PROBABILISTIC CRACK WIDTH VERIFICATION

The time-independent reliability analysis of the slab for the limit state of crack width is dealing with the probability p_{F1} of the random crack width $w(\boldsymbol{X})$ over-crossing the required constraint w_{lim} expressed by

$$p_{F1} = P\{\xi_{lim}\, w_{lim} - \xi_w\, w(\boldsymbol{X}) < 0\} \tag{8}$$

where $\boldsymbol{X}$ is a vector of basic variables and ξ_{lim}, ξ_w are the model uncertainties of the requirements on the crack width limit and the model of crack width, respectively. The structure is assumed to be reliable if the inequality is satisfied

$$p_{F1}\,(\boldsymbol{X}) \le p_{Ft} \tag{9}$$

where p_{Ft} = the target probability of failure that should not be exceeded during the design working life of the structure. Three different levels of the serviceability performance of structures should be distinguished—irreversible, reversible and long-term.

The probabilistic models of basic variables entering the equation (6) are based on recommendations of PMC (2014) and previous reliability analyses developed in the Klokner Institute CTU. Some of the models applied in the reliability analyses are assumed to be deterministic values, while the others are considered as random variables having the normal, lognormal, beta or gamma distribution. Statistical properties of basic variables are described using the moment characteristics (by mean, standard deviation), lower and upper bounds.

The theoretical models assume different probabilities of exceeding the characteristic value of crack width, or the maximum crack spacing. The probability of over-crossing the characteristic crack width w_k is 5% according to Eurocodes [1,3], Model Code 2010 and CSN 73 1201 (1986), 10% in CEP FIP Model Code 1990 and 20% in BS 8110 (1989). The limit state function is based on the theoretical model of mean crack width. Therefore, it is excluded that the theoretical models take into account different probability of over-crossing design value of crack width, or specified maximum distance of cracks.

The results of reliability analysis of the reinforced concrete slab for the limit state of crack width are illustrated in Figure 2.

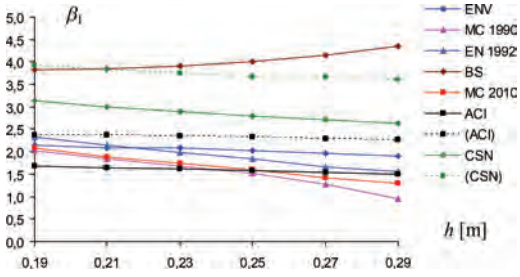


Figure 2. Reliability of reinforced concrete slab of height h for the limit states of crack width.

Analysis of the sensitivity factors α of basic variables indicates that the significant basic variables influencing the design crack width include permanent and variable loads, thickness of the cover of reinforcement, tensile strength of concrete and influence coefficients (e.g. expressing the bond strength, duration of the loading, shape of the strain across cross-section).

7 CREDIBILITY ANALYSIS OF CRACK WIDTH MODELS

The credibility (reliability) of the specified characteristic value of crack width w_k is verified. The probability of the random variable $w(\boldsymbol{X})$ exceeding the crack width w_k determined in accordance with relevant theoretical model of particular standard is expressed as

$$p_{F2} = \mathrm{P}\{w_k(\boldsymbol{x}_k) - \xi_w\, w(\boldsymbol{X}) < 0 \tag{10}$$

where $\boldsymbol{x}_k$ is the vector of characteristic values of basic variables and the coefficient ξ_w represents uncertainties of action effects and inaccuracy of the crack width model.

The probability p_{F2} of the random crack $w(\boldsymbol{X})$ over-crossing the crack width w_k according to relationship (7) for the slab expressed here by reliability index β_2 is illustrated in Figure 3.

Analysis of the credibility of specified crack width w_k indicates, that the reliability index β_2 determined for a reinforced concrete slab appears to be low for the theoretical models introduced in national American and British standards and rather high in Czech standard. The credibility of theoretical models seems to be sufficient in EN 1992–1-1 (2004), and a little bit higher than the credibility of the model provided in Model Code 2010.

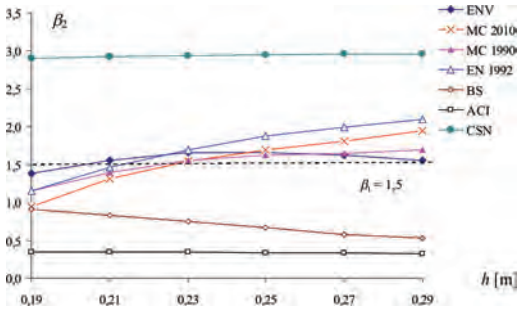


Figure 3. The credibility of the determined crack width w_k for selected theoretical models.

8 CONCLUDING REMARKS

Presently the basic Eurocode EN 1990 for the design of structures introduces general recommendations for the verification of the serviceability limit states. The development of supplementary provisions is needed including classification of structures based on the consequences of failure. Deterministic methods of structural analysis commonly used for the verification of structures in the serviceability limit states do not enable objective evaluation of structural reliability. The probabilistic methods facilitate comprehensive analysis of the reliability of a structure and assessment of credibility of used theoretical models.

An example of verification of a reinforced concrete slab with respect to the limit state of crack width indicates that the same limiting serviceability constraints (crack width limit) is compared with characteristic values of crack widths obtained on the basis of a broad range of normative recommendations and also having different statistical meaning.

The methods of structural reliability enable realistic analysis of concrete members with respect to the crack width. Reliability indices β assessed in the analysis of the credibility of the analytical crack width formulae and the reliability of reinforced concrete slab with respect to limit crack width have a significant scatter and in some cases seem to be rather low.

The credibility of theoretical models for crack width is independent on the previous design of a reinforced concrete slab for the ultimate limit states of Eurocodes or relevant national standards and may be applied for calibration purposes.

Analysis of sensitivity factors α indicates the significant basic variables influencing the reliability index and the credibility of the design crack width including permanent and variable loads, thickness of the cover of reinforcement, tensile strength of concrete and influence coefficients (expressing the bond strength, duration of loading).

It appears that the probabilistic methods can be effectively used for the development and calibration of new theoretical models applied in design and verification of structures. They may be applied for further harmonisation of parameters NDPs in Eurocodes.

ACKNOWLEDGEMENT.

This work has been supported by he Czech Science Foundation under Grant 16–04132S.

REFERENCES

ACI 318–89. 1989. Manual of Concrete Practice, reported by ACI Committee 301. American Concrete Institute.
BS 8110. 1989. Structural Use of Concrete. Part 2: Code of Practice for Design and Construction, British Standards Institution, London and Amendment 1.
Cervenka V., Markova J., 2017. Sykora M. et al. Uncertainties of Crack Width Models. In: *fib* 2017
CSN 73 1201. 1986. Design of concrete structures. ÚNMZ. pp. 93 (in Czech)
EN 1990. 2002. Basis of structural design. 2002
EN 1992–1–1. 2004. Design of Concrete Structures—Part 1: General Rules and Rules for Buildings.
fib Bulletin 34. 2005. Model Code for Service Life Design.
Holický, M. & Marková, J. 2007. Probabilistic Design of Structures for Durability. In: ESREL 07. pp. 1757–1762.
Holický, M. 2009. Reliability Analysis for Structural Design. Stellenbosch. pp. 199.
ISO 13822. 2010. Bases for design of structures—Assessment of existing structures.
ISO 13823. 2008. General principles on the design of structures for durability.
ISO 2394. 2015. General principles on reliability for structures
Marková J. & Sýkora M.. 2016. Uncertainties in Crack Width Verification of Reinforced Concrete Structures. In: ESREL 2016
Mlcoch J., Marková J., Sýkora M., 2017. Uncertainty in Crack Width Estimates According to fib Model Code 2010; In: *Transactions of the VSB—Technical University of Ostrava. Construction Series*, 17/1, 2017
Model Code 2010. 2012. *Design of concrete structures*. fib Bulletins 55 and 56.
Probabilistic Model Code. 2014. JCSS.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Sealing life evaluation of soft-packed power batteries based on ADT and modified CZM

W. Zhang, Y.M. Liu & Y.X. Chen
School of Reliability and Systems Engineering, Beihang University, Beijing, P.R. China

H. Sun
Contemporary Amperex Technology Limited, Fujian, P.R. China

ABSTRACT: The soft-packed power battery is becoming a promising power source because of its high energy density and light weight. The duration of sealing has a significant impact on the service life of the soft-packed battery. However, currently there is no effective test method or theoretical models to evaluate its sealing life. In this paper, an Accelerated Degradation Test (ADT) is designed and performed to investigate the sealing strength degradation. The testing is conducted to measure the degradation rates and residual tearing strengths of the standard specimens under the different accelerated stresses. Based on the testing data and membrane theory, a modified Cohesion Zone Model (CZM) is proposed to describe the tearing curve and strength degradation. This model is validated by the testing data well.

1 INTRODUCTION

The soft-packed power battery, as the heart part of electric vehicles, has become the focus of study because of its light weight and high energy density compared with the traditional hard-shell battery (Gallagher, K.C. 2016, Lu, L. 2013, Lee, H. 2014). However, the sealing of soft-packed battery is still has to face some challenges. The sealing duration becomes one of the key factors severely restricting the reliability and safety of the battery, because the sealing is always subjected to tearing load due to slow and constant increase of the inner gas pressure. So, this paper is devoted to investigating the tearing behavior of the sealing and developing an effective testing method and analytical model for its life prediction.

Many methods are proposed to describe film tearing behavior, such as Kim model (Kim, K.S. 1988), Wei-Hutehinson model (Wei, Y. 1998) and cohesion zone model (Barenblatt 1959, Dugdale 1960). Among them, cohesion zone model (CZM) is a more perfect and significant mathematical tool for characterizing crack propagation. In the study of elastic and plastic materials, it can effectively simulate plastic deformation and creep behavior ahead of the crack tip (Barenblatt 1959, Dugdale 1960, Needleman 1987, 1990). Needleman proposed a unified theoretical framework for the cohesion zone from initial degumming to complete detachment (Needleman 1987). Kent used the trapezoidal cohesion model to study the composite cracking process of the film adhesive layer (Kent 2008). Zhao used the CZM to simulate the tearing process between metal film and ceramic substrate (Zhao, H.F. 2008). However, these methods can just describe the instantaneous tearing strength. They do not consider the strength degradation due to viscous deformation of the polymer material under the constant load over a long time period. The theoretical analysis concerns about the tearing strength degradation with time and the size of any moment. Therefore, the models need to be modified considering degradation characteristics of product performance. It is deserved to develop a CZM-based method for the tearing strength degradation of sealing.

To obtain the information of degradation, accelerated degradation test (ADT) is adopted in this study. It is applicable to products with long life and strength degradation. ADT accelerates performance degradation by increasing the level of stress under the condition of keeping the degradation mechanism unchanged and collect performance degradation data (Fallou, B. 1979). These data are then used to estimate the reliability of the product and to predict the life of the product at normal stress levels. Nelson first studied the accelerated degradation test, analyzed the accelerated degradation test data of the insulation material, and obtained the service life of the insulation material under the normal stress (Nelson, W. 1990). Constant stress accelerated degradation test is widely applied in product life assessment, the operation of

which is simple and whose data statistics method is mature (Ma, X.B. 2011). This is obviously suitable for lithium-ion batteries that is a long-life and highly reliable product (Meeker, W.Q. 1995). It is considered that there is a significant degradation characteristic for the behavior that the residual strength of the sealing decreases, for the result that the polymer layer of hot-adhesive soft-packed will have a viscous deformation and be eventually torn under continuous stress. Most of the study still focuses on the correlation between strength and displacement of the sealing, which directly results in deficiency in the mature life evaluation method of the sealing. There is no study of the degeneration of the sealing tearing strength.

In this study, first, the behavior and mechanism of sealing tearing strength degradation are analyzed. A constant stress ADT is designed and implemented. Then, a modified CZM is developed to evaluate strength degradation using testing data. Finally, some conclusions and discussions are given based on the current research.

2 TESTING

2.1 *Specimens and experiment set-up*

The product tested in this study is outsourcing aluminum-plastic film of the soft-packed lithium-ion battery—PA/AL/CPP composite film. The innermost layer is CPP (Cast Polypropylene), which has a high thermal viscosity and is used for fusion when the upper and lower heads are pressurized at high temperature; The middle layer is AL (Aluminum Foil) layer, which is the carrier of the heat-sealing material and prevents water penetration; The outermost layer is PA (nylon), which has a certain anti-puncture performance and plays a decorative role. Fig. 1 and Fig. 2 show the product and product stress diagram respectively. The battery will produce the gas inside in the process of storage or application. The gas pressure causes the sealing of the aluminum plastic film to be torn, that results in the failure of the battery.

The standard specimens are cut out of the seal side with a spill area as shown in Fig. 3. The width W is 8 mm and the length L is 40 mm.

Test system consists of two parts: in situ fatigue test bench and optical measurement microscope, shown in Fig. 4. The test stage manufactured by Care Test Company is mounted on a stage of a measuring microscope and the sample was clamped by the stage. During loading/unloading, images were captured at high resolution using an industrial miniature camera. At the same time, experimental data was recorded and analyzed by Care-Test-fatigue system.

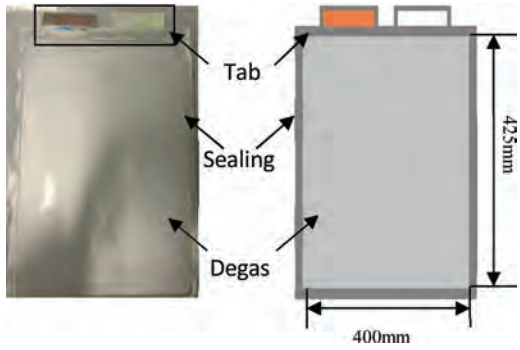


Figure 1. CPP soft-package lithium-ion battery.

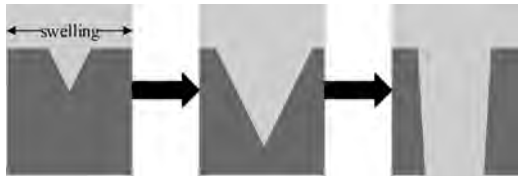


Figure 2. Sealing tearing process caused by the inner gas pressure.

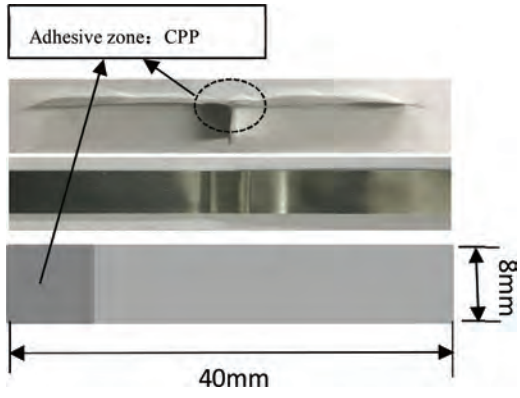


Figure 3. The illustration of specimen.

2.2 *Experimental procedure*

2.2.1 *Strength CURVE measurement*

In order to standardize the tearing strength of the sealing and define the stretch rate of the accelerated degradation test, perform direct tearing test first. The specimen is fixed as shown in Fig. 5. The displacement control of the fatigue stretch test machine was carried out, and several tests were performed at different stretch rates. Tearing direction is perpendicular to the adhesive area and tearing angle is 90 degree. It is known that the tearing process of polymeric materials is influenced by

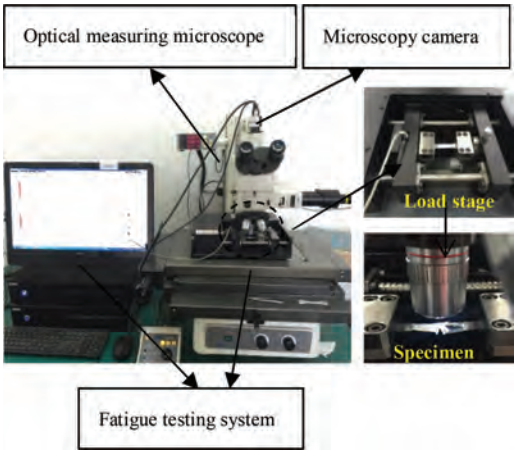


Figure 4. In-situ optical microscopy fatigue testing system.

Figure 5. Specimen installation method.

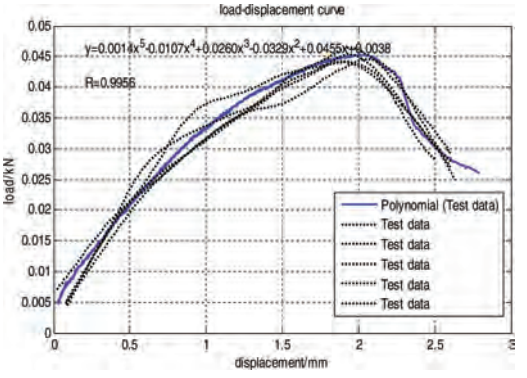


Figure 6. Schematic illustration of curve fit to test results.

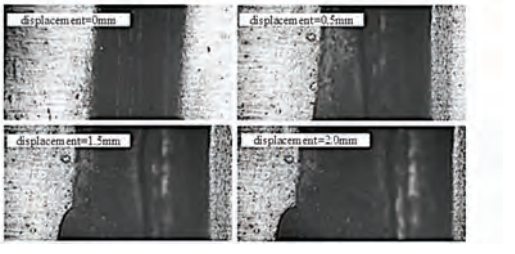


Figure 7. The physical photo of sealing tearing process.

temperature and load (Zhang, G.J. 1996). The test is usually carried out at room temperature and the load is a more sensitive factor. Therefore, only the tearing behavior at room temperature and strength degradation behavior under different constant loads are studied in this paper. The test temperature controlled at 25 degrees. Record the maximum tensile load. According to GB/T 22638.7-2008 ⟨⟨Test Method for Aluminum Foils—Part 7: Tearing strength⟩⟩, the tearing strength is defined as the maximum tensile load at a fixed width. Five specimens were tested repeatedly at each tearing speed, and taking the average value as the tearing strength at that speed.

The load-displacement curves are obtained by Care-Test-fatigue system. The tearing curves can converge when the stretch rate drops to 1 N/mm. Therefore, the ADT stretch rate is defined as 1 N/mm. In Fig. 6, the load-displacement curves obtained from the five sets of data are different. Fig. 7 indicates the process of the sealing tearing. Therefore, average the test results. The maximum in five sets of the load data are 45.37, 44.92, 44.96, 45.61, 44.14 N and the average is 45.00 N. The corresponding displacements are 2.03, 2.18, 1.99, 2.00, 1.80 mm and the average is 2.00 mm. In order to illustrate the general mechanics of the sealing adhesive interface, the five groups of curves are fitted to a standard curve using the least squares method, as shown in Fig. 6.

There is almost no change in the displacement when the load is less than 5 N; when the load is between 5 and 20 N, the displacement increase is approximately linear and can be regarded as the elastic stretching stage; when the load is more than 20 N, the displacement change rate gradually increases. After the load reaches a maximum of 45 N, the sealing is torn rapidly until it is completely torn. Therefore, this sealing tearing strength is calibrated as 45 N. According to the above test results, combined with the principle of degeneration, the constant load of ADT can be controlled between 5 and 20 N and the tests under different load conditions are carried out.

2.2.2 *Accelerated degradation test*

Without changing the failure mechanism of the product, the test conditions are increased to accelerate the failure of the specimens, and the performance degradation data under the high stress level are used to extrapolate the service life of the product under normal use stress. This is the basic principle of the ADT. The specific method is: under

different loads: 5, 10, 20 N; different duration: 20h, 40h, test sealing degradation rate of tearing strength. According to the relationship between time and tearing displacement, the regression curve equation and judgment coefficient are obtained. Under different constant loads, the curves of tearing strength degradation of sealing are shown in Fig. 8 when the duration is 20 h. It can be seen that the sealing is slowly torn with time even if the load is constant. This shows that the tearing strength has been degraded. As the constant load increases, the degradation rate gradually increases.

In order to measure the residual strength after degradation, displacement control was used to perform the tearing test based on the ADT. The result is shown in Fig. 10. And Fig. 9 indicates the process of the sealing tearing.

As it can be seen from the above figure, the maximum tearing strength gradually decreases as the constant load increases while the time is the same. That is when the load time is 20 h and the constant load is 5 N and 10 N respectively, the corresponding

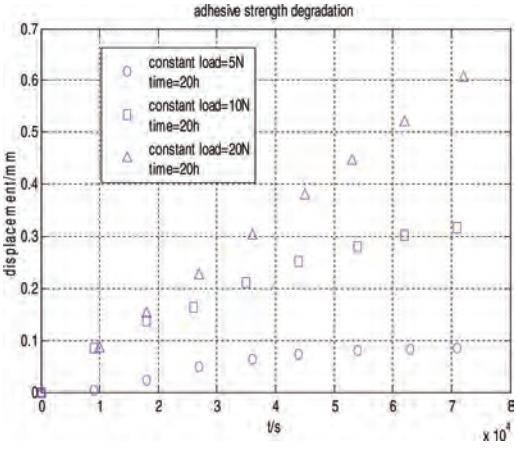


Figure 8. Schematic illustration of relationship between the displacement and time.

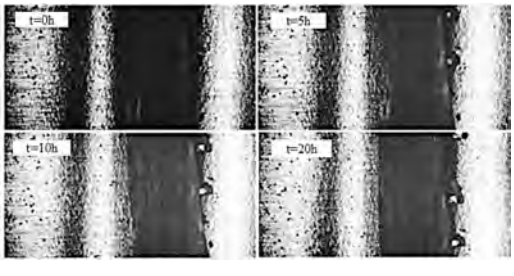


Figure 9. The physical photo of sealing tearing process.

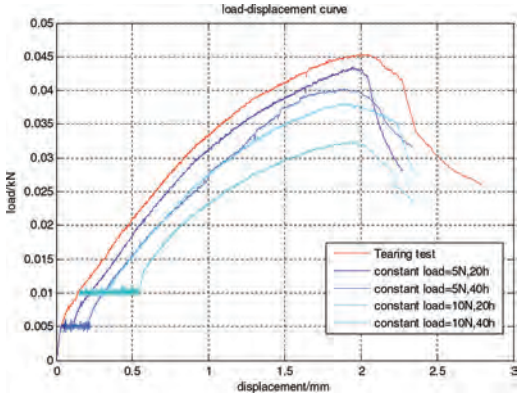


Figure 10. Schematic illustration of curve fit to test results.

tearing strength peak are: 43 N, 38 N; Maintaining a constant load unchanged, with the loading time gradually increases, the maximum tearing strength decreased. That is when the constant load is 10 N and the loading time is 20 h and 40 h respectively, the peak of the tearing strength corresponding to complete tearing is 38 N and 33 N respectively. This shows that the cumulative effect of load has a significant effect on the sealing tearing. It is that a constant load applied to the sealing for a period of time causes the tearing strength to decrease.

3 THEORETICAL ANALYSIS

3.1 *Exponent cohesion zone model*

The mechanical response of polymer materials always shows the combination of elasticity and viscous property, which is called viscoelasticity (Bower, D.I. 2002). CPP is subjected to thermomechanical treatment, such as stretching, at a temperature that is under the melting point and above a glass transition temperature (the glass transition temperature of the polypropylene material is –35°C). In this case, the orientation process of the polymer may occur. As a result of the orientation, the tensile strength and flexural fatigue strength of the polymer material significantly increase in the orientation direction. At the same time, the orientation perpendicular to the orientation direction significantly decreases. The macro performance is torn.

The CZM is based on elastic-plastic fracture mechanics and is used to investigate the elastic-plastic region of the crack tip. The constitutive relation of the cohesion zone is defined by the cohesion and the open displacement at the interface. On the basis of the test, the interface is characterized as a thin layer of the constitutive relation of the exponential cohesion zone model and the exponential

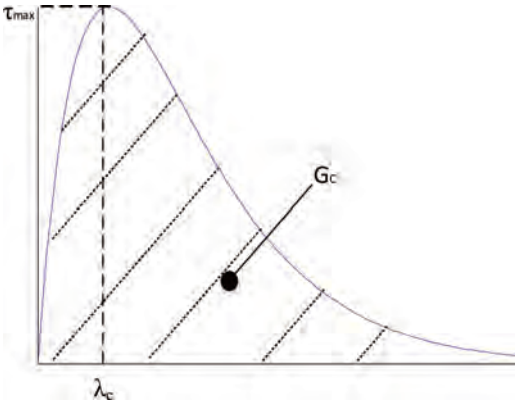


Figure 11. Tearing strength—displacement curve.

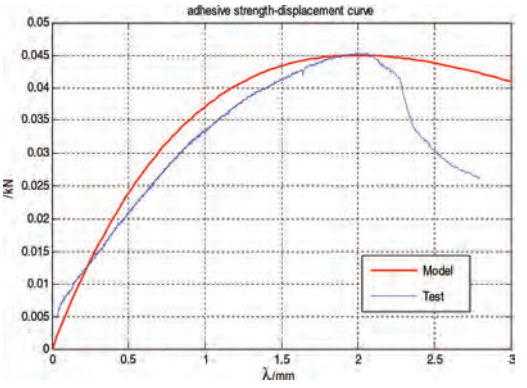


Figure 12. Comparison of tearing curve in CZM model and test result.

CZM is chosen to simulate the separation of the sealed boundary. The tearing strength is given by:

$$\tau = \tau_{max} \frac{\lambda}{\lambda_c} exp\left(1 - \frac{\lambda}{\lambda_c}\right) \tag{1}$$

In Fig 11, the displacement characteristic value λ and the tearing strength value τ are two independent parameters. However, the energy required to separate within a unit area is often used as a parameter. It is equal to the sum of the area under the tearing strength-displacement curve, expressed as follows:

$$G_c = \int_0^{\infty} \tau(\lambda) d\lambda \tag{2}$$

It can be determined by tearing test. If we use the energy required for separation G as a model parameter, then equation (1) can be expressed as:

$$\tau = \frac{G_c}{\lambda_c} \frac{\lambda}{\lambda_c} exp\left(-\frac{\lambda}{\lambda_c}\right) \tag{3}$$

In this paper, the required energy G (N/mm) and displacement eigenvalue λ /mm are two independent constitutive parameters. The relationship between tearing strength and displacement was measured by the test. The peak of strength was 45.00 N and the displacement was 2.0 mm. The calculated interfacial fracture energy was 244.66 N/mm. Fig. 12 shows the predicted results that experimental data on specimen are used to validate the CZM. Obviously, this CZM can give a satisfactory prediction. The sealing CZM expression is as follows:

$$\tau = \frac{0.24466}{2.0} \frac{\lambda}{2.0} exp\left(-\frac{\lambda}{2.0}\right) = f(\lambda) \tag{4}$$

3.2 *Modified CZM*

The change in the displacement of the adhesive area under a constant load is shown in Fig. 9. With the degradation of strength is considered, time and constant load are taken as degradation factors to modify the equation (4). Tearing process can be divided into three stages: to reach a constant load stage, to maintain a constant load loading stage, rapid tearing stage. Assumptions: the constant load is τ_0 /N: the displacement reached to constant load is $f'(\tau_0)$: the constant load loading time is t/s; the degradation rate is (determined by the degeneration rate curve)v; the displacement at the start of rapid tearing is $f'(\tau_0) + vt$. Then, the modified CZM can be written as:

A. to achieve a constant load phase: $\tau = f(\lambda)$
B. to maintain a constant load loading stage: $\tau = \tau_0$
C. rapid tearing stage: $\tau = f(\lambda) - f(vt)$

$$\tau(t, \tau_0, \lambda) = \begin{cases} f(\lambda) & 0 \le \lambda \le f'(\tau_0) \\ \tau_0 & f'(\tau_0) < \lambda \le f'(\tau_0) + v \times t \\ f(\lambda) - f(vt) & f'(\tau_0) + v \times t < \lambda \end{cases} \tag{5}$$

In ADT, different constant loads can change the rate of tearing strength degradation. To obtain the degradation rate under different constant loads, calculate the slope of the straight line in Fig. 8. The result is given in Table 1. Then, the simulation curve of the relationship between degradation rate and constant load is obtained as shown in Fig. 13 and expressed by the following equation:

Table 1. Tearing strength degradation rate under different constant load.

τ_0	v
N	$\times 10^{-6}$ mm/s
0	0
5	1.4
10	4.1

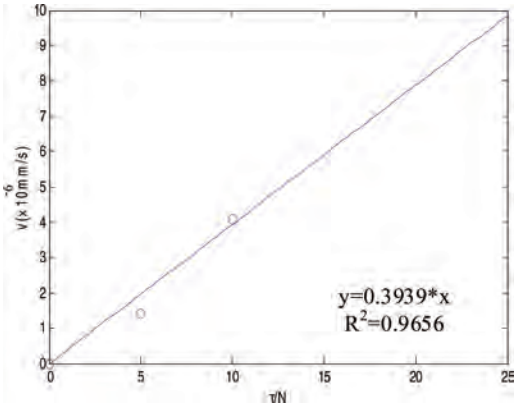


Figure 13. Schematic illustration of curve fit to test data.

$$v = 0.3939\tau_0 = g(\tau_0) \tag{6}$$

Substituting Equation (6) into Equation (5), the modified CZM can be expressed as:

$$\tau(t,\ \tau_0,\ \lambda) = \begin{cases} f(\lambda) & 0 \le \lambda \le f'(\tau_0) \\ \tau_0 & f'(\tau_0) < \lambda \le f'(\tau_0) + 0.3939\tau_0 \times t \\ \mathrm{f}(\lambda) - \mathrm{f}(0.3939\tau_0 \mathrm{t}) & f'(\tau_0) + 0.3939\tau_0 \times t < \lambda \end{cases} \tag{7}$$

4 MODEL VALIDATION

In this section, experiental datas on specimen with under different constant load and duration are used to validate the proposed modified CZM.

In Figure 14, the model predictions are compared to testing data under constant loading and duration. The horizontal axis is the tearing displacement and the vertical axis is the tearing strength. The red line is testing data and the blue line represent predictions of our proposed model. The constant load is 5 N and duration is 40 h in (a). And the (b) is a validation when the constant load is 10 N and duration is 40 h. The (c) is a validation when the constant load is 20 N and duration is 20 h. It is clear that the predictions are in good agreement with the testing data. Obviously, for the residual strength, the modified model can give satisfactory predictions. The strength degradation phenomenon due to constant loading is described well using the proposed model. Good agreements are observed between the model predictions and experimental data. Overall, this model can predict the strength degradation under constant loading and the duration.

5 CONCLUSION AND FUTURE WORK

In this paper, a modified CZM considering degeneration is proposed to evaluate the tearing strength of sealing and calculate the residual strength by ADT. The modified model and the testing data have a good match. It is beneficial to solve the effect of constant loading on the real-time tearing strength of the sealing. Then, the most use security problems will be avoided in a large part. Based on current research, the following conclusions can be given:

1. An ADT is designed to simulate the tearing process of the sealing by using standard specimen on the in-situ tensile machine. The tearing strength of the sealing degrades under a small static tensile load, and the sealing is torn over a long period of time. A positive correlation is observed between the degradation rate and load.
2. A modified exponent CZM considering degradation is proposed to depict the tearing strength degradation curve and residual strength curve of the sealing. The different strength degradation data are used for model validation. Good agreements can be seen between the testing data and model predictions.

The current study considers the stress only, and the effect of temperature and electrolyte will be investigated in the future.

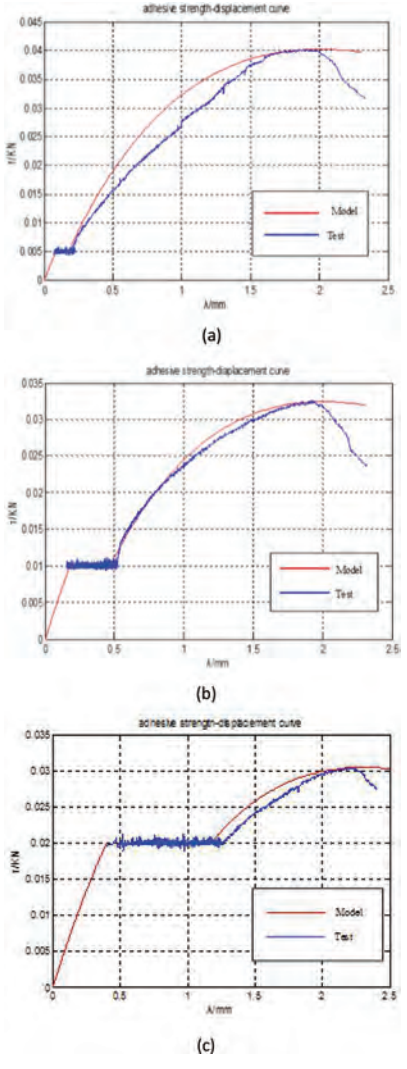


Figure 14. Comparison of tearing curve in modified CZM model and test result.

ACKNOWLEDGEMENT

The research is financially supported by National Key R&D Program of China (NO. 2016YFB0100400).

REFERENCES

Barenblatt, GI. 1959. The formation of equilibrium cracks during brittle fracture: general ideas and hypotheses, axially symmetric crack. *Applied Mathematics and Mechanics* 23: 622–636.

Bower, D.I. 2002. An Introduction to polymer physics. *Cambridge: Cambridge University Press* 2002: 162–219.

Cahmacho, GT. & Ortiz M. 1996. Computational modeling of impact damage in brittle materials. *International Journal of Solids and Structures* 33: 2899–2938.

Cotterell, B. & Hbaieb, K. & Wlliams, J.G. & Hadavinia, H. & Tropse, V. 2006. The root rotation in double cantilever beam and peel tests. *Mechanics of Material* 38(7): 571–580.

Dugdale, DS. 1960. Yielding of steel sheets containing slits. *Journal of the Mechanics and Physics of Solids* 8: 100–104.

Fallou, B. & Buruiere, C. & Morel, J.F. 1979. First approach on multiple stress accelerated life testing of electrical insulation. *NRC Conference on electrical insulation and dielectric Phenomena in Pocono*: 621–628.

Gallagher, K.C. & Trask, S.E. & Bauer, C. 2016. Optimizing areal capacities through understanding the limitations of lithium-ion electrodes. *Journal of the Electrochemical Society* 163(2): A 138-A 149.

Hadavinia, H. & Kawashita, L. & Kinloch, A. & Moore, D.R. & Williams, G. 2006. A numerical analysis of the elastic-plastic peel test. *Engineering Fracture Mechanics* 73(16): 2324–2335.

Lu, L. & Han, X. & Li, J. 2013. A review on the key issues for lithium-ion battery management in electric vehicles. *Journal of power sources* 226: 272–288.

Lee, H. & Yanilmaz, M. & Toprakci, O. 2014. A review of recent developments in membrane separators for rechargeable lithium-ion batteries. *Energy & Environmental Science* 7(12): 3857–3886.

Kent Salomonsson. 2008. Mixed mode modeling of a thin adhesive layer using a Meso-mechanical model. Mechanics of Materials 40: 665–72.

Kim, K.S. & Aravas, N. 1988. Elasto-plastic analysis of the peel test. *Solids Struct* 24: 417–435.

Ma, X.B. & Wang J.B., Zhao, Y. 2011. Reliability assessment using constant-stress accelerated degradation data based on pseudo life distribution. *System Engineering and Electronics* 33 (1): 228–232.

Meeker, W.Q. & Hamada, M. 1995. Statistical tools for the rapid development & evaluation of high-reliability products. *IEEE Transactions on Reliability* 44(2): 187–198.

Mettas, A. & Vassiliou, P. 2002. Modeling and analysis of time-dependent stress accelerated life data. *Proceedings Annual Reliability and Maintainability symposium* 343–348.

Needleman, A. 1987. A continuum model for void nucleation by inclusion deadhesiveing. *Journal of Applied Mechanics* 54: 525–531.

Needleman, A. 1990. An analysis of tensile decohesion along an interface. *Journal of the Mechanics and Physics of Solids* 38(3): 289–324.

Nelson, W. 1990. Accelerated Testing: Statistical Methods, Test Plans, and Data Analysis. *New York: John Wiley & Sons.*

Pardoen, T. & Ferracin, T. & Landis, C.M. & Delannay, F. 2005. Constraint effects in adhesive joint fracture. *Journal of the Mechanics and Physics of Solid* 53(9): 1951–1983.

Peck, D.S. 1986. Comprehensive model for humidity testing correlation. *Proceeding of 24th annual international reliability Physics symposium*: 44–50.
Tvergaard, V. & Hutchinson, JW. 1992. The relation between crack growth resistance and fracture process parameters in elastic-plastic solids. *Journal of Mechanics and Physics of Solids* 40: 1377–1397.
Wei, Y. & Hutchinson, J.W. 1998. Interface strength, work of adhesion and plasticity in the peel test. *Fracture* 93: 315–333.
Wei, Y. 2002. Thin layer splitting along the elastic-plastic solid surface. *International Journal of Fracture* 113(3): 233–252.
Wei, Y. 2004. Modeling nonlinear peeling of ductile thin films—Critical assessment of analytical bending models using FE simulations. *International Journal of Solids and Structures* 41(18): 5087–5104.
Yang, Q.D. & Thouless, M.D. & Ward, S.M. 1999. Numerical simulation of adhesively-bonded beams failing with extensive plastic deformation. *Journal of the Mechanics and Physics of Solids* 47: 1337–1353.
Yang, Q.D. & Thouless, M.D. & Ward.S.M. 2000. Analysis of the 900-peel test with extensive plastic deformation. *Journal of Adhesion* 72: 115–132.
Yang, Q.D. & Thouless, M.D. 2001. Mixed-mode fracture analyses of plastically-deforming adhesive joints. *International Journal of Fracture* 110: 175–187.
Yang, Q.D. & Cox, B.N. 2005. Cohesive models for damage evolution in laminated composites. *International Journal of Fracture* 113: 107–137.
Zhao, H.F. & Wei, Y.G. 2008. Inverse Analysis Determining Interfacial Properties Between Metal Film and Ceramic Substrate with an Adhesive Layer. *Acta Mechanica Sinica* 24(3): 297–303.
Zhang, G.J. & Pan, Z. & Lin, X.W. Yan, Y.J. 1996. Effect deformation orientation of polymer on crack propagation. *Materials of Mechanical Engineering* 3(20): 15~21.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Probabilistic analyses of existing power producing facilities

J. Markova & K. Jung
Klokner Institute, Czech Technical University in Prague, Czech Republic

K. Stastna
Institute of Applied Mechanics, Brno, Czech Republic

ABSTRACT: Assessment of existing components of power producing facilities is based on the probabilistic methods of the theory of structural reliability provided in Eurocodes and ISO standards. An example of quick-closing valves in a selected hydroelectric power plant indicates the assessment of reliability and prediction of remaining working life of a structural component for the considered model of corrosion. Moreover, assessment of reliability level of main steel structural members of existing industrial bridge for two alternatives of a cleaning machinery reveals rather low reliability level of some significantly deteriorated structural members of the industrial bridge in the power plant, and additional measures are to be applied.

1 INTRODUCTION

Various existing construction works in power-plants built in the Czech Republic are gradually deteriorating due to the effects of adverse factors having significant impact on their reliability. The main factors include adverse surrounding environment, chemical attacks, physical effects and repeated actions due to technological processes.

Actual technical state of power producing facilities needs to be regularly evaluated enabling optimal decision regarding their repairs, strengthening or replacement. Simplified conservative procedures for the reliability assessment of existing structures based on the methods commonly used in current codes for the design of new structures may lead to their expensive repairs.

That is why the assessment of existing structures of power-plants are based on the probabilistic methods of the theory of structural reliability given in the international standards ISO 13822 (2010) and ISO 13823 (2008).

The probabilistic assessment of existing energetics devices evoked by necessity of the assessment of their current reliability level, estimation of their remaining life time and also change of technological procedure evoked by need for more effective, however heavier machine in a selected hydroelectric power-plant in the Czech Republic given here as an example shows the procedure for the estimation of reliability, bearing capacity and residual life of existing structures for assumed corrosion models.

2 VERIFICATION OF WORKING LIFE

The reliability requirements for existing structures as well as for new ones may be expressed in terms of the failure probability P_f or reliability index β. The relationship between the both reliability indices is given as

$$\beta = \Phi^{-1}(P_f) \tag{1}$$

where $\Phi(.)$ denotes the standardized normal distribution function. The following reliability condition is required

$$P_f \le P_{ft} \text{ or } \beta \ge \beta_t \tag{2}$$

where P_{ft} and β_t are the target values of the failure probability and reliability index.

The target reliability level which should be used for the verification of existing structures may be based on calibrations taking into account the concept of the minimum expected costs and social risks. The recommended values of the target reliability index β_t for the verification of various limit states given in ISO 13822 (2010) are illustrated in Table 1.

More detailed target reliability indices depending on the consequences of failures and the relative costs for safe design are provided in ISO 2394 (2015).

EN 1990 (2002) gives principles of reliability differentiation in Annex B. Three reliability classes RC are recommended and the target values of reli-

Table 1. Indicative target reliabilities for existing structures.

Limit states	Target reliability index β_t
Serviceability—irreversible	1.5
Ultimate with failure consequences	
– very low	2.3
– low	3.1
– medium	3.8
– high	4.3

Table 2. Reliability differentiation of hydro-technical structures according to CSN EN 1990 (2011).

Reliability class	Examples of construction works
RC3	Dam, weir higher than 5 m, main conduit of potable water to town agglomeration
RC 2	Sewage disposal plant, accumulating pumping plant, water reservoir, aqueduct, hydro-electric power station, weir height up to 5 m, nautical channel
RC 1	Swimming pool, melioration structure

Table 3. Partial factors for actions.

	CSN 75 6303 (2012)		
Type of action	RC1	RC2	RC3
Hydrostatic pressure	1.0	1.1	1.2
Hydrodynamic pressure	1.2	1.3	1.4
Permanent action	1.2	1.35	1.4
Variable action	1.35	1.5	1.65

ability indices β_t proposed including examples of construction works. Presently hydro-technical construction works are not given in the scope of the current generation of Eurocodes and they need to be based on supplementary national prescriptive provisions.

Reliability differentiation of hydro-technical construction works is illustrated in Table 2 as recommended in the recently developed National Annex to CSN EN 1990 (2011) of the Czech Republic.

Partial factors for hydrostatic and hydrodynamic pressures, permanent and variable actions recommended in CSN 75 6303 (2012) for reliability classes RC1 to RC3 are given in Table 3.

Application of probabilistic methods for specification of working life of an existing structure

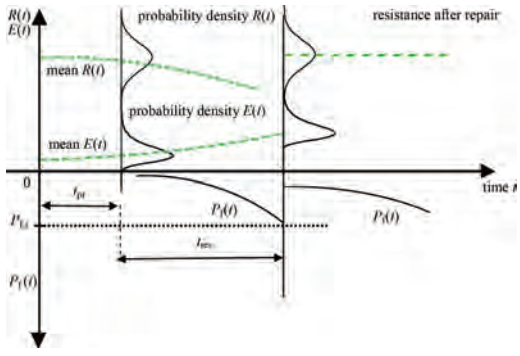


Figure 1. Probabilistic assessment of remaining working life.

is illustrated in Figure 1, Holický & Marková (2007).

It is assumed that the assessment (inspection) of an existing component of a power producing facility is performed in time t_{pr} from the beginning of the structure completion. In case that the time-dependent resistance $R(t)$ of a component and load effects $E(t)$ are known, the remaining working life of the component may be specified.

For estimation of the residual working life t_{res} of the component, the following expression is given as

$$P_f(t_{res}) = P\{R(t_{res}) - E(t_{res}) < 0\} \approx P_{f,t} \quad (3)$$

facilitating decision about its repair or replacing.

The general requirements for the assessment of structural reliability and safety for users are applied for reliability analysis of quick-closing steel valves and also for existing industrial steel bridge in the hydroelectric power plant operating about 50 years.

3 VERIFICATION OF QUICK-CLOSING STEEL VALVES

3.1 *Inspection of state of the quick-closing valves*

The inspection of the quick-closing valves of hydroelectric power plant revealed deterioration of the components including steel cover plates as illustrated in Figure 2.

3.2 *Reliability analysis based on partial factor method*

Firstly, the partial factor method given in EN 1990 (2002) and ISO 13822 (2010) is applied for verification of a cover plate of quick-closing steel valves which is one of main components.

The steel cover plate having thickness of 0,018 m is supported by stiffeners 0,8 m spaced.

Figure 2. Deterioration of a steel cover plate.

Hydrostatic and hydrodynamic components are considered for determination of water pressure on the plate given as

$$q = q_{hs} + q_{hd} = \rho g\, h + \rho Q\, v/A \tag{4}$$

where ρ = water density; h = depth; Q = flow rate; v = water speed; A = area of the cover plate.

The basic condition $M_{Rd} > M_{Ed}$ between the design value of bending moments for resistance and effects of actions should be satisfied. The reliability of the structural component is verified with respect to the ultimate limit state given as

$$M_{Rd} = \gamma_u\, b\, d^2\, f_{yk}/(4\, \gamma_M) \tag{5}$$

where γ_u = coefficient of model uncertainty of resistance, d = plate thickness, b = plate length, f_{yk} = characteristic yield strength, γ_M = material factor.

The characteristic yield strength is considered f_{yk} = 235 MPa, the partial factor γ_M = 1.15 and the coefficient γ_u = 0.85 according to CSN 75 6303 (2012).

The maximum water pressure is considered for determination of load effects for continuously supported cover plate given as

$$M_{Ed} = \delta\,(\gamma_{Qhs}\, q_{hs,k} + \gamma_{Qhd}\, q_{hd,k})\, L^2/12 \tag{6}$$

where γ_{Qhs} = partial factor for hydrostatic pressure, γ_{Qhd} = partial factor for hydrodynamic pressure which are considered here according to Table 3 for structures in class RC2. The dynamic amplification factor δ = 1.3 is considered here as recommended in CSN 75 6303 (2012).

The serviceability limit state of the component is also verified

$$M_{Ed} \le C_d \tag{7}$$

where C_d = limiting deformation.

Results of time dependent reliability analyses of a steel component with respect to both the ultimate (ULS) and serviceability (SLS) limit states are illustrated in Figure 3. The reliability of the cover plate significantly decreases with decrement Δd of plate thickness due to steel corrosion.

In case that the partial factor method is applied for verification of existing structural member in the common consequence class CC2, it is shown that its bearing capacity is exhausted after about 53 years and some measures are needed to be taken, e.g. strengthening or replacement.

3.3 *Application of probabilistic methods*

The probabilistic methods are applied for evaluation of the reliability level of the steel component, Holicky (2009).

The time-dependent corrosion $d_{corr}(t)$ of the steel cover plate is based on the relationship

$$d_{corr}(t) = A\, t^B \tag{8}$$

where A and B = parameters of analytical model which are considered for parameter A in a range from 0,03 to 0,13, and for parameter B from 0.6 to 0.7. It is assumed here for non-uniform corrosion

$$A = 0{,}06 \text{ mm};\ B = 0{,}7 \tag{9}$$

which is based on the results of inspection of the actual state of steel quick-closing valves. Pitting corrosion with pits up to 1 mm is shown in Figure 4.

It is considered that the steel cover plate of quick-closing valves (about 50 years old) is deteriorating due to non-uniform corrosion with average

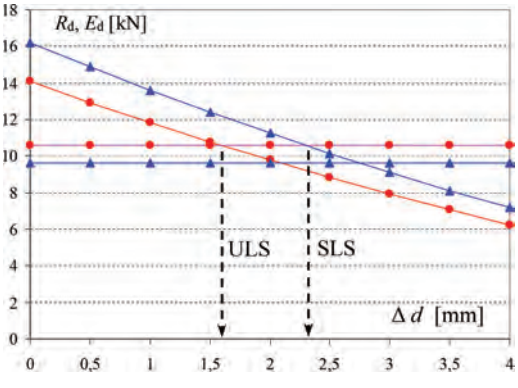


Figure 3. Progressive reduction of the component bearing capacity and serviceability with decrease of slab thickness Δd.

Figure 4. Pitting corrosion of a steel cover plate.

Table 4. Probabilistic models of basic variables.

Basic variable	Symbol	Distr.	Mean μ	C.V. V
Yield strength [MPa]	f_y	LN	280	0.08
Plate span [m]	L	DET	0.75	–
Plate thickness [m]	d	N	0.018	0.03
Plate width [m]	b	DET	1	–
Water pressure [kN/m^2]	q	N	157.6	0.1
Dynamic factor	δ	DET	1.3	–
Parameter A [mm]	A	N	0.06	0.10
Parameter B	B	DET	0.7	–
Resistance uncertainty	ξ_R	DET	0.85	–

one-side decrease up to 1 mm and about 30% probability of simultaneous weakening at opposite side of the cover plate. The model of corrosion given in equation (8) considering parameters in equation (9) leads after 50 years to actual average depth of corrosion of 1.3 mm.

The ultimate limit state is expressed as the difference $\Delta M(t)$ between the time-dependent resistance moment $M_R(t)$ and the bending moment M_E due to the water pressure q

$$\Delta M(t) = \xi_R b\,(d - 1{,}3 d_{\mathrm{corr}}(t))^2 f_y/4 - \delta q L^2/12 \qquad (10)$$

The probabilistic models of basic variables are given in Table 4 assuming normal (N) or lognormal (LN) distributions. Some of the basic variables are considered to be deterministic (denoted DET).

The reliability of the steel cover plate decreases in time due to the reduction of its thickness $d_{\mathrm{corr}}(t)$. The value of the target reliability index $\beta(80) = \beta_t = 3.7$ for considered 80 year working life of the cover plate is specified on the basis of required reliability index for a reference period of one year $\beta(1) = 4.7$ given as

$$\Phi(\beta(80)) = (\Phi(\beta(1))^{80} = (\Phi(4.7))^{80} = \Phi(3.7) \qquad (11)$$

The probabilistic reliability assessment of the steel member is based on own software tool developed in software Mathcad.

Initial reliability of the cover plate non-affected by corrosion (reliability index $\beta = 5.2$) satisfies the target reliability level $\beta_t = 3.7$ for considered reference period of 80 years. For one-side corrosion of 1 mm and cover plate 53 years old (current state), the reliability index decreases to $\beta = 4$, see Figure 5.

It appears that the working life of the cover plate may be estimated to approximately 70 years. Thus, the residual working life of a cover plate is considered to be approximately about next 20 years.

In case that a more effective protection level of cover-plate is provided (currently planned) then lower rate of corrosion may be considered (e.g. parameter $B = 0.6$ only, see case 1 in Figure 6).

New estimation of the residual working life is shown in Figure 6 when new protective coating of

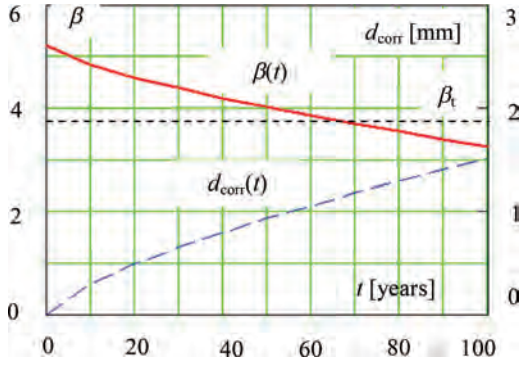


Figure 5. Time-dependent reliability index $\beta(t)$ and one-side corrosion depth $d_{\mathrm{corr}}(t)$.

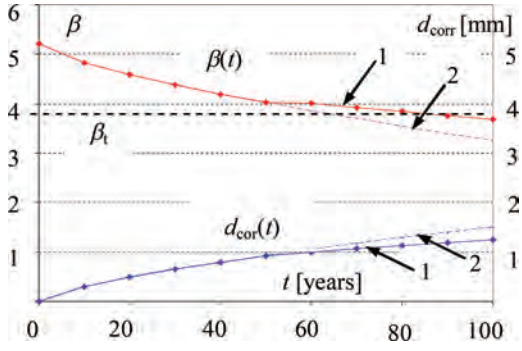


Figure 6. Time-dependent reliability index $\beta(t)$ and one-side corrosion depth $d_{corr}(t)$ for considered increased maintenance level (Case 1), and common maintenance level (Case 2).

steel member is provided leading to extended residual working life from 70 to 90 years.

4 VERIFICATION OF INDUSTRIAL BRIDGE

4.1 *Inspection of the state of bridge*

Existing industrial steel bridge supported by massive concrete columns being used for running of cleaning machine for trash racks in the power station are assessed for possibility of their further usage assuming two alternatives of a new machine (Fig. 7).

Industrial steel bridge is composed of two main structural members—steel longitudinal main rolled beams (I 280) stiffened by cross beams (I 140).

The longitudinal main beam is stiffened in the location of concrete columns by two supporting lateral diagonals (rolled U profiles) built in columns. The beams are supported at their ends by rather not enough suitable U profiles embedded in abutments. Visual inspection of steel structures revealed significant deterioration of load bearing members, mainly in locations of their joints and also in connections with steel bridge deck made from corrugated steel plate (Fig. 8).

Figure 7. Existing industrial bridge with a cleaning machine.

Figure 8. Significant deterioration of bridge structural members.

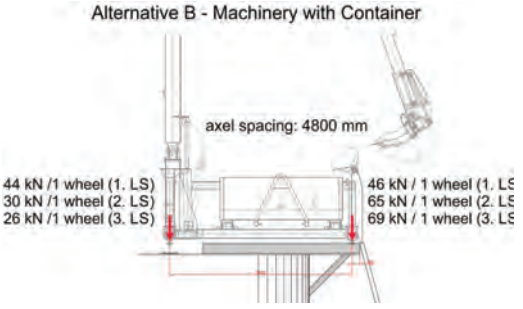


Figure 9. Illustration of new cleaning machinery, alt. B.

Table 5. Load effects and resistances of main structural members of the industrial bridge, in kNm.

Basic member	E	R	R_{red}
Alt. A – main beam	76.6	141.5	99
Alt. B – main beam	73.4	141.5	99
Alt. A – cross beam	14.6	19.8	13.9
Alt. B – cross beam	13.3	19.8	13.9

Reliability of the main structural members of the industrial bridge is verified for two alternatives of the new cleaning machinery for trash racks, see Fig. 9 where the alternative B is illustrated. Firstly, the reliability is verified by partial factor method and commonly applied procedure according to current standards ČSN ISO 13822 (2010), ČSN 73 0038 (2014) and ČSN EN 1990 (2004) (see also 3.1), secondly applying probabilistic methods and the Probabilistic Model Code (PMC 2014) of the Joint Committee on Structural Reliability (JCSS).

Results of reliability verification based on the partial factor method is given in Table 5 considering two alternatives of machinery with different self-weight, total load effects and geometry including supports.

4.2 *Verification by partial factor method*

The reliability of the existing industrial bridge is verified by the partial factor method given in ČSN ISO 13822 (2010) and ČSN 73 0038 (2014).

Self-weight, permanent loads and imposed loads evoked by new machinery are considered. The load bearing capacity of main structural members without any deterioration, and also for estimated actual 30% reduction of resistance is taken into account.

The results indicate that the structures are less loaded when the machinery B is used in comparison to machinery A. For non-deteriorated structure, and also for partly deteriorated structure

Table 6. Reliability of basic bridge steel members.

Basic member	β	$\beta_{red,1}$	$\beta_{red,2}$
Alt. A – main beam	6.5	5.6	4.7
Alt. B – main beam	6.7	5.8	4.9
Alt. A – cross beam	4.7	3.5	2.8
Alt. B – cross beam	5.1	3.9	3.3

(assumed reduction of resistance) the basic reliability condition is fulfilled.

However, low reliability level reveal inclined struts supporting longitudinal beams in regions of concrete columns. Requirement on limiting deflection of steel longitudinal main beams is also not satisfied.

4.3 *Probabilistic structural analyses*

Probabilistic methods are applied for structural verification based on the principles of the Probabilistic Model Code (PMC 2014) and expressions (1,2). The reliability analyses of two basic members, of a longitudinal main beam and cross beam for three cases: a structural member without any deterioration, and also for estimated 20% reduction ($\beta_{red,1}$) and 30% reduction ($\beta_{red,2}$) of resistance is taken into account for two alternative machinery A and B. The results of reliability analysis are given in Table 6.

In case that the main structural members of the industrial bridge are without any deterioration, the resulting reliability indices fulfil the required target reliability index given in EN 1990 (2002).

However, in case that 20%, resp. 30% reduction of resistance is assumed due to the corrosion of steel members, the reliability level of such deteriorated cross beams might be rather low and additional measures (repair, strengthening) should be undertaken.

Moreover, joints of steel members should be checked and repaired if needed.

5 CONCLUDING REMARKS

It is shown that application of partial factor method for the assessment of existing structures might lead in some cases to conservative estimations.

The probabilistic assessment of existing structures make it possible to effectively estimate remaining working life of structures and to plan their maintenance and required economic resources.

The assessment of the quick-closing valves has shown that their remaining life-time is about additional 20 years provided that regular maintenance will be made. Protective layers of steel components should be renovated and regularly inspected. When the reliability index would decrease below the target reliability index 3.7 (estimated to 20 years), a new reliability assessment of cover plates should be made on the basis of updated material characteristic.

Assessment of existing industrial bridge for new technology of cleaning reveals need for regular maintenance. Machinery B appears to be more suitable for cleaning of racks.

Probabilistic methods represent suitable tool for decision about alternative technological procedures and evaluation of actual conditions of structures.

ACKNOWLEDGEMENT

This is the partial outcome of the project TE01020068 Centre of Research and Experimental Development of Reliable Energy Production supported by the Technological Agency of the Czech Republic.

REFERENCES

CSN 75 6303. 2012. Basis of structural design and actions on hydrotechnical structures. ÚNMZ. pp. 45 (in Czech).
CSN EN 1990. 2011. Basis of structural design. ÚNMZ. pp. 95 (implemented in the Czech Republic).
EN 1990. 2002. Basis of structural design. CEN. pp. 87.
EN 1993-1-1. 2005. Eurocode 3: Design of steel structures—Part 1-1: General rules and rules for buildings. CEN. pp. 91.
Holický, M. 2009. Reliability Analysis for Structural Design. Stellenbosch. pp. 199.
Holický, M. & Marková, J. 2007. Probabilistic Design of Structures for Durability. In: ESREL 07. pp. 1757–1762.
Markova, J. & Holicky, M. 2014. Reliability of structures in national codes to Eurocodes, In: Safety and Reliability, Methodology and Applications. Wroclaw, 2015, p. 2207–2212.
fib Bulletin 34. 2005. Model Code for Service Life Design.
ISO 2394. 2015. General principles on reliability for structures.
ISO 13822. 2010. Bases for design of structures—Assessment of existing structures. ISO. pp. 35.
ISO 13823. 2008. General principles on the design of structures for durability. ISO. pp. 39.
Probabilistic Model Code. JCSS. 2014.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Thermal fatigue lifetime prediction of BGA solder joint via a novel fatigue crack propagation model

W. Men, Y. Chen & R. Kang
Reliability and System Engineering School, BeiHang University, Beijing, China

ABSTRACT: Solder joint thermal fatigue is a major mechanism of electronic products failure. Previous studies had proposed several physics-of-failure models for thermal fatigue life prediction based on stress, strain, or energy. However, few models considered fatigue crack propagation. From that point of view, this paper presented a novel fatigue crack propagation model for ball grid array package solder joint thermal fatigue life prediction. Due to the difficulty of crack length measurement, experiment was conducted to explore the relationship of fatigue crack propagation and daisy chains resistances growth, where the daisy chains were created by solder joint network. Furthermore, the finite element simulation method was used to extend this model for solder joints with different dimensions and materials and the final form of the fatigue crack propagation model was obtained.

1 INTRODUCTION

In electronic package structure, solder joints have the functions of electronic connection, mechanical connection between chips and substrate and thermal dissipation tunnel. Reliability of solder joint has a big impact on the function of the whole electronic product. Due to a mismatch of the coefficient of thermal expansion, electronic solder joint experiences cycle shear strain, which ultimately leads to fatigue failure. Predictions of solder joint thermal fatigue life are mainly based on statistical method or Physics-of-Failure (PoF) models. However, the statistical data is often confidential or difficult to obtain for aerospace electronic products. Consequently, the PoF models are widely used for predicting solder joint life in engineering.

The earliest low-cycle fatigue model was Coffin-Manson model, which presented an expression of fatigue lifetime by exploring the relationship of stress cycle damage and plastic deformation (Manson 1965). Smith et al. (1970) improved Coffin-Manson model by taking energy changes into account and raised model accuracy and effectiveness. Engelmaier-Wild model considered the combined influence of tress relaxation, plastic deformation, and creep, and had been widely used in standards IPC-SM-785 and IPC-D-279 (Engelmaier 1983, Socie 1987).

These models were mainly based on stress, strain, or energy. Darveaux model took hysteresis energy theory and fracture mechanics theory into account to predict crack growth rate, thus the failure cycle numbers could be obtained (Darveaux 1997, Darveaux 2002). However, the parameters in the model couldn't be obtained directly, which limited the applicability of this model (Akay et al. 1997, Perkins 2007). New strategies need to be developed. In this study, a fatigue crack propagation model has been developed for Ball Grid Array (BGA) package solder joint life prediction. Experimental and simulation methods were adopted to determine the parameters and to extend it for more general conditions.

The remainder of this paper is organized as follows. Section 2 presents the fatigue crack propagation model considering stress intensity factor. Section 3 discusses the experimental procedures and results of data analysis to form the complete crack propagation model. Section 4 investigates the relationship between solder joint dimension and material with the proposed fatigue crack propagation model via Finite Element Analysis (FEA) simulation. Section 5 provides conclusions as well as directions for future work.

2 FATIGUE CRACK PROPAGATION MODEL

The thermal failure of electronic devices is a kind of fault phenomenon that mainly occurs in the interconnected parts under the action of long-term temperature cycling. Typical failure mechanisms include thermal fatigue of solder joints and thermal fatigue of plated-through holes, which are mainly caused by cyclical switching of circuits and periodic changes of ambient temperature, which eventually lead to thermal fatigue of solder joints.

For the BGA package solder joints, fatigue can be divided into two phases during the failure process:

– During the forming phase of the fatigue source, the initial crack appears at the solder joint surface with the maximum stress;
– Under the shear stress, the initial crack tip experiences repeated plastic deformation and continuous expansion until fatigue fracture occurred.

Previous studies (Forman & Shivakumar 1986, Manson et al. 1966) have shown that there is an exponential integral relationship of crack length difference Δl and stress intensity difference Δk during the crack propagation stage, and the exponent is related to the cycle number to fatigue failure N_f:

$$\Delta l = \int_{l_0}^{l_f} (\Delta k)^{aN_f{}^b} dl \tag{1}$$

where l = crack length; l_0 = initial crack length; l_f = crack length when fatigue failure occurs; k = stress intensity factor; a and b are constants to be defined.

The stress intensity factor k is a parameter that reflects the strength of the stress field at the crack tip of an elastic object, which can be defined as follow:

$$k = \sigma\sqrt{M} \tag{2}$$

where σ = shear stress and M = bending moment of crack cross section. For BGA package solder joints, the tip shear stresses corresponding to different crack lengths are different; however, the bending moment for the same crack location is constant. Consequently, stress intensity difference can be defined as follow:

$$\Delta k = \Delta\sigma\sqrt{M} \tag{3}$$

Through a large number of statistical data fitting, it is found that $\Delta\sigma$, M, and failure cycle number N have a certain relationship:

$$\frac{N_l}{\sqrt{M}} = \Delta\sigma N_f \tag{4}$$

where N_l = cycle number when the crack length reaches l and N_f = cycle number to fatigue failure. Then Equation (1) can be replaced by:

$$\Delta l = \int_{l_0}^{l_f} \left(\frac{N_l}{N_f}\right)^{aN_f{}^b} dl \tag{5}$$

After the integral is simplified:

$$l = l_0 + (l_f - l_0)\left(\frac{N_l}{N_f}\right)^{aN_f{}^b} \tag{6}$$

When the solder joint falls off, the crack length is equal to the diameter of the solder joint, which denotes failure in principle. However, in order to leave a security margin, the solder joint is considered to have failed when the crack length reaches 70% of the diameter of the solder joint based on GB/T6398. Generally, initial crack length $l_0 = 10^{-8} \sim 10^{-10}$ m, which is small compared to the expansion crack length; consequently, we assumed that $l_0 = 0$. Then the general form of the fatigue crack propagation model can be defined as follow:

$$l = 0.7D\left(\frac{N_l}{N_f}\right)^{aN_f{}^b} \tag{7}$$

where D = solder joint diameter. N_l and N_f can be obtained via experiment or simulation; however, constants a and b are unavailable directly due to the difficulty of crack length measurement. Therefore, the following sections will show a new approach to solve this problem and the fatigue crack propagation model of Equation (7) is determined.

3 EXPERIMENT

This section mainly shows the experimental procedures for obtaining N_l and N_f, as well as the new approach to replace crack length measurement for data fitting to calculate coefficients a and b according to a real-life case.

3.1 *Temperature profile*

The temperature profile used in this study is used in an engine control system of a helicopter, which involves parking, taxi, takeoff, ascent, level-flight, descent, and landing phases. During hot day, the ground temperature can reach 55°C, and since the electronic device was installed inside, considering a temperature difference of 15°C, this would be 70°C. During the flight phase, it will drop to 20°C due to a combination of various factors that influence temperature. The resulting life temperature profile is shown in Figure 1, where the off-time is omitted in this case due to its small impact on system life.

The detailed values of the temperature profile are shown in Table 1. The lower dwell was at 20°C for 115 min, while the upper dwell was at 70°C for 30 min. The overall cycle duration was

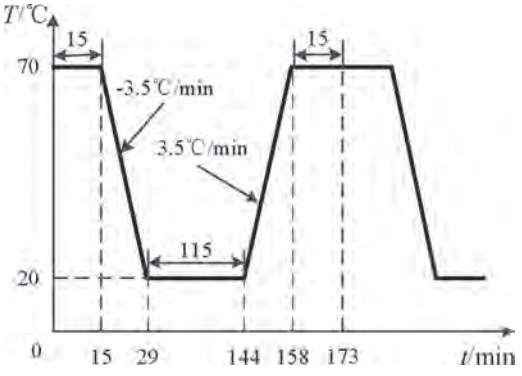


Figure 1. Temperature profile.

Table 1. Detailed values of the temperature profile.

	Temperature (°C)	Duration time (min)
1	70	15
2	Decreasing	14
3	20	115
4	Increasing	14
5	70	15

173 min, and the rate of temperature change was 3.5°C/min.

3.2 *Test specimen*

Due to the difficulty of crack length measurement, experiment was conducted for BGA package solder joints to explore the relationship of fatigue crack propagation and daisy chains resistances growth, where the daisy chains were created by solder joint network.

The first step was to produce test specimens. Chips were surface-mounted on printed circuit boards (PCBs) with 0.5 mm diameter Sn62Pb36 Ag2 solder joints. There were two test boards (A1 and A2) each had six BGA package chips (U1-U6) with an electrical resistance network that could be monitored for failure, which is shown in Figure 2. The solder joint network created three daisy chains named T1, T2, and T3. The resistance of the daisy chain was equal to the sum of the resistances of all solder joints contained within. The solder joints of different daisy chains were under different thermal expansion mismatch conditions. Among them, the daisy chain under the most severe condition would be the first one to fail and its resistances were representations of the solder joints' fatigue crack lengths.

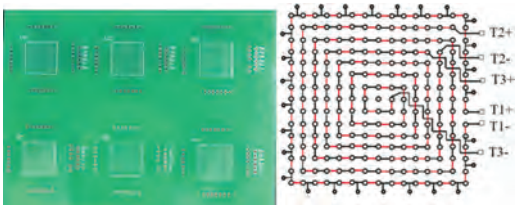


Figure 2. Test specimen and three daisy chains.

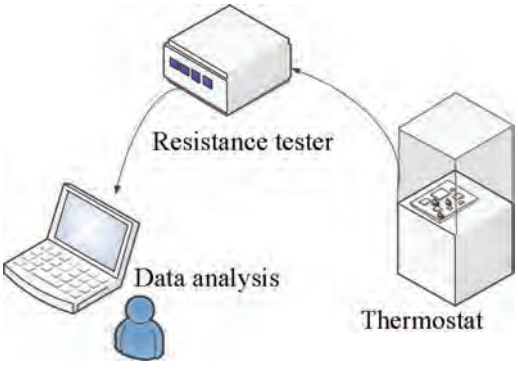


Figure 3. Devices of thermal fatigue experiments.

3.3 *Device setup and parameter conversion*

The temperature cycling profile was achieved with a temperature chamber. And the resistance values of daisy chains were measured with a resistance tester. The experiment and test devices were organized as shown in Figure 3.

Then the thermal fatigue failure cycle numbers N and daisy chains resistances r were obtained. IPC-9701 had shown that there was an exponential relationship of daisy chain resistance r and fatigue crack length l of BGA package solder joints, which is shown as follow after deduction and simplification:

$$r = mi(D - l)^{-n} \tag{8}$$

where i = solder joint number within the daisy chain; D = solder joint diameter; m and n are constants to be defined. Then the relationship of daisy chain resistance r and thermal fatigue failure cycle number N can be derived by combining Equations (7) and (8):

$$r = miD\left[1 - 0.7\left(\frac{N_r}{N_f}\right)^{aN_f{}^b}\right]^{-n} \tag{9}$$

where $N_r = N_l$.

3.4 *Results and discussion*

Twelve resistance values (two test boards with six chips each) of each type of daisy chain had been obtained within this experiment. The results of T1 daisy chains are shown in Figure 4.

Interconnecting failure was defined as a 20% increase in nominal resistance of the daisy chains, which means that at least one of the solder joints had failed ($l_f = 0.7D$). It can be found from Figure 4 that the resistance value of the T1 daisy chain reached 20% increase at about 500 cycles. Consequently, the subsequent data was useless for data fitting due to the overstepping of security margin. Generally, these twelve groups of data had good uniformity. The medium value of these twelve groups of data was used to analyze the fatigue crack propagation model.

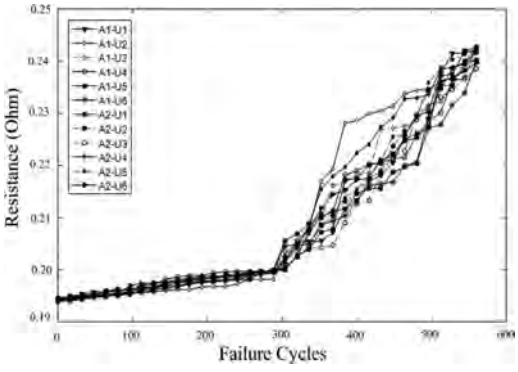


Figure 4. Resistance of the T1 daisy chain in twelve chip samples.

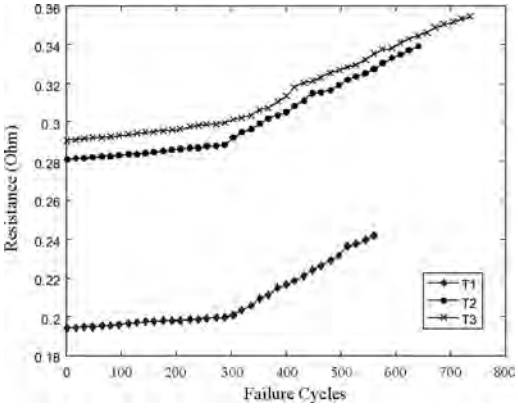


Figure 5. Resistance of daisy chains T1, T2, and T3.

Figure 5 shows the resistances of daisy chains T1, T2, and T3 of the experiment after integration.

The initial resistance values of three daisy chains were diverse due to the difference of solder joint numbers within. Figure 5 shows that the daisy chain T1 was the first one to fail. The reason is that daisy chain T1 was installed on the outermost side of the chip of Figure 2, which has maximal strain under the same thermal stress. As mentioned in section 3.2, daisy chain T1 was under the most severe thermal expansion mismatch conditions, and its resistances represented the thermal fatigue lives of solder joints of the chips. The resistances and failure cycles of daisy chain T1 had been recorded for data fitting purposes.

3.5 *Data fitting*

The experimental data were verified and N and r were picked up, which is shown in Table 2.

Table 2 shows that when resistance r reached a 20% increase for the first time, cycle number N was 508, which represents the fatigue failure cycles N_f in Equation (9). The diameter of solder joints is 0.5 mm and the solder joint number within daisy chain T1 is 56.

Data fitting has been processed with Matlab and a, b, m, and n in Equation (9) are shown in Table 3.

Here, SSE represents the sum of squares due to error, while R-square represents the coefficient of

Table 2. Failure cycles and resistance values of daisy chain T1.

	N (cycles)	r (Ohm)
1	0	0.1942
2	16	0.1946
3	32	0.1948
4	48	0.1950
5	60	0.1952
6	76	0.1955
7	100	0.1963
8	128	0.1969
9	160	0.1976
10	220	0.1985
11	240	0.1990
12	288	0.2000
13	304	0.2028
14	320	0.2031
15	352	0.2098
16	396	0.2197
17	424	0.2211
18	460	0.2298
19	508	0.2394

Table 3. Values of a, b, m, and n and goodness of fitting.

Parameter	Value
a	0.6637
b	0.4053
m	0.0032
n	0.1795
SSE	0.0240
R-square	0.9397
Adjusted R-square	0.9305
RSME	0.0736

determination. Adjusted R-square represents the degree-of-freedom adjusted coefficient of determination, and RMSE represents the root mean squared error.

Then, the relationship of daisy chain resistance r and thermal fatigue failure cycle number N of BGA package Sn62Pb36 Ag2 solder joints with 0.5 mm diameter could be integrated as:

$$r = 0.0032iD\left[1 - 0.7\left(\frac{N_r}{N_f}\right)^{0.6637N_f^{0.4053}}\right]^{-0.1795} \quad (10)$$

where r = resistance of the first failed daisy chain; i = solder joint number within the daisy chain; D = solder joint diameter; N_r = cycle number when the resistance reaches r and N_f = cycle number to fatigue failure. And the fatigue crack propagation model for the certain solder joint can be obtained:

$$l = 0.7D\left(\frac{N_l}{N_f}\right)^{0.6637N_f^{0.4053}} \quad (11)$$

where l = crack length and $N_l = N_r$ = cycle number when the crack length reaches l.

4 FEA SIMULATION

Experimental results were most realistic; however, both time and cost of the experiment were quite significant for aviation products due to the large failure cycle numbers. Therefore, simulation is still the main technology applied for engineering. This section is mainly to show the simulation method for solder joint thermal fatigue, and to explore the relationship between parameters a, b, m, and n with the material and dimension of solder joint; then, the final form of the fatigue crack propagation model could be obtained.

4.1 *Modeling*

A typical BGA package was modeled with the finite element method using ANSYS; the results are shown in Figure 6. The model contained solder joint, chip, substrate, encapsulated layer, and PCB layer. In ANSYS, element type Visco107 was used to describe the solder joint material.

To evaluate the impact of solder joint material and dimension on the crack propagation model, four types of solder joint materials were selected. These were Sn62Pb36 Ag2, Sn96.5 Ag3Cu0.5, Sn60Pb40, and Pb97.5Sn2.5, and the diameters of the solder joint were selected as 0.25 mm, 0.4 mm, 0.5 mm, and 0.6 mm.

4.2 *Simulation results*

The solder joint strain could be calculated via FEA. With the relationship of strain and thermal damage resistance, the daisy chain resistance could be obtained. Figure 7 shows the resistance of T1 obtained via experimental and FEA.

Figure 7 shows that the simulation result was not as smooth as the experimental result; however, the error was controlled within 5%, which was considered to be reasonable within 20%.

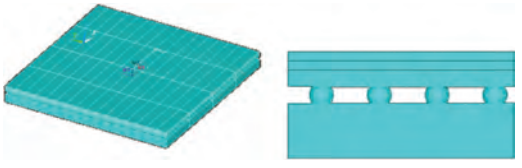


Figure 6. FEA model and solder joint model.

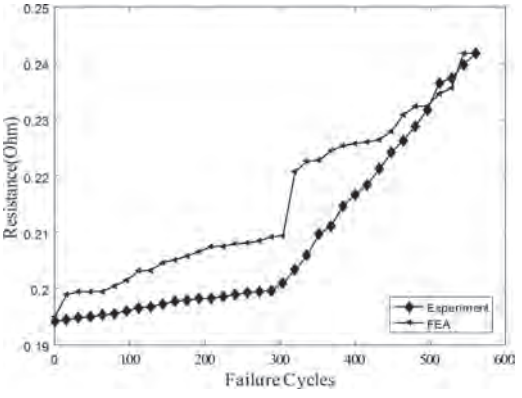


Figure 7. Resistance of the T1 daisy chain via experimental and FEA.

4.3 *Crack propagation model considering material and dimension*

Using these simulation results, the parameter *a*, *b*, *m*, and *n* in Equation (9) for four material types and four diameters of solder joints could be calculated, and the results are shown in Table 4. The material of the solder joint was expressed with Young's Module E and fatigue strength factor σ_f, which are the most critical parameters of the material for fatigue failure. Figure 8 shows the relationship between parameters *a*, *b*, *m*, and *n* depending on diameter and material of the solder joint.

Table 4. Parameter values of different solder joint diameters and materials.

	D (mm)	E (GPa)	σ_f (GPa)	*a*	*b*	*m*	*n*
Sn62	0.25	30	1.97	0.649	0.391	0.0033	0.181
Pb36	0.4	30	1.97	0.652	0.384	0.0031	0.179
Ag2	0.5	30	1.97	0.650	0.404	0.0036	0.180
	0.6	30	1.97	0.648	0.391	0.0032	0.179
Sn96.5	0.25	35	2.25	1.178	0.411	0.0262	0.183
Ag3	0.4	35	2.25	1.175	0.406	0.0262	0.179
Cu0.5	0.5	35	2.25	1.180	0.412	0.0259	0.178
	0.6	35	2.25	1.180	0.401	0.0258	0.182
Sn60	0.25	32	2.01	0.872	0.408	0.0110	0.179
Pb40	0.4	32	2.01	0.875	0.401	0.0114	0.178
	0.5	32	2.01	0.873	0.411	0.0112	0.178
	0.6	32	2.01	0.874	0.407	0.0111	0.182
Pb97.5	0.25	40	2.55	1.567	0.389	0.0448	0.180
Sn2.5	0.4	40	2.55	1.564	0.406	0.0450	0.180
	0.5	40	2.55	1.569	0.394	0.0451	0.182
	0.6	40	2.55	1.564	0.387	0.0450	0.180

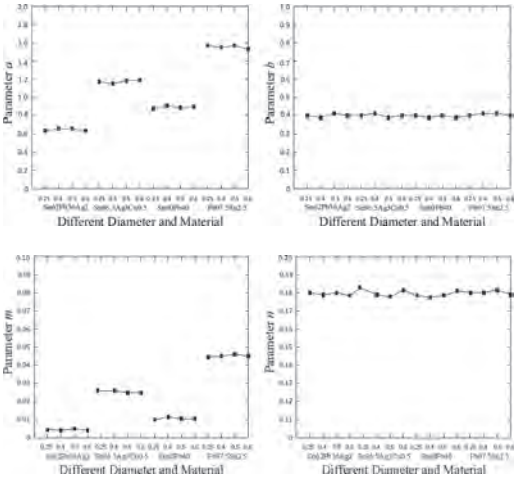


Figure 8. Relationships of parameters with diameter and material of solder joints.

Figure 8 shows that the material and dimension of solder joints have no effect on parameters *b* and *n*. The mean values are 0.400 and 0.180 respectively after testing. Figure 8 furthermore shows that parameter *a* and *m* only changed with changing material. With the simulation results of FEA, the relationships between parameter *a* and *m* with Young's Module E and fatigue strength factor σ_f were fitted via Matlab:

$$a = 0.1109E - 0.3123\sigma_f - 2.0448 \tag{12}$$

$$m = 0.0035E + 0.0118\sigma_f - 0.1246 \tag{13}$$

Consequently, Equation (9) can be expressed as:

$$r = miD\left[1 - 0.7\left(\frac{N_r}{N_f}\right)^{aN_f{}^{0.4}}\right]^{-0.18}$$

$$\begin{aligned} m &= 0.0035E + 0.0118\sigma_f - 0.1246 \\ a &= 0.1109E - 0.3123\sigma_f - 2.0448 \end{aligned} \tag{14}$$

And the final form of the fatigue crack propagation model can be obtained:

$$l = 0.7D\left(\frac{N_l}{N_f}\right)^{aN_f{}^{0.4}}$$

$$a = 0.1109E - 0.3123\sigma_f - 2.0448 \tag{15}$$

where Young's Module E and fatigue strength factor σ_f can be directly found within a material handbook.

5 CONCLUSION AND FUTURE WORK

This paper presented a fatigue crack propagation model for solder joint thermal fatigue life prediction from the perspective of stress intensity factor. To determine the constants within the model, experiment was conducted and a new approach with daisy chain was used to solve the problem of crack length measuring difficulty. The model for a certain solder joint was obtained. To study the relationship of the coefficients of the fatigue crack propagation model with the dimension and material of the solder joint, the Finite Element method was used to simulate more cases. And a more general form of the model was achieved.

A further study may focus on additional FEA simulations and study other factors that may affect fatigue crack propagation models, such as different

mounted styles of the chip and different conductivities of the solder joint.

ACKNOWLEDGMENTS

This work was funded by the National Natural Science Foundation of China under contract number 61503014 and 61573043.

REFERENCES

Akay, H.U., Zhang, H., & Paydar, N.H. 1997. Experimental correlations of an energy-based fatigue life prediction method for solder joints. *Advances in Electronic Packaging* 19(2): 1567–1574.

Darveaux, R. 1997. Solder joint fatigue life model. *Proc of the TMS Annual Meeting, Orlando*: 213–218.

Darveaux, R. 2002. Effect of simulation methodology on solder joint crack growth correlation and fatigue life prediction. *Journal of Electronic Packaging* 124(3): 147–154.

Engelmaier, W. 1983. Fatigue life of leadless chip carrier solder joints during power cycling. *IEEE Transactions on Components Hybrids & Manufacturing Technology* 6(3): 232–237.

Forman, R.G. & Shivakumar, V. 1986. Growth behavior of surface cracks in the circumferential plane of solid and hollow cylinders. *Fracture Mechanics: Seventeenth Volume. ASTM International*: 345–349.

Manson, S.S. 1965. Fatigue: a complex subject—some simple approximations. *Experimental Mechanics* 5(4): 193–226.

Manson, S.S., Freche, J.C., & Ensign, C.R. 1966. Application of a double linear damage rule to cumulative fatigue. *Fatigue Crack Propagation. ASTM International*: 384–412.

Perkins, A.E. 2007. Investigation and prediction of solder joint reliability for ceramic area array packages under thermal cycling, power cycling, and vibration environments. *Dissertation Abstracts International* 68(5): 3354–3358.

Smith, K.N., Watson, P., & Topper, T.H. 1970. A stress-strain function for the fatigue of metals. *Journal of Materials* 5(4): 767–778.

Socie, D. 1987. Multiaxial fatigue damage models. *Key Engineering Materials* 324–325(4): 747–750.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Reliability of the aircraft in the Polish operational aviation

M. Zieja & M. Woch
Division for IT Support of Logistics, Air Force Institute of Technology, Warsaw, Poland

J. Tomaszewska
Faculty of Aviation, Polish Air Force Academy, Dęblin, Poland

ABSTRACT: This paper is an outgrowth of an investigation to determine what is the main source of failure in the Polish operational aviation. The proper understanding of failure data is valuable in prediction of the future needs in a specified planning horizon or for specified operational hours. The major effort of the study was the reliability analysis for two types of aircraft, which are used in the Polish operational aviation. As the main results the value of the number of hours between failures is presented and this distribution is tested with some mathematical models. The description of the data by Weibull distribution, which is one of the most frequently used functions in reliability analysis, has been tested. Then, the same test has been performed for another mathematical distributions. It has been observed that for an aircraft of first type the best description can be given by different distribution as for the second type and that the standard Weibull distribution is not always the best reliability model. The potentially causes of this behavior are discussed in the article. Additionally, the change of the number of the failures as a function of time is presented. It is observed that this behavior varied for the different types of aircraft.

1 INTRODUCTION

1.1 *Formulation of the problem*

Aircrafts system may be designed with redundancy of components in order to improve their overall reliability. Such approach can increase the reliability of the whole system, however, the improved reliability is not the only factor that contributes to the effectiveness of a task performance. If the system fails and has to be repaired, then the time to repair is also an important factor in effectiveness. One of the most important factors of the high efficiency of the system usage is the appropriate estimation of the expected time to failure, proper planning and efficient performing of inspections and any possible repairs (Tloczynski 2017b). The time to failure can be estimated by knowledge of the behavior of a data from the exploitation process. Data can be described by the different statistical models. Proper selection of the model is externally important which will be presented in this contribution by comparison of data with the selected models. The advantage of this method over the calculation of mean and standard division is the reduction of reparation costs. Failure rate function shows exactly when early failures occurs, constant failure rate may describes random failures, whereas increasing failure rate tells about wear-out failures (Babiarz 2016).

2 FORMULATION OF THE PROBLEM

2.1 *Basics definitions*

A failure rate function is defined as the limit, if it exists, of the ratio of the conditional probability that the instant of time, T, of a failure of an item falls within a given time interval $t + \Delta t$ and the length of this interval, Δt, when Δt leads to zero, given that the item is in an up state at the beginning of the time interval, which can be described as (Koucky and Valis 2007, Valis et al. 2014):

$$\lambda(t) = \lim_{\Delta t \to 0^+} \frac{P\{t < T \le t + \Delta t \mid T > t\}}{\Delta t} \tag{1}$$

where T is a continuous positive random variable of device operation time.

If T has a density $f(t)$ and the distribution $F(t)$ equation (1) will take the form:

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \tag{2}$$

where $F(t) = \int_0^t f(u)\,du = P\{T \le t\} = 1 - P\{T > t\}$.

2.2 *The Weibull distribution*

The Weibull distribution is one of the most widely used lifetime distributions in reliability engineering

(Tloczynski 2017a). It is a versatile distribution, that can take on the characteristics of other types of distributions, based on the value of the shape parameter, β.

Weibull Probability Density Function can be described by the following formula (Hinz, Hienzsch, & Bracke 2017a):

$$f(t)=\frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1} e^{-\left(\frac{t-\gamma}{\eta}\right)^{\beta}} \tag{3}$$

where:

η - scale parameter, or characteristic life,

β - shape parameter (or slope),

γ - location parameter (or failure free life).

The Weibull Failure Rate Function is given by:

$$\lambda(t)=\frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1} \tag{4}$$

2.3 *The Burr distribution*

In probability theory, statistics and econometrics, the Burr Type XII distribution or simply the Burr distribution[1] is a continuous probability distribution for a non-negative random variable. It is also known as the SinghMaddala distribution[2] and is one of a number of different distributions sometimes called the generalized log-logistic distribution.

Burr Probability Density Function can be described by the following formula (Mueller, Hinz, & Bracke 2017):

$$f(t)=\frac{\alpha k\left(t-\gamma/\beta\right)^{\alpha-1}}{\beta\left[1+\left(t-\gamma/\beta\right)^{\alpha}\right]^{k+1}} \tag{5}$$

where:

k - continuous shape parameter ($k>0$),

α - continuous shape parameter ($\alpha>0$),

β - continuous scale parameter ($\beta>0$),

γ - continuous location parameter ($\gamma\equiv 0$ yields the three-parameter Dagum distribution).

The Burr Failure Rate Function is given by:

$$\lambda(t)=\frac{f(t)}{1-F(t)}=\frac{\dfrac{\alpha k\left(t-\gamma/\beta\right)^{\alpha-1}}{\beta\left[1+\left(t-\gamma/\beta\right)^{\alpha}\right]^{k+1}}}{1-\left[1+\left(t-\gamma/\beta\right)^{\alpha}\right]^{-k}} \tag{6}$$

2.4 *The Dagum distribution*

The Dagum distribution is a continuous probability distribution defined over positive real numbers. It is named after Camilo Dagum, who proposed it in a series of papers in the 1970s. The Dagum distribution arose from several variants of a new model on the size distribution of personal income and is mostly associated with the study of income distribution.

Dagum Probability Density Function can be described by the following formula (Vintr & Valis 2011):

$$f(t)=\frac{\alpha k\left(t-\gamma/\beta\right)^{\alpha k-1}}{\beta\left[1+\left(t-\gamma/\beta\right)^{\alpha}\right]^{k+1}} \tag{7}$$

where:

k - continuous shape parameter ($k>0$),

α - continuous shape parameter ($\alpha>0$),

β - continuous scale parameter ($\beta>0$),

γ - continuous location parameter ($\gamma\equiv 0$ yields the three-parameter Dagum distribution).

The Dagum Failure Rate Function is given by:

$$\lambda(t)=\frac{f(t)}{1-F(t)}=\frac{\dfrac{\alpha k\left(t-\gamma/\beta\right)^{\alpha k-1}}{\beta\left[1+\left(t-\gamma/\beta\right)^{\alpha}\right]^{k+1}}}{1-\left[1+\left(t-\gamma/\beta\right)^{-\alpha}\right]^{-k}} \tag{8}$$

2.5 *The log-logistic distribution*

In probability and statistics, the log-logistic distribution (known as the Fisk distribution in economics) is a continuous probability distribution for a non-negative random variable. It is used in survival analysis as a parametric model for events whose rate increases initially and decreases later, for example mortality rate from cancer following diagnosis or treatment. It has also been used in hydrology to model stream flow and precipitation, in economics as a simple model of the distribution of wealth or income, and in networking to model the transmission times of data considering both the network and the software.

The log-logistic distribution is the probability distribution of a random variable whose logarithm has a logistic distribution. It is similar in shape to the log-normal distribution but has heavier tails. Unlike the log-normal, its cumulative distribution-function can be written in closed form.

Log-logistic Probability Density Function can be described by the following formula (Hinz, Hienzsch, & Bracke 2017b):

$$f(t)=\frac{\alpha}{\beta}\left(\frac{t-\gamma}{\beta}\right)^{\alpha-1}\left[1+\left(\frac{t-\gamma}{\beta}\right)^{\alpha}\right]^{-2} \tag{9}$$

where:

α - continuous shape parameter ($\alpha > 0$),

β - continuous scale parameter ($\beta > 0$),

γ - continuous location parameter ($\gamma \equiv 0$ yields the two-parameter Log-logistic distribution).

The Log-logistic Failure Rate Function is given by (Valis, Zak, Pokora, & Lansky 2016):

$$\lambda(t) = \frac{f(t)}{1-F(t)} = \frac{\frac{\alpha}{\beta}\left(\frac{t-\gamma}{\beta}\right)^{\alpha-1}\left[1+\left(\frac{t-\gamma}{\beta}\right)^{\alpha}\right]^{-2}}{1-\left[1+\left(\frac{\beta}{t-\gamma}\right)^{\alpha}\right]^{-1}} \quad (10)$$

3 EXPERIMENTAL STUDY OF STATISTICAL METHODS

3.1 *Input data*

This analysis uses data on one type of aircraft used in Polish Air Force for a selected group of pilots. Data were obtained from the operation and maintenance. The analysed data are presented in the form of graph of average flight hours between failure as a function of years in Figures 1 and 2. These figures show the data for the years 2011 and 2017. It was noted that data for 2013 are concentrated at lower values, while for 2017 at higher. Such behavior may indicate that at some point there was an improvement of procedures to control the aircraft, which resulted in fewer aircraft failures. It may also be the result increasing the maintained.

Based on Figures 1 and 2, the division into two groups of data can be observed: the first one covers the years 2011–2015, for which the average time between failures is about 15 hours, second contains years 2016–2017 for which the average time between failures increases twice.

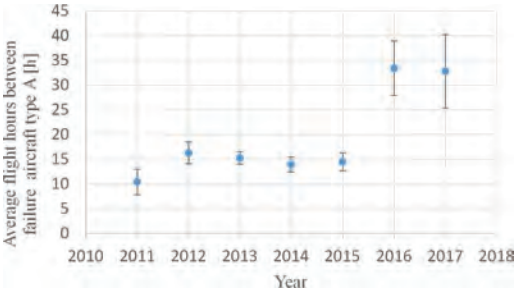


Figure 1. Average flight hours between failure for 4th generation fighter aircraft type A as a function of years.

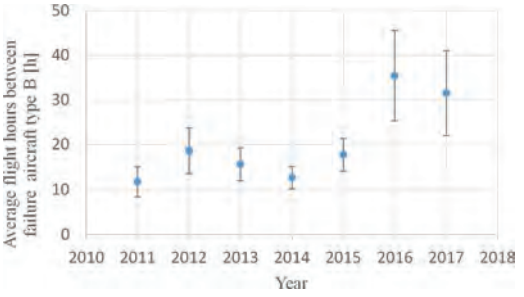


Figure 2. Average flight hours between failure for 4th generation fighter aircraft type A as a function of years.

Comparison of distributions of probability density function with data obtained during the operation process is shown in Figure 7. In addition, Figure 8 summarize the fitting parameters for failure rate function distributions. It can be observed that the for aircraft type A the shape is more narrow. The Weibull's distributions for both type of aircraft have similar shapes, but for type B the most expected value is shifted to lower value. Because no significant differences between the failure time for two types of aircraft have been observed, for better precision of the fit, has been done for all collected data together.

3.2 *Quality of the fit*

To determine the quality of the fit and to determine whether the analysis distribution describes the data, the following tests were used:

- Kolmogorov-Smirnov,
- Anderson-Darling,
- Chi-kwadrat,

which are based on the empirical distribution function (EDF).

3.2.1 *The Kolmogorov-Smirnov statistic*

The Kolmogorov-Smirnov statistic is defined as:

$$D = \sup_x \left| F_n(x) - F(x) \right| \quad (11)$$

The Kolmogorov-Smirnov statistic belongs to the supremum class of EDF statistics (Chakravarti, Laha, & Roy 1967). This class of statistics is based on the largest vertical difference between $F(x)$ and $F_n(x)$. The Kolmogorov-Smirnov statistic is computed as the maximum of D^+ and D^-, where D^+ is the largest vertical distance between the EDF and the distribution function when the EDF is greater than the distribution function, and D^- is the largest vertical distance when the EDF is less than the distribution function.

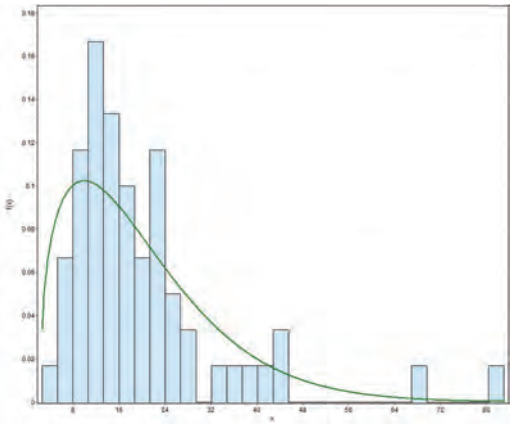


Figure 3. Weibull probability density function fit for 4th generation fighter aircraft type A.

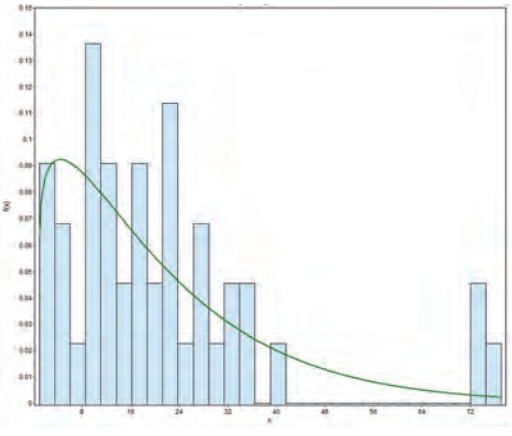


Figure 4. Weibull probability density function fit for 4th generation fighter aircraft type B.

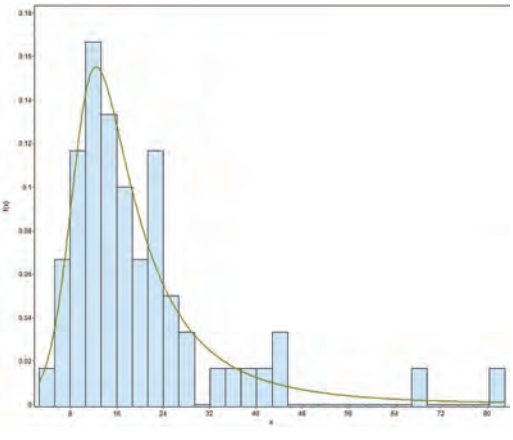


Figure 5. Burr probability density function fit for 4th generation fighter aircraft type A.

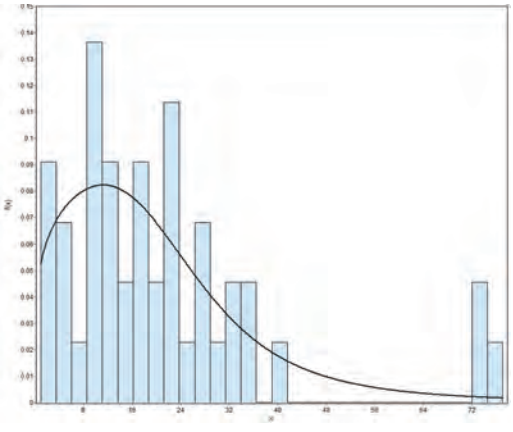


Figure 6. Dagum probability density function fit for 4th generation fighter aircraft type B.

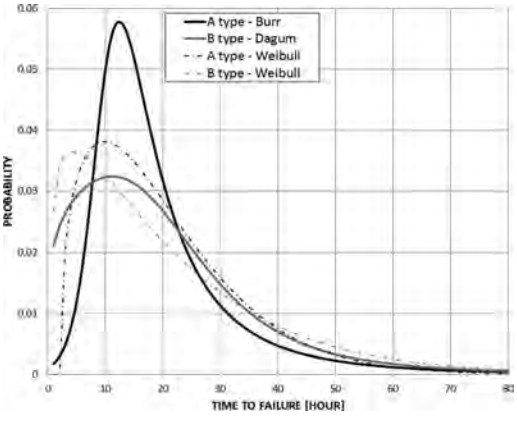


Figure 7. Probability density function comparison.

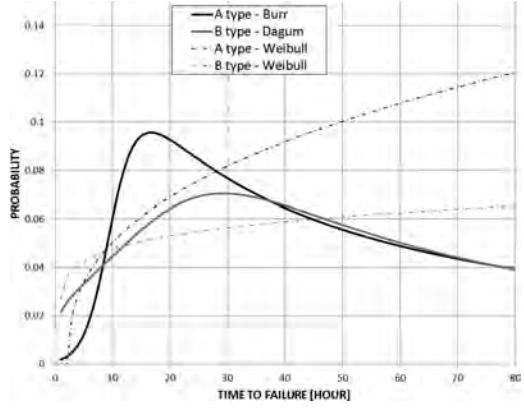


Figure 8. Failure rate comparison.

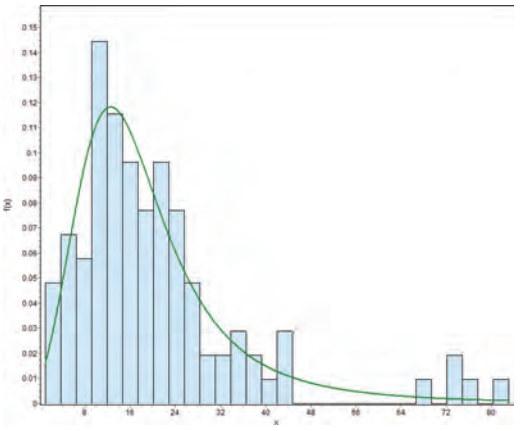


Figure 9. Log-logistic probability density function fit for 4th generation fighter aircraft.

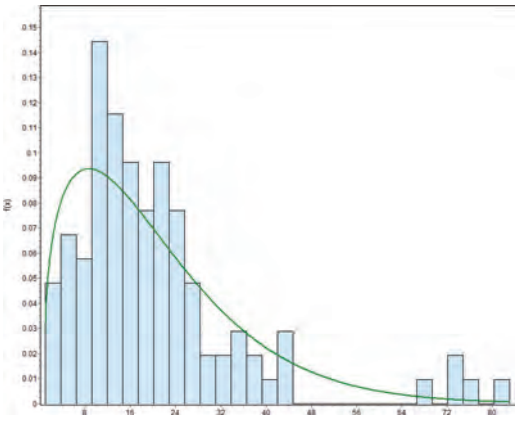


Figure 10. Weibull probability density function fit for 4th generation fighter aircraft.

Table 1. Goodness of fit—summary for 4th generation fighter aircraft type A.

		Test					
		Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
#	Distribution	*p-value*	Rank	*p-value*	Rank	*p-value*	Rank
1	Dagum	0.053	1	0.207	2	1.518	8
2	Gen.Logistic	0.062	2	0.212	3	2.240	16
3	Wakeby	0.062	3	0.187	1	3.230	28
4	Gen.Extreme Value	0.062	4	0.231	4	1.843	13
5	Burr	0.064	5	0.318	10	1.611	9
6	Log-Logistic(3P)	0.067	6	0.263	7	1.110	2
7	Frechet(3P)	0.068	7	0.250	5	1.645	10
8	Pearson5(3P)	0.068	8	0.259	6	2.648	22
19	Weibull(3P)	0.090	19	0.511	18	1.455	4
22	Weibull	0.094	22	0.481	16	2.812	25

$$\begin{aligned} D^{+} &= \max_i\left(\frac{i}{n}-U_{(i)}\right), \\ D^{-} &= \max_i\left(U_{(i)}-\frac{i-1}{n}\right), \\ D &= \max\left(D^{+},D^{-}\right). \end{aligned} \tag{12}$$

3.2.2 *The Anderson-Darling statistic*

The Anderson-Darling Statistic belongs to the quadratic class of EDF statistics (Stephens 1974). The Anderson-Darling statistic is defined as:

$$A^2 = n\int_{-\infty}^{+\infty}\left(F_n(x)-F(x)\right)^2 \left[F(x)\left(1-F(x)\right)\right]^{-1} dF(x) \tag{13}$$

Here the weight function is $\psi(x) = [F(x)(1-F(x)]$. The Anderson-Darling statistic is computed as

$$A^2 = -n-\frac{1}{n}\sum_{i=1}^{n}\left[\left(2i-1\right)\log U_{(i)} + \left(2n+1-2i\right)\log\left(1-U_{(i)}\right)\right] \tag{14}$$

3.2.3 *The Chi-Squared statistic*

The Chi-Squared statistics belongs to the quadratic class of EDF statistics (Snedecor & Cochran 1989). This class of statistics is based on the squared difference $(F)_n(x)-F(x))^2$. Quadratic statistics have the following general form:

Table 2. Goodness of fit—summary for 4th generation fighter aircraft type B.

		Test					
		Kolmogorov Smirnov		Anderson Darling		Chi-kwadrat	
#	Distribution	*p-value*	Rank	*p-value*	Rank	*p-value*	Rank
1	Burr(4P)	0.044	1	0.125	1	0.940	1
2	Burr	0.046	2	0.155	2	1.711	4
3	Dagum(4P)	0.047	3	0.163	4	1.576	3
4	Dagum	0.049	4	0.171	5	2.406	10
5	Wakeby	0.049	5	0.163	3	1.126	2
6	Log-Logistic(3P)	0.051	6	0.183	7	2.402	9
7	Gen.Logistic	0.052	7	0.181	6	2.093	5
8	Gen.ExtremeValue	0.056	8	0.219	8	2.155	6
26	Weibull	0.096	26	1.440	28	7.397	28
30	Weibull(3P)	0.110	30	1.174	26	7.517	30

Table 3. Goodness of fit—summary for 4th generation fighter aircraft.

		Test					
		Kolmogorov Smirnov		Anderson Darling		Chi-kwadrat	
#	Distribution	*p-value*	Rank	*p-value*	Rank	*p-value*	Rank
1	Log-Logistic(3P)	0.040	1	0.207	3	0.838	2
2	Wakeby	0.042	2	0.193	1	2.000	9
3	Gen.Logistic	0.042	3	0.203	2	0.913	3
4	Burr(4P)	0.043	4	0.224	4	0.680	1
5	Gen.ExtremeValue	0.048	5	0.309	7	1.097	4
6	Dagum(4P)	0.052	6	0.269	5	1.205	6
7	Pearson5(3P)	0.053	7	0.346	8	2.609	11
8	Pearson6(4P)	0.054	8	0.364	9	2.590	10
19	Weibull	0.078	19	1.180	20	5.379	17
26	Weibull(3P)	0.097	26	1.371	23	9.820	25

$$Q = n\int_{-\infty}^{+\infty} \left(F_n(x) - F(x) \right)^2 \Psi(x)\, dF \tag{15}$$

The function weights $\Psi(x)$ the squared difference $(F)_n(x) - F(x))^2$.

The Tables 1–3 presented that the goodness-of-fit tests reject the null hypothesis that the number of hours between failures can be approximated by any of the distribution. It is merely because the *p-value* is greater than 0.01.

4 CONCLUSION

This contribution stands at the beginning of research program projecting risk at the Air Force Institute of Technology for 4th generation fighter aircraft and subsequent development will be the subject of future papers. The risk profile of an aircraft is changing and, as such as, it is essential that we will continue to define and monitor desired performance outcomes in line with the risk profile. In order to be able to create such profile, it is necessary to select a probability distribution that will facilitate the construction of reasonably precise probability statements of the type that one wishes to make. In this analysis, the number of hours between failures has been used as the main variable for taking the conclusion about the dependability of the aircraft. It was concluded, from the quality of fitting, that the distribution of Dumm and Darum are the best descriptions of the gathered data. Commonly used in reliability the Weibull distribution, in this particular scenario, did not meet the expected requirements. This can be the result of the relatively small statistic of the incidents of the 4th generation fighter aircraft.

REFERENCES

Babiarz, B. (2016, Jul). Reliability analysis in subsystem of heat supply. In IEEE (Ed.), *2016 International Conference on Information and Digital Technologies (IDT)*, Rzeszow, Poland, pp. 11–16.

Chakravarti, I.M., R.G. Laha, & J. Roy (1967). *Handbook of methods of Applied Statistics*. John Wiley & Sons.

Hinz, M., F. Hienzsch, & S. Bracke (2017a). Detection of distinctions in car fleets based on measured and simulated data. In *RAMS 2017 63rd Annual Reliability and Maintainability Symposium*, Orlando, Florida, U.S.A.

Hinz, M., F. Hienzsch, & S. Bracke (2017b, Sep). Development of two methods for the characterisation of an automotive fleet behaviour based on the simulation of single car rides. In *RISK, RELIABILITY AND SAFETY: INNOVATING THEORY AND PRACTICE*, Glasgow, SCOTLAND, pp. 1593–1598.

Koucky, M. & D. Valis (2007, June). Reliability of sequential systems with a restricted number of renewals. In T. Aven and J.E. Vinnem (Eds.), *Proceedings and Monographs in Engineering, Water and Earth Sciences 2007*, Stavanger, Norway, pp. 1845–1849.

Mueller, A., M. Hinz, & S. Bracke (2017, Sep). Optimization of the dental implant testing based on fem simulation of fatigue and accelerated life. In *RISK, RELIABILITY AND SAFETY: INNOVATING THEORY AND PRACTICE*, Glasgow, SCOTLAND, pp. 16–22.

Snedecor, G.W. & W.G. Cochran (1989). *Statistical Methods*.

Stephens, M.A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of American Statistical Association 69*(347), 730–737.

Tloczynski, D. (2017a). Air transport service in academic research at polish airports. In G. Sierpinski (Ed.), *Advances in Intelligent Systems and Computing*, Volume 505, Katowice, Poland, pp. 23–32. Conference: 13th Scientific and Technical Conference on Transport Systems. Theory and Practice.

Tloczynski, D. (2017b). Security as a determinant of choice of air transport service and air carrier on the basis of research. *Scientific Journal of Silesian University of Technology-Series Transport 95*, 213–222.

Valis, D., L. Zak, & O. Pokora (2014). Engine residual technical life estimation based on tribo data. *Eksploatacja i Niezawodnosc Maintenance and Reliability 16*(2), 203210.

Valis, D., L. Zak, O. Pokora, & P. Lansky (2016). Perspective analysis outcomes of selected tribodiagnostic data used as input for condition based maintenance. *Reliability Engineering & System Safety 145*, 231–242.

Vintr, Z. & D. Valis (2011). A tool for decision making in kout-of-n system maintenance. *Applied Mechanics and Materials 110–116*, 5257–5264.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Buffered environmental contours

K.R. Dahl & A.B. Huseby
Department of Mathematics, University of Oslo, Norway

ABSTRACT: The main idea of this paper is to use the notion of buffered failure probability from probabilistic structural design, first introduced by Rockafellar & Royset (2010), to introduce buffered environmental contours. Classical environmental contours are used in structural design in order to obtain upper bounds on the failure probabilities of a large class of designs. The purpose of buffered failure probabilities is the same. However, in constrast to classical environmental contours, this new concept does not just take into account failure vs. functioning, but also to which extent the system is failing. For example, this is relevant when considering the risk of flooding: We are not just interested in knowing whether a river has flooded. The damages caused by the flooding greatly depends on how much the water has risen above the standard level.

1 INTRODUCTION

Environmental contours are widely used as a basis for e.g., ship design. Such contours allow the designer to verify that a given mechanical structure is safe, i.e, that the failure probability is below a certain value. A realistic model of the environmental loads and the resulting response is crucial for structural reliability analysis of mechanical constructions exposed to environmental forces. See Winterstein et al. (1993) and Haver & Winterstein (2009). For applications of environmental contours in marine structural design, see e.g., Baarholm et al. (2010), Fontaine et al. (2013), Jonathan et al. (2011), Moan (2009) and Ditlevsen (2002).

The traditional approach to environmental contours is based on the well-known *Rosenblatt transformation* introduced in Rosenblatt (1952). This transformation maps the the environmental variables into independent standard normal variables. Using the transformed environmental variables a contour with the desired properties can easily be constructed by identifying a sphere centered in the origin and with a suitable radius. More specifically, the sphere can be chosen so that any non-overlapping convex failure region has a probability less than or equal to a desired exceedence probability. The corresponding environmental contour in the original space can then be found by transforming the sphere back into the original space.

Alternatively, an environmental contour can be constructed directly in the original space using Monte Carlo simulation. See Huseby et al. (2013), Huseby et al. (2015a) and Huseby et al. (2015b). Contours constructed using this approach will always be convex sets. This yields a more straightforward interpretation of the contours. Another advantage of this approach is a more flexible framework for establishing environmental contours, which for example simplifies the inclusion of effects such as future projections of the wave climate related to climatic change. See Vanem & Bitner-Gregersen (2012).

In the present paper we introduce a new concept called buffered environmental contours. This concept is based on the notion of buffered failure probability from probabilistic structural design, first introduced by Rockafellar & Royset (2010). Contrary to classical environmental contours, this new concept does not just take into account failure vs. functioning, but also to which extent the system is failing. For example, this is relevant when considering the risk of flooding: We are not just interested in knowing whether a river has flooded. The damages caused by the flooding greatly depends on how much the water has risen above the standard level.

The structure of this paper is as follows: In Section 2, we recall the classic definition of failure probability in probabilistic structural design and compare this to the concept of buffered failure probability, as defined in Rockafellar & Royset (2010). Furthermore, we recall some of the arguments favoring the buffered failure probability over the regular failure probability. Then, in Section 3, we recall the concept of environmental contours and how such contours are used in structural design in order to find upper bounds on the failure probabilities of a large class of designs. In Section 4, we introduce the new concept of buffered environmental contours, and argue that these contours are better suited than the classical ones in

cases where the level of malfunctioning is important. Finally, in Section 5, we apply the proposed contours to a real life example, and compare the contours to the classical environmental contours.

2 STRUCTURAL DESIGN AND THE BUFFERED FAILURE PROBABILITY

In probabilistic structural design, it is common to define a *performance function*[1] $g(x,V)$ depending on some design variables $x=(x_1,x_2,\ldots,x_m)'$ and some environmental quantities[2] $V=(V_1,V_2,\ldots,V_n)'\in\mathcal{V}$, where $\mathcal{V}\subseteq\mathbb{R}^n$. The design variables can be influenced by the designer of the structure, and may respresent material type or layout. The quantities are usually random, and cannot be directly impacted by the designer. Hence, they may describe environmental conditions, material quality or loads. To emphasize the randomness of the quantities, we denote them by captial letters. In contrast, the design variables are controlled by the designer and hence denoted by small letters.

For a given design x s, $g(\boldsymbol{x},\boldsymbol{V})$ represents the performance of the structure, and is called the *state of the structure*. A given mechanical structure can withstand environmental stress up to a certain level. The *failure region* of the structure is the set of states of the environmental variables that imply that the structure fails. The performance function is defined such that if $g(\boldsymbol{x},\boldsymbol{V})>0$, the structure is *failed*, while if $g(\boldsymbol{x},\boldsymbol{V})\le 0$, the structure is *functioning*. Moreover, for a given $\boldsymbol{x}$ the set $\mathcal{F}(\boldsymbol{x})=\{\boldsymbol{v}\in\mathcal{V}: g(\boldsymbol{x},\boldsymbol{v})>0\}$ is called the *failure region* of the structure[3].

2.1 *The failure probability, reliability and approximation methods*

The failure probability, denoted by $p_f(\boldsymbol{x})$, of the structure is the probability that the structure is failed. That is, $p_f(x)=P(g(\boldsymbol{x},\boldsymbol{V})>0)$. If $f_{\boldsymbol{V}}(v)$ is the joint probability density function for the random vector $\boldsymbol{V}$, the failure probability is given by:

$$p_f(\boldsymbol{x})=\int_{\mathcal{F}(x)} f_{\boldsymbol{V}}(v)\,d\boldsymbol{v}. \tag{1}$$

For a given $\boldsymbol{x}$ the *reliability*, $R(\boldsymbol{x})$, of the system is defined as the probability that the system is functioning, i.e.:

$$R(\boldsymbol{x})=1-p_f(\boldsymbol{x}) \tag{2}$$

A classic problem is to compute the reliability of the system. In order to do so, we need to compute the integral (1). In many cases it is difficult to obtain and analytical solution to this. To overcome this issue various approximation methods have been proposed. Two traditional methods for doing this are the *first-order reliability method* (FORM) and the *second-order reliability method* (SORM). The basic idea of the first-order reliability method is to approximate the failure boundary at a spesific point by a first order Taylor expansion. The idea behind SORM is similar, but using a second order Taylor expansion instead. In both cases, the approximated failure probability can be used to optimize the structural design, i.e. determine a feasible design which has an acceptable failure probability.

2.2 *Return periods*

As is common in structural design models, we view $\boldsymbol{V}$ as representing the average value of the relevant environmental variables in a suitable time interval of length L. Based on this and knowledge of the performance function g it is possible to compute the so-called *return period*. This is done as follows:

We consider the environmental exposure of the given design from time $t\ge 0$. The time axis is divided into intervals of some specified length L, and we let $\boldsymbol{V}_i$ denote the average environmental quantity in the ith period, $i=1,2,\ldots$. It is common to assume that $\boldsymbol{V}_1,\boldsymbol{V}_2,\ldots$ are independent and identically distributed. This is a fairly strict assumption, but as it is so frequently used in structural design, we assume this as well. We then let $T:=\min\{i: g(\boldsymbol{x},\boldsymbol{V}_i)>0\}$. By the assumptions it follows that T is geometrically distributed with probability $p_f=P(g(\boldsymbol{x},\boldsymbol{V})>0)$. The *return period* is defined as $E[T]=1/p_f$. Thus, the return period can be interpreted as a property of the distribution of g($\boldsymbol{x}$, $\boldsymbol{V}$). Hence, it suffices to analyze this distribution, which is what we will focus on in this paper.

2.3 *The buffered failure probability*

The approximations made by FORM and SORM can sometimes be too crude and ignore serious risks. Therefore, we will consider the buffered failure probability, introduced by Rockafellar & Royset (2010) as an alternative to the failure probability. This concept relates closely to the conditional value-at-risk (also called expected shortfall, average

1. 1The performance function is sometimes called the limit-state function.
2. Environmental quantities should here be understood in a broad sense. E.g., for marine structures such quantities typically includes wave height and period. For other types of structures, one may consider e.g., material quality, effects of erosion or corrosion as environmental quantities.
3. In some papers, such as Huseby et al. (2013), the failed states are defined as the states such that $g(\boldsymbol{x},\boldsymbol{V})<0$. This is just a matter of choice of notation.

value-at-risk or expected tail loss), which is a notion frequently used in mathematical finance and financial engineering, see Pflug (2000), Rockafellar (2007) as well as Rockafellar & Uryasev (2000).

Recall that for any level of probability α, the α-quantile of the distribution of a random variable is the value of the inverse of its cumulative distribution function at α. For the random variable $g(\boldsymbol{x},\boldsymbol{V})$, we let $q_\alpha(\boldsymbol{x})$ denote its α-quantile. Similarly, for any probability level α, the α-superquantile of $g(\boldsymbol{x},\boldsymbol{V})$, $\bar{q}_\alpha(\boldsymbol{x})$, is defined as:

$$\bar{q}_\alpha(\boldsymbol{x}) = E\left[g(\boldsymbol{x},\boldsymbol{V}) \middle| g(\boldsymbol{x},\boldsymbol{V}) \rangle q_\alpha(\boldsymbol{x})\right]. \tag{3}$$

That is, the α-superquantile is the conditional expectation of $g(\boldsymbol{x},\boldsymbol{V})$ when we know that its value is greater than or equal the α-quantile. Rockafellar & Royset (2010) then define the buffered failure probability, $\bar{p}_f(\boldsymbol{x})$, as follows:

$$\bar{p}_f(\boldsymbol{x}) = 1-\alpha, \tag{4}$$

where α is chosen so that $\bar{q}_\alpha(x)=0$. Note that from the previous definitions we have:

$$\bar{p}_f(\boldsymbol{x}) = P(g(\boldsymbol{x},\boldsymbol{V}) > q_\alpha(\boldsymbol{x})) = 1-F\left(q_\alpha(\boldsymbol{x})\right) \tag{5}$$

where F denotes the cumulative distribution function of $g(\boldsymbol{x},\boldsymbol{V})$.

In order to show how to calculate the buffered failure probability $\bar{p}_f(\boldsymbol{x})$, we consider the plot shown in Figure 1. The curve in the plot represents the cumulative distribution function of the performance function, $g(\boldsymbol{x},\boldsymbol{V})$. As an example we have chosen a Gaussian distribution with mean value –2.5 and standard deviation 1.5. For this distribution we have $F(0) = 0.952$, as can also be seen in the figure by considering the right-most vertical dashed line starting at 0 on the x-axis, and the corresponding upper horizontal dashed line starting at

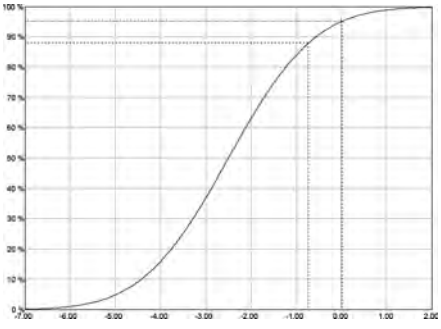


Figure 1. Buffered failure probability calculation where: $p_f(\boldsymbol{x}) = 0.048$, $q_\alpha(\boldsymbol{x}) = -0.743$, $\alpha = F\left(q_\alpha(\boldsymbol{x})\right) = 0.879$, and $\bar{p}_f(\boldsymbol{x}) = 1-\boldsymbol{\alpha} = 0.121$.

0.952. Hence, we get that $p_f(\boldsymbol{x}) = 1-F(0) = 0.048$. In the figure $p_f(\boldsymbol{x})$ is the distance between 100%-line and the upper horizontal dashed line.

Using e.g., Monte Carlo simulation it is easy to estimate $q_\alpha(\boldsymbol{x})$, and we find that $q_\alpha(\boldsymbol{x}) = -0.743$. In the figure $q_\alpha(\boldsymbol{x})$ is represented by the leftmost vertical dashed line. By following this line until it crosses the cumulative curve, we find that $\alpha = F\left(q_\alpha(\boldsymbol{x})\right) = 0.879$. Finally, the buffered failure probability is found to be $\bar{p}_f(\boldsymbol{x}) = 1-\alpha = 0.121$. In the figure $\bar{p}_f(\boldsymbol{x})$ is the distance between 100%-line and the lower horizontal dashed line.

It is easy to see that we always have $q_\alpha(\boldsymbol{x}) \le 0$, and thus, it follows that $\alpha = F\left(q_\alpha(\boldsymbol{x})\right) \le F(0)$. This implies that:

$$\bar{p}_f(\boldsymbol{x}) = 1-\alpha \ge 1-F(0) = p_f(\boldsymbol{x}).$$

Hence, it follows that the buffered failure probability is more conservative than the failure probability. See Rockafellar & Royset (2010) for a detailed discussion of this.

Rockafellar & Royset (2010) present several advantages of using the buffered failure probability instead of the regular failure probability. The following are some of the key arguments:

- In general, the failure probability $p_f(\boldsymbol{x})$ cannot be computed analytically, and the techniques commonly used to approximate it, such as FORM or Monte Carlo methods, can sometimes ignore serious risks. This makes it problematic to apply standard non-linear optimization algorithms in connection to structure design. In contrast, non-linear optimization algorithms are directly applicable when using the buffered failure probability instead.
- The buffered failure probability contains more information about the tail behaviour of the distribution of $g(\boldsymbol{x},\boldsymbol{V})$ than the failure probability.
- The buffered failure probability can lead to more computational efficiency in design optimization when the performance function $g(\boldsymbol{x},\boldsymbol{V})$ is expensive to evaluate.

The *buffered reliability*, $\bar{R}(\boldsymbol{x})$, of the structure is defined as $\bar{R}(\boldsymbol{x}) = 1-\bar{p}_f(\underline{\boldsymbol{x}})$. Since $p_f(\boldsymbol{x}) \le \bar{p}_f(\boldsymbol{x})$, it follows that $R(\boldsymbol{x}) \ge \bar{R}(\boldsymbol{x})$. That is, the reliability of the system is greater than or equal to the buffered reliability. Again, this essentially says that the buffered reliability is more conservative than the reliability.

3 ENVIRONMENTAL CONTOURS

Environmental contours are typically used during the early design phases where the exact shape

of the failure region is typically *unknown*. At this stage it it may not be possible to express a precise functional relationship between a set of design variables $\boldsymbol{x}$ and the performance of the structure. Instead we skip $\boldsymbol{x}$ in the notation and let the design options be embedded in the performance function $g(\boldsymbol{V})$ itself. In particular we denote the failure region simply by $\mathcal{F}$, while the corresponding failure probability, $P(\boldsymbol{V} \in \mathcal{F})$, is denoted by $p_f(\mathcal{F})$.

Although $\mathcal{F}$ is unknown, it may still be possible to argue that $\mathcal{F}$ belongs to some known family, $\mathcal{E}$, of failure regions. As in the previous sections we consider cases where the environmental conditions can be described by a stochastic vector $\boldsymbol{V} \in \mathbb{R}^n$ with a known distribution. An important part of the probabilistic design process is then to make sure that $P(\boldsymbol{V} \in \mathcal{F})$ is acceptable for all $\mathcal{F} \in \mathcal{E}$.

In order to avoid failure regions with unacceptable probabilities, it is necessary to put some restrictions on the family $\mathcal{E}$. This is done by introducing a set $\mathcal{B} \subseteq \mathbb{R}^n$ chosen so that for any relevant failure region $\mathcal{F}$ which do not overlap with $\mathcal{B}$, the failure probability $P(\boldsymbol{V} \in \mathcal{F})$ is *small*. The family $\mathcal{E}$ is chosen relative to $\mathcal{B}$ so that $\mathcal{F} \cap \mathcal{B} \subseteq \partial\mathcal{B}$ for all $\mathcal{F} \in \mathcal{E}$, where $\partial\mathcal{B}$ denotes the boundary of $\mathcal{B}$. This boundary is then referred to as an *environmental contour*. See Figure 2.

Following Huseby et al. (2017) we define the *exceedence probability* of $\mathcal{B}$ with respect to $\mathcal{E}$ as:

$$P_e(\mathcal{B}, \mathcal{E}) := \sup\{p_f(\mathcal{F}) : \mathcal{F} \in \mathcal{E}\}. \tag{6}$$

For a given *target probability* P_e the objective is to choose an environmental contour $\partial\mathcal{B}$ such that:

$$P_e(\mathcal{B}, \mathcal{E}) = P_e$$

We observe that the exceedence probability defined above represents an upper bound on the failure probability of the structure assuming that the true failure region is a member of the family $\mathcal{E}$. Of particular interest are cases where one can argue that the failure region of a structure is *convex*. That is, cases where $\mathcal{E}$ is the class of all convex sets which do not intersect with the interior of $\mathcal{B}$. In the remaining part of the paper we will assume that $\mathcal{E}$ satisfies this.

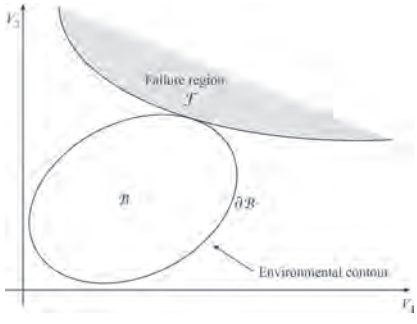


Figure 2. An environmental contour $\partial\mathcal{B}$ and a failure region $\mathcal{F}$

3.1 *Monte Carlo contours*

There are many possible ways of constructing environmental contours. In this paper we focus on the Monte Carlo based approach first introduced in Huseby et al. (2013), and improved in Huseby et al. (2015a) and Huseby et al. (2015b).

Let $\mathcal{U}$ be the set of all unit vectors in $\mathbb{R}^n$, and let $\boldsymbol{u} \in \mathcal{U}$ We then introduce a function $C(\boldsymbol{u})$ defined for all $\boldsymbol{u} \in \mathcal{U}$ as:

$$C(\boldsymbol{u}) := \inf\{C : P(\boldsymbol{u}'\boldsymbol{V} > C) \le P_e\} \tag{7}$$

Thus, $C(\boldsymbol{u})$ is the $(1-P_e)$-quantile of the distribution of $\boldsymbol{u}'\boldsymbol{V}$. Given the distribution of $\boldsymbol{V}$, the function $C(\boldsymbol{u})$ can easily be estimated by using Monte Carlo simulation. Thus, let $V_1, \ldots, V_N$ be a random sample from the distribution of $\boldsymbol{V}$. We then choose $\boldsymbol{u} \in \mathcal{U}$, and let $Y_r(\boldsymbol{u}) = \boldsymbol{u}'\boldsymbol{V}_r$, $r = 1, \ldots, N$. These results are sorted in ascending order:

$$Y_{(1)} \le Y_{(2)} \le \cdots \le Y_{(N)}$$

Using the sorted numbers we first estimate $C(\boldsymbol{u})$. Since $C(\boldsymbol{u})$ is the $(1-P_e)$-quantile in the distribution, a natural estimator is:

$$\hat{C}(\boldsymbol{u}) = Y_{(k)},$$

where k is determined so that:

$$\frac{k}{N} \approx 1 - P_e.$$

Note, however, that this estimator can be improved considerably by using importance sampling. See Huseby et al. (2015b) for details.

For each $\boldsymbol{u} \in \mathcal{U}$, we also introduce the halfspaces:

$$\Pi^-(\boldsymbol{u}) = \{\boldsymbol{v} : \boldsymbol{u}'\boldsymbol{v} \le C(u)\},$$
$$\Pi^+(\boldsymbol{u}) = \{\boldsymbol{v} : \boldsymbol{u}'\boldsymbol{v} > C(u)\}.$$

We then define the environmental contour as the boundary $\partial\mathcal{B}$ of the *convex set* set $\mathcal{B}$ given by:

$$\mathcal{B} := \bigcap_{u \in \mathcal{U}} \Pi^-(u) \tag{8}$$

It follows that the exceedence probability of $\mathcal{B}$ with respect to $\mathcal{E}$ is given by:

$$P_e(\mathcal{B},\mathcal{E}) = \sup\{p_f(\mathcal{F}) : \mathcal{F} \in \mathcal{E}\}$$
$$= \sup\{p_f(\Pi^+(\boldsymbol{u})) : \boldsymbol{u} \in \mathcal{U}\}$$
$$= \sup_{u\in\mathcal{U}} P(\boldsymbol{u}'\boldsymbol{V} > C(\boldsymbol{u})) = P_e,$$

where the second equality follows since we have assumed that $\mathcal{F}$ is convex and hence contained in $\Pi^+(u)$ for all $\mathcal{F} \in \mathcal{E}$ In fact for all $u \in \mathcal{U}$ we have $\Pi^+(u) \in \mathcal{E}$ as well, and these halfspaces are the maximal sets within ε. Moreover, the last equation follows by the definition of $C(\boldsymbol{a})$ given in (7). Thus, we conclude that the contour $\partial\mathcal{B}$ indeed has the correct exceedence probability with respect to ε. See Huseby et al. (2017) for further details regarding this.

4 BUFFERED ENVIRONMENTAL CONTOURS

In this section, we introduce a new concept called *buffered environmental contours*. This combines the ideas behind buffered failure probabilities and environmental contours. Before we introduce the main results we review a result on superquantiles which will be essential in our approach (See Rockafellar (2007).)

Proposition 1. Let g_1 and g_2 be two performance functions such that $g_1(V) \le g_2(V)$ *almost surely, and let* $\bar{q}_{1,\alpha}$ *and* $\bar{q}_{2,\alpha}$ *denote the* α*-superquantiles of* g_1 *and* g_2 *respectively. Then* $\bar{q}_{1,\alpha} \le \bar{q}_{2,\alpha}$.

As a corollary of this result we get the following result on buffered failure probabilities:

Corollary 2. *Let* g_1 *and* g_2 *be two performance functions such that* $g_1(V) \le g_2(V)$ *almost surely, and let* $\bar{p}_{1,f}$ *and* $\bar{p}_{2,f}$ *denote the buffered failure probabilities of* g_1 *and* g_2 *respectively. Then* $\bar{p}_{1,f} \le \bar{p}_{2,f}$.

For a given performance function g its failure probability, p_f, can be computed based on the failure region of g alone. In contrast, computing the buffered failure probability, $\bar{p}_f$, requires more detailed information about the distribution of g. We indicate this by expressing $\bar{p}_f$ as a function of g and denoted $\bar{p}_f(g)$.

Just as for classical environmental contours, a *buffered environmental contour* is the boundary $\partial\mathcal{B}$ of some suitable set $\bar{\mathcal{B}} \subseteq \mathbb{R}^n$. We shall now describe how the set $\bar{\mathcal{B}}$ can be constructed. As in the previous section we let $\mathcal{U}$ be the set of all unit vectors in $\mathbb{R}^n$, and let $u \in \mathcal{U}$ Moreover, we let P_e be a given target probability, and let $C(\boldsymbol{u})$ be defined by (7). In order to introduce buffering, we let:

$$\bar{C}(\boldsymbol{u}) := \mathrm{E}[\boldsymbol{u}'\boldsymbol{V} \,|\, \boldsymbol{u}'\boldsymbol{V} \rangle C(u)]. \tag{9}$$

Given the distribution of V, the function $\bar{C}(u)$ can easily be estimated by using Monte Carlo simulation. As in Subsection 3.1, we let $V_1, \ldots, V_N$ be a random sample from the distribution of $\boldsymbol{V}$, and choose $u \in \mathcal{U}$. Based on the sorted values $Y_{(1)} \le Y_{(2)} \le \cdots \le Y_{(N)}$ we first estimate $C(\boldsymbol{u})$ by $Y_{(k)}$ as previously explained. We then estimate $\bar{C}(u)$ by computing the average value of the sampled values which are greater than $Y_{(k)}$. Thus, we estimate $\bar{C}(u)$ by:

$$\hat{\bar{C}}(\boldsymbol{u}) = \frac{1}{N-k}\sum_{r>k} Y_{(r)}.$$

For each $u \in \mathcal{U}$, we also introduce the halfspaces:

$$\bar{\Pi}^-(\boldsymbol{u}) = \{\boldsymbol{v} : \boldsymbol{u}'\boldsymbol{v} \le \bar{C}(\boldsymbol{u})\},$$
$$\bar{\Pi}^+(\boldsymbol{u}) = \{\boldsymbol{v} : \boldsymbol{u}'\boldsymbol{v} > \bar{C}(\boldsymbol{u})\},$$

similar to what we did in the previous section. Finally, we define the buffered environmental contour as the boundary $\partial\bar{\mathcal{B}}$ of the *convex set* set $\bar{\mathcal{B}}$ given by:

$$\bar{\mathcal{B}} := \bigcap_{u\in\mathcal{U}} \bar{\Pi}^-(\boldsymbol{u}) \tag{10}$$

We observe that by (12) we obviously have that $\bar{C}(\boldsymbol{u}) > C(\boldsymbol{u})$. . By comparing (8) and (10), it is easy to see that this implies that:

$$\mathcal{B} \subset \bar{\mathcal{B}}.$$

Thus, given that the same target probability P_e is used to construct both contours, the buffered environmental contour is more conservative than the classical environmental contour.

The next step is to identify a family $\mathcal{G}$ of performance functions defined relative to the set $\mathcal{B}$ such that $\bar{p}_f(g) \le P_e$ for all $g \in \mathcal{G}$ We recall that for the classical environmental contour we chose to let ε be the family of all convex failure regions which do not intersect with the interior of $\mathcal{B}$. Thus, one might think that the natural counterpart for buffered environmental contours would be to let $\mathcal{G}$ be the family of performance functions with convex failure regions which do not intersect with the interior of $\bar{\mathcal{B}}$. In this case, however, we need more control over the distributions of the performance functions. In order to do so we choose $u \in \mathcal{U}$ and introduce the performance function $\Gamma(u,\cdot)$ given by:

$$\Gamma(\boldsymbol{u},\boldsymbol{V}) = \boldsymbol{u}'\boldsymbol{V} - \bar{C}(\boldsymbol{u})$$

By (12) we have:

$$\mathrm{E}\left[\Gamma(\boldsymbol{u},\boldsymbol{V})\Gamma(\boldsymbol{u},\boldsymbol{V}) > C(\boldsymbol{u}) - \bar{C}(\boldsymbol{u})\right]$$
$$= \mathrm{E}[\boldsymbol{u}',\boldsymbol{V} \mid \boldsymbol{u}',\boldsymbol{V} > C(\boldsymbol{u})] - \bar{C}(\boldsymbol{u}) = 0.$$

Moreover, by (7) we have:

$$\begin{aligned}\bar{p}_f\left(\Gamma(\boldsymbol{u},\cdot)\right) &= P(\Gamma(\boldsymbol{u},\boldsymbol{V}) > C(\boldsymbol{u}) - \bar{C}(\boldsymbol{u})) \\ &= P(\boldsymbol{u}',\boldsymbol{V} > C(\boldsymbol{u})) = P_e\end{aligned}$$

Since the unit vector u was arbitrarily chosen, we conclude that the performance function $\Gamma(u,\cdot)$ has the desired buffered failure probability P_e for all $u \in \mathcal{U}$

We will use these performance functions as a basis for constructing the family $\mathcal{G}$ where the $\Gamma(u,\cdot)$-functions serve as *maximal* elements in this family. Note that the $\Gamma(u,\cdot)$-functions now play a similar role as the halfspaces $\Pi^+(u)$ played in the construction of the family $\mathcal{F}$. Thus, we let $\mathcal{G}$ be the family of all performance functions g for which there exists a $u \in \mathcal{U}$ such that $g(\boldsymbol{v}) \le \Gamma(\boldsymbol{u},\boldsymbol{v})$ for all $v \in \mathcal{V}$. By the above discussion the following result is immediate:

Theorem 3. For all $g \in \mathcal{G}$ we have $\bar{p}_f(g) \le P_e$..

Proof: Assume that $g \in \mathcal{G}$. Then there exists a $u \in \mathcal{U}$ such that $g(\boldsymbol{V}) \le \Gamma(\boldsymbol{u},\boldsymbol{V})$ almost surely. Hence, by Corollary 4.2 and the above calculations we have:

$$\bar{p}_f(g) \le \bar{p}_f\left(\Gamma(\boldsymbol{u},\cdot)\right) = P_e.$$

□

Having constructed both the set $\bar{\mathcal{B}}$ and the family $\mathcal{G}$ we are now ready to introduce the *buffered exceedence probability* of $\bar{\mathcal{B}}$ with respect to $\mathcal{G}$ defined as:

$$\bar{P}_e\left(\bar{\mathcal{B}},\mathcal{G}\right) := \sup\{\bar{p}_f(g) : g \in \mathcal{G}\}. \tag{11}$$

We note that by the definition of $\mathcal{G}$ it follows that $\Gamma(\boldsymbol{u},\cdot) \in \mathcal{G}$ for all $u \in \mathcal{U}$. Hence, we get:

$$\begin{aligned}\bar{P}_e\left(\bar{\mathcal{B}},\mathcal{G}\right) &= \sup\{\bar{p}_f(g) : g \in \mathcal{G}\} \\ &= \sup\{\bar{p}_f\left(\Gamma(\boldsymbol{u},\cdot)\right) : \boldsymbol{u} \in \mathcal{U}\} = P_e,\end{aligned}$$

Thus, we conclude that the contour $\partial\bar{\mathcal{B}}$ indeed has the correct buffered exceedence probability with respect to $\mathcal{G}$.

If $g \in \mathcal{G}$ and $g(\boldsymbol{v}) \le \Gamma(\boldsymbol{u},\boldsymbol{v})$ for all $v \in \mathcal{V}$, we have:

$$\begin{aligned}\mathcal{F}(g) \subseteq \mathcal{F}\left(\Gamma(\boldsymbol{u},\cdot)\right) &= \{\boldsymbol{v} : \boldsymbol{u}'\boldsymbol{v} - \bar{C}(\boldsymbol{u}) > 0\} \\ &= \{\boldsymbol{v} : \boldsymbol{u}'\boldsymbol{v} > \bar{C}(\boldsymbol{u})\} = \bar{\Pi}^+(u)\end{aligned}$$

Thus, the failure region of a performance function $g \in \mathcal{G}$ does not overlap with the interior of the set $\bar{\mathcal{B}}$, but is contained within a halfspace supporting $\bar{\mathcal{B}}$. This is similar to the relation between failure regions in the family ε and the set $\mathcal{B}$ for the classical environmental contours. However, as already pointed out, knowledge about the failure region of a performance function is not sufficient to ensure that the performance function has the correct buffered failure probability.

It may be argued that the choice of the $\Gamma(\boldsymbol{u},\cdot)$-functions as maximal elements in the family $\mathcal{G}$ is too restrictive. In order to have a more flexible framework, it is possible to consider a slightly more general approach where we define:

$$\bar{C}_a(\boldsymbol{u}) := \mathrm{E}[a\boldsymbol{u}'\boldsymbol{V} \,|\, \boldsymbol{u}'\boldsymbol{V} \rangle C(\boldsymbol{u})] = a\bar{C}(\boldsymbol{u}), \tag{12}$$

where a is a positive constant. By increasing the a-factor, the contour may be inflated so that it can be used for steeper performance factors.

On the other hand it should be noted that to ensure that a given performance function g has the correct buffered failure probability, it is not necessary that $g(v)$ is dominated by some $\Gamma(\boldsymbol{u},\cdot)$-function for *all* $v \in \mathcal{V}$. It is sufficient that this holds for v-values corresponding to the upper tail area of g.

5 NUMERICAL EXAMPLE

In this subsection we illustrate the proposed method by considering a numerical example introduced in Vanem & Bitner-Gregersen (2015). More specifically, we consider joint long-term models for *significant wave height*, denoted by H, and *wave period* denoted by T. A marginal distribution is fitted to the data for significant wave height and a conditional model, conditioned on the value of significant wave height, is subsequently fitted to the wave period. The joint model is the product of these distribution functions:

$$f_{T,H}(t,h) = f_H(h) f_{T|H}(t \,|\, h)$$

Simultaneous distributions have been fitted to data assuming a three-parameter Weibull distribution for the significant wave height, H, and a lognormal conditional distribution for the wave period, T. The three-parameter Weibull distribution is parameterized by a location parameter, γ, a scale parameter α, and a shape parameter β as follows:

$$f_H(h) = \frac{\beta}{\alpha}\left(\frac{h-\gamma}{\alpha}\right)^{\beta-1} e^{-[(h-\gamma)/\alpha]^\beta}, \quad h \ge \gamma.$$

The lognormal distribution has two parameters, the log-mean μ and the log-standard deviation σ and is expressed as:

$$f_{T|H}(t \,|\, h) = \frac{1}{t\sqrt{2\pi}} e^{-\left[(\ln(t)-\mu)^2/(2\sigma^2)\right]}, \quad t \ge 0,$$

where the dependence between H and T is modelled by letting the parameters μ and σ be expressed in terms of H as follows:

$$\mu = E\left[\ln(T) \mid H = h\right] = a_1 + a_2 h^{a_3},$$

$$\sigma = SD\left[\ln(T) \mid H = h\right] = b_1 + b_2 e^{b_3 h}.$$

The parameters a_1, a_2, a_3, b_1, b_2, b_3 are estimated using available data from the relevant geographical location. In the example considered here the parameters are fitted based on a data set from North West Australia. We consider data for two different cases: *swell* and *wind sea*. The parameters for the three-parameter Weibull distribution are listed in Table 1, while the parameters for the conditional log-normal distribution are listed in Table 2. In all the examples we use a return period of 25 years. The models are fitted using sea states representing periods of 1 hour. Thus, we get 24 data points per 24 hours. Thus, the desired exceedence probability is given by:

$$P_e = \frac{1}{25 \cdot 365.25 \cdot 24} = 4.5631 \cdot 10^{-6}.$$

For more details about these examples we refer to (Vanem & Bitner-Gregersen 2015).

The classical environmental contours are estimated based on the methods presented in Huseby et al. (2013). More specifically, we have used Method 2 presented in this paper. The buffered environmental contours are estimated in exactly the same way, except that $\hat{C}(\boldsymbol{u})$ is replaced by $\hat{C}(\boldsymbol{u})$ for all $\boldsymbol{u} \in \mathcal{U}$.

In Figure 3 and Figure 4 the resulting environment contours are shown. As one expected, the classical environmental contours are located inside their respective buffered contours. Thus, since the target probability P_e is the same for both types of contours, the buffered contours are more conservative than the classical contours.

Table 1. Fitted parameter for the three-parameter Weibull distribution for signifcant wave heights.

	α	β	γ
Swell	0.450	1.580	0.132
Wind sea	0.605	0.867	0.322

Table 2. Fitted parameter for the conditional log-normal distribution for wave periods.

		$i = 1$	$i = 2$	$i = 3$
Swell	a_i	0.010	2.543	0.032
	b_i	0.137	0.000	0.000
Wind sea	a_i	0.000	1.798	0.134
	b_i	0.042	0.224	–0.500

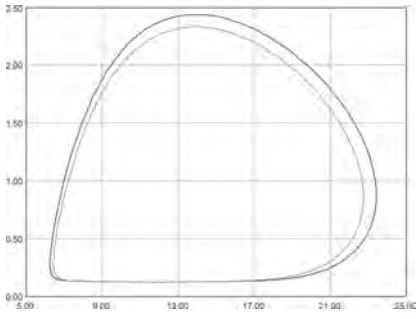


Figure 3. Buffered environmental contour (black) and classical environmental contour (gray) for North West Australia Swell with return period 25 years.

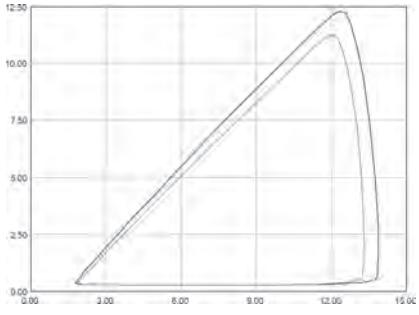


Figure 4. Buffered environmental contour (black) and classical environmental contour (gray) for North West Australia Wind sea with return period 25 years.

6 CONCLUSIONS AND FUTURE WORK

In the present paper we have introduced the concept of buffered environmental contours, and shown how such contours can be estimated using Monte Carlo simulations. Such contours do not just take into account the probability of failure, but also the consequences of a failure. This is relevant e.g., when analysing the risk of flooding at a given location. While it may not be possible to prevent floodings from occurring, the damage caused by such an event can vary a lot depending on how much the water has risen above the normal level. In some cases only minor damages may be the result. In other cases the consequences can be catastrophic.

For a given target probability, P_e buffered environmental contours are generally more conservative than the classical environmental contours. However, in cases where the consequences are more important than the triggering event itself, a higher target probability might be acceptable as long as the damages are manageable. Thus, in real-life applications a buffered environmental contour may not be so conservative after all. At the same time these contours provide much more information

about the tail area of the environmental variables. This may be very useful when a design is optimized.

The buffered environmental contours proposed in this paper are the natural extension of the Monte Carlo contours introduced in Huseby et al. (2013). In particular both contour types are boundaries of convex sets. Sometimes this restriction may lead to contours which include areas of very low probability. Thus, it would be of interest to investigate other ways of constructing buffered contours. In particular, it is possible to modify contours obtained by using the Rosenblatt transformation so that they include buffering. To make this work, however, evaluating the resulting contours becomes very important. The evaluation framework described in Huseby et al. (2017) may serve as a starting point.

Future work in this area also includes the use of buffered environmental contours in design optimization, but with additional design constraints. The question is how such additional constraints can be dealt with. An initial idea is to apply a Lagrange duality method in order to transform the problem into a previously known form.

It would also be interesting to compare buffered environmental contours to the conservative environmental contours defined by Leira (2008). The contours defined in Leira (2008) are typically larger sets than the environmental contours considered in Section 3, which means that they are more conservative when it comes to classifying structures as safe.

Another idea which requires further investigation is how time can be introduced into this model in a less restrictive way. As mentioned in Subsection 2.2, we consider average stochastic environmental conditions $V_1, V_2, \ldots$ over some specified time intervals and assume independence and identical distributions of the V_i s. A more realistic approach would be to introduce a stochastic process in continuous time modelling the environmental situation. It is interesting to see how this affects the model and what consequences this has for the design optimization.

ACKNOWLEDGEMENTS

This paper has been written with support from the Research Council of Norway (RCN) through the project *ECSADES* Environmental Contours for Safe Design of Ships and other marine structures.

REFERENCES

Baarholm, G., S. Haver, & O. Økland (2010). Combining contours of significant wave height and peak period with platform response distributions for predicting design response. *Marine Structures 23*, 147–163.

Ditlevsen, O. (2002). Stochastic model for joint wave and wind loads on offshore structures. *Structural Safety 24*, 139–163.

Fontaine, E., P. Orsero, A. Ledoux, R. Nerzic, M. Prevesto, & V. Quiniou (2013). Reliability analysis and response based design of a moored fpso in west africa. *Structural Safety 41*, 82–96.

Haver, S. & S. Winterstein (2009). Environmental contour lines: A method for estimating long term extremes by a short term analysis. *Transactions of the Society of Naval Architects and Marine Engineers 116*, 116–127.

Huseby, A.B., E. Vanem, & K. Eskeland (2017). Evaluating properties of environmental contours. In M. Cepin and R. Bris (Eds.), *Safety and Reliability, Theory and Applications. Proceedings of the European safety and reliability Conference*, pp. 2101–2109. CRC Press.

Huseby, A.B., E. Vanem, & B. Natvig (2013). A new approach to environmental contours for ocean engineering applications based on direct monte carlo simulations. *Ocean Engineering 60*, 124–135.

Huseby, A.B., E. Vanem, & B. Natvig (2015a). Alternative environmental contours for structural reliability analysis. *Structural Safety 54*, 32–45.

Huseby, A.B., E. Vanem, & B. Natvig (2015b). A new monte carlo method for environmental contour estimation. In T. Nowakowski, M. Mlynczak, A. Jodejko-Pietruczuk, and S.Werbinska-Wojciechowska (Eds.), *Safety and Reliability: Methodology and Applications. Proceedings of the European safety and reliability Conference*, pp. 2091–2098. Taylor & Francis.

Jonathan, P., K. Ewans, & J. Flynn (2011). On the estimation of ocean engineering design contours. *Journal of Offshore Mech. Arct. Eng 136(4)*, 8 pages.

Leira, B.J. (2008). A comparison of stochastic process models for definition of design contours. *Structural Safety 30*, 493–505.

Moan, T. (2009). Development of accidental collapse limit state criteria for offshore structures. *Structural Safety 31*, 124–135.

Pflug, G. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In S.P. Uryasev (Ed.), *Probabilistic Constrained Optimization. Methodology and Applications.*, Norwell, pp. 272–281. Kluwer Academic Publishers.

Rockafellar, R.T. (2007). *Coherent Approaches to Risk in Optimization Under Uncertainty.*, Chapter 3, pp. 38–61. INFORMS Tutorials in Operations Research.

Rockafellar, R.T. & J.O. Royset (2010). On buffered failure probability in design and optimization of structures. *Reliability Engineering and System Safety 95*, 499–510.

Rockafellar, R.T. & S.P. Uryasev (2000). Optimization of conditional value-at-risk. *Journal of Risk 2*, 21–42.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Stat. 23*, 470–472.

Vanem, E. & E. Bitner-Gregersen (2012). Stochastic modeling of long-term trends in wave climate and its potential impact on ship structural loads. *Applied Ocean Research 37*, 235–248.

Vanem, E. & E. Bitner-Gregersen (2015). Alternative environmental contours for marine structural design—a comparison study. *Journal of Offshore Mechanics and Arctic Engineering 137*, 051601–1–051601–8.

Winterstein, S., T. Ude, C. Cornell, P. Bjerager, & S. Haver (1993). Environmental parameters for extreme response: Inverse form with omission factors. In *Proc. 6th International Conference on Structural Safety and Reliability.*, Innsbruck, Austria, pp. 551–557. CRC Press, Taylor and Francis Group.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Technical service life prediction of deteriorating structures

O. Lukoševičienė & R. Kliukas
Department of Civil Engineering, Vilnius Gediminas Technical University, Vilnius, Lithuania

ABSTRACT: The technical service life as a quantitative durability parameter of deteriorating structural members is studied. The probabilistic analysis and prediction of durability of deteriorating structures subjected to recurrent extreme service and climate actions is discussed. The strategy of this prediction is based on the concept that not only a performance but also a safety margin of deteriorating members of load-carrying structures are time-dependent random variables. The effect of coincident recurrent extreme actions on their survival probabilities is analysed. The instantaneous and time-dependent survival probabilities of particular members may be assessed by the method of transformed conditional probabilities. The technical service life t_t, as a quantitative durability parameter of deteriorating members is related with the target value of generalized reliability index. The presented methodology on durability prediction of structures may help engineers to calculate the technical service life of deteriorating structures.

1 INTRODUCTION

The load-carrying structures of buildings and construction works must be designed, constructed, erected and operated in such a way that they maintain safety and all quality parameters during an explicit or implicit period of time without requiring unforeseen costs for their maintenance, repair and reconstruction. Higher serviceability and durability requirements are applied to structures which routine or preventive maintenance and repair require great efforts (Lukoševičienė and Kudzys 2009). Timely maintenance and repairs may prolong their technical service life effectively. Qualitative and quantitative inspections in a standard format present the performance of materials and components. The contemporary standard ISO/CD 15686 (1997) recommends using inspections data for assessment and prediction of the durability of deteriorating and ageing members by deterministic and semi-probabilistic methods.

Stewart and Val (1999), Czarnecki and Nowak (2008), Cheung *et al.* (2009) developed probabilistic design approaches and models, describing gradually deteriorating concrete and steel structures of construction works. The random gradual deterioration of load-carrying systems and decrease in their structural reliability are caused by heterogeneous actions of dynamically changing environmental service loads, and aggressive accumulation of chemicals on concrete or steel surfaces of system members. Besides, a predictive reliability analysis of deteriorating members and systems is required to prevent structures from premature damage and to avoid losses and accidents.

But the technical service life, t_t, as basic durability factor of structural members has a big random scatter and should be treated as a stochastic variable which values may be assessed by probability-based approaches and methods. Besides, using deterministic and semi-probabilistic approaches, it is inconceivable to fix a real reliability index of a deteriorating member the failure domain of which changes with time. A durability as time-dependent probabilistic reliability of elements may be prolonged by repairs of materials and components. The standard differentiation in the reliability of structures is based only on classes of failure consequences (EN 1990 2002). However, the methodology of sustainable durability predictions requires taking into account future repair and replacement abilities of deteriorating structural members. The minimum values for reliability index, β_T, associate with the structures or structural members.

Purpose of this study is to suggest approaches of the probability-based prediction of a time-dependent safety, reliability index and technical service life of deteriorating members of structures when they are affected by aggressive actions and recurrent extreme loads.

2 RELIABILITY ANALYSIS OF DETERIORATING STRUCTURES

2.1 *Deterioration of load-carrying structures*

The probability-based reliability analysis of structural systems (frameworks, truss, carcasses) or their subsystems i.e. structural members (beams, slabs,

columns, joints) as physically distinguishable part of a building or civil engineering work may be objectively assessed and predicted only knowing survival probabilities of particular members (normal, cross or oblique sections, connections, deflections). The reliability prediction of deteriorating members and their systems will account for all extreme action combinations. Predicted durability parameters for deteriorating structures depend on chemical or physical diagnosis and the acceptable risk of serviceability failure associated with the damage levels and losses.

Aggressive actions induced by concrete carbonation, chloride penetration and other chemical attacks are the main factors determining the deterioration of concrete members and their systems (Lukoševičienė and Kudzys, 2009). The aggressive environmental conditions cause the degradation of concrete covers and reinforced bars lead to the irreversible limit state of load-bearing members. A process of concrete carbonation leading to global depassivation of reinforcement bars is initiated by carbon dioxide. The highly alkalinous environment with the pH value larger than 12 protects steel members from corrosion, when carbon dioxide, CO_2, reaches the reinforcing bars and the corrosion product, $Fe(OH)_2$, begins to cover their surfaces (CEB Bulletin 238, 1997). The values of deterministic durability parameters of members due to their deteriorating are assessed and presented at design codes or standards by deterministic implicit recommendations and related with long-term experience of designers (EN 1990, 2002; ISO 2394, 1998; EN 12500, 2000).

According to JCSS (2000), in any case, it is expedient to divide the life cycle t_n of deteriorating structures into the initiation, t_{in}, and propagation, t_{pr}, periods (Figure 1). Using hierarchical models, the time-dependent resistance and action effect are presented in Figure 1 (b). The resistance of particular members in the propagation may be defined as (Mori and Ellingwood, 1993)

$$R(t) = \varphi(t) R_{in} \tag{1}$$

where $\varphi(t)$ denotes the deterioration function; R_{in} is the initial value of member resistance.

The deterioration function of particular members caused by corrosion may be presented in the form:

$$\varphi(t) = 1 - a(t - t_{in})^b \le 1; \tag{2}$$

where a is a degradation intensity factor; b defines a non-linearity of the deterioration function; t – time being considered; t_{in} – time of corrosion initiation (Mori & Ellingwood, 1993; Zhong & Zhao, 2005). Deterioration function is close to rectlinear ($b \approx 1$) and parabolic ($b \approx 2$) when corresponding degradation mechanisms are a steel corrosion and aggressive environmental attacks. Marine corrosion process of steel structures is not linear function of time (Melcher et al., 2008). The deterioration function by Eq. (2) analysis are shown in Figure 2, when b = 1 and $a = 0.00125, 0.0025, 0.00375$ (rectlinears *2*, *3*, *4*), respectively. The coefficient of variation of deteriorating member resistance may be expressed as $\boldsymbol{\delta} R(t) = \left[\boldsymbol{\delta}^2 R_{in} + \boldsymbol{\delta}^2 \varphi(t)\right]^{1/2}$. Its components $\boldsymbol{\delta} R_{in} = \sigma R_{in} / R_{in,m}$ and $\boldsymbol{\delta} \varphi(t) = \sigma \varphi(t) / \varphi_m(t) = a_m (t - t_{in}) \times \boldsymbol{\delta} a / \left[1 - a_m (t - t_{in})\right]$ characterize the variations of initial resistance R_{in} and rectilinear deterioration function from Eq. (2). When the parameters of a factor a of this function $a_m < 0.005$, $\boldsymbol{\delta} a = 0.3 - 0.5$ and $\boldsymbol{\delta} R_{in} = 0.08 - 0.20$ the random value $\boldsymbol{\delta} R(t_{pr})$ insignificantly exceeds the coefficient of variation $\boldsymbol{\delta} R_{in}$. Therefore the variances of time-dependent and initial resistances of deteriorating particular members are very close in their values, i.e. $\boldsymbol{\sigma}^2 R(t) \approx \boldsymbol{\sigma}^2 R_{in}$.

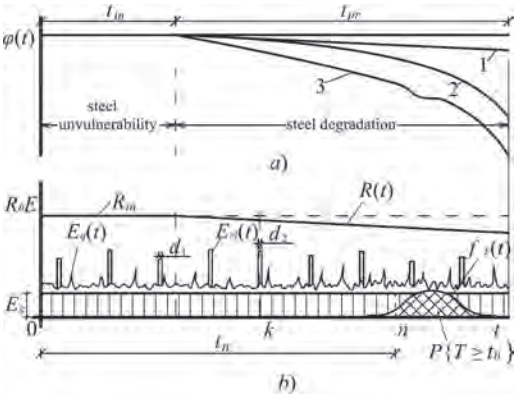


Figure 1. Degradation function $\varphi(t)$ (*a*) and dynamic model (*b*) for time-dependent reliability analysis: *1* – unloaded members, *2* – loaded columns, *3* – loaded beams (Kudzys & Lukoševičienė, 2009).

According to Stewart and Val (1999) also Enright and Frangopol (1998) studies of reinforcement corrosion of reinforced concrete bridge beams, it has led to propose several resistance deterioration functions (Figure 2) that relate to low, medium and high deterioration (curves 1, 5, 6), respectively:

$$\varphi(t) = 1 - at + a_1 t^2; \tag{3}$$

when

$$a = \begin{cases} 0.0005 & a_1 = 0 \;\; t_{in} = 10 \, years \;\; as \; low \\ 0.005 & a_1 = 0 \;\; t_{in} = 5 \, years \;\; as \; medium \\ 0.01 & a_1 = 0.00005 \;\; t_{in} = 2.5 \, years \;\; as \; high \end{cases}$$

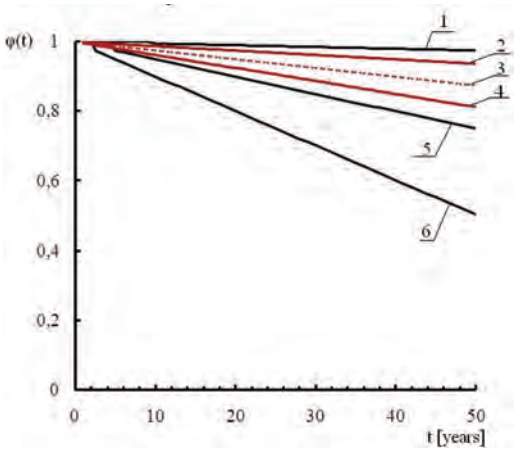


Figure 2. Deterioration function calculated using curves (*1*, *5*, *6*) from Eq. (3), rectlinears (*2*, *3*, *4*) from Eq. (2).

where a, a_1 – degradation intensity parameter. It is assumed that a_1 is a random variable with the coefficient of variation equal to 0.2.

The deterioration function in initiation and propagation periods of reinforced concrete and steel structures is shown in Figure 1 (a).

2.2 *Random safety margin of particular members*

Using the hierarchical dynamic model for time-dependent reliability analysis, the time-dependent safety margin of deteriorating particular members exposed to permanent and variable action effects may be written in the form:

$$Z(t) = g\left[\, \boldsymbol{X}(t), \boldsymbol{\theta} \right] = \theta_R R(t) - \theta_g E_g - \theta_q E_{q_s} - \\ - \theta_q E_{q_e}(t) - \theta_s E_s(t) - \theta_w E_w(t) \tag{4}$$

here $\boldsymbol{X}(t)$ and $\boldsymbol{\theta}$ are the vectors of basic physical and additional variables, representing random components (resistances and action effects) and their model uncertainties; $R(t)$ – time-dependent resistance of deteriorating members; $E_g, E_{q_s}, E_{q_e}, E_s$ and E_w are the action effects caused by permanent, sustained, extraordinary service (live), snow and wind loads, respectively; $\theta_R, \theta_g, \theta_{q_s}, \theta_{q_e}, \theta_s, \theta_w$ are additional random variables containing the design model uncertainties associated with resistances and action effects of particular members. The mean values and standard deviations of additional variables are: $\theta_{Rm} = 0.99 - 1.10$, $\boldsymbol{\sigma}\theta_R = 0.05 - 0.10$ and $\theta_{gm} = \theta_{q_s m} = \theta_{q_e m} = \theta_{sm} = \theta_{wm} \approx 1.00$, $\boldsymbol{\sigma}\theta_g = \boldsymbol{\sigma}\theta_{q_s} = \boldsymbol{\sigma}\theta_{q_e} = \boldsymbol{\sigma}\theta_s = \boldsymbol{\sigma}\theta_w \approx 0.10$ (Hong and Lind 1996, JCSS 2000);

The resistance as static or dynamic structural response for which a probability distribution can be described by normal or lognormal distribution laws (ISO 2394 1998; EN 1990 2002). According to JCSS (2000), EN 1990 (2002), ISO 2394 (1998), Mori and Kato (2003) recommendations, a Gaussian distribution law is to be used for permanent actions, when lognormal, Gaussian, Weibull and gamma distributions may be assumed for sustained live loads. Imposed intermittent extraordinary service and industrial actions may be assumed to be distributed by exponential law (Vrowenvelder 2002). The Gumbel cumulative distribution function is quite appropriate for the probability analysis of structures subjected to climate annual extreme wind and snow loads (ISO 4354 1998, JCSS 2000).

The structural safety analysis of deteriorating members may be based on the limit state criteria:

$$Z_k = R_k - E_k, k = 1, 2, ..., n-1, n, \tag{5}$$

where $R_k = \varphi_k \theta_R R_{in}$ is the resistance of deteriorating members at the sequence cut *k;* E_k is action effects at the same sequence cut k.

The means and variances of these components of the safety margin Z_k given by Eq. (5) may be expressed:

$$\left(\theta_R R_k\right)_m = \theta_{Rm}\left[1 - \alpha_m (t - t_{in})^{\beta}\right] R_{in,m} \tag{6}$$

$$\boldsymbol{\sigma}^2\left(\theta_R R\right) = \theta_{Rm}^2 \boldsymbol{\sigma}^2 R + R_m^2 \boldsymbol{\sigma}^2 \theta_R \tag{7}$$

$$\theta_{Em} E_m \tag{8}$$

$$\boldsymbol{\sigma}^2 E = \theta_{Em}^2 \boldsymbol{\sigma}^2 E_k + E_m^2 \boldsymbol{\sigma}^2 \theta_E. \tag{9}$$

When extreme action effects are caused by two stochastically independent variable actions, a failure of members may occur not only in the case of their coincidence but also when the value of one out of two effects is extreme. Therefore, three stochastically dependent safety margins should be considered as follows:

$$Z_{1k} = R_k - E_{1k}, k = 1, 2, ..., n_1, \tag{10}$$

$$Z_{2k} = R_k - E_{2k}, k = 1, 2, ..., n_2, \tag{11}$$

$$Z_{3k} = R_k - E_{12k} = R_{ck} - E_{1k} - E_{2k}, k = 1, 2, ..., n_{12}, \tag{12}$$

where n_{12} is the number of sequence cuts.

The durations of extreme floor and climate actions are: d_q = 1–14 days for merchant and 1–3 days for other buildings, d_s = 14–28 days and d_w = 8–12 hours. Renewal rates of annual extreme actions are equal to λ = 1/year. Therefore, the recurrence number of two joint extreme actions

during the design working life of structures, t_n in years, may be calculated by the formulae:

$$n_{12} = t_n(d_1 + d_2)\lambda_1\lambda_2 \tag{13}$$

where $\lambda_1 = \lambda_2 = 1/t_\lambda$ are the renewal rates of extreme loads (Lukoševičienė and Kudzys 2009).

2.3 *Instantaneous and time-dependent survival probabilities of particular members*

The instantaneous survival probability of particular members at k-th extreme situation, assuming that they were safe at the situations 1,2,…,k–1, may be expressed as:

$$P(S_k) = P(Z_k > 0) = P\{R_k - E_k > 0\},\ k = \overline{1,n} \tag{14}$$

The values of instantaneous survival probability of particular members may be calculated using analytical, numerical integration and Monte Carlo simulation methods. The resistance R_k and single extreme action effect E_k may be treated in the design of structural safety of ductile particular members as statistically independent variables of their random safety margins. Therefore, the instantaneous survival probability of deteriorating members can be expressed by convolution integral as:

$$P(S_k) = \int_0^\infty f_{R_k}(x) F_{E_k}(x)dx \tag{15}$$

where $f_{R_k}(x)$ is the density function of resistance of a member and $F_{E_k}(x)$ is the cumulative distribution function of its extreme action effect.

The computer program (Lukoševičienė and Kudzys, 2009) was written in the Matlab environment and adapted to predictions of time-dependent instantaneous survival probabilities of ductile autosystems of the reinforced concrete or steel members. So, the time-dependent survival probability of particular member as stochastic autosystem with n elements may be defined as follows:

$$P_{as}(S) = P(S_1) \times \prod_{k=2}^{n} \left(P(S_k) \times \left\{ 1 + \rho_{k|f}^{x_k} \left[\frac{1}{P(S_k)} - 1 \right] \right\} \right) \tag{16}$$

where n is the number of extreme situations;

$$\rho_{k|f}^{x_k} = \left(\frac{1}{k-1} \sum_{i=1}^{(k-1)} \rho_{ki} \right)^{x_k} \tag{17}$$

is the bounded conventional correlation factor of a quadratic matrix between safety margins of system elements as a member of conventional correlation vector written in the form:

$$\boldsymbol{\rho} = \left[1, \rho_{2|1}^{x_2}, \ldots, \rho_{k|1\ldots(k-1)}^{x_k} \right] \tag{18}$$

When the vector $\boldsymbol{\rho} = 0$ or $\boldsymbol{\rho} = 1$, the proposed Eqs. (16), (22) and (23) give accurate solution.

Basic correlation coefficients between system elements may be calculated by the equation:

$$\rho_{ki} = \rho(Z_k, Z_i) = \frac{Cov(Z_k, Z_i)}{\boldsymbol{\sigma} Z_k \cdot \boldsymbol{\sigma} Z_i} \tag{19}$$

where $Cov(Z_k, Z_i)$ and $\boldsymbol{\sigma} Z_k$, $\boldsymbol{\sigma} Z_i$ are covariance and standard deviations of safety margins Z_k and Z_i calculated by Eq. (5).

The bounded index of a conventional correlation factor may be expressed as:

$$x_k \approx P(S_1) \times \left[(4.5 + 4\rho_k)/(1 - 0.98\rho_k) \right]^{\nu_k} \tag{20}$$

where its index

$$\nu_k = \frac{\sum_{i=2}^{k} P(S_k)}{(k-1) \times (k + 2\beta_k^2 - 15\beta_k + 30)^{1/2}} \tag{21}$$

helps us evaluate an effect of survival probabilities of system elements $P(S_k) = \Phi(\beta_k)$ on the bounded index by Eq. (16) when a reliability index of k-th element is equal to β_k.

The relation between bounded correlation factor and correlation coefficient for systems consist of four and ten elements are presented in Figure 3.

2.4 *Technical service life prediction*

The technical service life as the lifetime at preset target reliability index of deteriorating members is the period for which it can actually perform, according to the service requirements based on an intended purpose, without major repairs. In any case, the technical service life, t_t, of members comes to an end before the beginning of concrete spalling or steel cracking process at an attack phase (Lukoševičienė and Kudzys, 2009). The design service life of structures is their working life used in the design process taking into consideration its probable dispersion since it is a time-dependent random quality. The design service life should ensure the required level of safety with respect to target service life (Poukhonto, 2003). Target service life of buildings and structures is specified by the client or owner in accordance with general rules (EN 1990, 2000; ISO 2394, 1998). The requirements for the technical service life

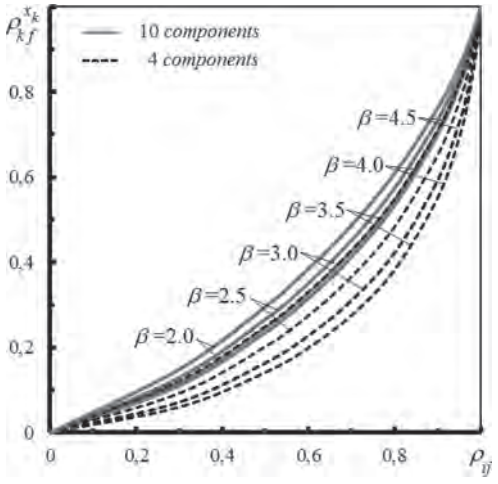


Figure 3. The bounded correlation factor versus basic correlation coefficient.

depending on the level under consideration, may be specified for structural integrity of the building, load carrying capacity and strength of the materials. Most of these requirements have been included in codes and standards (Trbojevic, 2009).

The durability prediction of structures should be considered for beams, columns, slabs, piles, joints and other structural members as auto systems representing their multicriteria failure mode due to various action effects and responses of particular members. Illustrations of series, series-parallel and parallel system are shown in Figure 4. For example, continuous beams are characterised using stochastically dependent conventional elements in series-parallel connections (Figure 4 (c)). Due to system redundancy, according to the research, limit state in any one normal section 1 or 2 of beams does not mean their failure. Besides, the failure of beams in any oblique section 3 implies the failure of auto system.

According to concepts of transformed conditional probabilities, the total survival probabilities of structural members (beams, columns, plates, trusses) as series, series-parallel and parallel systems may be expressed:

$$P_{ss}(S) \approx P(S_1)\times\prod_{k=2}^{m}\left(P(S_k)\times\left\{1+\rho_{k|f}^{x_k}\left[\frac{1}{P(S_k)}-1\right]\right\}\right) \quad (22)$$

$$P_{ps}(S) \approx 1-P(F_1)\times\prod_{k=2}^{m}\left(P(F_k)\times\left\{1+\rho_{k|f}^{x_k}\left[\frac{1}{P(F_k)}-1\right]\right\}\right) \quad (23)$$

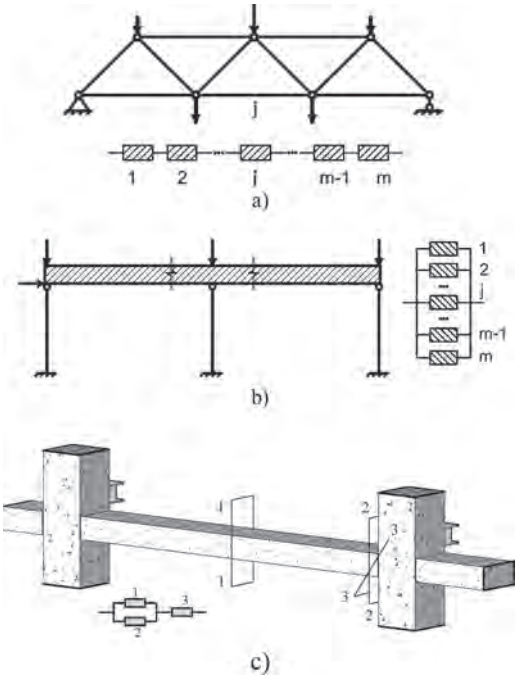


Figure 4. Illustrations of series (a), series-parallel (c) and parallel (b) systems.

$P(S_k)$ from Eq. (22) and $P(F_k)$ from Eq. (23) denote the survival and failure probabilities of ductile autosystem $k = 1\ldots m$; where m is the number of random discrete failure modes; $\rho_{k|f}^{x_k}$ is the bounded conventional correlation factor assessed by Eq. (17).

The generalized reliability indices of series and parallel systems are expressed by:

$$\beta_{ss} = \Phi^{-1}\left[P_{ss}(S)\right] \quad (24)$$

$$\beta_{ps} = \Phi^{-1}\left[P_{ps}(S)\right] \quad (25)$$

where $\Phi^{-1}[\bullet]$ is the inverse cumulative distribution function of the standard normal distribution of their survival probabilities $P_{ss}(S)$ by Eq. (22) and $P_{ps}(S)$ by Eq. (23) (Kudzys and Lukoševičienė 2016).

The design documents recommended by the Joint Committee on Structural Safety (JCSS, 2000) are acknowledged as the progressive probabilistic model code in the design of structural members and their systems. The reliability index of structures is related to the consequences of failure classes when the target value is equal to β_t = 3.3–4.3 (EN 1990, 2002; ISO 2394, 1998). Particular elements and members of the structure may be designed on

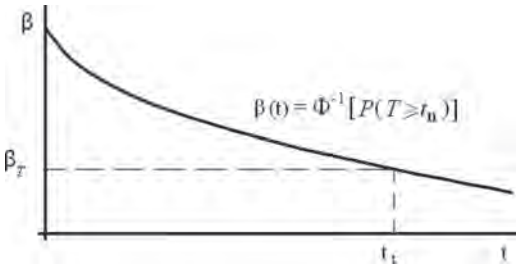


Figure 5. Determination of technical service life t_t of a structural member using the time-dependent reliability index curve.

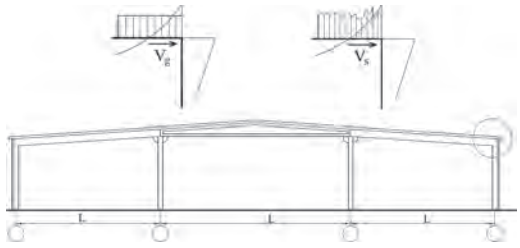


Figure 6. Scheme of single storey building.

Table 1. The failure probabilities in normal random vector space of the concrete structure under the effect of carbonization and chloride ions aggressive actions.

Sampling time	Failure probability		Coefficient of correlation	$P(F_2 \cap F_1)$	$P(F_2 \cup F_1)$	Method
	carbonization	Chloride ions				
The 20th year	198×10^{-7}	107×10^{-7}	0.93	57×10^{-7} 56.8×10^{-7}	248×10^{-7} 248.2×10^{-7}	Zhong & Zhao (2005) TCPM
The 40th year	0.04847	0.00914	0.95	0.00908 0.00709	0.04853 0.05052	Zhong & Zhao (2005) TCPM
The 60th year	0.4053	0.1335	0.94	0.1334 0.1263	0.4054 0.4125	Zhong & Zhao (2005) TCPM

the same higher or lower reliability index as for the entire structure. The reliability classes may be defined by the reliability index β concept. Three reliability classes RC1, RC2 and RC3 may be associated with the three consequences classes CC1, CC2 and CC3.

The technical service lives as a quantitative durability parameter of deteriorating structural members may be calculated from Eqs. (22) – (23) when t_t its value becomes like to their life cycle t_n. The computation is iterated until the value t_t that corresponds the target probability $P(T \geq t_n) = \Phi(\beta_T)$ (Figure 5).

The value of technical service life, t_t, at preset reliability index can be defined by analytic-graphic approaches.

3 NUMERICAL EXAMPLES

3.1 *The time-dependent structural safety of deteriorating reinforced concrete members*

The time-dependent structural safety of deteriorating reinforced concrete members are investigated Zhong and Zhao (2005). They proposed the algorithm for failure probability analysis. This method, interesting for practical use, helps engineers to determine approximately the probability of a structure with different failure modes. The sufficient accuracy of numerical values by method of TCP in comparison with this algorithm and MCS methods is demonstrated in Table 1.

3.2 *Technical service life prediction of single storey building members*

The procedure of technical service life prediction is applied to the knee-joint of single storey reinforced concrete building (Figure 6).

Their degradation process is caused by concrete carbonation and the function is: $\varphi(t) = 1 - 0.004(t - t_{in})$, where $t_{in} = 12$ years is the initiation period. The values of parameters of additional variables are: $\theta_{Rm} = \theta_{gm} = \theta_{sm} = 1.0$, $\sigma^2\theta_R = \sigma^2\theta_g == \sigma^2\theta_s = 0$.

The values of mean and variance of shear resistances in this period are: $R_{in,m} = 387.6\ kN$, $\sigma^2 R_{in} = \sigma^2 R(t) = (0.128 \cdot 387.6)^2 = 2461.4(kN)^2$.

The values of means and variances of shear forces caused by permanent and snow loads are: $V_{gm} = 77.7\ kN$, $\sigma^2 V_g = (0.10 \cdot 77.7)^2 = 60.4\ (kN)^2$; $V_{sm} = 30.4\ kN$, $\sigma^2 V_s = (0.60 \cdot 30.4)^2 = 332.7\ (kN)^2$.

The time-dependent survival probability $P(T \geq t)$ of deteriorating knee-joints was calculated using the method of transformed conditional

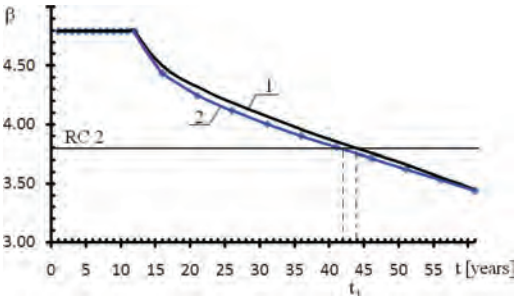


Figure 7. Technical service lives t_1 of knee-joints from the graph of time-dependent reliability indexes by TCPM (1) and numerical integration (2).

probabilities expressed by Eq. (22) and numerical integration methods. The time-dependent drop of reliability indexes $\beta(t) = \Phi^{-1}\left[P\left(T \geq t \right) \right]$ of knee-joints is demonstrated in Figure 7.

According to this Figure 7, the technical service live of considered knee-joints for structures RC2 class are equal about 42 or 44 years using the methods of transformed conditional probabilities (curve *1*) and numerical integration (curve *2*).

The illustrative examples show that the results of system analysis obtained by the proposed TCP method are very close to the probabilistic data computed by exact but sophisticated numerical integration.

4 CONCLUSIONS

The probabilistic technical service life concept related to target reliability indexes of deteriorating particular and structural members helps us to represent their quantitative durability. The probabilistic technical service life of structural members as series, series-parallel and parallel systems may be predicted by analytic-graphic method.

The probability-based analysis and durability prediction of deteriorating members as systems of extreme events may be related to the autosystem concept. The survival and failure probabilities of sustainable series, parallel and series-parallel systems may be calculated using the method of transformed conditional probabilities.

The presented probability-based approaches, design models on durability prediction of structural members may be convenient for many practitioners and can stimulate designers to predict the technical service life of deteriorating structures more actively and effectively.

REFERENCES

CEB Bulletin 238. 1997. New approach to durability design. Sprint Druck, Stuttgart: 138.

Cheung, M.M.S., Zhao, J. & Chan, Y.B. 2009. Service life prediction of RC bridge structures exposed to chloride environments, *Journal of Bridge Engineering* 14(3): 164–178.

Czarnecki, A.; Nowak, A.S. 2008. Time-variant reliability profiles for steel girder bridges. *Structural Safety* 30: 49–64.

EN 12500. 2000. Corrosion likelihood in atmospheric environment.

EN 1990. 2002. Eurocode-Basic of structural design. CEN, Brussels.

Enright, M.P.; Frangopol, D.M. 1998. Probabilistic analysis of resistance degradation of reinforced concrete bridge beams under corrosion. *Engineering Structures* 20(11): 960–971.

Hong, H.P.; Lind, N.C. 1996. Approximate reliability analysis using normal polynomial and simulation results. *Structural Safety* 18(4): 329–339.

ISO 2394. 1998. General principles on reliability for structures. Switzerland.

ISO/CD 15686. 1997. Buildings: Service life planning, Part 1-general principles.

JCSS 2000. Probabilistic model code: Part 1-Basis of design. Joint Committee on Structural Safety, p. 65.

Kudzys, A.; Lukoševičienė, O. 2009. On the safety prediction of deteriorating structures. *Mechanika* (78): 5–11.

Kudzys, A.; Lukoševičienė, O. 2016. The application of a compound model for predicting reliability indices of engineering systems. *Mechanika* 22(2): 143–148.

Lukoševičienė, O.; Kudzys, A. 2009. The durability prediction of deteriorating reinforced concrete members// Budownictwo i Inżynieria Środowiska: zeszyty naukowe Politechniki Rzeszowskiej. Rzeszow. 53(265), 109–118.

Melchers, R.E.; Li, C.Q.; Lawanwisut, W. 2008. Probabilistic modeling of structural deterioration of reinforced concrete beams under saline environment corrosion. *Structural Safety* 30: 447–460.

Mori, Y. & Kato, T. 2003. Multinormal integrals by importance sampling for series system reliability. *Structural Safety* 25: 363–378.

Mori, Y.; Ellingwood, B.R. 1993. Time-dependent system reliability analysis by adaptive importance sampling. *Structural Safety* 12(1): 59–73.

Poukhonto, L.M. 2003. Durability of concrete structures and constructions: silos, bunkers, reservoirs, water towers, retaining walls. A.A. Balkema publishers, 408 p.

Stewart, M.G.; Val, D.V. 1999. Role of load history in reliability-based decision analysis of aging bridges. *Journal of Structural Engineering*: 776–783.

Trbojevic, V.M. 2009. Another look at risk and structural reliability criteria. *Structural Safety* 31: 245–250.

Vrowenvelder, A.C.V.M. 2002. Developments towards full probabilistic design codes. *Structural Safety*: 24(2–4): 417–432.

Zhong, W.Q.; Zhao, Y.G. 2005. Reliability bound estimation for R.C. structures under corrosive effects. *Collaboration and harmonization in creative systems-hara(eds)*, Taylor & Francis group, London, 755–761.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Reliability quantitative analysis method for mechanical system by using extended fault tree

Tianxiang Yu, Yaxin Liu, Xinchen Zhuang & Bolin Shang
School of Aeronautics, Northwestern Polytechnical University, Xian, China

ABSTRACT: Fault tree analysis is an important method in safety engineering and reliability engineering, which is a top-down method and is able to map the relationship between complex events such as system-level failures and basic events such as component-level failures by creating a logic diagram of the overall system. For the reliability analysis of mechanical systems, it's extremely important to obtain the quantitative relationship between the failure and the design parameters except the logic relationship between the top event and the basic event in the design stage. However, the conventional fault tree analysis is hard to build such relationship. As a result, this paper extends the conventional fault tree by developing a new arborescence under basic events of fault tree. A kind of custom gate is defined within our new method. The custom gate is used to build the quantitative relation between the process variables (force, deformation and other performance feature), which can be used to characterize the basic failure mode, and basic random variables (dimension parameters, material parameters and load parameters, etc.). According to different failure models, the custom gate can be different. Then the limited state function of the basic event can be presented based on physical model. The procedure of this method is demonstrated in this paper, and different reliability models for mechanical system can be represented clearly. A Monte Carlo method involving dependent variables is provided to calculate probability of the top event. An illustrative example of a lock mechanism is presented to demonstrate the method in this work.

1 INTRODUCTION

With the rapid development of aerospace industry, the requirement of high-reliability mechanical system is increasing. The reliability evaluation of mechanical system is of great importance. However, the operation environment for mechanical system is getting more and more complex making reliability assessment of these systems harder.

Although reliability evaluation of electronic products is becoming much maturer based on probability and statistics theories, as well as the failure mechanism, it is still very hard when it comes to the reliability assessment of mechanical systems for the diversity of mechanical components and complexity of the failure mechanism. The conventional reliability assessment based on statistics cannot suit the mechanical system very well, as it is hard to associate the basic random variables (for example the component dimensions and the materials etc.) with the failure modes. As a result, the study of methods based on physical causal mechanisms focuses more on reliability of mechanical systems these years. These method combines probability theory with the physical modes [1–2], for instance, the method of Fault Tree Analysis (FTA).

In fault trees, the logical connections between faults and their causes are represented graphically. FTA is deductive in nature meaning that the analysis starts with a top event (a system failure) and works backwards from the top of the tree towards the leaves of the tree to determine the root causes of the top event [3]. FTA was first put forward in 1961 by Watson [4] of Bell Laboratory and was used in the development of Minuteman Missile. Then it was applied in nuclear industry in 1975 [5]. And in 1977, Lapp and Powers develop a method using computer to create fault trees automatically [6–7]. To overcome some shortcomings of conventional FTA, e.g. in handling the uncertainties, allowing the use of linguistic variables, and integrating human error in failure logic model, the fuzzy FTA [8] was developed. Fuzzy FTA provides a framework where basic notions such as similarity, uncertainty and preference can be modeled effectively. In 1994, Sawyer [9] published his paper on the fuzzy fault tree analysis of mechanical system. Lindhe [10] used fault tree analysis on an integrated level, and a probabilistic risk analysis of a large drinking water system in Sweden was carried out. Mao [11] applied fuzzy fault tree to analysis the automatic water supply system in fire control system. Then, dynamic fault tree (DFT) [12–13] is developed. DFT extend conventional FT by defining additional gates called dynamic gates to model complex interactions such as sequence and functional-dependent failures,

spares and dynamic redundancy management [14]. Yuan [15] applied a quantitative reliability analysis method to warm spare gate based on DFT. Nadjafi [16] investigated the reliability of Emergency Detection System, combining Fuzzy Monte Carlo Simulation and DFT.

Although plenty of work has been done on FTA, the present studies on FTA cannot comprehensively show the failure mechanism of mechanical systems. It is unable to track lowest level items that contributing to the failure modes. The probability statistical information of failure rate for bottom event is usually not sufficient, making the quantity analysis inaccurate. Besides, the correlation between the failure modes and the correlation between their causes are usually ignored. All these correlations make the reliability assessment intractable. Since the conventional fault tree is not accurate enough for complex mechanical system reliability assessment, it is necessary to develop an efficient method.

We improve the conventional FTA method and extend it into a new method. The method presented combines physical model with system reliability theory, and most importantly, can associate the basic variables with failure modes of mechanical systems in a systemic way. As a result, connections between faults and lowest level items are represented. In addition, the correlation of the failure modes can be easily identified with this method, and then correlation analysis can be carried out. Besides, this method is convenient to implement programmatically.

The rest of this paper is organized as follows: In section two, we provide an introduction on the conventional FTA method, and then our extended fault tree method is introduced. In section three, we provide the main process of mechanical system reliability assessment. In section four, we give out an engineering example to demonstrate our method. A brief closure, along with a summary of our future works is provided in section five.

2 FAULT TREE ANALYSIS

In this section, we will firstly give a general description about conventional fault tree on its structure and function. Then we will introduce our new method based on the conventional fault tree.

2.1 *Conventional fault tree*

Fault tree is a common method in system reliability analysis. Conventional fault tree is a kind of graphic method using logic gates to describe the relationship between a system-level failure and basic events such as component-level failures.

A fault tree contains four kinds of elements, which are described as follows:

1. The top event, which is on the root of fault tree, represents the failure of system. In mechanical system, the top event means the failure of the system. For example, the top event in fault tree shown in Figure 1 is "S–Re*l*"
2. The bottom events, which are on the tips of fault tree. The bottom events represent the kind of failure modes that can't be resolved any more. For example, in Figure 1, elements from $g_1 < 0$ to $g_3 < 0$ represent the bottom events.
3. The intermediate events. This kind of events is between the top event and bottom events, and represents the failure modes that can be the consequence of bottom events or other intermediate events. For example, in the fault tree shown in Figure 1, element M_1 is the intermediate event.
4. The logic gates. In fault tree, this kind of elements is used to connect events and describe logical relationship between them. The common used logic gates include AND gate and OR gate. The AND gate means that the event upon it happens if any one of the events below it happens. In another word, the events below AND gate are in series relationship.

The OR gate means that the event upon it happens only when all the events below it happens. In another word, the events below OR gate are in parallel relationship.

Using fault tree, we can estimate the importance degree of each event, and based on the assumption that value of each failure mode is certain, the probability of system failure can be evaluated ultimately.

2.2 *Extended fault tree*

Focusing on the complex characteristics of the failure modes in the mechanism system, we improve and extend the conventional FTA. The new method can be set into two parts: the system calculating tree and the failure mode calculating tree.

The system calculating tree is just the same as conventional fault tree. Each bottom event of system calculating tree is connected to the failure mode calculating tree as its top events.

For the system shown in Figure 1, the diagram of failure mode $g_1 < 0$ calculating tree is shown in Figure 2:

As is shown in Figure 2, the failure mode calculating tree includes four kinds of elements, which will be described in detail as follows:

1. The top node

The top node is put on the top of our new arborescence. This element represents a specific failure mode which is one of the bottom events in system

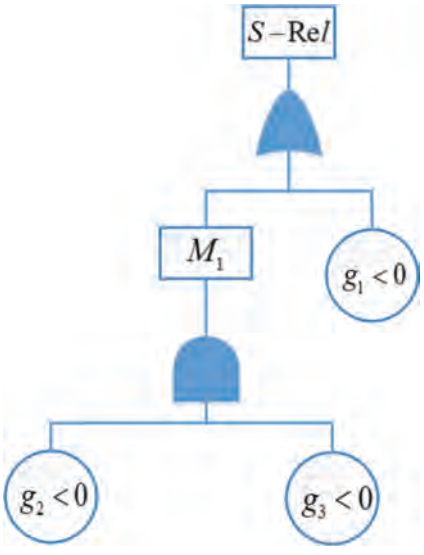


Figure 1. System calculating tree.

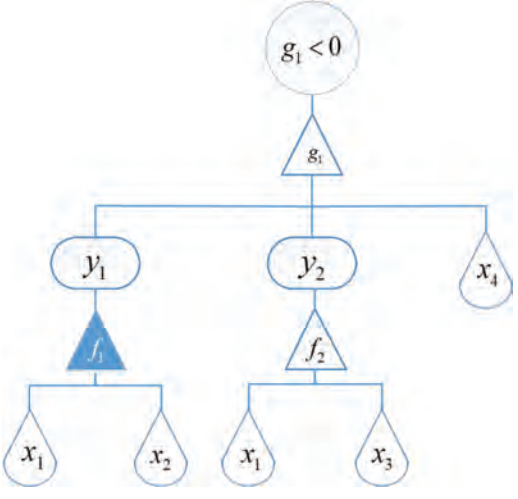


Figure 2. Failure mode calculating tree.

calculating tree. For example, in the failure mode calculating tree shown in Figure 2, the element $g < 0$ is the top node. For convenience of later calculation, we attach the reliability calculation methods to this node. These methods include FOSM, AFOSM, Monte Carlo, etc. As shown in Figure 3, we use a circle to represent the top node.

2. The bottom node

For mechanical system, this kind of elements represents the basic variables. The basic variables can be the geometry size or material of mechanical

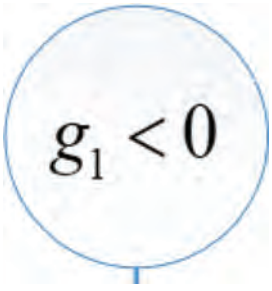


Figure 3. The top node.

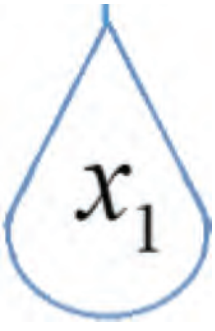


Figure 4. The bottom node.

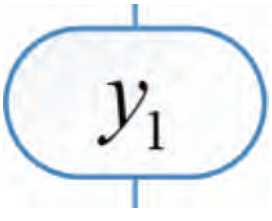


Figure 5. The intermediate node.

parts. In the failure mode calculating tree shown in Figure 2, the elements from x_1 to x_4 are the bottom nodes. Because basic design variables are random in actual engineering, we attach the statistical parameter to these bottom nodes. And in special condition, we will give the probability distribution function (CDF) and probability density function (PDF) of the basic design variables. As show in Figure 4, we use the shape of water drop to represent the bottom nodes.

3. The intermediate node

In Figure 3, the elements y_1 and y_2 are the intermediate nodes. For mechanical system, the intermediate nodes represent the process variables, which depend on the basic design variables. We use the shape in Figure 5 to represent the intermediate node.

4. The custom gate

The failure mode calculating tree is developed to establish connection between a specific failure mode and basic design variables. For this purpose, we put forward a kind of custom gate referring to the function of logic gate in conventional fault tree. We defined that a custom gate should have its input data, output data and a built-in function. The built-in function depends on the physical mechanism of the failure mode.

The custom gate is represented by triangle in the diagram and has different definition in different position.

1. For the connection between basic variables and a specific process variable, we define two kinds of custom gates, the explicit gate and the implicit gate, both of which use basic variables as its input data and process variables as its output data.

When the process variables are obtained by using specific software or algorithm to deal with basic variables, the implicit gate is available and represented by a blue triangle.

The build-in function of an implicit gate can be an instruction to call the proper software or algorithm. For the implicit gate in Figure 2, its build-in function is:

$$y_1 = f_1(x_1, x_2) \tag{1}$$

When there is analytic mathematical model between the process variable and basic variables, the explicit gate is available and represented by a white triangle. The build-in function of explicit gate in Figure 2 is:

$$y_2 = f_2(x_1, x_3) \tag{2}$$

The explicit gate and implicit gate is shown in Figure 6.

2. For the connection between process variables and a specific failure mode, we define a specific gate, the limit sate gate. As shown in Figure 7, this kind of gates use process variables and some basic variables as its input data and output the probability of the failure mode.

This gate contains a limit state equation. The limit state equation is defined by making the value

Figure 6. The implicit gate and explicit gate.

Figure 7. The gate of performance function.

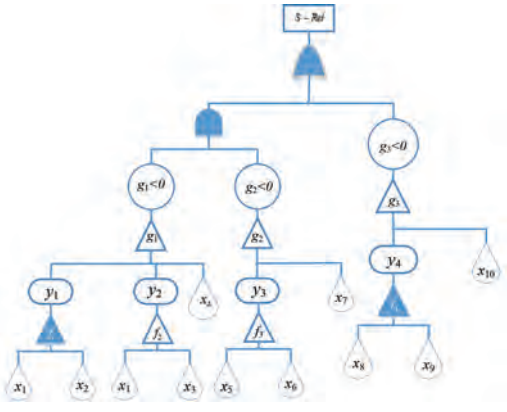


Figure 8. The extended fault tree.

of performance function based on the physical mechanism model of the failure mode equal to zero. For example, the limit state equation in Figure 2 can be expressed as:

$$g_1(y_1, y_2, x_4) = 0 \tag{3}$$

Attaching failure mode calculating tree to the bottom events of system calculating tree, we have our extended fault tree, as is shown in Figure 8.

Based on the extended fault tree described above, we can firstly use failure calculating tree to obtain the probability of each failure mode and then use system calculating tree to obtain the reliability of system.

We can see that there are two main advantages of our extended fault tree compared with the conventional one. First, using our new method, the relationship between basic variables and failure of system can be described in a clear and systematic way. Second, when it comes to a mechanical system with a plenty of complex failure modes, compared with the conventional fault tree, our method can obtain the probability of failure modes more accurately by using implicit gate and explicit gate.

3 PROCESS OF ANALYSIS

The quantitative analysis of the extended fault tree is presented in this section. For a complex mechanical system, in the process of quantitative reliability analysis, the main problem is correlation of the failure modes. To solve the problem of correlation and calculate the reliability of system, a method based on Monte Carlo simulation is provided. The main process is shown in Figure 9.

The process of analysis in detail is as follows:

1. Calculate the minimal cut sets

Cut sets are the unique combinations of failure modes that can cause system failure. Specifically, a cut set is said to be a minimal cut set if, when any basic event is removed from the set, the remaining events collectively are no longer a cut set [17]. The minimal cut sets can be obtained according to the method found by Fussel [18], which is based on the function of logic gates in system calculating tree. For the system shown in Figure 8, there are two minimal cut sets: $\{(g_1 < 0)\}$ and $\{(g_2 < 0),(g_3 < 0)\}$.

Once the minimal cut sets are acquired, then the structure function of system calculating tree can be expressed as:

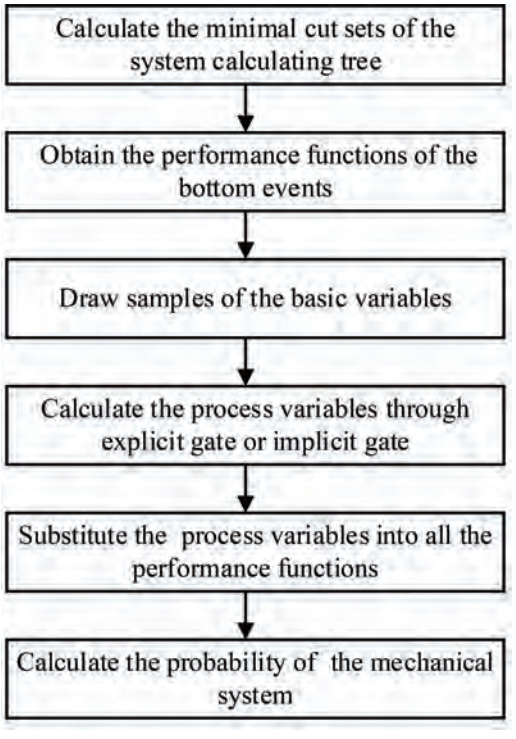


Figure 9. Flowchart of reliability assessment of the mechanical system using Monte Carlo simulation.

$$\Phi(X) = \sum_{j=1}^{N_K} K_j = \sum_{j=1}^{N_K} \prod_{i \in K_j} x_i \tag{4}$$

In Equation (4), N_K is the number of minimal cut sets. K_j represents the *jth* minimal cut, x_i means the *ith* bottom event in the *jth* minimal cut.

$$K_j = \prod_{i \in K_j} x_i \tag{5}$$

As a result, for the system shown in Figure 8, the probability of system failure can be expressed as:

$$P_f = P\{(g_1 < 0) \cup [(g_2 < 0) \cap (g_3 < 0)]\} \tag{6}$$

2. Obtain the performance function of the bottom event

The performance function of the bottom event is obtained by Failure mode calculating tree to associate the bottom nodes or process variables with the bottom event. The common methods used in this process include response surface method (RSM), Kriging method etc.

3. Probability calculation of top event in system calculating tree based on Monte Carlo simulation

The probability of top event in system calculating tree is calculated by

$$\begin{aligned} P(T) = P(K_1 \cup K_2 \cup \cdots \cup K_{N_k}) = \\ \sum_{i=1}^{N_k} P(K_i) - \sum_{i<j=2}^{N_k} P(K_i K_j) + \sum_{i<j<k=3}^{N_k} P(K_i K_j K_k) \\ + \cdots + (-1)^{N_k - 1} P(K_1 K_2 \cdots K_{N_k}) \end{aligned} \tag{7}$$

Considering the correlation between the bottom events and the correlation between minimal cut sets, Equation (7) cannot be solved analytically. Monte Carlo simulation is adopted. Each one of the bottom nodes is considered as a random variable with certain expected values and variances. We draw physical samples from these variables, and calculate the value of process variables by explicit gate or implicit gate. Then probability $P(K_1 K_2 \cdots K_N)$ in Equation (7) is acquired through process variables. At last, the probability of top event can be obtained.

4 CASE STUDY

In this section, we will use our new method to analysis the reliability of a kind of lock mechanism used to control the opening and closing of the aircraft landing gear cabin door.

The mechanism is shown in Figure 10. This system includes six components: a cylinder, a piston, a lock hook, a rocker arm and two connecting rods. There are six revolute joints in the system, which are represented by R0~R5 in Figure 10. The piston is connected to the rocker arm by a spring.

The system has three main failure modes: The start-up failure, the movement failure and the positioning failure.

Each of these failure modes can lead to the failure of system. As a result, the system calculating tree is shown in Figure 11.

In Figure 11, we use $g_1 < 0$ to represent the start-up failure, $g_2 < 0$ to represent the movement failure, $g_3 < 0$ to represent the positioning failure.

The start-up failure occurs when the starting force is smaller than a certain value. As a result, starting force F is considered as the failure index.

The performance function of start-up failure is shown in Equation (8)

$$g_1 = F - [F] \tag{8}$$

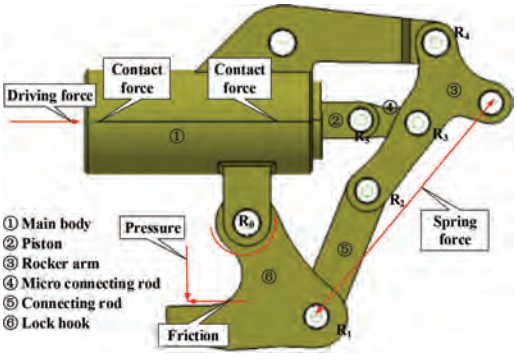


Figure 10. The lock mechanism.

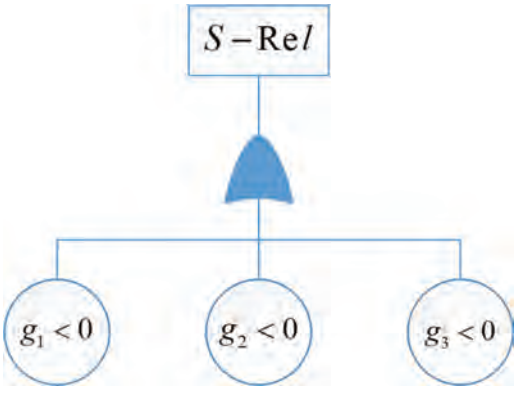


Figure 11. System calculating tree of the lock mechanism.

where F represents the starting force, $[F]$ represents the limit state value of starting force.

The movement failure occurs when time to unlock is shorter than a certain value. As a result, unlocking time is considered as the failure index and the performance function of the movement failure is shown in Equation (9)

$$g_2 = t - [t] \tag{9}$$

where t represents the unlocking time, $[t]$ represent the limit state value of unlocking time.

For both the failure of start-up and movement, there are five influence factors, which are considered as random variables. The distributions of these variables are shown in Table 1.

In Table 1, X_1 represents the elastic coefficient of the spring; X_2 represents the damping coefficient of the spring; X_3 represents the decoupling angle between the lock hook and the lock ring; X_4 represents the maximum contact force between the lock hook and the lock ring; X_5 represents the change rate of driving force attached on piston. There are no explicit functions for the two failure modes. As a result, the implicit functions of these two failure modes are obtained by the dynamics simulation software: LMS Virtual. Lab, as is shown in Figure 12.

As a result, the start-up force and unlocking time can be expressed as follows:

Table 1. The distributions of random variable in the mechanism.

Variables/Unit	Distribution	Mean value	Standard deviation
X_1/(N/m)	Normal	5620	100
X_2/(kg/s)	Normal	478	10
X_3/deg	Normal	46.5	0.5
X_4/N	Normal	4000	400
X_5	normal	14	0.02

Figure 12. Simulation model.

$$F = F(X_1, X_2, X_3, X_4, X_5) \tag{10}$$

$$t = t(X_1, X_2, X_3, X_4, X_5) \tag{11}$$

As a consequence, the failure mode calculating trees of start-up failure and movement failure is shown in Figure 13.

For the positioning failure, to establish the geometric model, we simplify Figure 8 into Figure 14.

Based on the geometric model shown in Figure 9, deflection h is considered as the index of positioning accuracy of the mechanism. The performance function of positioning failure can be expressed as

$$g_3 = h - [h] \tag{12}$$

where h represents the deflection, $[h]$ represents the limit state value of deflection.

According to Figure 12, the deflection h can be written as:

$$h(a,b,f) = a \cdot \sin\left(\frac{\arccos\left(a^2 + f^2 - b^2\right)}{2 \cdot a \cdot f}\right) \tag{13}$$

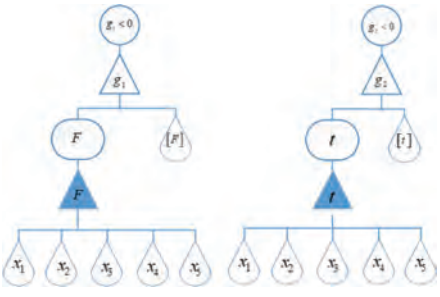


Figure 13. Failure mode calculating trees of start-up failure and movement failure.

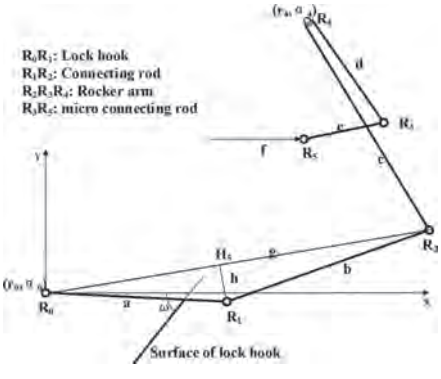


Figure 14. Schematic diagram of the system.

In Equation (13), there are three independent variables: the length of the lock hook represented by a, the length of connecting rod represented by b, the length of piston represented by f. These independent variables are random due to manufacture errors and assembly tolerance. The distributions of their error are shown in Table 2.

Then the failure mode calculating tree of positioning failure is shown in Figure 15.

According to the method mentioned in this paper, the extended fault tree of this lock mechanism is shown in Figure 16.

From Figure 16, it can be concluded that start-up failure and the movement failure are dependent. The positioning failure is independent with the other two failure modes.

There are three minimal cut sets: $\{(g_1 < 0)\}$, $\{(g_2 < 0)\}$ and $\{(g_3 < 0)\}$. According to the process of analysis in Section 3, 1,000,000 samples of the variables (X1, X2, X3, X4, X5, a, b, f) are drawn. Substitute the samples into Equation (10), (11), (13) corresponding the three failure modes. For the three failure modes, failures occur if the values are less than or equal to zero. The probabilities of the three

Table 2. The distributions of length error of parts.

		Error			
	Initial length	Lower limit	Upper limit	Mean value	Standard deviation
a/mm	47.566	–0.226	0.011	–0.1075	0.0395
b/mm	56	–0.240	0.048	–0.0960	0.0480
f/mm	0	–0.250	0.250	0	0.0833

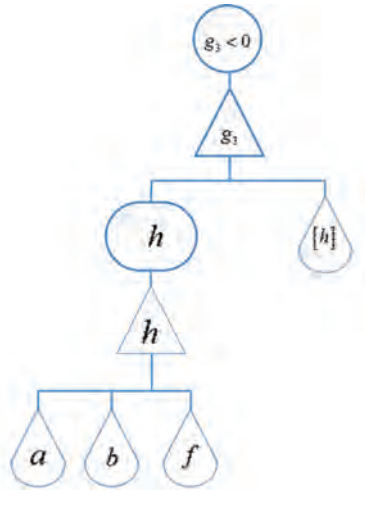


Figure 15. Failure mode calculating trees of positioning failure.

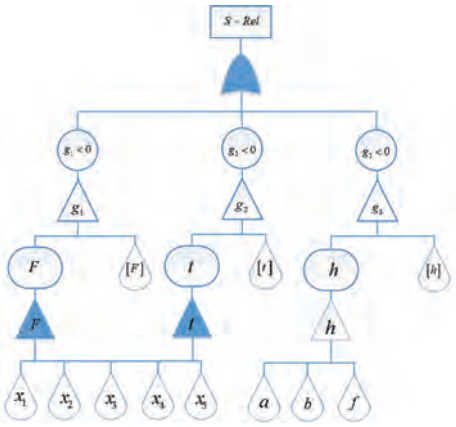


Figure 16. Extended fault tree of the lock mechanism.

failure modes are 0.0035, 0.0045 and 0.0082, respectively. The probability of the lock system is 0.0141.

5 CONCLUSIONS

In this paper, a new method of mechanical system reliability quantitative analysis using extended fault tree is introduced. Focusing on the failure modes of mechanism system, we improve and extend the conventional FTA. The new method can be set into two parts: the system calculating tree and the failure mode calculating tree. A kind of custom gate is defined in our new method to associate the basic random variables and the bottom events. A procedure for reliability evaluation of the mechanical system based on the extended fault tree is provided, in which the correlation of the bottom events is considered. A lock mechanism is investigated to demonstrate our method. The result shows that our extended fault tree can effectively be used to assess the reliability of the mechanical system.

In the process of our study, it is found that the correlation exists in mechanical system is extremely complex. As a consequence, in the future, we will focus more on correlation exists among failure modes, and develop a series of software based on our extended fault tree.

ACKNOWLEDGEMENTS

This work is financially supported by the National Natural Science Foundation of China (Grant No. 51675428).

REFERENCES

[1] T.W. Dakin. "Electrical insulation deterioration treated as a chemical rate phenomenon". AIEE Transctions, vol.67, pp.113, 1948.

[2] A.A. Johnson. "Low Cycle Impact fatigue properties of pearlitic plain carbon steels," Fatigue and Fracture of Engineering Materials and Structures, vol.8, pp.87–294, 1985.

[3] A. Volkanovski, M. Cepin, B Mavko. "Application of the fault tree analysis for assessment of power system reliability." *Reliability Engineering & System Safety*, vol.94: 1116–1127, 2009.

[4] W.S. Lee, D.L. Grosh, F.A. Tillman, et al. "Fault tree analysis, methods, and applications—a review," IEEE *Transactions on Reliability*, vol.34(3): pp.194–203, 1985.

[5] N.C. Rasmussen. "Nuclear power: Rasmussen on reactor safety: How nuclear power reactor risks are quantified; and nuclear sabotage, theft, shipping, and waste disposal risks put in perspective," IEEE Spectrum vol.12, pp.46–47, 1975.

[6] S.A. Lapp, G.J. Powers. "Computer-aided synthesis of fault tree," IEEE Transaction on *Reliability*[C], Pittsburgh, pp.2–12, 1977.

[7] S.A. Lapp, G.J. Powers. "Update of Lapp-Powers fault tree synthesis algorithm," IEEE Transactions on *Reliability*[C], Sewickley, pp.12–15, 1979.

[8] H. Tanaka, L.T. Fan, F.S. Lai, et al. "Fault-tree analysis by fuzzy probability," IEEE Transactions on *Reliability*, vol.32, pp.453–457, 1983.

[9] J.P. Sawyer, S.S. Rao. "Fault tree analysis offuzzy mechanical system" Microelectronics and *Reliability*, vol.34, pp.653–667, 1994.

[10] A, Lindhe, L. Rosen, T. Norberg. "Fault tree analysis for integrated and probabilistic risk analysis of drinking water systems." Water Research, 43: 1641–1653, 2008.

[11] G.Z. Mao, J.W. Tu, H.B. Du. "Reliability evaluation based on fuzzy fault tree," IEEE International Conference on Industrial Engineering and Engineering Management (IE&EM 2010), Xiamen, China, 2010.

[12] J.B. Dugan, S.J. Bavuso, M.A. Boyd. "Dynamic fault-tree for fault-tolerant computer systems," IEEE Transactions on *Reliability*, vol.41, pp.363–376, 1992.

[13] J.B. Dugan, K.J. Sullivan, D. Coppit. "Developing a low cost high-quality software tool for dynamic fault-tree analysis" IEEE Transactions on *Reliability*, vol.49, pp.49–59, 2000.

[14] K.D. Rao, V. Gopika, V.V. Rao, ect. "Dynamic fault tree analysis using Monte Carlo simulation in probabilistic safety assessment." *Reliability Engineering & System Safety*, 92: 872–883, 2009.

[15] L. Leng, Y. Liu. "Fault tree reliability analysis for passive medium pressure safety injection system in nuclear power plant."*Energy and Power Engineering,* vol.5: 264–268, 2013.

[16] M. Nadjafi, M.A. Farsi, H.J. Khamnei. "Dynamic fault tree analysis using fuzzy L-U bounds failure distributions," *Journal of Intelligent & Fuzzy Systems,* vol. 33, pp. 3275–3286, 2017.

[17] D. Kececioglu. Reliability Engineering Handbook, Vol.2, 1991.

[18] J.B. Fussell, E.B. Henry, N.H. Marshall. "MOCUS—a computer program to obtain minimal sets from fault trees," ANCR-1156, Aerojet Nuclear Company, Idaho Falls, Idaho, 1974.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Research on kinematic reliability of flapping mechanism for flapping wing flight

Z. Yang & J. Xuan
School of Aeronautics, Northwestern Polytechnical University, Xi'an, Shaanxi, China

ABSTRACT: As the core component, the kinematic reliability of flapping mechanism influences the flight accuracy and efficiency of flapping wing flight directly. The joint clearance caused by wear is the most important factor affecting the kinematic accuracy of the mechanism, which directly affects the control accuracy and flight. This paper analyzes the kinematic reliability of flapping wing mechanism with different clearance of hinges based on the dynamics of mechanisms with clearance. The variation of the flutter angle, flutter angular velocity and angular acceleration of wingtip with different clearance are calculated. The Multiple joints clearance are considered in the flapping wing mechanism. Synchronous accuracy of two wings also be analyzed. Results show the relationships of the flutter angle, flutter angular velocity, angular acceleration and the joints clearance. The results can be used to optimize the flapping mechanism structure parameters.

1 INTRODUCTION

Flapping wing flight (MAV) is a kind of aircraft that imitates the flying creature of nature and relies on flapping wings to generate lift and thrust power simultaneously. It has the unique advantages of high flight efficiency, high maneuverability and compact structure. As the core component, the kinematic reliability of flapping mechanism influences the flight accuracy and efficiency of flapping wing flight directly, such as the kinematic accuracy and synchronism of flapping mechanism. The joint clearance caused by wear is the most important factor affecting the kinematic accuracy of the mechanism, which directly affects the control accuracy and flight. Therefore, the kinematic accuracy and synchronization of flapping wing mechanism with hinge clearance should be investigated.

Tsai (2004) presents an effective method to analyze the transmission performance of linkages. Equivalent kinematical pairs were used to model the motion freedom caused by the joint clearances. The mechanism positions are solved. Ting (2000) analyzes the influence of joint clearance on accuracy of position. Flores (2012, 2012, 2010, 2007) public a series of articles for the mechanism modeling with joint clearance, and also develop algorithm to solve.

Liu (2006, 2007) analyze the force-displacement relationship of spherical joints with clearances based on the Hertz theory.

This paper aims to analyze the kinematic reliability of flapping wing mechanism with different clearance of hinges based on the dynamics of mechanisms with clearance. The flutter angle, flutter angular velocity and angular acceleration of wingtip with different clearance are calculated. The Multiple joints clearance are considered in the flapping wing mechanism. Synchronous accuracy of two wings are also analyzed. The relationships between the flutter angle, flutter angular velocity, angular acceleration and the joint clearance are determined.

2 EVALUATION OF MECHANISM SYNCHRONIZATION

2.1 *Synchronization*

Assuming that: achieving some function needs n mechanisms to conduct the design motions synchronously; the allowable maximum time difference for the mechanisms to finish the motion is ε; t_i is the time that the mechanism i needs to finish the motion. Then the expression for the synchronization requirement of the mechanisms is shown as the following equation:

$$\max(t_i) - \min(t_j) < \varepsilon, (1 \le i, j \le n) \tag{1}$$

If the mechanisms that require motion synchronization are parallel mechanisms, which means the

mechanisms have same design motion and working conditions, then the motion synchronization of the mechanisms can be expressed in another way: the displacement difference of these mechanisms are supposed not to be greater than the prescribed tolerance at the same moment. Assuming that: achieving some function needs n mechanisms to conduct the design motions synchronously; the allowable maximum position difference at the same moment is ε; q_i is the position where the mechanism i is. Then the expression for the synchronization requirement of the mechanisms is shown as the following equation:

$$\max(q_i) - \min(q_j) < \varepsilon, (1 \le i, j \le n) \tag{2}$$

2.2 *The evaluation method of the mechanism synchronization based on the interval estimation*

The traditional probability-based evaluation method of mechanism synchronization can solve the synchronization problems with same probability distribution types and parameters. However, the probability distribution types and parameters are rarely the same in real engineering problems. Therefore, it is better to apply the evaluation method of the mechanism synchronization which is based on the interval estimation.

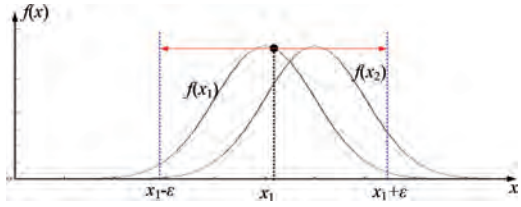


Figure 1. Interval schematic diagram of two mechanism synchronization.

Based on the analysis of two mechanisms, the motion synchronization of three mechanisms can be analyzed. If two mechanisms satisfy the synchronization requirement, then the motion time of the third mechanism must range between $\max(x_1,x_2) - \varepsilon$ and $\min(x_1,x_2) + \varepsilon$. If $x_1 > x_2$, the range of x_3 is $[x_1 - \varepsilon, x_2 + \varepsilon]$; if $x_1 < x_2$, the range of x_3 is $[x_2 - \varepsilon, x_1 + \varepsilon]$. The time interval of synchronization is shown in Fig. 2 and the synchronization probability can be expressed as the following formula:

$$R_{\text{syn}} = \int_0^{\infty} f_1(x_1)\left[\int_{x_1-\varepsilon}^{x_1} f_2(x_2)\left[\int_{x_1-\varepsilon}^{x_2+\varepsilon} f_3(x_3)\mathrm{d}x_3\right]\mathrm{d}x_2\right]\mathrm{d}x_1 + \int_0^{\infty} f_1(x_1)\left[\int_{x_1}^{x_1+\varepsilon} f_2(x_2)\left[\int_{x_2-\varepsilon}^{x_1+\varepsilon} f_3(x_3)\mathrm{d}x_3\right]\mathrm{d}x_2\right]\mathrm{d}x_1 \tag{4}$$

Accordingly, the formula of the synchronization probability of n mechanisms can be deduced:

$$R_{\text{syn}} = \int_0^{\infty} f_1(x_1)\left[\int_{x_1-\varepsilon}^{x_1+\varepsilon} f_2(x_2)\left[\int_{\max(x_1,x_2)-\varepsilon}^{\min(x_1,x_2)+\varepsilon} f_3(x_3)\ldots\int_{\max(x_1,x_2,\ldots,x_{n-1})-\varepsilon}^{\min(x_1,x_2,\ldots,x_{n-1})+\varepsilon} f_n(x_n)\mathrm{d}x_n\ldots\mathrm{d}x_3\right]\mathrm{d}x_2\right]\mathrm{d}x_1 \tag{5}$$

For a system with n mechanisms, the motion times of the mechanisms, which are expressed as x_i ($i = 1,\ldots,n$), are random variables. Besides, the probability density function of x_i is set as $f_i(x_i)$, $i = 1,\ldots,n$.

The allowable maximum time difference for the mechanisms to finish the motion is ε. For any two mechanisms in the system, this requirement can be transformed as a probability requirement: $P\{|x_i - x_j| \le \varepsilon\}$.

For the two mechanisms, once the motion time of one of the mechanism x_1 is determined, then the motion time of the second mechanism should range in $[x_1 - \varepsilon, x_1 + \varepsilon]$ to satisfy the synchronization requirement. As shown in Fig. 1, the probability for the two mechanisms' synchronization can be expressed as the following formula:

$$R = \int_0^{\infty} f_1(x_1)\left[\int_{x_1-\varepsilon}^{x_1+\varepsilon} f_2(x_2)\mathrm{d}x_2\right]\mathrm{d}x_1 \tag{3}$$

As can be seen in Formula (5), the synchronization probability of n mechanisms can be expressed by n-ple integrals. In the n-ple integrals, the upper and lower limits of the outermost layer variable x_1 is the range of the system's motion time, normally as $[0,\infty)$, while the upper and lower limits of every inner layer variable is determined according to the number size of the outer layer variables.

The inner layer integral cannot be solved directly, as its upper and lower limits is determined according to the number size of the outer layer variables. Therefore, the interval method is used to transform the non-deterministic upper and lower limits to deterministic upper and lower limits for the inner layer integral, then the Formula (5) can be solved.

As can be seen in the above analysis, when two mechanisms are considered, there is one independ-

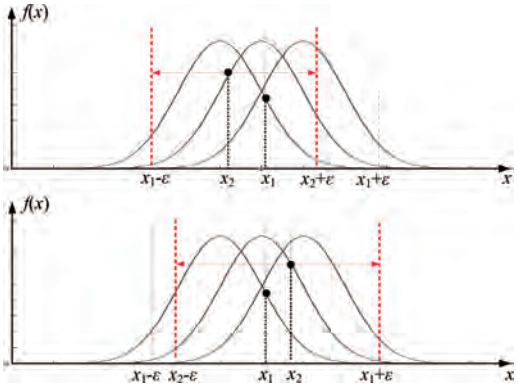


Figure 2. Interval schematic diagram of three mechanism synchronization.

Table 1. The relationship of mechanisms number n and independent integral interval number N.

Number of mechanism	2	3	4	5	…	n
x_{i-1} number of positon	1	2	3	4	…	n–1
Number of Independent interval	1	2	6	24	…	(n–1)!

ent integral interval; when three mechanisms are considered, there are two independent integral intervals; when four mechanisms are considered, there are six independent integral intervals; when n mechanisms are considered, the integral interval of every variable is determined by the number size of all variable before it. Therefore, for a system with n mechanisms, there are $N = (n - 1)!$ independent integral intervals. The relationship of mechanisms number n and independent integral interval number N is listed in Table 1.

From the above analysis, it can be seen that the synchronization probability of n mechanisms can be expressed as the sum of $(n - 1)!$ n-ple integrals. As the $(n - 1)!$ integrals are independent with each other, the synchronization probability of n mechanisms can be deduced:

$$R_n = \sum_{i=1}^{(n-1)!} R_{n,i} \tag{6}$$

For most of the probability distributions, though their probability density functions are elementary functions, their integrals are not elementary functions. Therefore, numerical methods are needed to calculate the synchronization probability of mechanisms, which mainly include numerical integration methods and Monte Carlo Method.

1. Solving the n-ple integrals by Numerical Integration Method

If the n-ple integrals cannot be expressed by elementary functions, then the Newton-Leibniz Equation is not able to solve the integrals. Therefore, the integral equation is transformed to the limit equation:

$$I = \int_a^b f(x)\mathrm{d}x = \lim_{h\to 0} f(\xi_i)h_i \tag{7}$$

In Formula (7), $h_i = x_i - x_{i-1}, x_{i-1} \le \xi_i \le x_i, a \le x_0 < x_1 < \ldots < x_n \le b$. Then an approximation method can be used as shown in the formula:

$$I = \int_a^b f(x)\mathrm{d}x \approx \sum_{i=0}^{n} A_i f(x_i) = I_n \tag{8}$$

In the above formula, $\{x_i\}_{i=0}^n$ are integral nodes, and $\{A_i\}_{i=0}^n$ are the multiplier parameters.

The frequently-used numerical integral methods are: interpolated quadrature formula, complex quadrature formula, Romberg quadrature formula and Gauss quadrature formula.

The n-ple integral can be solved by repeatedly solving the single integral using the above methods.

2. Solving the n-ple integrals by Monte Carlo Method

Assuming that D is a region in the n-dimensional space R_n, $f(x) \in D \subset R_n \to R$, the n-ple integrals on the region D can be expressed as:

$$I = \int_D f(p)\mathrm{d}p \tag{9}$$

where, I can be regarded as the result expectation of the region's measure multiplying function f on the region D. The fundamental Monte Carlo Method is to find a hypercube (with known measure M_c) which contains region D, generate n sample points which obey the uniform distribution in the hypercube, and counting the number of sample points in the region D. Assuming that there is m out n sample points in the regions D, then the measure of region D is approximated as:

$$M_D \approx \frac{mM_c}{n} \tag{10}$$

The expectation of the function f:

$$\bar{f} \approx \frac{1}{m}\sum_{i\in D} f(p_i) \tag{11}$$

2.3 *The analysis of mechanism synchronization*

For the system with n mechanisms, the motion time of every mechanism is random variable and expressed as x_i, $i = 1, \ldots, n$ respectively. The total load

of the system is random variable is also a random variable and expressed as z. As the motion time of mechanism i is related to several influence factors, such as mechanism parameters, working loads and environment conditions, assume that the function of the motion time and the influence factors is:

$$x_i = f(\mathrm{y}_{i1}, \mathrm{y}_{i2}, \cdots \mathrm{y}_{im}, \alpha_i \cdot \mathrm{z}) \tag{12}$$

In the function, $\mathrm{y}_{i1}, \mathrm{y}_{i2}, \cdots \mathrm{y}_{im}$ are m random variables that influence the motion time of the mechanism i, α_i is the load weight of the load on mechanism i to the total load and it is related to the mechanism parameters:

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n = 1 \tag{13}$$

The synchronization requirement of two mechanisms can be expressed as:

$$\max_{\substack{0<i,j<n \\ i \neq j}} (|\mathrm{x}_i - \mathrm{x}_j|) \leq \varepsilon \tag{14}$$

As for the synchronization evaluation of two mechanisms, Coupla function $\mathrm{C}_\theta(u,v)$ can be used, in which θ is the correlation parameter of two mechanisms. The Coupla function can be obtained by sample fitting, and the synchronization probability can be expressed as:

$$R = \int_0^\infty f_1(x_1) \left[\int_{x_1-\varepsilon}^{x_1+\varepsilon} \frac{\partial^2 C_\theta(u,v)}{\partial u \partial v} f_2(x_2) \mathrm{d}x_2 \right] \mathrm{d}x_1 \tag{15}$$

As for the synchronization evaluation of three mechanisms or more, firstly use the sample fitting to obtain the Coupla function, then obtain the joint probability density function $f(x_1, x_2, \ldots x_n)$ via the edge probability density function $f(x_i)$, then the synchronization probability of multi mechanisms can be figured out combined with interval evaluation method.

For the system in which the mechanisms are not independent, as the motion times are correlated with each other and the correlation can hardly be illustrated in mathematical forms, so it is very difficult to obtain the synchronization probability using only mathematical methods. For these non-independent cases, the method combined dynamic simulation and sampling calculation is widely used to obtain the synchronization probability of the mechanisms in the system.

3 SYNCHRONISM ANALYSIS OF FLAPPING MECHANISM

3.1 *Flapping mechanism*

The structure of flapping wing is shown as Fig. 3. The principle of flapping mechanism is shown as

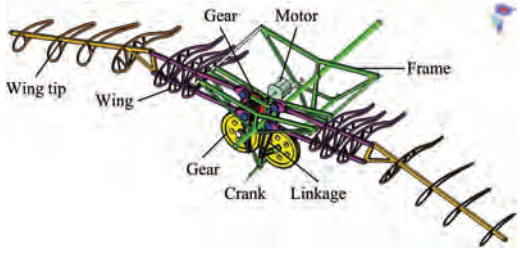


Figure 3. The structure of flapping wing.

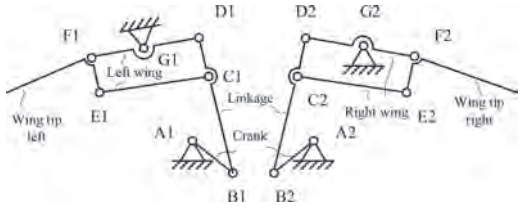


Figure 4. The principle of flapping mechanism.

Table 2. Dimension of revolute joint with clearance.

Revolute joint	Diameter of shaft	Diameter of sleeve
B_1	3	2.6
C_1	12	11.2
D_1	12	11.2
E_1	5.5	5

Fig. 4. The crank drive mechanism to maintain uniform rotation, and drive the wing tip under flutter through the connecting rod.

3.2 *Dynamic simulation of flapping wing mechanism*

LMS Virtual Lab is used to construct dynamic model. For the kinematic pair with joint clearance, combing spherical contact with plane pair to substitute the rotation pair, dimension of revolute joint with clearance are shown as Table 2. The variation of wingtip angle, angular velocity and angular acceleration with time can be calculated. The results are shown as Fig. 5 to Fig. 9.

We can conclude from those figures, the variations of wingtip angle with joint clearance are obvious, especially at the limit position; the variations of wingtip angle velocity with joint clearance are bigger, fluctuate around ideal position; the variations of wingtip angle velocity with joint clearance are extreme, the wingtip is under extreme shock load.

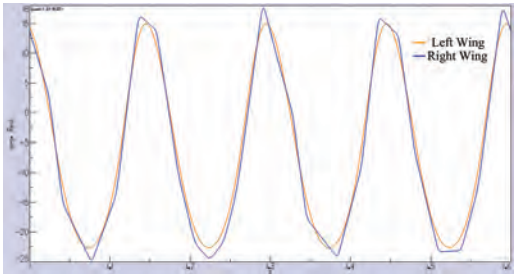


Figure 5. The variation of wingtip angle with time.

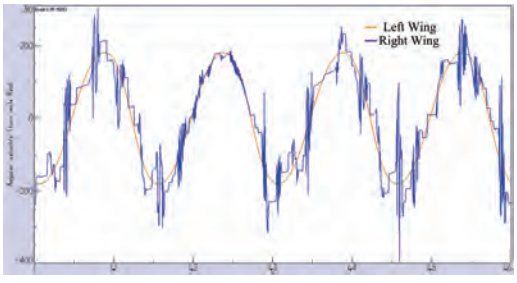


Figure 6. The variation of wingtip angular velocity with time.

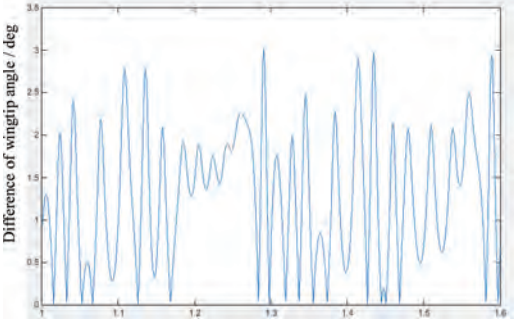


Figure 7. The variation of wingtip angular acceleration with time.

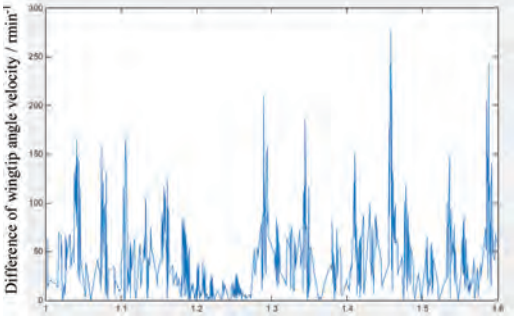


Figure 8. The velocity difference of left and right wing tip with time.

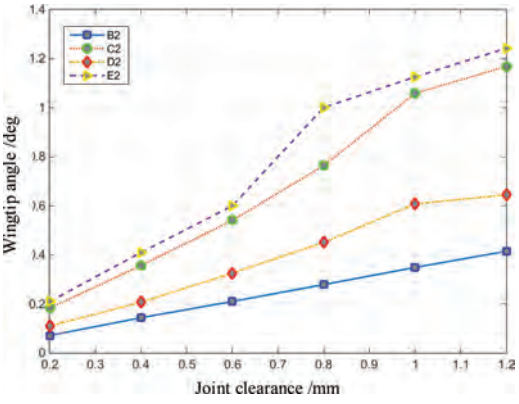


Figure 9. The difference of left and right wingtip angle with time and clearance position.

3.3 *Influence of clearance on motion synchronism*

To analyze the variation of wingtip angle with joint clearance, the joints clearance of B2, C2, D2, E2 are setup from 0.2 mm to 1.2 mm, the difference between left and right wingtips are calculate. The results are shown as Fig. 9.

From the Fig. 9, as the clearance increased, the difference of left and right wingtip is increased. The influence of value of revolute clearance E2 is biggest, and the influence of value of revolute clearance B2 is smallest.

4 CONCLUSIONS

This paper aims to analyze the kinematic reliability of flapping wing mechanism with different joint clearance based on the dynamics of mechanisms with clearance. The results can be used to optimize the flapping mechanism structure parameters. The conclusions as follows.

- The variations of wingtip angle with joint clearance are obviously, especially at the limit position; the variations of wingtip angle velocity with joint clearance are large, fluctuate around ideal position; the variations of wingtip angle velocity with joint clearance are extremely large, the wingtip is under extremely shock load.
- As the clearance increased, the difference of left and right wingtip is increased. The influence of value of revolute clearance E2 is largest, and the influence of value of revolute clearance B2 is smallest.

ACKNOWLEDGMENT

The authors gratefully wish to acknowledge the supported by National Natural Science Founda-

tion of China under grant No. 51505383, China Postdoctoral Science Foundation under grant No. 2017T100771.

REFERENCES

Flores P, Ambrsio J, Claro J.C.P, Lankarani H.M. (2007) Dynamic behavior of plan arrigid multi-body systems including revolute joints with clearance. Proceedings of the institution of mechanical engineers—Part K: Journal of multi-body dynamics. 221(2), 161–174.

Flores P, Lankarani H.M. (2010) Spatial rigid-multibody systems with lubricated spherical clearance joints: modeling and simulation. Nonlinear Dynamics. 60, 99–114.

Flores P, Lankarani H.M. (2012) Dynamic response of multibody systems with multiple clearance joint. Nonlinear Dynamics. 7, 031003.

Gilardi G. Sharf I. (2002) Literature survey of contact dynamics modeling. Mechanism and Machine Theory. 37, 1213–1239.

Liu C.S, Zhang K, Yang L. (2006) Normal Force-Displacement Relationship of Spherical Joints with Clearances. Transactions of the ASME: Journal of Computational and Nonlinear Dynamics. 1, 160–167.

Liu C.S, Zhang K, Yang R. (2007) The FEM analysis and approximate model for cylindrical joints with clearances. Mechanism and Machine Theory. 42, 183–197.

Machado M, Moreira P, Flores P, Lankarani H.M. (2012) Compliant contact force models in multibody dynamics: evolution of the hertz contact theory. Mechanism and machine theory. 53, 99–121.

Ting K.L, Zhu J.M, Watkins D. (2000) The effects of joint clearance on position and orientation deviation of linkages and manipulators. Mechanism and machine theory. 35(3), 391–401.

Tsai, M.J, Lai, T.H. (2004) Kinematic sensitivity analysis of linkage with joint clearance based on transmission quality. Mechanism and machine theory. 39, 1189–1206.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Subset simulation and global minimization: Any problems?

K. Breitung
Engineering Risk Analysis Group, Technical University of Munich, Munich, Germany

ABSTRACT: In the last fifteen years the subset sampling method has often been used in reliability problems as a tool for calculating small probabilities. This method is extrapolating from an initial Monte Carlo estimate for the probability content ofa failure domain found by a suitable higher level of the original limit state function. Then iteratively conditional probabilities are estimated for failures domains decreasing to the original failure domain. Here is explained why this concept is very problematic.

1 INTRODUCTION

A basic problem of structural reliability is the calculation of failure probabilities given by *n*-dimensional integrals in the following form

$$P = \int_{g(x)<0} f(x)\, dx \tag{1}$$

Here $f(x)$ is an n-dimensional PDF (probability density function) of a random vector X and $g(x)$ is the LSF (limit state function) giving the failure condition. During the development of structural reliability methods it was found to be favorable to transform the random vector X into a standard normal random vector U with independents components. So the problem was then in this standardized form:

$$P = (2\pi)^{-n/2} \int_{g(u)<0} \exp\left(-\frac{|u|^2}{2}\right) du \tag{2}$$

Such transformations into the standard normal space for random vectors with independent components have been described first by Rackwitz & Fiessler (1977). For random vectors with dependent components the Rosenblatt-transformation is proposed to obtain a transformation but with the exception of the example given in Hohenbichler & Rackwitz (1981) no applications of this transformation concept are known to the author. A practically applicable method seems to be the Nataf-transformation described in Der Kiureghian & Liu (1986).

A newer method for the calculation of failure probabilities is the subset simulation concept. It will be outlined here that there is an intrinsic flaw in it.

2 GLOBAL AND LOCAL EXTREMA

Shortly a few basic facts about local and global extrema are outlined here before proceeding. Let be given a function $f : D \to \mathrm{R}$. A local minimum (maximum) is a point $x_0 \in D$ such that there exists a neighborhood u of x_0 that for all $x \in U \cap D$ one has always

$$f(x) \geq f(x_0) \left(\text{resp.}\ f(x) \leq f(x_0)\right) \tag{3}$$

In a similar way a global minimum (maximum) for this function is defined as a point $x_1 \in D$ such that for all $x \in D$ always

$$f(x) \geq f(x_0) \left(\text{resp.}\ f(x) \leq f(x_0)\right) \tag{4}$$

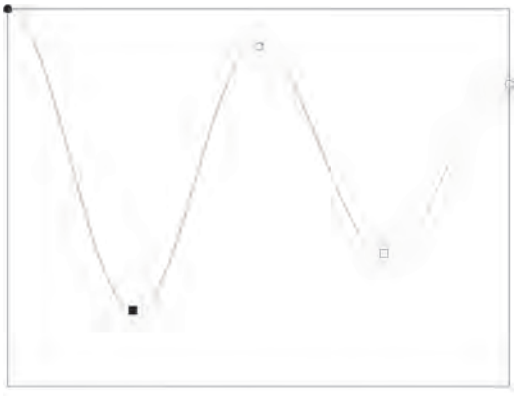
The first is a local property of the function, it depends only on its behavior in any arbitrary small neighborhood of the point. The second is global, one has to know the values of the function over the whole domain of definition. Therefore methods to find local extrema are plentiful and described in all textbooks about numerical analysis, whereas in these textbooks the problem of finding global extrema is covered only in a marginal way (see for e.g. Nocedal & Wright (1999)).

In handling optimization problems it is important to be aware if one searches a global or a local extremum. Mostly a global should be found. But this is often done by using methods for determining local extrema and then in some way a conclusion is made that the obtained local extremum is also a global.

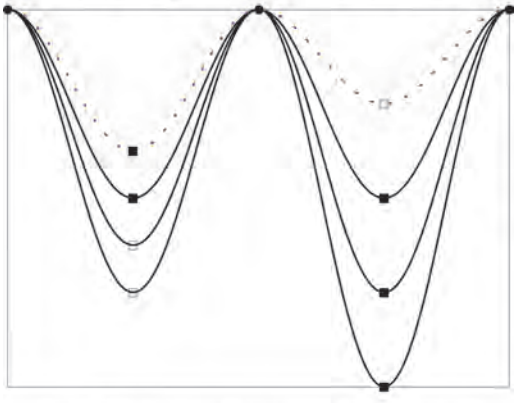
The most usual naïve way for this is to start a local minimum search several times from random starting points and record all obtained local minimum. The

smallest of these can then be considered a global one with some confidence, if enough runs have been made.

In finding global extrema a common misunderstanding is that if one has a function depending on a parameter in a continuous way that then with a continuous change of it also the positions of the global extrema move continuously. But in fact these points jump also in non-pathological cases. In Figure 1(a) for a simple one-dimensional example the difference between local and global extrema is shown. In Figure 1(b) the global minima for a function depending on a parameter are shown, the lower ones by solid curves and the upmost function is shown by a dotted curve. For the three lower functions global minima are on the left side, forthe



(a) Local and global extrema of a function



(b) The global minima of a function depending on a parameter

Figure 1. Global and local extrema of functions: minima are shown by squares, maxima by circles, filled symbols are global extrema.

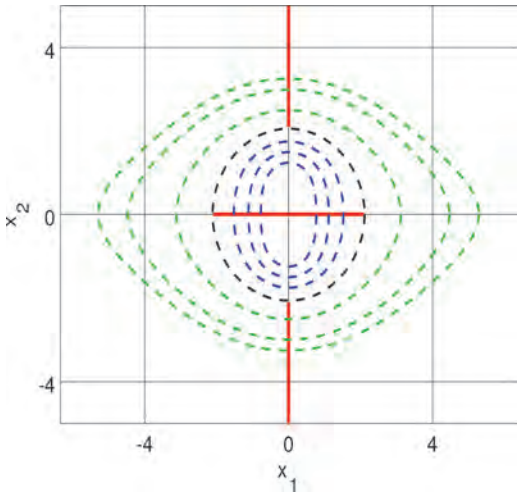


Figure 2. Constrained minimal distance points for the function given in equation (5).

two upper global minima are on the right side of the figure. The second function from above has two minima. The upmost function, denoted by a dotted curve, has its global minimum on the left side of the figure. This shows that the position of a global minimum of a function depending on a parameter does not necessarily vary continuously if the parameter is varied continuously.

Also in global optimization under constraints the minimum points tend to jump if the problem is non-trivial. This will be shown in the following simple example. Let be given a LSF defined by

$$g(u_1,u_2) = \beta - u_1^2 - \frac{u_1^2+u_2^2}{b^2}u_2^2 = \beta - u_1^2 - \frac{\left(u_1^2u_2^2+u_2^4\right)}{b^2} \tag{5}$$

In Figure 2 the positions of the global minima of $|u|$ under the constraint $g(u)=c$ are on the red line segments. Reaching the black circle they jump. In Zhigljavsky & žilinskas (2008) the problem of distinguishing between a local and a global maximum is explained in detail in the first chapter. This is one of the main problems in global optimization. Failing to understand this and accepting local extrema asglobal ones might lead to erroneous results.

3 ASYMPTOTIC ANALYSIS AND FAILURE PROBABILITIES

With concepts of asymptotic analysis one can get asymptotic approximation for failure probabilities. Here only an extremly shortened version is given,

for more details see Breitung (1994). This concept considers the failure probability as a function of the distance of the failure probability to the origin β, defined by

$$\beta = \min_{g(u)\le 0} |u| \tag{6}$$

For smooth LSF's β is the distance of one or more points on the limit state surface $\{g(u)=0\}$ to the origin. If the failure domain F is imbedded into a family of expanding surfaces one has asymptotically

$$\mathrm{P}(F) \approx \Phi(-\beta) \prod_{i=1}^{n-1} (1-\beta\kappa_i)^{-1/2} \tag{7}$$

with the κ_i the main curvatures of the limit state surface at the point u_0, called design point. In the case of several such points their contributions will be added. So looking at the example in Figure 3 the curvatures at the two points are needed for the asymptotic approximation. Does one need the curvatures? Not necessarily. The lemma of Hohenbichler (Breitung (1994), p. 53) says that an approximation can be obtained also if one can calculate the probability content of the neighborhoods of the design points. So one has Figure 4

$$\mathrm{P}(F) \approx \mathrm{P}(A_1) + \mathrm{P}(A_2) \tag{8}$$

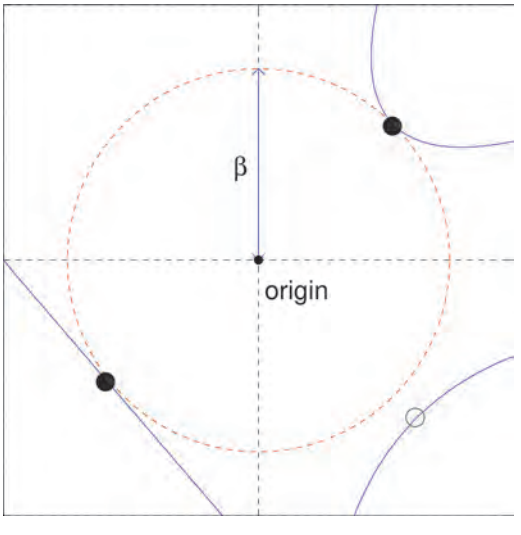


Figure 3. Asymptotic approximation with curvatures.

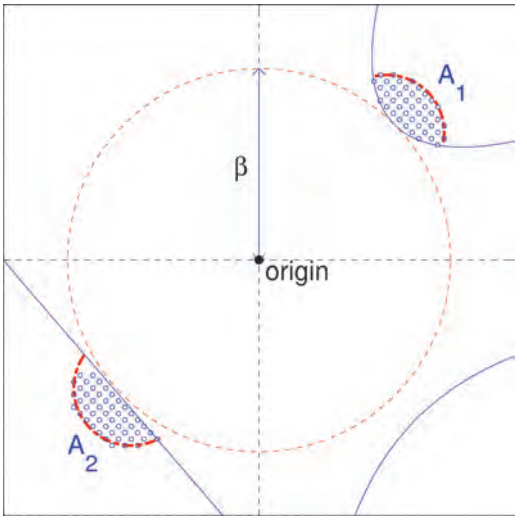


Figure 4. Asymptotic approximation calculating the probability content of the neighborhoods.

4 THE SUBSET SIMULATION CONCEPT

The subset simulation algorithm is a variant of Monte Carlo methods; it tries to avoid the large number of data points which are needed in standard Monte Carlo by using instead of it an iterative procedure. It can be subsumed under the generic term of stochastic optimization procedures.

While importance sampling methods try to improve the efficiency of Monte Carlo by identifying regions with high probability content and moving more date points there, SuS starts from an enlarged failure domain whose design points are much nearer to the origin and then moves step by step towards the original failure domain. These regions are defined here by domains in the form $F_i = \{g(u) < a_i\}$ with the a_i's being positive and $a_i \to 0$. The basic thought of the method (see (Au & Beck 2001) and (Au & Wang 2014)) is now to write the failure probability $P(F)$ as a product of conditional probabilities

$$\begin{aligned} \mathrm{P}(F_n) &= \mathrm{P}(F_1 \mid F_0)\mathrm{P}(F_2 \mid F_1)\ldots\mathrm{P}(F_n \mid F_{n-1}) \\ &= \prod_{k=0}^{n-1} \mathrm{P}(F_{k+1} \mid F_k) \end{aligned} \tag{9}$$

Here $\mathbb{R}^n = F_0 \supset F_1 \supset F_2 \supset \ldots \supset F_n = F$. In Figure 5 such a standard case is shown, the design points for the various LSF's are shown by black squares. Since the respective (suitably chosen) conditional probabilities are relatively large compared with the failure probability P(F) which has to be estimated, such an access to the problem has the advantage that these conditional probabilities can

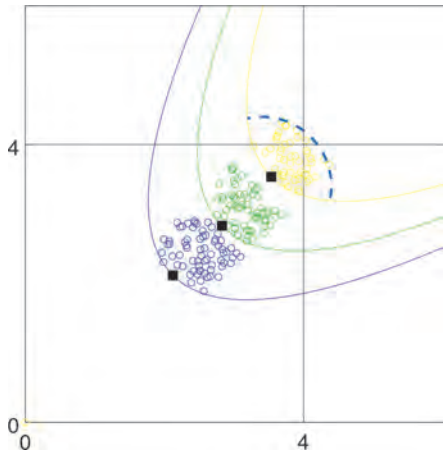


Figure 5. Typical SuS example.

be estimated more efficiently with much smaller sample sizes. The details how these samples are produced with Monte Carlo Markov chains can be found in the references given above. SuS can be seen as a stochastic continuation method (this concept is explained in Allgower & Georg (2003)). Such methods try to extrapolate from a point x which solves an equation system $F(x, \tau)$ for a given values of the parameter τ to find solutions for other values of τ.

5 SUBSET SIMULATION AND STOCHASTIC MINIMIZATION

It seems not to be quite clear how to classify this SuS in mathematics, it uses MCMC as a tool, but MCMC is not the objective. The goal of the approach is to find an estimator $\hat{p}$ for the unknown probability $p = \mathrm{P}(F_n)$. Now in a Bayesian setting estimation procedures are mostly based on using a loss function and then determining the estimator as a minimizer ofthe expected value of the loss function. This is outlined in Box & Tiao (1973), appendix A5.6, p. 308. For a loss function $L(.)$ then the estimator $\hat{\theta}$ for a parameter θ is chosen such that

$$IE\Big(L\big(\hat{\theta} - \theta\big)\Big) \to \min \tag{10}$$

Therefore a parameter estimation always is transformed into a minimization problem. Since there is a stochastic component, it is a stochastic minimization and since one wants the best estimator, it is a global stochastic minimization (see e.g. Spall (2004)).

In SuS the chosen loss function is the squared coefficient of variation for the estimator $\widehat{\mathrm{P}(F_n)} = \hat{p}$

$$IE\left[\left(\frac{\hat{p} - p}{p} \right)^2 \right] \to \min \tag{11}$$

Under the assumption of unbiasedness and independence of the estimators $\hat{p}_i = \widehat{\mathrm{P}(F_i \mid F_{i-1})}$ of this can be approximated by

$$IE\left[\left(\frac{\hat{p} - p}{p} \right)^2 \right] \approx \sum_i IE\left[\left(\frac{\widehat{p_i} - p_i}{p_i} \right)^2 \right] \tag{12}$$

Then the approximative optimal estimator is given by choosing for the $\hat{p}_i$'s the value of the empirical conditional distribution function at the points c_i, since these are unbiased minimal variance estimators for the p_i's (see e.g. Lehmann & Casella (1998)). This shows that the estimation in SuS is a minimization procedure. And the goal is clearly to find an estimator which achieves the global minimum value for the estimation error in equation (12).

6 EXAMPLES

All the examples with the exception of the last one were calculated with the SuS algorithm given in Li & Cao (2016). For the last example the algorithm given in Uribe (2016) was used, since it records the seeds. As parameters were taken 500 samples per step, an acceptance probability of 0.1 and a chain length of ten. This setup was used in Au & Wang (2014) for estimates of the context of two-dimensional domains. The design points are marked by red circles and the SuS data points by green points.

The focus here is on the geometric structure of the failure domains and their design points, on purpose the different probability estimates which can be computed are excluded. Only two-dimensional cases are considered, since only for them it is possible to produce informative diagrams showing the behavior of the algorithm.

The claim of SuS proponents is that the method is much more general than FORM/SORM and makes no use of design point concepts. But this should mean that for such simple two-dimensional cases, where it is essential to find the design points for getting useful estimates of the reliability index and/or the failure probability, also the SuS approach should give *a fortiori* good results.

In the examples cases are shown where the data points of the SuS algorithm move not to the design point but to points further away from the origin

than the design point leading to an overestimation of the reliability index and an underestimation of the failure probability. In Breitung (2016) and Breitung (2017) similar examples were studied.

6.1 *Basic example*

A simple example will show the misunderstanding in the SuS algorithm about global and local minimization. Let be given two LSF's g_1 and g_2 and together with them define a third g as minimum of both:

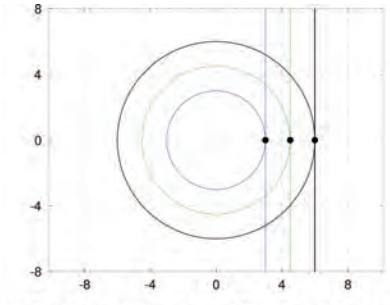
$$\begin{aligned} g_1(u_1,u_2) &= 6-\frac{u_1}{6} \\ g_2(u_1,u_2) &= 5+u_1, \\ g(u_1,u_2) &= \min(g_1,g_2) \end{aligned} \tag{13}$$

Consider now two reliability problems. In the first the failure domain is given by

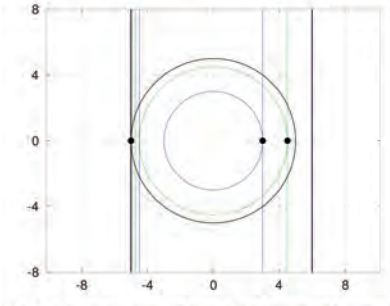
$$F_1 = \{g_1(u_1,u_2) < 0\} \tag{14}$$

In the second the failure domain F is given by

$$F = \{g(u_1,u_2) < 0\} \tag{15}$$

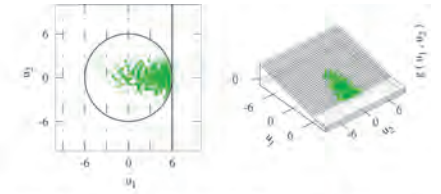


(a) The design points (black filled circles) for the LSF's $g_1(u_1,u_2) = 0.5, 0.25, 0$

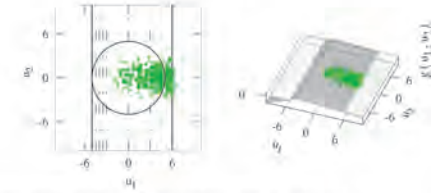


(b) The design points (black filled circles) for the LSF's $g(u_1,u_2) = 0.5, 0.25, 0$

Figure 6. The design points for the LSF's g_1 and g.



(a) SuS algorithm for the LSF g_1, a) solid line contour $g_1 = 0$, dash-dotted lines further contours for $g_1 > 0$, b) graph of g_1; green dots show SuS data points



(b) SuS algorithm for the LSF g, a) contour line $g = 0$, b) graph of g; green dots show SuS data points

Figure 7. SuS for the LSF's g_1 and g.

In both cases it is a series system which fails if at least one of the two functions is less than zero. The LSF's are shown in Figure 7. The SuS algorithm works correctly in the first case, see Figure 7a, but fails for the second, see Figure 7b.

6.2 *Invariance*

An important reason that the Hasofer-Lind index was adopted as a measure for reliability is its invariance under reformulations or reparametrizations of the underlying reliability problem (see Hasofer & Lind (1974)). Also the convergence proofs for the beta point search algorithms do not depend on the specific form of the LSF. But clearly one has to start the search algorithm from different points to obtain all global minimal distance points, i.e. design points. Consider now a series system consisting of two independent components, so failure occurs if at least one fails. The first component fails if $u_1 > 5$ and the second component if $u_2 < -4$. Now, this limit state surface can be the zero set of different LSF's. For example, one has

$$g(u_1,u_2) = \min\begin{cases} 5-u_1 \\ 4+u_2 \end{cases} \tag{16}$$

Here both LSF's are linear, with SuS one obtains as expected an estimate for the asymptotic failure probability approximation $\widehat{P(F)} \approx \Phi(-4) \approx 3.17 \cdot 10^{-5}$. The performance of the SuS method is shown in Figure 8.

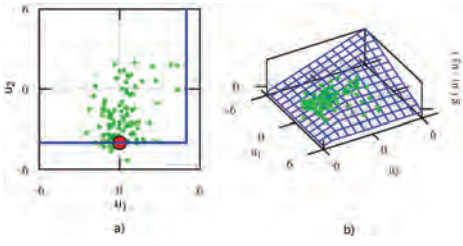


Figure 8. Series system defined by LSF in equation (16).

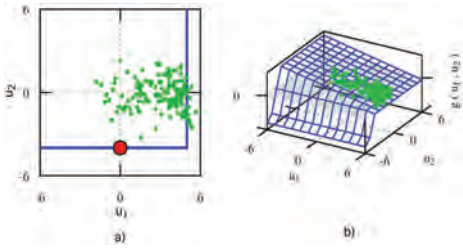


Figure 9. Series system defined by LSF in equation (18).

Assume now that the LSF for the second random variable is given not by a linear but by a logistic function in the form:

$$g_2^*(u_2) = \frac{1}{1+exp\,(-2(u_2+4))} - 0.5 \tag{17}$$

Then the LSF $g^*(u_1,u_2)$ given by

$$g^*(u_1,u_2) = \min\begin{cases} 5-u_1 \\ \dfrac{1}{1+exp\,(-2(u_2+4))} - 0.5 \end{cases} \tag{18}$$

defines the same limit surface as before, but the shape of the LSF is different and the contour lines of these functions are different in the safe and unsafe domain, both LSF's have only the contour of zero level set in common. Here, with the LSF defined in equation 18, the points in SuS converge towards the point (5,0) and one gets as probability estimate a value of $\Phi(-5) \approx 2.87 \cdot 10^{-7}$ whereas the true failure probability is approximately equal to $\Phi(-4) \approx 3.17 \cdot 10^{-5}$ as shown in Figure 9. So, here the different forms of the LSF's influence the result of the method. The reason is that the structure of the LSF in the neighborhood of the origin is different from its form near the limit state surface.

The same limit state surface can be described by a plethora of different LSF's. Their specific forms will influence the behavior of the SuS algorithm. Especially for more complicated LSF's for series or parallel systems it might be useful to clear inasmuch this can create convergence problems or lead to incorrect results. Certainly there will be cases where the result will not depend on the changing structures of the LSF's, but as the example above shows, it would be overoptimistic to assume that this is generally so.

The same limit state surface can be described by a plethora of different LSF's. Their specific forms will influence the behavior of the SuS algorithm. Especially for more complicated LSF's for series or parallel systems it might be useful to clear inasmuch this can create convergence problems or lead to incorrect results. Certainly there will be cases where the result will not depend on the changing structures of the LSF's, but as the example above shows, it would be overoptimistic to assume that this is generally so.

6.3 *Several design points*

Let the LSF be

$$g_\beta(u_1,u_2) = \frac{\beta^2}{2} - |u_1 \cdot u_2|. \tag{19}$$

Due to the symmetry of the LSF there are four beta points. In a FORM/SORM analysis one obtains using the results found in the following asymptotic approximation one has for the failure probability

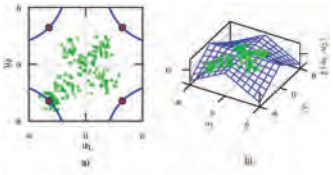
$$P(g_\beta(u_1,u_2) < 0) \sim 2\sqrt{2} \cdot \Phi(-\beta), \beta \to \infty. \tag{20}$$

Here clearly in a SORM analysis the beta point search algorithms have to be started several times to find all beta points.

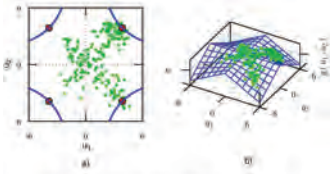
If this problem is examined now with SuS the possible outcomes of runs are shown in Figure 10. In fifty runs of SuS in one case only one beta point was detected, in 11 two, in 29 three and only in nine cases all four were found. This might lead to a systematic underestimation of the failure probability when not all beta points are found. If now several runs are combined, there will still be a bias, the failure probability will be underestimated. It is unclear to the author how to get a good estimator of the failure probability here without making some sort of geometric analysis similar to FORM/SORM. Also in the FORM/SORM approach the search algorithm for design points can end in local minimum distance points. But this is well known and by running the search from many different starting points one can expect to find all relevant design points (see e.g. Ditlevsen & Madsen (1996), section 5.2 and Melchers & Beck (2018), section 4.3.7).

6.4 *Stationarity of seeds*

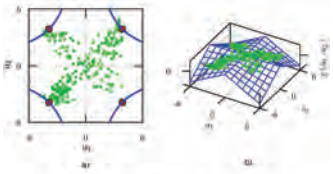
A recurring claim in the SuS literature is that the seeds for the next step which lie in the failure domain F_i already have a stationary distribution over this



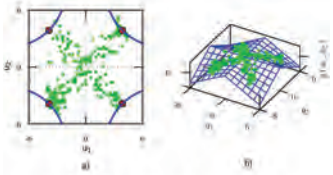
(a) SuS detects one design point



(b) SuS detects two design points



(c) SuS detects three design points



(d) SuS detects four design points

Figure 10. SuS for the LSF $g(u_1,u_2)=15-|u_1\cdot u_2|$.

domain, i.e. perfect sampling. This is true only for the first step, in all further steps the seeds do not have this stationary distribution. This was discussed in Botev & Kroese (2012), section 7 and in remark 4.4.3 in Botev (2009), but it seems that these findings remained unnoticed. The claim is true only for the first step in the algorithm where the seeds are coming from a Monte Carlo sample. This can be shown using theorem 2.4.1, p. 24 in Arnold et al. (2008). As explained in Geyer (2011) for proving that a MCMC produces exactly stationary data points, the random mechanism creating those must be studied.

Here a simple argument is given that the seeds in higher steps do not have this stationary distribution. Consider the simple case that F_i is a semi infinite interval $[b_i,\infty)$. An explanation that this cannot be correct is the fact that in one Markov chain generated with the MCMC algorithm proposed there, the next point can be the same point as before if the point found by the iteration algorithm is rejected. So the probability that the smallest seed on the interval F_i is equal to the left limit point b_i of the interval is larger than zero. But for a stationary distribution it is zero.

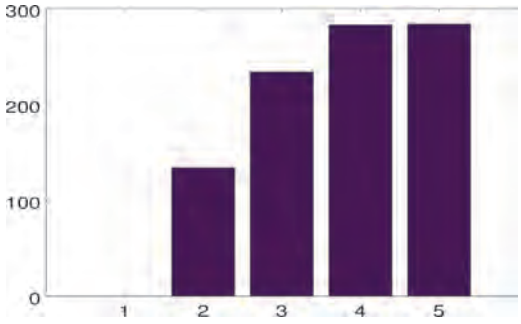


Figure 11. Occurence of seeds with distance zero from the lower end point.

To demonstrate this for the problem with $g(u)=5-u$ one thousand runs SuS runs were made with 500 points per step and an acceptance probability of 0.1 and 10 samples per chain. In the following Figure 11 it is recorded how often thesmallest seed had distance zero from the next lower order statistic. For the first step this never happens, but from the second step it happens quite often. This contradicts the claim that the seeds have a stationary distribution over $F_i=[c_i,\infty)$ for $i>1$. Important here is to note that if the distributions were stationary, a double would occur with probability zero, so even one such event happening during a numerical simulation makes the claim of stationarity invalid. The fact that for different SuS implementations and points in steps the number might be lower does not change the conclusions; since it is a qualitative not a quantitative statement. If at least one such point appears in a numerical simulation the claim of stationarity is disproved.

7 PROBLEM STRUCTURE AND SUS

Another important point which is problematic in SuS is the ignoring of the problem structure. This is advertised as a special feature. For example in Zuev et al. (2012) it is written:

> *Subset Simulation provides an efficient stochastic simulation algorithm for computing failure probabilities for general reliability problems without using any specific information about the dynamic system other than an inputoutput model. This independence of a systems inherent properties makes Subset Simulation potentially useful for applications in different areas of science and engineering where the notion of "failure" has its own specific meaning,…*

Contrasting this opinion there is the following quote from Monahan (2011) p. 394:

> *For MCMC, an extremely naive user can generate a lot of output without even under- standing the problem. The lack of discipline of learning about the problem that other methods require can lead to unfounded optimism and confidence in the results.*

8 CONCLUSIONS

The subset simulation approach to structural reliability analysis ignores the problem how to distinguish between local and global constrained mimimum distance points on the limit state surface. It finds minimum points without checking if there are other minimum points which give a smaller value. Certainly this analysis is restricted to cases where design points exist, but also for these cases SuS should give correct results. This problem of local minima is known in FORM/SORM methods and is avoided by running design point searches from different starting points. Therefore there is a considerable danger of obtaining erroneous results, since SuS ignores the possibility of finding only local minima instead of global. This misunderstanding of the basic problem can be seen in all papers about SuS as far as the author knows, it is always assumed that there is no need to ascertain that the found solution is not based on a local minimum. But this is the essential problem of global minimization to avoid local minima and to escape from them towards global minima.

It is obvious that similar problems will appear for examples in higher dimensions. The simple examples here were chosen, because they allow an illustration of the behavior of the method in a graphical way, which is not possible in higher dimensions. Maybe each of these primitive examples can be solved by adding to SuS some specific adhockeries for the example, but in this way the principal problem of avoiding local extrema is not addressed. And this has to be resolved for SuS to be accepted as a meaningful reliability method furthermore.

REFERENCES

Allgower, E. L. & K. Georg (2003). *Introduction to Numerical Continuation Methods*. SIAM. SIAM Classics in Applied Mathematics 45.

Arnold, B. C., N. Balakrishnan, & H. N. Nagaraja (2008). *A First Course in Order Statistics*. Philadelphia, PA, USA: SIAM.

Au, S. K. & J. L. Beck (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics 16*, 263–277.

Au, S.-K. & Y.Wang (2014). *Engineering Risk Assessment with Subset Simulation*. New York: JohnWiley & Sons, Ltd.

Botev, Z. (2009). Splitting methods for efficient combinatorial counting and rare-event probability estimation. http://espace.library.uq.edu.au/view/UQ:178513, School of Mathematics and Physics, The University of Queensland. Technical report.

Botev, Z. & D. Kroese (2012). Efficient Monte Carlo simulation via the generalized splitting method. *Stat Comput 22*, 1–16.

Box, G. & G. Tiao (1973). *Bayesian Inference in statistical analysis*. Reading, Mass.: Addisson-Wesley.

Breitung, K. (1994). *Asymptotic Approximations for Probability Integrals*. Berlin: Springer. Lecture Notes in Mathematics, Nr. 1592.

Breitung, K. (2016). Extrapolation, Invariance, Geometry and Subset Sampling. In R. Caspeele et al. (Eds.), *14th International Probabilistic Workshop*, Cham, CH, pp. 33–44. Springer Nature.

Breitung, K. (2017). The Geometry of Limit State Function Graphs and Subset Simulation . https://arxiv.org/pdf/1705.04453.pdf.

Der Kiureghian, A. & P. Liu (1986). Structural Reliability under Incomplete Probability Information. *Journal of Engineering Mechanics 112*(1), 85–104.

Ditlevsen, O. & H. Madsen (1996). *Structural Reliability Methods*. Chichester: Wiley.

Geyer, C. J. (2011). Introduction to Markov Chain Monte Carlo. In S. Brooks et al. (Eds.), *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC.

Hasofer, A. & N. Lind (1974). An exact and invariant firstorder reliability format. *Journal of the Engineering Mechanics Division ASCE 100*(1), 111–121.

Hohenbichler, M. & R. Rackwitz (1981). Non-normal dependent vectors in structural safety. *Journal of the Engineering Mechanics Division ASCE 107*(6), 1227–1241.

Lehmann, E. & G. Casella (1998). *Theory of Point Estimation* (2nd ed.). New York: Springer.

Li, H.-S. & Z.-J. Cao (2016). Matlab codes of Subset Simulation for reliability analysis and structural optimization. *Structural and Multidisciplinary Optimization*, 1–20.

Melchers, R. & A. T. Beck (2018). *Structural Reliability, Analysis and Prediction* (Third ed.). New York: Wiley.

Monahan, J. (2011). *Numerical Methods in Statistics* (second ed.). Cambridge University Press.

Nocedal, J. & S. J. Wright (1999). *Numerical Optimization*. Springer Series in Operations Research. New York: Springer.

Rackwitz, R. & B. Fiessler (1977). Structural Reliability under Combined Random Load Sequences. *Computers and Structures 9*, 489–494.

Spall, J. C. (2004). Stochastic Optimization. In J. Gentle, W. Härdle, and Y. Mori (Eds.), *Handbook of Computational Statistics*. Heidelberg: Springer.

Uribe, F. (2016). Matlab-code for subset simulation. https://de.mathworks.com/matlabcentral/fileexchange/57947-monte-carlo-andsubset-simulation-example.

Zhigljavsky, A. & A. žilinskas (2008). Stochastic Global Optimization. Springer US.

Zuev, K., J. Beck, S. K. Au, & L. Katafygiotis (2012). Bayesian post-processor and other enhancements of Subset Simulation for estimating failure probabilities in high dimensions. *Computers and Structures 92–93*, 283–296.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Two-dimensional approach towards a probabilistic model of fatigue cracking of an industrial pipeline

M. Zieja
Air Force Institute of Technology, Warsaw, Poland

M. Jasztal, S. Stępień & M. Ważny
Military University of Technology, Warsaw, Poland

ABSTRACT: The subject addressed in this paper concerns the constantly developing methods of determining the fatigue strength of industrial pipelines. The description of semi-elliptical crack propagation proposed in the paper applies a deterministic model based on a modified Paris' formula, which became the basis for developing a probabilistic model, being the subject of this publication. Based on a differential equation, a Fokker-Plank parabolic partial equation was derived, which describes a crack development in a probabilistic sense, at the same time taking into account deterministic model relationships. A solution of the equation is a two-dimensional normal distribution of crack propagations. A manner of estimating the distribution parameters is also stated. Having a distribution of probability with known parameters, in the case of an assumed probability of exceeding the permissible crack length, the fatigue life of model elements cut-out from industrial pipelines was estimated.

1 INTRODUCTION

Forecasting the propagation of fatigue cracks in elements of industrial pipelines is an important and still not fully addressed issue, which is of fundamental significance due to severe outcomes of a potential damage. The subject approached in this paper concerns the constantly developing methods of determining the fatigue strength of such objects (Mazur 2002, Song et al. 2002, Śnieżek & Goss 2006, Śnieżek et al. 2007, Śnieżek & Stępień 2007). Current solutions in this field consider the growth of surface cracks, outgoing from free edges of the body and propagate along the pipeline surface, as well as into the material, towards the pipe wall thickness (Boukharouba & Pluyinage 1999, Kim & Hwang 1997, Song et al. 2002, Śnieżek & Stępień 2007, Śnieżek et al. 2006). The occurring randomness of the load, material properties, geometry, etc., prompts many scientists to apply probabilistic crack propagation models (Ahammed 1997 & 1998, Ahammed & Melchers 1997, Caleyo 2002, Kocańda et al. 1999). However, some of them (Baranowski & Małachowski 2015, Kocańda et al. 1999, Śnieżek & Stępień 2007, Śnieżek et al. 2006, Tomaszek et al. 2008), when applying a probabilistic description of a crack growth dynamics, take into account the relationships of a deterministic model, which physically reflects the aspects of fatigue crack growth. Therefore, the description of semi-elliptical crack propagation suggested in the paper, which includes a probabilistic model based on a deterministic model published by the authors (Zieja et al. 2017). A physical description of a crack growth utilizes a modified Paris' formula, taking into account two mutually dependent crack direction, i.e., along the small and large ellipses. The development of the model was based on experimental tests on overhead industrial pipelines after thirty years of operation (made from 1H18 N9T steel), as well as sections of new pipelines, prior to handing them over to operation (made from 1.4541 steel). Samples for fatigue tests were cut out from these pipelines and then subjected to flat (non-zero-pulsating) bending at different value of stress amplitudes. An analysis of many variants of mutual interconnection of cracking velocity in both directions, provided a surprising conformity of the results of interconnection between quotients of cracking velocity in both direction and the quotient of crack length (for different samples and load values) (Zieja et al. 2017). The conducted analysis indicates that in the case of surface length of the crack, it is possible to apply the classic Paris' formula, whereas in the case of cracking towards the depth, the correction factor had to be applied. Next, fatigue life calculations were conducted based on the proposed modifications variants of the Paris' formula, with the obtained fatigue lives being similar, and at the same time "safer" in

relation to the actual fatigue life obtained during fatigue tests. It confirmed the possibility of utilizing the presented variants of the deterministic model. The deterministic model developed in the paper (Zieja et al. 2017) became the base for developing the probabilistic model, which is the main subject of this article.

2 BASIC ASSUMPTIONS OF THE PROBABILISTIC MODEL

In order to describe fatigue cracking in the general case of a random load, an equation proposed by prof. Henryk Tomaszek was used (Kocańda et al. 1999, Śnieżek & Stępień 2007, Tomaszek et al. 2008):

$$\begin{aligned}&\frac{\partial U(a,c,t)}{\partial t} = -C(a,c)U(a,c,t) - \\ &-\frac{\partial b_1(a,c)U(a,c,t)}{\partial a} - \frac{\partial b_2(a,c)U(a,c,t)}{\partial c} + \\ &+\frac{\partial^2 \mu(a,c)U(a,c,t)}{\partial a \partial c} + \frac{1}{2}\frac{\partial^2 d_1(a,c)U(a,c,t)}{\partial a^2} + \\ &+\frac{1}{2}\frac{\partial^2 d_2(a,c)U(a,c,t)}{\partial c^2}\end{aligned} \tag{1}$$

where $U(a,c,t)$ = a function of length c and depth a densities of cracks at a time of t; $C(a,c)$ = a factor characterizing the possibility of a catastrophic crack event, when the crack depth is a, and length is c; $d_1(a,c)$ and $d_2(a,c)$ = mean squares of crack increment in relevant directions, in adopted units of time; $b_1(a,c)$ and $b_2(a,c)$ = mean values of crack increment in relevant directions, in adopted units of time; and $\mu(a,c)$ = correlation moment in an assumed unit of time:

$$\mu(a,c) = r\sqrt{d_1(a,c)}\sqrt{d_2(a,c)},$$

where as r is a crack length and depth correlation factor.

Further analysis of the equation (1) is a difficult issue, since it lacks an analytical solution. However, by adopting certain assumptions, it is possible to simplify the above equation. The assumptions for the probabilistic model have the following form:

a. there are such lengths of fatigue cracks (in both, mutually perpendicular directions) that within a certain interval (or for a given number of load cycles), the probability of a catastrophic element crack is equal to zero;
b. in the deterministic approach, the velocities of fatigue cracking are often defined by appropriately modified Paris' formulas (Zieja et al. 2017);
c. load cycles with a duration of Δt may not appear in a continuous manner, but rather randomly, with an intensity of λ i.e. $\lambda \Delta t \leq 1$.

The following differential equation describing a two-dimensional crack growth in the probabilistic sense, may be formed under these assumptions:

$$U_{a,c,t+\Delta t} = (1-\lambda\,\Delta t)U_{a,c,t} + \lambda\,\Delta t\,U_{a-\Delta a,c-\Delta c,t} \tag{2}$$

where $U_{a,c,t}$ = a probability that at a time t, the crack length is c, and the depth is a; Δa and Δc = crack increments into and along the surface of a material, over a time interval of Δt; and λ = load cycle appearance intensity.

3 THE DETERMINATION OF CRACK LENGTH DISTRIBUTION PROBABILITY

From the equation (2), after a transition to functional notation, we obtain:

$$\begin{aligned}&U(a,c,t+\Delta t) = (1-\lambda\Delta t)U(a,c,t) + \\ &+\lambda\Delta t\,U(a-\Delta a, c-\Delta c, t)\end{aligned} \tag{3}$$

where $U(a,c,t)$ = crack length density function.

Expanding the equation terms into a Taylor series and taking into account, for t, the terms of expansion to the first derivative, and for a and c, the terms of expansion to the second derivatives, the following was obtained:

$$\begin{aligned}&U(a,c,t+\Delta t) = U(a,c,t) + \frac{\partial U(a,c,t)}{\partial t}\,\Delta t \\ &U(a-\Delta a, c-\Delta c, t) = U(a,c,t) - \frac{\partial U(a,c,t)}{\partial c}\,\Delta c - \\ &-\frac{\partial U(a,c,t)}{\partial c}\,\Delta c + \frac{1}{2}\frac{\partial^2 U(a,c,t)}{\partial a^2}\,\Delta a^2 + \\ &+\frac{1}{2}\frac{\partial^2 U(a,c,t)}{\partial c^2}\,\Delta c^2 + \frac{\partial^2 U(a,c,t)}{\partial a \partial c}\Delta a \Delta c.\end{aligned} \tag{4}$$

The equation (2), after conversions, will assume the following form:

$$\begin{aligned}&\frac{\partial U(a,c,t)}{\partial t} = -\lambda\frac{\partial U(a,c,t)}{\partial a}\Delta a - \lambda\frac{\partial U(a,c,t)}{\partial c}\Delta c + \\ &+\frac{1}{2}\lambda\frac{\partial^2 U(a,c,t)}{\partial a^2}\Delta a^2 + \frac{1}{2}\lambda\frac{\partial^2 U(a,c,t)}{\partial c^2}\Delta c^2 + \\ &+\lambda\frac{\partial^2 U(a,c,t)}{\partial a \partial c}\Delta a \Delta c\end{aligned} \tag{5}$$

The Δa and Δc increments should be determined as time functions from the relevant relationships of a deterministic model, presented in a previous

paper (Zieja et al. 2017). For known increments, we can write:

$$\frac{\partial U(a,c,t)}{\partial t} = -b_1(t)\frac{\partial U(a,c,t)}{\partial a} - b_2(t)\frac{\partial U(a,c,t)}{\partial c} + \\ + \frac{1}{2}d_1(t)\frac{\partial^2 U(a,c,t)}{\partial a^2} + \frac{1}{2}d_2(t)\frac{\partial^2 U(a,c,t)}{\partial c^2} + \\ + \mu(t)\frac{\partial^2 U(a,c,t)}{\partial a \partial c} \tag{6}$$

where $b_1(t) = \lambda \Delta a(t)$; $b_2(t) = \lambda \Delta c(t)$; $d_1(t) = \lambda \Delta a^2(t)$; $d_2(t) = \lambda \Delta c^2(t)$; and $\mu(t) = \lambda \Delta a(t) \Delta c(t)$.

Whereby, the dot over the variables means a derivative after time. The solution of the equation (6) has the following form:

$$U(a,c,t) = \frac{1}{2\pi}\frac{1}{\sqrt{d_1(t)d_2(t)}} \times \\ \times \frac{1}{\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left[\frac{(a-b_1(t))^2}{d_1(t)} - 2r\frac{(a-b_1(t))(c-b_2(t))}{\sqrt{d_1(t)d_2(t)}} + \frac{(c-b_2(t))^2}{d_2(t)}\right]} \tag{7}$$

where $b_1(t) = \int_0^t \lambda\, \Delta a(w)\, dw$; $b_2(t) = \int_0^t \lambda\, \Delta c(w)\, dw$;

$d_1(t) = \int_0^t \lambda\, \Delta a^2(w)\, dw$; $d_2(t) = \int_0^t \lambda\, \Delta c^2(w)\, dw$;

$\mu(t) = \int_0^t \lambda\, \Delta a(w) \Delta c(w)\, dw$; and $r = \frac{\mu(t)}{\sqrt{d_1(t)d_2(t)}}$.

4 THE DETERMINATION OF CRACK LENGTH DISTRIBUTION PARAMETERS

Defining the parameters of presented distributions requires the determination of the increments of length "c" and depth "a" of a crack, within one load cycle. For this purpose, we shall utilize the relationships in the previously developed deterministic model (Zieja et al. 2017). Based on the relationships (8), (9), (10) describing the cracking velocity in three variants, and taking into account the solutions of the deterministic models shown below, the determination of required crack length and depth increments was attempted.

The 1st variant—common is only the exponent m of the Paris' formula for both cracking directions:

$$V_a = \frac{da}{dN} = C_a \Delta K_a^m = C_a \left(M_{ka} \Delta\sigma \sqrt{\pi a}\right)^m \\ V_c = \frac{dc}{dN} = C_c \Delta K_c^m = C_c \left(M_{kc} \Delta\sigma \sqrt{\pi c}\right)^m \tag{8}$$

where a = crack depth; c = half the length of a surface crack; ΔK_a = the range of the stress intensity factor for crack depth a; ΔK_c = the range of the stress intensity factor for surface length c; N = number of load cycles; $\Delta\sigma$ = stress range; C_a, C_c, m = Paris formula factors; M_{ka} and M_{kc} = correction factors, taking into account the finiteness of element dimensions, determined for the crack length c and depth a, respectively.

The 2nd variant—correction of the stress intensity factor with a c/a quotient, for the velocity in the direction of the crack depth:

$$V_a = \frac{da}{dN} = C_a \Delta K_a^m = C_a \left(M_{ka} \frac{c}{a} \Delta\sigma \sqrt{\pi a}\right)^m \\ V_c = \frac{dc}{dN} = C_c \Delta K_c^m = C_c \left(M_{kc} \Delta\sigma \sqrt{\pi c}\right)^m. \tag{9}$$

The 3rd variant—correction of the stress intensity factor with a square of the c/a quotient, for the velocity in the direction of the crack depth:

$$V_a = \frac{da}{dN} = C_a \Delta K_a^m = C_a \left(M_{ka} \left(\frac{c}{a}\right)^2 \Delta\sigma \sqrt{\pi a}\right)^m \\ V_c = \frac{dc}{dN} = C_c \Delta K_c^m = C_c \left(M_{kc} \Delta\sigma \sqrt{\pi c}\right)^m. \tag{10}$$

The below notations include the probability P_{th} of an event in which a stress range $\Delta\sigma$ exceeds the threshold value and a crack grows. Moreover, it was assumed that $N = \lambda t$ and additional factors were introduced:

$$\alpha_a = C_a M_{ka}^m P_{th} \Delta\sigma^m \pi^{\frac{m}{2}} \\ \alpha_c = C_c M_{kc}^m P_{th} \Delta\sigma^m \pi^{\frac{m}{2}}. \tag{11}$$

Below you can find the determined two-dimensional crack length and depth increments. The solutions were presented in two variants: for the values of the exponent $m = 2$ and $m \neq 2$ (conducting calculations for specific values of the exponent is only dictated by limitation of the mathematical apparatus and does not have physical justification).

For $m = 2$:

Variant 1:

$$\Delta a = a_0 \alpha_a e^{\alpha_a \lambda t} \\ \Delta c = c_0 \alpha_c e^{\alpha_c \lambda t}; \tag{12}$$

Variant 2:

$$\Delta a = \frac{\alpha_a c_0^2 e^{2\alpha_c \lambda t}}{\sqrt{\frac{\alpha_a}{\alpha_c} c_0^2 \left(e^{2\alpha_c \lambda t} - 1\right) + a_0^2}}; \\ \Delta c = c_0 \alpha_c e^{\alpha_c \lambda t} \tag{13}$$

Variant 3:

$$\Delta a = \frac{\alpha_a c_0^4 e^{4\alpha_c \lambda t}}{\left[\frac{\alpha_a}{\alpha_c} c_0^4 \left(e^{4\alpha_c \lambda t} - 1\right) + a_0^4\right]^{\frac{3}{4}}}.$$
$$\Delta c = c_0 \alpha_c e^{\alpha_c \lambda t}. \tag{14}$$

For $m \neq 2$:
Variant 1:

$$\Delta a = \alpha_a \left(\frac{2-m}{2} \alpha_a \lambda t + a_0^{\frac{2-m}{2}}\right)^{\frac{m}{2-m}}$$
$$\Delta c = \alpha_c \left(\frac{2-m}{2} \alpha_c \lambda t + c_0^{\frac{2-m}{2}}\right)^{\frac{m}{2-m}}; \tag{15}$$

Variant 2:

$$\Delta a = \frac{\alpha_a \left(\frac{2-m}{2} \alpha_c \lambda t + c_0^{\frac{2-m}{2}}\right)^{\frac{2m}{2-m}}}{\left\{\frac{\alpha_a}{\alpha_c}\left[\frac{2-m}{2} \alpha_c \lambda t + c_0^{\frac{2-m}{2}}\right]^{\frac{2+m}{2-m}} + a_0^{\frac{2+m}{2}} - \frac{\alpha_a}{\alpha_c} c_0^{\frac{2+m}{2}}\right\}^{\frac{m}{2+m}}}$$
$$\Delta c = \alpha_c \left(\frac{2-m}{2} \alpha_c \lambda t + c_0^{\frac{2-m}{2}}\right)^{\frac{m}{2-m}}; \tag{16}$$

Variant 3:

$$\Delta a = \frac{\alpha_a \left(\frac{2-m}{2} \alpha_c \lambda t + c_0^{\frac{2-m}{2}}\right)^{\frac{4m}{2-m}}}{\left\{\frac{\alpha_a}{\alpha_c}\left[\frac{2-m}{2} \alpha_c \lambda t + c_0^{\frac{2-m}{2}}\right]^{\frac{2+3m}{2-m}} + a_0^{\frac{3m+2}{2}} - \frac{\alpha_a}{\alpha_c} c_0^{\frac{2+3m}{2}}\right\}^{\frac{3m}{3m+2}}}$$
$$\Delta c = \alpha_c \left(\frac{2-m}{2} \alpha_c \lambda t + c_0^{\frac{2-m}{2}}\right)^{\frac{m}{2-m}}. \tag{17}$$

Next, the parameters of relevant probability distributions were determined. It required the calculation of integrals present in the formula (7). The results obtained for the considered calculation variants are presented below.

For $m = 2$:
Variant 1:

$$b_1(t) = a_0 e^{\alpha_a \lambda t};$$
$$b_2(t) = c_0 e^{\alpha_c \lambda t};$$
$$d_1(t) = \frac{1}{2} a_0^2 \alpha_a e^{2\alpha_a \lambda t};$$
$$d_2(t) = \frac{1}{2} c_0^2 \alpha_c e^{2\alpha_c \lambda t}; \tag{18}$$

Variant 2:

$$b_1(t) = \sqrt{\frac{\alpha_a}{\alpha_c} c_0^2 \left(e^{2\alpha_c \lambda t} - 1\right) + a_0^2};$$
$$b_2(t) = c_0 e^{\alpha_c \lambda t};$$
$$d_1(t) = \frac{\alpha_c}{2}\left[\frac{\alpha_a}{\alpha_c} c_0^2 \left(e^{2\alpha_c \lambda t} - 1\right) + a_0^2 + \left(\frac{\alpha_a}{\alpha_c} c_0^2 - a_0^2\right) \ln\left[\frac{\alpha_a}{\alpha_c} c_0^2 \left(e^{2\alpha_c \lambda t} - 1\right) + a_0^2\right]\right]$$
$$d_2(t) = \frac{1}{2} c_0^2 \alpha_c e^{2\alpha_c \lambda t} \tag{19}$$

Variant 3:

$$b_1(t) = \sqrt[4]{\frac{\alpha_a}{\alpha_c} c_0^4 \left(e^{4\alpha_c \lambda t} - 1\right) + a_0^4};$$
$$b_2(t) = c_0 e^{\alpha_c \lambda t};$$
$$d_1(t) = \frac{\alpha_c}{2}\left[\sqrt{\frac{\alpha_a}{\alpha_c} c_0^4 \left(e^{4\alpha_c \lambda t} - 1\right) + a_0^4} - \frac{\left(\frac{\alpha_a}{\alpha_c} c_0^4 - a_0^4\right)}{\sqrt{\frac{\alpha_a}{\alpha_c} c_0^4 \left(e^{4\alpha_c \lambda t} - 1\right) + a_0^4}}\right];$$
$$d_2(t) = \frac{1}{2} c_0^2 \alpha_c e^{2\alpha_c \lambda t}; \tag{20}$$

For $m \neq 2$:
Variant 1:

$$b_1(t) = \left[\frac{2-m}{2} \alpha_a \lambda t + a_0^{\frac{2-m}{2}}\right]^{\frac{2}{2-m}};$$
$$b_2(t) = \left[\frac{2-m}{2} \alpha_c \lambda t + c_0^{\frac{2-m}{2}}\right]^{\frac{2}{2-m}};$$
$$d_1(t) = \frac{2\alpha_a}{2+m}\left[\frac{2-m}{2} \alpha_a \lambda t + a_0^{\frac{2-m}{2}}\right]^{\frac{2+m}{2-m}};$$
$$d_2(t) = \frac{2\alpha_c}{2+m}\left[\frac{2-m}{2} \alpha_c \lambda t + c_0^{\frac{2-m}{2}}\right]^{\frac{2+m}{2-m}}; \tag{21}$$

Variant 2:

$$b_1(t)=\left\{\frac{\alpha_a}{\alpha_c}\left[\frac{2-m}{2}\alpha_c\lambda t+c_0^{\frac{2-m}{2}}\right]^{\frac{2+m}{2-m}}+a_0^{\frac{2+m}{2}}-\frac{\alpha_a}{\alpha_c}c_0^{\frac{2+m}{2}}\right\}^{\frac{2}{2+m}};$$

$$b_2(t)=\left[\frac{2-m}{2}\alpha_c\lambda t+c_0^{\frac{2-m}{2}}\right]^{\frac{2}{2-m}};$$

$$d_1(t)=\frac{2}{2+m}\alpha_a^{\frac{4}{2+m}}\alpha_c^{\frac{m-2}{2+m}}\left[\frac{2-m}{2}\alpha_c\lambda t+c_0^{\frac{2-m}{2}}\right]^{\frac{2+m}{2-m}};$$

$$d_2(t)=\frac{2\alpha_c}{2+m}\left[\frac{2-m}{2}\alpha_c\lambda t+c_0^{\frac{2-m}{2}}\right]^{\frac{2+m}{2-m}};\tag{22}$$

Variant 3:

$$b_1(t)=\left[\frac{\alpha_a}{\alpha_c}\left[\frac{2-m}{2}\alpha_c\lambda t+c_0^{\frac{2-m}{2}}\right]^{\frac{2+3m}{2-m}}+a_0^{\frac{3m+2}{2}}-\frac{\alpha_a}{\alpha_c}c_0^{\frac{2+3m}{2}}\right]^{\frac{2}{3m+2}};$$

$$b_2(t)=\left[\frac{2-m}{2}\alpha_c\lambda t+c_0^{\frac{2-m}{2}}\right]^{\frac{2}{2-m}};$$

$$d_1(t)=\frac{2}{2+m}\alpha_a^{\frac{4}{3m+2}}\alpha_c^{\frac{3m-2}{3m+2}}\left[\frac{2-m}{2}\alpha_c\lambda t+c_0^{\frac{2-m}{2}}\right]^{\frac{2+m}{2-m}};$$

$$d_2(t)=\frac{2\alpha_c}{2+m}\left[\frac{2-m}{2}\alpha_c\lambda t+c_0^{\frac{2-m}{2}}\right]^{\frac{2+m}{2-m}}.\tag{23}$$

Determining the variant $d_1(t)$ for any value of the exponent m, in the 2nd and 3rd variants of the model required the determination of integral values, with their analytical calculation significantly difficult. It was decided to simplify the integrands through omitting the constants, with their value (especially for large t) has a little impact on the calculation results. In addition, such a simplification will increase the variance, which translates to a more conservative estimation of the fatigue life. As a result of the conducted calculations, expressions were obtained (18÷23), which define the expected value and crack length and depth variance for individually considered calculation variants.

5 ESTIMATING THE FACTORS OCCURRING IN EXPRESSIONS DEFINING DISTRIBUTION PARAMETERS

In order to present distribution parameters in an explicit form, it is necessary to estimate the value of coefficients α_a, α_c, m and a correlation coefficient r. The actual crack growth during operation depends on many factors of a random character, e.g. the course of a random load. Determining an element fatigue life in the probabilistic approach requires the knowledge of the crack growth, covering a full probabilistic characteristic of the crack propagation, therefore, providing information on random cracking factors. Let it be a general notation, in the following form:

$$\begin{matrix}\left[(a_0,t_0),(a_1,t_1),(a_2,t_2),...,(a_n,t_n)\right]\\ \left[(c_0,t_0),(c_1,t_1),(c_2,t_2),...,(c_n,t_n)\right].\end{matrix}\tag{24}$$

Based on experimental data concerning crack growth (24), the values of required factors should be estimated. The value of the m coefficient was decided to be determined according to the commonly adopted manner in a deterministic system, from a relationship binding the velocity of a crack over the surface of material V_c with the value of the stress intensity coefficient. Whilst, the values of the remaining coefficients were determined based on a method with the biggest credibility. Due to the fact that the method is commonly known, its description was omitted. The form of the obtained distribution parameters coefficient estimation results is presented below.

The correlation coefficient relationship for the adopted experimental data has the following form:

$$r^*=\frac{\frac{1}{n}\sum_{k=0}^{n-1}\frac{\left[(a_{k+1}-a_k)-b_1^*(\lambda t_{k+1}-\lambda t_k)\right]\times\left[(c_{k+1}-c_k)-b_2^*(\lambda t_{k+1}-\lambda t_k)\right]}{\lambda t_{k+1}-\lambda t_k}}{\sqrt{d_1^*}\sqrt{d_2^*}},\tag{25}$$

where:

$$b_1^*=\frac{a_n}{\lambda t_n};$$

$$b_2^*=\frac{c_n}{\lambda t_n};$$

$$d_1^*=\delta_1^2=\frac{1}{n}\sum_{k=0}^{n-1}\frac{\left((a_{k+1}-a_k)-b_1^*(\lambda t_{k+1}-\lambda t_k)\right)^2}{\lambda t_{k+1}-\lambda t_k};$$

$$d_2^*=\delta_2^2=\frac{1}{n}\sum_{k=0}^{n-1}\frac{\left((c_{k+1}-c_k)-b_2^*(\lambda t_{k+1}-\lambda t_k)\right)^2}{\lambda t_{k+1}-\lambda t_k}.$$

The expressions used to estimate factors α_a, α_c have the following form:

For $m = 2$:

Variant 1:

$$\alpha_a = \frac{1}{\lambda t_n} \ln \frac{a_n}{a_0}; \quad \alpha_c = \frac{1}{\lambda t_n} \ln \frac{c_n}{c_0}; \tag{26}$$

Variant 2:

$$\alpha_a = \frac{1}{\lambda t_n} \frac{a_n^2 - a_0^2}{c_n^2 - c_0^2} \ln \frac{c_n}{c_0}; \quad \alpha_c = \frac{1}{\lambda t_n} \ln \frac{c_n}{c_0}; \tag{27}$$

Variant 3:

$$\alpha_a = \frac{1}{\lambda t_n} \frac{a_n^4 - a_0^4}{c_n^4 - c_0^4} \ln \frac{c_n}{c_0}; \quad \alpha_c = \frac{1}{\lambda t_n} \ln \frac{c_n}{c_0}. \tag{28}$$

For $m \neq 2$:

Variant 1:

$$\alpha_a = \frac{1}{\lambda t_n} \frac{2}{m-2} \left(\frac{1}{a_0^{\frac{m-2}{2}}} - \frac{1}{a_n^{\frac{m-2}{2}}} \right); \quad \alpha_c = \frac{1}{\lambda t_n} \frac{2}{m-2} \left(\frac{1}{c_0^{\frac{m-2}{2}}} - \frac{1}{c_n^{\frac{m-2}{2}}} \right); \tag{29}$$

Variant 2:

$$\alpha_a = \frac{1}{\lambda t_n} \frac{2}{m-2} \left(\frac{1}{c_0^{\frac{m-2}{2}}} - \frac{1}{c_n^{\frac{m-2}{2}}} \right) \frac{a_n^{\frac{2+m}{2}} - a_0^{\frac{2+m}{2}}}{c_n^{\frac{2+m}{2}} - c_0^{\frac{2+m}{2}}}; \quad \alpha_c = \frac{1}{\lambda t_n} \frac{2}{m-2} \left(\frac{1}{c_0^{\frac{m-2}{2}}} - \frac{1}{c_n^{\frac{m-2}{2}}} \right); \tag{30}$$

Variant 3:

$$\alpha_a = \frac{1}{\lambda t_n} \frac{2}{m-2} \left(\frac{1}{c_0^{\frac{m-2}{2}}} - \frac{1}{c_n^{\frac{m-2}{2}}} \right) \frac{a_n^{\frac{3m+2}{2}} - a_0^{\frac{3m+2}{2}}}{c_n^{\frac{3m+2}{2}} - c_0^{\frac{3m+2}{2}}}; \quad \alpha_c = \frac{1}{\lambda t_n} \frac{2}{m-2} \left(\frac{1}{c_0^{\frac{m-2}{2}}} - \frac{1}{c_n^{\frac{m-2}{2}}} \right). \tag{31}$$

6 FATIGUE LIFE ESTIMATION METHOD

Using the developed form of the crack length distribution with known parameters, it is possible to determine the probability of not exceeding the permissible crack lengths $R(t)$:

$$R(t) = P\left(a \le a_d, c \le c_d, t\right) = \int_{-\infty}^{a_d} \int_{-\infty}^{c_d} U(a,c,t)\, da\, dc. \tag{32}$$

According to the assumptions of the probabilistic model, the permissible crack lengths a_d and c_d should be determined in such a manner, so that the risk of immediate element destruction is sufficiently low. Mutual interconnection of both cracking directions hinders the calculations associated with a two-dimensional normal distribution. The information about linking random variables of the crack length is provided by the correlation coefficient value. In the case of a correlation coefficient value, estimated acc. to the previous equation (25), the coordinate system should be converted so that the correlation coefficient for the converted random variables was equal to zero. The coordinates are converted according to the relationship:

$$\begin{aligned} a' &= a\cos\alpha + c\sin\alpha \\ c' &= -a\sin\alpha + c\cos\alpha, \end{aligned} \tag{33}$$

where the angle α is determined from the formula:

$$tg2\alpha = \frac{2r^* \sqrt{d_1^*} \sqrt{d_2^*}}{d_1^* - d_2^*}. \tag{34}$$

With such a conversion, the modified crack lengths a' and c' are independent random variables for each determined t. The distribution parameters should also be converted according to the above rule:

$$\begin{aligned} b_1'(t) &= b_1(t)\cos\alpha + b_2(t)\sin\alpha; \\ b_2'(t) &= -b_1(t)\sin\alpha + b_2(t)\cos\alpha; \\ d_1'(t) &= \left(\sqrt{d_1(t)}\cos\alpha + \sqrt{d_2(t)}\sin\alpha\right)^2; \\ d_2'(t) &= \left(-\sqrt{d_1(t)}\sin\alpha + \sqrt{d_2(t)}\cos\alpha\right)^2. \end{aligned} \tag{35}$$

In the new coordinate system, the relationship (32) shall assume the following form:

$$\begin{aligned} R(t) &= P\left(a \le a_d, c \le c_d, t\right) = \\ &= P_1\left(a' \le a_d', t\right) \cdot P_2\left(c' \le c_d', t\right), \end{aligned} \tag{36}$$

where: $P_1\left(a' \le a_d', t\right)$ and $P_2\left(c' \le c_d', t\right)$ are determined from one-dimensional normal

distribution. Using normal distribution tables requires additional standardization of random variables. Assuming that $R(t) \leq R_0$, where: R_0 is the minimum permissible probability of not exceeding the permissible crack lengths a_d and c_d, for which an element's fatigue life may be determined. At the same time, probability $1 - R_0$ will be the probability of exceeding permissible operational durations.

7 CALCULATION EXAMPLE

The example shall utilize the data from experimental tests presented in papers (Śnieżek & Stępień 2007, Zieja et al. 2017). The sample designations assumed in the aforementioned elaborations were kept. As per the previously presented assumptions for three variants of the model, calculations of factors present in expressions for crack length distribution parameters were made, and their results are shown in Table 1 below.

Having the values of calculated coefficient (as per the previously presented procedure), the estimation of the fatigue life for individual samples was commenced. It was decided to present the obtained results in groups concerning a given material and a given loading manner. For all samples, a minimum probability of not exceeding the permissible crack parameters $R_0 = 0{,}9$ was assumed.

For Group I samples cut out from a new pipeline, prior to including the operation, made from 1.4541 steel (acc. to EN 10088-3), subject to flat non-zero-pulsating bending with a frequency of 20 Hz (cycle asymmetry coefficient R = 0) with subsequent stress nominal amplitude values given in Table 2. Permissible crack lengths in individual directions: $a_d = 3$ mm; $c_d = 9$ mm.

For Group II samples cut out from a pipeline after 30 years of operation, made from 1H18N9T steel (acc. to PN-71/H-86020), subject to flat non-zero-pulsating bending with a frequency of 20 Hz (cycle asymmetry coefficient R = 0) with subsequent stress nominal amplitude values given in table 4. Permissible crack lengths in individual directions: $a_d = 3$ mm; $c_d = 9$ mm.

The obtained fatigue life calculation results, based on the three developed model variants are similar to the actual life, obtained during fatigue tests of bent samples. At the same time, in the second and third calculation variant, "safe" fatigue lives were achieved in comparison to the actual life, however less conservative from the point of

Table 2. Values of nominal stress amplitudes for 1.4541 steel samples.

Sample designation	Stress range $\Delta\sigma$ [MPa]
Sample 1	170
Sample 2	180
Sample 3	200
Sample 4	200
Sample 5	250
Sample 6	250

Table 3. Fatigue life calculation results acc. to the developed probabilistic model.

Sample designation	Actual life N_R [cycles]	Projected life N_T [cycles]		
		Variant 1	Variant 2	Variant 3
Sample 1	3 689 125	3 543 791	3 657 527	3 657 848
Sample 2	2 705 250	2 477 217	2 573 436	2 673 900
Sample 3	2 217 000	2 186 162	2 212 251	2 212 263
Sample 4	1 025 375	974 936	1 013 023	1 013 197
Sample 5	411 875	354 515	402 535	402 742
Sample 6	433 625	415 798	430 378	430 408

Table 1. Calculated values of distribution parameters' factors.

Sample	r	m	Variant 1		Variant 2		Variant 3	
			α_a	α_c	α_a	α_c	α_a	α_c
1	0.855	5.8417	1.82×10^{-7}	4.51×10^{-8}	7.16×10^{-10}	4.51×10^{-8}	1.51×10^{-12}	4.51×10^{-8}
2	0.888	5.8358	2.48×10^{-7}	6.16×10^{-8}	9.82×10^{-10}	6.16×10^{-8}	2.08×10^{-12}	6.16×10^{-8}
3	0.853	7.8811	2.51×10^{-7}	2.84×10^{-8}	1.54×10^{-10}	2.84×10^{-8}	3.76×10^{-14}	2.84×10^{-8}
4	0.819	5.3271	6.96×10^{-7}	2.12×10^{-7}	4.4×10^{-9}	2.12×10^{-7}	1.06×10^{-11}	2.12×10^{-7}
5	0.821	5.3677	1.73×10^{-6}	5.17×10^{-7}	1.08×10^{-8}	5.17×10^{-7}	3.86×10^{-11}	5.17×10^{-7}
6	0.852	6.1564	1.49×10^{-6}	3.27×10^{-7}	4.4×10^{-9}	3.27×10^{-7}	6.64×10^{-12}	3.27×10^{-7}
7	0.868	4.8690	6.13×10^{-6}	1.76×10^{-6}	4.8×10^{-8}	1.76×10^{-6}	2.94×10^{-10}	1.76×10^{-6}
8	0.878	5.0368	2.72×10^{-6}	7.21×10^{-7}	1.8×10^{-8}	7.21×10^{-7}	9.25×10^{-11}	1.76×10^{-6}
9.18	0.877	5.0133	1.36×10^{-6}	3.51×10^{-7}	8.82×10^{-9}	3.51×10^{-7}	4.57×10^{-11}	3.51×10^{-7}

Table 4. Values of nominal stress amplitudes for 1H18N9T steel samples.

Sample designation	Stress range $\Delta\sigma$ [MPa]
Sample 7	300
Sample 8	250
Sample 9.18	200

Table 5. Fatigue life calculation results acc. to the developed probabilistic model.

Sample designation	Actual life N_R [cycles]	Projected life N_T [cycles]		
		Variant 1	Variant 2	Variant 3
Sample 7	152 000	148 731	150 088	150 162
Sample 8	338 400	329 867	333 899	333 934
Sample 9.18	712 670	698 102	704 266	704 319

view of preventive forecasting of fatigue life, in comparison to the fatigue life results for the first variant. It confirms the earlier remarks regarding the possibility of applying the probabilistic model to estimate fatigue life.

8 CONCLUSIONS

The probabilistic description of semi-elliptic crack propagations proposed in the paper utilizes a previously developed deterministic model (Zieja et al. 2017), based on the modified Paris' formula. On the basis of the differential equation, the Fokker-Plank parabolic partial equation, which describes a crack development in a probabilistic sense, was derived, at the same time taking into account the deterministic model relationships. A solution of the equation is a two-dimensional normal distribution of crack propagations. A manner of estimating the distribution parameters is also stated. Having the distribution of probability with known parameters, in the case of an assumed risk of exceeding the permissible crack length, the fatigue life of model elements cut-out from industrial pipelines was estimated.

The conducted analyses took into account the interconnection of cracking directions, which is the main advantage of the two-dimensional model. The forecast fatigue life depends on two permissible crack lengths. Taking into account the critical conditions allows to ensure a higher safety level in relation to a single condition.

Experimental tests regarding the fatigue life of model elements with a propagating semi-elliptical crack, enabled to provide the calculation model with necessary data, describing the development character of the considered two-dimensional cracks. The forecast and recorded during the experimental tests fatigue lives of samples show satisfying compliance. However, the fatigue lives forecast on the basis of developed mathematical models are lower than in the case of the experimental ones, hence, "safe" and, in addition, less conservative than the results obtained on the basis of a deterministic model and presented in the paper (Zieja et al. 2017).

REFERENCES

Ahammed, M. 1997. Prediction of remaining strength of corroded pressurised pipelines. *International Journal of Pressure Vessels and Piping* 71: 213–217.

Ahammed, M. 1998. Probabilistic estimation of remaining life of a pipeline in the presence of active corrosion defects. *International Journal of Pressure Vessels and Piping* 75: 321–329.

Ahammed, M. & Melchers, R.E. 1997. Probabilistic analysis of underground pipelines subject to combined stress and corrosion. *Engineering Structures* 19(12): 988–994.

Baranowski, P. & Małachowski, J. 2015. Numerical study of selected military vehicle chassis subjected to blast loading in terms of tire strength improving. Bulletin of The Polish Academy of Sciences: Technical Sciences, 63(4): 867–878.

Boukharouba, T. & Pluyinage, G. 1999. Prediction of semi-elliptical defect form. Case of a pipe subjected to internal pressure. *Nuclear Engineering and Design* 188: 161–171.

Caleyo, F. & Gonzalez, J.L. & Hallen, J.M. 2002. A study on the reliability assessment methodology for pipelines with active corrosion defects. *International Journal of Pressure Vessels and Piping* 79: 77–86.

Kim, J.H. & Hwang, I.S. 1997. Crack shape evolution of surface flaws under fatigue loading of austenitic pipes. *Nuclear Engineering and Design* 174: 17–24.

Kocańda, D. & Kocańda, S. & Miller, K.J. & Tomaszek, H. 1999. Experimental and theoretical investigations of short fatigue crack growth in laser hardened medium carbon steel. *Engineering Against Fatigue:* 501–507. Rotterdam-Brookfield: Balkema.

Kocańda, D. & Kocańda, S. & Tomaszek, H. 1999. Probabilistic approach to the short and long fatigue crack growth description in a notched member. *Fatigue'99; Proc. Intern. Fatigue Congress*: 2673–2678. Beijing: Higher Education Press Beijing, EMAS, Cradley Heath.

Mazur, A. 2002. Mechanical performance of 1H18N9T steel chemical pipeline. *Doctoral dissertation*. Warsaw: Military University of Technology.

Śnieżek, L. & Goss, Cz. 2006. Investigation Into Fatigue Life of Welded Chemical Pipelines. *ECF 16-16th European Conference of Fracture*, Alexandroupolis, Greece, July 3–7. Abstract: 173–174 (CD-6 pages).

Śnieżek, L. & Goss, Cz. & Mazur, A. 2007. Experimental and Theoretical Investigations of Fatigue Life of Chemical Pipelines. *Archives of Civil Engineering* l.

Śnieżek, L. & Stępień, S. 2007. *Fatigue life of the chemical pipeline with a propagating semi-elliptical crack.* Warsaw: Military University of Technology.

Śnieżek, L. & Stępień, S. & Kulec, P. 2006. The probabilistically approached forecast on fatigue life of notched members. *Technical Sciences* 9: 79–94.

Song, P.S. & Sheu, B.C. & Shieh, Y.I. 2002. Prediction of semi-elliptical surface crack growth in 2024-T4 aluminium alloy. *International Journal of Pressure Vessels and Piping* 79: 273–278.

Tomaszek, H. & Jasztal, M. & Zieja, M. 2013. Application of the Paris formula with m = 2 and the variable load spectrum to a simplified method for evaluation of reliability and fatigue life demonstrated by aircraft components. *Maintenance and Reliability* 15(4): 297–303.

Tomaszek, H. & Jasztal, M. & Zieja, M. 2011. A simplified method to assess fatigue life of selected structural components of an aircraft for a variable load spectrum. *Maintenance and Reliability* 4: 29–34.

Tomaszek, H. & Żurek, J. & Jasztal, M. 2008. *Forecasting of damage hazardous for the aircraft flight safety*. Warsaw: Wydawnictwo Naukowe Instytutu Technologii Eksploatacji—PIB.

Woch, M. & Kurdelski, M. & Matyjewski, M. 2015. Reliability at the checkpoints of an aircraft supporting structure. Ek-sploatacja i Niezawodnosc—Maintenance and Reliability 17(3): 457–462.

Zieja, M. & Jasztal, M. & Stępień, S. & Ważny, M. 2017. The analysis of the fatigue crack growth rate in pipeline elements in two-dimensional depiction. Safety and Reliability—Theory and Applications. 27th Annual Conference on European Safety and Reliability (ESREL 2017). London: CRC/Balkema: 2147–2154.

Żurek, J. & Smalko, Z. & Zieja, M. 2010. Methods applied to identify causes of air events. *Reliability, Risk and Safety: Theory and Applications*. CRC Press-Taylor and Francis Group: 1817–1822.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Environmental contours for design of ice-capable vessels

Wei Chai, Bernt J. Leira & Chana Sinsabvarodom
Department of Marine Technology, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: The environmental contour method is an established method for design of ships and marine structures. In this method, environmental contours for a given return period are developed in order to identify the environmental conditions which are associated with the most critical structural responses, and then the long-term extreme response for the same return period can be effectively estimated based on the response statistics of the most critical (short-term) responses. In this work, the environmental contour method is described for design of ice-capable vessels in Arctic regions. In particular, for vessels sailing in Arctic regions, the most dangerous condition is associated with ice ridges. Based on the probability distributions of key parameters of the first-year ice ridges which determine the ice ridge loads, the environmental contours are developed. The effects of the correlation between different environmental variables are discussed. A simple case is proposed to illustrate application of the environmental contour method for design of ice-capable vessels.

1 INTRODUCTION

For ships and marine structures subjected to environmental loads, such as wind, wave and ice forces, etc., evaluation of the extreme response over their specific lifetime is necessary and important at the design stage. The full long-term analysis which accounts for the structural response from each short-term environmental condition and the occurrence rate of each short-term condition is recognized as the most reliable and accurate approach. However, such long-term analysis is time consuming, especially for large and complex systems (Leira, 2008).

In order to improve the computational efficiency, the long-term analysis can be simplified either by reducing the computation cost of short-term analysis (Chai et al., 2016) or by developing approximate method that requires a lower number of short-term analysis, such as the IFORM (inverse first order reliability method) (Giske et al., 2017) and the environmental contour method (Haver and Winterstein, 2009).

The environmental contour method offers a simplified and fast way to estimate the long-term extreme response and has been widely used for design of ships and marine structures at the early design stage (Li et al., 2016, DNV, 2010, Baarholm et al., 2010). An environmental contour is the locus of all environmental parameters correspond to a given annual exceedance probability along which extreme responses with the corresponding return period should lie. Traditionally, establishment of such environmental contour for a given return period is based on the IFORM and the joint distribution of the environmental parameters.

The main advantage for the environmental contour method lies in the fact that the structural response is uncoupled from the description of the environmental variables. As a result, the long-term extreme response for a given return period can be approximately estimated on the basis of response statistics associated with the most critical short-term environmental conditions along the environmental contour, which corresponds to a given return period. Therefore, in this method, short-term analyses are performed only for a few conditions on the environmental contour and then numerical simulations or experiments are performed in order to identify the most critical short-term response for further application (Winterstein et al., 1993).

In this work, the concept of environmental contour is applied for design of ice-capable vessels operating in Arctic regions. For ships sailing in sea ice areas without icebergs, ice ridges are assumed to pose a major threat to the vessel since they determine and govern the design loads for the vessel (Høyland, 2014). Key parameters of the ice ridge which determine the main loads of the ship-ice ridge interaction process are identified at first. Then, probabilistic models are applied to describe the distributions of these key parameters in order to develop the environmental contours.

The procedure of extreme response prediction based on the environmental contour method is simply described in this work. Finally, a simple case is given to illustrate the development of a

desired environmental contour used for design of ice-capable vessels.

2 ENVIROMENTAL CONTOUR METHOD

In this section, construction of the environmental contour for a given return period and application of the environmental contour method are presented. Assume that for the N-year return period, the corresponding failure probability of the structure response is given as P_f. As for the traditional IFORM to generate the environmental contour with a N-year return period, a n-dimensional (n is the dimensionality of the environment parameter vector) hypersphere with radius β_F is first created in the U space, with the value of β_F being given as (Haver and Winterstein, 2009):

$$\beta_F = \Phi^{-1}(1 - P_f) \tag{1}$$

where Φ represents the cumulative density function of the standard normal distribution.

Then, the n-dimensional hypersphere is transformed into the physical parameter space by consideration of the joint distribution of the environmental parameters in order to get the environmental contour. Figure 1 presents an example of an environmental contour in the physical parameter space.

After establishment of the environmental contour with the N-year return period, a limited set of design conditions, i.e. short-term environmental conditions, are selected on the environmental contour (e.g. see Figure 1). Time consuming numerical calculations (or expensive experiments) for structural responses are only needed for these selected environmental conditions. The most critical short-term response is identified in order to estimate the long-term extreme response with the N-year return period. The main process of applying the environmental contour method for long-term extreme response approximation is presented in Figure 2.

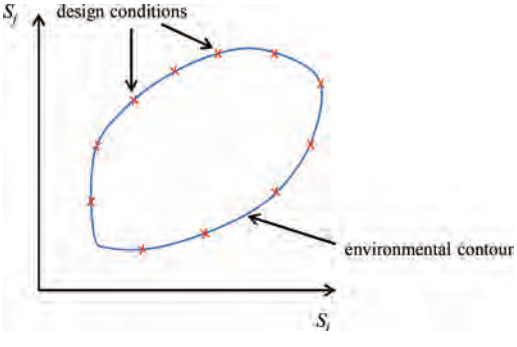


Figure 1. Environmental contour in the physical parameter space and selected environmental design conditions.

3 FIRST-YEAR ICE RIDGE

A ridge is a line or wall of broken ice features forced by pressure or shear. When first formed, an ice ridge is simply a pile of unconsolidated ice blocks. Then, these blocks may become to some extent consolidated by refreezing processes and form the ice ridge. Figure 1 illustrates a typical ice ridge, which consists of two parts: the sail and the keel. The sail part is above the water and has pores filled with air and snow. The keel is the underwater part and can be further separated into an upper completely frozen layer called the consolidated layer, which is always thicker than the surrounding level ice and a lower unconsolidated part that has loose blocks partially refrozen together with water trapped between the blocks (ISO, 2010).

Generally, ice ridges are subdivided into first-year versus older ice ridges. During its first winter and summer, an ice ridge is called a first-year ice ridge (e.g. Figure 3). The consolidation process in the keel part progresses with time and the keel part is closed to fully consolidated if the ridge has survived one summer's melt. Ridges that survive one or more summers are referred to as old ice ridges.

In this work, first-year ice ridges are considered for design of ice-capable vessels. For one thing, the ice conditions along the commercial Arctic shipping routes, such as the Northern Sea Route, are mostly first-year and few ice appears in summer seasons. For another, fewer studies have been made on old ice ridges than the first-year ice ridges and studies for the mechanical and physical properties of the multi-year ice ridges are very limit.

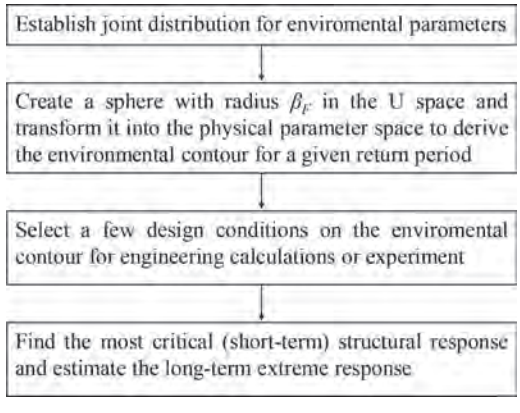


Figure 2. Flowchart of the environmental contour method used for approximate the long-term extreme response.

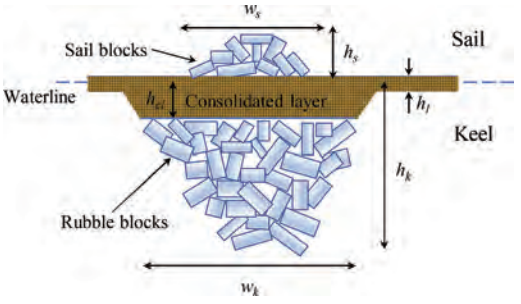


Figure 3. First-year ice ridge with some key parameters: sail draft h_s, sail width ws, consolidated layer thickness h_{cl}, surrounding level ice thickness h_l, keel draft h_k and keel width w_k.

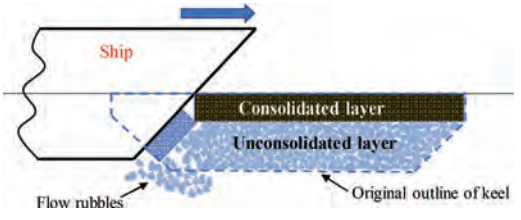


Figure 4. Illustration of ship interacting with a typical first-year ice ridge.

For the scenario of an ice-capable ship interacting with a first-year ice ridge, the ship structure can be simplified as a downward sloping structure. The effects of ridge sail can be neglected since the volume of the sail is small compared to that of the keel part (ISO, 2010). Former studies on a ridge failure against a confederation bridge, whose piers are designed as (slope) ice-breaking cones, have shown that failure of the consolidated layer is the dominant term in determining the keel loads due to the ice ridge interaction with the slope structure. Also, there is no correlation between the keel depth, h_k, and the keel loads. Based on former studies for first-year ice ridge interacting with slope structures, the ship-ice ridge interaction process is illustrated in Figure 4. Flow rubbles from the unconsolidated layer would be cleared by the local water current. The action due to the consolidated layer part can be approximated as level ice with an equal thickness that interacts with the vessel and the mechanical properties of the consolidated layer are assumed to be close to those of level ice. Therefore, the ship-ridge interaction process is further simplified as a ship-level ice interaction event for the purpose of preliminary design.

The ship-level ice interaction process is initiated by a localized crushing of the ice edge and then the contact area between the ship and the ice sheet as well as the crushing force both increase when the ship is advancing and penetrating the ice features. The ice sheet eventually deflects and the bending stresses imply a flexural failure at a certain distance from the crushing region (Jordaan, 2001). Therefore, the ice thickness, the crushing strength and the flexural strength of the (equivalent) level ice are considered as the key parameters for determining the loads due to ship-ice ridge interaction process at the early design stage.

4 PROBABILISTIC MODELS

In this section, probabilistic models are applied in order to describe the distributions of the abovementioned key parameters which determines the loads due to ship-ice ridge interaction. These key parameter are believed to be dependent on geographical location and season, etc. In this work, the first-year ice ridges in the Barents Sea are considered since many field experiments have been performed in this area by relevant Norwegian and Russian research institutes.

The average thickness of the consolidated layer, based on measurements which are collected by mechanical drilling, can be described by a Gamma distribution (Strub-Klein and Sudom, 2012):

$$f(h_{cl}) = \frac{1}{\Gamma(k)\theta^k} h_{cl}^{k-1} \exp\left(-\frac{h_{cl}}{\theta}\right) \tag{2}$$

where k and θ are the shape and scale parameters for the Gamma distribution, respectively. Based on the experimental data, these two values are determined to be 2.97 and 0.54.

The fitted and sample probability density function of the consolidated layer thickness are plotted in Figure 5. However, sample data for the flexural

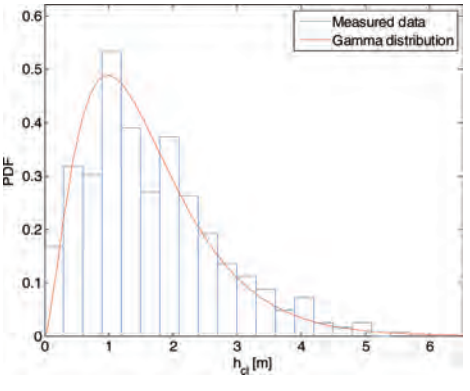


Figure 5. Probability distribution for the consolidated layer thickness h_{cl}.

strength σ_f and the crushing strength σ_c of the consolidated layer is very limited. The mechanical properties of the surrounding level sea ice can serve as effective alternatives (Timco et al., 2000). On the basis of the experimental data for the flexural strength of the level ice in the Barents Sea, the marginal PDF (probability density function) of the flexural strength is described by a two-parameter Weibull distribution (Krupina and Kubyshkin, 2007):

$$f(\sigma_f) = \frac{\beta}{\alpha}\left(\frac{\sigma_f}{\alpha}\right)^{\beta-1}\exp\left(\left(\frac{\sigma_f}{\alpha}\right)^{\beta}\right) \tag{3}$$

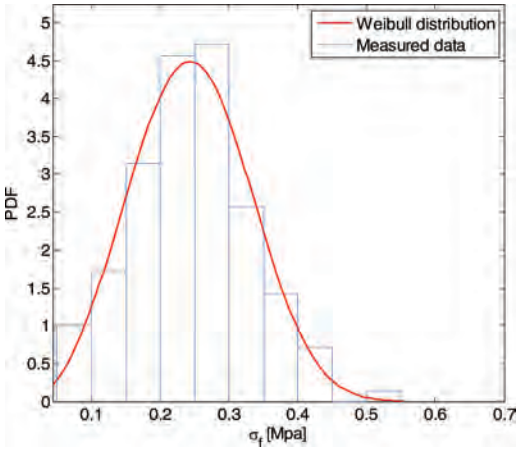


Figure 6. Probability distribution for the flexural strength σ_f.

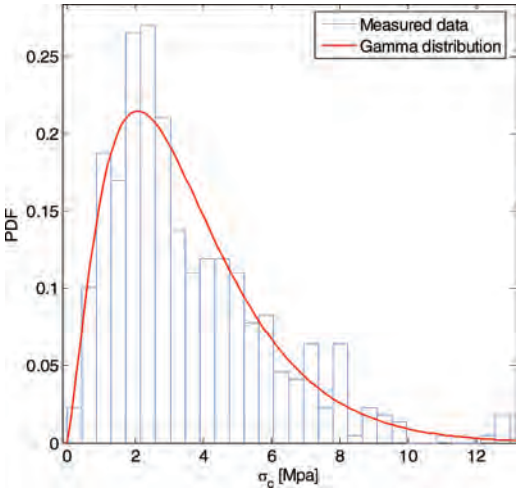


Figure 7. Probability distribution for the crushing strength σ_c.

where the scale parameter $\alpha = 0.274$ and the shape parameter $\beta = 3.167$ are obtained from actual data. The marginal PDF and the empirical histogram of the flexural strength are presented in Figure 6.

Full-scale measurements have been performed to collect the data of the crushing (or compressive) strength of the level ice in the Barents Sea. The distribution of the crushing strength for the vertically loaded samples can be described by the Gamma distribution given in equation (1) with a shape parameter of 2.5 and a scale parameter of 1.40 (Strub-Klein, 2017). The marginal PDF of the of the crushing strength and the corresponding empirical histogram based on experimental data are shown in Figure 7.

5 ENVIROMENTAL CONTOUR

Assume that the desired ice-capable ship mainly sails and operates in the Barents Sea with an annual voyage length of 5000 km in ice ridge fields. The ice ridge density is 2/km along the route. Therefore, for a 50-year return period, the failure probability P_f is determined as:

$$P_f = 1/(50 \cdot 5000 \cdot 2) \tag{4}$$

Correspondingly, the radius of the sphere in the U space, β_F is equal to 4.611 according to equation (1).

Transformation of the sphere in the U space into the environmental contour in the physical parameter space is generally based on the inverse Rosenblatt transformation (DNV, 2010). However, in order to perform such a transformation, the joint distribution of the environmental parameters as described by the conditional modeling approach is required. This necessitates a great amount of sampled data. Due to the limitation of experimental data for the first-year ice ridge statistics, only the marginal PDFs of the abovementioned key parameters can be obtained. Nevertheless, the joint distribution of the environmental parameters with consideration of the correlations between these environmental variables can be approximated by the Nataf distribution model (Silva-González et al., 2013). Then, the environmental contours can be obtained by the Nataf transformation.

Let the variables S_1, S_2 and S_3 represent the consolidated layer thickness, the flexural strength and the crushing strength, respectively. The symbols ρ_{12}, ρ_{13} and ρ_{23} denote the correlation coefficients between these variables. The Nataf transformation model is given as:

$$F_{S_1}(s_1) = \Phi(u_1) \tag{5}$$

$$F_{S_2}(s_2) = \Phi(u_2\sqrt{1-\rho_{12}'^2} + \rho_{12}'\Phi^{-1}(F_{S_1}(s_1))) \tag{6}$$

$$\begin{aligned} F_{S_3}(s_3) = \Phi(&\frac{u_3}{\sqrt{1-\rho_{12}'^2}}\sqrt{1-\rho_{12}'^2-\rho_{13}'^2-\rho_{23}'^2+2\rho_{12}'\rho_{13}'\rho_{23}'} \\ &+\frac{1}{1-\rho_{12}'^2}(\rho_{13}'-\rho_{12}'\rho_{23}')\Phi^{-1}(F_{S_1}(s_1)) \\ &+\frac{1}{1-\rho_{12}'^2}(\rho_{23}'-\rho_{12}'\rho_{13}')\Phi^{-1}(F_{S_2}(s_2))) \end{aligned} \tag{7}$$

where U_1, U_2 and U_3 represent independent standard normal variables and their values can be obtained based on the following equation:

$$\sum_{i=1}^{3} u_i^2 = \beta_F^2 \tag{8}$$

The coefficients ρ_{ij}' ($i, j = 1, 2, 3$; $i \neq j$) are the corresponding (equivalent) correlation coefficients used in the Nataf transformation and their relationship with ρ_{ij} can be approximated by a semi-empirical equation, which is given as:

$$\rho_{ij}' = \xi \cdot \rho_{ij} \tag{9}$$

where relevant description for determining the function ξ can be found in Ref. (Liu and Der Kiureghian, 1986).

There exist few previous studies on the correlation between these three key parameters for sea ice material. For simplicity, we assume that the correlation coefficients ρ_{ij} ($i, j = 1, 2, 3$; $i \neq j$) are equal to 0.5. Correspondingly, the 50-year contour surface for the three key parameters is obtained by the Nataf model described by equations (5)–(9) and it is plotted in Figure 8.

The correlation coefficients in this simple case are chosen somewhat arbitrarily due to a imitated amount of information in the literature. Therefore, the influence of the correlation coefficient on the shape of the environmental contour is studied in a parametric way. For simplicity, a two-dimensional contour line on the contour surface is selected for this sensitive study. The consolidated layer is fixed as its mean value 1.59 m, and the corresponding two-dimensional contour lines for different values of the correlation coefficient between the flexural strength and the crushing strength are plotted in Figure 9. In addition, for the sensitive study, ρ_{12} and ρ_{13} are kept at a fixed value equal to 0.5.

It is seen in Figure 9 that the variation of the correlation coefficient has a quite important effect on the shapes of the contour lines. The maximum values of the flexural strength and the crushing strength on the contour line do not change with the correlation coefficient. However, the critical

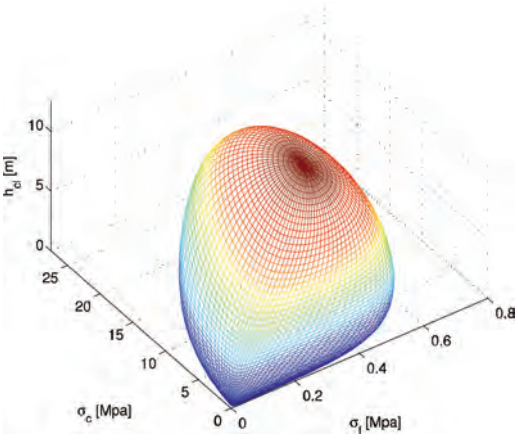


Figure 8. The 50-year contour surface for the consolidated layer thickness, flexural strength and crushing strength.

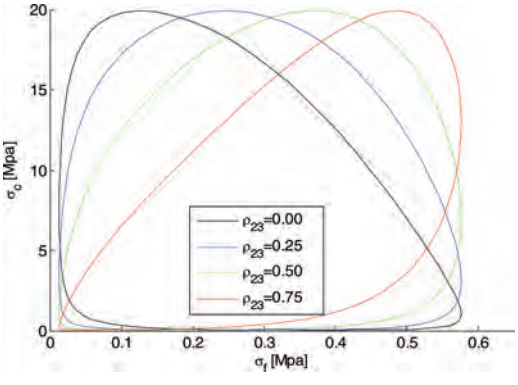


Figure 9. Influence of the correlation coefficient on the two-dimensional contour lines.

region with large values of the flexural strength and crushing strength becomes narrow as the correlation coefficients increase. Narrower critical region implies that large values of the flexural strength are more easily (or more possibility) to be accompanied by large values of the crushing strength, which would cause serious structural response. Therefore, the correlation coefficients has significant influence on the extreme response for the ice-capable vessel sailing in the ice ridge field.

6 CONCLUSIONS

In this work, the environmental contour method was proposed for design of ice-capable vessels sailing in areas with ice ridges. Environmental

contours for the key parameters of the first-year ice ridge have been established by applying the Nataf model.

Based on the abovementioned simple case, it is found that variation of the correlation coefficients between the environmental parameters has a very important influence on the shapes of the environment contour as well as on the extreme response of the vessel. Therefore, a reliable numerical model is important in order to investigate this influence.

Furthermore, better knowledge for the correlations between these key parameters, either provided by experiments or numerical simulations, will promote the development of reliability-based design of ice-capable vessels in Arctic regions.

ACKNOWLEDGEMENTS

This work is supported by Research Council of Norway (RCN project number: 249272/O80). The authors wish to thank Prof. Knut Vilhelm Høyland for discussion in connection with ice ridges.

REFERENCES

Baarholm, G.S., Haver, S. & Økland, O.D. (2010) Combining contours of significant wave height and peak period with platform response distributions for predicting design response. *Marine Structures,* 23, 147–163.

Chai, W., Naess, A., Leira, B.J. & Bulian, G. (2016) Efficient Monte Carlo simulation and Grim effective wave model for predicting the extreme response of a vessel rolling in random head seas. *Ocean Engineering,* 123, 191–203.

DNV (2010) Recommended practice DNV-RP-C205: Environmental conditions and environmental loads. *Norway: Det Norske Veritas.*

Giske, F.-I.G., Leira, B.J. & Øiseth, O. (2017) Full long-term extreme response analysis of marine structures using inverse FORM. *Probabilistic Engineering Mechanics*.

Haver, S. & Winterstein, S.R. (2009) Environmental contour lines: A method for estimating long term extremes by a short term analysis. *Transactions of the Society of Naval Architects and Marine Engineers,* 116, 116–127.

Høyland, K.V. (2014) Ice ridge characteristics and engineering concerns regarding ice ridges. *Proc IAHR Int Symp on Ice.* Singapore.

ISO (2010) 19906: Petroleum and Natural Gas Industries–Arctic offshore structures. *Geneva: ISO.*

Jordaan, I.J. (2001) Mechanics of ice–structure interaction. *Engineering Fracture Mechanics,* 68, 1923–1960.

Krupina, N.A. & Kubyshkin, N.V. (2007) Flexural Strength of Drifting Level First-year Ice in Barents Sea. *International Journal of Offshore and Polar Engineering,* 17.

Leira, B.J. (2008) A comparison of stochastic process models for definition of design contours. *Structural Safety,* 30, 493–505.

Li, Q., Gao, Z. & Moan, T. (2016) Modified environmental contour method for predicting long-term extreme responses of bottom-fixed offshore wind turbines. *Marine Structures,* 48, 15–32.

Liu, P.-L. & Der Kiureghian, A. (1986) Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics,* 1, 105–112.

Silva-González, F., Heredia-Zavoni, E. & Montes-Iturrizaga, R. (2013) Development of environmental contours using Nataf distribution model. *Ocean Engineering,* 58, 27–34.

Strub-Klein, L. & Sudom, D. (2012) A comprehensive analysis of the morphology of first-year sea ice ridges. *Cold Regions Science and Technology,* 82, 94–109.

Strub-Klein, L. (2017) A Statistical Analysis of First-Year Level Ice Uniaxial Compressive Strength in the Svalbard Area. *Journal of Offshore Mechanics and Arctic Engineering,* 139, 011503.

Timco, G., Croasdale, K. & Wright, B. (2000) An overview of first-year sea ice ridges. *PERD/CHC report*, 5–112.

Winterstein, S.R., Ude, T.C., Cornell, C.A., Bjerager, P. & Haver, S. (1993) Environmental parameters for extreme response: Inverse FORM with omission factors. *Proceedings of the ICOSSAR-93, Innsbruck, Austria*, 551–557.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Partial factors for fatigue loads in the Eurocode system for road bridge design

S.B. Hashemi, J. Maljaars & H.H. Snijder
Department of the Built Environment, Eindhoven University of Technology, Eindhoven, The Netherlands

ABSTRACT: In the Eurocode system, for fatigue design of bridges, the recommended partial factor for fatigue traffic loads is set to 1. In this paper, the adequacy of this approach is investigated by performing a reliability analysis on two types of welded joint in a main girder of a steel motorway bridge. For this purpose, a weigh in motion measurement dataset belonging to a main Dutch motorway has been compared with the fatigue load model 4 of Eurocode EN 1991-2 with respect to the stress spectrum and the fatigue damage of two structural steel details. Several structural schemes have been considered to study the effect of the shape and length of the influence line. The distributions of the stochastic variables such as dynamic amplification, accuracy of the structural model, and future traffic trends have been estimated or taken from literature. Partial factors for fatigue loads have then been calibrated in such a way that the target reliability is obtained. The influence of each stochastic variable on partial factors has been studied by derivation of the sensitivity factors. The results show that a considerably higher fatigue partial factor is required for fatigue loads on road bridges than the value of 1 currently recommended in EN 1991-2.

1 INTRODUCTION

Most of the parameters governing the load effects on a structure as well as the structure's response are uncertain in nature. To avoid the complexity of dealing with these random variables, standards such as Eurocode provide a deterministic model for structural design wherein characteristic values are provided for the load and the resistance and these values need to be corrected by partial factors to arrive at the design values, the latter giving the desired reliability level. In the other word, partial factors are introduced to link the deterministic models used for practical design to the required reliability level. The general equation for design of a structural component according to the Eurocode standard (EN1990, 2002) is;

$$\frac{R_c}{\gamma_M} - \gamma_F \times E_c \geq 0 \quad \rightarrow \quad R_d \geq E_d \tag{1}$$

where R_c and R_d are the characteristic and design values of the material resistance respectively, E_c and E_d are the characteristic and design values of the load effect, γ_M is the partial factor for the resistance and γ_F is the partial factor for the load effect. In the Eurocode system for fatigue design of bridges (EN1991-2, 2003) (EN1993-1-9, 2005), a partial factor larger than 1 is recommended only on the resistance side of the limit state (γ_{Mf}) and the partial factor on the load side (γ_{Ff}) is recommended as 1. The factor γ_{Mf} depends on the choice of fatigue assessment method as well as the consequences of failure. For the safe-life fatigue assessment method and a structure with high consequence of failure, γ_{Mf} is recommended as 1.35. In this paper, the adequacy of these partial factors is investigated in a probabilistic approach where all influential random variables are attempted to be modelled as close to reality as possible. For this purpose, several structural schemes are defined and for each one of them, two structural details are designed to meet the Eurocode's safety requirements. Subsequently, the reliability of each designed case at the end of service life (100 years) is evaluated by the probabilistic approach and is compared with the safety requirement of the Eurocode (EN1990, 2002). Partial factors are calibrated in such a way that the deterministic design with partial factors arrives at the target reliability level.

2 METHODS

To consider the effects of the shape and length of the influence lines on the reliability analysis outcomes, four different structural schemes (Figure 1) with three different span lengths (L = 5 m, 20 m and 100 m) are considered. These are the same structural schemes which have been used for calibration of fatigue load models in Eurocode. Furthermore, two common structural details in

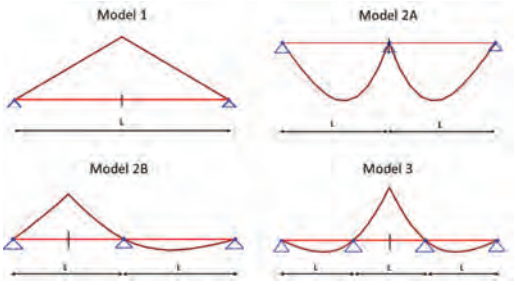


Figure 1. Structural schemes used in the analyses.

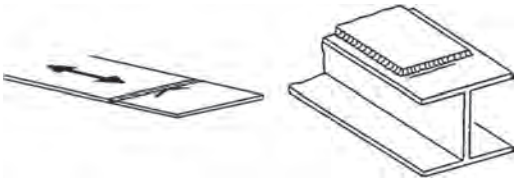


Figure 2. Left: Transverse butt welded joint (Category 80), Right: Cover plate (Category 50).

bridges, transverse butt weld joint and cover plate (Figure 2), are chosen to study the effect of the different S-N curves on the partial factors.

2.1 *Deterministic approach*

The first step in the reliability analysis is to design a structure for fatigue according to the Eurocode system. For this purpose, the fatigue load model 4 (FLM4) for road bridges has been applied on the selected structural scheme (influence line) and for a fictitious cross section (section modulus) the characteristic stress spectrum will be obtained by using the rain flow cycle counting method. FLM4 is selected for this purpose because it is the most accurate fatigue load model in the Eurocodes. It consists of a set of five heavy vehicles, each having a certain axle configuration, load distribution and fraction of the total traffic volume. These are pulled over the influence line, resulting in a stress history. A rain flow counting procedure subsequently provides the characteristic stress spectrum, which is multiplied by a partial factor γ_{Ff} so as to arrive at the design stress spectrum. The design stress spectrum is to be compared to the fatigue resistance. However, because the recommended value of γ_{Ff} is set equal to 1, the characteristic and design spectra are the same. The cumulative stress spectrum obtained for the structural scheme 3 with span length of 100 m is shown with solid black curve in Figure 3.

The fatigue resistance is a term referring to the capability of a specific structural detail to withstand the repetitive loads. This feature can be studied by laboratory tests and be presented by the so called S-N curve. The number of the load cycles (N) that the detail can resist under repetitive loading with stress range $\Delta\sigma$ is recorded during the experiment and the S-N curve is the curve fitting these points. In Eurocode (EN1993-1-9, 2005), this curve is tri-linear in log-log scale and can be presented by the following equation;

$$\log(N) = \log(a_i) - m_i \log(\Delta\sigma) \tag{2}$$

where a – the N-axis-intercept – and m – the negative inverse slope – are properties depending on the detail type and the index i indicates two different branches of the trilinear curve. The first line (i = 1) descends with $m_1 = 3$ until the point known as the constant amplitude (CA) fatigue limit (CAFL) which is defined at $N = 5 \cdot 10^6$ cycles. To consider the effect of the variable amplitude (VA) loading, the second line (i = 2) is extended from the CAFL with $m_2 = 5$ until the fatigue cut-off limit at $N = 10^8$ cycles. Damage below this cut-off value is ignored. The characteristic values of a_1 and a_2 can be calculated using (EN1993-1-9, 2005) for each detail type.

The characteristic S-N curves are divided by γ_{Mf} to obtain the design curves. These S-N curves are shown in blue in Figure 3 for the cover plate.

Having the design stress spectrum and design S-N curves, the fatigue damage (D) can be calculated using the Palmgren Miner damage accumulation rule (Miner, 1945). According to this rule, all stress cycles cause proportional fatigue damage which is linearly additive:

$$D_n = \sum_i d_i = \sum_i n_i / N_i \tag{3}$$

where D_n is the damage due to $n = \sum_i n_i$ cycles, d_i is the damage caused by all stress cycles n_i in the design stress spectrum that have the same range $\Delta\sigma_i$ and N_i is the number of cycles to failure for

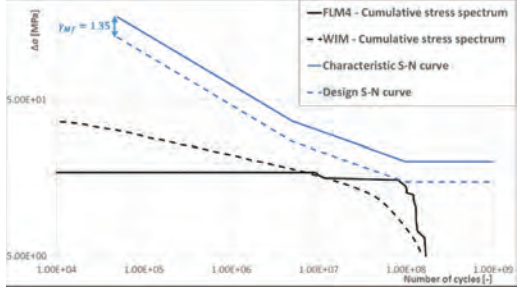


Figure 3. Deterministic cumulative stress spectra and S-N curves for cover plate.

that same stress range obtained from the S-N relation in equation (2).

For the structures designed by Eurocode's regulations, D_n should be equal to or smaller than 1 at the end of life. By adjusting the section modulus in such a way that D_n is equal to 1, a structure is obtained that just meets the requirement for fatigue.

In the next step, to simulate the reality more accurately, a Weigh In Motion (WIM) data set belonging to a main Dutch highway (A16) is used instead of FLM4. WIM is a traffic measurement system which is able to record the speed, the number of axles, the axle loads and the axle distance of a passing vehicle as well as the distance between consecutive vehicles. In the considered WIM dataset, these traffic properties are recorded for a traffic flow in one direction and for the duration of one month. Thus, to simulate the traffic flow for the structure's entire service life of 100 years, the recorded number of $2 \cdot 10^5$ heavy vehicles should be multiplied by 1200 if trends are absent. The analysis is proceeded by applying the measured traffic on the previously adjusted influence line for each structural scheme and detail. This resembles the situation when a bridge is designed following Eurocode system and is loaded by actual traffic. The stress spectra for actual traffic are calculated as explained for FLM4 and in this paper are referred to as 'WIM' spectra. The WIM cumulative stress range spectrum for the case of structural scheme 3 with span length of 100 m is shown with dashed black curve in Figure 3. The stress range spectrum of both load models are indicated in Figure 3 for the structural scheme 3 and span length of 100 m.

2.1.2 *Probabilistic approach*

There are several sources of uncertainty on the load side such as dynamic amplification factor (DAF), trend amplification factor (t) and load effect model uncertainty (B). On the material resistance side, uncertainties can be considered by scatter of the fatigue test data. In the following, these random variables are described in more detail.

A DAF should be introduced to compensate for the absence of the dynamic vehicle-bridge interaction in the WIM database. Theoretical studies to determine the DAF are usually aimed at the ultimate limit state and give an exaggerated effect for a fatigue assessment. In practice, this factor can therefore be calculated as the ratio between the maximum stress recorded at the crossing of a test vehicle with high speed (e.g. 80 km/h) and with low speed (e.g. 20 km/h). Based on the strain gauge measurements carried out on some Dutch motorway bridges, DAF in this study is assumed to be distributed according to Table 1. However, there is room for improvement of this distribution by collecting more measurement data.

The trend amplification factor takes into account possible change in load over time. The design life of important bridges is usually 100 years and in that period vehicle's number and weight can change significantly. Clearly, there is a large uncertainty in estimating trends and nevertheless, it should be estimated based on either the assumptions or extrapolation of very limited available data. The Dutch standard for assessment of existing structures (NEN8701, 2011) suggests an increase of 20% on axle loads in 100 years, i.e. a linear trend in time with an average annual increase of 0.2% with respect to the design year (Figure 4). An uncertainty over this trend increasing with time with a standard deviation of 0.05 after 100 years is assumed. The linear trend is approximately equivalent to an average increase of 10% in axle loads over the entire life with a standard deviation of 9%. The annual loads are thus multiplied by factor (t) which is statistically distributed as

Table 1. Distribution of random variables.

Variable		Distribution	*mean*	STD
DAF	Dynamic amplification factor	Log-normal	1	0.05
t	Trend amplification factor	Normal	1.1	9E-4
B	Load effect model uncertainty	Log-normal	1	0.1
$log_{10}(a_1)^*$	Material parameter-1st line (Butt weld joint)	Normal	12.42	0.25
$log_{10}(a_2)^*$	Material parameter-2nd line (Butt weld joint)	Normal	16.07	0.32
$log_{10}(a_1)^*$	Material parameter-1st line (Cover plate)	Normal	11.69	0.18
$log_{10}(a_2)^*$	Material parameter-2nd line (Cover plate)	Normal	15.02	0.3
D_{cr}	Critical fatigue damage	Log-normal	1	0.3

*a_1 and a_2 are fully correlated.

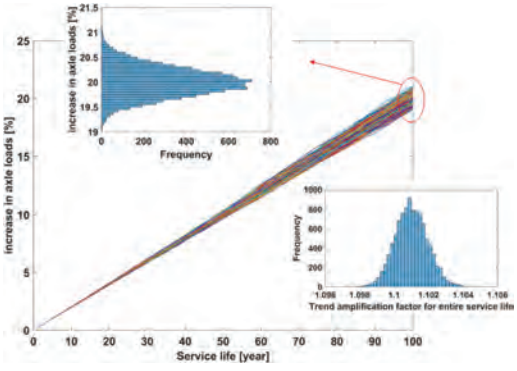


Figure 4. Trend amplification factor distribution.

shown in Figure 4 and is presented in Table 1. A possible change in number of vehicles is not considered. Background of this assumption is that the considered highway—containing 3 lanes per traffic direction—is believed to have reached is maximum capacity. A higher number of vehicles is expected to result into significant traffic jam and thus no larger number of passing vehicles.

Section 2.1.1 demonstrates that an influence line is required to determine the stress ranges. In practice, such an influence line is subtracted from a structural model of the bridge. The model may contain approximations and errors, so this influence line might be different from reality. The uncertainty related to this issue is known as the load effect model uncertainty. By comparing the calculated influence line with the influence line derived from the measurement, this uncertainty can be evaluated. In this study, the JCSS (JCSS, 2001) recommendation is used for the distribution of the load effect model uncertainty (Table 1).

The fatigue resistance is known for its significant scatter. This is mainly due to the uncontrollable differences in test samples. For simplicity, the slopes of the fitted lines m_1 and m_2 are assumed to be constant and the scatter is presented by the random variables a_i in Equation (2). EN 1993-1-9 includes the characteristic lines having 95% probability of survival and the scatter is not mentioned. Therefore, in this research, the standard deviations of the logarithm of $a_i (Log(a)_{STD,i})$ are taken from the British standard (BS7608, 2014). Starting with the characteristic values a_1 and a_2 from Eurocode (EN1993-1-9, 2005), their mean values can be obtained as;

$$Log(a)_{mean,i} = Log(a)_{characteristic,i} + 1.645 Log(a)_{STD,i} \quad (4)$$

Having the distribution parameters for all random variables, reliability analysis can be performed to evaluate the safety status of the designed details at the end of the service life. Figure 5 shows the WIM cumulative stress spectrum and its lower (5% fraction) and upper (95% fraction) bounds for the structural scheme 3 with span length of 100 m as well as the mean S-N curve and S-N curves with 5% and 95% probability of survival for a cover plate detail including all random variables introduced above.

2.2 *Reliability analysis*

The limit state function is defined as;

$$g(\boldsymbol{X}) = D_{cr} - D_n(\boldsymbol{X}) \quad (5)$$

where the random variable D_{cr} is the critical damage, D_n is the fatigue damage at the end of life and $\mathbf{X}$ is the vector of all previously mentioned random variables. D_{cr} follows a lognormal distribution with parameters according to Table 1 (JCSS, 2001). Failure can be defined as the situation wherein $g(\boldsymbol{X}) < 0$. Therefore, probability of failure is defined as;

$$P_f = P\left[g(\boldsymbol{X}) < 0 \right] = \int_{g(x)<0} f_x(\boldsymbol{X})\, dx \quad (6)$$

where $f_x(\boldsymbol{X})$ is the multivariable probability density function of $\mathbf{X}$. First Order Reliability Method (FORM) is used to approximate the integral in equation (6). FORM, developed by (Hasofer & Lind, 1974) is based on the idea that in the standard normal space, the reliability index, β, is the shortest distance from origin to the limit state surface $g(\boldsymbol{U}) = 0$, where $\mathbf{U}$ is the vector of normalized random variables. The point $\mathbf{u}^*$ on the limit state surface which has the shortest distance to the origin is known as the design point or the most probable failure point. Finding the design point is an iterative process and once calculated, the reliability index is

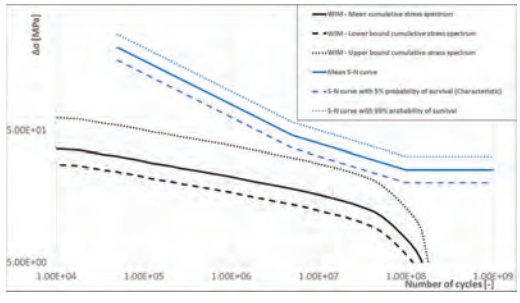


Figure 5. Probabilistic cumulative stress spectra and S-N curves for cover plate including all random variables.

obtained. Another direct outcome of the FORM is the sensitivity factor of each random variable, α_i, which is a measure for the relative importance of the standard deviation of the ith random variable to the reliability index. It can be calculated as;

$$\alpha_i = -u_i^* / \beta \tag{7}$$

Sensitivity factors are used for derivation of partial factors according to the following equation;

$$\gamma_{fi} = \frac{\left[1 - \alpha_i \beta_t STD(E_i)\right] \times mean(E_i)}{characteristic(E_i)} \tag{8}$$

where β_t is the target reliability index (Section 2.1.4), E_i is the ith random variable and γ_{fi} is the required partial factor for E_i.

Fatigue design factor (FDF) is defined as the multiplication of partial factors of resistance and load effect random variables;

$$FDF = \prod_i \gamma_{Mfi} \times \prod_j \gamma_{Ffj} = \gamma_{Mf} \times \gamma_{Ff} \tag{9}$$

The accuracy of FORM is checked by a Crude Monte Carlo (CMC) simulation with $4 \cdot 10^6$ number of samples. In this method, the probability of failure can be calculated as the ratio of the number of failure cases $\left(g(X) < 0\right)$ over the total number of CMC samples. Having the probability of failure, the reliability index can be calculated as;

$$\beta = -\Phi^{-1}\left(P_f\right) \tag{10}$$

where $\Phi^{-1}(.)$ is the inverse cumulative normal distribution function.

2.3 *Target reliability index*

Target reliability index (β_t) is the answer to the question "What level of safety is sufficient?". Several factors play a role in this answer, including the consequence of failure in terms of both loss of human life and economical aspects, the required cost for improving safety, the structure's planned service life and the type of considered limit state. In Eurocode, β_t for fatigue limit state is ranging between the target values of the reliability indices for the ultimate and serviceability limit states. For the structures such as bridges which can be categorized into the consequence class 3 (EN1990, 2002), with details having large consequences of failure and that are designed according to the safe life concept, the target reliability index for fatigue is set equal to the ultimate limit state value of 4.3.

3 RESULTS AND DISCUSSIONS

Having in mind that the results of this study are widely dependent on the assigned distributions of the random variables, the reliability indices at the end of life for both cover plate and transverse butt welded joint and for the all structural schemes are shown in Figure 6. For all considered cases, the reliability index is lower than the target value when applying the recommended values of partial factors. In addition, it can be observed that the value of β depends on the shape and length of the influence line. This is caused by the approximations and simplifications in the fatigue load model FLM4. Furthermore, it can be observed that the cover plate results in slightly higher reliability at the end of life than the transverse butt weld joint, due to the differences in scatter of the fatigue resistance (Table 1).

The corresponding fatigue design factors for each detail and structural scheme and under the WIM spectra (structure designed with characteristic FLM4 and loaded by WIM measured traffic) are shown in Figure 7.

For the span length of 100 m and especially for Model 2A, the values of FDF are much higher than the value resulting from the recommended

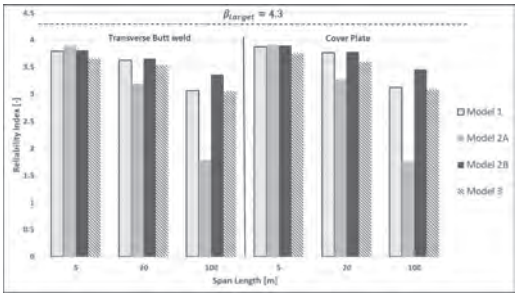


Figure 6. Reliability indices at the end of life.

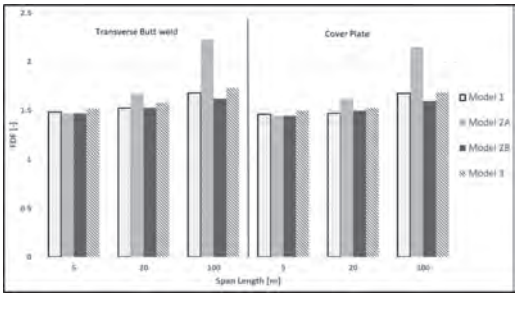


Figure 7. Required fatigue design factors to reach the target reliability index at the end of service life.

partial factors and these are not in line with the other values. The reason is the deficiency of FLM4 to simulate the real traffic flow for the bridges with very long span. In the Dutch national annex to EN 1991-2 (EN1991-2/NB, 2011), this flaw is fixed by introducing a second heavy vehicle on the bridge in 20% of the number of heavy vehicle crossings for continuous positive or negative influence lengths larger than 60 m. The center to center distance between these two vehicles is set as 50 m.

Repeating the reliability analysis with FLM4a according to the Dutch NA, the values of FDF for the span length of 100 m are reduced to the values presented in Figure 8. These values can still be as high as 1.92 for model 2 A but usually fatigue is not the dominant failure mode at location of the intermediate support of a two span bridge. This example—and the still significant differences between the various structural schemes—demonstrate the necessity of updating the fatigue load models in the Eurocode.

Now that the FDF is determined, the next step is to distribute this over γ_{Mf} and γ_{Ff}. A direct determination of the partial factors based on the FORM sensitivity factors is problematic, because:

- for fatigue the three-branch S-N curve uncertainty is related to the number of cycles and not to the stress range, and;
- the entire stress range spectrum influences the reliability instead of a single stress value. As demonstrated in Figure 3, the shape of the stress range spectrum of FLM4 is not in agreement with that of the WIM data.

For this reason, an intermediate step is taken where the section modulus is determined using the S-N curve as resistance model and the WIM database as load model. In this case, the uncertainty on the load side is captured in the multiplication of the DAF, t and B. This results in a single multiplication factor for the stress spectrum and hence

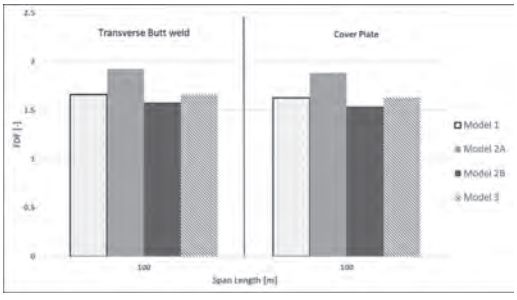


Figure 8. Required FDF after modification of FLM according to the Dutch national annex.

Table 2. Sensitivity factors of random variables.

Variable	DAF	t	B	$Log_{10}(a_1)$	$Log_{10}(a_2)$	D_{cr}
α	–0.268	–0.0044	–0.534	0.0739	0.728	0.329
$\sum \alpha^2$	0.357			0.643		
γ_{Ffi}*	1.06	1.1	1.23			

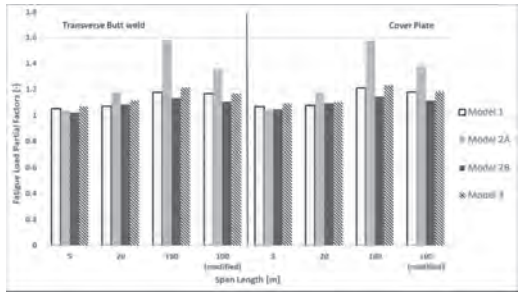


Figure 9. Fatigue load partial factors.

enables the determination of a modified partial factor for the load side, γ_{Ff}*, via the influence factors α_t, α_{DAF} and α_B. Note that γ_{Ff}*, considers the uncertainties related to the load side but not the inaccuracy of FLM 4 to represent the WIM measurement. The same analysis gives the modified fatigue design factor FDF* and the partial factor on the resistance side can now be determined as γ_{Mf} = FDF*/γ_{Ff}*. Finally, the partial factor on the load side including the difference between the actual WIM data and the FLM4 model is determined with the FDF of the original analysis of Figures 7 and 8, via γ_{Ff} = FDF/γ_{Mf}.

As expected, the factors FDF*, α_t, α_{DAF}, α_B and γ_{Mf} are almost independent on the structural models—variation of less than 2%. This demonstrates the consistency of the approach.

Table 2 presents the sensitivity factors of all random variables as well as γ_{Ffi}* values for the transverse butt welded joint located on model 2B with span length of 20 m.

For the case of transverse Butt welded joint, the value calculated for γ_{Mf} is 1.42 (i.e. 5% higher than Eurocode's recommendation) and for the case of cover plate it is equal to 1.37 (1.8% higher than Eurocode's recommendation). The calculated values of γ_{Ff} are presented in Figure 9. It can be observed that for all considered cases, γ_{Ff} has a value larger than Eurocode's recommended value of 1. Considering all studied structural schemes with all assigned span lengths and for both structural details, γ_{Ff} takes the average value of 1.16 with standard deviation of 0.13.

4 CONCLUSION

In this paper the necessity of increasing the values of partial factors for fatigue traffic load models in Eurocode to design a steel bridge with the safe-life assessment approach has been studied. For this purpose, two different structural details located in twelve structural schemes have been designed to meet the Eurocode's fatigue safety requirements. After studying the random variables which influence the fatigue limit state and assigning a distribution to each one of them, each designed detail has been loaded by measured traffic and reliability analysis has been performed to evaluate the safety status of the detail at the end of its life. It has been observed that in all cases, the reliability index is lower than the minimum acceptable value set by the standard. The reliability depends on the structural detail as well as the shape and length of influence line. The study has been proceeded by calculating the values of the partial factors that are needed to reach the standard's target reliability level. The calculated fatigue resistance partial factors are 1.37 and 1.42 for the two steel details considered and these values are slightly higher than the recommended factor of 1.35 in the Eurocodes. The main finding of this study is that the value of fatigue load partial factor recommended by the Eurocodes is too low for fatigue load model FLM4. For all studied structural schemes and for both chosen details, the required load partial factors γ_{Ff} are higher than the factor 1 as recommended by the Eurocodes. The factor γ_{Ff} strongly depends on the shape and length of influence lines and varied between 1.02 and 1.58 for the considered systems.

The values presented should not be considered as 'the truth' because they are subjected to the choices of the distributions of load related variables. Nonetheless, it is clear that the recommended partial factor of 1 in the current Eurocodes is too small and the large differences in the calculated factors demonstrate the necessity of improving the fatigue load models in the Eurocode.

A sensitivity analysis to study the effect of the most influential load related random variables on the partial factors is another outcome of this study. It has been found that load effect model uncertainty has a large impact on the results, followed by the load trend amplification factor. Dynamic amplification factor has a lower influence than the other variables. Based on this analysis, future studies can be planned to evaluate the distribution parameters of these random variables.

ACKNOWLEDGEMENTS

The authors would like to thank the Dutch infrastructure asset owners ProRail and Rijkswaterstaat as well as the Netherlands Organization for Applied Scientific Research (TNO) for their supports.

REFERENCES

BS7608, 2014. *BS 7608, BSI Standards Publication: Guide to fatigue design and assessment of steel products,* s.l.: s.n.

EN1990, 2002. *EN 1990, European Committee for Standardization. Eurocode: Basis of structural design,* s.l.: s.n.

EN1991-2/NB, 2011. *Nationale bijlage bij NEN-EN 1991-2+C1. Eurocode 1: Belasting op constructies—Deel 2: Verkeerbelasting op bruggen,* s.l.: s.n.

EN1991-2, 2003. *EN 1991-2, European Committee for Standardization. Eurocode 1: Actions on structures,* s.l.: s.n.

EN1993-1-9, 2005. *EN1993-1-9, European committee for standardization. Eurocode 3: design of steel structures,* s.l.: s.n.

Euler, M. U.K., 2013. *Statistical intervals for evaluation of test data according to Eurocode 3 part 1-9,* s.l.: s.n.

Hasofer A., L.N., 1977. *An exact and invariant first order reliability format,* s.l.: Proc. ACSE, J. Eng. Mech. Div.

Hasofer, A. & Lind, N., 1974. *An exact and invariant first order reliability format,* s.l.: Proc. ASCE, J. Eng. Mech. Div.

JCSS, 2001. *JCSS, Joint committee of structural safety. Probabilistic model code,* s.l.: s.n.

Miner, M., 1945. *Cumulative Damage in Fatigue, Journal of Applied Mechanics,* s.l.: Journal of Applied Mechanics.

NEN8701, 2011. *NEN 8701, Assessment of existing structures in case of reconstruction and disapproval actions,* s.l.: s.n.

System reliability

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Reliability analysis in the presence of Aleatory uncertainty

L.G. Crespo, S.P. Kenny & D.P. Giesy
Dynamic Systems and Control Branch, NASA Langley Research Center, Hampton, Virginia, USA

ABSTRACT: This paper proposes a method for characterizing a system's response given data. This response might prescribe the failure domain needed to assess the reliability of such a system. We focus on the case in which not all uncertain parameters affecting the response are observable and the measurements are corrupted by noise. In this setting, the system response is not given by a function but instead by a random process. In this paper we use a staircase random predictor model to characterize such a process. Consequently, the resulting failure probability is not a scalar but a random variable. This variable accounts for the aleatory contributions of the model-form uncertainty and the measurement noise affecting the system's response. Furthermore, we propose a framework that enables trading off the system's performance, measured by the extension of an acceptable range of operating conditions, against the system's reliability, measured by an admissible range of failure probabilities. The risk incurred by such a practice is ignoring a (small) percentage of the predicted worst-case system responses. These ideas are illustrated by performing the reliability analysis of an aeroelastic structure subject to flutter instability. Furthermore, this paper puts forth a means to quantify the error resulting from having a dataset of limited size when performing the above analysis.

1 INTRODUCTION

Metamodeling (Simpson, Peplinski, Koch, & Allen 2001) refers to the process of creating a mathematical representation of a phenomenon based on input-output data. This paper uses a metamodeling technique for constructing computational models describing the distribution of a continuous output variable. These models are called *Random Predictor Models* (RPMs) because the predicted output corresponding to any given input is a random variable. One common example of an RPM is a Gaussian Process (GP) model (Rasmussen & Williams 2006). In contrast to GP models, which only lead unimodal and symmetric responses, we focus on RPMs having a bounded support set and prescribed values for the first four moments. The manipulation of these functions enables the generation of predictors that accurately describe possibly skewed and multimodal responses typical of many physical phenomena.

This paper extends the developments on RPMs made by the authors to account for sampling error in the moment estimates. As an application, we use RPMs for the reliability and risk analysis of a flexible structure. To make the paper self-contained, essential concepts are presented. Supplemental information is available at (Crespo, Giesy, & Kenny 2017a).

2 PROBLEM STATEMENT

A DGM is postulated to act on a vector of input variables, $x \in \mathbb{R}^{n_x}$, to produce an output, $y \in \mathbb{R}^{n_y}$. In this article the focus will be on the single-output ($n_y = 1$) multi-input ($n_x \geq 1$) case. The dependency of the output on the input is arbitrary. This covers the case in which y is a function of x with all components of x available (so there is only one output value for each available input), the case in which y is a function of x but not all components of x are available (so there might be infinitely many outputs for each measured input, and the case in which y is an arbitrary random process of x. Assume that N Independent and Identically Distributed (IID) input-output pairs are obtained from a stationary DGM, and denote by $\mathbb{D} = \left\{x^{(i)}, y^{(i)}\right\}$, for $i = 1, \quad N$ the corresponding data sequence. The main objective of a predictor model is to generate a computational representation of a DGM based on the data in $\mathbb{D}$. Two types of predictors will be used hereafter. An Interval Predictor Model (IPM) yields a bounded interval of output values at any value of the input. The desired IPM is a narrow interval wherein unobserved data will fall with high probability. Conversely, a Random Predictor Model (RPM) yields a random variable at any value of the input. The desired RPM accurately describes the distribution governing the DGM.

3 PRELIMINARIES

Consider the continuous random variable z with support set $\Delta_z = [z_L, z_U]$ Probability Density Function (PDF) $f_z : \Delta_z \subset \mathbb{R} \to \mathbb{R}^+$, and Cumulative Distribution Function (CDF) $F_z : \Delta_z \to [0,1]$. Denote by m_r the r-th central moment of z, which is defined as

$$m_r = \int_{\Delta_z} (z-\mu)^r f_z(z)\, dz, r = 0,1,2,\ldots \tag{1}$$

where μ is the expected value of z. Note that $m_0 = 1$, $m_1 = 0$, m_2 is the variance, m_3 is the third-order central moment, and m_4 is the fourth-order central moment. Where reference is made to the r-th moment of a random variable, we assume that the corresponding integral in (1) converges for that distribution.

The random variables of interest will be constrained to have a bounded support set and given values for μ, m_2, m_3, and m_4. The bounded support constraint is $\Delta_z \subseteq \Omega_z$, where $\Omega_z = [\underline{z}, \overline{z}]$ with $\overline{z} \geq \underline{z}$ given, whereas the moment constraints are given by (1). The parameters of these constraints will be grouped into the variable $\theta_z \in \mathbb{R}^6$ given by

$$\theta_z = [\underline{z}, \overline{z}, \mu, m_2, m_3, m_4]. \tag{2}$$

Any random variable z having a support set contained by $[\underline{z}, \overline{z}]$ with moments μ, m_2, m_3, and m_4 must satisfy the feasibility conditions $g(\theta_z) \leq 0$ given in (Crespo, Giesy, & Kenny 2017b). The realizations of θ satisfying these conditions constitute the θ-feasible domain, Θ, defined as

$$\Theta = \{\theta : g(\theta) \leq 0\}. \tag{3}$$

A member of Θ will be called θ-feasible. Determining membership in Θ is a distribution-free assessment applicable to possibly infinitely many random variables satisfying the desired constraints.

A particular class of random variables that can realize most of Θ is proposed in (Crespo, Giesy, & Kenny 2017b). This class is called *Staircase* because the PDF of its members is piecewise constant over bins of equal width. Staircase variables, are calculated by solving the convex optimization program

$$\min_{\ell \geq 0} \{ J(\theta, n_b) : A(\theta, n_b)\ell = b(\theta), \theta \in \Theta \}, \tag{4}$$

where J is the cost function used for optimization, n_b is the number of bins partitioning Ω_z, $\ell \in \mathbb{R}^{n_b}$ are the values of the PDF at the bin centers, and $A\ell = b$ are moment matching constraints. Staircase variables enable modeling complex phenomena efficiently. Staircase variables will be denoted as

$$z \sim S_z(\theta_z, n_b, J). \tag{5}$$

When the cost is chosen to be the entropy, E, we obtain a maximal entropy staircase variable. The points $\theta \in \Theta$ for which a staircase variable with n_b bins exists constitutes the staircase feasible domain, $\mathcal{S}(n_b)$. As expected, $\mathcal{S}(n_b) \subset \Theta$. Supplemental information is available at (Crespo, Giesy, & Kenny 2017b).

3.1 *Staircase estimation*

This section focuses on the estimation of the hyperparameter θ_z of a staircase variable S_z from the samples $z^{(1)}, \ldots z^{(N)}$. Expert opinion should be used to prescribe n_b and the bound to the support Ω_z. Whereas Ω_z must[1] contain $\dot{\Delta} = [\min\{z^{(j)}\}, \max\{z^{(j)}\}]$, the moments μ, m_2, m_3, and m_4 can be chosen to be the sampling moments $\dot{\mu}, \dot{m}_2, \dot{m}_3$ and $\dot{m}_4$. The developments that follow account for the error incurred by using these estimates.

Lets first focus on the error incurred by using $\Omega_z \supset \dot{\Delta}$. Finite values of N often make $\dot{\Delta}$ a subset of the support set Δ. Scenario optimization (Campi & Garatti 2008) enables bounding the probability of the tails of the PDF of z extending beyond $\dot{\Delta}$. In particular,

$$\mathbb{P}\left[z \notin \dot{\Delta} \right] = \kappa q^\top \ell \leq \hat{\epsilon}, \tag{6}$$

where

$$\hat{\epsilon} = 1 - e^{\log(\beta)/(N-1)}, \tag{7}$$

β is the confidence parameter, and $q \in \mathbb{R}^{n_b}$ is:

$$q_i = \begin{cases} 1 & \text{if } (z_i, z_{i+1}] \subset \Omega_z \setminus \dot{\Delta}, \\ 0 & \text{if } (z_i, z_{i+1}] \subset \dot{\Delta}, \\ \dfrac{min\{z^{(j)}\} - z_i}{z_{i+1} - z_i} & \text{if } z_i \leq \min\{z^{(j)}\} \leq z_{i+1}, \\ \dfrac{z_{i+1} - max\{z^{(j)}\}}{z_{i+1} - z_i} & \text{otherwise.} \end{cases}$$

Equation (6) ensures that the staircase variable conforms to the tails probability bound from convex scenario theory.

We now focus on the error in the sample moments. This error, fully prescribed by the corresponding sampling distributions, can be quantified

[1] Sampling estimates will be denoted with a dot-superscript.

by using bootstrapping techniques. Alternatively, an asymptotic approximation to the sampling distribution, grounded in the central limit theorem, can be used instead. Such a distribution is given by the normal variables

$$\begin{aligned}
\mu &\sim \mathcal{N}_{\mu}\left(\dot{\mu}, \sqrt{\frac{\dot{m}_2}{N}}\right), \\
m_2 &\sim \mathcal{N}_{m_2}\left(\dot{m}_2, \sqrt{\frac{\dot{m}_4 - \dot{m}_2^2}{N}}\right), \\
m_3 &\sim \mathcal{N}_{m_3}\left(\dot{m}_3, \sqrt{\frac{t_1}{N}}\right), \\
m_4 &\sim \mathcal{N}_{m_4}\left(\dot{m}_4, \sqrt{\frac{t_2}{N}}\right),
\end{aligned} \tag{8}$$

conditional on $\theta \in \Theta$, where $\dot{m}_k$ for k = 5,6,7,8 are the sample fifth, sixth, seventh and eighth central-order moments and $t_1 = \dot{m}_6 - \dot{m}_3^2 - 6\dot{m}_4\dot{m}_2 + 9\dot{m}_2^3$, $t_2 = \dot{m}_8 - \dot{m}_4^2 - 8\dot{m}_5\dot{m}_3 + 16\dot{m}_2\dot{m}_3^2$. These expressions correspond to an arbitrarily distributed variable z for a sufficiently large value of N (Kendall & Stuart 1969). For small values of N, bootstrapping techniques often yield a more accurate approximation.

To account for sampling error in the calculation of a staircase variable, the moment matching constraints are replaced by the polynomial inequality constraints

$$\underline{\mu} \le \mu(\ell) \le \bar{\mu}, \tag{9}$$

$$\underline{m}_2 \le m_2(\ell) \le \bar{m}_2, \tag{10}$$

$$\underline{m}_3 \le m_3(\ell) \le \bar{m}_3, \tag{11}$$

$$\underline{m}_4 \le m_4(\ell) \le \bar{m}_4, \tag{12}$$

where $\mu(\ell) = r_2\ell, m_2(\ell) = r_3\ell - \mu^2, m_3(\ell) = r_4\ell - \mu^3 - 3\mu m_2$ and $m_4(\ell) = r_5\ell - 4\mu m_3 - 6\mu^2 m_2 - \mu^4$ are the moments realized by the staircase variable, r_i is the i-th row vector of A in (4), and the moment bounds are the $1 - \alpha$ confidence intervals corresponding to the sampling distributions, e.g., $\dot{\mu} - 1.96\sqrt{\dot{m}_2/N} \le \mu(\ell) \le \dot{\mu} + 1.96\sqrt{\dot{m}_2/N}$ for a 95% confidence interval. Note that the box of moments defined by (9–12) might not be fully contained in Θ.

Sampling error can be accounted for by solving for a maximal entropy staircase variable constrained by Equations (6), and (9–12). The resulting staircase variable will not account for the manner in which the sampling distribution allocates probability within the box of moments. This consideration can be taken into account by using a likelihood-dependent cost, such as

$$J(\ell) = -E(\ell) - \log\{L(\ell)\}, \tag{13}$$

where $L = \mathcal{N}_{\mu(\ell)}\mathcal{N}_{m_2(\ell)}\mathcal{N}_{m_3(\ell)}\mathcal{N}_{m_4(\ell)}$ is the likelihood function corresponding to the sampling distribution.

In summary, staircase variables provide (i) the ability to represent a wide range of density shapes, (ii) the ability to represent most of the feasible space Θ, (iii) the ability to account for the effects of having a limited number of observations, and (iv) the low-computational cost required to efficiently perform various uncertainty quantification tasks.

4 PREDICTOR MODELS

4.1 *Interval predictor models*

This section presents a means to calculate the support set of an RPM. This will be carried out by finding a baseline IPM using the same data sequence $\mathbb{D}$ that will be used to construct the RPM.

An IPM assigns to each instance vector $x \in X \subseteq \mathbb{R}^{n_x}$ a corresponding outcome interval in $Y \subseteq \mathbb{R}$. That is, an IPM is a set-valued map, $I_y : x \to I_y(x) \subseteq Y$, where $I_y(x)$ is the prediction interval. Depending on context, the term IPM will refer to either the function I_y or its graph $\{(x,y) : x \in X, y \in I_y(x)\}$ in $X \times Y$. A nonparametric IPM is given by

$$I_y(x) = \left\{\left[\underline{y}(x), \bar{y}(x)\right], \bar{y}(x) \ge \underline{y}(x)\right\}. \tag{14}$$

where the functions $\underline{y}(x)$ and $\bar{y}(x)$ are the lower and upper boundaries of the IPM respectively. A parametric IPM is obtained by associating to each $x \in X$ the set of outputs y that result from evaluating the function $y = M(x,p)$ at all values of p in the set P, so

$$I_y(x,P) = \{y = M(x,p), p \in P\}. \tag{15}$$

Attention will be limited to the case in which the output depends linearly on p and arbitrarily on x, so $y = p^T\varphi(x)$, where $\varphi(x) \in \mathbb{R}^{n_p}$ is an arbitrary basis. Several IPM types can be calculated within this framework. In thispaper we use the technique used in (Crespo, Kenny, & Giesy 2016).

Example 1: Next we use an analytically described DGM of which we have full knowledge. Figure 1 shows the corresponding one percentiles. Note that the support set, moments and modality of the DGM are strongly nonlinear functions of the input. The high concentration of percentile lines at the edge of the support indicates a bimodal structure. N = 1000 IID observations were drawn from the DGM

to form the data sequence $\mathbb{D}$. This data, shown in Figure 1, was then used to construct an IPM with $n_p = 20$ terms and a Gaussian basis structure. That is, $\varphi_i(x) = e^{-(x-u_i)^2/v_i}$, where $u_i \in \mathbb{R}$ is a center and $v_i \in \mathbb{R}$ is a length-scale parameter, for $i = 1, \ldots, n_p$. Centers are uniformly distributed over $X = [-\pi, \pi]$, whereas the length scale parameters are made all equal to $\pi/5$. Figure 1 also shows the resulting IPM. Notice that the IPM tightly encloses all the data as intended. An IPM will be used to describe the support of the DGM, whereas RPMs, introduced next, will be used to characterize its distribution.

4.2 *Random predictor models*

An RPM is a mapping that assigns to each input vector $x \in X$ a corresponding random variable $R_y(x)$. A non-parametric RPM is the random variable-valued map given by

$$R_y(x) = \left\{ f_{y(x)}(y(x)), y(x) \in \Delta_y(x) \right\}, \tag{16}$$

where $f_{y(x)}$ is the PDF of y at $x \in X$ having the support set $\Delta_y(x) = \left[\underline{y}(x), \bar{y}(x) \right] \subseteq Y$. By contrast, a parametric RPM is obtained by associating to each $x \in X$ the set of outputs y corresponding to all values of p described by a random vector with joint PDF $f_p(p)$ supported in Δ_p, so

$$R_y(x, f_p) = \left\{ y = M(x, p), p \sim f_p(p), p \in \Delta_p \right\}. \tag{17}$$

The RPMs used below assume a staircase structure. As such, they require prescribing an input-dependent hyper-parameter $h(x) = [n_b(x), \theta_{y(x)}]$ for all $x \in X$. Hereafter we will assume that $n_b(x)$ is a fixed constant, and focus on the prescription of $\theta_{y(x)}$. Define as

$$\tilde{\theta}_{y(x)} = \left[\tilde{\underline{y}}(x), \tilde{\bar{y}}(x), \tilde{\mu}_{y(x)}, \tilde{m}_{2,y(x)}, \tilde{m}_{3,y(x)}, \tilde{m}_{4,y(x)} \right],$$

a set of target functions prescribed according to the data, and by $\theta_{y(x)}$ the set functions realized by an RPM. An RPM that accurately represent the DGM must make $\theta_{y(x)}$ close to $\tilde{\theta}_{y(x)}$. The strategy for prescribing $\theta_{y(x)}$ according to the data sequence $\mathbb{D}$ presented in (Crespo, Giesy, & Kenny 2017b) will be used here. $\tilde{\theta}_{y(x)}$ is based on a weighted average of values $y^{(i)}$ for $x^{(i)}$ close to x.

A non-parametric RPMs with a staircase structure can be readily calculated from the target functions $\tilde{\theta}_{y(x)}$ by making the prediction at any input value $x \in X$ realize the corresponding target. Once a set of staircase-feasible target functions $\tilde{\theta}_{y(x)}$ is obtained, the RPM $S_{y(x)}\left(\tilde{\theta}_{y(x)}, n_b, J\right)$ can be readily evaluated at any value of the input. The resulting RPM will conform to the target regardless of the choice of J.

Example 2: Next we use the data sequence of the DGM in Example 1 to derive an RPM. Figure 2 shows the functions in $\theta_{y(x)}$ corresponding to the DGM (solid lines) along with the target functions $\tilde{\theta}_{y(x)}$ (dashed lines). Note that the target functions approximate the DGM well in spite of only using N = 1000 observations. The difference between the two sets of functions is caused by the limited amount of data available and by using neighboring data to calculate $\tilde{\theta}_{y(x)}$. The targets $\tilde{\theta}_{y(x)}$ in Figure 2 were used to build the RPM $S_{y(x)}\left(\tilde{\theta}_{y(x)}, 500, E\right)$. This RPM is staircase-feasible throughout X. This was also the case for values of n_b as small as 100. The plot at the top of Figure 3 shows the 1-percentiles of this RPM. This figure was generated by calculating staircase

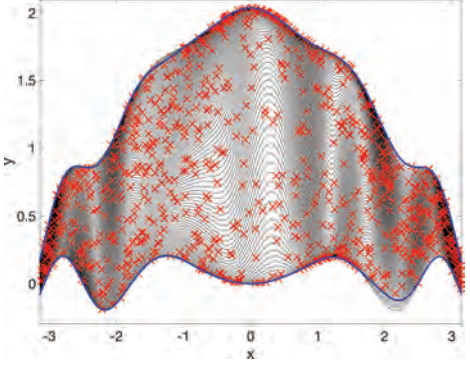


Figure 1. One-percentiles of the DGM (black lines), N = 1000 observations (red ×), and IPM limits (blue lines).

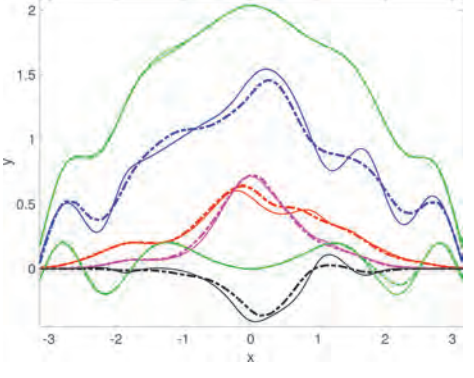


Figure 2. Support set (green), mean (blue), variance (red), third-order- (black), and fourth-order-central moment (magenta) functions corresponding to the DGM (solid) and to the target (dashed-dotted).

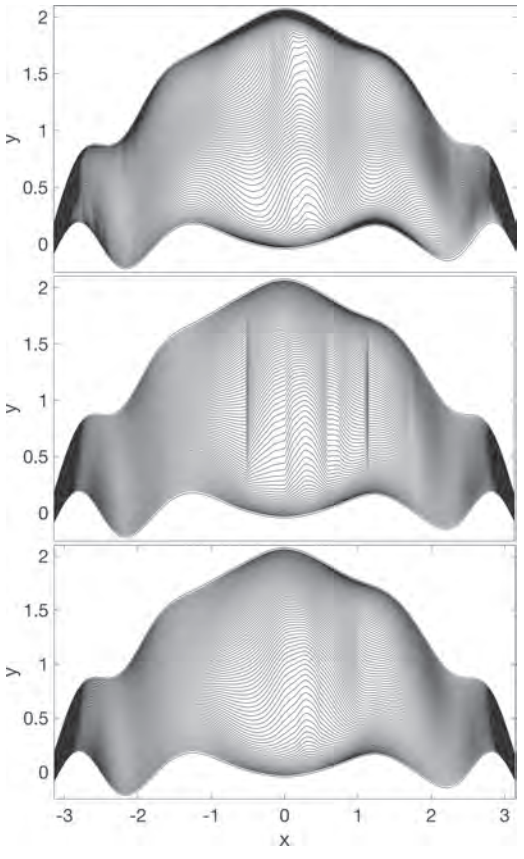


Figure 3. Moment-matching RPM (top), moment-bounded RPM of maximal entropy (middle), and moment-bounded RPM (bottom) based on Equation (13).

variables over a uniform grid of input values in X, sampling them, smoothing the corresponding empirical CDF using a Gaussian kernel (Silverman 1986), and grouping the points belonging to the same percentile line. The moment functions attained by the RPM are indistinguishable from the targets shown in Figure 2. The comparison between the DGM, shown in Figure 1, and the moment-matching RPM indicates excellent agreement despite only using $N = 1000$ data points. Note that RPM describes well the bimodal structure of the DGM by replicating the regions where probability is highly concentrated, i.e., the regions in the upper and lower limit of the support where many percentile lines coalesce. Furthermore the skewness of the probability mass in the interior of the support set follows the same oscillatory patterns present in the DGM.

Example 3: Next we study the effects of the sampling error on the empirical target $\tilde{\theta}_{y(x)}$, and on the resulting staircase RPM. The observations prescribing the target functions at x are weighted according to their separation from such a point (Crespo, Giesy, & Kenny 2017a). The weight is the greatest when the datum is at x, and it approaches zero as its separation from x increases. For the functions shown in Figure 2, the number of observations having a non-negligible weight ranges from 110 to 261. To quantify the sparsity of the dataset we define the equivalent number of observations, n_e, as

$$n_e(x) = \sum_{i=1}^{N} w\left(x^{(i)}, x\right). \tag{18}$$

The effects of the sparsity in the data will be quantified using the developments in Section 3.1. In particular, we will generate a maximal entropy RPM satisfying the constraints (6) and (9–12) for $N = n_e(x)$. The value of n_e is inversely proportional to the size of the box of moments. For this data sequence, the value of n_e ranges between 47.15 and 112. This indicates that the dataset is sparse. In contrast to the RPM at the top of Figure 3, the resulting RPM will not match moments estimated from the data but instead, it will realize moment functions bounded by their sample uncertainty. As such, we will refer to this RPM as a moment-bounded RPM. Figure 4 shows the 95% confidence intervals of the four moment functions. Note that the range of these intervals exhibits oscillations, reaching their largest spread near $x = 0$. These regions contain the moments realized by staircase variables comprising the moment-bounded RPM, which are shown as solid lines. The first three

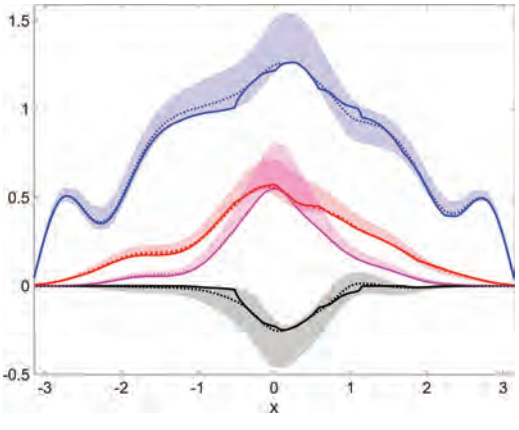


Figure 4. Optimal mean (blue), variance (red), third-order- (black), and fourth-order-central moments (magenta) corresponding to a moment-bounded RPM using maximal entropy (solid lines), Equation (13) (dotted lines), and corresponding sampling error ranges (shaded regions).

moments vary in the interior of their confidence intervals whereas the fourth moment stays on the lower limit. All but the fourth moment function take on values that vary within the intervals. The uncertainty in the sample moments increases the expected entropy $\mathbb{E}_x[E]$ of the RPM from –0.3642 to 0.7484.

Figure 3 shows the moment-matching RPM as well as moment-bounded RPMs of maximal entropy. Note that the most prominent features of the process, such as the peaks at the boundaries of the support set and the patterns of the lines in its interior, have faded in the latter predictor. Furthermore, note that the derivative discontinuities in the moment functions of Figure 4 yield derivative discontinuities in the percentile lines. Such discontinuities can be eliminated by using a Kernel smoother or by using another cost function. Alternatively, we can calculate a moment-bounded RPM having the cost function in Equation (13). The corresponding moment functions are shown as dotted lines in Figure 4. In contrast to the moments for the maximal entropy formulation, the new moments have continuous derivatives throughout X. The resulting RPM, shown at the bottom of Figure 3, exhibits smooth percentile lines. This is achieved at the expense of a minor entropy reduction to $\mathbb{E}_x[E] = 0.73$. As n_e increases, the width of the confidence intervals reduces making moment-bounded RPMs converge to the moment-matching RPM. The PDF-matching or the maximum-likelihood staircase formulations are preferable when n_e is sufficiently large.

Example 4: Next we consider the reliability analysis (Rackwitz 2001) of an airfoil subject to aeroelastic flutter (Mahler, Touze, Doare, Habib, & Kerschen 2017). During flutter the pitch and plunge dynamics are coupled yielding a self-sustaining limit cycle oscillation that might compromise the structural integrity of an aircraft. The onset of flutter depends on the free stream airflow speed v, as well as inertial, geometrical, and material properties of the wing. These parameters include the plunge and pitch stiffnesses, the aerodynamic lift, the location of the center of mass, and the location of the elastic axis. In this context, a reliability analysis seeks to quantify the probability of flutter instability (i.e., probability of failure) given probabilistic prescriptions for the parameters.

The objectives of this example are two-fold. First, we use RPMs to characterize the system response. Measurement error and model-form uncertainty make the data and the response aleatory, thereby justifying such a modeling choice. A deterministic response model along with a probabilistic description of the parametric uncertainty would enable the calculation of the failure probability. However, when the response is intrinsically aleatory, the failure probability can only be determined to lie within a range of values. We then explore the reduction in the range of failure probabilities resulting from ignoring a (small) percentage of the responses predicted by the RPM.

The stability of the system is evaluated by calculating the damping coefficient, y (i.e., the output in the context of this paper) of the time response to a given flow speed. As in (Canor, Caracoglia, & Denoel 2015), damping is related to the real part of the eigenvalues of a linear dynamic model. Non-negative values of y denote an unstable system, whereas negatives values correspond to a stable system. Hence, the failure domain is defined as

$$\mathcal{F} = \{x : y(x \mid v) \geq 0\}. \tag{19}$$

The measured output y depends on measurement errors, as well as epistemic and aleatory uncertainties. In this example we will divide the uncertain parameters into two groups. The primary group consists of measurable parameters having a strong influence on the output (e.g., stiffnesses), whereas the secondary group consist of parameters that are either weakly important, unmeasurable, or unknown to the analyst (e.g., measurement error, unsteady aerodynamics, etc.). In this study the primary group of parameters constitute the input x. Note that variations in the secondary parameters make $y(x)$ aleatory. Hence, in contrast to a standard reliability analysis for which a parametric model explicitly prescribes the dependency of the limit state on the uncertain parameters (Sun, Wang, Rui, & Tong 2017), the limit state we are aiming to identify will not be a deterministic function of all the parameters, but instead a random process depending on the primary parameters. This implies that the failure probability will range on an interval whose spread depends upon the manner in which the RPM describing $y(x)$ crosses zero (failure boundary).

For simplicity in the analysis, x will be assumed to be a single non-dimensional parameter describing the ratio of the pitch and plunge stiffnesses. Independent input-output pairs $\mathbb{D} = \{x_i, y_i\}$ for $i = 1, \quad N$, were obtained by simulating the flutter dynamics of N = 2500 individual airfoils. The randomness in the output stems from variability of not only x but also of the secondary aerodynamic and structural parameters affecting y. It is expected that such parameters mostly vary within 10% of their nominal value.

Figure 5 shows the input-output data for the free stream airflow speeds v = 0.75, v = 0.79, v = 0.83, v = 0.87, v = 0.91, v = 0.95. Each chosen airfoil was evaluated at these speeds. Note that the number of data points falling into the failure domain $y \geq 0$

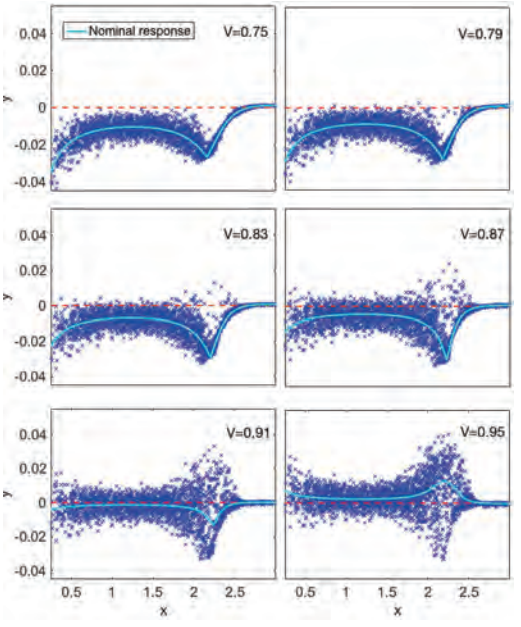


Figure 5. Nominal system response (solid line), and $N = 2500$ observations (blue ×) for several airflow speeds.

increases with v. In all cases, however, the $y = 0$ manifold is crossed by the nominal response near $x = 2.8$. The response of a calibrated deterministic model, to be referred to as a nominal response, is shown as a solid curve. Whereas the nominal system for all but the greatest speed crosses into the instability region at $x = 2.8$, somewhere in $v \in [0.91, 0.95]$ the curve flips to the opposite side of $y = 0$ (see the bottom plots in Figure 5). Reliability analyses using the nominal responses as the limit states will yield an abrupt discontinuity in the failure probability at that speed. This sudden change in the response is caused by a bifurcation. The manner by which the system transitions into instability (e.g, the region in x becoming unstable first) cannot be inferred from studying the nominal system.

The data for all 6 speeds was processed and the resulting moment functions were calculated. Figure 6 shows these functions and their corresponding uncertainty ranges. Note the high sensitivity of the functions to v. For instance, the variance varies considerably throughout x converging to a small value when x is large. Furthermore, the third-order central moment at $x = 2.15$ goes from being practically zero to being large, first positive and then negative. Features like these, driven by the dynamics of the system, can not be accurately described by a GP model. These functions were then used to calculate a maximal entropy RPM with $n_b = 300$ bins. Figure 7 shows

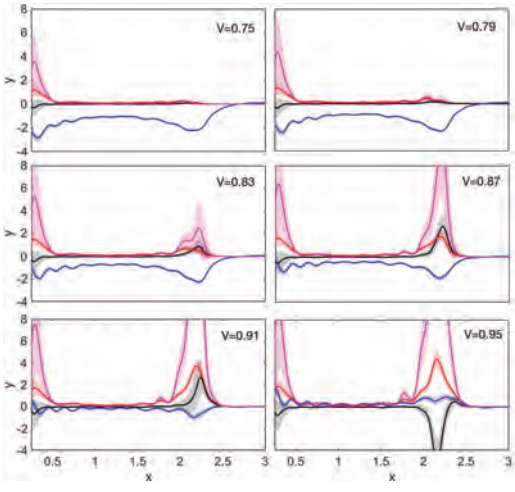


Figure 6. Mean (blue), second- (red), third-(black), and fourth-order (magenta) central moments along with their sampling error ranges (shaded areas).

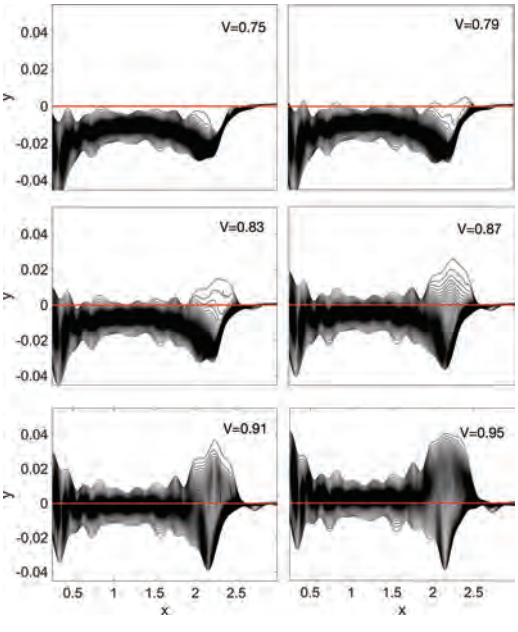


Figure 7. RPMs for several airflow speeds.

the one-percentile curves for the resulting RPMs. The excursion of individual percentiles into the instability region prescribes the severity (i.e., what is the failure probability for a fixed value of x and v) and the manner (i.e., which x region transitions to instability first) by which flutter occurs.

The crossing $y(x \mid v) = 0$ occurs over a range of x values. Let's formalize this notion by defining the τ-percentile of the RPM as

$$y_\tau(x \mid v) = F^{-1}_{y(x|v)}(\tau/100), \tag{20}$$

where $F_{y(x|v)}$ is the distribution of the RPM for speed v, and $\tau \in [0,100]$ is the percentile of interest. Hence, $y_{100}(x \mid v)$ is the upper limit of the RPM and $y_0(x \mid v)$ is the lower limit.

The failure domain associated with the τ-percentile at airspeed v is $\mathcal{F}_\tau = \{x : y_\tau(x \mid v) > 0\}$, whereas the corresponding failure probability is

$$\mathbb{P}[\mathcal{F}_\tau] = \int_{\mathcal{F}_\tau} f_x(x)\,dx, \tag{21}$$

where $f_x(x)$ is the PDF of x. $f_x(x)$ can be prescribed according to the available data or to expert opinion[2]. Modeling the response as an RPM yields the failure probability range

$$r(v) = \left[\ \mathbb{P}[\mathcal{F}_0], \mathbb{P}[\mathcal{F}_{100}]\ \right]. \tag{22}$$

This range can be readily computed for any $f_x(x)$. For instance, if x is a Beta random variable with hyper-parameters 3 and 3 and support [0, 3], we obtain $r(0.75) = [0.0040, 0.0187]$, $r(0.77) = [0.0059, 0.8814]$, $r(0.79) = [0.0024, 0.9948]$. These ranges account for all predicted outputs regardless of their likelihood. The upper limit of some of these ranges is distressingly close to one. However, it is possible that a very small portion of the predicted responses is responsible for most of the spread in the failure probability range. At this point the analyst might want to contemplate the following questions: If we are willing to ignore some of the worst predicted outputs what will be the corresponding reduction in the failure probability range? How large should be such a reduction (if any) to justify taking such a risk? By worst-case outputs we mean those leading to the largest decrease in the upper limit of the failure probability range. In this context, the risk, to be denoted as $\zeta \in [0,100]$, is the percentage of the worst-case predicted responses the analyst is willing to neglect. The analyst might be willing to accept a small risk providing that the corresponding reduction in the failure probability is sufficiently large. This will be the case when the limits of $r(v)$ are prescribed by extreme, low-probability events occurring at the long tail of a distribution. For a given ζ, the failure probability range is given by

$$r(v,\zeta) = \left[\ \mathbb{P}[\mathcal{F}_0], \mathbb{P}[\mathcal{F}_{100-\zeta}]\ \right]. \tag{23}$$

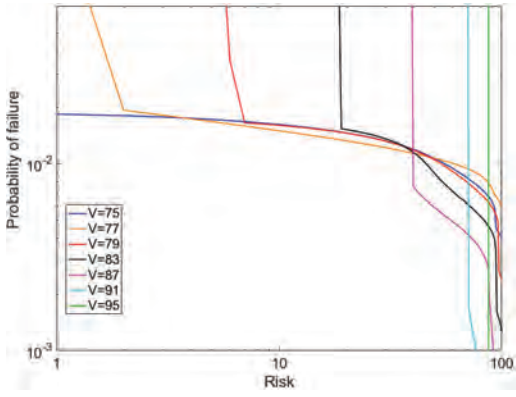


Figure 8. Failure probability vs. risk for several airflow speeds.

Figure 8 shows $\mathbb{P}[\mathcal{F}_{100-\zeta}]$ as a function of the risk ζ for several airflow speeds. This figure enables making informed risk-based decisions regarding the reliability assessment of the system. For instance, the range of failure probabilities corresponding to a zero risk for $v = 0.77$ is $r(0.77, 0) = [0.0059, 0.8814]$. These two values correspond to the points on the $v = 0.77$ curve for which the risk is 100 (smallest failure probability) and zero (largest failure probability). This level of risk implies that all predicted outcomes are accounted for. If the analyst is willing to ignore the worst 2 percent of the outputs, Figure 8 leads to $r(0.77, 2) = [0.0059, 0.0193]$. These two values correspond to the points on the $v = 0.77$ curve for which the risk is 100 (smallest failure probability) and 2 (largest failure probability). Therefore, a risk of 2 percent decreases the largest failure probability by 0.8621. This illustrates that a small percentage of the predicted responses is responsible for most of the spread in the range of failure probabilities, and that a risk averse approach might yield an overly conservative prediction. Whereas the zero-risk interval contains *all* predicted responses, the small-risk interval contains *most* of the responses more tightly. This information enables the analyst to avoid making overly conservative assessments driven by extreme low-probability events rarely seen in practice.

[2]If the realizations of x in $\mathbb{D}$ were controlled to ensure a good coverage of the response function over the domain of interest X, they should not be used to prescribe a naturally occurring $f_x(x)$, e.g., variations resulting from a manufacturing process.

5 CONCLUSIONS

This paper illustrates the use of staircase RPMs by applying them to the reliability analysis of an aeroelastic structure. The ability of the staircase variables to describe skewed and multimodal responses over an input-dependent interval makes them well suited for structural dynamics and controls applications. We consider the case in which the predictor is designed to match sample moments exactly (a setting applicable to large datasets), as well as the case in which the predictor accounts for the uncertainty in those estimates (a setting applicable to sparse datasets). The versatility and low computational cost of the proposed framework makes it appropriate for a wide range of applications in science and engineering.

REFERENCES

Campi, M., G. Calafiore, & S. Garatti 2009. Interval predictor models: Identification and reliability. *Automatica 45*(2), 382–392.

Campi, M. C. & S. Garatti 2008. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization 19*(3), 1211–1230.

Canor, T., L. Caracoglia, & V. Denoel 2015. Application of random eigenvalue analysis to assess bridge flutter probability. *Journal of wind engineering and industrial aerodynamics 140*, 79–86.

Crespo, L. G., D. P. Giesy, & S. P. Kenny 2017a, June. Nonparametric random predictor models with a staircase structure. In *ESREL 2017, Portoroz, Slovenia*.

Crespo, L. G., D. P. Giesy, & S. P. Kenny 2017b, June. On the calculation and shaping of staircase random variables. In *ESREL 2017, Portoroz, Slovenia*.

Crespo, L. G., S. P. Kenny, & D. P. Giesy 2016. Interval predictor models with a linear parameter dependency. *ASME Journal of verification, validation and uncertainty quantification 1*(2), 1–10.

Kendall, M. & A. Stuart 1969. *The advanced theory of statistics*. London, 3rd edition: Charles, Griffin and Co.

L. Munoz-Gonzales, M. Lazaro-Gredilla, A. F.-V. 2016. Gaussian processes for nonstationary regression. *IEEE transactions on pattern analysis and machine intelligence 38*(3), 618–623.

Mahler, A., C. Touze, O. Doare, G. Habib, & G. Kerschen 2017. Flutter control of a two degrees of freedom airfoil using a nonlinear tuned vibration absorber. *Journal of Computational nonlinear dynamics 12*(5).

Rackwitz, R. 2001. Reliability analysis, a review and some perspectives. *Structural Safety 23*, 365–395.

Rasmussen, C. E. & C. K. Williams 2006. *Gaussian Processes for Machine Learning*. MIT Press.

Silverman, B. W. 1986. *Density Estimation for statistics and data analysis*. 11 New Fetter Lane, London, England: Chapman and Hall.

Simpson, T., J. Peplinski, P. Koch, & J. Allen 2001. Metamodels for computer-based engineering design: survey and recommendations. *Engineering with Computers 17*(1), 129–150.

Sun, Z., J. Wang, L. Rui, & C. Tong 2017. LIF: a new kriging based learning function and its application to structural reliability analysis. *Reliability engineering and system safety 157*, 152–165.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Reliability aspects of a series load–sharing system

V.V. Krivtsov
Ford Motor Company and University of Maryland

S.V. Amari
BAE Systems

V.I. Gurevich
Israel Electrical Corporation

ABSTRACT: In reliability engineering, load–sharing is typically associated with a system in parallel configuration. Examples include bridge support structures, electric power supply systems, multiprocessor computing systems, etc. We consider a reliability maximization problem for a high–voltage commutation device, wherein the total voltage across the device is shared by the components in series configuration. Here, the increase of the number of load–sharing components increases component–level reliability (as the voltage load per component reduces) but may decrease system–level reliability (due to the increased number of components in series). We review optimal solutions for the proportional hazard and accelerated life models with the underlying exponential & Weibull distributions and elaborate on the log–linear, power, and Eyring laws used in the life–load models.

1 INTRODUCTION

A load–sharing system is typically associated with a system in parallel configuration. In their renowned text on System Reliability Theory, Rausand and Hoyland (2009) write "Consider a parallel system with two identical components. The components share a common load." Another famous text by Kapur and Lamberson (1977) also treats load–sharing systems exclusively as parallel systems.

Examples of parallel load–sharing systems include but are not limited to: civil engineering [e.g., structures (Chen & Lui 2005)], materials engineering [e.g., composites with fiber bundles (Phoenix & Tierney 1983)] power engineering [e.g., distributed generation systems (Marwalli & Keyhani 2004)], computer/network engineering [multiprocessor computing systems (Eager et al. 1986)]. Load–sharing system are also often discussed in the context of (still parallel) systems with k–out–of–n configuration, e.g., Huang & Xu (2010) and Amari & Bergman (2008).

In many applications of electrical and power engineering, namely in high voltage power equipment, one often runs into problem of switching high voltages by solid–state devices, where the system voltage (10–100 kV) well exceeds the operating voltages of individual switching components (1–3 kV). One of the commonly practiced ways to increase the voltage capability of high voltage switching devices is to put single switching components in a series configuration. Shown in Figure 1 is a high voltage thyristor switch (Gurevich & Krivtsov 1991), wherein the total switching voltage is shared by the serially connected thyristors.

In this case, the increase of the number of thyristors increases thyristor–level reliability (as the voltage load per thyristor reduces) but may decrease system–level reliability (due to the increased number of components in series). Clearly, the system reliability function in this case should have a maximum associated with an optimal number of thyristors in series.

We derived (Krivtsov et al. 2017) optimal solutions to the two popular life–load models: the

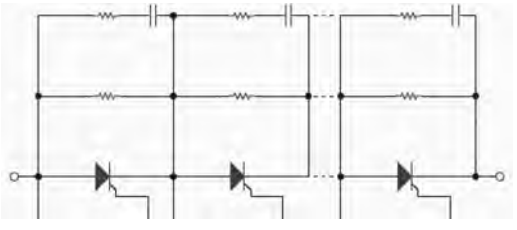


Figure 1. A serial load–sharing system, wherein total commutating voltage is shared by the serially connected SCRs (thyristors).

proportional hazard and the accelerated failure time with the underlying exponential and Weibull distributions. In this paper, we elaborate on the log–linear, power and Eyring laws used in the aforementioned life–load models.

2 PRELIMINARIES

Let $h(t;L)$ be the load–dependent failure rate of a component. Two commonly used models relating failure time to a load level are: the Proportional Hazard Model (PHM) and the Accelerated Failure Time Model (AFTM).

2.1 *Proportional hazard model*

Under the PHM, we have:

$$h(t;L) = \psi(L)\cdot h(t;L_0) = \psi(L)\cdot h_0(t), \tag{1}$$

where $h_0(t)$ is the baseline failure rate and $\psi(L_0)=1$.

The commonly used model for $\psi(L)$ is the log–linear (exponential) law: $\psi(L)=\exp(\alpha_0+\alpha_1 L)$. In fact, the exponential term can be replaced by any known positive, non–decreasing function. For example, for the log–linear law:

$$\psi(L)=\exp(\alpha_0+\alpha_1 L)=\delta\cdot\exp(\alpha_1 L) \tag{2}$$

For the power law:

$$\psi(L) = \delta\cdot L^{\alpha} \tag{3}$$

For the linear law:

$$\psi(L) = c+\alpha L \tag{4}$$

For the Eyring law:

$$\psi(L)=L^{-1}\exp(\alpha_0+\alpha_1/L) \tag{5}$$

The baseline failure rate can follow any time–varying function. Further, for the cumulative hazard and the reliability functions, we have:

$$H(t;L) = \psi(L)\cdot H_0(t) \tag{6}$$

$$R(t;L)=\exp\{-H(t;L)\}=[R_0(t)]^{\psi(L)} \tag{7}$$

2.2 *Accelerated failure time model*

Under this model, the effect of the load is multiplicative in time. The reliability function is expressed as:

$$R(t;L) = R_0(t\cdot\phi(L)) \tag{8}$$

where $R_0(.)$ is the reliability function at the baseline load. Function $\phi(L)$ represents the acceleration factor at load L. Without loss of generality, we can assume that $\phi(L_0)=1$. When there is only one type of load, commonly used forms of $\phi(L)$ include the power law:

$$\phi(L) = \delta\cdot L^{\alpha} \tag{9}$$

the log–linear law:

$$\phi(L) = \delta\cdot\exp(\alpha L) \tag{10}$$

and the Eyring law:

$$\phi(L)=L^{-1}\exp(\alpha_0+\alpha_1/L) \tag{11}$$

Finally, for the hazard functions we have:

$$H(t;L) = H_0(t\cdot\phi(L)) \tag{12}$$

$$h(t;L) = \phi(L)\cdot h_0(t\cdot\phi(L)) \tag{13}$$

It must be noted that if the baseline distribution is Weibull (or Exponential) and the multiplicative factor (acceleration factor) follows the power law, then the AFTM and PHM coincide. However, in general, there is no direct duality between the models.

2.3 *System reliability under series load–sharing*

Consider a load–sharing series system with n components. The total load on the system is $L_T \equiv V$. The load is equally distributed between the components. Hence, the load on each component is:

$$L \equiv \frac{L_T}{n} = \frac{V}{n} \tag{14}$$

and the system's reliability is:

$$R_s(t;V,n)=[R(t;V/n)]^n \tag{15}$$

Note that for fixed t, function $R(t;V/n)$ is an *increasing* function in n. However, function R^n is a *decreasing* function in n. Hence, there is an optimal value of n that maximizes the system reliability.

Alternatively, one can consider a logarithmic function of system reliability:

$$\ln R_s(t;V,n) = \mathrm{n}\cdot\ln R(t;V/n) \tag{16}$$

The right hand side of the above equation has two product terms. The first term increases with n and the second term decreases with n. Further, it implies that

$$H_s(t;V,n) = n \cdot H(t;V/n) \tag{17}$$

To maximize the system reliability, we need to minimize the corresponding cumulative hazard function. Again, it has two product terms. The first term increases with the number of components and the second term decreases with the number. Hence, there exist an optimal value that minimizes the cumulative hazard function and maximizes the system reliability.

3 OPTIMAL NUMBER OF COMPONENTS IN A SERIES LOAD–SHARING SYSTEM

3.1 *Load–sharing under PHM*

The system reliability in this case is:

$$\begin{aligned} R_s(t;V,n) &= \left[R(t;L) \right]^n \equiv \left[R\left(t;\frac{V}{n} \right) \right]^n \\ &= \left[\left[R_0(t) \right]^{\psi(V/n)} \right]^n = \left[R_0(t) \right]^{n\cdot\psi(V/n)} \end{aligned} \tag{18}$$

It follows that for fixed t, $R_0(t)$ is also fixed. Hence, for fixed t, to maximize the system reliability, we need to minimize $g(n) \equiv n \cdot \psi(V/n)$. Thus, the optimal n is independent of the form of the underlying failure time distribution $R_0(t)$ and is also independent of mission time t.

Under the log–linear law, we have:

$$\psi(L) = \exp(\alpha_0 + \alpha_1 L) = \delta \cdot \exp(\alpha_1 L) \tag{19}$$

For notational simplicity, hereafter we'll use α in the place of α_1. Hence,

$$\psi(L) \equiv \psi(V/n) = \delta \cdot \exp(\alpha \cdot V/n) \tag{20}$$

$$g(n) \equiv n \cdot \psi\left(\frac{V}{n} \right) = n\delta \cdot \exp(\alpha \cdot V/n) \tag{21}$$

The minimum of $g(n)$ can be obtained as

$$\begin{aligned} &\frac{d}{dn} g(n) \equiv \delta \cdot \exp\left(\alpha \cdot \frac{V}{n} \right) + n\delta.\exp\left(\alpha \cdot \frac{V}{n} \right)\left(-\frac{\alpha V}{n^2} \right) \\ &\frac{d}{dn} g(n) = 0 \Rightarrow n = \alpha \cdot V \end{aligned} \tag{22}$$

Thus, the optimal n that maximizes system reliability is

$$n = \|\alpha \cdot V\| \tag{23}$$

where ||.|| is the *nearest integer function.*

Figure 2 shows system–level reliability as a function of the number of the load–sharing components in series (with $\delta = 1$, $\alpha = 2$ and $V = 3$) for various values of component reliability. As expected, the optimal number does not depend on the latter.

It can be shown that for a PHM in the power law form, depending on model parameters, the optimal n will be either 1 or ∞. The linear law has a similar behavior. In both cases, the cost function can be used as regularization.

3.2 *Series load–sharing under AFTM*

The system reliability in this case is:

$$\begin{aligned} R_s(t;V,n) &= \left[R(t;L) \right]^n \\ &\equiv \left[R_0\left(t \cdot \phi(L) \right) \right]^n = \left[\exp\left\{ -H_0\left(t \cdot \phi(L) \right) \right\} \right]^n \end{aligned} \tag{24}$$

$$\begin{aligned} &R_s(t;V,n) \\ &= \exp\left\{ -n \cdot H_0\left(t.\phi(L) \right) \right\} = \exp\left\{ -H_s(t;V,n) \right\} \end{aligned} \tag{25}$$

where

$$H_s(t;V,n) = n \cdot H_0\left(t \cdot \phi(L) \right) \tag{26}$$

From the equation above, it follows that a) maximizing the system reliability is equivalent to minimizing the cumulative hazard function and b) unlike the PHM, the optimal value *does*

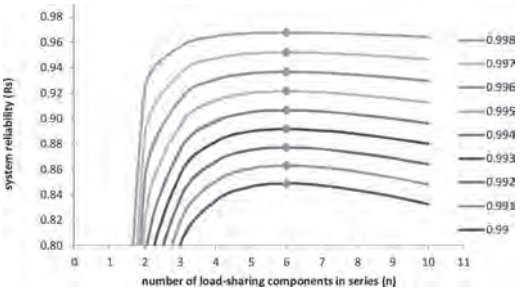


Figure 2. System–level reliability as a function of the number of the load–sharing components in series (with $\delta = 1$, $\alpha = 2$ and $V = 3$) for various values of component reliability. Optimal n is marked with the asterisk.

depends on the form of the underlying failure time distribution.

3.2.1 *AFTM with power law and the underlying exponential distribution*

For the exponential distribution, we have:

$$H_0(t) = \lambda t \tag{27}$$

Hence,

$$H_s(t;V,n) = n \cdot \lambda t \cdot \phi(L) = n \cdot \lambda t \cdot \phi(V/n) \tag{28}$$

In this case, maximizing the system reliability is equivalent to minimizing $g(n) \equiv n\phi(V/n)$. The optimal n is independent of the mission time or even the baseline hazard function.

Now, for the power law:

$$\phi(L) = \delta \cdot L^{\alpha} \tag{29}$$

and

$$g(n) = n\phi\left(\frac{V}{n}\right) = n\delta\left(\frac{V}{n}\right)^{\alpha} \tag{30}$$

Similar to the PHM with the power law, depending on model parameters, the optimal n will again be either 1 or ∞. The cost function can be used to regularize this case.

3.2.2 *AFTM with log–linear law and the underlying exponential distribution*

For the log–linear law:

$$\phi(L) = \delta \cdot \exp(\alpha L) \tag{31}$$

and

$$g(n) = n\phi\left(\frac{V}{n}\right) = n\delta \cdot \exp\left(\alpha \cdot \frac{V}{n}\right) \tag{32}$$

Note that the functional form of $g(n)$ is the same as in the PHM model with the log–linear law. Thus, the optimal n that maximizes system reliability is

$$n = \|\alpha \cdot V\| \tag{33}$$

3.2.3 *AFTM with Eyring law and the underlying exponential distribution*

$$\phi(L) = L^{-1}\exp(\alpha_0 + \alpha_1 / L) \tag{34}$$

and

$$g(n) = n\phi\left(\frac{V}{n}\right) = \frac{n^2}{V}\exp\left(\alpha_0 + \alpha_1 \frac{n}{V}\right) \tag{35}$$

The minimum of $g(n)$ is $\arg\left(\frac{d}{dn}g(n) = 0\right)$.

Let $X(n) \equiv \exp(\alpha_0 + \alpha_1 \frac{n}{V})$. Then with $\frac{d}{dn}g(n) \equiv \frac{X(n)}{V}(2n + \alpha_1)$, it becomes easy to show that the optimal n that maximizes system reliability is

$$n = \left\| -\frac{1}{2} \right\| \tag{36}$$

3.2.4 *AFTM with power law and the underlying Weibull distribution*

For the Weibull distribution, we have:

$$H_0(t) = \left(\frac{t}{\eta}\right)^{\beta} \tag{37}$$

Hence,

$$H_s(t;V,n) = n \cdot \left(\frac{t}{\eta}\right)^{\beta} \cdot \left[\phi(V/n)\right]^{\beta} \tag{38}$$

From the equation above, it follows that a) maximizing the system reliability is equivalent to minimizing $g(n) \equiv n \cdot [\phi(V/n)]^{\beta}$, and b) the optimal value of n is independent of the mission time and the scale parameter of the Weibull distribution.

For the power law:

$$\phi(L) = \delta \cdot L^{\alpha} \tag{39}$$

and

$$g(n) \equiv n\left[\phi\left(\frac{V}{n}\right)\right]^{\beta} = n\left[\delta \cdot \left(\frac{V}{n}\right)^{\alpha}\right]^{\beta} \tag{40}$$

Again, depending on model parameters, the optimal n will again be either 1 or ∞. The cost function can be used to regularize this case.

3.2.5 *AFTM with log–linear law and the underlying weibull distribution*

For the log–linear law:

$$\phi(L) = \delta \cdot \exp(\alpha L) \tag{41}$$

and

$$g(n) \equiv n\left[\phi\left(\frac{V}{n}\right)\right]^{\beta} = n \cdot \delta^{\beta} \cdot \exp\{\alpha\beta V/n\} \tag{42}$$

Hence,

$$\frac{d}{dn}g(n) = 0 \Rightarrow n = \alpha \cdot \beta \cdot V \tag{43}$$

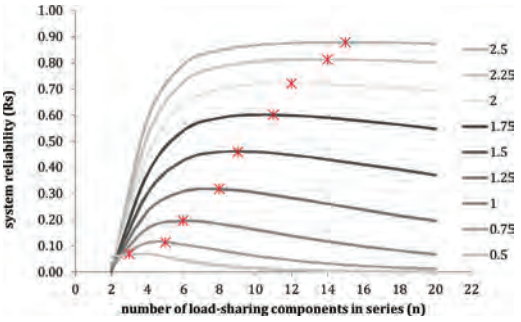


Figure 3. System–level reliability as a function of the number of the load–sharing components in series (with $\alpha = 2$ and $V = 3$) for various values of Weibull shape parameter with the scale parameter of $\eta = 1$ and mission time $t = 0.1$. Optimal n is marked with the asterisk.

Thus, the optimal n that maximizes the system reliability is:

$$n = \|\alpha \cdot \beta \cdot V\| \tag{44}$$

Figure 3 shows system–level reliability as a function of the number of the load–sharing components in series (with $\alpha = 2$ and $V = 3$) for various values of Weibull shape parameter with the scale parameter of $\eta = 1$ and mission time $t = 0.1$.

REFERENCES

Amari S.V. & Bergman R. 2008. Reliability analysis of k–out–of–n load–sharing systems. In Proc. *Annual Reliability and Maintainability Symposium*. January 2008, 440–445.

Chen W.F. & Lui, E.M. 2005. *Principles of Structural Design*. London: CRC Press.

Eager D.L., Lazowska E.D., Zahorjan J. 1986. Adaptive load sharing in homogenous distributed systems. *IEEE Trans. Software Engin.*, 12(5): 662–675.

Gurevich V.I & Krivtsov V.V. 1991. A High Voltage Solid–State Switching Device—Invention # SU–1728945–A1.

Gurevich V.I. 2003. *Protection Devices and Systems for High–Voltage Applications*. New York: Marcel Dekker.

Hoyland, A. & Rausand M. 2009. *System Reliability Theory: Models and Statistical Methods*. New York: Wiley.

Huang L. & Xu Q. 2010. Lifetime Reliability for Load–Sharing Redundant Systems with Arbitrary Failure Distributions, *IEEE Trans. Rel*. 59(2): 319–330.

Kapur K.C. & Lamberson L.R. 1977. *Reliability in Engineering Design*. New York: Wiley.

Krivtsov V.V., Amari S.V. Gurevich V.I. 2017. Load sharing in series configuration. *Qual Reliab Engng Int.* 34(1): 15–26.

Marwali M.N. & Keyhani A. 2004. Control of distributed generation systems, *IEEE Trans. Power Electron.* 19(6): 1541–1550.

Phoenix S.L & Tierney L.J. 1983. A statistical model for the time dependent failure of unidirectional composite materials under local elastic load–sharing among fibers. *Engineering Fracture Mechanics*. 18(1): 193–215.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

An evidential network-based method for common-cause failure analysis under uncertainty

S. Qiu & Henry X.G. Ming
School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China
Shanghai Key Laboratory of Advanced Manufacturing Environment, Shanghai, China

Y. Hou
INSA Centre Val de Loire, Bourges, France

ABSTRACT: In an engineering system, multiple components may fail simultaneously due to a shared cause or Common Cause (CC). This kind of failure is referred to as a Common-Cause Failure (CCF), and it contributes greatly to the system unreliability. Due to the insufficiency of historical data and system complexities, uncertainty is an inevitable problem in system reliability analysis. This paper proposes a Valuation-Based System (VBS) method to incorporate CCFs explicitly in system reliability analysis considering parametric uncertainty (related to reliability data of components) and model uncertainty (related to the system structure). The method can model different kinds of uncertainties, have no limitation on the type of failure distributions of system components, and allow the relationship between CCs being s-independent, s-dependent or mutually exclusive.

1 INTRODUCTION

Systems can be subject to Common-cause Failures (CCFs), where multiple components fail or malfunction simultaneously due to a shared cause or Common Cause (CC). For example, redundancy technique is often used to improve the system reliability, but it also has negative effects on the system reliability because identical components may fail simultaneously due to a CC. The presence of CCF makes the system be more prone to fail. Therefore, it is important to take into account the contributions of CCFs in system reliability analysis. There are two types of CCFs existing in engineering systems: external CCFs that are caused by external factors such as extreme weather and human errors, and internal CCFs that caused by propagated failures within the system such as the functional dependencies or physical dependencies between components (Misra 2008, Xing & Levitin 2013). In this paper, only external CCFs are taken into account.

CCFs have received considerable attention in system reliability analysis (Vaurio 2003, Misra 2008). Xing et al. (2007) summarized the limitations of existing CCF models, including being concerned with a specific system structure (Xie, Zhou, & Wang 2005), being applicable only to exponential time-to-failure distributions (Anderson & Agarwal 1992), having combinatorial explosion problems (Dai, Xie, Poh, & Ng 2004), limiting components belonging to at most one single Common-Cause Group (CCG) (Vaurio 1998), having a single CC (Amari, Dugan, & Misra 1999), defining CC as being s-independent or mutually exclusive (Vaurio 2003). Motivated by the limitations of existing methods, Xing *et al.* (Xing, Meshkat, & Donohue 2007, Xing 2008, Xing, Shrestha, Meshkat, & Wang 2009) proposed a Fault Tree (FT)-based Efficient Decomposition and Aggregation (EDA) approach which decomposes an original reliability problem with CCFs into a number of reduced reliability problems without considering the effect of CCFs using the Total Probability Theorem. However, in the EDA approach, models of some reduced reliability problems share common sub-models, and these sub-models are stored and evaluated several times. To improve the efficiency of the EDA approach, Mo and Xing (2013) extended the EDA approach by proposing an enhanced Decision Diagram-based (DD-based) method in which a single compact DD is generated to model all reduced reliability problems sharing their isomorphic sub-DDs.

Uncertainty analysis is challenging in reliability and risk analysis of complex systems (Qiu et al. 2014, Qiu et al. 2015, Qiu et al. 2018). Different kinds of uncertainties present in reliability studies because of many reasons, such as the randomness in phenomena and the insufficiency of data. As summarized in Pate-Cornell (1996), uncertainty is usually clas-

sified into two types: aleatory uncertainty and epistemic uncertainty. Aleatory uncertainty is due to the inherent variation in physical systems. It is also called irreducible uncertainty because it cannot be reduced. Epistemic uncertainty arises from the lack of knowledge. It is also called reducible uncertainty because it can be reduced by acquiring knowledge. Reliability parameters of components generally come from statistics, experiments, expert's opinions, similar components, etc. Therefore, uncertainties related reliability parameters of components may be aleatory or epistemic. In the previous CCF studies, only aleatory parametric uncertainty quantified by probability is taken into account. When it is difficult to measure the probabilities of CCs accurately, parametric models such as beta-factor model, binomial failure rate model, have been widely used to quantitatively analyze CCFs (Misra 2008). However, except the aleatory uncertainty coming from the randomness, there may exist epistemic uncertainty coming from the insufficiency of data and information in real systems. Because the probability cannot distinguish aleatory and epistemic uncertainties, probabilistic methods are not appropriate to analyze CCFs when there are aleatory and epistemic uncertainties. Therefore, non-probabilistic methods need to be developed for the CCF analysis considering aleatory and epistemic uncertainties.

Our proposed non-probabilistic method is based on Valuation-Based System (VBS). VBS was first introduced by Shenoy (1989) as a framework for representation of and reasoning with knowledge under uncertainty. Different types of uncertainties can be modeled and quantified in the framework of VBS using different uncertainty theories, including probability theory, imprecise probability theory, belief functions theory, possibility theory, etc. The proposed VBS method is suitable to analyze CCFs considering both of aleatory and epistemic uncertainties, have no limitations on the type of failure distributions of system components and allow the relationship between CCs being s-independent, s-dependent, or mutually exclusive. In the proposed VBS method, n CCs are modeled as n basic events and their three kinds of relationships are also considered in the model.

The structure of this paper is organized as follows. Section 2 presents the basic notions of the VBS. Section 3 presents an illustrative computer system. Section 4 presents the proposed VBS method and applies it to the illustrative example. Section 5 gives some conclusions and perspectives.

2 VALUATION-BASED SYSTEM

Shenoy (1989) introduced VBS as a framework for representation of and reasoning with knowledge under uncertainty. This framework belongs to the family of graphical models. In a VBS, a set of variables and a set of valuations affected to the subsets of the variables are used to represent knowledge. Reasoning with this knowledge means to find the marginals of the joint valuation for the variables of interest using two operators called combination and marginalization. VBS can represent different types of uncertainties in different domains including probabilities, basic probability assignments (bpas), possibility values, etc. The detailed definitions and features of VBS in reliability analysis based are discussed as follows (Qiu et al. 2017, Qiu et al. 2017).

2.1 *Basic probability assignments (bpas)*

A reliability analysis problem can be modeled using a finite set of variables that represent the components and system states. For a variable X, the frame of discernment Ω_X denotes the set of all the possible values that X can take. The mapping $m^{\Omega_X}: 2^{\Omega_X} \to [0,1]$ is called bpa if $\forall A \in 2^{\Omega_X}, \sum_{A \subseteq \Omega_X} m^{\Omega_X}(A) = 1$.

The set A can be an event or a subset of events. Indeed, a bpa m^{Ω_X} is assigned to each subset of 2^{Ω_X} such that $m^{\Omega_X}(A)$ represents the subjective probability assigned to the information which exactly supports A.

The lower bound of the probability over a set A on Ω_X represents the sum of all bpas of subsets that imply A. It is computed as follows (Shafer 1976)

$$\underline{P}(A) = \sum_{B|B \subseteq A} m^{\Omega_X}(B) \quad A, B \subseteq \Omega_X \tag{1}$$

The upper bound of the probability over A on Ω_X is defined as the total amount of bpas of subsets that are consistent with A. It is computed as follows

$$\overline{P}(A) = \sum_{B|B \cap A \neq \varnothing} m^{\Omega_X}(B) \quad A, B \subseteq \Omega_X \tag{2}$$

The following example explains the meaning of m, $\underline{P}$ and $\overline{P}$. Suppose that variables X and Y represent the states of Component 1 and Component 2. These variables may have two qualitative values: failed or working. The frames of discernment Ω_X and Ω_Y of X and Y are given by $\Omega_X = \Omega_Y = \{F, W\}$.

An expert assigns bpas to the two values of X as follows: $m_1^{\Omega_X}(\{W\}) = 0.7, m_1^{\Omega_X}(\{W, F\}) = 0.2, m_1^{\Omega_X}(\{F\}) = 0.1$. The expert assigns also bpas to the two states of Y as follows: $m_2^{\Omega_Y}(\{W\}) = 0.6, m_2^{\Omega_Y}(\{W, F\}) = 0.4$.

According to Eq.1 and Eq. 2, the probability of the event A: "*Component 1 is in the working state*", is included in

$[\underline{P}(A), \overline{P}(A)] = [m_1^{\Omega_X}(W), m_1^{\Omega_X}(\{W\}) + m_1^{\Omega_X}(\{W,F\})]$ = [0.7,0.9]. The value 0.7 represents all the information that implies the event A, whereas the value 0.9 represents all the information that is consistent with the event A according to the expert. The length of the interval $\overline{P}(A) - P(A) = 0.2$ represents the imprecision about A.

2.2 *Joint basic probability assignments*

Let X and Y be two variables defined on frames Ω_X and Ω_Y. Let $\Omega_X \times \Omega_Y$ be the cartesian product of Ω_X and Ω_Y. A bpa $m^{\Omega_X \times \Omega_Y}$ defined on $\Omega_X \times \Omega_Y$ is called a joint bpa, and can be seen as an uncertain relation between variables X and Y. To evaluate the VBS, joints bpas defined on the cartesian product of variables are used to show the degree (or strength) of the relationship between variables.

We retake the same example, and we aim to evaluate the probability of the event: "The system composed of components 1 and 2 is in the working state" which is represented by a variable S. The variable S may have two qualitative values: failed or working, and its frame of discernment is given by $\Omega_S = \{W,F\}$. The variable S depends on the variables X (state of Component 1) and Y (state of Component 2). This relationship will be represented by a joint bpa m_3. The expert gives the following joint bpa m_3 in order to define the degree of relationships between the variables:

$$\begin{cases} m_3^{\Omega_X \times \Omega_Y \times \Omega_S}(\{(W_1,W_2,W_S),(F_1,F_2,F_S)\}) = 0.75 \\ m_3^{\Omega_X \times \Omega_Y \times \Omega_S}(\{F_1,W_2,W_S),(W_1,F_2,W_S)\}) = 0.2 \\ m_3^{\Omega_X \times \Omega_Y \times \Omega_S}(\{\Omega_X \times \Omega_Y \times \Omega_S\}) = 0.05 \end{cases}$$

Fig. 1 represents the VBS of the example. As we can see, we have three nodes which represent the variables X, Y and S. We have also three bpas: m_1, m_2, and m_3 which respectively represent the bpas assigned to X, Y, and the relationship between X, Y and S.

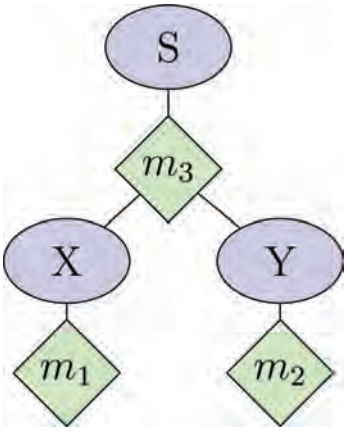


Figure 1. VBS example.

2.3 *Inference in VBS*

After constructing the VBS and assigning the bpas of events (variables) and joints bpas of relationship between events (variables), we have to use inference in order to obtain the probability of the variable of interest. Making inference means to find marginals for the variables of interest using combination and marginalization operators.

The combination operation of two bpas $m_i^{\Omega_X}$ and $m_j^{\Omega_Y}$ respectively defined on Ω_X and Ω_Y is performed as follows (Smets & Kennes 1994)

$$m_{i \oplus j}^{\Omega_X \times \Omega_Y}(H) = \frac{\sum_{A \cap B = H, \forall A,B \subseteq \Omega_X \times \Omega_Y} m_i^{\Omega_X \times \Omega_Y}(A)\, m_j^{\Omega_X \times ga_Y}(B)}{1 - \sum_{A \cap B = \varnothing, \forall A,B \subseteq \Omega_X \times \Omega_Y} m_i^{\Omega_X \times \Omega_Y}(A)\, m_j^{\Omega_X \times \Omega_Y}(B)} \quad (3)$$

Note that in order to combine two bpas respectively defined on Ω_X and Ω_Y, they must be defined on the same frame of discernment $\Omega_X \times \Omega_Y$. The operation which allows one to extend a bpa m^{Ω_X} defined on Ω_X to $\Omega_X \times \Omega_Y$ is called the vacuous extension and defined as follows

$$m^{\Omega_X \uparrow (\Omega_X \times \Omega_Y)}(B) = \begin{cases} m^{\Omega_X}(A) & \text{if } B = A \times \Omega_Y, \forall\ A \subseteq \Omega_X \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

On the other hand, the marginalization operator is used to focus the information contained by a valuation onto a smaller domain. Indeed, the marginal (or projection) of the bpa $m^{\Omega_X \times \Omega_Y}$ on the frame Ω_X is defined by

$$m^{(\Omega_X \times \Omega_Y)\downarrow \Omega_X}(A) = \sum_{B \subseteq \Omega_X \times \Omega_Y / Proj(B \downarrow \Omega_X) = A} m^{\Omega_X \times \Omega_Y}(B), \forall A \subseteq \Omega_X \quad (5)$$

where $Proj(B \downarrow \Omega_x) = \{x \in \Omega_x / \exists y \in \Omega_y, (x,y) \in B\}$.

To summarize, for a variable of interest Z that depends on two other variables X and Y, $(\oplus m)^{\downarrow \Omega_Z}$ is computed by marginalizing (projection) on Ω_Z the joint valuation $\oplus m$ obtained by the combination of the bpas $m^{\Omega_X}, m^{\Omega_Y}$, and the joint bpa $m^{\Omega_X \times \Omega_Y \times \Omega_Z}$.

We retake the same example in order to compute the upper and lower probabilities of the event: "the system composed of components 1 and 2 is in the working state". We first combine the bpas $m_1^{\Omega_X}, m_2^{\Omega_Y}$, and the joint bpa $m_3^{\Omega_X \times \Omega_Y \times \Omega_S}$

using Eq. 4 (After vacuous extension of each bpa using Eq. 5). Then, we marginalize the obtained bpa $(m_1 \oplus m_2 \oplus m_3)^{\Omega_X \times \Omega_Y \times \Omega_S}$ on the frame Ω_S using Eq. 6. Finally, using Eqs. 1 and 2, we obtain upper and lower probabilities of the event "the system composed of components 1 and 2 is in the working state" from the bpa $m_3^{(\Omega_X \times \Omega_Y \times \Omega_S)\downarrow \Omega_S}$ as follows: $\left[\underline{P}(W_S), \overline{P}(W_S) \right] = \left[0.9,1 \right]$.

3 AN ILLUSTRATIVE EXAMPLE

Here we use an example of a computer system subject to CCFs proposed by Mo and Xing (2013). This system consists of three processors (P_1, P_2, and P_3), two buses (B_1 and B_2), and three memory units (M_1, M_2, and M_3). It works if at least two of the three processors, at least one of the two buses, and at least one of the three memory units work. Fig. 2 shows the reliability block diagram of the example computer system.

Two CCs are supposed to affect the components in the example computer system. CC_1 affects processor P_1 and memory unit M_1. CC_2 affects processor P_1, bus B_2 and memory unit M_2. Thus, the two CCGs are $CCG_1 = \{P_1, M_1\}$ and $CCG_2 = \{P_1, B_2, M_2\}$. There are three kinds of relationships between CC_1 and CC_2: s-independent, s-dependent, or mutually exclusive.

Assume that all components and the system are binary, and their frames of discernment are $\Omega = \{0,1\}$ where '0' denotes failed state and '1' denotes working state. The frame of discernment of CCs is $\Omega_{CC_i} = \{0_i, 1_i\}$ where '0_i' denotes the non-occurrence of CC_i and '1_i' denotes the occurrence of CC_i.

Probabilities used in this example are given as follows (Mo & Xing 2013):

- Failure probabilities of components: $q_{P_1} = q_{P_2} = q_{P_3} = 0.002$, $q_{B_1} = q_{B_2} = 0.001$, $q_{M_1} = q_{M_2} = q_{M_3} = 0.003$. Reliabilities of components are hereinafter denoted by p.
- Occurrence probabilities of CCs: $\Pr\{CC_1\} = 0.001$ and $\Pr\{CC_2\} = 0.0015$.
- If CCs are s-dependent, the conditional occurrence probabilities are $\Pr\{CC_2 \mid CC_1\} = 0.0025$ and $\Pr\{CC_2 \mid \leftarrow CC_1\} = 0.0015$.

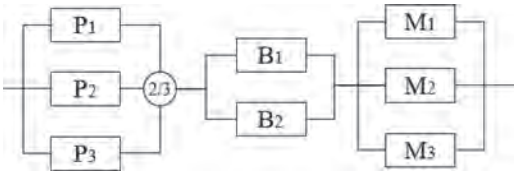


Figure 2. Reliability block diagram of the example computer system.

4 PROPOSED VBS METHOD

This section presents the explicit VBS method to incorporate CCFs considering only aleatory uncertainty represented by probability, and applies it to the example computer system given in Section 3.

4.1 *Proposed method*

The proposed VBS method contains the following five steps:

Step 1: Develop the VBS model of the studied system. Each CC is modeled by a basic event, and the relationship between CCs is modeled by a basic event.

Step 2: Establish the truth table that represents the relationship between CCs and transform the truth table into a bpa. There are three kinds of relationships between CCs: s-independent, s-dependent, or mutually exclusive. In this work, n CCs occurring s-dependently means that CC_i depends on CC_{i-1}. A variable R is used to describe the three kinds of relationships. Its frame of discernment is $\Omega_R = \{1_R, 2_R, 3_R\}$, where '$1_R$' denotes s-independent, '2_R' denotes s-dependent, and '3_R' denotes mutually exclusive. The occurrence of CCs is influenced by their relationship, e.g., if n CCs are mutually exclusive, no CCs can occur simultaneously. Table 1 is the truth table for CCs under different relationships with the corresponding conditional probabilities.

Step 3: Establish truth tables that represent the relationships between CCs and their affecting components, and transform truth tables into bpas. A component in the system may be affected by several CCs. For example, if a component x is affected by $CC_l, CC_{l+1}, \ldots, CC_k$, Table 2 is the truth table representing the relation between x and these CCs.

Step 4: Establish truth tables that represent the relationships between components, subsystems, the system, and transform truth tables into bpas. For example, for a system consisting of n components in parallel, its truth table is shown in Table 3.

Step 5: Declare bpas for leaf nodes according to their reliabilities, calculate the joint bpa and marginalize it over the frame of *System*.

4.2 *Illustrative example*

In this subsection, the proposed VBS method is used to model and evaluate the reliability of the example computer system.

Step 1: Fig. 3 shows the valuation network of the example computer system in the proposed VBS

Table 1. Truth table for CCs under different relationships.

R	CC_1	CC_2	…	CC_{n-1}	CC_n	$\Pr\{CC_1 \,\&\, \ldots \,\&\, CC_n \mid R\}$
1_R	0_1	0_2	…	0_{n-1}	0_n	$\Pr\{-CC_1\}\Pr\{-CC_2\}\ldots\Pr\{-CC_{n-1}\}\Pr\{-CC_n\}$
1_R	0_1	0_2	…	0_{n-1}	1_n	$\Pr\{-CC_1\}\Pr\{-CC_2\}\ldots\Pr\{-CC_{n-1}\}\Pr\{CC_n\}$
1_R	0_1	0_2	…	1_{n-1}	0_n	$\Pr\{-CC_1\}\Pr\{-CC_2\}\ldots\Pr\{CC_{n-1}\}\Pr\{-CC_n\}$
1_R	0_1	0_2	…	1_{n-1}	1_n	$\Pr\{-CC_1\}\Pr\{-CC_2\}\ldots\Pr\{CC_{n-1}\}\Pr\{CC_n\}$
…	…	…	…	…	…	…
1_R	1_1	1_2	…	1_{n-1}	1_n	$\Pr\{CC_1\}\Pr\{CC_2\}\ldots\Pr\{CC_{n-1}\}\Pr\{CC_n\}$
2_R	0_1	0_2	…	0_{n-1}	0_n	$\Pr\{-CC_1\}\Pr\{-CC_2\|-CC_1\}\ldots\Pr\{-CC_{n-1}\|-CC_{n-2}\}\Pr\{-CC_n\|-CC_{n-1}\}$
2_R	0_1	0_2	…	0_{n-1}	1_n	$\Pr\{-CC_1\}\Pr\{-CC_2\|-CC_1\}\ldots\Pr\{-CC_{n-1}\|-CC_{n-2}\}\Pr\{CC_n\|-CC_{n-1}\}$
2_R	0_1	0_2	…	1_{n-1}	0_n	$\Pr\{-CC_1\}\Pr\{-CC_2\|-CC_1\}\ldots\Pr\{CC_{n-1}\|-CC_{n-2}\}\Pr\{-CC_n\|CC_{n-1}\}$
2_R	0_1	0_2	…	1_{n-1}	1_n	$\Pr\{-CC_1\}\Pr\{-CC_2\|-CC_1\}\ldots\Pr\{CC_{n-1}\|-CC_{n-2}\}\Pr\{CC_n\|CC_{n-1}\}$
…	…	…	…	…	…	…
2_R	1_1	1_2	…	1_{n-1}	1_n	$\Pr\{CC_1\}\Pr\{CC_2\|CC_1\}\ldots\Pr\{CC_{n-1}\|CC_{n-2}\}\Pr\{CC_n\|CC_{n-1}\}$
3_R	0_1	0_2	…	0_{n-1}	0_n	$1-\sum_{i=1}^{n}\Pr\{CC_i\}$
3_R	0_1	0_2	…	0_{n-1}	1_n	$\Pr\{CC_n\}$
3_R	0_1	0_2	…	1_{n-1}	0_n	$\Pr\{CC_{n-1}\}$
…	…	…	…	…	…	…
3_R	0_1	1_2	…	0_{n-1}	0_n	$\Pr\{CC_2\}$
3_R	1_1	0_2	…	0_{n-1}	0_n	$\Pr\{CC_1\}$

Table 2. Truth table of the relation between x and CCs.

CC_l	…	CC_{k-1}	CC_k	x	$\Pr\{x \mid CC_l \,\&\, \ldots \,\&\, CC_k\}$
0_l	…	0_{k-1}	0_k	1_x	p_x
0_l	…	0_{k-1}	0_k	0_x	q_x
0_l	…	0_{k-1}	1_k	0_x	**1**
0_l	…	1_{k-1}	0_k	0_x	**1**
0_l	…	1_{k-1}	1_k	0_x	**1**
⋮	⋮	⋮	⋮	⋮	⋮
1_l	…	1_{k-1}	1_k	0_x	**1**

Table 3. Truth table of a parallel system.

x_1	…	x_{n-1}	x_n	*System*	$\Pr\{\text{System} \mid x_1 \,\&\ldots\&\, x_n\}$
0_{x_1}	⋯	$0_{x_{n-1}}$	0_{x_n}	0_S	**1**
0_{x_1}	⋯	$0_{x_{n-1}}$	1_{x_n}	1_S	**1**
0_{x_1}	⋯	$1_{x_{n-1}}$	0_{x_n}	1_S	**1**
0_{x_1}	⋯	$1_{x_{n-1}}$	1_{x_n}	1_S	**1**
⋮	⋮	⋮	⋮	⋮	⋮
1_{x_1}	⋯	$1_{x_{n-1}}$	1_{x_n}	1_S	**1**

method. 15 variables and 14 bpas are used to model the example computer system. Two CCs and the relationship between CCs are modeled separately as basic events.

Step 2: The relationship between two CCs can be s-independent, s-dependent, or mutually exclusive. Table 4 is the truth table for CCs under different relationships with the corresponding conditional probabilities.

The truth table in Table 4 is transformed into the following bpa $m_{13}^{\Omega_R\Omega_{CC1}\Omega_{CC2}}$:

$$m_{13}\left(\{1_R,0_1,0_2\},\{2_R,0_1,0_2\},\{3_R,0_1,0_2\}\right)=0.9975$$
$$m_{13}\left(\{1_R,0_1,1_2\},\{2_R,0_1,1_2\},\{3_R,0_1,1_2\}\right)=0.0014985$$
$$m_{13}\left(\{1_R,0_1,0_2\},\{2_R,0_1,0_2\},\{3_R,0_1,1_2\}\right)=0.0000015,$$
$$m_{13}\left(\{1_R,1_1,0_2\},\{2_R,1_1,0_2\},\{3_R,1_1,0_2\}\right)=0.0009975$$
$$m_{13}\left(\{1_R,1_1,0_2\},\{2_R,1_1,1_2\},\{3_R,1_1,0_2\}\right)=0.000001$$
$$m_{13}\left(\{1_R,1_1,1_2\},\{2_R,1_1,1_2\},\{3_R,1_1,0_2\}\right)=0.0000015$$

Step 3: P_1 is affected by CC_1 and CC_2. Table 5 shows the truth table representing the relationship between P_1 and two CCs.

The truth table in Table 5 is transformed into the following bpa $m_5^{\Omega_{CC1}\Omega_{CC2}\Omega_{P_1}}$:

$$m_5(\{0_1,0_2,1_{P_1}\},\{0_1,1_2,0_{P_1}\},\{1_1,0_2,0_{P_1}\},\{1_1,1_2,0_{P_1}\})=0.998$$
$$m_5(\{0_1,0_2,0_{P_1}\},\{0_1,1_2,0_{P_1}\},\{1_1,0_2,0_{P_1}\},\{1_1,1_2,0_{P_1}\})=0.002$$

The same for $B_2(m_9)$, $M_1(m_{10})$ and $M_2(m_{11})$.

Step 4: The relationships between components, subsystems and the system depend on the system structure. For example, the processor subsystem works if at least two out of the three

Table 4. Truth table for two CCs under different relationships.

R	CC_1	CC_2	$\Pr\{CC_1 \,\&\, CC_2 \mid R\}$
1_R	0_1	0_2	$\Pr\{-CC_1\}\Pr\{-CC_2\} = 0.9975015$
1_R	0_1	1_2	$\Pr\{-CC_1\}\Pr\{CC_2\} = 0.0014985$
1_R	1_1	0_2	$\Pr\{CC_1\}\Pr\{-CC_2\} = 0.0009985$
1_R	1_1	1_2	$\Pr\{CC_1\}\Pr\{CC_2\} = 0.0000015$
2_R	0_1	0_2	$\Pr\{-CC_1\}\Pr\{-CC_2 \mid -CC_1\} = 0.9975015$
2_R	0_1	1_2	$\Pr\{-CC_1\}\Pr\{CC_2 \mid -CC_1\} = 0.0014985$
2_R	1_1	0_2	$\Pr\{CC_1\}\Pr\{-CC_2 \mid CC_1\} = 0.0009975$
2_R	1_1	1_2	$\Pr\{CC_1\}\Pr\{CC_2 \mid CC_1\} = 0.0000025$
3_R	0_1	0_2	$1 - \Pr\{CC_1\} - \Pr\{CC_2\} = 0.9975$
3_R	0_1	1_2	$\Pr\{CC_2\} = 0.0015$
3_R	1_1	0_2	$\Pr\{CC_1\} = 0.001$

Table 5. Truth table of the relationship between P_1 and CCs.

CC_1	CC_2	P_1	$\Pr\{P_1 \mid CC_1 \,\&\, CC_2\}$
0_1	0_2	1_{P_1}	$p_{P_1} = 0.998$
0_1	0_2	0_{P_1}	$q_{P_1} = 0.002$
0_1	1_2	0_{P_1}	**1**
1_1	0_2	0_{P_1}	**1**
1_1	1_2	0_{P_1}	**1**

Table 6. Truth table of the processor subsystem.

P_1	P_2	P_3	P	$\Pr\{P \mid P_1 \,\&\, P_2 \,\&\, P_3\}$
0_{P_1}	0_{P_2}	0_{P_3}	0_P	**1**
0_{P_1}	0_{P_2}	1_{P_3}	0_P	**1**
0_{P_1}	1_{P_2}	0_{P_3}	0_P	**1**
0_{P_1}	1_{P_2}	1_{P_3}	1_P	**1**
1_{P_1}	0_{P_2}	0_{P_3}	0_P	**1**
1_{P_1}	0_{P_2}	1_{P_3}	1_P	**1**
1_{P_1}	1_{P_2}	0_{P_3}	1_P	**1**
1_{P_1}	1_{P_2}	1_{P_3}	1_P	**1**

processors work, thus the truth table of the processor subsystem is given in Table 6.

The truth table in Table 6 is transformed into the following bpa $m_2^{\Omega_{P_1}\Omega_{P_2}\Omega_{P_3}\Omega_P}$:

$$m_2(\{0_{P_1},0_{P_2},0_{P_3},0_P\},\{0_{P_1},0_{P_2},1_{P_3},0_P\},\{0_{P_1},1_{P_2},0_{P_3},0_P\},\{0_{P_1},1_{P_2},1_{P_3},1_P\},\{1_{P_1},0_{P_2},0_{P_3},0_P\},\{1_{P_1},0_{P_2},1_{P_3},1_P\},\{1_{P_1},1_{P_2},0_{P_3},1_P\},\{1_{P_1},1_{P_2},1_{P_3},1_P\}) = 1$$

The same for the bus subsystem (m_3), the memory subsystem (m_4), and the entire system (m_1).

Step 5: m_6, m_7, m_8, m_{12} represent the knowledge about the states of P_2, P_3, B_1, M_3. According to their failure probabilities, we have

$$m_6^{\Omega_{P_2}}(\{0_{P_2}\}) = m_7^{\Omega_{P_3}}(\{0_{P_3}\}) = 0.002$$
$$m_6^{\Omega_{P_2}}(\{1_{P_2}\}) = m_7^{\Omega_{P_3}}(\{1_{P_3}\}) = 0.998$$
$$m_8^{\Omega_{B_1}}(\{0_{B_1}\}) = 0.001 \quad m_8^{\Omega_{B_1}}(\{1_{B_1}\}) = 0.999$$
$$m_{12}^{\Omega_{M_3}}(\{0_{M_3}\}) = 0.003 \quad m_{12}^{\Omega_{M_3}}(\{1_{M_3}\}) = 0.997$$

m_{14} represent the knowledge about the relationship of CCs. There are three kinds of relationship between CCs:

- $m_{14}^{\Omega_R}(\{1_R\}) = 1$ if CCs are s-independent;
- $m_{14}^{\Omega_R}(\{2_R\}) = 1$ if CCs are s-dependent;
- $m_{14}^{\Omega_R}(\{3_R\}) = 1$ if CCs are mutually exclusive.

The system reliability is computed by $(m_1^{\Omega_P\Omega_B\Omega_M\Omega_S} \otimes m_2^{\Omega_{P_1}\Omega_{P_2}\Omega_{P_3}\Omega_P} \otimes \ldots \otimes m_{13}^{\Omega_R\Omega_{CC1}\Omega_{CC2}} \otimes m_{14}^{\Omega_R})^{\downarrow\Omega_S}$. If CCs are s-independent, the obtained system unreliability is 2.44843e-05. If CCs are s-dependent, the obtained system unreliability is 2.44883e-05. If CCs are mutually exclusive, the obtained system unreliability is 2.44859e-05.

The obtained results are different from the results in Mo and Xing (2013), because the conditional probabilities $\Pr\{Systemfailure \mid CCE_i\}$ in the compared paper are incorrect.

The proposed explicit VBS method is more efficient than the explicit FT-based methods Vaurio 2003, Wang, Xing, & Levitin 2014). In the explicit FT-based methods, if there are n CCs in which CC_i affects m_i components, then $\sum_{i=1}^{n} m_i$ basic events and logical gates are needed to model CCs in the expanded FT models. In the proposed explicit VBS method, only n basic events are needed to model CCs. And as explained in Wang et al. (2014), the explicit FT-based method is only applicable to systems subject to s-independent CCs, whereas the proposed explicit VBS method is applicable to systems subject to s-independent, s-dependent, or mutually exclusive CCs by introducing a variable representing the relationship between CCs.

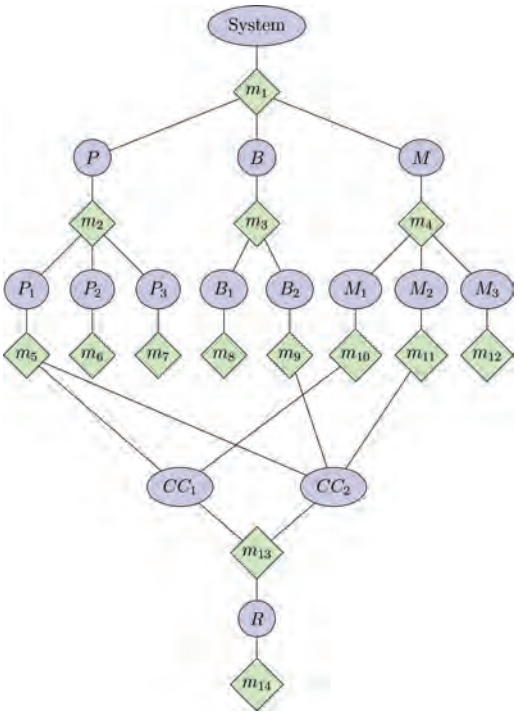


Figure 3. Valuation network of the example computer system subject to CCFs in the proposed VBS method.

Table 7. Truth table of the processor subsystem under model epistemic uncertainty.

P_1	P_2	P_3	P	$\Pr\{P\|P_1 \,\&\, P_2 \,\&\, P_3\}$
0_{P_1}	0_{P_2}	0_{P_3}	0_P	**1**
0_{P_1}	0_{P_2}	1_{P_3}	$\mathbf{0_P, 1_P}$	**1**
0_{P_1}	1_{P_2}	0_{P_3}	$\mathbf{0_P, 1_P}$	**1**
0_{P_1}	1_{P_2}	1_{P_3}	1_P	**1**
1_{P_1}	0_{P_2}	0_{P_3}	$\mathbf{0_P, 1_P}$	**1**
1_{P_1}	0_{P_2}	1_{P_3}	1_P	**1**
1_{P_1}	1_{P_2}	0_{P_3}	1_P	**1**
1_{P_1}	1_{P_2}	1_{P_3}	1_P	**1**

4.3 *Epistemic model uncertainty*

This subsection takes the epistemic model uncertainty into account. Epistemic model uncertainty means that the analyst cannot give the precise system structure due to the lack of knowledge or information. For example, in the example computer system, the analyst has difficulty in judging the structure of the processor subsystem due to the lack of relative knowledge. He/she cannot judge whether it is a parallel system or a 2-out-of-3: G system (the subsystem works if and only if at least two out of the three processors work). The epistemic uncertainty of the analyst can be described by the truth table in Table 7. Items in bold represent the epistemic model uncertainty.

The truth table in Table 7 is transformed into the following joint bpa $m_2^{\Omega_{P_1}\Omega_{P_2}\Omega_{P_3}\Omega_P}$:

$$m_2\Big(\{0_{P_1},0_{P_2},0_{P_3},0_P\},\{0_{P_1},0_{P_2},1_{P_3},0_P\},\{0_{P_1},0_{P_2},1_{P_3},1_P\},\{0_{P_1},1_{P_2},0_{P_3},0_P\},\{0_{P_1},1_{P_2},0_{P_3},1_P\},\{0_{P_1},1_{P_2},1_{P_3},1_P\},\{1_{P_1},0_{P_2},0_{P_3},0_P\},\{1_{P_1},0_{P_2},0_{P_3},1_P\},\{1_{P_1},0_{P_2},1_{P_3},1_P\},\{1_{P_1},1_{P_2},0_{P_3},1_P\},\{1_{P_1},1_{P_2},1_{P_3},1_P\}\Big)=1$$

All other bpas remain unchanged. The proposed VBS method is applied to model and evaluate the system unreliability under this epistemic model uncertainty. Finally, the system unreliability obtained by the VBS method is [2.57036e-06,2.44843e-05].

5 CONCLUSION

In this paper, a VBS method is proposed to analyze CCFs considering aleatory and epistemic uncertainties. In the proposed VBS method, n CCs are modeled as n basic events and their three kinds of relationships are also considered in the model. Our proposed VBS method is suitable to analyze CCFs considering both of aleatory and epistemic uncertainties, have no limitations on the type of failure distributions of system components and allow the relationship between CCs being s-independent, s-dependent, or mutually exclusive. In the future, the proposed VBS approach will be extended to analyze the Belief CCF which means that the occurrence of a CC may result in failures of different components with different degrees of belief.

ACKNOWLEDGMENT

This work was sponsored by Shanghai Pujiang Program under grant number 16PJ1404500 and National Science Foundation of China under grant number 71632008.

REFERENCES

Amari, S.V., J.B. Dugan, & R.B. Misra (1999). Optimal reliability of systems subject to imperfect fault-coverage. *IEEE Transactions on Reliability 48*(3), 275–284.

Anderson, P.M. & S.K. Agarwal (1992). An Improved Model for Protective-System Reliability. *IEEE Transactions on Reliability 41*(3), 422–426.

Dai, Y., M. Xie, K. Poh, & S. Ng (2004). A model for correlated failures in N -version programming. *IIE Transactions 36*(12), 1183–1192.

Misra, K.B. (2008). *Handbook of Performability Engineering*. International Journal of Performability Engineering.

Mo, Y. & L. Xing (2013). An enhanced decision diagram-based method for common-cause failure analysis. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 227*(5), 557–566.

Pate-Cornell, M.E. (1996). Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering & System Safety 54*, 95–111.

Qiu, S., N. Rachedi, M. Sallak, & F. Vanderhaegen (2017). A quantitative model for the risk evaluation of driver-ADAS systems under uncertainty. *Reliability Engineering & System Safety 167*, 184–191.

Qiu, S., R. Sacile, M. Sallak, & W. Schön (2015). On the application of Valuation-Based Systems in the assessment of the probability bounds of Hazardous Material transportation accidents occurrence. *Safety Science 72*, 83–96.

Qiu, S., M. Sallak, W. Schön, & Z. Cherfi-Boulanger (2014). Availability assessment of railway signalling systems with uncertainty analysis using Statecharts. *Simulation Modelling Practice and Theory 47*, 1–18.

Qiu, S., M. Sallak, W. Schön, & Z. Cherfi-Boulanger (2017). Application of Valuation-Based Systems for the availability assessment of systems under uncertainty. *Control Engineering Practice 66*, 39–50.

Qiu, S., M. Sallak, W. Schön, & H.X. Ming (2018). Extended LK heuristics for the optimization of linear consecutivek-out-of-n: F systems considering parametric uncertainty and model uncertainty. *Reliability Engineering and System Safety*, Accepted. 10.1016/j.ress.2018.01.016.

Shafer, G. (1976). *A mathematical Theory of Evidence*. New Jersey: Princeton University Press.

Shenoy, P.P. (1989). A valuation-based language for expert systems. *International Journal of Approximate Reasoning 3*(5), 383–411.

Smets, P. & R. Kennes (1994). The transferable belief model. *Artificial Intelligence 66*, 191–234.

Vaurio, J. (2003). Common cause failure probabilities in standby safety system fault tree analysis with testing-scheme and timing dependencies. *Reliability Engineering & System Safety 79*(1), 43–57.

Vaurio, J.K. (1998). An implicit method for incorporating common-cause failures in system analysis. *IEEE Transactions on Reliability 47*(2), 173–180.

Wang, C., L. Xing, & G. Levitin (2014). Explicit and implicit methods for probabilistic common-cause failure analysis. *Reliability Engineering and System Safety 131*, 175–184.

Xie, L., J. Zhou, & X. Wang (2005). Data mapping and the prediction of common cause failure probability. *IEEE Transactions on Reliability 54*(2), 291–296.

Xing, L. (2008). An Efficient Binary-Decision-Diagram-Based Approach for Network Reliability and Sensitivity Analysis. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 38*(1), 105–115.

Xing, L. & G. Levitin (2013). BDD-based reliability evaluation of phased-mission systems with internal/external common-cause failures. *Reliability Engineering and System Safety 112*, 145–153.

Xing, L., L. Meshkat, & S.K. Donohue (2007). Reliability analysis of hierarchical computer-based systems subject to common-cause failures. *Reliability Engineering and System Safety 92*(3), 351–359.

Xing, L., A. Shrestha, L. Meshkat, & W. Wang (2009). Incorporating common-cause failures into the modular hierarchical systems analysis. *IEEE Transactions on Reliability 58*(1), 10–19.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

A mathematical model for preliminary reliability and maintainability allocation

Z. Vintr & K. Hasilová
University of Defence in Brno, Czech Republic

M. Vintr
Independent Reliability Consultant, Brno, Czech Republic

ABSTRACT: The article deals with a simple way to allocate reliability and maintainability requirements which enables us to specify single system parts requirements preliminarily in early phases of system development and design. The method might be used for the system made of subsystems arranged into a serial structure when a failure of any subsystem leads to a whole system fault. For the whole system, however, there are no separately set requirements for its reliability and maintainability level, but overall it is necessary to achieve a certain level of availability. The method is based on a rational presumption that the requirements for a single subsystems reliability and maintainability level should be determined so that the subsystems in which a higher failure occurrence might be expected could have stricter maintainability level requirements than the systems with a higher reliability level. The suggested method is illustrated through a practical example.

1 INTRODUCTION

From a dependability point of view, a decisive property of numerous technical systems is their availability defined as the ability to be in a state to perform as required. What is really important then is not the system reliability or maintainability level itself, but their mutual combination expressed by availability.

Availability is of major importance in the systems used in non-stop operation (manufacturing technology, electric power generation and distribution, etc.), or they are involved in relatively long missions and their interruption is undesirable (means of transport, weapon systems, etc.). During designing a complex weapon system, there was a requirement specified for its availability level, and while forming the system conception, the allocation of single subsystems reliability and maintainability was already required.

The performed allocation was expected to be only preliminary and the results were to serve as a part of input information to decide about the conceptual solution of the whole system and all its subsystems. Therefore, it was required that the method should be easy to apply and provide only general information about single subsystems. Based on the overall requirement for the whole system availability level, the aim of the method was to determine general requirements for a single subsystems reliability and maintainability level. The method was also expected to be easy and fast to apply since during a development and design process a number of possible system conceptions are assumed to be formulated, therefore the method should enable us to assess quickly their advantages and disadvantages also from the dependability point of view.

In the available literature there are different methods and procedures dealing with the allocation of availability requirements. A certain part of these methods is based on the application of differently defined importance coefficients (Jigar et al. 2016; Barabady & Kumar 2007), or they use purposefully selected weighted parameters (Chang et al. 2009; Hagmark & Virtanen 2006). Also a generic algorithm and its modifications which enable a multi-criteria decision process to be made, are used when allocating availability requirements (Elegbede & Adjallah 2003; Huang et al. 2009; Oliveto 1999). The multi-criteria approach is also applied by other methods (Chiang & Chen 2007; Mohamed et al. 2002), which, when allocating the requirements, observe whether certain optimizing aims are fulfilled, e.g. minimization of system ownership costs (Kumar et al. 2007). There are also the methods based on the application of fuzzy sets in combination with a hierarchy process (Wang et al. 2012), for example, or in the form of intuitionistic fuzzy optimization (Song et al. 2015).

All the introduced approaches are based on the assumption that relatively accurate information about the system design and functions is available. In view of this fact, the possibilities of using these methods and procedures during the period of designing a basic system conception are very limited since in this phase of system development and design all necessary information has not been available yet.

Another problem is that the methods are relatively complex and their practical application is by no means trivial and might be time consuming if it is carried out for a bigger amount of system arrangement versions.

In view of these facts, we have suggested a simple method of availability requirements allocation which enables us to specify very quickly the requirements for single subsystems reliability and maintainability with only limited input information about the system design.

The requirements allocation performed with the use of the suggested method is only general and its results serve as the basis for evaluating different conceptions of the suggested system and selecting the best version.

After the system conception is clarified, it is necessary during the system development and design to perform repeated allocation of availability requirements using sophisticated allocation methods which will result in achieving final requirements for single subsystems and their components reliability and maintainability.

2 SOLUTION ASSUMPTIONS

Solution assumptions result from the above mentioned mission and the nature of the system for whose development and design the allocation method was suggested. The analysed system is characterized by the following properties:

- The considered system consists of n inter-independent subsystems (from a reliability point of view).
- All subsystems are arranged into a serial structure (from a reliability point of view).
- The failure of any subsystem results in the fault of all the system and its operation is interrupted. At any moment only one subsystem can be in a fault state.
- The system gets into an available state after the restoration of a relevant subsystem.
- Possible administrative, logistic and technical delays are not considered.
- Times between failures and times to restoration have exponential distributions.

The inherent availability of the system characterized this way might be expressed by the following formula:

$$A = \frac{MTBF}{MTBF + MTTR} = \frac{\mu}{\lambda + \mu} \tag{1}$$

where $MTBF$ is Mean Operating Time Between Failures, $MTTR$ is Mean Time to Restoration, λ is Failure Rate and μ is Repair Rate, and then the following formulas apply:

$$\lambda = \frac{1}{MTBF} \tag{2}$$

$$\mu = \frac{1}{MTTR} \tag{3}$$

For a mutual relation between the failure rate of a whole system and the failure rates of single systems for the system defined above the following formula applies (Vintr & Holub 2001):

$$\lambda = \sum_{i=1}^{i=n} \lambda_i \tag{4}$$

where λ_i is failure rate of the i-th subsystem and n is the total number of subsystems. Also the relation between repair rate μ of a whole system and repair rates of single subsystems can be expressed in a similar way. If the failure rate of single subsystems is known, the formula for a given system might be expressed in the following manner (Vintr & Holub 2001):

$$\mu = \frac{\sum_{i=1}^{i=n} \lambda_i}{\sum_{i=1}^{i=n} \frac{\lambda_i}{\mu_i}} \tag{5}$$

where μ_i is Repair Rate of the i-th subsystem.

The final equation characterizing the relation between whole system availability and the measures of single subsystems reliability and maintainability can be obtained by substituting from the equations (4) and (5) to the equation (1):

$$A = \frac{1}{1 + \sum_{i=1}^{i=n} \frac{\lambda_i}{\mu_i}} \tag{6}$$

This equation might also be transformed and then it can express the dependence of the whole system availability A on the level of single subsystems availability A_i (Vintr & Holub 2001):

$$A = \frac{1}{1 - n + \sum_{i=1}^{i=n} A_i^{-1}} \tag{7}$$

3 BASIC ALLOCATION PRINCIPLE

The nature of the solved task required that within the allocation not only the requirements for single systems availability would be specified, but also the requirements for each subsystem reliability and maintainability level would be determined. These requirements made the task solution rather complicated since at the time of preliminary allocating the requirements only general systems conception and general information about functions to be carried out by the system were known.

This information provides at least a general idea of the position and the task of each subsystem and its presumed design complexity. When it comes to the maintainability of single systems, however, it is very difficult to draw any conclusions from this information. In order to overcome this obstacle, a general principle expressing the required relation between a reliability and maintainability subsystems level has been formulated.

This principle is based on a rational requirement that the subsystems with a lower reliability level (more frequent failure occurrence might be expected) should have a higher maintainability level (they will be repaired more often) than the subsystems with a higher reliability level (they will be repaired less often). Generally, this principle says that the mean time to recovery as a subsystem maintainability measure should be directly proportional to the mean operating time between failures as a subsystem reliability measure. When applying this principle, the following should apply:

$$\frac{MTTR_i}{MTBF_i} = \frac{\lambda_i}{\mu_i} = s \tag{8}$$

where s is a required ratio between $MTTR$ and $MTBF$ in all subsystems.

After adapting the equation (6) with the use of the relation expressed by the equation (8), we obtain the following formula:

$$A = \frac{1}{1+ns} \tag{9}$$

After adapting the introduced equation, we get the formula which enables us to calculate the ratio s when knowing a required system availability level A and the number of subsystems n:

$$s = \frac{1-A}{nA} \tag{10}$$

4 ALLOCATION PROCEDURE

The initial step of the whole process is the expert determination of the requirements for a whole system reliability level in the form of a specific value $MTBF$, or a failure rate λ. This is based on the nature of the suggested system, the way of its application, and the evaluation of the consequences of a possible operation interruption caused by a failure. Also the information about the behaviour of similar systems can be used.

The next step is to introduce the weight factor which represents the ratio between single subsystems failure rate and the required failure rate of the whole system:

$$\omega_i = \frac{\lambda_i}{\lambda} \tag{11}$$

By substituting from the equation (11) to the equation (4) we then obtain basic condition which has to be fulfilled by the sum of single weight factors:

$$\sum_{i=1}^{i=n} \omega_i = 1 \tag{12}$$

Next, the weight factors have to fulfil the following condition:

$$\omega_i \geq \frac{\lambda_{i(\min)}}{\lambda} \tag{13}$$

where $\lambda_{i(\min)}$ represents a maximum acceptable requirement for a subsystem reliability level. This requirement is again determined by an expert decision and enables the subsystem reliability requirements to be rationally performable with respect to the used technologies. In fact, the maximum acceptable requirement for system reliability has to be determined so that $n\lambda_{i(\min)} < \lambda$ applies.

The introduced method requires that in the next step the weight factors for all subsystems are to be determined using an expert decision. However, the weight factors have to fulfil the requirements resulting from the equations (12) and (13), therefore this is a rather difficult task. In order to make it simple, a method of subsystems point aided estimation has been used (Vintr & Holub 2001).

When applying this method, to each subsystem a point value $a_i \in [1; a_{i(\max)}]$ is allocated. This point value characterizes a required system reliability level. The higher the required subsystem reliability level when compared with other subsystems, the higher the point value allocated to the system. The range of the used point scale, or the maximum point value $a_{i(\max)}$, is determined so that it could enable single subsystems requirements to be finely differentiated.

For the transformation of the determined point values into the weight factors value, the following equation is used:

$$\omega_i = a_i\, k + q \tag{14}$$

where real numbers k and q are linear transformation coefficients. Using this formula ensures that the determined weight factors will fulfil the conditions expressed by the equations (12) and (13). If the formula (14) is to be applied practically, it is necessary to determine the values of coefficients k and q.

With respect to the relation between the point evaluation of a relevant subsystem and its reliability level described above, the following equation must apply:

$$\omega_{i(\min)} = a_{i(\max)}\, k + q \tag{15}$$

With respect to the formula (11), this equation might be further adapted as follows:

$$\frac{\lambda_{i(\min)}}{\lambda} = a_{i(\max)}\, k + q \tag{16}$$

If we substitute into the equation (12) for ω_i the expression from the equation (14) and modify it appropriately, we get the following equation:

$$k \sum_{i=1}^{i=n} a_i + n\, q = 1 \tag{17}$$

The equations (16) and (17) form a system of two equations of two unknown variables, and after solving them, we can determine the values of coefficients k and q:

$$k = \frac{n\, \lambda_{i(\min)} - \lambda}{\lambda \left(a_{i(\max)}\, n - \sum_{i=1}^{i=n} a_i \right)} \tag{18}$$

$$q = \frac{\lambda\, a_{i(\max)} - \lambda_{i(\min)} \sum_{i=1}^{i=n} a_i}{\lambda \left(a_{i(\max)}\, n - \sum_{i=1}^{i=n} a_i \right)} \tag{19}$$

The resulting relation for determining the weight factor might be obtained by substituting the coefficients expressed this way into the equation (14):

$$\omega_i = \frac{a_i\, (n\, \lambda_{i(\min)} - \lambda) + \lambda\, a_{i(\max)} - \lambda_{i(\min)} \sum_{i=1}^{i=n} a_i}{\lambda \left(a_{i(\max)}\, n - \sum_{i=1}^{i=n} a_i \right)} \tag{20}$$

When knowing the weight factor, it is possible to determine both the requirement for each subsystem reliability level:

$$\lambda_i = \omega_i\, \lambda \tag{21}$$

and also the maintainability level requirement, using the equations (8) and (10):

$$\mu_i = \lambda_i \frac{n\, A}{(1 - A)} \tag{22}$$

5 EXAMPLE OF THE METHOD APPLICATION

The practical application of the suggested method will be demonstrated by the following example. It is required that the preliminary allocation of reliability and maintainability requirements for the system consisting of eight subsystems is to be performed. The system fulfils the initial prerequisites for the application of this method (see Chapter 2).

For the whole system, the $A = 0.997$ level availability is required and the required failure rate was determined as $\lambda = 2.5 \cdot 10^{-4}$. The maximum acceptable reliability level for single subsystems was determined in the form of a minimum acceptable value of subsystem failure rate $\lambda_{i(\min)} = 1.25 \cdot 10^{-5}$.

In the system, a point evaluation of a required single subsystems reliability level was performed when applying the point scale $a_i \in [1; 10]$. The results of the point evaluation are put in Table 1. Coefficients k and q have been calculated using the equations (18) and (19) and the following results have been obtained $k = -0.01875$ and $q = 0.2375$.

The requirements for the single subsystems reliability and maintainability level were determined on the basis of the point evaluation using the equations (20), (21) and (22). The results of the performed calculations are also put in Table 1.

Table 1. Results of calculation.

Subsystem number i	Point value a_i	Weight factor ω_i	Failure rate λ_i	Repair rate μ_i
1	1	2.19E–01	5.47E–05	1.45E–01
2	10	5.00E–02	1.25E–05	3.32E–02
3	7	1.06E–01	2.66E–05	7.06E–02
4	8	8.75E–02	2.19E–05	5.82E–02
5	6	1.25E–01	3.13E–05	8.31E–02
6	5	1.44E–01	3.59E–05	9.55E–02
7	2	2.00E–01	5.00E–05	1.33E–01
8	9	6.88E–02	1.72E–05	4.57E–02

6 CONCLUSIONS

The suggested method enables us to perform relatively simple preliminary allocation of reliability and maintainability requirements with very limited information about the analysed system. Therefore, the application of this method is suitable mainly at early stages of the system development and design when different system conceptions are taken into consideration. In this situation the method enables us to specify very quickly general requirements for single subsystems reliability and maintainability and provides inputs used for evaluating different conceptions of the suggested system and selecting the most appropriate variant.

All the allocation model can be easily made in any spreadsheet processor enabling us to evaluate very easily how the change of allocation input parameters influences the results.

A drawback of this method is that most of the input parameters have to be determined using an expert decision, and the expert's knowledge and his experience level inevitably influence the results of the performed allocation. Therefore, when determining allocation input parameters, it is advisable not to involve an individual, but a group of experienced experts.

In conclusion it is necessary to emphasize that the suggested allocation method is by its nature only preliminary and if we are to clarify the systems conception, it is necessary during its development and design to perform repeated availability requirements allocation using sophisticated allocation methods.

ACKNOWLEDGMENTS

This paper has been prepared with the support of the Ministry of Defence of the Czech Republic, Partial Project for Institutional Development, MOBAUT, University of Defence, Brno.

REFERENCES

Barabady, J. & Kumar, U. 2007. Availability allocation through importance measures. *International Journal of Quality & Reliability Management* 24 (6): 643–657.

Chang, Y.C., Chang K.H. & Liaw, C.S. 2009. Innovative reliability allocation using the maximal entropy ordered weighted averaging method. *Computers & Industrial Engineering* 57(4): 1274–1281.

Chiang, C.H. & Chen, L.H. 2007. Availability allocation and multi-objective optimization for parallel–series systems. *European Journal of Operational Research* 180(3): 1231–1244.

Elegbede, C. & Adjallah, K. 2003. Availability allocation to repairable systems with genetic algorithms: a multi-objective formulation. *Reliability Engineering and System Safety* 82(3): 319–330.

Hagmark, P.E. & Virtanen, S. 2006. Specification and allocation of reliability and availability requirements. In *Proceedings of Annual Reliability and Maintainability Symposium, RAMS '06*: 304–309. Washington: IEEE Computer Society.

Huang, H.Z., Qu, J. & Zuo, M.J. 2009. Genetic-algorithm-based optimal apportionment of reliability and redundancy under multiple objectives. *IIE Transactions* 41(4): 287–298.

Jigar, A.A., Haskins, C. & Lundteigen, M.A. 2016. Availability Allocation Using Systems Engineering Principles. In *Proceedings—International Conference on Industrial Engineering and Operations Management:* 1488–1497. Singapore: IEOM Society International.

Kumar, D., Ramirez-Marquez, J.E., Nowicki, D. & Verma, D. 2007. Reliability and Maintainability Allocation to Minimize Total Cost of Ownership in a Series-Parallel System. *International Journal of Risk & Reliability* 221(2): 133–140.

Mohamed, A.A. Ravindra, A. & Leemis, L. 2002. An interactive multicriteria availability allocation algorithm. *International Journal of Operations and Quantitative Management* 8(1): 1–19.

Oliveto, F.E. 1999. An Algorithm to Partition the Operational Availability Parts of an Optimal Provisioning Strategy. In *Proceedings of Annual Reliability and Maintainability Symposium, RAMS '99*: 310–316. Washington: IEEE.

Song, C.H., Guo, L.H., Wang N.C. & Ma, L. 2015. Availability Optimization and Allocation for Repairable Systems Using Intuitionistic Fuzzy Optimization. In *Prognostics and System Health Management Conference*. New York: IEEE.

Vintr, Z. & Holub, H. 2001 R&M Requirements Allocation in Upgrading a System. In *Proceedings of Annual Reliability and Maintainability Symposium*: 258–263. Philadelphia: IEEE.

Wang, S, Li,S., Zhou, J., Li, Q., & Kang, L. 2012. Reliability Allocation for CNC Machine Based on Improved Fuzzy Analytic Hierarchy Process. *Advances in Information Sciences and Service Sciences* 4(1): 320–327.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Methodology for the preparation of accelerated reliability testing of electronic components in combat vehicles

X.P. Cu
Faculty of Military Technology, University of Defence, Czech Republic

H.A. Bui
Department of Foreign Languages, University of Fire Fighting and Prevention in Hanoi, Vietnam

ABSTRACT: Accelerated Reliability Testing (ART) consists of tests designed to determine the reliability information of a product such as failure intensity, failure probability or survival, or time to failure of a particular product. ART is divided into three phases: preparation phase, implementation phase, evaluation phase. This paper describes the methodology for the preparation of ART of electronic components used in combat vehicles. Before ART is performed, it is extremely important to do test planning including sample selection, test condition selection, ART method selection, determination of stresses level and test duration, etc. The successful design of an ART requires good knowledge of the intended use condition, environmental impact profile, operation and design of the product. Electronic components in combat vehicles often operate in harsh environments and they are generally expected to be subject to higher levels of stress than common commercial electronic components.

1 INTRODUCTION

Reliability is ability to perform as required, without failure, for a given time interval, under given conditions. The time interval duration can be expressed in units appropriate to the item concerned, e.g. calendar time, operating cycles, distance run, etc., and the units should always be clearly stated. Given conditions include aspects that affect reliability, such as: mode of operation, stress levels, environmental conditions, and maintenance (IEC 60050-192). Because of the reliability of a system or a device is mainly dependent on the reliability of its components, the evaluation of the reliability of the components is very important to understand the reliable life of the overall systems and devices.

Reliability of electronic component is estimated by using the standard handbook, statistical analysis of operation & maintenance data or performing reliability testing experiments (Varde 2010). In general, the use of standard handbooks, like MIL-HDBK-217 approach, has some inherent limitations as it does not allow simulations with projected component load profiles; relatively large uncertainties that are associated with various parameters and hence, in the final results; no provisions to assess the root cause of component failure; it is not effective in predicting the reliability of new components or new design of conventional components, etc. Hence, another approach to determine the reliability of electronic components should be using accelerated test methods.

Accelerated test is testing in which the stress level, or rate of stress application, exceeds that occurring under specified operational conditions, to reduce the duration required to produce a stress. The object of the methods is to either identify potential design weakness or provide information on item dependability or to achieve necessary reliability/availability improvement, all within a compressed or accelerated period of time response (IEC 62506).

Accelerated testing methods help investigate the reliability of electronic components as regards certain dominant failure mechanisms under normal operating conditions, provide details about the various degradation mechanisms and thereby improved understanding of the root cause of the failure (Varde 2010).

Accelerated tests are commonly used in the automotive and electrotechnical industries to assess or demonstrate component and system reliability, to detect failure modes which may occur during the life of product, to compare different manufacturers, etc. For military purposes, accelerated tests are possible to apply for the electronic components in combat vehicles.

Common reasons for failure in electronic components are environmental contaminants and conditions, such as temperature, thermal cycles,

humidity and other failures deriving (e.g. vibration, ripple voltage, and overvoltage). Based on the cumulative damage model, expected test information, and assumptions about product usage, accelerated test methods can be divided into three types (IEC 62506):

- Qualitative accelerated tests (type A);
- Quantitative accelerated tests (type B);
- Quantitative time and event compressed tests (type C).

Qualitative accelerated tests (type A) are used primarily to identify failures and failure modes without attempting to make any predictions as to the product's life under use conditions (Reliasoft Corporation 2001). Commonly used qualitative testing models are HALT (Highly accelerated limit tests), HAST (Highly accelerated stress test), HASS/HASA (Highly accelerated stress screening/audit). Qualitative tests are performed with the specimens subjected to a single severe level of stress, to multiple stresses, or to a time-varying stress (e.g., stress cycling, cold to hot, etc.). The results of the tests are then used to increase the margin of strength of the design. Moreover, they provide valuable information such as the types and levels of stresses to employ during a subsequent quantitative test.

Quantitative accelerated tests (type B) are designed to quantify the life characteristics of the product component or system under using conditions, and thereby provide reliability information (Reliasoft Corporation 2001). The purpose of quantitative accelerated testing is to estimate one or more reliability information such as failure intensity, failure probability or survival, or time to failure. It can also be used to assist in the performance of risk assessments or design comparisons.

Quantitative time compressed tests (type C1) are achieved by excluding the "shutdown time" i.e. by focusing the test just on the switch-on time. However, only focusing on the operating time can ignore damage during downtime. Quantitative event compressed tests (type C2) are used to repeat events with greater intensity than the product usage in practice. With this type of testing, some negative consequences can be created by using continuous stress, in some way causes failures that do not occur under normal conditions.

Accelerated testing type A is used in product design phase (new electronic component development, modernization of existing components applying new technologies). Accelerated testing type B and C is used to quantify the reliability parameters of the electronic components, thereby minimizing the lack of reliability information of electronic components used in combat vehicles.

This paper presents the process of designing a quantitative accelerated test applied to electronic components in combat vehicles.

2 ELECTRONIC COMPONENT IN MODERN COMBAT VEHICLES

Today's modern combat vehicles achieve the parameters of the major military characteristics (tactical-technical parameters) through a high proportion of electronic components with digital control. The digitization of military technologies is one of the key requirements of individual armaments, newly purchased combat vehicles, or weapon systems. These electronic components contribute significantly to the achievement of the reliability of vehicles. In-vehicle electronic systems can be divided into control systems, protection systems and information systems (Chloupka 2012). Electronic control systems control mechanical, hydraulic, and other functions (e.g. fuel injection, steering, braking). In modern combat vehicles, engine, transmission, active suspension and brake systems, are controlled by electronic control unit (ECU).

Protection systems work with digital and analog information, and they do not directly connect to control links, for example, diagnostic system, navigation system, security and detection systems. Data collected by protection systems will be shared with other systems, or directly trigger protective elements. The function of these systems is controlled by the control algorithm.

Information systems work with data unrelated to control links. The crew can track and share information from the different vehicles that are aggregated and displayed on the display (e.g., displaying the vehicle's own position on a map). The crew also can distribute information to defined systems in combat vehicle (weapon systems, protection systems) and then can make decisions in combat activities.

Electronic components in-vehicle electronic systems are generally expected to be subject to higher levels of stress than commercial electronic components, mainly temperature, vibrations, shocks, voltage, dustiness and humidity. The basic requirements for electronic components in combat vehicles are shown in Table 1.

Electronic components act in the same way no matter whether for military or civilian purposes; therefore, the standardized procedures of accelerated tests have been already available for commercial electronic components, they can be applied or modified for use on electronic components of combat vehicles (Vintr et al. 2013).

Table 1. The basic requirements for electronic components in combat vehicles.

Parameter	Value
MTTF	15 years
Power supply	18 V to 33 V
Ambient Operating Temperature	–40°C to 70°C
Storage temperature	–40°C to 80°C
Vibration	10 G, 40 to 500 Hz
Mechanical shock	500 G, 0.2 ms to 2 ms
Repeated mechanical shock	20 G, 1 ms to 15 ms
Relative humidity	98% at 25°C
Chemical resistance	1 mg/m^3
Dust resistance	2 g/m^3
Impulse voltage resistance	70 V/3 ms

3 MODELS FOR ACCELERATED RELIABILITY TESTING

These are the models of stress during the life of a product when the damage after a unit of test time is appropriately accelerated by increasing the stress level. The underlying assumption when using any of these models is that the components operating under normal conditions experience the same failure mechanism as those occurring at the accelerated stress conditions. It is assumed that the time-scale transformation or acceleration factor is constant and hence implies linear acceleration (Chaluvadi 2008).

3.1 *Arrhenius model*

Temperature is commonly used as an environmental stress for testing of electronic devices. This is generally modeled using the Arrhenius reaction rate, which is used for constant temperature stresses and is based on the assumption that absolute temperature is due to the emergence of certain mechanisms of failure. The Arrhenius reaction rate equation is given by (IEC 62506):

$$R(T) = Ae^{-\frac{E_a}{k_B T}} \tag{1}$$

where $R(T)$ is the speed of reaction; A is a constant (which is not a function of temperature); E_a is the activation energy (eV); k_B is the Boltzman's constant, $k_B = 8{,}617385.10^{-5}$ (eV/K); T is the absolute temperature (K). The Arrhenius life-stress relationship is given by:

$$L(T) = Ce^{\frac{D}{T}} \tag{2}$$

The relationship is linearized by taking the natural logarithm of both sides in the Arrhenius equation or:

$$\ln[L(T)] = \frac{D}{T} + \ln(C) \tag{3}$$

where $L(T)$ represents a quantifiable life measure such as *MTTF*, characteristic life, etc.; $\ln(C)$ is the intercept of the line; D is the slope of the line ($= E_a/k_B$).

Since the Arrhenius is a physics-based model derived for temperature dependence, it is used for temperature accelerated tests. For the same reason, temperature values must be in absolute units (Kelvin), even though the Arrhenius equation is unitless.

Acceleration factor is the ratio of the stress response rate of the test specimen under the accelerated conditions, to the stress response rate under specified operational conditions.

$$A_{F\_Arr} = \frac{R(T_{Test})}{R(T_{Use})} = \frac{Ce^{-\frac{E_a}{k_B T_{Test}}}}{Ce^{-\frac{E_a}{k_B T_{Use}}}} = e^{\frac{E_a}{k_B}\left(\frac{1}{T_{Use}} - \frac{1}{T_{Test}}\right)} \tag{4}$$

where T_{Use} and T_{Test} are use temperature and test temperature. Arrhenius model is applied to a plurality of statistical distributions in reliability analysis. Its applicability is determined by conditions where exposure to thermal stress is expected by a constant temperature. The model is not applicable for damage caused by low temperature. For such types of damage, it is recommended that a failed test be performed to determine a specific model.

3.2 *Eyring model*

This model is most often used when thermal stress (temperature) is the acceleration variable, which is similar to the Arrhenius model. However, the Eyring model is also used for stress variables other such as humidity. The relationship is given by (IEC 62506):

$$L(S_E) = \frac{1}{S_E} e^{-\left(A - \frac{B}{S_E}\right)} \tag{5}$$

where $L(S_E)$ represents a quantifiable life measure, such as *MTTF*, characteristic life, median life, etc.; S_E is the stress level (temperature values are in absolute units Kelvin). A and B are model parameters to be determined by the test or approximated by the literature values. The acceleration factor in this model is:

$$A_{F\_E} = \frac{L(S_{E\_Test})}{L(S_{E\_Use})} = \frac{S_{E\_Use}}{S_{E\_Test}} e^{B\left(\frac{1}{S_{E\_Use}} - \frac{1}{S_{E\_Test}}\right)} \quad (6)$$

where S_{E_Use} and S_{E_Test} are the use stress level and accelerated stress level.

The Eyring model can be used for all types of mathematical probability distributions that are normally used in reliability analysis. Mathematical levels of reliability can be determined for individual parameters or functions based on appropriate statistics.

3.3 *Inverse Power Law (IPL) model*

This model is commonly used to test electronic devices when the stress is dynamic stresses such as shock, vibration or climatic stresses such as temperature cycles, temperature changes, humidity. The relationship is given by (IEC 62506):

$$L(S) = \frac{1}{CS^n} \quad (7)$$

where: $L(S)$ represents a quantifiable life measure, such as *MTTF*, characteristic life, etc.; S is the stress; C is the model parameters to be determined; n is the model parameter dependent on the behavior of the stress to be determined.

The parameter n in the inverse power relationship is a measure of the effect of the stress on the life. As the absolute value of n increases, the greater the effect of the stress. Negative values of n indicate an increasing life with increasing stress. An absolute value of n approaching zero indicates small effect of the stress on the life, with no effect (constant life with stress) when $n = 0$.

The IPL appears as a straight line when plotted on a log-log paper. The equation of the line is given by:

$$\ln[L(S)] = -nL(S) - \ln(C) \quad (8)$$

For the IPL relationship, the acceleration factor is given by:

$$A_{F\_IPL} = \frac{L(S_{Use})}{L(S_{Test})} = \frac{\frac{1}{CS_{Use}^n}}{\frac{1}{CS_{Test}^n}} = \left(\frac{S_{Test}}{S_{Use}}\right)^n \quad (9)$$

where S_{Use} and S_{Test} are the use stress level and accelerated stress level.

3.4 *Coffin-Manson model*

This model is used to test electronic devices when the stress is a thermal cycle. The failure mechanism is thermal cracking. The equation of the model is (Cui 2005):

$$N = \frac{C}{\Delta T^{\gamma}} \quad (10)$$

where $\Delta T = T_{\max} - T_{\min}$ is the temperature range; $T_{\max}$ is the high extreme temperature; $T_{\min}$ is low extreme temperature; N is the number of cycles to failure; C and γ are properties of the material and test setup. The acceleration factor is given by:

$$A_{F\_CM} = \left(\frac{\Delta T_{test}}{\Delta T_{use}}\right)^{\gamma} \quad (11)$$

where ΔT_{use}, ΔT_{test} are temperature range under normal operation and temperature range under test operation.

4 DESIGN OF ACCELERATED TEST OF ELECTRONIC COMPONENTS

Before performing tests, it is fundamental to plan and determine the characteristic of the sample and the stress-scheme (De Carlo et al. 2014). Planning activities are performed respecting the specifications, the budget and the time constraints. The purpose of ART is to provide objective and reproducible data on the reliability of the object. This requires that the test conditions described in the test plan and the methods used to process the test results are as reproducible as possible and that the test samples used are representative. The steps for designing the ART of electronic components in combat vehicles are presented below.

- Select electronic components in combat vehicles;
- Determine operating conditions of electronic components;
- Decide stress types, levels and numbers;
- Decide sample numbers;
- Calculate acceleration factors and test duration.

4.1 *Electronic components selection in combat vehicles*

If electronic components have not been identified before, based on the results of the analysis below, it is possible to identify representative electronic component for the implementation of ART.

- Analyze overview about each electronic component used in the combat vehicle;
- Analyze the importance of electronic components for the overall functions of the combat vehicle;

- Analyze tactical and technical data of individual electronic components, design, technological level of electronic components (processor technology, processor boards, and displays);
- Analyze the location of electronic components in combat vehicles;
- Specify the applicability of the results obtained by the accelerated test to other similar electronic components.
- Analyze the reliability data of an electronic component specified by the manufacturer, or it may be routed for the purpose of an accelerated test.

4.2 *Operating conditions of electronic components*

The conditions for performing the ART are typically determined on the basis of normal operation of the combat vehicles. Modern combat vehicles have the diagnostic system that archives the operating time for selected components, allowing for the setting of accelerated test parameters. Military equipment is often used as follows (Vintr et al. 2013):

- Free state, during which the electronic components are switched off – up to 90% of operating life time;
- Operating state, during which the electronic components are switched on – up to 10% of operating life time.

4.3 *Stress types, level and number*

In the automotive and electrotechnical industries, when performing accelerated tests, the most commonly used stresses are temperature, temperature cycling, humidity, vibrations/shocks, voltage (overvoltage/low voltage) or combination of these stresses. The different types of stress have different effects on the load of the electronic component and its failure rate, which is reflected in the value of acceleration factor and accelerated test time. By combining different types of stress during an accelerated test, it is possible to achieve higher overall acceleration factor, which means a substantial shortening of cumulative test time. Accordingly, it is necessary to understand the behavior of electronic components with different types of stress by performing tests on statistical samples. In general, procedures designed for commercial electronic components can also be applied to electronic components of combat vehicles. However, it must be remembered that military elements do not behave similarly to commercial electronic components. For this reason, it is also necessary to analyze the accelerated test procedure of commercial electronic component before it is applied to electronic components of combat vehicles.

When performing the ART of electronic components in combat vehicles, there are four types of stress should be focused, namely temperature cycling, temperature, humidity and vibration. The temperature is the basic stress of electronic components. It is sometimes said that temperature is the enemy of reliability of all electronic components. Whether they are used for commercial or military, these components are stressed by the environmental temperature when turned off and the operating temperature when turned on. The temperature cycling which particularly stresses on the solder joints of the electronic components results in a total failure of the electronic components (for example breaking of solder joints, loss of conductive contact, and deformation of the processor plate). The basic factors that affect the failure rate of electronic components by temperature cycling are:

- Type of solder is used and its properties;
- Temperature rise in electronic components;
- Minimum temperature difference (e.g., on/off).

During the switch off state the electronic components of combat vehicles are stressed by ambient temperature. Depending on the season of the year, the idle state does not have a profound effect on an average annual temperature.

During the operation of military equipment, the electronic components are switched on and by their function they generate their own heat. After switch of the equipment it reaches its operating temperature. The resulting temperature of an electronic part depends then on the ambient temperature at which the electronic part is switched on. In order to find out the operating temperatures of electronic parts placed in combat vehicles, an experiment has to be performed.

When determining temperature cycling, it is necessary to find out the time and the number of the switch on/off of electronic components during operation. Modern combat vehicles with electronic components are able to record and archive the time of switch on/off including an hour, a day and a year.

For other stresses such as humidity and vibration, the stress level is also based on the operating conditions and operating environment of combat vehicles. Then, the experiments are performed with selected military electronic components. Based on this information, it is possible to specify the types of stresses or combinations which have a major influence on the reliability or the failure rate of the electronic components. The determination of stress level in test depends on operating limits of electronic components and technical parameters of test equipment such as temperature range, temperature rate of change, humidity range.

The required information for the test calculation is shown in Table 2.

4.4 *The number of samples in test*

The goal of ART affects the determination of sample numbers, the extent of accelerated test, types of damages, possibilities of failure diagnosing, the method of failure resolution.

In qualitative accelerated tests (type A), the sample selection range is determined by the number of stresses and the number of failures detected. For example, in the classic HALT, it requires one sample for low temperature, one for high temperature, one for vibration, one for temperature cycles and one for the combined test by temperature cycles and vibrations; therefore, total is 5 samples. In order to account for more than one failure mode, it is recommended to use another 2 to 5 samples, so the recommended total selection range is 7 to 10 objects (IEC 62506).

In quantitative accelerated tests (type B and C), the number of samples is determined mainly by estimating the constant failure rate or time to failure. A typical selection range for accelerated testing is 77 samples for 1000 h (JESD 47G). In the case of exponential distribution, standards for test plans such as IEC 61123 and IEC 61124 can be used. For a test using Weibull's distribution, at least 5 to 10 failures are expected. Because Weibull's test is often stopped, when one-third of the objects under test fails, the selection range is 15 to 30 objects (IEC 62506).

Table 2. Operating conditions and types of stresses.

Parameter	Symbol	Value (unit)
Operation time		
Switch on	t_{on}	h
Switch off	t_{off}	h
Temperature		
Switch on	T_{on}	K
Switch off	T_{off}	K
Test temperature	T_{test}	K
Temperature change	ΔT_{use}	K
Temperature change in test	ΔT_{test}	K
Total number of cycles	N_{use}	
Speed of temperature change	ξ_{use}	K/min
Speed of temperature change in test	ξ_{test}	K/min
Vibration	W_{use}	m/s^2
Vibration in test	W_{test}	m/s^2
Relative humidity	RH_{use}	%
Humidity in test	RH_{test}	%
Required lifetime	t_0	h
Required reliability	$R_0(t_0)$	
Activation energy	E_a	eV

4.5 *Acceleration factor and test duration*

4.5.1 *Temperature*

When stress is temperature, it is possible to use the Arrhenius reaction model or the Eyring model to calculate the acceleration factor. The Eyring model contains constants A and B, and for military electronic elements, it has to be determined by other tests. To determine them, extensive experiments are needed (on a statistically significant number of electronic components). For products of medium complexity, precision acceleration tests may become problematic because of the different components and materials have different values of constants B; therefore, the accelerated tests of electronic components often use the Arrhenius reaction model. According to the standard GS 95003-1-2010 prescribed in vehicle electrical and electronic components using the Arrhenius reaction model. When using the Arrhenius reaction model is necessary to set the value of activation energy E_a, which directly affects the magnitude of the acceleration factor. The values of activation energy with different standards are shown in Table 3. Figure 1 shows the effect of activation energy E_a on the acceleration factor.

The acceleration factor is given by Arrhenius reaction model:

$$A_T = e^{\frac{E_a}{k_B}\left(\frac{1}{T_{on}}-\frac{1}{T_{test}}\right)} \tag{12}$$

The time with the temperature at switch off state is normalized to the temperature at switch on state:

$$t_{on\_N} = t_{on} + t_{off} e^{-\frac{E_a}{k_B}\left(\frac{1}{T_{off}}-\frac{1}{T_{on}}\right)} \tag{13}$$

The test duration of the ART is:

$$t_{T\_test} = k\frac{t_{on\_N}}{A_T} = k t_{on\_N} e^{-\frac{E_a}{k_B}\left(\frac{1}{T_{on}}-\frac{1}{T_{test}}\right)} \tag{14}$$

Table 3. The values of activation energy (Chloupka 2012).

Electronic components	MIL217F (1996)	RDF (1996)	MIL217 plus (1996)	FIDES (2009)
Bipolar logic	0.4 eV	0.4 eV	0.8 eV	0.7 eV
CMOS logic	0.35 eV	0.3 eV	0.8 eV	0.7 eV
BICmos logic	0.5 eV	0.4	0.8 eV	0.7 eV
Linear	0.65 eV		0.8 eV	0.7 eV
Memories	0.6 eV		0.8 eV	0.7 eV
VHSIC	0.4 eV		0.8 eV	0.7 eV

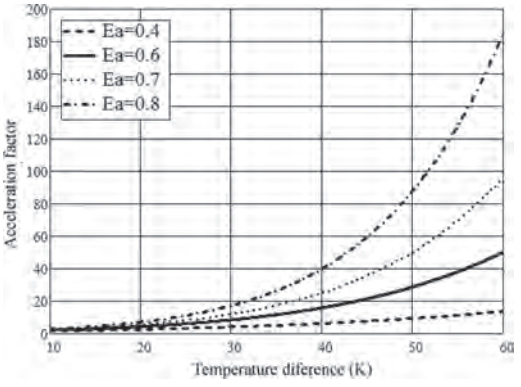


Figure 1. Acceleration factor with different values E_a.

where k is the multiplier of the actual duration of the stress, is determined from the graph in the standard IEC 62506.

4.5.2 *Temperature cycling*
The acceleration factor is given by Coffin-Manson model:

$$A_{TC} = \left(\frac{\Delta T_{use}}{\Delta T_{test}}\right)^{-n} \left(\frac{\xi_{use}}{\xi_{test}}\right)^{-1/3} \tag{15}$$

where n is the temperature cycle exponent (n typical value is around 2).

Number of cycles in test is given by:

$$N_{test} = k\frac{N_{use}}{A_{TC}} = kN_{use}\left(\frac{\Delta T_{use}}{\Delta T_{test}}\right)^{n} \left(\frac{\xi_{use}}{\xi_{test}}\right)^{1/3} \tag{16}$$

Figure 2 shows the schematic of temperature cycle profile. The temperature cycle profile can be characterized by: High extreme temperature (T_{max}), low extreme temperature (T_{min}), temperature change (ΔT), Ramp rates and dwell times at extreme temperatures (Cui 2005). The test duration of a temperature cycle is:

$$t_{TC} = 2\frac{\Delta T_{test}}{60\xi_{test}} + \frac{t_{T\_test}}{N_{test}} + t_{T\_low} \tag{17}$$

where t_{T_low} is dwell time at the lower temperature extreme.

4.5.3 *Humidity*
During the humidity test, the acceleration is achieved by increasing the relative humidity during the test as well as by increasing the test temperature above the expected values during use.

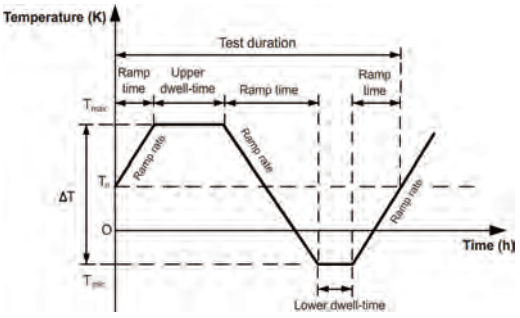


Figure 2. Schematic of temperature cycle profile.

Thermal acceleration is the same as the temperature exposure test for the full equivalent time using the t_{on_N} for the switching temperature. The acceleration factor for the humidity test is given by Hallberg-Peck model:

$$A_{RH} = \left(\frac{RH_{use}}{RH_{test}}\right)^{-h} e^{\frac{E_a}{k_B}\left(\frac{1}{T_{on}} - \frac{1}{T_{RH}}\right)} \tag{18}$$

where T_{RH} is temperature with humidity test, h is the exponent for the acceleration factor caused by humidity according to IPL model.

The test duration of exposure in humidity is given by:

$$t_{RH\_test} = k\frac{t_{on\_N}}{A_{RH}} = kt_{on\_N}\left(\frac{RH_{use}}{RH_{test}}\right)^{h} e^{-\frac{E_a}{k_B}\left(\frac{1}{T_{on}} - \frac{1}{T_{RH}}\right)} \tag{19}$$

4.5.4 *Vibration*
With stress is vibration, it is possible to use the IPL model to calculate the acceleration factor.

$$A_W = \left(\frac{W_{use}}{W_{test}}\right)^{-w} \tag{20}$$

where w is the exponent for the acceleration factor caused by vibration according to IPL model. In the absence of the test-specific constant, it is usually assumed that $w = 4$.

The test duration is given by:

$$t_{W\_test} = k\frac{t_{W\_use}}{A_W} = kt_{W\_use}\left(\frac{W_{use}}{W_{test}}\right)^{w} \tag{21}$$

where t_{W_use} is operation time in vibration test. Typically, for electronic components in automo-

Table 4. Calculation results for ART.

Stress Test: Temperature Acceleration Model: Arrhenius						
T_{on} [K]	T_{off} [K]	t_{on} [h]	t_{off} [h]	T_{test} [K]	A_T [–]	t_{T_test} [h]
Stress Test: Temperature cycling Acceleration Model: Coffin Manson						
ΔT_{use} [K]	ΔT_{test} [K]	ξ_{use} [K/min]	ξ_{test} [K/min]	N_{use} [–]	A_{TC} [–]	N_{test} [–]
Stress Test: Temperature Humidity Bias Acceleration Model: Hallberg-Peck						
RH_{use} [%]	RH_{test} [%]	T_{RH} [K]	A_{RH} [–]	t_{RH_test} [h]		
Stress Test: Vibration Acceleration Model: Inverse power law						
W_{use} [m/s^2]	W_{test} [m/s^2]	t_{w_use} [h]	A_W [–]	t_{w_test} [h]		
Combined stresses						
t_0 [h]	$R_0(t_0)$ [–]	A [–]	t_{test} [h]			

tive, one hour of non-accelerated vibration test is estimated with distance traveled approximately 1600 km; therefore, $t_{W\_use} = D/1600$ (h), where D is total distance traveled in kilometer of automotive.

4.5.5 *Combined stresses*

To determine the overall acceleration factor, it will be assumed that temperature cycles and vibration cause the same failure mode while temperature and humidity cause another failure mode. The overall acceleration factor is given by:

$$A = \frac{A_{TC}A_W + A_T A_{RH}}{4} \tag{22}$$

Then the total test duration is:

$$t_{test} = -\frac{t_0}{\ln(R_0(t_0))A} \tag{23}$$

The calculation results for ART of electronic components are shown in the Table 4.

5 CONCLUSIONS

This article presents the steps for designing a quantitative ART of electronic components in combat vehicles. Commonly used stress types for electronics components are temperature, temperature cycling, vibration, humidity or combination of stresses. However, in practice, it is necessary to analyze the type of electronic components, operating conditions, operating environments, and capabilities of the existing test equipment to select the type of characteristic stress for the failure mode of electronic components. The ART method of electronic components provided in the commercial industry can be applied for electronic components in combat vehicles after the detailed analysis. The next step is conducting the ART and evaluate the results under actual operating conditions.

ACKNOWLEDGMENTS

This paper has been prepared with the support of the Ministry of Defence of the Czech Republic, Partial Project for Institutional Development, MOBAUT, University of Defence, Brno.

REFERENCES

Chaluvadi, V. 2008. *Accelerated Life Testing of Electronic Revenue Meters*. All Theses. South Carolina: Clemson University.

Charki, A., Laronde, R., Guérin, F., Bigaud, D. & Coadou, F. 2011. Robustness evaluation using highly accelerated life testing. *The International Journal of Advanced Manufacturing Technology* 56(9): 1253–1261.

Chloupka, J. 2012. *Accelerated reliability testing of electronic elements of military vehicles (in Czech)*. PhD Thesis. Brno: University of Defence.

Cui, H. 2005. Accelerated Temperature Cycle Test and Coffin-Manson Model for Electronic Packaging. *Reliability and Maintainability Symposium* 2005: 556–560.

De Carlo, F., Borgia, O., & Tucci, M. 2014. Accelerated degradation tests for reliability estimation of a new product: a case study for washing machines. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 228(2): 127–138.

Escobar, L.A. & William, Q.M. 2006. A review of accelerated test models. *Institute of Mathematical Statistics* 21(4): 552–577.

GS 95003-1. 2010. *Electrical/Electronic assemblies in Motor vehicle, General information*. München: BMW Group Standard.

Hobbs, G.K. 2000. *Accelerated reliability engineering: HALT and HASS*. New York: John Wiley & Sons.

IEC 60050-192. 2015. *International electrotechnical vocabulary – Part 192: Dependability*. IEC.

IEC 62506. 2013. *Methods for product accelerated testing*. IEC.

JEDEC Standard. 2009. *Stress-Test-Driven Qualification of Integrated Circuits, JESD47G*. Arlington: JEDEC Solid State Technology Association.

MIL-HDBK-217F. 1991. *Reliability prediction of electronic equipment.* Washington, D.C.: US Department of Defense.

Nelson, W.B. 2009. *Accelerated testing: statistical models, test plans, and data analysis*. New York: John Wiley & Sons.

Qingchuan, H., Wenhua, C., Jun, P. & Ping, Q. 2012. Improved step stress accelerated life testing method for electronic product. *Microelectronics Reliability* 52(11): 2773–2780.

Reliasoft Corporation. 2001. *Accelerated life testing reference*. Tucson: ReliaSoft Publishing.

Varde, P.V. 2010. Physics-of-failure based approach for predicting life and reliability of electronics components. *Barc Newsletter* 313: 38–46.

Vintr, Z. 2010. Accelerated reliability test for the automobile electronic system (in Czech). In 39-th Meeting of the Expert group on reliability, Praha. *Czech Society for quality*: 21–26.

Vintr, Z., Vališ, D. & Chaloupka, J. 2013. Accelerated tests time of electronic parts used in military vehicles. *Journal of Science of the Gen. Tadeusz Kosiuszko Military Academy of Land Forces* 169: 145–151.

Yu, Z., Ren, Z., Tao, J. & Chen, X. 2014. Accelerated testing with multiple failure modes under several temperature conditions. *Mathematical Problems in Engineering* 2014: 1–8.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Approximation method for reliability of one-unit repairable system with time redundancy

Xiaoyue Wu & Haiyue Yu
School of Systems Engineering, National University of Defense Technology, Changsha, China

ABSTRACT: For some engineering repairable systems like the spaceflight Telemetry, Tracking and Control (TTC) system, mission can be executed within a given time interval and been regarded as successful so long as the system remains in operational state for a minimum length of time. Thus, such system has time redundancy in mission execution and higher mission reliability. This paper presents a discrete approximation method to numerically calculate the mission reliability of such kind of system with one repairable unit and binary states. The time window is divided into a number of time slices and then the mission reliability of the system is derived by solving two groups of recursive discrete time equations. All these equations are established by decomposing the corresponding random event into a number of disjoint events at discrete time points with reduced time lengths. A numerical example is provided for a one unit system. The results of the example are compared with those previously obtained by other solution methods. It was shown that the proposed methods are computationally efficient, and the approximated missionreliability converges to the analytical results when the width of the divided time slices decreases.

1 INTRODUCTION

In engineering practice, there are systems that have time redundancy in mission execution. Specifically, such system needs to work continuously for a minimum length of time to accomplish prescribed critical missions within a given time interval (called time window). A typical example of such system is the space flight Telemetry, Tracking and Control (TT&C) system that provide services such as orbit tracking, remote control and data transmission to spacecrafts during their fly over the ground stations (Wu 2014, Guest 2013). Sometimes, the TT&C service requires only a short time duration within the time window when the spacecraft fly over, so time redundancy exists in such cases.

Although many research works have been done on mission reliability of systems, not enough attention has been given to mission reliability of system with time redundancy. Many existing works are on phased-mission system(PMS) which accomplishes missions in consecutive phases and requires that the system must keep in operational state for all the time duration of the mission phase (Xing & Amari 2008). The existing methods for mission reliability generally includes analytical methods, and simulation methods. Models based on Boolean algebra, Binary Decision Diagrams (BDD), fault trees, and Continuous Time Markov Chains (CTMC) are among the most commonly used analytical methods (Kim & Park 1994, Xing & Amari 2008). To avoid the strict restriction on the modeling assumptions and computational complexity, simulation models like Petri nets, Monte Carlo can provide alternative options (Wu & Wu 2015, Wu & Hillston 2016), but may have precision and computation time issues.

To evaluate the mission reliability of such kinds of systems, the effect of time redundancy should be taken into consideration to avoid underestimating the mission reliability. For repairable one-unit systems, Wu presented a model by decomposing the mission success event into a number of disjoint events, and gave general formulas for its computation (Wu 2014). Later, for semi-Markov systems with multiple units, Wu and Hillston gave an analytical approach based on matrix integral equations for mission reliability evaluation (Wu & Hillston 2015). The integral matrix equations are numerically computed by time discretization approximation. However, both of these analytical methods involving complicated integral expressions and numerical computations. Monte Carlo methods have also been applied for reliability evaluation of such systems (Wu & Guo 2017, Wu & Hillston 2016). The main purpose of this paper is to present a new analytical approach for evaluating mission reliability of repairable one-unit systems with time redundancy. As the first step for mission reliability evaluation, we divided the time window into a number of time slices, and then build recursive equations for numerical computation. Thus, it is different from our previous approaches in the

order of discretizationas we do time discretization firstly and then build reliability evaluation model afterwards.The new approach is expected to be more effective and more computationally efficient.

2 MISSION RELIABILITY MODEL

2.1 *Basic assumptions*

In this paper, we make the following assumptions:

- the system has one unit that is repairable and has two states: operational state (or up state) and failure state (or down state).
- the time to failure and repair time both follow exponential distribution.
- the unit after repair is "as good as new".
- the system is required to accomplish its mission within the time window $[0,T]$.
- the mission is successfully accomplished if the system can keep in up state continuously for time duration no less than T_d.

2.2 *Equations for mission reliability*

Let the interval $[0,T]$ be divided into N intervals as follows.

$$0 =< w_0 < w_1 < \cdots < w_{n-1} < w_N = T \tag{1}$$

where

$$\begin{aligned} w_i &= i\delta, \quad i = 1,\cdots,N \\ \delta &= T / N \end{aligned} \tag{2}$$

Let

$$d = \min_i |T_d - i\delta| \tag{3}$$

Assume that the time to failure of the system W and the repair time Z follow exponential distribution with failure rate λ, and repair rate μ respectively, namely $W \sim Exp(\lambda), Z \sim Exp(\mu)$.

Now we introduce the following notations

P_{ud}^k: the probability that the system starts from up state and remains in up state continuously until at time w_k enters into down state.

P_{du}^k: the probability that the system starts from down state and remains in down state continuously until at time w_k enters into up state.

$P_u(n,d)$: the probability that the system has an operational span not less than T_d within time interval $[0,T_n]$, and the system starts at up state.

$P_d(n,d)$: the probability that the system has an operational span not less than T_d within time interval $[0, T_n]$, and the system starts at down state.

Thinking from renewal points of view, we obtain the following equations

$$P_u(n,d) = \sum_{k=d}^{\infty} P_{ud}^k + \sum_{k=1}^{d-1} P_{ud}^k \cdot P_d(n-k,d) \tag{4}$$

$$P_d(n,d) = \sum_{k=1}^{n-d} P_{du}^k \cdot P_u(n-k,d) \tag{5}$$

Obviously, the boundary conditions are

$$P_u(n,d) = 0 \quad n < d \tag{6}$$

$$P_d(n,d) = 0 \quad n \le d \tag{7}$$

The mission reliability can be obtained by calculating $P_u(N,d)$, which is the probability that the system starts in up state and has at least a continuous time duration T_d of operational state during time interval $[0,T], T = \delta N$.

2.3 *Probabilities of sojourn time*

For implementing the previously established model, it is necessary to give methods for calculating the sojourn time probabilities corresponding to different starting states of the system.

Since both the working time and repair time of the system follows exponential distribution, we can describe the behavior of the system with a binary state continuous time Markov chain (CTMC) model.

By the CTMC theory (Ross 2010), for $k > 0$, we have

$$\begin{aligned} P_{ud}^k &= exp(-\lambda w_k)\left(1 - exp(-\lambda\delta)\right) \\ &= exp(-\lambda k\delta)\left(1 - exp(-\lambda\delta)\right) \end{aligned} \tag{8}$$

$$\begin{aligned} P_{du}^k &= exp(-\mu w_k)\left(1 - exp(-\mu\delta)\right) \\ &= exp(-\mu k\delta)\left(1 - exp(-\mu\delta)\right) \end{aligned} \tag{9}$$

3 ALGORITHM FOR SOLVING EQUATIONS

3.1 *Recursive equations*

For brevity, we define the following notations:

$$G_n = P_u(n,d) \tag{10}$$

$$F_n = P_d(n,d) \tag{11}$$

$$A_k = P_{ud}^k \tag{12}$$

$$B_k = P_{du}^k \tag{13}$$

$$C = \sum_{k=d}^{\infty} A_k = exp(-\lambda w_d) = exp(-\lambda d\delta) \tag{14}$$

Then, the equations (4) and (5) can be rewritten as

$$G_n = C + \sum_{k=1}^{d-1} A_k \cdot F_{n-k} \tag{15}$$

$$F_n = \sum_{k=1}^{n-d} B_k G_{n-k} \tag{16}$$

The associated boundary conditions are

$$G_n = 0 \quad n < d \tag{17}$$

$$F_n = 0 \quad n \le d \tag{18}$$

which make equations (15),(16) more specific

$$G_n = C + \sum_{k=1}^{M(n,d)} A_k \cdot F_{n-k} \tag{19}$$

$$F_n = \sum_{k=1}^{n-d} B_k G_{n-k} \tag{20}$$

where $M(n,d) = \min\{d-1, n-d\}$.

Thus, we obtain recursive equations for find the mission reliability $G_N = P_u(N,d)$.

3.2 *Solution procedure*

Based on the previous results, the mission reliability of the system can be solved with the following computational procedure.

Step 1 : Initialization

Set N, δ, d.

Set $G_n^{old} = 0, F_n^{old} = 0, \quad n = 0, \cdots, N$.

Step 2 : Calculate C, and $A_k, B_k, k = 1, \cdots, N$.

Step 3 : from $n = 1$ to $n = N$, calculate

$$G_n^{new} = C + \sum_{k=1}^{M(n,d)} A_k \cdot F_{n-k}^{old} \tag{21}$$

$$F_n^{new} = \sum_{k=1}^{n-d} B_k G_{n-k}^{old} \tag{22}$$

Step 4 : if the difference between G^{new} and G^{old} is small enough, then stop. The value of G_N^{new} will be the mission reliability.

Otherwise, set $G^{old} \leftarrow G^{new}$, $F^{old} \leftarrow F^{new}$, go back to Step 3.

4 NUMERICAL EXAMPLE

To verify the proposed model and the solution algorithm, we use an example of one-unit system (Wu 2014), the time to failure and repair time of the unit are both of exponential distributions. The failure rate and repair rate of the unit are $\lambda = 1/60$ and $\mu = 1/10$ respectively. The system is required to work continuously for $T_d = 60$ within the given time interval [0, T], $T = 100$. For this example, the mission reliability obtained by other methods is 0.533, including analytical and simulation methods (Wu 2014, Wu & Hillston 2015).

Table 1. Mission reliability for different N.

N	$R = G_N$	N	$R = G_N$
100	0.5093	2000	0.5322
200	0.5208	2500	0.5324
500	0.5283	3000	0.5326
800	0.5302	3500	0.5327
1000	0.5308	4000	0.5328
1200	0.5313	4500	0.5329
1500	0.5317	5000	0.5329

Table 2. Mission reliability after different iterations.

Iteration	3	5	8	10	15
$R = G_N$	0.5100	0.5313	0.5328	0.5329	0.5329

The proposed method is used to solve the mission reliability. To study the approximation precision of the time discretization, different value of N for the fixed time window [0, T] is used to compute the mission reliability. The results are shown in Table 1.

In order to see the convergence performance of our recursive algorithm for solving equations (19) and (20), the mission reliability calculated after different number of iterations are provided in Table 2, where the number of discretization is set as $N = 4500$.

From Table 2, we can see that our algorithm has fast convergence speed. After about 8 iterations of the recursive equations, the value of the mission reliability sufficiently approaches the analytical result 0.533.

By comparison of the mission reliability results in the two tables, we can find that the number of discretization of time window has more influence on the precision of the obtained results. When the length of time slice becomes about 1/50 of the length of the time window, it can be expected that the obtained mission reliability reach satisfied precision.

5 CONCLUSIONS

For one-unit repairable systems with time redundancy that requires only a minimum length of operational time within a time window for its mission

success, we present an analytical approximation approach to calculate the mission reliability. Based on the system behaviors at discrete time points, the recursive equations for mission reliability are established for different system starting states and remaining time durations. Moreover, an iterative numerical computation procedure is presented for obtaining the system mission reliability. This approach is different from our previous approach by making time discretization before building the recursive equations. Results of numerical example shows that the proposed method provides an efficient way for mission reliability evaluation of systems with time redundancy. As future research work, we will try to extend the model and algorithm to more complicated cases.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (grant No.71671185).

REFERENCES

Guest, A.N. (2013). *Handbook of Satellite Applications*, Volume 1, Chapter Telemetry, Tracking, and Command (TT&C), pp. 1067–1078. New York: Springer.

Kim, K. & K.S. Park (1994, Jun). Phased-mission system reliability under Markov environment. *IEEE Transactions on Reliability 43*(2), 301–309.

Ross, S.M. (2010). *Introduction to probability models* (10 ed.). Academic press.

Wu, X. (2014, July). Mission reliability model for repairable system with minimum operation time requirement. In *Proceedings for 8th IMA International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR)*, University of Oxford,UK, pp. 348–351. Institute of Mathematics and its Applications. 10–12 July 2014.

Wu, X. & B. Guo (2017, July). An object-oriented simulation model for reliability of PMS with time redundancy. In *2017 IEEE International Conference on Software Quality, Reliability and Security*, pp. 185–189. IEEE Reliability Society:

IEEE Computer Society Conference Publishing Services (CPS).

Wu, X. & J. Hillston (2015). Mission reliability of semi-Markov systems under generalized operational time requirements. *Reliability Engineering & System Safety 140*(Supplement C), 122–129.

Wu, X. & J. Hillston (2016, July). Monte carlo simulation for reliability estimation of phased-mission systems with minimum operational time requirement. In H.-Z. Huang, H. Xu, and D. Meng (Eds.), *The Proceedings of 2016 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering & 2016 World Congress on Engineering Asset Management*, Volume 1, JiuZhaigou, Sichuan, China, pp. 99–104.

Wu, X.Y. & X.Y.Wu (2015). Extended object-oriented Petri net model for mission reliability simulation of repairable PMS with common cause failures. *Reliability Engineering & System Safety 136*, 109–119.

Xing, L. & S.V. Amari (2008). *Handbook of Performability Engineering* (1 ed.)., Chapter Reliability of Phased-mission Systems, pp. 349–368. Springer.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A reliability analysis method for complex mechanical systems containing probabilistic-interval information

W.S. Peng, M. Xu, C.H. Zeng & Z. Bian
China Aero-Polytechnology Establishment, Beijing, China

J.G. Zhang
Beihang University, Beijing, China

ABSTRACT: In this paper, a new reliability analysis method for complex mechanical system which contains hybrid uncertainties is proposed. The hybrid uncertainties containing randomness and intervals are considered in the limit-state function of reliability. Middle Point Limit-State (MPLS) model is first proposed for calculating the reliability index of hybrid uncertainties, and this model could make the index interval smaller and accurate. For complex mechanical systems, system hierarchy technique is used in reliability modeling, and composite limit-state function is then obtained to calculate the system reliability. By constructing the optimization model of the composite limit-state function, the reliability index interval can be obtained, and artificial bee colony algorithm is employed to solving these optimization models. The reliability index interval obtained through the method proposed by this paper could be more precise comparing to existing techniques. The reliability indexes containing the interval and MPLS index may be a more useful analysis tool to assess the mechanical systems reliability with hybrid uncertainties. A numerical case is carried out to illustrate this method and an engineering case of the satellite driving system is used as a typical complex mechanical system to demonstrate the analysis method proposed in this paper.

1 INTRODUCTION

The complexity of modern mechanical system generally emerges in two sides: first, the system structure and its components are complex, mechanical systems are typical multi-level systems, and there are many elements in every level. Second, the uncertain information of mechanical system is complex. Depending on the amounts of the statistic information, some variables or parameters have sufficient statistical information and they are probabilistic variables; some other variables lack enough statistical information, so they could only be given upper and lower values of interval.

While dealing with the hybrid uncertainties of interval and probability, some researches have been emerged in the structure reliability analysis field. Techniques as: transforming the interval uncertainties to random uncertainties, then calculating the reliability with classical random reliability theory such as First Order Reliability Method (FORM), Second Order Reliability Method (SORM) (Hurtado et al 2012, Hurtado 2013); assuming the interval variables obey uniform distribution in the interval range, and the non-probabilistic reliability index is used to estimate the reliability of structures (Jiang et al. 2013, Wang et al 2010, Qiu et al. 2008). These methods are based on the assumption that the interval variables obey random distributions. This assumption has neglected the objective fact that interval variables have no enough information but only the upper and lower boundaries. In order to overcome the above deficient, some other techniques like tow-step method has been developed for the hybrid uncertainties structural system (Jiang et al. 2011, Du 2007, Guo et al. 2009), this technique is not suitable for the system reliability analysis, the main problem is the interval extension because of the Interval Arithmetic Method (IAM). With the complexity of the mathematic process, the range of the output interval will become too large to accept. Interval-truncation approach (Lu et al. 2002, Zhao et al. 2008) is developed to reduce interval extension, and it could make sense in simple interval arithmetic equations, but is not suitable for the complex problems, especially for complex mechanical systems. Other epistemic uncertainty theory such as fuzzy possibility (Singh et al. 2008) and dempster-shafer evidence theory (Flage et al. 2011, Zio et al. 2012) could also deal with the interval uncertainty, but these theories are suitable for specific uncertainty structure as fuzzy set, basic belief assignment. The uncertainty measures such as possibility, belief and plausibility are used for reliability assessment.

This paper aims to develop an efficient and relative accurate reliability analysis method for complex mechanical system with hybrid uncertainties.

A Middle Point Limit-State (MPLS) model is first introduced in this paper to deal with the hybrid uncertainties of interval and probability, and the composite limit-state function and optimization model for complex mechanical system reliability analysis are also proposed by this paper. First the system reliability model is built with the hierarchy model of the system, and then the composite limit-state function of the system is obtained according to different levels and logic gates of the hierarch systems. The optimization model for calculating the system reliability index is constructed and Artificial Bee Colony (ABC) algorithm is employed to solve it. The optimization model method has been used in the fuzzy industrial system analysis (Fales 2010, Harish et al. 2012, Sharma et al. 2009, Garg et al. 2013). The main contribution of this method is that the Optimization Model Based Method (OMM) instead of IAM, is used to calculate the system reliability, and the final results of reliability interval will be more precise. It should be noted that the mechanical systems in the paper are considered with two states and the basic logic gates "AND" and "OR" are used.

The remainder of this paper is organized as follows. Random distribution with interval parameters are employed to quantify the uncertainty, and the MPLS model is introduced for this case in section 2; section 3 are reliability model establishment of complex mechanical system; section 4 demonstrates the reliability analysis method based on the optimization model and ABC algorithm; a numerical case and an engineering case are carried out in section 5 to demonstrate this method. Finally, a conclusion is given in section 6.

2 MIDDLE POINT LIMIT-STATE MODEL FOR HYBRID UNCERTAINTIES

Define the limit-state function Z of mechanical system as follows:

$$Z = g(\boldsymbol{X}, \boldsymbol{Y}) \tag{1}$$

where $\boldsymbol{X} = \{x_1, x_2, \cdots, x_n\}$ are random variables, $\boldsymbol{Y} = \{y_1, y_2, \cdots, y_m\}$ are interval variables, and the upper and lower boundary values of $\boldsymbol{Y}$ are given:

$$\boldsymbol{Y} = [\boldsymbol{Y}_{\min}, \boldsymbol{Y}_{\max}] \tag{2}$$

The mean value and deviation of $\boldsymbol{Y}$ are defined as:

$$\boldsymbol{Y}_{\mathrm{m}} = \frac{\boldsymbol{Y}_{\min} + \boldsymbol{Y}_{\max}}{2} \tag{3}$$

$$\boldsymbol{Y}_{\mathrm{r}} = \frac{\boldsymbol{Y}_{\max} - \boldsymbol{Y}_{\min}}{2} \tag{4}$$

Because of the interval variables $\boldsymbol{Y}$, the limit-state function Z is an area composed of two surfaces in the probabilistic space. It is shown as in Figure 1.

For reliability analysis of the hybrid uncertainty limit-state function Z, this paper constructs a "middle point limit-state (MPLS)" model to calculate the reliability index β. In the MPLS model, a new limit-state function is constructed when all the interval variables are chosen as the middle point in the interval ranges. And this limit-state function is defined as:

$$Z_{\mathrm{m}} = g(\boldsymbol{X}, \boldsymbol{Y}_{\mathrm{m}}) = 0 \tag{5}$$

When using the FORM or SORM, The equation (1) could be rewritten as:

$$Z = g(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{g}(\mathrm{T}(\boldsymbol{U}, \boldsymbol{Y})) = G(\boldsymbol{U}, \boldsymbol{Y}) \tag{6}$$

where $\boldsymbol{U} = \{u_1, u_2, \cdots, u_n\}$ is the standard normal variables transferred from $\boldsymbol{X}$.

The reliability index β^{m} of the Z_{m} can be obtained by solving the follow optimization model:

$$\begin{cases} \beta^{m} = \min\limits_{U} \|\boldsymbol{U}\| \\ s.t. \quad G(\boldsymbol{U}, \boldsymbol{Y}_{\mathrm{m}}) = 0 \end{cases} \tag{7}$$

For reliability index interval $\beta^{\mathrm{I}} = [\beta^{\min}, \beta^{\max}]$, the following two optimization problems can be solved.

$$\begin{cases} \beta^{\min} = \min\limits_{U} \|\boldsymbol{U}\| \\ s.t. \quad \min\limits_{Y} G(\boldsymbol{U}, \boldsymbol{Y}) = 0 \end{cases} \tag{8}$$

and:

$$\begin{cases} \beta^{\max} = \min\limits_{U} \|\boldsymbol{U}\| \\ s.t. \quad \max\limits_{Y} G(\boldsymbol{U}, \boldsymbol{Y}) = 0 \end{cases} \tag{9}$$

The redefinition of reliability index interval β^{I}. Through the above work, three reliability indexes

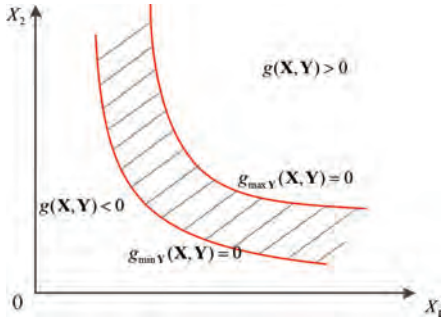


Figure 1. State space of performance function including interval variables.

$\beta^{\min}, \beta^{\max}$ and β^{m} is obtained, then the reliability index interval β^{I} of Z can be defined according to the following cases:

case 1:

If $\beta^{\min} \geq \beta^{m}$, then $\beta^{I} = [\beta^{m}, \beta^{\max}]$.

case 2:

If $\beta^{\max} \leq \beta^{m}$, then $\beta^{I} = [\beta^{\min}, \beta^{m}]$.

case 3:

If $\beta^{\min} < \beta^{m} < \beta^{\max}$, then $\beta^{I} = [\beta^{\min}, \beta^{\max}]$, and β^{m} is an accessory reliability index. The final reliability index is defined as:

$$\beta = \begin{cases} \beta^{I} \\ \beta^{m} \end{cases} \tag{10}$$

The reliability is:

$$R = \begin{cases} R^{I} = \Phi(\beta^{I}) \\ R^{m} = \Phi(\beta^{m}) \end{cases} \tag{11}$$

3 COMPOSITE LIMIT-STATE FUNCTION METHOD FOR MECHANICAL SYSTEM RELIABILITY ANALYSIS

3.1 *Hierarchy model of mechanical system*

For complex mechanical system, the reliability analysis is full of challenge because of the complexity both in structure and uncertainty information, especially when the amount of its elements is too large. Due to this reason, system hierarchy technique is an efficiency way for complex mechanical reliability analysis.

The system hierarchy method is as shown in Figure 3. In this method, the mechanical system is divided into different levels according to its structure and components. Generally, the top level is the system and the bottom level contains the variables and parameters of basic elements. Middle levels are sub-systems or parts of the mechanical system. Different levels are jointed by the logic gates.

In the hierarchy model of the system reliability, logical operations of the gates in the hierarchy model are appointed as follows.

For the "AND" logical gate, the reliability is:

$$R_{\text{AND}} = \prod_{i} R_i \tag{12}$$

For the "OR" logical gate, the reliability is:

$$R_{\text{OR}} = 1 - \prod_{i} (1 - R_i) \tag{13}$$

It should be noted that all the components or failure modes are independent with each other.

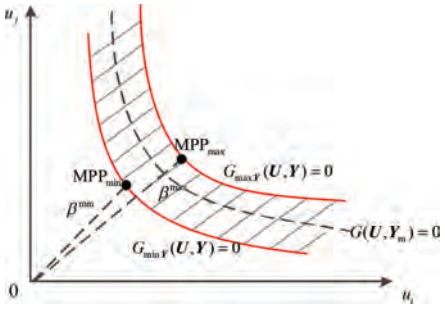


Figure 2. Reliability index of hybrid uncertainty.

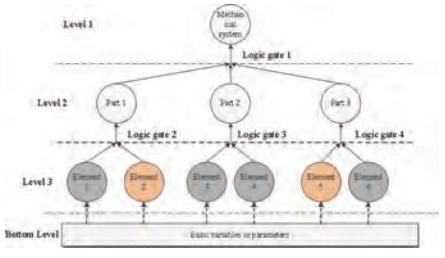


Figure 3. The hierarchy model of mechanical system.

3.2 *Composite limit-state function and calculation of system reliability*

In section 3.1, the complex mechanical system is modeled by different levels and logic gates according the system hierarchy method. Then this section proposes the construction of the composite limit-state function and how to calculate the system reliability. The Figure 3 is used to illustrate this method. Through this figure, we could see that the bottom levels are the basic variables and parameters, and the variables are both probabilistic and interval in this paper. Then every element in level 3 could have the limit-state functions based on its failure modes and variables, and these limit-state functions are defined as:

$$\boldsymbol{Z}_{\text{level3}} = \left\{ Z^{1}_{level3}, Z^{2}_{level3}, \cdots, Z^{6}_{level3} \right\} \tag{14}$$

$$Z^{i}_{level3} = g\left(\boldsymbol{X}_i, \boldsymbol{Y}_i\right) \tag{15}$$

The index Z^{i}_{level3} represents limit-state of the ith element in level 3. The indexes $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$ represent the probabilistic variables and interval variables of the ith element.

For level 2, the limit-state function of the part 1, part 2 and part 3 could be obtained based on the index $\boldsymbol{Z}_{\text{level3}}$ and the logic gate. And these limit-state functions are:

$$\boldsymbol{Z}_{\text{level2}} = \left\{ Z^1_{level2}, Z^2_{level2}, Z^3_{level2} \right\} \tag{16}$$

$$Z^1_{level2} = L\left(Z^1_{level3}, Z^2_{level3} \right) \tag{17}$$

In which $L(\bullet)$ is a logic function. For level 1, the limit-state function of the mechanical system could be obtained through the index $\boldsymbol{Z}_{\text{level2}}$ and the logic gate 1, and the limit-state function is as:

$$Z_{sys} = L\left(Z^1_{level1}, Z^2_{level2}, Z^2_{level3} \right) \tag{18}$$

and Z_{sys} is defined as the composite limit-state function of the mechanical system. However, the composite limit-state function is not suitable for using directly, especially for the large systems with too much variables, and the limit-state function in this paper are considered as the hybrid uncertainty limit-state function. So this paper employs the reliability index instead of the limit-state function. According to MPLS in section 2, every limit-state function of element in the level 3 could get a reliability index interval β^{I} and R^{I}, then, the reliability of the composite limit-state function Z_{sys} could be defined as:

$$R^{\text{I}}_{sys} = L\left(R^{\text{I}}_1, R^{\text{I}}_2, \cdots, R^{\text{I}}_i \right) \tag{19}$$

where the function principle of L is the same as the equation (19).

After calculation, the index R^{I}_{sys} is also an interval and defining it as:

$$R^{\text{I}}_{sys} = \left[R^{\min}_{sys}, R^{\max}_{sys} \right] \tag{20}$$

where $R^{\min}_{sys}$ and $R^{\max}_{sys}$ are the lower and upper boundary of the interval respectively.

4 RELIABILITY ANALYSIS OF COMPLEX MECHANICAL SYSTEM

4.1 *Optimization models for calculating the reliability*

For the reliability computation of R^{I}_{sys}, the optimization model based method (OMM) is utilized instead of the interval arithmetical method (IAM). The boundary values of R^{I}_{sys} in equation (20) could be obtained by solving the optimization problems as follows:

For $R^{\min}_{sys}$

$$\begin{aligned} &\text{Minimize} && g\left(R^{\text{I}}_1, R^{\text{I}}_2, \cdots, R^{\text{I}}_i \right) \\ &\text{Subject to} && R^{\min}_i \le R^{\text{I}}_i \le R^{\max}_i \end{aligned}$$

For $R^{\max}_{sys}$

$$\begin{aligned} &\text{Maximize} && g\left(R^{\text{I}}_1, R^{\text{I}}_2, \cdots, R^{\text{I}}_i \right) \\ &\text{Subject to} && R^{\min}_i \le R^{\text{I}}_i \le R^{\max}_i \end{aligned}$$

According to (Garg & Sharma 2012), the smaller of the interval range, the more meaningful of the system reliability decision. For the complex mechanical system, the optimization model is obviously a high nonlinear problem, thus a high efficient optimization algorithm is needed here. Artificial bee colony (ABC) algorithm is proved to be an effective algorithm (Peng et al. 2013), and this paper will use it to calculate the optimization models to get the system reliability.

The whole steps of the reliability analysis method proposed in this paper are as follows:

Step 1: analyze the structure and components of the mechanical system, build the reliability model with system hierarchy technique;

Step 2: identify the bottom elements of the system, collect and extract the uncertainty information of these elements;

Step 3: build the reliability limit-state functions of these elements, if it is implicit problem, the RSM is needed. Then the composite limit-state function of the system is identified;

Step 4: calculate the reliability index interval β^{I} and R^{I} of the elements based on the MPLS method proposed in the section 2, and then construction the optimization model for solving the boundary of the R^{I}_{sys} of the composite limit-state function Z_{sys};

Step 5: optimize the model with ABC algorithm, the mechanical system reliability index R^{I}_{sys} and R^{m}_{sys} are obtained.

Artificial Bee Colony (ABC) algorithm was proposed by Karaboga for optimizing numerical problems (Karaboga 2007, 2008). The algorithm simulates the intelligent foraging behavior of honeybee swarms. It is a very simple, robust and population based stochastic optimization algorithm.

5 CASE STUDY

5.1 *The numerical case*

Assume the hierarchy model of a mechanical system as shown in Figure 4, after system hierarchy, there are 5 levels in the mechanical system S, the bottom level is the basic variables. There are 9 elements in the level 4, and all these elements are considered with hybrid uncertainties. Their basic variables and limit-state function are given. The 5 logic gates G_1~G_5 are as follows: G_1, G_2 and G_4 are "OR" gates, G_3 and G_5 are "AND" gates. Now we calculate reliability index of S with the method proposed in this paper.

Denote the reliability performance functions of the bottom elements E_1~E_9 as follows:

$$\begin{cases} g(\boldsymbol{X},\boldsymbol{Y})_{E_1} = x_1 \cdot y_1 / x_3^2 - 0.8\, x_2 / y_1 \\ g(\boldsymbol{X},\boldsymbol{Y})_{E_2} = 567 \cdot x_4 \cdot x_5 - 0.5 y_2^2 \\ g(\boldsymbol{X},\boldsymbol{Y})_{E_3} = x_6 \cdot x_7^2 + \sqrt{2} / x_7 - y_3 \\ g(\boldsymbol{X},\boldsymbol{Y})_{E_4} = x_8 - \sqrt{300 x_9^2 + 1.92 y_4^2} \\ g(\boldsymbol{X},\boldsymbol{Y})_{E_5} = \exp\left[0.4(y_5 + 2) + 5.02\right] \\ \quad - \exp\left[0.3 y_6 + 500\right] - 200 x_{10} x_{11}^2 \\ g(\boldsymbol{X},\boldsymbol{Y})_{E_6} \\ \quad = y_7 - 6 x_{11} \cdot y_8 / (x_{12}^2 \cdot x_{13}) - 6 x_{11} \cdot y_9 / (x_{12} \cdot x_{13}^2) \\ g(\boldsymbol{X},\boldsymbol{Y})_{E_7} \\ \quad = x_{14}^3 \cdot y_{10} \cdot y_{11} / x_{15} + \sqrt{2} x_{15}^2 / \sqrt{y_{10}} - 326 \\ g(\boldsymbol{X},\boldsymbol{Y})_{E_8} \\ \quad = 0.01 - (48 x_{16} + 32) / (18 x_{17} + 3) \cdot y_{12} / (y_{13} y_{14}) \\ g(\boldsymbol{X},\boldsymbol{Y})_{E_9} = \sqrt{x_{18}} \cdot x_{19} / y_{15} - x_{20} / y_{16} \end{cases} \tag{21}$$

The uncertainties of all variables are as shown in Table 1 and Table 2. Based on the MPLS model proposed in this paper, the reliability index β_i. of the bottom elements of the system could be obtained, and the results are listed in Table 3.

Construction of the composite limit-state function of the system S. According to the system structure and its logic gates, the composite limit-state function can be obtained as:

$$\begin{cases} G_{sys}(\boldsymbol{X},\boldsymbol{Y}) = L_{\mathrm{OR}}(\mathrm{g}_{M_1}, \mathrm{g}_{M_2}) \\ \mathrm{g}_{M_1} = L_{\mathrm{OR}}\left(g_{E_1} \sim g_{E_4}\right) \\ \mathrm{g}_{M_2} = L_{\mathrm{AND}}\left(\mathrm{g}_{M_3}, \mathrm{g}_{M_4}\right) \\ \mathrm{g}_{M_3} = L_{\mathrm{OR}}\left(g_{E_5}, g_{E_6}\right) \\ \mathrm{g}_{M_3} = L_{\mathrm{AND}}\left(g_{E_7}, g_{E_8}, g_{E_9}\right) \end{cases} \tag{22}$$

where L is the function according to the logic operation (12) and (13). Similarly, the system reliability R_S^{I} could be obtained according the equation (23)

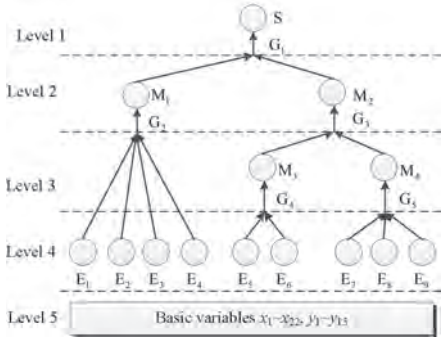


Figure 4. Hierarchal model of the system S.

$$\begin{aligned} R_{sys} &= R_1 \cdot R_2 \cdot R_3 \cdot R_4 \\ &\cdot \left[1 - (1 - R_5 \cdot R_6) \cdot (1 - R_7) \cdot (1 - R_8) \cdot (1 - R_9)\right] \end{aligned} \tag{23}$$

The system reliability R_{sys}^{m} is calculated directly according to equation (23), it is: $R_{sys}^{\mathrm{m}} = 0.94960402$.

The lower boundary value of the system reliability R_{sys}^{I} is

$$\begin{aligned} &\text{minimize} && R_{sys} \\ &\text{subject to} && R_{E_i} \in \left[R_{E_i}\right] \end{aligned}$$

The upper boundary value of the system reliability R_S^{I} is

$$\begin{aligned} &\text{maximize} && R_{sys} \\ &\text{subject to} && R_{E_i} \in \left[R_{E_i}\right] \end{aligned}$$

The ABC is used to optimize the above models, it has been implemented in Matlab (MathWorks) and the program has been run on a T6400@2GHz Intel Core (TM) 2 Duo processor with 2 GB of Random Access Memory (RAM). The selected values of all the parameters for ABC are given as:

Colony size (CS) = 20 × number of subunits.
Limit for scout = (CS × D)/2, where D is dimension (number of variables)
Number of generation = 500.

The termination criterion has been set to a maximum number of generations of order of relative error equal to 10^{-6}, whichever is achieved first.

The results of the ABC algorithm are: 0.93429182 (the lower boundary) and 0.96392416 (the upper boundary). The optimization process is as shown in Figure 5 and Figure 6.

6 RESULTS AND DISCUSSION

According to the reliability analysis method proposed in this paper, when all variables are the mean values, the reliability of system is $R_{sys}^{\mathrm{m}} = 0.94960402$. The index R_{sys}^{I} is as follows: if using IAM directly, the reliability interval R_{sys}^{I} is [0.91771156, 0.96653623], if using OMM, the reliability interval R_{sys}^{I} with ABC algorithm is [0.93429182, 0.96392416]. Comparing to the IAM result, the interval range obtained through the OMM is smaller, especially is the lower boundary of the interval. The lower boundary value is reduced about 0.01658026. The upper boundary value is almost the same with the IAM result, which is reduced about 2.61207e-3. This is because the reliability interval of all the bottom elements are smaller

Table 1. Uncertain information of random variables.

Variables	Distribution	Mean	Variance	Variables	Distribution	Mean	Variance
x_1	Normal	26	0.1	x_{11}	Normal	1000	100
x_2	Normal	3.2	0.05	x_{12}	Normal	100	15
x_3	Weibull	7.2	0.07	x_{13}	Lognormal	200	20
x_4	Lognormal	0.6	0.131	x_{14}	Weibull	10	0.1
x_5	Lognormal	2.18	0.03	x_{15}	Normal	4	0.085
x_6	Normal	2.2	0.01	x_{16}	Extreme I	0.3	0.003
x_7	Normal	12.5	0.1	x_{17}	Lognormal	0.1	0.001
x_8	Extreme I	48	3	x_{18}	Normal	152	0.01
x_9	Normal	1.0	0.16	x_{19}	Normal	60	0.085
x_{10}	Weibull	1.0	0.001	x_{20}	Normal	1.6	0.02

Table 2. Uncertain information of interval variables.

Variables	Upper boundary	Lower boundary	Variables	Upper boundary	Lower boundary
y_1	3	2.8	y_9	27	23
y_2	33.78	31.72	y_{10}	4.6	4.2
y_3	222	230	y_{11}	1.2	0.8
y_4	22	18	y_{12}	20.5	20
y_5	2	0	y_{13}	2.003e6	2e6
y_6	1.5	0.5	y_{14}	0.021	0.018
y_7	380	360	y_{15}	23	20
y_8	53	47	y_{16}	2.4	2

Table 3. Reliability of different parts.

Part	Reliability index β^I	Reliability index β^n
E_1	2.4119,3.1729	2.8012
E_2	1.7269,2.3996	2.0679
E_3	1.8211,1.9646	1.8925
E_4	4.4907,6.1194	5.3166
E_5	2.9614,3.4732	3.3013
E_6	5.8321,5.8923	5.8772
E_7	3.0761,3.4638	3.3587
E_8	2.7731,3.4253	2.8196
E_9	2.8845,3.2076	3.0313

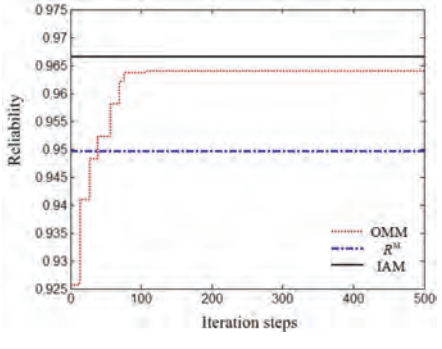


Figure 5. Optimization process of the upper boundary.

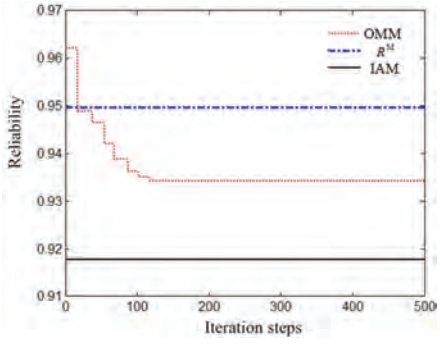


Figure 6. Optimization process of lower boundary.

than 1. It also could be seen that the interval reduction in this case is not obvious in this case, this can be explain that the system is not complex enough.

The spread width to the index R_{sys}^{m} is as follows: through IAM, the left spread width is 0.03189246, and the right spread width is 0.01693221. Through OMM, the left spread width is 0.0153122, and the right spread width is 0.01432014. Suppose the index R_{sys}^{m} is the most sensitivity point in the interval of the index R_{sys}^{I}, then interval obtained by OMM is more sensitive than the result obtained by IAM.

In order to illustrate the advantages of the ABC algorithm, other intelligence optimization algorithms as particle swarm optimization (PSO) algorithm and genetic algorithm (GA) are also used for these same optimization models. The results are listed in Table 4. It shows that ABC algorithm and PSO algorithm could get almost the same results. The results demonstrate that ABC and the PSO are both good at the global searching ability in the optimization process. But PSO need more steps (177) than ABC (102) to convergence to a stable solve. The GA convergences very quickly (only 77 steps), but the optimization results is much rougher than the ABC and PSO.

6.1 *Engineering case*

A satellite driving system is composed of a stepper motor, a harmony reducer, a photoelectric encoder and a microswitch. The CAD model of the satellite driving system is as shown in Figure 7. After structure analysis, the hierarchy model of the driving system as shown in Figure 8. The event in top level is the failure of driving system.

Through the calculation and statistical experiments, in the driving system, the reliability indexes of the bottom elements are shown in Table 5.

Based on the hierarchy model in Figure 9, the system reliability is:

$$R_S = R_{E_1} \cdot R_{E_4} \cdot R_{E_5} \cdot R_{E_6} \cdot R_{E_7} \cdot \left[1-\left(1-R_{E_2}\right)\cdot\left(1-R_{E_3}\right)\right] \cdot R_{E_8} \cdot R_{E_9} \cdot R_{E_{10}} \cdot R_{E_{11}} \tag{24}$$

The upper and lower boundary values of the system reliability could be obtained by solving follow optimization models.

$$\begin{aligned} &\text{minimize/maximize} && R_S \\ &\text{subject to} && R_{E_i} \in \left[R_{E_i}\right] \end{aligned}$$

Based on the data in Table 5, the reliability R_S^{m} could be obtained directly and it is 0.97313495. For R_S^{I}, the result of the OMM is $R^{\mathrm{I}} = [0.9485038, 0.9911342]$ and the result of the IAM is $R^{\mathrm{I}} = [0.9338912, 0.9933270]$. It can be shown that the interval obtained by OMM is obviously smaller than the interval obtained by IAM, it means that the OMM can reduce the interval extension in this problem. Through the Figure 10, we can see that ABC algorithm could maintain in the optima margin stably for a high nonlinear problem. The OMM based on the ABC algorithm could get precise interval result for a complex system, and it is meaningful for the decision making.

The left spread width and right spread width to the Rm is as shown in Table 6. The spread width of the OMM is smaller than the IAM. Compared to the IAM, the left spread width of the OMM reduce about 0.0145918(37.1825%), and the right spread width reduce 0.0021858(10.82878%).

Table 4. Results of different algorithm.

Methods	Lower boundary value	Upper boundary value	Iteration steps
ABC	0.93429182	0.96392416	102
PSO	0.93420751	0.96398842	177
GA	0.92682667	0.96738521	74

Figure 7. CAD model of the driving system and the key components.

Figure 8. Hierarchy model of the driving system.

Table 5. Reliability of the bottom elements.

Elements	R^I	R^M
P_1	0.9998	0.9998
P_2	[0.9998,0.9999]	0.99984
P_3	[0.9998,0.9999]	0.999835
P_4	[0.99972,0.99986]	0.99981
P_5	[0.9980,0.9985]	0.999826
P_6	[0.99982,0.99987]	0.999858
P_7	[0.9999,0.99998]	0.99994
P_8	0.999	0.999
P_9	[0.999,0.9998]	0.9996
P_{10}	[0.980,0.999]	0.9906
P_{11}	[0.9575,0.9975]	0.9845

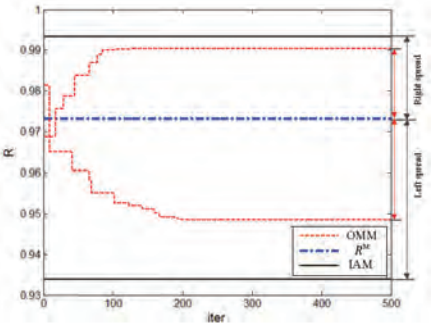


Figure 9. Optima results and the interval math results.

Table 6. Reliability of the bottom elements.

Methods	Left spread width	Right spread width
IAM	0.0392437	0.0201851
OMM	0.0246311	0.0179993

7 CONCLUSION

In this paper, a reliability analysis method is proposed for complex mechanical system. This method overcomes the complexity difficult of uncertain information brought by the completeness difference of the statistic information. The reliability index β^{I} as well as β^{m} of the middle point limit state function is introduced to comprehensively assess of the hybrid uncertainty system reliability. This method could avoid both subjective assumption to interval variables and information waste of random variables.

Composite limit-state function is used to build the reliability model of the complex mechanical system, and it could deal with the complex of large variables and information. Optimization model based technique is employed to solve the system reliability index interval, and advance intelligence ABC algorithm is utilized in the optimization models. This optimization model based method could conquer the difficulties brought by the complexity of the system structure, and also could avoid the interval extension induced by the IAM. Through this method, the higher sensitivity zone of the reliability interval will be achieved; this is very useful and meaningful for reliability estimation and decision making for complex mechanical system.

REFERENCES

Du, X.P. 2007. Interval reliability analysis, *in ASME 2007 design engineering technical conference and computers and information in engineering conference (DETC2007)*, Las Vegas, Nevada, USA.

Fales, R. 2010. Uncertainty modeling and predicting the probability of stability and performance in the manufacture of dynamic systems. *ISA Transactions*, 49(1):528–534.

Flage, R. Aven, T. Baraldi, P. & Zio, E. 2011. On imprecision in relation to uncertainty importance measures *Proc European Safety & Reliability Conference.* Troyes, France.

Garg, H. & Rani, M. et al. 2013. Predicting uncertain behavior of press unit in a paper industry using artificial bee colony and fuzzy Lambda–Tau methodology. *Applied Soft Computing*, 13:1869–1881.

Guo, J. & Du, X.P. 2009. Reliability sensitivity analysis with random and interval variables. *International journal for numerical methods in engineering*, 78:1585–1617.

Harish, G. & Sharma, S.P. 2012. Stochastic behavior analysis of complex repairable industrial systems utilizing uncertain data. *ISA Transactions*, 51:752–762.

Hurtado, J.E. 2013. Assessment of reliability intervals under input distributions with uncertain parameters. *Prababilistic Engineering Mechanics*, 32:80–92.

Hurtado, J.E. & Alvarez, D.A. 2012. The encounter of interval and probabilistic approaches to structural reliability at design point. *Applied Mechanics and Engineering*, 225(1):74–94.

Jiang, C. & Li, W.X. et al. 2011. Structural reliability analysis based on random distributions with interval parameters. *Computer and Structures*, 89(1):2292–2302.

Jiang, C. & Long, X.Y. et al. 2013. Probability-interval hybrid reliability analysis for cracked structures existing epistemic uncertainty, *Engineering Fracture Mechanics*, 112:148–164.

Karaboga, D. & Basturk, B. 2007. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of Global optimization*, 39(3):459–471.

Karaboga, D. & Basturk, B. 2008. On the performance of artificial bee colony (ABC) algorithm. *Appl. Soft Comput*, 8:687–697.

Lu, Z.Z. & Feng, Y.W. et al. 2002. An advanced interval-truncation approach and non-probabilistic reliability analysis based on interval analysis. *Chinese Journal of Computational Mechanics*, 19(3):260–264.

Peng, W.S. & Zhang, J.G. et al. 2013. The mechanical reliability optimization based on the improved artificial bee colony algorithm. *Chemical Engineering Transactions*, 33:505–510.

Qiu, Z.P. & Yang, D. et al. 2008. Probabilistic interval reliability of structural systems. *International Journal of Solids and Structures*, 45(10):2850–2860.

Sharma, S.P. & Kumar, D. 2009. RAM analysis of the press unit in a paper plant using genetic algorithm and Lambda-Tau methodology. *Applications of Soft Computing*, Springer Berlin Heidelberg:127–137.

Singh, M & T, Markeset. 2008. Fuzzy reliability analysis of corroded oil and gas pipes. *Proc European Safety & Reliability Conference.* Valencia, Spain.

Wang, J. & Qiu, Z.P. 2010. The reliability analysis of probabilistic and interval hybrid structural system. *Applied Mathematical Modelling*, 34(3):3648–3658.

Zhao, M.H. & Jiang, C. et al. 2008. Non-probabilistic reliability analysis of retaining walls based on interval theory. *Chinese Journal of Geotechnical Engineering*, 30(4):467–472.

Zio, E. 2012. Dempster-Shafer theory of evidence to handle maintenance models tainted with imprecision, *Proc European Safety & Reliability Conference*. Helsinki, Finland.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Common cause failures and cascading failures in technical systems: Similarities, differences and barriers

L. Xie, M.A. Lundteigen & Y.L. Liu
NTNU, Trondheim, Norway

ABSTRACT: Many technical systems continue to increase in size and complexity, with more interactions and interdependencies between components. Dependent failures, such as common cause failures and cascading failures, are becoming important concerns to system reliability. Both failure types may lead to the unavailability of multiple components at the same time or within a short time interval. Although many researchers have studied common cause failures and cascading failures respectively, there is little comparison of the two concepts. This paper investigates the similarities and differences of these two failure groups, with focus on the conditions and nature of initiations and propagation of such failures. Moreover, a comparison is also made about suitable barrier strategies that can either prevent or reduce the consequences of failure. The paper concludes the study with a demonstration of reliability modeling for common cause- and cascading failures.

1 INTRODUCTION

Technical systems, such like railway systems, processing systems in chemical and petroleum plants, and power grids, are becoming increasingly complex. These systems include many physical components, with a huge number of interaction and interdependencies. Sometimes, those failures occurring in multiple components are resulted from the interconnections. We refer to such failures as dependent failures. Within the category of dependent failures, there are two sub-categories that are of specific interest: common cause failures (CCFs) and cascading failures (Rausand and Lundteigen, 2014). In the chemical and process industry, cascading processes are called as domino effects (Abdolhamidzadeh et al., 2010, Abdolhamidzadeh et al., 2009, Landucci et al., 2016).

Past accidents and near misses have shown that dependent failures are one of main threats to a complex system. For example, CCFs are main contributors of failures in safety systems of the oil and gas industry (Smith and Simpson, 2004, Lundteigen and Rausand, 2007). Fires in the chemical and process industry highlight the severe cascading consequences (Landucci et al., 2016, Cozzani and Reniers, 2013). The blackouts in United States, Canada in 2003, and Europe in 2006 are also the examples of cascading failures (Kotzanikolaou et al., 2013, Andersson et al., 2005). Many other infrastructure systems, like water distribution networks, transportation, also often suffer from cascading failures (Lin et al., 2014, Shuang et al., 2014, Ouyang, 2014).

So far, it seems like most attention has directed to CCFs and in specific for safety-critical systems where redundancy is used actively to enhance reliability (Paula et al., 1991, Humphreys and Jenkins, 1991, Lundteigen and Rausand, 2007, IEC61508, 2010, A. Mosleh, 1998). There have been two main strategies suggested for incorporating defenses against CCFs in design. One is to carry out analyses to identify and remove causes, and the other is to introduce measures to reduce the effects of CCFs in case they occur. Suggested methods include cause-defense matrices, common cause analysis, and zonal analysis (Humphreys and Jenkins, 1991, Paula et al., 1991).

The defenses to CCFs are typically identified in design, however, measures in the operational phase are also important (Lundteigen and Rausand, 2007). Even for an excellent system design, there will always remain a risk of CCFs. It is therefore required to include the contribution of CCFs in quantitative analyses used to demonstrate adequate reliability. A high number of models has been introduced for this purpose (Vesely, 1977, Fleming, 1975, Evans et al., 1984, Mosleh and Siu, 1987). The standard beta factor model is perhaps the most widely adopted, due to its simplicity (Fleming, 1975, IEC61508, 2010). The PDS method (Hauge et al., 2015) is an extension of the standard beta factor, where a second parameter is added to account for voting, e.g. 2-out-of-3 and 1-out-of-3.

As for cascading failures, it is of interest to consider efficient means to avoid or reduce the vulnerability of the failures in the system design, and to quantify cascading failures. An important

task in these analyses is to study interdependencies, and many analyzing approaches in literature are based the topology of complex network (Motter and Lai, 2002, Wang, 2012, Albert and Barabási, 2002). One kind of cascading failures are the failures when a heavily load component fails, and its load is redistributed to other components, resulting in loads on that exceed their capacities. State-based approaches, such as Markovian process, approaches based on the Bayesian network models, and Monte Carlo Simulation have been used to analyze cascading failures (Iyer et al., 2009, Calviño et al., 2016, Erp et al., 2017).

In fact, many technical systems can be subject to both CCFs and cascading failures, thus it is important to consider both failure categories in reliability analysis. Unfortunately, very limited attention has been directed comparing the two types of dependent failures, and their corresponding defense strategies. Kotzanikolaou et al. (2013) highlight that CCFs may have cascading effects, but do not go into much detail.

The objective of this paper is therefore to make a comprehensive comparison on the concepts, causes, and mechanisms of the two failures, and provide some suggestions on the analysis and defense strategies. In this paper, we use the term of barrier to denote a specific defense measure.

The rest of the paper is organized as follows: In section 2, we discuss the definitions and interpretations of CCFs and cascading failures. Sections 3 and 4 present the similarities and distinctions of the two failures. In section 5, we clarify the barriers against the two failures. A small example is then employed in section 6, to illustrate that the effects of CCFs and cascading failures. Conclusions and discussions occur in section 7.

2 DEFINITIONS AND INTERPRETATIONS

According to Humphreys and Jenkins (Humphreys and Jenkins, 1991), *dependent failures refer to the failures whose probability cannot be expressed by unconditional probability of the individual event*. Dependencies in a technical system may derive from the sameness of the types of components, exposure from the same environment, the use of shared resources, functionality, the common shocks and the incapability to resist certain hazardous events (Rausand, 2013).

People in different industrial sectors define CCFs in their own ways. Nuclear sector defines it as *two or more component fault states exist at the same time, or with a short interval, because of a shared cause* (Mosleh et al., 1988). The generic standard on design and operation of electric, electronic, and programmable electronic safety-related systems, IEC 61508, defines a CCF as a *failure that is the result of one or more events, causing concurrent failures of two or more separate channels in a multiple channel system, leading to system failure* (IEC61508, 2010). Both definitions emphasize that CCFs involve at least two failures that are due to a shared or common cause.

Cascading failure may be *multiple failures, where initiated by the failure of one component in the system that results in a chain reaction, the so-called domino effect* (Rausand and Øien, 1996). In power systems, cascading failure is referred to *a sequence of dependent failures of individual components that successively weakens the systems* (Baldick et al., 2008). It differs from the definition in infrastructures that limit the cascading failure to the propagation of failures between components (Rinaldi et al., 2001). Generally, we can find some same elements in the definitions that cascading failures are multiple failures initiated by one, and a sequential effect occurs.

From the perspective of failure causes, both CCFs and cascading failures result from some common vulnerabilities of more than one component. These two types of failures are interrelated in some cases (Laprie et al., 2007, Kotzanikolaou et al., 2013). However, they are still two distinctive categories of dependent failures. As Smith and Watson explained, CCFs emphasize that failures are located in 'first in line', which means that the failure are only dependent on the causes, but not on each.

In the following sections, we try to elaborate similarities and difference between the two failures.

3 SIMILARITIES

We categorize the similarities between CCFs and cascading failures into three: *multiplicity*, *timeliness* and *classification of causes*.

3.1 *Multiplicity*

Both CCFs and cascading failures obviously involve more than one components. We are concerned with the *effect of failure* of several components and *functions* for two categories of failures.

3.2 *Timeliness*

For both CCFs and cascading failures, the time from the first failure to the existence of multiple failures is often short. In case of insufficient mitigation measures, the collapse of an entire system may occur very soon. For example, in the Three Mile Island accident caused by CCFs in 1979, the radiation level in the primary coolant water

was around 300 times of the expected level after only 2 hours (Hasani, 2017). The power blackout in India in 2012 due to cascading failures, spread across 22 states within 12 hours and affected more than 620 million people (Russel, 2012).

3.3 *Root causes*

Root causes of both CCFs and cascading failures are the common vulnerability of more than one components in a system. Coupling factors between components can explain why multiple components are destroyed by a common hazardous event, e.g. cold temperature, extreme snowfall or electrical failure. Meanwhile, for cascading failures, couplings also can explain why multiple components are affected by the faults of relevant components. For example, the unavailability of one processing unit increases the workload of another unit.

4 DIFFERENCES

For differences between two types of failures, we categorize them into two: *initiation* and *propagation* of failures, as shown in Table 1. Initiation of failures.

As seen in Table 1, the initiating event of a CCF can be either replicated or occur simultaneously for several components. The effect of CCFs arises from shared causes, may be simultaneous failures or failures with some time apart. A cascading failure always starts with a single preceding component failure, as the effect of an initiating event.

Table 1. Differences between CCFs and cascading failures.

Difference	Characteristics	CCFs	Cascading failures
Initiation	Triggering condition	Shared causes	Conditional on preceding failures
	Occurrence	Simultaneously or during a critical time of interest	Sequence
Propagation	Sequence	First in line	Series
	Consequence	Finite	Possibly infinite
	Pathway	Cause-components	Connected/ dependent components

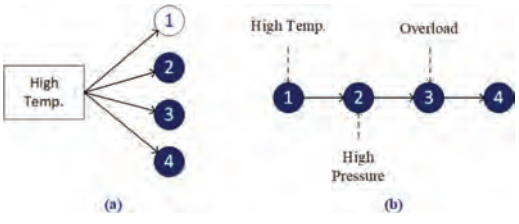


Figure 1. CCF and cascading failures.

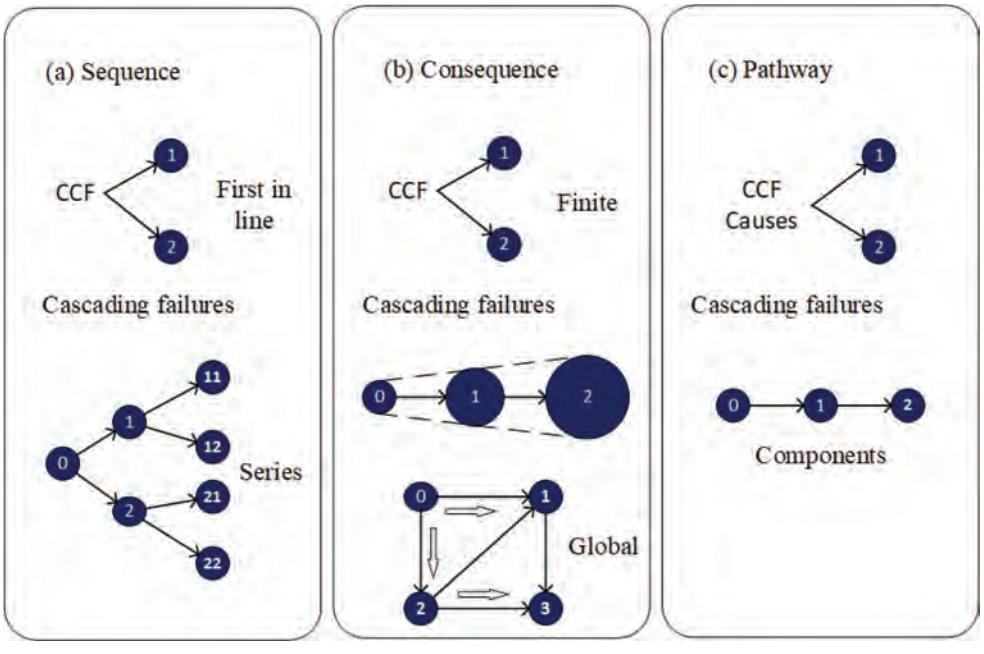


Figure 2. Comparisons of CCFs and cascading failures in terms of impact and effect.

To illustrate these differences, we introduce two small examples, as shown in Figure 1. High temperature is the initiating event of both a CCF and a cascading failure in this case. In Figure 1(a), all the four components expose themselves to high temperature, and so all or some of the components fail simultaneously or in a short interval. However, in the case of a cascading failure of Figure 1(b), only component 1 is exposed to high temperature, and fails due to this initiating event. Then, the failure of component 1 trigger the failures of other components due to diverse reasons. Even in the same cascading sequence, the failure causes can be different for the different components.

4.1 *Propagation of failures*

Propagation of failure means in this context the evolvement of multiple failures, with the initiating event already manifested. Figure 2 illustrates the differences in the propagation of CCFs and cascading failures. CCFs are *first in line* failures that delineate the exclusion of dependent failures from CCF definition (Smith and Watson, 1980), which implies that CCFs are directly linked to the failure causes. On the contrary, the propagation of a cascading failure follows a series of interactions. CCFs are most different from cascading failures in terms of the approaches of propagation. As shown in Figure 2(a), for CCFs, the first in line failure only occurs on component 1 and 2. For the consequence of failure propagation, as shown in Figure 2(b), a cascading failure can escalate and result in worse impacts on the other parts of a system, such as more serious disruptions, overload to neighbors and longer recovery time etc. CCFs highlight a direct cause-effect relationship between the cause and the failed components (Rausand and Lundteigen, 2014), whereas the pathway of cascading failures involve the interactions or dependencies between relevant components, see in Figure 2(c).

5 BARRIERS

Barriers are employed to prevent, control or mitigate undesired events or accident (Sklet, 2006). Sometimes, barriers are also called defenses, protection layers or countermeasures. In general, a barrier function can be realized by many different means, such as by a technical or physical system, human actions and procedural deficiencies.

In the design phase of a system, it is possible to introduce barriers against potential failures, like separation, diversity, quality control, simplicity of design etc. Some of them are effective to reduce the probability of CCFs, and some of them are more functional for protecting the system from cascading failures. Considering the similarities and differences of CCFs and cascading failures, we can categorize barriers into three groups: *barriers against both failures*, *barriers against CCFs* and *barriers against cascading failures*.

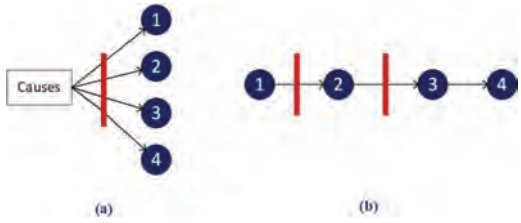


Figure 3. Barriers for CCF and cascading failures.

- Barriers efficient for both failures: Such kind of barriers should be designed in consideration of the similarities of CCFs and cascading failures, such like their root causes and coupling factors. One way of barrier design is therefore to mitigate and reduce the vulnerability to root causes. Simplicity can be regarded as a barrier, for example, to reduce system complexity that is one important source of vulnerability. Another way of barrier design is to decrease the coupling degrees among components. Spatial and temporal separations are examples of decreasing coupling degrees. In practices, we can find that firewalls in a process plant are effective barriers to prevent fire disasters.
- Barriers against CCFs: The effectiveness of such barriers is to isolate failure causes and components, as shown in Figure 3(a). One example is diversity of the design. Diverse components will often have different failure modes, and are therefore less likely to be affected by the common cause. However, diversity is not effective to mitigate cascading failures. When the failure of one component brings higher workload to its neighbors and their failure probabilities, no matter the components are identical or not.
- Barriers against cascading failures: The main purposes of this kind of barriers are to stop or slow down failure propagation, as shown in Figure 3(b). An example for this class of barriers is a process shutdown valve that can isolate related process segments. In case abnormal events have occurred in the upstream facility, the shutdown valve can stop or limit the flow between two facilities, and thereby cease the failure propagation.

In the next section, we will use a small example to illustrate the quantitative analyses for CCFs and cascading failures, and the effects of barriers.

6 CASE STUDY

Suppose a system comprising two parallel components. The effects of failures and corresponding barriers for the two dependent failures are studied separately, as illustrated in Figure 4.

For modeling CCFs, a new *independent* "CCF" event is added in the standard beta model with beta-factor β. The parameter β can be interpreted as the conditional probability that a failure of a channel is in fact a common-cause failure:

$$\beta = \Pr(CCF \,|\, Failure\ of\ channels) \tag{1}$$

With inclusion of CCFs, the total system reliability can be obtain as:

$$R(t) = 2R - R^{(2-\beta)} \tag{2}$$

where $R = 0.8$ and $\beta = 0.1$.

For modeling cascading failures, it is necessary to consider the effects of functional dependency between the two components, and Bayesian network model is an approach we used here. The conditional failure probability is a measure of dependency that differ from the conditional probability β for CCFs. The conditional probability for cascading failures can be defined as:

$$\Pr(Comp.B\ fails \,|\, comp.A\ fails) = \frac{F_D}{F_A} \tag{3}$$

Here, F_A and F_B denote the individual failure probability for component A and B. F_D denotes the failure probability for component A on the condition of component A has failed. The total system reliability with cascading failures can be obtained as:

$$\begin{aligned} R(t) &= 1 - F_A(F_B + F_D - F_B F_D) \\ &= 1 - (1-R)^2 - (1-R)^2 R P_r \end{aligned} \tag{4}$$

where P_r denotes conditional probability between component A and B and is assigned as $0.1 (\Pr = 0.1)$.

As shown in Figure 5, the total system reliability with CCFs becomes 0.946, but it is 0.957 with the effects of cascading failures at that time. This

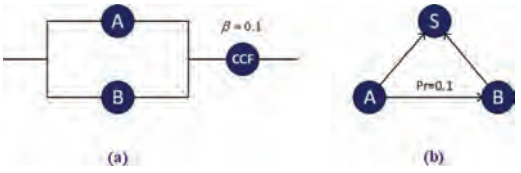


Figure 4. Case study for CCF and cascading failures.

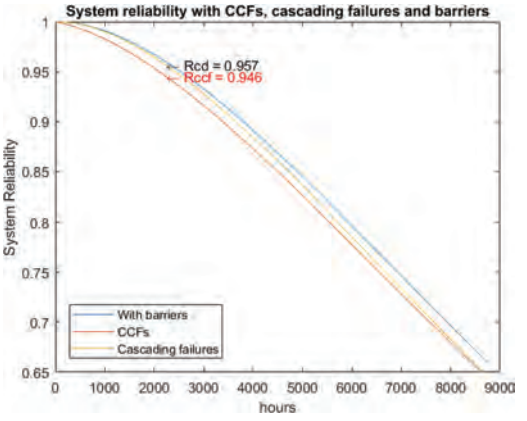


Figure 5. Reliability with cascading failures & CCFs.

implies that CCFs may have more influence on the reliability performance than cascading failures in this case, when using similar assumptions about the probability of having additional failures, when a first failure has occurred.

We now introduce time-dependent probabilities for reliability analysis, and assume that the time to failure is exponentially distributed, with failure rate of 1E-04 per hour for each component. For the system with CCFs, the total system reliability can be obtain as:

$$R(t) = \left[2e^{-(1-\beta)\lambda t} - e^{-2(1-\beta)\lambda t}\right] e^{-\beta\lambda t} \tag{5}$$

For the system with cascading failures, the total system reliability can be obtain as:

$$R(t) = 1 - (1 - e^{-\lambda t})^2 - (1 - e^{-\lambda t})^2 e^{-\lambda t} P_r \tag{6}$$

Figure 5 illustrates calculated system reliability considering the effects of the two failures as a function of time. We can see that, in this case, the two failures seems to have comparable effects on the system reliability.

For CCFs, the function of barriers is to separate shared root causes from the components. The function of the barriers against cascading failures is to prevent propagation of the failures between component A and B. Reliability of the system with barriers is illustrated in the blue line in Figure 5, implying that the system reliability will increase when performing barriers function against the failures.

7 CONCLUSION AND FURTHER WORK

Exploring similarities and difference between CCFs and cascading failures facilitate us to answer

the following questions: 1) why such dependent failures initiate, 2) how dependent failures contribute to disruptions in the systems, and 3) what kind of barriers are needed and how they should be implemented. In this paper, we find that CCFs and cascading failures may have comparable influences on the performance of a simple system. More probabilistic and quantitative analyses are required, to evaluate the impacts of cascading failures in a larger and more complex system (Erp et al., 2017).

Our further work will involve modeling the interdependent systems with cascading failures and CCFs, and developing tools to evaluate reliability for complex systems. It is also of interest to identify different failure modes and perform barrier analysis for both of the failures, which can help to allocate barriers and thereby optimize barrier functions.

REFERENCES

Abdolhamidzadeh, B., Abbasi, T., Rashtchian, D. & Abbasi, S.A. (2010) A new method for assessing domino effect in chemical process industry. *Journal of hazardous materials,* 182, 416–426.

Abdolhamidzadeh, B., Rashtchian, D. & Ashuri, E. (2009) A new methodology for frequency estimation of second or higher level domino accidents in chemical and petrochemical plants using monte carlo simulation. *Iranian Journal of Chemistry and Chemical Engineering (IJCCE),* 28, 21–28.

Albert, R. & Barabási, A.-L. (2002) Statistical mechanics of complex networks. *Reviews of modern physics,* 74, 47.

Andersson, G., Donalek, P., Farmer, R., Hatziargyriou, N., Kamwa, I., Kundur, P., Martins, N., Paserba, J., Pourbeik, P. & Sanchez-Gasca, J. (2005) Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance. *IEEE Transactions on Power Systems,* 20, 1922–1928.

Baldick, R., Chowdhury, B., Dobson, I., Dong, Z., Gou, B., Hawkins, D., Huang, H., Joung, M., Kirschen, D. & Li, F. (2008) Initial review of methods for cascading failure analysis in electric power transmission systems IEEE PES CAMS task force on understanding, prediction, mitigation and restoration of cascading failures. *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE.* IEEE.

Calviño, A., Grande, Z., Sánchez-Cambronero, S., Gallego, I., Rivas, A. & Menéndez, J.M. (2016) A Markovian–Bayesian network for risk analysis of high speed and conventional railway lines integrating human errors. *Computer-Aided Civil and Infrastructure Engineering*, 31, 193–218.

Cozzani, V. & Reniers, G. (2013) Historical background and state of the art on domino effect assessment. *Domino Effects in the Process Industries: Modelling, Prevention and Managing.* Elsevier, Amsterdam, The Netherlands.

Erp, N.V., Linger, R., Khakzad, N. & Gelder, P.V. (2017) Report on risk analysis framework for collateral impacts of cascading effects. *RAIN—Risk Analysis of Infrastructure Networks in Response to Extreme Weather.* TU Delft.

Evans, M., Parry, G. & Wreathall, J. (1984) On the treatment of common-cause failures in system analysis. *Reliability engineering,* 9, 107–115.

Fleming, K. (1975) Reliability model for common mode failures in redundant safety systems. *Modeling and simulation. Volume 6, Part 1.*

Hasani, F. (2017) Calculation and Analysis of Reliability with Consideration of Common Cause Failures (CCF)(Case Study: The Input of the Dynamic Positioning System of a Submarine). *International Journal of Industrial Engineering & Production Research,* 28, 175–187.

Hauge, S., Hoem, A., Hokstad, P., Habrekke, S. & Lundteigen, M.A. (2015) Common Cause Failures in Safety Instrumented Systems. SINTEF Technology and Society Trondheim.

Humphreys, P. & Jenkins, A.M. (1991) Dependent failures developments. *Reliability Engineering & System Safety,* 34, 417–427.

Iec61508 (2010) Functional safety of electrical/electronic/programmable electronic safety related systems. *International Electrotechnical Commission*.

Iyer, S.M., Nakayama, M.K. & Gerbessiotis, A.V. (2009) A Markovian dependability model with cascading failures. *IEEE Transactions on Computers,* 58, 1238–1249.

Kotzanikolaou, P., Theoharidou, M. & Gritzalis, D. (2013) Cascading effects of common-cause failures in critical infrastructures. *International Conference on Critical Infrastructure Protection.* Springer.

Landucci, G., Argenti, F., Spadoni, G. & Cozzani, V. (2016) Domino effect frequency assessment: The role of safety barriers. *Journal of Loss Prevention in the Process Industries,* 44, 706–717.

Laprie, J.-C., Kanoun, K. & Kaâniche, M. (2007) Modelling interdependencies between the electricity and information infrastructures. *Computer Safety, Reliability, and Security*, 54–67.

Lin, Y., Li, D., Liu, C. & Kang, R. (2014) Framework design for reliability engineering of complex systems. *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2014 IEEE 4th Annual International Conference on.* IEEE.

Lundteigen, M.A. & Rausand, M. (2007) Common cause failures in safety instrumented systems on oil and gas installations: Implementing defense measures through function testing. *Journal of Loss Prevention in the process industries,* 20, 218–229.

Mosleh, A., D.M. Rasmuson & F.M. Marshall (1998) Guidelines on modeling common cause failures in probablistic risk assessment.

Mosleh, A., Fleming, K., Parry, G., Paula, H., Worledge, D. & Rasmuson, D.M. (1988) Procedures for treating common cause failures in safety and reliability studies: Volume 1, Procedural framework and examples. Pickard, Lowe and Garrick, Inc., Newport Beach, CA (USA).

Mosleh, A. & Siu, N. (1987) A multi-parameter common cause failure model. *Transactions of the 9th international conference on structural mechanics in reactor technology. Vol. M*.

Motter, A.E. & Lai, Y.-C. (2002) Cascade-based attacks on complex networks. *Physical Review E,* 66, 065102.

Ouyang, M. (2014) Review on modeling and simulation of interdependent critical infrastructure systems. *Reliability engineering & System safety,* 121, 43–60.

Paula, H.M., Campbell, D.J. & Rasmuson, D.M. (1991) Qualitative cause-defense matrices: Engineering tools to support the analysis and prevention of common cause failures. *Reliability Engineering & System Safety,* 34, 389–415.

Rausand, M. (2013) *Risk assessment: theory, methods, and applications*, John Wiley & Sons.

Rausand, M. & Lundteigen, M.A. (2014) *Reliability of safety-critical systems: theory and applications*, John Wiley & Sons.

Rausand, M. & Øien, K. (1996) The basic concepts of failure analysis. *Reliability Engineering & System Safety,* 53, 73–83.

Rinaldi, S.M., Peerenboom, J.P. & Kelly, T.K. (2001) Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems,* 21, 11–25.

Russel, H.S. a. R. (2012) 620 million withow power in india after 3 pewer grids fail.

Shuang, Q., Zhang, M. & Yuan, Y. (2014) Node vulnerability of water distribution networks under cascading failures. *Reliability Engineering & System Safety,* 124, 132–141.

Sklet, S. (2006) Safety barriers: Definition, classification, and performance. *Journal of loss prevention in the process industries,* 19, 494–506.

Smith, A.M. & Watson, I.A. (1980) Common cause failures—a dilemma in perspective. *Reliability Engineering,* 1, 127–142.

Smith, D.J. & Simpson, K.G. (2004) *Functional Safety: A straightforward guide to applying IEC 61508 and related standards*, Routledge.

Vesely, W. (1977) Estimating common cause failure probabilities in reliability and risk analysis: Marshall-Olkin specializations. *Nuclear systems reliability engineering and risk assessment,* 2.

Wang, J. (2012) Mitigation of cascading failures on complex networks. *Nonlinear Dynamics,* 70, 1959–1967.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Industry 4.0 and complexity: Markov and Petri net based calculation of PFH for designated architectures and beyond

M. Albert
SICK AG, Waldkirch, Germany

M. Dorra
Institute for Occupational Safety and Health of the German Social Accident Insurance (IFA), Sankt Augustin, Germany

ABSTRACT: Industry 4.0 opens up for a new area of industrial automation and challenges established methods for the assurance of safety. Traditionally, the architecture is a major criterion for the determination of safety levels and so-called designated architectures serve as a basis for the calculation of the *PFH* (EN ISO 13849-1, 2015, IEC 62061, 2015). This easy approach is, however, limited to systems which can be directly mapped to the assumptions. For modern, software-intensive systems the extended possibilities for online diagnostics makes this mapping particularly difficult. To make the *PFH* calculation more flexible and transparent, we have derived Markov-based analytic equations for the designated architectures. In this paper, we will sketch our approach for a single-channel system with diagnostics. We compare the result to Petri net simulations, analyze the influence of individual parameters and argue how these models can be extended to more complex systems even beyond the realm of the designated architectures.

1 INTRODUCTION

In machinery the demand for safety and flexibility at the same time has led to the development of a large variety of safety functions. According to the risk associated with particular hazards in the application, these functions are intended to assure an appropriate risk reduction. Accordingly, the integrity of the safety function is typically specified as a Safety Level, e.g. as the "Safety Integrity Level" (SIL) of IEC 61508 (IEC 61508-1, 2010), the SIL claim of IEC 62061 (IEC 62061, 2015), or the Performance level (PL) of ISO 13849-1 (EN ISO 13849-1, 2015).

These standards directly link the achievable safety level to a target probability of dangerous failures. For safety in machinery (usually operated in high demand mode), this is the average frequency of dangerous failures of the safety function *PFH* (IEC 61508-4, 2010, Innal et al. 2010).

Most implementations of safety functions involve electrics and/or electronics but the complete functional chain often also employs pneumatics, hydraulics or mechanics. Irrespective of the technologies, the safety-related reliability in terms of the *PFH* value can be kept sufficiently low by using diagnostics and redundancy. The standards IEC 61508, IEC 62061 and ISO 13849 (IEC 61508-1, 2010, IEC 62061, 2015, EN ISO 13849-1, 2015) governing the functional safety of machinery address a couple of simple and generic system architectures but pursue different ways in evaluating the *PFH* value: IEC 61508 and IEC 62061 by provision of equations for calculation of the *PFH*, ISO 13849-1 by tables. However, the underlying approaches of these standards as well as the assumptions made result in several shortcomings concerning the applicability of these simple approaches to real systems. On the other hand, in all of the standards, stochastic modeling is suggested as a suitable method to determine the *PFH* value of systems which cannot directly be mapped to their dedicated architectures.

Among these methods, Markov models have proven to be a suitable tool for the calculation of safety-related reliability measures (Brissaud & Oliveira 2012, Dutuit et al. 2008).

Unlike numerical methods (stochastic Petri nets, Monte Carlo simulation), they enable the derivation of analytic equations (Dutuit et al. 2008). Their major drawbacks are the limitation to exponentially distributed processes (constant transition rates) and the (exponential) explosion of the state space with the number of elements. These drawbacks make the application to more complex systems with non-exponentially distributed processes and many states difficult (IEC 61508-6, 2010).

In the first part of this paper, we will argue, how the limitation to exponentially distributed processes can be overcome for a single channel system

with diagnostics (1oo1D) without significantly reducing the precision of the results. Subsequently, a very general analytic formula for the calculation of the *PFH* for the 1oo1D systems will be derived.

The second part will show, how the analytic results can be reproduced by Petri net based Monte-Carlo simulations and will argue how the corresponding Petri net can be extended to more complex systems.

2 SINGLE-CHANNEL SYSTEM WITH DIAGNOSTICS

Figure 1 shows the block diagram of the single channel system with diagnostics (1oo1D) considered in this paper.

The functional channel F with the dangerous failure rate λ_{FD} performs the safety function. The diagnostic channel M (monitor) with the dangerous failure rate λ_{MD} tests F with the diagnostic coverage *DC* and the diagnostic rate r_t. With respect to F "dangerous failure" means loss of the safety function whereas with respect to M "dangerous failure" means loss of the diagnostic function which M should execute for F. β is the "common cause factor" as defined by IEC 61508-6, 2010, Annex D. It represents a measure for the susceptibility of F and M to dangerous failures due to the same causes. If a dangerous failure of F is detected, M initiates a safe state. The diagnostic channel M is not tested by any diagnostics. Failures of the diagnostic channel will hence not be detected. The system is subject to regular demands of the safety function that occur at random points in time. r_d represents the mean demand rate. At least one demand occurs per year (high demand mode of operation, see IEC 61508-4, 2010).

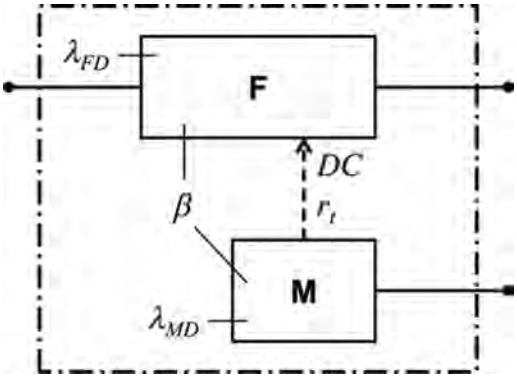


Figure 1. Block diagram of a single channel system with diagnostic channel.

2.1 *State transition model of the single channel system*

Modelling the above-noted features of the 1oo1D systems leads to the state transition model depicted in Figure 2. All system states shaded in light grey are dangerous states because the safety function cannot be executed due to a failure of channel F. A demand on the safety function will lead to a hazardous event if the system is in one of these states.

Assuming time-constant failure rates, the transitions representing those failures are exponentially distributed functions (solid arrows). Testing of F and demand of the safety function are supposed to be uniformly distributed (dashed arrows) which corresponds to a constant frequency of occurrence. These processes are either driven by automatic periodic procedures (especially the tests) or their frequency is estimated (especially the demands). In the latter case the estimation typically comprises an average frequency only and there is no basis for the assumption of any other distribution than the uniform distribution as an adequate representation of regularity. On the contrary the repair process complies with a jump distribution (dotted arrows, labeled with the repair rate r_r). Failures of the functional channel F and of the monitor M due to the same cause ("common cause failures") are associated with the common cause failure rate λ_{CC}. A common cause failure can only occur in the OK state from where it leads directly to the state "F DU, M D".

Since not all of the transitions have exponential distribution functions, this graph does not represent a Markov model.

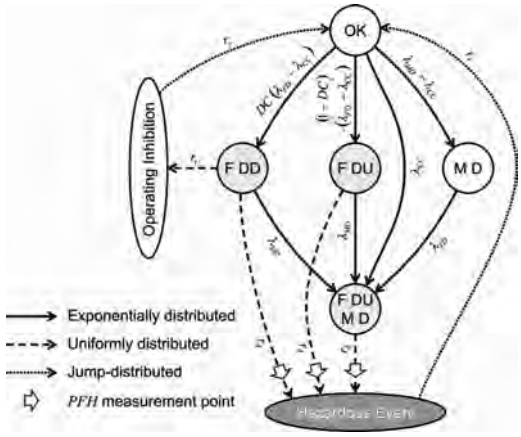


Figure 2. State transition model for the single channel system with diagnostics; "F DD": F failed dangerously detectable, "F DU": F failed dangerously undetectable, "M D": M failed dangerously (unable to detect faults of F).

Applying a demand of the safety function while the system is in one of the dangerous states will result in a "hazardous event". Therefore, according to Figure 2, the *PFH* is made up of the absolute flow (events per time unit) associated with the demand-driven transitions. This is where the *PFH* measurement points are located.

2.2 *Principles for model simplification*

In order to transform the model into a simpler and Markov-compatible form, several simplifications can be carried out.

The "F DU" state can be merged with the "F DU, M D" state without any influence on *PFH*, because the sum of their probabilities is not affected by the transition between them. It is, however, the probability sum only which determines the *PFH* contribution of these states or, respectively, of the merged state, since there is no other exit than via the demand process. Additionally, the ratio of the transition rates allows for some significant model simplifications without considerable loss of *PFH* accuracy. For instance, by neglecting the Mean Repair Time (*MRT*) the "operating inhibition" and "hazardous event" states can be merged with the "OK" state.

On the other hand, the most interesting state of the model is "F DD". It is fed by the very low rate *DC* ($\lambda_{FD} - \lambda_{CC}$) (detectable failures of F without common cause failures) but cleared by the diagnosis or the demand process on a regular basis with a much higher rate. As a consequence, "F DD" is nearly empty at any time and the probability of the state can be neglected. (It will though be used to calculate the outflow rates of this state, see below.) Moreover, the "F DD" state is additionally discharged by failures of the monitor M (rate λ_{MD}). The "F DD" state can be eliminated by making use of its very low probability and by determining the outflow rates as a function of the inflow rate.

This situation is sketched in Figure 3, where the related state is labelled "FPN", as it serves as "flow partitioning node".

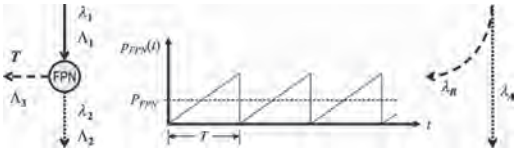


Figure 3. Flow partitioning node; λ_1 and λ_2 are nominal rates of exponentially distributed failure processes, T is the mean time between two clearings of FPN, Λ_1, Λ_2 and Λ_3 denote the corresponding absolute mean flow rates, λ_A and λ_B are the resulting mean surrogate outflow rates of the FPN node.

At "FPN" phases of duration T with feeding by the absolute rate Λ_1 and a small discharge by the rate λ_2 are interrupted by a complete clearing. "Absolute rate" means probability flow per time unit. In case of an exponentially distributed process, the absolute rate Λ is given by the probability of the source state, multiplied by the nominal rate λ of that process. During the interval [0, T] of the time t, for the probability of the state it holds true that

$$p_{FPN}(t+\Delta t) = p_{FPN}(t) + \Lambda_1 \cdot \Delta t - p_{FPN}(t) \cdot \lambda_2 \cdot \Delta t \tag{1}$$

For $\Delta t \to 0$ this yields:

$$\dot{p}_{FPN}(t) = \Lambda_1 - \lambda_2 \cdot p_{FPN}(t) . \tag{2}$$

With $0 \le t \le T$ and the initial condition $p_{FPN}(0) = 0$, the solution of this differential equation is:

$$p_{FPN}(t) = \frac{\Lambda_1}{\lambda_2}\left(1 - e^{-\lambda_2 t}\right) \tag{3}$$

The mean probability is calculated by integration over a time interval of duration T and division by T:

$$P_{FPN} = \frac{1}{T}\int_0^T p_{FPN}(t)\,dt = \frac{\Lambda_1}{\lambda_2} + \frac{\Lambda_1}{\lambda_2^2 \cdot T}\left(e^{-\lambda_2 T} - 1\right) \tag{4}$$

By replacing the absolute inflow rate Λ_1 by the nominal rate λ_1, Equation 4 allows to calculate the partitioning of the nominal inflow rate λ_1 of "FPN" to the two fractions λ_A and λ_B:

$$\lambda_A = \lambda_1\left(1 - \frac{1 - e^{-\lambda_2 T}}{\lambda_2 T}\right) \tag{5}$$

$$\lambda_B = \frac{\lambda_1}{\lambda_2 T}\left(1 - e^{-\lambda_2 T}\right) \tag{6}$$

Note that the splitting of λ_1 to λ_A and λ_B according to Equations 5 and 6 considers the exponential distribution function associated with λ_2 and the uniform distribution function associated with the clearing process with rate $1/T$.

The flow partitioning node "F DD" of Figure 2 is cleared by testing as well as by demand on the safety function. Hence, the following substitution in Equations 5 and 6 is necessary

$$T = \frac{1}{r_t + r_d} \tag{7}$$

According to Figure 2, λ_1 and λ_2 are given by $DC\,(\lambda_{FD} - \lambda_{CC})$ or λ_{MD}, respectively.

2.3 *Resulting model and analytic result*

Applying the above-mentioned simplifications to the model of Figure 2 yields the model shown in Figure 4.

The simple two-state model of Figure 4 is finally a Markov model since all occurring transitions are exponentially distributed. The related differential equation system can be solved for the initial conditions $p_{OK}(0) = 1$ and $p_{MD}(0) = 0$ whereupon, according to Figure 4, the instantaneous *PFH* value can be calculated by

$$\begin{aligned} pfh(t) \;=\; & \lambda_B \,\frac{r_d}{r_t + r_d}\, p_{OK}(t) \\ & + \left[(1 - DC)\,(\lambda_{FD} - \lambda_{CC}) + \lambda_A + \lambda_{CC} \right] p_{OK}(t) \\ & + \lambda_{FD}\; p_{MD}(t) \end{aligned} \tag{8}$$

Calculating the mean value of $pfh(t)$ over the mission time T_M eventually yields

$$\begin{aligned} PFH \;=\; & \lambda_{FD} \\ & - \left\{ DC \cdot TRTE \cdot \frac{(\lambda_{FD} - \lambda_{CC})}{(\lambda_{FD} + \lambda_{MD} - \lambda_{CC})^2\, T_M} \right. \\ & \cdot \left[\lambda_{FD}\,(\lambda_{FD} + \lambda_{MD} - \lambda_{CC})\, T_M \right. \\ & \left. \left. + (\lambda_{MD} - \lambda_{CC})\left(1 - e^{-(\lambda_{FD} + \lambda_{MD} - \lambda_{CC})\, T_M}\right) \right] \right\} \end{aligned} \tag{9}$$

where the common cause failure rate of the channels F and M is given by

$$\lambda_{CC} \;=\; \beta \cdot \min(\lambda_{FD}, \lambda_{MD}) \tag{10}$$

And *TRTE* is the time-related test efficiency. Making use of $\lambda_{MD} \ll r_t + r_d$ this quantity can be calculated by

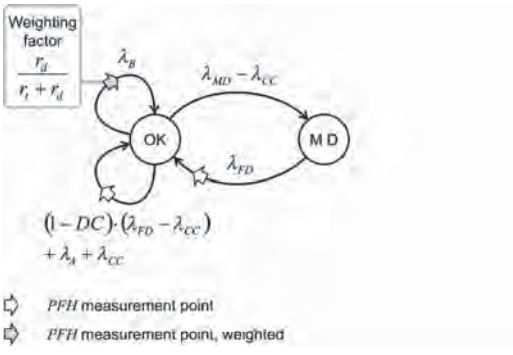


Figure 4. Graph of the Markov model for the single channel system with diagnostics.

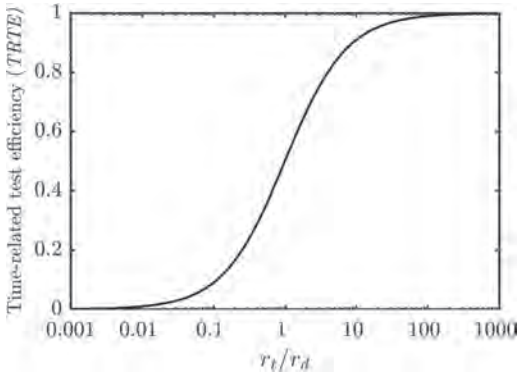


Figure 5. Time-related test efficiency as a function of rt/rd.

$$TRTE \;=\; \frac{r_t}{r_t + r_d} \tag{11}$$

TRTE is a measure for the test effectiveness under consideration of a temporal "race" of testing and demand on the safety function. System design should aspire to avoid such a race so as not to diminish the effect of testing. This is accomplished if the test is always executed in time and in this case *TRTE* can be set to its maximum value of 1. $TRTE = 0$ as well as $DC = 0$ lead to an untested channel F and $PFH = \lambda_{FD}$.

If time-optimal testing is not feasible, *TRTE* should be made as high as possible by implementing an adequate test rate. The model of Figure 2 assumes the absence of any synchronization of test and demand. Provided this, Equation 11 states the dependence of the time-related test efficiency *TRTE* on the ratio of the test rate r_t and the demand rate r_d. This dependence is depicted in Figure 5.

IEC 61508, IEC 62061 and ISO 13849 recommend a ratio r_t/r_d of ≥100 (IEC 61508-2, 2010, IEC 62061, 2015, EN ISO 13849-1, 2015). According to Figure 5 or Equation 11, this ensures a time-related test efficiency of >0.99.

2.4 *Other architectures*

The designated architectures of machine controls addressed by the standards for functional safety also comprise the two-channel architecture. Therefore, by applying the principles described in section 2.2, an analytic Markov model-based solution for the *PFH* was also derived for the asymmetric redundant 1oo2D structure. Due to limited space it is not reported here. A manuscript regarding this matter has been proposed for usage within standardization (Dorra, M., 2017).

3 PETRI NET BASED MONTE CARLO SIMULATION

Finite state machines are an efficient way to model the behavior and the dynamics of safety related systems (EN 61508-6, 2010). Since the 1960s Petri nets have been used for this purpose and since then have proven to be useful and versatile tool. While in the beginning (timed) Petri nets where mostly used to synthesize Markov graphs, Monte Carlo simulation of such nets became increasingly popular to investigate their dynamic behavior and to overcome the problem of the exploding state space of the Markovian approach (EN 61508-6, 2010). In case of Petri nets, the size of the model scales linear with the number of modeled elements. Beside their scaling properties, their easy graphical representation and the possibility to simulate the system graphically makes Petri nets a powerful tool for the modelling of safety-related systems. Results for the calculation of the average probability of dangerous failures on demand in case of the low demand mode (<1 demand/year) obtained with Markov models and Petri nets have proven to be similar (Brissaud & Oliveira 2012).

3.1 *Model*

The basic Petri net model of the 1oo1D architecture of Figure 1 is depicted in Figure 6. The model consists of:

- Two places "F_{OK}" and "F_D" representing the working and the failed state of the functional channel, respectively.
- Two places "MOK" and "MD" representing the working and the failed state of the monitor channel.
- The transitions "Failure F" and "Failure M" representing the failures of the functional and the monitor channel, respectively. Whenever the "Failure F" transition is fired, an assertion is used to generate a random number to classify detectable and undetectable failures (See description of the transition "Failure Detected").
- A transition "Failure CC" representing the common-cause failures of function and monitor channel.
- The transition "Demand" representing the demand of the safety functions (after a failure of the function channel). Any firing of "Demand" will reset the Petri net, assuming that the system will be brought to an "as-good-as-new-state" if a dangerous failure has occurred prior to a demand of the safety function. If the monitor channel is also in the failed state, it will be reset by the transition "Repair M on demand".
- The cyclic transition "Failure Detected" representing the periodic testing of the functional channel by the monitor. Predicates are used to distinguish detectable and undetectable failures by comparing the random number generated by the "Failure F" transition to the DC.
- Additional assertion and predicates are used, to assure the correctness of the model, while keeping the graphical representation as simple as possible.

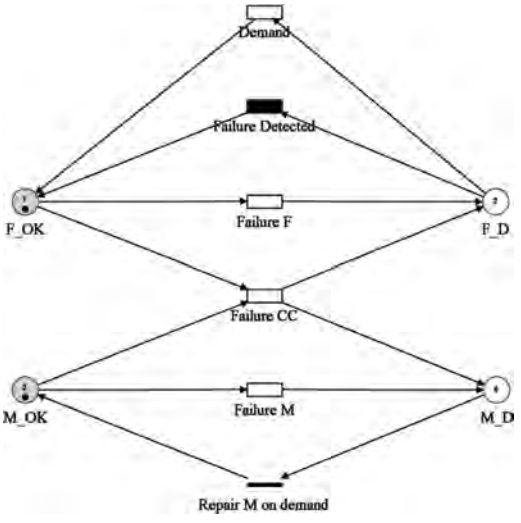


Figure 6. Petri net for single Channel system with diagnostics. For further details see text.

The Petri net model is investigated using the Monte-Carlo simulation engine of the GRIF software-framework (Satodev, 2017).

As argued earlier in this paper, any demand on the safety function from a failure state of "F" will contribute to the *PFH*. In case of the Petri net, the *PFH* of the system can hence be determined by the average frequency at which the "Demand" transition fires (Innal et al. 2010).

As the monitor channel itself is not monitored, a failure of this channel will not be detected and a subsequent failure of the functional channel will likewise remain undetected.

3.2 *Comparison*

Figure 7 shows the results of the Monte-Carlo simulations of the Petri net and compares them to the solution of the analytic Markov Modell (eq. 9) for the same sets of parameters. The results of the two methods are very similar and reproduce the table values of ISO 13849-1, Annex K, both for $DC = 90\%$ and for $DC = 60\%$ (EN ISO 13849-1, 2015). For the calculation, the following assumptions where made: Demand rate: $r_d = 360\ h^{-1}$, rate of diagnosis: $r_t = 100 \cdot r_d$, same $MTTF_D$, for

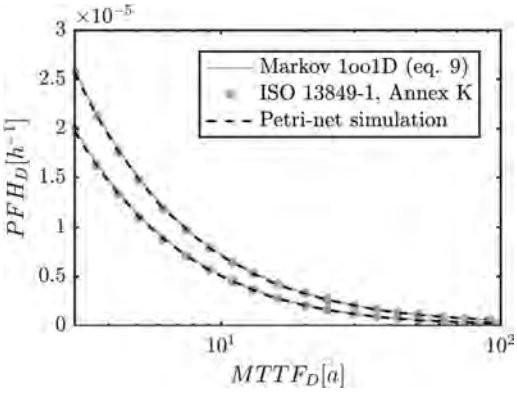


Figure 7. Comparison of *PFH* values calculated with the analytic Markov-Formula (see Equation 9) to the values of the table in ISO 13849-1, Annex K (EN ISO 13849-1, 2015), and the Petri net based Monte Carlo simulation. The two sets of curves correspond to $DC = 90\%$ (upper) and $DC = 60\%$ (lower). The results of the Markov and the Petri net approach are very similar and reproduce reasonably well the values of ISO 13849-1, Annex K.

function channel and monitor channel: $MTTF_{MD} = MTTF_{FD}$, common cause failure rate: $\lambda_{CC} = \beta \cdot \min(1/MTTF_{FD}, 1/MTTF_{MD})$, with $\beta = 2\%$, mission time: $T_M = 20$ a. These assumptions are in accordance with the prerequisites and assumptions of ISO 13849-1 and allow for a direct comparison of the results to this standard (EN ISO 13849-1, 2015).

3.3 *Failure sequences*

The properties of the 1oo1D System were further studied by a systematic investigation of the relevant parameters of the model. The Petri net model allows for the investigation of the relevance of the failure sequences to the overall frequency of dangerous failures.

In the simulation, the order of the fringing of the transitions was tracked by assertions. The clusters of firing sequences listed in the following list are investigated, where the order of the firing of transitions is indicated by arrows. Detectable and Undetectable failures are distinguished by assertions and predicates on the transition, as described above. To simplify reading, this implicit distinction is indicated by adding detectable/undetectable to the Failure transition in brackets:

1. F_{DD}-sequences:
   - Failure F(detectable)→Demand
   - Failure F(detectable)→Failure M→Demand
2. F_{DU}-sequences:
   - Failure F(undetectable)→Demand
   - Failure F(undetectable)→Failure M→Demand
3. M_D-sequences:
   - Failure M→Failure F(detectable)→Demand
   - Failure M→Failure F(undetectable)→Demand
4. CCF-sequence:
   - CCF→Demand

The demand is explicitly modeled in order to investigate the residual contribution of detectable failures to the overall failure rate due to the finite frequency of the diagnosis by the monitor channel.

Figure 8 shows the relative contribution of the major failure sequences to the *PFH* as a function of the $MTTF_D$ values for different *DC* values. The following assumptions were made for the system: The mean time to failure of the monitor channel was assumed to be half the $MTTF_D$ of the function channel (This ratio is suggested as a lower bound for the $MTTF_D$ of the monitor channel in (EN ISO 13849-1, 2015)). The fraction of common-cause failures of the functional channel and the monitor channel is assumed to be $\beta = 2\%$. The rate of demands of the safety function is $r_D = 360\ h^{-1}$, corresponding to a demand every 10 seconds. The mission time is $T_M = 20$ a.

While for a $DC = 60\%$, the *PFH* is dominated by F_{DU} for $MTTF_D$ values above approx. 48 years, the $M_D \rightarrow F_{D,D} \rightarrow$Demand sequence is dominant for lower $MTTF_D$ for DC = 60% and for all $MTTF_D$ values up to 100 years for $DC = 90\%$ and $DC = 99\%$.

Beside the influence of the finite frequency of the diagnosis on the *PFH* discussed earlier in this paper (see Equation 11) this shows how big the impact of failures of the monitor channel is and how important it is to integrate these failures correctly into *PFH* calculations. Furthermore, one could improve the system by introducing a diagnostic on the monitor channel itself, e.g. by the Functional channel. The necessary extension to the model and the results of the simulation will be discussed in the following chapter.

3.4 *Complex systems beyond designated architectures*

The properties of monitor channels become even more important for more complex systems which cannot easily be mapped onto designated architectures underlying the simplified formulas of established standards (EN ISO 13849-1, 2015, IEC 62061, 2015).

Traditionally, the architecture of a system was the basis criterion for the achievable safety level for Industrial safety systems. (EN 954, 1996). Even though the target measures for dangerous failures have been established as the defining criterion for Safety Levels (EN 61508-1, 2010) and have been introduced in the generic standards for safety of machinery (EN ISO 13849-1, 2015, IEC 62061,

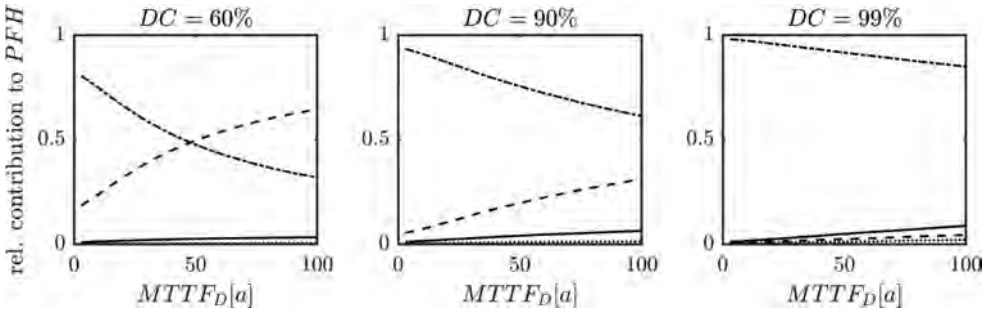


Figure 8. Relative contribution of the individual failure channels. Dotted line: F_{DD}, dashed line: F_{DU}, solid line: Common-Cause failure, dotted dashed line: M_D followed by F_D.

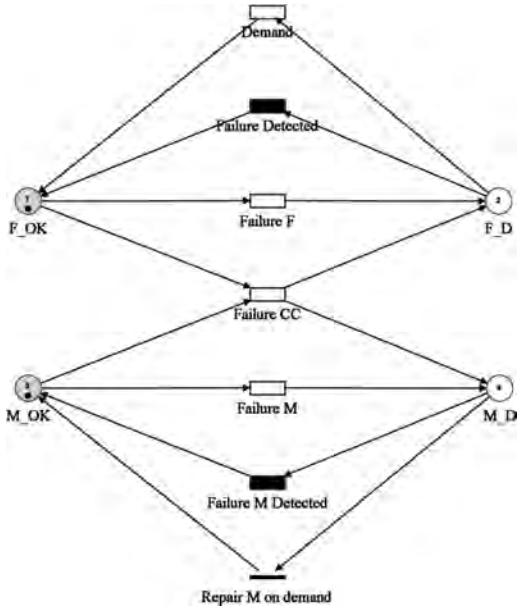


Figure 9. Petri net model of a system with a monitor channel which is monitored by the functional channel.

2015), the architecture is still an important criterion.

However, mapping modern software-intensive systems to the designated architectures of the established standards is not always easy. Programmable electronics introduce various levels of diagnostics into the systems, which can be hardly mapped to the assumptions of the designated architectures. Some standards (EN ISO 13849-1, 2015, IEC 62061, 2015, EN 61508-2, 2010) suggest the application of stochastic methods for these cases. However, due to the explosion of the state space for complex systems with many constituents, Markov Models are difficult to apply. Furthermore, the timing of monitor channels frequently does not follow exponential probability distributions. Monitor channels may in some cases even be synchronized to the potential demands of the safety function—especially in systems with cyclic working processes and the stochastic modelling of these systems becomes increasingly complex. Petri nets offer a very versatile tool to model even complex scenarios, where various monitor channels work concurrently—possibly on different time scales, with multiple different ranges of coverage and test conditions.

An example of such a system is depicted in Figure 9. The model is a slight extension to the established designated architecture of a 1oo1D system. The system comprises a monitor channel with a relatively high diagnostic coverage ($DC_F = 99\%$), which is in return monitored by the functional channel itself with a diagnostic coverage of $DC_M = 99\%$. The $MTTF_D$ of both channels is assumed to be equal.

Monte-Carlos simulation of the system reveals a substantial reduction of the *PFH* by the additional diagnosis of the monitor channel itself. In Figure 10 the relative reduction of the *PFH* is depicted as a function of the $MTTF_D$ (assumed to be equal for function and monitor channel). For low $MTTF_D$ values, the reduction exceeds 90%, while for a high $MTTF_D$ value of 100 a, the reduction still amounts to 69%.

The rate of the diagnosis of the functional channel is assumed to be 100 times the demand rate $r_d = 360\ \mathrm{h}^{-1}$. The monitor channel itself is checked with a frequency of $r_{t,M} = 1/24$ h.

The variation of this frequency is, however, less critical for the *PFH* value than the frequency of the diagnosis of the functional channel by the monitor channel. While the latter frequency has to be substantially higher than the demand rate to ensure the detection of failures of the function channel before any demand, the test of the monitor channel is efficient as long as the test rate is substantially lower than the failure rate of the functional

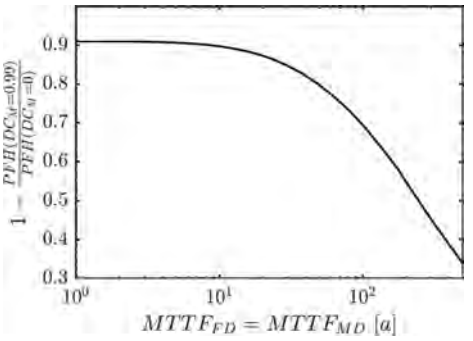


Figure 10. Effect of the additional diagnosis of the monitor channel itself. Depicted is the relative reduction of the *PFH* by the additional diagnosis on the monitor for a diagnostic coverage of the monitor of $DC_M = 99\%$ as compared to $DC_M = 0\%$ (no diagnostics on the monitor channel).

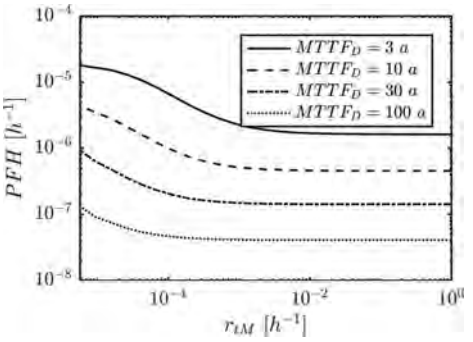


Figure 11. Dependency of the *PFH* on the rate at which the monitor channel is tested, r_{tM}, The four curves correspond to various $MTTF_D$ value (assumed to be equal for function and monitor channel): $MTTF_D = 10$ a (dashed line), $MTTF_D = 30$ a (dashed-dotted line), $MTTF_D = 3$ a (solid line) and $MTTF_D = 100$ a (dotted line).

channel. The dependency of the *PFH* on the rate at which the monitor channel is tested, r_{tM}, is depicted in Figure 11 for various $MTTF_D$ values (assumed to be equal for function and monitor channel). It is obvious, that the diagnosis of the monitor is efficient even for a very low $MTTF_D = 3$ a and a test rate as low as $r_{tM} = 10^{-2}$ h^{-1}. Higher test rates will not substantially decreased the *PFH* for the case of $MTTF_D = 3$ a.

4 SUMMARY

In this paper we presented two approaches for the determination of the *PFH* for single channel systems with diagnostics, one based on a Markov model, the second using Petri net based Monte-Carlo simulation. We introduced a detailed (non-Markovian) state transition model of the 1oo1D architecture with non-exponential distributions and sketched how this model can be transformed to an approximate 2-state Markov model by valid approximations and appropriate merging of states, and by use of a so-called flow partitioning node. For this 2-state Markov model we presented a general analytic formula for the calculation of the *PFH*, which includes important features of 1oo1D systems, e.g. it takes into account the effect of non-time-optimal testing of the functional channel performing the safety function.

We furthermore demonstrated how the same system can be modelled by a Petri net and obtained excellent conformity of the calculation results of the *PFH* between the Markov-based analytic formula and a Monte-Carlo simulation of the Petri net. Both approaches are also in excellent agreement with the *PFH* values given in ISO 13849-1 (ISO 13849-1, 2015) as table values.

The Petri net model was further used to investigate the importance of the contributions of the various failure sequences to the overall *PFH*. It revealed the large influence of sequences in which the Monitor fails prior to the Functional channel.

In the last part of the paper we argued how the Petri net model of 1oo1D could be extended to deal with architectures which are not covered by the dedicated architectures of standards on functional safety. We showed how additional diagnostics on the monitor itself could be used to substantially improve the *PFH* of the modeled system.

With Industry 4.0, modern, software-intensive systems with their versatile possibilities for online diagnostics will play an important role for the safety of machinery. Our results open up for a transparent and flexible determination of the *PFH* for these systems.

REFERENCES

Brissaud, F. & Oliveira, L., 2012. Average probability of a dangerous failure on demand: Different modelling methods, similar results. Proceedings of the 11th International Probabilistic Safety Assessment and Management Conference & the Annual European Safety and Reliability Conference. Helsinki, Finland: 6073–6082.

Dorra, M., 2017. Markov model-based calculation of the PFH_D of safety functions for machines. Unpublished manuscript, proposed for usage within standardization.

Dutuit, Y., Rauzy, A.B. & Signoret, J.P., 2008. A snapshot of methods and tools to assess safety integrity levels of high-integrity protection systems, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 222, no. 3: 371–379.

EN 954-1, 1996. Safety of machinery—Safety-related parts of control systems—Part 1: General principles for design. Berlin: Beuth Verlag.

EN ISO 13849-1, 2015. Safety of machinery—Safety-related parts of control systems—Part 1: General principles for design. Geneva: International Organization for Standardization.

IEC 61508-1, 2010. Functional safety of electrical/electronic/programmable electronic safety-related systems—Part 1: General requirements. Brussels: Comité Européen de Normalisation Electrotechnique.

IEC 61508-2, 2010. Functional safety of electrical/electronic/programmable electronic safety-related systems—Part 2: Requirements for electrical/electronic/programmable electronic safety-related systems. Brussels: Comité Européen de Normalisation Electrotechnique.

IEC 61508-4, 2010. Functional safety of electrical/electronic/programmable electronic safety-related systems—Part 4: Definitions and abbreviations. Brussels: Comité Européen de Normalisation Electrotechnique.

IEC 61508-6, 2010. Functional safety of electrical/electronic/programmable electronic safety-related systems—Part 6: Guidelines on the application of IEC 61508-2 and IEC 61508-3. Definitions and abbreviations. Brussels: Comité Européen de Normalisation Electrotechnique.

IEC 62061, 2015. Safety of machinery—Functional safety of safety-related electrical, electronic and programmable electronic control systems. Geneva: International Electrotechnical Commission.

Innal, F., Dutuit, Y., Rauzy, A. & J.P. Signoret, 2010. New insight into the average probability of failure on demand and the probability of dangerous failure per hour of safety instrumented systems. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability,* vol. 224, no. 2: 75–86.

Satodev, 2017. GRIF—GRaphical Interface for reliability forecasting. Accessed 13 December 2017. <http://grif-workshop.com/grif/>.

Slovák, R.; Kassev, K.; Stoytcheva, N.; Ivanov, E. & Schnieder, E. 2007. General Stochastic Modelling for Quantitative Safety Analysis Using Markov Chains and Petri Nets. *Information Technologies and Control. (2)*: 17–30.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Failure rates of safety critical equipment based on inventory attributes

S. Håbrekke & S. Hauge
SINTEF Technology and Society, Trondheim, Norway

L. Xie & M.A. Lundteigen
NTNU, Trondheim, Norway

ABSTRACT: Reliable failure rate estimates of safety critical equipment is crucial for verifying performance requirements and for trending the safety performance of the equipment. Joint industry efforts, like OREDA, Exida and PDS handbooks, supported by international standardization on data collection such as ISO 14224, publish generic failure rates for selected equipment commonly used in the oil and gas industry. Such generic data builds on field experience and ensures a transfer of knowledge from the operational phase to new design projects. Currently, most generic data is generated for a specific equipment group, and not for separate inventory attributes, such as size, type/fabricate, service and flow medium. Some standards and methods, like MIL-HDBK-217F, suggest how to encounter effects of inventory attributes, but these approaches are also generic, and does not account for sector specific experience, i.e. from operation of oil and gas facilities. In light of the increased focus on digitalization, it is expected that the access to data will be improved, and it is therefore important to utilize these data more efficiently in the business sector in question. The PDS forum in Norway, who gathers most actors involved in Norwegian oil and gas industry, initiated a study to analyze operational data of safety critical equipment, with the purpose to study more specific and recent effects of inventory attributes. Data from several oil and gas facilities in Norway, both offshore and onshore, have been systematized and analyzed. The purpose of this paper is to present the approach used to analyze these data, including data collection and statistical methods, as well as the final results of the study. The starting point was inventory attributes suggested by expert judgments, and their effects were investigated with basis in the collected data. Information from the operating companies' maintenance system has been too sparse to support all suggested inventories, and the choice of inventory attributes were narrowed down for some selected equipment groups; fire and gas detectors, level transmitters, shutdown valves and pressure safety valves.

1 INTRODUCTION

1.1 *Background*

The Petroleum Safety Authority (PSA) in Norway requires (in Management Regulations, section 19) that the operators shall collect, process and use field-based reliability data to ensure that the safety systems perform according to specified requirements. A key task of safety management is therefore to register equipment failures and to use this information to verify that the systems are sufficiently reliable. For the oil and gas industry, it is vital to share field experience with new projects, to make realistic assumptions about the performance of new equipment in known operating environment. Equipment failures are therefore collected from several facilities under the framework of ISO 14224 (ISO 2016a), and generic failure rates based on operational experience are presented in PDS handbooks (SINTEF 2013a), OREDA handbooks (OREDA 2015a, OREDA 2015b) and Exida handbooks (Exida 2015).

Safety Instrumented Systems (SISs) perform safety critical functions such as to shut down the plant or isolate ignition sources. IEC 61508 (IEC 2010) and IEC 61511 (IEC 2016), which are mandatory standards to use for design and operation of SISs, suggest a risk-based approach to the formulation of reliability requirements. The starting point is a risk analysis, which defines the necessary risk reduction for each Safety Instrumented Function (SIF) carried out by a SIS. This risk reduction is translated into a Safety Integrity Level (SIL) requirement. Four different SIL levels are defined (SIL 1 – SIL 4), and for each level it is specified a required reliability performance interval. Probabilistic calculations are needed to demonstrate that each SIF meets the given SIL requirement. Since the risk analysis considers what is acceptable risk at a given facility it is important to use "realistic" reliability data when the performance

of the SIFs are estimated. Realistic in this context means to consider historic field experience data, rather than data obtained strictly from analyses and/or from testing in a laboratory with controlled environment—not covering all possible failure causes experienced in operation. In other words, the reliability calculated in design should as far as possible reflect the reliability that is experienced under typical conditions in the operational phase. This is a requirement that is also emphasized and strengthened in the new edition of IEC 61511 (IEC 2016); operators must ensure that data are both credible, traceable, documented, and justified.

A practical challenge with generic data, is the time from data are collected to publishing of updated handbooks and data bases. The time lag is often five years or more, and it is therefore of interest for operators to collect and systemize their own operational experience. Many operators in Norway now carry out regular (e.g. annual) reviews of reported failures. The results from these reviews are then used to monitor reliability performance in operation, to give feedback to manufacturers (about problems experienced) and to make decisions about changes in functional test intervals (Hauge & Lundteigen 2008).

Reviews of failures reported on various Norwegian oil and gas facilities, indicate that failure rates for similar types of equipment can be quite different between facilities, even if the operating environment is more or less the same. This result may be explained by some variations in technology used (e.g. detection principle for gas detectors), process medium, the external environment, quality and frequency of maintenance and inspection, etc. Care should therefore be taken to use generic failure rates in studies for reliability demonstration without considering such influencing factors. Consequently, it is desirable to supplement generic failure rates (that represent an "average performance" for comparable equipment) with equipment characteristic parameters that can identify more specific values of failure rates, i.e. *inventory attributes* of the equipment (e.g. size of valve, type of detector, etc.). The term inventory attribute is introduced for equipment attributes particularly important for the reliability performance. For example, it may be of interest to distinguish between failure rates for different sizes of shutdown valves.

We foresee at least two applications for failure rates based on specific inventory attributes: 1) Monitoring the reliability of existing SIFs, allowing for the specific characteristics of the equipment, and 2) Calculating the reliability of new SIFs, or existing SIFs considering the influences of design, operation, environment and maintenance characteristics.

1.2 *Objective and scope of paper*

The main objective of this paper is to identify which inventory attributes that may be relevant for the four equipment groups; fire and gas detectors, level transmitters, shutdown valves and pressure safety valves, and to analyze which of the suggested inventory attributes (if any) are significant based on systemized field experience gathered by SINTEF in the period 2006–2016.

The content of this paper is based on work performed as a continuation of a research project funded by the Norwegian Research Council and the members of the PDS forum (www.sintef.no/pds). SINTEF has previously systematized failure data for six offshore and onshore oil and gas facilities (a total of more than 13000 maintenance notifications) in Norway. These failure reports, supplemented by additional expert judgements, have been used for selecting possible inventory attributes influencing the failure rates of some selected equipment types. Then, data for these inventory attributes have been collected (as complete as possible) to be able to analyze the possible impact of selected inventory attributes.

2 FAILURE DATA COLLECTION AND DATA FORMAT

2.1 *Generic data sources*

Generic failure rates are mainly derived from data collected by an organization and published in handbooks or as computerized databases (Rausand 2014). The failure rates can often be regarded as an average of the experienced performance for specific equipment groups.

The oil and gas industry has collected failure data over many years and for several offshore facilities, mainly on the Norwegian continental shelf. Relevant generic data sources are the OREDA handbooks, ref. OREDA (2015a) and OREDA (2015b), PDS handbooks, ref. SINTEF (2013a) and SINTEF (2013b), and the safety equipment reliability handbook (SERH) published by Exida (2015).

2.2 *Operator's data*

ISO 20815 (ISO 2008) on production performance assurance, emphasizes that the systematic collection and treatment of operational experience is considered as an investment and means for improvement of production and safety critical equipment. The oil and gas industry has been in the forefront of developing international standards on data collection with ISO 14224 (ISO 2016a).

Reliability data can help operators to plan the preventive maintenance, e.g. to optimize test

intervals, avoid unscheduled stops and reduce the amount of corrective maintenance. Aggregated data, used to determine generic values of failure rates, represents an experience transfer from operation to analyses needed for new facilities and for installations of new systems.

Several operators are continuously working on systemizing their failure records, to have their own "preferred" data set. This data set can be used to estimate an average performance of equipment for a single facility or for several facilities with the same operator. If the amount of data is extensive, one can also estimate separate failure rates for specific equipment attributes.

2.3 *Failure data from operational reviews*

Failures revealed during operation and maintenance are reported by maintenance notifications. A notification allows some free text description of the failure and about the measures implemented to correct the failure. In addition, it is also possible to characterize the failure, by ticking off in lists of predefined classes of failure causes, failure modes, and detection methods. Most maintenance systems are aligned with ISO 14224 for data collection, and additional effort is needed to further classify the failures into Dangerous Detected (DD) failures, Dangerous Undetected (DU) failures and safe (S) failures, for alignment with IEC 61508 and IEC 61511 taxonomy. Information about inventory attributes of interest may be partly available in notifications and partly in SIS related documents, such as the Safety Requirements Specification (SRS), safety manuals, and Safety Analysis Reports (SARs).

The operating companies themselves can perform regular operational reviews, or they can use assistance from consultants or research institutes. In either case, it is important to involve personnel from key disciplines such as automation, safety and maintenance from the specific facility and company in question. The main purpose of the reviews is to verify the performance of SIL rated equipment and to give recommendations related to maintenance and testing. A secondary purpose of the review, as suggested in this paper, is to analyze such data in more detail to investigate the performance for various inventory attributes.

2.4 *Selection of equipment groups*

Several types of safety critical equipment are used in SIFs on an oil and gas facility. Operational reviews of safety critical equipment covers about twenty different equipment groups, however, some groups consist of few equipment units. To limit the scope of the analyses in the PDS project, it was decided to extract groups of equipment where:

- a certain amount of data (both failures and a certain amount of aggregated operational time) has been gathered.
- the equipment group is represented on several facilities.
- the equipment group is represented in several SIFs on a facility.
- the equipment types have some attributes that are considered as significant with respect to the failure rate.

The selection of equipment groups and inventory attributes was consulted with experts within the PDS forum participants in an experts meeting. The recommendation was to focus on fire and gas detectors, level transmitters, shutdown valves and pressure safety valves, since these groups both contain a significant amount of equipment and contain possible significant inventory attributes.

2.5 *Uncertainty & data collection challenges*

A major challenge when splitting up the failure rates according to inventory attributes is to obtain sufficient statistical confidence. If no DU failures have been experienced in an observation period, the statistical confidence is lower even if the observation time is quite extensive.

Quality of data is another challenge. It is important that we can rely on the information given in the notification, e.g. that the failure mode and the detection method have been classified correctly. Data collection is both time consuming and demanding; it is seldom straightforward to identify all relevant information from the maintenance records. To obtain correct information about the actual failure, it is often necessary to discuss individual notifications with operators and maintenance personnel. Such work is time consuming, but nevertheless rewarding, e.g. to avoid repeating (and thereby costly) failures. Many operational reviews reported repeating failures, where seemingly insufficient measures had been implemented to remove the cause of failure. For the purposes of analyzing inventory attributes, we have removed repeating failures, to avoid that these are given too high weight in the overall results.

From the operational reviews, we saw that the effects of local facility conditions were important and that based on our experience, should be considered. One facility may experience specific problems (e.g. icing) for some particular type of equipment, which are not observed at other facilities. Local conditions seem to be of particularly interest for the occurrence of Common Cause Failures (CCFs), i.e. failures that are dependent due to a shared cause and which occur close in time. An example of a local problem which turned out to be defined as a CCF, was a number of failures

for shutdown valves caused by wrong type (here viscosity) of hydraulic oil. Some of the failures related to specific problems at one facility (such as the hydraulic oil problem) which are not likely to occur at other facilities, have been removed from our data set for the analyses of inventory attributes. Other local problems, that were defined as CCFs, were icing problems. Icing is often more challenging for facilities in the Barents Sea compared to the North Sea, however, unfortunate design solutions may also allow for icing to occur. Failures related to such conditions that may occur on several facilities, have *not* been removed from our data set for the analyses of inventory attributes.

3 STATISTICAL METHODS

For the analyses of data from operational reviews, the focus has been on DU failures since these failures will influence the most important performance requirements related to SIF equipment, such as the Probability of Failure on Demand (PFD) for a SIF. The total data set for analyses comprises all equipment units that have been involved in the operational reviews, i.e. has been part of equipment groups considered in the reviews. For each unit, data on inventory attributes has been collected together with the information about if the unit has experienced a DU failure or not in a predefined observation period.

It was decided to introduce a Generalized Linear Model (GLM) based on a binomial distribution, where only two possible outcomes are considered. The response variable is a discrete variable with two possible outcomes; 0 and 1, i.e. "No DU failure" or "DU failure", such that GLM can predict failure probability and assess effect on failure probability from the predefined inventory attributes. Other methods, such as lifetime modelling was also considered. However, since the observation period for some of the facilities does not cover the entire lifetime of the item and the time an item was put into operation is unknown for most of the components, such methods were disregarded.

GLM describes the statistical relationships between *response variables* Y_1, Y_2, …, Y_N and *explanatory variables* x_1, x_2, … x_k by estimating the corresponding inventory attributes β_1, β_2, …, β_k. An explanatory variable is a type of independent variable that can affect response variables, which may be fixed by the experimental design. GLM are mostly based on maximum likelihood estimation and allows for regression modeling when response variables are distributed as one of the members of the exponential family. The model is given by:

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_{ik} x_{ik} \quad (1)$$

and

$$y_i = g(\mu_i), \mu_i = E(Y_i). \quad (2)$$

Here, $g(\mu)$ is called the link function. Further, let $Yi \sim Binomial(n, p_i)$ express the response variables with failure probability p_i, i.e. the likelihood for an item to fail at a given time. Then the GLM model based on binomial distribution is given as:

$$\log(\frac{p_i}{1-p_i}) = \beta_{i0} + \beta_{i1} x_{i1} + ... + \beta_{ik} x_{ik}. \quad (3)$$

In this GLM, inputs are related to inventory attributes as well as failure data. Outputs of the model are related to failure probabilities that are used to check whether there is statistical significance. The formula of failure probability is:

$$p_i = \frac{\exp(\beta_{i0} + \beta_{i1} x_{i1} + ... + \beta_{ik} x_{ik})}{1 + \exp(\beta_{i0} + \beta_{i1} x_{i1} + ... + \beta_{ik} x_{ik})}. \quad (4)$$

We identify the parameters that significantly impact on the reliability performance by checking variables in the regression model with small p-values and large coefficients. A small p-value suggests that changes in the explanatory variable are associated with the changes in the response variable, i.e. the inventory attribute is significant and does influence the DU failure rate (based on our data and under the given assumptions). The exponential coefficient represents the change in the response variable when changing the categories of one inventory attribute holding the other explanatory variables constant.

4 DATA COLLECTION OF INVENTORY ATTRIBUTES

4.1 *General*

Collection of data for inventory attributes turned out to be a rather time-consuming activity. Only parts of the relevant information were (easily) found in the operator's maintenance system. It was necessary to supplement with information from other sources, such as process and instrument diagrams (P&IDs), data sheets and manufacturer specifications together with discussions with technical advisors and process engineers.

The manufacturer name was the most straightforward inventory attribute to obtain from the maintenance system. However, within an equipment group we would find some ("small") manufacturers that had not delivered more than a few equipment units each. Thus, to be able to achieve

some rational results, the number of manufacturers were kept to a minimum by grouping all the "small" manufacturers into an "other" group. Grouping the outcomes represented by a small number of units into a common "other" category, was also performed for other attributes with several outcomes/categories.

4.2 *Fire and gas detectors*

The expert review meeting suggested the following inventory attributes for fire and gas detectors: *Manufacturer*, *measuring principle*, i.e. which physical principle the detection is based on, and *model type*. Table 1 shows in more detail the inventory attributes for point gas detectors. The same types of inventory attributes were selected for line gas detectors, smoke detectors and flame detectors.

For the analyses of detectors, we were in the fortunate situation to access more data than from the six facilities where SINTEF was involved in operational reviews. For this particular equipment group, it was possible to add the inventory attribute "facility", due to the extensiveness of data, to allow comparison in failure rates and effects of inventory attributes between different facilities.

4.3 *Level transmitters*

Level transmitters are often placed in a group called "process transmitters", together with temperature transmitters and pressure transmitter. In our analyses, we wanted to focus on the level transmitters alone, mainly because they are more dependent on measuring principle and operating conditions (various medium, foaming, calibration challenges, etc.) than the other transmitters. Measuring principles for level transmitters are divided into the categories; displacer, pressure, radar (guided wave radar) and others (nuclear, ultrasonic, servo, capacitance and magnetostrictive).

Table 2 shows the list of inventory attributes agreed upon among the experts to include in the analyses. However, due to lack of details in collected data, we were left with three credible inventory attributes; *manufacturer*, *measuring principle* and *number of medium phases*. Regarding the measuring principle, we made the following assumptions based on the type of vessels for which the transmitters were installed: Level transmitters for 1st stage separators and test separators normally measure three types of medium (e.g. oil, MEG/water and gas) while level transmitters in scrubbers, 2nd stage separators and 3rd stage separators are supposed to be used to measure for two types of medium, e.g. liquid/gas. In case this information about vessel type was not evident, the number of vessel outlet lines was checked against e.g. P&IDs to obtain number of fluid phases inside the vessels.

Despite the effort, we were left with several transmitters where information was missing. E.g., measuring principle was not identified for 10% of the transmitters and manufacturer was not identified for 11%.

Table 1. Inventory attributes—Point gas detectors.

Attributes	Incl.	Examples of categories
Manufacturer	YES	Autronica, Dräger, Simtronics…
Measuring principle	YES	IR, Wireless, Acoustic…*
Model	YES	HC200, PIR 7000, GD10…

*Catalytic gas detectors have been removed from the data set due to significant more DU failures than the rest of the measuring principles.

Table 2. Inventory attributes—Level transmitters.

Attributes	Incl.	Examples of categories
Manufacturer	YES	Vega, Fisher-Rosemount…
Measuring principle	YES	Displacer, Pressure, Radar…
No. of medium phases	YES	1, 2 or 3
Type of medium	NO	Hydrocarbon, Water, Chemical…
Type of vessel	NO	Separator, Scrubber, Tank…
Special problems	NO	Foaming, Sand, Scale…

4.4 *Shutdown valves*

The data collected through operational reviews contains in total 1245 Emergency Shutdown (ESD) and Process Shutdown (PSD) valves. Table 3 shows the inventory attributes selected in the expert review meeting. Unfortunately, it was necessary to remove all data from one of the facilities due to very sparse information on the inventory attributes manufacturer and (valve) size.

Categories for *manufacturer*, *size* and *type* of valve were obtained from the maintenance system and equipment and facility specific information such as data sheets and P&IDs.

The process *medium* exposing the valves was assessed by experts at one of the facilities: It was suggested that for each system at the facility a corresponding (typical) medium could be assumed, and this "mapping" between system and medium was adopted for the rest of the facilities.

To avoid too many categories for valve size, group size intervals were decided together with an expert. Criteria for these intervals were based on the valve and process characteristics, rather than

Table 3. Inventory attributes—PSD and ESD valves.

Attributes	Incl.	Examples of categories
Actuation principle	NO	Electric, Hydraulic, Pneumatic…
Manufacturer	YES	Tai Milano, Swagelok, BIS…
Medium	YES	Gas, HC liquid, Water…
Size	YES	0–1″, 1–3″, 3–18″ and >18″
Special problems	NO	Corrosion, Icing…
Type	YES	Ball, Gate, Butterfly, Other

Table 4. Inventory attributes—Pressure safety valves.

Attributes	Incl.	Examples of categories
Actuation principle	NO	Pilot, Spring, Pressure-vacuum…
Dirty or clean service	YES	Yes or No
Manufacturer	YES	Petrolvalves, O.M.S.,…
Medium	NO	Gas, HC liquid…
Size	NO	

having equally sized categories: E.g., valves less than 1″ has been defined as a separated category since they normally are water-based and attached with lower risk compared to bigger valves. Thus, the number of valves in each size category varies.

The categories for inventory attribute "actuation principle" was not straight forward to retrieve. This would require time-consuming manual information, and was only performed for one of the facilities. Hence, the actuation principle inventory attribute was removed from the analyses. Also, the inventory attribute "specific problems" (corrosion, icing and temperature changes) was removed, as this information required manual and rather time-consuming effort.

Table 3 lists the examples of categories assigned to selected inventory attributes. Also for this equipment group, information was missing. For example, we found that 14% of the valves had unknown manufacturer, 13% had unknown size and 8% had unknown type.

4.5 *Pressure safety valves*

The inventory attributes that were selected for pressure safety valves (PSVs) based on expert meeting are presented in Table 4. *Manufacturer* and size of valve were obtained from the maintenance system and equipment and facility specific information such as data sheets and P&IDs. However, we faced major problems with missing category information. E.g. for about 50% of the PSVs, information about the valve size was not found in the maintenance system. Thus, the inventory attribute "size" was not part of the PSV analyses. Unfortunately, we also had to remove data for PSVs from one of the facilities due to missing information.

Dirty or clean service, i.e. if the medium flowing through a PSV is "dirty" (e.g. including sand, crude oil, etc.) or "clean" (e.g. pure gas), was together with the actuation principle pointed out by the experts as a possible significant inventory attribute for the PSVs. To simplify the analyses, it was assumed that all PSVs installed in the same system had the same category (either "dirty" or "clean"). The inventory attribute "medium" was not included, since it would be partly correlated to the inventory attribute "service".

The actuation principle of the PSVs was identified to some extent in the maintenance system. As for ESD and PSD valves, it was very time-consuming to extract this information and this has not yet been performed. Hence, the actuation principle was not part of the analyses.

Table 4 summarizes the selected inventory attributes. Note that examples of categories are not provided for the inventory attribute "size" since it was decided to omit this one from the analyses.

5 ANALYSES AND RESULTS

5.1 *Assumptions*

The data for all the finally selected inventory attributes were for the analyses combined with information about how many DU failures that had been registered for the equipment group and the aggregated time in operation. In the data set, we removed DU failures that had been repeated for the same equipment, to avoid double counting of the same failure event.

For some of the inventory attributes, e.g. medium for PSD and ESD valves and dirty or clean service for PSVs, the categories are based on the assumption that all equipment installed in one particular system share the same medium and service; e.g. all valves in system number 43 (Flare system) is assumed to share the medium "gas" and "clean" service.

The number of predefined categories and of course how they are defined, e.g. size intervals and which categories belonging to the "other" category, will also impact the results.

Some assumptions have also been made regarding the analyses and for the data to fit the statistical analyses as described in section 3. E.g., the DU failures are assumed to be identically distributed and to occur stochastically independent. It is also assumed that inadequate information and missing data do not have any effect on the results of the analyses.

The observation periods from each facility is not equal and varies from two to 11 years for those facilities included in the data set. Thus, some assumptions about observation periods had to be made to get observations periods as equal as possible and to utilize all DU failures: The final periods should not be too short such that there would be very few observation periods with failure compared to observation periods without failures—then it would be more difficult getting significant results. On the other hand, the periods should not be too large such that multiple failures of the same component often would occur within the same period—then we would not utilize all the DU failures. Also, different observation period intervals were concerned in data analyses for fire and gas detectors compared to other equipment groups:

For those equipment groups with data from five or six facilities (PSD and ESD valves, level transmitters and PSVs) three–four years was regarded as one observation period. Then, for a facility with observation period between three and four years, each equipment unit was counted once in the total data set. For a facility with 11 years of operational experience, the inventory data was counted for three times in the total data set (and the DU-failures were distributed on the correct observation period based on the notification date).

For fire and gas detectors, where data from several facilities was included and many of those with shorter observation periods, two–three years was regarded as one observation period.

The GLM model was implemented in software R, which is a free software for statistical computing and graphics.

5.2 *Results*

The aim of the analyses was for each equipment group to identify which inventory attributes and related categories that became statistical significant (if any).

Table 5 shows the results of the analyses for each equipment group, listing the most significant inventory attributes and associated categories—with respect to the DU failure rate. Note that not all attributes and categories have been found to be significant, and they are therefore not listed. Neither are those categories less significant compared to two or more other categories. "Significant" implies that the inventory attribute and its associated category(s) either contribute to significantly higher failure rate or significantly lower failure rate compared to the other attributes/categories.

From Table 5 we see that the manufacturer, typically represented by one or two of the largest manufacturers, is significant for most of the equipment groups. For ESD and PSD valves the largest valves seem to have a higher failure rate compared to small valves. Also, the medium may be important for the failure rate for ESD/PSD valves.

Table 5. List of most significant inventory attributes and categories.

Equipment group	Attribute	Category
ESD/PSD valves	Size	>18″
	Medium	Gas, Water, …
	Manufacturer	*Confidentially*
Line gas detectors	Manufacturer	*Confidentially*
Point gas detectors	Measuring principle	IR
PSVs	Manufacturer	*Confidentially*
	Dirty or clean service	Dirty, clean

One inventory attribute suggested by experts, and that we was able to analyze, was not found to be of significant in our analyses: "measuring principle" for level transmitters. This may be partly explained by inadequate level of details concerning inventory attributes in the applied data.

For the fire and gas detectors where data was available from several facilities, also the facility was included as a separate attribute. The results showed that some facilities turned out to contribute to significant higher or lower failure rates compared to the others.

6 RECOMMENDATIONS AND FURTHER WORK

Measures and means for improving the quality of data that are recorded into the maintenance system is an important area for further research. Today, the recording is mainly manual, and there lack a systematic way for consistent recording of information for more automatic extraction and analyses. Based on the results of this study, it is possible to suggest more specific categories of information to be recorded. It may be necessary to further investigate the implications of assumptions that were made for our analyses. Both those that were made to overcome practical obstacles, e.g. due to lack of information related to selected inventory attributes, and those made to simplify the selection of categories, e.g. about the relationship between categories for inventory attributes (e.g. clean service) and system number (e.g. flare system). It is also possible to perform other types of analyses, e.g. "big data" analyses, to identify significant inventory attributes particularly when the amount of data, inventory attributes and categories increases.

Operational experience indicates that similar equipment performs differently between facilities with a comparable operating environment. It is therefore desirable to supplement the generic data

with inventory attributes that can explain the varying performance, and enable the reliability analyst to better predict the variations. Due to the limited information about inventory attributes, it is recommended that the operators increase the amount of relevant inventory information in their maintenance systems, in particular for safety critical equipment part of operational reviews. Then, the failure data and data for inventory attributes can be combined to perform in depth analyses.

Data collection is becoming increasingly important both with respect to quantity and quality. It is an important activity to provide feedback on experience from the operational phase to designers of new systems and for monitoring the operational performance of safety barriers. It is also an important activity seen in relation to the increasing trend of lifetime-extension for existing facilities. SINTEF and PDS forum is also working on means for enhancing the digitalization of failure reporting, classification, and analyses, to update the generic failure rates more frequently and to reduce the manual effort in this process. A higher level of automatic analyses of data can help when prioritizing resources needed to improve the overall quality of data.

ACKNOWLEDGEMENT

Thanks to the members of the PDS forum (www.sintef.no/pds) that have contributed with valuable information, expert judgements and meaningful discussions. Special thanks to personnel from the operating companies that have contributed with data from operational experience, expert judgements, participation in workshops and valuable input and comments.

REFERENCES

Exida 2007. Safety Equipment Reliability Handbook. Third Edition.

Exida 2015. Safety Equipment Reliability Handbook. Fourth Edition.

Hauge S., Hokstad P., Håbrekke S., Lundteigen M.A. 2015. Common cause failures in safety-instrumented systems: Using field experience from the petroleum industry. *Reliability Engineering & System Safety* volume 151: pages 34–45.

Hauge S., Lundteigen M.A. 2008. Guidelines for follow-up of Safety Instrumented Systems (SIS) in the operating phase. SINTEF report A8788.

IEC 2010. IEC 61508:2010 Functional safety of electrical/electronic/programmable electronic safety-related systems.

IEC 2016. IEC 61511:2016 Functional safety—Safety instrumented systems for the process industry sector.

ISO 2008. ISO 20815:2008 Petroleum, petrochemical and natural gas industries—Production assurance and reliability management.

ISO 2016a. ISO 14224:2016 Petroleum, petrochemical and natural gas industries—Collection and exchange of reliability and maintenance data for equipment. Edition 3.0.

ISO 2016b. CEN ISO/TR 12489:2016 Petroleum, petrochemical and natural gas industries—Reliability modelling and calculation of safety systems.

Norwegian oil and gas association 2004. 0*70 – Norwegian oil and gas application of IEC 61508 and IEC 61511 in the Norwegian petroleum industry.*

OREDA 2015a. Offshore and Onshore Reliability Data. 6th Edition. Volume 1 – Topside Equipment.

OREDA 2015b. Offshore and Onshore Reliability Data. 6th Edition. Volume 2 – Subsea Equipment.

Rausand M. 2014. *Reliability of Safety-Critical Systems: Theory and Applications.* Wiley.

SINTEF 2013a. *Reliability Data for Safety Instrumented Systems. PDS data handbook 2013 edition.*

SINTEF 2013b. *Reliability Prediction Method for Safety Instrumented Systems. PDS method handbook 2013 edition.*

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Availability modeling of a virtualized IP multimedia subsystem using non-Markovian stochastic reward nets

M. Di Mauro, G. Galatro, M. Longo & F. Postiglione
Department of Information and Electrical Engineering and Applied Mathematics (DIEM)
University of Salerno, Fisciano (SA), Italy

M. Tambasco
Research Consortium on Telecommunications (Co.Ri.TeL)
University of Salerno, Fisciano (SA), Italy

ABSTRACT: A significant fraction of current research in telecommunications is investigating new paradigms to guarantee high flexibility in deploying network infrastructures. Network Function Virtualization (NFV) is probably the most effective paradigm: it adapts tothe telecommunication networks the virtualization concepts originally conceived in the computer world. According to NFV specifications, network elements as switches or routers can be implemented as virtual machines called Virtual Network Functions (VNFs), and deployed in field by timely and cost-effective operations. Some Telco operators are already taking advantage of these virtualized resources, by aggregating more VNFs in order to provide new services. One example is IP Multimedia Subsystem (IMS), providing multimedia delivery services, that can be suitably deployed by means of interconnected VNFs. We consider a virtualized implementation of an IMS system (vIMS) that we characterize by an availability standpoint. First, we describe a vIMS system as a chain of virtualized network elements modeled by three components: hardware, hypervisor, and application. Subsequently, we model the probabilistic behavior of the network nodes by Stochastic Reward Nets, that account for failure and repair events characterizing each node. Innovating on previous formulations, part of the analysis is carried out by adopting non-Markovian models, thus allowing for more realistic (non-exponentially distributed) times between some state transitions. As final results, we determine the optimal redundant vIMS configuration able to guarantee a steady-state availability not less than 0.99999, and we provide a sensitivity analysis useful to evaluate the system robustness to variation of parameters from their nominal values.

1 INTRODUCTION

Network Function Virtualization (NFV) (ETSI 2012) represents one of the most innovative paradigms within the fifth generation (5G) of telecommunication systems. Basically, it has been designed to boost the deployment of new network services by exploiting the virtualization concepts. Within an NFV domain, traditional network equipments (e.g. firewalls, routers, switches, etc.) are replaced by their virtual counterparts named Virtualized Network Functions (VNFs). According to NFV logic, every VNF relies on a decoupled structure typically made of: *i)* a hardware part accounting for physical components; *ii)* a hypervisor part acting as an abstraction layer between hardware and application *iii)*; an application part that includes the software logic and runs on top of a VNF. An architecture that can profitably benefit from an NFV environment is the IP Multimedia Subsystem (IMS) (3GPP 2001). IMS exploits the all-IP-based paradigm to provide a plethora of multimedia services (audio/video sessions, online messaging, presence, IP TV, etc.) by taking advantage of Session Initiation Protocol (SIP) (Rosenberg et al. 2002). In the present work, the new virtualized IMS (vIMS) framework (namely, IMS in an NFV environment) has been characterized in terms of its availability. Accordingly, we find the optimal system configuration respecting the so-called "five nines" availability requirements. It consists in tolerating a maximum system downtime of 5 minutes and 26 seconds per year. The paper is structured as follows. Section 2 contains a brief *excursus* about relevant and pertinent works. Section 3 offers an overview about possible vIMS deployments. An availability analysis is then presented in section 4, where details about Stochastic Reward Nets (SRNs) and Reliability Block Diagram (RBD) methodologies are provided. Section 5 presents a numerical experiment, where characteristic system parameters as repair and failure rates of the vIMS

components have been set in accordance to scientific literature and expert hints. Finally, concluding remarks along with considerations about possible future work are drawn in Section 6.

2 RELATED WORK

Recently, the dependability and availability of novel telecommunication infrastructures are becoming critical issues since network operators are committed to strict Service Level Agreements. Accordingly, such issues have got attentions from scientific andtechnical literature. The work presented in (de S. Matos et al. 2012) is one of the first papers devoted to cope with the availability issues of virtualized infrastructures. In particular, the authors propose a three-level model of a generic system (hardware, software, hypervisor) analyzed by combining Continuous Time Markov Chains (CTMCs) and fault trees formalisms. A combination of approaches is also used in (Fernandes et al. 2012), where a dependability assessment of virtual networks is presented by exploiting both combinatorial and state-based models. A mathematical framework to model a restoration mechanism of virtual resources caused by failures is instead proposed in (Taleb et al. 2016). A stochastic model-driven approach to evaluate the availability of an Infrastructure-as-a-Service (IaaS) cloud system is presented in (Ghosh et al. 2014). The failure events have been managed by considering migration of physical machines among three types: hot (running machine), cold (turned off machine), and warm (turned on, but not yet ready machine). Some algorithms aimed at solving the Minimum Total Failure Removal (MTFR) problem have been proposed in (Liu et al. 2016), where a reliability evaluation of an NFV environment has been faced. The present work takes inspiration from some recent contributions proposed by the authors, see (Di Mauro et al. 2016), (Di Mauro et al. 2017), and, basically, provides two original developments. The first one concerns the stochastic modeling of a virtualized IMS (called vIMS) framework by considering a typical telecommunication network scenario. The second one pertains to the proposal of a more realistic model relaxing the assumption of exponentially distributed failure and repair times, characterizingMarkovian model: the choice of some non exponentially-distributed transition times has some influence on the transient analysis of systems.

3 OVERVIEW OF THE IMS ARCHITECTURE

IP Multimedia Subsystem was born as a framework able to providing access to a plethora of multimedia IP-based services with guaranteed quality of service (Camarillo and Garcia-Martin 2008). IMS architecture supports a broad range of services by exploiting the flexibility of SIP protocol, among which, multimedia and real-time sessions (such as phone calls), web messaging and enriched communications. The signaling flows are managed by the CallSession Control Function (CSCF) servers, that communicate by exchanging mainly SIP messages. The CSCF functionalities are distributed among three servers. The *Proxy* CSCF (P-CSCF) is a SIP proxy, and acts as an interface between the user device and the IMS network. The *Interrogating* CSCF (I-CSCF) forwards SIP requests or responses within the domain. The *Serving* CSCF (S-CSCF) is in charge of performing some core functions as session and routing control and user registration management. Another key element of the IMS infrastructure is the Home Subscriber Server (HSS), an advanced database containing users' profiles that can be queried by means of a specific protocol called Diameter. Such nodes are interconnected among them to provide basic and advanced services. An example is offered in Fig. 1, where a (simplified) Registration procedure (needed before exploiting IMS services) is depicted. Initially, a user device contacts P-CSCF via Register message (1). Such a message is passed to I-CSCF (3) that, in turns, sends it to HSS (3) in order to retrieve the address of S-CSCF in charge of current registration. Once obtained the information from HSS (4), the message is forwarded to the correct S-CSCF (5). Finally, an OK message indicating a correct device registration is back-propagated to the user device (6), (7), (8). Once completed the registration procedure, the device is ready to use IMS services as a real-time audio/video session for example. It is worth noting that, in Fig. 1 is depicted the signalling flow, but, typically media flows between two user devices (i.e. the content of an audio/video call) traverse a different path.

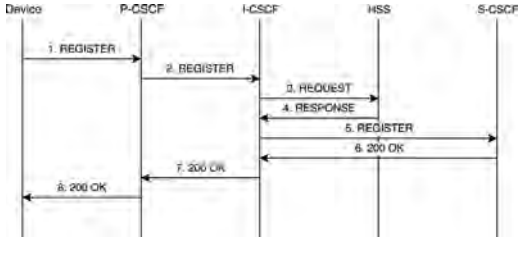


Figure 1. A simplified registration procedure in IMS domain. The *Register* message is propagated from device to S-CSCF. A 200 *OK* message is back-propagated to device if the procedure ends correctly.

3.1 *Virtualized IMS domain*

We recall that, in our proposal, the IMS framework is assumed as integrated in an NFV environment. Such a hybrid solution is getting recently attention either by technical literature (Duan et al. 2017), and by the industrial world (ETSI 2015). Accordingly, the involved nodes (P-CSCF, I-CSCF, S-CSCF, HSS) are modeled by VNFs composed, in turn, by three layers:

- *Hardware*: represents the aggregate of physical subsystems (CPU, RAM, Storage, etc.) that are often deployed in server farms;
- *Application*: represents the software logic deployed on top of a specific VNF (P-CSCF, HSS, etc.);
- *Hypervisor (or Virtual Machine Monitor)*: an intermediate (software-based) level allowing to deploy one (or more) virtual nodes on the same hardware and to manage the resource consumption of each VNF.

In our setting, the hardware and hypervisor layers are supposed to be the same for CSCF servers and for the HSS node, while the application layer is modeled separately for CSCF nodes and HSS.

4 AVAILABILITY MODEL

The system availability analysis is performed by taking into account a two-level hierarchy model combining two formalisms: Reliability Block Diagrams (RBDs) and Stochastic Reward Networks (SRNs). The former is useful to model the system in terms of interconnections among nodes (subsystems) of the considered vIMS infrastructure, and constitutes the first level. Figure 2 shows the RBD representation derived from the Registration scenario reported in Fig. 1, where a series model is necessary to represent the IMS, since all network functionalities must be active to guarantee the Registration service to users, whereas a parallel configuration for each node (replicas) are usefulto ensure a certain degree of redundancy in case of failures. Furthermore, we assume that HSS node is deployed in a *k-out-of-n:G* configuration, where

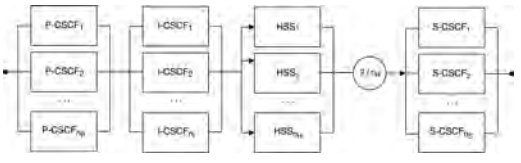


Figure 2. The Reliability Block Diagram representation of virtualized IMS domain, where HSS is deployed in a 2-out-of-n_H redundancy configuration.

k represents the number of HSS replicas that must work to make the HSS node to work. Henceforth, we assume $k = 2$, in accordance with most actual deployments.

On the other hand, the second level of the hierarchical model based on SRN is used to describe the internal behavior of a single node by characterizing the relationships among three layers (hardware, hypervisor, application). In the following subsection, after a brief description of the SRN formalism, we provide a specific model of a generic vIMS node. For further readings on the use of SRN for availability evaluation, refer to (Muppala et al. 1994)..

4.1 *Stochastic reward networks approach*

The SRN model derives from Markov Reward Model (MRM) which enhances the traditional Continuous-Time Markov Chain (CTMC) by adding a reward rate to each state. With state-space based models (such as MRMs), a classic problem of modeling real-world systems is related to the growth of state space. On the contrary, an SRN-based representation allows a more compact description of the underlying system, by identifying repetitive structures. In this way, it is possible to automatically generate the underlying MRMs (Bolch et al. 1998). SRN model adopts a bipartite directed graph representation where: *i) places* (represented by circles) specify a condition (e.g., the system is *down* or *up*), and *ii) transitions* (represented by rectangles) denote actions (e.g., a system crash). Places and transitions are connected by *arcs* denoted by directed edges. In the SRN formalism, the transition times are assumed as exponentially distributed since they take into account a probabilistic delay. A place typically contains a number called *token*, that represents an holding condition. In case of a condition change, a transition is *fired*, and the token is moved from one place to another. A measure of interest is the distributionof tokens, called *marking*, that denotes the possible assignment of tokens to all places of the underlying Markov model, and is useful to capture the dynamics of the overall system.

In the SRN context, the reward function, say $X(t)$, has a crucial role. It is a (non-negative) random process that represents system conditions, namely $X(t)$ varies with time t in accordance to the desired measures, for instance availability, dependability, or performance (Muppala et al. 1996). Being interested in availability evaluation, $X(t)$ is defined as: $X(t) = 1$ in case of working system (up condition) at t, and $X(t) = 0$ otherwise (down condition). Accordingly, it is possible to define the instantaneous availability $A(t)$ as the expected reward function at time t, i.e.

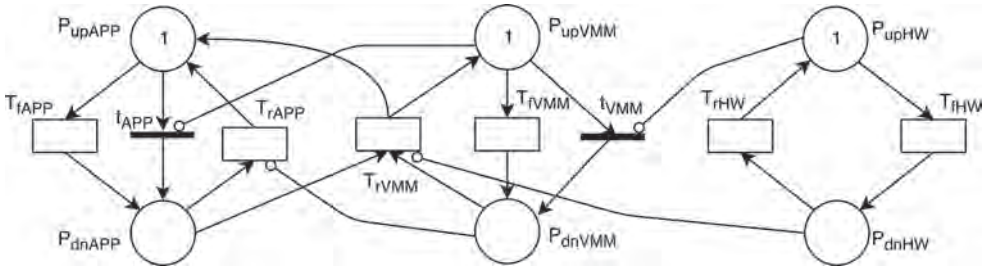


Figure 3. SRN representation of one generic node replica of vIMS network infrastructure.

$$A(t) = Pr\{X(t) = 1\} = E(X(t)) = \sum_{i \in S} r_i \cdot p_i(t), \qquad (1)$$

where S denotes the state space (namely, the set of markings within SRN) that can be broken in a subset of up states S_u (reward rate $r_i = 1$), and a subset of down states S_d (reward rate $r_i = 0$), whereas $p_i(t)$ is the probability of the system being in state i.

The SRN of a generic vIMS node replica (CSCF or HSS) is shown in Fig. 3, while the overall vIMS infrastructure is described by the RDB model presented in Fig. 2.

By inspection of Fig. 3 it is possible to distinguish the following entities:

- *Places* (circles): the group of places P_{upHW}, P_{upVMM}, and P_{upAPP}, takes into account the working conditions of hardware, hypervisor (VMM subscript stands for Virtual Machine Monitor), and application layers respectively. The numbers inside the three places indicate the corresponding initial (working) conditions. Conversely, the group of places P_{dnHW}, P_{dnVMM}, and P_{dnAPP}, represents the failure conditions of hardware, hypervisor, and application layers respectively.
- *Timed Transitions* (unfilled rectangles): such transitions take into account the various layers behavior; in particular, $T_{fAPP}[T_{rAPP}]$, $T_{fVMM}[T_{rVMM}]$, and $T_{fHW}[T_{rHW}]$ denote the failure [repair] events of the application, the hypervisor, and the hardware, respectively.
- *Immediate Transitions* (thin and filled rectangles): such transitions take into account the instantaneous actions, namely, actions characterized by zero transition time. In the proposed SRN, two immediate transitions appear: t_{APP} and t_{VMM}.

4.2 *Evolutionary model of SRN*

In this section we analyze the dynamics of the system, namely, the conditions arising when events such as failures or repairs emerge. In particular, we focus on the evolution of the SRN model of a single vIMS node. For the sake of simplicity, it is useful to consider an initial fully working condition for the node characterized by a token in each P_{up} place of the SRN. If an application failure happens, it means that the software function representing the logic of a vIMS node (a CSCF node or the HSS node) breaks. In this case, the transition T_{fAPP} is fired, and the token leaves place P_{upAPP} to enter place P_{dnAPP}. Once the application gets repaired (sometimes a trivial reboot procedure could solve the problem), the transition T_{rAPP} is fired, and the token comes back to initial place P_{up}. In case of hypervisor failure, instead, the transition T_{fVMM} is fired, and the token is moved from P_{upVMM} place to P_{dnAPP} place. It is worth noting that the place P_{upVMM} is connected to immediate transition t_{APP} by an *inhibitory* arc (the segment with a small circle close to t_{APP}). Such an arc forces t_{APP} to get fired in order to model an application failure. The application layer, in fact, needs a working hypervisor layer to work correctly. On the contrary, when the hypervisor gets repaired, the token is again moved from P_{dnAPP} to P_{upVMM} place, and the inhibitory arc is now ineffective. A similar reasoning holds for the failure of hardware. The token passes from P_{upHW} to P_{dnHW} place as transition T_{fHW} gets fired. In such a case, an inhibitory arc (between P_{upHW} and t_{VMM}) forces t_{VMM} to move the token from P_{upVMM} to P_{dnVMM} since the hypervisor layer needs an underlying functioning hardware to work properly. It is interesting to note that, another inhibitory arc connects P_{dnHW} and T_{rVMM}. Such an arc inhibits the firing of transition T_{rVMM} (and consequently the firing of transition T_{rAPP}), until the hardware layer gets repaired.

Let now be $r_{i,j}$ the reward rate assigned to marking i (i-th distribution of tokens), and $p_{i,j}(t)$ the probability of a generic node replica j modeled by the SRN depicted in Fig. 3 to be in marking i at time t. Being the markings mutually exclusive, it is possible to express the instantaneous availability $A^{(j)}(t)$ according to (1), namely

$$A^{(j)}(t) = \sum_{i \in I} r_{i,j} \cdot p_{i,j}(t), \qquad (2)$$

with I identifying the set of so-called *tangible markings* (markings with no immediate transitions enabled). The reward rate $r_{i,j}$ associated to the tangible marking i is given by:

$$r_{i,j} = \begin{cases} 1 & \text{if } \left(\#P_{upAPP} = 1\right) \\ 0 & \text{otherwise.} \end{cases}$$

It is worth noting that such a reward rate condition does not need to account for up condition of hypervisor and hardware because it is intrinsically embedded in the SRN model in Fig. 2, where inhibitory arcs avoid having a working application layer coupled with failed hypervisor and/or hardware layers. A steady-state analysis for the SRN-based model of the generic vIMS node replica j can be simply obtained by (2) for long runs ($t \to \infty$) and the corresponding steady-stateavailability $A^{(j)}$ has the form:

$$A^{(j)} = \lim_{t \to +\infty} A^{(j)}(t) = \sum_{i \in I} r_{i,j} \cdot p_{i,j}, \tag{3}$$

where $p_{i,j}$ represents the steady-state probability given by $p_{i,j} = \lim_{t \to +\infty} p_{i,j}(t)$. .

We recall that the overall vIMS system can be modeled as a series/parallel of *independent* subsystems as shown in Fig. 2 and the availability of all subsystems is derived from (3). Accordingly, by exploiting the power of RBD representation, the overall vIMS steady-state availability can be expressed as

$$A_{vIMS} = \left[1 - \prod_{j=1}^{n_P}\left(1 - A_P^{(j)}\right)\right] \cdot \left[1 - \prod_{j=1}^{n_S}\left(1 - A_S^{(j)}\right)\right]\left[1 - \prod_{j=1}^{n_I}\left(1 - A_I^{(j)}\right)\right] \cdot \sum_{j=2}^{n_H}\binom{n_H}{j} A_H^j \left(1 - A_H\right)^{n_H - j}, \tag{4}$$

where $A_P^{(j)}$, $A_S^{(j)}$, and $A_I^{(j)}$ are the steady-state availabilities of j-th replica of nodes P-CSCF, S-CSCF and I-CSCF, respectively, while the steady-state availability of HSS node replicas is $A_H^{(j)} = A_H, \ \forall j$; the numbers of redundant subsystems of each network functionality are n_P, n_S, n_I, and n_H, respectively. Being the vIMS a series system of network functionalities, vIMS steady-state availability (4) is a product of single node availability, where P-CSCF, S-CSCF and I-CSCF nodes are in parallel configuration and provide the first three terms in (4). The last factor, instead, takes into account the *k-out-of-n:G* configuration of the HSS node, where $k = 2$ and $n = n_H$. .

5 NUMERICAL ANALYSIS

This section presents a numerical analysis of the proposed IMS framework over an NFV environment by exploiting two software tools: SHARPE (Symbolic Hierarchical Automated Reliability and Performance Evaluator) (Sahner and Trivedi 1987) and TimeNet (German et al. 1995). It aims to single out the minimal-cost redundant configuration of the virtualized IMS system able to guarantee the "five nines" requirement for telecommunication systems availability. The parameters used for this analysis, and, representative of the mean time of failures and repairs as regards the various components (hardware, hypervisor, application) are reported in Table 1. The adopted values are derived in part from the experience of telecommunication experts, and in part from technical literature. For the sake of simplicity we assume that: *i)* hardware and hypervisor are supposed to be the same for all the nodes; *ii)* all the software instances running on CSCF nodes (P-CSCF, I-CSCF, S-CSCF) are characterized by the same failure and repair times. On the contrary, the software instance running on the HSS is supposed to have a different mean time of failure, being the database a more delicate and prone to failures element. As common in literature, the symbols λ and μ denote failure and repair rates, respectively. To characterize the stationary availability of the vIMS system in long runs, a steady-state analysis is carried out by considering some exemplary settings as shown in Table 2. The first column of Table 2 indicates the setting identifier $\left(S_1, \ \ldots, S_5\right)$. The second column indicates the considered redundancy level; for example, the setting S_3 is characterized by two (whatever) CSCF nodes having redundancy of 2, the remaining CSCF node having redundancy of 3, and the HSS node having redundancy of 3. The third column indicates the steady-state availability value of the whole virtual-

Table 1. Input parameters for the hardware subsystems.

Parameter	Description	Value
$1/\lambda_{HW}$	mean time for hardware failure	60000 hours
$1/\lambda_{VMM}$	mean time for hypervisor failure	5000 hours
$1/\lambda_{CSCF}$	mean time for CSCF node failure	3000 hours
$1/\lambda_{HSS}$	mean time for HSS node failure	2000 hours
$1/\mu_{HW}$	mean time for hardware repair	8 hours
$1/\mu_{VMM}$	mean time for hypervisor repair	2 hours
$1/\mu_{CSCF}$	mean time for CSCF software repair	1 hour
$1/\mu_{HSS}$	mean time for HSS software repair	1 hour

Table 2. Availability results of the whole virtualized IMS.

Setting	Redundancy Level	A_{vIMS}
S_1	$CSCF = [2,2,2]$, $HSS = 3$	0.99999416
S_2	$CSCF = [2,2,2]$, $HSS = 4$	0.99999756
S_3	$CSCF = [2,2,3]$, $HSS = 3$	0.99999497
S_4	$CSCF = [3,3,3]$, $HSS = 3$	0.99999659
S_5	$CSCF = [2,3,3]$, $HSS = 4$	0.99999918

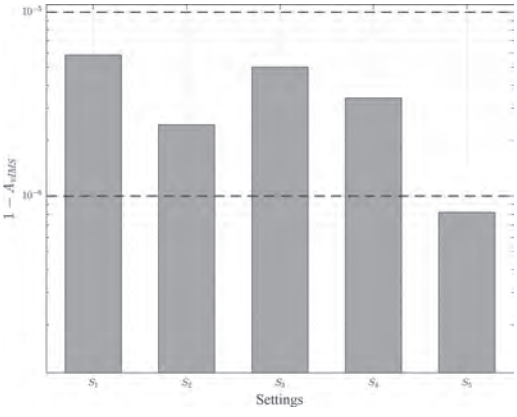


Figure 4. Steady-state unavailability for different settings $S_1, S_2, \ldots, S_5$. The above horizontal dashed line represents the required unavailability: $1 - A_{vIMS} = 10^{-5}$.

ized IMS system in Fig. 2. In order to visualize in a more comfortable manner the steady-state availability results, we show in Fig. 4 the unavailability of the virtualized IMS system $1 - A_{vIMS}$, and it is possible to observe that each setting S_i satisfies the "five nines" requirement, namely, each bar lies below the horizontal dashed line at $1 - A_{vIMS} = 10^{-5}$. In the case of setting S_5, the system is even able to satisfy the more challenging "six nines" requirement, being the corresponding bar lying belowthe horizontal dashed line at $1 - A_{vIMS} = 10^{-6}$. Among the considered settings, S_1 entails the minimum number of deployed replicas, namely, 2 replicas for each CSCF node and 3 replicas for the HSS node. Consequently, setting S_1 represents the optimal redundant configuration in terms of minimum number of deployed replicas while fulfilling the desired high availability requirement.

5.1 *Transient non-Markovian analysis*

The performed regime analysis is useful to evaluate the system behavior as $t \to \infty$, but it cannot capture the dynamics of the system when, for example, some node instances are changing their states from failed to repaired. We approach this issue analyzing the dynamics of the system when software instances (namely the application parts) of both P-CSCF node replicas are down and ready to be repaired. This is the typical case of a contemporary (and often unplanned) update of operating system that forces both software instances to be rebooted.

Actually, in order to consider a more realistic scenario, we replace the classical hypothesis of exponentially distributed software repair transition times with a Weibull-distributed one. In particular, the transient analysis evaluates the behavior of theinstantaneous availability $A_{vIMS}(t)$ and the interval availability $\overline{A_{vIMS}(t)}$ of vIMS system in (0, t], that is defined as

$$\overline{A_{vIMS}(t)} = \frac{1}{t}\int_0^t A_{vIMS}(u)\,\mathrm{d}u. \tag{5}$$

The interval availability represents the time average of the instantaneous availability function over the interval (0, t]. Figures 5 and 6 show the behavior of $A_{vIMS}(t)$ and of $\overline{A_{vIMS}(t)}$ when a Markovian process (all exponential transition times) and a non-Markovian process (Weibull software repair transition times) are considered, respectively. In the case of Weibull software repair transition times, we consider the same mean repair time used in the case of exponentially distributed software repair transition times. Therefore, we set the shape parameter of the Weibull to $\alpha = 3$ (as reported in (Guida et al. 2013)), while the scale parameter β turns out to be 1.1198. Besides, Fig. 7 highlights the different behavior of

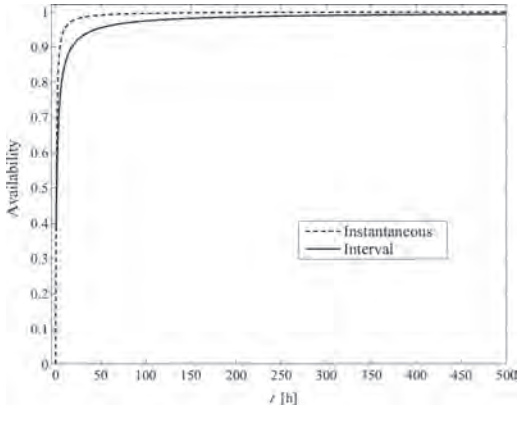


Figure 5. Instantaneous ($A_{vIMS}(t)$) and interval ($\overline{A_{vIMS}(t)}$) availability in case of exponential repair times of P-CSCF application part.

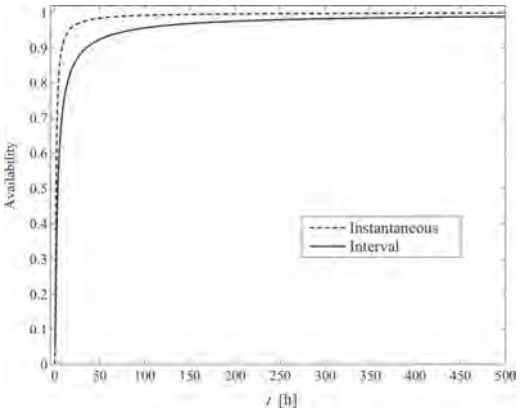


Figure 6. Instantaneous ($A_{vIMS}(t)$) and interval ($\overline{A_{vIMS}}(t)$) availability in case of non-Markovian process with a Weibull repair time of P-CSCF application part.

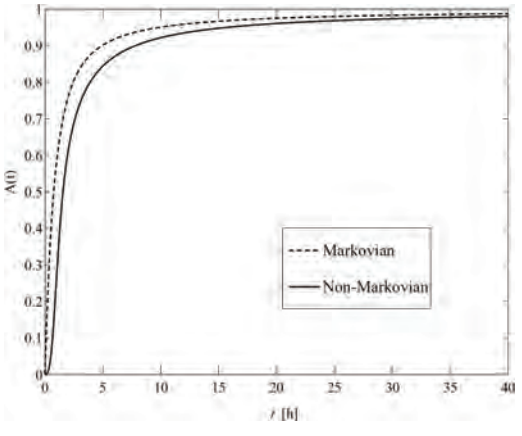


Figure 7. Instantaneous availability comparison in case of Markovian process (all exponential transition times) and non-Markovian process (Weibull repair time of P-CSCF application part).

instantaneous availability $A_{vIMS}(t)$ during the transient, in the case of exponential and Weibull failure rate (of application part). It is worth noting that the transient of $A_{vIMS}(t)$ in the Weibull case is slower than the transient of $A_{vIMS}(t)$ in the exponential case. This behavior is compatible with real-world effects (Ayers 2012). In both cases, as expected, the interval availability $\overline{A_{vIMS}}(t)$ in (5) converges more slowly to the steady-state availability with respect to the instantaneous availability $A_{vIMS}(t)$.

5.2 *Sensitivity analysis*

As far as the last part of the considered experiment, we perform a sensitivity analysis with respect to deflections of two system parameters from their nominal values: λ_{P-CSCF} and λ_{HSS}. Such an analysis has been carried out by considering the minimal cost setting S_1 obtained as a result of the regime analysis. In Fig. 8, the influence of the P-CSCF application part failure time has been considered. The nominal value amounts to 3000 hours (see Table 1), but if we relax such a value to about 1500 hours, we are still able to satisfy the "five nines" requirement. At $1/\lambda_{P-CSCF} = 1500$, in fact, the value A_{vIMS} lies approximately around to 0.9999935. Figure 9, instead, shows the influence of the HSS application part failure time. In this case, the nominal value amounts to 2000 hours, and no side effects are notable unless $1/\lambda_{HSS}$ undergoes the value 1000 (approximately). In this case, the horizontal dashed line at 0.999990 is useful to identify the *breakpoint* across the high availability requirement.

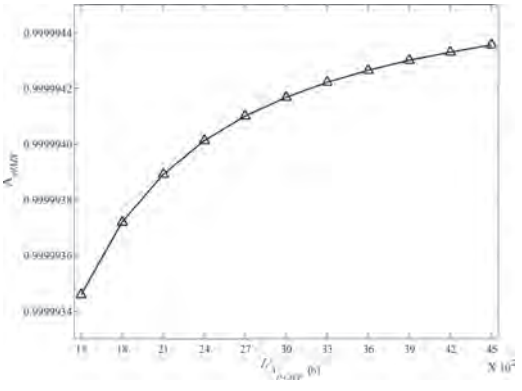


Figure 8. Influence of $1/\lambda_{P-CSCF}$ failure rate over system availability, in case of optimal setting S_1.

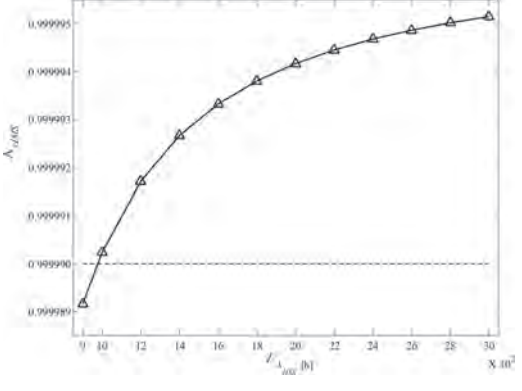


Figure 9. Influence of $1/\lambda_{HSS}$ failure rate over system availability, in case of optimal setting S_1.

6 CONCLUSIONS

Network Function Virtualization and IP Multimedia Subsystem represent two fundamental concepts within the 5G telecommunication world. The former is exploited to virtualize network functions into building blocks, that can be connected or chained in severalways. The latter is designed to provide advanced and IP-based multimedia services. By combining these two paradigms it is possible to derive a virtualized IMS (vIMS) infrastructure that, in this work, has been characterized in terms of availability. Eachvirtual block of vIMS framework has been modeled as a three-layer structure considering: hardware, software (or application) and hypervisor. From a macroscopic viewpoint, the vIMS has been described by exploiting the Reliability Block Diagram (RBD) representation, useful to capture the relationships among the blocks. On the other hand, each virtual block has been modeled by a Stochastic RewardNets (SRN), able to characterize the probabilistic functioning of the underlying system in terms of failure and repair events. Such a modeling phase is preparatory to perform: *i)* a steady-state availability analysis aimed at finding out the minimal-cost redundant configuration that guarantees the "five nines" availability requirement; *ii)* a transient analysis focused on capturing the vIMS dynamics, where a more realistic repair distribution (Weibull) for software layer has been taken into account; *iii)* a sensitivity analysis aimed at evaluating the robustness of vIMS in case of some fluctuation of the critical parameters with respect to nominal values or estimation errors in real data sets. Future work will be devoted to the availability analysis of a more realistic vIMS infrastructure, where some network nodes are co-located, and common mode failures arise.

REFERENCES

3GPP (2001). TS 23.228. IP Multimedia Subsystem (IMS). Ayers, M. (2012). *Telecommunications System Reliability Engineering, theory, and practice*. Ed. John Wiley and Sons.

Bolch, G., S. Greiner, H. De Meer, and K. Trivedi (1998). *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*. New York, NY, USA: Wiley-Interscience.

Camarillo, G. and M. Garcia-Martin (2008). *The 3G IP Multimedia Subsystem* (3rd ed.). John Wiley and Sons.

de S. Matos, R., P.R.M. Maciel, F. Machida, K. Dong-Seong, and K.S. Trivedi (2012). Sensitivity analysis of server virtualized system availability. *IEEE Transactions on Reliability 61*(4), 994–1006.

Di Mauro, M., G. Galatro, M. Longo, F. Postiglione, and M. Tambasco (2017). Availability evaluation of a virtualized IP Multimedia Subsystem for 5G network architectures. In M. Cepin and R. Bris (Eds.), *Safety and Reliability - Theory and Applications*, pp. 2203–2210. Taylor & Francis Group.

Di Mauro, M., M. Longo, F. Postiglione, R. Restaino, and M. Tambasco (2016). Availability evaluation of the virtualized infrastructure manager in network function virtualization environments. In L. Walls, M. Revie, and T. Bedford (Eds.), *Risk, Reliability and Safety: Innovating Theory and Practice*, pp. 2591–2596. Taylor & Francis Group.

Duan, J., C. Wu, F. Le, A. Liu, and Y. Peng (2017). Dynamic scaling of virtualized, distributed service chains: A case study of IMS. *IEEE Journal on Selected Areas in Communications 35*(11), 2501–2511.

ETSI (2012). Network Functions Virtualisation: An introduction, benefits, enablers, challenges and call for action. ETSI (2015). VoLTE service based on vEPC and vIMS Architecture.

Fernandes, S.F.L., E. Tavares, M.A. Santos, V. Lira, and P.R.M. Maciel (2012). Dependability assessment of virtualized networks. In *Proc. IEEE ICC 2012*, pp. 2711–2716.

German, R., C. Kelling, A. Zimmermann, and G. Hommel (1995). TimeNET: a toolkit for evaluating non-Markovian stochastic Petri nets. *Performance Evaluation 24*(1–2), 69–87.

Ghosh, R., F. Longo, F. Frattini, S. Russo, and K.S. Trivedi (2014). Scalable analytics for iaas cloud availability. *IEEE Transactions on Cloud Computing 2*(1), 57–70.

Guida, M., M. Longo, F. Postiglione, K. Trivedi, and X. Yin (2013). Semi-Markov models for performance evaluation of failure-prone IP multimedia subsystem core networks. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 227*(3), 290–301.

Liu, J., Z. Jiang, N. Kato, O. Akashi, and A. Takahara (2016). Reliability evaluation for NFV deployment of future mobile broadband networks. *IEEE Wireless Communications 23*(3), 90–96.

Muppala, J., M. Malhotra, and K. Trivedi (1996). *Markov Dependability Models of Complex Systems: Analysis Techniques*. Springer Berlin Heidelberg.

Muppala, J.K., G. Ciardo, and K.S. Trivedi (1994). Stochastic Reward Nets for reliability prediction. In *Communications in Reliability, Maintainability and Serviceability*, pp. 9–20.

Rosenberg, J.D., H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler (2002). Session Initiation Protocol (SIP). IETF RFC 3261, http://www.ietf.org/rfc/rfc3261.txt.

Sahner, R.A. and K. Trivedi (1987). Reliability modeling using SHARPE. IEEE *Transactions on Reliability 36*(2), 186–193.

Taleb, T., A. Ksentini, and B. Sericola (2016). On service resilience in cloud-native 5G mobile systems. *IEEE Journal on Selected Areas in Communications 34*(3), 483–496.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

AltaRica 3.0 code generation from SysML models

Nga Nguyen
Quartz Laboratory, EISTI, Cergy, France

Faida Mhenni & Jean-Yves Choley
Quartz Laboratory, SUPMECA, Saint-Ouen, France

ABSTRACT: In order to bridge the gap between Model-Based Systems Engineering and Model-Based Safety Assessment, we propose in this paper a language transformation between SysML semi-formal models and the formal language AltaRica 3.0. Meta-data of SysML Block Definition Diagram and Internal Block Diagram that describe system architecture as well as meta-data of SysML State Machine Diagram that represent system behavior (in a limited formalism with respect to AltaRica's Guarded Transitions System) are used to generate AltaRica classes, blocks, events, transitions, etc. Flow port directions and connectors are used to create flow propagation assertions. The object and prototype-oriented paradigm of AltaRica 3.0 with class, composition, inheritance, etc. will be respected since SysML and AltaRica's System Structure Modeling Language share commonalities in structuring constructs. It is obvious that one modeling language cannot be replaced by the other because their goals and domains are different, but the mapping between languages such as SysML and AltaRica allows better understanding and communication between systems engineers and safety experts. Once the preliminary AltaRica 3.0 code is generated with structural and behavioral information, safety experts will complete and validate the code with stochastic models, synchronization, common cause failures and redundancy mechanism to carry out safety assessment based on the expressive power of the language, thanks to its mathematical framework.

1 INTRODUCTION

Model-Based Systems Engineering (MBSE) (INCOSE 2015) is a common approach supporting the design and management of complex systems. Various modeling tools and languages can be used according to the different domains involved in the system, the level of detail, the system aspects to be modeled, etc. SysML (OMG 2015) is a general-purpose modeling language adapted to systems engineering since it allows to express the main concepts inherent to the different aspects of system development. It provides a unified standard for specifying, analyzing, designing, and validating complex systems. The language also allows a multi-viewpoint model and building traceability links.

Model-Based Safety Assessment (MBSA) aims at using high level modeling languages to integrate risk analysis with system architecture. Classical safety modeling formalisms such as Fault Trees, Blocks Reliability Diagrams, Event Trees, Markov Chains and Stochastic Petri Nets can provide efficient assessment algorithms and/or expressive power but suffer from the distance with the architecture of the system. AltaRica (Point 2000) is a formal language that supports MBSA approach. An earlier version of the language, AltaRica Data Flow (ADF) (Arnold et al. 2000) has already been embedded in some commercial integrated modeling environments such as Safety Designer, Cecilia OCAS and Simfia. In ADF, a model is composed of nodes that are characterized by their reachable states, in and out flows, events, transitions and assertions. Once a system model is specified in the AltaRica language, it can be compiled into a lower level formalism such as finite-state machines, fault trees and stochastic Petri Nets and different safety assessments can be performed.

A more recent version of AltaRica, AltaRica 3.0 (Prosvirnova 2014) (Batteux et al. 2015) improves the language with the underlying mathematical framework Guarded Transition Systems (GTS) and new structure constructs via System Structure Modeling Language (S2ML). GTS (Rauzy 2008) is a general states/events formalism which handles looped systems by using fix-point calculation mechanism. Compositions such as free product and synchronization between GTS allow to build hierarchical and modular systems. Meanwhile, S2ML assembles constructs coming from object and prototype-oriented modeling languages such as classes, prototypes, composition, inheritance, etc. A flattening algorithm is needed to collapse the hierarchy of nested blocks and instances of

classes into a single program in order to compile and execute AltaRica 3.0 models.

The first objective of this paper is to study the existing work concerning the automatic translation between SysML and AltaRica in order to integrate the reliability analysis with the design process. Secondly, we will introduce our approach to generate AltaRica 3.0 code from SysML models that takes into account the new features of the language while being consistent with the previous work to not reinvent the wheel. The paper is organized as follows. Section 2 presents some related work on SysML and AltaRica mapping. The new version of AltaRica is introduced briefly in Section 3 with an example taken from AltaRica 3.0 training material. Section 4 describes SysML elements used for AltaRica 3.0 mapping as well as the generated code for the given example. Discussions and future work are given in Section 5.

2 RELATED WORK

The link between SysML and AltaRica has been studied in several research work such as David et al. (2009), Belmonte and Soubiran (2012), Ruin et al. (2012), Yakymets et al. (2013) and Hecht et al. (2015).

David et al. (2009) proposed, as a continuation of their MéDISIS methodology integrating systems engineering and safety analysis, a mapping between SysML models and AltaRica Data Flow (ADF) language, so that existing tools to quantify reliability indicators such as the global failure rate, the mean time to failure, etc., can be used directly on the failure modes identified in the previous step of the methodology. An architecture translation was first carried out by exploiting SysML architectural view through Block Definition Diagram (BDD) and Internal Block Diagram (IBD). A SysML block with its properties is translated into an ADF node with its sub nodes (parts), flow variables (ports) and state variables (values). The synchronization of flow exchange is performed by IBD analysis. Limited variable types and unidirectional flow directions in ADF obliged the authors to propose some adaptations to perform an automatic translation. For the behavioral part, an exhaustive list of SysML elements that depict the comportment (operations, actions, messages, etc) is proposed to guide the creation of events, transitions, guards, assignments or assertions in ADF. However, SysML state-charts diagrams whose mathematical framework is close to the transitions/events mechanism of AltaRica are not considered in this work that can facilitate the automatic translation.

Belmonte and Soubiran in (Belmonte & Soubiran 2012) proposed a translation from Obeo Designer's Domain Specific Modeling Language (DSML) for Preliminary Hazard Analysis (PHA) and Failure Mode and Effects Analysis (FMEA) as well as SysML for system functional models into AltaRica to enable formal verification. A node in an Altarica model can be expressed as an octuple (l, SI, F, S, E, I, T, A) where the components correspond to respectively identifier, subnode instances, flow definitions, state variable definitions, events, initial state, guarded transitions and boolean assertions. Different transformation rules concerning the current operational context, the environment nodes (functional activities or operations), the FMEA nodes (dysfunctional specifications)and the PHA node (top level node) are given. However, no real system has been studied yet in order to prove the scalability of the method.

In the context of Complex Maintenance Program Quantification, Ruin et al. (2012) proposed a framework using SysML to model the static and interaction part of production systems, and AltaRica Data Flow formal language to model the concept behavior part. The model building language and the model execution language are related to each other by a transformation step. The SysML Sequence Diagram representing interactions between blocks and the Parametric Diagram describing relationships between attributes of blocks are used to generate ADF parts made up with nodes, states, events, initial states and transitions. According to their algorithm: a node is created for each lifeline in the sequence diagram; an event corresponds to a reflexive message (looped back message) to the same lifeline; and when a transition takes place, the state variable will change from one "condition mark" to another. Synchronization is made via messages sent from one lifeline to the other and the looped back messages. However, no algorithm is given for the parametric diagram and the case study in the paper is not an industrial scale system. As cited in Ruin et al. (2012), ADF presents some restrictions like the impossibility to model looped systems and acausal connection where the direction of the flow propagation depends on the states.

Yakymets et al. (2013) presented a safety modeling framework for fault tree generation and analysis SMF-FTA. This framework includes meta-models, profiles, model transformation, verification, and Fault Tree Analysis (FTA) tools. In this approach, several steps are needed. First, the system to be analyzed is designed and its structural models are built using the SysML BDD and IBD diagrams. Second, these models are annotated with failure behavior. Then the entire model is converted into AltaRica language by using transfor-

mation rules. A SysML block is translated into an AltaRica node containing flows that correspond to the block's ports. Expressions showing how output port errors can be derived from internal failures of the block and/or possible deviations in the input ports are stored in SysML opaque expressions and used to generate AltaRica internal component transitions. Connection between components via the connectors are used to create assertions. An algorithm already existing in the ARC tool analyzes the AltaRica model and derives the different minimal cut sets from the model. These cut sets are assembled to form the final fault tree. The resulting fault tree can be represented either with open-Probabilistic Safety Assessment (PSA) or a SysML dedicated profile

In Hecht et al. (2015), Hecht et al. generated FMEA from AltaRica code, and this code is created from a series of interacting SysML state machines. However, no detail is given about the mapping elements, except an example of text export from SysML models showing: i) a block and its ports coming from an IBD and ii) states, events and transitions deriving from a state machine diagram of the corresponding block. This SysML output parameter file is transformed into an AltaRica input file where the specific block becomes a node with its own state variable, events and transitions.

No related work concerns yet the latest version of AltaRica with which new theoretical framework and structural constructs have been established. Generating AltaRica 3.0 code directly from SysML models while being consistent with existing related work would be the next step to verify the feasibility of the approach.

3 ALTARICA 3.0 LANGUAGE

AltaRica 3.0 is currently under specification in the framework of the OpenAltaRica project (http://openaltarica.fr). In this paper, we are only focused on the main concepts of the language. Advanced notions such as functions, operators and records are not considered at the moment.

A *block* in AltaRica is used to encapsulate a finite state automaton whose state variables will have type and initial value. A *transition* is made of an *event*, a guard and an action to be performed when the transaction is fired. To illustrate the language, an example taken from Open AltaRica training document is given in Fig. 1. The system has an environment with a true input value and an observer which is the output of the system. The five components *a*, *b*, *c*, *d* and *e* are identical, each one having an input and an output flow connected to each other. Each component can fail and be repaired and the state machine is given in Fig. 2.

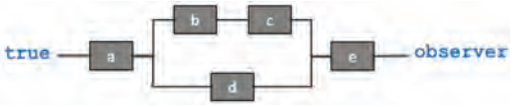


Figure 1. Case study from OpenAltaRica training session.

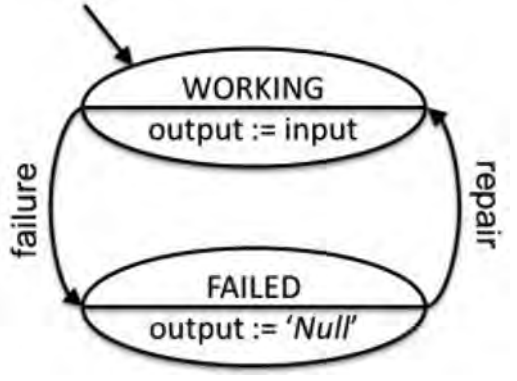


Figure 2. Component finite state machine.

We can define a *class* in AltaRica code that represents a generic component as in Fig. 3. The class Component has a boolean state variable "working" and two flow variables "input" and "output". The two events "failure" and "repair" will change the value of the component state from true to false and inversely. The assertion represents the internal transfer function of the component: if the component is working, the output is equal to the input. Otherwise, it is false.

The case study system can be defined by using the instances of the class Component. Since the model of the whole system is unique, it is denoted by a structural construct *block* that represents a prototype (Fig. 4). An observer is added to model the system output. The assertion part describes the relations between components. Since there are redundant sub-systems, component *e* fails only if there is no output from both *c* and *d* (*e:input:= c:outputord:output*).

For the events, we can add stochastic or deterministic delays as well as parameters used in the corresponding law to define different failure rates for the components as in the updated version in Fig. 5. This allows to associate with each event a delay, a memory policy and a weight in order to have stochastic timed model. To generate fault trees from the AltaRica model, we need to declare the top event with the observer. Also, a common cause failure (CCF) on components *b* and *c* can be defined with its dedicated parameter and used

```
class Component
 Boolean working (init = true);
 Boolean input, output (reset = false);
 event failure,  repair;
 transition
  failure: working -> working := false;
  repair: not working -> working := true;
 assertion
  output := if working then input else false;
end
```

Figure 3. Component code: version 1.

```
block CaseStudy
 Component a, b, c, d, e;
 observer Boolean out = e.output;
 assertion
  a.input := true;
  b.input := a.output;
  c.input := b.output;
  d.input := a.output;
  e.input := c.output or d.output;
end
```

Figure 4. System code: version 1.

```
class Component
 Boolean working (init = true);
 Boolean input, output (reset = false);
 parameter Real pLambda = 1.0e-5;
 parameter Real pMu = 1.0e-2;
 event failure (delay = exponential(pLambda));
 event repair (delay = exponential(pMu));
 transition
  ...
end
```

Figure 5. Component code: version 2.

```
block CaseStudy
 Component a, b ,c , d, e;
 Component d (pLambda = 1.0e-6);
 parameter Real pLambdaCCF = 1.0e-6;
 event eventCCF(delay=exponential(pLambdaCCF));
 observer Boolean topEvent = e.output==false;
 transition
  eventCCF: ?b.failure & ?c.failure;
 assertion
   ...
end
```

Figure 6. System code: version 2.

in the transition to synchronize the components. The prefix ? or ! mean that the event is optional or mandatory, respectively. The updated code for the system is given in Fig. 6. With these information, we can compute the different minimal cut sets, the probability of the top-event for a given mission time, etc. with the OpenAltaRica platform.

AltaRica 3.0 supports also more advanced concepts such as inheritance between classes, acausal connection for bidirectional flows and cold redundancy. The mechanism to update flows uses a fix-point algorithm that allows to detect modeling errors at run time if no fix point is reached.

4 SYSML MODELS AND ALTARICA CODE GENERATION

Table 1 shows our proposal of basic elements mapping from SysML to AltaRica. This model-to-model transformation is based on previous work in Section 2 and takes into account new concepts of the current version of AltaRica.

We have modeled the AltaRica 3.0 case study with SysML. Figures 7, 8 and 9 show respectively the Block Definition Diagram, Internal Block Diagram and State Machine Diagram of the example.

To generate code for the Component class, the different states in the state machine will become an enumeration ComponentState {working, failed}, with a specific initial value. The two flow ports "input" and "output" become automatically typed flow variables in AltaRica. Actually, they have a boolean type, with *reset* made to false by default.

Table 1. SysML and AltaRica 3.0 mapping.

SysML	AltaRica
generic block with several instances	class
specific block with unique occurrence	block
block inheritance	extends
flow port	flow variable
states	a state variable with enumeration domain
initial state	init
transition	event
guard	guard
connector	assertion

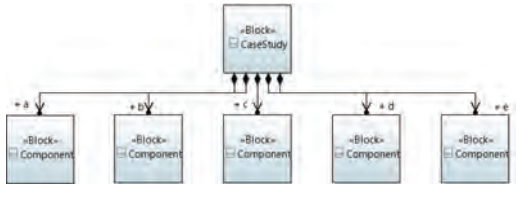


Figure 7. Case study block definition diagram.

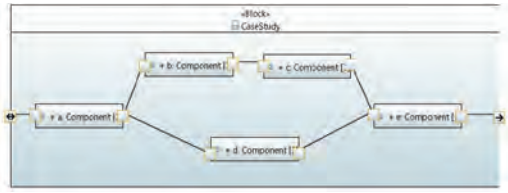


Figure 8. Case study internal block diagram.

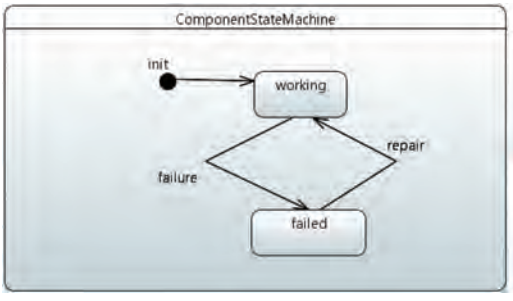


Figure 9. Component state machine diagram.

```
domain ComponentState {working, failed}

class Component
 ComponentState stateVar (init = working);
 Boolean input, output (reset = false);
 event failure,  repair;
 transition
  failure:stateVar==working->stateVar:=failed;
  repair:stateVar==failed->stateVar:=working;
 assertion
  output:=if (stateVar==working) then input
```

Figure 10. Generated component code.

The events and the transitions in AltaRica will be created automatically according to the information in the state machine. Boolean expressions for the guard conditions are added if available. The transfer function of the component that calculates the value of the output flow variable from the value of the state variable and the input flow variable is generated automatically in the *assertion* part. The generated code for the Component class is given in Fig. 10.

The information from IBD is used to generate code for the whole system which is made of 5 components. The interactions between the components are modeled through the connectors and the flow direction (in, out and inout). The input flow of a component depends on the output flow of another component if there is a directed link between them. Since AltaRica 3.0 supports acausal components, i.e. components for which inputs and

```
block CaseStudy
 Component a, b, c, d, e;
 observer Boolean out = e.output;
 assertion
   a.input := true;
   b.input := a.output;
   c.input := b.output;
   d.input := a.output;
   e.input := c.output or d.output;
end
```

Figure 11. Generated system code.

outputs are decided at run time like the inout port of SysML, we can also create an acausal connection in the *assertion* part. For an input port that receives information from different output ports (the case of the component *e* that is linked to two components *c* and *d*), we have to verify if they come from redundant sub-systems or not. We have proposed in another paper (Nguyen, Mhenni, & Choley 2016) the Redundancy Profile, a SysML extension which allows integrating redundancy-relevant properties in the system model in order to better represent system architecture. So these data can be used to generate the assertion *e:input:= c:outputord:output*. For the case when there is no redundancy, the equation should be *e:input:= c:outputandd:output*. An *observer* has been created to observe the final output of the whole system that takes the value of *e:output*. The generated code for the case study is showed in Fig. 11.

5 DISCUSSION AND FUTURE WORK

Currently, our transformation handles just simple concepts of AltaRica 3.0 language with what we can perform a direct mapping. The next step of our work is the completion of the Safety Profile (Mhenni et al. 2016) that contains the Redundancy Profile in order to integrate information concerning stochastic models (delays, parameters, expectations) as well as redundancy mechanism. AltaRica code can be generated with more precise data. The model-to-model transformation will be implemented in a proof of concept tool to validate the approach.

REFERENCES

Arnold, A., A. Griffault, G. Point, & A. Rauzy (2000). The AltaRica language and its semantics. *Fundamenta Informaticae 34*, 109–124.

Batteux, M., T. Prosvirnova, & A. Rauzy (2015). AltaRica 3.0 language specification.

Belmonte, F. & E. Soubiran (2012). A model based approach for safety analysis. In F. Ortmeier and P. Daniel (Eds.), *Computer Safety, Reliability, and Secu-*

rity, Volume 7613 of *Lecture Notes in Computer Science*, pp. 50–63. Springer Berlin Heidelberg.

David, P., V. Idasiak, & F. Kratz (2009, September). Automating the synthesis of AltaRica Data-Flow models from SysML. In M.G.S. Briš (Ed.), *ESREL 2009*, Prague, Czech Republic, pp. 8. Taylor & Francis Group.

Hecht, M., E. Nguyen, A. Chuidian, J. Pinchak, & E. Dimpfl (2015). Creation of Failure Modes and Effects Analyses from SysML. In *SAE International Aertoech*.

INCOSE (2015, August). Systems *Engineering Handbook: A Guide for System Life Cycle Processes and Activities* (John Wiley and Sons ed.). Hoboken, NJ, USA: Internationl Council of Systems Engineering.

Mhenni, F., N. Nguyen, & J.-Y. Choley (2016). SafeSysE: A safety analysis integration in systems engineering approach. *IEEE Systems Journal PP*.

Nguyen, N., F. Mhenni, & J.-Y. Choley (2016). Redundancy handling with model-based systems engineering,. In *26th European Safety and Reliability Conference (ESREL)*.

OMG (2015, September). Systems Modeling Language, version 1.4.

Point, G. (2000). Alta-Rica: *Contribution á l'unification des méthodes formelles et de la sûreté de fonctionnement*. Ph. D. thesis, Université de Bordeaux I.

Prosvirnova, T. (2014). AltaRica 3.0: a Model-Based approach for Safety Analyses.

Rauzy, A. (2008). Guarded transition systems: a new states/events formalism for reliability studies. In *Proceedings of the Institution of Mechanical Engineers Part O Journal of Risk and Reliability*.

Ruin, T., E. Levrat, & B. Iung (2012, November). Modeling framework based on SysML and AltaRica data flow languages for developing models to support complex maintenance program quantification. In *2nd IFAC Workshop on Advanced Maintenance Engineering, Service and Technology, A-Mest'12*, Sevilla, Spain.

Yakymets, N., H. Jaber, & A. Lanusse (2013). Model-based system engineering for fault tree generation and analysis. In S. Hammoudi, L.F. Pires, J. Filipe, and R.C. das Neves (Eds.), *Proceedings of the 1st International Conference on Model-Driven Engineering and Software Development, Barcelona, Spain, 19–21 February*, pp. 210–214.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Failure behavior analysis of hot standby system based on BDD method

Z. Wang, Y. Chen, W. Men & R. Kang
Reliability and System Engineering School, Beihang University, Beijing, China

ABSTRACT: Hot standby technology becomes a popular method for improving the reliability of a crucial component, and a significant system. Considering the failure mechanisms of components, this paper proposes a binary decision diagram-based method for modeling the failure behaviors of hot standby systems. The failure probability of the component what fails firstly among the primary and hot standby component is discussed. We find that under the condition of that one component fails firstly, as time goes by, the transient and cumulative failure probabilities of that the primary and hot standby component fail firstly change. Besides, after one component fails firstly, the reliability of the system sharply decreases and totally depends on the statuses of those remaining functional components. With the help of the method proposed in this paper, more practical work about the maintainability and design optimization can be done.

1 INTRODUCTION

In some safety or mission-critical applications, such as nuclear power plants, aerospace, telecommunications and so on, the redundancy technology is a main method for ensuring the reliability or safety of the systems. The redundancy technology can be divided into 3 types, hot, cold, and warm standby, due to the differences of working conditions of redundancy components (Levitin et al. 2015). In hot standby systems, a hot standby component is exposed to the same operational and environmental stresses as the primary component's. And when the primary one fails, a hot standby component will be switched into the system immediately, and start to output its function and guarantee the mission finished successfully. With the applications of hot standby systems in engineering products, the study about how the hot standby system fails becomes more and more critical.

When the lifetime distributions of all the system's components follow exponential distributions or the failure rates are known, Ren & Zhang (2009) modeled and analyzed the hot standby system's structure by using Markov process method. Though the Markov process method can be applied for the reliability analysis easily, it still has a strict limitation on the lifetime distribution of the components (Xing et al. 2015). Ebrahimipour et al. (2010) obtained the reliability expression of k-out-of-n multi-state series-parallel systems by using the universal generating function. Levitin et al. (2015) analyzed the reliability of 1-out-of-n heterogeneous standby system, and a corresponding mathematical expression is derived. Besides, Levitin also considered two different kinds of failure propagation, selective and global propagation, into the reliability modeling. But when the system's scale is large or its logical relations are complicated, that kind of analyzing method will be extremely difficult for applying.

The modeling or analyzing method for a hot standby system is mainly based on the assumption that the probability density functions (pdfs) or the cumulative distribution functions (cdfs) of components are known (Levitin et al. 2015, Ardakan & Hamadani 2015, Levitin et al. 2014). And most papers pay attentions to the final results of systems' reliability, not how the components' failure develop and affect the entire system's failure. To overcome the difficulty of collecting components' lifetime data and find the nature of failure, physics of failure (PoF) is proposed (Hassan & Aldemir 1990). Because the lifetimes of failure mechanisms simulated by PoF method are constant, Probabilistic Physics of Failure (PPoF) is developed for obtaining the probabilistic distributions of those lifetimes (Zoran & Vlado 2008, Hall & Strutt 2003). Based on those key theories of PoF and PPoF, Chen Y et al. (2015) proposed five correlations among failure mechanisms, which are competition, trigger, acceleration, inhibit, and accumulation. And Chen also provided an effective way, failure mechanism tree (FMT), for analyzing the failure behaviors of components and systems.

In chapter 2, this paper will introduce the failure behaviors of a hot standby system and give a series of expressions for all behaviors. In chapter 3, a hot standby system in the supply module of an

electronic control system will be analyzed as an example for showing the characteristic of the hot standby system.

2 FAILURE BEHAVIOR OF A HOT STANDBY SYSTEM

2.1 *Equations for the reliability of a hot standby system*

Failure behavior is used for describing how the failure mechanisms of components develop with the effect of correlations among themselves, and eventually cause the failure of the entire system. This paper focuses on analyzing failure behaviors of the hot standby system consisting one primary component and one hot standby component. And the switch what is used for switching the hot standby component to be operational after the primary component fails is supposed to be perfect. Assume that the failure probabilities of components are irrelevant with usage or switch intensity.

Because of the same operational conditions that both primary and hot standby components are exposed to, the failure mechanisms of the components develop and influence each other since the system starts to work. The failure sequences of those components are probabilistic and not determined till some components fail, which also means each component has the possibility to fail earlier than the other. So, according to whether the failures of components happen, the working phase of a hot standby system can be divided into two parts, phase I and II.

In phase I, no component fails. The failure mechanisms start to develop since the system begins its working, but none of them reaches its failure threshold. So, the reliability of the entire system is influenced by the developments of failure mechanisms and in-time reliability of both components. Only if both of the components fail, the failure of a hot standby system will occur. The reliability of the system can be described as (1).

$$\begin{aligned} R(t) &= 1 - F_P(t) \bullet F_S(t) \\ &= 1 - \int_0^t f_P(t_P)dt_P \int_0^t f_S(t_S)dt_S \end{aligned} \tag{1}$$

where f_P(t) and f_S(t) are the probability density functions of the failures of primary and hot standby components, respectively.

In phase II, the primary or hot standby component fails at time t_f, but the other one is survived and switched to be operational for ensuring the system finish its function. The entire system will fail if the survived one fails. So, the reliability of the system is equal to the survived one's, which also shown in (2).

$$R(t) = \begin{cases} 1 - F_P(t) & \text{a hot standby component} \\ = 1 - \int_0^t f_P(t_P)\,dt_P & \text{fails at } t_f \\ 1 - F_S(t) & \text{a primary component} \\ = 1 - \int_0^t f_S(t_S)\,dt_S & \text{fails at } t_f \end{cases} \tag{2}$$

From the discussion above, we know that the component failing at t_f is a primary or hot standby component will lead to a different reliability level of the system. So, the study on the distribution function of t_f is a significant job for learning the changes of the system's reliability and adjusting maintenance strategy. t_f equals the minimum failure time among a primary and hot standby component's, which is shown as (3).

$$t_f = \begin{cases} t_P & t_P \le t_S \\ t_S & t_P > t_S \end{cases} \tag{3}$$

where t_P and t_S are the failure times of the primary and hot standby component, respectively.

When the primary component fails firstly, which means $t_P \le t_S$, t_f equals the failure time of primary component. And at time t, the transient probability of that the primary component fails, but the hot standby component is survived can be obtained as (4). Besides, the cumulative probability of that the primary component fails firstly is (5).

$$\begin{aligned} P_P(t) &= P\left(t_f = t_P = t, t_S \ge t\right) \\ &= f_P(t)dt \cdot \int_t^\infty f_S(t_S)dt_S \end{aligned} \tag{4}$$

$$\begin{aligned} F_P(t) &= \int_0^t P_P(t_{FP})dt_{FP} \\ &= \int_0^t f_P(t_{FP})dt_{FP} \cdot \int_{t_{FP}}^\infty f_S(t_S)dt_S \end{aligned} \tag{5}$$

When the hot standby component fails firstly, which means $t_P > t_S$, t_f equals the failure time of the hot standby component. So at time t, the transient probability of that the hot standby component fails, but the primary is survived can be derived as (6). And the cumulative probability of that the hot standby component fails firstly is (7).

$$\begin{aligned} P_S(t) &= P\left(t_f = t_S = t, t_P > t\right) \\ &= f_S(t)dt \cdot \int_t^\infty f_P(t_P)dt_P \end{aligned} \tag{6}$$

$$\begin{aligned} F_S(t) &= \int_0^t P_S(t_{FS})dt_{FS} \\ &= \int_0^t f_S(t_{FS})dt_{FS} \cdot \int_{t_{FS}}^\infty f_P(t_P)dt_P \end{aligned} \tag{7}$$

For further discussion, if there is a component fails firstly at time t, the probability of that the failed component is the primary component is $P_{P|F}(t)$, and the probability of that the failed component is the hot standby component is $P_{S|F}(t)$, which can be calculated as (8) and (9).

$$P_{P|F}(t) = P_P(t) / \left[P_P(t) + P_S(t) \right] = \frac{f_P(t)\,dt \cdot \int_t^{\infty} f_S(t_S)\,dt_S}{f_P(t)\,dt_P \cdot \int_t^{\infty} f_S(t_S)\,dt_S + f_S(t)\,dt \cdot \int_t^{\infty} f_P(t_P)\,dt_P} \tag{8}$$

$$P_{S|F}(t) = P_S(t) / \left[P_P(t) + P_S(t) \right] = \frac{f_S(t)\,dt \cdot \int_t^{\infty} f_P(t_P)\,dt_P}{f_P(t)\,dt_P \cdot \int_t^{\infty} f_S(t_S)\,dt_S + f_S(t)\,dt \cdot \int_t^{\infty} f_P(t_P)\,dt_P} \tag{9}$$

When one component fails, the reliability of the system will decreases and equals to the survived one's reliability. Whether the system's reliability after t_f reaches the desired level or not will influence the optimal choice of actions for preventing and reducing the losses caused by the system's failure. So, the cumulative probability of t_f can be obtained as (10).

$$F_f(t) = \int_0^t f_P(t_P)\,dt_P \int_{t_P}^{\infty} f_S(t_S)\,dt_S + \int_0^t f_S(t_S)\,dt_S \int_{t_S}^{\infty} f_P(t_P)\,dt_P \tag{10}$$

2.2 *Binary Decision Diagram (BDD) for simulation*

The binary decision diagram (BDD) is proposed on the Shannon decomposition theorem (Xing L. et al. 2012). As a matter of fact, the BDD model has been widely applied for simplifying logical analysis on large static fault trees. For a Boolean formulation, f is defined upon a set of Boolean variables: $x_1, x_2, \ldots, x_n$, and it can be equivalently decomposed by the Shannon expansion rule as (11).

$$f = x_i \cdot f_{x_i=1} + \overline{x_i} \cdot f_{x_i=0} = ite(x_i, f_{x_i=1}, f_{x_i=0}) \quad (1 \le i \le n) \tag{11}$$

where $f_{xi=b}$ represents f when $x_i = b$, and b is a Boolean constant 0 or 1; *ite* denotes the compact if-then-else format. Equation (11) also can be illustrated as Fig. 1.

To build a BDD model of a fault tree, an efficient recursive generating algorithm through a depth first left traversal of the fault tree is presented as (12).

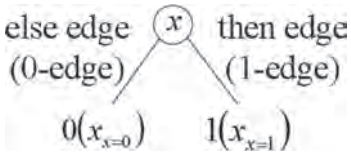


Figure 1. Shannon decomposition and *ite* format of x.

$$G \lozenge H = ite(x, G_1, G_0) \lozenge ite(y, H_1, H_0) = \begin{cases} ite(x, G_1 \lozenge H_1, G_0 \lozenge H_0) & index(x) = index(y) \\ ite(x, G_1 \lozenge H, G_0 \lozenge H) & index(x) < index(y) \\ ite(x, G \lozenge H_1, G \lozenge H_0) & index(x) > index(y) \end{cases} \tag{12}$$

where G and H are two Boolean formulations representing structure functions of the traversed subtrees, and G_i and H_i (i = 1,0) are the subfunctions of G and H. $\lozenge$ indicates logic operation of AND or OR, and *index*(x) and *index*(y) denote an argument index of the variable x and y.

This paper will model the failure behavior of the hot standby system by adjusted BDD models. The basic and important issue is dealing with the correlations among failure mechanisms. According the method of modeling a FMT model (Chen et al. 2015), the FMT models of all components in hot standby systems can be generated, which can be transformed into BDD models with the approaches listed in Table 1.

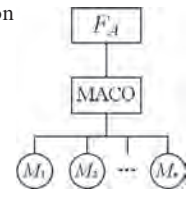
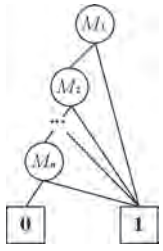
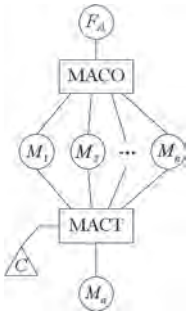
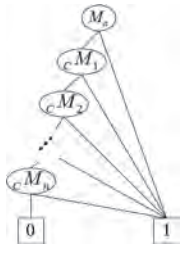
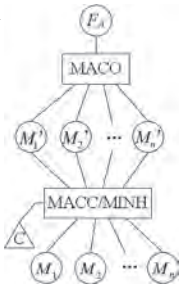
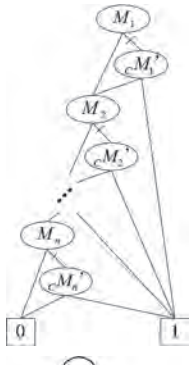
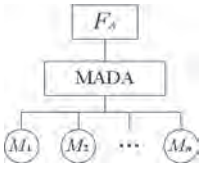
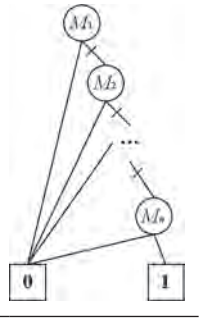
In hot standby systems, only if both the primary and hot standby components fail, the entire systems will fail, which means the logic relations between the primary and hot standby ones can be described as the logic AND. On the other hand, the logic OR may exists among different components. So, the fault tree (FT) model of the different components can be generated, and the exact methods about how to build the FT and transform it into a BDD model are shown in Table 2 (Levitin et al. 2015).

In Table 2, A_i (i = 1, 2,...,n) represents the components of a system.

To analyze the BDD model, a Monte-Carol based method is used in this paper. Firstly, a large number of random lifetimes of failure mechanisms should be generated. Secondly, the logical expressions about the failure times of components and the entire systems must be derived. The exact method is listed in Table 3. Finally, lots of the random failure time will be computed and some corresponding reliability curves can be generated.

In Table 3, T_a is the lifetime of the failure mechanism M_a, T_i (i = 1,2,…,n) is the lifetime of the failure mechanism M_i (i = 1,2,…,n), T_i' (i = 1,2,…,n)

Table 1. FMT and BDD models of mechanism correlations.

Mechanism correlation	FMT model	Corresponding BDD model
Competition		
Trigger		
Acceleration or inhibit		
Accumulation		

is the lifetime of the failure mechanism M_i (i = 1,2,…,n) after being accelerated or inhibited, T_{Ai} (i = 1,2,…,n) is the lifetime of the component A_i (i = 1,2,…,n), t_C is the time when event C happens.

Table 2. FT and BDD models of logic AND and OR.

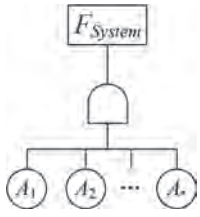
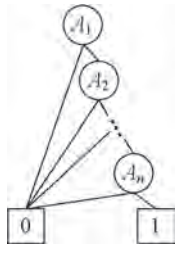
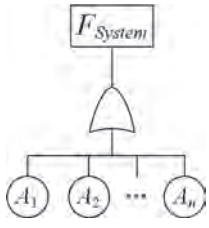
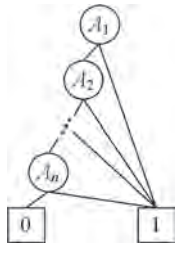
FT model	Corresponding BDD model
	
	

Table 3. Logical expressions for the mechanism correlations, logic AND and logic OR.

Mechanism/logic correlations	Logical expressions
Competition	$t_A = \min\{T_1, T_2, \cdots, T_n\}$
Trigger	$t_A = \min\{T_a, t_C + T_1, t_C + T_2, \cdots, t_C + T_n\}$
Acceleration or inhibit	$t_A = t_C + \min\{T_1' - \frac{T_1'}{T_1} t_C, T_2' - \frac{T_2'}{T_2} t_C, \cdots, T_n' - \frac{T_n'}{T_n} t_C\}$
Accumulation	$t_A = \frac{1}{\frac{1}{T_1} + \frac{1}{T_2} + \cdots + \frac{1}{T_n}}$
Logic AND	$t = \max\{T_{A1}, T_{A2}, \cdots, T_{An}\}$
Logic OR	$t = \min\{T_{A1}, T_{A2}, L, T_{An}\}$

3 A CASE STUDY

3.1 *Description of the case*

This section takes a supply system in an electronic control system of an aircraft as an example for discussing the failure behavior of a hot standby system. The supply system contains two identical supply modules consisting 6 components, which is

simplified and shown in Fig. 2. The exact meanings of all components in Fig. 2 are shown in Table 4.

The components in primary and hot standby supply systems are the same component, but different in their lifetime distribution's parameters. According to the FMMEA result of the example supply system, main failure mechanisms, correlations among those mechanisms, and the effect for every component are shown in Table 5.

According historical data or the PPoF method, the distributions and parameters of those failure mechanisms' lifetimes in different components are shown in Table 5.

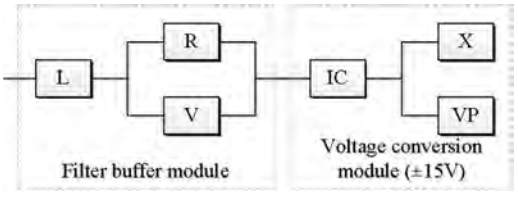


Figure 2. The supply module in the supply system.

Table 4. Components in the supply module.

Component symbol	Definition and details
L	Inductance
R	Resistance
V	Thyristor
IC	DC-DC supply converter (PWM switching mode power supply)
VP	Optocoupler
X	Perfect socket

In Table 5, TF is thermal fatigue, EC is electrolytic corrosion, MM is metal migration, HCI is hot carrier injection, VF is vibration fatigue, EM is electrical migration, TDDB is time-dependent dielectric breakdown, and EB is electrical breakdown.

3.2 *Failure behavior modeling*

With the method of generating a FMT model (Chen Y. et al. 2015), the FMT model of the supply system can be generated as shown in Fig. 3.

After transforming the FMT into BDD model, the failure behavior model is obtained as shown in Fig. 4.

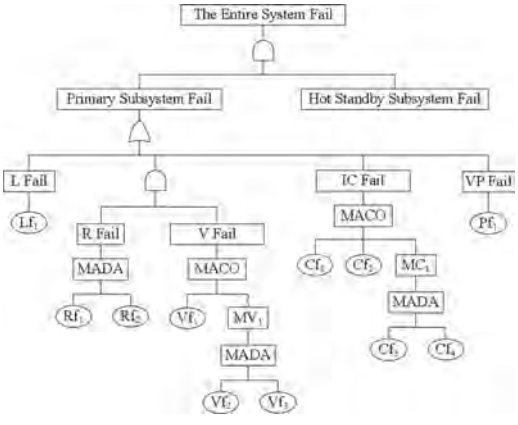


Figure 3. The FMT model of the entire system.

Table 5. Failure mechanisms and correlations in components, and distribution types and parameters of mechanisms.

Component	Mechanism	Mechanism symbol	Failure effect	Failure mechanism correlations	Effect symbol	Failure mechanism correlations	Distribution type	Distribution parameter: Primary subsystem β(θ)	Primary subsystem η(σ)	Hot standby subsystem β(θ)	Hot standby subsystem η(σ)
L	TF	Lf_1	Open	/	/	/	Weibull	2.1	6225	1.7	3172
R	EC	Rf_1	Resistance increase	Parameter union	MR1	/	Lognormal	8.38	0.7	9.71	1.3
	MM	Rf_2	Resistance increase				Weibull	1.3	4715	2.4	7863
V	HCI	Vf_1	Short	/	/	Competition	Weibull	1.73	6429	7.3	4217
	VF	Vf_2	Open	Damage accumulation	MV1		Weibull	6.4	8523	5.2	9726
	TF	Vf_3	Open				Weibull	6.7	9318	4.9	8179
IC	EM	Cf_1	Open	/	/	Competition	Weibull	3.1	5621	2.72	6923
	TDDB	Cf_2	Short	/	/		Lognormal	11.37	1.5	12.79	1.4
	VF	Cf_3	Open	Damage accumulation	MC1		Weibull	5.7	8374	2.91	6741
	TF	Cf_4	Open				Weibull	4.9	7956	3.57	5357
VP	EB	Pf_1	Short	/	/	/	Weibull	5.6	4682	3.4	7423

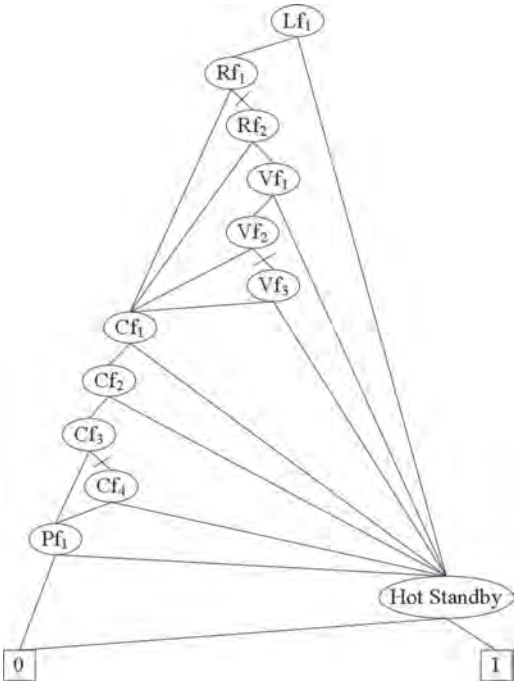


Figure 4. The BDD model of the failure behavior of the example supply system.

3.3 *Simulation*

Using the Monte-Carlo method, 2,000,000 random lifetimes of all mechanisms are generated. From the BDD model of the example system in Fig. 4, the logical expressions of the primary, hot standby subsystems and the entire system are shown in (13)-(15).

$$T_{\text{primary}} = T_{\text{standby}} = \min\{\max\{T_{\text{MR1}}, \min\{T_{\text{Vf1}}, T_{\text{MV1}}\}\}, T_{\text{Lf1}}, T_{\text{Cf1}}, T_{\text{Cf2}}, T_{\text{MC1}}, T_{\text{Pf1}}\} \tag{13}$$

$$T_{\text{fail-firstly}} = \min\{T_{\text{primary}}, T_{\text{standby}}\} \tag{14}$$

$$T_{\text{system-failing}} = \max\{T_{\text{primary}}, T_{\text{standby}}\} \tag{15}$$

Calculating the random lifetimes by those logical expressions, the exact failure times of the system are obtained. The reliability curve of primary and hot standby subsystem, and the entire system are shown in Fig. 5.

In Fig. 5, the reliability of primary subsystem is bigger than the hot standby one's at the same time, and both of them are smaller than the supply system's, which means the hot standby technology used in this system works and the hot standby subsystem may fail firstly.

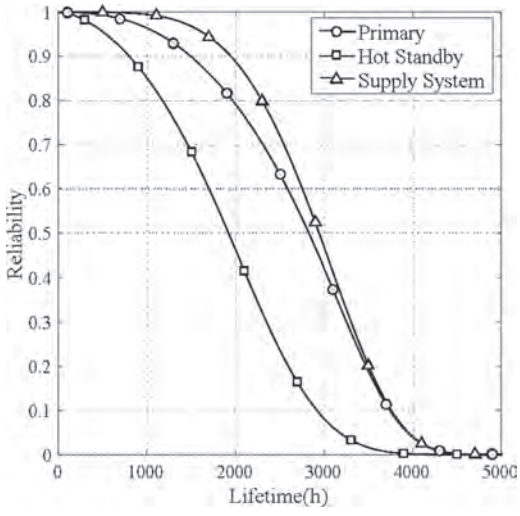


Figure 5. The reliability curves of the primary and hot standby subsystems and the entire system.

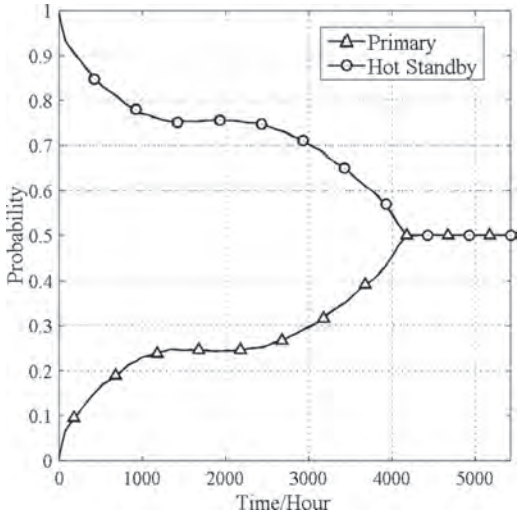


Figure 6. The transient failure probability curves of that the primary and hot standby system fail firstly.

Besides, if there is a subsystem fails firstly, the transient failure probability curves of that the primary and hot standby system fail firstly are shown in Fig. 6. And the cumulative failure probability curves of that the primary and hot standby system fail firstly are shown in Fig. 7.

In Fig. 6, we can learn that during the early period of the system's lifetime ($0 < t < 1000$ hours), $P_{S|F}(t)$ is nearly 75% ~ 100%, which decreases

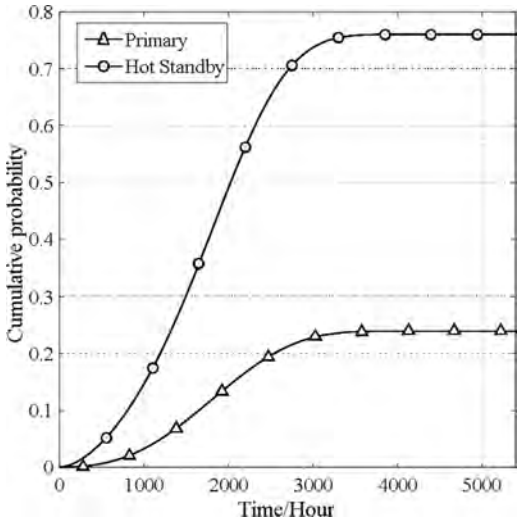


Figure 7. The cumulative failure probability curves of that the primary and hot standby system fail firstly.

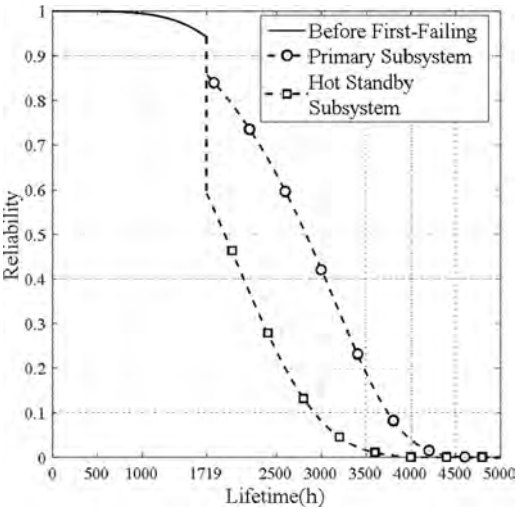


Figure 9. The reliability of the supply system after one subsystem's failing at t = 1719 hour.

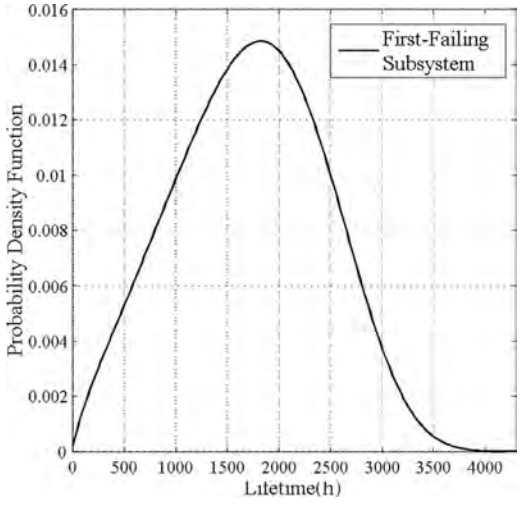


Figure 8. The pdf probability curve of the first-failing subsystem.

as time goes by. And if there is a supply module fails firstly, the probability of that the failed module might be the hot standby module. During 1000 < t < 2500 hours, $P_{S|F}(t)$ is nearly stable at 75%, and decrease after 2500 hours. During the last period of the system's lifetime (t > 4000 hours), $P_{S|F}(t) = P_{P|F}(t) = 50\%$.

In Fig. 7, we learn that in summary, the hot standby supply system is more likely to fail firstly at the probability of nearly 77%. Moreover, the pdf curve of that one supply module fails firstly is shown in Fig. 8.

From Fig. 8, the probability of one of the subsystems fails at about t = 2000 hour is biggest. So at that moment, the reliability of the supply system should be paid enough attentions, and the frequency of checking the working conditions of the two subsystems should be greater. Besides, the mean time of the failure time of first-failing subsystem is about 1719 hours. Suppose that there is a subsystem fails at t = 1719 hour, the reliability of the entire system will have totally different results as shown in Fig. 9.

If primary subsystem fails at t = 1719 hour, the reliability will equal to the hot standby subsystem's; if hot subsystem fails at t = 1719 hour, the reliability will equal to the primary one's, which has a more dramatic decline and may not reach the requirement of the reliability and the hot standby subsystem should be replaced.

4 CONCLUSION AND FUTURE WORK

This paper discussed the failure behavior of a hot standby system and proposed a BDD-based method for evaluating the hot standby system's failure behavior. We derived some functions for describing the transient and cumulative failure probabilities of that the primary or hot standby component failed firstly, which may be helpful for the optimizations of the maintenance allocation and functional design. The sharply decrease of the

reliability of a hot standby system after some components fail firstly are analyzed.

Next, we will optimize this kind of analyzing method, and try to apply it to the failure behavior analysis of more complex systems, like 1-out-of-n hot standby system.

ACKNOWLEDGMENTS

This work was funded by the National Natural Science Foundation of China under contract number 61503014 and 61573043.

REFERENCES

Ardakan M.A. & Hamadani A.Z. 2014. Reliability Optimization of Series-Parallel Systems with Mixed Redundancy Strategy in Subsystems. *Reliability Engineering and System Safety* 130: 132–139.

Chen Y., Yang L., Ye C., Kang R. 2015. Failure mechanism dependence and reliability evaluation of non-repairable system. *Reliability Engineering and System Safety* 138: 273–283.

Ebrahimipour V., Sheikhalishahi M., Shoja B.M., & Goldansaz M. 2010. A Universal Generating Function Approach for Redundancy Optimization for Hot-Standby Multi-State Series-Parallel k-out-of-n Systems. *2010 Fourth UKSim European Symposium on Computer Modeling and Simulation*: 235–239.

Hall P.L. & Strutt J.E. 2003. Probabilistic Physics-of-Failure Models for Component Reliabilities Using Monte Carlo Simulation and Weibull Analysis: A Parametric Study. *Reliability Engineering & System Safety* 80(3): 233–242.

Hassan M. & Aldemir T. 1990. A data base oriented dynamic methodology for failure analysis of closed loop control systems in Process Plants. *Reliability Engineering* 27: 275–322.

Levitin G. Xing L., Ben-Haim H., & Dai Y. 2015. Effect of Failure Propagation on Cold vs. Hot Standby Tradeoff in Heterogeneous 1-Out-of-N:G Systems. *IEEE Transactions on Reliability* 64: 410–419.

Levitin G., Xing L., & Dai Y. 2014. Cold vs. Hot Standby Mission Operation Cost Minimization for 1-Out-of-N Systems. *European Journal of Operational Research* 234: 155–162.

Levitin G., Xing L., & Dai Y. 2015. Reliability and Mission Cost of 1-Out-of-N:G Systems with State-Dependent Standby Mode Transfers. *IEEE Transactions on Reliability* 64: 454–464.

Ren S., & Zhang C. 2009. Study on the Reliability of Hot Standby Repairable Supply System Based on Markov Model. *6th International Conference on Service Systems and Service Management*: 318–322.

Xing L., Shrestha A., & Dai Y. 2011. Exact Combinatorial Reliability Analysis of Dynamic Systems with Sequence-Dependent Failures. *Reliability Engineering and System Safety* 96: 1375–1385.

Xing L., Tannous O., & Dugan J.B. 2012. Reliability Analysis of Nonrepairable Cold-Standby Systems Using Sequential Binary Decision Diagrams. *IEEE Transactions on System, Man, and Cybernetics—Part A: Systems and Humans* 42(3): 715–726.

Zoran M.é. & Vlado S. 2008. The Physics-of-Failure Approach in Reliability Engineering. *30th International Conference on Information Technology Interfaces*: 745–750.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Dependability analysis of a product line using its model

B. Chieb, V. Idasiak & F. Kratz
Laboratoire PRISME, INSA Centre Val de Loire, Bourges, France

ABSTRACT: The objective of this study is to introduce safety studies as soon as the engineering information is available. Safety studies require the use of formalisms. SysML begins to be a good vector of System Engineering activities, and Feature Model seems like an excellent candidate for the product line description. In order to perform a safety analysis, the required information is extracted from the Feature Model/SysML models of the product line. To reduce the number of studies, we provide FMECA (Failure Modes, Effects and Criticality Analysis) of product line (parametric FMECA) type analysis support, which allows conducting an analysis at the level of the product line, and provides rapid analysis synthesis for each product. In this study, we introduce a new process dedicated to product line in the MéDISIS method. We design a meta-model of System Engineering to help the information management related to the product line variability from the functional and organic point of view. We define the parametric FMECA which carries all relevant information from the models. It allows the decisions and choices capitalization during the safety analysis, especially the impact of variability of the product in terms of dependability. The new MéDISIS process automatically generates from both models a parametric FMECA of the product line. Finally, the process is finalized by the Dependability Engineer using the consolidation tool from MéDISIS in order to generate the final FMECA. The synthesis method of a parametric FMECA from models will be presented. In particular, we will discuss how a variability by its presence or absence can influence a dependability analysis, and how the rules used to define variabilities are taken into account for the final FMECA synthesis.

1 INTRODUCTION

In the current commercial and industrial context, the industrial strategies are often based on the definition of line of products. The objective is the decrease of the costs and of the delays while favoring a better offer adaptation to the expectations of the market, and an increase of the products or systems quality. Such an industrial strategy impacts naturally on organizations, in particular on the processes and on the methods of System Engineering (SE). Indeed, one of SE qualities is to allow the capitalization of the convergent validated solutions of the product creation. Thus, the SE so offers a natural support to the approaches by product line characterizing a set of products having common elements and variabilities.

A product line defines a set of products having common characteristics and architecture related to components and functions, with which are associated by variabilities, which is necessary for the satisfaction of a range of needs.

The capitalization of common parts knowledge is then one of engines of studies rationalization, while artifacts stemming from the management of the variabilities and from their coherences constitute a brake.

The works made during the definition of the method MéDISIS (Method of Integration safety analyses in the process of System Engineering) and its platform (David et al. 2010) and (Cressent et al. 2013), established the bases of an exploitation of the Model Based Systems Engineering (MBSE) to generate partially models of operating safety analysis. More recently, (Kajdan & Idasiak 2015) allowed by the definition of minimal set information of MBSE, the rationalization of the techniques of generation of pre-FMECA. Furthermore, this rationalization gave the possibility of following the evolution in a unique one FMECA of the modifications carried out on a system model. The variant part of SE models of product line escapes this type of generation process. Thus, our objective consists in adapting the methodology to products designed in the form of a product line, and in allowing so to build generic studies of the operating safety.

A product line, for the SE analysis and design steps, is to identify the common elements of all products in the range in order to make significant gains through the capitalization and reuse of artifacts produced by the SE processes implemented: cost reduction, development time reduction and, testing, and in our case, dependability analyzes. But every product in the product line has to adapt

to a wide variety of needs, which are then differentiated by specificities. The set of specificities characterizes the product line, and constitutes the variabilities. An example of use cases in section 4 will support our explanations.

If the common part requires a few changes in practices in modeling techniques and methods, it is not the same for variability. Product line models provide a solution to clearly identify the set of common elements of all variant elements and their dependencies. In fact, the variation models also introduce a logic allowing the configuration expression of each product that only uses a subset of variants.

In this article, we will propose two approaches to dependability analysis based on product line models. We will extend our reflection on the generation of pre-FMECA process of the MéDISIS method in order to provide for each product line model a FMECA declinable for each product. Section 2 presents the engineering system concepts used for product line model. Section 3 presents the standard processes of MéDISIS and the proposed processes. Section 4 presents a case study conducted to evaluate the tow processes. Section 5 presents the conclusion.

2 THE PRODUCT LINE AND ENGENEERING SYSTEM MODEL

The concepts of the product line modeling are known in the computer under the name of SPL (Software Product Line) (Lee et al. 2009), they introduce Feature Model. The Feature Model (FM) modeling technique is based on logical concepts between the model's features. It is used to define similarity and variation in models and provide support for consistent variability management.

Some authors have defined their own variability management (Berrebi 2013) in the field of aeronautics. In particular, at the level of the architectural organization of the model. The variant management proposed is an alternative to the variation point that has less logic operator than their equivalent with FM. Moreover, (Sierla et al. 2014) shows the applicability of FM in areas other than software, including the HiP-HOPS approach (based on the dysfunctional model) for product line safety analysis (De Oliveira et al. 2014). This variant management is explained in the approach proposed by (De Oliveira et al. 2014), which uses the Hephaestus/Simulink tool (Steiner et al. 2013) to manage variability in a product line model.

(Le put 2016) presents a proposal for integrating the definition steps and the conceptualization of the product line in the field of the ES. Here the product line is presented as a generic product reusable and scalable according to the need. This method definition is based on the properties present in the FM.

Some ES tools and MéDISIS use a modeling paradigm like SysML (Friedenthal 2014) to represent the model of a product or system. Among the works that make it possible to use the SysML language to model a product line, we find in particular (Grönniger et al). This work presents an approach to modeling an automobile system using FM and an internal organic modeling defined by the internal block diagram IBD (SysML). The authors of (Grönniger et al. 2008) propose a translation of a product line model using the FM technique, in SysML language, which makes it possible to model the variables of automobile systems in internal block model in order to be able to analyze the generic system and its variability. In (Höfig et al. 2014), the authors propose a structuring of data by instantiation with a meta-model to obtain a reusable FMECA. However, they do not define the constraints between the variant elements. Through this analysis, we are interested in using the FM and the SysML modeling language to represent a product line model in different technology areas. We must now define the nature of the information needed to build a FMECA.

The authors of (Kajdan & Idasiak 2015) introduce a system model analysis support, defining sets of artifacts (needs, functions, components, and requirements) and dependency relationships between them. This hierarchical view, necessary for a functional or organic analysis, must be adapted to take into account the artifacts introduced by FM. Similarly (David et al. 2010) defines the meta-model supporting the generation of pre-FMECA, it will be adapted to build the parametric FMECA associated with a product line.

(Mazo 2014) shows the possibilities offered by the FM syntax. Indeed, FM can be likened to the superimposition of different product models for which are highlighted the common artifacts and variants, as well as the relationships between the latter which are at the point of variation according to (Le put 2016). These relationships are constraints on the configurations that a product can have.

2.1 *Model concepts*

Figure 1 shows the different concepts needed in system model for realizing dependability analysis. These concepts are identified as follows: the need that is represented in the system need view, the function realized by system represented in functional view, the component that makes up the system and that is represented in component view and the requirement system which can be applied

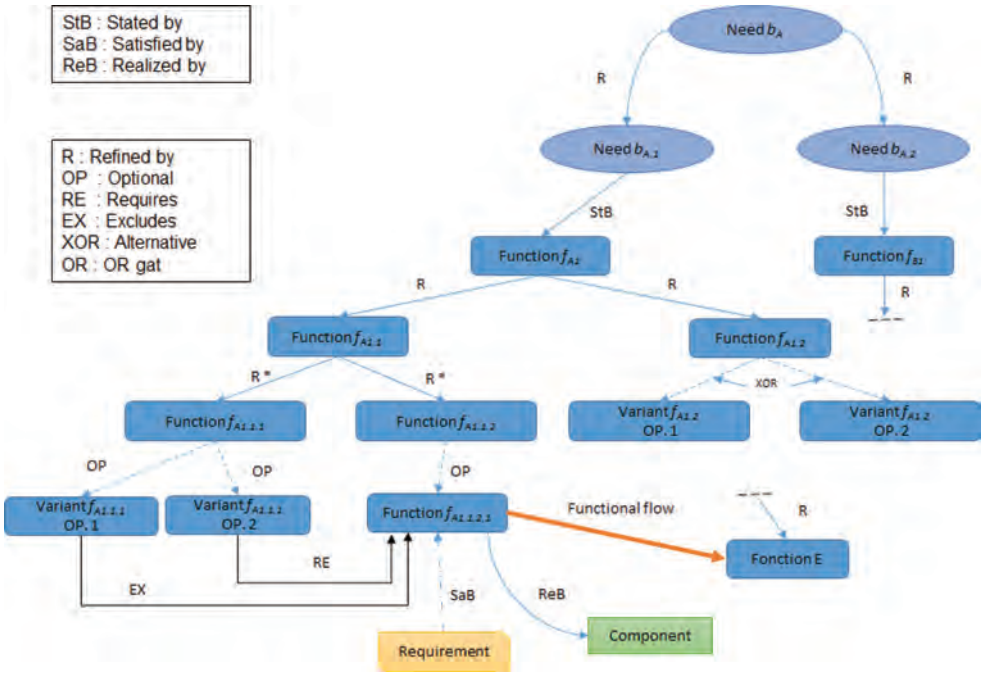


Figure 1. SE mathematical model for product line.

to it (requirement view). They are identified and matched when a model is being established thanks to the following relationships: Stated by (*Need **StB** Function*), Satisfied by (*Function **SaB** Requirement*), Realized by (*Component **ReB** Function*), Refined by (*Concept(x) **R** Concept(x)*). In addition to previous relationships, we introduced or modify six relationships to define the product line SE model.

Either the functions used in the model: $f_i, f_j \in F$:

- Refined by (R or MA): $f_i \xrightarrow{R} f_j$ (noted also f_i R f_j): means that f_i is necessarily refined by f_j.
- Optional (OP): $f_i \xrightarrow{OP} f_j$ (noted also f_i OP f_j): means that f_i is maybe refined by f_j.
- Requires (RE): $f_i \Rightarrow f_j$ (noted also f_i RE f_j): means that the use of f_i implies the use of f_j. Note: f_i do note refine f_j and f_j do note refine f_i.
- Excludes (EX): $f_i \Rightarrow \overline{f_j}$ (noted also f_i EX f_j): means that the use of f_i excluded the use of f_j. Note: $f_i \Rightarrow \overline{f_j}$ is equivalent to $f_j \Rightarrow \overline{f_i}$, or $\overline{(f_i \wedge f_j)}$.
- Alternative (XOR): $f_i \xrightarrow{XOR} f_j$ (noted also f_i XOR f_j): means that only one of f_i and f_j maybe are used. Note: $f_i \Rightarrow \overline{f_j}$ and $f_j \Rightarrow \overline{f_i}$
- OR gate (OR): $f_i \xrightarrow{OR} f_j$ (noted also f_i OR f_j): means that among f_i and f_j minimum one is used.

After identifying the SE relationships needed between the concepts that make up the product line model to conduct a dependability analysis on that model. These new relations will allow describing a new meta-model of the database which helps to generate a PL's FMECA and to formalize its semantics.

This meta-model Figure 2 builds on that presented by (Cressent et al. 2013) and coordinates three categories of information useful for a product line FMECA construction, the information elements derived from the activities of: product SE, product line SE and Safety Analysis. The failure mode part makes it possible to capitalize the decisions and choices of the dependability analysis associated with the studied elements.

2.2 *Product line composition element*

System variants modeling is a basic technique for model based systems engineering for the product line. We need to model variants to: analyze design alternatives; evaluate variants via configurations; product family modeling. So the challenge is to separate the variant from the common part and manage the dependencies.

The product line elements are decomposed into three parts as shown in Figure 3:

– *Core part*: contains the invariant elements of the system and contains all elements that are used in all system configurations. These elements do not depend on any variant element;

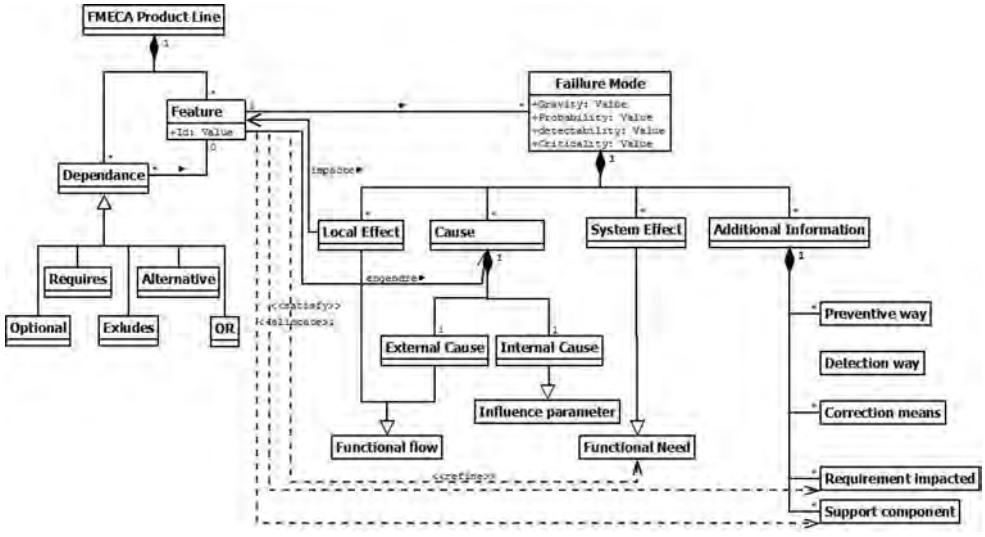


Figure 2. FMECA meta-model of product line.

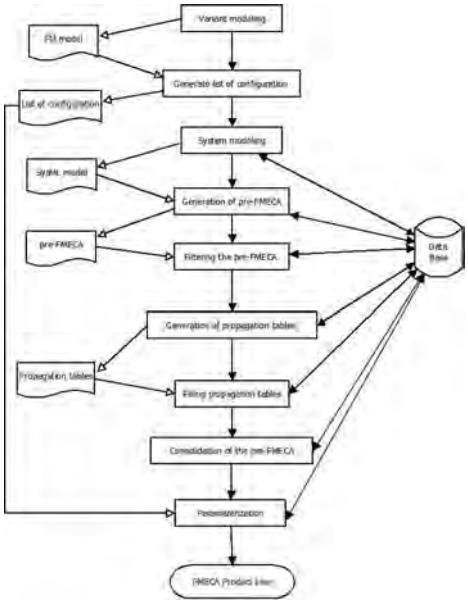


Figure 3. Flowchart describing the proposed process in MéDISIS methodology.

– *Variant-Core Part*: contains the invariant elements of the system which are related at least on variant element;
– *Variants Part*: contains the variant elements of the system, only occurs in some configurations and is part of the system.

This architecture makes it possible to set up a safety analysis strategy on the product line model, with the prioritization of data-flow present in the architecture.

2.3 *Product line safety analysis*

From the dependability standpoint, the system has three types of elements which represents the triplet of a line in the FMECA:

– E_s: Studied element;
– E_c: failure Cause element;
– E_e: failure Effect element.

The variation in product line model may change the analysis for each product configuration because it changes the functional or organic structure of a product. For study the impact of variations in a product, it is necessary to classify the FMECA's lines according to the type triplet studied. Indeed, some of this line can be done earliest in the design process, that is linked to the variability can't be done only when the variability choose is done.

The product line FMECA contains 8 types of line. Each line type in FMECA of product line has a specific characteristic represented in Table 1. The type lines T1 and T8 are composed of the same type of elements set. Such line is analyzed with the same method, which will allow a saving of time in the safety analysis using the standard methodology MéDISIS.

The variation in the product line is presented in line type T2, T3, T4, T5, T6 and T7. This line may include in his triplet (E_c,E_s,E_e) a variant element and core element. T1 and T8 FMECA lines

Table 1. Product line FMECA composition lines.

E_s	E_c	E_e	Line type	Set type
Core	Core or IC	Core	T1	{*Core*}
Core	Core or IC	Variant	T2	{*Core*} ∪ {*Variant*}
Core	Variant	Core	T3	{*Core*} ∪ {*Variant*}
Core	Variant	Variant	T4	{*Core*} ∪ {*Variant*}
Variant	Core	Core	T5	{*Core*} ∪ {*Variant*}
Variant	Core	Variant	T6	{*Core*} ∪ {*Variant*}
Variant	Variant or IC	Core	T7	{*Core*} ∪ {*Variant*}
Variant	Variant or IC	Variant	T8	{*Variant*}

Note. *IC*: Internal Cause (component-specific).

can be analyzed as soon as the core elements or variant elements are designed, even before the end of the total definition of the product line. Whereas, the lines T2 to T7 cannot be analyzed only when the association of the elements variant with the core elements.

3 PRODUCT LINE DEPENDABILITY ANALYSIS PROCESS

This section focuses on the comparison of two dependability analysis processes based on a product line model. For that, we must deploy a sequence of information processing available allowing a progressive contribution of the information held by the system engineer and the dependability engineer, this is what is called a MéDISIS process. These processes are done to make possible the following principles:

– As soon as information is available during an ES step, it must be able to be used in a step of the dependability MéDISIS process;
– As soon as the information is required, the process clearly identifies it so that the expert can provide it.

3.1 *Standard process*

The description of the product line model must be complete in FM modeling to begin the SE process. The FM concepts represent the variation in the model. Our ES process show on Figure 3 begins with FM modeling for the product line which represents a variant model. Among the results obtained of this model, we find a configuration list of the possible product model in the line. The configuration list is composed of two sets of elements, a set of common elements that are represented by the '*Mandatory*' relation in the FM model and a set of variant elements represented by the '*Optional*' relation and the orthogonal relations like '*OR*', '*XOR*' and '*EX*' relations. Selecting variant elements allows you to specify a valid model among the models in the product line. After this step, we will use these configurations for the system modeling of the products of the line in SysML (it is a modeling language specific to system engineering to model different views of the system). A model produced in SysML corresponds to one configuration (one product) in the FM model. The SysML model must respect to some basic modeling rule in (Kajdan & Idasiak 2015) to ensure consistency between the artifacts used in the model.

The next step is to generate a preliminary FMECA (pre-FMECA) from each product model (the SysML "*Functional, Structural and Behavioral View*" model). The pre-FMECA produced is generated automatically thanks to MéDISIS, it is a transformation of the product model.

Each pre-FMECA undergoes a set of operations directed by the dependability engineer in order to obtain the final FMECA. Indeed, pre-FMECA is then finalized by using the MéDISIS consolidation tool. The consolidation process consists of linking several lines of the pre-FMECA. Each merged line makes it possible to follow the propagation of a failure, starting from the internal cause of a studied function to final system effect. This process contains several tasks:

– Annotation and filtering of the pre-FMECA;
– Generation of failures propagation tables;
– Filling failures propagation tables;
– Consolidation of the pre-FMECA.

The first task is to filter the pre-FMECA. For this, the System Engineer chooses what the lines he considers relevant are or not, and he may make comments to record the reasons for his choices or to provide additional information. In order to be taken into account in the other tasks of the consolidation (Versioning tool).

The second task is to generate the failures propagation tables from the pre-FMECA annotated and filtered.

These correspond to the propagation tables of each element of the pre-FMECA. The dependability engineer must choose from the database a failure mode that will be propagated to studied element linked with a functional flow with an upstream element.

The final task of the process is carried out using the propagation tables and the pre-FMECA annotated and filtered. In fact, to do the consolidated pre-FMECA, the process performs two main steps:

– search for possible paths of failure propagation, i.e. consolidated FMECA lines are one of propagation graph paths;
– for each paths (i.e. FMECA lines) each step of the path allows to concatenate the corresponding pre-FMECA lines.

The MéDISIS database has been set up in the processing process. The database allows consistency of dependability analyzes thanks to the architecture of its meta-model that links functional or dysfunctional information to one another. This database interacts during all stages of the process (from filtering, the propagation tables until the consolidation of the FMECA).

The database is used to complete the pre-FMECA of a new product in the product line, a previous product of which had already allowed FMECA generation. Indeed, the database is developed to be initialized during the first generation pre-FMECA after completion by the expert. Subsequently, with each successive generation, this database will allow:

– to restore the filtering of the expert for the elements of the functional analysis that have not been modified;
– highlight the data that has been modified since the previous version.

This database joins the main principals of ISO 26550 'Asset Base', thus makes it possible to offer consistent information throughout the evolution of the products in the line, specifically designed to facilitate the work of the operating safety expert and to guide the realization of fault analysis.

At the end of the process, we get a specific FMECA for each product in the line.

3.2 *Proposed process*

This process is also based on FM model of the product line. The model will be able to represent the different elements of the line product set. It thus makes it possible to identify the common and the variant parts which constitutes the product line.

However, this model does not represent the flow of data between the elements that are necessary to perform a functional analysis on the system that represents the product line. Although this modeling allowed managing variation in the product line, these limitation of FM modeling require completing the FM modeling with SE concepts (c.f. Figure 1).

To ensure a good basis of the study, we used the SysML language to complete the product line model. The product line model in SysML contains all the elements defined in the FM model and describes them from functional and/or organic point of view. This model makes it possible to represent the system model of the product line containing all the variations identified in the FM model. This model is composed in two parts, a Core part which represents the common elements of the set of products of the range which are identified by the relation '*MA*' in the FM model and the variant part which represents the elements varying from the set of products of the range and which are identified by the relation '*OP*' and orthogonal relations ('Alternative' and '*OR*').

The next step is to generate a preliminary FMECA (pre-FMECA) from the product line model. The pre-FMECA has generated automatically thanks to MéDISIS. The latter must be supplemented by the relations resulting from the FM model in order to present the identified constraints of the modeling of the product line. So, the pre-FMECA obtained will contain all the variant part of the product line. This part is identifiable by the relation '*OP*' in the column "Hierarchical relationship" in pre-FMECA. Similarly, the common part is to identify by the relation '*MA*'.

Therefore, the added concepts from the FM model of the product line present in the last two columns of the parametric pre-FMECA underline the association between the variants and the valid configuration scenarios. These concepts present the dependence relationships (hierarchical and orthogonal), in order to be able to build a pre-FMECA which corresponds to a model of the product line. The hierarchical relationship column defines the optionality '*OP*' or the refine '*MA*' (mandatory). These two properties make it possible to identify the invariant elements presented by the sign '*MA*' belonging to the Core part and the variant elements presented by the sign '*OP*' belonging to the variation part. As for the orthogonal relations column, this one represents the relations of implication '*RE*' and exclusion '*EX*' between the elements to be studied (the variable elements). These two properties make it possible to expose the dependencies between the elements. In particular, we have the relations '*OR*' and '*XOR*' which are also considered as orthogonal relations.

Through this overall representation of the FMECA, we will be able to perform the dependability analysis on products derived from the product line without using their model, on which we could apply the existing FMECA MéDISIS process. However for each product, it would be necessary to redo an identical analysis part. However, we seek to capitalize FMECA type analyzes in the case of a product line. These relationships make it possible to represent constraints between a set of elements in the same point of variations.

Parallel elements that have the same input elements and output elements can be implicitly identified. Given that the principle of redundancy that

is presented by the parallel structure makes it possible to provide several resources to achieve the same function. This principle is a sufficient condition for variant modeling. On the other hand, the '*XOR*' relation complements this principle to become a necessary and sufficient condition in the analysis and modeling of the product line.

The next step is to generate the FMECA from the product line, we will follow the same tasks as the first process. The filtering task consists of selecting the FMECA lines used for the analysis (The elements that make up the lines of this pre-FMECA will all be taken into account in the filtering task). After filtering the product line pre-FMECA, the next task is devoted to generating the propagation tables. The same conditions are taken with respect to the standard process. The association between failure modes and the element under study as well as their causes and effects may be specific to each of these configurations. After obtaining the product line FMECA, the final step can be done. This last step which is setting the FMECA using configuration lists to get a FMECA for a specific product.

4 CASE STUDY

The Winch-PL is a lifting system that has an Operational Safety Brake for Hoist Winch (FOST) (Chieb et al. 2017). This system was chosen to evaluate the two approaches based on a proposed product line model to support the automatic analysis of the dependability of several products in the line. The results of applying the proposed approach to Winch-PL products were used as proof of concept.

The system consists of three parts: the transmission part subsystem which includes the two elastic couplings 1 and 2 with torque limiter which is an alternative with the elastic couplings 1, the chain reducer and the drum. The control subsystem includes the electric motor, two speed sensors (one for the drum and another for the motor) this part contains no variation. The safety subsystem is composed by a safety automata to ensure the braking when there will be a shift of speed between the drum and the engine, and finally, the standard brake and the safety brake.

From the product line model of the Figure 4, we obtain 6 valid possible configurations. The six configurations are obtained using the constraints and relationships applied in the product line model. We focus our study only for the three main configurations.

We began our case study by applying the standard process to the study system. We define the Winch-PL product line model in FM, then create system models for each configuration of the three selected from the configuration list obtained. For each model produced, we generate a pre-FMECA in order to carry out the consolidation tasks (c.f. section 3.1). In the end, we obtained FMECA that corresponds on each product in the product line. Using the database we generate a product line FMECA that brings together all FMECA products.

The saving of time during the realization of the first FMECA of the product is the order 50% thanks to the MéDISIS tools. During the realization of the following FMECA, the versioning tool allows the recovery of the information and the elements present in both lines type 1 and 8. The type 2 to 7 of line in FMECA depends on the number of different variability between the two products and of their functional or organic connection with the core elements.

In the second part of our case study, we applied the new process. We followed every step of this process. At the end of the process, a parametric FMECA of the product line FMECA is obtained. Table 2 is the overview of the parametric FMECA of the Winch-PL system. It contains the Id column which represents the line's identifier and the Element column which contains the name of the studied element, the cause, local effect, and failure mode column. This parametric FMECA thus comprises the note column for the selection of the valid lines, the HR and OR columns for the relationships related to the variability and at the end the column type of line for the classification of the lines of the product line.

To obtain a FMECA of a product from the parametric FMECA of the product line, we made the selection of the lines that are product specific and then proceed to the step of propagation of the failures.

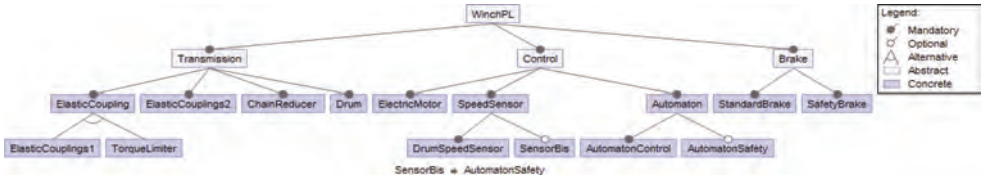


Figure 4. The Winch-PL system in FM.

Table 2. A product line FMECA overview.

Id	Element	Cause	Local Effect	Failure mode	Note	HR	OR	Line type
01	brake	IC	drum	LF	ok	MA	/	T1
02	e_coupling1	IC	ch_reducer	LF	ok	OP	ec1 *xor* tl	T7
03	sensor_bis	IC	saf_automat	LF	ok	OP	as = >sb	T8
04	spd_sensor	IC	automat	LF	ok	MA	/	T1
05	ele_motor	IC	e_coupling1	LF	ok	OP	ec1 *xor* tl	T2
06	trq_limiter	ele_motor	ch_reducer	LF	ok	OP	ec1 *xor* tl	T6

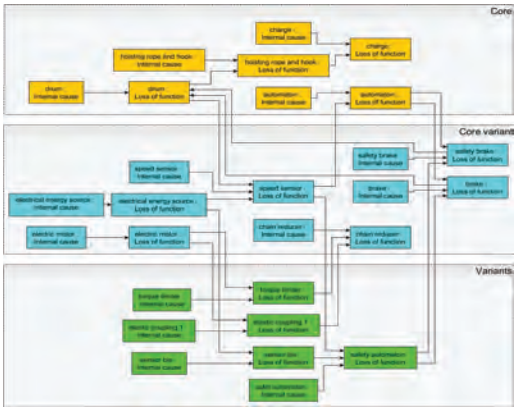


Figure 5. Failure propagation graph for the Winch-PL system.

For example, if we analyze, the "Require" relation who is used to manage the variation of the optional elements 'safety automaton' and 'sensor bis'. The impacted FMECA lines have several types (T6, T7, and T8). These elements must be optional in order to apply this relationship. The alternative choice between two elements is presented in the pre-FMECA of our case study by the relation "*XOR*". For example, the presence of the element '**elastic coupling 1**' excludes the presence of the element '**torque limiter**' and vice versa. Indeed, the analysis is done in the case of a product configuration on a single element among the set of alternative elements. The alternative relationship is applied to optional elements.

Figure 5 shows the failure prognosis graph for Winch-PL system. For our study, we chose only one failure mode that is "loss of function", over the internal failure of the element that is represented by its internal cause. A color code was established to distinguish each part. This graphic representation is automatically generated from pre-FMECA or consolidated FMECA. This representation is synchronized to the standard tabular representation of the FMECA and help the analyst to make his decisions faster.

5 CONCLUSION

When products are defined in the product line, it is difficult to not repeat the same FMECA type study for the common elements of these products. In order to optimize the time needed to produce FMECA for each product, we first studied how product line models influence information needed for SE modeling and then their dependability analysis.

We have adapted the MéDISIS method to take into account a product line model. We compared two automated analysis processes by the MéDISIS tools, the first uses the ability of the MéDISIS versioning tools but uses product models. The second allows to start analyze with a product line model, which makes it possible to realize the dependability analysis earlier.

These results allowed us to have a very important gain in time of analysis time and to be able to reuse parts of this analysis for another product system of the same product line.

In this first experimentation, we note an acceleration of the dependability studies from standard FMECA. Without other feedback, it's difficult to quantify this gain because it depends on the number and the nature of each variability. Moreover, make a FMECA at product line stage allow an earlier starting of dependability studies.

ACKNOWLEDGMENT

The research leading to these results has received funding from the French single inter-ministry fund, sixteenth framework program (F1407032) for the FOST project.

REFERENCES

Berrebi, J. (2013). Contribution à l'intégration d'une liaison avionique sans fil. L'ingénierie système appliquée

à une problématique industrielle. 2013. Doctoral dissertation. Ecole Polytechnique X.

Chieb, B., Idasiak, V. & Kratz, F., (2017). Analyse de sûreté de fonctionnement d'une ligne de produits à l'aide de son modèle. QUALITA. 2017. Bourges.

Cressent, R. et al. (2013). Designing the database for a reliability aware Model-Based System Engineering process. *Reliability Engineering & System Safety*, 111, 171–182.

David, P. et al. (2010). Reliability study of complex physical systems using SysML. *Reliability Engineering & System Safety*, 95(4), 431–450.

De Oliveira et al. (2014, November). A model-based approach to support the automatic safety analysis of multiple product line products. In Computing Systems Engineering (SBESC), 2014 Brazilian Symposium on (pp. 7–12). IEEE.

Friedenthal, S. et al. (2014). A practical guide to SysML: the systems modeling language. Morgan Kaufmann.

Grönniger, H. et al. (2008, March). Modeling variants of automotive systems using views. In Proceedings of Workshop Modellbasierte Entwicklung von eingebetteten Fahrzeugfunktionen (MBEFF) (pp. 76–89).

Höfig, K. et al. (2014). MetaFMEA-A framework for reusable FMEAs. In *Model-Based Safety and Assessment* (pp. 110–122). Springer, Cham.

Kajdan, R. & Idasiak, V. (2015). Model-based systems engineering and failure analysis: Experience feedback. ESREL 2015, Zürich.

Le Put, A., L'Ingénierie Système d'une Ligne de Produits, Ouvrage collectif AFIS, 1ere Edition, 2016.

Lee, K. et al. (2002, April). Concepts and guidelines of feature modeling for product line software engineering. In International Conference on Software Reuse (pp. 62–77). Springer, Berlin, Heidelberg.

Mazo, R. (2011). A generic approach for automated verification of product line models. 2011. Doctoral dissertation. Université Panthéon-Sorbonne-Paris I.

Sierla, S. et al. (2014). Safety analysis of mechatronic product lines. *Mechatronics*, 24(3), 231–240.

Steiner, E. et al. (2013). Managing SPL variabilities in UAV Simulink models with Pure: variants and Hephaestus. *CLEI Electronic Journal*, 16(1), 7–7.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Preliminary safety assessment of circular variable nacelle inlet concepts for aero engines in civil aviation

S. Kazula, D. Grasselt, M. Mischke & K. Höschler
Brandenburg University of Technology, Cottbus, Germany

ABSTRACT: A safe design process and its application are introduced to a concept study for circular variable aero engine inlets. The paper highlights the tasks of inlets, the compromise in designing them and how using variable inlets could solve this compromise and allow for faster and more efficient commercial aircraft. However, high safety and reliability requirements bring up disadvantages. Tackling these disadvantages, a systems engineering approach is complemented by a safety assessment process, according to Aerospace Recommended Practice ARP 4754A. Safety methods that are applicable during early phases of the product development process are presented and applied to develop feasible variable inlet concepts. Hence, safety requirements, potential failure events and resulting failure modes are systematically identified, assessed and mitigated. The mitigation of a failure condition by the means of redundancy within the adjustment control system is presented.

NOMENCLATURE

AMC	Acceptable Means of Compliance
ARP	Aerospace Recommended Practice
CCA	Common Cause Analysis
CMA	Common Mode Analysis
CS	Certification Specification
DD	Dependence Diagram
EASA	European Aviation Safety Agency
FAST	Function Analysis System Technique
FHA	Functional Hazard Assessment
FMEA	Failure Modes & Effects Analysis
FTA	Fault Tree Analysis
MA	Markov Analysis
PRA	Particular Risk Analysis
PSSA	Preliminary System Safety Assessment
RTO	Rejected Take-off
SAE	Society of Automotive Engineers
SSA	System Safety Assessment
TRL	Technology Readiness Level
US	United States
VDI	Verein Deutscher Ingenieure/ The Association of German Engineers
ZSA	Zonal Safety Analysis

1 INTRODUCTION

Improvement of efficiency and thus emissions as well as travelling speed are major goals in civil aviation (European Commission 2001, European Commission 2011). These goals also depend on the aero engine and its subsystems. One of these subsystems is the inlet that supplies the aero engine with air. The geometry of these inlets is designed as a fixed compromise regarding aerodynamic drag. Significant work on the optimisation of the surface geometry of rigid inlets has been carried out (Luidens et al. 1979, Pierluissi et al. 2011, Albert & Bestle 2014, Schnell & Corroyer 2015), however, a rigid inlet can only achieve the best compromise between optimal geometries for different flight conditions.

Using aero engine inlets with a variable lip and duct geometry for different flight conditions is perceived as a possibility to reduce aerodynamic drag and therefore to have a positive effect on aircraft efficiency and speed (Baier 2015). Therefore, studies, e.g. (Kondor & Moore 2004), da Rocha-Schmidt et al. (2014) and Ozdemir et al. (2015), have been conducted on concepts for variable inlets for commercial aircraft. Additionally, first patents for variable inlets, e.g. US 4075833 and US 5000399, exist.

The only commercial aircraft that used variable inlet systems are the Concorde and the Tupolev Tu-144. Their variable inlet systems consisted of movable ramps and flaps for supersonic flight. Compared to subsonic aircraft both aircraft models had huge deficits concerning range, efficiency and noise, which is why they have been retired.

Within the scope of this study, in contrast to inlets with ramps and flaps, variable inlets with a closed contour that can adjust the lip and duct geometry are investigated. From the beginning, the development process is accompanied by a safety

and reliability process. The aim of this study is to achieve technical feasible concepts for this kind of variable inlets, related to Technology Readiness Level (TRL) 3.

Although those variable inlets have been investigated in first research studies they did not find its way in modern civil aviation yet. This can be justified by the trade-off between the positive and negative effects of the usage of variable inlets on commercial aircraft. Besides additional weight and production costs, negative effects are the increased complexity and thus potential safety and reliability issues. These issues are related to the numerous requirements and boundary conditions regarding inlets.

In contrast to earlier studies, it is reasonable to supplement the utilised systems engineering approach for product development with a safety assessment process to identify and fulfil the safety requirements in aviation right from the beginning. Thusly, possible events as well as resulting failure conditions can be identified and possible safety issues remedied during early phases of the product development process. Such an approach has been carried out successfully within a project concerning coupled actuation systems for thrust reversers and variable nozzles of aero engines (Grasselt & Höschler 2015, Grasselt et al. 2017). As a result, variable aero engine inlets could be utilised in civil aviation, whereby these aircraft could be more efficient and faster, while maintaining high safety and reliability.

Therefore, this paper deals with the safety assessment process within concept development and preliminary design of variable inlets in civil aviation. First, the inlet system and its necessary functions are described. Afterwards, the utilised safety assessment process referring to ARP4761 is introduced and applicable safety analysis methods, e.g. Functional Hazard Assessment (FHA) and Fault Tree Analysis (FTA), are presented. Some selected results of the applied methods are shown and discussed. These results contain the identification of safety-relevant requirements, failure conditions and events. Finally, the influence of this assessment on the design of the investigated concepts is shown.

2 AERO ENGINE INLETS

2.1 *Tasks and implementation*

The main purpose of nacelle inlets for aero engines is to divide the free stream in front of the aero engine, depending on the capture stream tube as a function of operating conditions, into an internal and an external airflow. The external airflow shall flow over the nacelle surface while avoiding flow separation and other sources of drag (Mattingly 2006). The objective of the internal airflow is to supply the aero engine during each operating condition with the correct quantity of air at a desired flow velocity (Rolls-Royce Plc 2015). The axial flow velocity that is required by the compressor system of the aero engine is around Mach 0.5 (MacIsaac & Langton 2011). Hence, a deceleration of the internal airflow is required at flight speeds above Mach 0.5 to ensure a highly efficient and safe operation of the compressor system. Thus, the inner contour of the inlet duct has to be designed as a diffuser (Farokhi 2014). The efficiency and operational stability of the compressor system also depends on the uniformity of the airflow. Therefore, flow separations should be avoided under all conditions, as they can lead to vibration excitations, rotating stall and engine surge concatenated by a loss of thrust and reduced aero engine durability.

Moreover, the fan and compressor induced noise emissions have to be reduced by the inlet, which is achieved by integrating acoustic treatment into the diffuser wall, see Figure 1.

Furthermore, probes for pressure and temperature measurement at the fan level can be part of the inlet. Hence, these measured data must be provided to a flight control system.

Additionally, the inlet has to protect itself and the compressor system from icing and its consequences, e.g. impact damage and flow separation.

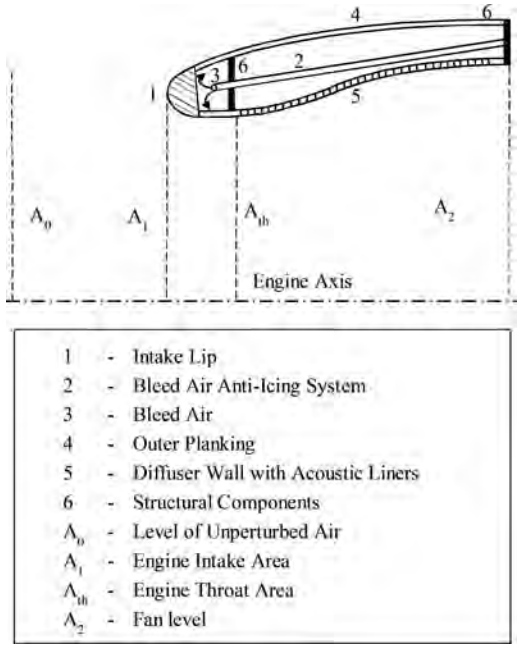


Figure 1. Typical design of a rigid subsonic nacelle inlet.

An anti-icing system is installed in the inlet to ensure this. Most commonly, electrical or bleed air anti ice systems are used (Rolls-Royce Plc 2015). Bleed air anti ice systems transfer hot air from the compressor to the inlet lip to prevent icing. Aluminium is typically used for the inlet lip, due to its good heat conductivity. Furthermore, aluminium is light and resilient to foreign object damage, sand erosion, hail and bird strikes. It is to prove during certification that thrust can be maintained to a certain level to assure that the flight can be safely continued after a single bird strike (Hedayati & Sadighi 2016).

The outer planking as well as the inner boundary with the acoustic linings is made of composites, which minimises weight. Thereby, the inlet should withstand stipulated loads and be robust against damage (EASA 2016).

This way, incidents like the Air France Flight AF-66, where the fan and the engine inlet were separated from the aircraft during flight, should occur at a very low rate to minimise the risk for passengers and crew (Aviation Herald 2017).

2.2 *Trade-off in design and variable inlets*

Ensuring reliable operation during all flight phases must be unified with aerodynamic requirements during the geometric design process of nacelle inlets (Seddon & Goldsmith 1999). On the one hand, the inlet should be highly efficient at high flight velocities above Mach 0.8 during cruise operation. On the other hand, it is necessary to avoid flow separations and hazardous events during take-off and climb operation up to Mach 0.3.

Figure 2 presents, optimal aerodynamic contours of an inlet for different flight conditions. Optimal efficiency at high flight velocities can be achieved by a thin or sharp lip contour combined with a small entry area A_1 to minimise wave and spillage drag (Farokhi 2014). As the entry area is reduced, a longer diffuser is required at high velocities to avoid flow separations. Sharp inlet lips can be used for flight Mach numbers up to 1.6 without significant losses (Farokhi 2014). However, at low aircraft velocities, where high angles of incidence and crosswind can occur, a sharp or thin lip contour is sensitive to flow separations and its potential negative consequences. For these operating conditions, a round and thick inlet lip with a large inlet area is optimal. Such a 'blunt' lip geometry causes higher drag and thus less efficiency during operation at higher flight Mach numbers (Bräunling 2015). Hence, conventional rigid subsonic inlets can only accomplish a compromise geometry that produces increased losses during high flight velocities.

Using a variable inlet, which applies the optimal contour for each flight condition, can improve efficiency and maximum aircraft speed during cruise

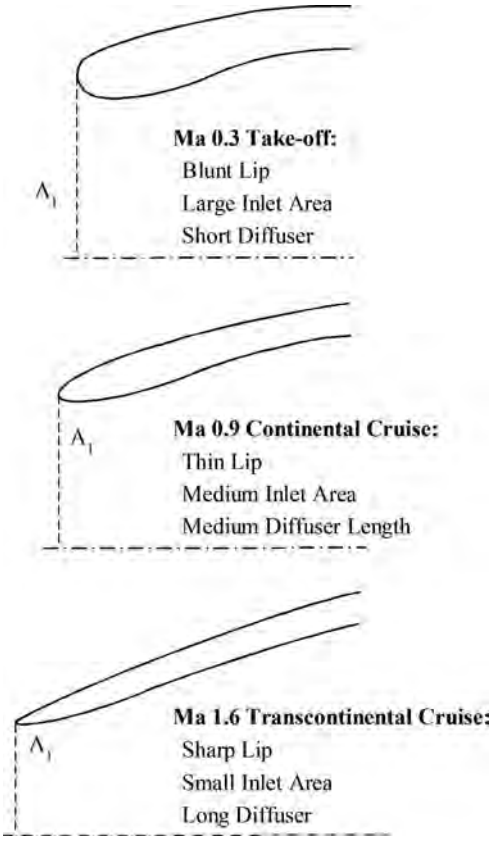


Figure 2. Tendencies of optimal inlet contours for different flight phases and velocities.

flight, while ensuring reliable operation during take-off and climb. However, it needs to be considered that the variation of the inlet presents an additional function, which can entail further reliability and safety issues (SAE Aerospace 2010).

Therefore, variable nacelle inlets for Mach numbers up to 1.6 are investigated focussing on safety and reliability in the context of an internal research project at the chair of Aero Engine Design at the Brandenburg University of Technology. Within the scope of that project, a methodical safe design approach is developed and utilised to perform a feasibility study for concepts for variable inlets up to TRL 3. The safe design approach for variable inlet concepts has been presented in Kazula & Höschler (2017). These concepts can be divided into three geometry adjusting mechanism groups: movement of rigid segments, deformation of elastic surface material and boundary layer control. The preliminary safety assessment in chapter 4 focusses primarily on the first group.

3 SAFE DESIGN APPROACH

3.1 *Systems engineering*

When designing complex systems, it is favourable to utilise a systems engineering approach, e.g. Design for Six Sigma and VDI Guideline 2221. Methodical design approaches allow for improved requirements, interface and risk management, complexity reduction as well as more efficient solving of the design task. Furthermore, weaknesses during development can be minimised. Most methodical design approaches are based on a common iterative structure:

- starting with analyses concerning requirements and functions of the desired product,
- continuing with the allocation of solution principles to functions, preliminary design and preselection of potential solution architectures
- and concluding with detailed design, as well as validation and evaluation of the design.

As the desired product functions in modern industries become increasingly complex, particular safety efforts, such as safety assessments and tests, should be considered to ensure safety and reliability (Bertsche & Lechner 2004).

3.2 *Safety and reliability engineering*

The safety and the reliability of a product play important roles in various industries for reasons of efficiency and business sustainability up to social acceptance. These industries, all of them using a separate safety approach, include the power, rail, shipping, automotive and aviation industry (Verein Deutscher Ingenieure 2000). The most effective time to improve product safety and reliability is during early development phases (Bertsche & Lechner 2004). On the one hand, this can be achieved by using a mature design approach according to design guidelines, which contains a precise and complete requirements document and early testing. On the other hand, analytical methods can be utilised to forecast reliability and to find weaknesses in design. Analytical methods are divided into qualitative methods and quantitative methods. Whereas quantitative methods, e.g. Markov Analysis are used to predict probability of faults, qualitative methods like the FHA are utilised to identify and assess potential failure events and resulting conditions. The safety and reliability of a product can be positively influenced by utilising appropriate safety and reliability methods during each step of the development process. For this purpose, multiple authors, e.g. Bertsche & Lechner (2004) and Meyna & Pauli (2010), as well as organisations, e.g. International Organization for Standardization (2011) in ISO 26262 for automotive industry, propose safety processes for different areas of application.

3.3 *Safety process in aviation*

In aviation, failures could lead to fatal accidents. The risk for accidents can be reduced by improvements in the areas of airplane design, flight operations, maintenance, air traffic management, regulations and design methodologies (Hasson & Crotty 1997). Since the Chicago Convention in 1944, local aviation authorities have been publishing regulations to ensure a safe operation. The European Aviation Safety Agency (EASA), for instance, releases Certification Specifications (CS), e.g. CS-25 – Large Aeroplanes. Paragraph CS-25 AMC 25.1309 describes the safety assessment process in aviation that is based on the process in ARP 4754A (SAE Aerospace 2010) and the methods in ARP 4761 (SAE Aerospace 1996). The design approach for variable inlets of Kazula & Höschler (2017) utilises safety methods of the ARP 4761, see Figure 3.

3.4 *Safety methods for variable inlet development*

The first of the methods presented in Figure 3 is the preparation of a requirements document that is as complete and accurate as possible. Therefore, all requirements that could be introduced by the different stakeholders should be identified and quantified (Sadraey 2013). A product with high safety and reliability as well as low development costs can be achieved during early stages of development by focussing on the Type Certificate Program and its entailed requirements. These requirements are set by the aviation authorities and have been considered within this study. However, the creation of a requirements document is an iterative process as some safety requirements are only identifiable during the safety assessment.

To comply with a requirement, a product must fulfil a function. A function is defined as the conversion of input material, energy or data into desired output (Roth 2001). Within the scope of the development process it is usual to perform a functional structure analysis. This way, the necessary main and secondary functions are identified and broken down up to elementary functions like ‘convert’ or ‘increase’ (Koller & Kastrup 1998). A detailed functional structure can contribute to a high product safety by preventing design flaws, however, a too detailed breakdown should be avoided, as this limits the solution variety during development (Verein Deutscher Ingenieure 1993). Hence, it is reasonable to start with a simple functional structure and iteratively increase the level of detail within the development process (Roth 2000). The most common means to create a functional

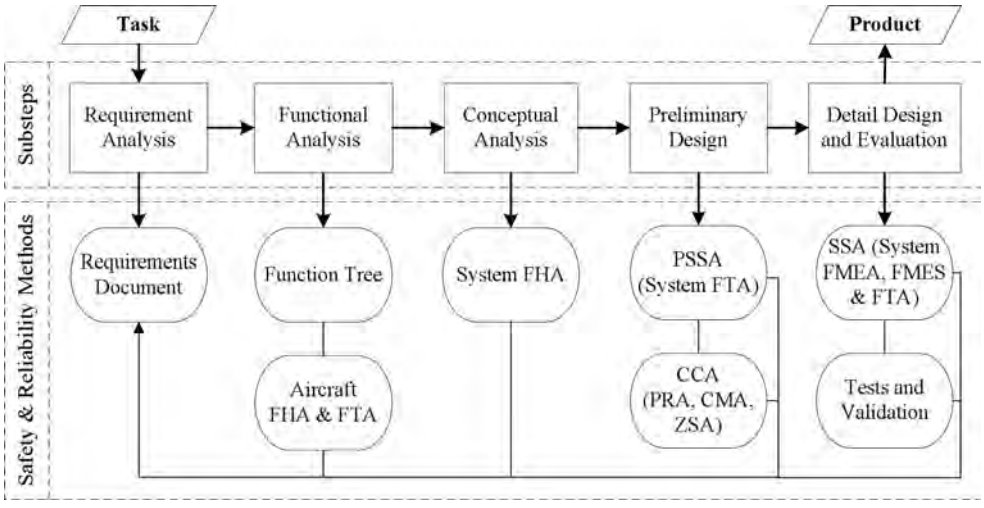


Figure 3. Suitable safety and reliability methods for the separate phases of the methodical design approach.

structure are the FAST-diagram (Function Analysis System Technique), the function net and the function tree. While all of these methods are quite similar, function trees are the recommended choice, because of their structure that synergises well with the Functional Hazard Assessment (FHA) (Verein Deutscher Ingenieure 2000).

The FHA is a qualitative method that should be performed at the beginning of the safety process. At this point, it is reasonable to carry out qualitative safety methods, as they support the systematic investigation of failure conditions, causes and consequences. During later phases of the development process these qualitative methods can be replaced by quantitative methods to investigate reliability in more detail (SAE Aerospace 2010). The main objective of the FHA is to systematically assess functions of a systems and to identify and classify failure conditions as well as their effects. It is performed on Aircraft, System and if required subsystem level. Similar to the function tree, a disadvantage of this method is that it can be difficult for inexperienced users to apply this method appropriately as it allows for an easy deployment of enormous tables and as the classification of hazards can be rather subjective. Kritzinger (2016) describes the advantages of an FHA, one of them being the provision of top level events for the following Preliminary System Safety Assessment (PSSA).

The PSSA is used to analyse which single or multiple system, subsystem or component failures lead to the functional hazards that have been identified within the FHA. This way, safety related design requirements can be determined and concept designs can be evaluated. A PSSA is performed by utilising the Fault Tree Analysis (FTA), the Dependence Diagram (DD) or the Markov Analysis (MA), supplemented by a Common Cause Analysis (CCA) (SAE Aerospace 1996). While the most commonly used FTA, which is an iterative top down method, presents the relationship between failures through logic gates, the DD uses paths and the MA time dependant probability functions. On the one hand, FTA, DD and MA have many advantages, which are discussed for instance in Kritzinger (2016) and SAE Aerospace (1996). On the other hand, all of these methods lack a systematic that assures completeness (Verein Deutscher Ingenieure 2000).

Therefore, within the System Safety Assessment (SSA) it is reasonable to combine the top down method FTA with the bottom up method Failure Modes and Effects Analysis (FMEA), that is simple to use but iterative and time consuming (Kritzinger 2016). The SSA evaluates the compliance of the investigated system with the safety requirements from the PSSA by applying analyses and test methods. In addition to the FMEA, the SSA includes quantitative FTAs and a CCA. A CCA comprises of a Zonal Safety Analysis (ZSA), a Particular Risk Analysis (PRA) and a Common Mode Analysis (CMA). A PRA is utilised to identify external events and a ZSA to determine individual failure modes that can cause hazards. A CMA is used to verify independence of functions, as this is not done within the FHA (SAE Aerospace 1996). Finally, tests must be performed to validate the results from earlier analyses and to comply with the requirements set by the aviation authorities. It is reasonable that the less experience

the product developer has, the more tests should be performed for a successful certification.

4 SELECTED RESULTS

4.1 *Function trees*

The Aircraft has to fulfil different 1st level or top level functions according to SAE Aerospace (2010) and Kritzinger (2016), one of them being 'control thrust', see Figure 4. Thrust control is achieved by generating, adjusting, ensuring and determining thrust. The inlet system influences all of these functions, e.g. to generate thrust, the inlet system has to provide the aero engine with an airflow.

The function tree that is presented in Figure 5 contains a few important functions of the nacelle inlet system. Different flight phases require for different geometries to achieve minimal drag and an internal airflow with high uniformity and the appropriate velocity. An inlet geometry adjustment system can improve the achievability of these functional requirements. Additionally, an adjustment system could be used to prevent negative effects of icing by detaching ice. Furthermore, it could utilise existing systems for data and energy transfer. The influence of the variability on other inlet subsystems has to be investigated, as the inlet would not be a rigid enclosed structure anymore and the installation space for acoustic treatment

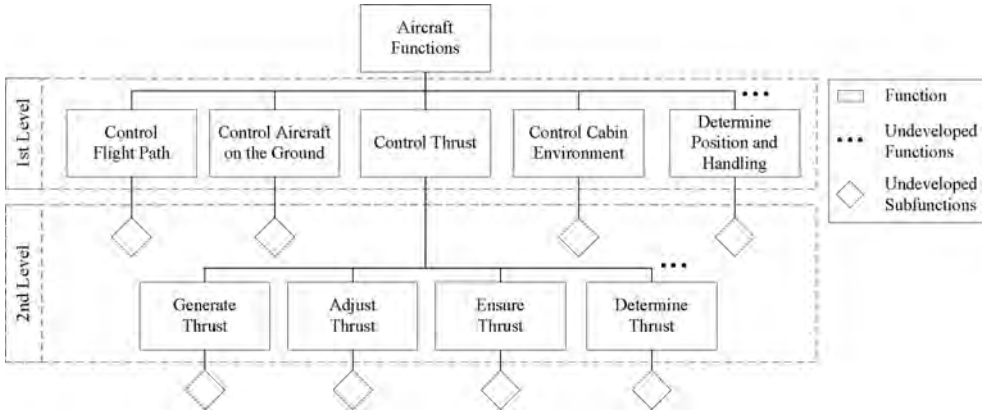


Figure 4. Simplified aircraft function tree.

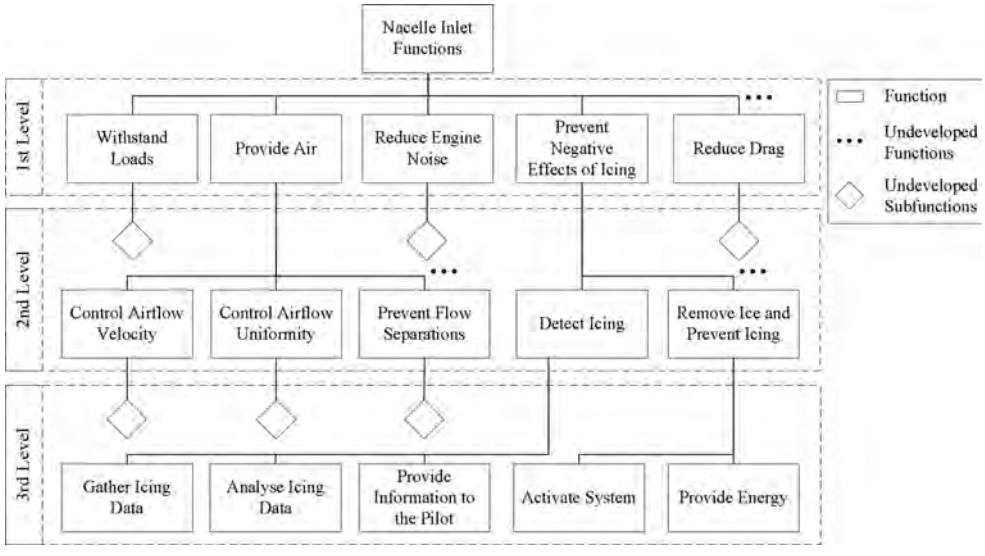


Figure 5. Simplified aero engine nacelle inlet system function tree.

could be reduced. The essential functions of an inlet adjustment system are change and locking of the inlet geometry as well as gathering, transferring and processing geometry data.

4.2 *FHA*

The function trees can be used to determine the input for the FHA. For clarity reasons, only the function 'adjust inlet geometry' and its impact on the aircraft is presented in the following. After identifying the functions, associated failure conditions and their effects must to be determined regarding single and multiple failures during normal, e.g. take-off, climb and cruise, as well as special conditions, e.g. windmilling. The main failure condition of the adjustment system is that an undesired geometry is adjusted, what could affect a single engine or all engines.

A further categorisation is possible concerning:

- which geometry is adjusted,
- is the system either incapable to maintain the desired geometry or to adjust the geometry on time or in general,
- is the failure occurring either announced or unannounced, either abruptly or anticipated.

Then, the effects of these failure conditions must be identified and classified. Possible effects of the presented failure conditions are 'loss of thrust' or 'reduced efficiency'. The classification is conducted according to CS-25 AMC 25.1309 regarding the severity of an effect on aircraft, crew or occupants and is divided into 'catastrophic', 'hazardous', 'major', 'minor' and 'no safety effect'. Each of these classifications entails a specific probability requirement concerning occurrence of a failure mode. Table 1 presents the failure classification for an undesired, abrupt but announced

Table 1. Failure classification for the event 'undesired geometry is adjusted' on a single engine.

		Probability requirement
Flight phase	Classification	Events per flight hour
Start	No Safety Effect	No Probability Requirement
Idle	No Safety Effect	No Probability Requirement
Taxi	No Safety Effect	No Probability Requirement
Take-off	Minor	<1.0E-03
RTO	No Safety Effect	No Probability Requirement
Climb	Minor	<1.0E-03
Cruise	No Safety Effect	No Probability Requirement
Descent	No Safety Effect	No Probability Requirement
Approach	No Safety Effect	No Probability Requirement
Go-around	Minor	<1.0E-03
Landing	No Safety Effect	No Probability Requirement
Brake	No Safety Effect	No Probability Requirement

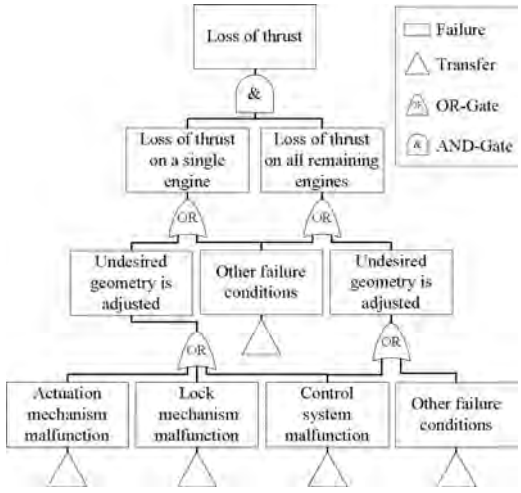


Figure 6. Simplified fault tree for the inlet adjustment system.

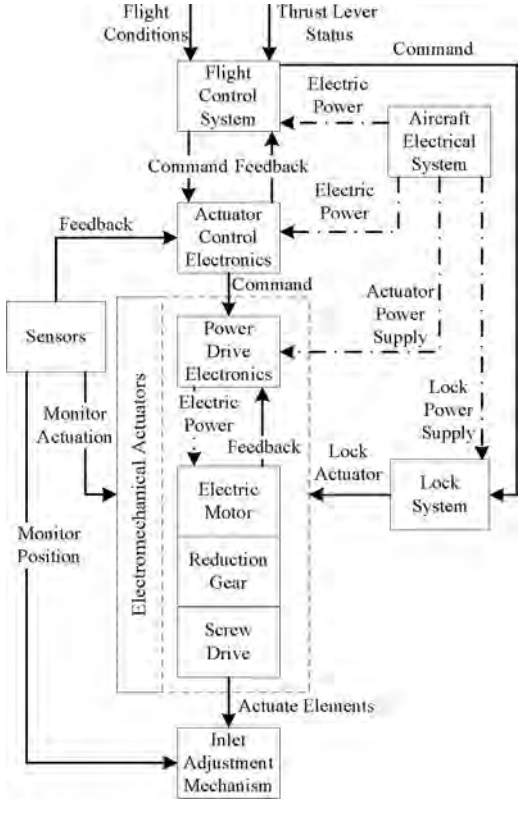


Figure 7. Control flow chart for a variable inlet system.

adjustment of a sharp thin cruise geometry on a single engine. Commercial aircraft must be able to fly with one engine inoperative. A failure of the adjustment system of a single engine can only cause surge and therefore loss of thrust on a single engine. Hence, this event only leads to a slight reduction of the aircraft functional capability and a slight increase in workload for the flight crew. However, a failure of the adjustment system on multiple engines could lead to loss of thrust on more than one engine. For instance, during take-off this can result in a rejected take-off (RTO). This is classified as a hazardous mode with a probability requirement of less than 1.0E-07 events per flight hour (EASA 2016).

4.3 *FTA and failure prevention*

The fault tree in Figure 6 shows how the hazardous event 'loss of thrust' is based on the loss of thrust on all engines. This is caused by the incapability to provide air sufficiently due to the adjustment of an undesired inlet geometry. This could be caused by a single control system failure.

Single failure conditions can be avoided by a redundant design. This can be achieved by design modifications. Therefore, the adjustment system is protected by a locking system, which is independent from the actuator control electronics, see Figure 7. Furthermore, the individual engines should be controlled separately.

5 CONCLUSIONS

While being able to improve the speed and efficiency of modern aircraft, variable inlets did not find its way in commercial aviation yet due to potential safety and reliability issues. These issues can be reduced by applying a coupled design and safety process that is integrated in the early phases of the product development process. Moreover, the methods of this process are discussed and an FHA as well as an FTA are applied to variable inlet concepts. This results in the identification of failure probability requirements and design adjustments.

It is still impossible to ensure that every failure is identified during the development, but the probability of design faults can be reduced significantly. During the investigation of variable inlets, the bottom up method FMEA should be applied to complement the top down approach of the FTA. A CCA, especially a PRA, should be performed to investigate independence and external events. Finally, validation and verification tests must be carried out in consultation with the aviation authorities.

It should be considered that some safety regulations, guidelines and expert opinions could be too conservative and thus limit the solution variety of new technologies. Furthermore, safety and reliability is often neglected during academic studies.

Due to the safe design approach of this study, function trees, an FHA and an FTA have been described in this paper. This results in the identification of safety requirements and design faults. These faults can be mitigated by the means of redundancy within the adjustment control system. This way, the negative effects of variable inlets can be reduced during early stages of the product development. Hence, variable inlets could be applied in commercial aviation. This could allow for faster and more efficient aircraft, while maintaining safety and reliability. Moreover, the safe design approach could be applied to other product developments in aviation or other industries, as there is currently no general safe design approach.

REFERENCES

Albert, M. & Bestle, D. 2014. Automatic Design Evaluation of Nacelle Geometry Using 3D-CFD. *15th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference. Atlanta, GA.*

Aviation Herald 2017. *Incident: France A388 over Greenland on Sep 30th 2017, uncontained engine failure, fan and engine inlet separated.*

Baier, H. 2015. *Morphelle—Project Final Report.* München.

Bertsche, B. & Lechner, G. 2004. *Zuverlässigkeit im Fahrzeug- und Maschinenbau.* Berlin: Springer.

Bräunling, W.J.G. 2015. *Flugzeugtriebwerke.* Berlin: Springer.

da Rocha-Schmidt, L. et al. 2014. Progress towards Adaptive Aircraft Engine Nacelles. *29th Congress of the International Council of the Aeronautical Sciences, St. Petersburg, Russia.*

EASA 2016: *CS-25. Certification Specifications and Acceptable Means of Compliance for Large Aeroplanes—Amdt 18.*

European Commission 2001. *European Aeronautics. A vision for 2020.* Luxembourg: Off. for Off. Publ. of the Europ. Communities.

European Commission 2011. *Flightpath 2050. Europe's vision for aviation.* Luxembourg: Publ. Off. of the Europ. Union.

Farokhi, S. 2014. *Aircraft propulsion.* Chichester: Wiley.

Grasselt, D. & Höschler, K. 2015. Safety Assessment of Aero Engine Thrust Reverser Actuation Systems. *22nd International Symposium on Air Breathing Engines.*

Grasselt, D. et al. 2017. A Design Approach for a Coupled Actuator System for Variable Nozzles and Thrust Reverser of Aero Engines. *ASME 2017 Fluids Engineering Division Summer Meeting.*

Hasson, J. & Crotty, D. 1997. Boeing's safety assessment pro-cesses for commercial airplane designs. *Avionics Systems Conference AIAA/IEEE.*

Hedayati, R. & Sadighi, M. 2016: *Bird strike*. Cambridge: Woodhead Publishing.

International Organization for Standardization 2011: *ISO 26262*. Switzerland: International Organization for Standardization.

Kazula, S. & Höschler, K. 2017. A Systems Engineering Approach to Variable Intakes for Civil Aviation. *7th European Conference for Aeronautics and Space Sciences.*

Koller, R. & Kastrup, N. 1998. *Prinziplösungen zur Konstruktion technischer Produkte.* Berlin: Springer.

Kondor, S. & Moore, M. 2004. Experimental Investigation of a Morphing Nacelle Ducted Fan. *NASA/CP-2005.*

Kritzinger, D. 2016. *Aircraft system safety.* Duxford: Woodhead Publishing.

Luidens, R.W. et al. 1979. An Approach to Optimum Subsonic Inlet Design. *ASME Conference.*

MacIsaac, B. & Langton, R. 2011. *Gas Turbine Propulsion Systems.* West Sussex: Wiley.

Mattingly, J. 2006. *Elements of Propulsion.* Reston: AIAA.

Meyna, A. & Pauli, B. 2010. *Taschenbuch der Zuverlässigkeitstechnik.* München: Hanser.

Ozdemir, N. et al. 2015. Morphing nacelle inlet lip with pneumatic actuators and a flexible nano composite sandwich panel. *Smart Mater. Struct.* 24 (12), S. 125018.

Pierluissi, A. et al. 2011. Intake Lip Design System for Gas Turbine Engines for Subsonic Applications. *49th AIAA Aerospace Sciences Meeting.*

Rolls-Royce Plc 2015. *The jet engine*. Chichester: Wiley.

Roth, K. 2000. *Konstruieren mit Konstruktionskatalogen. Band 1: Konstruktionslehre.* Berlin: Springer.

Roth, K. 2001. *Konstruieren mit Konstruktionskatalogen. Band 2: Kataloge.* Berlin: Springer.

Sadraey, M.H. 2013. *Aircraft design.* Chichester: Wiley.

SAE Aerospace 1996. *ARP 4761.* Warrendale: SAE Aerospace.

SAE Aerospace 2010. *ARP 4754A.* Warrendale: SAE Aerospace.

Schnell, R. & Corroyer, J. 2015. Coupled Fan and Intake Design Optimization for Installed UHBR-Engines with Ultra-Short Nacelles. *22nd International Symposium on Air Breathing Engines*.

Seddon, J. & Goldsmith, E.L. 1999. *Intake aerodynamics.* Reston: AIAA.

Verein Deutscher Ingenieure 1993. *VDI 2221. Methodik zum Entwickeln und Konstruieren*. Düsseldorf: Beuth.

Verein Deutscher Ingenieure 2000. *VDI-Berichte 1546. Sicherheit komplexer Verkehrssysteme.* Düsseldorf: VDI.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

A PMS-MMDD model for reliability assessment of multi-state phased-mission system

Xiang-Yu Li
Center for System Reliability and Safety, University of Electronic Science and Technology of China, Chengdu, China
Chair System Science and the Energy Challenge, Fondation Électricité de France (EDF), CentraleSupélec, Université Paris Saclay, Cedex, France

Yan-Feng Li, Hong-Zhong Huang & Junyu Guo
Center for System Reliability and Safety, University of Electronic Science and Technology of China, Chengdu, China

Enrico Zio
Chair System Science and the Energy Challenge, Fondation Électricité de France (EDF), CentraleSupélec, Université Paris Saclay, Cedex, France
Department of Energy, Politecnico di Milano, Milano, Italy

ABSTRACT: Multi-State Phased-Mission Systems (MS-PMSs) are Multi-State Systems (MSSs) that accomplish different missions in a series of consecutive and non-overlapping time durations, called phases. Many practical PMSs are non-repairable, such as the spacecraft working in the outer space. In the non-repairable MS-PMS, the dependence among phases are more complicated, like the components' states cannot be better in the latter phases. The paths of the MMDD model generated by traditional MMDD manipulation rules cannot consider this kind of dependency and part of the paths may have the self-conflict problems. To solve this problem, a MMDD algorithm for MS-PMS and PMS-MMDD model are proposed. By the PMS-MMDD model, the system MMDD model can be generated without any additional steps. At last, the Monte Carlo simulation method is used to certify the proposed PMS-MMDD model.

1 INTRODUCTION

The PMSs are systems that need to complete different missions in multiple, consecutive, non-overlapping time durations, known as phases (Xing and Amari, 2008). The PMSs are commonly seen in the aerospace industry, like the spacecraft, whose lifetime can be divided into several phases: launching, orbit-transfer, on-orbit operation, back-to-earth. The systems/components in the PMSs can exhibit multiple performance levels or states, called MS-PMS. The reliability of the MS-PMS is defined as the probability that the system state stays above the failure states in all phases. The challenges in analyzing the MS-PMS are mainly due to two aspects (Shrestha et al., 2011):

1. The phase dependence, the states at the end of one phase should be equal to the beginning of the consecutive phase and the components failed in one phase will remain in the failure state in the following phases;
2. The dynamic behaviors, the working components, system structure and the working environment are different in different phases, so the distinct model for each phase are necessary.

The existing PMS analysis methods can be divided into two classes: the analytical methods and the simulation methods. The analytical methods can be further categorized into three types: (1) the combinatorial methods, like the Binary Decision Diagram (BDD), (Zang et al., 1999), or MMDD method (Shrestha et al., 2011); (2) state space model, such as the Continuous Time Markov Chain (CTMC) (Wu et al., 2012); (3) modular method (Ou and Dugan, 2004), which combines the combinatorial methods and the state space model and possess advantages of both. But most of the existing methods are limited into the binary state systems/components on the other hand, many research efforts have been developed in the analysis of the Multi-State Systems (MSSs) (Liu et al., 2008, Liu and Huang, 2010), but all of these works do not consider the system multi-phased behaviors.

The researches on the MS-PMS are very few until now due to the complexity of both PMS and MSS. And the multi-state behaviors renders more

complicated the phase dependence and the dynamics leading to the state explosion problem. Recently, the MMDD method is used to assess reliability of MS-PMS with multi-state repairable components (Shrestha et al., 2011). The system MMDD model is generated by the general MMDD manipulation rules and all the path probabilities are evaluated by the CTMC. In the paper by (Shrestha et al., 2011), only the repairable MS-PMS is considered and only the first kind of the phase dependence (the state of component at the beginning of one phase should be identical to the state at the end of last phase) is considered. So the paths generated by this method may have the self-conflict problems in the non-repairable PMS. In this paper, a MMDD algorithm for the MS-PMS and PMS-MMDD model are proposed to evaluate system state probabilities of the MS-PMS. By the MMDD algorithm for the MS-PMS, the self-conflict paths can be cancelled automatically in the model generation process without any additional steps, which makes the system modelling more efficiently.

This paper is organized as follows. Sections 2 introduces the basic concepts of MMDD model. Section 3 introduces a simple example in detail. In section 4, the proposed MMDD algorithm and the model generation process by the PMS-MMDD model are introduced. The path probability evaluation method and the result certification are shown in section 5. The section 6 gives the conclusions and future works.

2 BASIC CONCEPTS OF MMDD MODEL

The MMDD model, an extension of the BDD, is proposed in (Shrestha et al., 2011) and applied to analyze the MSS and MS-PMS. Like BDD, MMDD is also based on the Shannon's decomposition and manipulation of the multi-valued logic function. The multi-valued logic function F of component A with m states is,

$$\begin{aligned} F &= case\left(A, F_1, F_2, \cdots\cdots, F_m\right) \\ &= A_1 \bullet F_{x_A=1} + A_2 \bullet F_{x_A=2} + \cdots\cdots + A_m \bullet F_{x_A=m} \end{aligned} \tag{1}$$

There are two kinds of nodes in the MMDD model: (1) the non-sink node, which is labeled with a multivalued variable x_A, which representing the behaviors of component A, and its corresponding multiple outgoing edges; (2) the sink node, '0' and '1', representing the system is in one specific state or not. And each disjoint path from the root node to the sink node '1' represent the system is in one specific state.

The generation of the MMDD model is based on the manipulation rule (Xing and Dai, 2009),

$$\begin{aligned} g \lozenge h &= case\left(x, G_1, \cdots G_m\right) \lozenge casey\left(y, H_1, \cdots H_m\right) \\ &= \begin{cases} case\left(x, G_1 \lozenge H_1, \cdots, G_m \lozenge H_m\right), index\left(x\right) = index\left(y\right) \\ case\left(x, G_1 \lozenge h, \cdots, G_m \lozenge h\right) \quad , index\left(x\right) < index\left(y\right) \\ case\left(y, g \lozenge H_1, \cdots, g \lozenge H_m\right) \quad , index\left(x\right) > index\left(y\right) \end{cases} \end{aligned} \tag{2}$$

where $\lozenge$ represent a logic operation (OR, AND) and the *index* represents the predefined variable orders.

3 A SIMPLE EXAMPLE

In this paper, the non-repairable MS-PMS is studied, which is common seen in the aerospace industry. A simple example system is used to show the proposed method. The system consists of three components, A, B and C and their state transition graphs are shown in Figure 1. The $\lambda_{i,j}$ represents the transition rate of the component transit from state i to state j. And the parameters are shown in Table 1.

The example PMS consists of three consecutive missions phases: τ_1, τ_2 and τ_3. Mission τ_1 needs component A in state 3 or component C in state 2; Mission τ_2 needs component A above state 2 or component B in state 3; Mission τ_3 needs component B above state 2 or component C in state 2. The entire mission is successful if all the three consecutive missions are completed successfully. The system structure function describing can be specified as,

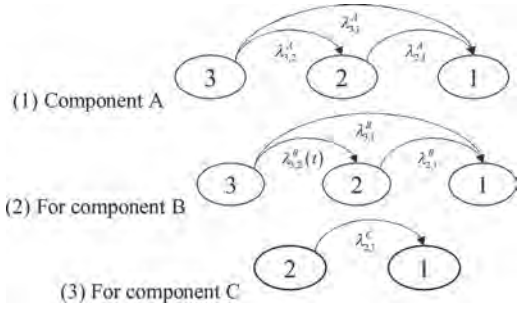


Figure 1. The state transition graphs for the components in the example system.

Table 1. The parameters of components in the example system.

	A	B	C
$\lambda_{3,2}$	1/100	1/120	
$\lambda_{3,1}$	1/150	1/160	
$\lambda_{2,1}$	1/200	1/200	1/160

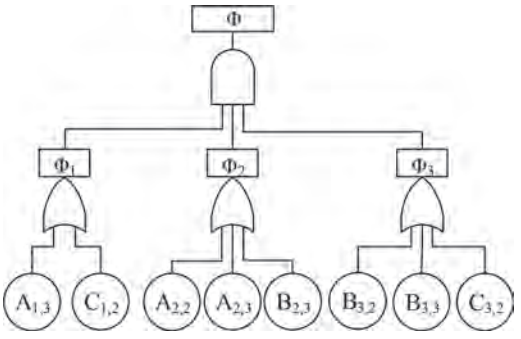


Figure 2. The MFT model for the example system.

$$\Phi = \Phi_1 \cdot \Phi_2 \cdot \Phi_3 = \left(A_{1,3} + C_{1,2}\right)\left(A_{2,2} + A_{2,3} + B_{2,3}\right)\left(B_{3,2} + B_{3,3} + C_{3,2}\right) \quad (3)$$

According to the descriptions of the system structure above, the multistate fault tree (MFT) model for the example system is shown in Figure 2.

Each of the bottom event in Figure 2 represents one component in one specific phase and state. For example, $A_{1,3}$ means component A stays in state 3 in phase 1.

4 MMDD MODEL FOR MS-PMS

4.1 *The MMDD algorithm for PMS*

In the paper (Shrestha et al., 2011), only the repairable MS-PMS is studied and the MMDD model is generated by the general MMDD manipulation rules, but part of paths have self-conflict problems in the non-repairable PMS. For example, in the path '$N_{10} \to N_9 \to N_4 \to N_1 \to 1$' of Figure 6 in the paper (Shrestha et al., 2011), the component c is in state 1 in phase 3 and in state 0 in phase 1 simultaneously, which cannot happen in the non-repairable MS-PMS. To address this problem, a MMDD algorithm for the MS-PMS extended from the BDD algorithm for the binary nonrepairable PMS (Zang et al., 1999) is proposed. Firstly, a phase algebra for the MS-PMS is proposed based on the phase dependency among phases in the non-repairable MS-PMS, shown in Table 2.

The physical meaning of the elements in Table 2 are,

1. $A_{i,m} \bullet A_{j,m} \to A_{j,m}$, component A is in the perfect state m both in phase i and phase j implies component A is in state m in phase j.
2. $A_{i,1} \bullet A_{j,1} \to A_{j,1}$, component A is in worst state 1 both in phase i and phase j implies component A is in worst state 1 in phase i.

Table 2. The phase algebra for the MS-PMS.

$A_{i,m} \bullet A_{j,m} \to A_{j,m}$	$A_{i,m} + A_{j,m} \to A_{i,m}$
$A_{i,1} \bullet A_{j,1} \to A_{i,1}$	$A_{i,1} + A_{j,1} \to A_{j,1}$
$A_{i,S_i} \bullet A_{j,S_j} \to 0$ $(1 \le S_i < S_j < m)$	$A_{i,S_i} + A_{j,S_j} \to 1$ $(1 \le S_j < S_i < m)$

3. $A_{i,S_i} \bullet A_{j,S_j} \to 0$ $(1 \le S_i < S_j < m)$, component A is in state S_i in phase i and is in state S_i in phase j simultaneously is not exist.
4. $A_{i,m} + A_{j,m} \to A_{i,m}$, component A is in the perfect state m in phase i or phase j implies component A is in state m in phase i.
5. $A_{i,1} + A_{j,1} \to A_{j,1}$, component A is in state 1 in phase i or phase j implies component A is in state 1 in phase j.
6. $A_{i,S_i} + A_{j,S_j}$, has no physical meaning if the components are not repairable.

4.2 *The MMDD operation for the MS-PMS*

With the above phase algebra for MS-PMS, the phase dependence operation (PDO) for the MS-PMS can be derived for both the forward PDO and backward PDO: (1) forward PDO, the order of the variables are the same as the phase order; (2) backward PDO, the order of the variables is reverse of the phase order (Zang et al., 1999).

Assume that component A with m states works both in phase i and phase j $(i < j)$, and the case format of component in phase i and phase j, E_i and E_j, can be written as,

$$E_i = case\left(A_i, \left(E_i\right)_{A_{i,m}}, \cdots, \left(E_i\right)_{A_{i,1}}\right) = case\left(A_i, G_m, \cdots, G_1\right) \quad (4)$$

$$E_j = case\left(A_j, \left(E_j\right)_{A_{j,m}}, \cdots, \left(E_j\right)_{A_{j,1}}\right) = case\left(A_j, H_m, \cdots, H_1\right) \quad (5)$$

where $A_{i,m}, G_m = 1$ means component A is in state m in phase j and $A_{i,m}, G_m = 0$ means component A is not in state m in phase j.

For the forward PDO $\left(index\left(A_i < A_j\right)\right)$, the proposed MMDD manipulation rule for the MS-PMS is,

$$case\left(A_i, G_m, \cdots, G_1\right) \Diamond case\left(A_j, H_m, \cdots, H_1\right) = case\left(A_i, G_m \Diamond E_{j,1}, G_{m-1} \Diamond E_{j,2}, \cdots, G_1 \Diamond H_1\right) \quad (6)$$

where

$$E_{j,n} = case\left(A_j, \underbrace{0, \cdots, 0}_{n-1}, H_{m-n+1}, \cdots, H_2, H_1\right), \ 1 \le n \le m$$

For the backward PDO $\left(index\left(A_i > A_j\right)\right)$, the proposed MMDD manipulation rule for the MS-PMS is,

$$\begin{aligned} &case\left(A_i, G_m, \cdots, G_1\right) \Diamond case\left(A_j, H_m, \cdots, H_1\right) \\ &= case\left(A_j, G_m \Diamond H_m, E_{i,2} \Diamond H_{m-1}, \cdots, E_{i,m} \Diamond H_1\right) \end{aligned} \tag{7}$$

where

$$E_{i,n} = case\left(A_i, G_m, G_{m-1}, \cdots, G_{m-n+1}, \underbrace{0, \cdots, 0}_{m-n} \right),\ 1 \le n \le m$$

5 SYSTEM EVALUATION

In this section, the system evaluation process by the proposed MMDD model and CTMC are shown in detail, with the following two steps: model generation and path probabilities evaluation.

5.1 *Model generation*

Firstly, the basic events of the MFT model in Figure 2 are transferred into MMDD basic events. For example, $case\left(A_i,1,0,0\right)$, $case\left(A_i,0,1,0\right)$, $case\left(A_i,0,0,1\right)$ represent component A in state 3, state 2 and state 1 in phase i, respectively. Then, with the MMDD basic events and the system structure shown in Figure 2, the MMDD models for each phase are generated by the general MMDD manipulation rules. The MMDD models for phases of the example system are shown in Figure 3. Then, the system MMDD model is generated by the proposed MMDD model for the MS-PMS, shown in Figure 4.

The grouped solid lines and dashed lines in Figure 4 represent the working states and failure states of the corresponding non-sink nodes. According to the definition of the MMDD model, each path from the root node to the sink node '1' represent that the system stays in one specific state in all phases. From Figure 4, we can get all the paths that represent the example system stays in the working state in all 3 phases. There are totally 9 paths, shown as,

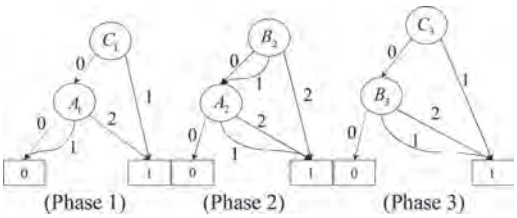


Figure 3. The MMDD model for each phase.

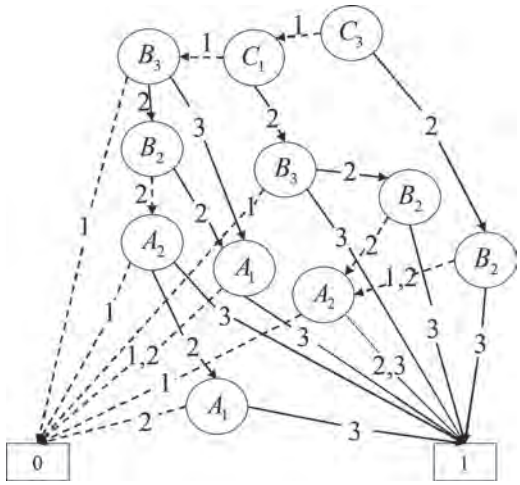


Figure 4. The grouped MMDD model for all three phases of the example PMS.

$$\begin{aligned} &\eta_1 = C_{3,(2)}B_{2,(3)},\ \eta_2 = C_{3,(2)}B_{2,(1,2)}A_{2,(2,3)} \\ &\eta_3 = C_{3,(1)}C_{1,(2)}B_{3,(3)},\ \eta_4 = C_{3,(1)}C_{1,(2)}B_{3,(2)}B_{2,(3)} \\ &\eta_5 = C_{3,(1)}C_{1,(2)}B_{3,(2)}B_{2,(2)}A_{2,(2,3)} \\ &\eta_6 = C_{3,(1)}C_{1,(1)}B_{3,(3)}A_{1,(3)} \\ &\eta_7 = C_{3,(1)}C_{1,(1)}B_{3,(2)}B_{2,(3)}A_{1,(3)} \\ &\eta_8 = C_{3,(1)}C_{1,(1)}B_{3,(2)}B_{2,(2)}A_{2,(3)} \\ &\eta_9 = C_{3,(1)}C_{1,(1)}B_{3,(2)}B_{2,(2)}A_{2,(2)}A_{1,(3)} \end{aligned} \tag{8}$$

5.2 *Path probability evaluation*

In this section, the probabilities of the disjoint paths generated in the last section are evaluated by the CTMC. All the components are independent on each other, so the probability of each path can be evaluated by multiplying the components' transition probabilities in the different phases. For example, the probability of path η_6 can be evaluated as,

$$\begin{aligned} \Pr\left(\eta_6 = 1\right) &= \Pr\left(C_{3,(1)}C_{1,(2)}B_{3,(3)}A_{1,(3)}\right) \\ &= \Pr\left(C_{1,(2)}C_{3,(1)}\right)\Pr\left(B_{3,(3)}\right)\Pr\left(A_{1,(3)}\right) \end{aligned} \tag{9}$$

In this paper, only the non-repairable components are studied, so the state of one component cannot be better in the latter phases and the state at the beginning of one phase should be equal to the end of last phase. Let S_j represent the state of one component in phase $i\left(1 \le i \le m\right)$. Due to the memoryless property of the homogeneous CTMC (Lisnianski and Levitin, 2003), the path probability of the component can be evaluated as,

$$\begin{aligned}&\Pr\left(S_1=c_1,S_2=c_2,\cdots S_m=c_m\right)\\&=\Pr_{N,c_1}\left(T_1\right)\Pr_{i_1,c_2}\left(T_1\right)\cdots\Pr_{i_{m-1},c_m}\left(T_m\right)\end{aligned}\tag{10}$$

where c_i means the specific state of component A at the end of phase i and $c_i\in[1,2,\cdots,N]$.

For example, in path η_6 component C is in state 2 in phase 1 and transit to state 1 in phase 3,

$$\Pr\left(C_{1,(2)}C_{3,(1)}\right)=p^c_{2,1}\left(T_1\right)\cdot p^c_{1,1}\left(T_3\right)\tag{11}$$

Component B is in state 3 in phase 3 and it works both in phase 2 and phase 3, so component B is in state 3 in both phase 2 and phase 3,

$$\Pr\left(B_{3,(3)}\right)=p^B_{3,3}\left(T_2+T_3\right)\tag{12}$$

Component A is in state 3 in phase 1,

$$\Pr\left(A_{3,(3)}\right)=p^A_{3,3}\left(T_1\right)\tag{13}$$

$\boldsymbol{P}(t)$, in which $P_{i,j}(t)$ represent the probability of component A transit from state i to state j during the time interval $[0,t]$. $p_{i,j}(t)$ can be evaluated by the CTMC with the state transition rate $\lambda_{i,j}$,

$$p^A_{i,j}(t)=\exp(\boldsymbol{Q}t)\tag{14}$$

$$\boldsymbol{Q}=\left|\lambda_{i,j}\right|\tag{15}$$

Then, the probability of path η_6 can be computed as,

$$\begin{aligned}\Pr\left(\eta_6=1\right)&=\Pr\left(C_{1,(2)}C_{3,(1)}\right)\Pr\left(B_{3,(3)}\right)\Pr\left(A_{1,(3)}\right)\\&=0.0228\end{aligned}\tag{16}$$

Similarly, the probabilities of other eight paths can be evaluated, and the reliability of the example system at the end of phase 3 can be computed as,

$$R_{sys}=\sum_{i=1}^{9}\Pr\left(\eta_i=1\right)=0.5819\tag{17}$$

5.3 *MC simulation validation*

In the previous sections, the reliability of the example PMS is evaluated by the PMS-MMDD model and the CTMC. To certify the correctness of these methods, the MC simulation method (Zio, 2013) is applied. The components' transition time and system failure time realizations simulation procedure is shown in Figure 5.

Then, the system reliability can be evaluated by statistically analyzing the system failure time realizations. In this paper, the amount of the realization is $N_{\max}=2\cdot10^6$. And the system reliability evaluated by the simulation method is,

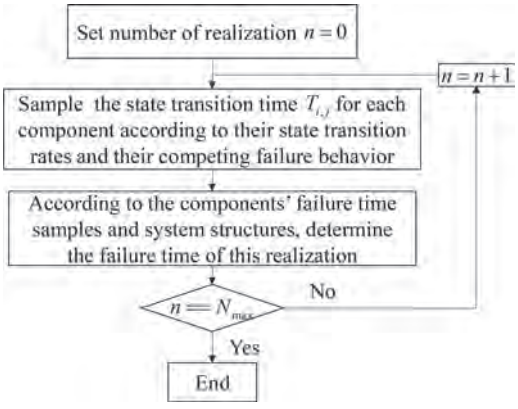


Figure 5. The grouped MMDD model for all three phases of the example PMS.

$$R^{sim}_{sys}=0.5817\tag{18}$$

Furthermore, the calculation time of the analytical approach and simulation approach are 0.07s and 74s, respectively, which illustrate that the analytical approach is more computationally efficient.

6 CONCLUSIONS AND FUTURE WORKS

In this paper, the non-repairable MS-PMSs are studied. In the non-repairable MS-PMS, the components' states cannot be better in the latter phases. To address this dependency in the system modelling, a MMDD algorithm for PMS and MMDD-PMS model are proposed to analyze the reliability of the MS-PMS with non-repairable components. By the proposed method, the self-conflict paths can be cancelled automatically and the system MMDD model can be generated without any additional steps. With the paths from the MMDD model, the system reliability is evaluated by adding all the path probabilities evaluated by the CTMC. At last, the result is certified by the MC simulation method and the comparison also illustrate the computational efficiency of the proposed analytical method.

Similar to the PMS-BDD model, the scale of the MMDD model is also heavily dependent on the predefined variable orders. As a part of the future works, the relationship between variable orders and the model scale of the PMS-MMDD model will be explored. On the other hand, the non-exponential transition time distribution is more general

in practice, the reliability evaluation considering non-exponential transition time distribution will be studied.

ACKNOWLEDGMENT

This research was supported by the Fundamental Research Funds for the Central Universities under the contract number ZYGX2014Z010.

REFERENCES

Lisnianski, A. & Levitin, G. 2003. *Multi-state system reliability: assessment, optimization and applications,* World Scientific Publishing Co Inc.

Liu, Y. & Huang, H.-Z. 2010. Reliability assessment for fuzzy multi-state systems. *International Journal of Systems Science,* 41, 365–379.

Liu, Y., Huang, H.-Z. & PHAM, H. Reliability evaluation of systems with degradation and random shocks. Reliability and Maintainability Symposium, 2008. RAMS 2008. Annual, 2008. IEEE, 328–333.

Ou, Y. & Dugan, J.B. 2004. Modular solution of dynamic multi-phase systems. *IEEE Transactions on Reliability,* 53, 499–508.

Shrestha, A., Xing, L. & Dai, Y. 2011. Reliability analysis of multistate phased-mission systems with unordered and ordered states. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans,* 41, 625–636.

Wu, X., Yan, H. & Li, L. 2012. Numerical method for reliability analysis of phased-mission system using Markov chains. *Communications in Statistics-Theory and Methods,* 41, 3960–3973.

Xing, L. & Amari, S.V. 2008. Reliability of phased-mission systems. *Handbook of performability engineering*, 349–368.

Xing, L. & Dai, Y. 2009. A new decision-diagram-based method for efficient analysis on multistate systems. *IEEE Transactions on Dependable and Secure Computing,* 6, 161–174.

Zang, X., Sun, N. & Trivedi, K.S. 1999. A BDD-based algorithm for reliability analysis of phased-mission systems. *IEEE Transactions on Reliability,* 48, 50–60.

Zio, E. 2013. *The Monte Carlo simulation method for system reliability and risk analysis,* Springer.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Reliability modeling for dependent competing failure processes between component degradation and system performance deterioration

Yugang Zhang, Jingyi Liu, Bifeng Song & Tianxiang Yu
School of Aeronautics, Northwestern Polytechnical University, Xi'an, China

ABSTRACT: Mechanical systems are usually subjected to multiple dependent competing failure processes. The processes were categorized two type of failures: soft failure caused by continuous component degradation together with additional abrupt degradation due to a shock process, and hard failure caused by the instantaneous stress from the same shock process. The models in open literature have simulated the aforementioned processes. However in practice, some mechanical systems have phenomena that all the components still in the good states, but the system functions cannot meet the requirements of performance. This paper studies reliability for multi-components system subject to dependent competing risks of components degradation, components random stress shocks and system performances deterioration. These failure processes are dependent in three respects: 1) the component stress shocks over certain critical threshold affects the component degradation process, 2) the component degradation impacts the component catastrophic failure threshold level, and 3) the component degradation influences the system performances. Based on component degradation, component random shocks and system performance modeling, a reliability model for systems considering three dependent failure patterns is derived. Then an application example using an electro-mechanical system illustrates the effectiveness of the proposed approach with sensitivity analysis. The results whether considering system performance deterioration or not are compared and discussed.

1 INTRODUCTION

The mechanical system is subjected to external random loads during the working process. At the same time there are internal wear, fatigue, aging degradation processes. The system is under multiple failure processes including catastrophic failures and degradation failures due to external shock loads and degradation processes. The two failure processes are dependent and competing against each other. Correlation means that there is a mutual influence between the two failures and competition means that any type of failure can result in system failure. The catastrophic failure mode (fracture, yielding, plastic deformation, etc. also called hard failure mode) and degradation failure mode (wear, corrosion, fatigue, and etc. also called soft failure mode) are competing with each other and whichever occurs first may make system fail. It is impossible that both risk modes occurs simultaneously due to the competing (Li & Pham 2005).

In the literature, there mainly exist five typical categories of random shock models and lots of mixed models based on them. An extreme shock model (Gut & Hüsler 1999, Cirillo & Hüsler 2011) supposes system failure occurs when the magnitude of any shock load exceeds a critical value. A cumulative shock model (Sumita & Shanthikumar 1985, Montoro-Cazorla & Pérez-Ocón 2015) assumes system failure occurs when the cumulative damage from extern shock loads exceeds a specified threshold. A run shock model (Mallor & Omey 2001, Eryilmaz 2016) supposes system failure occurs when there is a consecutive run of *m*-shocks exceeding a critical magnitude. A δ-shock model (Rangan & Tansu 2010, Wang & Peng 2017) assumes system failure occurs when the time interval between two successive shocks is less than a specified threshold δ. A mixed shock model (Gut 2001, Rafiee et al. 2015, 2017) includes two or more basic shock models in which the system is supposed to fail either because of one large shock, or as a result of many smaller ones.

Gut et al. expanded the shock model further and considered that some shocks may harm the system and lead to the threshold change. Accordingly, the failure rate would also change (Gut & Hüsler 2005). Jiang et al. supposes shocks with different magnitudes might have different impacts on the failure of the system and proposed a zone shock model (Jiang et al. 2015).

The degradation models can be classified into disperse process and continuous process. Xue et al. extended 2-state reliability analysis method to multi-state. The system multi-state reliabilities are analyzed by combining Markov processes and

s-coherent multistate system structure functions (Xue & Yang 1995). Li et al. developed a generalized multi-state degraded system reliability model which operating condition of the multi-state systems is characterized by a finite number of states (Li & Pham 2005). Zhang et al. introduced multiple discrete function theory considering the monotone and coherence of the multi-state system to describe the structure function of system state. The law of Demogen and a new block diagram algorithm are developed to simplify the expression for the system reliability (Zhang et al. 2017).

For continuous degradation model, there are two categories to describe the process: degradation path model and stochastic process model.

Bagdonavičius et al. used a general nonparametric, nonlinear path model to simulate degradation process in which supposing that the intensities of failures depend on degradation level (Bagdonavičius et al. 2005). Xu et al. proposed a class of general path models to incorporate time-varying usage and environmental variables for modeling of degradation paths. They used nonlinear functions to describe the degradation paths, and random effects to describe unit-to-unit variability (Xu et al. 2016).

Yan et al. proposed an improved time-variant reliability model considering stochastic process of strength degradation based on semi-stochastic-process. The parameters are determined based on the model of Gamma degradation process with independent increments. This method can avoid the redundant and complex calculations and obtain accurate reliability results (Yan et al. 2016). Huang et al. proposed an adaptive skew-Wiener model to simulate the degradation drift of industrial devices and predict the remaining useful life. A two-stage algorithm is adopted to estimate unknown parameters as well (Huang et al. 2017).

There may exist competing between system catastrophic failure and degradation failure.

Peng et al. developed a reliability model based on degradation and random shock modeling for systems subject to Multiple Dependent Competing Failure Processes (MDCFP). Two dependent failure processes, soft failures and hard failures, were considered in the model. Hard failure occurs when the size of external shock exceeds its threshold. Degradation is composed by pure degradation and instantaneous increase lead by shocks. Soft failure happens if the degradation exceeds its specified threshold. The two failures are competing with each other based on which the system reliabilities are analyzed (Peng et al. 2010). Jiang et al. supposed that the system withstanding shocks is deteriorating, and its resistance to failure is weakening. They introduced shift failure thresholds to the competing failure model (Jiang et al. 2012). Similarly, Rafiee proposed a reliability model for systems subject to dependent competing failure processes of degradation and random shocks with a changing degradation rate. The degradation rate will change when the system becomes more susceptible to fatigue and deteriorates faster (Rafiee et al. 2014). Besides, Song studied the reliability of multi-components system subject to multiple dependent competing failure processes considering dependent shock damage on failure processes among components. They developed reliability models considering different patterns for dependent shock damage for more realistic and accurate prediction of system reliability (Song et al. 2016).

There are many other researchers studied the competing failure models. For all we know, the degradation threshold is only for the soft failure of the component. There is no constraint on system performance parameters, supposing that system failure occurs when component's degradation exceeding a critical magnitude. This assumption is reasonable for a system composed of structural parts, but there are some problems for the system contains mechanisms.

Firstly, that a component of system occurs soft failure does not mean system must fail. For example, a simple slider crank mechanism contains frame, crank, connecting rod, and slider. Usually, there exist multiple wear degradation processes in the kinematic pairs of the mechanism. The performance parameters of the system are affected by each wear degradation process. When a wear degradation value has exceed its specified threshold, the performance parameter value of the system may exceed the system requirement, or it may not exceed the requirement and be in a normal state. The system's state is only determined by the performance of the system.

Secondly, some mechanisms have phenomena that all the components still in the good states, but the system functions cannot meet the requirements of performance. The degradation of wear may increase the coefficient of friction. Although the soft failure of all components of the system has not occurred, the increase of friction coefficient can lead to stagnation failure of the mechanism.

Finally, the degradation of some system components is difficult to measure directly, especially for mechanism, but the performance value of the mechanism output is easy to acquire.

Based on the above problem, we will study the reliability for multi-components system (especially for mechanism) subject to dependent competing degradation of components, random stress shocks of components and performances deterioration of system. A reliability model for systems considering three dependent failure patterns is derived based on component degradation, component random shocks and system performance modeling. Then an application example will illustrate the effectiveness of the proposed approach with sensitivity analysis. The results whether considering system performance deterioration or not will be compared and discussed.

2 SYSTEM FAILURE DESCRIPTION

As shown in Figure 1, the component of a mechanical system are in the process of degradation. The degradation trajectory may be linear or non-linear. In addition, the stress shocks of component *i* over certain critical threshold will affect the component degradation process with additional abrupt degradation damage. At the same time, the degradation of component will impact the component hard failure threshold level. So the hard failure threshold of the component will degrade accordingly. When the strength of the component is less than the load stress, the component catastrophic failure will occur.

For a mechanism, the degradation of components will influence the system performances based on topological relationship of mechanical composition. As shown in Figure 2, the system performance of the mechanism will degrade according to the degradation value of component 1, 2, …, *n* and their topological function. When the system performance exceeds the threshold range, the system performance degradation failure will occur.

The component catastrophic failure and the system performance degradation failure are dependent because of the same load stress environment. There also exists a competition for the two types of failures. Any type of failure occurs, the mechanism will fail.

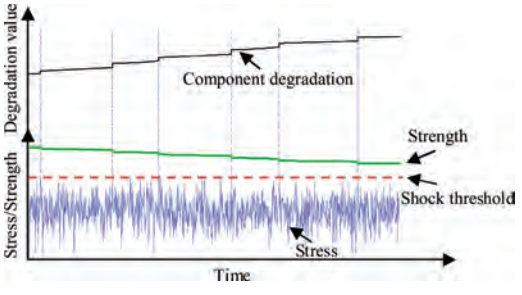


Figure 1. Dependent degradation processes for component i.

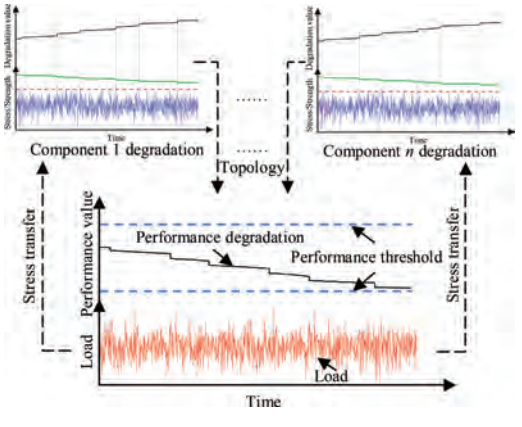


Figure 2. Dependent competing failure processes between components' catastrophic failure and mechanism's performance degradation failure.

3 RELIABILITY MODELLING FOR MULTI-COMPONENTS SYSTEM

3.1 *System load stress analysis*

Typically speaking, the mechanism system is always under its necessary working loads. However, the magnitude of loads are not fixed values due to several reasons including randomness of shock, vibration of system and etc. The system load will be transferred to component level by specific ways. The transmitting ways are determined by the basic structure form and total degradation of all the components for mechanical system.

Supposing the transfer laws is expressed by

$$S_i = f_i(B, \boldsymbol{L}, \boldsymbol{F}, t) \qquad i = 1,2,\cdots,n \tag{1}$$

where S_i is the stress size of the *i*th component; $f_i()$ is the transfer function between the *i*th component stress and system loads; B indicates the basic structure form of the mechanical system; $\boldsymbol{L}$ is the total degradation vector of all the components; $\boldsymbol{F}$ means the system load spectrums vector; t is time variable.

The transfer laws may be explicit or implicit form. For a simple mechanism can get explicit transfer function, the analytical method can be used to get component level stress. For a complex mechanism cannot get explicit transfer function, we can use multi flexible body dynamics method to get the stress.

The oversized stress is more attractive because the larger the stress, the more dangerous the component. A larger stress may cause the component additional abrupt degradation.

Due to the randomness of system external loads, the component stress with magnitude larger than a fixed threshold is subjected to a Poisson process. Concretely, the *i*th component stress shocks arrive with a Poisson process $\{N_i(t)|S_i(t) > H_i, t \geq 0\}$ with rate λ_i. The probability that the *i*th component stress shocks equal k is expressed as

$$P(N_i(t) = k) = \frac{\exp(-\lambda_i t)(\lambda_i t)^k}{k!} \tag{2}$$

where H_i is the abrupt degradation threshold of component *i*.

The magnitude of component stress larger than the fixed threshold can be statistically obtained. The type and parameters of stress shocks distribution may obtain by using the maximum likelihood method.

3.2 *Reliability modeling for component*

Components are subjected to two kinds of processes, the catastrophic failure process and degradation process as we mentioned above. The catastrophic failure happens when the size of stress of component exceeds its threshold. The threshold is changing with time. It is related to the degradation of the component.

3.2.1 *Degradation process modeling*

The component degradation is consisted by two parts, the pure degradation $X_i(t)$ and abrupt degradation Z_i. The exponential degradation path is used to model the pure degradation. The degradation path is expressed by

$$X_i(t) = \varphi_i + \gamma_i t^{\zeta} + \varepsilon(t) \tag{3}$$

where φ_i is the initial value; γ_i and ζ are the degradation parameters; ε is a standard Brown motion $\varepsilon \sim B(0, \sigma^2 t)$.

The cumulative abrupt degradation Z_i caused by random stress shocks is

$$Z_i(t) = \sum_{j=1}^{N_i(t)} Y_{ij} \tag{4}$$

where $N_i(t)$ is the total number that the *i*th component stress shocks occur; Y_{ij} is the abrupt degradation value of component *i* stress shock *j*.

The abrupt degradation value Y_{ij} is dependent on the component stress shocks. If the shock is larger, the abrupt degradation value is more likely deteriorating. We assume a positive linear correlation between the abrupt degradation value and the component stress shock. Their relation is expressed as

$$Y_{ij} = \alpha_i S_{ij} \tag{5}$$

where S_{ij} is the component stress shocks greater than H_i until time *t*. Which can be determined by equation (2). α_i is the degradation conversion coefficient.

Thus, the total degradation of component *i* is expressed by

$$\begin{aligned} L_i(t) &= X_i(t) + Z_i(t) \\ &= \varphi_i + \gamma_i t^{\zeta} + \varepsilon(t) + \sum_{j=1}^{N_i(t)} \alpha_i S_{ij} \end{aligned} \tag{6}$$

3.2.2 *Catastrophic failure process modeling*

Catastrophic failure occurs once a stress of component exceeds its strength threshold. The probability that no catastrophic failure happens is

$$P(S < D_i) = F_S\left(D_i(t)\right) \tag{7}$$

where S is a random variable of component stress; $D_i(t)$ is the current strength threshold which changes with time due to the degradation of the component.

The current strength threshold is related to the total degradation of the component. The greater the total degradation of the component, the smaller the residual strength is. So we suppose the current strength threshold is a linear function of the total degradation of the component, as showed in Figure 3.

$$D_i(t) = D_0 + \beta_i L_i(t) \tag{8}$$

where β_i is the threshold conversion coefficient; D_0 is initial threshold value.

The randomness of Poisson process indicates that there may be any shock times, from 0 to ∞, until time *t*. For a better illustration, there is an assumption that there are *j* shocks until time *t*. Besides, the coming moment of the *j*th shock is indicated by T_j. Thus, the probability that no catastrophic failure occurs for the *i*th component is expressed by

$$\begin{aligned} &F_S\left(D_i(t)\right) = P\left(N_i(t) = 0\right) \\ &+ \sum_{j=1}^{\infty} \int_0^t \int_{T_1}^t \cdots \int_{T_{j-1}}^t \left(\prod_{k=1}^{j} P\left(S_{ik} < D_i(T_k)\right) \middle| \right. \\ &\left. N_i(t) = j, t_1 = T_1, t_2 = T_2, \cdots, t_j = T_j \right) \cdot P\left(N_i(t) = j\right) \end{aligned} \tag{9}$$

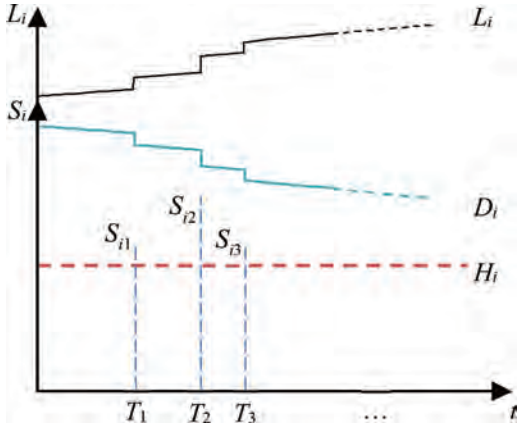


Figure 3. Catastrophic failure process for component *i*.

3.3 *Reliability modeling for system*

3.3.1 *System performance degradation modeling*

Mechanism is designed for specific performance such as load transmission and displacement shifting. The achievement of system performance is combined action of all the units of including all components, all pairs, all loads and etc.

Typically speaking, the system performance value is a function of relevant units:

$$P_e = f_e\left(B, \mathbf{L}, \mathbf{F}, t\right) \tag{10}$$

where P_e is the system performance value; $f_e()$ is the function between the components parameters and system performance.

This function is the same as the system load transfer function (1). Its result needs to be solved iteratively with the load transfer function.

3.3.2 *System performance degradation failure modeling*

As we have analyzed above, there are some degradation in components. The degradation may deduce the system performance deterioration. In other words, the performance value will also degrade due to the degradation of components. Thus, performance value is also a variable of time t which is expressed by $O_s(t)$.

If the mechanical system performance wants to meet requirements, the performance value need to conform to the ideal value. Namely, the error of performance value and its ideal value is under a threshold. The performance reliability is indicated by

$$P_s = P(\left|O_s - O_{si}\right| < \delta) \tag{11}$$

where O_{si} is the ideal value of performance and δ is the maximum allowable performance error.

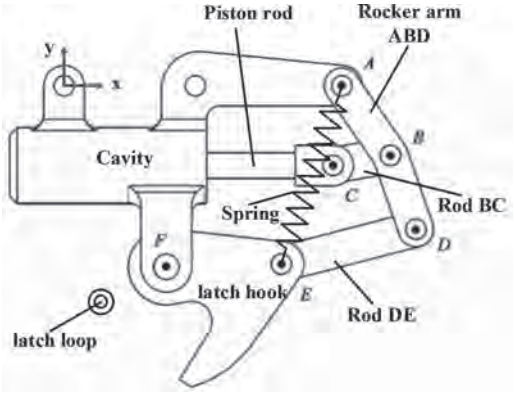


Figure 4. The composition of GDLM.

3.4 *System reliability modeling*

All in all, mechanical system may fail due to two kinds of failure types, the catastrophic failure of components and the degradation failure of system performance. This two failures are dependent and competing with each other.

If the system keeps in good state, neither the catastrophic failure nor the performance failure can happen. The reliability is expressed by

$$R_s = P_s \cdot \prod_{i=1}^{n} F_S\left(D_i(t)\right) \tag{12}$$

4 NUMERICAL EXAMPLES AND RESULTS

An aircraft gear door lock mechanism (GDLM) example is used to illustrate the competing reliability model. The GDLM consists of seven parts, which are a cavity, piston rod, latch hook, rocker arm ABD, rod BC, rod DE and spring, as illustrated in Figure 4. They are connected by six revolute joints (A, B, C, D, E and F). In the opening process, the piston rod experiences the pressure of the hydraulic oil moves towards the right. The rocker arm ABD rotates anti-clockwise. Meanwhile, the latch hook rotates anti-clockwise until the latch hook is separated from a latch loop. The opening process is investigated in this paper. There are mainly two kinds of failure modes in the opening process: component fracture and mechanism seizure.

The latch hook experiences the gear door load from the latch loop when the mechanism is in closed state. In the opening process, the latch hook rotates anti-clockwise. The friction force and friction moments are caused by the joints. The friction force and friction moments are determined by the contact stress and friction coefficients. The magnitude of the contact stress can be determined by the dynamics of the mechanism.

When the driving force from the piston rod cannot overcome the resistance from the mechanism, the seizure of the mechanism will happened.

Hence, the failure criterion for seizure is:

$$O_s - O_{si} < 0 \tag{13}$$

where O_s donates the maximum force that the piston rod can provide; O_{si} presents the resistance of the mechanism, which is 2700 N.

It is difficult to present the explicit expression for the performance function O_s due to the complexity of the GDLM and the diversity of the influence factors. Therefore, based on Virtual.Lab software, a multi-body dynamic simulation model of the

GDLM is established. The relationship between the unlocking force O_s and the influence factors is constructed by the response surface method.

$$\begin{aligned} O_s = & -4416.10 + 187.16X_1 + 0.3378X_2 + \\ & 1382.69X_3 + 0.0617X_4 - 2.02X_1^2 + 4.9339 \times 10^{-8} \\ & X_2^2 - 261.53X_3^2 + 2.7067 \times 10^{-9} \\ & X_4^2 - 0.00239X_1X_2 - 23.93X_1X_3 - 3.9136 \\ & \times 10^{-4} X_1X_4 + 0.3236X_2X_3 - 1.062 \times 10^{-6} \\ & X_2X_4 - 0.0238797X_3X_4 \end{aligned} \tag{14}$$

where X_1, X_2, X_3, X_4 are influence factors, the meanings and values of the parameters are listed in Table 1.

Typically, total load is distributed to each component according to the construction of mechanism system. For this GDLM, the load distributed to each component can be obtained by dynamic analysis. For simplicity, we assume that the proportion of stress per component is constant. The value is listed in following Table 2.

According to equation (9), (11), (12), the reliability of GDLM is calculated at various time. Numerical method based on Monte Carlo is used for calculation due to the lack of analytic solution.

The reliability and failure probability density results are showed in Figures 5 and 6. The x-label is the working time which is stated by number of working circles. The y-label is reliability value and failure probability density respectively for GDLM with performance deterioration considered.

From Figure 6 we can obtain that component catastrophic failure is less probable than system performance failure. The system total reliability is dependent on the competing results of component catastrophic failure and system performance failure. When considering system performance deterioration, the reliability reducing rate is quite fast at high reliability level.

This appearance manifest that the deterioration influences equipment with high reliability seriously. For mechanical system requiring precision and dependability, it is of vital importance to control the deterioration and accident shock load.

Figure 7 presents sensitivity analyses of latch hook friction coefficient on system total reliability. From Figure 6, by increasing the value (γ_3 increases from 3×10^{-7} to 7×10^{-7}), $R(t)$ shifts to the left. At

Table 1. Values of the parameter used in the GDLS reliability analysis.

Parameters	Distribution	Mean	Standard deviation	Degradation parameters
Angle when latch hook contact loop X_1	normal	45°	0.5°	
Latch hook load shock X_2	normal	10000 N	300 N	$\lambda = 2.5 \times 10^{-5}$;
Latch hook friction coefficient X_3	normal	0.05	0.001	$\varphi_3 = 0$; $\alpha_3 = 1.0$; $\gamma_3 = 5 \times 10^{-7}$; $\zeta_3 = 1.05$; $\sigma_3 = 5.5 \times 10^{-5}$
Stiffness of spring X_4	normal	7558.6 N/m	378 N/m	$\varphi_4 = 0$; $\alpha_4 = -1.0$; $\gamma_4 = -9 \times 10^{-4}$; $\zeta_4 = 1.05$; $\sigma_4 = 0.4$
Threshold of joint A of rock arm X_5	normal	1.2 GPa	0.06 GPa	$\beta_5 = -1.0$; $\alpha_5 = 0.1$; $\gamma_5 = 3.2 \times 10^{-7}$; $\varphi_5 = 0$; $\zeta_5 = 1.05$; $\sigma_5 = 5.5 \times 10^{-5}$
Threshold of joint B of rod BC X_6	normal	1.2 GPa	0.06 GPa	$\beta_6 = -1.0$; $\alpha_6 = 0.1$; $\gamma_6 = 1.2 \times 10^{-7}$; $\varphi_6 = 0$; $\zeta_6 = 1.05$; $\sigma_6 = 5.5 \times 10^{-5}$
Threshold of joint C of rod BC X_7	normal	1.2 GPa	0.06 GPa	$\beta_7 = -1.0$; $\alpha_7 = 0.1$; $\gamma_7 = 1.2 \times 10^{-7}$; $\varphi_7 = 0$; $\zeta_7 = 1.05$; $\sigma_7 = 5.5 \times 10^{-5}$
Threshold of joint D of rock arm X_8	normal	1.2 GPa	0.06 GPa	$\beta_8 = -1.0$; $\alpha_8 = 0.1$; $\gamma_8 = 2.8 \times 10^{-7}$; $\varphi_8 = 0$; $\zeta_8 = 1.05$; $\sigma_8 = 5.5 \times 10^{-5}$
Threshold of joint E of latch hook X_9	normal	1.2 GPa	0.06 GPa	$\beta_9 = -1.0$; $\alpha_9 = 0.1$; $\gamma_9 = 2.8 \times 10^{-7}$; $\varphi_9 = 0$; $\zeta_9 = 1.05$; $\sigma_9 = 5.5 \times 10^{-5}$
Threshold of joint F of latch hook X_{10}	normal	1.2 GPa	0.06 GPa	$\beta_{10} = -1.0$; $\alpha_{10} = 0.1$; $\gamma_{10} = 5.6 \times 10^{-7}$; $\varphi_{10} = 0$; $\zeta_{10} = 1.05$; $\sigma_{10} = 5.5 \times 10^{-5}$

Table 2. Ratio of each component stress.

Components	Joint A	Joint B	Joint C	Joint D	Joint E	Joint F
Stress ratio	0.4	0.15	0.15	0.35	0.35	0.7

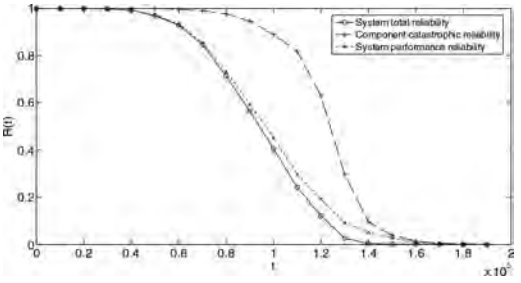


Figure 5. The reliability of GDLM at various time.

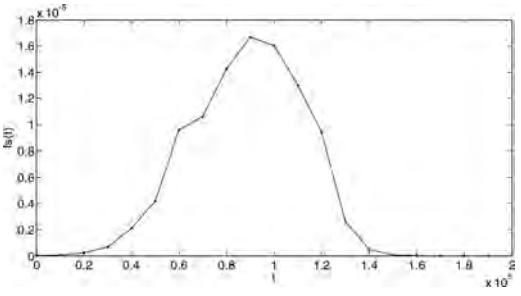


Figure 6. The failure probability density of GDLM.

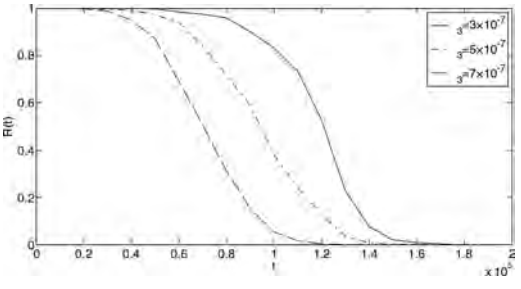


Figure 7. Sensitivity analysis of $R(t)$ on latch hook friction coefficient.

reliability level 0.8, The number of working circles decreases from 1.04×10^5 to 0.55×10^5 by nearly half.

It can be inferred that increased degradation coefficient γ_3 decreases the system reliability. For the comparisons, the degradation rate parameters have a very obvious impact on the reliability of the system.

5 CONCLUSIONS

The reliability model for multi-components system subject to dependent competing risks of components catastrophic failure and system performances deterioration failure has been developed. These failure processes are dependent in three respects: 1) the component stress shocks over certain critical threshold affects the component degradation process, 2) the component degradation impacts the component catastrophic failure threshold level, and 3) the component degradation influences the system performances.

This reliability model mainly for mechanism considered the relationship between the system and its component. Through this relationship and the degradation models of components, the system performance deterioration failure model has been proposed. The system performance deterioration failure process and the components catastrophic failure process are dependent and competing. The system total reliability is the competing result of the above two processes.

It is more realistic and difficult due to the failure processes being dependent not only due to a common load but also due to constraints between components. We presented a numerical example to demonstrate the effectiveness of the proposed approach and observe the reliability with sensitivity analysis for an aircraft gear door lock mechanism. The results whether considering system performance deterioration or not are compared and discussed.

For more accurate, the reliability model need consider the actual latch hook friction coefficient decreasing rate. This may be extended for future.

ACKNOWLEDGEMENT

This study was based in part upon work supported by the Fundamental Research Funds for the Central Universities grant No.3102015 BJ (II)JL01 and the National Natural Science Foundation of China grant No.51675428.

REFERENCES

Bagdonavičius, V., Haghighi, F. & Nikulin M. 2005. Statistical analysis of general degradation path model and failure time data with multiple failure modes. *Communications in Statistics – Theory and Methods*, 34(8):1771–1791.

Cirillo, P. & Hüsler, J. 2011. Extreme shock models: An alternative perspective. *Statistics & Probability Letters*, 81(1):25–30.

Eryilmaz, S. 2016. Computing optimal replacement time and mean residual life in reliability shock models. *Computers & Industrial Engineering*, 103(1):40–45.

Gut, A. & Hüsler, J. 1999. Extreme shock models. *Extremes*, 2(3):295–307.

Gut, A. & Hüsler, J. 2005. Realistic variation of shock models. *Statistics & Probability Letters*, 74(2):187–204.

Gut, A. 2001. Mixed shock models. Bernoulli, 7(3):541–555.

Huang, Z., Xu, Z. & Ke, X. 2017. Remaining useful life prediction for an adaptive skew-Wiener process model. *Mechanical Systems & Signal Processing*, 87 A:294–306.

Jiang, L., Feng, Q. & Coit, D.W. 2012. Reliability and maintenance modeling for dependent competing failure processes with shifting failure thresholds. *IEEE Transactions on Reliability*, 61(4):932–948.

Jiang, L., Feng, Q. & Coit, D.W. 2015. Modeling zoned shock effects on stochastic degradation in dependent failure processes. *IIE Transactions*, 47(5):460–470.

Li, W. & Pham, H. 2005. Reliability modeling of multi-state degraded systems with multi-competing failures and random shocks. *IEEE Transactions on Reliability*, 54(2):297–303.

Mallor, F. & Omey, E. 2001. Shocks, runs and random sums. *Journal of Applied Probability*, 38(2):438–448.

Montoro-Cazorla, D. & Pérez-Ocón, R. 2015. A reliability system under cumulative shocks governed by a BMAP. *Applied Mathematical Modelling*, 39(23–24):7620–7629.

Peng, H., Feng, Q. & Coit, D.W. 2010. Reliability and maintenance modeling for systems subject to multiple dependent competing failure processes. *IIE Transactions*, 43(1): 12–22.

Rafiee, K., Feng, Q. & Coit, D.W. 2014. Reliability modeling for dependent competing failure processes with changing degradation rate. *IIE Transactions*, 46(5): 483–496.

Rafiee, K., Feng, Q. & Coit, D.W. 2015. Condition-based maintenance for repairable deteriorating systems subject to a generalized mixed shock model. *IEEE Transactions on Reliability*, 64(4):1164–1174.

Rafiee, K., Feng, Q. & Coit, D.W. 2017. Reliability assessment of competing risks with generalized mixed shock models. *Reliability Engineering & System Safety*, 159:1–11.

Rangan, A. & Tansu, A. 2010. Some results on a new class of shock models. *Asia-Pacific Journal of Operational Research*, 27(4):503–515.

Song, S., Coit, D.W. & Feng, Q. 2016. Reliability analysis of multiple-component series systems subject to hard and soft failures with dependent shock effects. *IIE Transactions*, 48(8): 720–735.

Sumita, U. & Shanthikumar J.G. 1985. A class of correlated cumulative shock models. *Advances in Applied Probability*, 17(2):347–366.

Wang, G.J. & Peng, R. 2017. A generalised δ-shock model with two types of shocks. *International Journal of Systems Science Operations & Logistics*, 4(4):372–383.

Xu, Z., Hong, Y. & Jin, R. 2016. Nonlinear general path models for degradation data with dynamic covariates. *Applied Stochastic Models in Business & Industry*, 32(2):153–167.

Xue, J. & Yang, K. 1995. Dynamic reliability analysis of coherent multistate systems. *IEEE Transactions on reliability*, 44(4): 683–688.

Yan, M., Sun, B. & Li, Z. 2016. An improved time-variant reliability method for structural components based on gamma degradation process model// *Prognostics and System Health Management Conference (PHM-Chengdu)*, Chengdu, 1–6. doi: 10.1109/PHM.2016.781980.

Zhang, Y.J., Sun, Y.C. & Zhang, Y.J. 2017. Reliability analysis for multi-state coherent system with monotonic components based on pivotal boundary points of clustering states. *Acta Aeronautica et Astronautica Sinica*, 38(8): 220868.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Quantitative reliability analysis method for repairable systems with multiple correlations based on goal-oriented method

X.J. Yi & B. Xu
China North Vehicle Research Institute, Beijing, China
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

J. Shi
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

P. Hou & H.N. Mu
Beijing Institute of Technology, Beijing, China

ABSTRACT: This paper proposes a new quantitative reliability analysis method for repairable systems with multiple correlations based on Goal Oriented (GO) method. First, the solving methods for repairable systems with shutdown correlation, maintenance correlation, standby correlation and common cause failure in GO method are presented. And the reliability analysis process of this paper's GO method is formulated. Then, the hydraulic oil supply system of heavy vehicle is taken as an example to conduct reliability analysis by this paper's GO method. Finally, in order to verify the feasibility and advantage of the proposed GO method, its analysis result compared with those by GO method for system without considering correlations and system with considering a single kind correlation. All in all, this study not only improves the theory GO methodology; but also provides a new reliability analysis approach for repairable systems with multiple correlations.

1 INTRODUCTION

The reliability of repairable systems is a prerequisite for its normal operating. Correlations are universal characteristic in the repairable system, and affect the system reliability directly, such as shutdown correlation, maintenance correlation, standby correlation and common cause failure. If the repairable systems with correlations are not considered these correlations, the reliability analysis result will have a large bias. Different from Fault Tree Analysis (FTA), Failure Mode, Effects Analysis (FMEA), and Monte-Carlo Simulation (MCS), Goal Oriented (GO) methodology is a success-oriented method for system reliability analysis. It has three obvious advantages [1], as follows:

i. GO model is directly developed according to system principle diagram, flowchart or engineering drawing, so it is more objective.
ii. GO method can combine with other technologies to improve the GO method easily so that it can solve various kinds of practical engineering problems, such as multi-fault modes [2–4], Closed-Loop Feedback [5–7], multifunction [8, 9], multi-state [10–12], and so on.
iii. Both of the accurate quantitative analysis result and qualitative analysis result can be obtained by GO method.

GO method has become increasingly popular in recent years because of its advantages in aspects of establishing system model and its stronger analysis power. Indeed, a large number of engineering applications have established its value [1, 13]. It is an efficient technology approach to conduct reliability analysis [14], reliability optimization allocation [15, 16] and reliability assessment [17, 18]. Although the basic theory of GO method has been improved so that it can solve some single correlation, such as standby correlation [19], the existing GO methods are not suitable for repairable systems with multiple correlations, which are shutdown correlation, maintenance correlation, standby correlation and common cause failure. Above all, the main contributions of this study are as follows:

i. The new quantitative reliability analysis method for repairable systems with multiple correlations, which are redundancy structure with shutdown correlation, maintenance correlation, standby correlation and common cause

failure based on GO method, is expounded in detail.

ii. The reliability analysis process of this paper's GO method is formulated.

iii. The hydraulic oil supply system of heavy vehicle is taken as an example to conduct reliability analysis by the proposed GO method firstly.

The remainder of the paper is organized as follows. The GO method for quantitative reliability analysis of repairable systems with multiple correlations is proposed in Section 2. Section 3 illustrates a practical example, which is a hydraulic oil supply system of heavy vehicle, based on the proposed GO method. Section 4 provides results analysis in order to verify the feasibility, advantage and reasonability of this paper's GO method. Section 5 provides some conclusions on the findings of the research.

2 GO METHOD FOR REPAIRABLE SYSTEMS WITH MULTIPLE CORRELATIONS

The quantitative reliability analysis result by using GO method is obtained based on GO model by adopting GO algorithm according to the reliability analysis process of GO method. So, the GO algorithm and the reliability analysis process of GO method are the key factors for conducting GO analysis. In this section, the GO algorithm for redundancy structure with multiple correlations and the reliability analysis process of this paper's GO method are proposed, respectively.

2.1 *GO algorithm for repairable systems with multiple correlations*

The redundancy structures are often used in complex repairable systems, and the shutdown correlation, maintenance correlation, standby correlation and common cause failure affect the reliability of such structure directly.

1. GO algorithm for redundancy structure with shutdown correlation, maintenance correlation and standby correlation

In quantitative GO analysis, the redundancy structure with shutdown correlation, maintenance correlation and standby correlation is equivalent to a unit, which is represented by Type 1 operator, which is used to describe two-state unit. And the reliability parameters of such unit are obtained by GO algorithm for redundancy structure with shutdown correlation, maintenance correlation and standby correlation [20], as shown in Eq. (1)–(5).

$$\begin{cases} a_i = \begin{cases} (M-i+1)\lambda, & J=0 \text{ OR } \quad M-i+1<K \\ K\lambda, & J=1 \text{ OR } \quad M-i+1\geq K \end{cases} \\ b_i = \begin{cases} i\mu, & i \leq L \\ L\mu, & i > L \end{cases} \end{cases} \tag{1}$$

$$P_i = P_0 \prod_{j=1}^{i} \frac{a_j}{b_j} \tag{2}$$

$$P_R = \sum_{i=0}^{M-K} P_i \Big/ \sum_{i=0}^{I} P_i \tag{3}$$

$$\lambda_R = P_{m-k} \cdot a_{m-k+1} \Big/ \sum_{i=0}^{M-K} P_i \tag{4}$$

$$\mu_R = P_{M-K+1} b_{M-K+1} \Big/ \sum_{i=M-K+1}^{I} P_i \tag{5}$$

where M is the unit number of redundancy structure, K is the operating unit number of redundancy structure, I is the faulting unit number of redundancy structure, J is the standby indicator, $J = 0$ represents no standby unit in redundancy structure, $J = 0$ represents the redundancy structure has standby units, P_0 is the success probability of redundancy structure at the condition of all unit operating, P_i is the state probability of redundancy structure, P_R, λ_R and μ_R are the success probability, failure rate and maintenance rate of equivalent unit for redundancy structure, respectively.

2. GO algorithm for repairable systems with common cause failure

In quantitative GO analysis, the quantitative analysis result can be obtained by using GO algorithm for repairable systems with common cause failure [21], as shown in Eq. (6).

$$R_S = R_I + \sum_{m=1}^{M} C_m (R_{00\ldots} - R_{11\ldots}) \tag{6}$$

where R_S is the system availability by GO operation according to the basic GO algorithm [22], R_I is the system availability without considering common cause failure by GO operation according to the basic GO algorithm [22], C_m is the common cause failure probability of *mth* redundancy structure with common cause failure, $m = 1, 2, \cdots, M$, $R_{00\ldots}$ and $R_{11\ldots}$ are the system availabilities at the situation of the availabilities of all units in each redundancy structure with common cause failure as 0 and 1, respectively.

2.2 *Reliability analysis process of proposed GO method*

The reliability analysis process is the criterion and prerequisite for conducting quantitative GO

analysis. While, the reliability analysis process of basic GO method is not suitable for repairable systems with multiple correlations [23], so the reliability analysis process of repairable systems with multiple correlations based on this paper's GO method is formulated, and its process diagram is shown in Fig. 1.

1. Analyzing system
   Besides the contents of system analysis in the existing GO methodology [21], the redundancy structure with multiple correlations should be determined.
2. Developing GO model
   According to the results of system analysis, the function GO operator and logical GO operator are selected to represent the unit and the logical relationship in system. Then, the GO model of system is established by using signal flow to connect the GO operators.
3. Obtaining reliability parameters
   a. To obtain the reliability parameters of unit. They mainly contain the failure rate, maintenance rate and availability of unit.
   b. To obtain the reliability parameters of redundancy structure with shutdown correlation, maintenance correlation and standby correlation according to Eq. (1)-(5).
   c. To obtain the reliability parameters of redundancy structure with common cause failure. They mainly contain the parameters of common cause failure model, common cause failure probability, and availability.
4. Conducting quantitative GO analysis
   According to Eq. (6), the system availability can be obtained. If the shared signal does not exists in GO model, the GO operation can be conducted by the direct algorithm. If the shared signal exist in GO model, the GO operation can be conducted by the exact algorithm with shared signal [22].

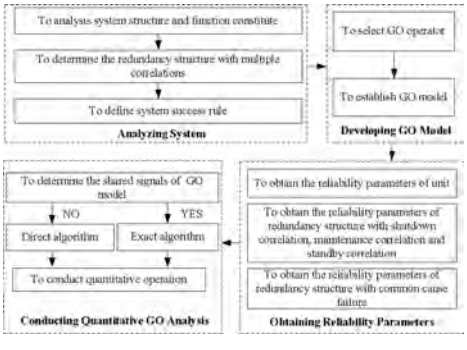


Figure 1. Reliability analysis process diagram of proposed GO method.

3 EXAMPLE

In order to illustrate the usage of this paper's GO method, the quantitative analysis of a hydraulic oil supply system for a military vehicle is conducted based on the proposed GO method.

3.1 *Analyzing hydraulic oil supply system*

1. To analysis structure and function constitute of hydraulic oil supply system
   The hydraulic oil supply system is used to supply working oil for variable speed control system, steering control system, hydraulic torque converter and lubrication system. And its schematic diagram is shown in Fig. 2.
2. To determine the redundancy structure with multiple correlations
   In hydraulic oil supply system, LF1 group and LF2 group are redundancy structures with shutdown correlation, maintenance correlation and standby correlation. The P1 group, LF1 group and LF2 group exist common cause failure because of external disturbance, such as extremely heavy impact, oversize oleo contaminated particle, and so on.
3. To define success rule of hydraulic oil supply system
   According to above analysis, the success rule of hydraulic oil supply system can be defined as that system can supply working oil to variable speed control system, steering control system, hydraulic torque converter, and lubrication system of a military vehicle under high steering speed condition without considering overload protection.

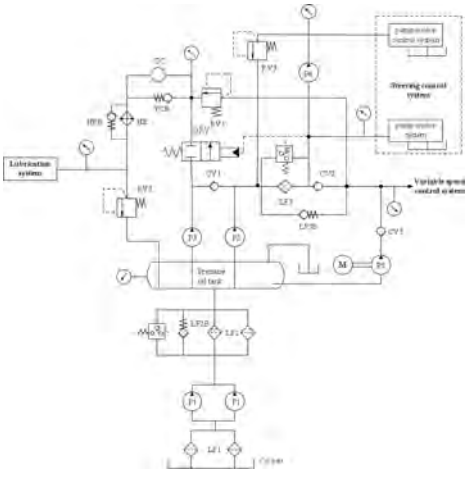


Figure 2. Schematic diagram of hydraulic oil supply system.

3.2 *Developing GO model of hydraulic oil supply system*

According to above analysis, the GO model of hydraulic oil supply system is developed, as follows:

1. To select GO operator
   There are 5 kinds of GO operators. And the GO operators corresponding units and logical relations are presented in Table 1.
2. To develop GO model
   According to analysis result of hydraulic oil supply system and Table 1, the GO model of hydraulic oil supply system is established, as shown in Fig. 3. In operators, the former number and latter number are type and serial number of GO operator, respectively. And the number on a signal flow is serial number of signal flow. The signal flow 37 is system output.

3.3 *Obtaining reliability parameters*

1. To obtain the reliability parameters of unit, as presented in Table 2. In Table 2, the failure rate and maintenance rate of unit are determined according to engineering statistic.
2. To obtain the reliability parameters of LF1 group and LF2 group by Eq. (1)-(5). In LF1 and LF2 group, $M = 2$, $K = 2$, $I = 2$, $L = 2$, $J = 0$. The reliability parameters of LF1 group are illustrated in detail, as presented in Table 3.
   In the same way, the reliability parameters of LF2 group can be obtained, i. e. its availability, failure rate and maintenance rate are 0.999981809, 0.000018341533 and 1.008264462, respectively.
3. To obtain the reliability parameters of redundancy structure with common cause failure, as presented in Table 4. In this case, we adopt the common cause failure β model [23].

Table 1. Operator type of unit.

No.	Unit	Type	Description
1	Oil pan	5	Input unit
2, 3	LF1	1	Two-state unit
5, 6	P1	6	Unit controlled by two control signals
7	Pump group power	5	Input unit
9, 10	LF2	1	Two-state unit
12, 18	LF2B, LF3B	1	two-state unit
14	Pressure oil tank	1	two-state unit
15	P2	6	Unit controlled by two control signals
16	LF3	1	Two-state unit
17, 24	CV2, CV3	1	Two-state unit
20	P4	6	Unit controlled by two control signals
21, 35, 36	RV1, RV2, RV3	1	Two-state unit
22	P5	6	Unit controlled by two control signals
23	P5 power	5	Input unit
26	P3	6	Unit controlled by two control signals
27	DRV	1	Two-state unit
29	TC	1	Two-state unit
30	TCB	1	Two-state unit
32	HE	1	Two-state unit
33	HEB	1	Two-state unit
4, 8,11, 25, 28		2	"OR" logical relation
37		10	"AND" logical relation
13, 31, 34, 19		18	"Standby" logical relation

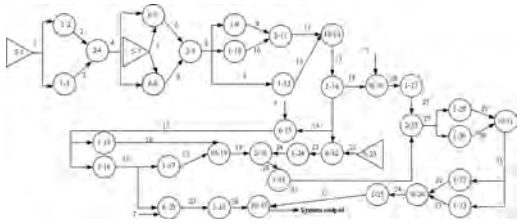


Figure 3. Reliability analysis process diagram of proposed GO method.

Table 2. Reliability parameters of unit.

Unit	Failure rate/ hour	Repair rate/ hour	Availability
1	0.00075	3.0000	0.999750062
2, 3	0.00605	1.0021	0.993999202
5, 6	0.00075	1.4933	0.999497999
7	0.02180	1.6274	0.986781433
9, 10	0.00305	1.0083	0.996984123
12, 18	0.00119	0.8865	0.996984123
14	0.00005	0.5000	0.999900010
15	0.00075	1.4933	0.999497999
16	0.00015	1.2000	0.999875015
17, 24	0.00075	1.0995	0.999318341
20	0.00075	1.4933	0.999497999
21, 35, 36	0.00060	1.5652	0.999616814
22	0.00075	1.4933	0.999497999
23	0.02180	1.6274	0.986781433
26	0.00075	1.4933	0.999497999
27	0.00110	1.0009	0.998902171
29	0.00050	0.0600	0.991735537
30	0.00119	0.8865	0.998659395
32	0.00040	0.0500	0.992063492
33	0.00119	0.8865	0.998659395

Table 3. Reliability parameters of LF1 group.

State number	0 (all units operating)	1 (one unit faulting)	2 (all units faulting)
Fault unit number	0	1	2
Operating unit number	2	1	0
Standby unit number	0	0	0
a_i	–	0.0121	0.00605
b_i	–	1.002149395	1.002149395
Pi	1	0.012074048	0.000072891
P_R (LF1 group)		0.9999279830	
λ_R (LF1 group)		0.0000721765	
μ_R (LF1 group)		1.0021493950	

Table 4. Reliability parameters of redundancy structure with common cause failure.

Structure	Failure rate	β	A_I	C
P1	0.00075	0.3	0.999648546	7.5332e-05
LF1	0.0000721765	0.13	0.999937345	3.7407e-04
LF2	0.0000183415	0.024	0.999982246	2.4794e-05

Table 5. Quantitative analysis by GO method.

R_1	Non-R_I item (P1 group)
0.98449496	–7.4169e-05
Non-R_I item (LF1 group)	Non-R_I item (LF2 group)
–3.6829e-04	–3.2723e-08
R_S	
0.9840524683	

3.4 *Conducting quantitative GO analysis of hydraulic oil supply system*

According to the definition of shared signal [22], the signal flow 1, 4, 7, 8, 14, 15, 16, 25 and 31 are shared signals in Fig. 3. Thus, the exact algorithm with shared signals is adopted to conduct GO operation. And the system availability is obtained by Eq. (6), as presented in Table 5.

4 RESULT ANALYSIS

In order to verify the feasibility and advantage of the proposed GO method, its analysis result compared with those by GO methods for system without considering correlations, system considering redundancy structure with shutdown correlation, maintenance correlation and standby correlation and system considering common cause failure. And the quantitative analysis results are presented in Table. 6.

Table 6. Quantitative analysis results by different GO methods.

Method	System availability
This paper's method	0.9840524683
Method without considering correlations	0.9845210823
Method considering redundancy structure with shutdown correlation, maintenance correlation and standby correlation	0.9840869188
Method considering common cause failure	0.9844857454

According to Table 6, we can see

1. The analysis result by this paper's method is smaller than the result by GO method without considering correlations. It meets engineering practice. It shows that if the correlations in systems are neglected, the reliability analysis result will have a bias.
2. The analysis result by this paper's method is smaller than the result by GO method considering redundancy structure with shutdown correlation, maintenance correlation and standby correlation. It meets engineering practice. It shows that the shutdown number, maintenance man, and standby structure affect the system reliability directly.
3. The analysis result by this paper's method is smaller than the result by GO method considering common cause failure. It meets engineering practice. It shows that the reliability analysis for system with redundancy structure should consider the common cause failure.

The reliability analysis process of example shows that the GO model is developed according to the system principle, system structure and function constitute, and the quantitative analysis result is obtained by multiple GO operations. It indicates that the GO method has obvious advantages in terms of establishing system model and conducting quantitative analysis.

5 CONCLUSION

This study proposes a new quantitative reliability analysis method for repairable systems with multiple correlations based on GO method. First, the solving methods for repairable systems with multiple correlations in GO method are presented, which are GO algorithm for repairable systems with common cause failure, and GO algorithm for redundancy structure with shutdown correlation, maintenance correlation and standby correlation,

respectively. On this base, the reliability analysis process of the proposed GO method is formulated. Then, the hydraulic oil supply system of military vehicle is taken as an example to conduct reliability analysis by this paper's GO method firstly. Finally, its quantitative analysis result compared with those by GO method for system without considering correlations and system with considering a single kind correlation. The analysis results show that the correlations affect the system reliability directly. And the reliability analysis process of example shows that the GO method has obvious advantages in establishing system model and conducting quantitative analysis. All in all, this study not only improves the theory GO methodology; but also provides a new reliability analysis approach for repairable systems with multiple correlations.

ACKNOWLEDGEMENTS

This paper is supported by National Natural Science Foundation of China (NSAF Joint Funds, U1530135) in the years 2016–2018, the Ministry of Industry and Information Technology (China) in the years 2011–2014 (ZQ092012B003). We are grateful to the editors and reviewers for the suggestions that improve the draft of this paper.

REFERENCES

Dong X.Y. Song H.W., Tian Y. B, 2000. Analysis of the Basic Models for Common Cause Failure, Journal of Beijing Institute of Technology, 20(2): 145~149.

Shen Z.P., Huang X.R., 2004. Principle and Application of GO Methodology, Tsinghua University press, Beijing, China.

Shen Z.P., Tang H., 2006. System Reliability Analysis with Common Cause Failures Using the GO Methodology, Journal of Tsinghua University, 46(6): 829–832.

Shen Z.P., Wang Y., Huang X.R., 2003. A quantification algorithm for a repairable system in the GO methodology, Reliability Engineering and System Safety, 80(3): 293~298.

Zhou, L.G. H.P. Dong, X.J. Yi, et al, 2015. Reliability Analysis of Retracting Actuator with Multi-State Based on Goal Oriented Methodology, *Shanghai Jiaotong Univ. (Sci.)*, **20(3)**: 307–311.

Yi, X.J. H.P. Dong, Q.F. Wang & Z. Zhang, 2015. A new system reliability analysis method: the current development of GO methodology in China, *WIT Transactions on Engineering Sciences*, **109**, 222~229.

Yi X.J., B.S. Dhillon, Shi J., et al., 2017. Quantitative Reliability Analysis of Repairable Systems with Closed-Loop Feedback Based on GO Methodology, *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, **39: 1845**, doi: 10.1007/s40430-016-0665-9.

Yi X.J., Dhillon B.S., Mu H.N., et al., 2016. Reliability Analysis Method for Multi-State Repairable Systems Based on Goal Oriented Methodology, *Proceedings of ASME 2016 International Mechanical Engineering Congress & Exposition*, Phoenix, USA, November 11–17, 2016, IMECE2016–65380.

Yi X.J., Dhillon B.S., Shi J., et al., 2015. Reliability Analysis Method on Repairable System with Standby Structure Based on Goal Oriented Methodology, Quality and Reliability Engineering International, 32(7), Doi: 10.1002/qre.1953.

Yi X.J., Dhillon B.S., Shi J., et al., 2016. A New Reliability Analysis Method for Heavy Vehicle Systems Based on Goal Oriented Methodology, *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, Doi: 10.1177/0954407016671276.

Yi X.J., Dhillon B.S., Shi J., et al., 2016. Reliability Optimization Allocation Method for Multifunction Systems Based on Goal Oriented Methodology, *Proceedings of ASME 2016 International Mechanical Engineering Congress & Exposition*, Phoenix, USA, November 11–17, IMECE2016–65383.

Yi X.J., Dong H.P., Jiang J.P., et al, 2014. Reliability Analysis of Hydraulic Transmission Oil Supply System of Power-Shift Steering Transmission Based on GO methodology, *Journal of Donghua University (Eng. Ed.)*, **31(6)**: 785–788.

Yi X.J., Hou P., Shi J., et al., 2017. A New Reliability Assessment Method for Complex Systems Based on Goal Oriented Methodology, *27th annual European Safety and Reliability Conference, Slovenia*, Portoroz, June 18–22.

Yi X.J., Lu M.C., Shi J., et al., 2017. A New Reliability Assessment Method for Complex Nuclear Power Equipment Based on Goal Oriented Methodology, *ASME International Conference on Nuclear Engineering 25*, Shang Hai, China, July 2–6.

Yi X.J., Mu H.N., Hou P., Lai Y.H., 2016. A reliability optimization allocation method considering differentiation of functions Based on Goal Oriented method, *7th International Conference on Computational Methods (ICCM2016)*, August 1st–4th, Berkeley, CA, USA.

Yi X.J., Shi J., and Dong H.P., et al, 2016. Reliability Analysis of Repairable System with Multiple Fault Modes Based on GO Methodology. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, **2**, DOI: 10.1115/1.4030971.

Yi X.J., Shi J., and Dong H.P., et al., 2014. Reliability Analysis of Repairable System with Multiple Failure Modes Based on GO Methodology, *Proceedings of ASME 2014 International Mechanical Engineering Congress & Exposition*, Montreal, Canada, November 14–20 2014, **14**, IMECE2014–36198.

Yi X.J., Shi J., and Mu H.N., et al., 2015. Reliability Analysis of Hydraulic Steering System with DICLFL Considering Shutdown Correlation Based on GO Methodology, *In: The First International Conference on Reliability Systems Engineering & 2015 Prognostics and System Health Management Conference-Beijing*, Beijing, China, October 21–23 2015, paper no. PR0101.

Yi X.J., Shi J., and Mu H.N., et al., 2016. A Reliability Analysis Method on Repairable Systems with Dual Input Closed-Loop Feedback Link Considering Shutdown Correlation Based on GO Methodology. *Journal of Donghua University (Eng. Ed)*, **33**, pp. 25~29.

Yi X.J., Shi J., Dhillon B.S., Reliability Analysis Method for Repairable Systems with Multi-Function Modes Based on Goal Oriented Methodology, *Quality and Reliability Engineering International*, doi: 10.1002/qre.2180.

Yi X.J., Shi J., Mu H.N., et al, 2015. Reliability Analysis of Repairable System with Multiple-Input And Multi-Function Component Based on GO Methodology, *In: ASME 2015 International Mechanical Engineering Congress & Exposition*, Houston, Texas, Nov13–19, paper no. IMECE2015–51289.

Yi X.J., Shi J., Mu H.N., et al, 2016. Reliability Analysis of Repairable System with Multiple-Input And Multi-Function Component Based on GO Methodology, *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, doi:10.1115/1.4034744.

Yi X.J. & Dhillon B.S. et al., 2016. Reliability Analysis Using GO Methodology: A Review, *The 22nd ISSAT International Conference Reliability and Quality in Design,* Los Angles, California, August 4–6 2016, paper no. RQD-53.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Reliability based topology optimization design of the network system: A case study on a sewage treatment system

X.J. Yi
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China
China North Vehicle Research Institute, Beijing, China

P. Hou & H.N. Mu
Beijing Institute of Technology, Beijing, China

Y.H. Lai
College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing, China

ABSTRACT: The network topological optimization, developing the scheme of pipelined connections needs to find the best layout of all links in the urban sewage system, makes the whole system cost-effective with the satisfaction of some special criteria. In this paper, the application of constraints based on some priori knowledge, including the requirement to link the components in network systems, reduces the solution space of the optimization substantially, unlike most literatures focus on improving the efficiency of optimization algorithms through the modification of the algorithm itself. In order to evaluate it, the network topological optimization of a sewage system, serving two different cities, is investigated. Performance of the solution based on the application of priori knowledge is validated through the sewage system. It indicates that the application of priori knowledge can be efficient for the solution.

1 INTRODUCTION

Urban sewage systems are important infrastructures for modern cities. A successful design of the urban sewage system is a guarantee for its availability, economy, and reliability. As the critical content of the design, developing the scheme of pipelined connections needs to find the best layout of all links in the urban sewage system, which makes the whole system cost-effective with the satisfaction of some special criterions. This design problem is defined as the optimization of network topology [1].

The objective of network topological optimization is to find the optimal layout of links among components in the system, and some requirements should be met such as the performance of networks, transmission delay. This paper considers the network topology optimization of the urban sewage system. In particular, its design should have a minimal cost under the constraint that the reliability of the system based on the design is not less than a given system reliability.

In general, the study of network topological optimization concentrates on three aspects: the choice of a measure of reliability, the algorithm adopted to calculate the system reliability and the solution of the optimization model. The measures of reliability in network systems include the connectivity measure [2], the connectivity measure with the consideration of the performance [3], and the measure based on capacity [4]. At the same time, the exact algorithms based on these measures of system reliability are widely investigated [5]. However, the increased size of network systems poses a new challenge to the computation of system reliability. Approximation algorithms are adopted to address this challenge [6–8].

The solution of the optimization model, unlike the choice of a measure of reliability and the algorithm adopted to calculate the system reliability, is mainly about how to solve the optimization model effectively. In this aspect, many optimization algorithms are proposed to the solution of network topological optimization [9–17]. However, most literatures focus on improving the efficiency of optimization algorithms through the modification of the algorithm itself.

In this paper, the application of constraints based on prior knowledge, including the requirement to link the components in network systems, reduces the solution space of the optimization substantially. In order to evaluate it, the network topological optimization of a sewage system, serving two different cities, is investigated. The constraints in optimization are that the system reliability should be larger than a given threshold, and the total cost of the sewage system is minimized. In

addition, performance of the solution based on the application of priori knowledge is validated through the sewage system. It indicates that the application of priori knowledge can be helpful for the solution.

2 MODELING FOR RELIABILITY BASED TOPOLOGY OPTIMIZATION DESIGN PROBLEM

In most research about reliability based the topology optimized design, it is often given that: (a) location of sewage treatment plants in different levels, (b) the reliability of sewage treatment plants, (c) the cost of pipelines that connect sewage treatment plants, (d) the constraint of sewage treatment system reliability. The layout of the sewage treatment system is actually determined by the connection relationship among cities and sewage treatment plants in different levels. The connection relationship is often defined by topological relation, and it is represented by graph in graph theory. The objective of reliability based topology optimization design is to find the optimal graph that enables the sewage treatment system has the minimal cost and the required reliability. Therefore, the optimization is adapted to model for reliability based topology optimization design. In the optimization model, the decision variable is the graph representing the topological relation of the sewage treatment system. The objective function is to evaluate the cost of the sewage treatment system; the constraint is used to guarantee the reliability of the sewage treatment system that meets the requirement.

2.1 *Representation of the topological characteristic of the sewage treatment system*

This paper adopts graph to represent the layout of the sewage treatment system. Graph may reveal the topological characteristics of the sewage treatment system, in particular the connection relationship among cities and sewage treatment plants in different levels. In addition, the adjacent matrix, a representation of graph, is a very useful data structure for Genetic Algorithm to solve the problem.

In the research about reliability based topology optimization design, the graph is used to represent the layout of systems, such like social networks, power networks. A graph g consists of the nodes set V and the arcs set E. The elements in the nodes set are nodes in graph, and the components of the sewage treatment system, such as cities and sewage treatment plants, are usually represented by nodes. The elements of the set E represent the pipelines in the sewage treatment system.

$$Adjacent\ Matrix = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \quad (1)$$

Although graphs are very intuitive to show the layout of sewage treatment system, it is hard for a graph to participate in an operation as the part of Optimization Algorithm. To address the challenge, the adjacent matrix, a representation of graph, is adopted in this paper. The adjacent matrix is a very useful data structure, and it enables the graph to participate in the operation of Optimization Algorithm to solve the problem. The dimension of an adjacent matrix is determined by the amount of nodes in the graph. The row number or column number is consistent with the node number. For example, the adjacent matrix of the graph in Fig. 1 is shown in Eq. (1). The elements in an adjacent matrix can be 0 or 1. The element 0 means that

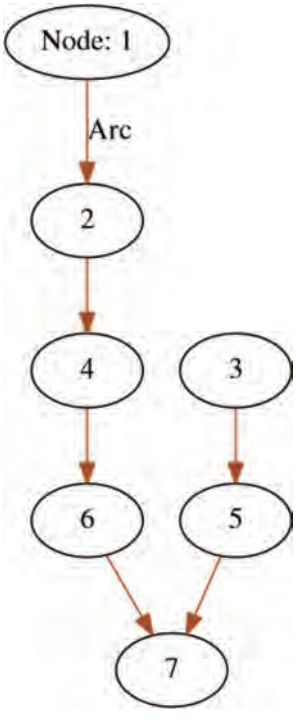


Figure 1. The graph referred to adjacent matrix in Eq. (1).

there is no connection between the node whose number equals to the row number and the node whose number equals to the number of columns number, and the element 1 means that there is a link between them. Therefore, the adjacent matrix is equal to the graph, and it can be adopted to represent the layout of the sewage treatment system.

2.2 *Solution space in optimization model*

In fact, the objective of solving the optimization model is to find an optimal solution that satisfies the constraint in solution space. Therefore, defining a reasonable solution space is helpful for optimization algorithm to solve efficiently. In this paper, the solution space is all the possible layout of sewage treatment system, and these layouts meet the design requirements. It consists of the necessary components of the whole sewage treatment system and the join sequence of cities and sewage-treatment plants. The normal sewage treatment system of a city includes a primary sewage treatment plant, a secondary sewage treatment plant and a tertiary sewage treatment plant. Therefore, the sewage treatment system must contain all the three plants. In addition to this, the join sequence of cities and sewage-treatment plants must be correct. The sewage of cities must be disposed first by the primary sewage treatment plant, and then the secondary sewage treatment plant is used for the process. The process of sewage treatment finishes at the tertiary sewage treatment plant.

According to the design requirements of sewage treatment system, this paper proposes a method that restricts the value of elements in adjacent matrix to limiting solution space. It is often given that the locations of cities and sewage treatment plants, therefore the amount of nodes in graph is known. Based on that, the dimension of the adjacent matrix can be determined. At the other hand, there is no connection between sewage-treatment plants and themselves, therefore diagonal elements in the adjacent matrix is 0. Based on the information above, the amount of unknown elements in the adjacent matrix is decreased. The adjacent matrix can be represented like Eq. (2).

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ n-2 \\ n-1 \\ n \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & \cdots & n-2 & n-1 & n \end{array} \\ \begin{pmatrix} 0 & a_{1,2} & a_{1,3} & \cdots & a_{1,n-2} & a_{1,n-1} & a_{1,n} \\ a_{2,1} & 0 & a_{2,3} & \cdots & a_{2,n-2} & a_{2,n-1} & a_{2,n} \\ a_{3,1} & a_{3,2} & 0 & \cdots & a_{3,n-2} & a_{3,n-1} & a_{3,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n-2,1} & a_{n-2,2} & a_{n-2,3} & \vdots & 0 & a_{n-2,n-1} & a_{n-2,n} \\ a_{n-1,1} & a_{n-1,2} & a_{n-1,3} & \vdots & a_{n-1,n-2} & 0 & a_{n-1,n} \\ a_{n,1} & a_{n,2} & a_{n,3} & \vdots & a_{n,n-2} & a_{n,n-1} & 0 \end{pmatrix} \end{array} \tag{2}$$

In fact, the amount of unknown elements is decreased from n^2 to $n^2 - n$, which means that the solution space is limited. Therefore, the solution space is determined by the $n^2 - n$ elements, and they can be represented by $X = (x_1, x_2, \cdots, x_{n^2-n})$.

2.3 *Objective function in optimization model*

The objective function is to calculate the cost of the sewage treatment system. In this paper, the construction cost of the sewage treatment system is ignored; therefore the cost of the sewage treatment system is determined by the connections among the plants. The length of pipelines for connections mainly affects the cost. According to the requirements of design, the primary sewage treatment plant is near the cities, but the secondary sewage treatment plant and the tertiary sewage treatment plant are far from the cities. Fortunately, the cost of different connections among them is usually known, and they are often represented by a matrix that is called cost matrix **A**, as shown in Eq. (3).

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ n-2 \\ n-1 \\ n \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & \cdots & n-2 & n-1 & n \end{array} \\ \begin{pmatrix} c_{1,1} & c_{1,2} & c_{1,3} & \cdots & c_{1,n-2} & c_{1,n-1} & c_{1,n} \\ c_{2,1} & c_{2,2} & c_{2,3} & \cdots & c_{2,n-2} & c_{2,n-1} & c_{2,n} \\ c_{3,1} & c_{3,2} & c_{3,3} & \cdots & c_{3,n-2} & c_{3,n-1} & c_{3,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{n-2,1} & c_{n-2,2} & c_{n-2,3} & \vdots & c_{n-2,n-2} & c_{n-2,n-1} & c_{n-2,n} \\ c_{n-1,1} & c_{n-1,2} & c_{n-1,3} & \vdots & c_{n-1,n-2} & c_{n-1,n-1} & c_{n-1,n} \\ c_{n,1} & c_{n,2} & c_{n,3} & \vdots & c_{n,n-2} & c_{n,n-1} & c_{n,n} \end{pmatrix} \end{array} \tag{3}$$

The cost of different connections can be obtained by the cost matrix **A** and the connections that exist in layout are obtained by the decision variable **X**. Therefore, the objective can be represented by the Eq. (4).

$$Min: c = F(\mathbf{A}, \mathbf{X}) \tag{4}$$

2.4 *Constraint in optimization model*

The constraints in optimization model are used to guarantee the layout that is solved by the model meets the reliability requirement and the design requirements.

When the layout of the sewage treatment system is determined, the reliability of the sewage treatment system can be calculated using the computational method of the network reliability. At present, there are four computational methods of the network reliability: 1) calculate the probability of connectivity, 2) calculate the probability of connectivity with the consideration of the network capacity, 3) calculate the probability of connectivity with the consideration of the network performance, 4) calculate the reliability based on mission. This paper adopts the method of calculating the probability of connectivity to calculate the reliability of the sewage treatment system. This result of this method is also called s-t reliability, and it is

determined by the probability of the connectivity of all the path from the source node to the terminal node. In this paper, the source node is the city, and the terminal node is the tertiary sewage treatment plant. In addition, it is not necessary to consider the reliability of pipelines; therefore the reliability of the sewage treatment system is determined by the components of the sewage treatment system. In this paper, the reliability of the components is represented by $R=(R_1,R_2,\cdots,R_n)$, and they are actually given in most study. Using the reliability of components R and the connection information X, the constraint is defined in Eq. (5). The system reliability R^* is the threshold in requirement.

$$s.t. \quad g(R,X)\geq R^* \tag{5}$$

According to the content above, the optimization model is defined in Eq. (6).

$$\begin{aligned} &Min: c=F(\mathbf{A},\mathbf{X}) \\ &s.t. \quad g(R,X)\geq R^* \end{aligned} \tag{6}$$

3 A CASE STUDY

A sewage treatment system, serving for two cities, is proposed in this paper to be analyzed. There are two primary sewage treatment plants, two secondary sewage treatment plants and one tertiary sewage treatment plant. Their information about the node number and reliability are shown in Table 1. It should be noted that there are some pressurization equipment in cities, therefore the cities also have the reliability. Because the location of cities and sewage treatment plants are given in Fig. 2, the cost matrix is also known, and it is shown in Eq. (7). At the same time, the threshold R^* is 0.8.

$$A=\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}\begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{pmatrix} 0 & 100 & 11 & 43 & 30 & 55 & 60 \\ 100 & 0 & 39 & 14 & 42 & 28 & 58 \\ 11 & 39 & 0 & 50 & 28 & 37 & 42 \\ 43 & 14 & 50 & 0 & 25 & 19 & 32 \\ 30 & 42 & 28 & 25 & 0 & 34 & 37 \\ 55 & 28 & 37 & 19 & 34 & 0 & 35 \\ 60 & 58 & 42 & 32 & 37 & 35 & 0 \end{pmatrix} \end{array} \tag{7}$$

3.1 *Building the optimization model*

There are seven nodes in this case, therefore the dimension of adjacent matrix is 7×7. In addition, there are some more design requirements about the join sequence of cities and sewage-treatment plants, and these requirements are: 1) city can only connect with the other city or the primary sewage treatment plants, 2) the primary sewage treatment plant can only connect with the other primary sewage treatment plant or the secondary sewage treatment plants, 3) the secondary sewage treatment plant can only connect with the other secondary sewage treatment plant or the tertiary sewage treatment plant. Thus, the adjacent matrix is defined in Eq. (8). And the decision variable $\mathbf{X}$ in this case is defined as $X=(x_1,x_2,\cdots,x_{16})$.

Table 1. The node number and reliability information.

Component name	Node number	Reliability
City A	1	0.9895
City B	2	0.9882
Primary sewage treatment plant A	3	0.9535
Primary sewage treatment plant B	4	0.9886
Secondary sewage treatment plant A	5	0.9328
Secondary sewage treatment plant B	6	0.9761
Tertiary sewage treatment plant	7	0.9585

Figure 2. The location cities and components in the sewage system.

$$Adjacent\ Matrix=\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}\begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{pmatrix} 0 & x_1 & x_2 & x_3 & 0 & 0 & 0 \\ x_4 & 0 & x_5 & x_6 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_7 & x_8 & x_9 & 0 \\ 0 & 0 & x_{10} & 0 & x_{11} & x_{12} & 0 \\ 0 & 0 & 0 & 0 & 0 & x_{13} & x_{14} \\ 0 & 0 & 0 & 0 & x_{15} & 0 & x_{16} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \tag{8}$$

The cost is calculated by the objective function through Hadamard product between the cost matrix and the adjacent matrix. The Hadamard product is new matrix that its dimension is still 7×7. These non-zero elements are the cost of connections

which exist in the designed layout. The sum of these non-zero elements is the total cost of the layout.

There are some more constraints in this optimization model except the reliability requirement. It should guarantee: 1) the sewage of city A and B must be discharged, 2) the sewage discharged by city A and B should be disposed by the primary sewage treatment plants, 3) the sewage discharged by primary sewage treatment plants should be disposed by secondary sewage treatment plants, 4) the sewage discharged by secondary sewage treatment plants should be disposed by the tertiary sewage treatment plant. All the requirements are defined in Eq. (9). Therefore, based on all above, the Optimization Model is defined in Eq. (10).

$$\begin{aligned} &x_1 + x_2 + x_3 \geq 1 \\ &x_4 + x_5 + x_6 \geq 1 \\ &x_2 + x_3 + x_5 + x_6 \geq 1 \\ &x_8 + x_9 + x_{11} + x_{12} \geq 1 \\ &x_{14} + x_{16} \geq 1 \end{aligned} \tag{9}$$

$$\begin{aligned} Min: \quad & c = F(A, X) \\ s.t. \quad & g(R, X) \geq R^* \\ & x_1 + x_2 + x_3 \geq 1 \\ & x_4 + x_5 + x_6 \geq 1 \\ & x_2 + x_3 + x_5 + x_6 \geq 1 \\ & x_8 + x_9 + x_{11} + x_{12} \geq 1 \\ & x_{14} + x_{16} \geq 1 \end{aligned} \tag{10}$$

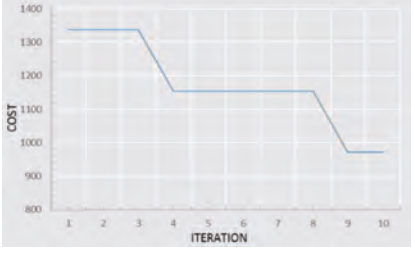


Figure 3. Cost in iterations.

Figure 4. The layout of the sewage treatment system.

3.2 *Solving the optimization model*

Genetic Algorithm is adopted to solve the optimization model. The population size in algorithm is 200. The crossover rate and mutation rate are 0.8 and 0.5 respectively. The stopping criterion of Genetic Algorithm is set to execute 200 iterations. In order to evaluate the performance of the application of constraints based on prior knowledge, the set of parameters is chosen for an efficient solution. The result shown in Fig. 3 indicates that the cost converges to a value, and the value is 971. The layout of the sewage treatment system is shown in Fig. 4, according to the decision variable **X**. Obviously, the cost in previous generations such as 1338 in the third generation and 1154 in the eighth generation shows that the global optimized result is obtained.

At the same time, the solution without these constraints in Eq. (8) is obtained. Although the final solution is the same with the solution with the application of constraints based on prior knowledge, the process of solving the optimization model should execute many times to find the solution meets which the requirements of the design.

4 CONCLUSION

This study has formally defined a reliability based topology optimization design problem through a sewage treatment system to find the optimal layout of the sewage treatment system that serves for two cities. The optimal layout of the sewage treatment system should have the minimum cost subject to the reliability constraint R^*. A modeling and analyzing method is proposed in this paper for the reliability based topology optimization design problem. The method can be used to build the objective function and the constrains. In addition, the priori knowledge about the design requirements is used to determine the elements in adjacent

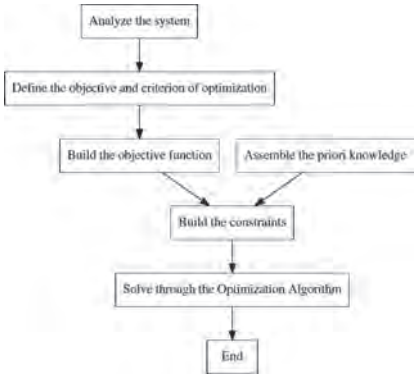


Figure 5. The procedure of the method.

Figure 6. The original design of the sewage system.

Table 2. The comparison between the original design and the optimal design.

Type	Decision variable
The original design	1 0 0 0 0 1 0 0 0 0 1 0 0 1 0 0
The optimal design	0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 1

matrix, which limits the solution space, and this makes the algorithm efficient for solving the problem. Finally, Genetic Algorithm is adopted to solve the optimization model. According to the procedure above shown in the Fig. 5, the optimal layout of the sewage treatment system is obtained. The cost based on the optimal layout is 971, and its reliability 0.8266 that meets the reliability requirement. Compared with the original design shown in Fig. 6, the significant change in the optimal decision variable is presented in Table 2. Obviously, the method makes the design change toward the direction of more feasibility. In particular, the application of constraints based on prior knowledge makes the solution efficient. Therefore, this method is proved to be helpful for the reliability based topology optimization design problem.

ACKNOWLEDGEMENTS

This paper is supported by National Natural Science Foundation of China (NSAF Joint Funds, U1530135) in the years 2016–2018. We are grateful to the editors and reviewers for the suggestions that improve the draft of this paper.

NOMENCLATURE

g	The graph
V	The set of nodes in graph
E	The set of arcs in graph
A	The cost matrix
X	The decision variable
$F(.)$	The objective function
R	The reliability of components
$g(.)$	The function of system reliability
c	The total cost of a sewage treatment
R^*	The threshold in reliability requirement

REFERENCES

Abo ElFotoh H.M.F. and Al-Sumait L.S., 2001. A neural approach to topological optimization of communication networks, with reliability constraints. *IEEE Transactions on Reliability*, 50: 397–408.

Altiparmak F., Dengiz B., and Belgin O., 2010. Design of reliable communication networks: A hybrid ant colony optimization approach for the design of reliable networks. *IIE Trans.*, 42: 273–287.

Atiqullah M. and Rao S., 1993. Reliability optimization of communication networks using simulated annealing, *Microelectron. Rel.*, 33: 1303–1319.

Chelouah R. and Siarray P., 2015. Tabu search applied to global optimization, *Eur. J. Oper. Res.*, 123: 256–270.

Chen A., Yang H., Hong K.L., et al., 2002. Capacity reliability of a road network: an assessment methodology and numerical results, *Transportation Research Part B Methodological*, 36(3): 225–252.

Deeter D. and Smith E., 2016. Economic design of reliable networks, *IIE Trans.*, 30: 1161–1174.

Dengiz B., Altiparmak F. and Smith A.E., 1997. Local search genetic algorithm for optimal design of reliable networks. *IEEE Trans. Evol. Comput.*, 3(1): 179–188.

Dengiz B., Altiparmak F., and Smith A.E., 1993. Efficient optimization of all-terminal reliable networks. *IEEE Transactions on Reliability*, 33: 18–26.

Fishman G.S., 2007. A Comparison of Four Monte Carlo Methods for Estimating the Probability of s-t Connectedness, *IEEE Transactions on Reliability*, 35(2): 145–155.

Hsu Steen J, Yuang Maria., 1998. Efficient Computation of Terminal-pair Reliability Using Triangle Reduction in Network Management, *IEEE International Conference on Communications*: 281–285.

Jan R.H., Hwang F.J., 1993. Topological optimization problem of communication networks subject to a reliability constraint. INFOCOM '90, *Ninth Joint Conference of the IEEE Computer and Communication Societies*, 2: 487–494.

Konak A., Smith A.E., 2006. Network Reliability Optimization, *Handbook of Optimization in Telecommunications*, 735–760.

Papagianni C., Papadopoulos K., and Pappas C., 2008. Communication network design using particle swarm optimization, *In Proc. Int. Multiconf. Comput. Sci. Inf. Technol.*, 13–15.

Pierre S., Hyppolite M.A., Bourjolly J.M., and Dioume O., 1995. Topological design of computer communication networks using simulated annealing, *Eng. Applicat. Artificial Intell.*, 8: 61–69.

Rothemund P W K, Winfree E., 2000. The program-size complexity of self-assembled squares, *ACM Symposium on Theory of Computing*, 459–468.

Satitsatian S., Kapur K.C., 2006. An algorithm for lower reliability bounds of multistate two-terminal networks, *IEEE Transactions on Reliability*, 55(2): 199–206.

Shier D.R., 1991. Network reliability and algebraic structures. Clarendon Press.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Towards a systematic evaluation of supplementary protective measures and their quantification for use in functional safety

J. Zehetner & U. Weber
Furtwangen University of Applied Science (HFU), Germany

I. Häring & W. Riedel
Fraunhofer Institut für Kurzzeitdynamik (EMI), Germany

ABSTRACT: Risk reduction can be conducted through constructive, technical, organizational or personnel measures. State of the art is to determine the necessary risk reduction, taking into account control dependent protective measures in the sense of functional safety. Complementary protective measures are a special subgroup of protective measures.

Such additional safety functions extend the usual quantified sensor logic actuator chain of safety functions by the operator and the necessary human-machine interface. Standard methods of quantification only apply to the reliability of the technical part of such complementary safety functions. This paper deals with the challenge of taking into account large extent through the human-machine interface. It proposes a systematic approach using a combination of methods offered by standards and literature to determine and quantify the risk reduction of the overall complementary safety instrumented functions. The paper summarizes the proposal of an extension for the assessment of complementary protective functions.

1 INTRODUCTION

1.1 *Motivation*

The ongoing technological changes affect traditional workplaces. Workers change their role from part of the production chain to part of the control system as the level of automation is ever increasing. This affects how safety is ensured. Therefore it is important that human-machine interfaces fulfil the requirements of safety-related functions.

Some of these interfaces are represented by mechanical control elements that initiate electronic safety functions. These functions are grouped together under complementary protective measures. Typically these measures are not operated as often as a control-dependent safety function, and are considered to be a complementary safety device.

ISO 12100 provides details on complementary protective measures. These are measures which are neither inherently safe design nor safeguarding, but are required due to intended use or reasonably foreseeable misuse of the machine (ISO 12100, 2010).

Nevertheless, these human-machine interfaces must be taken into account for an overall risk assessment of machinery and systems.

Within the scope of the IEC 61508 is the overall risk assessment of systems for which electrical/electronic/programmable electronic safety-related systems significantly contribute to the overall risk reduction. This standard also defines human error as a systematic failure which has to be considered in the overall risk assessment (IEC 61508, 2011). Hence, if complementary safety functions are part of the risk reduction strategy in integration, operation or maintenance they have to be considered in the overall risk assessment and evaluation. This is independent of the decision whether complementary safety functions are to be treated as standard safety functions as well as their risk reduction effect. For instance, their risk reduction could be considered as a safety buffer (safety factor, back up safety function) in addition to more automated safety functions.

Other standards like the EN 61511 also require that the design of a safety instrumented function shall take into account human error (IEC 61511, 2004). This means that the assessment of the control elements of human-machine interfaces are part of the overall assessment of safety functions.

Nonetheless, the available standards do not give a sufficient answer to the question of how asses the human machine interfaces. For instance, it has to be discussed if the assessment should be qualitative or quantitative.

For this reason, there is a need for research to asses such interfaces as part of the safety chain.

1.2 *Challenge/main topic*

The Machinery Directive 2006/42/EC prescribes a hierarchy of measures for risk reduction (Machinery Directive 2006/42/EC, 2006):

- Inherently safe design measures,
- technical protective measures, and
- information for users.

Technical protective measures could be divided into control independent and control dependent measures. If risks are then reduced by control dependent measures, a validation must be conducted to proof sufficient risk reduction.

For this verification, the assessment results usually are summarized in a Reliability Block Diagram (RBD) (IEC 61508, 2011). Based on stochastic data, calculations and estimation procedures, a quantitative reliability value (e.g. called SIL_{claim}) is determined for each safety function block and the overall system.

However, complementary protective measures take a special position. They are initialized by an operator and not by a sensor. Hence, in order to evaluate complementary protective measures, the reliability block diagram is proposed to be extended as shown in Figure 1 to include the operator and the human-machine interface.

It is necessary to review in which form the operator and the mechanical interface are included in the standards of reliability assessment. Therefore states of the art for both areas are reviewed in the paper, with focus on quantitative approaches:

- State of practice to assess the operator is to use the human reliability analysis approaches.
- State of practice to assess electrical/electronic/programmable electronic safety-related systems is to use component failure values, calculation and estimation methods.

It has to be discussed which of both method sets can be used to assess the design of the human-machine interface properly.

To do so, complementary protective measures are first of all introduced in section 2 and subdivided into different technologies and operating procedures.

In section 3 a taxonomy is developed to choose appropriate methods and requirements for needed extensions are listed.

Section 4 then reviews the appropriate methods presented by standards and literature and discusses approaches to extend existing methods, i.e. methods that are not yet mentioned in standards and thus are beyond current best practices.

This extension should enable determining and quantifying the reliability of mechanical human-machine-interfaces. Aim is to support the determination of a total reliability value for complementary protective measures.

In section 5, the most suitable methods for determining human reliability analysis and technical safety for complementary protective measure interfaces are listed. Finally suitable additions to the existing methodology to assess safety functions with human-machine interfaces are described.

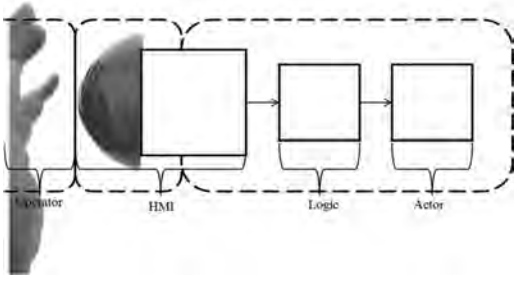


Figure 1. Extended reliability block diagram.

2 HUMAN MACHINE INTERFACE

Complementary protective measures are used as one of the few safety human-machine interfaces. Directory 2006/42/EG stipulates that any machine must be equipped with one or more complementary protective measures, unless the risk cannot be reduced without complementary protective measures.

Areas of application of complementary protective measures include machines, process engineering or emergency braking devices on trains. Disconnectors as used in the automotive sector. Emergency shut downs in laboratories or control rooms are integral part of such facilities.

Since these control elements initialize a safety function, they need to be highly available. For this reason they have to be clearly recognizable, easily visible and quickly accessible (ISO 13850, 2014; Machinery Directive 2006/42/EC, 2006).

The design of the control elements depends on the requirements of the application. Due to this, they can be designed as pushbuttons, wires, ropes, rails, handles or handlebars (Gehlen and Rudnik, 2015). In addition, there are also large scale control elements or specially designed foot switches.

The selection depends also on the type of actuation, e.g. by hand or foot. However, other extremities such as elbows or knees are less common but also used (ISO 13850, 2014; IEC 60947, 2015; Gehlen and Rudnik, 2015).

Special designs are becoming increasingly important in order to ensure adaptation to the operator or the working environment, for example as compensation for the physically disabled or applications on control elements to prevent accidental actuation.

Due to that variety of common control elements, an extension of the methodology is necessary in order to allow to inclusion of control elements when determining reliability.

In doing so, section 3 defines requirements to choose suitable methods. These requirements are needed to categorize and select promising existing and emerging methods of human reliability and technical safety.

3 CLASSIFICATION TAXONOMY OF METHODS FOR HUMAN-MACHINE INTERFACE ASSESSMENT

The necessity to assess the human-machine interface was described in section 1. This section develops a taxonomy for choosing appropriate assessment methods and points out requirements for the needed extensions.

To support an overall risk assessment in the sense of IEC 61508, valid input data are necessary. Due to the fact that IEC 61508 does not list many methods for risk assessment, those listed by ISO/IEC 31010 are discussed.

In order to use the output of the methods within the reliability block diagram, the output should be able to be transferred into quantitative values.

For the human-machine interface of complementary protective measures there is a lack of this kind of data.

The method or a combination of methods to be selected should be able to submit an essential contribution to the quantification of such interfaces.

The method should consider design requirements as defined in directive 2006/42/EG. Interface devices must be

- clearly identifiable,
- clearly visible and
- quickly accessible.

This has to hold for different product types and all kinds of actuations mentioned in section 2.

Furthermore performance shaping factors like design, temperature, noise or workflow should be measurable by the selected set of methods.

The set of methods should build on at least partially validated approaches as documented in reviewed literature and be already accepted or acceptable by experts as well as practitioners.

In summary, the methods are examined with regard to the applicability of the following items:

1. Delivery of quantitative data;
2. Existing scientific acceptance and level of validation to build on;
3. The method set should allow to link to the concept of Safety Integrity Levels (SIL) as defined in IEC 61508 and related standards;
4. The methods should allow the assessment of the fulfillment of existing design requirements as of Machinery Directive 2006/42/EC;
5. High expected scientific acceptance and practical feasibility of methods.

4 REVIEW OF METHODS

This section reviews and discusses methods, techniques and measures offered by major standards in general and in the domain of functional safety and machine safety. First methods for human reliability prediction will be reviewed and discussed. In doing so all mentioned aspects of section 3 are considered.

4.1 *Standard of human reliability analysis*

Human Reliability Analysis (HRA) is commonly divided into three categories. The first generation focuses on quantification in terms of success/failure of actions. The second generation focuses on cognitive aspects that cause errors by taking into account performance shaping factors. The third generation focuses on human performance factor relations and dependencies (Di Pasquale et al., 2015).

Methods of all generations are divided into analytical and expert estimation procedures, as characterized by (Sträter, 1997).

Examples of methods for expert estimation methods are Human Error Assessment and Reduction Technique (HEART) (Swain and Guttmann, 1983) and SLIM and Standardised Plant Analysis Risk-Human Reliability Analysis (SPAR-H). In these procedures, the cognitive area is the main focus. The employed data bases on studies, surveys and extensive literature research.

Exemplarily analytical methods are Technique for Human Error Rate Prediction (THERP) and Accident Sequence Evaluation Programme (ASEP) (Swain and Guttmann, 1983) as well as Méthode d'Evaluation de la Réalisations des Missions Opérateur pour la Sureté (MERMOS), Connectionism Assessment of Human Reliability (CAHR), Cognitive Reliability and Error Analysis Method (CREAM) (Zhang and Tan, 2018), Systematic Human Error Reduction and Prediction Approach (SHERPA) (Bligård and Osvalder, 2014) and A Technique for Human Error Analysis (ATHEANA). In such methods, the assessment of the human error probability (HEP) is based on lists of error probabilities and uncertainty factors (Kirwan, 1998). Values for these lists were determined by observations and tests in simulation and fake-real situations.

To evaluate human errors scenarios, fault tree analysis can be used as it was done with maintenance

procedures of a pump in (Noroozi *et al.*, 2014). *Quantitative* data of human error prediction tables as it was first introduced by THERP are the empirical basis of this method. Also the output of these methods can be presented quantitative.

However, the data presented by human error prediction tables consider only the operator and not the human machine interface, in particular not the interface of complementary protective measures.

Nonetheless, these methods deliver interesting approaches to quantify qualitatively estimated data and interpret data from test scenarios. They can furthermore be used to define test settings. Pursuing the project it is proposed to further scrutinize the HRA methods (Petruni *et al.*, 2017).

Design requirements are not in the scope of these methods, because of their focus on human error prediction. Because of this focus, they are of limited applicability for the assessment of the mechanical part of the human-machine interface and its interaction with humans.

4.2 *Standard technical safety methods*

ISO 31010 provides risk assessment methods. They are categorized in look-up methods, statistical methods, control assessment methods, function analysis, scenario analysis, supporting methods and others (IEC/ISO 31010, 2009; Tixier *et al.*, 2002).

Look-up methods basically use hazards lists to generate input for further analysis. It is common to use these methods in early phases of risk identification, construction or design processes. Therefore, they deliver only indications if risks are possible or not. Exemplarily methods are check-lists, preliminary hazard analysis (PHA) and brainstorming (IEC/ISO 31010, 2009).

Delivering quantitative output is not their purpose. This leads to no approaches to quantify the output. Through their simple structure they are not qualified for the proof of reliability of complementary protective measures. Nevertheless, in an assessment process they are seen as a valid starting point and to achieve completeness of risk assessments.

Control assessment methods take into account all layers of protection to assess risks (IEC/ISO 31010, 2009). Exemplarily methods are bow-tie analysis and layer of protection analysis (LOPA). These methods are also used for risk assessment in the field of functional safety as in IEC 61508.

These methods evaluate systems and their different protection measures concluding the wide range of technically, intercompany, organization, emergency response and so on (Gowland, 2006).

Typical output data are management recommendations and priorizing of risk measures. The kind of output cannot be used for SIL quantification. Therefore a transfer to evaluate complementary protective measures is not indicated.

Due to their alignment on global protection layers, the methods do not focus on single design factors.

Function analysis deals with the functional units or main functions of technical systems. Exemplarily methods are Hazard and Operability study (HAZOP) and Hazard Analysis and Critical Control Points (HACCP) (IEC/ISO 31010, 2009).

The approach of these methods is that the considered system is divided into components, units or subsets. Nominal functions are assigned to these subgroups.

Subsequently, parameters are formulated that lead to a deviation of the nominal function. The qualitative discussion takes place on the extent to which the identified deviations lead to an increase of risks (IEC/ISO 31010, 2009).

These methods are also used for risk assessment in the field of functional safety as in IEC 61508. However, the focus is on the identification of risks that could be minimized later on by control-independent protection measures. This approach therefore determines the necessary degree of risk reduction and thus the requirements for standard safety integrated functions. They do not assess the achieved value of the actual technical implementation or the human-machine interface.

Finally in functional analysis, a level of the necessary risk reduction is given. That leads to further requirements of the assessment process, like to use statistical data in the reliability block diagram describing the proposed architectures.

According to the discussion given, this method type cannot be used to proof the achieved level of reliability of the human-machine interface. Due to this, also the scope of the method does not cover design and performance shaping factors.

Other methods are e.g. consequence/probability matrix and risk indices.

These methods assign the values of a system in matrixes or tables to determine a risk index.

Delivering quantitative values depends on quantitative input data. When using convertible data a quantitative overall value can be determined. Normally semi-quantitative scales are used (IEC/ISO 31010, 2009).

Due to their general approach these methods are not specified, instead they only rank input data. In general these methods are no stand-alone method but in combination with other methods they will be interesting for complementary protective measures. In this way it would be possible to rank the design data.

Scenario analysis summarizes scenario based methods. To define solutions these methods define possible effects and causes for defined scenarios.

In doing so, top event can be analyzed but not all risks can be identified.

Event tree analysis (ETA) follows an inductive approach to identify consequences which result from an initiating event. Due to this approach, this method is not useable for risk evaluation (Stapelberg, 2009).

Fault tree analysis (FTA) is a deductive process. By subdividing a system into its basic components this method identifies factors which contribute to an undesired event (Vesely, Goldberg, Roberts, Haasl, 1981). Beside software and hardware failures also human error failures can be considered.

In order to be able to create fault-trees there must be detailed knowledge of the system. To obtain additional detailed quantitative values, quantitative input data is required. For complementary protective measures these cannot be provided.

By using quantitative input data FTA is able to predict failure probability for a specific event. However, used stand-alone, it cannot be used to predict SIL values.

Due to its general approach, it is possible to use FTAs for many application areas. The scope can be laid on design and performance shaping factors. In this way, fault tree analysis could be used to assess design failures which can lead to a delayed response.

Supporting methods are methods guided by a survey manager. In contrast to normal survey methods, negative group effects can be counteracted with a survey manager. Exemplarily methods are structured or semi structured interviews and the Delphi technique (IEC/ISO 31010, 2009).

These methods deliver no quantitative data but they can be used to provide input data for other methods. In combination with fault tree analysis, using supporting methods could identify factors which contribute to an undesired event.

Supporting methods deliver no data to proof the achieved SIL level.

Supporting methods are applicable to every area. Through the survey manager, the scope can be laid on design and performance shaping factors.

Besides contributing input for fault tree analysis, supporting methods could be used in test scenarios to determine subjective opinions by the test persons.

Statistical methods use mathematical models. By using reliability data (e.g. failure rates), these methods are able to predict the possible development of a protective measure. Exemplarily methods are Markov chains, Monte Carlo simulation and Bayes nets (IEC/ISO 31010, 2009).

In the field of functional safety, methods like parts-count and parts stress are also used in Failure Modes Effects and Diagnostic Analysis (FMEDA) (Smith, 2017) and fault tree applications. Often such methods highly resolve, at least for simple components (as opposed to complex components such as ASICs or microcontrollers). With respect to failure modes, e.g. short cut, drift, and interruption failures. With such very standardized methods an overall reliability value of the safety functions is calculated by single values form each component as well as diagnostic coverage, safe failure fraction and hardware failure tolerance values (Smith, 2017).

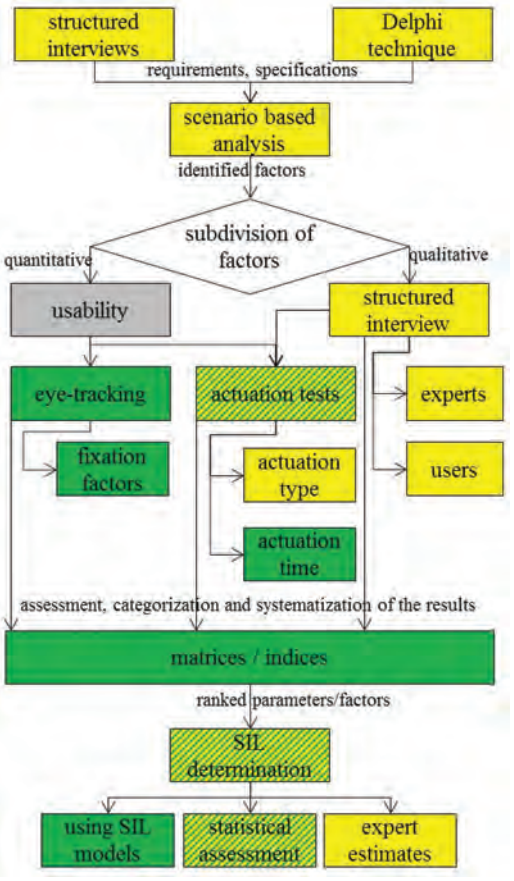


Figure 2. Planned method chain.

Statistical methods cannot be used in the first instance, if there is no quantitative reliability data available. However, if such values can be determined by other methods, system state analysis and reliability prediction methods can be used to determine the listed overall reliability values.

Due to the generic mathematical model these inductive and deductive methods are suitable for almost every area of application.

A special coverage of design or performance shaping factors is not possible, but if corresponding input data is connected to these mathematical models, they can consider these factors. For

instance, depending on the design of the physical interface, certain failure modes are feasible.

5 EXTENSION BY MEASURING METHODS

As shown in section 4, the influence of the physical design of the mechanical human-machine interfaces for complementary protective measures is neither sufficiently taken into account by methods for human reliability analysis nor for technical safety.

This fact requires broadening the reviewed methods with a new method of measurement in order to evaluate the interface. The result should allow for a more precise evaluation of this type of protective measures and for the development of the silk models.

The following method chain is proposed for assessing the physical human machine interface of complementary protective measures. The listed methods are subdivided into quantitative and qualitative (cf. Figure 3) assurance methods and their use for the assessment is explained. Subsequently, the possibility of transferring the gained data to SIL values is explained.

5.1 *Qualitative methodological approaches*

First, qualitative methods are used to evaluate requirements and specifications that influence the reliability of the mechanical part of a human machine interface as complementary protective measures. These have to adhere to legal requirements as defined in directive 2006/42/EG and normative regulations.

With these, input scenario based analyses are conducted to identify contributing factors which influence fault free and highly available usability of complementary protective measures.

Figure 3. Exemplarily heatmap for emergency stop button.

Exemplary factors are size, shape and color. In the further procedure, these factors are subdivided in qualitative- and quantitative measureable factors.

Qualitative factors are analyzed e.g. by structured expert interviews and test persons of the test procedure. In addition, types of actuation are video—monitored during the test procedure.

The aim is to evaluate the subjective impressions of the evaluators and the test persons and with the measured values from the usage test. Compliance with legal and normative regulations is part of the evaluation process.

5.2 *Quantitative methodological approaches*

The quantitative measurable factors determined from the first steps will be examined in depth.

One auspicious approach to measure these parameters is eye-tracking (Harezlak et al., 2014; Guo et al., 2016; Khalighy et al., 2015). This video analysis approach from the area of usability can be used to measure different parameters include (Khalighy *et al.*, 2015):

- Appropriateness, which indicates what design elements are preferred by consumers for a certain function
- Novelty, which includes unexpected and unexperienced design elements for a certain function

Eye-tracking also provides quantitative output such as number of fixations, standard deviation of fixation or common area of fixation (Khalighy *et al.*, 2015; Takahashi *et al.*, 2017). An example can be seen in Figure 3.

Easy to use in simulations, fake test scenarios and real test scenarios (Hahn and Lüdtke, 2013) are further advantages of this method.

Eye tracking can be used to measure the reaction and the behavior of humans but also their interaction with the mechanical elements of human-machine interface (Khalighy *et al.*, 2015; Kim *et al.*, 2017). Due to this eye tracking is good suitable to measure the defined design requirements identifiability, visibility and accessibility (cf. Machinery Directive 2006/42/EC, 2006).

In addition to the parameters that can be measured in the eye tracking procedure, the actuation time is also measured during the test settings, since it can be used to assess accessibility, for example.

Depending on their importance, certain qualitative and quantitative factors are evaluated, categorized and systematized. Matrices or indexes can be used for this purpose. The final goal is to develop a way to make use of the parameters examined in SIL assessments.

Table 1. Comparison table based on IEC 61508, 2011.

Actuations time	No. of fixations	Area of fixation	Factor
12 ms	16	3 cm²	0,01
20 ms	10	5 cm²	0,05
….	….	….	….

5.3 *Transfer to SIL*

Transferring the measured data, e.g. reaction and response times, to values of functional safety can be achieved in several ways.

- One to one transfer (using SIL models)
- Statistical assessment
- Expert estimations

An attempt can be a one-to-one transfer into existing SIL values. But so far, there is no indication that such a transfer is easily possible beyond the assessment of the fulfillment of specified requirements of safety functions, e.g. reaction within 1 sec.

It is more likely that statistical assessment of multiple eye tracking experiments is necessary. In this way, probabilities of failure of the human or the physical human-machine interface are feasible.

The measurements can be used to determine comparison tables similar to the table for calculating the beta factor according to IEC 61508. In this way, an analogical transfer of the results can be proposed. A simplified example table is shown in Table 1.

Experts can estimate frequencies of the observed reaction types, thus resulting in an overall reliability assessment. This approach can be further supported by historical accident data analysis.

6 CONCLUSION

Complementary protective measures need a proof of reliability according to IEC 61508. It is needed for overall risk assessment as well as overall safety evaluation. To deliver a proof of reliability quantitative data is required.

It was shown that there is a lack of such data for complementary protective measures.

After reviewing typical methods from the field of human reliability analysis and technical safety no directly applicable method is identified. However, several methods are supporting the quantification of the reliability of the mechanical human-machine interface.

To deliver quantitative data the use and extension of standard existing methods are described. To this end a tool chain was proposed with the main methods intended to be used.

In particular, the video analysis based tool eye tracking is identified as promising key method, to deliver the searched data.

Since it does not directly generate reliability data, it was also discussed how the measured values can be further evaluated and combined with standard reliability values from IEC 61508.

In conclusion, a way forward the present work sketches a way forward how to address the challenge of the reliability quantification for physical human machine interfaces as complementary safety functions for safety relevant systems.

REFERENCES

Bligård, L.-O. and Osvalder, A.-L. (2014), "Predictive use error analysis—Development of AEA, SHERPA and PHEA to better predict, identify and present use errors", *International Journal of Industrial Ergonomics*, Vol. 44 No. 1, pp. 153–170.

Di Pasquale, V., Miranda, S., Iannone, R. and Riemma, S. (2015), "An HRA-based simulation model for the optimization of the rest breaks configurations in human-intensive working activities", *IFAC-PapersOnLine*, Vol. 48 No. 3, pp. 332–337.

Gehlen, P. and Rudnik, S. (2015), *Not-Halt oder Not-Aus?: Eine Erläuterung unter Berücksichtigung von DIN EN 60204-1 (VDE 0113-1) und DIN EN ISO 13850, VDE-Schriftenreihe—Normen verständlich*, Vol. 154, VDE-Verl, Berlin, Offenbach.

Gowland, R. (2006), "The accidental risk assessment methodology for industries (ARAMIS)/layer of protection analysis (LOPA) methodology: a step forward towards convergent practices in risk assessment?", *Journal of hazardous materials*, Vol. 130 No. 3, pp. 307–310.

Guo, F., Ding, Y., Liu, W., Liu, C. and Zhang, X. (2016), "Can eye-tracking data be measured to assess product design?: Visual attention mechanism should be considered", *International Journal of Industrial Ergonomics*, Vol. 53, pp. 229–235.

Hahn, A. and Lüdtke, A. (2013), "Risk Assessment of Human Machine Interaction for Control and eNavigation Systems of Marine Vessels", *IFAC Proceedings Volumes*, Vol. 46 No. 33, pp. 368–373.

Harezlak, K., Kasprowski, P. and Stasch, M. (2014), "Towards Accurate Eye Tracker Calibration—Methods and Procedures", *Procedia Computer Science*, Vol. 35, pp. 1073–1081.

IEC 60947 (2015), *Low-voltage switchgear ans controlgear—Part 1:General rules (IEC 60947-1 +A1:2010 + A2:2014); German version EN 60947-1:2007 + A1:2010 + A2:2014* No. 60947-1, VDE Verlag, Berlin.

IEC 61508 (2011), *Functional safety of electrical/electronic/programmable electronic safety-related systems* No. IEC 61508, 2nd ed, International Electrotechnical Commission, Geneva, Switzerland.

IEC 61511 (2004), *Functional safety—Safety instrumented systems for the process industry sector* No. IEC 61511.

IEC/ISO 31010 (2009), *Risk management—Risk assessment techniques* No. IEC/ISO 31010.

ISO 13850 (2014), *Sicherheit von Maschinen—Not-Halt—Gestaltungsleitsätze (ISO/DIS 13850:2014);*

Deutsche Fassung prEN ISO 13850:2014 No. 13850, Beuth Verlag Gmbh, Berlin.

ISO 12100 (2010), *Safety of machinery—General principles for design—Risk assessment and risk reduction; German version EN ISO 12100:2010* No. ISO 12100, Beuth.

Khalighy, S., Green, G., Scheepers, C. and Whittet, C. (2015), "Quantifying the qualities of aesthetics in product design using eye-tracking technology", *International Journal of Industrial Ergonomics*, Vol. 49, pp. 31–43.

Kim, A.R., Park, J., Kim, Y., Kim, J. and Seong, P.H. (2017), "Quantification of performance shaping factors (PSFs)' weightings for human reliability analysis (HRA) of low power and shutdown (LPSD) operations", *Annals of Nuclear Energy*, Vol. 101, pp. 375–382.

Kirwan, B. (1998), "Human error identification techniques for risk assessment of high risk systems—Part 1: review and evaluation of techniques", *Applied Ergonomics*, Vol. 29 No. 3, pp. 157–177.

Machinery Directive 2006/42/EC (2006), *The European Parliament and the Council of the European Union*.

Noroozi, A., Khan, F., MacKinnon, S., Amyotte, P. and Deacon, T. (2014), "Determination of human error probabilities in maintenance procedures of a pump", *Process Safety and Environmental Protection*, Vol. 92 No. 2, pp. 131–141.

Petruni, A., Giagloglou, E., Douglas, E., Geng, J., Leva, M.C. and Demichela, M. (2017), "Applying Analytic Hierarchy Process (AHP) to choose a human factors technique: Choosing the suitable Human Reliability Analysis technique for the automotive industry", *Safety Science*.

Smith, D.J. (2017), *RELIABILITY, MAINTAINABILITY AND RISK: Practical methods for engineers*, ELSEVIER BUTTERWORTH-HEIN, [Place of publication not identified].

Stapelberg, R.F. (2009), *Handbook of Reliability, Availability, Maintainability and Safety in Engineering Design, SpringerLink Bücher*, Springer London, London.

Sträter, O. (1997), *Beurteilung der menschlichen Zuverlässigkeit auf der Basis von Betriebserfahrung*, Gesellschaft für Anlagen—und Reaktorsicherheit, Köln.

Swain, A. and Guttmann, H.E. (1983), *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications: NUREG/CR- 1278*.

Takahashi, R., Suzuki, H., Chew, J.Y., Ohtake, Y., Nagai, Y. and Ohtomi, K. (2017), "A System for Three-Dimensional Gaze Fixation Analysis Using Eye Tracking Glasses", *Journal of Computational Design and Engineering*.

Tixier, J., Dusserre, G., Salvi, O. and Gaston, D. (2002), "Review of 62 risk analysis methodologies of industrial plants", *Journal of Loss Prevention in the Process Industrie*, No. 15, pp. 291–303.

Vesely, Goldberg, Roberts, Haasl (1981), *Fault Tree Handbook: Systems and Reliability Research*, NUREG-0492, U.S. Government Printing Office, Washington.

Zhang, R. and Tan, H. (2018), "An integrated human reliability based decision pool generating and decision making method for power supply system in LNG terminal", *Safety Science*, Vol. 101 No. Supplement C, pp. 86–97.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Simulation analysis of aerodrome CNS system reliability

M. Kozłowski, J. Skorupski & A. Stelmach
Faculty of Transport, Warsaw University of Technology, Poland

ABSTRACT: Ensuring safety, regularity and continuity of air traffic are priority objectives of aeronautical regulations and procedures. They concern, among others, Air Traffic Control services (ATC), Communications, Navigation and Surveillance systems (CNS), Visual/Instrumental Flight Rules (VFR/IFR), minimal Meteorological Flight Conditions (FMC) and air traffic procedures. This paper focuses on aerodrome CNS systems reliability in relation to aerodrome traffic operations and ATC procedures. The aerodrome CNS system has complex technical and functional structure. The elements forming the system are very reliable and accurate. However, the possibility of their operational use depends not only on the technical state. Aerodrome ATC unit (TWR) can issue a clearance for performing an operation (approaching, landing, take-off and climbing) only when the requirements meet the actual conditions specified by FMC, ATC procedures and air traffic rules, approved Flight Plan and operational status of CNS devices. This means that different combinations of operational situations and reliability status of CNS devices should be considered. In this paper Petri nets are proposed to model the aerodrome CNS system reliability structure. The model was used to perform simulation analysis of CNS system reliability for different air traffic and meteorological conditions. Applicability of the method has been shown on the example for selected scheduling season.

1 INTRODUCTION

Ensuring safety, regularity and continuity of aerodrome operations is a fundamental requirement of aviation law. This results in specific tasks for the aerodrome managing body. They include appropriate planning and maintenance of airport infrastructure elements as well as implementation of appropriate technical and operational procedures. The realization of these tasks is connected with the necessity of incurring costs, which are an important element of the budget of the aerodrome operator. Aviation law regulations do not specify the requirements for the categories of procedures and navigational aids that must be implemented and installed at a given aerodrome. This is the decision of the aerodrome operator. These decisions are related to the safety and economic objectives taken into account in the aerodrome's investment plans. Usually, investment decisions are based on cost-benefit analyzes, which consider the issue of aerodrome continuity. It can be effectively analyzed on the basis of reliability theory with the use of operational readiness measures (Kozłowski, 2015, 2016). The obtained results allow defining adequate investment processes including cost aspects. This also applies to the decision to leave actual or increase the category of procedures and navigation aids, which entails significant costs but gives the opportunity to increase revenues from serviced air traffic and transport. An element of these analyzes is the assessment of the reliability of the CNS system in relation to meteorological flight conditions and planned air traffic.

This work aims at finding an effective and objective method of assessing the adequacy of CNS infrastructure to FMC and ATC procedures and traffic rules. This will help identification of necessary changes and verification of investment plans.

2 AERODROME INFRASTRUCTURE

2.1 *Airside infrastructure*

Aerodrome is an area on a land, including any buildings, installations and equipment intended to be used for the arrival, departure and surface movement of aircraft (Annex 14 ICAO). The aerodrome infrastructure includes two basic groups of elements:

– the movement area,
– communications, navigation and surveillance (CNS) devices.

All elements of the aerodrome infrastructure has specific characteristics which must meet the requirements and be consistent with the specifications (nominal values and tolerances of parameters) given in legal regulations. These characteristics define technical, operational and

reliability parameters, and their values must be consistent with the established requirements specified, e.g. in: Annex 10 ICAO, Annex 11 ICAO, Annex 14 ICAO.

Movement area is that part of an aerodrome to be used for the take-off, landing and taxiing of aircraft, consisting of the maneuvering area (including runways and taxiways) and the aprons (Annex 14 ICAO). A runway is rectangular area on the aerodrome intended for the landing and take-off aircraft operations. Depending on the types of installed CNS and their parameters, two types of runways are established: non-instrument runways (NI) intended for the operation of aircraft using visual approach procedures and instrument runways for instrumental approach procedures. The latter may be divided into following types:

- non-precision (I-NP) runway, served by visual aids and a non-visual aid providing directional guidance adequate for a straight-in approach,
- category I (I-PI) runway, served by ILS or MLS and visual aids intended for operations with a decision height (DH) not lower than 60 m (200 ft) and either a visibility not less than 800 m or a runway visual range (RVR) not less than 550 m,
- category II (I-PII) runway, served by ILS or MLS and visual aids intended for operations with a DH lower than 60 m (200 ft) but not lower than 30 m (100 ft) and a RVR not less than 300 m,
- category III (I-PIII A/B/C) runway, served by ILS or MLS and: (A) — intended for operations with a DH lower than 30 m (100 ft), or no decision height and a RVR not less than 175 m, (B) — intended for operations with a DH lower than 15 m (50 ft), or no DH and a RVR less than 175 m but not less than 50 m, (C) — intended for operations with no DH and no RVR limitations.

2.2 *CNS infrastructure*

CNS systems are three technologies that are used at the Air Traffic Management (ATM) to perform air operations.

1. Aerodrome control radio station (COM) is the standard aerodrome air-ground communication device ensuring two-way voice communication between TWR and aircraft.
2. The aerodrome navigation systems include the following device groups:
   - visual aids for navigation (VAN),
   - radio aids for navigation (NAV).

   The standard aerodrome VAN aids consist of:
   - indicators and signaling devices,
   - markings,
   - signs (mandatory and information),
   - lights: approach lighting systems (ALS) - Calvert or ALPA-ATA, visual approach slope indicator systems—VASIS or PAPI, lights installed on runways, taxiways and aprons.

   The standard aerodrome NAV aids consist of:
   - non-directional radio beacon (NDB),
   - the VHF omnidirectional radio range (VOR),
   - distance measuring equipment (DME),
   - instrument landing system (ILS) or microwave landing system (MLS).
3. The surveillance radar (SUR) is an equipment used by air traffic control (ATC) to determine the position of an aircraft in range and azimuth. The standard aerodrome SUR systems used by TWR are:
   - primary surveillance radar, which uses reflected radio signals (PSR),
   - secondary surveillance radar, which uses transmitters/receivers (interrogators) and transponders (SSR).

All of the CNS facilities, in accordance with the requirements of legal regulations (Annex 10 ICAO), are characterized by very high technical reliability. However, their operational status depends on many various factors. These factors determine operational reliability considered as operational readiness. The CNS facility failure is understood as any unanticipated occurrence which gives rise to an operationally significant period during which a facility does not provide service within the specified tolerances.

Aerodrome CNS devices reliability depend and is achieved by a combination of factors, especially:

- flight meteorological conditions (FMC), expressed in terms of visibility, distance from cloud, and ceiling,
- aerodrome infrastructure maintenance programs and procedures (AMP),
- level of redundancy in the reliability structure.

The general formula of the CNS facility (for which the failures follow a Poisson distribution) reliability (as a percentage) – R, defined as a probability that the facility will be operative within the specified tolerances for a time t formula is expressed as:

$$R = 100e^{-t/MTBF} \tag{1}$$

where:
e = base of natural logarithms,
t = time period of interest,
$MTBF$ = mean time between CNS device failures.

The following reliability values of CNS devices were adopted, as shown in Table 1.

Table 1. Exemplary values of CNS facility reliability.

COM	NAV	VAN	SUR
R_{COM} = 0,9995	R_{NDB} = 0,9950 R_{VOR} = 0,9900 R_{DME} = 0,9900 R_{ILS} = 0,9970	R_{ALS} = 0,9990 R_{PAPI} = 0,9980 R_{RWYI} = 0,9995 R_{TWYI} = 0,9990	R_{SUR} = 0,9800

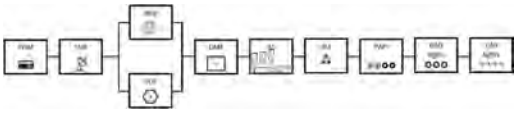


Figure 1. Operational structure of aerodrome CNS system.

CNS devices are needed in accordance with established procedures to ensure safe operations in aerodrome traffic, their functional structure is shown in Figure 1. This structure has been mapped in the model presented in Section 4.

3 AERODROME TRAFFIC

3.1 *Aircraft operations and procedures*

Aerodrome traffic is defined as all traffic on the aerodrome maneuvering area and all aircraft flying in the vicinity of an aerodrome (Annex 11 ICAO). Aircraft may flight in accordance with the visual flight rules (VFR) or instrument flight rules (IFR), and conduct approach to landing, landing, taxiing, take-off and departure operations. The largest range of operational requirements is assigned to approach and landing operations.

In this paper, special attention is paid to approach and landing procedures, in particular to instrument approach procedure (IAP), i.e. a series of predetermined maneuvers from the initial approach fix or from the beginning of a defined arrival route to a point from which a landing can be completed (ICAO Doc 4444). Instrument approach procedures are classified as follows:

- non-precision approach (NPA), it utilizes lateral guidance but not the vertical guidance,
- approach with vertical guidance (APV), it utilizes lateral and vertical guidance but does not meet the requirements established for precision approach and landing operations,
- precision approach (PA) using precision lateral and vertical guidance with minima as determined by the category of operation.

Some standard procedures are usually implemented in aerodrome traffic (Annex 11 ICAO):

- standard instrument arrival (STAR) - an IFR arrival route linking a point on an ATS route with a point from which an instrument approach procedure can be commenced,
- standard instrument departure (SID) - an IFR departure route linking the aerodrome with a specified point on a designated ATS route.

Analysis presented in this paper applies to a controlled aerodrome, where the air traffic control service is provided by TWR (ICAO Doc 4444). To perform an operation in controlled aerodrome traffic, the aircraft needs TWR clearance, i.e. authorization to proceed under specified conditions. Before issuing the clearance, TWR checks the compliance between the requirements and specifications, procedures, operational status of CNS devices, traffic situation, current FMC and aerodrome operating minima (AOM).

3.2 *Aerodrome operating minima*

Aerodrome operating minima express the limits of usability of an aerodrome for:

- take-off, expressed in terms of RVR or visibility and cloud conditions,
- landing, expressed in terms of RVR and DH or minimum descent height (MDH) and cloud conditions as appropriate to the category of the operation (Annex 6 ICAO).

The Aerodrome Operator (AO) establishing the AOM should take into account and consider the following factors:

- types, performance and handling characteristics of the aircraft,
- dimensions, characteristics and categories of the runways,
- categories and performance of the available CNS devices and systems,
- aerodrome ATC/TWR procedures,
- obstacles locations and dimensions and obstacle clearance height (OCH) and necessary MDH,
- shape and sizes of obstacle free zone (OFZ) and normal operating zone (NOZ),
- descent profile determined for vertical guidance during a final approach (glide path—GP).

To ensure the adequate level of safety the Aerodrome Operator shall specify incremental values for height of cloud base (CB) and visibility (RVR), to be added to the operator's established AOM. The following AOM were adopted, as shown in Table 2.

For the given RWY and approach procedure categories and the for aerodrome operating minima, the minimal alternative path of operational readiness were determined, as shown in Figure 2.

Table 2. Controlled aerodrome operating minima.

RWY Cat/ PA Cat	min RVR	min CB (DH)
NI	5000 m	450 m
I-NP	550 m	150 m (OCH)
I-PI	550 m	60 m
I-PII	300 m	30 m
I-PIIIA	175 m	30 m
I-PIIIB	175 m	15 m
I-PIIIC	0 m	0 m

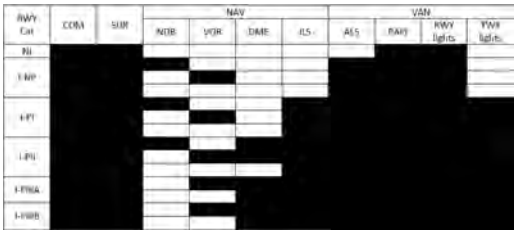


Figure 2. Identified minimal paths of operational readiness at controlled aerodrome Cat IIIB.

Table 3. Time percentage of FMC occurrence in SS and SW.

FMC	Scheduling period SS	WS
RVR > 5000 m and CB ≥ 450 m	18%	6%
5000 m > RVR ≥ 550 m; 450 m > CB ≥ 150 m	27%	12%
5000 m > RVR ≥ 550 m; 150 m > CB ≥ 60 m	18%	19%
550 m > RVR ≥ 300 m; 60 m > CB ≥ 30 m	13%	23%
300 m > RVR ≥ 175 m; 60 m > CB ≥ 30 m	10%	16%
300 m > RVR ≥ 175 m; 30 m > CB ≥ 15m	11%	13%
175 m > RVR and 15 m > CB	3%	11%

3.3 *Operation of the aerodrome*

The subject of research presented in this paper is in fact the airport i.e. aerodrome intended for service commercial air transport.

The operation process of the airport is carried out in two scheduling period i.e. either the summer (SS) or winter (WS) scheduling season as used in the schedules of air carriers. The operating conditions and structure of the air traffic are different in the scheduling seasons.

In subsequent studies, the following data and assumptions were adopted, as shown in Tables 3 and 4.

Table 4. Percentage of IFR and VFR operations in SS and SW.

FR	Scheduling period SS	WS
IFR	90%	95%
VFR	10%	5%

4 MODEL OF AERODROME CNS SYSTEM

As indicated above, calculation of the operational readiness of the aerodrome depends not only on the reliability of the CNS system, but also on other factors described by random variables, the nature of which has been identified. The form of the density function of these random variables indicates the need to use simulation methods to study the operational readiness of the aerodrome.

Model of aerodrome CNS system for reliability and operational readiness analysis was created using Petri nets (Jensen, 1997, Marsan et al, 1999, Reisig, 2013). Originally, they were developed for modeling computer systems working synchronously. However, their high versatility has resulted in their having many other applications in recent years, including modeling and support of air traffic management processes (Davidrajuh & Lin, 2011, Oberheid & Söffker, 2008, Skorupski, 2011, 2015a, 2016, Werther et al., 2007) and systems reliability (Vismari & Camargo, 2011, Pinna et al., 2013, Song et al., 2017, Nývlt et al., 2015, Skorupski, 2015b).

The model of aerodrome CNS system has been implemented as a colored, hierarchical, priority Petri net. The network hierarchy is represented in the form of so-called "pages" responsible for different parts of the model: *I-NP* or *I-PIIIB*, *CNS*, *FMC*.

4.1 *General characteristics of Petri nets*

The basis for building a Petri net is a bipartite graph containing two disjoint sets of vertices called places (designated by ellipses) and transitions (rectangles). The arcs in this graph are directed. A characteristic feature of the graph used in Petri nets is that the arcs have to combine different types of vertices.

The set of places P corresponds to CNS systems operational states or FMC parameters observed at the airport. The set of transitions T corresponds to the generators of these values. The set of arcs defines the process of changing CNS and FMC parameters in subsequent experiments.

4.2 *Petri net for modeling airport operations*

The airport traffic model presented in this paper can therefore be written as

$$S_{AT} = \{P, T, I, O, H, M_0, \Gamma, C, G, E, B\} \quad (2)$$

where:
$M_0: P \to \mathbb{Z}_+ \times R$ – initial marking,
Γ – nonempty, finite set of colors,
C – function determining what color of tokens can be stored in a given place: $C: P \to \Gamma$,
G – function defining the conditions that must be satisfied for the transition before it can be fired; these are the expressions containing variables belonging to Γ, for which the evaluation can be made, giving as a result a Boolean value,
E – function describing the so-called weight of arcs, i.e. expressions containing variables of types belonging to Γ for which the evaluation can be made, giving as a result a multiset over the type of color assigned to a place that is at the beginning or the end of the arc,
$B: T \to \mathbb{R}_+$ – function determining the priority of transition t; this function applies only for transitions that are simultaneously active; in this situation a free choice of transition to be fired is possible.

4.3 *Mapping the aerodrome CNS system reliability*

The model of aerodrome CNS system reliability represents operational states of individual CNS devices, FMC parameters (visibility and height of cloud base) and aerodrome operating minima through the use of colored Petri nets. The main advantage of such approach lies in that the tokens located in places which determine the system states allow one to easily calculate the whole aerodrome CNS system reliability.

For example, in Figure 3 the reliability structure of the CNS system implemented for controlled aerodrome of category Cat IIIB is presented.

Places *COM*, *SUR*, *NDB*, *VOR*, *DME*, *ALS*, *PAPI*, *RWY*, *ILS*, *TWY* contain a token describing operational state of the corresponding system. Places *CNS*1 and *CNS*2 are responsible for storing the current reliability status of the entire CNS system. The determination of this state takes place in the *rel*() procedure. In addition, there is a place *S* whose task is to synchronize the control in the simulation program.

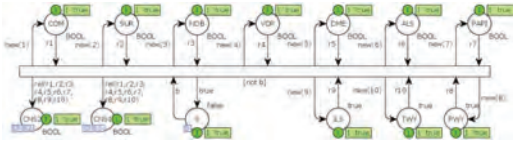


Figure 3. Petri net (*CNS* page) representing reliability structure of the CNS system of controlled aerodrome Cat IIIB.

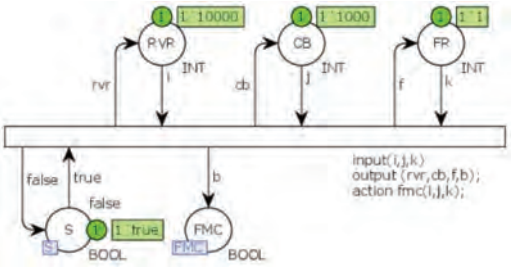


Figure 4. Petri net (*FMC* page) representing the flight meteorological conditions for controlled aerodrome Cat IIIB.

Figure 4 shows the part of the model responsible for simulating atmospheric conditions. It is represented by the *FMC* website.

Places *RVR*, *CB* and *FR* represent visibility, height of cloud base and flight conditions (IFR or VFR) respectively. The place *S* is used to synchronize the control (similarly as on the *CNS* page) and the place *FMC* stores information on whether the current meteorological conditions are sufficient to perform the landing operation in accordance with the current flight conditions. The determination of compliance of these conditions is implemented in the *fmc*() procedure.

4.4 *Computer tool in CPN Tools environment*

The presented model has been implemented as a computer software using CPN Tools 4.0 environment (Jensen et al. 2007). It is a very convenient tool because it allows at the same time creating a model, simulating at different input parameters and simultaneously analyze the results in the state space.

The model of aerodrome CNS system reliability is implemented as the hierarchical Petri net in which different parts of the model are created independently and during the simulation are synchronized by means of special mechanisms.

Figure 5 shows the page *I-PIIIB* responsible in the model hierarchy for calculating the results. Places marked with *CNS*1, *CNS*2 and *FMC* labels in the bottom left corner, called "fused places", are used to synchronize with relevant pages representing the remaining levels in the hierarchy. All fused places marked with the same label are identical, regardless of which part of the model they are placed in. The mechanism of fused places also

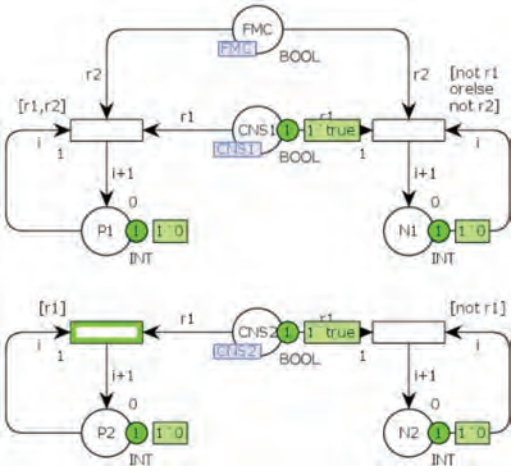


Figure 5. Petri net for calculation of the results.

gives the ability to synchronize the elements within one page of the model.

Place $P1$ counts those simulation experiments in which the landing operation could be performed, and the place $N1$ counts those experiments in which the meteorological conditions or the reliability state of the CNS system were insufficient to perform the operation. Similarly, the $P2$ and $N2$ places register situations of full reliability of all CNS system components and partial failure of the system.

The model and the computer tool have been validated with the use of real data by comparing the data obtained from measurements with the probabilities obtained from the model. Due to the volume of the paper, details of the validation are not shown.

5 SIMULATION EXPERIMENTS

Using the developed model and computer tool, a series of simulation experiments were performed to determine both the reliability of the CNS system at the airport and the possibility of performing air operations, i.e. operational readiness of the airport.

The first experiment was made for a poorly equipped airport of category Cat NP. The following elements of the CNS system must be able to perform the landing operation under IFR conditions: COM, SUR, ALS, PAPI, RWY lights and at least one of the following: NDB, VOR, DME. In addition, the FMC conditions must be at least 550 m visibility and 150 m cloud base. Landing in VFR conditions requires visibility of at least 5000 m and cloud base of at least 450 m.

The experiment consisted in simulating 10^7 landing operations taking into account the real parameters of random variables. The reliability of the CNS system obtained from the simulation is 0.976 and operational readiness is only 0.41.

The second experiment concerned a well-equipped airport of category Cat IIIB. The following elements of the CNS system must be able to perform the landing operation under IFR conditions: COM, SUR, DME, ILS, ALS, PAPI, RWY lights, TWY lights. In addition, the FMC conditions must be at least 175 m visibility and 15 m cloud base. As before, landing at VFR requires at least 5000 m visibility and at least 450 m cloud base.

The experiment consisted in simulating 10^7 landing operations taking into account the real parameters of random variables. The reliability of the CNS system obtained from the simulation is 0.962 and operational readiness is 0.857.

The reliability of the CNS system is slightly lower for the Cat IIIB airport, which is due to the larger number of necessary devices. At the same time, these devices allow for much more precise navigation, which allows to significantly increase operational readiness.

6 CONCLUSIONS

Comparison of the obtained results of conducted simulation experiments indicates the lack of dependence between the reliability of the CNS system and the aerodrome operational readiness. It proves that decisions regarding the modernization of the CNS infrastructure, and thus the change of the airport category and investment projects, should be made in the aspect of ensuring business continuity.

Obtained results and conclusions from the use of the Petri nets model will be used for further research on the determination the airport minimum business continuity and other parameters, such as minimum recovery time or maximum tolerable period of disruption (ISO 22301).

ABBREVIATIONS AND ACRONYMS

ALS – Approach lighting system.
AMP – Aerodrome maintenance program.
AO – Aerodrome operator.
AOM – Aerodrome operating minima.
APV – Approach with vertical guidance.
ATC – Air traffic control.
ATM – Air traffic management.
ATS – Air traffic services.
Cat. – Category.

CB – Cloud base.
CNS – Communications, navigation and surveillance.
COM – Communications.
DH – decision height.
DME – Distance measuring equipment.
FMC – flight meteorological conditions.
FR – Flight rules.
ft – Feet (dimensional unit).
ICAO – International Civil Aviation Organization.
IFR – Instrument flight rules.
ILS – Instrument landing system.
I-NP – Instrumental Non-precision approach.
I-P – Instrumental Precision approach.
MTBF – Mean time between failures.
NDB – Non-directional radio beacon.
NI – Non-instrumental approach.
PAPI – Precision approach path indicator.
R – reliability.
RVR – runaway visual range.
RWY – Runway visual range.
SID – Standard instrument departure.
SS – Summer schedule season.
STAR – Standard instrument arrival.
SUR – Surveillance.
TWR – Aerodrome control tower.
TWY – Taxiway.
VFR – Visual flight rules.
VOR – VHF omnidirectional radio range.
WS – Winter schedule season.

REFERENCES

Annex 6 ICAO *Operation of Aircraft,* Vol I *International Commercial Air Transport—Aeroplanes*.

Annex 10 ICAO *Aeronautical Telecommunications:* Volume I *Radio Navigation Aids,* VOL III *Communication Systems,* VOL IV *Surveillance and collision Avoidance Systems*.

Annex 11 ICAO *Air Traffic Services*.

Annex 14 ICAO *Aerodromes* Volume I *Aerodrome Design and Operations*.

Davidrajuh R. & Lin, B. 2011. Exploring airport traffic capability using Petri net based model, *Expert Systems with Applications*, 38 (9), 10923–10931.

ICAO Doc 4444 *Air Traffic Management*.

ICAO Doc 8400 *Procedures for Air Navigation Services—ICAO Abbreviations and Codes*.

ISO 22301:2012 *Societal security—Business Continuity Management Systems—Requirements*.

Jensen, K., 1997. *Coloured Petri Nets. Basic Concepts, Analysis Methods and Practical Use*. Berlin: Springer Verlag.

Jensen, K., Kristensen, L.M., & Wells, L. 2007. Coloured Petri Nets and CPN Tools for Modelling and Validation of Concurrent Systems, *International Journal on Software Tools for Technology Transfer* 9(3-4): 213–254.

Kozłowski M., 2015. *Aspect of reliability in airport business continuity management*, Journal of KONBiN 3(35)2015: 43–50.

Kozłowski M., 2016. *The concept of method for determining the minimum level of airport business continuity*, Journal of KONBiN 1(37)2016: 5–21.

Marsan, M.A., Balbo, G., Conte, G., Donatelli, S. & Franceschinis G., 1999. *Modelling with generalized stochastic Petri Nets*. Torino: Universita degli Studi di Torino, Dipartamento d'Informatica.

Nývlt, O., Haugen, S., and Ferkl, L., 2015. Complex accident scenarios modelled and analysed by Stochastic Petri Nets. *Reliability Engineering & System Safety* 142: 539–555.

Oberheid, H. & Söffker, D., 2008. Cooperative Arrival Management in Air Traffic Control—A Coloured Petri Net Model of Sequence Planning, in: *Applications and Theory of Petri Nets*, vol. 5062, K. van Hee and R. Valk, Eds. Springer Berlin/Heidelberg: 348–367.

Pinna, B., Babykina, G., Brinzei, N., & Petin, J.-F., 2013. Using Coloured Petri Nets for Integrated Reliability and Safety Evaluations. in: *4th IFAC Workshop on Dependable Control of Discrete Systems*: 19–24.

Reisig, W., 2013. *Understanding Petri nets. Modeling Techniques, Analysis Methods, Case Studies*. Berlin: Springer Verlag.

Skorupski, J., 2011. Method of analysis of the relation between serious incident and accident in air traffic, in: *Advances in Safety, Reliability and Risk Management*, G. Soares, Ed. London: CRC Press/Taylor & Francis: 2393–2401.

Skorupski, J., 2015a. The risk of an air accident as a result of a serious incident of the hybrid type, *Reliability Engineering & System Safety*, 140: 37–52.

Skorupski, J., 2015b. Airport operations safety assessment with the use of colored Petri nets. in: *Safety and Reliability of Complex Engineered Systems—Proceedings of the 25th European Safety and Reliability Conference, ESREL*, Zurich: CRC Press/Taylor & Francis.

Skorupski, J., 2016. The simulation-fuzzy method of assessing the risk of air traffic accidents using the fuzzy risk matrix. *Safety Science* 88: 76–87.

Song, H., Liu, J., & Schnieder, E., 2017. Validation, verification and evaluation of a Train to Train Distance Measurement System by means of Colored Petri Nets. *Reliability Engineering & System Safety* 164: 10–23.

Vismari, F.L., & Camargo, J.B., 2011. A safety assessment methodology applied to CNS/ATM-based air traffic control system. *Reliability Engineering and System Safety*, 96(7): 727–738.

Werther, N., Moehlenbrink, C., & Rudolph, M. 2007. Colored Petri Net based formal airport control model for simulation and analysis of airport control processes, in: *Proceedings of the 1st international conference on Digital human modeling* (ICDHM'07): 1027–1036.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Digitalization of the power business: How to make this work?

A.B. Svendsen & T. Tollefsen
Promaps Technology, Norway

T. Gjengedal
UiT—The Arctic University of Norway, Norway

M. Goodwin
UiA—The University of Agder, Norway

S. Antonsen
NTNU—The Norwegian University of Science and Technology, Norway

ABSTRACT: As a result of the digitalization of the power business in Norway and Europa, a lot of new possibilities and challenges arise. In 2014 an expert committee one outlined a proposal for the future grid company structure in Norway (Reiten, 2014). In addition, new technologies are being implemented in the system. Wind power, solar power, un-regulated small hydro power production, battery storage domestic and industrial and electrification of transport. Transmission System Operators (TSOs) have a responsibility to supply industry and communities with reliable electric power. However, the operators have been virtually blind to slowly occurring changes in the load profile that reduce the expected regularity of the power supply. This paper will focus on the possibilities and challenges the power business are facing. The paper will describe what technologies is needed i.e Real time probabilistic risk calculations, artificial intelligence, machine learning and smart grid technology. The main question is: can the power business and the introduction of new system tools manage without probabilistic risk calculation for making use of the digitalization and the corresponding big data?

1 INTRODUCTION

1.1 *History of the electric grid*

Modern Norway was built and industrialized by the fact that we managed to utilize rivers and waterfalls for power generation. Hydropower is still the cornerstone of the Norwegian power system, but wind power and solar energy is becoming an increasing part of the energy system. The grid has been developed over 150 years since the first small hydro plants were installed to supply small local industries. Hydro power plants were constructed over time as the industrial development moved forward. Initially the generation supplied local and regional consumers, but as transmission technology developed regions were connected via high voltage lines.

Now, the main grid is the most important part of the grid system, as failure here could mean power outage for very many consumers. The main grid was built largely from the 1950s to the 1980s. However, regional islands existed until 1994 when the main grid was finally established throughout Norway. Deregulation and competition was introduced in 1991 with the purpose of improving the socio-economic efficiency in the energy sector Now, the Norwegian main grid is aging and in the process of being replaced and upgraded by construction of new 420 kV lines in combination with digitalization of the power system. (Statnett, 2017).

1.2 *The change in the power system, smart grid, solar, wind, battery etc.*

The energy system is a critical part of a well-functioning society. Norway is largely electrified and power transmission is an important prerequisite for value creation.

Although the power grids are largely built as before, the power system changes at a rapid pace. Hence, the Transmission System Operator (TSO) must be an enabler and be prepared for the future. The Norwegian aging main grid is in the process of being replaced and upgraded. The power grid takes a long time to plan and build, and have a long lead time, which contrasts strongly with an energy sector in a rapid change. The load is increasing and more generation is being installed. The Green Certificate Scheme provides incentives to expand renewable power generation, including small scale

hydropower, wind power and solar power. Implementation of Automatic Meter Reading and Control Systems at the consumer level will allow for activation of consumer flexibility. Consumption patterns in the energy sector are changing rapidly (NVE, 2016).

Hence, the growing expansion of renewable energy and activation of flexible loads increases the complexities in balancing generation and demand in the power system. The energy-shifting and fast-ramping capability of energy storage has led to increasing interests in batteries to facilitate the integration of renewable resources (Amrouche, 2016).

The future power system will become even more dynamic and the need for real time information for monitoring the system status and for taking the proper control actions are increasing

1.3 *The need for a solution*

Global challenges regarding energy and climate change, the environment, safety, technology and renewable solutions, use and conservation of energy, use of batteries and the connection of electrical vehicles requires greater effort. The changing landscape of the power and utilities industry is resulting in new expectations for IT. Transmission system operators are struggling to fulfil their traditional mission of maintaining security of supply in a rapidly evolving environment driven by digitalization. Digital transformation is something that has become a common trend, and it is one that has reached the Power & Utilities sector moving rather quickly (Digitalization & Energy, 2016). The physical power system cannot function effectively without a well-functioning power market with smart ICT systems. The power system must be able to cope with the increasing variability in load and generation. Hence, in a complex power system, new solutions in the field of ICT and new market models are required to ensure the reliability and security of supply, to ensure that the transmission capacity is optimally utilized and that control actions are taken when needed.

2 DIGITALIZATION OF THE POWER BUSINESS—THE SOLUTION?

2.1 *The goal with digitalization*

The concept of a Digital Power System (DPS) has been discussed for many years. The DPS may be defined like the digital power system being the digital, figuration and real-time description and reappearance of physical structure, technical characteristic, management system as well as personal information system of a real power system which is in operation. The DPS will be able to make a significant contribution to administrating and decision-making more scientifically (Chakrabortty, 2017).

The share unregulated renewable power generation is rising and the power system is changing rapidly. Changes like this must be able to handle tomorrow's energy system. Hence, the reasons and goals for implementing the digital power systems are multiple:

- Monitoring g and controlling all components by equipping them by sensors
- Measure the condition of the power system flows, angle differences, stability margins, and hence the reliability and security
- improving security and stability online, online making and implementing economical operation strategy and carrying out emergency and anti-fault control, etc.
- Better utilization of the facilities
- Precise state information results in increased capacity and fewer faults
- More efficient maintenance and increased lifetime

In the end, the primary goal is to increase the value creation while maintaining the reliability and security of the system.

2.2 *Big data*

Data has always been an important asset in every industry. Since the early days of the information age, business intelligence and descriptive statistics have been used as the standard tools for extracting information and make important decisions from all kinds of collected data. However, as the cost of collecting, storing, and processing data has been dropping exponentially, the amount and the diversity of the data has reached the point where traditional approaches are no longer feasible. The term Big Data is often used to refer to any data that requires new techniques and tools in order for it to be processed and analyzed. Big Data could also be looked from the point of view of the new set of technologies that are helping to solve the challenges in collecting, managing, and analyzing Big Data. These technologies include cloud computing and cluster computing for data storage and manipulation, Artificial Intelligence (AI) and machine learning for data analysis (L'Heureux, 2017).

As in many other sectors big data analytics and machine learning are also getting involved in the energy sector and tools are being developed. They are for example used to forecast electricity demand at substation level, segment customers based on their power consumption patterns, implement demand response strategies, for power system condition monitoring and controls.

The value of big data may come from several use cases: as a source of analytics, as a source for control actions and as an enabler for new products and serv-

ices. An energy company could e.g. track, collect, and store all available data from their system from customers, from components, from system data, from GPS trails to geographical and meteorological data, then combine them together and use big data analytics to produce high value actionable insights and controls.

Use of big data technologies may also open up completely new business models and introduce new products and services in the energy sector.

2.3 *Smart grid*

The smart grid would be an enhancement of the electrical grid, using two-way communications and distributed intelligent devices (Smart Grids European Technology Platform, 2011). Two-way flows of electricity and information could improve the delivery network. A smart grid would allow the power industry to observe and control parts of the system at higher resolution in time and space. One of the purposes of the smart grid is real time information exchange to make operation as efficient as possible. It would allow management of the grid on all time scales from high-frequency switching devices on a microsecond scale, to wind and solar output variations on a minute scale, to the future effects of the carbon emissions generated by power production on a decade scale.

The management system in smart grid is the subsystem that provides advanced management and control services. Most of the focus aim to improve energy efficiency, demand profile, utility, flexibility, cost, based on the infrastructure by using optimization, machine learning and game theory. Within the advanced infrastructure framework of smart grid, more and more new management services and applications are expected to emerge and eventually revolutionize consumers' daily lives.

The protection system of a smart grid provides grid reliability analysis, failure protection, and security and privacy protection services. While the additional communication infrastructure of a smart grid provides additional protective and security mechanisms, it also presents a risk of external attack and internal failures (Pandey, 2017).

3 POWER SYSTEM OPERATION

3.1 *Balancing the system*

Electricity must be produced at the same time as the power is consumed. In addition, the production must be equal to the power consumed. This is called the instantaneous balance of the power system. The power market is the central tool for balancing supply and demand for power. The results of the daily pricing calculation in the day-ahead market are the basis for the Norwegian TSO Statnett's planning and maintenance of current balance in the following operating day. The continuous balancing of production and consumption is very important for the reliability of the system. In case of imbalances, system administrators implement measures to restore the balance, such as adjusting output or consumption.

Statnett has been given the system responsibility in the Norwegian power system. System Requirements in the Power System (Regulations on system responsibility in the power system, 2002) emphasize that the system operator shall provide frequency regulation, ensure instantaneous balance in the power system, develop market solutions that contribute to the efficient development and utilization of the power system, and to the greatest extent possible use of instruments based on market principles. The System Responsible company coordinates the operation of the power system, provides for the determination of capacity for the market, bottleneck handling and trade with other countries.

A well-designed power system has the following characteristics:

- Provide all consumption regardless of geographical location
- Provide consumption at all times
- Must be able to handle variability in consumption and production
- Supply must be of good quality and meet defined quality requirements
- Must be based on economic 'optimal' principle
- Must meet required and defined security goals

The delivered power must meet certain minimum delivery quality requirements. The following determines the quality:

- System frequency must be kept around the specified 50 Hz with variation within + –0,1 Hz
- The voltages are kept within narrow, prescribed limits around the normal value. Generally, the voltage variation should be within + –10% (or 5% in some systems)

To ensure that voltages and frequencies are kept within their limits, voltage and frequency regulation is required for efficient operation of the power system (Gjengedal, 2017).

3.2 *Traditionally operation and planning N-1*

Operating the network according to the N-1 criterion means that failure of a component does not result in interruptions in the supply to the end user. It is referred to as reduced reliability when the N-1 criterion is no longer met, in cases where it should normally be met. Statnett as a system administrator has the means through the system liability regulation (Regulations on system responsibility in the power system, 2002) to be able to change the grids configurations, as well as demand up-or

down regulation of production. Such means can help ensure operation according to the N-1 criterion. However, the authorities are not requiring that the main network should meet the N-1 operation safety at all times.

3.3 *Power system operation challenges*

In power system operation, traditionally slow-changing and predictable parameters are now changing fast. In addition, new system parameters are being introduced in the power system. This will influence the inherent properties of the system as well as the risk for outages. While the power system complexity is increasing, the operational available response time is decreasing fast. Example: The introduction of solar and wind production represents a power production which is difficult to manage when the wind stops or clouds cover the sun.

Combine this with a high degree of automation and smart grid technology, and the existing operational «know how» may not be sufficient to deal with the properties of the current and future power system.

What the neighboring power system is doing, will affect the current power system in regards of dynamics and risk. Until now power system operators have evaluated the system risk for loss of load qualitative, based on experience. This "gut feeling" based on such an experience approach, will not be valid in the future without nurturing new skills and competencies.

The first step is to be able to assess the system risk level equally for each power company. The only way to achieve this is by assessing this quantitatively with probabilistic approach.

The power business has lacked tools to evaluate quantitatively the system risk level in near real-time, and to assess possible risk reducing actions. This challenge was addressed by the GARPUR project 2017 (GARPUR, 2017).

3.4 *How will the future power system risk develop?*

New production with challenging properties, like solar and wind are put into the power system in an increasing rate. These alone will increase the risk in the system due to the characteristic intermittency property. It is expected more severe weather affecting the power system, resulting in higher risk. It is expected higher peak load as consequence of electrifying transportation and petroleum production, and this will increase risk in period of peak loads. New smart grid technology like disconnection of loads when needed, will reduce the risk in the system. The IoT (internet of things) technology may also increase the challenge of balancing the system due to rapid in and out connection of load driven by new sets of criteria not known by the system operators. This technology alone represents a threat to the system based on the possibility that third parties can hack into the equipment and connect/disconnect technology without anybody noticing, resulting possible outage of large areas.

Furthermore, battery storage is being introduced. This will most likely reduce the risk of outages in the system. In addition, artificial intelligence and machine learning is already on our door step. This can both help the system if done right, or increase the risk done wrong.

All these factors combined is a large order for the human mind to process in real-time. It is often seen that the risk driver is the combination of many seemingly unrelated factors and events, and not a single cause and effect scenario.

Example: A given power system has a unique dynamic characteristic and inherit risk property. The power system will typically be in a N-0 situation in periods with high peak load. Depending on the size of the load(s) affected by the reduced power delivery reliability, the total system risk is affected (see Figure 1). Above the red line the expected "not delivered energy" exceeds the acceptable level in regards of expected cost due to outage or the power companies' goals. Increase the system load for the same power system and the risk for outage will increase (see Figure 2). By operating the same

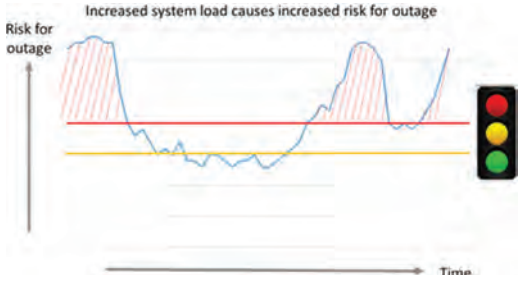


Figure 1. System risk for a given power system.

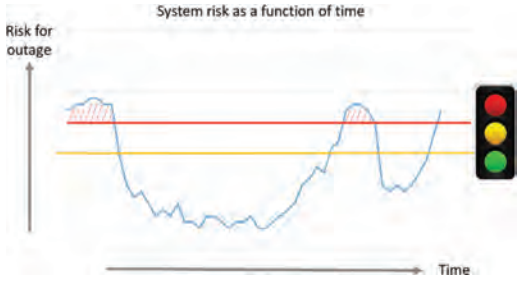


Figure 2. Increase only the load and the risk for outage increases.

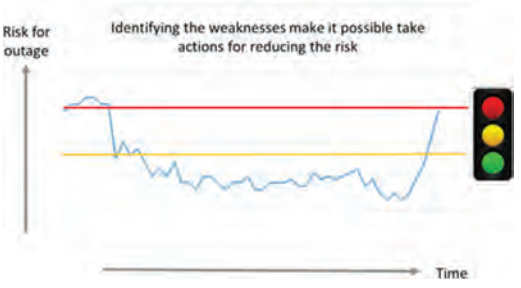


Figure 3. Operate the power system optimally, and the same increase in load can me managed in terms of risk for outage.

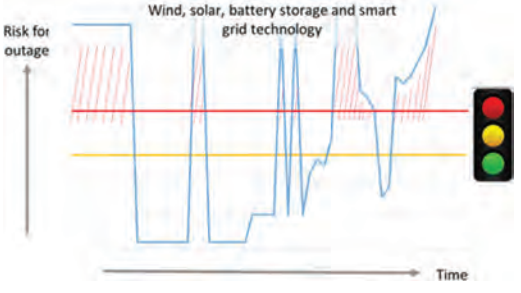


Figure 4. Prediction for future system risk development.

power system with the increased load more optimally, the total risk can be reduced to being within the acceptable risk level (see Figure 3).

With all the changes that are being introduced in the power system, our prediction is that the power system will experience system risk levels that goes from a very high level to the theoretical lowest level, and changing continuously by the minute (see Figure 4).

4 PROBABILISTIC RISK ASSESSMENT

4.1 *The starting point for a consistent risk management for the whole value chain*

To have a consistent risk management throughout to value chain, it is vital to be able to calculate the risk level of the current state of the power system. The current state will as it is in any power system analysis, be the base case for the next step analysis and evaluation. Furthermore, everything that is being done and planned, is being done for keeping the power system in operation. Therefore, the risk assessments done in the operation of a power system by operators or planning of operation, should be the foundation of risk evaluation to be communicated in the power grid company as input for all future evaluations and analysis.

4.2 *Identifying the power system inherit risk properties*

To be able to identify the power system's inherent risk properties, the following objects and parameter has to be included in a mathematical representation of the power system:

- Production – Spinning reserve – Location
- Production type
- Power system configuration
  - load flow
  - load demand
  - system dynamics
- Component reliability
- Maintenance interval and prioritizing
- Weather influence
- Energy storage possibilities and Smart grid technology
- System operators action and strategy
- Influence of other grid company's actions in their own power system

These factors have the characteristic of slow changing and fast changing properties. Since a power system is changing every minute of the year's 8760 hours, the total system risk graph will also vary by the minute. It is therefore, important to model the power system in great detail so that every change that occurs is reflected by the mathematical model. Furthermore, since historical data of the power system is stored (i.e. configuration, production, load flow, load level and failure rate for each component in the system) is available, validation of the mathematical model is possible to test against previous recorded risk levels.

4.3 *Calculation of the probabilistic risk level in near real time*

A short description of the calculation sequence in PROMAPS:

1. Calculate reliability of each grid segments. Each component in a grid segment can be described with multiple possible states, for instance Functioning, Intermediate fault or Lasting fault. PROMAPS use Markov models to represent each individual component in the grid segments:

$$\dot{p}_i = A_i p_i \tag{1}$$

   where A_i is a Markov model containing fault rates and repair rates.
2. It is possible to build reliability models of whole grid segments by simply combining all the

Markov models of each individual component as Kronecker sums, as follows:

$$A = A_1 \oplus A_2 \oplus ... \oplus A_n \quad (2)$$

is combined into common states, thus reducing the number of states in the grid segment model to a few unique states.

3. Calculate the probability of each system states. A system reliability model is calculated by combining all grid segments models using Kronecher sums, as described in the last step.
4. Discard all system states with probability below some probability threshold. The probability threshold is dependent on how many states should remain in the set for further assessments.
5. Calculate maximum power transmission capacity for each state in set.
6. Calculate expected power shortage at each load point
7. Calculate expected power supply reliability, and mean time between loss of supply
8. Calculate various auxiliary variables including economic data.

The analysis can be performed for various load profiles. For online reliability assessments, parts of the calculation sequence are repeated whenever new online data is available.

Promaps risk assessment principles for online calculations has been presented in detail in PMAPS2012 (Svendsen, 2012). The concept has also been researched in the recently completed pan-European project GARPUR (GARPUR, 2017). Common for these methodologies for real-time risk assessments is that they consist of two main parts:

1. Calculate the probability of all sequences of events in the power grid
2. Calculate the consequence of these events.

Since there is a "infinite" number of possible events in a power grid, there also need to be some principle of discarding events with negligible risk. The simplest approach is to discard all contingency with probability below some probability threshold. The uncertainty of the risk assessments is related to the sum of risk of all events that has been discarded.

5 AI AND MACHINE LEARNING IN OPERATION OF POWER SYSTEM

5.1 *What data is needed*

There are numerous data available as input for AI and machine learning such as e.g.: current and historical: load, production, spinning reserve, sensor data (current, voltage, frequency) load flow, configuration of the power system, component data (type, characteristic, age), component health indexes, all previous outages and causes, weather type at the point of outage, dynamic data of the strength of the system form PMUs, protection schemes and other functionalities.

In addition, also near real time probabilistic risk calculation are available as input (Tollefsen, 2015). This represents big calculated data sets that gives a new insight in the inherit property of the power system.

Essential data for operation of the system: Voltage, frequency, production, spinning reserve and regulation possibilities.

5.2 *AI agents assigned to perform tasks in the power system*

Artificial Intelligence in general and supervised deep learning in particular tends to work well with large amounts of data (Schmidhuber, 2015). In supervised scenarios, the deep learning algorithms learn from known correct examples, and pick up trends and patterns that depict specific scenarios. In these cases, there is a trade-off between data quality and data size. The lower the quality of the data, the more data is needed extract the correct patterns. The most common example where this works is social media such as Facebook which contains enormous data of varying quality, enabling complex artificial intelligence algorithms.

The same basic concept is true for power systems, and it is therefore crucial that large amounts of data from power systems, including smart meters, is collected. The Norwegian power system manager Statnett has a particularly important role here. A concrete example of an application area artificial intelligence is expected to play a pivotal role is predicting electrical consumption peaks to avoid power outages (Goodwin, 2016). Over consumption may have serious consequences such as power outage. By predicting future peaks in the consumption, techniques such as load balancing could be carried out to avoid the problems. This clearly has to be carried out before the consumption peak happens, but knowing the consumption before occurrence is difficult. For this particular case, positive and negative examples should be collected, which in this case is examples of normal power flow, and over consumption. The artificial intelligence networks are trained with the data, and learns to understand which consumption trends lead to peak in the data. After the training phase, the network is put into practice and predicts future peaks which could either be used directly in an automated system to initiate load balancing, or as input to a decision support system. Other examples where artificial intelligence could play

a similar role are operation of power system and production prediction.

The operation of a power system is based on a set of rules and constraints that the power system operator must operate the system within. The constraints are at set of limits that is related to the physical properties for the different power system components. This can e.g. be thermal limits for power lines/transformers or other components, the normal frequency deviation should be within ± 0,1 Hz, voltage variations within ± 10% and the angle difference in three-phase current should be within maximum limits when reconnecting different part of the power system.

The rules and constraint connected to operation of a power system is notably different than the rules of board games such as chess. A power system is a stochastic system influenced by physical laws and human behavior such as consumption. The rules in a chess board and the behavior of the game are deterministic which means that future states are easily predictable, albeit many. However, there are similarities as well. The number waste amounts of states and complex behavior is identifiable in both complex board games and power systems. We can imagine that a machine learning also can be applied for operation of power systems based on the principle applied by Deep Mind with the new chees Alpha Zero program (Silver, 2017) and by the improved operation strategy obtained for Googles data center by use of Machine Learning Applications for Data Center Optimization (Gao, 2016).

5.3 *How can we evaluate the AI actions and gain trust?*

The artificial intelligence techniques vary from being statistically based on probabilistic induction, to knowledge based, and neuron based deep learning. For deep learning, which is undoubtedly, the most promising artificial intelligence technique in use, a confidence level is available as part of the supervised classification output. This confidence is very different from a probability, but can in any case be used as part of a trust schemes. If a deep learning network were to predict future problems, whenever it outputs an expected problem it can at the same time output how confident it is that is an actual problem. If this is part of a decision support system, the confidence can be used to inform a human decision maker in a decision support control room.

6 HUMANS, DECISION SUPPORT SYSTEM AND AI

6.1 *Decision support for system operation*

The ability to deal with the real-time fluctuations of the power system is not only a question of creating new technology and algorithms. Humans are still in the loop and the energy system is thus not only a technical system. It is a socio-technical system where the sense making, decisions and interventions of control room operators play an important part in the reliability of the system as a whole. This means that decision-support technology can play a key role in upholding the security of supply, but also that we need to take into account the human part of decision-making in control rooms. New decision-support systems will meet existing competence and experience, both at the individual and team level. In order to make sure that decision-support systems have the intended effects, the human perspective must be included in the development of the systems to ensure a good match between humans, technology and the organization of decision-making.

Advances in modelling and machine learning allow for information processing and problem solving that surpasses the capacity of an individual human decision-maker. Nevertheless, there is a need to find a balance between man and machine in the distribution of decision-making functions. Also, any period of technological transition will face challenges related to the competence of the existing workforce in the use of new technology, as well as a warranted level of trust into what new technology can and cannot do.

The importance of taking into account the relationship between human decision-making and algorithms can be illustrated by an example from New Scientist Volume 236, describing the domination of robots in the financial markets resulting in the human trader era is fading: “There are still a lot of unanswered questions surrounding the last bond market ‘flash crash’. On 15 October 2014”, “the US Treasury market crashed for about 10 minutes. Experts hypothesized that “activities of electronic trading algorithms” bore part of the blame, but reserved judgement for when they had more information. Three years later, no one is any wiser” (Adee, 2017).

This can also be the case for the future power system where smart grid technology, IoT and machine learning is put into operation. In order to avoid similar algorithm-induced “flash crashes” in the power system, it is vital to ensure the understanding of the system dynamics, the inherent risk properties and applying deep learning approaches as decision support.

7 CONCLUSIONS

The power system is rapidly changing towards the digital power system by using advanced ICT solutions, big data, smart grid, AI, machine learning and other advanced instruments.

The digitalization of the power business in Norway, Europe and other parts of the world, arise numerous new possibilities but also challenges.

To gain trust in the machine learning technology being introduced to power system, and to avoid similar problems in the power system as the financial markets experienced with the 'flash crash' from 2014, new insight is needed. The need for understanding and tracking in near real time the power systems inherent system property in regards of power system dynamics and risk level, becomes evident.

REFERENCES

Adee S. 2017. The money machine. New Scientist, Volume 236, Issue 3147, 14 October 2017, Pages 22–23.

Aranya Chakrabortty, Alex Huang: 'Digital Grid: Transforming the Electric Power Grid into an Innovation Engine for the United States', North Carolina State University, Computing Catalyst Consortium 2017.

'Digitalization & Energy', IEA, Sept 17, 2017.

Gao J. 2016. Machine Learning Applications for Data Center Optimization. Google 2016.

GARPUR 2017. Collaborative R&D project co-funded by the European Commission (7th Framework Programme).

Gjengedal T.: *Power System Operation and Control, UiT-The Arctic University, 2017.

Goodwin M., A. Yazidi 2016. Journal: Integrated Computer-Aided Engineering, vol. 23, no. 2, pp. 101–113, 2016.

L'Heureux A., K Grolinger, H F. Elyamany, M A. M. Capretz 'Machine Learning With Big Data: Challenges and Approaches'. Machine Learning With Big Data: Challenges and Approaches' IEEE Access, April 2017.

NVE: 'New technologies and demand response'. NVE 2016.

Ould Amrouche S., D.Rekioua, T. Rekioua, S.Bacha: 'Overview of energy storage in renewable energy system'. International Journal of Hydrogen Energy, Volume 41, Issue 45, 7 December 2016, pages 20914–20927.

Rajendra Kumar Pandey, Mohit Misra 'Cyber Security threats-Smart grid infrastructure', Power Systems Conference (NPSC), 2016 National, India, IEEE Xplore: 20 February 2017.

Regulations on system responsibility in the power system (FOS) 2002, Ministry of Petroleum and Energy, Norway.

Reiten E., L. Sørgard, K. Bjella 2014. A better organized power grid, Ministry of Petroleum and Energy, Norway.

Schmidhuber J. 2014. Deep Learning in Neural Networks: An Overview. Journal reference: Neural Networks, Vol 61, pp 85–117, Jan 2015.

Silver D., T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, D. Hassabis 2017. Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm. Cite as: arXiv:1712.01815 [cs. AI].

"Smart Grids European Technology Platform" www.smartgrids.eu.smartgrids.eu. 2011. Retrieved 2011-10-11.

Statnett Grid Development Plan 2017 (in Norwegian). www.Statnett.no.

Svendsen A.B., J. Eman, T. Tollefsen, Y Aabø, T. Digernes, S. Løvlund, J. O. Gjerde 2012. Online reliability assessment of power system, PMAPS 2012.

Tollefsen T., A.B. Svendsen, R.F. Pedersen, P. Skeie, T.M. Lunde, J. Mælan 2015. Online Reliability Calculations of Power Systems with Forecasted and Real Time Weather Influence, Esrel 2015.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Network analysis of the European natural gas infrastructure to quantify its performance in long-duration pipeline shutdown scenarios

P. Lustenberger & W. Kim
Future Resilient Systems (FRS), Swiss Federal Institute of Technology (ETH) Zurich, Zurich, Switzerland
Singapore-ETH Centre (SEC), Singapore

F. Schumacher, M. Spada, P. Burgherr & S. Hirschberg
Laboratory for Energy Systems Analysis, Paul Scherrer Institute (PSI), Villigen PSI, Switzerland

B. Stojadinović
Institute of Structural Engineering, Swiss Federal Institute of Technology (ETH) Zurich, Zurich, Switzerland

ABSTRACT: Nowadays, a fundamental requirement for a prosperous society is a reliable energy supply. The complex network theory provides an excellent basis to explore the functionality of such systems in response to severe component failures. In this case study, the European natural gas system is analyzed. The actual natural gas consumption is geospatially allocated to the infrastructure network. The network is abstracted and the flow capacity of the network is computed. A scenario analysis is conducted in order to identify the impact of storage facilities on the actual maximum possible flow. Furthermore, the natural gas supply shortage caused by each pipeline in case of a potential pipeline shut-down or failure is estimated. Finally, potential strategic locations of storage facilities for a more reliable natural gas network are identified.

1 INTRODUCTION

The consistent and continuous flow of goods and energy is an essential requirement for a reliable and prosperous economy in today's world. Looking at Europe, for instance, a notable share of more than 20% of the primary energy consumption is covered by natural gas (Eurostat, 2017). However, this causes also dependency on reliable supply, which cannot always be guaranteed to the full extent. In the last decade, several unforeseen natural gas disruptions with consequences on the economy occurred (Austvik, 2016). For example, the terrorist attack on the Amenas natural gas plant caused a reduction of more than 10% of the natural gas production of Algeria (Chrisafis et al., 2015). A very recent example is the explosion happened in Austria, where the accident disrupted natural gas flows toward Croatia, Italy and Slovenia. (Tirone and Wabl, 2017). Not only actual disruptions can cause damage, but also the awareness of a potential supply shortage can trigger uncertainty about business continuity. This happened for example in March 2013 in the United Kingdom, when natural gas supply equivalent to only six hours of supply was available in the storage facilities (Plimmer and Chazan, 2013).

Natural gas is transported and distributed by a well-developed system. The design of such a system requires long-term planning and large infrastructure investments. This kind of investments locks the capital in long-term contracts involving often policy decisions and agreements on national or regional levels (e.g. Carvalho et al., 2014, Mišík and Nosko, 2017). The European natural gas system became over time a large infrastructure network with many components, such as compressor stations, storage facilities, gas processing plants, Liquefied Natural Gas (LNG) terminals, LNG liquefaction and regasification facilities, aiming at assuring high reliability of the supply system. These components should be well planned and coordinated to guarantee a continous and adequate natural gas flow.

The natural gas demand cannot be covered by the European countries' available natural gas resources (BP, 2016). Therefore, it is necessary that natural gas is transported by pipelines from the East and South to Europe. Complementary, natural gas is also imported via LNG terminals.

The natural gas infrastructure network has to be able to compensate for potential pipeline shutdowns or failures, among others. This can be achieved through the construction of strategic

storage facilities and by increasing the network connectivity by means of new pipelines or diversification of entry points (e.g. LNG terminals).

Because of its importance, the European natural gas network is under constant analysis in order to improve the network's security of supply capacity. Besides natural hazard analysis (Poljansek et al., 2012, Lustenberger et al., 2017), a transmission network model, the so-called ProGasNet was developed (Praks et al., 2015) and complex network methodologies employed for analysis of the network properties (Carvalho et al., 2009). All the applied methodologies are based on graph theory. Moreover, Monte-Carlo-based reliability and vulnerability assessments of the network are applied (Praks et al., 2017), different flow algorithms and ways to allocate the natural gas in case of a disruption are tested (Carvalho et al., 2014). All the developed methodologies enable the identification of bottle-necks in case of a network component failure. A methodology to systematically analyse the impact of natural gas storage facilities on the network, however, is not developed yet. Furthermore, the impact of natural gas storage facilities on the entire European natural gas network has not been previously investigated.

The aim of this study is to develop a methodology to analyse the European natural gas infrastructure network's capacity to cover the demand not only during full operational conditions, but also in case of a long-duration pipeline shutdown or failure. Because storage facilities can be considered as consumption or supply infrastructure components in the network, the impact of feeding natural gas from natural gas storage facilities into the network in case of a potential long-term pipeline shutdown or failure is examined. The analytical framework employed in the current study comprises four steps to achieve this goal:

1. extending the natural gas infrastructure network model with geo-referenced storage facilities,
2. allocating natural gas demand according to geospatial consumption distribution,
3. analysing the extended natural gas infrastructure network by applying complex network theory, and
4. conducting scenario analysis with and without storage facilities to analyse the potential impact of pipeline failures on the overall network flow.

This study builds on Lustenberger et al. (2017). The focus is on the European Network of Transmission System Operators for Gas (ENTSOG) because sufficiently detailed data was available for the analysis. The network is abstracted, and the impact in case of a pipeline failure causing supply losses is estimated for each pipeline segment. Based on this, the specific impact for each pipeline segment is quantified. Moreover, the effect of storage facilities, to compensate such a potential failure, is analysed.

2 METHOD

In this section, the considered network system, including its components, is described, and the methods applied are explained.

2.1 *Geospatial natural gas infrastructure network model and its components*

The European natural gas infrastructure network consists of several thousand components including pipeline segments, entry and exit points, storage facilities and LNG terminals. In the present study, these infrastructure components are taken into account.

The natural gas infrastructure network data on geospatial basis can be abstracted as graph (e.g. Carvalho et al., 2014, Praks et al., 2015), while important features as the network topology are retained. This enables to employ a maximum flow algorithm well known in graph theory (Heineman et al., 2016). The graph can be defined as $G = (V, E, c)$, whereas G is an undirected graph with the vertices V and the edges E and c denotes the capacity of each edge. The vertices V can be sinks (e.g. consumer vertices) or sources (e.g. natural gas drilling platforms, LNG import terminals, and import pipelines).

A special case are the storage facilities and the natural gas power plants, which are discussed in section 2.2.

$E(u,v)$ represents the pipeline, connecting V_u and V_v. Whereas u and v are unique vertex identities. The flow capacity of $E(u,v)$ is defined as $c(u,v)$, while the natural gas flow through $E(u,v)$ is $f(u,v)$. If V_u and V_v are connected by multiple edges, $E(u,v)$ represents the sum of these edges. In this case, $c(u,v)$ is the sum of the involved pipeline capacities. The graph is then built according to Equation 1 and Equation 2.

$$E(u,v) = E(v,u) \tag{1}$$

$$c(u,v) = c(v,u) \tag{2}$$

With this abstraction, the natural gas infrastructure network can be transformed into an undirected (bidirectional edges) graph.

2.2 *Allocation of natural gas demand*

The actual natural gas consumption is allocated to the identified sink vertices. This is done considering

the geospatial distribution of the population density. It is assumed that the population density represents the geospatial distribution of the natural gas consumption of all the demand sectors, such as households or industries. The exception is the actual consumption of the natural gas power plants, which are identified as additional large consumers. The latter distinction is done because a linear allocation of the natural gas consumption at the country level to the population density on raster grid cell resolution does not necessarily reflect the spatial distribution of the natural gas power plants. Moreover, natural gas power plants constitute large consumers, which need to be taken into account separately for the spatial natural gas consumption allocation. The consumption allocation is conducted according to the following steps.

First, the population density for each country is estimated based on the corresponding population maps (CIESIN, 2016) and the country boundaries (zones) (GADM, 2016).

Second, the natural gas consumption of the natural gas power plants is estimated for each country. The consumption for each power plant is derived according to its CO_2 emissions provided by Enipedia (2016) and summed up at the country level.

Third, the natural gas consumption at the country level for the other demand sectors (e.g. industries and households) is estimated. This consumption (OD-Consumption) can be represented by subtracting the natural gas consumed by natural gas power plants from the total annual natural gas consumption at the country level (IEA, 2017). The resulting OD-Consumption is normalized per capita at the country level. The per capita consumption can then be allocated to the population density for each grid cell (~1 km) resulting in the OD-Consumption map of natural gas.

Fourth, the European area is split into demand areas applying a Voronoi algorithm (also called Thiessen algorithm) considering the identified sink vertices. The Voronoi algorithm cuts the connection line to the nearest neighbour of a vertex into half and connects the cutting point to an area around the corresponding vertex (Aurenhammer, 1991). The resulting Voronoi diagram is a partitioning of Europe into demand areas, in such a way that all points within a given Voronoi cell are closer to the corresponding vertex than to any other vertex. With this procedure, consumption can be allocated to the network, while the topology of the network remains the same.

Fifth, demand areas are spatially joined with the the OD-Consumption map. For this, all the raster grid cells are summed up within each demand area.

Finally, the natural gas consumption of the power plants is added to the corresponding demand areas.

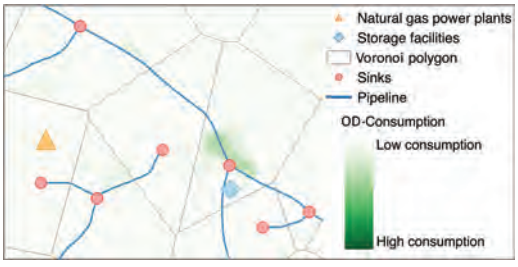


Figure 1. Map of the network infrastructure, demand areas (Voronoi polygons) and other demand sectors' consumption (OD-Consumption) on raster grid for illustration purpose.

This procedure provides an estimate of the natural gas consumed in each demand area on an annual basis. The data used and the spatial distribution are illustrated in Figure 1.

The power plants are connected to the network by assigning their consumption to the corresponding demand area. Storage facilities can serve as sinks or sources depending on the actual operation of the facility.

The storage facilities are assigned to the corresponding demand area. Depending on the scenario definition (see section 2.4), the corresponding demand area vertex is considered as source and/or sink.

The result of this procedure is a graph defined as $G(V,E,c)$ with $s,t \in V$, where s represents the source vertices and t denotes the sink or targeted vertices. Moreover, storage facilities can be switched on or off by turning the corresponding vertex from a sink to a source.

2.3 *Complex network theory—maximum flow estimation*

The abstract network poses a multi-source/sink problem because there are more than one identified source and sink. The multi-source/sink problem is commonly solved introducing a virtual super source and a virtual super sink, connecting all the corresponding sources and sinks as illustrated in Figure 2 (e.g. Ford and Fulkerson, 1956).

Moreover, flow capacity can be assigned representing the sink or source. This can be done by allocating flow capacities to the resulting virtual edges connecting the virtual super sink or the virtual super source to the corresponding vertex. In this study, the virtual edges connecting the source vertices to the virtual super source are assumed to be infinite, but the virtual edges connecting the sink vertices to the virtual super sink are limited by the actual consumption estimated at each corresponding vertex.

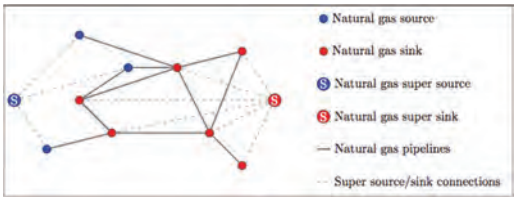


Figure 2. Illustration of the graph transformation from a multi-sink/source problem to a single sink/source problem.

Table 1. Ford-Fulkerson algorithm as applied in this study.

Algorithm: Ford-Fulkerson algorithm (Ford and Fulkerson, 1956)

Data: $G(V,E,c)$ $s,t \in V$
Result: F_{max}
while $p(s,t) \mid f(s,t) \leq c(s,t)$ exists **do**
1. Find path with unexploited flow
   $p_{unExp} = p(s,t) \mid f(s,t)$;
2. Find unexploited capacity
   $c_{unExp} = min(c(u,v)-f(u,v))$ for $u \in p_{unExp}$ and $v \in p_{unExp}$
3. Add unexploited capacity to flow on p_{unExp}
   $f(u,v) = f(u,v)+c_{unExp}$ for $u \in p_{unExp}$ *and* $v \in p_{unExp}$

end

$F_{max} = \sum_{u \in V} f(u,s^*) = \sum_{u \in V} f(t^*,u)$

The maximum possible flow between the virtual super sink and source is computed as described in Table 1 is derived from Heineman et al. (2016).

2.4 *Scenario analysis*

To estimate the impact in case of a pipeline failure, each pipeline is once deleted resulting in $\bar{G} = (V,\bar{E})$. Whereas $\bar{E} = E \backslash E_k$ with k representing the number of removed edges, which is in this case $E_k = E_i$. With this the impact in case of a single pipeline loss on the maximum possible flow is estimated. Hence, cascading failures in the network are not investigated. The dependencies of the network components in the sense of pipeline capacity to compensate a potential flow loss, however, are considered. As the maximum possible flow is not only limited by the pipeline capacity, but also by the actual consumption, the resulting difference can be considered as the loss of service of the European natural gas infrastructure network. This impact is estimated according to Equation 3.

$$\Delta F_{\max,i} = \frac{F_{\max} - F_{\max,i}}{F_{\max}} \cdot 100\% \quad (3)$$

where $\Delta F_{max,i}$ is representing the impact of pipeline i in percentage in case of a failure of pipeline i. F_{max} stands for the initial maximum flow and $F_{max,i}$ for the computed maximum flow without the pipeline i.

To estimate the impact of a storage facility considering a potential pipeline failure, a scenario analysis is conducted. For this purpose, the above-mentioned simulation is carried out twice, once without and once with storage facilities as natural gas sources. The storage facilities are added to the vertex in the corresponding demand area. The identified vertices in the corresponding demand areas are then connected to the virtual super source, representing not only a sink but also a source vertex.

The defined scenarios are with and without storage facilities. In this way, the impact of a storage facility on the network, and the service loss in case of a potential pipeline failure can be calculated.

3 DATA

In this section the datasets employed and the data acquisition is described. The used data is listed in Table 2. All data were available in a format directly usable in this study, except for the European natural gas infrastructure network data. The acquisition of this data and its processing is described in section 3.1.

3.1 *European natural gas infrastructure network*

The natural gas infrastructure data was extracted from three map layers, each consisting of 1024 high-resolution raster map tiles (images) from ENTSOG (2016). The three map layers represent three different ranges of pipeline diameters to which approximate flow capacities can be assigned.

First, the high-resolution images were georeferenced. Second, the raster cells were reclassified and the raster-grids vectorized. This resulted in a pipeline network in a digital map format (Lustenberger et al., 2017). To each edge a flow capacity c can be assigned according to the given diameter of the corresponding pipeline using the information from Carvalho et al. (2009).

The identification of the sources was done according to the provided map of ENTSOG (2016). Drilling platforms and LNG terminals were identified as source vertices. In addition, the entry points of the pipelines running from the East and from the South to Europe were identified as source vertices. This was concluded from the natural gas import statistics according to BP (2016).

In addition, natural gas storage information is processed from (ENTSOG, 2016). The storage facilities' information, including facility name and ENTSOG-ID, are downloaded via the ENTSOG Transparency Platform's API (Application Programming Interface). The point coordinates

Table 2. Datasets used.

Dataset	Format	Source	Description
Population density	Raster	(CIESIN, 2016)	Gridded population of the world on raster grid with cell size of ~1 km.
Country boundaries	Polygon	(GADM, 2016)	Worldwide country boundaries (Administrative level 0).
National natural gas consumption	Table	(IEA, 2017)	Annual natural gas production, imports and exports on country level in ktoe (2015).
Natural gas pipelines	Polyline	(ENTSOG, 2016)	European natural gas pipelines in three different diameter classes (9611 pipelines, 284641 km).
Natural gas storage facilities	Point	(ENTSOG, 2016)	European natural gas storage facilities with location, facility name (152 entries).
Natural gas sources	Point	(ENTSOG, 2016)	European natural gas sources (LNG ports, drilling platforms and system boundary crossing points - 152 entries).
Natural gas power plants	Point	(Enipedia, 2016)	Natural gas power plant coordinates and associated annual CO_2 emissions (618 entries, 565 used, 53 incomplete (not used)).

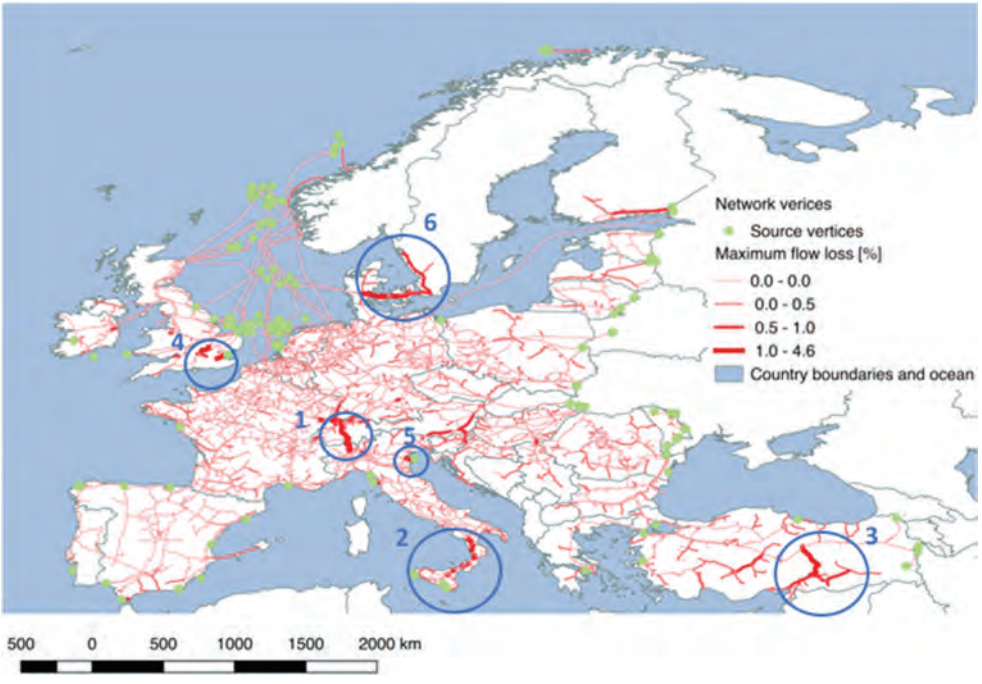


Figure 3. Impact of a potential pipeline failure or shutdown on the maximum flow for the scenario without storage facilities.

were then georeferenced according to the map provided from ENTSOG (2016).

4 RESULTS AND DISCUSSION

Figure 3 displays the resulting maximum flow loss ($\Delta F_{max,i}$ [%]) without storage facilities. The maximum flow loss can also be interpreted as loss of service. For six pipelines (labeled 1 to 6) the estimated loss is higher than 1% (~4.5 10^9 m^3/year). Similarly, Figure 4 shows the resulting maximum flow loss considering the storage facilities in addition as source vertices. The cases shown in Figures 3 and 4 are discussed in Table 3.

As illustrated in Table 3, the introduction of storage facilities is an effective measure to reduce the system service loss in case of a long-duration pipeline failure or shutdown. This is because storage facilities can compensate a potential failure and provide natural gas during the time of reduced flow capacity of the network.

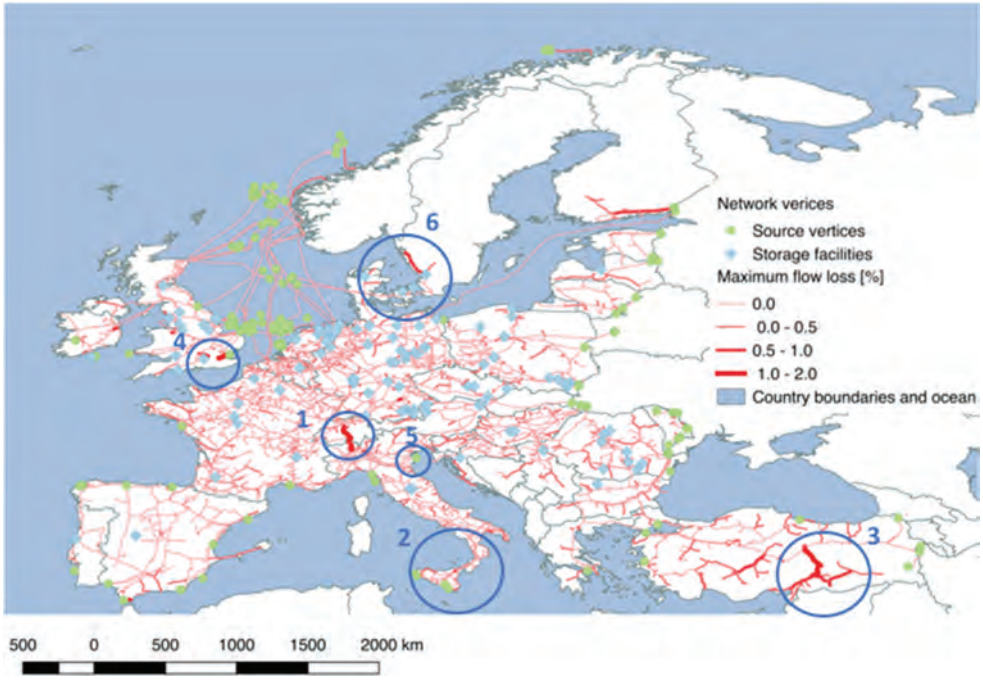


Figure 4. Impact of a potential pipeline failure or shutdown on the maximum flow for the scenario with storage facilities.

Table 3. Description of the pipelines with $F_{\max,i}$ higher than 1% resulting from the scenario without natural gas storages. The index is referring to Figure 3 and Figure 4.

	Scenario without storage facilities (Figure 3)		Scenario with storage facilities (Figure 4)	
Index	$\Delta F_{\max,i}$ [%]	Description	$\Delta F_{\max,i}$ [%]	Description
1	*4.6*	The Transit Gas pipeline and its southern extensions have the highest impact on the maximum flow in case of a failure. The pipelines connect northern Italy via Gries Pass through Switzerland with the German and French natural gas networks (Transitgas, 2017). The high maximum flow loss results from the fact that this pipeline is the only one crossing the central Alps providing a North/South connection.	*2.0*	With the introduction of the storage facilities the loss in case of a failure of the Transit Gas pipeline can be reduced from 4.6% to 2.0%. Several storage facilities are available in Germany and France to cover the remaining demand at the northern end of the pipeline, while two storage facilities are identified in central Italy (ENTSOG, 2016) to absorb demand reduction at the southern part of the pipeline.
2	3.5	The pipelines connecting Sicily and Calabria have the second highest impact on the maximum flow in case of a failure. The reason is that those pipelines are a part of the Trans-Mediterranean pipeline system (Transmed, 2017) and the Green-Stream project (GreenStream, 2017) connecting Algeria and Tunisia with Continental Europe. These pipelines provide a major share of the natural gas consumed in Europe (Carvalho et al., 2009).	0.9	With storage facilities, the loss can be substantially reduced. In case of a pipeline failure, Sicily is supplied by marine pipelines coming from northern Africa, while the supply reduction on the Italian mainland is partly absorbed by two natural gas storage facilities in central Italy.

(*Continued*)

Table 3. (*Continued*).

Index	Scenario without storage facilities (Figure 3)		Scenario with storage facilities (Figure 4)	
	$\Delta F_{\max,i}$ [%]	Description	$\Delta F_{\max,i}$ [%]	Description
3	1.8	The Turkish pipeline between the cities Silvas and Malatya connects the south-eastern part of Turkey with the Trans-Anatolian pipeline bringing natural gas from Georgia and Iran to Ankara (TANAP, 2017). The entire natural gas consumption of south east Turkey is provided with this pipeline, which amounts to ~1.8% of the consumption of the investigated system.	1.8	This loss remains the same because no storage facilities were identified in south eastern Turkey.
4	1.8/1.8	Two critical pipelines in the region of London were identified. The high loss of the south-eastern pipeline can be explained as it is the only high capacity pipeline delivering natural gas to London and the surrounding area. For the other pipeline in the north-western part the high loss is due to its high capacity as it delivers natural gas to the administrative regions of south west England.	1.8/0.0	The south eastern high capacity pipeline shows an unchanged loss as it is a dead-end pipeline. The loss of the other pipeline could be fully compensated due to compensation from storage facilities.
5	1.8	The pipeline near Venice connects the LNG port Cavarzere Porto Levante with the Italian mainland (AdriaticLNG, 2017). This indicates that this port plays a major strategic role for the continuous natural gas supply to continental Europe.	0.2	The loss of this pipeline can be reduced due to compensation of supply from storage facilities in the center of Italy.
6	1.1	This pipeline connects the European mainland with the Danish Islands Funen and Zealand. The importance of this pipeline is due to the fact that it is the only pipeline supplying Copenhagen and southern Sweden with natural gas.	0.0	Because two storage facilities are located close to Copenhagen, the loss caused by a potential failure or shutdown of this pipeline could be reduced to 0.

In this study, it is generally assumed that pipeline failures or shutdowns have a duration of one year, which is likely to be highly conservative for most cases. This does not affect the identification of critical pipelines as it is based on topological analysis and is thus independent of the duration of the loss of supply. On the other hand, the absolute loss of supply on the yearly basis as estimated here is expected to be much lower due to the expected typically much shorter duration of the interruptions.

The dependencies of the different components are assessed to the extent that in case of a pipeline shutdown, the natural gas flow is rerouted according to available pipeline capacities. Hence, in case there is no additional pipeline capacity available to compensate, there is a shortage of natural gas supply and a reduction in the maximum flow of the network.

Given the relatively conservative assumption of long duration of unavailability of critical pipelines after disruptions, the estimated losses are significant, but relatively limited. Moreover, the losses can to a certain extent be mitigated by the storage facilities.

5 CONCLUSIONS

In this study, the European natural gas infrastructure network is analysed considering the impacts of postulated failures of critical pipelines, and the potential mitigation of capacity losses by storage facilities.

Considering that the network capacity is not only limited by actual pipeline flow capacities, but also the consumption, leads to more realistic maximum flow analysis results. This is in contrast to the often-proposed maximum flow application with infinite capacities of the identified sources and sinks.

The maximum flow loss analysis showed that the failure of a single pipeline in a scenario without natural gas storage facilities can cause service losses >1% for six pipelines, with a maximum of 4.6% for the Transit Gas pipeline. Taking storage facilities into account can significantly reduce losses for all pipelines, except for south-east Turkey where no storage facilities are identified and a dead-end pipeline supplying the London area.

The Transit-Gas pipeline, crossing the Alps from Switzerland to Italy results in the highest maximum flow loss for both scenarios (with and without storage facilities).

For long-duration pipeline failures, the storage volume is required to be sufficiently large to act as a compensating source in the system throughout the whole disruption time. It is assumed that the storage capacity is not a limiting factor in reducing a pipeline supply loss. Additionally, the flows and consumption are aggregated on an annual level. This means that this study does not consider seasonal variabilities of natural gas consumption. The natural gas consumption in Europe is expected to be higher in winter than in summer (Hauser et al., 2017). Furthermore, this study assumes deterministically that the critical pipelines will be unavailabele during the whole year, which is very conservative. In summary, this study focuses on investigation of topological properties of the network and does not attempt to analyse the likely duration of disruptions.

In practice, not only the strategic placement of additional sources like storage facilities or LNG terminals, but also a temporal increase of the pipeline capacity, e.g. through increase of pressure in a pipeline, could provide sufficient natural gas to cover the demand (Su et al., 2017). This rerouting with temporal capacity increase of pipelines in case of a single pipeline failure or shutdown is partially considered. The pipeline capacities were quite roughly allocated and rather overestimated. It can be assumed that all pipelines are running at the corresponding allocated maximum capacity. This overestimation could possibly represent the credit for a temporal capacity increase of the pipelines in the system.

Although, the implementation of the findings of this study at large requires some caution due to current limitations, the described approach and results could provide the relevant stakeholders with useful indications for further enhancements of the reliability of natural gas infrastructure networks in general and of the European natural gas network in particular.

The results, and especially the approach developed in this study, could provide potentially helpful inputs for governmental agencies, the private industry and policy–makers.

It is recommended that the model could be further improved in terms of the time resolution. This would take into account the seasonal variation of the actual natural gas consumption and the actual storage capacities and storage use (source/sink). Furthermore, a network analysis methodology considering not only edge removal, but also storage capacities and the given temporal limitations could be developed.

Considering future work, the network model presented in this study will be combined with probabilistic treatments of potential losses of critical pipelines due to technical failures and selected natural events. This will provide an anlysis of the actual risks caused by potential pipeline failures considering not only consequences but also frequencies. Moreover, future work will have a risk and resilience prespective on the system level and not a reliability perspective on the component level.

ACKNOWLEDGEMENTS

The research was conducted at the Future Resilient Systems at the Singapore-ETH Centre, which was established collaboratively between ETH Zurich and Singapore's National Research Foundation (FI 370074011) under its Campus for Research Excellence and Technological Enterprise programme.

The natural gas network data are from the Transparency platform of ENTSOG and/or its members.

REFERENCES

AdriaticLNG (2017) Cavarzere porto levante. http://www.adriaticlng.it/en/home/.

Aurenhammer, F. (1991) Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR),* 23, 345–405.

Austvik, O.G. (2016) The Energy Union and security-of-gas supply. *Energy Policy,* 96, 372–382.

BP (2016) Statistical Review of World Energy. IN PBP (Ed.

Carvalho, R., Buzna, L., Bono, F., Gutiérrez, E., Just, W. & Arrowsmith, D. (2009) Robustness of Trans-European Gas Networks. *Physical Review E.*

Carvalho, R., Buzna, L., Bono, F., Masera, M., Arrowsmith, D.K. & Helbing, D. (2014) Resilience of Natural Gas Networks during Conflicts, Crises and Disruptions. *PLoS ONE,* 9, e90265.

Center for International Earth Science Information Network—Columbia University CIESIN (2016) Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals. Palisades, NY, NASA Socioeconomic Data and Applications Center (SEDAC).

Chrisafis, A., Borger, J., McCurry, J. & Macalister, T. (2015) Algeria hostage crisis: the full story of the kidnapping in the desert. *The Guardian.*

Enipedia (2016) http://enipedia.tudelft.nl/wiki/Main_Page.
ENTSOG (2016) ENTSOG Transparancy Platform. http://www.entsog.eu/.
Eurostat (2017) Gross inland consumption. http://ec.europa.eu/eurostat/statistics-explained/index.php/Consumption_of_energy.
Ford, L.R. & Fulkerson, D.R. (1956) Maximal flow through a network. *Canadian journal of Mathematics,* 8, 399–404.
GADM (2016) GADM database of Global Administrative Areas (v2.8). http://www.gadm.org/.
GreenStream (2017) The GreenStream Pipeline. http://www.greenstreambv.com/en/pages/home.shtml.
Hauser, P., Hobbie, H. & Möst, D. (2017) Resilience in the German natural gas network: Modelling approach for a high-resolution natural gas system. *2017 14th International Conference on the European Energy Market (EEM).*
Heineman, G.T., Pollice, G. & Selkow, S. (2016) *Algorithms in a nutshell: A practical guide*, O'Reilly Media, Inc.
IEA (2017) World energy balances. http://dx.doi.org/10.1787/data-00512-en.
Lustenberger, P., Sun, T., Gasser, P., Kim, W., Spada, M., Burgherr, P., Hirschberg, S. & Stojadinović, B. (2017) Potential impacts of selected natural hazards and technical failures on the natural gas tranmission network in Europe. *European Safety and Reliability Conference.* Portoroz, Slovenia, Taylor & Francis.
Mišík, M. & Nosko, A. (2017) The Eastring gas pipeline in the context of the Central and Eastern European gas supply challenge. *Nature Energy,* 2, 844–848.
Plimmer, G. & Chazan, G. (2013) UK gas supply six hours from running out in March. *Financial Times.*
Poljansek, K., Bono, F. & Gutiérrez, E. (2012) Seismic risk assessment of interdependent critical infrastructure systems: The case of European gas and electricity networks. *Earthquake Engineering & Structural Dynamics,* 41, 61–79.
Praks, P., Kopustinskas, V. & Masera, M. (2015) Probabilistic modelling of security of supply in gas networks and evaluation of new infrastructure. *Reliability Engineering & System Safety*.
Praks, P., Kopustinskas, V. & Masera, M. (2017) Monte-Carlo-based reliability and vulnerability assessment of a natural gas transmission system due to random network component failures. *Sustainable and Resilient Infrastructure,* 2, 97–107.
Su, H., Zhang, J., Zio, E., Yang, N., Li, X. & Zhang, Z. (2017) An integrated systemic method for supply reliability assessment of natural gas pipeline networks. *Applied Energy*.
TANAP (2017) Trans anatolian natural gas pipeline project. http://www.tanap.com/.
Tirone, J. & Wabl, M. (2017) Austrian Gas Pipeline Explosion Disrupts Key EU Supply Hub. *Bloomberg Markets.*
Transitgas (2017) The pipeline system. http://www.transitgas.org/EN/.
Transmed (2017) Gas transportation system. http://www.transmed-spa.it/?lingua=2.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

An efficient reliability analysis on complex non-repairable systems with common-cause failures

G. Feng
Department of Engineering Mathematics, University of Bristol, Bristol, UK

H. George-Williams
Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK
Institute of Nuclear Engineering and Science, National Tsing Hua University, Hsinchu, Taiwan

E. Patelli
Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK

F.P.A. Coolen
Department of Mathematical Sciences, Durham University, Durham, UK

M. Beer
Institute for Risk and Reliability, Leibniz University Hannover, Hannover, Germany
Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK
School of Civil Engineering and Shanghai Institute of Disaster Prevention and Relief, Tongji University, China

ABSTRACT: Common-Cause Failures (CCF) impose severe consequences on a complex system's reliability and overall performance. A more realistic assessment, therefore, of the survivability of the system requires an adequate consideration of these failures. The survival signature approach opens up a new and efficient way to compute system reliability, given its ability to segregate the structural and probabilistic attributes of the system. Traditional survival signature-based approaches assume the failure of one component to have no effect on the survival of the others. This assumption, however, is flawed for most realistic systems, given the existence of various forms of couplings between components. This paper, therefore, presents a novel and general survival signature-based simulation approach for non-repairable complex systems. We have used Monte Carlo Simulation to enhance the easy propagation of CCF across the complex system, instead of an analytical approach, which currently is impossible. In real application world, however, due to lack of knowledge or data about the behaviour of a certain component, its parameters can only be reported with a certain level of confidence, normally expressed as an interval. In order to deal with the imprecision, the double loop Monte Carlo simulation methodology which bases on the survival signature is used to analyse the complex system with CCF. The numerical examples are presented in the end to show the applicability of the approach.

1 INTRODUCTION

Common-Cause Failures (CCFs) are failure events that affect multiple components simultaneously. The origin of common cause events can be outside the system components they affect, or they can originate from the components themselves, causing the other components to fail. The proper consideration and modelling of CCFs is essential in complex systems reliability analysis, as they may have a significantly adverse effect on the system's overall functionality. They have been shown by many studies (Dhillon & Anude 1994) to decrease the reliability and availability of multi-component systems. They are, therefore, extremely important in reliability assessment and must be given adequate treatment, to minimise overestimation (Modarres 2006).

The CCF event can either impact the overall system operation or only affect specific components within the system (Wierman et al. 2007). Aldemir (1987) haa given an overview of parametric Common-Cause Failure models. To be specific, for component level, the CCF event is a component level failure. Rasmuson and Kelly reviewed the basic concepts of modelling CCFs in reliability and risk

studies (Rasmuson & Kelly 2008). One of the most commonly used single parameter models defined by Fleming (1975) is the β-factor model, which is the first parameter model applied to common cause failures in risk and reliability analysis. He then generalised the β-factor model to the multiple Greek letter model in 1986 (Fleming et al. 1986). The α-factor model which is proposed by Mosleh et al. (1988) develops CCFs from a set of failure ratios and the total component failure rate. Based on the α-factor model, Kelly & Atwood (2011) presented a method for developing Dirichlet prior distributions that have specified marginal means, but which are otherwise minimally informative. The binomial failure rate model (Atwood 1986) on the other hand, estimates the failure frequency of two or more components in a redundant system. This is computed as the product of the CCF shock arrival rate and the conditional failure probability of the components given the shock.

At for system level, the CCF event is a system functional level failure. A number of models have been developed recently. For instance, George-Williams & Patelli (2017) proposed an efficient load-flow simulation approach to assess the availability of reconfigurable multi-state systems with inter-dependencies. A robust Bayesian approach to the α-factor model for common cause failures has also been proposed by Troffaes et al. (2014). Coolen & Coolen-Maturi (2015b) presented a non-parametric predictive inference for system reliability following a common cause failure. However, there are mainly two problems within the above research works: (1) either recognise the components within the system as exchangeable single type; or (2) evaluate the system configuration for every reliability estimation trial, which is time consuming. Therefore, an extension of above works is needed. To be specific, it is necessary to perform reliability analysis on systems susceptible to CCFs, because these realistic complex systems always consist of components which belong to different types. Survival signature provides a good way to solve this problem.

Survival signature was first proposed by Coolen & Coolen-Maturi (2012) in 2012. It is a powerful methodology which can not only hold the merits of the former system signature (Samaniego 2007), but can be used in complex system with components belong to multiple types. In essence, it does not have the assumption that components of different types are exchangeable, which overcomes the long-standing limitation of the system signature. This is useful when a system consists of components that belong to different types, which means their failure times follow different probability distributions characters (Coolen & Coolen-Maturi 2015a). Therefore, survival signature is a promising method for application to complex systems. Based on the former work, Aslett et al. (2015) analysed system reliability within the Bayesian framework of statistics. Feng et al. (2016) dealt with the imprecision within the system by analytical and numerical ways respectively, what is more, new component importance measures were presented in this paper. An imprecise Bayesian non-parametric approach by using sets of priors to system reliability with multiple types of components was developed by Walter et al. (2017). Patelli et al. (2017) proposed efficient simulation approaches which based on survival signature for reliability analysis on large system. Reed (2017) put forward an efficient algorithm for exact computation of survival signature using binary decision diagrams.

This paper is organised as follows. Section 2 gives a brief conceptions about the survival signature and α-factor parameter. The survival signature-based simulation reliability approach is proposed in Section 3, in addition, this Section introduces imprecision within the components failure times. The applicability and performance of the proposed approaches is presented in Section 4. Finally Section 5 closes the paper with conclusions.

2 SURVIVAL SIGNATURE AND α-FACTOR PARAMETER

2.1 *Survival signature*

Suppose there is a complex system with m components which belong to $K \geq 2$ component types, with m_k components of type $k \in \{1,2,...,K\}$ and $\sum_{k=1}^{K} m_k = m$. Assume that the random failure times of components of the same type are exchangeable, while full independence is assumed for components belong to different types (*iid*), the survival signature which can be denoted by $\Phi(l_1,l_2,...,l_K)$, with $l_k = 0,1,...,m_k$ for $k = 1,2,...,K$. It defines the probability that the system functions given that l_k of its m_k components of type k work, for each $k \in \{1,2,...,K\}$. There are $\binom{m_k}{l_k}$ state vectors $\underline{x}^k$ with $\sum_{i=1}^{m_k} x_i^k = l_k$ (k = 1, 2, …, K), where $\underline{x}^k = (x_1^k, x_2^k, ..., x_{m_k}^k)$. Let $S_{l_1,l_2,...,l_K}$ denote the set of all state vectors for the whole system, and it can be known that all the state vectors $\underline{x}^k \in S_{l_k}^k$ are equally likely to occur. Therefore, the survival signature can be expressed as:

$$\Phi(l_1,...,l_K) = \left[\prod_{k=1}^{K} \binom{m_k}{l_k}^{-1}\right] \times \sum_{\underline{x} \in S_{l_1,...,l_K}} \phi(\underline{x}) \tag{1}$$

where $\phi = \phi(\underline{x}) : \{0,1\}^m \rightarrow \{0,1\}$ is the system structure function, i.e., the system status based on all

possible state vectors $\underline{x}$. ϕ is 1 if the system functions for state vector $\underline{x}$ and 0 if not.

Let $C_k(t) \in \{0,1,...,m_k\}$ denote the number of k components working at time t. Assume that the components of type k have a known cumulative distribution function (*CDF*) $F_k(t)$ and the components failure times of different type are assumed independent, then:

$$P(\bigcap_{k=1}^{K}\{C_k(t)=l_k\}) = \prod_{k=1}^{K} P\left(C_k(t)=l_k\right) = \prod_{k=1}^{K}\binom{m_k}{l_k}\left[F_k(t)\right]^{m_k-l_k}\left[1-F_k(t)\right]^{l_k} \quad (2)$$

Hence, the survival function of the system with K types of components becomes:

$$P(T_s > t) = \sum_{l_1=0}^{m_1}\cdots\sum_{l_K=0}^{m_K}\Phi(l_1,...,l_K)\,P\left(\bigcap_{k=1}^{K}\{C_k(t)=l_k\}\right) \quad (3)$$

Equation 3 shows that the structure of the system is separated from the its components failure times, which is the typical advantage of the survival signature. The survival signature is a summary of structure functions and only needs to be calculated once for the same system. As a result, it is an efficient method to perform system reliability analysis on complex systems with multiple component types.

2.2 α-*factor model*

The α-factor model is particularly useful in the practical engineering world as the alpha factor parameters can be got through experts' judgement of the system or past data on the system.

The parameter, α_r, of the model, is the fraction of the total component failure events causing the simultaneous failure of an additional $r-1$ components.

Let us assume there is a system with three exchangeable components, α_1 means the failure of one component cannot influence the status of the other components. α_3 denotes the failure of one component can lead to the other two components fail simultaneously, which means CCFs occur. For $\alpha_2 = p$, it expresses that there is a probability p of one additional component failing, following the failure of a component in this system. It can be drawn that $\sum_{r=1}^{3}\alpha_r = 1$.

Similarly, for the complex system with multiple component types, the alpha parameters α_r^k denotes that if one component of type k fails due to an common cause event, the probability that the other $r-1$ components fail simultaneously. If there are m_k components in this group, it can conclude that $\sum_{r=1}^{m_k}\alpha_r^k = 1$.

Based on the definition of α-factor parameter, the probability of a common cause basic event involving failure of k components in a system of m components can be calculated by Equation 4.

$$Q_k = \frac{k}{\binom{m-1}{k-1}}\frac{\alpha_k}{\alpha_t}Q_t \quad (4)$$

where, $k = 1,2,...,m$ and $\alpha_t = \sum_{k=1}^{m} k\alpha_k$. Q_t is the total probability of failure accounting both for common cause failures and independent failures. The alpha parameter estimator can be expressed as:

$$\alpha_k = \frac{n_k}{\sum_{i=1}^{m} n_i} \quad (5)$$

where, n_k is the number of events with k failed components.

The α parameter estimator represents the probability that exactly k of the m components fail, given that at least one failure has occurred. It can be seen from Equation 5 that the sum of all the α_k will be 1. The advantage of the α-factor model is its distinction between the total failure rate of a component Q_t, for which we generally have a lot of information, and common cause failures modelled by α_k, for which we generally have very little information (Troffaes et al. 2014).

3 EFFICIENT METHOD FOR ANALYSING SYSTEM RELIABILITY WITH COMMON CAUSE FAILURES

3.1 *The proposed approach*

In order to perform reliability assessment on any kind of systems without introducing simplifications or unjustified assumptions, this Section proposes a simulation method to analyse system reliability after common cause failures.

Suppose there is a complex system with m components which belong to X_c different component groups, and there are m_k components of type $k \in \{1,2,...,K\}$ and $\sum_{k=1}^{K} m_k = m$. The common cause group matrix can be expressed as M_{CCG}, the α factor parameters of each component group, $\alpha_{m_k}^k$, are stored in the matrix, recall $\sum_{k=1}^{K}\alpha_{m_k}^k = 1$.

The number of failing component of each type is depended on the number of component still functioning, therefore, it is necessary to use the α-factor model to provide probabilities for any combinations of number of components that would fail

when a common cause failure event occurs. Here has an assumption that if one component of type k fails, it can only influence the components within the same component group X_c under the CCF model. This is reasonable as the components of the same type tend to be influenced by the same common cause failure event, this is also the reason why they are grouped in the same type. Then looking at how many of the components of each type still function, and assume exchangeability within them with regard to the CCF failure model.

The reliability of the system after common cause failures can be estimated adopting the following simulation procedure:

Step 1. Initialise the counter V to store the output, define the mission time as t_m and number of samples as N;

Step 2. Define the component groups as X_c, and the common cause group matrix M_{CCG}, the α factor parameters $\alpha^k_{m_k}$ are stored in the matrix;

Step 3. Sample the failure time of each component as $t_i \le t_{i+1}$, where $i = 1, 2, \ldots, m$, and set $t_{old} = 0$, at this time the survival signature (production level) is equal to 1;

Step 4. Set the current time $t_{current} = min(t_i)$;

Step 5. At time t_i, finding out which component fails and which common cause group it belongs to. The components affected by a failure event due to CCF can be expresses as V_{comp};

Step 6. Upgrade the number of working components of each component group after the CCF, and then get the survival signature (production level) Φ_{t_i} after the corresponding failure time t_i;

Step 7. Set the failure time of the components (the failure component and its common cause failure components) as infinite;

Step 8. Repeat Steps 4 through 7 until $t_m > t_{old}$;

Step 9. Store the production level of the system over the time by $V(j) = V(j) + \Phi_{t_i}$;

Step 10. Repeat Steps 3 through 9 for N times.

Therefore, the survival function of the complex system after common cause failures is obtained by averaging the vector collecting the production level of the system over the number of samples: $P(T_s > t \mid CCF) = V_t / N$.

The algorithm of the proposed simulation method can be seen follows:

Algorithm 1 The Proposed Approach's Algorithm

Require: N, t_m, X_c, M_{CCG}, N_t: number of discretisation step
Set $V(1:N_t) = 0$ ▷ Initialise counters
Set Φ =Survival Signature ▷ Load survival signature
for $n = i : N$ **do** ▷ Loop over number of samples
  Sample t_i ▷ Sample failure time of every component in each type
  Set $[t_{current}, X_c] = min(t_i)$ ▷ Minimum failure time and component index
  while $[t_{current} \le t_m$ **do**
    $Vch = ceil([t_{old}, t_{current}]/timestep) + 1$ ▷ Define channels
    $V(Vch(1) : Vch(2) - 1) = V(Vch(1) : Vch(2) - 1) + \Phi(t_{old})$ ▷ Update counters
    if M_{CCG} is not empty **then** ▷ CCF within system
      $CCF_{CumProb} = cumsum(\alpha^k_{m_k})$ ▷ Cumsum CCF probability of the failed component group
      $r = find(CCF_{CumProb} \ge rand, 1)$ ▷ Find the number of components in the CCG to fail
      if $r > 1$ **then**
        Propagate CCF
      end if
    else ▷ No CCF within the system
      $V_c = X_c$ ▷ V_c contains affected components
    end if
    $\Phi_{t_{old}} = \Phi_{t_{current}}$ ▷ Upgrade survival signature
    $t_{old} = t_{current}$ ▷ Upgrade new old time
    $[t_{current}, X_c] = min(t_i)$ ▷ Minimum failure time and component index after CCF
  end while
  if $t_m > t_{old}$ **then**
    $V(Vch(1) : Vch(2)) = V(Vch(1) : Vch(2)) + \Phi(t_{old})$ ▷ Update counters
  end if
end for

3.2 *Imprecision in consideration*

In the engineering applications, if there exist imprecision within the components failure time distributions, or empirical distribution of components failure times are used, no analytical methods can be used without resorting to some degree of simplification or approximation (Beer et al. 2013) (Aven 2017). Instead, the proposed simulation methods can be applied to any systems irrespectively to the probability distribution for the component failure time used.

To be specific, the system reliability performance after common cause failures can be simulated using survival signature-based Monte Carlo method. This double loop simulation method (Du & Chen 2004) not only has the advantage of survival signature to handle complex system reliability problems, but can recur to Monte Carlo simulation to deal with the uncertainties within the system.

Double loop sampling involves two layers of sampling: the outer loop is called the parameter

loop since it concerns sampling different values for the set of distribution parameters for all of the uncertain quantities; while the inner loop goes by the name of probability loop because it involves sampling from precise probability distribution functions. As a matter of fact, double loop sampling implicates sampling from an analytical distribution whose parameters have been generated by sampling.

To solve the parameter epistemic imprecision within components, it is just need to add an optimization loop around the survival signature-based simulation method cited in Section 3.1 to estimate the bounds. In other words, it can be done by adding a simple Monte Carlo loop and sampling the values of components parameters from uniform distributions.

4 NUMERICAL EXAMPLE

Shown in Figure 1 is an arbitrary 13-component complex system, which components are arranged into five groups. The number within each box denotes which group the component belongs to while the number outside defines the index of the component in the system. The system is assumed to be non-repairable and components of the same group have the same failure time distribution, as defined in Table 2. In the table, an exponential distribution is defined by its mean (in hours) while a Weibull distribution is defined by a set which first element is its scale parameter (in hours).

The system is first analysed without CCF using the proposed simulation model with the data presented in Table 2 and compared to its analytical solution.

It is then re-analysed considering common cause failures with all common cause groups are active. For this system, the common cause group failure matrix M_{CCG}, with and without CCF, can be expressed in Equations 6 and 7 respectively. The results obtained are shown in Figure 2.

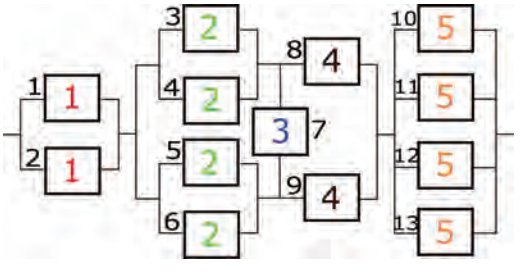


Figure 1. Complex system with thirteen components which belong to four types. The number inside the component box represents the type, while the number outside the box expresses the component index.

Table 1. Component failure data with precise distribution parameters.

Component type	Distribution type	Distribution parameters	CCF parameters
1	Weibull	(1.8,2.2)	{0.95, 0.05}
2	Exponential	1.2	{0.8, 0.1, 0.05, 0.05}
3	Weibull	(2.3,1.6)	{1}
4	Weibull	(3.2,2.6)	{0.9, 0.1}
5	Exponential	2.1	{0.75, 0.1, 0.1, 0.05}

Table 2. Component failure data with imprecise distribution parameters.

Component type	Distribution type	Distribution parameters	CCF parameters
1	Weibull	([1.68,1.86], [2.08,2.32])	{0.95, 0.05}
2	Exponential	[1.07,1.33]	{0.8, 0.1, 0.05, 0.05}
3	Weibull	([2.12,2.51], [1.38,1.72])	{1}
4	Weibull	([2.99,3.41], [2.51,2.79])	{0.9, 0.1}
5	Exponential	[2.01,2.28]	{0.75, 0.1, 0.1, 0.05}

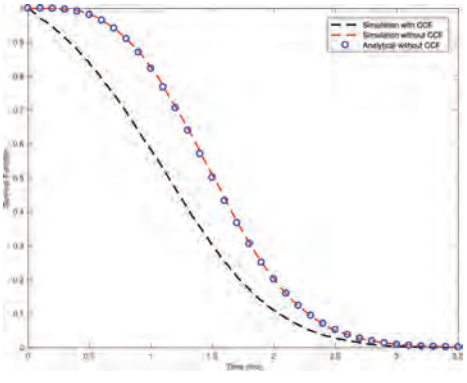


Figure 2. Survival function of the system in Figure 1 with CCF and without CCF through simulation method, along with the system reliability without CCF got by analytical solution.

$$M_{CCG} = \begin{pmatrix} 0.95 & 0.05 & 0 & 0 \\ 0.8 & 0.1 & 0.05 & 0.05 \\ 1 & 0 & 0 & 0 \\ 0.9 & 0.1 & 0 & 0 \\ 0.75 & 0.1 & 0.1 & 0.05 \end{pmatrix} \quad (6)$$

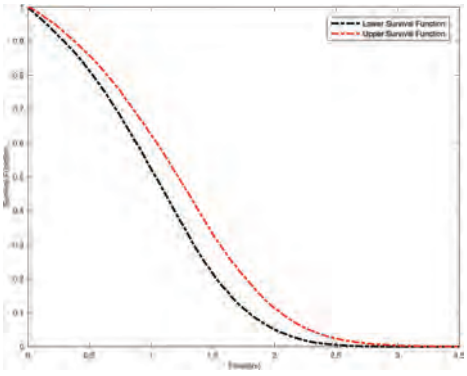


Figure 3. Lower and upper survival function bounds of the system in Figure 1 with CCF.

$$M_{CCG} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{7}$$

The accuracy and generality of the proposed simulation approach are validated by the plots in Figure 2, given the agreement between the simulation and analytical results. As shown, the reliability of the system reduces drastically when the effects of CCF are factored into the analysis. It exemplifies the need to consider this realistic aspect of a system's operation in its reliability evaluation.

To deduce the effects of imprecision in the failure distribution parameters of components on the system survival function, the system is analysed using the data presented in Table 2. Instead of a single curve, the survival function, in this case, could be any of an infinite number of curves lying within the bounds shown in Figure 3.

5 CONCLUSIONS

Common-Cause Failures (CCF) have an adverse effect on the reliability and performance of multi-component systems. They are normally a consequence of functional couplings between a group of components due to a variety of possible reasons. Thus, there is an inevitability about the susceptibility of realistic multi-component engineering systems to these failures. The need, therefore, to incorporate CCF considerations into system analysis is overwhelming, as the alternative may lead to overestimating the reliability of the system.

This paper puts forwards an efficient simulation method which bases on the survival signature to perform reliability analysis on complex systems with common cause failures. In fact, this approach extends the applicability of the survival signature approach to systems susceptible to common cause failures. More importantly, it holds the merits of both survival signature methodology and Monte Carlo simulation. Therefore, this approach is general and allows to know the survival function of the system after common cause failures at each time. What is more, the probabilistic uncertainty and imprecision in components parameters are taken into consideration by resorting this general simulation method. The effectiveness and feasibility of the proposed approach hasbeen demonstrated by the numerical example.

REFERENCES

Aldemir, T. (1987). Computer-assisted markov failure modeling of process control systems. *IEEE Transactions on reliability 36*(1), 133–144.

Aslett, L.J., F.P. Coolen, & S.P. Wilson (2015). Bayesian inference for reliability of systems and networks using the survival signature. *Risk Analysis 35*(3), 1640–1651.

Atwood, C.L. (1986). The binomial failure rate common cause model. *Technometrics 28*(2), 139–148.

Aven, T. (2017). Improving the foundation and practice of reliability engineering. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 1748006X17699478.

Beer, M., S. Ferson, & V. Kreinovich (2013). Imprecise probabilities in engineering analyses. *Mechanical systems and signal processing 37*(1), 4–29.

Coolen, F.P. & T. Coolen-Maturi (2012). Generalizing the signature to systems with multiple types of components. In *Complex Systems and Dependability*, pp. 115–130. Springer.

Coolen, F.P. & T. Coolen-Maturi (2015a). Modelling uncertain aspects of system dependability with survival signatures. In *Dependability Problems of Complex Information Systems*, pp. 19–34. Springer.

Coolen, F.P. & T. Coolen-Maturi (2015b). Predictive inference for system reliability after common-cause component failures. *Reliability Engineering & System Safety 135*, 27–33.

Dhillon, B. & O. Anude (1994). Common-cause failures in engineering systems: A review. *International Journal of Reliability, Quality and Safety Engineering 1*(01), 103–129.

Du, X. & W. Chen (2004). Sequential optimization and reliability assessment method for efficient probabilistic design. *Journal of mechanical design 126*(2), 225–233.

Feng, G., E. Patelli, M. Beer, & F.P. Coolen (2016). Imprecise system reliability and component importance based on survival signature. *Reliability Engineering & System Safety 150*, 116–125.

Fleming, K. (1975). Reliability model for common mode failures in redundant safety systems. In *Modeling and simulation. Volume 6, Part 1*.

Fleming, K.N., A. Mosleh, & R.K. Deremer (1986). A systematic procedure for the incorporation of common cause events into risk and reliability models. *Nuclear Engineering and Design 93*(2–3), 245–273.

George-Williams, H. & E. Patelli (2017). Efficient availability assessment of reconfigurable multi-state systems with interdependencies. *Reliability Engineering & System Safety 165*, 431–444.

Kelly, D. & C. Atwood (2011). Finding a minimally informative dirichlet prior distribution using least squares. *Reliability Engineering & System Safety 96*(3), 398–402.

Modarres, M. (2006). *Risk analysis in engineering: techniques, tools, and trends*. CRC press.

Mosleh, A., K. Fleming, G. Parry, H. Paula, D. Worledge, & D.M. Rasmuson (1988). Procedures for treating common cause failures in safety and reliability studies: Volume 1, procedural framework and examples: Final report. Technical report, Pickard, Lowe and Garrick, Inc., Newport Beach, CA (USA).

Patelli, E., G. Feng, F.P. Coolen, & T. Coolen-Maturi (2017). Simulation methods for system reliability using the survival signature. *Reliability Engineering & System Safety 167*, 327–337.

Rasmuson, D.M. & D.L. Kelly (2008). Common-cause failure analysis in event assessment. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 222*(4), 521–532.

Reed, S. (2017). An efficient algorithm for exact computation of system and survival signatures using binary decision diagrams. *Reliability Engineering & System Safety*.

Samaniego, F.J. (2007). *System signatures and their applications in engineering reliability*, Volume 110. Springer Science & Business Media.

Troffaes, M.C., G. Walter, & D. Kelly (2014). A robust Bayesian approach to modeling epistemic uncertainty in common cause failure models. *Reliability Engineering & System Safety 125*, 13–21.

Walter, G., L.J. Aslett, & F. Coolen (2017). Bayesian nonparametric system reliability using sets of priors. *International Journal of Approximate Reasoning 30*, 67–88.

Wierman, T.E., D.M. Rasmuson, & A. Mosleh (2007). *Common-cause failure database and analysis system: event data collection, classification, and coding*. Division of Risk Assessment and Special Projects, Office of Nuclear Regulatory Research, US Nuclear Regulatory Commission.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

LLVM-based stochastic error propagation analysis of manually developed software components

A. Morozov & K. Janschek
Institute of Automation, Faculty of Electrical and Computer Engineering, Technische Universität Dresden, Germany

Y. Zhou
Altran Deutschland S.A.S. & Co. KG, Powertrain Automotive, Munich, Germany

ABSTRACT: Modern industrial trends such a Cyber-Physical Systems and System of Systems lead to the continuously increasing complexity and heterogeneity of components and interfaces, as well as more and more advanced software parts. Classical reliability evaluation methods, recommended in nowadays standards, such as Fault Tree Analysis (FTA) and Failure Mode and Effect Analysis (FMEA), fail to describe system behavioral aspects in a sufficiently deep manner. Therefore, additional, sophisticated and highly specialized methods for the analysis of the effects of unavoidable faults are required. Recently introduced Dual-graph Error Propagation Model (DEPM) is a stochastic framework that captures system properties relevant to error propagation processes such as control and data flow structures and reliability characteristics of single components. The DEPM helps to estimate the impact of a fault of a particular component on the overall system reliability, e.g. to compute the mean number of erroneous values in a critical system output during given operation time. A DEPM can be automatically generated from various semi-formal system representations such as UML/SysML, AADL, or Simulink/Stateflow. However, despite the common trend towards model-based system development the functional software parts usually incorporate manually programmed code. The error propagation properties of this manual code also need to be analyzed and considered during the reliability evaluation of the complete system. This paper presents a new method, based on the Low-Level Virtual Machine (LLVM) compiler framework, that allows the automatic transformation of C-code or another LLVM supported front-end into a DEPM. The source code is compiled into the LLVM Intermediate Representation and instrumented in order to analyze control and data flow structures of LLVM instructions and control flow transition probabilities. The obtained information is transformed into the formal DEPM XML for further analysis. The paper describes the transformation method and its application to a low-level flight control software of a UAV system.

1 INTRODUCTION

1.1 *Motivation*

Model-Based System Engineering (MBSE) plays an important role in the development of modern safetycritical systems. Advanced MBSE toolchains support the system development starting from high-level design up to deployment and testing. MATLAB Simulink (MathWorks 2017a) and Stateflow (Math-Works 2017b) dominate in the field of control software development. Nowadays trends such a Cyber-Physical Systems and System of Systems lead to the continuously increasing complexity and heterogeneity of components and interfaces, as well as more and more advanced software parts. The increasing complexity brings new challenges for system analysis. Classical reliability and safety evaluation methods, recommended in nowadays industrial standards, such as Fault Tree Analysis (FTA) and Failure Mode and Effect Analysis (FMEA), fail to describe system behavioral aspects in a sufficiently deep manner. Therefore, additional, sophisticated and highly specialized methods for the analysis of the effects of unavoidable faults are required. Recently introduced Dual-graph Error Propagation Model (DEPM) (Morozov & Janschek 2014) is a stochastic framework that captures system properties relevant to error propagation processes. The DEPM helps to estimate the impact of a fault of a particular component on the overall system reliability, e.g. to compute the mean number of erroneous values in critical system outputs.

A DEPM can be automatically generated from various semi-formal system representations such as Simulink/Stateflow (Morozov et al. 2016), UML/SysML (Ding et al. 2016), and AADL (Morozov

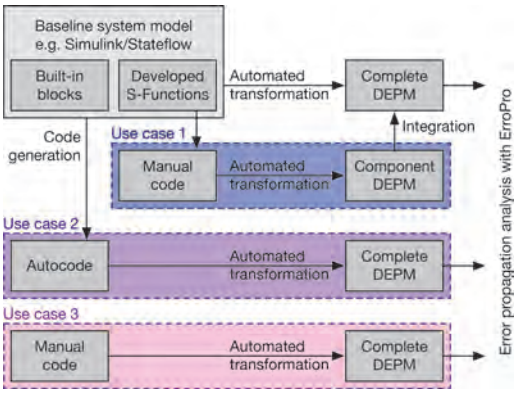


Figure 1. Three possible use cases of the proposed transformation method.

et al. 2018). Despite the common MBSE trend, the functional software parts usually incorporate manually programmed code. The error propagation properties of this code should be analyzed with a DEPM and considered during the reliability evaluation of the complete system. This paper presents a new method, based on the Low-Level Virtual Machine (LLVM) compiler framework, that allows the automatic transformation of C-code or another LLVM supported front-end into a DEPM. Figure 1 demonstrates three possible use cases:

- Use case 1: a Simulink model contains manually developed S-Function blocks, we generate individual DEPMs separately for these blocks using the proposed method and integrate them into a top-level DEPM generated from the Simulink model using the method presented in (Morozov, Janschek, Krüger, & Schiele 2016).
- Use case 2: we apply the proposed method to the automatically generated C-code of the entire Simulink model.
- Use case 3: we generated DEPMs for manually developed code in case of none-MBSE approach. An example of this use case is shown in Section 3.

The rest of the paper is organized as follows. The two following subsections give the overview of the background DEPM and LLVM technologies. Section 2 provides technical details of the proposed transformation method. Section 3 demonstrates the method with a case study, showing the transformation of lowlevel flight control software of a UAV into a DEPM.

1.2 *Dual-graph error propagation model*

The DEPM is a mathematical model that captures control and data flow aspects of a system, described using the following set-based mathematical notation:

$$DEPM := \langle E, D, A_{CF}, A_{DF}, C \rangle \tag{1}$$

E is a non-empty set of executable system elements;
D is a set of data storages;
A_{CF} is a set of directed control flow arcs, extended with control flow decision probabilities;
A_{DF} is a set of directed data flow arcs;
C is a set of conditions of the elements.

Fig. 2a shows a simple DEPM example. An *Element*, e.g. *A*, *B*, or *C*, represents an executable part of a system. Each element may receive input data and provide output data. A *Data*, e.g. *d1*, *d2*, or *d3*, represents a variable that can be read or written by an *Element*. For instance, *Element C* reads *d2* and *d3*, and writes *output*. A *ControlFlow* arc (black lines in Fig. 2a), weighted with an attribute probability, represents a control flow transition between *Elements*. For instance, after the execution of *A*, *B* will be executed with probability 0.7 and *C* with probability 0.3. A *DataFlow* arc (purple lines in Fig. 2a) describes data transfer between the elements. A *DataFlow* connects an *Element* with a *Data* or vice versa. The *DataFlow* arcs are considered to be the paths of the data error propagation.

The fault activation and the error propagation can be specified using probabilistic *Conditions* in the elements. During the execution of an element, faults can be activated and occurred errors propagate to its outputs. The elements, which can activate faults, are highlighted in red. For instance, in the element *A*, highlighted in red in Fig. 2a, faults can be activated with probability 0.1, defined in

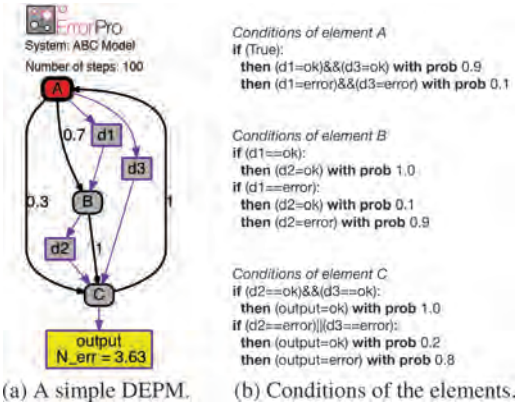


Figure 2. An example of a DEPM with specified conditions of the elements.

the conditions of *A* (see conditions *A* in Fig. 2b), and occurred errors propagate to its output data *d1* and *d3*.

The error propagation or error correction probabilities for each element are defined also using probabilistic *Conditions* (see Fig. 2b). The errors can propagate from the data inputs to the outputs. For instance, the conditions of the element *B* specify that the element *B* does not activate faults, but the errors can propagate from *d1* to *d2* with the probability 0.9.

ErrorProTM (Morozov et al. 2015) is our software tool for stochastic error propagation analysis that allows users to create and compute DEPMs. Discretetime Markov chain models that describe system execution and error propagation processes are automatically generated and computed using an interface with the PRISM model checker (Kwiatkowska, Norman, & Parker 2011). In the DEPM, the *mean number of errors* of all data, marked with yellow color, can be computed. The reliability metric, *mean number of errors*, is the average number of cumulative occurred errors in the data during the system execution. For instance, the *mean number of errors* in the data storage *output* during 100 steps (execution of one element is one step) is equal to 3.63, as shown in Fig. 2a. A DEPM model can be stored in an XML file. A fragment example of an XML file is shown in Listing 1.

1.3 *LLVM compiler infrastructure*

The LLVM Project (Lattner & Adve 2004) is a collection of modular and reusable compiler and toolchain technologies. The LLVM libraries provide a source and target-independent optimizer, along with code generation support for many popular CPUs. These libraries are built around an assembly-like low-level code representation

```
<?xml version="1.0" encoding="utf−8"?>
<model n_steps="100" name="ABC Model" version="3.0">
 <element initial ="true" name="A"/>
  ...
 <data name="d1"/>
  ...
 <control_flow from="A" prob="0.7" to="B"/>
  ...
 <data_flow from="A" to="d1"/>
  ...
 <conditions element_name="A">
  <if statement="True">
   <then prob="0.9" update="d1 = ok; d3 = ok;"/>
   <then prob="0.1" update="d1 = error ; d3 = error ;"/>
  </if>
 </conditions>
  ...
</model>
```

Listing 1. An example of a DEPM XML file.

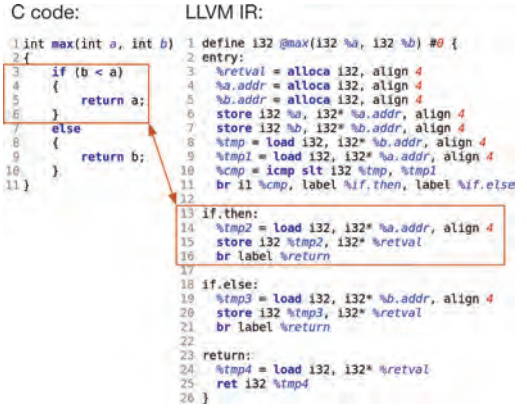


Figure 3. C-code example and its LLVM IR.

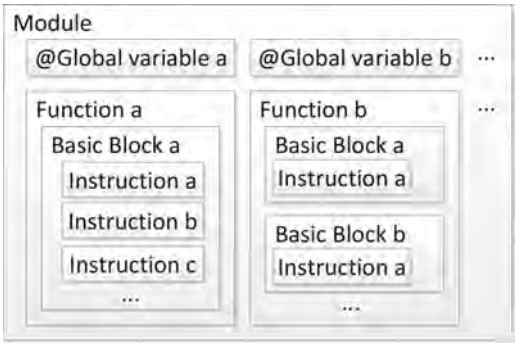


Figure 4. The hierarchical structure of LLVM IR.

known as the LLVM Intermediate Representation (LLVM IR). The LLVM IR is basically a representation in-between a high-level language and a low-level machine code. Figure 3 shows a small piece of C-code (left) and the corresponding LLVM IR code (right). The LLVM IR uses Static Single Assignment (SSA) format: Each defined variable cannot be re-defined. There are only explicit strongly typed variables in the IR code.

The LLVM IR code is structured into modules that contain functions as it is shown in Figure 4. The functions are decomposed into basic blocks. A basic block contains a sequence of single instructions. The conditional jumps between basic blocks form the control flow structure.

Figure 5 shows a typical LLVM-based workflow. An input source code, e.g. a C-code or any other of more than 20 LLVM supported languages, can be compiled with *clang* into the LLVM IR. Using the LLVM API, an optimizer can be built that produces an optimized version of the compiled LLVM

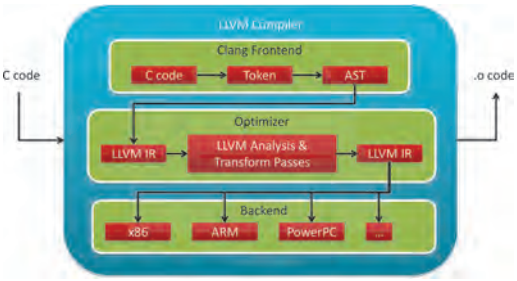


Figure 5. Common LLVM-based workflow.

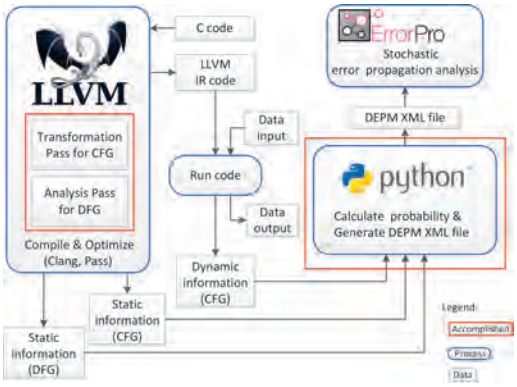


Figure 6. Top-level architecture of the C-to-DEPM transformation method.

IR, which, in turn, can be further compiled with *llc* for x86, ARM, PowerPC or other platforms. The optimizer can be built with so-called analysis or transformation passes. An analysis pass does not change the input IR, but provides analytical information, for instance, counts the number of instructions or builds a call graph. In contrast, a transformation pass changes and optimizes the original LLVM IR.

2 TRANSFORMATION METHOD

Figure 6 shows the architecture of the proposed transformation method. Rounded rectangles with blue borders represent activities and gray rectangles represent data. The red frames highlight our contribution: We have developed two LLVM passes and the python script that processes the outputs of the passes and generates DEPM XML file for further analysis with ErrorPro™. All the steps are automated and can be run with a single shell script that calls LLVM tools (compilers, linkers etc), as well as our passes, and the DEPM generation python script.

2.1 *Transformation pass for control flow analysis*

The first, transformation, pass helps to identify elements of the future DEPM, control flow structure, and control flow transition probabilities. The overview of the transformation pass is shown in Figure 7. The pass takes the LLVM IR of the original C-Code generated with *clang* as input and performs two tasks. (1) The first task is to generate the list of DEPM elements that represent single LLVM IR instructions. Using the LLVM API, the pass iterates over all LLVM IR instructions and stores the information into the *elements.txt* file. The pass gives a unique name for each instruction taking it account its location according to the LLVM IR hierarchy. The structure of *elements.txt* is shown in the right-hand side of Figure 7. (2) The

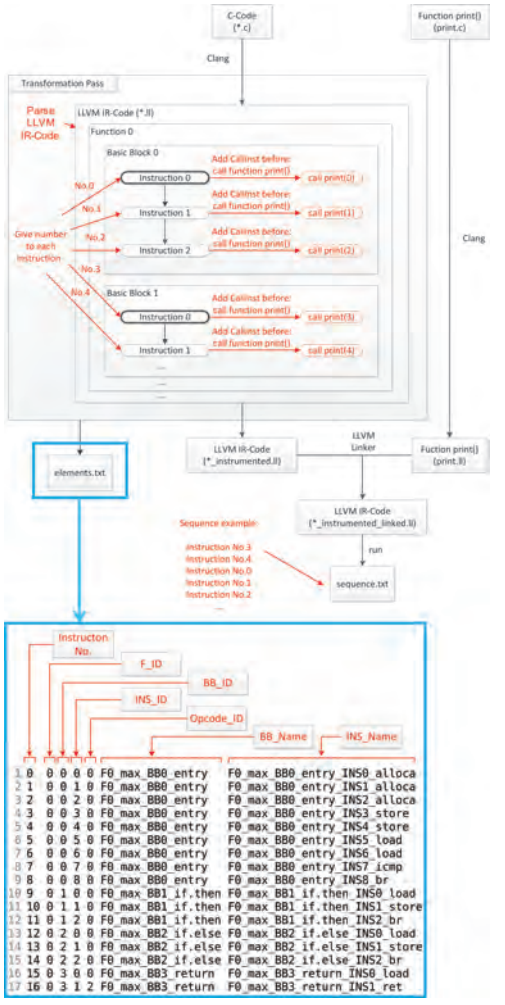


Figure 7. Structure of the transformation pass.

second task is the generation of the instructions execution sequence in order to define the structure of the DEPM control flow as well as control flow transition probabilities. In order to achieve this, we link an external C-code of a "print" function with the original LLVM IR using *llvm-link* tool and append calls to this function after each instruction of the original LLVM IR. After this instrumentation, we compile and run the program. The instructions execution sequence is stored in *sequence.txt*.

2.2 *Analysis pass for data flow analysis*

The second, analysis, pass makes no changes in the original LLVM IR but analyzes the code in order to identify variables and their relations with the LLVM instructions. The overview of the analysis pass is given in Figure 8. This pass iterates over the instructions and identifies related input and output variables (operands) and generates unique names for the future DEPM data storages. The pass also classifies the variables as local or global, and stores this information (data records) into the *datas.txt* file. An example of the data records is shown on the right-hand side of Figure 8. The general challenge here is to define rules for different types of LLVM instructions.

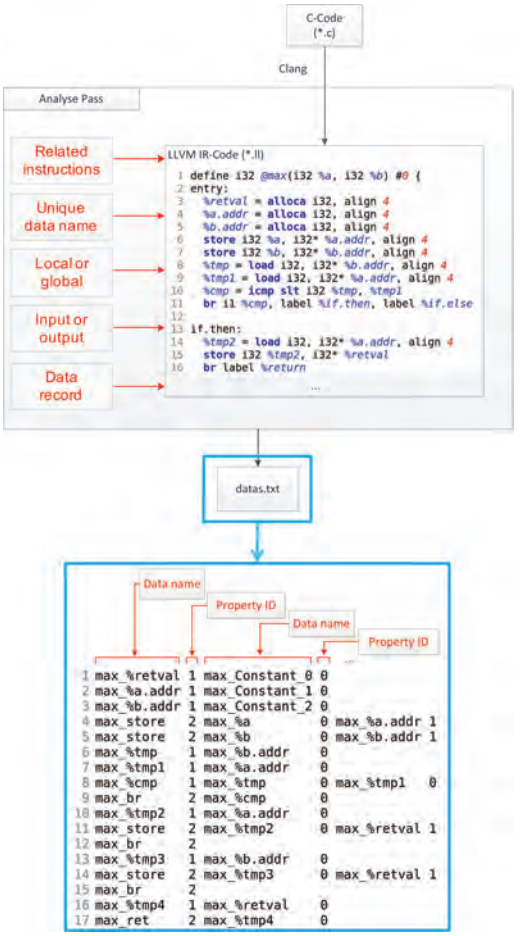


Figure 8. Structure of the analysis pass.

2.3 *DEPM generation*

The last step is the generation of a DEPM model in XML format. The python script parses the *elements. txt*, *sequence.txt*, and *data.txt* files. The information from *elements.txt* is transformed into a set of <element> tags of the DEPM XML file. The *sequnce.txt* file is processed in order to count the number of control flow transitions between the elements and transform this information into control flow arcs and transition probabilities for <cf_arc> tags. The *data.txt* file helps to create a set of data storages using the <data> tags and connect them with the elements using <data_flow> tags. The generation of DEPM conditions is left out of the scope of this method and will be considered later.

2.4 *Hierarchy*

The DEPM supports hierarchical models. A DEPM element can be compound and contain

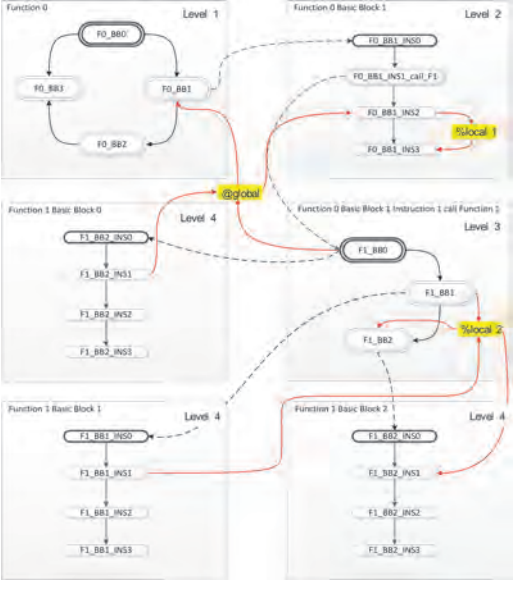


Figure 9. Composition of DEPM elements and data storages in different hierarchical levels.

another DEPM inside. This helps to map the LLVM hierarchy to the DEPM like it is shown in Figure 9. The *main* function of an LLVM module is modeled with the toplevel DEPM, see Level 1 in Figure 9. The elements of this DEPM represent basic blocks of the *main* function. Each of these blocks contains a sequence of instructions that is modeled with the elements of Level 2 DEPMs. Some of these instructions might call other functions. Basic blocks of these functions and their instructions are modeled using the Level 3 and Level 4 DEPMs respectively. The local variables are modeled with the corresponding DEPM data storages and the global variables with the data storages of the toplevel DEPM.

2.5 *Limitations*

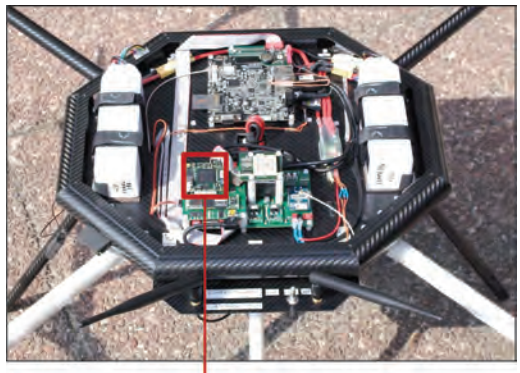
The current software implementation of the transformation method is a proof-of-concept and has a number of technical restrictions:

- A function can be called only in one place.
- A function can not call itself—no recursion.
- Pointers and arrays are not supported.
- All the functions have to be defined in a single module.

The elimination of these restrictions requires some effort, but even now the method can be applied for the safety-critical control software analysis that is usually developed or auto-coded according to standards like MISRA-C (MISRA Ltd 2004) that considerably limits the capabilities of C/C++.

3 CASE STUDY

A part of embedded flight control software of an octocopter UAV has been selected as a case study for the introduced transformation method, see Figure 10. This case study was also discussed in (Morozov & Janschek 2016). The flight vehicle contains a number of onboard computers with guidance, navigation, and control software. A part of the control software, responsible for attitude and rate control, was selected as one of the most critical part of the entire UAV. The *main* function of the selected flight control software is shown in Figure 10. The software is written in C and contains approximately 800 lines of code. It is decomposed into six functions. The functions "read_input", "rate_control", and "ecg" are invoked in each iteration of the main loop, the functions "err_quat" and "attd_ctrl" are executed in each second iteration, and the function "eul_to_quat" in each forth iteration. The software reads sensor data and external inputs ("read_input"), processes them ("eul_to_quat" and "err_quat"), performs attitude and rate control ("attd_ctrl" and "rate_ctrl"), and generates engine commands ("ecg"). The output variable "mtr cmd" is a critical system output.



```
int main(int argc, char *argv[])
{
    input_file = fopen("data/input.txt","r");
    output_file = fopen("data/output.txt","w");
    static uint32_t step = 0;
    while(step<100)
    {
        read_input();
        if(step % 2 == 0)
        {
            if(step % 4 == 0)
                eul_to_quat();
            err_quat();
            attd_ctrl();
        }
        rate_ctrl();
        ecg();
        step++;

        for(uint8_t i = 0; i < 8; i++)
        {
            fprintf(output_file, "%i;",mtr_cmd[i]);
        }
       fprintf(output_file,"\n");
    }
    fclose(output_file);
    fclose(input_file);
    return 0;
}
```

Figure 10. An octocopter UAV system. Top: an embedded computer with the low-level flight control software highlighted with the red frame. Bottom: Ccode of the main functions of the low-level flight control software.

This software was successfully transformed into a set of hierarchical DEPM models using the introduced method. The top-level DEPM is shown in Figure 11. The DEPM control flow branching correctly represents the *if-structure* of the discussed *main* function. Table 1 gives the numerical evaluation of the case study size.

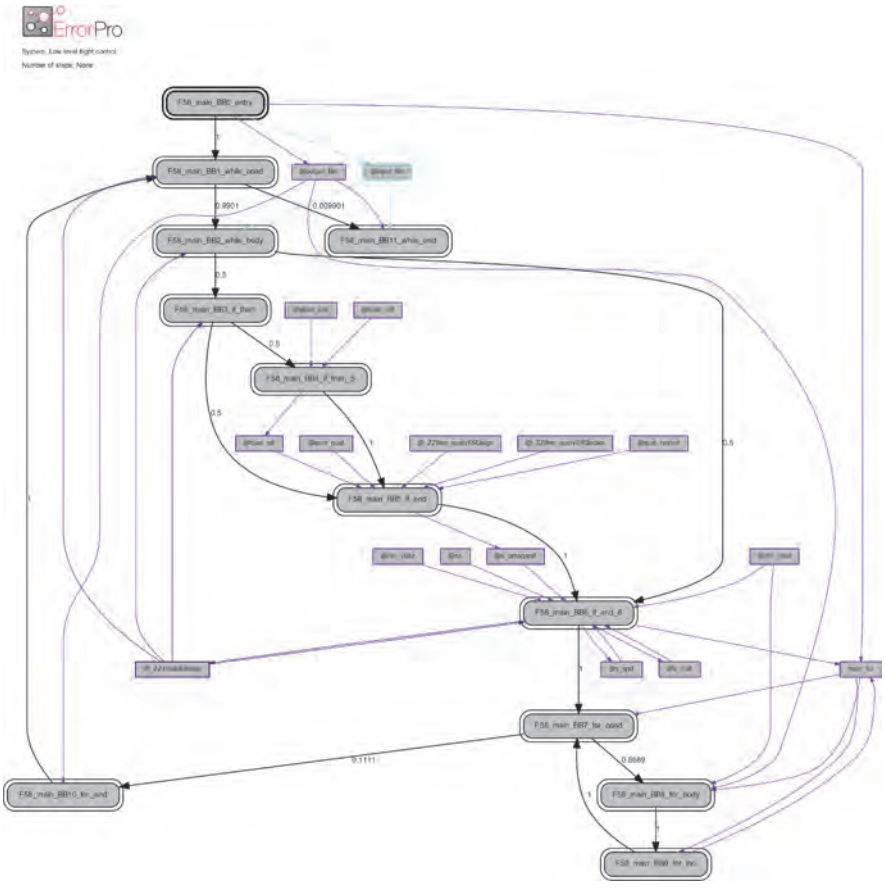


Figure 11. The top-level DEPM model, automatically generated using the introduced transformation method.

Table 1. Numerical evaluation of the case study size.

Number of LLVM instructions	2070
Number of LLVM basic blocks	906
Number of entries in control sequence	507136
Total number of DEPMs	250
Total number of DEPM elements	2976
Total number of DEPM data storages	2104
Total number of DEPM control flow arcs	1882
Total number of DEPM data flow arcs	5099

4 CONCLUSION

A new LLVM-based method for automatic transformation of manually developed code to the stochastic Dual-graph Error Propagation Model (DEPM) has been presented in this paper. The DEPM helps to estimate the impact of a fault of a particular system component or input on the overall system reliability, e.g. to compute the mean number of erroneous values in a critical system output during given operation time.

The DEPM was originally developed in order to support MBSE development approaches. However, the functional software parts usually incorporate manually programmed code and the presented method enables the analysis of the error propagation properties of the manual code for enhanced reliability evaluation of the complete system.

The introduced method maps control and data flow structures based on the LLVM IR. The further development requires the mapping of reliability properties for single DEPM elements using the discussed DEPM conditions. The current software implementation of the transformation method is a proof-of-concept with a number of serious technical limitations that shall be eliminated. The performance of the method will be also optimized.

REFERENCES

Ding, K., T. Mutzke, A. Morozov, & K. Janschek (2016). Automatic transformation of uml system models for model-based error propagation analysis of mechatronic systems. *IFACPapersOnLine 49*(21), 439–446. 7th IFAC Symposium on Mechatronic Systems MECHATRONICS 2016.

Kwiatkowska, M., G. Norman, & D. Parker (2011). Prism 4.0: Verification of probabilistic real-time systems. In *International Conference on Computer Aided Verification*, pp. 585–591. Springer.

Lattner, C. & V. Adve (2004, Mar). LLVM: A compilation framework for lifelong program analysis and transformation. San Jose, CA, USA, pp. 75–88.

MathWorks (2017a). Matlab & simulink: Simulink users guide r2017a.

MathWorks (2017b). Matlab & simulink: Stateflow users guide r2017a.

MISRA Ltd (2004, October). MISRA-C:2004 Guidelines for the use of the C language in critical systems.

Morozov, A. & K. Janschek (2014). Probabilistic error propagation model for mechatronic systems. *Mechatronics 24*(8), 1189–1202.

Morozov, A. & K. Janschek (2016). Flight control software failure mitigation: Design optimization for softwareimplemented fault detectors. *IFAC-PapersOnLine 49*(17), 248–253. 20th IFAC Symposium on Automatic Control in Aerospace ACA 2016.

Morozov, A., K. Janschek, T. Krüger, & A. Schiele (2016). Stochastic error propagation analysis of model-driven space robotic software implemented in simulink. In *Third Workshop on Model-Driven Robot Software Engineering, Leipzig, Germany*.

Morozov, A., R. Tuk, & K. Janschek (2015). Errorpro: Software tool for stochastic error propagation analysis. In *1st International Workshop on Resiliency in Embedded Electronic Systems, Amsterdam, The Netherlands*, pp. 59–60.

Morozov, A., T. Mutzke, B. Ren, & K. Janschek (2018). Aadlbased stochastic error propagation analysis for reliable system design of a medical patient table. *Accepted to the Proceedings of the 64th Annual Reliability & Maintainability Symposium (RAMS)*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Verification of timing properties of a medical patient table case study using probabilistic model checking

T. Mutzke & J. Braun
Diagnostic Imaging, Components and Vacuum Technology, Siemens Healthcare GmbH, Kemnath, Germany

A. Morozov, K. Ding & K. Janschek
Institut für Automatisierungstechnik, Technische Universität Dresden, Germany

ABSTRACT: The analysis and verification of the timing behavior is an important aspect in the model-based development of mechatronic systems. Components of a mechatronic system typically communicate and share data over a network which originates real-time requirements. The network introduced imperfections like package loss and delay have a stochastic nature and thus may lead to the violation of a real-time requirement which is considered as a timing error and has an impact on the system performance and reliability. The model-based timing analysis is a valuable task for the evaluation and prediction of the system reliability. Recently we introduced a method for the model-based analysis of timing errors. It comprises a mapping of a semi-formal baseline system model into a probabilistic model, namely a discrete-time Markov chain model. In this paper, we discuss the analysis of the Markov chain model with respect to the timing requirements with probabilistic model checking techniques. We use the probabilistic model checking tool PRISM and the probabilistic computation tree logic for the property specification. Model checking requires the appropriate transformation of the informal timing requirements to the temporal logic expressions. The verification results show whether a specific timing property is satisfied by the model or not and reveal design flaws in the early design phase. A model of a mobile medical patient table serves as a demonstrative case study.

1 INTRODUCTION

Model-based system development allows the analysis and verification of the system behavior and its conformity to the requirements early in the design phase. A challenging task is the verification of reliability properties. Our research group is focused on the model-based dependability analysis of mechatronic systems and recently introduced a method for the analysis of data errors and their propagation with the goal to identify weak design parts with respect to reliability (Morozov and Janschek 2014). A significant effort was spent in the automatic transformation of baseline system models, specified e. g. with UML and the Architecture Analysis & Design Language (AADL, see (Feiler and Gluch 2012)), into an appropriate formal model for the error propagation analysis (Ding et al. 2016) and (Morozov et al. 2018). Components of a mechatronic system typically communicate over a network. In a networked distributed system with concurrent processes that exchange data periodically, real-time requirements arise between the processes which provide data (sender) and the processes which receive the data (receiver). The violation of the real-time requirements may lead to degradation of the system behavior and has an impact on system reliability. Thus, reliability analysis in early design phases, in particular, the analysis of the timing, is a valuable but challenging task.

Recently we introduced a method for stochastic timing analysis (Mutzke et al. 2016) and extended it in (Mutzke et al. 2018). A semi-formal baseline system behavioral model which is annotated with stochastic timing properties is mapped into an intermediate representation, a formal timed Petri net model. The analysis of the Petri net model considering the stochastic timing properties results in a probabilistic Markov chain model.

In this paper, we discuss the analysis of the Markov chain model with respect to the timing requirements with probabilistic model checking techniques (the highlighted lower part in Figure 1). Verification techniques like simulation and test typically do not cover all possible system execution scenarios. In comparison to this, model checking allows verifying the correctness of the model whether a particular property is satisfied or not. We use the probabilistic model checking tool PRISM (Kwiatkowska et al. 2011) and the proba-

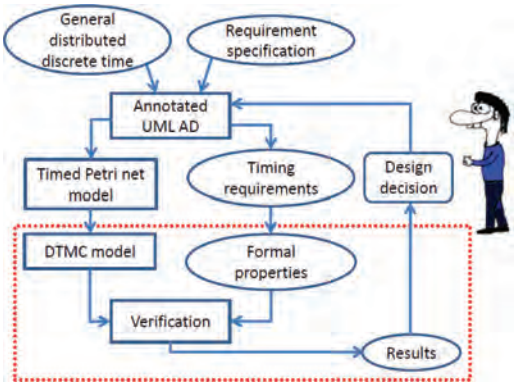


Figure 1. The top-level workflow of the methodology.

bilistic computation tree logic for the property specification. This requires the appropriate transformation of the informal timing requirements to the temporal logic expression and its mapping to the property language of the PRISM tool. The verification results provide valuable information in the early design phase whether a property is satisfied or not and support design decisions which help to achieve an acceptable level of reliability. A model of a mobile medical patient table is used as a demonstrative case study.

This article is organized as follows: Section 2 gives an overview of the related work and empathizes the contribution of this paper. In Section 3, a representative case study model of a mobile medical patient table is introduced. The analysis of the Markov chain model with probabilistic model checking techniques is discussed in Section 4. Finally, Section 5 presents the results and gives an outlook on future research topics.

2 RELATED WORK AND CONTRIBUTION

In (Mutzke et al. 2016) we introduced a method for the model-based, stochastic timing analysis based on a Discrete-Time Markov Chain (DTMC) model. The DTMC model is generated from an annotated baseline system model. In (Mutzke et al. 2016) the method is extended so that the assumptions regarding the distribution of the execution times can be relaxed. However, the scope was on proposing the methodology whereas in this paper the analysis of the DTMC model is discussed in detail.

Several authors proposed valuable methods targeting the mapping of semi-formal UML/SysML diagrams into formal models for the analysis and verification of timing properties. In Ali et al. (2015) a formal verification of SysML internal block diagrams with discrete time constraints is proposed. It is based on the mapping of the model and the requirements into a probabilistic timed automaton and probabilistic computational tree logic expression, respectively. The verification is done with the PRISM model checking tool. However, the annotated discrete time constraints are non-probabilistic. The specified properties derived by the user requirements thus are also non-probabilistic. In Baouya et al. (2015) a probabilistic and timed verification framework of SysML state machine diagrams, annotated with time and probability properties, is introduced. However, the timing properties, which are assigned to the states, are specified as an interval with minimum and maximum values and the probabilistic properties are attached to the decision nodes and represent the control flow probabilities. In Jarraya et al. (2007) an approach for the automatic verification and performance analysis of time-constrained SysML Activity Diagrams is presented. The SysML Activity Diagram is mapped directly to the corresponding DTMC models and verified with the PRISM model checking tool.

3 CONTRIBUTION

In this paper, we discuss the analysis of the Markov chain model. Probabilistic model checking is used to verify the correctness of the model with respect to the timing requirements. A model of a mobile medical patient table serves as a representative case study. We present a set of properties derived by the timing requirements, which are verified using probabilistic model checking. The properties are formalized using the probabilistic computation tree logic. We use the probabilistic model checking tool PRISM to verify whether the Markov chain model, expressed with the PRISM modeling language, satisfies the defined properties, expressed with the PRISM property language. The verification results help revealing design flaws regarding the timing behavior and estimating the system reliability to achieve an acceptable level of reliability.

4 CASE STUDY

A concept study of a mobile medical patient table (MPT, Figure 2), introduced in Mutzke et al. (2018), is used as a demonstrative example. A typical use case scenario for a mobile MPT is the patient transport between the preparation room and the examination room. In the examination room, the MPT needs to be precisely positioned relative to specific modalities, e. g. a magnetic resonance imaging device (MRI). The MPT is

Figure 2. A design concept of a mobile medical patient table (Siemens Healthcare GmbH).

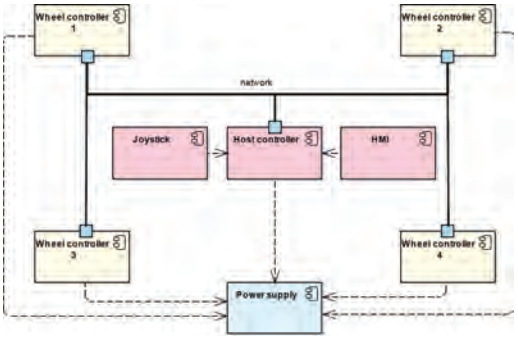


Figure 3. The UML components diagram of the MPT.

equipped with four electrical driven omnidirectional wheels. This allows autonomous movements in arbitrary directions, including lateral movements and rotations around an arbitrary vertical axis, without a steering mechanism. The movement may be commanded by an operator or a modality, e. g. position correction based on the evaluation of MRI images.

The mechatronic system MPT comprises four motion controllers for each individual omnidirectional wheel and the main controller which provides the interfaces and computes the movement trajectory based on the target position. The main controller communicates via a network with the motion controllers. Figure 3 shows the UML component diagram that describes the interfaces of the discussed MPT components. The dashed lines represent the dependency from the components to the power supply on board of the MPT.

Network induced imperfections may lead to a degradation of the system performance and reliability, e. g. a delayed position update may lead to discontinuity of position, movement, velocity or acceleration, resulting in leaving the planned movement path and fail to reach the target position.

4.1 *Baseline system model*

An abstract behavioral model, the UML Activity Diagram in Figure 4, serves as a baseline model of the MPT for the timing analysis (Mutzke et al. 2018). It represents a behavioral model for the automated positioning task. It comprises the control flow as well as the data flow of the system. The atomic executable elements, the *actions* are annotated with stochastic timing properties listed in Table 1. This abstract model is subject to refinement in follow-up design iterations, however, this level is sufficient for the timing analysis.

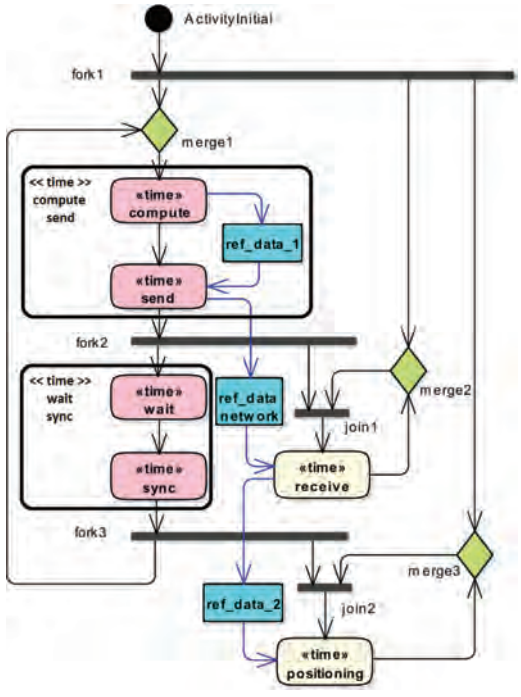


Figure 4. The UML activity diagram of the MPT.

Table 1. Annotated timing properties of the activity diagram in Figure 4.

Action	Time	Probability
compute	(6, 7, 8)	(0.05, 0.9, 0.05)
send	(3, 4, 5)	(0.9, 0.09, 0.01)
wait	(39)	(1.0)
sync	(1, 2, 3)	(0.99, 0.009, 0.001)
receive	(10, 20, 30, 40)	(0.9, 0.05, 0.03, 0.02)
positioning	(5, 10, 15)	(0.99, 0.009, 0.001)
(compute, send)	(9, 10, 11, 12, 13)	(0.045, 0.8145, 0.1265, 0.0135, 0.0005)
(wait, sync)	(40, 41, 42)	(0.99, 0.009, 0.001)

The actions on the left hand side of the activity diagram, namely *compute*, *send*, *wait* and *sync* represent the main controller process P_M (red). The actions *receive* and *positioning* represent one of the wheel controller processes P_W (yellow). For the sake of simplicity, the model contains only one of the four wheel controller processes. The Activity starts at the *ActivityInitial* node. The black horizontal bars represent *fork* and *join* nodes indicating the begin and the end of parallel execution paths. The black edges represent the control flow and the blue edges the data flow respectively. The action *compute* plans the movement trajectory, time-equidistant steps of the movement path, for every wheel and provides the data *ref_data*_1 which are used by the action *send*. The data *ref_data_network* are an output of the action *send* and input of the action *receive* of the wheel controller process. The *fork*2 node ensures that the action *receive* starts after the action *send* is finished and both, *wait* and *receive* start at the same time. The *join*1 node ensures that action *receive* need to be finished processing the data *ref_data_network*[*k*] before processing the data *ref_data_network*[*k* + 1]. The action *sync* initiates the activation of the transmitted data *ref_data*_2 from action *receive* to action *positioning*. The behavioral model implies the timing requirements (i) that action *receive* has to be finished before the action *sync* is finished which can be expressed by the statement $t_{wait} + t_{sync} > t_{receive}$ and (ii) that the action *positioning* has to be finished before the action *send* is finished which can be expressed by the statement $t_{compute} + t_{send} > t_{positioning}$. The baseline model is annotated with the discrete timing properties listed in Table 1. The values and its distributions are chosen arbitrarily and are subject to refine while the design is enhanced. However there are three types of distributions: deterministic for the action *wait*, monotonically decreasing for the actions *send,sync,receive* and a variance around a characteristic mean for the action *compute*. Figure 5 shows a timing chart for the MPT representing about three cycles of the nominal, error-free system run. The horizontal bars visualize the action execution times and their order of execution in each process P_M and P_W, respectively.

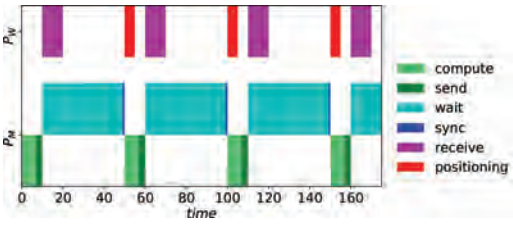


Figure 5. The nominal, error-free timing chart of the MPT.

4.2 *Stochastic timing analysis*

According to Mutzke et al. (2018), in this section, the approach for the stochastic timing analysis, applied to the case study baseline system model, is briefly demonstrated.

The analysis comprises the steps:

- Reducing the baseline system model and computing probabilistic timing properties;
- Mapping of the UML AD to a timed Petri net model;
- Generation of the DTMC model.

4.2.1 *Baseline system model reduction*

The baseline system model elements which are sequentially executed, namely *compute*, *send* and *wait*, *sync* are subject to model reduction. For the reduced single elements (*compute*, *send*) and (*wait*, *snyc*), which substitutes the elements in the baseline model, we have to compute the sum of the execution times. The execution times are assumed to be independent random variables with a discrete distribution. The sum is also a random variable and its distribution can be determined by the convolution of the distributions of each random variable. For two discrete distributions m_A, m_B it follows

$$(m_A * m_B)(j) = \sum_k m_A(k) \cdot m_B(j-k), \quad k, j \in \mathbb{Z}$$

where *j*,*k* iterates over the values of m_A and m_B, respectively. The asterisk symbol (∗) represents the convolution operator (Bronstein et al. 2005). The last two rows in Table 1 show the computed results for the reduced elements (*compute*, *send*) and (*wait*, *sync*) of the case study baseline system model.

4.2.2 *Computing timing properties*

The timing requirements now can be reformulated: (i) $t_{waitsync} > t_{receive}$ and (ii) $t_{computesend} > t_{positioning}$. The violation of a timing requirement is considered as the occurrence of a timing error. Thus, the probability of the occurrence of a timing error with respect to the timing requirements can be expressed as: (i) $Prob\left(t_{receive} \geq t_{waitsync}\right)$ and (ii) $Prob\left(t_{positioning} \geq t_{computesend}\right)$.

The probability of the occurrence of the timing error is obtained by computing the joint probabilities which satisfy the condition $t_{receive} \geq t_{waitsync}$. For (i) with $t_{waitsync} = t_A$ and $t_{receive} = t_B$ this can be computed with

$$Prob\left(t_A \leq t_B\right) = \sum_{u_2 = t_{Bmin}}^{t_{Bmax}} \sum_{u_1 = u2}^{t_{Amax}} p\left(u_1, u_2\right)$$

Where $p(u_1,u_2) = Prob\left(t_{Au_1}\right) \cdot Prob\left(t_{Bu_2}\right)$ is the joint probability of a single realization of t_A and t_B. The subscripts *min* and *max* indicate the border of the individual interval respectively.

Finally, we obtain the following results for the probability of the occurrence of a timing error for the case study model:

i. $Prob\left(t_{receive} \geq t_{waitsync}\right) = 0.0198$,
ii. $Prob(t_{positioning} \geq t_{computesend}) = 0.008735$, respectively

The results show that (i) approximately two out of one hundred data *ref_data_network* won't be received in time by the wheel controller process and (ii) approximately in nine out of one thousand execution cycles the *positionin* task exceeds the *compute* and *send* tasks.

4.2.3 *Mapping into a formal Petri net model*

A semi-formal UML Activity Diagram can be mapped into a formal Petri net model by applying the rules defined in Störrle and Hausmann (2004) and adapted in Mutzke et al. (2016). The Petri net represents the control flow of the baseline system model. Figure 6(a) shows the corresponding Petri net model of the baseline system model, the UML Activity Diagram in Figure 4, where the immediate transitions are represented by thin black bars while the timed transitions are represented by a black rectangle. The *ActivityInitial* of the Activity Diagram is mapped to the place p_1 in the Petri net model and is initially marked with a token.

Table 2 shows the assignment of the UML Activity Diagram nodes to the nodes of the Petri net model.

The timing requirements following the notation of the Petri net can be expressed with (i) the transition *t*6 fires before the transition *t*5 and (ii) the transition *t*9 fires before the transition *t*2. A precondition is that the pre-places of the transitions *t*6, *t*5 and *t*9, *t*2, namely *p*7, *p*8 and *p*2, *p*11, are marked simultaneously. The Petri net reveals that this is ensured by design.

4.2.4 *Generation of a discrete-time Markov chain model*

A reachability analysis of the generated Petri net model reveals all markings reachable from the initial marking $\boldsymbol{m}_0$. The result is a directed graph *RG* with the state space *S*. Considering the timing requirements (i) and (ii), we can identify the states $S_{nom} \subset S$ and $S_{err} \subset S$ which represents the nominal operational states and the erroneous states, respectively.

Simultaneously activated timed transitions in the Petri net model are represented by nondeterministic state transitions in the *RG* characterized by places with two or more outgoing edges. To this edges we can assign the state transition probabilities, computed

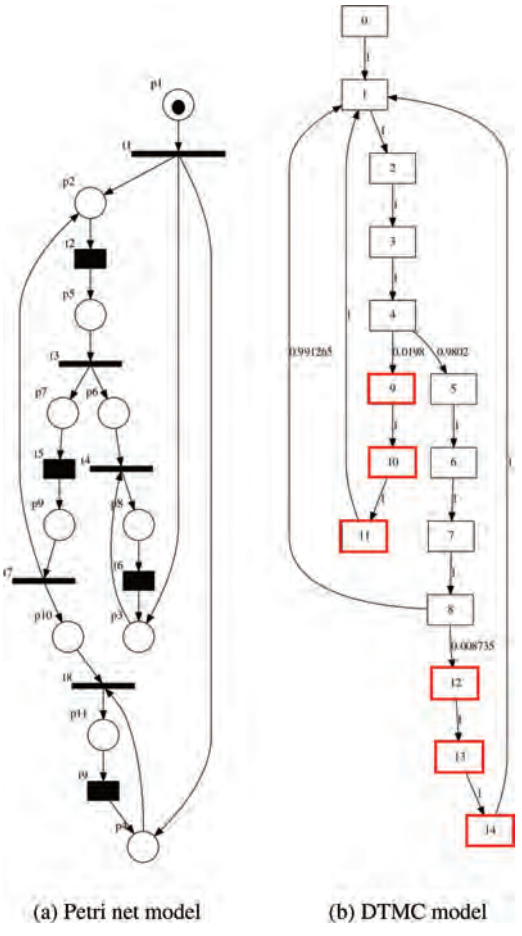


Figure 6. Graphical representations of the Petri net model and its corresponding DTMC model.

Table 2. The assignment of the activity diagram nodes in Figure 4 to the Petri net nodes in Figure 6(a).

PN nodes	AD nodes
p1	ActivityInitial
t(1, 3, 7)	fork(1, 2, 3)
p(2, 3, 4)	merge(1, 2, 3)
p(5, 6, ..., 11, 12)	auxiliary
t(4, 8)	join(1, 2)
t2	action *compute*, *send*
t5	action *wait*, *sync*
t6	action *receive*
t9	action *positioning*

in Section 3.2.2. The reachability graph attached by state transition probabilities represents a discrete time Markov chain model (DTMC). Figure 6(b) shows the obtained DTMC model with states in nominal oper-

ation S_{nom} and the erroneous states S_{err}, highlighted in red, generated out of the Petri net model that is shown in Figure 6(a). Table 3 contains the assignment of the DTMC states to the Petri net places. The occurrence of a timing error following the notation of the DTMC can be identified with (i) the state s_9 is visited and (ii) the state s_{12} is visited, respectively. The DTMC model comprises 65 states and 145 edges, however in Figure 6(b) only the states in nominal operation path with a cyclic behavior and two path fragments, each consisting of three erroneous states, are shown, which are highlighted inred.

5 ANALYSIS OF THE MARKOV CHAIN MODEL

The generated Markov chain model is the foundation for the analysis of specific properties related to the timing behavior. Formally, according to (Baier & Katoen 2008), a DTMC is a tuple

$$\mathcal{M} = \left(S, P, \iota_{init}, AP, L \right)$$

$S = \left(s_1, s_2, ..., s_n \right)$	is a finite set of states;
$\boldsymbol{P} : S \times S \to \left[0,1 \right]$	is the transition probability function such that for all states s: $\sum_{s' \in S} P(s,s') = 1$;
$\iota_{init} : S \to \left[0,1 \right]$	is the initial distribution such that $\sum_{s \in S} \iota_{init}(s) = 1$;
$AP = \left(nom, err \right)$	is a set of atomic propositions
$L : S \to 2^{AP}$	a labeling function.

The states of the DTMC model are labeled with the atomic propositions *nom* and *err*, respectively (Table 3). The analysis of the DTMC model is done with the probabilistic model checking tool PRISM. Listing 1 shows the DTMC specified in the PRISM modeling language.

Lines 31 to 41 in Listing 1 defining rewards, namely visiting the states s_9, s_{12} and the disjunction $s_9 \vee s_{12}$, that indicate the occurrence of timing error. This allows to compute the expected number of errors during specific number of execution steps.

To specify the properties, we use the probabilistic computation tree logic (PCTL). This is a temporal logic based on the computation tree logic (CTL). A PCTL formula formulates conditions on states of a Markov chain. The notation used in this paper is according to (Baier and Katoen 2008). Compared to CTL which uses universal and existential path quantifiers ∀ and ∃, the PCTL uses the probabilistic operator $\mathbb{P}_J(\varphi)$, where φ is a path formula, and J is an interval of [0, 1]. The syntax for a PCTL state formula is given by:

```
1  dtmc
2
3  label "nom" = s=0 | s=1 | s=2 | s=3 | s=4 | s=5 | s=6 | s=7 | s=8;
4  label "err" = s=9 | s=10 | s=11 | s=12 | s=13 | s=14;
5
6  module mpt
7
8  s :[0..14]  init  0;  // states
9
10 [] s=0 -> 1:(s'=1);
11 [] s=1 -> 1:(s'=2);
12 [] s=2 -> 1:(s'=3);
13 [] s=3 -> 1:(s'=4);
14 [] s=4 -> 0.9802:(s'=5) + 0.0198:(s'=9);
15 [] s=5 -> 1:(s'=6);
16 [] s=6 -> 1:(s'=7);
17 [] s=7 -> 1:(s'=8);
18 [] s=8 -> 0.991265:(s'=1) + 0.008735:(s'=12);
19
20 [] s=9 -> 1:(s'=10);
21 [] s=10 -> 1:(s'=11);
22 [] s=11 -> 1:(s'=1);
23
24 [] s=12 -> 1:(s'=13);
25 [] s=13 -> 1:(s'=14);
26 [] s=14 -> 1:(s'=1);
27
28 endmodule
29
30
31 rewards "num_error_9_12"
32 [] s = 9 | s = 12 : 1;
33 endrewards
34
35 rewards "num_error_9"
36 [] s = 9 : 1;
37 endrewards
38
39 rewards "num_error_12"
40 [] s = 12 : 1;
41 endrewards
```

Listing 1. The DTMC model expressed in the PRISM modeling language.

Table 3. Assignment of the PN places to the DTMC states.

DTMC states	PN places	DTMC label
0	(1)	"nom"
1	(2, 3, 4)	"nom"
2	(3, 4, 5)	"nom"
3	(3, 4, 6, 7)	"nom"
4	(4, 7, 8)	"nom"
5	(3, 4, 7)	"nom"
6	(3, 4, 9)	"nom"
7	(2, 3, 4, 10)	"nom"
8	(2, 3, 11)	"nom"
9	(4, 8, 9)	"err"
10	(2, 4, 8, 10)	"err"
11	(2, 8, 11)	"err"
12	(3, 5, 11)	"err"
13	(3, 6, 7, 11)	"err"
14	(7, 8, 11)	"err"

$$\Phi ::= true \left| a \right| \Phi_1 \wedge \Phi_2 \left| \neg \Phi \right| \mathbb{P}_J \left(\varphi \right)$$

Where $a \in AP$ is a atomic proposition. The syntax for a PCTL path formula is given by:

$$\varphi ::= \bigcirc \Phi \,|\, \Phi_1 \cup \Phi_2 \,|\, \Phi_1 \cup^{\leq n} \Phi_2$$

where $\bigcirc$ is the next operator. Formula $\bigcirc\Phi$ holds for a path if Φ holds for the next state in the path. $\Phi_1 \cup \Phi_2$ holds for a path if there is a state in the path where Φ_2 holds and Φ_1 holds in all states visited before. $\Phi_1 \cup^{\leq n} \Phi_2$ is the step-bounded variant meaning that Φ_2 will hold within at most n steps.

The PCTL is used to specify the following properties relevant for the analysis of the timing:

```
const int k;

// (a1) probability of a timing error (i) in k steps
P=? [ F<=k (s=9) ]

// (a2) probability of a timing error (ii) in k steps
P=? [ F<=k (s=12) ]

// (a3) probability of a timing error (i) or (ii) in k steps
P=? [ F<=k (s=9|s=12) ]

// (b1) cumulative reward: the expected number of errors (i)
// within k steps
R{"num_error_9"}=? [ C<=k ]

// (b2) cumulative reward: the expected number of errors (i)
// within k steps
R{"num_error_12"}=? [ C<=k ]

// (b3) cumulative reward: the expected number of errors
// (i) or (ii) within k steps
R{"num_error_9_12"}=? [ C<=k ]

// (c) infinitely often "nom"
P>=1 [ G F "nom" ]

// (d) always an error state implies within at most three
// steps a nominal state is reached
P>=1 [ G "err" => (X X X "nom") ]
```

Listing 2. The properties specified in the PRISM property language.

a. What is the probability of the occurrence of a timing error (i), (ii) and the disjunction (i) $\vee$ (ii) within k execution steps?
b. Cumulative reward: What is the expected number of errors within k execution steps?
c. The system will always recover from a timing error and reach a nominal operation state.
d. Whenever an error state occurs, in at most three steps a nominal operational state is reached.

To verify the properties, they have to be formalized. In Listing 2 the properties are specified in the PRISM property language.

6 RESULTS AND CONCLUSIONS

Table 4 lists the formal expressions of the defined properties in PCTL as well as the PRISM property language.

Figures 7 and 8 show the verification results for the properties (a) and (b), respectively. The graphs represent the the timing error (i), (ii) and the disjunction (i)$_\vee$ (ii) highlighted green, red and blue, respectively.

For k = 1000 steps almost sure one of the timing errors occur, but the (ii) is less probable. Hence design enhancements should be focused to decrease the probability of (i). The verification result of property (c) means that a nominal operational state is visited infinitely often. Hence the system will not remain in an erroneous state. The verification result of property (d) states that the system always recovers from an erroneous state within at most three steps.

Table 4. Formal property specification and results.

Property	PCTL	PRISM	Result
(a1)	$\mathbb{P}_{=?}(\Diamond_{<=k}\ (s=9))$	$P=?\,[F <= k (s=9)]$	see Figure 7
(a2)	$\mathbb{P}_{=?}(\Diamond_{<=k} (s=12))$	$P=?\,[F <= k (s=12)]$	see Figure 7
(a3)	$\mathbb{P}_{=?}(\Diamond_{<=k} (s=9 \vee s=12))$	$P=?\,[F <= k (s=9 \mid s=12)]$	see Figure 7
(b1)	–	$R\{"num\_error\_9"\}=?\,[C <= k]$	see Figure 8
(b2)	–	$R\{"num\_error\_12"\}=?\,[C <= k]$	see Figure 8
(b3)	–	$R\{"num\_error\_9\_12"\}=?\,[C <= k]$	see Figure 8
(d)	$\mathbb{P}_{=1}(\Box\Diamond\ "nom")$	$P>=1[G\ F\ "nom"]$	true
(e)	$\mathbb{P}_{=1}(\Box\ "err" \rightarrow (\bigcirc\bigcirc\bigcirc\ "nom"))$	$P>=1\,[G\ "err" => (XXX\ "nom")]$	true

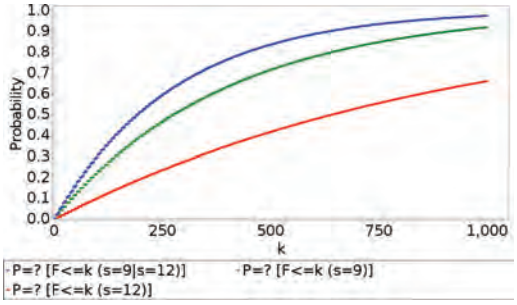


Figure 7. The probability of a timing error within k steps.

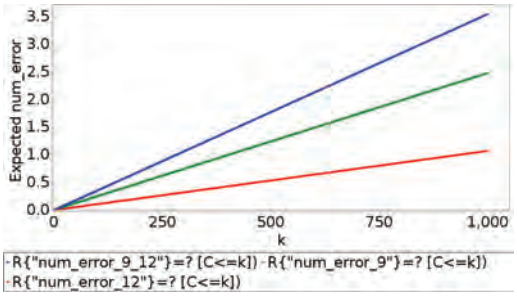


Figure 8. The expected number of timing errors within k steps.

In this paper, we discussed the analysis of the Markov chain model with respect to the timing requirements with probabilistic model checking techniques. This requires the appropriate transformation of the informal timing requirements to the temporal logic expression and its mapping to the property language of the PRISM tool. The results are valuable for the assessment of the system reliability and help to achieve an acceptable level of reliability.

A challenging task is to formalize the appropriate timing requirements from an informal requirement specification or to derive the implicit timing requirements from the baseline system model, where the latter is done manually in this paper. Thus a future research topic is the automatic recognition of the timing requirements and their mapping to a temporal logic expression.

REFERENCES

Ali, S., M. A. Basit-Ur-Rahim, & F. Arif (2015, June). Formal verification of internal block diagram of sysml for modeling real-time system. In *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pp. 1–6.

Baier, C. & J.-P. Katoen (2008). *Principles of Model Checking*. The MIT Press.

Baouya, A., D. Bennouar, O. A. Mohamed, & S. Ouchani (2015, April). A probabilistic and timed verification approach of sysml state machine diagram. In *2015 12th International Symposium on Programming and Systems (ISPS)*, pp. 1–9.

Bronstein, I. N., K. A. Semendjajew, & G. Musiol (2005). *Taschenbuch der Mathematik*. Deutsch (Harri).

Ding, K., T. Mutzke, A. Morozov, & K. Janschek (2016). Automatic transformation of uml system models for model-based error propagation analysis of mechatronic systems. *IFACPapersOnLine 49*(21), 439–446.

Feiler, P. H. & D. P. Gluch (2012). *Model-Based Engineering with AADL*. Upper Saddle River, NJ: Addison-Wesley.

Jarraya, Y., A. Soeanu, M. Debbabi, & F. Hassa¨ıne (2007). Automatic verification and performance analysis of timeconstrained SysML activity diagrams. In *Proc. 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems* (ECBS'07), pp. 515–522. IEEE.

Kwiatkowska, M., G. Norman, & D. Parker (2011). PRISM 4.0: Verification of probabilistic real-time systems. In G. Gopalakrishnan and S. Qadeer (Eds.), *Proc. 23rd International Conference on Computer Aided Verification (CAV'11)*, Volume 6806 of LNCS, pp. 585–591. Springer.

Morozov, A. & K. Janschek (2014). Probabilistic error propagation model for mechatronic systems. *Mechatronics 24*(8), 1189–1202.

Morozov, A., T. Mutzke, B. Ren, & K. Janschek (2018). Aadlbased stochastic error propagation analysis for reliable system design of a medical patient table. RAMS - *The Annual Reliability and Maintainability Symposium*. Accepted paper.

Mutzke, T., K. Ding, A. Morozov, K. Janschek, & J. Braun (2016, Sept). Model-based analysis of timing errors for reliable design of mechatronic medical devices. In *2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol)*, pp. 233–238.

Mutzke, T., A. Morozov, K. Ding, K. Janschek, & J. Braun (2018). Stochastic model-based analysis of timing errors for mechatronic systems with user-defined general discrete-time distributions. *10th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*. Submitted paper.

Störrle, H. & J. Hausmann (2004). Semantics of uml 2.0 activities. *Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Direct integration of safety analysis in a model based system engineering process: Lessons learned from Ariane 6 control bench family RAMS studies

R. López, A. Guillén & J. Sanmartí
GTD, Barcelona, Spain

C. Canart & J. Masfrand
CNES/DLA, Paris, France

ABSTRACT: As of today, two of major topics in Reliability, Availability, Maintainability and Safety (RAMS) field are how to deal with the growing complexity of new systems and how to reduce the time required to perform RAMS analysis. Complexity is perceived as the most challenging factor when developing "safe proved" critical systems. During the last decade Model Based System Engineering has been broadly deployed in the industry in order to manage complexity. Complexity is also managed by adopting incremental and/or iterative development lifecycles. It is therefore imperative to integrate RAMS analyses with such means and processes. This implies dealing with a live, changing design baseline. To deal with the two major topics detailed above, a new practical RAMS modelling methodology is presented in this paper, based on a generic Model Based System Engineering (MBSE) tool with an Engineering Model (EM) shared between Design and RAMS teams. Requirements, Functional Trees, System Architecture and Detailed Design are included in this model. Furthermore the model includes Functional Failure Mode and Effect Analyses (FMEAs), Components FMEAs, Feared Events and related Fault Tree Analysis. In this way, safety and engineering models are intrinsically linked and RAMS teams can fully exploit traceability (established and maintained by the design team) between Functions and Components & Interfaces. An automated export of RAMS related data, (Reliability Data & Fault Tree topologies) allows numerical calculations in a reliability tool. This methodology can deal with both software and hardware technologies. It is valid for highly critical systems (safety related) but also for less critical systems with complex availability modelling (production means). A real use case for this methodology is presented; the Ariane 6 Control Bench Family (A6 CBF). A6-CBF is a family of 9 control benches, covering all the ground control needs of future Ariane 6 launcher in terms of production, validation, training and launch operations.

1 INTRODUCTION

Model Based Safety Analysis (MBSA) has become a trending topic in the dependability field since the mid-2000s (Lesage & Kruse, 2011; Joshi, Keimdahl, Miller, & Whalen, 2005). Formal methods will be the RAMS analyst's Swiss army tool to perform safety case of complex systems (Rauzy & Chaire, 2014). They already have been successfully used in highly regulated industries where safety certification is mandatory (e.g. certification of the flight control system of the aircraft Falcon 7X, Bieber et al., 2008). On the other hand, the use of formal methods or specific languages for dysfunctional modeling (AADL, Altarica, eventB, Safety Architect from ALL4TEC, Safety Designer from Dassault) requires a specific knowledge and a considerable amount of time which is usually only economically viable in systems to be produced in large series.

The proposed methodology in this paper is an intermediate strategy between traditional and innovative techniques. It allows the use of well-known methods (Fault Trees Analysis (FTA), Failure Mode Effects Criticality Analysis (FMECA), etc.) with very complex systems. Our approach ensures the completeness and consistency of the RAMS studies, the most challenging parts within "manual" safety analyses. It also allows to trace and link elements proper to RAMS analysis with any other element of an Engineering Model (EM). That may sound obvious but it adds a big value for design and exploitation teams. It results in something as simple but powerful as identifying which diagnostic cover a particular failure mode and in which test case the expected behavior has been validated with two mouse clicks.

2 PROPOSED METHODOLOGY

To integrate RAMS in MBSE lifecycle, system and dependability engineering should share a common EM. RAMS analysis shall be part of the whole system development cycle using information available on EM at every moment from Requirements Elicitation to Site Acceptance phases.

2.1 *Functional Analysis (FA)*

Functional Analysis (FA) is achieved by defining functions to be performed by the system. Main functions are decomposed into several subfunctions recursively until a sufficiently detailed functional tree is obtained. This decomposition allows functional specification of the system in a complete way. Every single functional requirement shall be linked to a function. Function performance requirements shall be identified usually as a non- functional requirement.

2.2 *Functional FMEA (F-FMEA)*

For each function established in the functional tree, five generic Functional Failure Modes (F-FM) based on (Betancourt, Birla, Gassino, & Regnier, 2011) are created. These 5 failure modes are defined as follows:

F-FM1: Fail to perform the function at the required time
F-FM2: Fail to perform the function with correct value
F-FM3: Performance of an unwanted action
F-FM4: Interference or unexpected coupling with another function
F-FM5: Unable to remove the function

Once these 5 F-FM are associated to each function, the effects of the failure modes of a function are studied in a systematic and recursive way and the relevant relationships are established with the failure modes of the higher functions following the functional arborescence defined. That builds up a Functional Failure modes Fault Tree. For that, OR and AND gates can be implemented defining an appropriate kind of aggregation/decomposition link between the functions. A crucial element is provided by F-FM4 which establishes relationships between F-FM whose functions are not directly related in the functional hierarchy tree but a possible failure in one of them could impact the correct performance of the other. This feature is very useful while analyzing transversal Uses Cases requiring a sequence of particular functions.

2.3 *Feared Event (FE) identification*

Feared Events (FE) are identified mainly by the requirements defined by the client based on previous Preliminary Risk Assessment (PRA). Each FE is assigned a certain severity level as defined in ECSS Q30 based on the impact that it may have on the system.

2.4 *Link between Functional Failure Modes (F-FM) and Feared Events (FE)*

Every FE shall be related to one or more F-FM, which become the cause for the occurrence of this FE. The "higher" the Functional Tree is related to a FE, the more F-FM and consequently the more C-FM are potentially involved. For each FE a Fault Tree Analysis (FTA) is generated whose result is the assignation of a criticality level to each function. This functional criticality level allows RAMS requirements allocation usually by using function performance requirements.

2.5 *Traceability between Functional Analysis (FA) and System Component Hierarchy (SCH)*

During the Preliminary Design Phase a system Architecture is defined and high level components in the Product Breakdown Structure (PBS) are identified. As the Detailed Design Phase progresses

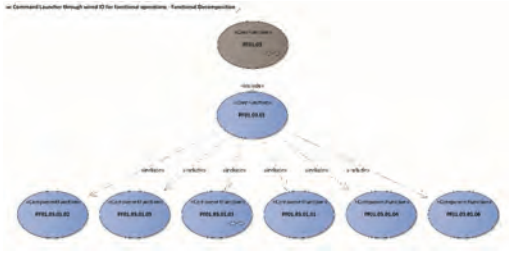


Figure 1. Functional analysis tree.

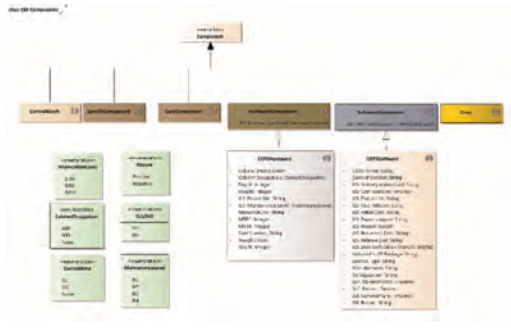


Figure 2. HW/SW component domain model.

Functional Analysis and component design is refined producing sub-functions implemented by subcomponents. In this way the function tree is related to a hierarchy of components in the PBS. The relationships between the components and the functions they implement must be established at the required level of detail. SCH decomposition should arrive until the Line-Replaceable Unit level the lowest system element to be exchangeable on the field. At the end, each LRU component shall implement one or various functions of the lowest level of the functional tree.

2.6 *Identification of Component Failure Modes (C-FM)*

For every component a set of C-FM is identified. For HW type components the use of a standard Failure Mode Distribution database (RMQSI Knowledge Center, 2016) is recommended. Similar HW_Components can be grouped in HW_Families sharing a list of common Failure Modes.

SW Failure modes are identified in relation to the functionalities of the SW component being analyzed. For SW components generics SW failure modes are available on the literature (Reifer, 1979). One of more general sets of SW failure modes is defined in. (ECSS, 2017) Three SW-FM are proposed:

SW-FM1: functionality not performed;
SW-FM2: functionality performed wrongly (e.g. wrong/null data provided, wrong effect);
SW-FM3: functionality performed with wrong timing (e.g. too late).

The persistence of the error causing the failure mode and the nature of the software application to be analyzed has to be taken into account carefully. In contrast to HW C-FM where the random nature of physical ageing plays a major role, most of the SW-FM can be covered/prevented by some level of testing (unit/integration/validation testing). EM allows traceability between these failure modes and the related diagnostic and tests to be implemented in further phases.

Figure 3. RAMS domain model.

2.7 *Link between Component Failure Modes (C-FM) and Functional Failure Modes (F-FM)*

Each Software/Hardware component (SW/HW) has its own C-FM which are identified following the procedure established in the previous sections. These C-FM can be linked to those F-FM that belong to the function or functions that the component implements. To achieve accurate results this analysis shall be performed by the RAMS team together with the team responsible for the component design.

2.8 *Results from RAMS studies implemented on the engineering model*

2.8.1 *Component criticality analysis*

Following the traces "Feared Event > Functional Failure Mode > Functional Failure Mode > … > Component Failure Mode > Component" it is possible to identify all components leading to a Feared Event. Components are then classified according to the criticality of the Feared Event. A set of Non-Functional requirements is imposed to each component depending of its criticality level.

2.8.2 *Quantitative and qualitative RAMS requirements verification*

Exportation from EM to BlockSim of related Fault Trees diagrams (via dedicated script generating an XML file per diagram) is possible. This exportation is limited to simple Fault Trees (managing OR & AND gates). However, the script manages "repeated" events and events present in different points of the same Fault Tree. Probabilities are then computed based on reliability parameters of component failure modes. This reliability data is available on the EM, and exported previously to BlockSim. For availability Fault Trees, a single fault mode per component is used, aggregating all failure mode distributions. Fault Trees in BlockSim can be used also to calculate Minimal Cut Sets allowing validation of qualitative requirements like the absence of Single Point of Failure, Fail Operational, Fail Safe or similar criteria.

3 USE CASE: ARIANE 6 CONTROL BENCH FAMILY

The Ariane 6 program, approved in 2014, aims to reduce the cost of space launch by half compared to Ariane 5. Europe will remain number one in the

commercial launch services market while responding to the needs of European institutional missions. In the particular case of Ariane 6, a MBSE approach is incorporated in the design of its control benches (Cherqui, Comery, & Lesens, 2016). Moreover, some MBSA efforts were taken within Ariane 5 environment (Ercilbengoa, Schoenig, & Hutinet, 2010). A6-Control Bench Family is the family of nine Control Benches (seven to be deployed in continental Europe, two in European Spaceport in Kourou), covering all Ground Control needs of Ariane 6 launcher in terms of production, validation, training and launch. A6-CBF will be developed by a consortium led by GTD, formed by GTD and CLEMESSY.

3.1 *A6-CBF: Core and instances*

A6-CBF its developed in an innovative iterative incremental life cycle. It is built around a "catalogue" of components with a reference architecture implementing a suite of main functions. This catalogue of components is called the Core. This Core has to be instantiated for a particular production site bench implementing components that perform the functions required in this site. Four versions of A6-CBF Core are planned, with the first bench (based on CORE 1) to be deployed operationally in the ArianeGroup Le Mureaux site in 2018.

3.2 *Selected MBSE tools*

3.2.1 *Enterprise Architect (EA)*

The A6-CBF engineering process is carried out with a commercial Model Based System Engineering tool, Enterprise Architect (EA). An Engineering Model is shared between Design and RAMS teams. This Engineering Model uses a SysML language subset to model system design and stores all data and elements needed to perform RAMS analyses.

3.2.2 *BlockSim/XFMEA by Reliasoft*

Two tools specialized in reliability engineering industry, BlockSim and Xfmea from Reliasoft, are selected to perform specific RAMS calculations. RAMS related data (FTA and reliability data) is exported from the EA model via XML format. BlockSim is used in the project to perform numerical and minimal cut sets computations. By using Fault Trees and Reliability Blocks Diagrams it is possible to validate the reliability of the system and the operational availability requirements.The size of the project implies a large amount of data that needs to be processed. Hence, Xfmea is used in the project as a tool to facilitate data management and reporting during the development of FMECAs.

3.2.3 *Enterprise architect scripting*

Since the design of the system is conceived in an iterative process, the design is expected to be changing continuously and information and elements needed for RAMS analysis will be updated continuously. Regarding the fact that RAMS analysis are performed outside the engineering model from EA and they are performed at same time the system is being developed, there is a need to constantly update the ReliaSoft database with the most recent information. In order to do so, scripts are developed to import information like reliability data, Fault Tree topologies or system architecture to ReliaSoft from EA and keep them up to date whenever it is needed.

3.3 *A6-CBF RAMS modelling objects*

An overview of RAMS modelling objects used is presented on Figure 4:

1. Client Requirement [YELLOW]: Requirement coming from any applicable documents.
2. CBF Requirements [GREEN]: Requirements declined by project team.
3. Feared Event [RED]: Technical Risk related to one CBF requirement that shall be analyzed by Fault Tree
4. Fatal Alarm [RED]: Alarm generated by Monitor Control that will put the bench in a predefined state. A FearedEvent Fatal_Alarm is used to show which conditions are required to activate this alarm.
5. Trigger [BLUE]: Condition that activates a StateMachine transition. (BENCH_Operational > BENCH_Failed trigger by Fatal Alarm)
6. Function (CORE or Component) [ROUND]: function defined in CBF Functional Analysis tree.
7. Functional Failure Mode [LIGHT ORANGE]: describing on the five generics failure mode of function
8. Component Level 1 [PINK]: CBF component to be developed & validated autonomously.
9. Component (HW or SW) [PINK]: decomposition of ComponentL1
10. COTS (HW or SW) [part of an ILS Family] [OLIVE]: Commercial off-the-shelf component
11. Component Failure Mode [DARK ORANGE: describing failure mode of component
12. MCS Diagnostic [LIGHT YELLOW]: permitting the diagnostic of a Component/Functional Failure Mode.
13. Open Issue [linked to IBTReq]: OpenPoint to discuss.
14. Assumption: [within OpenIssue] linked to open issues

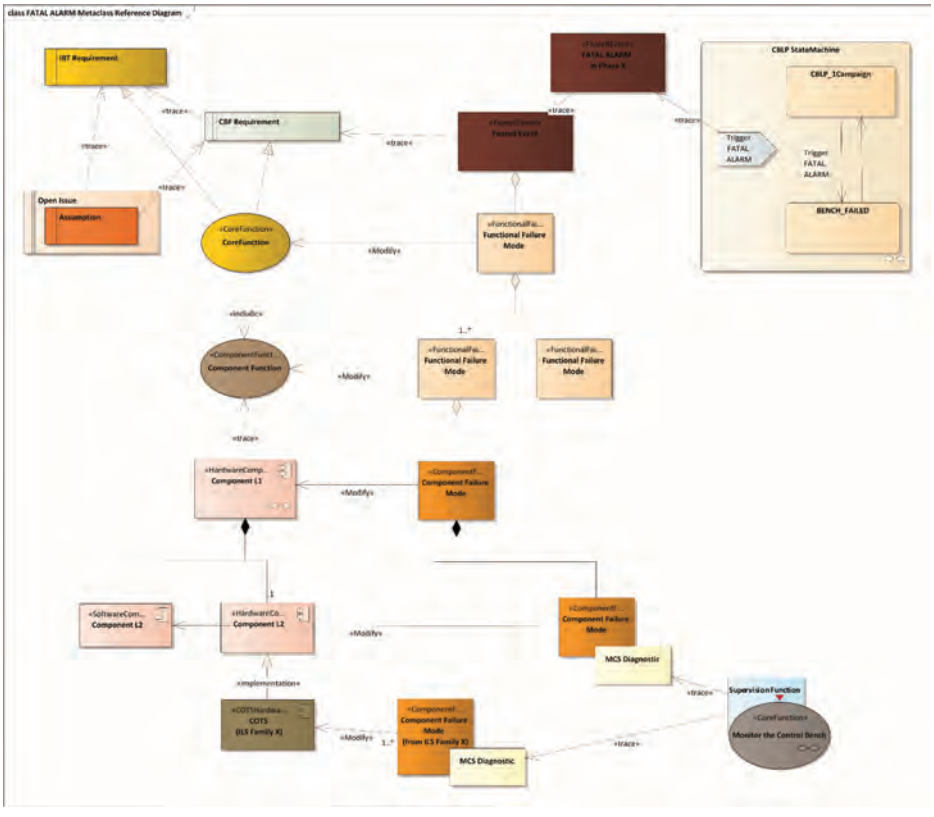


Figure 4. RAMS modelling objects in EA.

3.4 *A6-CBF functional analysis*

In the A6-CBF functional decomposition has been initially performed for Core Functions. Through the functional definition it is possible to distinguish two kind of functions: Core Functions (Core-F) and Component Functions (Comp-F). Additionally, some Instance Functions (Inst-F) may have been used to model specific functionalities not covered by the Core. The criteria for the use of Core-F or Comp-F Functions is based on the participation of one or more components in the function performance. In this way, for those functions where it is expected that only a single component will participate, Comp-F are used, whilst for the rest CORE-F are used. Comp-F will be located at the lowest levels of the functional tree and CORE-F, resulting from the integration of various Comp-F, at higher levels of the tree. A total of 21 main functions (PF) have been defined (eg: acquisition, commanding, data archiving, ...). They have been decomposed following the methodology exposed in section 2.1.

This results in more than 250 Core-F, more than 1200 Comp-F and about 10000 related requirements. Complementary, transversals Use Cases has been used to model system temporal behavior involving functions not directly linked on the functional tree.

3.5 *A6-CBF feared event identification*

Around 50 Feared Events have been identified. Feared Events are classified in five severity levels.

3.6 *A6-CBF Functional AMDEC*

Export of System hierarchy, FA, F-FM, C-FM, an diagnostics from EM model allows to interact with FMECA via a classical template, allowing a navigation around information in a practical way, easy to share with other non specialized stakeholders.

3.7 *Traceability between functional analysis and system components*

Around 70 different Components L1 (to be developed and validated unitary) are currently defined.

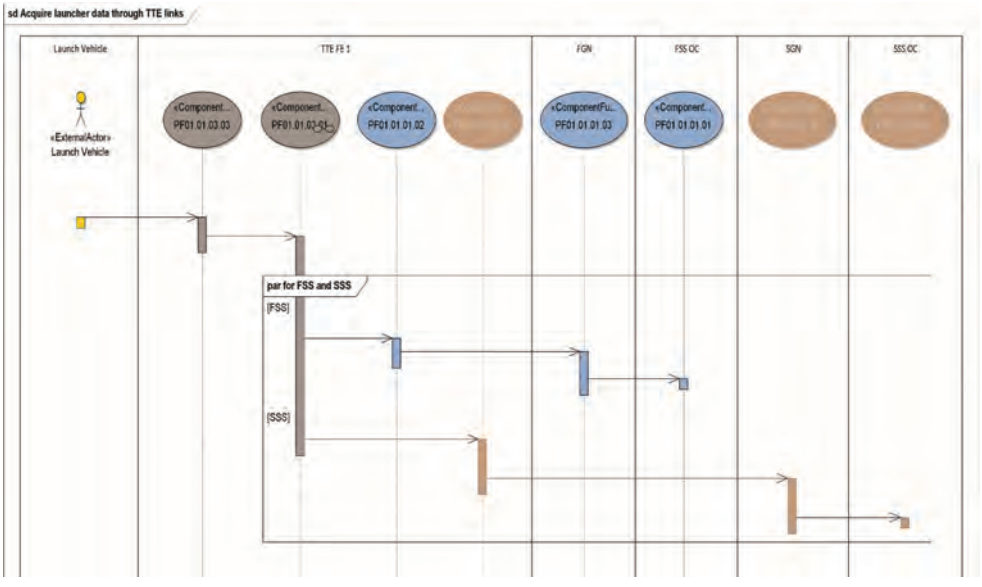


Figure 5. Example of CBF transversal function.

Recursive decomposition has been produced hundreds of component L2 (SW & HW). Each component implement a small subset of 1200 component functions, allowing Criticality Analysis.

3.8 *Identification of component failure modes*

As of today around 40 HW families have been identified. The Failure Mode/Mechanism Distributions database (FMD) by Quanterion is used to select appropriate failure modes for each HWFamily.

3.9 *Link between component failure modes and functional failure modes*

A stereotype object of type "Diagnostic" has been used to identify detection mechanisms associated to a C-FM or F-FM. In this way, it is easy to identify the Failure Modes that are not detectable. The diagnoses are identified with the nomenclature used by the supervision component (log code errors).

The general behavior of the A6 Control Bench defined by general state machine defined by Supervision component. States (the transitions of those states are Trigger type stereotypes, associated with feared Events. The dynamic effects of system behavior can also be modeled with activity diagrams or state machines, allowing triggers. In this way it is possible to know in what conditions a certain failure mode will produce the transition from the state bank Fail to the state (FatalAlarm).

Figure 6. A6-CBF feared events severity.

Figure 7. Components Vs Function traces.

These operating ways are used for Supervision components as well.

3.10 *A6-CBF RAMS results*

3.10.1 *Component criticality analysis*

A6-CBF Components L1 has been classified regarding criticality levels. Three components have been identified as the most critical regarding safety Feared Events. For each of this L1 components, a refinement work is perform in order to identify and isolate subcomponents with the high criticality.

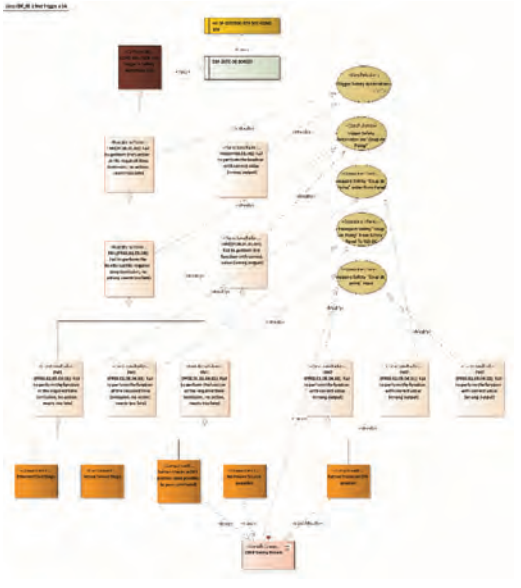


Figure 8. Feared event modelling diagram in EA.

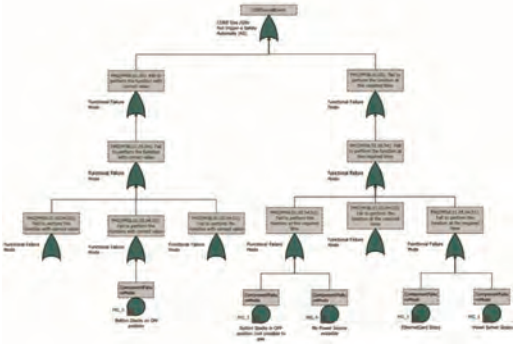


Figure 9. Feared event fault tree in BlockSim.

3.10.2 *Quantitative and qualitative RAMS requirements verification*

In this project phase, export and computation of FearedEvents Fault Tree from EA to BlockmSim allows to produce preliminary results, taking into account the current design definition. As design evolves, recursive iterations are performed to assure design complaint against RAMS requirements.

3.10.3 *Changes requested safety assessment*

A6-CBF shall operate for at least 30 years, so a specific Integrated Logistic Support program has been established to ensure operational capabilities during this period. Most of the components of the system have a useful life of less than 30 years which is why it is essential to manage the obsolescence of the material.

A component taxonomy has been created (HW Families) which allows to identify ILS Recommendations and defining RAMS requirements for each product. Each HWFamily can be implemented by different pre-qualified Part Numbers. The required reliability characteristics are defined at the HWFamiliy level (MTBFmin, MTTRmax, etc.). Failure mode distributions are also defined at this level based on the data available in the different reliability databases (NTNU, 2017). Even in this initial phase, during the three years between the first and the last Control Bench delivery, it is highly possible that some PartNumbers evolve due to obsolescence issues generating specific configurations for each bank.

Having defined the minimum RAMS requirements at the HWFamiliy level ensures that the system will continue to comply with requirements regardless of the PartNumber used.

4 OUTLOOK

This methodology can easily be extended to include cyber risk and human factors. It is possible to model realistic reliability and operational availability taking into account functions required for each of the operational phases of the bench (e.g. launch campaign, launch chronology, ignition sequence, post-launch revalidation phase).

4.1 *Cyber risk modelling*

In order to include cybersecurity analyses, functional analysis must be carried out until reaching a decomposition level with all required information transport functions (data flows). Data flows routing (components, ports and physical links involved) would have to be defined.

For each of these flows and any component where data is treated/storage following failure modes would have to be created:

– loss of data integrity,
– loss of data confidentiality,
– unavailability of data

Following the method presented in 2.4, these dataflow/component failure modes will induce functional failure modes which in turn could be part of an existing safety feared event Fault Tree (FE: To send an unexpected command). If specific feared events regarding cyber risks could be defined, (e.g. theft of information, deterioration of public image) every HW/SW component could be classified according to criticality of cyber risks induced. Cyber risk requirements could be allocated to each component according to this criticality.

4.2 *Human error factors modelling*

Following the same logic used previously for each interface between the system and a human-type actor Human Error Failure Modes (HE-FM) could be created. Different HE-FM taxonomies are used in the literature (e.g. THERP, SHERPA, HEART, see Pocock, S., Harrison, M. D., Wright, P. C., & Johnson, P. 2001). These HE-FM, can be linked to the F-FM of functions performed by components in interface with humans (usually IHMs). Components could be then classified according to criticality of human errors. Human reliability requirements could be allocated for each component according to this criticality.

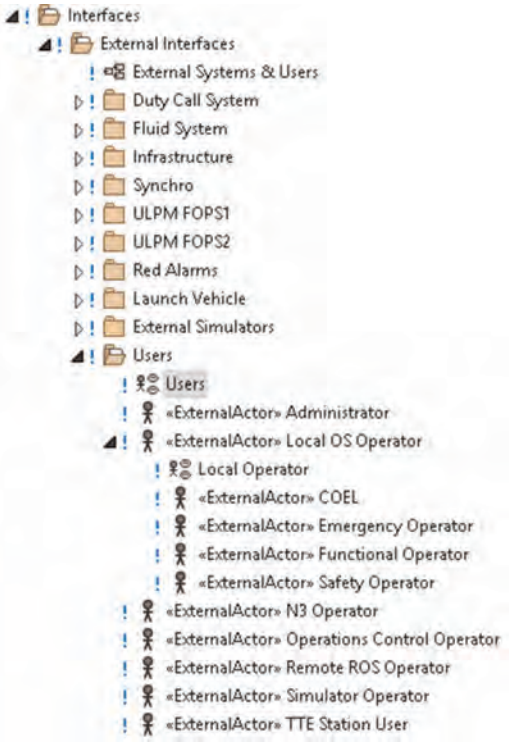


Figure 10. External actors, human type.

4.3 *Reliability and operational availability by operation phase*

The profile mission of each A6 Control Bench site is unique. In the case of Ariane 6 Control Bench Launch Pad (CBLP) it is used during 10 days and its exploitation is oriented to ensure safety during operations and availability at H0 (launch time). Production control benches are exploited to ensure continuous availability during the production checkout phase. The development of functional analysis by means of use cases enables the addition of "tag values" identifying which functions are required by phases.

5 CONCLUSIONS

A new methodology for direct integration of safety analysis in a MBSE process has been presented. System and safety engineering activities are carried out in a common Engineering Model. In this way functional FMECAs tasks can be started at a very early stage of the project identifying critical functions. Particularly for SW components, a detailed functional decomposition allows to identify and isolate high from low criticality components. Following design advancement (architecture & detailed design), FMECA's can be refined by performing criticality analysis of components/subcomponents. This kind of incremental approach reduces the required period for the RAMS analysis.

Coherence between RAMS and design configuration is always ensured since a single and centralized engineering model is the source of both types of information. Model scripting capabilities export the selected information to reliability tools generating according Fault Trees and FMECA in a more traditional format. This facilitates the exchange with other project actors. Likewise, quantitative and qualitative requirements can be computed numerically. The ability to manage permits, locks and warnings about design modifications on the model allows the RAMS engineers to have an accurate view of every design change. This guarantees a correct impact analysis.

Traditional system safety analysis is usually based on informal specifications and highly dependent on the skills of the RAMS analyst. Lastly, but not less important, another advantage provided by this methodology is the reduction of manual effort and related error-prone tasks. Again, this reduces cost and time and it increases the quality of the outcome.

REFERENCES

Betancourt, L, Birla, S, Gassino P. & Regnier. P, 2011 NUREG/IA 0254 Suitability of Fault Modes and Effects Analysis for Regulatory Assurance of Complex Logic in Digital Instrumentation and Control Systems, Nuclear Regulatory Commission.

Bieber, P., Blanquart, J.P., Durrieu, G., Lesens, D., Lucotte, J., Tardy, F., Turin, M., Seguin, C., and Conquet, E. 2008. Integration of formal fault analysis in assert: Case studies and lessons learnt. *In Proceedings of 4th European Congress Embedded Real Time Software*, ERTS 2008. Toulouse (France).

Cherqui, S. Comery, P. & Lesens, D. 2016. MBSSE used in the Ariane 6 launcher development Model-Based System and Software Engineering - Future directions ESA/ESTEC.

Ercilbengoa A.-E., Schoenig R., Hutinet T., 2010. Analyse dysfonctionnelle sous l'outil safety designer d'une boucle de pilotage du lanceur Ariane 5. *Proceedings of 17ème Colloque National de Fiabilité et maintenabilité – λμ 17*–La Rochelle, France, vol. 4C-7, pp. 1–11, 2010. [EVE 06] EVERDIJ M., BLOOM H.

European Cooperation for Space Standardization, 2017. *Software dependability and safety*. (ECSS-Q-HB-80–03 A).

Joshi, A., Heimdahl, M.P.E., Miller, S.P. & Whalen.P.W. 2005. *Model Based Safety Analysis*. Minneapolis: University of Minnesota.

Lesage, J. J., & Kruse, R. (2011). Qualitative and Quantitative Formal Model-Based Safety Analysis.

Norwegian University of Science and Technology (NTNU). 2017. Reliability Data Sources. Available at: https://www.ntnu.edu/ross/info/data. [Accessed 3 December 2017].

Pocock, S., Harrison, M. D., Wright, P. C., & Johnson, P. 2001. THEA: A Technique for Human Error Assessment Early in Design. *In Interact* (Vol. 1, pp. 247–254).

Rauzy, A. & Chaire B.F. 2014 Model-Based Safety Assessment: Rational and trends. Mecatronics (MECATRONICS), 2014 10th France-Japan/8th Europe-Asia Congress on. IEEE, 2014.

Reifer, D.J. (1979), Software failure modes and effects analysis" *IEEE Transactions on Reliability*, Volume R-28, #3, pp. 247–249.

RMQSI Knowledge Center 2016. Failure Mode/Mechanism Distributions (FMD-2016). Utica: Quanterion Solutions Incorporated.

Wagner W. & P.H.A.J.M Van Gelder, Applying RAMSSHEEP analysis for risk-driven maintenance. In: Steenbergen et al. (eds), Safety, Reliability and Risk Analysis: Beyond the Horizon. London: Taylor & Francis Group, 2014, pp. 703–713.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Masked data analysis for storage reliability model with initial failures

M. Zhao
Faculty of Engineering and Sustainable Development, University of Gävle, Gävle, Sweden

Y.J. Zhang
School of Mathematics and Physics, Anhui University of Technology, Maanshan, China

J.F. Yang
Faculty of Information Engineering, Guizhou Institute of Technology, Guiyang, China

ABSTRACT: Storage reliability is of importance for the products that largely stay in storage in their total life-cycle such as warning systems for harmful radiation detection, rescue systems, many kinds of defense systems, etc. The storage reliability of a product is commonly defined as the probability that the product can perform its specific function for a period of specific storage time under specific storage environment. Logically, the failures of the product in storage should be identified with the same criteria as in its operation process. However, the failure data in storage may be observed indirectly through the maintenance or inspection activities. Nevertheless, when the storage reliability is concerned in general, the reliability model should take into consideration the possibility that the operational reliability does not start at 100%, for example, the one-shot product may have only 96% operational reliability when they are newly produced. In this paper, the storage reliability model with possibly initial failures, which are usually neglected at the beginning of storage in most of storage models, is studied on the statistical analysis method when the masked data are observed. The parametric estimation procedure, based on the Least Squares method, is developed generally by applying an EM-like (Expectation and Maximization) algorithm for the storage data in which some information about which components have caused the system failures is not known, namely the failure data are masked. The estimates of the model parameters including the initial reliability are formalized. In the case of exponentially distributed storage lifetime and series system, a numerical example is provided to illustrate the method and procedure though the method is not limited to such case. The results should be useful for planning a storage environment, decision-making concerning the maximum length of storage, maintenance strategy optimization and identifying the production quality.

1 INTRODUCTION

Storage reliability is generally referred to the ability of a product that can still be able to perform its required functions after it has been in the storage state for a certain period of time. There are many examples of such products or systems whose life-cycle is mostly dominated by their storage time. The failure mechanisms of the product in storage and operation are completely different, and this implies that the storage reliability model should be considered by taking the operation reliability into account. There have been some storage models proposed under different assumptions [1–6].

In the literatures, the storage reliability models are commonly to assume that the products are perfect at the beginning of the storage, but as a result of the slow deterioration process, the operation reliability will degrade as the their storage time goes [5–6]. However, the operational reliability at the beginning of storage is not always 100% for many kinds of one-shot products [3]. Specifically, some one-shot products cannot be evaluated about if they are functioning or not, instead, they are only assessed to be good to fulfill the requirements on their performance measures under the periodical maintenance [7–14]. Note that a one-shot product is referred to those devices or equipment that can be used for only one time. After the use, they are destroyed or should be extensively rebuilt. Some examples are missiles, fire extinguishers, airbags in cars, etc.

To study the storage reliability with initial failures, a generalized storage reliability model was proposed by Zhao & Xie [15–16]. The model is expressed as

$$R(t) = R_0 R_s(t) \tag{1}$$

where R_0 can be interpreted as the initial reliability of the product and may not be completely known for the products in storage. $R_s(t)$ is called the inherent storage reliability [15] and entirely reflects the effect of the storage on the products. Note that the storage reliability can also be studied based on the knowledge of Physics of Failure (PoF), and various PoF based models can be found in [17–23].

Recently, the estimation methods for the model expressed by (1) have been considered by Zhang *et al.* [5–6]. In the study presented in [5], an integrated approach is proposed to estimate the storage reliability based on the combinational estimates of the failure numbers and current reliabilities at each testing times when groups of binomial-type failure data are available. The E-Bayesian estimation of the failure probabilities is further applied by Zhang *et al.* [6] into the integrated approach proposed in [5] for the same types of the testing data.

In general, the failure data in storage may be observed indirectly through the maintenance or inspection activities and can also be obtained through the measures of the product performance particularly based on their components. Furthermore, the field testing data can be available, but the failure causes cannot be always known due to various reasons such as high cost, difficulty to diagnosis, lack of enough information, etc. This is the case where the system lifetime data is masked for the causes of some failures are unknown or the components resulting in the system failures cannot be identified [24–26]. The masked data analysis has been considered by several authors. Miyakawa [24] proposed both parametric and nonparametric estimators in the case of a simple two-component series system by assuming that the lifetimes of components are exponentially distributed. For a general series system, the expression of the likelihood function was derived by Usher & Hodgson [25–26]. The Maximum Likelihood Estimation (MLE) was considered in the cases of two and three-component series systems and it was shown that the closed-form maximum likelihood estimators are algebraically intractable. The masked data analysis was also studied by Hansen & Thyregod [27] for a superimposed renewal process. The underlying lifetime distribution of the components are assumed to be identically mixed-exponential and each failure is masked with the same probability p for all components. The parameter p is unknown and has to be estimated from the field data. In software reliability, the MLE estimation of software reliability were also considered for superimposed nonhomogeneous Poisson processes in the case of the masked data [27–28]. To the author's knowledge, however, there has not been the study on the masked data analysis under the storage model with possible initial failures described in (1).

In this paper, the storage reliability model with the possible initial failures is studied for the masked failure data of systems. The masked data are the groups of binomial-type system failure data that may not be identified on which components caused the system failures. A general method of the parametric estimates is developed by applying a modified EM (Expectation and Maximization) algorithm for parametric ML estimation. In the case of exponentially distributed storage lifetime and series system, the method and procedure are formulized in detail. A numerical example is also provided to illustrate the method and procedure though the method is not limited to such case. The results should be useful for planning a storage environment, decision-making concerning the maximum length of storage, maintenance strategy optimization and identifying the production quality.

2 MODEL DESCRIPTION AND DATA STRUCTURE

2.1 *Model formulation*

Following the common definition, the storage reliability of products in this paper is defined as the probability that the product can perform its required function for a period of specific storage time under the defined storage environment [15]. Note that the definition given here implies that the product should be functioning in operation if it is said to be good. Consequently, it may not be possible for one-shot products to be identified as good or not by taking only the maintenance inspection.

Let T be the storage lifetime of a product, although its value can be hard to observe in reality for some types of products, it is a genuine random variable and often known to be greater, smaller or between some storage time points [5–6]. The storage reliability function given in (1) can be derived as

$$\begin{aligned} &R(t) = P(T > t) = \\ &P(T > t | \textit{product is good before storage}) * \\ &P(\textit{product is good before storage}) = R_0 R_s(t) \end{aligned} \tag{2}$$

Note that "*product is good before storage*" is referred to the random event that the product can be functioning when it is new without of storage.

The meaning of R_0 can simply be the successful probability for one-shot products to be applied. For electronic components, it will represent the proportion of non-defectives in the population, but mostly, it will be 100% for common products.

For the inherent storage reliability function $R_s(t)$, some common lifetime distributions, such as exponential, Weibull or lognormal distributions, can be applied to model the failure process of the products due to the material deterioration in storage. To simplify the method presented in this paper, the exponential distribution is applied for the inherent storage reliability although the method can be valid for a general lifetime distribution. The model expressed in (1) is therefore rewritten as

$$R(t) = R_0 e^{-\lambda t} \tag{3}$$

For a series system with k components, the storage reliability of the system, for the sake of simplicity, is given as

$$R(t) = R_1(t)\,R_2(t)\ldots R_k(t) = R_{01}R_{02}\ldots R_{0k}e^{-\lambda t} \tag{4}$$

where $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_k, \lambda_i$ is the failure rate of ith component, $i = 1,2,\cdots,k$. Note that the system reliability given by (4) assumes that all components have the same type of lifetimes that are exponentially distributed with different failure rates.

2.2 *Masked failure data*

Suppose that the systems in storage have the binomial-type failure data at sequential observation times $t_1 < t_2 < \cdots < t_m$. The masked data can have the form as shown in the following table.

In Table 1, n_j and n_{ij} are the numbers of the system and ith component tested at observation time t_j, f_j and f_{ij} are the failure numbers of system and ith components, respectively, $i = 1,2,\ \ ,k;\ j = 1,2,\ \ ,m$.

Note that if there is not masked data, the storage reliability of each component at observation times can be simply estimated as

Table 1. Binominal-type masked failure data of a series system with k components.

Causes of failures	Observation times		⋮	
	t_1	t_2		t_m
*System	(n_1, f_1)	(n_2, f_2)		(n_m, f_m)
Component 1	(n_{11}, f_{11})	(n_{12}, f_{12})		(n_{1m}, f_{1m})
Component 2	(n_{21}, f_{21})	(n_{22}, f_{22})		(n_{2m}, f_{2m})
⋮				
Component k	(n_{k1}, f_{k1})	(n_{k2}, f_{k2})		(n_{km}, f_{km})

*The system failures are not identified on which component's failures caused its failures and then called to be masked.

$$\hat{R}_i(t_j) = \frac{n_{ij} - f_{ij}}{n_{ij}},\ i = 1,2,\ldots,k;\ j = 1,2,\ldots,m \tag{5}$$

It is also easy to write the likelihood function of the parameters of initial reliability and failure rate for each component, see e.g. [5–6] for detail. The parametric estimation of the component reliability can separately be made by ordinary methods. However, when the masked data are present, the likelihood function of all parameters of the components have to be considered and becomes a complex multivariable function with a very high dimension. It is therefore difficult to find out the MLE (Maximum Likelihood Estimate) of these parameters.

2.3 *Least squares estimation with non-masked data*

When the exponential distributions are applied for the inherent storage lifetimes of components, it is easier to obtain the initial reliability and failure rate of each component if the data are not masked. The MLE of these parameters can be obtained by solving the ML equations numerically, see [5–6] for detail. Nevertheless, the Least Squares (LS) estimates are more convenient in the case of exponential distributions since the analytic formulas can be written.

According to the reliability function of each component, the following linear equations hold:

$$Y_{ij} = a_i + b_i t_j, \tag{6}$$

where

$$ln\left(R_i(t_j)\right) = ln(R_{0i}) - \lambda_i t_j,$$
$$Y_{ij} = \ln\left(R_i(t_j)\right); a_i = \ln(R_{0i}); b_i = -\lambda_i;$$
$$i = 1,2,\ldots,k;\ j = 1,2,\ldots,m.$$

For a fix i, by using the estimates given in formula (5) to replace $R_i(t_j), j = 1,2,\ \ ,m$, it can easily be seen that the Least Squares estimates of a_i and b_i, therefore R_{0i} and λ_i, can be calculated as

$$\hat{R}_{0i} = \exp(\bar{R}_{iln} + \hat{\lambda}_i \bar{t}), \tag{7}$$

$$\hat{\lambda}_i = -\frac{\sum_{j=1}^{m}\left[\ln\left(\hat{R}_i(t_j)\right) - \bar{R}_{iln}\right]\left[t_j - \bar{t}\right]}{\sum_{j=1}^{m}\left(t_j - \bar{t}\right)^2}, \tag{8}$$

$$\bar{R}_{iln} = \frac{1}{m}\sum_{j=1}^{m} ln\left(\hat{R}_i(t_j)\right), \bar{t} = \sum_{j=1}^{m} t_j, \tag{9}$$

$i = 1, 2, \ldots, k$.

Note that the LS estimates of the initial reliability and failure rate for each component cannot be obtained from formulas (7), (8) and (9) if the masked data as displayed in Table 1 are applied. The ML estimates are also extremely difficult to obtain since the ML equations will contain too many parameters.

3 PARAMETRIC ESTIMATION USING MASKED DATA

3.1 *EM algorithm and its modification*

The EM (Expectation-Maximization) algorithm has recently been applied to solve the ML parametric estimation problem in software reliability when the masked data are presented, see e.g. [27–28]. The properties of this algorithm can also be seen in [30–33].

The general idea of the EM algorithm in the context of the masked data is to give the initial values of the parameters to be estimated, for example, θ^0, and calculate the expectations of the missing data. The ML estimates θ^1 can be obtained as if the complete data are available, and then repeat the procedure taking θ^1 as the new initial values until the stable values are obtained. The applicability of the EM algorithm should be that the ML estimates can easily be obtained. Unfortunately, the ML estimates of the component parameters in the storage reliability model considered here do not have analytic forms. This implies that the application of the EM algorithm is also complicated.

Note that the LS estimates of the component parameters have close analytic form as given in formulas (7), (8) and (9). A modified EM algorithm, called ELS algorithm, is therefore proposed and generally presented as follows:

E-step: Give the initial values of the parameters θ^0, find out the expectations of the missing data;

LS-step: Using the estimated missing data and unmasked data as if they are complete, find out the LS estimates of the parameters θ^1 that will be used as the new initial values of the E-step for the iterations until the convergence is reached.

To the author's knowledge, there is no similar study on the ELS algorithm found in the literature. In principle, the ELS algorithm can simply be applied for any model if the LS estimates can easily be obtained, for examples, in the case of Weibull, exponential or log-normal distributions of the storage lifetimes. In the following, the ELS algorithm is presented in detail in the case of exponential distributions of component storage lifetimes.

3.2 *ELS algorithm for binominal-type data*

For the masked data type shown in Table 1, the masked data are the numbers of system failures that are not identified as which components are the causes of these failures. To realize the ELS algorithm, one needs to find out what are the expected number of failures for each component at each observation time point when the system has f failures. For simplicity, the following discussion is given without considering the variable time.

Let n be the number of tested systems and f the failure number. The components have their storage reliability as $R_i, i = 1,2,\ \ ,k,$ then the system has the reliability $R = R_1 * R_2 * \cdots * R_k$. If it is assumed that these system failures are not identified due to which component, it can simply be proved that conditional on (n, f), the expected number of failures for ith component, f_i^E, is given by

$$f_i^E = \frac{1-R_i}{1-\mathrm{R}} f, i = 1,2,\ldots,k. \tag{10}$$

Note that the sum of all f_i^E is generally larger than the failure number f of the system since one system failure can be caused by more components. Concerning on the ELS algorithm, the formula (10) will be used to convert the system binomial data (n, f) into component binomial data $\left(n, f_i^E\right), i = 1,2,\ \ ,k.$

The ELS algorithm can now be specified for the masked data given in Table 1 on the storage reliability model (4):

E-step: Give the initial values of the component initial reliabilities $\left(R_{01}^0, R_{02}^0\ \ R_{0k}^0\right)$ and failure rates $\left(\lambda_1^0, \lambda_2^0\ \ \lambda_k^0\right)$, calculate the followings:

- Component and system reliabilities at observation times $t_1 < t_2 < \cdots < t_m$:

$$R_{ij} = R_{0\mathrm{i}}^0 e^{-\lambda_i^0 t_j},\ R_j = R_{1j} R_{2j} \cdots R_{kj};$$
$$i = 1,2,\ldots,k;\ j = 1,2,\ldots,m.$$

- For the masked system data (n_j, f_j), calculate the expected number of failures for each component using formula (10):

$$f_{ij}^E = \frac{1-R_{ij}}{1-R_j} f_j, i = 1,2,\ldots,k,\ j = 1,2,\ldots,m.$$

- Update the component binomial data in Table 1 as

$$n_{ij}^* = n_{ij}, +n_j, f_{ij}^* = f_{ij} + f_{ij}^E,$$
$$i = 1,2,\ \ ,k,\ j = 1,2,\ \ ,m. \tag{11}$$

LS-step: Use the estimated component binomial data calculated by formula (11), also apply formulas (5), (7), (8) and (9), the LS estimates of the initial reliabilities and failure rates of the components are obtained as

$$\hat{R}_i^*\left(t_j\right)=\frac{n_{ij}^*-f_{ij}^*}{n_{ij}^*},\ i=1,2,\ldots,k;\ j=1,2,\ldots,m \tag{12}$$

$$\hat{R}_{0i}^1=\exp(\bar{R}_{iln}^*+\hat{\lambda}_i\bar{t}), \tag{13}$$

$$\hat{\lambda}_i^1=-\frac{\sum_{j=1}^{m}\left[\ln\left(\hat{R}_i^*\left(t_j\right)\right)-\bar{R}_{iln}^*\right]\left[t_j-\bar{t}\right]}{\sum_{j=1}^{m}\left(t_j-\bar{t}\right)^2}, \tag{14}$$

$$\bar{R}_{iln}^*=\frac{1}{m}\sum_{j=1}^{m}\ln\left(\hat{R}_i^*\left(t_j\right)\right),\bar{t}=\sum_{j=1}^{m}t_j, \tag{15}$$

$i = 1, 2, \ldots, k$.

When the LS-step is finished, the E-step will be repeated, but the initial values $\left(R_{01}^0,R_{02}^0\quad R_{0k}^0\right)$ and $\left(\lambda_1^0,\lambda_2^0\quad\lambda_k^0\right)$ will be replaced by the LS estimates $\left(R_{01}^1,R_{02}^1\quad R_{0k}^1\right)$ and $\left(\lambda_1^1,\lambda_2^1\quad\lambda_k^1\right)$. The iteration of the E-step and LS-step can be terminated when the stable LS estimates are received.

4 NUMERICAL EXAMPLE

To illustrate the ELS algorithm to estimate the storage reliability, a series system of 2 components is considered. The binominal-type masked data created manually for the purpose of the illustration, is listed in Table 2:

Note that there are four parameters to be estimated in this example: two initial reliability R_{01}, R_{02}, and two failure rates λ_1, λ_2 of component 1 and component 2, respectively. By applying the ELS algorithm, the estimates of these parameters can be easily obtained numerically.

In Figure 1 and Figure 2, the original observations of reliability estimated by the successful ratio, the estimated reliability by the least squares method (without system data) and the ELS algorithm (with system data), respectively, are displayed for Component 1 and 2. The differences are obviously visible, but reasonable since the system failure data should be used when the component reliability is analyzed.

Table 2. Binominal-type masked failure data of a series system with 2 components.

Observation times 1000 h	Causes of failures		
	System	Component 1	Component 2
5	(40,3)	(55,3)	(45,2)
10	(40,4)	(50,4)	(45,3)
15	(20,2)	(60,5)	(55,6)
20	(20,3)	(70,7)	(65,5)
25	(20,3)	(60,7)	(55,10)
30	(15,2)	(55,9)	(50,11)
35	(12,2)	(55,12)	(50,10)
40	(15,3)	(60,14)	(60,12)
45	(20,4)	(58,15)	(53,14)
50	(15,3)	(50,13)	(45,13)

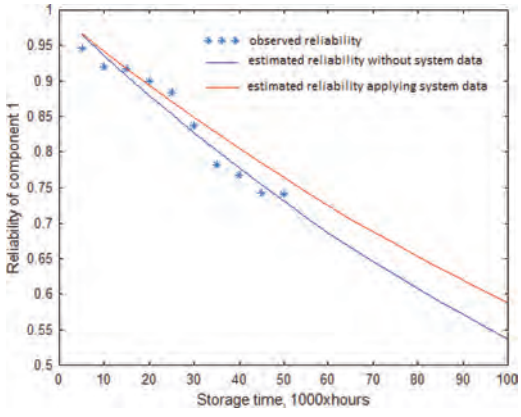


Figure 1. The observed and estimated reliability functions of Component 1.

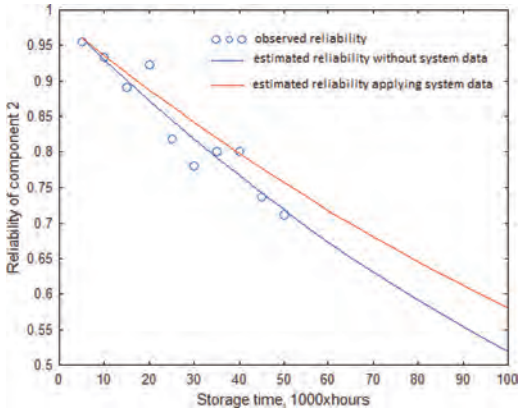


Figure 2. The observed and estimated reliability functions of Component 2.

The blue curves in Figures 1 and 2 represent the estimated storage reliability functions when the model parameters are estimated by the least squares method without applying the system data. The red curves represents the estimated storage reliability functions by the ELS method.

In Table 3, the compared results are also given to see the differences between the estimated parameters in the models. The estimates of the failure rates of the components with and without using the masked system data differ to each other. This indicates that it is important and necessary to apply the masked system data for component parameters.

Table 3. Comparison results of parametric estimates.

Methods	Component 1		Component 2	
	R_{01}	λ_1	R_{02}	λ_2
Estimates without system data	0.9949	0.0062	0.9927	0.0065
Estimates using system data	0.9920	0.0052	0.9861	0.0053

5 CONCLUSION

In this paper, the masked data analysis is considered for the storage reliability model with initial failures. To solve the problem that the estimation of the component reliability becomes difficult when the cause of system failures are hidden, an EM-like method (algorithm), the so-called ELS method is proposed for the series system with binomial-type failure data although the ELS algorithm can also be easily applied to other system configuration. A numerical example is provided to illustrate the method. By the illustrated example, it can be seen that the parameter estimations are quite different by the methods without and with the system failure data when the component reliability is estimated. The proposed ELS method has greatly simplified the parametric estimation in the case of the masked data.

ACKNOWLEDGEMENT

This work is supported by Centre of Logistics and Innovative Production[1]; MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Project No. 17YJA630135)[2], Science and Technology Foundation of Guizhou (No.: QianKeHeJZi[2015]2064)[3].

REFERENCES

[1] Mense AT, Gullo L, Thomas J, Shedlock P. Models and methods for determining storage reliability. In: Proceedings of the IEEE international reliability physics symposium (IRPS). Anaheim (CA), 2013, ER.3.1–ER.3.6.

[2] Merren GT. Dormant storage reliability assessments-data based. IEEE Transactions on Components, Hybrids, and Manufacturing Technology, 1981, 4(4): 446–454.

[3] Mirzahosseinian H, Piplani R. A study of repairable parts inventory system operating under performance-based contract. European Journal of Operational Research, 2011, 214(2): 256–261.

[4] Verdingovas V, Denmark L, Jellesen MS, Ambat R. Impact of NaCl contamination and climatic conditions on the reliability of printed circuit board assemblies. IEEE Transactions on Device and Materials Reliability, 2014, 14(1): 42–51.

[5] Zhang YJ, Sun YC, Zhao M. A combinatorial estimation approach for storage reliability with initial failures based on periodic testing data. Communications in Statistics-Simulation and Computation, 2017, 46(4): 3319–3340.

[6] Zhang YJ, Zhao M, Zhang ST, Wang, JM, Zhang YH. An integrated approach to estimate storage reliability with initial failures based on E-Bayesian estimates. Reliability Engineering and System Safety, 2017, 159: 24–36.

[7] Cui LR, Zhao X, Shen JB, Xu Y. An availability model for storage products under periodical inspections. International Journal of Reliability, Quality and Safety Engineering, 2010, 17(02): 89–103.

[8] Su C, Zhang YJ, Cao BX. Forecast model for real time reliability of storage system based on periodic inspection and maintenance data. Eksploatacja i Niezawodność, 2012, 14: 342–348.

[9] Kim HW, Yun WY. Reliability analysis for one-shot systems with periodic inspection. Journal of Korean Institute of Industrial Engineers, 2016, 42(1): 20–29.

[10] Martinez EC. Storage reliability with periodic test. Reliability and Maintainability Symposium, 1984. Proceedings. Annual. IEEE, 1984: 181–185.

[11] Ito K, Nakagawa T. Optimal inspection policies for a storage system with degradation at periodic tests. Mathematical and Computer Modelling, 2000, 31(10–12): 191–195.

[12] Nakagawa T, Mizutani S, Chen M. A summary of periodic and random inspection policies. Reliability Engineering & System Safety, 2010, 95(8): 906–911.

[13] Kitagawa T, Yuge T, Yanagi S. Periodic and non-periodic inspection policies for a one-shot system with minimal repair. Journal of Japan Industrial Management Association, 2016, 66(4): 387–95.

[14] Zhao M, Xie M. A model of storage reliability with possible initial failures. Reliability Engineering & System Safety, 1994, 43(3): 269–273.

[15] Zhao M, Xie M, Zhang YT. A study of a storage reliability estimation problem. Quality and reliability engineering international, 1995, 11(2): 123–127.

[16] Li Y, Agyakwa PA, Johnson CM. Physics-of-failure lifetime prediction models for wire bond interconnects in power electronic modules. IEEE Transactions Device Mater Reliability 2013, 13(1):9–17.

[17] Wise LJ, Schrimpf RD, Parks HG, Galloway KF. A generalized model for the lifetime of microelectronic components, applied to storage conditions. Microelectronics Reliability, 2001, 41:317–22.

[18] Cao R, Chen Y, Kang R. Storage reliability evaluation of engine control circuit module. In: Proceedings of the IEEE international conference on quality and reliability. Bangkok (Thailand), 2011: 473–476.

[19] Feng J, Sun Q, Jin TD. Storage life prediction for a high-performance capacitor using multi-phase Wiener degradation model. Communications in Statistics-Simulation and Computation, 2012, 41(8): 1317–1335.

[20] Liu ZY, Ma XB, Zhao Y. Storage reliability assessment for missile component with degradation failure mode in a temperature varying environment.

Acta Aeronautica et Astronautica Sinica, 2012, 33(9): 1671–1678.
[21] Pelzera R, Nelhiebelb M, Zinka R, Wöhlerta S, Lassnigc A, Khatibic G. High temperature storage reliability investigation of the Al–Cu wire bond interface. Microelectronics Reliability, 2012, 52(9–10):1966–1970.
[22] Nishad P, Diganta D, Estelle S, Michael P. Long term storage reliability of antifuse field programmable gate arrays. Microelectronics Reliability, 2013, 53(12):2052–6.
[23] Miyakawa M. Analysis of incomplete data in competing risk model. IEEE Transactions on Reliability, 1984, 33: 293–296.
[24] Usher JS, Guess, FM. An iterative approach for estimating component reliability from masked system life data. Quality and Reliability Engineering International, 1989, 5: 257–261.
[25] Usher, JS, Hodgson, TJ. Maximum likelihood analysis of component reliability using masked system life data. IEEE Transactions on Reliability, 1988, 37: 550–555.
[26] Hansen, CK, Thyregod, P. On the analysis of masked field failure data. Proc. European Safety and Reliability Conference, June 10–12, Copenhagen, 1992, Denmark, pp. 840–850.
[27] Zhao M, Xie M. EM algorithms for estimating software reliability based on masked data. Microelectronics Reliability, 1994, 34(6): 1027–1038.
[28] Yang JF, Zhao M. Maximum likelihood estimation for software reliability with masked failure data. In Chinese, Systems engineering and electronics, 2013, 35 (12): 2665–2669.
[29] Dempster, AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society - Series B, 1977, 39:1–38.
[30] Wu CFJ. On the convergence properties of the E-M algorithm. Annual of Statistics, 1983, 11: 95–103.
[31] Meng, XL, Rubin, DB. Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. Journal of the American Statistical Association, 1991, 86: 899–909.
[32] McLachlan GJ, Krishnan T. The EM algorithm and extensions. 2008, New York: Wiley.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Probability-based reliability and availability assessments for a lane at a signalised intersection

M. Maslak & K. Ostrowski
Cracow University of Technology, Cracow, Poland

ABSTRACT: An original computational approach allowing the reliability and the availability assessments of the lane located at an intersection with traffic signals is presented and discussed in detail. Operation of the lane of this type is described using a probabilistic model of an alternating renewal process. The occurrence of queues with vehicles waiting for the change of the lights, but only in the cases when the number of these vehicles is observed to be at least equal to the arbitrarily accepted number relating to the critical congestion level, is treated as the lane's failure. The proposed formal model is calibrated based on the empirical intensity function being a rate of occurrence of failures. The availability of the lane is interpreted as the probability that such lane will be serviceable at a predetermined time-point t, whereas its reliability—as the probability that it will be serviceable in the whole analysed time interval.

1 INTRODUCTION

A reliable assessment of lane's availability at a signalised intersection should be unambiguously linked to a detailed analysis of the queues which form randomly on the lane, whenever arriving vehicles do not encounter the green signal (Chodur, 2011; Tracz 2012). Of course, the formation of this type of queue is inevitable in a process of controlling the individual traffic flows, but the point is to ensure that the queue does not turn out to be unduly long (Ostrowski, 2014; Ostrowski, 2015). In the proposed calculation procedure, the authors determine a certain, arbitrarily set by the evaluator, number of queued vehicles on the analysed lane, subsequently designated as Q_{cr}, and treat it as a cut-off between a critical queue associated with a formal lack of the lane's availability for effective use, which the authors designate as a failure ("0" state), and a queue of an acceptable length, with a number of vehicles lower than Q_{cr}, which does not result in termination of the lane's serviceability (state "1"). In this way the lane performance is seen as a bi-polar relationship. In a given event, being equivalent to a single observation, it can be assigned to only one of two mutually exclusive assessments: serviceable ("0") or unserviceable ("1"). In such an approach, the process of use of the considered lane, viewed over a 24-hour cycle, can be modelled as a so-called alternative renewal process with alternating time-intervals of its serviceability and non-serviceability of random durations. In this model, each observation corresponds to a full cycle of signal changes of the lane's traffic signals. Usually, the cycle features a green-yellow-red sequence of a fixed length. Nevertheless, the constant length of the entire sequence does not mean the constant duration of each colour signal. In each sequence, these intervals can have random durations, which is typical of so-called accommodative signalling. Knowing the unified duration of the whole single sequence, it is easy to determine the number of observations, i.e. of complete sequences, specified per hour t. This number is referred to as n_t. Of course, in accordance with the above assumptions, it does not depend on the time at which the observations were carried out, thus $\forall_t n_t = n$. The authors subsequently relate with this value the numbers $n_t^{fail} = n_t^{(0)}$ and $n_t^{serviceable} = n_t^{(1)}$, where $n_t^{(0)} + n_t^{(1)} = n_t = n$, determined for each hour of the observations corresponding to the fixed number of signal change sequences. The first of these two values is the measure of the number of events when the lack of the lane's serviceability was noted in a given hour t while the second one—the measure of the number of observations evidencing the complete serviceability of such lane. Both values mentioned above are used here to create the empirical histograms—in the first case the one relating to a random number of the lane's failures and, in the second case, the other one associated with a random number of the lane's serviceability states. It can be noticed that the model proposed by us does not go into detail as for the structure of traffic on the lane, neither does it distinguish among types of vehicles nor their speed parameters. The authors are merely interested in the measured values of a random failure rate, specified in

each hour, and also in its empirical distribution, identified taking into account a whole 24-hour observation cycle.

2 EMPIRICAL FUNCTIONAL CHARACTERISTICS DETERMINING THE RELIABILITY OF THE LANE UNDER CONSIDERATION

2.1 *Histograms relating to the random numbers of the lane's failures and to the random number of the lane's serviceability states*

In order to enhance the clarity of the argument, the proposed procedure is presented here on a computational example containing an interpretation of the results of an experiment conducted by the authors over 100 observation days on a selected lane of one of Cracow's major signalised intersections. The signalling selected for the experiment had a sequence of exactly $n = 25$ full signal change cycles during each hour, which translates into $N = 24\,n$=600 such cycles observed each day. The results obtained by us on each observation day were used in histograms displaying a random number of $n_t^{(0)}$ the lane's failures and, separately, in histograms displaying a random number of the lane's serviceability states $n_t^{(1)}$, with a one-hour width of the class interval $(\Delta t = 1h)$ and with a half hour $t = 0$ *h* 30 *m*; 1 *h* 30 *m*; …; 23 *h* 30 *m* being a centre of each interval. The most probable daily histograms of a random number of the lane's failures are shown in detail in Figure. 1a while the most probable histograms of a random number of the lane's serviceability states in Figure. 1b. They were developed on the assuming that $Q_{cr} = 7$. It is easy to notice that at night, or specifically, between 20 *h* 30 *m* and 6 *h* 30 *m* of the next day, the traffic on the analysed lane was so small that in principle there were no failures. Such failures could happen sporadically, on separate days, but this was not reflected in the histogram averaging the results compiled throughout the duration of the experiment. On the other hand, in the morning hours, between 9 *h* 30 *m* and 10 *h* 30 *m*, the analysed lane was unavailable on each day of the study.

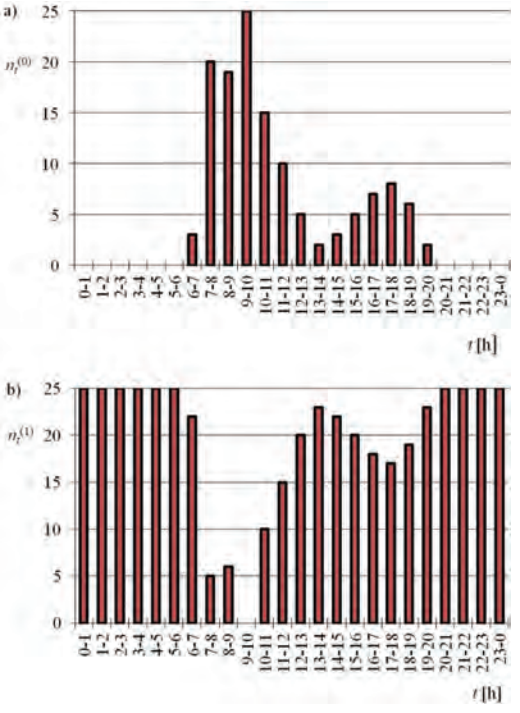


Figure 1. The most probable empirical histograms displaying on a day-by-day basis the random number of the lane's availability failures (Figure. 1a) and the random number of its serviceability states (Figure. 1b), obtained from the study of the analysed lane. The length of the critical queue was assumed at $Q_{cr} = 7$.

2.2 *Empirical functions of the lane's reliability and of the lane's unreliability*

The data in the histogram shown in detail in Figure. 1a allow assigning to the analysed lane the empirical function of its unreliability $\hat{S}(\tau)$ while the data given in Figure. 1b form the empirical function of the lane's reliability $\hat{R}(\tau)$. In this approach time τ means a selected value of variable *t*. The appropriate mathematical formulae look as follows (Migdalski, 1982):

$$\hat{S}(\tau) = \frac{\sum_{t=0h30m}^{\tau} n_t^{(0)}}{N} \tag{1}$$

$$\hat{R}(\tau) = 1 - \frac{\sum_{t=0h30m}^{\tau} n_t^{(0)}}{N} \tag{2}$$

The graphs of these functions, for the data considered in the example (and compiled in Tables 1 and 2), are presented in Figure. 2a and in Figure. 2b, respectively.

2.3 *Empirical probability density functions of the random occurrence of the lane's availability and failure states*

As the next step of our analysis the appropriate empirical probability density functions (*pdf*-s) are developed: first—for a random occurrence of the lane's availability states and then—for a random

Table 1. Numbers of cases, obtained experimentally for each class interval, with the availability and with the lack of the availability of the analysed lane.

$t = \tau$	$n_t^{(0)}$	$n_t^{(1)}$	$t = \tau$	$N_{\tau+\Delta t}^{(1)}$	$\overline{N_\tau}$
0h30 m	0	25	600	600	600
1h30 m	0	25	600	600	600
2h30 m	0	25	600	600	600
3h30 m	0	25	600	600	600
4h30 m	0	25	600	600	600
5h30 m	0	25	600	597	598.5
6h30 m	3	22	597	577	587
7h30 m	20	5	577	558	567.5
8h30 m	19	6	558	533	545.5
9h30 m	25	0	533	518	525.5
10h30 m	15	10	518	508	513
11h30 m	10	15	508	503	505.5
12h30 m	5	20	503	501	502
13h30 m	2	23	501	498	499.5
14h30 m	3	22	498	493	495.5
15h30 m	5	20	493	486	489.5
16h30 m	7	18	486	478	482
17h30 m	8	17	478	472	475
18h30 m	6	19	472	470	471
19h30 m	2	23	470	470	470
20h30 m	0	25	470	470	470
21h30 m	0	25	470	470	470
22h30 m	0	25	470	470	470
23h30 m	0	25	470	470	470

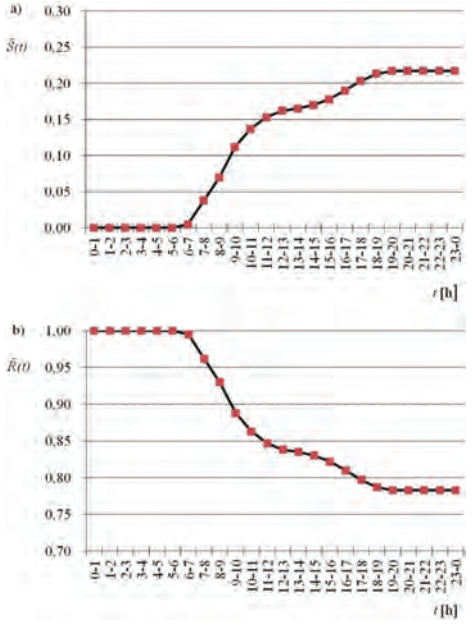


Figure 2. The unreliability (Figure. 2a) and the reliability (Figure. 2b) functions determined experimentally for the analysed lane.

occurrence of the lane's lack of availability. These functions are marked as $\hat{f}(\tau)$ and $\hat{g}(\tau)$, respectively. Based on (Migdalski, 1982) we have:

$$\hat{f}(\tau) = \frac{n - n_t^{(0)}(\tau)}{N \cdot \Delta t} \tag{3}$$

$$\hat{g}(\tau) = \frac{n_t^{(0)}(\tau)}{N \cdot \Delta t} \tag{4}$$

The graphs of these functions obtained for the data considered in the example and compiled in Table 2 are shown in Figure. 3a and in Figure. 3b, respectively.

2.4 *Empirical values relating to the lane's renewal intensity and other ones—determining its failure rate*

The next pair of the empirical dependencies determined by the authors for the lane considered in the example involves respectively the intensity of the lane's renewals $\hat{\upsilon}(\tau)$, when the lane returns to its availability state, and the lane's failure rate $\hat{\lambda}(\tau)$. In order to determine these values for the selected

Table 2. Detailed values determining the functions characterising the availability or the lack of the availability of the analysed lane, calculated on the basis of the empirical data.

$t = \tau$	$\hat{R}$	$\hat{F}$	$\hat{f}$ [h^{-1}]	$\hat{g}$ [h^{-1}]	$\hat{\upsilon}$ [h^{-1}]	$\hat{\lambda}$ [h^{-1}]
0h30 m	1.000	0.000	0.042	0.000	0.042	0.000
1h30 m	1.000	0.000	0.042	0.000	0.042	0.000
2h30 m	1.000	0.000	0.042	0.000	0.042	0.000
3h30 m	1.000	0.000	0.042	0.000	0.042	0.000
4h30 m	1.000	0.000	0.042	0.000	0.042	0.000
5h30 m	1.000	0.000	0.042	0.000	0.042	0.000
6h30 m	0.995	0.005	0.037	0.005	0.038	0.005
7h30 m	0.962	0.038	0.008	0.033	0.009	0.035
8h30 m	0.930	0.070	0.010	0.032	0.011	0.035
9h30 m	0.888	0.112	0.000	0.042	0.000	0.048
10h30 m	0.863	0.137	0.017	0.025	0.020	0.029
11h30 m	0.847	0.153	0.025	0.017	0.030	0.020
12h30 m	0.838	0.162	0.033	0.008	0.040	0.010
13h30 m	0.835	0.165	0.038	0.003	0.046	0.004
14h30 m	0.830	0.170	0.037	0.005	0.044	0.006
15h30 m	0.822	0.178	0.033	0.008	0.041	0.010
16h30 m	0.810	0.190	0.030	0.012	0.037	0.015
17h30 m	0.797	0.203	0.028	0.013	0.036	0.017
18h30 m	0.787	0.213	0.032	0.010	0.040	0.013
19h30 m	0.783	0.217	0.038	0.003	0.049	0.004
20h30 m	0.783	0.217	0.042	0.000	0.053	0.000
21h30 m	0.783	0.217	0.042	0.000	0.053	0.000
22h30 m	0.783	0.217	0.042	0.000	0.053	0.000
23h30 m	0.783	0.217	0.042	0.000	0.053	0.000

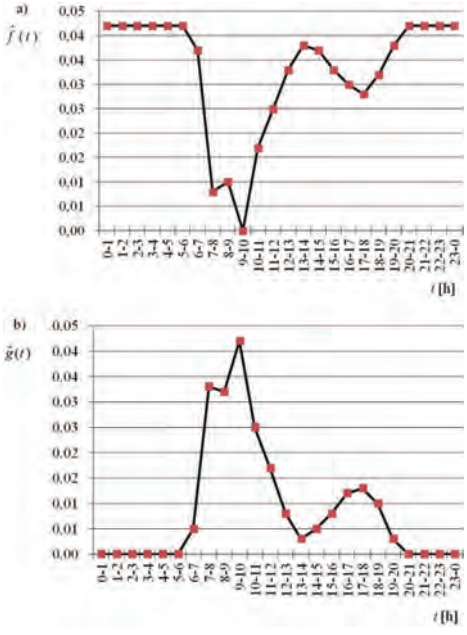


Figure 3. Empirical probability density functions of the occurrence of the state of the lane's availability (Figure. 3a) and of the state of the lane's unavailability (Figure. 3b).

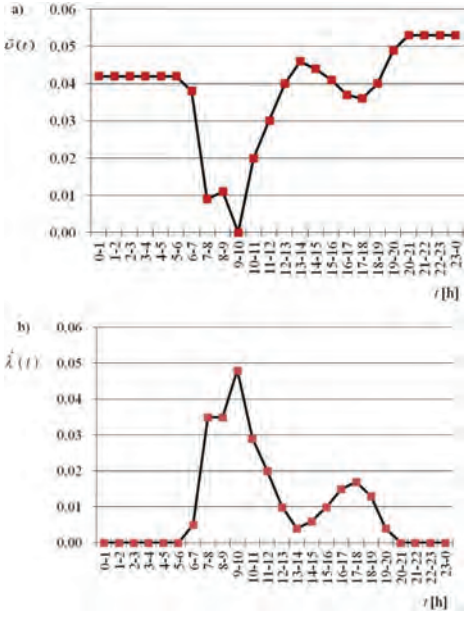


Figure 4. Empirical functions of the lane's renewal intensity (Figure. 4a) and of the lane's failure rate (Figure. 4b) relating to the analysed lane.

argument $t = \tau$, the authors first need to prepare the following auxiliary parameters:

$$N_{\tau}^{(1)} = N - \sum_{t=0h30m}^{\tau} n_t^{(0)} \tag{5}$$

$$N_{\tau+\Delta t}^{(1)} = N_{\tau}^{(1)} - n_t^{(0)}(\tau) \tag{6}$$

for which:

$$\overline{N_{\tau}} = 0.5\left(N_{\tau}^{(1)} + N_{\tau+\Delta t}^{(1)}\right) \tag{7}$$

Based on the above, the following is derived, respectively:

$$\hat{v}(\tau) = \frac{n - n_t^{(0)}(\tau)}{\overline{N_{\tau}}\Delta t} \tag{8}$$

and:

$$\hat{\lambda}(\tau) = \frac{n_t^{(0)}(\tau)}{\overline{N_{\tau}}\Delta t} \tag{9}$$

The detailed values of both these functions, calculated for subsequent values of argument $t = \tau$, are compiled in Tables 1 and 2. These functions are also shown in detail in Figure. 4a and in Figure. 4b, respectively.

3 RENEWAL FUNCTION AND RENEWAL DENSITY SPECIFIED FOR THE WEIBULL-TYPE RENEWAL PROCESS

3.1 *Idea of an alternative renewal process modelling the use of the lane*

As it was indicated in Chapter 1, the manner in which the lane is used during successive full cycles of signal change of a fixed duration of a single cycle can be formally described by a scheme of an alternative renewal process in which the time-intervals of lane's serviceability (availability) of random length T_i alternate with the time-intervals of lane's non-serviceability (failure) which are also of random length Θ_i. As a result of such assumption, we are dealing with a sequence $T_1, \Theta_1, T_2, \Theta_2, ..., T_i, \Theta_i, ...$ (Figure. 5).

3.2 *Modelled probability distributions for the random durations of the states with the lane's serviceability and of the states with the lane's non-serviceability.*

Careful analysis of the graph of the empirical probability density functions of respectively $\hat{f}(\tau)$

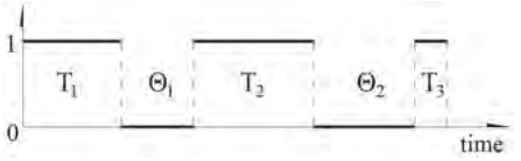


Figure 5. Scheme of an alternative renewal process considered in the example.

- related to a random occurrence in time τ of a state with the lane's availability, and of $\hat{g}(\tau)$ - related to a random occurrence of a state with the lane's failure, as shown in Figure. 3, indicates that it is difficult to assign the uniform probability distribution for the whole day to the random durations of these states. This conclusion is further corroborated by the non-monotonic graphs of functions $\hat{\upsilon}(\tau)$ and $\hat{\lambda}(\tau)$. Consequently, the full daily cycle is recommended to be broken down into shorter component time-intervals for which the fitted probability distribution modelling the use of the lane in such a short period becomes sufficiently reliable. Under this approach, in the night, when during the entire time of the lane use its availability seems to be fully ensured, the random duration of the state of the lane's serviceability should be undoubtedly modelled by a uniform probability distribution. Much more interesting, however, are the lane's states ("availability" or "unavailability"), with random durations, which alternate with each other during a day. Considering the fact that the values of a lane's failure rate did not turn out to be constant in any time-period being sufficiently long, the random durations of the lane's serviceability states T_i and also the random durations of the lane's non-serviceability states Θ_i are recommended by the authors to be assigned by the two-parameter Weibull probability distributions, for which the intensity of a failure occurrence is determined from the formula (Kaminskiy, 2013):

$$\lambda(\tau;\alpha_\lambda,\beta_\lambda)=\frac{\beta_\lambda}{\alpha_\lambda}\left(\frac{\tau}{\alpha_\lambda}\right)^{\beta_\lambda-1} \tag{10}$$

whereas the intensity of a successive renewal, leading to the lane's serviceability state, is derived from a similar formula:

$$\upsilon(\tau;\alpha_\upsilon,\beta_\upsilon)=\frac{\beta_\upsilon}{\alpha_\upsilon}\left(\frac{\tau}{\alpha_\upsilon}\right)^{\beta_\upsilon-1} \tag{11}$$

In this approach, parameters α_λ and α_υ are the scale parameters while the parameters β_λ and β_υ are interpreted as the shape coefficients. In time-intervals when function $\hat{\lambda}(\tau)$ rises non-linearly it is usually assumed that $\beta_\lambda \geq 3$. On the other hand, for time-intervals with a non-linearly decreasing function $\hat{\lambda}(\tau)$, a value $\beta_\lambda < 1$ should be assumed.

3.3 *Determining the renewal function relating to the considered process*

The choice of a renewal function that will fit well all the parameters of the considered formal model is difficult, especially in the case of a non-monotonic empirical failure intensity function $\hat{\lambda}(\tau)$ relating to this model. In our example this function during certain periods of a day turns out to be a growing function while during others, a decreasing function. It is recommended in the professional literature to use for such the complex design cases a mixture of the probability distributions (Kaminskiy, 2013). In the simplified model proposed by the authors in this paper the Weibull probability distribution is chosen in this field to describe both the random times of the lane's serviceability T_i and the random times of the lane's non-serviceability Θ_i. With such assumption, the conventional approach used to determine the appropriate renewal function, based on the use of the Laplace transform technique (Bobrowski, 1985), becomes rather ineffective. Some proposals of the way in which this function could be determined as accurately as possible in the case when the Weibull-type renewal process is considered were given for example in (Yannaros, 1994) and in (Lomnicki, 1966). In other papers, e.g. in (Smeitink & Dekker, 1990) and in (Jiang, 2008), it was recommended to use in this field some more or less simplified estimates. In our analysis we have chosen for practical use the approximate formula given in (Jiang, 2010). Such a formula is appropriate, however, only in time-intervals for which the failure rate can be estimated as growing. This formula is constructed as follows:

$$H(\tau)=\xi F(\tau)+(1-\xi)\Lambda(\tau) \tag{12}$$

As one can see, the renewal function $H(\tau)$ is calculated here by an appropriate combination of the cumulative distribution function (*cdf*) $F(\tau)$, being specific for the Weibull probability distribution, and of the other function being the measure of the cumulative frequency of the lane's failures $\Lambda(\tau)$. The combination coefficient ξ, estimated using the least-squares method, is derived in such an approach from the formula:

$$\xi=\xi(\beta)=1-\exp\left[-\left(\frac{\beta-1}{0.8731}\right)^{0.9269}\right] \tag{13}$$

Obviously, its value depends on the value of the shape coefficient, i.e. $\beta=\beta_\lambda$ or $\beta=\beta_\upsilon$, respectively.

It should be emphasised that the above mentioned formula was calibrated on the assumption that the scale coefficient was set at the level $\alpha = 1$. The cumulative distribution function for the Weibull probability distribution, described both according to (10) and according to (11), is expressed as:

$$F(\tau) = 1 - \exp\left[-\left(\frac{\tau}{\alpha}\right)^{\beta}\right] \tag{14}$$

The cumulative failure intensity function for the same distribution can be calculated directly based on the definition, which gives:

$$\Lambda(\tau) = -\ln\left[1 - F(\tau)\right] = \left(\frac{\tau}{\alpha}\right)^{\beta} \tag{15}$$

Detailed values of this function can also be derived directly from the empirical data, on the basis of the histograms presented in Chapter 2 of this paper. It is essential that the approach described above allows an efficient determination of the renewal function $H(\tau)$ only for those time-intervals of a day when the empirical function of failure intensity $\hat{\lambda}(\tau)$ is observed to be monotonically increasing. In future research the authors will try to determine a similar-type renewal function for time-intervals of decreasing failure intensity. In order to attain this objective, we want to apply a procedure proposed in (Smith & Leadbetter, 1963), according to which:

$$H(\tau) = \sum_{i=1}^{\infty} (-1)^{i-1} \eta_i \frac{z^i}{\gamma_i i!} \tag{16}$$

where:

$$z = \tau^{\beta} \tag{17}$$

$$\eta_1 = \gamma_1 \tag{18}$$

$$\eta_i = \gamma_i - \sum_{j=1}^{i-1} \gamma_i \eta_{i-j} \tag{19}$$

$$\gamma_i = \frac{\Gamma(i\beta + 1)}{i!} \tag{20}$$

$$\Gamma(k) = \int_0^{\infty} x^{k-1} e^{-x} dx \tag{21}$$

3.4 *Determining the renewal density*

Setting out from equation (12) defined for time-intervals with increasing values of functions $\hat{\lambda}(\tau)$, it is easy to determine the renewal density accompanying this phase of the lane's use. It can be expressed as:

$$h(\tau) = \frac{dH}{d\tau} = \xi f(\tau) + (1 - \xi)\frac{f(\tau)}{R(\tau)} \tag{22}$$

In this case, the lane's reliability function $R(\tau)$ approximates the empirical function described in formula (2), whereas the probability density function of the lane's failure $f(\tau)$ approximates the corresponding empirical function defined by equation (3). Application of formula (16) for time-intervals with decreasing values of function $\hat{\lambda}(\tau)$ leads to the dependence given in (Smith & Leadbetter, 1963):

$$h(\tau) = \frac{\beta}{\tau} \sum_{i=1}^{\infty} (-1)^{i-1} \eta_i \frac{i z^i}{\gamma_i i!} \tag{23}$$

4 AVAILABILITY AND RELIABILITY OF THE ANALYSED LANE

4.1 *Lane's availability in the model of the alternative renewal process*

So far, the random durations of the time of the lane's serviceability state T_i were described by a probability distribution characterised by the probability density function $f(\tau)$ (approximating the empirical function $\hat{f}(\tau)$) and by the cumulative distribution function $F(\tau)$. Similarly, the lane's random non-serviceability time-intervals Θ_i were described by a probability distribution with the probability density function $g(\tau)$ and with the cumulative distribution function $G(\tau)$. With such assumptions, according to the conventional renewal theory, the distribution of the probability of random time-intervals between consecutive renewals (interpreted as the restoration of the lane to the serviceability state), i.e. of the time-intervals arranged in sequence

$$T_1 + \Theta_1, T_2 + \Theta_2, ..., T_i + \Theta_i, ..., \tag{24}$$

can be described by the cumulative distribution function:

$$\Phi(\tau) = \int_0^{\tau} F(\tau - t) dG(t) \tag{25}$$

for which the Laplace transform is as follows:

$$\tilde{\Phi}(s) = s\tilde{F}(s)\tilde{G}(s) \tag{26}$$

A similar formula is true for the Laplace transform of the probability density functions, giving:

$$\tilde{\phi}(s) = \tilde{f}(s)\tilde{g}(s) \tag{27}$$

Based on the above, the authors compute the Laplace transform of the renewal function of an alternative renewal process analysed in the experiment described in the example, which produces:

$$\tilde{H}(s) = \frac{\tilde{f}(s)\tilde{g}(s)}{s\left[1-\tilde{f}(s)\tilde{g}(s)\right]} \tag{28}$$

Function $H(\tau)$ computed by the appropriate re-transform in this case is interpreted as the anticipated number of the lane's renewals within the (0, τ) time-interval. Likewise, for the same lane, the expected number of the lane's failures can be expressed as a renewal function $H_f(\tau)$, for which the Laplace transform is expressed by the formula:

$$\tilde{H}_f(s) = \frac{\tilde{f}(s)}{s\left[1-\tilde{f}(s)\tilde{g}(s)\right]} \tag{29}$$

The lane's availability $K(\tau)$ in this model expresses the probability that at a certain time τ such lane will be observed in a serviceability state. Therefore, this availability can be calculated as the sum of the probabilities of two fully separable random events formulated as follows:

- the first event—that the serviceability time-interval until the first lane's failure is longer than τ, which means that $T_1 > \tau$,
- the second event—that the n-th lane's renewal occurs within the $(t, t+\Delta t)$ time-interval, while within the (t, τ) time-interval the next lane's failure does not occur.

Hence:

$$K(\tau) = R(\tau) + \int_0^{\tau} R(\tau - t)\, dH(t) \tag{30}$$

The Laplace transform for such availability is expressed by the following formula:

$$\tilde{K}(s) = \left[1-\tilde{f}(s)\right]\left[\frac{1}{s} + \tilde{H}(s)\right] \tag{31}$$

which, after a substitution of (28), yields:

$$\tilde{K}(s) = \frac{1-\tilde{f}(s)}{s\left[1-\tilde{f}(s)\tilde{g}(s)\right]} \tag{32}$$

Ultimately:

$$\tilde{K}(s) = \tilde{H}(s) - \tilde{H}_f(s) + \frac{1}{s} \tag{33}$$

hence:

$$K(\tau) = H(\tau) - H_f(\tau) + 1 \tag{34}$$

This means that the probability that the analysed lane will be observed at time τ in a state of its non-serviceability, expressed as the difference 1 – $K(\tau)$, in the model of the alternative renewal process presented by the authors in this paper is equal to the difference between the expected number of the lane's failures and the number of the lane's renewals that have occurred until that time.

4.2 *Reliability of the analysed lane*

The basic difference between the availability $K(\tau)$ and the reliability $\Omega(\tau, \Xi)$ determined for the lane considered in the example is such that the first of these values is determined for a selected time-point $t = \tau$, while the second one—for the entire time of the observation, for example for $\Xi = 24$ h. In this approach, the reliability of the lane is interpreted as the probability that such lane will be observed in a serviceability state throughout the whole day. Therefore the following formula can be applied for practical use:

$$\Omega(\tau, \Xi) = R(\tau + \Xi) + \int_0^{\tau} R(\tau + \Xi - t)\, dH(t) \tag{35}$$

5 CONCLUDING REMARKS

The main aim of the research undertaken by the authors was to verify the possibility of describing the process of the use of the lane at a signalised intersection, having unambiguously random characteristics, using the classical formal model of an alternative renewal process. In this analysis, failures of such lane corresponded to the time-intervals when it was non-serviceable because the queue of vehicles awaiting before the red traffic signal turned out to be too long. However, in subsequent signal cycles with a fixed duration of a single cycle, the lane was renewed after each occurrence of a non-serviceability state, so that it became serviceable until the next failure. According to the authors, the choice of the formal model itself seems to be quite straightforward, although its authoritative and reliable mathematical description is difficult. The experiment considered in the example revealed that the use of the lane analysed in a 24-hour cycle is associated with complex, multimodal probabil-

ity distributions of subsequent failures and subsequent renewals of random duration. Therefore, a precise description of the process of this type seems to require an appropriate combination of the probability distributions, which significantly complicates both the mathematical formulae themselves and the resulting inferences. In the presented analysis the authors attempted to apply in practice certain, seemingly acceptable, simplifications. Due to the fact that the function defining the lane's failure intensity did not turn out to be a uniform with respect to the observation time at any stage of the cycle analysed daily the authors were unable to use in this field the simple formal model developed for the renewal processes of the exponential-type, well described theoretically. Therefore it has been proposed a different, more complex, calculation procedure, constituting a set of the analyses conducted separately at certain time-intervals, for which one could assume an unambiguously monotonically increasing or monotonically decreasing graph of the failure intensity function. This approach allowed us to describe the considered renewal process by using a two-parameter Weibull probability distribution, reduced further to the only one-parameter process due to the assumption that the scale parameter was set at a level $\alpha = 1$. The composition of the appropriate failure rate functions, where each of which will be specified in a given time-interval, should allow for the calculation of the reliability being accurate for the analysed lane. The basic problem involved in describing the actual process of the lane's use by means of a formal model of an alternative renewal process with random serviceability and non-serviceability time-intervals characterised by the Weibull probability distributions is that in this case it is impossible to effectively use the Laplace transform to specify the renewal function. In fact, the precise determination of this function in a strict manner becomes possible only on the basis of complex and time-consuming numerical calculations. For this reason, the authors recommend to use in this field a simplified estimate, presented in a mathematically-closed form, which seems to be acceptable, even if prone to quite a significant error of approximation for high values of the shape coefficient β. This estimation, however, can only be used during time-intervals with increased failure intensity function. Wherever the intensity of the failure occurrence is diminishing, the authors suggest using a more complex recursive procedure, which is well described in professional literature. The authors intend to verify the accuracy of this type of approach in the future. Undoubtedly, if the well-verified renewal function determining the proposed formal model will be unambiguously identified, for the available experimental data, it should enable to definitively calculate first the lane's availability, specified at a given time-point $t = \tau$, and subsequently its reliability, specified for the whole observation time-interval Ξ. As the functions of this type, proposed by the authors so far, do not seem to be determined accurately enough, in the final part of this study, the authors limit themselves to presenting subsequent steps of the planned future procedure, without providing numerical results of the applied formulae. Nevertheless, the authors hope that this description of the process of the lane's use at a signalised intersection will prove useful in the practical design of intersections of this type, even if the mathematical model itself seems to be relatively complex.

REFERENCES

Bobrowski, D. 1985. Modele i metody matematyczne teorii niezawodności w przykładach i zadaniach. Wydawnictwa Naukowo – Techniczne. Warszawa.

Chodur, J. Ostrowski, K. and Tracz, M. 2011. Impact of saturation flow changes on performance of traffic lanes at signalized intersections. *Proceedings of the 6th International Symposium on Highway Capacity and Quality of Service.* Stockholm. Sweden. pp. 600–611.

Jiang, R. 2008. A gamma-normal series truncation approximation for computing the Weibull renewal function. *Reliability Engineering & System Safety*. Vol. 93: 616–626.

Jiang, R. 2010. A simple approximation for the renewal function with an increasing failure rate. *Reliability Engineering & System Safety*. Vol. 95: 963–969.

Kaminskiy, M.P. 2013. Reliability models for engineers and scientists. CRC Press. Taylor & Francis Group. Boca Raton.

Lomnicki, Z.A. 1966. A note on the Weibull renewal process. *Biometrika*. Vol. 53. No. 3 and 4: 375–381.

Migdalski, J. (ed.) 1982. Poradnik niezawodności. Wydawnictwa Przemysłu Maszynowego WEMA. Warszawa.

Ostrowski, K. 2014. Attempt to apply the theory of reliability to assessment of signalised lane operation. Proc. of European Safety and Reliability Conference ESREL, Safety and Reliability, Methodology and applications, CRC Press/Balkena, Taylor and Francis Group. pp. 335–341. Wrocław. Poland.

Ostrowski, K., Tracz, M. 2015. Availability and reliability of a signalised lane. *Proc. of the 6th International Symposium on Transportation Network Reliability*. Japan.

Smeitink, E. & Dekker, R. 1990. A simple approximation to the renewal function. *IEEE Transactions on Reliability*. Vol. 39. No 1. April: 71–75.

Smith, W.L. & Leadbetter, M.R. 1963. On the renewal function for the Weibull distribution. *Technometricks*. 5 (3): 393–396.

Tracz, M. Ostrowski, K. 2012. Impact of capacity variability in different weather conditions on reliability of signalised intersections. *The 5th International Symposium on Transportation Network Reliability*. Hong Kong.

Yannaros, N. 1994. Weibull renewal processes. *Annals of the Institute of Statistical Mathematics*. Vol. 46. No. 4: 641–648.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Application of failure classification schemes to technology qualification

T. Myhrvold, A. Hafver, S. Eldevik, F.B. Pedersen, O.I. Haugen,
K. Kvinnesland & D. McGeorge
DNV GL, Norway

ABSTRACT: Current methods for Technology Qualification (TQ) rely to a large degree on traditional reliability methods developed for hardware systems. However, as technology becomes more dependent upon software, leading to more complex systems, there is a need to better reflect the system perspective. One aspect of complexity is that the system constituents is heavily interconnected and dependent upon each other, i.e. the system shows emergent behavior. Systemic failures, not caused by component failures, but caused by how the constituents interact (i.e. emergence), need to be addressed differently than previous practices. This paper presents a consistent way of classifying failures to help discovering the critical failures and to guide how to handle the identified failures in a system TQ setting. To ensure that the identification process is consistent and as complete as possible, we advocate that failures should be classified according to different perspectives, where the classification within each perspective is Mutually Exclusive and Collectively Exhaustive (MECE). We show how this can be used to guide how evidence is collected in the technology qualification process.

1 INTRODUCTION

1.1 *Background*

Technology development can either enable a project to be realized or it can enhance the value of it. Either way the technology developer has to build the operator's confidence in the technology. The operator again needs to build the confidence of the other stakeholders in the project before a decision to implement the technology can be taken. To build this confidence, a systematic risk-based qualification process must be performed.

Technology qualification comprises activities to assess, improve and safeguard technology. It aims at providing evidence that the technology will function within specified limits with an acceptable level of confidence (DNVGL-RP-A203, 2017).

To qualify a system, different perspectives are needed to identify possible ways in which the system may fail, and to select appropriate ways of collecting evidence for the qualification process. There exist many different definitions of failure and failure types and many ways of classifying failures. This makes it challenging to get an overview of the completeness and consistency with respect to whether most failures have been identified and addressed or not. In this paper, we suggest a method that provides a structured and consistent way of classifying and handling failures, and relate it to the TQ process as described in DNVGL-RP-A203.

1.2 *Objective*

The purpose of this work is to develop a systematic method to identify an as-complete-as-possible set of relevant failures, that allows for an effective failure identification and that increase the confidence that most failures have been identified and addressed. The method is generic and can be used in combination with both new and conventional failure assessment methods in a system TQ setting.

2 DEFINITION OF FAILURE

The literature is abundant with definitions of failure. In this work we use a broad definition:

Failure is the loss of a function (fully, in part, or erroneous delivery).

We explicitly avoid to mention where or how the loss of the function materializes, to highlight that the failure is on the abstraction level of the *function*. This makes it easier to consider failures without limiting it to actual resources, or combinations thereof. The term resources is used here to refer to an individual part, component, device, functional unit, equipment, subsystem, software, people, environment, organizational structures, etc. or combinations of one or more of the above. This is closely related to how an item is defined in IEC 60050-192 (2015).

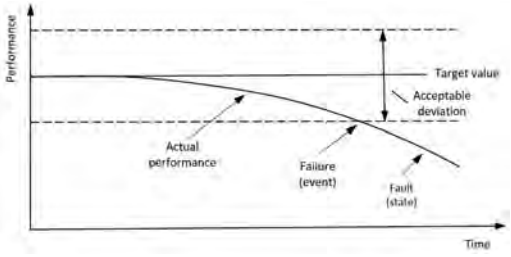


Figure 1. Illustration of the relationship between a failure event and a fault state (based on Rausand & Øien 1996).

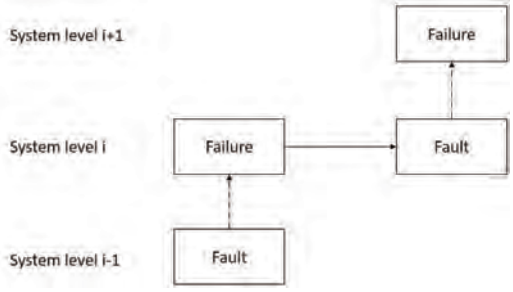


Figure 2. Failure model illustrating the relationship between failure at one level in a system hierarchy and fault at another level.

If failure is considered as the event when a function was lost, the failed state can then be related to the term *fault* state used by IEC 60050-192 (2015).

One illustration of the relationship between a failure event and a fault state is shown in Figure 1.

Figure 1 shows the relationship between a failure and a fault with respect to a specific function. Note that in Figure 1, a failure leads to a fault, and not the other way around. Figure 2 illustrates how a fault due to a failure at one level in a system hierarchy may lead to a failure at a higher level. Note, however, that a failed function on one level does not necessarily propagate to the next (e.g. if that level has redundancy or the failed function was non-critical for the next-level function), and that a failed function on one level does not necessarily imply a failure on a previous level (e.g. emergent failures).

3 EXISTING FAILURE CLASSIFICATION SCHEMES

As can be seen from the definition of failure given above, any method for failure identification and analysis strongly depends on the ability to identify all the required functions of the system subject to analysis. In this work it is assumed that these functions are known and will focus on ways that failures can be classified.

Various standards classify failures in different ways. IEC 61508-4 (2010) applies two main classification schemes, categorizing failures by its cause or by its mode. Failures categorized by its cause are either random (in hardware) or systematic (in hardware or software), where the definition of systematic failure are the same as given by IEC 60050-191 (1990).

IEC 61508-4 classify all (random hardware) failure by its mode as: dangerous undetected (DU), dangerous detected (DD), safe undetected (SU), and safe detected (SD).

ISO 14224 (2015) classifies failure after its failure mechanism, failure cause, and failure mode and provides a detailed list of failure modes down to component level.

ISO/TR 12489 (2013) provides three different classification schemes for use in reliability modelling and calculation of safety systems reliability: According to randomness, according to occurrence (state of the failing item: when running, on stand-by, due to demand), and by the way the failure is detected (revealed, hidden, or due to demand). Classification of failure according to randomness is divided into random and non-random (systematic) failure and are further broken down into hardware, software and human failure.

Blache and Shrivastava (1994) introduced a generic classification by using modified definitions from IEC-60050-191, which is shown in Figure 3.

Rausand and Øien (1996) stated that failure modes may be classified in three main groups related to the function of the item: total loss of function, partial loss of function, and erroneous function.

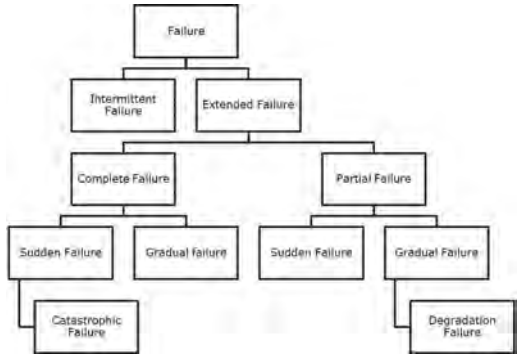


Figure 3. Failure classification scheme for failure modes by Blache and Shrivastava (1994).

Rausand and Øien (1996) also described an alternative classification scheme by failure cause as an option, where failure cause is "the circumstance during design, manufacture or use that have led to failure". The failure cause is a necessary information to avoid failures or re-occurrence of failures. Failure causes may be classified in relation to the life-cycle of an item or a functional block as illustrated in Figure 4. They also described classification of failure by its effect and severity (From MIL-STD 882): catastrophic, critical, marginal, negligible.

Håbrekke et al. (2013) classified failures into two main schemes based on the failure definitions given in IEC 61508-4 as mentioned above: either by its mode or by its cause. When classifying by its cause, Håbrekke et al. (2013) use the definitions given by the IEC 61508-4 standard (differentiate between random hardware failures and systematic failures), but give a more detailed breakdown of the systematic failures, as shown in Figure 5.

When classifying by its mode, they categorize failure (not limited to random hardware failures only) as (Håbrekke et al., 2013): DU, DD, SU, SD and non-critical (NC).

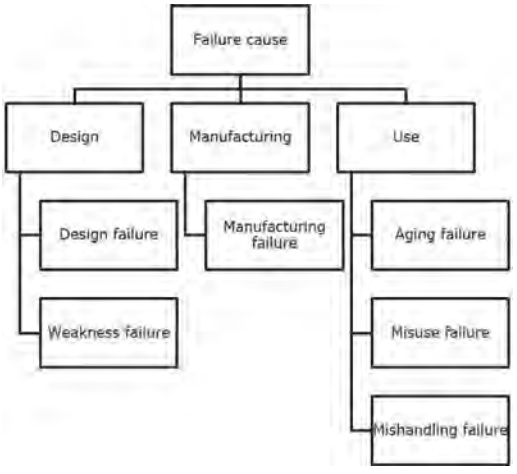


Figure 4. Failure classification after its cause (Rausand & Øien, 1996).

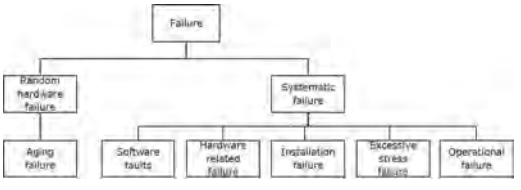


Figure 5. Failure classification after its cause by the PDS forum (Håbrekke et al., 2013).

The above show a wide variety of failure classification schemes, but no method to structure them or give guidance on use seems to exist.

In the current work we propose a simple and structured method for classification of failures and exemplify how it can be used for more consistent failure identification in a technology qualification context.

4 A STRUCTURED METHOD FOR IDENTIFYING FAILURES THROUGH PERSPECTIVES

A challenge with the classification structures described in the previous section is that it is difficult to know whether the schemes identify and address all possible and relevant failures and if some failures have been overlooked. Many of the classification schemes build elaborate hierarchies where failure classes are broken down into sub classes. A danger of following such a hierarchical approach is that the branches of the classification tree may differ with respect to how detailed they are and which criteria are used for the subdivision.

Rather than following a hierarchical approach, which may tend to narrow the perspective of the analyst, we propose a method where failures are identified by viewing the system from different perspectives, and alternating between these views, rather than creating one unified failure class hierarchy.

Within each perspective, a Mutually Exclusive and Collectively Exhaustive (MECE) set of categories are established. The MECE categories are intended to help the assessor structure the failure identification process by introducing complete sets of failure categories so that all aspects of a perspective is covered. This will increase the confidence that the failure identification method captures the relevant failures.

4.1 *Commingled structures*

Each of the perspectives and its MECE categories should trigger identification of different failures. That is, one should be able to identify and place failures in the MECE categories of each and every perspective.

This is fundamentally different than the hierarchical approaches described in Section 3. However, with the suggested approach, a variety of hierarchical classification schemes can be reproduced by combining the MECE categories of the different perspectives, and are thus fully compatible with previous methods.

This approach enables us to concentrate on N perspectives with M_N individually relevant MECE

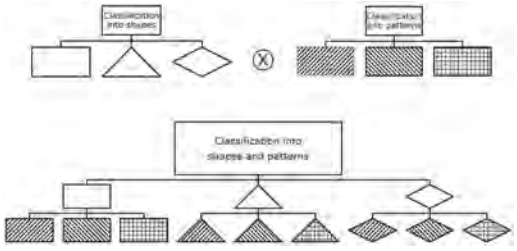


Figure 6. Independent classification schemes (upper panel) used to build a full classification structure (lower panel).

categories rather than complex structures of categories that may not be MECE. Figure 6 shows how two perspectives (here geometric form and pattern) can be assessed individually (upper panel), and if relevant can be combined to a commingled structure (lower panel).

The commingled classification structure will remain MECE, and thus complete. Note that all combinations may not be relevant, however this should not be regarded as a problem. Rather than assuming that some combination is impossible, the possibility of every combination is kept open. By considering the views independently, one avoids debates such as whether software failures can be random or not. This may help the analyst avoid limitations given by preconceived ideas.

5 FAILURE CLASSIFICATION IN ONE-DIMENSIONAL VIEWS

Technology qualification involves executing a work process comprising of a set of steps. In brief, the first four steps are:

- Establish a qualification basis identifying the technology, its functions, its intended use, as well as the expectations to the technology and the qualification targets.
- Assess the technology by categorizing the degree of novelty to focus the effort where the related uncertainty is most significant and identify the key challenges and uncertainties.
- Assess threats and identify failures and their risks.
- Develop a plan containing the qualification activities necessary to address the identified risks.

Part of the technology assessment is to break the top level system functions into sub-functions. The loss (partial or complete) of a function on any level in this functional hierarchy constitutes a failure of that function. Hence, what constitutes a failure, or which functions that fail is not the focus of this chapter. Rather, the focus here is to identify how functions may fail, by relating the functions to the resources and processes used to deliver the specific functions.

TQ is a process of assessing whether a particular solution, composed from a specific set of resources and involving a set of processes, is able to deliver the desired functions with sufficient confidence. This involves to identify how a function may fail and the associated risks. Based on this, the next step in the TQ process is to develop a plan of qualification activities necessary to address the identified risks. To accommodate this, we propose an approach where the system is analyzed from a set of different perspectives. Some relevant perspectives may be:

- Resource perspective
- Life-cycle perspective
- Randomness perspective
- System level perspective

In addition, two perspectives which may be used to guide the assessor on how to handle the identified failures may be:

- Controllability perspective
- Risk perspective

Each of these perspectives are described below, and we show how the individual perspectives may be used in TQ. For each perspective, we propose a list of MECE categories to ensure completeness and avoid confusion. Looping through a set of perspectives is intended to instruct the assessor to consecutively change her point-of-view, and thus help her identify a wider range of failures than if conventional methods were used. This is an alternative to using more elaborate hierarchical failure classification schemes that may be ambiguous or contain an unmanageable number of failure classes.

Note that other perspectives can also be relevant in TQ. This paper exemplify the use of relevant perspectives and categories, focusing on the principle of building unambiguous perspectives that can help the assessor in the failure identification process, and give guidance on how to handle the identified failures.

5.1 *Resource perspective*

The resource perspective provides a classification of failure based on its relation to the particular hardware, software or human components of a system (including their interactions), see Figure 7.

The core motivation of the resource view is to elicit which of the resources of the system are involved in the failure.

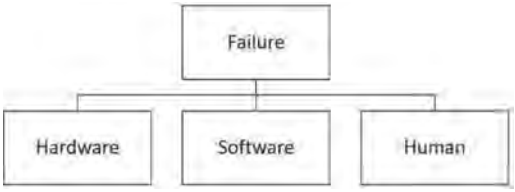


Figure 7. Classification of failure based on its origin.

Note that Figure 7 uses a coarse classification of resources into the groups hardware, software and human. This is only one possible set of MECE categories used for exemplification of the resource perspective. Other categories or granularity may be used as long as it is kept MECE.

Note that although the categories themselves are MECE (for consistency), a failure does not necessarily belong to only one category. The failure of a function may for example involve both hardware and software, or the interaction between them, even if neither of the individual resources have failed.

The resource perspective can help the assessor to identify the resources that need to be qualified to ensure that functions are adequately provided.

5.2 *Life-cycle perspective*

Classification of failure may be done in relation to time, i.e. when in the life cycle the failure is introduced s shown in Figure 8. We call this the life-cycle perspectives.

This is essentially similar to Rausand and Øien (1996) who described an alternative classification scheme by failure cause as an option, see Figure 4. In our case, we have also added failure occurring in concept, installation, and decommissioning of the system. The granularity of the categories is project dependent, depending on the nature of the system development and associated activities.

The importance of the life-cycle view in a TQ setting is that failures, or precursors to failures, may occur in any phase of the project. Some failures will materialize immediately while some may not materialize until later in the system's life-cycle. The TQ process is by nature iterative and flexible, and is used to reduce the possibility of unidentified failures that may propagate into later stages of the system life-cycle. Thus, taking a life-cycle perspective can help the assessor identify failures introduced at certain stages in the life-cycle, but that may not materialize until later stages. By identifying when a failure might be introduced and when it might materialize can help in finding ways of how it could be prevented or handled. For example, failures in specification, design, or manufacturing

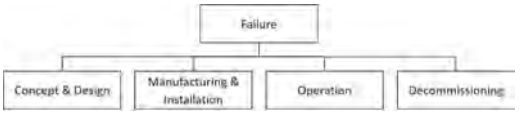


Figure 8. Classification of failure according to when in the life cycle the failure is introduced.

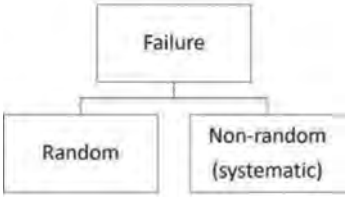


Figure 9. Classification of failure according to its randomness.

may be detected and handled through quality control, third party verification and testing. Other failures may develop and occur in operation and the assessor can then identify required contingency plans, monitoring, inspection, testing or maintenance plans, etc.

5.3 *Randomness perspective*

A commonly used failure classification scheme is to classify into random and systematic failure. Here we suggest to use random and non-random (see Figure 9) since this clearly states if a failure is random or not.

Note that we deliberately do not use the term random hardware failure, since random hardware failure is a subset of random failure. Both IEC 61508-4 (2010) and the PDS forum (Håbrekke et al., 2013) classify failures by its cause to be either random hardware failure or systematic. According to our understanding, none of these schemes are MECE.

In a TQ setting, whether a failure is considered a random (stochastic) process, or if it is systematic, will be very relevant with respect to how the failure is handled. If, for example, the assessed system function is to withstand a weather-related load, e.g. wave or wind loads, the maximum load the system will see throughout its lifetime, is stochastic. In this scenario, the usual approach is to estimate a maximum load distribution, and to require the structure to have the capacity to withstand some percentile of this maximum load distribution (e.g. a 100 yr load). On the other hand, if the assessed system is a small bridge designed for smaller cars, the bearing capacity (the system function) will always fail if a heavy-duty truck tries to cross. This

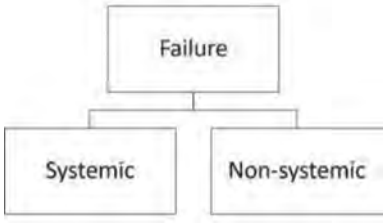


Figure 10. Classification of failure according to systemic behavior.

is a systematic, or non-random failure, and a TQ process would identify if the bridge may be subjected to such loads, and the bridge should either be re-designed, or limitations to vehicle weights could be imposed.

5.4 *System-level perspective*

Traditional reliability engineering methods derive failures on system level from failure on component level. However, a system function may fail even if none of its sub-functions or resources have failed. Such *systemic* failures are typically related to how different resources in a system interact. It may for instance occur based on interaction between the system and its environment, unintended or unfortunate interactions, or conflicting actions among the system resources. This perspective is shown in Figure 10.

Here we have used the ISO/TR 12489 (2013) definition of systemic failure:

"*Systemic failure* is failure at system level which cannot be simply described from the individual component failures of the system." Asking the question: Does the failure depend on failures on a lower system level, or is it a feature that only manifests itself on this level? A good example is the Mercedes A-class failure of the "Elk-test" (Teknikens Värld 2017). Here, the new Mercedes A-class was put through an evasive maneuver test, and above a certain speed, the car tipped over when performing the evasive maneuver. The steering performed within its intended range and the speed was within the car's limitations. Thus, no function was operated outside its intended range, however, the result was a total failure of the system's function, i.e. being able to drive and steer.

In a TQ setting, the distinction between systemic and non-systemic failures becomes important with respect to how they are identified and handled: A non-systemic failure can be addressed by ensuring that no subsystems or components fail. However, handling systemic failures requires that the entire system is considered as a whole (including architecture, life cycle, system control structures, etc.). Systemic failures cannot be identified by analyzing each of the system's resources separately. Some of the systemic failures may be identified through existing methods like system FMECA (Subburaman, 2010), STAMP/STPA (Leveson, 2011, 2013), FRAM (Hollnagel, 2004) or through system integration testing and system pilot testing. However, the assessor should, where relevant, also utilize virtual testing using e.g. simulations to discover systemic failures that cannot be anticipated based on the understanding of individual components, and which would be too expensive or dangerous to test in real-life. For the Mercedes A-class this would e.g. have been to explore the limits of the steering, the speed and the road friction.

5.5 *Controllability perspective*

According to control theory, the ability to control a process depends on the ability to observe the process, the ability to influence the process, having a defined objective, and sufficient understanding of how to influence the system, based on the above. In the controllability perspective we categorize failures as controllable or not. In reality, there might be degrees of controllability ranging from full control to no control at all. The granularity of the categorization can be modified according to project specific needs.

In a TQ setting, this perspective is important to identify which failures that can, or need to be, controlled, and which cannot. The latter must then either be accepted based on risk evaluations (see risk perspective below), or the system cannot be qualified.

For those failures that can be controlled, the control can be exercised at different stages in the life cycle of a system. Some can be controlled by design, e.g. according to standards, best practices, quality assurance, etc. Other failures may be controlled in fabrication through e.g. testing, quality assurance, procedural control, etc. Or similarly in other phases like installation, operation etc.

Sometimes it is possible to choose which phase control is implemented, e.g. through robust design, or alternatively operational monitoring and mitigation strategies.

5.6 *Risk perspective*

It is common to relate failures to risk. By risk we mean the consequences of the failure and its associated uncertainty (ISO 31000 2010, Society for Risk Analysis 2015). Different risk metrics may be used to measure the risk (either quantitatively or qualitatively). Here we only introduce the categories acceptable, and not acceptable. As with the other

perspectives, it is possible to use a more fine-grained set of categories if that is expedient for the project.

To define the border between acceptable and not acceptable risk, the ALARP principle (as low as reasonably practicable), is applied in many industries.

ALARP implies that risk should be reduced until the sacrifices necessary to reduce it further become disproportionate to the risk to be avoided (Health and Safety Executive UK 2017). Note that the risk perspective can be seen in conjunction with the controllability perspective as the ability to reduce risk is related to controllability. This is in essence the aim of the TQ process, i.e. to control, eliminate, or reduce the effect of, failures that are in the unacceptable risk category.

6 CONCLUDING REMARKS

This paper has presented a consistent way of classifying failures to help discovering the critical failures and to ensure that all types of failure are addressed, and to guide handling of identified failures in a system TQ setting. For failure identification, assessment and management, it is useful to explore the failures from different perspectives. To ensure that the identification process is consistent and as complete as possible, we suggest to use Mutually Exclusive and Collectively Exhaustive (MECE) failure classification schemes. These perspectives, or failure classes, can then be used to guide how evidence is collected in the qualification process.

The benefits of such an approach is that the failure identification process will become more practical and tractable by employing simple one-dimensional failure classification schemes and viewing them one at a time, rather than investigating hierarchical failure structures.

In addition, the method guides the assessor towards exploring the entire failure space, and do not exclude any possibilities from the start.

It should be noted that the method proposed in the current work is fully compatible with previous methods. For instance, by combining the one-dimensional MECE schemes, a more complicated structure is easily created, see Figure 6.

One of the perspectives suggested here explicitly focus on identifying and addressing system-level failures (also called systemic failures) that are not easily addressed by conventional methods.

In TQ this is useful since it gives guidance to identification of different types of failures and how to handle them, i.e. what the best method will be to collect the evidence that the critical failures will be managed in the best possible way.

ACKNOWLEDGEMENT

The work in this paper is partially based on work supported by ECSEL JU under the program HORIZON 2020 through SafeCOP (project number 270338).

REFERENCES

Blache, K.M. & Shrivastava, A.B. 1994. Defining failure of manufacturing machinery & equipment. *Proc. Annual Reliability and Maintainability Symp*., pages 69–75.

DNVGL-RP-A203. 2017. *Technology Qualification.* Recommended practice.

Health and Safety Executive UK. 2017. "ALARP at a glance." Available: www.hse.gov.uk/risk/theory/alarpglance.htm. [Accessed 15 December 2017].

Hollnagel, E. & O. Goteman. 2004. The functional resonance accident model. *Proceedings of cognitive system engineering in process plant*, 155–161.

Håbrekke, S., Hauge, S., & Onshus, T. 2013. *Reliability Data for Safety Instrumented Systems*. PDS Handbook.

IEC 60050-192. 2015. International electrotechnical vocabulary – Part 192: Dependability.

IEC 60050-191. 1990. International electrotechnical vocabulary – Part 191: Dependability and quality of service.

IEC 61508-4. 2010 Functional safety of electrical/electronic / programmable electronic safety related systems – Part 4: Definitions and abbreviations.

ISO 14224. 2015. Petroleum, petrochemical and natural gas industries. Collection and exchange of reliability and maintenance data for equipment.

ISO/TR 12489. 2013. Petroleum, petrochemical and natural gas industries — Reliability modelling and calculation of safety systems.

ISO 31000. 2010. Risk management.

Leveson, N.G. 2011. *Engineering a safer world: Systems Thinking Applied to Safety*. Mit Press.

Leveson, N.G. 2013. An STPA Primer, version 1, August 2013.

Rausand, M. & Øien, K. 1996. The basic concepts of failure analysis. *Reliability Engineering and System Safety*, 53:73–83.

Society for Risk Analysis, "SRA Glossary". 2015. Available: http://sra.org/sites/default/files/pdf/SRA_glossary_20150622.pdf [Accessed 15 December 2017].

Subburaman, K. 2010. A Modified FMEA Approach to Enhance Reliability of Lean Systems. Master's Thesis, University of Tennesee," http://trace.tennessee.edu/utk_gradthes/664. [Accessed 15 December 2017].

Teknikens Värld. 2017. Dagen då Mercedes A-klass slog runt. Available: http://teknikensvarld.se/dagen-da-mercedes-a-klass-slog-runt-550219/ [Accessed 15 December 2017].

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Imprecise reliability analysis of complex interconnected networks

J. Behrensdorf & M. Broggi
Institute for Risk and Reliability, Leibniz Universitt Hannover, Hannover, Germany

M. Beer
Institute for Risk and Reliability, Leibniz Universitt Hannover, Hannover, Germany
Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK
International Joint Research Center for Engineering Reliability and Stochastic Mechanics (ERSM), Tongji University, Shanghai, China

ABSTRACT: The effect of natural and man made disasters on critical infrastructures are substantial, as evident from recent history. Break downs of critical systems such as electrical power grids, water supply networks, communication networks or transportation can have dire consequences on the availability of aid in such a crisis. That is why, reliability analyses of these networks are of paramount importance. Two important factors must taken into consideration during reliability analysis. First, the networks are subject to complex interdependencies and must not be treated as individual units. Second, the reliability analysis is typically based on some form of data and or expert knowledge. However, this information is rarely precise or even available. Therefore, it is important to account for different kinds of uncertainties, namely aleatory uncertainty and epistemic uncertainty. Aleatory uncertainty represents the natural randomness in a process, while epistemic uncertainty represents vaguness or lack of knowledge in the model. In this work we present an approach to the numerical reliability analysis of complex networks and systems extending a previously developed method based on Monte Carlo simulation and survival signature. The extended method treats both kinds of uncertainties, thus, yielding better results. We show how Monte Carlo simulation controls aleatory uncertainty and apply sets of distributions (probability boxes) to treat epistemic uncertainties in component failures. In this framework, dependencies are modelled using copulas. Copulas possess the unique property of decoupling the odelling of the univariate margins from the modelling of the dependence structure for continuous multivariate distributions. Analoguous to the p-boxes we use sets of copulas to include imprecision in the dependencies. Finally, the method is applied to an example system of coupled networks.

1 INTRODUCTION

Modern infrastructure systems are highly complex and subject to a multitude of different dependencies. Disasters in recent years have shown how critical the impact of these dependencies can be. Failures in one network such as a power outage will surely impact other dependent systems. In worst case scenarios these dependencies can lead to cascading effect ultimately breaking down entire networks (Buldyrev et al. 2010). This highlights the need for methods of reliability analysis that can deal with these complexities.

Recently, the survival signature (Coolen and Coolen-Maturi 2013) has gained in popularity as a tool to aid with this task. The survival signature allows to decouple the structural evaluation from the probabilistic analysis, allowing for highly efficient simulation. In Behrensdorf et al. 2017 we introduced a method for the reliability analysis of complex interdependent networks. Other efficient algorithms can be found for example in Patelli et al. 2017.

A secondary task during the reliability analysis is the accurate modelling of component failures and dependencies. Typically, this is done based on data or expert assessments. However, both are subject to two kinds of imprecisions, namely, aleatory and epistemic uncertainty (Beer et al. 2013). Aleatory uncertainty represents the randomness inherent in a process, such as component degradation and external forces affecting the system (natural hazards, earthquakes, etc.), while epistemic uncertainty describes the uncertainty in the model due to a lack of or vagueness of knowledge about the system. The latter is usually regarded as reducible through acquiring of additional data and information.

In this work we expand our previously developed technique by inclusion of imprecision. The method is based on Monte Carlo simulation and as such already deals with aleatory uncertainty. In this extension the modelling of component failures is refined by applying probability-boxes (p-boxes) to account for epistemic uncertainty. Feng et al. 2016 have shown the advantages of using p-boxes

in reliability analysis. Additionally, the rather simple dependency modelling of the initially developed method is replaced by imprecise copulas (Montes et al. 2015). Copulas split continuous multivariate distributions in a dependence structure and univariate marginals, which in turn allows for separate flexible modelling of the two (Joe 2014).

This paper is outlined as follows. First we introduce the previously developed method for the reliability analysis of networks. Then, after presenting basic theory and notation on copulas, we discuss how to model dependencies with copulas and how to translate these methodologies into an imprecise setting. Finally, we apply the developed techniques to a simple example. The paper closes with some concluding remarks and an insight into future work.

2 RELIABILITY ANALYSIS

This section presents the survival signature based method to calculate the reliability of a system, as first presented in Behrensdorf et al. 2017.

The survival signature was developed as an extension to the system signature (Samaniego 2007), overcoming the limitations that restrict the system signature to systems of one single component type (Coolen and Coolen-Maturi 2013). The main function of the survival signature is to separate the structural information of a network from its probabilistic characteristics.

2.1 *Survival signature*

Considering a system with m components, the survival signature for l out of m components working is defined as

$$\Phi(l) = \binom{m}{l}^{-1} \sum_{\underline{x} \in S_l} \varphi(\underline{x}), \tag{1}$$

where $\underline{x} = (x_1, \ldots, x_m)$ denotes the state vector of the system with $x_i = 1$ and $x_i = 0$ representing a working or failed component respectively and $\varphi(\underline{x})$ is the structure function returning the state of the full system with $\varphi(\underline{x}) = 1$ indicating a working system and $\varphi(\underline{x}) = 0$ indicating a failed system.

Extending the survival signature to systems with K component types and m_k components for each type $k (k = 1, \ldots, K)$ and l_k out of m_k components working results in

$$\Phi(l_1,\ldots,l_k) = \left[\prod_{k=1}^{K} \binom{m_k}{l_k}^{-1} \right] \times \sum_{\underline{x} \in S_{l_1,\ldots,l_k}} \varphi(\underline{x}). \tag{2}$$

An efficient algorithm to compute the survival signature can be found in Aslett 2012.

2.2 *Survival function*

The next step in calculating the reliability of a system is the definition of the survival function. This function uses the survival signature to calculate the probability that a system is working at time t and as such calculates the reliability. The survival function is defined as

$$P(T_s > t) = \sum_{l_1=0}^{m_1} \cdots \sum_{l_k=0}^{m_k} \Phi(l_1,\ldots,l_k) \cdot P\left(\bigcap_{k=1}^{K} \{ C_t^k = l_k \} \right). \tag{3}$$

Note especially the separation of structural information (left) and probabilistic information (right). This means, that the structural evaluation of the system must occur only once for the entire reliability analysis. In the last remaining step, the probabilistic part of the survival function is approximated using Monte Carlo Simulation in order to be able to include imprecisions and interdependencies in the analysis.

2.3 *Monte Carlo simulation*

The simulation starts by selecting a sufficient number of samples N_{MC} and small time step followed by the sampling of component failure times from the assumed copula (see section 3). Next, in two nested loops over all combinations $l_1, \ldots, l_k$ where $\Phi(l_1,\ldots,l_k) > 0$ and all time steps t the number of samples in the same configuration are counted as $N_{l_1,\ldots,l_k}(t)$. Then, the probabilistic part of the survival function is approximated by

$$P\left(\bigcap_{k=1}^{K} \{ C_t^k = l_k \} \right) \approx \frac{N_{l_1,\ldots,l_k}(t)}{N_{MC}} \tag{4}$$

Finally, the partial reliabilities obtained in the previous step are multiplied by their probability from the survival signature and summed up, yielding the full reliability of the network.

3 MODELLING DEPENDENCIES

This section introduces the necessary notation of copulas and how to apply them to model dependencies in and between networks. In more detail, section 3.4 presents how to use copulas to model common causes of failure while section 3.5 shows how to model interdependencies. This is a very basic introduction, for a thorough discussion of copulas see Nelsen 2006.

3.1 *Copulas*

The basis of copulas is built by what is today known as Sklar's theorem (Sklar 1959). The theorem states that any multivariate distribution H (in dimensions $d \geq 2$) can be separated into its univariate marginal distributions F_i and a copula function $C:[0,1]^d \rightarrow [0,1]$.

Theorem 1. Sklar's theorem *Let H be a d-dimensional distribution function with margins $F_1, \ldots, F_d$. There exists an n-dimensional copula C such that for all* $\mathbf{x}$ in $\mathbb{R}^d$

$$H(x) = C\left(F_1(x_1), \ldots, F_n(x_d)\right). \tag{5}$$

If the marginals $F_1, \ldots, F_d$ are continuous, then C is unique; otherwise, C is unique on Range$(F_1) \times \ldots \times$ Range(F_d).

Conversely, if C is a d-copula and $F_1, \ldots, F_d$ are distribution functions, then the function H defined by Eq. 5 is an d-dimensional distribution function with margins $F_1, \ldots, F_d$

This facilitates separate modelling of the marginal distributions from modelling of the dependence structure, in turn allowing for effective treatment of imprecisions in both parts (see section 4).

In this work we apply three distinct copula families, namely the Gaussian copula, the Independence copula, and the Clayton copula. For an encompassing discussion of copula families the reader is referred to Nelsen 2006 and Joe 2014.

The d-dimensional Gaussian copula is defined as

$$C_R(u_1, \ldots, u_d) = \Phi_d\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right), \tag{6}$$

where $\mathbf{R} \in [-1,1]^{d \times d}$ is a positive definite correlation matrix and $\Phi_d(\cdot; \mathbf{R})$ is the d-variate cumulative distribution of a $\mathbb{N}_d(0, \mathbf{R})$ random vector. Φ^{-1} denotes the inverse of the univariate standard Gaussian cdf (Joe 2014).

The latter two copulas belong to the class of Archimedean copulas. This family is particularly popular due to their easy construction and wide range of applications (Nelsen 2006). The Archimedean copulas used in this work are one parameter families, which allow for easy treatment of imprecision as seen in the subsequent section. Any d-dimensional Archimedean copula is constructed using a so called *generator* function $\varphi:[0,\infty] \rightarrow [0,1]$ and its inverse φ^{-1} according to

$$C_\varphi(u_1, \ldots, u_d) := \varphi\left(\varphi^{-1}(u_1) + \ldots + \varphi^{-1}(u_d)\right), \tag{7}$$

where $u_1, \ldots u_d \in [0,1]$ (Mai and Scherer 2012). Table 7 shows the generators and parameter ranges for the Independence and Clayton copula families.

3.2 *Dependence measure*

Studies have shown, that correlation is not a suitable measurement of dependence for copulas (Schirma Schirmacher and Schirmacher 2008). Therefore, in this work, Kendall's tau is selected as the preferred dependence measure. Kendall's tau is based on *concordance*. A pair of random variables is said to be concordant if "large" values are associated with "large" values "small" values with "small". Formally, two observations (x_i, y_i) and (x_j, y_j) from a vector $(X,Y)(X,Y)$ are *concordant* if $(x_i - x_j)(y_i - y_j) > 0$ and discordant if $(x_i - x_j)(y_i - y_j) < 0$.

Then, Kendall's tau for a sample of n observations $\{(x_i, y_i), \ldots, (x_n, y_n)\}$ from a vector of continuous random variables (X,Y) can be defined as

$$t = \frac{c-d}{c+d} = (c-d) \Big/ \binom{n}{2}, \tag{8}$$

where c and d represent the number of concordant and discordant pairs among all possible pairs of observations. This value may also be interpreted as the probability of concordance minus the probability of discordance for a random pair of observations (x_i, y_i) and (x_j, y_j). Based on this fact, Kendall's tau for the random variables X and Y can be defined by

$$\begin{aligned} \tau(X,Y) = {} & P[(X - \tilde{X})(Y - \tilde{Y}) > 0] \\ & - P[(X - \tilde{X})(Y - \tilde{Y}) < 0], \end{aligned} \tag{9}$$

where $(\tilde{X}, \tilde{Y})$ is an independent copy of (X,Y) (Schirmacher and Schirmacher 2008).

Kendall's tau is not only used to measure dependence but also to find copula parameters representing a desired strength of dependence.

3.3 *Vine copulas*

In order to analyse the reliability of complex networks it one must build hight dimensional copulas. However, in comparison to bivariate copulas, the available literature on multivariate copulas is scarce (Mai and Scherer 2012). For this reason, we employ a technique called *pair copula construction* to break down multivariate copulas into combinations of bivariate copulas. More accurately, we use vine copulas as a graphical tool to model pair cop-

Table 1. Generators, generator inverses and parameter ranges for the Clayton and Independence copula.

Name	Generator $\varphi_\theta(\mathbf{t})$	Generator Inverse $\varphi_\theta^{-1}(\mathbf{t})$	Parameter θ
Clayton	$\frac{1}{\theta}(t^{-\theta} - 1)$	$(1+\theta t)^{-1/\theta}$	$\theta \in [-1,\infty) \setminus \{0\}$
Independence	$-\log(t)$	$\exp(-t)$	

ula constructions as sets of trees. An example of a five-dimensional vine copula is shown in Fig. 1.

A regular vine (R-Vine) $\mathcal{V}$ is defined as a set of trees $T_1,\ldots,T_{d-1}$, where T_1 consists of nodes $N_1 = \{1, \quad ,d\}$ and edges E_1. Every subsequent Tree T_j uses the edges E_{j-1} as nodes and connects them with the edges E_j. The last property needed to define a regular vine is the *proximity property*, stating that if a and b are connected by an edge in $T_j, j \geq 2$, then a and b must share a common node in T_{j-1}.

A plethora of different structures exist for a d-dimensional R-Vine based on the definition. Because of this and the fact, that a regular vine possesses 2^{n-1} sampling order we apply D-vines instead, which have a much more restrictive structure. As seen in Fig. 2, a D-Vine is characterized by each node $n \in N_i$ having a maximum degree of 2. Sampling from D-Vines is a lot simpler and in this work performed by using the MATLAB toolbox VineCopulaMatlab (Kurz 2016).

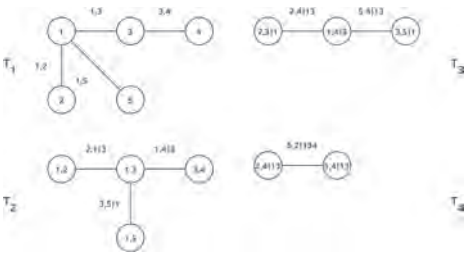


Figure 1. Five-dimensional copula represented as a regular vine.

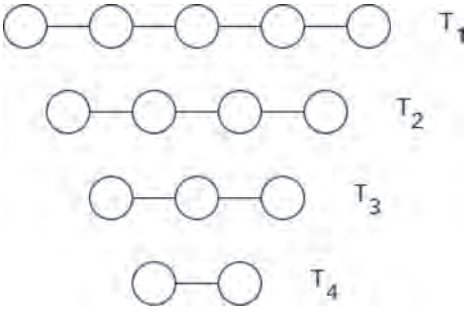


Figure 2. Structure of a five-dimensional D-Vine.

3.4 *Common cause of failure*

One type of dependent failure tackled in this work is common cause of failure. It is defined as two or more components failing at the same time due to common defects or weaknesses. Causes include but are not limited to: errors in manufacturing, errors during maintenance or operation, and environmental causes such as earthquakes or tsunamis (Hanks 1998). We model common cause of failure by application of Clayton copulas. This family possesses a property called lower-tail dependence, meaning that dependence is stronger in the lower-left quadrant of $[0,1]^2$. By modelling the failures this way, the dependence is much stronger in early component life and as such brings us closer to the traditional bathtub shape or component failure probabilities. Figure 3 shows samples drawn from a Clayton copula. Note, how the stronger dependence in the lower-tail is clearly visible in the scatter plot.

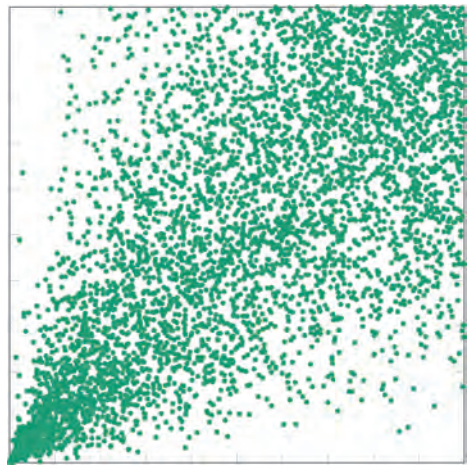


Figure 3. Samples drawn from a Clayton copula with the parameter chosen so $\tau = 0.5$.

3.5 *Interdependencies*

Interdependencies between nodes and networks are handled by application of Gaussian Copulas. Gaussian copulas possess no tail-dependence and show good results, although other families could be investigated for the same application in the future. In addition to the copula there is one more step required to accurately model the dependencies. We understand interdependencies as the phenomenon of one component failing due to the failure of another. As such, interdependencies imply causality. In order to represent this causality in the model, dependence is introduced in the marginals. During the transformation of the failure times sampled from the copula by the inverse transformation method, the marginal distributions are aggregated from the dependent marginals using Kendall's tau of the random variables u_1 and u_2 as

$$U_1 = (1-\tau)\cdot F_1^{-1}(u_1) + \tau \cdot F_2^{-1}(u_1), \tag{10}$$

where F_1 and F_2 are the marginals of a copula C.

4 HANDLING IMPRECISION

Two types of uncertainties must be taken care of during the reliability analysis, namely, aleatory and epistemic uncertainties. Aleatory uncertainty describes the natural randomness inherent in a process, while epistemic uncertainty represents the uncertainty due to vagueness in information or a lack thereof.

Aleatory uncertainty can automatically handled by our reliability analysis technique. Through assuming failure time distributions for the component failures and sampling these during Monte Carlo simulation, the randomness that our model is subject to is fully included. However, the selection appropriate failure time distributions is typically based on either data or expert knowledge, neither of which yield perfect results, in turn introducing epistemic uncertainty into the model. This uncertainty can be reduced by using *probability-boxes* (p-boxes) (Feng et al. 2016).

P-boxes are defined as bounds on the cumulative distribution function of a random variable. The left and right bounds can be found by for example selecting an appropriate distribution and giving the parameters as intervals. As such, a p-box comprises boththe aleatory and the epistemic uncertainty. An example of an exponential p-box with parameters $\lambda \in [1.2,2.2]$ is shown in Fig. 4.

By feeding the bounds of the p-box into the reliability analysis, the epistemic uncertainty propagates into the result. Thus, instead of one survival function, we obtain an upper and lower bound. Figure 5 shows an example of theupper and lower bounds obtained by performing a reliability analysis of a simple system of two parallel components of the same type, assuming the p-box of Fig. 4 for the failure time distributions.

Similarly to the application of p-boxes to handle epistemic uncertainty in the marginals, we can define the copula parameters as intervals and obtain *imprecise copulas* for the dependencies (Montes et al. 2015). This works especially well since all copula families, including the bivariate Gaussian copula, we apply in the vine copula are defined by a single parameter.

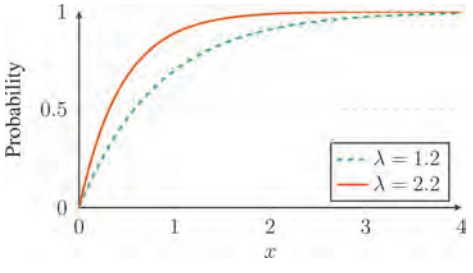


Figure 4. Example of an exponential p-box with $\lambda \in [1.2,2.2]$.

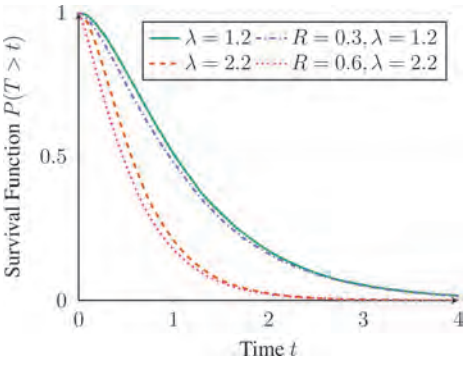


Figure 5. Upper and lower bounds of the reliability resulting from applying the p-box in Fig. 4 to a simple system of two parallel components.

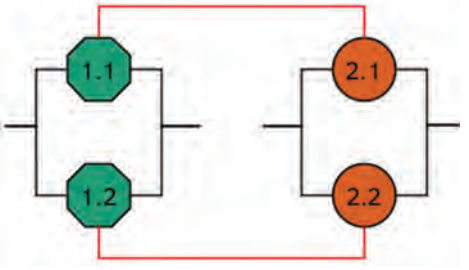


Figure 6. Structure of the example network. The red lines represent interdependencies between the two subsystems.

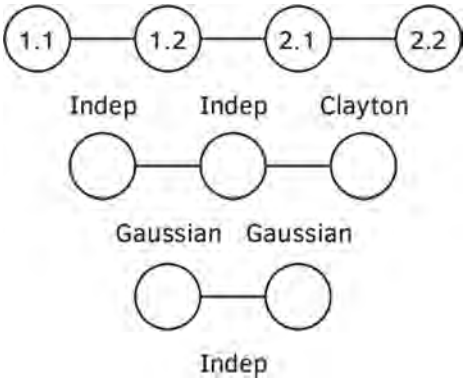


Figure 7. D-Vine used to model the common cause of failure in system 2 and the interdependencies between the two networks.

Finally, the bounds of the p-boxes as well as the bounds of the copula parameters are fed into the reliability analysis. Returning to the simple system of two parallel components and linking the components with an imprecise Gaussian copula $R \in [0.3, 0.6]$ results in the upper and lower bounds for the reliability as seen in Fig. 5.

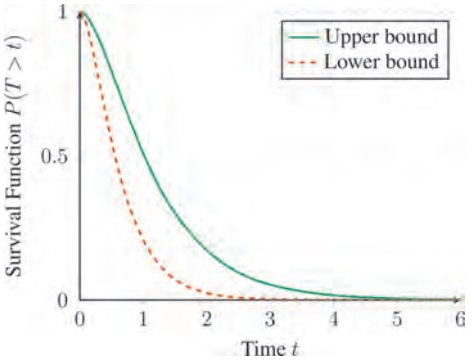


Figure 8. Bounds on the reliability of system 2 based on the assumed imprecisions.

5 NUMERICAL EXAMPLE

The methods that were introduced in the previous sections will now be applied to a simple toy example. All marginals and dependencies will be considered as imprecise in order to account for all aleatory and epistemic uncertainties. Figure 6 presents the example system build from two systems of two parallel components where the first and second components respectively are interconnected.

A four-dimension D-Vine copula, including the interdependencies and a common cause of failure shared among the components in system 2, is built in order to sample the component failure times. The structure of the vine is shown in Fig. 7.

The component failure times for systems 1 and 2 are assumed to be exponentially distributed with $\lambda_1 \in [1.5,1.7]$ and $\lambda_2 \in [0.7,1.1]$. The copula parameters are chosen such that $\tau \in [0.2,0.4]$ for the Clayton copula and $\tau \in [0.4,0.6]$ for the Gaussian copulas. The resulting bounds on the reliability of system 2 are plotted in Fig. 8.

6 CONCLUSION AND OUTLOOK

In this work we have presented how to perform reliability analyses of complex interdependent networks in a highly imprecise setting. The necessary theory on copulas and vine-copulas and applied to the modelling of dependencies in and between networks. Finally, the modelling of dependencies and a previously introduced method for the reliability analysis of networks were extended to account for both aleatory and epistemic uncertainties. As a result, we obtained bounds on the network reliability. The method was applied to a numerical toy example to prove the functionality.

It is obvious that this paper only serves as a short introduction into future work. The methods must be validated further and applied to complex real world networks in order to ensure usability. Especially the construction of the D-Vine copula for sampling of the component failure times must be further investigated. There exist a plethora of possible vine structures for a given problem and effective automatic construction techniques have to be created.

REFERENCES

Aslett, L. (2012). Reliabilitytheorey: Tools for structural reliability analysis, r package. http://www.louisaslett.com. Accessed: 11.12.2017.

Beer, M., S. Ferson,&V. Kreinovich (2013). Imprecise probabilities in engineering analyses. *Mechanical systems and signal processing 37*(1), 4–29.

Behrensdorf, J., M. Broggi, S. Brandt, & M. Beer (2017). Numerically efficient reliabilty analysis of interdependent networks. In *The 27th European Safety and Reliability Conference*.

Buldyrev, S. V., R. Parshani, G. Paul, H. E. Stanley, & S. Havlin (2010). Catastrophic cascade of failures in interdependent networks. *Nature 464*(7291), 1025–1028.

Coolen, F. P. & T. Coolen-Maturi (2013). Generalizing the signature to systems with multiple types of components. In *Complex systems and dependability*, pp. 115–130. Springer.

Feng, G., E. Patelli, M. Beer, & F. P. Coolen (2016). Imprecise system reliability and component importance based on survival signature. *Reliability Engineering & System Safety 150*, 116–125.

Hanks, B. (1998). An appreciation of common cause failures in reliability. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering 212*(1), 31–35.

Joe, H. (2014). *Dependence modeling with copulas.* CRC Press.

Kurz, M. (2016). Vinecopulamatlab toolbox. http: // maltekurz.github.io/VineCopulaMatlab/. Accessed: 11.12.2017.

Mai, J.F. & M. Scherer (2012). *Simulating copulas: stochastic models, sampling algorithms and applications.* World Scientific.

Montes, I., E. Miranda, R. Pelessoni, & P. Vicig (2015). Sklar's theorem in an imprecise setting. *Fuzzy Sets and Systems 278*, 48–66.

Nelsen, R. B. (2006). *An Introduction to Copulas.* Springer Science & Business Media.

Patelli, E., G. Feng, F. P. Coolen, & T. Coolen-Maturi (2017). Simulation methods for system reliability using the survival signature. *Reliability Engineering & System Safety 167*, 327–337.

Samaniego, F. J. (2007). System signatures and their applications in engineering reliability. *110.*

Schirmacher, D. & E. Schirmacher (2008). Multivariate dependence modeling using pair-copulas. Technical report, Citeseer.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris 8*, 229–231.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Communication failure analysis for a fleet formation flight of drones based on absorbing Markov chain

R. Abdallah
FEMTO-ST Institute, Université Bourgogne Franche-Comté, France
Université de Technologie de Belfort-Montbéliard, Belfort, France

C. Sarraf
Holy Spirit University of Kaslik, Kaslik, USEK, Lebanon

R. Kouta
Laboratory of Systems Modelling and Dependability (LM2S), Université de Technologie de Troyes, Troyes, France

J. Gaber & M. Wack
FEMTO-ST Institute, Université Bourgogne Franche-Comté, France
Université de Technologie de Belfort-Montbéliard, Belfort, France

ABSTRACT: Several unmanned aerial vehicles (UAVs) flight in a formation fleet are used to improve the effectiveness of civilian missions such as firefighting, searching, rescuing, etc., as well as the success of military applications. The increase use of these cooperative systems in hazardous environments makes the reliability improvement essential in order to prevent any catastrophic event. In this article, we aim to ensure a successful communication between the drones from one side, and between the drones and the ground station from the other side. We propose to identify, the different fault states and their probabilities during a communication. An Absorbing Markov Analysis approach is developed for these states. This framework can be used to find the riskiest scenarios and elements that need to be addressed in order to improve the reliability.

1 INTRODUCTION

Nowadays, the interest of using the unmanned aerial vehicles (UAVs), recognized as drones, has increased due to their use in several civilian applications such as searching, border surveillance, natural disaster monitoring, and firefighting, etc (Rabbath & Léchevin, 2010). They are known for 3D missions that are 'dirty, dull or dangerous' (Hattenberger, 2008). Lately, the focus is shifted toward the cooperative UAV fleet formation due to their mission in a large hazardous environment, since a single UAV has a limited energy and payload. The multi UAV system needs to ensure several properties such as robustness, cooperativeness and scalability. These proprieties can be attained by assuring the navigation of each UAV, the control of the whole fleet as well as constant and reliable communication between the drones on one side and between the drones and the ground station control (GSC) on the other side. Their different size and payloads, their flight times, the distance between two UAVs and the communication ranges are the causes that affect the overall performance of the fleet formation flight. The essential role of these cooperative systems reflects the importance of enhancing the reliability in order to avoid the failure of the communication between the aerial vehicles. Coordination between them should always be guaranteed despite the uncertainties of the environment, the network and simple failure in the hardware of a vehicle. In hence, the detection of the anomalous aircraft prevents the collisions between the aerial vehicles and the degradation of the team performance. The information flow between UAVs can be collected by an entity on ground, Ground Station Control (GSC), that controls the mission and makes decisions for the aircrafts; or alternatively, they share the information between them and make collective decisions.

Considering the importance of the reliability of the communication system, this paper presents, based on Markov model, a novel approach to evaluate it. The proposed framework takes in consideration the internal elements, both hardware and software, of the system as well as the surrounding environment.

The paper is organized as follows. Section 2 presents the related work on the reliability of the aerial vehicles focusing on the Markov chain. Section 3 provides a description of the proposed model of state diagram for the communication failure between UAVs. Section 4 gives a brief description of the proposed framework. The conclusion is attributed in Section 5.

2 RELATED WORK

Since the UAV's accidents and failure rates are higher than the manned aircraft, the reliability analysis of these systems presents an important focus for the researchers. The fault-tolerant system and the redundancy hardware do not always represent the efficient solution for this formation fleet flight due to incurred costs and weight. Different methods like the Fault Tree Analysis (FTA) (Abdallah, Kouta, Sarraf, Gaber, & Wack, December 2017), Failure Modes and Effects Analysis (FMEA) has been used to improve the reliability of the helicopters. In some cases, various FTA are needed to represent the different failure conditions of a complex system.

The evaluation of reliability of a system considers the state-space models, such as Markov Chain (Frattini, Bovenzi, Alonso, & Trivedi, 2010), that handle the failure/repair of its components and surrounding elements that might impact the reliability model. The Markov chain defines the derogation states of operation, where the functions are not all performed or where the state functions are absolutely stopped. In (Kitchin, 1988), distinct techniques used for establishing Markov models for the reliability of systems are provided emphasizing on the exponential model. It is devised to detect the failure and the method to recover it.

The reliability of the flight computer system (FCS) components including the flight computer and the navigation system is discussed in (Pashchuk, Salnyk, & Volochiy). It enquires a fault tolerant model considering two cases: the case where no additional standby microprocessors are implemented and the case of inherent standby microprocessor. A mathematical model based on Markov chain is applied to improve the reliability for the FCS components. An explanation of the Markov chain and Markov process is given in (Fuqua, 2003). It clarifies the powerful relation between the Markov chain and the reliability, maintainability and safety engineering (RMS) insisting of the International Standards that deal with this approach such as IEC 61165 and IEC 61508 that estimate the probability of failure of a critical system.

The issue of packet dropout for the drones' communications via wireless is investigated in (Zhou, Li, Lamont, & Rabbath, 2012). The authors proposed a two state Markov model in order to model the wireless channels taking into consideration the impacts of the Ricean fading. Their computer simulations are better than those of the most known models for wireless channels, the Gilbert-Elliott model (Gilbert, 1960), (Elliott, 1963), since their approach simulate the non-stationary errors.

A distributed computing system (DCS) is multiple processors that are interconnected via a network. In DCSs, the information is spread out among the nodes that consist of the data files, the processing elements, the shared resources and programs. In order to ensure the exchange of the information and the control of the data, the reliability of this system is important to be studied. It focuses on the analysis of the distributed program reliability (DPR) and the distributed system reliability (DSR). (Wang, 2004) suggested two reliability stochastic measures for these distributed systems: Markov-chain distributed program reliability (MDPR) and Markov-chain distributed system reliability (MDSR). The article describes the employment of one absorbing state for this problem and the probability of transition between the states. An Adaptive Markov Model Analysis (AMMA) is proposed in order to isolate the faults in the critical components. This proposed approach serves to make better the robustness and the availability of the UAV autopilot by incorporating the Fault Detection Isolation (FDI) approach (Krishnaprasad, Nanda, & Jayanthi, 2016). In (Kumar & Jackson, 2009), the paper discusses the reliability models based on the stochastic approach of Markov analysis, merged with the probabilistic approach of Weibull distribution in order to approximate the failure attributes of wear out components. This method is used since the components with wear out failure are characterized with variable failure rates depending on the operation time of the components. For this issue, a state transition diagram for six components optical telescope calibration system (OTCS) is shown. The partially observable Markov decision processes (POMDPs) is used in (Ragi & Chong, 2013) in order to determine a path planning for the UAVs to track different targets. The failure analysis of the flight control system of Air Force Institute of Technology (AFIT) UAV based on Markov analysis is elaborated in (Okafor & Eze, 2016). It shows the failure states and the probability of being in these states.

3 PROPOSED APPROACH

An Absorbing Markov Chain (AMC), where there is at least one absorbing state, is considered. An

absorbing state is characterized by the fact of once it is entered, it cannot be left. Each state in the transition diagram can be taken as an absorbing state. The transition between states can have multiple steps in order to attain the absorbing state. Two important variables should be calculated: the mean time t_{mean} in addition to the length of the path until the state is absorbed. We aim to evaluate the probability of being in each transient state leading to the absorbing state. Transitions between states are based on the probabilities that are function of the failure rates, of internal components as well as the occurrences of related events within the surrounding environments.

The main focus is to maintain a communication between the drones although all the uncertainties that can occur. We propose an Absorbing Markov Chain to model the problem and show the transition between the events that affect the communication.

First, the exchange of information and the communication is considered in a normal state. However, several causes can affect this state. The causes can be divided in internal causes at the level of the software and hardware failures and the external causes that are related to the human and the environment.

3.1 *Hardware failure*

The hardware failure can attack the engine, the power, the propellers and the antenna of transmission and reception. The issue is that during a flight, a hardware failure cannot be repairable and lead to an absorbing state of communication failure (Figure 1).

Figure 1 shows the causes of the hardware failure of a drone in addition to the transition between the transient states. The antenna failure can directly lead to a communication failure. The drones cannot send and receive anymore the information between them or to/from their ground station control. Moving to the power failure, it can be caused from the ventilation default and the disruption of the cables that induce an overheating of the drone and consequently a power failure. It is also attributed to an overcurrent/undercurrent, physical damage, overheating or exhaustion of the battery. The loss of the UAV transceiver affects the servomotor which its failure involves the actuator default and in hence the engine failure. So on, the

Table 1. Failure rates of the hardware events.

Events	Minimum Failure Rate (λ_{min}) ($\times 10^{-6}$)	Maximum Failure rate (λ_{max}) ($\times 10^{-6}$)	Mean Failure Rate (λ_{mean}) ($\times 10^{-6}$)
Loss of UAV transceiver ($\lambda_{ns\text{-}lt}$)	7,285E-01	3,740E+00	9,635E-01
Servomotor default ($\lambda_{ns\text{-}sd}$)	3,575E-01	3,952E+00	2,961E+00
Actuator default ($\lambda_{ns\text{-}ad}$)	1,218E-01	8,569E-01	3,444E-01
Disruption of cables ($\lambda_{ns\text{-}dc}$)	4,000E-05	4.087E+00	4.800E-05
Ventilation Default ($\lambda_{ns\text{-}vd}$)	3,483E-01	4.087E+00	2,891E+00
Overheating ($\lambda_{ns\text{-}oh}$)	12,617E+00	778,2E+00	44,421E+00
Power Failure ($\lambda_{ns\text{-}pf}$)	2,352E+00	9,104E+00	4,813E+00
Default assembly of propellers ($\lambda_{ns\text{-}da}$)	8,992E+01	6,296E+02	2,07E+02
Antenna failure ($\lambda_{ns\text{-}af}$)	3.662E+00	11.886E+00	20.467E+00

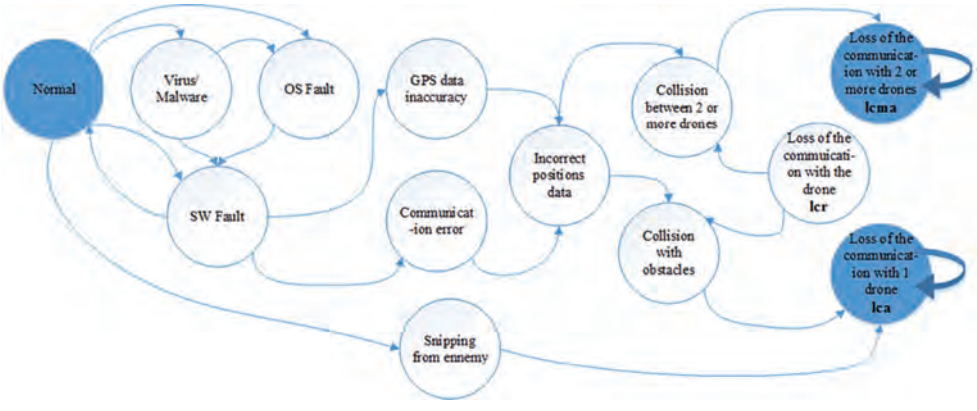


Figure 1. Hardware failure.

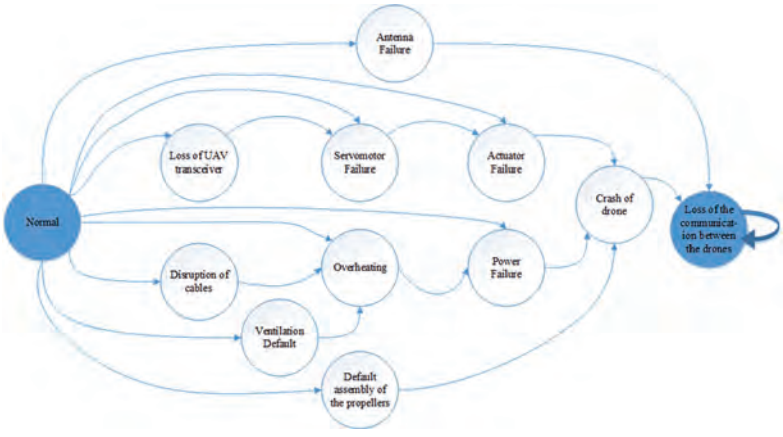


Figure 2. Software failure and collision events.

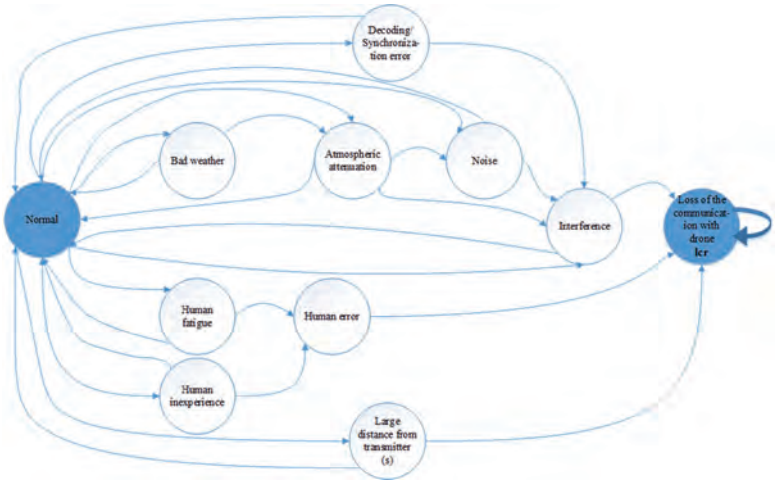


Figure 3. Exterior factors events.

engine failure, the power failure and the default assembly of the propellers can lead to the damage of a drone. The crash of the leader of the fleet formation flight is the most risky case since the leader controls the exchange of information between the fleet. Since the crash of a drone cannot be repairable, it leads to an absorbing communication failure between the drones. The failure rates of these events are known from the Nonelectronic Parts Reliability Data Publication (NPRD-2016) database.

Table 1 exhibits the failure rates of these hardware events considering the previous state as normal state ns. The annotation of the failure rates is λ from the previous status of transition to the next status of transition. The database gives only the failure rates from the normal states. The other transition failure rates are not known.

3.2 *Software failure and collision events*

The normal situation can be affected by a software failure bringing a disruption to the communication between the drones (Figure 2). An infected virus or malware represents an important reason for a mal-functioning of the UAV. It disturbs the

operating system OS of the drone in a manner that the two causes engender a software fault. The virus/malware and the OS fault can also attack the ground station control (GSC). The reliability of the software is ensured by regular updates of the operating system in addition to a good antivirus. The software faults affect the GPS data leading to a communication error between the drones or between one drone and the base station in a manner that the communication is not lost but there is an error in exchanging the information. The GPS data inaccuracy permits a confusion of the position of other's drones that influences the coordination of the fleet formation flight.

The wrong positions' data received from other drone might lead to collision between two or more UAVs or even to collision of the drones with an obstacle such as a building, trees, birds, etc. From the one hand, the collision between two drones leads to an absorbing state that cannot be avoided and repairable, the loss of communication between two or more drones *lcma.* On the other hand, the collision with obstacles in addition to the snipping of a drone from an enemy cause the loss of communication with only one drone *lca*, since it will not be presented in the fleet. This event is also an absorbing state.

3.3 *Exterior factors*

Different exterior factors influence the communication of the fleet formation flight. It englobes the animals, the weather, the obstacles and the human. The bad weather is an important state to prevent it. It includes temperature, wind, clouds, rain, ice, thunderstorms and fog. The transmitted signals might be subjected to an atmospheric attenuation due to a bad weather or other environmental conditions. An attenuation involves an interference and a noise that contributes to a bad signal transmitted between the aerial vehicles. If the medium of communication is exposed to jamming, echoes and noise, then that might interfere with what is transmitted, affecting the overall communication. Synchronization and the decoding of the message is an essential mechanism in the transmission of the message that can be affected by the interference phenomenon.

The human plays an important role in the communication especially with the GSC. The exhaustion of the GSC operator and his lack of experience and qualification in flying a certain type of drones contribute to the human error. The distance between the source and the destination can influence the quality of the received information as when the distance increases, the transmitted signal can be exposed the attenuation and atmospheric noises.

In order to avoid these events, the operator should chose the typical environment for the fleet. He might take in consideration the weather, the season and the time of flight. However, although all these events lead to the loss of communication with the drone, but this state is not absorbing. It is repairable since we can change the environment, chose the right persons to flight the fleet, the interference could affect the signal for a certain time then the fleet continues its mission. Figure 3 resumes the external factors.

On the contrary of hardware failure, the loss of communication with the drone lcr is a repairable state. It could be caused by software faults or environmental effects. A software fault in the communication could be repaired through alternative channels or by the ground station. The environment effects can be controlled by making the drones flying in close distances or alternatively planning the mission in some other time with better weather conditions. Once the lcr occurs, it might attribute to the collisions between the drones or with an obstacle.

4 BRIEF DESCRIPTION OF THE STATES

Table 2. Brief description of the states.

ns	Normal State	This is the normal situation where the communication system, between drones and with GSC, is functioning normally.
vm	Virus or malware	The system has been infected by a virus or malware leading to mal-functioning and abnormal behavior
OSf	Operating System fault	The operating system of a drone or GSC is not properly functioning due to being infected by virus or malware or due to some internal fault or error
swf	Software fault	A fault in the software that handles GPS data or the communication system within the drone or the communication system within the GSC
gpsf	GPS data inaccuracy	GPS data of one or more drones are inaccurate
wpd	Incorrect positioning data	Wrong positions of one or more drones have been communicated to other drones and/or GSC

(*Continued*)

Table 2. (*Continued*).

cf	Communication error	No communication between 2 or more drones and/or between 1 or more drones and GSC
cd	Collision between drones	Collision between 2 or more drones have occurred
co	Collision with obstacle(s)	A drone has collided with obstacle(s)
se	Snipping from enemy	A drone or more have been shut down by enemies or third parties
bw	Bad weather	A bad weather that might have impacts on the communications between drones and/or between the drones and GSC
aa	Atmospheric attenuation	Transmitted signals might be subjected to attenuation due to bad weather or other environmental conditions
no	Noise	Transmitted signals might be subjected to noise
in	Interference	Transmitted signals might be subjected to interference
de	Decoding/synchronization errors	Transmitted data might be subjected to decoding and/or synchronization errors
ld	Large distance	One or more drones have flown away from transmitters of other drones and GSC
hf	Human fatigue	GSC operator is experiencing exhausted and tired
he	Human error	GSC operator(s) committed error(s), due to fatigue or lack of experience and qualification, that might affect the communication system
af	Antenna failure	Hardware fault affecting the transmitter and/or receiver antennas of one or more drones of GSC
lt	Loss of UAV transceiver	Loss of an electronic device that transmits and receives the signal
sd	Servomotor default	Hardware default of the motor that permits the control of the position, the acceleration and velocity.
ad	Actuator default	Hardware default of an electronic speed controllers that is linked to the engine, servomotors and propellers UAV actuators
dc	Disruption of cables	Internal incident that cut the cables
vd	Ventilation default	The cooling system is in failure
oh	Overheating	The temperature of the drone is high due to a disruption of cables or due to a default in the cooling system
pf	Power failure	Due to short-circuit, overcurrent/undercurrent, battery damage, overheating
da	Default assembly of propellers	Loss of more than two propellers
l1d	Loss of one drone	Loss of one drone due to collision with obstacles, snipped by enemies, and/or due to internal operation failure
lmd	Loss of 2 or more drones	Loss of 2 or more drones due to collision between them
lcr	Loss of communication with a drone	Loss of communication with a drone due to environmental conditions or software faults. This state is repairable and the system could go back to its normal state (*ns*)
lca	Loss of communication with a drone due to the loss of the drone	Loss of communication with a drone due to the loss of the drone itself caused by collision, snipping with enemies, environmental conditions and/or some internal faults. This state is not repairable and therefore it is an absorbing state.
lcma	Loss of communication with multiple drones	Loss of communication with multiple drones due to collision between them. This state is not repairable and therefore it is an absorbing state.

5 CONCLUSION

Different state diagram are presented in this paper showing the causes of loss of communication between the drones or between the drones and the ground station control. It includes the hardware failure, the software failure in addition to the external factors that affect transmitted signals. The software failure is a repairable state in addition to the external factors that we can prevent them by choosing the suitable environment, season and time. On the contrary, as they are not recoverable, hardware failures will lead to an absorbing state for the loss of communication. We aim to consider in our future works the failure rates of the external and software events, the failure rates of transitions in addition to the repairable rates taking in consideration the different strategies of fleet formation flight.

REFERENCES

Abdallah, R., Kouta, R., Sarraf, C., Gaber, J., & Wack, M. (December 2017). Fault Tree Analysis for the Communication of a Fleet Formation Flight of UAVs. *2017 2nd International Conference on System Reliability and Safety.* Milano: IEEE.

Elliott, E.O. (1963). Estimates of error rate for codes of burst-noise channels. *The Bell System Technical Journal, 42*(5), 1977–1997.

Frattini, F., Bovenzi, A., Alonso, J., & Trivedi, K. (2010). Reliability indices. *Wiley Encyclopedia of Operations Research and Management Science*.

Fuqua, N.B. (2003). The applicability of markov analysis methods to reliability, maintainability, and safety. *Selected Topic in Assurance Related Technologies (START), 2*(10), 1–8.

Gilbert, E.N. (1960). Capacity of a Burst-Noise Channel. *Bell Labs Technical Journal, 39*(5), 1253–1265.

Hattenberger, G. (2008). *Vol en Formation sans Formation: contrôle et planification pour le vol en formation des avions sans pilote*. Thèse de doctorat, Université Paul Sabatier-Toulouse III.

Kitchin, J.F. (1988). Practical Markov modeling for reliability analysis. *Reliability and Maintainability Symposium Proceedings* (pp. 290–296). IEEE.

Krishnaprasad, R., Nanda, M., & Jayanthi, J. (2016). Adaptive Markov Model Analysis for Improving the Design of Unmanned Aerial Vehicles Autopilot. *Intelligent Systems Technologies and Applications*, 259–271.

Kumar, R., & Jackson, A. (2009). Accurate reliability modeling using Markov Analysis with non-constant hazard rates. *Aerospace conference* (pp. 1–7). IEEE.

Okafor, E.G., & Eze, I.H. (2016, January). Failure analysis of a UAV flight control system using Markov Analysis. *Nigerian Journal of Technology, 35*(1), 167–173.

Pashchuk, Y., Salnyk, Y., & Volochiy, S. (n.d.). *Reliability Synthesis for UAV Flight Control System*.

Rabbath, C.A., & Léchevin, N. (2010). *Safety and reliability in cooperating unmanned aerial systems*. World Scientific.

Ragi, S., & Chong, E.K. (2013). UAV path planning in a dynamic environment via partially observable Markov decision process. *IEEE Transactions on Aerospace and Electronic Systems, 9*(4), 2397–2412.

Wang, J.-L. (2004). Markov-chain based reliability analysis for distributed systems. *Computers & Electrical Engineering, 30*(3), 183–205.

Zhou, Y., Li, J., Lamont, L., & Rabbath, C.-A. (2012). Modeling of packet dropout for UAV wireless communications. *International Conference on Computing, Networking and Communications Invited Position Paper Track* (pp. 677–682). IEEE.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Analyzing the reliability for connected vehicles using qualitative approaches and quantitative methods

A. Dabboussi
OMNI, CNRS Femto Lab, (UTBM), Université Bourgogne Franche-Comté (UBFC), Belfort, France

R. Kouta
Laboratoire de Modélisation et Sûreté des Systèmes (LM2S), Institut Charles Delaunay, Troyes, France

J. Gaber & M. Wack
OMNI, CNRS Femto Lab, (UTBM), Université Bourgogne Franche-Comté (UBFC), Belfort, France

Bachar EL Hassan & Lina Nachabeh
Electrical Department, Faculty of Engineering Branch 1, Lebanese University, Tripoli, Lebanon

ABSTRACT: Connected vehicles such as Vehicular Ad hoc Networks (VANET), a subset of Mobile Ad hoc Networks (MANETs), is a wireless communication technology applied to transportation, referring to a set of smart vehicles used on the road. These vehicles provide communication services among one another (V2V) or with Road Side Infrastructure (V2I). The main benefits of VANET are enhancing road safety, reducing energy use and emissions, and giving information services. Reliability is one of the most critical issues related to VANET since the information transmitted is distributed in an open access environment. We focused in this paper, on the reliability of VANET as a function of reliable hardware and their functionality taking into consideration the needed security equipment. Reliability Block Diagrams (RBD) and Fault Tree (FT) were used to analyze the reliability of the vehicles and the Road Side equipment (RSU). In order to prove our result, a simulation was occurred using the RBD and FT probability and it has been conducted to validate the proposed approach. The data (Failure ratio) used were from professional database concerning the type of components presented in the system. Our scientific approach was structured with methods that combine qualitative approaches (such as functional analysis, Failure Modes and Effects Analysis (FMEA),...) and quantitative methods (Fault tree, probabilistic models of degradation, etc.) for the VANET. From this data an exponential model of reliability was proposed. The probability calculation was performed in relation to a reference time of use. Thereafter a sensitivity analysis was suggested concerning the reliability parameters and redesign proposals are developed for the components.

1 INTRODUCTION

Each year 1.25 million people die worldwide as a result of road traffic accidents according to the WHO's Global status report on road safety 2015 [1]. Connected Vehicles such as vehicular ad hoc networks (VANET) and the autonomous vehicles are proposed as solutions to improve road safety [2].

On-Board Unit (OBU) computers give a smart vehicle the ability to communicate with other vehicles (V2V) and with intelligent Road Side Unit (RSU) infrastructure (V2I). In this context, wireless vehicular network technologies will allow significant reduction of vehicular accidents [2]. Since 90% of accidents are caused by humans, connected vehicle systems will help the driver to avoid these accidents and to divide the percentage accident ratio by ten [3].

These smart vehicles must comply with the safety integrity obligations. Indeed, safety is a major concern for the user of automotive vehicles. Reliability is one of the most critical issues related to the connected vehicle since the information transmitted is distributed in an open access environment and also due to the high mobility, dynamic topology, delay constraints, varying environments and different traffic patterns in VANET [4].

With the evolving of the Internet of Things, ensuring the dependability and the reliability of vehicular networks is becoming critical, and any lack of this requirement can lead to accidents and catastrophic results [5].

This paper presents a design of reliability block diagrams for: (1) the OBU based on Dedicated Short Range Communication (DSRC) Standard,

(2) the RSU and (3) the devices that are especially interesting for security purposes such as Trusted Platform Module (TPM). TPM is often mounted on vehicles to offer reliable storage (e.g. user credentials and keys) and to compute cryptography. TPM hardware is assumed to be tamper resistant so that hackers can't gain access, even with physical presence [6].

In VANET, the reliability of networks must be given special attention since the system aims at safeguarding road safety and accomplishing secured communication. Currently, in VANET, most of the work carried out relates to the evaluation of the communication performance and for routing protocols, without taking into consideration the dependability and the operational safety of the system. Network reliability is greatly reliant on the availability of hardware components and their life time cycle. Given that this problem is relatively new for applications in the mobile environment using ad-hoc networks, hence the development of methods and models to assess the reliability of smart vehicles by the approaches of reliability analysis that combine qualitative approaches (such as functional analysis) and quantitative methods such as reliability block diagrams (RBD), and fault trees (FT) for the VANET, has become indispensable. Evaluating these problems leads to a sensitivity analysis concerning reliability parameters in order to propose a redesign for the components that aims to increase the reliability and the availability for the whole system.

The paper is organized as follows: Section 2 presents the related work, section 3 describes the analytical approach and the hardware architecture of VANET: OBU, RSU, TPM. Using the reliability block diagrams and the fault tree we calculate the reliability of the system in section 4. Section 5 describes the simulation performed and the results. Finally, Section 6 concludes the paper with lessons learned from this work, and points out future research directions on VANET.

2 RELATED WORK

Generally, reliability is defined as the ability of a functional unit to perform a required function under given conditions for a given time interval [7].

Technically speaking, reliability is the probability that the VANET systems will work without failure during its running time under wireless environment operating conditions.

It is important to differentiate between availability and reliability concepts; reliability refers to failure-free operation during an interval, while availability refers to an operation free of failure at a given instant of time [8].

Many research works have been proposed in literature to define the reliability for VANET and focus on protocols such as the DSRC and Wireless Access in Vehicular Environments (WAVE) protocols [9,10]. Moreover, most of the proposed approaches emphasize on reliability in terms of successful delivery and the optimization of the number of packet drops. For example, in [9] communication reliability metrics are defined such as packet delivery ratio, and distribution of consecutive packet drops are reduced. DSRC based scenarios are proposed. In [10], the authors investigate additional metrics, such as reliability, packet reception ratio, and effective range of coverage.

Le Lann in [11] presents the safety and reliability for Intelligent Vehicular Networks (IVN) for platoons and cohorts. To avoid failure, he introduces diversified functional redundancy, giving a replacement in case of telemetry failures as an example.

A lot of researches focus on finding a reliable routing in connected vehicles, and on classifying routing protocols [12,13,14].

In [12], the authors present a new algorithm with 2 notations: virtual equivalent node (VEN) and differentiated reliable path (DRP) to solve problem of link failures, RSU will play the role of VEN if the route failed. In [13], authors introduce a routing protocol ROVER (Robust Vehicular Routing) based on positions of local-aware vehicles to define the zones. The protocol broadcasts control packets and works as Ad hoc On-Demand Distance Vector (AODV) protocol to discover the best routing path. A classification is described in [14] to classify the existing VANET routing protocols into five categories according to their used routing metrics.

In [15], the authors calculate the reliability of OBU in function of hardware reliability and channel availability without taking hardware security into consideration and by using simulations instead of real data.

Waqar *et al*, lists all the papers working on the reliability of communication networks and classify them according the modeling and analysis techniques. For each paper listed in their work, Waqar *et al*, provides a survey of each application in communication networks mentioning the strong and weak points of different approaches [16]. Concerning VANET applications, only Ref. [15] mentioned above is listed.

Further, the authors in [17] proposed an evaluation method of the performance reliability of the Mobile Ad hoc Networks (MANET) and studied the effect of interference. However, reliable communication in VANET is more complex compared to MANET, due to its characteristics, such as the high mobility and the dynamic topology.

In all mentioned articles earlier, we notice that the reliability of hardware and the quantitative and qualitative reliability analysis for the connected vehicle system are ignored in realizing their performance. Hence, we propose to analyze the

reliability of VANET, in terms of network reliability, as function of hardware reliability and functionality while taking into consideration the needed security equipment. Several approaches are applied to evaluate the reliability and the availability for communication systems. In this paper, the investigation is directed using RBDs, for both V2V and V2R communications. RBD analysis is used to conduct both qualitative and quantitative analysis of the systems. RBD deals with the identification of critical or weak components of the VANET system. The reliability of the overall system is calculated on the basis of reliability of the individual components in the OBU. The exponential distributions are used to model the failure characteristics of the OBU equipment nodes.

The California Department of Motor Vehicles (CA DMV) is the state agency that registers motor vehicles and issues driver's licenses in the U.S. state of California, and is responsible for permitting and monitoring the testing of autonomous vehicles. DMV published reports concerning AV failures or disengagement. Disengagement is the failure event where the autonomous of the car (OBU algorithms) fail to take the right decision. In these situation the control should reverts to the human driver.

Studying this report for 2616 disengagements events from September 2014 to December 2016, show that 52% of this disengagement were due to system failure, which refers to a hardware, and/or software failure of the vehicle technology that causes the impossibility for the vehicle to continue the autonomous operations. System failure is a broad category that includes issues such as a discrepancy between onboard GPS systems, incorrect perception of external objects, incorrect prediction of other traffic behavior, and so forth.

The valuation of these problems is the object of quantitative and qualitative reliability analysis for the connected cars. As shown later, the reliability prediction helps us to precise the redundant component in the OBU and in the RSU in order to get a higher acceptable reliability level.

3 THE PROPOSED APPROACH

3.1 *The analysis approach*

In this paper, we focus on reliability and availability related to hardware failure issues because of their importance and wide utilization in the area of communication networks.

Using the Reliability analysis methods, we can identify the problems in VANET that lead to the improvement of the system design and avoid future problems. Reliability analysis plays an important role in the prediction of the behavior of wireless networks versus time, and gives us a clear view to take the right decision concerning the redesigning of the system in order to have an efficient communication network [18].

Fig. 1 shows the reliability procedure analyses approach; this approach begins with the functional analysis and the development for conceptual behavioral model of the system.

In the first step, we describe our mode of communication (wireless in our case) and the desired network behaviors, such as network protocols (e.g. DSRC, WAVE), network topologies (adhoc, broadcast) and fault tolerance of the connected vehicular network.

For this purpose, we tackled the preliminary risk analysis of the smart vehicles. The functional analysis studies how each of the parameters identified above are ensured by the components of the system. Based on this analysis of the functioning of the VANET, dysfunctions and failures could be identified.

The failure analysis of VANET research sources (causes) of system and their consequences

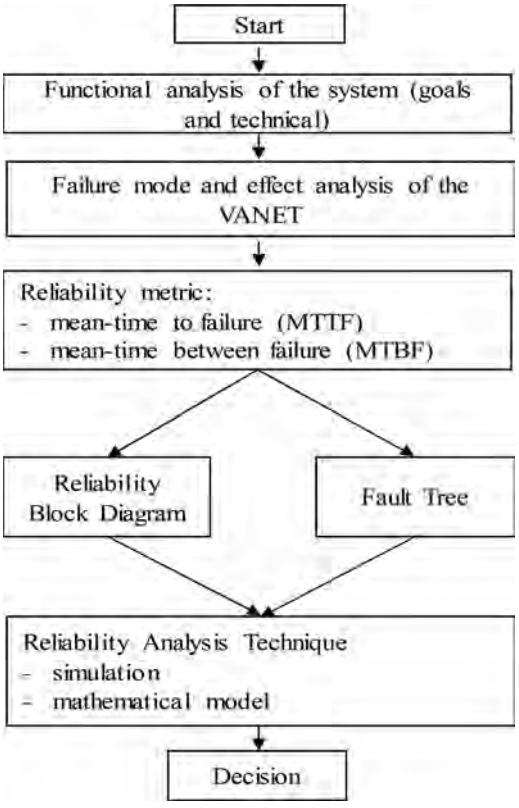


Figure 1. Reliability analyses procedure.

(effects). Two large families of analyses are commonly used. Some inductive, such as hazard and operability study (HAZOP) or analysis of failure modes (FMEA) used in our study, and others are deductive (such as fault tree analysis). The inductive analysis distinguishes systematically the causes and effects of failures. Deductive analysis, meanwhile, are top-down approaches. (step 2).

The third step is the calculation of basic metrics of reliability and availability, such as Mean-Time to Failure (MTTF), Mean-Time Between Failures (MTBF) [19], for each equipment used in VANET. In our case, we obtain these metrics from a professional data base called Quanterion Automated Databook (QAD) that uses Electronic Parts Reliability Data (EPRD). The MTTF and MTBF are measured in hours. These metrics can be obtained by statistically calculating the failure rates of the GPS that are embedded in OBU. The failure rate of components is calculated below:

$$\lambda_{failure/hour} = f / (n.t)$$

where f represents the number of failures occurred during the observations; n: represents the number of components observed and t: represents the total time of observation.

This data is used in order to calculate the reliability of OBU and RSU and then for the system as a whole.

RBD and FT are the most usually used formalisms in system reliability modeling. They are used through two different approaches: in a reliability block diagram, the OBU and the RSU are represented by components connected according to their function or reliability relationships, while fault trees show which combinations of the components failures will result in a system failure. (step 4).

In step 5 the reliability analysis technique and the simulation were carried out, since the understanding and choosing the right probability distributions is very important at this stage. The exponential distribution is widely used to describe events recurring at random points in time, such as the time between failures of electronic equipment of the OBU and the RSU or the time between arrivals at a service booth. It is related to the Poisson distribution, which describes the number of occurrences of an event in a given interval of time.

3.2 *Connected vehicles: Vanet*

The VANET system is composed of several sub-systems.

3.2.1 *On Board Unit (OBU)*

An OBU is a mobile or portable wireless device that is located inside intelligent vehicles [20]. It works as a communication device and allows DSRC communications with other OBUs or RSUs. OBU includes other communication systems (e.g. GPS.), and other subcomponents like: application unit hardware, Human Machine Interface (HMI) and power supply. Each vehicle equipped with an OBU collects data and information, (such as vehicle's speed, position, brake status, signal status, etc…) analyzes, processes and encrypts the data in order to send it as a safety message to other vehicles (V2V) or RSUs (V2R) through the wireless medium;

An automated vehicle is equipped with an OBU system designed to ensure a number of functions. In addition to processing, inputs/outputs, and storage functions, an OBU system provides other major functions such space-time localization and scene recognition, longitudinal telemetry and short-range omnidirectional communications.

The set of system components required for OBU are:

- Resource Command Processor (RCP): It directs the operation of the other units by providing timing and control signals. All other resources are managed by the RCP which is the processing unit of the system. It must permit a local elaboration of the data gathered from the infrastructure and from the smart vehicle.
- GPS system: 360° positioning and global time keeping in order to allow the vehicle to communicate its own position to perform geo-location.
- Wireless communication: VANET uses DSRC to provide an omnidirectional 360° radio wireless communication between moving vehicles. DSRC refers to 802.11p which is an improvement of IEEE 802.11a.
- Antenna: used with a transmitter or a receiver in order to achieve the dedicated range of wireless networks.
- HMI: an electronic display screen that is used for driver assistance in collision avoidance applications. HMI shows awareness messages such as: indications, warnings, and advices using different ways of interaction like visual (flashing light, image) and auditory (sound, alarm).
- Vehicle services: interacts directly with the body chassis systems of the car doing a tactile and kinesthetic functions such as the vibrations of the driver's chair or the steering wheel.

3.2.2 *Trusted Platform Module (TPM)*

In VANET, data is broadcasted over shared communication media: a malicious node may easily intercept, modify or inject data [21, 22]. Data injection can provoke collisions in a vehicular system.

Weak security leads to many traffic problems putting human lives at risk. In Vehicle-Based Security System (VBSS), the OBU generates the

Basic Safety Messages (BSM) that collect vehicle and road conditions data. These BSM should be certificated and signed in order to preserve privacy and enhance essential security services, such as authentication, integrity, confidentiality and nonrepudiation.

A secured vehicle needs some hardware requirements, such as TPM which can be integrated into the OBU; TPM is the hardware module that forms the security issues such as encryption/decryption, hashing and digital signature. It is able to protect and store data and keys in shielded locations.

From hardware point of view, a TPM contains the below components:

- Assembled controller: A TPM contains a controller bus, for the connection and coordination among its memory and peripherals.
- NVRAM: Non-volatile random-access memory is used for permanent storage of the startup configuration that is writeable. It is also used for permanent storage of hardware revision, identification information and the cryptographic keys.
- Crypto unit: as its name indicates, it is responsible for random number generation, public-key cryptographic algorithm, cryptographic hash functions, symmetric-key algorithms, digital signature generation and verification, and Elliptic Curve Cryptography (ECC).

3.2.3 *Road Side Unit (RSU)*

RSU is installed at the road side [23]. It includes communication hardware (e.g. Wi-Fi, UMTS, ITS G5, etc.), and serves as a gateway between OBUs and the communications infrastructure. It could provide location based services and Internet access for mobile devices to improve the communication connectivity. The main functions of RSUs are as follows:

– Network coverage extension of the Ad Hoc network and communication medium between OBUs and RSUs.
– Source of safety and awareness information like weather status.
– Prioritize management messages to and from the OBU.
– Gateways that allow vehicles to establish connection with the internet.

The RSU is connected to the V2I communications network. Prioritization of messages, is also managed by the RSU to and from the vehicle.

Similar to OBU, an RSU is composed of a communication transceiver (802.11 & 1609), GPS and a processor. RSU contains a router that acts as an interface to the V2I cloud network. RSU is also connected to a local safety processor which is related to traffic light signal controller.

4 THE RELIABILITY

In this section we used RBD and FT, the 2 most effective techniques for modeling reliability and availability of communication networks [16].

4.1 *RBD*

RBD for the OBU (Fig. 2) are graphical structures consisting of blocks representing the system components and the connection to these components. The system is functional, if at least one path of properly functional components from input to output exists, otherwise it fails [24].

Thus, the reliability of an OBU, represented by R_{OBU}, is given below:

$$R_{OBU} = R_{Anl} R_{DSRC} R_{GPS} R_{PS} R_{RCP} \left[\left(R_{HMI} + R_{VC} \right) - R_{HMI} R_{VC} \right] R_{TPM} \quad (1)$$

where the reliability of the TPM is set below:

$$R_{TPM} = R_{Cryp} R_{Mem} R_{Cont} \quad (2)$$

In this section we used Reliability Block Diagram (RBD) and the Fault tree (FT), the 2 most effective techniques for modeling reliability and availability of communication networks [16]

The RBD of RSU will be as show in Figure 3.

Consequently, the reliability of an RSU, represented by R_{RSU}, is given by:

$$R_{RSU} = R_{Anl} R_{DSRC} R_{GPS} R_{PS} R_{Process} R_{STL} \quad (3)$$

4.2 *Fault tree of OBU failure*

Fault Tree analysis is a graphical technique performed to know the probability of occurrence of the top event; i.e., a DSRC fail event can cause the whole system to fail. These causes of system failure are represented in the form of a tree rooted by the top event as depicted in Figure 4. Logic

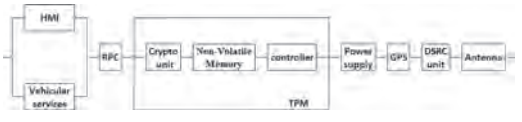


Figure 2. Reliability block diagram for OBU including the TPM.

Figure 3. Reliability block diagram for RSU.

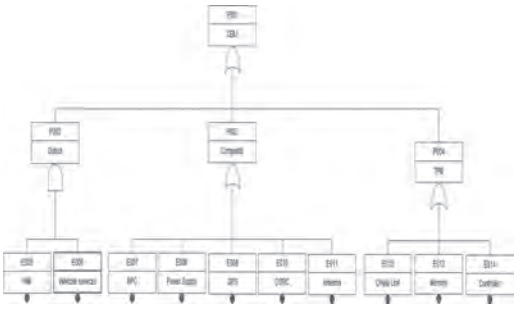


Figure 4. Fault tree event of OBU.

Boolean gates are used to link two or more cause events provoking one fault. The OR gate is presented when one fault from any node is enough to cause a fault. While the AND gate is used when the fault (output) occurs when all inputs fail (inputs are independent).

Many quantitative data are needed in order to achieve this analysis; hence the necessity that the data should be accurately targeted. QAD contains field failure rate data on a variety of electrical, mechanical, electromechanical, and microwave parts and assemblies. It is used as a source of reliability failure rate data. Data contained in EPRD reflects industry average failure rates, especially the summary failure rates which were derived by combining several failure rates on similar parts/assemblies from various sources. At t = 0, the population of parts has not experienced operation. As operating time increases, parts in the original population are replaced and the failure rate increases.

Probability of events is modelled from component failure rates. These are defined by the average of failure rate (Failure rate Mean (λ_{Mean})), (Failure rate Lower (λ_{Min} Failure Rate Upper (λ_{Max})), and by standard deviation of failure rate (Failure rate SD (λ_{SD})).

T The component failures of the system follow an exponential distribution on a total duration of 1year (8760 hours) to take into account the history of degradation in the estimation of the fault. During this time, we consider that the OBU system and the RSU system are operating in real conditions (Figure 4)

5 PROBABILISTIC MODELLING OF THE FAULT TREE EVENTS

The objective of this section is to present the probabilistic approach applied on data from the event tree. The top event is the OBU Failure. Reliability data expressed in failure rate are random.

The reliability information of these events are: λ_{Min} = minimum degradation rate, and λ_{Max} = maximum degradation rate, and λ_{Mean} = mean degradation rate, and λ_{SD} = standard deviation of degradation rate.

The exponential distributions, with failure rate $\boldsymbol{\lambda}$ and time-to-failure random variable, are used in order to express the reliability or availability of these individual components of the OBU. The dependability of each component is then used in order to determine the reliability of the overall VANET system by utilizing the mathematical expressions that are presented in eq. 1. The exponential model is the wieldiest used for electronics components such in the OBU and the RSU, even for a pessimistic scenario. The failure λ rate is considered as a random variable which is defined between two limits. Its mean and standard deviation are known.

Table 1 depicts the failure events data for all the components of the OBU; the power supply (E008) has the greater failure rate, and has a significant influence on the reliability of the OBU, since the power supply is implemented in series in the RBD, that means definitely when the power supply of OBU fails the whole system will fail. As an improvement we suggest to use a dual power supplies in the OBU.

The redundant power supply, shown in Figure 5, might be connected in parallel. After the redesigned RBD for the OBU, we noticed an improvement in reliability between the OBU having a single power supply and the redesigned OBU having dual power supplies.

Thus, the reliability of an OBU, after adding a redundant power supply R_{PS2}, represented by R_{OBU2}, is given below:

$$R_{OBU2} = R_{Ant} R_{DSRC} R_{GPS} \left[\left(R_{PS1} + R_{PS2} \right) - R_{PS1} R_{PS2} \right] R_{RCP} \left[\left(R_{HMI} + R_{VC} \right) - R_{HMI} R_{VC} \right] R_{TPM} \quad (4)$$

As shown in Figure 6, After 8760 hours (1 year) of running, the OBU1 powered by a single

Table 1. Failure events data.

Basic event	Label	λ_{Min}	λ_{Max}	λ_{Mean}	λ_{SD}
HMI	E005	1.4E-06	1.8E-06	1.6E-06	1.8E-07
Vehicular services	E006	4.2E-07	2.3E-06	1.2E-06	8.08E-07
RCP	E007	2.8E-06	3.7E-06	3.3E-06	3.7E-07
Power Supply	E008	4.7E-06	9.1E-06	6.3E-06	2.01E-06
GPS	E009	1.4E-06	1.8E-06	1.6E-06	1.8E-07
DSRC	E010	1.05E-06	1.2E-06	1.1E-06	8.18E-08
Antenna	E011	4.8E-07	6.2E-07	5.5E-07	6.18E-08
Crypto Unit	E012	2.8E-06	3.7E-06	3.3E-06	3.7E-07
Memory	E013	4.6E-07	2.5E-06	1.1E-06	9.04E-07
Controller	E014	1.2E-06	3.3E-06	2.1E-06	9.6E-07

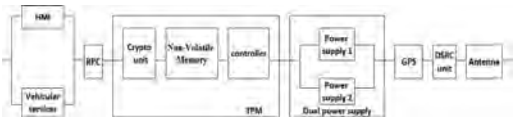


Figure 5. RBD for OBU2 using dual power supply.

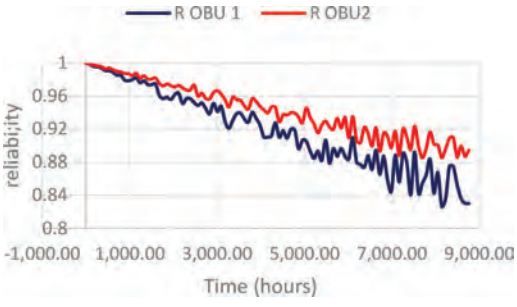


Figure 6. Reliability comparison between OBU1 and OBU2.

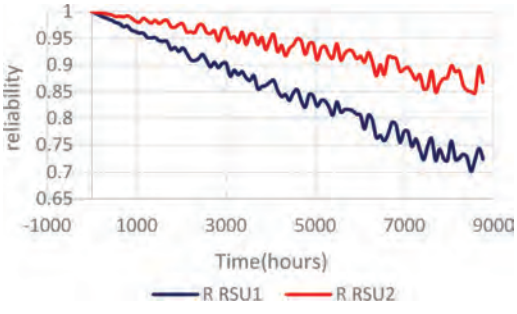


Figure 7. Reliability comparison between RSU1 and RSU2.

power supply reaches a reliability value equal to 82%, however the OBU2 powered by a dual power supply reaches 88% of reliability, that means an increase of at least 6%. This improvement becomes more evident and visible as time passes.

Concerning the RSU, and since all the components are implemented in serial, the remodeling proposed to implement the RSU with dual power supplies increases the reliability for more than 17%. In (Figure 7) it is clear that the degradation in reliability in RSU2 powered by dual power supplies is less severe than RSU1 that powered by a single power supply.

6 CONCLUSION AND FUTURE WORK

Connected vehicles are classified as "safety critical" because they are directly related to human safety. For this reason, in this paper we used the exponential model which is very conservative and pessimistic. A failure occurs whenever some function is lost, and no other function can supersede the lost function in due time.

Any OBU failure can lead to crashes. In similar domains (e.g. in aircraft), the control system and communication parts may be triplicated or quadrupled [25]. The model used in the aviation system, where the system reliability is very high and failure ratio should be in range of 10^{-9} per hour, can be used as a reference. The solution used in aircraft could be cloned into vehicle context in order to provide distinct redundancy of hardware. However, a trade off between the disadvantages of redundancy (cost, weight, power saving and design complexity...) and the reliability should always be considered.

In this article we demonstrated the importance of reliability analysis using mixed approach between qualitative and quantitative methods. As a result, we detected that the reliability of a system with a single point of failure such as the OBU system with serial subcomponents—will immensely degrade with time. Using a very pessimistic exponential model as showed in Figure 6, resulted in the degradation of the OBU 's reliability by 16% over a short period of one year—the failure rates of the OBU electronic components were in the range of E-07 to E-06. Moreover, one of the most important advantages of vehicular networks is that there are no energy constraints, unlike wireless sensor networks and other types of mobile devices used in MANETs where limited battery life is a major concern. Taking into consideration this feature we can implement the redundancy components for the VANET without worrying about energy consumption.

Autonomous Vehicles could also operate in isolation from other vehicles using internal sensors. A combination between the autonomous vehicle and the connected vehicle is called Automated Vehicle (CAV), which leverages autonomous and connected vehicle capabilities thus making the system more complex. For this reason, it is very important to analyze each component functionality to identify its failure rate ratio in order to redesign for redundancy.

The paper mainly emphasized on the reliability of the connected vehicle networks primarily with respect to hardware components and the needed security equipment involved in communications. However, a future study might tackle the reliability degradation using other models such as Weibull model which is classified as realistic model. Another future study will also tackle the reliability of the whole system in the operation mode and evaluate the reliability of communication from sender to receiver to include the reliability of data transmission while using real network and traffic conditions on the proposed modeling technique.

REFERENCES

[1] World Health Organisation. Road safety report. 2015. http://www.who.int/mediacentre/news/releases/2015/road-safety-report.

[2] R. Naja. Wireless Vehicular Networks for Car Collision Avoidance. Eds Springer. 2013.

[3] G. Le Lann. Safe Fully Automated Driving on Roads and Highways. Séminaire System X, Palaiseau, France 20 oct. 2015.

[4] S. Medetov, M. Bakhouya, J. Gaber, K. Zinedine, M. Wack, P. Lorenz. A Decentralized Approach for Information Dissemination in Vehicular Ad hoc Networks. Journal of Network and Computer Applications, Elsevier, Volume 46, November 2014, Pages 154–165.

[5] J. Petit, S. Shladover. Potential Cyberattacks on Automated Vehicles. IEEE Trans. *Intell.Transp. Syst.* 2015, 16, 546–556.

[6] I. Sumra, H. Halabi, J. Manan, M. Rehmam. Trust and Trusted Computing in VANET. Computer Science Journal Volume 1, Issue 1, April 2011.

[7] P. Hoang. System Software Reliability. Eds Springer Series in Reliability Engineering. 2006.

[8] K.S. Trivedi. Probability & statistics with reliability, queuing and computer science applications, Eds John Wiley & Sons, 2008.

[9] B. Fan, K. Hariharan. (2006). Reliability Analysis of DSRC Wireless Communication for Vehicle Safety Applications. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. 355–362. 10.1109/ITSC.2006.1706767.

[10] X. Ma, X. Yin, K.S. Trivedi. On the reliability of safety applications inVANETs. *Int J Perform Eng* 2012;8(2):115–30.

[11] G. Le Lann. Integrated Safety and Efficiency in Intelligent Vehicular Networks: Issues and Novel Constructs. Panos Papaioannou. TRA 2012 – Transport Research Arena Europe, Apr 2012, Ath_enes, Greece. Elsevier, 48, pp.951–961, 2012, ScienceDirect; Procedia—Social and Behavioral Sciences.

[12] R. He, H. Rutagemwa, X. Shen. Differentiated reliable routing in hybrid vehicular ad-hoc networks. May 2008, pp. 2353–2358.

[13] M. Kihl, M. Sichitiu, T. Ekeroth, M. Rozenberg. Reliable Geographical Multicast Routing in Vehicular Ad-Hoc Networks. Springer Berlin Heidelberg, 2007.

[14] Y. Gongjun, N. Mitton, X. Li. Reliable Routing in Vehicular Ad hoc Networks. The 7th International Workshop on Wireless Ad hoc and Sensor Networking (WWASN 2010), Jun 2010, Genoa, Italy, 2010.

[15] S. Dharmaraja, R. Vinayak, K.S. Trivedi. Reliability and survivability of vehicular ad hoc networks: An analytical approach. *Reliability Engineering & System Safety*, Volume 153, September 2016, pp. 28–38.

[16] W. Ahmad, O. Hasan, U. Pervez, J. Qadir. Reliability modeling and analysis of communication networks. *Journal of Network and Computer Applications*, Volume 78, 15 January 2017, pp. 191–215.

[17] S. Xiang, J. Yang. Performance reliability evaluation for mobile ad hoc networks, *In Reliability Engineering & System Safety,* 2017, ISSN 0951–8320.

[18] S. Bernardi, J. Merseguer, D.C. Petriu. Model-driven dependability assessment of software systems, Springer, 2013, Ch. Dependability analysis techniques.

[19] S. Stanley. MTBF, MTTR, MTTF & FIT explanation of terms, IMC Network (2011) 1–6.

[20] M. Petracca, P. Pagano, R. Pelliccia, M. Ghibaudi, C. Salvadori, C. Nastasi. On-Board Unit hardware and software design for Vehicular Ad-hoc Networks. In: Roadside Networks for Vehicular Communications: Architectures, Applications and Test Fields by Alexey Vinel, Robil Daher. Eds IGI Global. 2012.

[21] I. Sumra, H. Halabi, J. Manan Comparative study of security hardware modules (EDR, TPD and TPM) in VANET. 3rd National Information Technology Symposium (NITS 2011).

[22] R.G. Engoulou, M. Bellaïche, S. Pierre, A.Quintero. VANET security surveys. Computer Communications 44 (2014) 1–13.

[23] K. Tanuja, T.M. Sushma, M. Bharathi, K.H. Arun. A Survey on VANET Technologies. *International Journal of Computer Applications* (0975–8887) Volume 121 – No.18, July 2015.

[24] W. Ahmed, O. Hasan, S. Tahar. Formalization of reliability block diagramsin higher-order logic, *Journal of Applied Logic* 18 (2016) 19–41.

[25] J.R. Sklaroff. Redundnacy Mangement Technique for Space Shuttle Computers, *IBM Journal of Research and Development.* Volume: 20, Issue: 1, Jan. 1976.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Bayesian networks with imprecise datasets: Application to oscillating water column

H.D. Estrada-Lugo, E. Patelli & M. de Angelis
Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK

Daniel D. Raj
Department of Ocean Engineering, Indian Institute of Technology Madras, Chennai, India

ABSTRACT: The Bayesian Network approach is a probabilistic method with an increasing use in the risk assessment of complex systems. It has proven to be a reliable and powerful tool with the flexibility to include different types of data (from experimental data to expert judgement). The incorporation of system reliability methods allows traditional Bayesian networks to work with random variables with discrete and continuous distributions. On the other hand, probabilistic uncertainty comes from the complexity of reality that scientists try to reproduce by setting a controlled experiment, while imprecision is related to the quality of the specific instrument making the measurements. This imprecision or lack of data can be taken into account by the use of intervals and probability boxes as random variables in the network. The resolution of the system reliability problems to deal with these kinds of uncertainties has been carried out adopting Monte Carlo simulations. However, the latter method is computationally expensive preventing from producing a real-time analysis of the system represented by the network. In this work, the line sampling algorithm is used as an effective method to improve the efficiency of the reduction process from enhanced to traditional Bayesian networks. This allows to preserve all the advantages without increasing excessively the computational cost of the analysis. As an application example, a risk assessment of an oscillating water column is carried out using data obtained in the laboratory. The proposed method is run using the multipurpose software OpenCossan.

1 INTRODUCTION

Nowadays, experimental research in engineering deals with systems with high complexity due to the number of components used in the procedures. Occasionally, the limited capacity of the experimental arrangements to test different configurations and obtain a number of relevant measurements, hinders the impact of the study. In addition, the effects of epistemic uncertainty derived from the procedure and, the uncontrollable conditions under which the study is carried out can provide results with limited applicability or lack of meaning. There are different methods developed for modeling the dependability and evaluation of large engineering systems allowing to take into consideration both qualitative and quantitative information. Among the most used methods, the reliability block diagrams, fault trees and, event trees, can be identified as the techniques with reliable results providing a robust mathematical background (Hamada et al. 2008). However, several assumptions are made with these techniques.

The Bayesian Network (BN), is a probabilistic method to study and analyze the genuine dependencies or independences of variables that make up a system. This concept was proposed by Judea Pearl in 1988, originally for the artificial intelligence area (Pearl 1991). Currently, the BNs have many more applications ranging from system dependability (Castillo et al. 1997) and risk analysis (Hudson et al. 2002), to system maintenance (Kang and Golay 1999). It is worth noticing that this method has attracted an increasing interest, reaching 800% according to (Weber et al. 2012), during the last 20 years. The success of Bayesian networks rests on the graphical representation of the system, which renders them intuitive and easy to understand even by for non-experts. In addition, this method can be used to provide a diagnostic or predictive reasoning, a combination of both (Korb and Nicholson 2004) and also they accept new evidence that can be used to update the network and to adapt the model to the new parameters. Moreover, information of different types (e.g. expert judgment, experimental data, historical records, feedback experience, theoretical models, etc.) can be merged in the same network, inside structures called probability tables (or conditional probability tables in the case of children nodes). These tables are filled with crisp probability values, providing a global dependability estimation (Jensen and Nielsen 2007).

On the other hand, the high acceptance of the traditional Bayesian networks for uncertainty

reasoning is limited to the use of only crisp probabilities (Spiegelhalter 1987). This type of probabilities leads to discretization methods or hard assumptions, impoverishing the quality of the analysis (Tolo et al. 2016a). In order to work with continuous probabilities that can take into account the uncertainty of the variables in the network and avoid discretization of the input information, Daniel Straub and Armen Der Kiureghian (Straub and Der Kiureghian 2010) proposed to enhance BNs with structural reliability methods since these techniques support the use of continuous random variables. This approach embraces all the advantages of BNs and furthermore allows working with discrete, continuous, as well as small probability variables (ideal for low-probability high-impact events). Some applications have been done with this method, focusing on the risk assessment of technological facilities considering climate change (Tolo et al. 2016a).

Although the enhanced Bayesian networks consider a broader spectrum of variables, usually the information available rarely meets the requirements of the method. For example, often expert beliefs do not agree in an exact same probability value or the information is scarce, in which cases data have to be averaged or the analyst makes assumptions to perform the study. In engineering, it is common to perform measurements during an experiment, the results obtained will have attached an epistemic uncertainty that cannot be eliminated and underestimated. Consequently, the incorporation of imprecise probabilities becomes an imperative need that can improve the employability of BNs.

The proposed method attempts a naive implementation of Credal sets and p-boxes as a way to characterize imprecise probabilities in discrete and continuous variables, respectively. This approach is expected to overcome two main problems when dealing with uncertainty in Bayesian networks. The method is implemented on the multipurpose software OpenCossan, (Patelli et al. 2018). On one hand, the use of all the advantages of parametric and non-parametric p-boxes to work with continuous imprecise random variables and obtain the quantile bounds of the final distribution representing the system under analysis so they can be used after with the structural reliability methods. The network reduction process can be done with the *advanced line sampling* method, since it is capable of approximating the upper and lower bounds of the failure probability under the assumption of a monotone system. Once the network is reduced and the uncertainty of continuous variables propagated to the reduced network through the bounded probability of failure, the result will be a Credal network with only discrete but bounded variables. The method used, allows to compute the exact bounds of the query probability in the absence of evidence. In the case of evidence introduced in the network the method will provide the intervals, such that the true bounds of the query probability are located.

2 THEORETICAL BACKGROUND

2.1 *Bayesian network*

A Bayesian network, as established before, is presented in the form of a directed acyclic graph (DAG) made of *nodes* and arrows (called *links*) connecting those nodes. Each node represents a random variable with information about observable quantities or hypothesis of the system, whilst the links show the dependency among the nodes. Nodes can be differentiated as *parent* and *child*. A child node depends directly on another node, called the parent node and graphically they are connected by a link starting in the parent and ending in the child. Nodes with no parents are called *roots*. The dependence of nodes is ruled the d-separation concept. Two variables, namely, A and B, are d-separated if there is an intermediate variable, C, different from A and B, such that in a serial or diverging connection C is instantiated (with a specific probability value, i.e. evidence). In the case of a converging connection, if neither C nor any of its descendants are instantiated they are d-separated (Jensen and Nielsen 2007).

The arrangement of parents and children nodes connected by links allows performing diagnostic and predictive reasoning. The first approach can be done to know the causes of an event by querying a parent node given the information in children nodes. The predictive reasoning follows the direction of the links and uses the causes (information in the parents) to predict the effects (children). The probability values denoting the dependency between a child node with its parents is stored in a conditional probability table. In this arrangement, the probability of each state of the child is provided given each of the states of the parents. The total dependability of the network is quantified by the *joint probability distribution* which is defined as the product of all the conditional and unconditional probabilities specified in the network. This is governed by the chain rule for Bayesian networks (Jensen and Nielsen 2007) and, it is given as follows,

$$P\left(X_i\right)=\prod_{i=1}^{n}P[X_i \mid pa\left(X_i\right)] \tag{1}$$

where X_i represents each of the random variables of the network and, pa(X_i) is the probability of the parents of X_i. As an example, the joint probability of the Bayesian network presented in Figure 1 is given by the expression,

$$\begin{aligned} P\left(X_1X_2,X_3,X_4\right) &= P\left(X_1\right)P\left(X_2\right) \\ &\cdot\ P(X_3 \mid X_1,X_2)P(X_4 \mid X_1) \end{aligned} \tag{2}$$

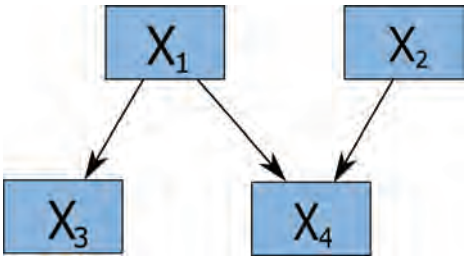


Figure 1. Example of a traditional Bayesian network.

The joint probability distribution function can be used to calculate the probability of any individual variable (Hosseini and Barker 2016). For instance, to calculate the probability only of the node X_3, the rest of the nodes are marginalized out from Eq.(2). So, $P(X_3)$ will be given as follows,

$$P(X_3) = \sum_{X_1,X_2,X_4} P(X_1)P(X_2) \times P(X_3 \mid X_1, X_2)P(X_4 \mid X_1) \quad (3)$$

Marginalization is a distributive process to calculate the total probability of a variable of interest by the summation of the products of all the possible combinations of local joint probabilities. This process allows to isolate the probability of the parameter of interest and to remove the rest of the variables from the joint probability distribution (Jensen and Nielsen 2007).

It is possible to remove variables that are not relevant to know the probability of the variable of interest, namely A. This process is called *variable elimination*. It consists of simply removing from the joint probability, variables that are outside the Markov blanket of A (i.e. variables that are parents or children of A, or sharing a children with A). The eliminated variables (those out of the Markov blanket), do not influence the probability measures of the variable of interest.

The type of information that can be adopted in this method involves real probability values (discrete nodes) or Gaussian distribution functions. The latter works with crisp value probabilities (Tolo et al. 2016b). However, this characteristic turns into the main drawback of this technique when the data comes in the way of continuous distributions. Nevertheless, this disadvantage is overcome with the use of structural reliability methods.

2.2 *Enhanced Bayesian networks*

Structural Reliability methods (SRMs) are used to work out the conditional probability tables of a BN containing both discrete and continuous random variables (enhanced Bayesian network) resulting in the reduction of the network to a traditional BN. Suppose the nodes in the network on Figure 2

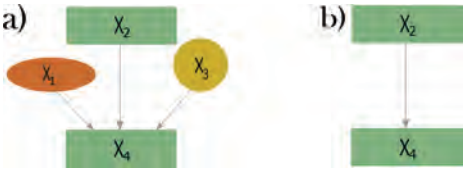


Figure 2. a) Simple enhanced Bayesian network with discrete nodes (rectangular shaped), probability function nodes (circle shaped) and, interval discrete nodes (elliptical shaped). b) Reduced Bayesian network with crisp probabilities.

a) correspond to independent random variables. Here X_1 is an interval node (with $f(X_1)$ its cumulative distribution function, CDF), X_2 and X_1, are discrete nodes (representing probability mass functions, $P(X_2)$, $P(X_4)$, respectively) and X_3 a continuous node (representing a CDF $f(X_3)$). According to (Straub and Der Kiureghian 2010) the enhance Bayesian network joint probability can be computed by approximating the equation below,

$$P(X_i) = \int_{X_1,X_3} f(X_1)f(X_3)P(X_2) \times P(X_4 \mid X_1, X_2, X_3)dX_1dX_3 \quad (4)$$

if the Markov properties are considered, node X_3 and X_1 are d-separated from X_2 since X_4 has not received any evidence yet. So, the joint probability of X_4 given X_2, from the equation above, can be written as,

$$P(X_4 \mid X_2) = \int_{X_1,X_3} f(X_1)f(X_3) \times P(X_4 \mid X_1, X_2, X_3)dX_1dX_3 \quad (5)$$

It has to be noticed that each entry in the conditional probability table of X_4 is defined by the domain $\Omega^{X_4}_{X_4,x_2}$ in the continuous space of X_1 and X_3 for a given value x_2 of the variable X_2. So the Eq. 5 can be further reduced as,

$$P(X_4 \mid X_2) = \int_{\Omega^{X_4}_{X_4,x_2}} f(X_1)f(X_3)dX_1dX_3 \quad (6)$$

The integral shown in Eq. 6 is equivalent to that of a reliability problem (Tolo et al. 2016b). Solving a structural reliability problem, i.e. approximating the system failure probability, a reduction of a network with continuous nodes to one with only discrete probability values is obtained.

There are certainly several approaches for the solution of problems like that shown in Eq. 6. These methods range from numerical approximations like Monte Carlo simulations, to the well-known and widely used *first-order* and *second-order reliability methods* (Hasofer and Lind 1974). Moreover, some advanced sampling techniques like Importance Sampling, Stratified Sampling or Advanced Line

Sampling (Hasofer and Lind 1974), among others, have been used as an alternative to the computationally expensive numerical approximations.

2.3 *Credal networks*

Credal networks can be referred as an extension of Bayesian networks to manage intervals of discrete probability values representing the lack of information and uncertainty about the variables involved. A Bayesian network constructed exclusively with discrete nodes such that, only one probability value is associated with the state of the variable. Such a state can belong to the variable itself, in the case of roots, or can be conditioned on the parents of that node, in the case of children. However, in a Credal network, probabilities are presented in the form of intervals that are associated with probabilistic inequalities. In this manner, a Credal network will represent the different variable states, each of them associated with one specific probability value inside the interval, of the same Bayesian network (Tolo et al. 2018). The graphical structure of such a network is the same as the Bayesian case, as well as the Markov blanket concept and d-separation of nodes. Nevertheless, the probability of a variable x is indicated in the form of the so-called credal set $K(X)$, whilst the set of joint probability measurements $P(X_i \mid pa(X_i))$ is named a joint credal set, given by $K(X_i \mid pa(X_i))$ (Cozman 2000). So, two different interval probabilities of variables X and Y (where Y is the complement of X) can be characterized by their upper and lower bounds, $[\overline{p}(X), \underline{p}(X)]$ and $[\underline{Y} = 1 - \overline{p}(X), \overline{Y} = 1 - \underline{p}(X)] \epsilon [0\ 1]$, respectively.

2.4 *Probability boxes*

Probability boxes (or p-boxes) allow making fewer assumptions about the values used in the study when correlations of the variables employed in the study are ignored due to the effect of aleatory and epistemic uncertainties. A p-box is specified for a random variable X by the interval bounds $\left[\underline{F}, \overline{F}\right]$ on a cumulative distribution function F with values between 0 and 1, such that $\underline{F}(X) \le F(X) \le \overline{F}(X)$ (Ferson et al. 2003). If a probability measure $\underline{p}$ (since it is the lower bound of that measure) for the random variable X_1 is given, the lower, $\underline{F}(X_1)$, and upper $\overline{F}(X_1)$, bounds of the p-box can be computed as follows (Walley 1991),

$$\underline{F}(X_1) = \underline{p}(X_i \le X_1), \overline{F}(X_1) = 1 - \underline{p}(X_i > X_1) \quad (7)$$

The p-box has a dual interpretation, i.e. $\overline{F}$, represents the probability (CDF axis) upper bound and quantile (x-values axis) lower bound. The opposite happens with $\underline{F}$. Therefore, this concept is applicable in cases of imprecise continuous probabilistic distributions and two types can be differentiated, parametric non-parametric p-boxes. A parametric p-box is defined when the shape of the probability distribution is known, but there is no precise information about the parameters of that distribution. The non-parametric case is rarer but can exist especially when an experiment has been performed and a set of measurements was obtained. It occurs when parameters regarding the probabilistic distribution, e.g. mean and variance, of a variable, are known but no information about the type of distribution is available (Ferson et al. 2003).

2.5 *Computational tool*

The open source software OpenCossan exploits the Object-Oriented programming paradigm that Matlab offers with the use of *classes* (entities containing the data and the functions or methods that the members of the same class have in common) and *objects* (instances of a class). This basis is employed in order to efficiently provide solutions to problems regarding uncertainty quantification, sensitivity and reliability analysis, robust design, among others (Patelli 2015). The object-oriented methodology allows reutilizing parts of code to create more complex objects in a systematic and condensed way. OpenCossan offers a wide flexibility to integrate new methods that enhance, improve or complement the current tools available in this software. This opens the gate for new developments that enrich the software robustness to provide solutions.

Within the framework of OpenCossan, three main toolboxes can be used for Bayesian networks. These are, *BayesianNetwork*, *EnhancedBayesianNetwork* and, *CredalNetwork*. The first one can be used in cases were variables only correspond to crisp probability values whilst the second one, also considers continuous probability distributions and bounded variables. The third toolbox is the one chosen for this study, since it allows working with continuous probabilities and interval variables representing imprecise probabilities. However, the graphical display process is done with the same method, *makeGraph* which is a class of *EnhancedBayesianNetwork*.

3 CASE STUDY: OSCILLATING WATER COLUMN

An Oscillating Water Column (OWC) is a type of the so-called wave energy converters that capture the energy that sea waves deposit once reaching the named structure. The popularity of this type of energy converters has increased over the last few decades, since it is an alternative for the clean energy production (Falcão 2010). The structure of an OWC consists of a chamber, partially

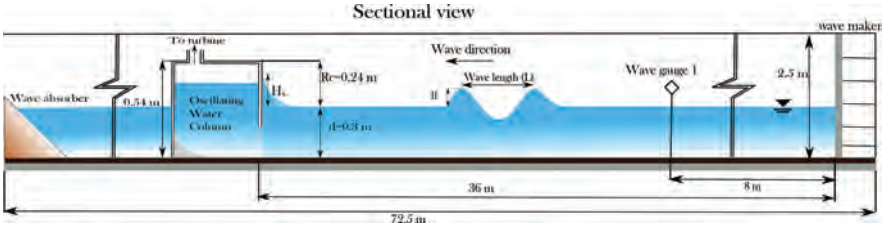


Figure 3. Sectional view wave flume with OWC on the left-hand side.

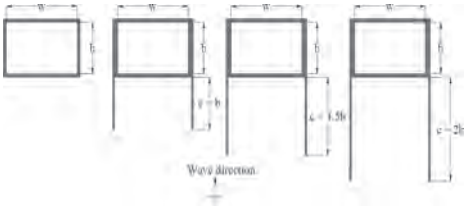


Figure 4. Schematic layout from a top view of harbour walls with different lengths (Daniel Raj et al. 2016a).

submerged in the sea, generally with two orifices. One is typically at the top of the chamber inside which turbine is placed, and the other one is below the water line facing the coming sea waves, see Figure 3. An entombed mass of water column formed inside the chamber oscillates as a result of the wave inside in the structure. This, in turn, drives an air flow through the orifice coupled to a turbine, thus generating electricity (Cruz 2008).

Since the system is always surrounded by sea waves (especially when it is built away from the shoreline), it is susceptible to overtopping events that can cause serious damages to the structure, particularly in the operational mode. Horizontal sliding, overturning, scouring and collapsing of the structure may be possible during extreme sea conditions (Cruz 2008). It is preferred to study such overtopping possibilities to be more on the conservative side. Generally, a conventional-OWC device is constructed with adequate height such that the overtopping event is quite not possible. As the addition of harbor walls increases the wave amplification, the possibility of an overtopping event should be carefully addressed.

A very simple Bayesian network was built to test the inclusion of imprecise probability nodes in the computational toolbox based on the OpenCossan software. The network represents the components involved in an experimental work carried out by Daniel Raj and his team (Daniel Raj, D. et al, 2016) at the Indian Institute of Technology Madras in India, to study the influence of harbour walls of an OWC on its energy efficiency characteristics. The present network is used to provide an assessment of the risk of structure overtopping triggered by the waves generated in the laboratory.

3.1 *Description of the experiment*

The experimental arrangement was 72.5 m long, 2 m wide, and with a deep wave flume of 2.5 m, please refer to Figure 3. The scaled OWC model was 0.540 *m* height, in a 1:20 ratio from a real prototype. To generate random water waves (it is able to reproduce either shallow or deep water waves) the flume is equipped with a wave maker system to generate waves with steepness characteristics within the limits of the operational range of the system. The generated waves for this experiment covered a range of relative water depths, d/L, from 0.074 to 0.23 and a wave steepness, H/L, from 0.0074 to 0.065. Here, d denotes water depth, H corresponds to the wave height (in metres) registered from the first wave gauge 1 (situated at 8 *m* from the water generator), and L denotes wavelength (in metres). The crest periods, T_p, adopted were from 1 to 2.5 *s* in 0.25 *s* intervals. For more information about this experiment please refer to (Daniel Raj et al. 2016a), (Daniel Raj et al. 2016b).

The experiment was carried out in two stages; the first stage involved the identification of efficient resonating length of harbour wall, as seen in Figure 4, which enhances the energy conversion capacity of the system. Four testing criteria have been chosen in the first stage, so as to investigate the effect of projecting sidewalls length on the efficiency of the OWC. Among them, one is without the sidewalls (conventional) and rest with the projecting sidewalls in three criteria such as c/b of 1, 1.5 and 2. In the second stage, the effect of the harbour walls on each side of the OWC was studied by varying the angle of the harbour walls within the range of $[4\pi/8, 7\pi/8]$ at intervals of $\pi/8$ with respect to the front lip wall of OWC, as seen in Figure 5. This angle variation is called *wall inclination* from now on. The wall length, c, was maintained constant in this stage at its optimal configuration identified from the first stage of experiment.

3.2 *The Bayesian network*

The goal of the experiment was to analyse the influence of different configurations of harbour walls in order to modify the resonance frequency of the water wave coming towards the structure so

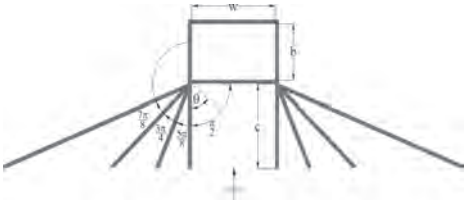


Figure 5. Schematic layout different harbour wall inclinations (top view) (Daniel Raj et al. 2016b).

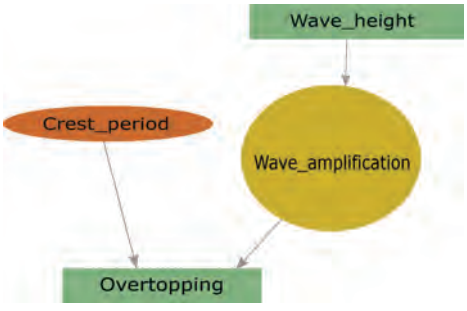


Figure 6. Enhanced Bayesian network for OWC experiment.

that the wave amplification was maximized. The maximization of wave amplification poses a potential risk of structure overtopping that can damage the rearward face of the experiment equipment or to bring unexpected consequences. The probability of having this event is quantified through the network presented in Figure 6 that takes into account the epistemic uncertainty characterizing the measurements of the amplified waves running up the structure. In addition to that, the wave properties, as well as the OWC configurations studied in the experiment, were considered as influencing factors for the occurrence of structure overtopping.

Since imprecise variables are employed in this simple study, the network for this problem is defined in the toolbox for Credal networks node by node, in the form of computational objects. Once all the variables are defined, an object of the class *CredalNetwork* is created to store the information regarding the nodes on the network, so further calculations can be performed. Then, the method *makeGraph* is invoked to display the network shown in Figure 6. The nodes are defined as follows.

Different harbour wall inclinations and lengths were tested to study their effects on wave amplification. In order to select the case to be studied with the BN, the *Experiment_case* node was defined as an interval node containing the maximum and minimum length ratio, for the variation of harbour wall lengths experiment, or the maximum and minimum angle for the remaining case. The values in the intervals are defined in such way that they can cover all the possible values used in the experiment.

Wave properties were used in the Bayesian network as follows, the *Crest_period* interval node contains the information of the mean level-up-to-down-crossing time of the incident waves. The crest periods were changed from 1 to 2.5 *s* intervals of 0.25 *s*. Meanwhile, the *Wave_height* node has the information of the wave height used in the experiments. The waves studied were 0.03, 0.06 and 0.09 *m* high for each of the harbour wall configurations. It has to be noted that the *Wave_height* will influence directly how high the amplified wave is.

According to some authors, as shown by Allsop review (Allsop et al. 1985), wave amplification phenomenon (or wave run-up) in steep structures slopes follows approximately Rayleigh distribution. For this reason, the *Wave_amplification* is assumed to follow this probability distribution with a scale parameter based on the experimental results obtained from the wave amplification measurements. However, the parameters to define this variable in the experiment are uncertain due to the lack of probabilistic data (i.e. only one measurement was taken for a given value of wave crest period, wall inclination or length). So, the use of p-boxes becomes handy. The *Wave_amplification* p-boxes were defined in such manner that all the possible values tested experimentally were enclosed in the p-box of each case. These values are presented in the Table 1, for the given harbour wall length and inclination configurations as well as each of the different wave heights, respectively.

In this work, the Owen equation proposed, by Mase (Mase et al. 2013), for overtopping discharge ratio Q is used to describe the *Overtopping* node and the probability of exceedance, $P(Z)$, of the maximum admissible wave overtopping is given as:

$$Z = Q_{max} - \left(A T_p g \, H_s \right) e^{-B \frac{R_c}{T_p \sqrt{g \, H_s}}} \tag{8}$$

where A and B are dimensionless empirical coefficients depending on the slope ratio of the structure. In this case, A and B are given as 0.0079 and 20.12 (Owen 1980), respectively. H_s is considered as the significant wave height of the amplified waves in meters, g the gravitational acceleration and, R_c the structure freeboard. If the condition

Table 1. Wave amplification p-boxes defined with the Rayleigh-distribution scale parameter for harbour-wall length (a_{length}) and inclination ($a_{inclination}$) experiments.

Wave height (m)	a_{length}	$a_{inclination}$	Distribution
0.03	[0.038, 0.077]	[0.045, 0.08]	Rayleigh
0.06	[0.082, 0.142]	[0.11, 0.157]	Rayleigh
0.09	[0.121, 0.213]	[0.183, 0.237]	Rayleigh

Figure 7. Credal network after reduction process for OWC experiment.

Table 2. Occurrence of structure overtopping for wall length case. The bounded values in $Overtopping_{a,b,c}$, correspond to the *Wave_amplification* p-boxes in each of the *Wave_height* cases, i.e. 0.03, 0.06 and 0.09 *m*, respectively.

Period (*s*)	$Overtopping_a$	$Overtopping_b$	$Overtopping_c$
1	[0, 0.013]	[0.016, 0.273]	[0.165, 0.555]
1.25	[0, 0.011]	[0.19, 0.279]	[0.165, 0.563]
1.5	[0, 0.01]	[0.021, 0.272]	[0.164, 0.551]
1.75	[0, 0.012]	[0.2, 0.269]	[0.159, 0.557]
2	[0, 0.01]	[0.02, 0.268]	[0.156, 0.562]
2.25	[0, 0.013]	[0.2, 0.275]	[0.161, 0.555]
2.5	[0, 0.009]	[0.2, 0.263]	[0.164, 0.557]

$P(Z \leq 0)$ is exceeded the event is considered a failure of the system meaning a wave overtopping of the OWC for a given maximum overtopping rate Q_{max}. Once all the nodes are defined in the toolbox, the network is reduced using the Adaptive Line Sampling method. In the Figure 7 can be observed the reduced network used for this study. The overtopping events are given in the continuous space of the *Wave_amplification* node for each given state of *Experiment_case*.

In this simple network there is only one p-box employed per simulation (regarding the *Wave_amplification* node), so the lower bound would correspond to the minimum amplification that a wave under those conditions can have. Thus, overtopping probability resulting from this calculation will be the minimum probability that this variable can have, assuming monotonicity in the system. In order words, given the lower bound of *Wave_amplification* the lower bound of *Overtopping* will be found. The same reasoning is used for the case of the upper bounds.

The lower and upper bounds of the overtopping probability are displayed in the Table 2, for the case of harbour wall length experiment, and in Table 3, for the case of harbour wall inclination experiment. In each *Wave Height* column in the tables are stored the overtopping probability bounds computed for each height, i.e., 0.03, 0.06 and 0.09 m for each of the different crest periods (T_p). This was done in order to provide a combinatorial study of the overtopping probability change for all the experimental set ups.

It can be observed that T_p does not influence significantly the structure overtopping occurrence. In fact, changes in overtopping results obtained from the different crest periods may be due to the randomness of the simulation functions used. A major influence comes mainly from the *Wave_height* and *Experiment_case*. It is logical that higher waves will increase the height of the amplified wave, in consequence, the probability of an overtopping event will increase as well. Comparing the results from both experiment arrangements, it can be resolved that the wall inclination factor increases the most the probability of overtopping occurrence. The wave amplification factor of 0.237 corresponding to a wall inclination of $3\pi/4$ with 0.09 *m* wave height can be referred as the worse case scenario. This can be useful when the values are only scaled up to prototype dimensions. For instance, a structure overtopping assessment can be provided without performing any experimental work, as long as the physical behavior of the variables preserves the same probabilistic distribution.

Table 3. Occurrence of structure overtopping for wall inclination case. The bounded values in $Overtopping_{a,b,c}$, correspond to the *Wave_amplification* p-boxes in each of the *Wave_height* cases, i.e. 0.03, 0.06 and 0.09 *m*, respectively.

Period (*s*)	$Overtopping_a$	$Overtopping_b$	$Overtopping_c$
1	[0, 0.017]	[0.115, 0.34]	[0.455, 0.623]
1.25	[0, 0.015]	[0.109, 0.342]	[0.451, 0.623]
1.5	[0, 0.015]	[0.111, 0.337]	[0.456, 0.631]
1.75	[0, 0.017]	[0.114, 0.343]	[0.453, 0.629]
2	[0, 0.015]	[0.11, 0.341]	[0.449, 0.621]
2.25	[0, 0.017]	[0.111, 0.34]	[0.455, 0.623]
2.5	[0, 0.017]	[0.113, 0.341]	[0.455, 0.621]

4 CONCLUSIONS

From the case study presented here, an approximation to the worst case scenario was found with the use of a very simple Bayesian network with the implementation of p-boxes and interval variables (credal sets). The original network containing continuous, interval and discrete variables was reduced using structural reliability methods (adaptive line sampling) to a simpler network containing only crisp probabilities. Since the values used as input in the p-boxes and in the interval variables contain all the possible cases in the experiment, none of the those should overcome the maximum probability given. So, epistemic uncertainty affecting this experiment can be quantified with this method. If specific data (different from that considered here) regarding any of the variables in the experiment, inside the bounds, are given, the overtopping probability results can be provided for that specific case and they will remain below or be equivalent than the maximum value achieved.

The implementation of imprecise data in Bayesian networks is a very necessary tool in engineer-

ing. However, the methods currently available are computationally expensive. Most importantly, when systems with a large number of variables are studied, the algorithms may suffer combinatorial explosion. Adaptive line sampling will be further studied to deal with bounded failure probabilities. In addition to that, random set theory (combined with Subset Simulation to improve calculation speed) appears as a good option to the limitations of the current tool, since the latter method is specially useful for small probability cases and good performance in both low- and high-dimensional spaces as well as in nonlinear limit functions.

ACKNOWLEDGEMENTS

The first author gratefully acknowledges the *Consejo Nacional de Ciencia y Tecnología (CONACyT)* for the award of a scholarship from the Mexican government for graduate studies. Also, special thanks to the European Unions Research and Innovation funding programme (Framework Programme) under the PLENOSE project (PIRSES-GA-2013–612581) for partially supporting this work.

REFERENCES

Allsop, N.W.H., L. Franco, & P.J. Hawkes (1985). Wave run-uP on steep slopes-A literature review. *Hydraulics Research 1*, 1–28.

Castillo, E., C. Solares, & P. Gomez (1997). Tail uncertainty analysis in complex systems. *Artificial Intelligence 96*(2), 395–419.

Cozman, F.G. (2000). Credal networks. *Artificial Intelligence 120*(2), 199–233.

Cruz, J. (2008). *Ocean Wave Energy Current Status and Future Perspectives*. Bristol, U.K.: Springer Berlin Heidelberg.

Daniel Raj, D., V. Sundar, & S. Sannasiraj (2016a). Experimental Investigation on Optimizing the Projecting Sidewalls of an Oscillating Water Column. *Proc., 9th Int. Conf. on Coastal Port Engineering in Developing Countries (Pianc-Copedec IX) IX*(c), 1–11.

Daniel Raj, D., V. Sundar, & S. Sannasiraj (2016b). Optimizing the harbor wall inclination in an oscillating water column. In *3rd International Conference on Coastal Zone Engineering And Management In The Middle East*, Dubai,UAE.

de Angelis, M., E. Patelli, & M. Beer (2015). Advanced Line Sampling for efficient robust reliability analysis. *Structural Safety 52*(PB), 170–182.

Falcão, A.F.d.O. (2010). Wave energy utilization: A review of the technologies. *Renewable and Sustainable Energy Reviews 14*(3), 899–918.

Ferson, S., V. Kreinovich, L. Ginzburg, D.S. Myers, & K. Sentz (2003). Constructing Probability Boxes and DempsterShafer Structures. *Sandia National Laboratories* (January), 143.

Hamada, M.S., A.G. Wilson, C.S. Reese, & H.F. Martz (2008). *Bayesian Reliability*. Springer Series in Statistics. New York, NY: Springer New York.

Hasofer, a. M. & N.C. Lind (1974). An exact and invariant first order reliability format. *Journal of the Engineering Mechanics Division ASCE 100*(August), 111–121.

Hosseini, S. & K. Barker (2016). A Bayesian network model for resilience-based supplier selection. *International Journal of Production Economics 180*, 68–87.

Hudson, L.D., B.S. Ware, K. Blackmond, & S.M. Mahoney (2002). An Application of Bayesian Networks to Antiterrorism Risk Management for Military Planners. *Digital Sand-box, Inc.*, 8.

Jensen, F.V. & T.D. Nielsen (2007). *Bayesian Networks and Decision Graphs*. Information Science and Statistics. New York, NY: Springer New York.

Kang, C. & M. Golay (1999, jul). A Bayesian belief network-based advisory system for operational availability focused diagnosis of complex nuclear power systems. *Expert Systems with Applications 17*(1), 21–32.

Korb, K.B. & A.E. Nicholson (2004). *Bayesian Artificial Intelligence*. Boca Raton, U.S.A.: Chapman & Hall/CRC.

Mase, H., T. Tamada, T. Yasuda, T.S. Hedges, & M.T. Reis (2013). Wave Runup and Overtopping at Seawalls Built on Land and in Very Shallow Water. *Journal of Waterway, Port, Coastal, and Ocean Engineering 139*(October), 346–357.

Owen, M.W. (1980). Design of seawalls allowing for wave over-topping. *Report Ex 924*(June), 39.

Patelli, E. (2015). COSSAN: A Multidisciplinary Software Suite for Uncertainty Quantification and Risk Management. In R. Ghanem (Ed.), *Handbook of Uncertainty Quantificatio*, pp. 1–69. Switzerland: Springer International Publishing.

Patelli, E., S. Tolo, H. George-Williams, J. Sadeghi, R. Rocchetta, M. de Angelis, & M. Broggi (2018). OpenCossan 2.0: an efficient computational toolbox for risk, reliability and resilience analysis. In *Proceedings of the joint ICVRAM ISUMA UNCERTAINTIES conference*.

Pearl, J. (1991). Probabilistic Reasoning in Intelligent Systems.

Spiegelhalter, D.J. (1987). A unified approach to imprecision and sensitivity of beliefs in expert systems. In *Uncertainty in Artificial Intelligence*, Seattle, WA, USA, pp. 199–208.

Straub, D. & A. Der Kiureghian (2010). Bayesian Network Enhanced with Structural Reliability Methods: Methodology. *Journal of Engineering Mechanics*.

Tolo, S., E. Patelli, & M. Beer (2016a, jun). Risk Assessment of Spent Nuclear Fuel Facilities Considering Climate Change. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 3*(2), G4016003.

Tolo, S., E. Patelli, & M. Beer (2016b). Robust vulnerability analysis of nuclear facilities subject to external hazards. *Stochastic Environmental Research and Risk Assessment*, 1–24.

Tolo, S., E. Patelli, & M. Beer (2018, jan). An open toolbox for the reduction, inference computation and sensitivity analysis of Credal Networks. *Advances in Engineering Software 115*, 126–148.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities* (1st ed.). London, U.K.: Champan and Hall.

Weber, P., G. Medina-Oliva, C. Simon, & B. Iung (2012). Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence 25*(4), 671–682.

Uncertainty analysis

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Interval-based parameters for stress diffusion in granular medium

D. Boumezerane
Middle East Technical University, Northern Cyprus, METU
Formerly Norwegian University of Science and Technology, NTNU, Norway

ABSTRACT: According to Bourdeau (1986), diffusion of stresses in a granular medium can be described using a probabilistic approach. A point load applied on the surface of a granular media will follow an erratic path, depending on the probability of transition between the grains. The diffusion of the expected vertical stress in the granular medium can be described by a Fokker-Planck type equation. In terms of expected vertical stresses, an equation of diffusion is obtained and the parameter of diffusion is shown to approximate the coefficient of lateral pressure of the material at a given depth z. The coefficient of lateral pressure of the material can be expressed in terms of intervals with upper and lower values to account for uncertainty.

In the present approach, we propose to solve the diffusion equation using interval-based parameters to account for uncertainty. Uncertain parameters are considered as discretized fuzzy numbers; they are combined with finite difference method to solve the diffusion equation. Comparisons are made with experimental and available data.

1 INTRODUCTION

Diffusion of stresses in granular media is a phenomenon that provokes rearrangement in grains and thus settlements. Soils as granular media are constituted of an assembly of particles of different sizes, mineralogy and morphologies. The particle sizes vary from less than 0.002 mm in clays to some tens of millimeters in gravel materials. Despite their granular aspect, they are considered, from the soil mechanics viewpoint, as a continuum. Theories of continuous media are often applied to model the behavior of cohesive materials such as clays. Cohesionless soils behavior on the other hand is difficult to capture using continuum approaches. Their granular aspect, especially when they are in loose states, makes their behavior complex to predict using conventional theories of continuum media.

Harr (1977) proposed an approach to estimate the expected vertical stress in a granular medium subjected to surficial loading. A concentrated load applied on the surface of a semi-infinite ground will follow an uncertain path between the grains. The resulting stress in one point is a random variable. Its distribution will reflect the composition of the media. Following a binomial law of diffusion on one side or the other given a reference mark, the transmission of the load in the granular medium will be expressed as an equation of diffusion of stresses. The main parameter characterizing the equation is the coefficient of lateral pressure in soils. Bourdeau (1986) extended this approach and proposed a formulation in terms of displacements diffusion in loose cohensionless granular medium.

We revisited the formulation in terms of stress diffusion to account for parameter uncertainty using intervals which can be combined and expressed also as fuzzy discretized numbers. A numerical approach based on finite difference method is combined with interval-based parameters to simulate diffusion of stresses in granular media. Interval-based parameters are used to account for uncertainty and comparisons are made with experimental data. The aim of this research is to study parameter uncertainty and its quantification in the context of stress diffusion in a granular medium.

2 DIFFUSION OF STRESSES IN A GRANULAR MEDIUM

Diffusion of stresses in soils is a fundamental aspect of soil mechanics. It is observed in all soil-structure interaction problems as a consequence of applied loads. Harr (1977) build a model for stress diffusion in a granular medium based on the idea of expected stress distribution at a point. A unit force applied on the surface will follow a random path between the grains. The vertical normal stress acting at a point within a medium is the total accumulated effect of many random variables; shape and distribution of particles, the spatial distribution of voids as well as their local configurations.

The central limit theorem assures that the distribution of stress will converge to the normal distribution as the number of particles becomes large (Harr, 1977).

In Figure 1 is shown, schematically, the transmission of vertical forces between particles. The effect of the boundary force will spread laterally in the positive and negative directions. At a representative particle, the input stress can be taken to divide in the left and right directions consistent with a Bernoulli trial. The division of stresses is expected to be equal. For a random distribution the frequency of moving to the left is the same as moving to the right. Figure 2 shows the distribution of stresses within a homogeneous random medium. If Δx is the average spacing of the stresses at row, with the rows taken to be Δz apart, the expected stress will follow a binomial distribution with the recurrence equations (Harr, 1977):

$$\bar{S}_z[x, z+\Delta z] = \frac{1}{2}\left\{\bar{S}_z[x-\Delta x, z] + \bar{S}_z[x+\Delta x, z]\right\} \tag{1}$$

Subtracting the expected vertical stress $\bar{S}_z[x, z]$ from each side of the expression and dividing by Δz, we get:

$$\frac{\bar{S}_z[x, z+\Delta z] - \bar{S}_z[x, z]}{\Delta z} = \frac{(\Delta x)^2}{2\Delta z} \left\{\frac{\bar{S}_z[x+\Delta x, z] - 2\bar{S}_z[x, z] + \bar{S}_z[x-\Delta x, z]}{(\Delta x)^2}\right\} \tag{2}$$

In the limit as Δx and Δz become very small, the equation can be expressed as a differential one:

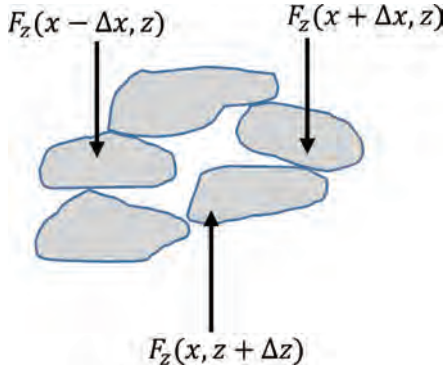


Figure 1. Transmission of vertical forces.

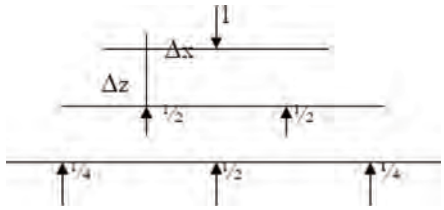


Figure 2. Unit stress distribution in terms of probabilities.

$$\frac{\partial \bar{S}_z}{\partial z} = D\frac{\partial^2 \bar{S}_z}{\partial x^2} \tag{3}$$

where $D = \lim_{\Delta x, \Delta z \to 0}\left\{\frac{1}{2}\frac{\Delta x^2}{\Delta z}\right\}$

The ratio $\frac{\Delta x^2}{\Delta z}$ is seen to reflect a characteristic of the particulate medium. Harr (1977) has shown that ($D = v.z$) is dependent on the depth z and the coefficient of lateral stress v. The author also showed that v can be approximated by the coefficient of earth pressure at rest K (using Jaky's formula for example).

Equation (3) is an equation of diffusion of expected vertical stresses in a granular media, with D the parameter of characterization.

The main objective of our research is to study the variability of the coefficient D and its influence on the diffusion of stresses in a granular medium. Uncertainty analysis will be carried out based on interval analysis and fuzzy representations of D. Different sources of information are used to consider the coefficient of diffusion.

2.1 *Uncertainty in the diffusion coefficient*

Uncertainties in geotechnics are generally classified into two categories, "epistemic" related to lack of data or knowledge and "aleatory" related to natural randomness (Baecher & Christian, 2003). The diffusion parameter D is subjected to epistemic uncertainties which have influence on the diffusion process in the granular medium. The main sources of uncertainty of D come from evaluation of lateral pressure coefficient K as $= K.z$. In situ and laboratory measures are used to evaluate lateral pressure coefficient in soils via empirical relations (Jaky's formula for example). Empirical correlations combining Jaky's formula $(1-\sin\varphi')$ with OCR from laboratory are also used to determine the lateral pressure coefficient at rest K_0 (Cai et al. 2011). When experimental results are given in terms of intervals the average values are generally used.

Figure 3 shows variation of lateral earth pressure based on different tests (Chen & Fang, 2008).

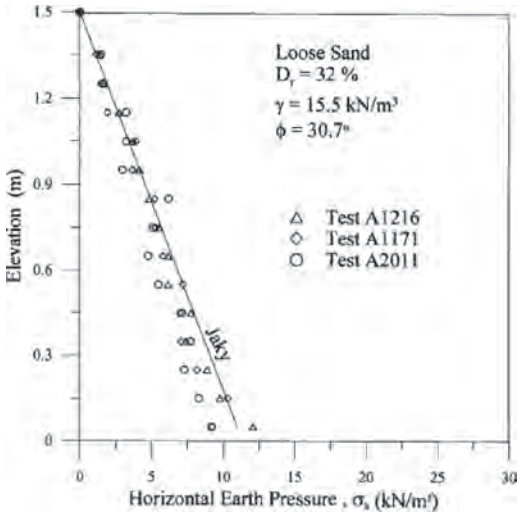


Figure 3. Distribution of horizontal earth pressure against model wall (Cheng & Fang, 2008).

Chen & Fang (2008) studied variation of lateral earth pressure in a block of sand. They presented experimental data on the variation of lateral earth pressure against a non-yielding retaining wall due to soil filling and vibratory compaction.

Schmertmann et al. (2005) used a K-Box device to describe stress diffusion in sand under a surface circular plate loading. The experiment intended to improve understanding of stress distribution in a particulate medium. Comparisons were made with obtained stresses from the probabilistic approach (Harr 1977). According to the authors the particulate-probabilistic theory seems approximately correct when $K = K_a$.

As we can observe, uncertainties are inherent to the testing method, in other terms using interval values can be suitable to handle the perturbations. Intervals are often used as a tool to handle uncertainties which arise during experiments. In different types of tests uncertainties propagate during measuring procedures and from the use of different devices. Using Jaky's Formula for example uncertainty lies essentially in measuring the material angle of friction φ′. Whatever the techniques used for measuring such a parameter, there are uncertainties to consider.

3 FORMULATION OF INTERVAL-BASED PARAMETERS FOR DIFFUSION

Interval analysis was introduced by Moore (1966) and is considered as a mathematical discipline that deals with quantities expressed as intervals which are common in engineering problems. Intervals are a convenient tool to deal with uncertainty when there is lack of data. As probabilistic approaches require important amounts of data, when in geotechnics information is scarce and small data available in general, the use of interval analysis could be of interest. In practice, it may be difficult sometime to get a large number of experimental data so we need an alternative method in which we may handle the uncertainty with few experimental data. In our problem, the diffusion parameter D will be considered as interval-based. Applied pressure on the ground can also be considered as an uncertain parameter expressed in terms of interval.

Finite difference schemes are usually applied to solve the type of diffusion equation we have in our case. We can either use a forward or backward scheme to solve it when dealing with deterministic values of the parameter D.

$$\frac{\partial \bar{S}_z}{\partial z} = D\frac{\partial^2 \bar{S}_z}{\partial x^2}$$

where

$$D = K.z$$

(For simplicity we use S instead of $\bar{S}_z$)

In finite difference method, the forward scheme can be expressed by:

$$S_j^{i+1} = S_j^i + Kz^i \frac{\Delta z}{(\Delta x)^2}\left(S_{j+1}^i - 2S_j^i + S_{j-1}^i \right) \qquad (4)$$

Index i is used in z direction and j in x direction.

The formulation of the problem to deal with interval valued parameters is conducted following interval arithmetic and rules. In Moore et al. (2009), Hanss (2005) we can find more details on derivation and application of interval arithmetic.

3.1 *Reminder on intervals*

Interval valued approach gained significant use in engineering especially when information is uncer-

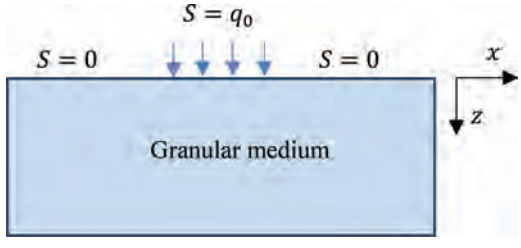


Figure 4. Scheme and Boundary Conditions.

tain. Studies were conducted in different fields, such as thermal conductivity to account for uncertainty using interval approaches and fuzzy sets (Wang 2014). Wang (2014) proposed a new numerical technique named as fuzzy finite difference method to solve the heat conduction problems with fuzzy uncertainties in both the physical parameters and initial/boundary conditions. The α level-cut method is used to study the problem in terms of interval equations. Kermani & Saburi (2007) presented a numerical method for solving "Fuzzy Partial Differential Equation" (FPDE) with some examples.

In soil mechanics the solutions for stress diffusion are based on continuum mechanics which considers the media as a continuum. Even constituted of particles, the granular aspect of the soil is not considered. The probabilistic approach from Harr (1977) considers the soil as a particulate media. The diffusion equation (1) is used with a parameter of diffusion D as a deterministic value. Deterministic values of D are not common, but we tend to use approximations. In the following the equation of diffusion of stresses in a granular medium will be considered in terms of intervals. The parameter of diffusion D and initial/boundary conditions are taken as intervals to account for uncertainty.

Using $\tilde{S}$, equation (3) can be written in terms of intervals as follows;

$$\frac{\partial \tilde{S}}{\partial z} = \tilde{D}\frac{\partial^2 \tilde{S}}{\partial x^2} \tag{5}$$

$\tilde{S}$ is replaced by S_α^I. We notice S_α^I for a given interval I of α level. $\underline{S}(\alpha), \overline{S}(\alpha)$ are the lower and upper values of the interval at α level

$$\frac{\partial \tilde{S}}{\partial z} = \frac{\partial \left[\underline{S}(\alpha), \overline{S}(\alpha) \right]}{\partial z}$$
$$\frac{\partial S_\alpha^I}{\partial z} = \frac{S_\alpha^{I.i+1} - S_\alpha^{I.i}}{\Delta z} \tag{6}$$

$$\frac{\partial^2 \tilde{S}}{\partial x^2} = \frac{\partial^2 S_\alpha^I}{\partial x^2} = \frac{S_{\alpha,j+1}^I - 2S_{\alpha,j}^I + S_{\alpha,j-1}^I}{(\Delta x)^2} \tag{7}$$

Then the forward scheme of finite difference approach using interval-based parameters can be written;

$$S_{\alpha,j}^{I,i+1} = S_{\alpha,j}^{I,i} + D_{\alpha,j}^{I,i}\frac{\Delta z}{(\Delta x)^2}\left[S_{\alpha,j+1}^{I,i} - 2S_{\alpha,j}^{I,i} + S_{\alpha,j-1}^{I,i} \right] \tag{8}$$

where

$$D_{\alpha,j}^{I,i} = K_{\alpha,j}^{I,i} z_j^i \tag{9}$$

α level cuts are used to characterize a fuzzy number.

A triangular fuzzy number, denoted by $u = \langle a, b, c \rangle$ where $a \le b \le c$ has α-cuts

$$[u]_\alpha = [a + \alpha(b-a), c - \alpha(c-b)], \alpha \in [0,1]$$

And membership function

$$\mu_{Tri}(x) = \begin{cases} \dfrac{x-a}{b-a} & if\ a \le x \le b \\ \dfrac{c-x}{c-b} & if\ b \le x \le c \\ 0 & otherwise \end{cases}$$

Elementary operations of interval arithmetic are used (Hanss, 2005).

$$[a_1, b_1] + [a_2, b_2] = [a_1 + a_2, b_1 + b_2]$$
$$[a_1, b_1] - [a_2, b_2] = [a_1 - b_2, b_1 - a_2]$$

$$[a_1, b_1] \times [a_2, b_2] = [\min(M), \max(M)]$$
$$M = \{a_1 a_2, a_1 b_2, b_1 a_2, b_1 b_2\}$$

4 CASE STUDY

Turedi and Ornek (2017) performed laboratory experiment on sand to investigate the stress and bearing capacity. The vertical stresses resulting from strip and rectangular footings are measured at different depths of the tank model of dimensions 1.25m length, 1.0 m width and 1.0 m depth. The loading is considered under a model footing of B = 0.1 m breadth and L = 5B length. The measured average peak friction angles ϕ were 36° and 42° for loose and dense sands, respectively. The results of measured vertical stress (kPa) are given in Figure 5.

The model was calibrated using the results from Turedi and Ornek (2017). Figure 6 shows the vertical stress distribution using the probabilistic approach when we consider a deterministic diffusion parameter $D = K.z$. Jaky's formula is used, $K = 1 - \sin\phi'$. As the experiment was performed in loose sand we used for our simulation $\phi' = 36°$ as given by the authors.

The purpose of our study is to take into account uncertainties using interval-valued parameters. The proposed model permits to consider the parameter of diffusion D in terms of intervals, or as triangular fuzzy number. The loading on the surface can be taken into account as interval also.

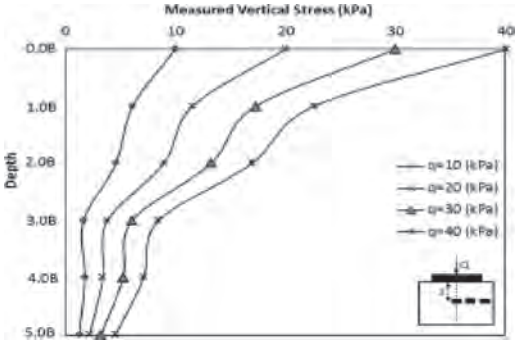


Figure 5. Vertical stress distribution along the depth at different loading levels (Experiment Turedi & Ornek, 2017).

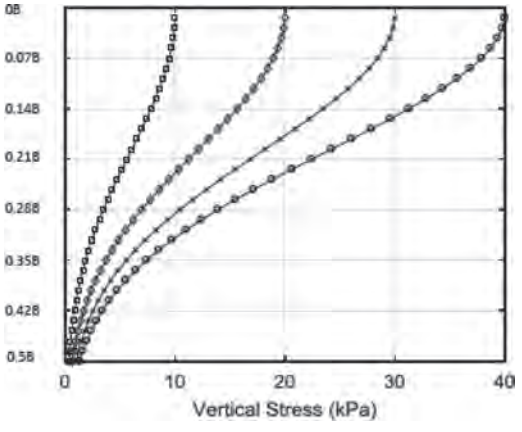


Figure 6. Vertical stress distribution along the depth at different loading levels (Probabilistic approach, $K = 1 - \sin\varphi'$).

Considering ranges of values for the coefficient of lateral pressure K around the friction angle of sand φ', the used interval values of K are given at different α level cuts. The applied load is taken as point valued q = 40 kPa. The triangular fuzzy number for K can be written in terms of intervals at level α, $[K]_\alpha = [0.35 + 0.03\alpha, 0.41 - 0.03\alpha]$.

And the membership function is given as

$$\mu_K(x) = \begin{cases} \dfrac{x - 0.35}{0.03} & \text{if } 0.35 \le x \le 0.38 \\ \dfrac{0.41 - x}{0.03} & \text{if } 0.38 \le x \le 0.41 \\ 0 & \text{otherwise} \end{cases}$$

Figure 7 shows distribution of vertical stress at α cut levels; $\alpha 0 = 0, \alpha 1 = 1/3, \alpha 2 = 2/3$ and $\alpha 3 = 1$.

Parameter D being dependent on depth, it is noticed that uncertainty is more pronounced with z increasing. Uncertain parameter of diffusion has important impact on stress distribution, especially at deeper levels. The observed dispersion suggests that using deterministic values for D gives only an approximation of the stress. It can underestimate or overestimate the real level of stress in the medium. Using interval values for the diffusion parameter helps estimating the evolution of stress spreading with depth.

One can also consider uncertainty in the loading using intervals for the surficial charge q. Influence of interval values of q is shown in Figure 8 for $q = [\underline{q}, \overline{q}] = [38\text{ kPa}, 40\text{ kPa}]$. The parameter of diffusion is kept point valued in this case with $K = 0.38$.

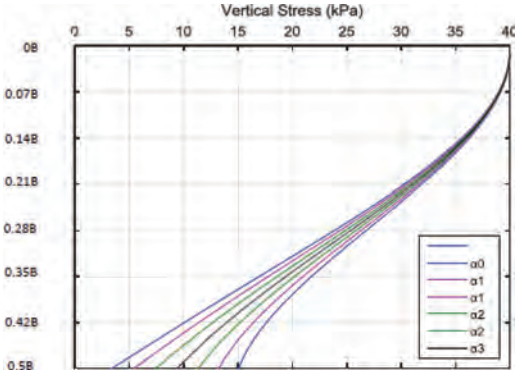


Figure 7. Vertical stress distribution under axis for different level cuts of K.

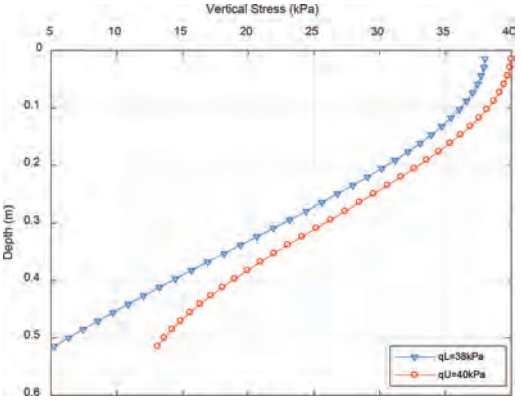


Figure 8. Vertical stress distribution under axis for inter-val surficial pressure $[q, \overline{q}]$.

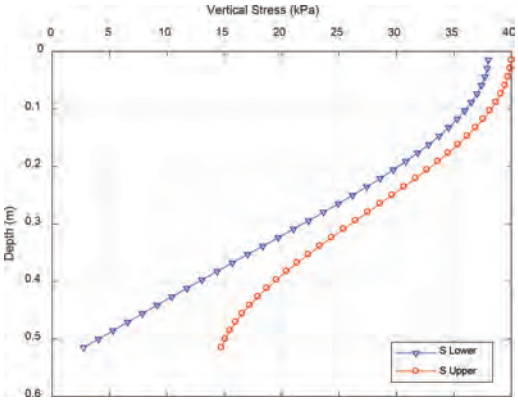


Figure 9. Vertical stress distribution for inter-val surficial pressure $\left[\, q, \overline{q} \,\right]$ and interval coefficient $\left[\, \underline{K}, \overline{K} \,\right]$.

The uncertainty on the load has effect on the distribution as the diffusion shows the vertical stress evolving as an interval from the surface, and it continues with the depth. It is slightly augmented for z deeper than half of the domain.

Combination of uncertainty due to K and q has higher effect on the evolution of uncertainty in vertical stress as it is illustrated in Figure 9. The surficial pressure is kept as $\mathrm{q} = \left[\, 38\ \mathrm{kPa}, 40\ \mathrm{kPa} \,\right]$ and the lateral pressure coefficient $K = \left[0.35, 0.41\right]$.

As uncertainty is combined between the surficial pressure q and the coefficient K we notice fast dispersion in the vertical stress S with depth as it was illustrated in Figure 7 already, but the effect is more pronounced because uncertainty originates from both the load and the soil parameter.

5 CONCLUSION

The probabilistic approach from Harr (1977) needed only one parameter for characterizing the particulate medium. It was shown in (Bourdeau, 1986) that lateral pressure coefficient can be a good estimate of the parameter for vertical stress diffusion in a granular soil. We revisited the theory from Harr to account for uncertainty in the diffusion parameter D and in the surficial loading. The purpose was to show the ability of interval valued method to handle uncertainty due to lack of knowledge and data. The finite difference scheme is practical for this type of equations. The construction of such a scheme for interval analysis is not straight forward as it obeys number of specific conditions. It was shown that uncertainty in the diffusion parameter affects significantly the distribution of vertical stress in the granular medium. And, when combined with the uncertainty from surficial pressure the stress distribution shows more dispersion with depth. Interval valued parameters are suitable when dealing with lack of data and uncertainty.

REFERENCES

Baecher GB. & Christian JT. 2003. Reliability and Statistics in Geotechnical Engineering. Wiley.

Bourdeau PL. 1986. Analyse probabiliste des tassements d'un massif de sol granulaire. These de Doctorat es Sciences Techniques, No. 628, Swiss Federal Institute of Technology.

Cai G, Liu S, Puppala AJ, Tong L. 2011. Assessment of the coefficient of lateral earth pressure at rest (K0) from in situ seismic tests. *Geotech Test J 2011;34(4):1–11*.

Chen TJ & YS Fang. 2008. Earth Pressure due to Vibratory Compaction. *Journal of Geotechnical and Geoenvironmental Engineering. 2008.134:437–444*.

Hanss M.2005. Applied Fuzzy Arithmetic, An Introduction with Engineering Applications. Springer Verlag.

Harr ME. 1977. Mechanics of particulate media: A probabilistic approach. McGraw-Hill.

Kermani MA & F. Saburi. 2007. Numerical Method for Fuzzy Partial Differential Equations. *Applied Mathematical Sciences, Vol. 1, 2007, no. 27, 1299–1309*.

Moore, RB Kearfott & MJ Cloud. 2009. Introduction to interval analysis. Society for Industrial and Applied Mathematics SIAM, Philadelphia.

Nayak S. & S. Chakraverty. 2015. Numerical solution of moving plate problem with uncertain parameters. *arXiv:1503.07809v1*.

Schmertmann JH. 2005. Stress Diffusion in Sand. *Journal of Geotechnical and Geoenvironmental Engineering. Vol.131, No1, January 2005*.

Turedi Y. & M. Ornek. 2017. Stress Analyses of Strip and Rectangular Footings Rested on Loose Sands. *NESciences, 2017, 2 (3): 93–112*.

Wang C. & Qiu ZP. 2014. Fuzzy finite difference method for heat conduction analysis with uncertain parameters. *Acta Mechanica Sinica (2014) 30(3):383–390. DOI 10.1007/s10409-014-0036-7*.

Wang C. & Qiu ZP. 2014a. Interval finite difference method for steady-state temperature field prediction with interval parameters. *Acta Mechanica Sinica (2014) 30(2):161–166. DOI 10.1007/s10409-014-0020-2*.

Zureigat HH & AM Ismail. 2016. Numerical Solution of Fuzzy Heat Equation with Two Different Fuzzifications. *SAI Computing Conference 2016.July 13-15, 2016. London, UK*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Uncertainty sensitivity assessment on the optimization of the design and operation of complex energy systems: A comprehensive approach

A. Nadal, A. Ruby & C. Bourasseau
CEA, LITEN, DTBH, Université Grenoble Alpes, Grenoble, France

D. Riu
G2Elab, CNRS, Grenoble Institute of Engineering, Université Grenoble Alpes, Grenoble, France

C. Bérenguer
GIPSA-lab, CNRS, Grenoble Institute of Engineering, Université Grenoble Alpes, Grenoble, France

ABSTRACT: For the optimization of renewable energy systems, uncertainties associated to technical and economic parameters are scarcely taken into account, which may lead to a weak confidence in results. In this paper, we propose and investigate a 4-steps methodology for uncertainty sensitivity assessment, with the objective to improve confidence in the assessment results and to support decision-making process following techno-economic optimization of the design and the operation of an autonomous power system. The methodology is applied to off-grid system including photovoltaic production, battery and hydrogen components (electrolyser, pressure storage and fuel cell). This energy system is modelled and optimized with Odyssey—a simulation software developed by CEA-LITEN since 2010. We focus on static parametrical uncertainties, linked to the energy system parameters.

1 INTRODUCTION

1.1 *Optimization of complex energy systems*

Energy systems are getting more and more complex, and difficult to assess because of (i) the variability of the renewable power sources and of the demand, (ii) the resultant necessity of storage and (iii) the presence of different and new energy vectors. The modelling and simulation software Odyssey (Guinot 2013) enables the realization of techno-economic optimizations of such energy systems design and operation. However, many parameters used to simulate the systems are uncertain (e.g. static component performances or economic properties, but also time series of production or load profiles) and it is necessary to evaluate the impact of these uncertainties on the design and operation arising from the optimization process to help decision-making about these systems.

1.2 *Problem statement*

Up to now techno economic studies carried out with Odyssey, as with most other similar simulation tools, have not taken into account uncertainties, but only have provided sensitivity analysis on uncertain key input parameters. Thus, the objective of our work is to develop a comprehensive approach to enhance the platform with capacities of uncertainty management, from the identification of the main sources of uncertainty to results analysis and support to decision making.

We identified two main ways to account for the uncertainty influence on the results of a techno-economic optimization. The first one consists in optimizing the system taking into account the uncertain parameters so as to get results robust to the considered uncertainties. The second way consists in optimizing the system and then apply the uncertainties to evaluate the sensitivity of this optimized design to uncertainties; this paper presents an application of this second one. In the first part, we will present a 4-step methodology for uncertainty assessment. In the second part, we will present a representative example of energy system optimization, which allows dealing with problematic of competition between technologies, the problematic of energy storage in off-grid power-system and the optimization of this system design and operation without uncertainty. In the third part, we will apply our methodology on the described study case, showing how we modelled uncertainties on selected parameters and how to assess the sensitivity of results to these uncertainties. Finally, the fourth part will expose our conclusions and give directions for our future research work.

2 PROPOSED METHODOLOGY

The methodology of uncertainty treatment proposed and implemented in this work is based on the four-steps approach described by de Rocquigny (de Rocquigny 2006a, b), schematized in Figure 1. This methodology is summarized below.

2.1 *Model of the system*

For step A, we assume that the considered energy system model is available and implemented in the software Odyssey, used as a black box.

The entries of this black box are technical and economic parameters divided in two types: design variables and uncertain parameters. The uncertain parameters are arbitrarily decided by the decision-maker or by the software user. On the contrary, the design variables can be set, and even optimized, like the optimization variables in Table 1. The next part will describe with precision the model of our case study.

The outputs of the black box are technical and economic indicators, which assess the performances of the design of the system. These indicators will be described later.

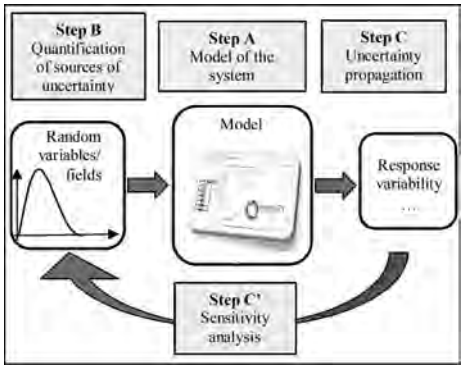


Figure 1. Uncertainty analysis common framework.

Table 1. Optimization variables.

Variable	Units	Optimization borders Minimum	Maximum
Number of Modules PV*	–	1	no
Number of Battery Units**	–	1	150
Number of electrolyze cells	–	5	no
Fuel Cell Stack Max Power	W	1	no
Volume of pressure tank	m^3	1	no

*Each module has a peak power of 1 kWp.
**Each unit has a rated capacity of 10 kWh.

2.2 *Uncertainty modelling*

Regarding the quantification of sources of uncertainty, at step B, we model the sources of uncertainty in a probabilistic framework, with probabilistic density laws. In this study, we assume that all the considered uncertainties are independent.

The three classical probabilistic laws are consid-ered: uniform, beta and Weibull, detailed in Table 3, in paragraph 4.1 where the whole uncertainty modelling is applied to our study case.

2.3 *Uncertainty propagation*

At step C, the propagation of uncertainties allows us to see how the outputs of the model respond to the uncertainties: we achieved this by coupling a Monte Carlo launcher provided by the Uranie software (Bouloré 2012) and the executable Odyssey, as schematized in Figure 2.

2.4 *Sensitivity analysis*

Finally, at 'step C', the sensitivity analysis permits to identify the uncertainties that have the strongest influence on the outputs of the model. This identification gives us the possibility to try to reduce the uncertainty of the most influent sources, in order to reduce the uncertainty of the outputs and facilitate decision making (Borgonovo 2016). Among the different methods of sensitivity analysis, we chose the Morris method and the Sobol indexes computation, which are closely complementary.

The Morris method (Morris 1991) allows first to classify the uncertain parameters in three categories:

- the parameters with negligible effects,
- the parameters with linear effect and without interaction,
- the parameters with nonlinear effects and/or interactions (without distinction of these two effect types).

Table 2. Optimized study cases design and indicators.

Case	0	01	05	1
Number of Modules PV (–)	735	735	660	600
Number of Battery Units (–)	146	145	135	138
Number of electrolysis cells (–)	8	5	5	5
Fuel Cell Stack Max Power (W)	43,500	10,500	5000	5000
Volume of pressure tank (m^3)	31	16	3.5	3.5
Unsatisfied load (%)	0	0.1	0.5	1
LEC (€/MWh)	404.9	336.1	295.5	280.2
Unused Primary Production (%)	39.2	39.7	33.1	26.8

Table 3. Uncertain parameters and associated probability distributions (Uniform, Beta or Weibull).

Component Parameter Unit	Law	Reference
PV		
CAPEX *€/Wp*	β [α = 1.8; β = 6; Min = 0.374; Max = 3.165]	IRENA (2016)
OPEX *% CAPEX*	U [2; 10]	i.d.*
Battery bank		
CAPEX *€/Wh*	β [α = 1.31; β = 3.5; Min = 0.102; Max = 0.354]	Battke et al. 2013
OPEX *% CAPEX*	U [2; 10]	i.d.*
Capacity loss *Wh/h*	U [1.4E-5; 4.2E-5]	Riffonneau et al. 2007
Self-discharge *W*	U [3.75E-5; 1.4E-4]	IRENA 2017
Charge efficiency –	β [α = 1; β = 4; Min = 0.8; Max = 0.9]	Battke et al. 2013
Discharge efficiency –	β [α = 1; β = 4; Min = 0.8; Max = 0.9]	Battke et al. 2013
Electrolyser		
CAPEX *€/W*	U [6.5; 13.1]	i.d.*
OPEX *% CAPEX*	U [2; 10]	i.d.*
Degradation *µV/h*	U [0.4; 15]	Bertuccioli et al. 2014
Cell voltage**	U [1.39; 1.54]	i.d.*
FC		
CAPEX *€/W*	U [2.2; 8]	i.d.*
OPEX *% CAPEX*	U [2; 10]	i.d.*
Degradation *µ%/h*	U [0.45; 1.35]	Kurtz et al. 2015
Efficiency**	U [0.30; 0.34]	FutureE Fuel Cell Solutions GmbH
H_2 tank		
CAPEX**	U [18,055; 28,239]	i.d.*
OPEX % CAPEX	U [2; 10]	i.d.*

*internal data.
**see 4.1.1.

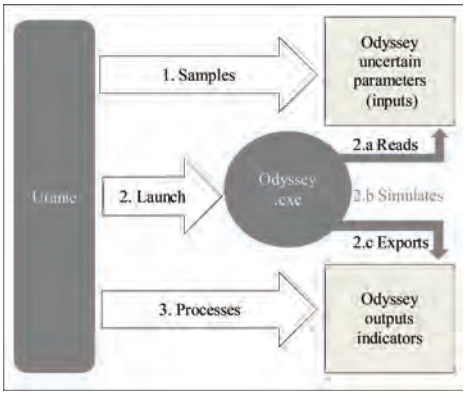


Figure 2. Odyssey/Uranie coupling.

This screening method presents the advantage to sort the uncertain parameters with a limited calculation cost. In fact, the Morris method requires N code computations, with:

$$N = r*(d+1) \quad (1)$$

with:

– r∈ $[\![4;10]\!]$,
– d: number of uncertain parameters.

We used the Morris method to eliminate the uncertain parameters with negligible effects on the output indicators, in order to calculate the Sobol indexes (Sobol 1993). This second part of the sensitivity analysis requires many more code computations, which explains why the Morris method is relevant to use before. Indeed, the calculation on the Sobol indexes required N code computations, with:

$$N = n*(d+2) \quad (2)$$

with:

– n: size of the sample,
– d: number of uncertain parameters.

The Morris method and the Sobol indexes computation combination compose the sensitivity analysis step.

3 MODELLING AND OPTIMIZATION OF THE SYSTEM WITHOUT UNCERTAINTY

3.1 *Case study description*

The case study investigated in this paper is a standalone power system located in Nigeria and is shown on Figure 3. It includes:

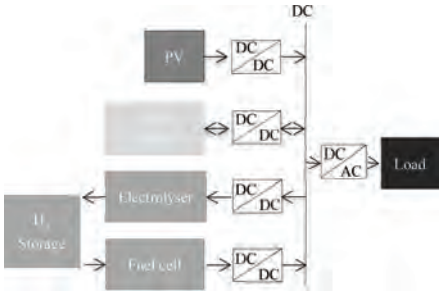


Figure 3. Architecture of the case study.

– an electrical load (Load),
– a photovoltaic (PV) plant,
– a bank of Lead-acid batteries,
– a complete hydrogen chain made of: a PEM electrolyser, a pressurized tank to store the hydrogen and a PEM fuel cell.

This example is representative of (i) the operating competition occurring between batteries and a hydrogen chain, (ii) the problematic of energy storage in off-grid power-system, and (iii) the PV over-sizing linked to the load satisfaction research.

The implemented power management strategy is based on the on/off switches of the electrolyser (ELY) and the fuel cell (FC), as was originally developed by Ulleberg (Ulleberg 2004), and exploited on a similar case by Guinot et al. (Guinot et al. 2015). The operation depends on the state of charge (SOC) of the battery and on levels fixing switching operations the fuel cell and the electrolyser (FC+, FC-, ELY- and ELY-, i.e. the operation parameters) given in Figure 3.

In this case study, the replacement of the components is not considered.

3.2 *Optimization criteria and variables*

The operation parameters are considered constant during the whole exploitation simulation. The optimization of the system operation consists in finding the best suited operation parameters to minimize both electrical cost and unsatisfied load, as for any other design parameters. No distinction is made between plant and controller optimization problems, as it would have been necessary if the operation parameters had been evolving according to the dynamic of the system (Fathy et al. 2001). However, in this study the operation are not optimized.

We selected as optimization variables the five dimensioning variables shown in Table 1.

Odyssey multicriteria optimization process uses a genetic algorithm in order to minimize the standard Levelized Electricity Cost (LEC) in €/MWh on the one hand and to minimize the unsatisfied load (UL) in %, i.e. the energy based percentage of unmet electrical load, on the other hand. Therefore, two objective functions are in competition. It is often observed that lowering the load satisfaction, by reducing the storage system size for example, leads to a lower cost of the system and thus the cost of the produced electricity. While on the contrary, improving the satisfaction of the load by oversizing the system tends to increase the cost of the produced electricity.

The simultaneous optimization of design and operation parameters allows to take maximum benefit from each optimized design.

The multicriteria algorithm used in this work is the Strength Pareto Evolutionary Algorithm 2 (Zitzler et al. 2001).

3.3 *Optimization results*

Due to the competition between both optimization criteria LEC and UL, the optimization results take the shape of a Pareto front as on Figure 5. On this Pareto front, four different design points were selected corresponding to different indicators values (LEC and UL). We selected the point according to the UL and we defined four different cases named from their UL value and with the designs given in Table 2.

The resulting cost distribution for the four selected cases given in Figure 6 illustrate the relative importance of the component costs within the overall system cost. We noticed that the component influencing the most the price was the PV array, which contributes for more than the half of it, in every case, followed by the battery bank (electrical storage). The fuel cell has a significant part in the price only in the case 0, which is intuitive regarding to the maximal power of the fuel cell stack (Table 2).

This part describes the way we identified and selected the optimal system designs without uncertainty consideration. In the following, we will

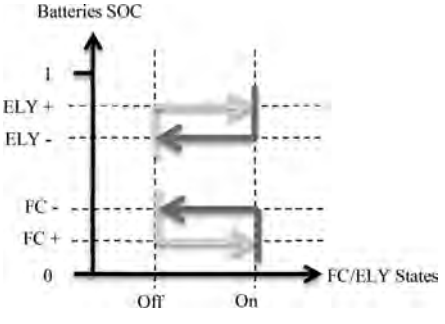


Figure 4. Power management strategy of hydrogen chain.

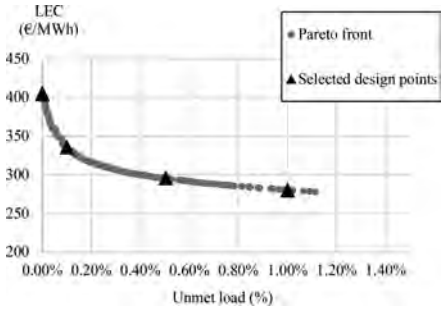


Figure 5. Pareto front resulting from the optimization of the system.

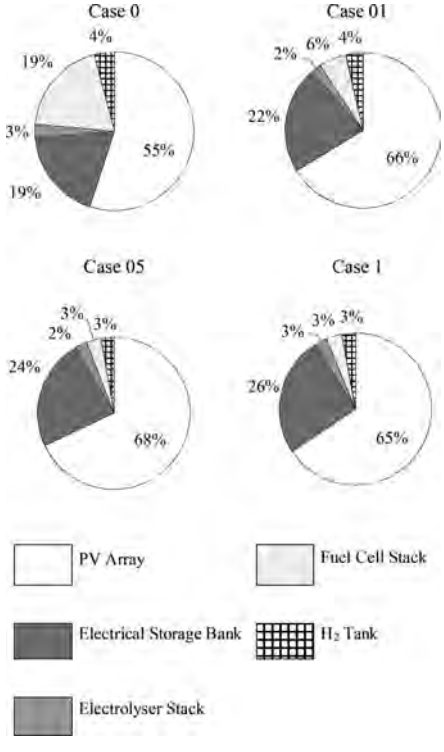


Figure 6. Cost distributions for the 4 different cases.

investigate the influence of uncertainties on the selected cases (optimized without uncertainties) and through them on the Pareto front. In fact, these four cases show similar cost distributions and thereby, it is interesting to observe if applying uncertainties may modify the comparison between them. As we optimized operating parameters, it will however not be possible to check whether operating parameters may counter-balance the effect of uncertainties on the design or not.

4 APPLICATION OF THE PROPOSED METHODOLOGY TO THE CASE STUDY

4.1 *Uncertainty modelling*

In this paper, we decide to focus on static parametrical uncertainties, linked to the energy system parameters. We have identified 25 parameters sources of uncertainty, all with an epistemic nature. In fact, the parameters of components that are not completely mature (such as the electrolyser and the fuel cell) are not well known. Moreover, even the mature components do not have parameters with perfectly known values.

An extensive literature research has been carried out to identify existing, validated or accepted uncertainty probabilistic models for components of energy systems. Table 3 summarizes the different uncertain characteristics of the system components considered in the study, with their associated probability distribution, and with the reference for the chosen uncertainty model. Uniform probability distributions have been associated to the uncertain characteristics of the innovative components (in order to report the equiprobability between the possible values) and to the uncertain parameters of mature components when no other "better" (e.g. from expert judgements) distribution is available.

There are too many uncertain parameters to be de-scribed exhaustively, but the following subsections will focus on three particularity types.

4.1.1 *Uncertain parameters modelled by polynomial models*

Three uncertain parameters, i.e. the cell voltage of the electrolyser, the efficiency of the fuel cell and the CAPEX of the hydrogen pressure tank are modelled by polynomials, respectively functions of the current density in the electrolyser, the pressure (P/Pnominal) of the fuel cell and the volume of the hydrogen pressure tank. We assumed that only the constant coefficient was uncertain, thus generating an area of possible values instead of a curve.

4.1.2 *Uncertain parameters of the photovoltaic panels*

No technical parameter of the photovoltaic panels was considered as uncertain in this paper. This is due to the fact that the solar production is defined directly by a time series data representing the electrical production, not considered as uncertain in this paper.

4.2 *Uncertainty propagation through Odyssey*

The immediate effect of the uncertainties are observed thanks to the Monte Carlo approach using the Odyssey model. Then the dispersions of the indicators (LEC and unmet load) are analyzed. They are not the same for these two indicators and

they depend on the case, i.e. on the design of the system. A realization of all the uncertain parameters is sampled, and based on this realization the system is simulated using its Odyssey model to propagate the uncertainty on the model output performance indicators (UL and LEC). This simulation is iterated for 300 Monte Carlo history. The results are given in Figure 7.

We observe the relative dispersion of the LEC and the unmet load, calculated by the ratio of the standard deviation and the average (Fig. 8). The relative dispersion of the unmet load decreases significantly from case 0 to case 1, i.e. inversely to the nominal unmet load characteristic of the design. While the relative dispersion of the LEC increases slowly from case 0 to case 1, i.e. also inversely to the nominal LEC characteristic of the design.

4.3 *Sensitivity assessment*

After propagating uncertainties, the sensitivity analysis described below aims to identify the most influent uncertain parameters on the output variance.

4.3.1 *Application of Morris method*

The Morris method permits us to select those uncertain parameters that have a non negligible influence on the output indicator (marked with + in the Table 4). Both for the LEC and the UL, whatever the design configuration, the method eliminates the same parameters.

However, the eliminated parameters are not the same for LEC and the UL. Indeed, the UL is decorrelated from the economic parameters, so we do not keep them for the Sobol indexes calculation. On the contrary, the LEC is not influenced only by economic parameters, since it depends also on the electricity production. For the Sobol indexes calculation related to the LEC, we also keep the more influent technical parameters based on the Morris method.

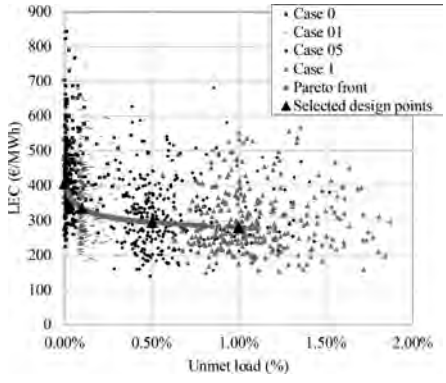


Figure 7. LEC and UL indicators for the four selected design configurations with uncertainties.

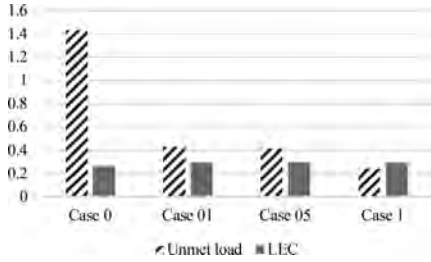


Figure 8. Relative dispersion of LEC and UL for the four study cases.

4.3.2 *Analysis with Sobol indexes*

The Sobol indexes give with precision the contributions of the variance of one output indicator due to the considered parameters.

4.3.2.1 Unmet load

Considering the unmet load variance, the Sobol indexes represented in the Figure 9 indicate that the most influencing uncertain parameter, whatever the case, is the capacity loss of the battery, followed by the discharge efficiency of the battery. The importance of these two parameters, linked to the battery bank shows the major role played by this component in the load satisfaction. The discharge efficiency is much more influent than the charge efficiency, because the PV panel installation is oversized and therefore the solar production is in excess, limiting the role of the charge efficiency. The charge efficiency takes a bigger importance only in case 05 and case 1 (responsible of respectively 3 and 5% of the unmet load variance) where the PV panel installation size is smaller (Table 2).

The ascendancy of the battery on the hydrogen chain is due to the design and the control of those. The powers delivered by the battery on one side and by the hydrogen chain (i.e. by the fuel cell) on the other side illustrate that the hydrogen chain supplies a negligible electric power, even in the case 0, in which the fuel cell has the biggest design, i.e. in which the hydrogen chain production is the most favorable (Fig. 10).

4.3.2.2 Levelized electricity cost

The Sobol indexes indicate that whatever the case, the most influent uncertain parameter on the LEC variance is the PV CAPEX, far before the PV OPEX and to a lower degree the battery bank CAPEX.

We notice that the hydrogen chain plays a significant role in the unmet load variance only in the case 0, i.e. with its largest design: 26% of the total cost in its integrality and 19% for the fuel cell (Fig. 6).

We can observe that if the Sobol index of a given parameter is linked to the cost weight of the corresponding component (studied in Sec-

Table 4. Morris method results.

Component Parameter	Unit	LEC	Unmet load
PV			
CAPEX	€/Wp	+	–
OPEX	% CAPEX	+	–
Battery bank			
CAPEX	€/Wh	+	–
OPEX	% CAPEX	+	–
Capacity loss	Wh/h	+	+
Self-discharge	W	+	+
Charge efficiency	–	–	+
Discharge efficiency	–	+	+
Electrolyser			
CAPEX	€/W	+	–
OPEX	% CAPEX	+	–
Degradation	µV/h	–	+
Cell voltage	V	–	+
FC			
CAPEX	€/W	+	–
OPEX	% CAPEX	+	–
Degradation	µ%/h	–	+
Efficiency	–	–	+
H_2 tank			
CAPEX	€/m^3	+	–
OPEX	% CAPEX	+	–

tion 2), there is however no direct proportional relation, because of the influence of the probability distribution of the input parameters values. For instance, the battery bank that plays an important role in the system cost (between 19% and 26%) has a relatively small impact (inferior than 8%) on the LEC variance. While the PV panel installation, which is the main contributor to the system cost, but no more than 68%, represents (CAPEX and OPEX unified) the overwhelmingly part (between 88% and 94%) of the LEC variance cause.

5 CONCLUSIONS AND FUTURE RESEARCH WORK

In this paper, we investigated a comprehensive 4-steps methodology to evaluate the impact of uncertainties, in order to improve the confidence in the assessment results of a complex autonomous power system modelled and optimized with Odyssey. We first chose to optimize the system before considering the uncertainties that were considered uncertainties as independent.

The results of the uncertainty propagation and of the sensitivity analysis teach us that the most influencing uncertain parameters are linked to the design of the system and in our case study are:

– the capacity loss followed by the discharge efficiency of the battery for the unmet load,
– the PV CAPEX followed by the PV OPEX and the battery CAPEX for the LEC.

There are several interesting points that still have to be thoroughly investigated. We want to investigate now (i) the incidence of the choice of the

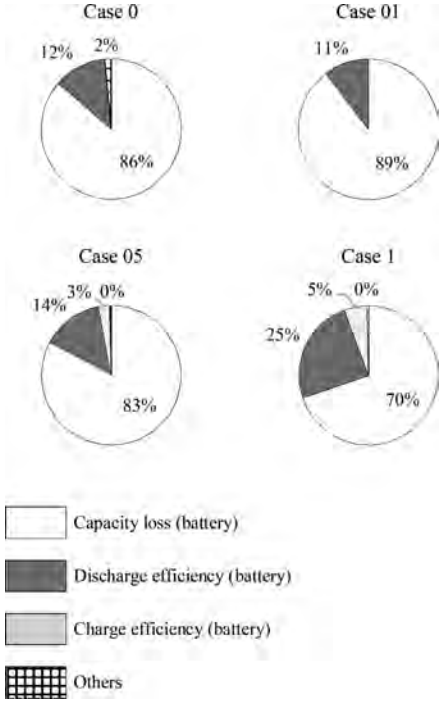


Figure 9. Normalized Sobol indexes (total order) for the four different cases, related to the unmet load.

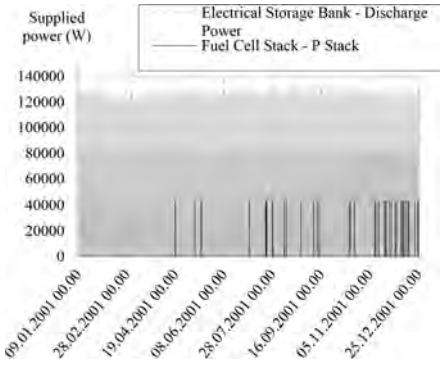


Figure 10. Powers supplied by the hydrogen chain and the battery bank in the case 0.

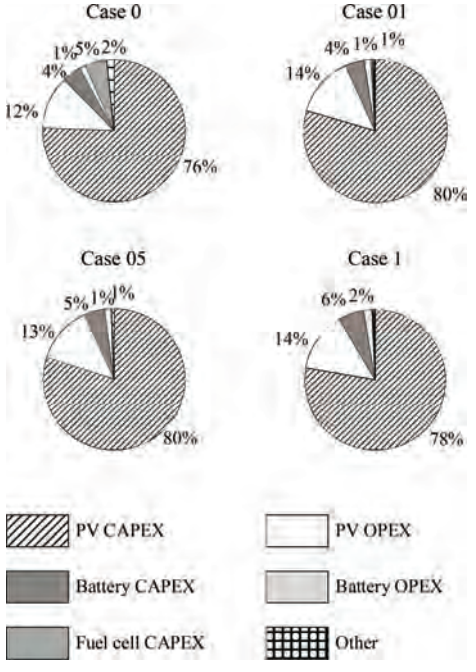


Figure 11. Normalized Sobol indexes (total order) for the four different cases, related to the LEC.

probabilistic distributions associated to the uncertain parameters, (ii) the inclusion of the uncertainty relative to time series, (iii) the optimization of operation parameters as a way to counter-balance uncertainties on the design of the system, and mainly (iv) the optimization taking directly into consideration the uncertainties.

REFERENCES

Battke, B. & Schmidt, T.S. & Grosspietsch, D. & Hoffmann, V.H., 2013. A review and probabilistic model of lifecycle costs of stationary batteries in multiple applications. *Renewable and Sustainable Energy Reviews* 25: 240–250.

Bertuccioli, L. & Chan, A. & Hart, D. & Lehner, F. & Madden, B., Standen, E., 2014. *Study on development of water electrolysis in the European Union.*

Borgonovo, E. & Plischke, E., 2016. Sensitivity analysis: A review of recent advances. *European Journal of Operational Research* 248: 869–887.

Bouloré, A. & Struzik, C. & Gaudier, F., 2012. Uncertainty and sensitivity analysis of the nuclear fuel thermal behavior. *Nuclear Engineering and Design,* SI: CFD4NRS-3 253: 200–210.

Chardonnet, C. & De Vos, L. & Genoese, F. & Roig, G. & Bart, F. & De Lacroix, T. & Ha, T. & Van Genabet, B., 2017. *Study on early business cases for H2 in energy storage and more broadly power to H2 applications.*

De Rocquigny, É., 2006a. La maîtrise des incertitudes dans un contexte industriel. 1re partie/: une approche méthodologique globale basée sur des exemples. *Journal de la société française de statistique* 147: 33–71.

De Rocquigny, É., 2006b. La maîtrise des incertitudes dans un contexte industriel. 2nde partie/: revue des méthodes de modélisation statistique physique et numérique.

Fuel Cells & Hydrogen Joint Undertaking, 2014. *Multi—Annual Work Plan 2014–2020.*

FutureE Fuel Cell Solutions GmbH, n.d. *Jupiter Product Family Sophisticated Fuel Cell Systems.*

Guinot, B., 2013. *Evaluation multicritère des technologies de stockage couplées aux énergies renouvelables/ conception et réalisation de la plateforme de simulation ODYSSEY pour l'optimisation du dimensionnement et de la gestion énergétique (phdthesis).* Université Grenoble Alpes.

Guinot, B. & Champel, B. & Montignac, F. & Lemaire, E. & Vannucci, D. & Sailler, S. & Bultel, Y., 2015. Techno-economic study of a PV-hydrogen-battery hybrid system for off-grid power supply: Impact of performances' ageing on optimal system sizing and competitiveness. *International Journal of Hydrogen Energy* 40: 623–632.

International Energy Agency, 2014. *Technology Roadmap Solar Photovoltaic Energy.*

IRENA, 2017. *Electricity Storage and Renewables: Costs and Markets to 2030. Abu Dhabi.*

IRENA, 2016. *The Power to Change: Solar and Wind Cost Reduction Potential to 2025.*

Kurtz, J. & Huyen, D. & Ainscough, C. & Saur, G., 2015. *Fuel Cell Technology Status: Degradation.*

Morris, M.D., 1991. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics* 33: 161–174.

Riffonneau, Y. & Barruel, F. & Seddik, B., 2007. Problématique du stockage associé aux systèmes photovoltaïques connectés au réseau.

Sobol, I.M., 1993. *Sensitivity estimates for non linear mathematical models.*

Ulleberg, Ø., 2004. The importance of control strategies in PV–hydrogen systems. *Solar Energy, Solar World Congress 2001* 76: 323–329.

Weckend, S. & Wade, A. & Heath, G., 2016. *End-of-Life Management: Solar Photovoltaic Panels.*

Zitzler, E. & Laumanns, M. & Thiele, L., 2001. *SPEA2: Improving the strength pareto evolutionary algorithm (Working Paper).* Eidgenössische Technische Hochschule Zürich (ETH), Institut für Technische Informatik und Kommunikationsnetze (TIK).

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Modular global uncertainty analysis of event-driven indicators of system's availability

Pawel M. Stano & Michal Spirzewski
National Centre for Nuclear Research, Swierk, Poland

ABSTRACT: The purpose of this manuscript is to propose a novel methodology for conducting uncertainty analysis for event-driven indicators of system's availability. In this approach the system's availability depends on the frequency of components' failures and their duration. For all the components the parameters that determine their behavior are obtained from statistical analysis, which means they are given with certain uncertainty, hence the system's availability is also uncertain. In this paper we present a numerically effective algorithm that assesses uncertainty about system's availability. A modular approach has been adopted, in which the uncertainty about the complete system's availability is estimated by combining the information obtained for separate subsystems using the knowledge about the functional relationships between them. Furthermore, to assure numerical efficiency, a novel approach has been adopted, which combines screening approach with quasi-random Sobol sampling approach. The functionality of the proposed method is visualized on an illustrative example.

1 INTRODUCTION

In this paper we develop a methodology to investigate complex input-output stochastic system used in reliability studies. We consider systems with the inputs defined as stochastic renewal processes that model occurrences of the so-called basic events, i.e., the events that cause the system's failure and the outputs of the systems defined as RAMI metrics.

There are two types of failures that might occur to the system: direct, or indirect, e.g., by igniting a sequence of chain events that leads to system failure. The events and their corresponding parameters are usually identified during a standard Failure Mode and Effects Analysis (FMEA) (Stamatis, 2003), which combines a detailed analysis of technical specifications of plant components, in order to identify their failure modes, with a detail analysis of operational interactions within the plant subsystems to discover the effects of previously identified failure modes. Thus, performing FMEA allows for efficient description of systemic structural dependencies (in a form of logic trees or functional block diagrams). Properly performed FMEA is considered as a prerequisite of the analysis described in this paper. In other words, we assume that the complete collection of basic events together with the paths of failure propagation are already identified. With such assumptions in place, it is possible to analyze operational metrics of the system. This is normally done by performing full Reliability, Availability, Maintainability and Inspectability (RAMI) analysis (Stapelberg, 2009). The full RAMI analysis is out of scope of this paper. Instead, the focus is put on investigating what impact the uncertainty in input parameters has on selected availability metrics of the system, which, in general, measure the proportion of time the system in operational with respect to the total system's life time. We adopt block-wise approach, similar to Dynamic Reliability Block Diagrams (DRBDs) (Distefano & Xing, 2006), in which a complex system is broken down into several layers of smaller subsystems. The operational conditions of components are modelled as stochastic processes whose evolution is determined by occurrences of failure events. Each component is associated with an effect function, which determines whether the failure of the component leads to system's deterioration (non-critical failure) or complete shutdown to conduct necessary repairs (critical failure).

The paper is organized in the following manner: Section 2 briefly covers preliminaries necessary to understand the content of the paper, Section 3 describes the general procedure to conduct modular global uncertainty analysis for event-driven indicators. In Section 4 the algorithm described in Section 3 is applied to an illustrative case study coming from the examination of an injector designed for International Fusion Materials Irradiation Facility (IFMIF) (Bargallo Font, 2014). Section 5 concludes the paper with a discussion about the advantages of the proposed method but also about its limitation in the current state of development.

2 PRELIMINARIES

The analysis presented in this paper is a two-level probabilistic approach. In this approach the prior knowledge Priors(θ) about uncertainty associated to basic events e_i identified in FMEA is fed to the quasi-random uncertainty sampler. This higher level algorithm generates a representative sample of the uncertain parameters $\theta_i,\ldots,\theta_j$ of the stochastic input variables, which are sent in parallel to the lower level algorithm. In these blocks probabilistic Monte Carlo simulations of the system are performed to compute the RAMI metrics for a given set of the uncertain parameters $\theta_i,\ldots,\theta_j$ of the stochastic input variables. These metrics are combined in a non-parametric filter for post-processing that leads to final global RAMI analysis. The structure of such a two-level algorithm is presented in Figure 1.

The structure of the higher level algorithm (Uncertainty Sampler) is presented in detail in Section 3. In this section we shall briefly describe how the lower level algorithm works.

2.1 *Probabilistic Monte Carlo RAMI simulations*

The lower level algorithm is designed to study the effects of the basic events mentioned in the previous section on the functionality of the system. In this work this was done by running a probabilistic Monte Carlo simulations of the DRBD model of the system using the Availsim 2.0 program (Bargallo Font, 2014). With this engine it was possible to repeatedly simulate the operation of a system over long periods of time with fixed parameters but random initial conditions, which in our case were times of failure associated with identified basic events. Thus, each simulation run was associated with different failure sequences. When a failure occurs, the performance of the system is affected by a certain factor. As failures accumulate, the performance decreases until it drops below a critical threshold when it is necessary to shut the system down for time necessary to perform corrective repairs that raise the system's performance above its minimal threshold. Note that in case of critical components a single failure leads to immediate shut down of the system. After the system is restarted, a new operational cycle begins until new failures, possibly of the same components, cause the system to shut down again. After the simulation is over, the total down time due to failures is calculated by adding all the down times caused by individual events. Then, the expected down time to the system is computed by averaging the total down times obtained in individual Monte Carlo runs of the simulation. Due to the high complexity of the considered systems, the time to complete a single run of the simulation can be significant. Therefore, it is advisable to stop the further Monte Carlo runs as soon as the expected down time to the system (computed over all the Monte Carlo runs) stabilizes.

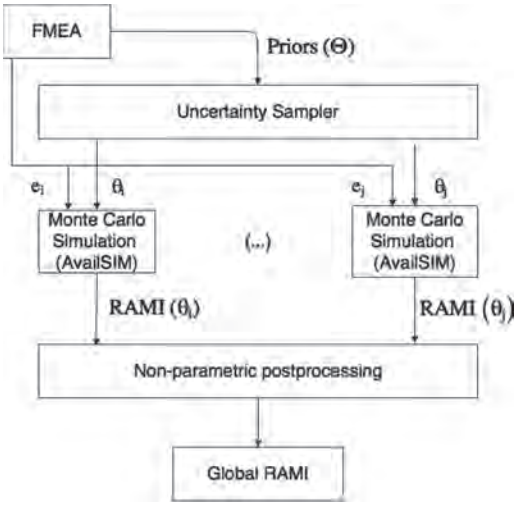


Figure 1. The schematic depiction of the complete two-level Global Uncertainty Analysis (GUA) algorithm.

2.2 *Output indicators of availability*

With every single run of the simulation a realization of a collection of independent stochastic renewal processes E_i is obtained. Each E_i models the occurrence of failure basic events e_i, where the failure rate follows exponential distribution with expectation defined by the Mean Time Between Failure (MTBF) parameter. The time required for the process to renew, the down time T_{down}, is stochastic itself, as it is determined by the sum of independent random variables:

$$T_{down} = T_{access} + T_{repair} + T_{recovery}. \quad (1)$$

where the variables are defined as follows:

a. T_{access} - time required to access the component to be repaired. The variable follows exponential distribution with the Mean Access Time (MAT);
b. T_{repair} - time required to perform repairs of the broken components. The variable follows exponential distribution with the Mean Time To Repair (MTTR);
c. $T_{recovery}$ - time required for the system to recover to the nominal operating conditions. The variable follows exponential distribution with the Mean Recovery Time (MRT).

An important distinction between the variables listed above is that the parameters of the variable T_{repair} usually depend only on the characteristics of the components the basic events are associated with, thus, normally, they vary from event to event. On the other hand, the parameters of the variables T_{access} and $T_{recovery}$ usually depend directly on the system and are often the same for multiple components (e.g., those that share the same location within a subsystem).

The stochastic realizations of variables T_{down}, T_{access}, T_{repair}, $T_{recovery}$ are used to define the output indicators of availability, which are functions of the basic events that cause system failure (inputs). We consider standard system's availability indicators (Stapelberg 2009, Lie & Hwang & Tillman, 1977) such as:

1. Inherent Availability (IA), which is defined as:

$$IA = \frac{\mathrm{T}\sum_{up}}{\mathrm{T}\sum_{up} + \mathrm{T}\sum_{down}}. \tag{2}$$

2. Achieved Availability (AA), defined as:

$$AA = \frac{\mathrm{T}\sum_{up}}{\mathrm{T}\sum_{up} + \mathrm{T}\sum_{down} + \mathrm{T}\sum_{maintenance}}. \tag{3}$$

3. Operational Availability (OA), defined as:

$$OA = \frac{\mathrm{T}\sum_{up}}{\mathrm{T}\sum_{total}}. \tag{4}$$

where:

- $\mathrm{T}\Sigma_{up}$ counts the total time the system is operational;
- $\mathrm{T}\Sigma_{down}$ counts the total time the system is shut down due to failures;
- $\mathrm{T}\Sigma_{maintenance}$ counts the total time the system is shut down due to routine maintenance;
- $\mathrm{T}\Sigma_{total}$ counts the total life time of the system, which includes $\mathrm{T}\Sigma_{up}$, $\mathrm{T}\Sigma_{down}$, $\mathrm{T}\Sigma_{maintenance}$, but also potential logistic delays typically associated with the operating cycle (insufficient service personnel, lack of spare parts, etc.).

Depending on a particular application, the uncertainty of any of the three indicators of the system availability can be investigated.

2.3 *Uncertainty problem formulation*

With the probabilistic simulations described in previous subsections it was possible to determine the average availability of the system for a given set of input parameters. Unfortunately, this information is often not sufficient because for reliable assessment of system's functionality the impact of the uncertainty of parameters such as MTBF, MTTR, MAT, MRT, on the system availability must be analyzed as well. To determine the exact nature of this impact, a thorough uncertainty and sensitivity analysis should be performed. The latter type of analysis, which aims to identify the basic events that introduce the highest sensitivity to the system's availability (Saltelli et al., 2008) is beyond the scope of this paper but is discussed in detail in (Stano, 2018). Instead, the goal of this paper is to describe a method to conduct the uncertainty analysis that aims to describe how the uncertainty in inputs propagate to the outputs of the system. Note that although the uncertainty about inputs is parametrized (the priors about MTBF, MTTR, etc., are given in parametric form, usually lognormal distributions), the uncertainty about outputs is usually nonparametric.

3 GLOBAL UNCERTAINTY ANALYSIS FOR EVENT-DRIVEN INDICATORS OF RAMI

This section presents the higher level algorithm mentioned in Section 2 that aims to generate the representative distribution of prior parameters.

The structure of the algorithm is schematically depicted in Figure 2. The algorithm is composed of three sequential steps, where the first one is of qualitative nature and the following two are of quantitative nature. Firstly, in Section 3.1 the sources of uncertainty associated with basic events are identified; secondly, in Section 3.2 the

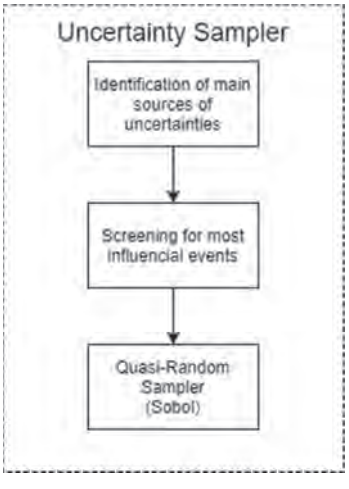


Figure 2. The schematic depiction of the uncertainty sampler.

most influential events are selected; and thirdly, in Section 3.3 the prior distribution of the input parameters is generated via quasi-random Sobol sampler. Finally, the description of modular GUA approach is presented.

3.1 *Identification of main sources of uncertainty*

Each basic event corresponds to a failure of a component or a group of components that cause system's performance deterioration (and eventual shut down) or immediate shut down to perform corrective maintenance. In this paper we assume the components to be within their Normal Life cycle, which means that they fail with a constant rate over time. With some extra effort, it is possible to adopt our methodology to events with time-varying failure rates but this is out of the scope of this paper.

In Section 2.2, we have described four main sources of uncertainty: MTBF, MTTR, MAT, MRT. Out of these two, the MTBF and MTTR are directly associated with the inputs to the system and are aleatory uncertainties with known parametric distributions. On the other hand, the MAT and MRT are structural uncertainties associated only with the system construction. Thus, because the stated purpose of GUA is to study the impacts the inputs have on the outputs, only the uncertainty about MTBF and MTTR parameters shall be investigated. The uncertainty in parameters MAT and MRT should be studied in the context of structural uncertainty analysis, which is devoted to studies of intrinsic system uncertainties (Schueller, 2009).

The prior uncertainties in MTBF and MTTR are modelled by the lifetime distributions derived from reliability tables of appropriate components. Sometimes it is enough to analyze impact of one parameter only (e.g., when the MTTR is very short compared to MAT and MRT, the only important uncertainty about inputs lies in MTBF).

The general model for the uncertainty analysis for the event-driven indicators is given by the following equation:

$$Y = F\left(X_1, \quad , X_d\right), \tag{5}$$

where:

- output Y is a system availability defined by (2–4);
- inputs X_s are random variables with lifetime distributions reflecting uncertainty in failure parameters of basic events (MTBF, MTTR).

It is important to distinguish the uncertainty in failure rates, which is modeled by variables X_s, from the uncertainty of failure occurrences, which are modelled by the renewal processes E_i i = 1,…,d. Thus, the model (5) should be understood as an expectation of a function of d stochastic processes E_i, conditional on realizations of d independent random variables X_s.

3.2 *Screening for most influential events*

The first quantitative part of the analysis is to identify failures that are most detrimental to system's availability. This is done by computing for all events the Fraction Contribution (FC), which is a ratio of unavailability due to an event divided by total unavailability due to all events:

$$FC_i = \frac{unavailability\ due\ to\ event\ e_i}{total\ unavailability}. \tag{6}$$

To compute such defined screening factor, the plant is simulated with nominal parameter settings for all events, i.e., failure and repair rates are fixed at the mean values derived from the reliability tables. Then, the n_0 events with cumulative FCs above a predefined threshold 1-ε are labeled as the most significant and selected for further analysis, i.e., we select those events for which the following holds:

$$\sum_{i=1}^{n_0} FC_i > 1 - \varepsilon. \tag{7}$$

With such restriction introduced, the numerical complexity of the problem is reduced significantly by not taking into consideration events with very low expected contributions to facility's unavailability due to failures. On a downside, by excluding the least influential events from the further analysis a certain amount of information about the system is lost. Thus, the cut-off threshold 1-ε should be selected very carefully so that the price of computational efficiency is not too high. On the top of that, the quantitative screening should be accompanied with the qualitative analysis in which one should investigate whether any events with critical importance to the functionality of the analyzed system are among the discarded events with low FCs. If that is the case it might be advisable to re-incorporated these critical events into GUA despite them having low values of quantitative indicator FCs.

3.3 *Sobol sampling sequences*

Although the dimensionality of the problem was greatly reduced by the screening procedure described in Section 3.2, for complex systems the dimensionality of RAMI simulations remains a numerical challenge. Therefore, we apply Sobol

approach in which the uncertain parameters are generated using quasi-random low-discrepancy Sobol sequences. Thanks to this approach it is possible to obtain a sample that densely covers the sampled space with smaller number of sampling points than those required by the orthogonal grid or the Monte Carlo method. This is because the Sobol sequences satisfy the so-called uniformity properties A and A' (Kucherenko et al., 2015) and at the same time avoid clustering of the samples, which commonly occurs with Monte Carlo samplers. The samples generated for n-dimensional unit cube $[0,1]^n$ are transformed, via inverse cumulative distribution function to get an accurate approximation of system's priors. This approach is visualized on an example of 512 Sobol 75-dimensional points projected on a two-dimensional plane defined by uncertain MTBFs of Acquisition modules and PLC (see Figure 3). Note that the Sobol points cover densely not only the full space (75 dimensions) but also all the lower-dimensional hyper spaces.

3.4 *Modular global uncertainty analysis*

Despite all the steps undertaken above to reduce the numerical load of the GUA algorithm, if it is executed on the complete system it might still be very expensive numerically. Therefore, to conduct numerically effective GUA we propose the following approach: firstly, we divide the system into modules that correspond to the most important parts of the system; secondly, we conduct the GUA described in Sections 3.1–3.3 independently for each module assuming that all the components of the remaining modules are 100% operational; thirdly, we combine the information obtained from separate modules using the knowledge about the functional relationships between events from different modules. This procedure allows for parallelization of the computations which significantly reduces the computational load of the algorithm that uses sequential architecture.

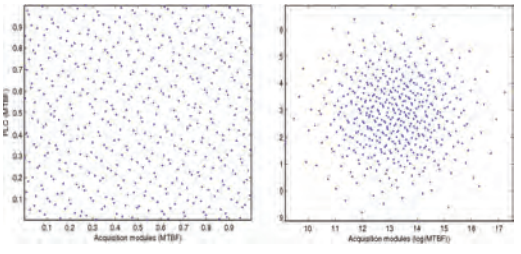


Figure 3. Sample of 512 Sobol sequences projected on two-dimensional plane approximating: uniform distribution (left) and lognormal distribution (right).

It is useful to introduce the following algebraic notation. Let us define System Unavailability (SU) by:

$$SA = 1 - SU \tag{8}$$

where SA denotes random variable defined by (2–4). Then let us introduce the following algebraic operation:

$$SA_1 \oplus SA_2 \underline{\underline{\text{def}}} 1 - SU_1 - SU_2, \tag{9}$$

where SA_1 and SA_2 follow the general form described in (9). The next step necessary to complete the uncertainty analysis for the whole system is to observe that:

$$SA_{SYS} = SA_1 \oplus \cdots \oplus SA_N + Residual, \tag{10}$$

where:

- $SA_{i=1\ldots N}$ are indicators of availability for N individual modules;
- $Residual$ is the term that accounts for interactions between the modules.

Note that with availability of the system assessed by (8–10), the *Residual* is always a nonnegative term. Indeed, in the modular approach we double count all the down times of modules that happen while some other module is already down. Thus, the *Residual* term measures the influence of all these almost-simultaneous occurrences of failure events from distinctive modules with the formula:

$$Residual = \sum_{j=1}^{N} \sum_{1 \le i_1 < \ldots < i_j \le N} \sum_{\substack{1 \le k^{i_1} \le N_{i_1} \\ \vdots \\ 1 \le k^{i_j} \le N_{i_j}}} SA\left(e_{k^{i_1}}^{i_1} \wedge \ldots \wedge e_{k^{i_j}}^{i_j}\right), \tag{11}$$

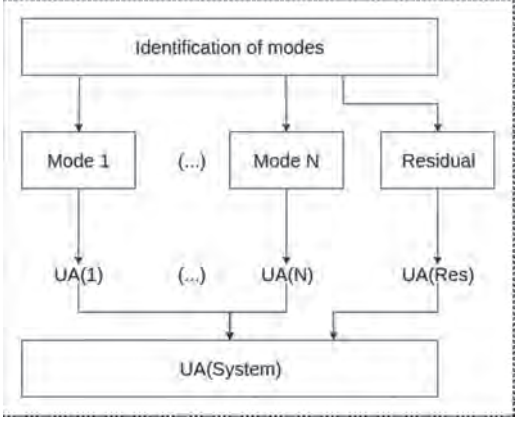


Figure 4. Schematic depiction of the modular GUA approach of a typical figure.

This means that even without analyzing the Residual, with (10) we obtain a conservative estimate of the availability of a system.

$$SA_{SYS} \geq SA_1 \oplus \cdots \oplus SA_N. \tag{12}$$

The schematic depiction of the modular approach described above is shown in Figure 4.

4 CASE STUDY

The practical implementation of the proposed GUA procedure is illustrated on an example, which involves the analysis of the model of the Injector System within the International Fusion Materials Irradiation Facility (IFMIF). The IFMIF is an accelerator-based neutron source, developed jointly by Europe and Japan, which is conceived for fusion materials testing. The main purpose of the Injector System of the linear accelerator under study is to deliver sufficient beam current to the first accelerating cavity (RFQ—Radio Frequency Quadrupole) (Bargallo Font, 2014), and thus to achieve a 125 mA RFQ output current. The Injector is composed of four subsystems: Source and Extraction System (SES); Low Energy Beam Transport (LEBT); Local Control System (LCS); and Auxiliaries System (AUX), which are connected sequentially reliability-wise. Thus, in case of the analyzed Injector, the GUA decomposition into modes naturally overlaps with the decomposition into subsystems.

In case of the Injector, all the identified basic events ignite failure sequences that cause immediate shutdown of the system so that the accelerator vault can be accessed by the repair crews to conduct corrective maintenance operations. Out of considered availability indicators (2–4), we have selected the IA because it measures the direct impact of system failures on accelerator's availability whereas the remaining two accounts also for other events (scheduled maintenance time, logistics delay, etc.).

4.1 *Identification of main sources of uncertainty in the Injector system*

The analysis of the Injector system within the IFMIF accelerator revealed that the main sources of uncertainty in the system lay in failure rates (MTBF parameters). Indeed, the MTTR associated with the basic events are relatively low and dominated by constant access time (MAT) and recovery time (MRT), which are determined by the accelerator technical specification. Consequently, the variations in MTTR do not influence the IA output indicator significantly. The uncertainties in MTBF are parametrized by lognormal distributions with parameters defined in the reliability tables of events detected for the IFMIF accelerator (for details see Bargallo Font (2014)).

4.2 *Screening for the most influential events in the Injector system*

In the analyzed system, according to (Bargallo Font, 2014), there are overall 508 basic events identified for the Injector, each characterized by the uncertainty about the MTBF. It has been established experimentally that for the nominal setting of uncertain parameters, the mean downtime of the Injector system stabilizes after 200 Monte Carlo runs of AvailSIM, which is depicted in Figure 5.

Thus, for a grid of n samples from a prior associated with each event and 200 Monte Carlo runs, the computational load of the rough GUA algorithm is:

$$n^{508} \cdot \ 200 \cdot \ \textit{single run}, \tag{13}$$

which is unfeasible even for a small n. Therefore, following the modular approach of Section 3.4, the space of basic events have been screened independently to identify these events that are the most detrimental to the functionality of each module. The events that contribute cumulatively to more than 90% to the inherent unavailability of each mode are selected for further analysis. These, ordered by the strength of their contribution are presented in Figure 6.

4.3 *Sobol sampling sequences in the Injector system*

For the events detected in the previous section, the uncertainties in MTBF need to be sampled from their lognormal priors. This is done with high-dimensional low-discrepancy Sobol sequences.

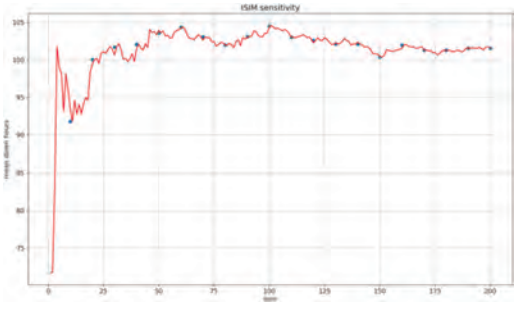


Figure 5. Mean downtime versus number of Monte Carlo runs (nominal parameters).

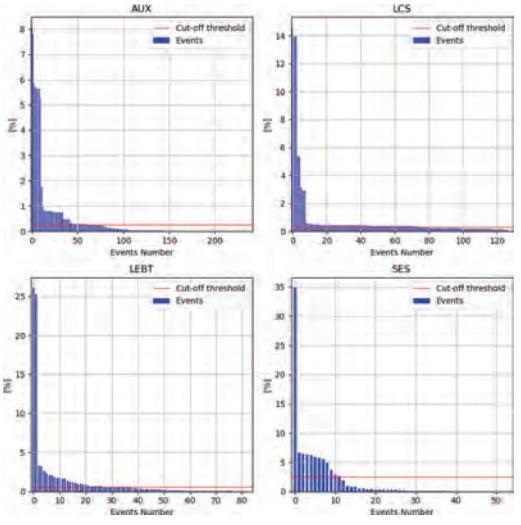


Figure 6. Fractional contributions of the most influential events for each mode considered (SES, LEBT, LCS, AUX).

Following the suggestions of (Kucherenko et al., 2015), the number of Sobol points is set to 512, which, as they argue, give lower discrepancy coverage than both pseudo-random Monte Carlo and Latin Hyper-cube methods. Thus, the computational load of the algorithm is reduced from (13) to:

$$512 \cdot \quad 200 \cdot \quad \textit{single run}, \tag{14}$$

which is manageable even for standard machines.

4.4 *Results of GUA for individual modules*

Four subsystems have been simulated with corresponding number of variables, identified in the screening process, on AvailSIM code. The computational load defined in eq. (14) was evenly distributed on two Xeon E5-2680 v3 processors with total number of 96 cores. There was no significant difference in computation time between subsystems, it took 15 minutes to finish on average, apart from AUX subsystem which required almost 12 hours of calculations to finish.

Since lognormal distribution was applied in the Sobol sequence generation, a non-parametric distribution of the Inherent Unavailability will be fitted to a lognormal distribution. The IU estimators of mean and standard deviation of the log(IU), denoted by μ and σ, respectively, are calculated in twofold.

First method is based on direct calculation of estimators of the μ and σ from standard equations:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{15a}$$

$$\hat{\sigma} = \left(\frac{1}{N}\sum_{i=1}^{N} \left(x_i - \mu \right)^2 \right)^{1/2} \tag{15b}$$

Second method is based on application of the python module *scipy.stats*, specifically *lognorm.fit* function. This function returns, apart from μ and σ estimators, the "loc" parameter, which add an extra degree of freedom in the estimator by allowing for shifts in the data. For each scenario simulated, the Kolmogorov-Smirnov test was performed on both estimators in order to check the goodness of each fit. Based on obtained p-values the better fit was selected to represent the empirical data. All estimators are summarized in Table 1.

For the estimators that give the closest fits, the probability density functions (PDF) and the cumulative distribution functions (CDF) were plotted to see how they compare to the data. The PDFs are compared with histograms of IU data for each considered scenario (see Figure 7) while the CDFs

Table 1. Lognormal estimators of IU.

Mode	$\hat{\mu}$	$\hat{\sigma}$	'loc'
AUX	–5.8	0.44	0.001
LCS	–8.12	0.46	5e-5
LEBT	–7.24	0.28	–
SES	–6.53	0.50	–

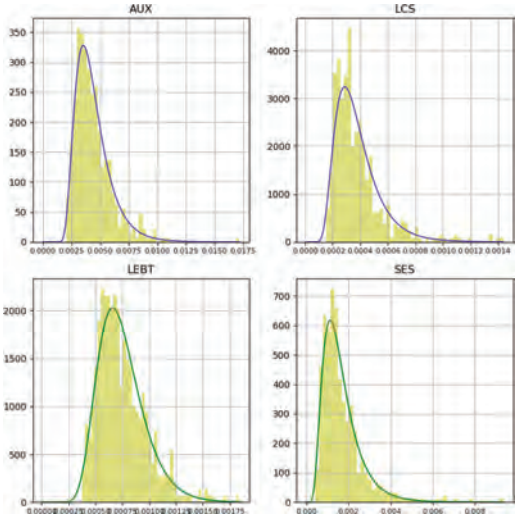


Figure 7. The PDFs fitted to simulated IU (histogram) obtained with lognormal.fit (blue) or classic estimators (green) for each mode considered (SES, LEBT, LCS, AUX).

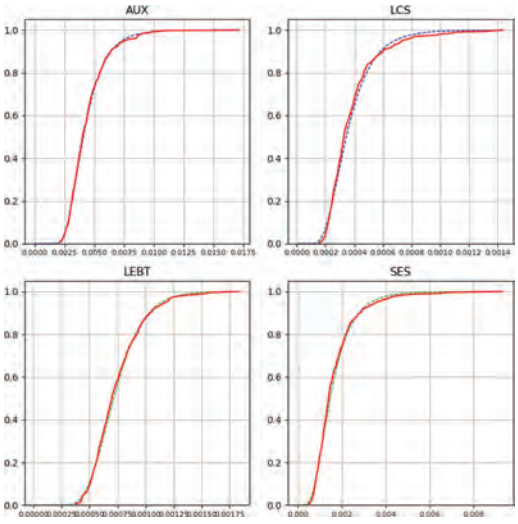


Figure 8. The CDFs fitted to simulated IU (histogram) obtained with lognormal.fit (blue) or classic estimators (green) for each mode considered (SES, LEBT, LCS, AUX).

are compared with empirical CDF of the dataset for each simulated scenario (see Figure 8).

4.5 *Complete GUA for the overall Injector system*

After the procedure was performed on all the four modes, the subsystem data were then combined in order to describe the complete Injector system of the IFMIF in the modular fashion described in Section 3.4.

To complete the analysis it is first necessary to estimate the impact of the Residual term on the uncertainty of the Injector system. To do that it is required to estimate the probability of the coincidental occurrences of the events from distinctive four subsystems. With some simple but tedious combinatorial calculations it can be shown that the probability of the coincidence of the multiple failure events (i.e., of the components coming from distinctive subsystems of the Injector) is bounded by 5e-5, where the boundary is very conservative by taking maximum failure probabilities across all the considered events. This means that the impact of *Residual* is similar to the impact of a single failure event. Thus, given that there are more than 500 events considered in the complete Injector system, the impact of *Residual* on the uncertainty of the IA can be assessed as negligible.

Consequently we estimate the IA of the complete Injector with the lower bound given in (12). Note that for each module considered the lognormal uncertainties in the input propagate into

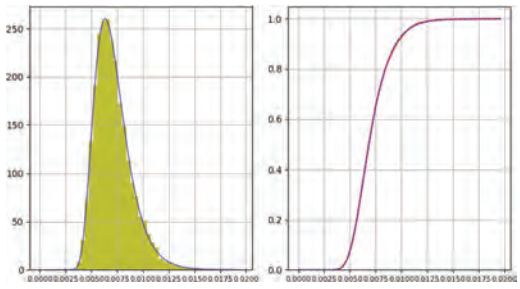


Figure 9. The PDF and CDF fitted to simulated IU (histogram, red) obtained with parametric estimators for the complete Injector.

Table 2. Summary statistics for the IA.

Mode	Mean [%]	Std [%]	1st percentage [%]	Error factor [%]
AUX	99.56	0.17	99.03	1.0057
LCS	99.96	0.019	99.89	1.0008
LEBT	99.93	0.023	99.85	1.0008
SES	99.83	0.104	99.39	1.0047
INJ	99.29	0.177	98.75	1.0057

quasi-lognormals in the outputs. Thanks to this property we were able to approximate the posterior IU distribution of the complete Injector in a parametric way, i.e., the sample distribution of IU for the complete Injector system is obtained by summing samples generated independently from lognormals obtained in previous section. The empirical distribution of 10,000 such generated system data is presented in Figure 9, which shows that the posterior uncertainty distribution of the complete Injector system can also be well approximated with lognormals.

To complete the analysis Table 2 presents summary statistics for individual modes and for the complete Injector system.

5 CONCLUSIONS & DISCUSSION

In this paper a general framework to conduct Modular Global Uncertainty Analysis (GUA) for event-driven RAMI indicators has been introduced. The method proposed is based on a two-level probabilistic simulations conducted in a modular manner, which makes it applicable to a broad range of RAMI problems. The main advantages of the proposed method lie in:

a. numerical efficiency: when compared with standard simulation technics based on Monte

Carlo methods, the proposed algorithm significantly reduces the computational complexity of the problem;

b. completeness: performing GUA makes it possible to estimate the posterior uncertainty distribution of systems' outputs from marginal distributions of systems' inputs;
c. modularity: with the proposed method it is possible to estimate the posterior distributions of systems' modules independently and combine them to produce the complete posterior distribution of the systems' outputs.

We have demonstrated that these three properties are satisfied by applying the proposed Modular GUA method to study the Injector system within the International Fusion Materials Irradiation Facility (IFMIF). For this system, the indicator of inherent availability (IA) can be well approximated with IA indicators obtained independently for four modules: SES, LEBT, LCS, AUX. This is possible due to a weak influence of the Residual that characterizes the impact of the coincidental failures of at least two events from different modules.

We finish the paper with the discussion on limitations of the proposed Modular GUA procedure. Firstly, the presented approach is focused on the analysis of the impacts of the uncertainty in the inputs to the system and does not deal with the impacts of structural uncertainties present in the system (e.g., uncertainties about MAT, MRT, logistic processes). Secondly, within the proposed GUA algorithm it is assumed that the uncertain failure rates of the components are within a Normal Life Cycle, which means that the failure rates are fixed for the whole duration of the simulated system life cycle. Consequently, further research is required to make the method applicable to problems with time-varying failure rates that are associated with infant or wear out exploitation stages of system's components. Thirdly, in general, the impact of Residual might not be negligible as was the case for the Injector system and without analyzing it in detail the right-hand side of (12) will underestimate the true uncertainty in system output. There are several possible ways to analyze the uncertainty associated with non-negligible residual, e.g., if the residual had an impact of another module, it could be handled by incorporating a virtual module of equivalent characteristics into online simulations to update global uncertainty profile. Another possible approach would be to perform (offline) screening on interactions to identify these that matter the most and incorporate them as an interaction module into the simulations—this could be useful in the analysis of sensitivity of the system. Finally, the events rejected during the screening phase due to their marginal contribution on system's availability, might have influential impact taken cumulatively. Therefore, it is desirable to incorporate into the existing procedure a routine that would quantify the uncertainty impact of the least influential events taken collectively. With such procedure we could validate the correctness of cut-off threshold setting.

All the three aforementioned concerns regarding the proposed Modular GUA procedure are a subject of ongoing research.

REFERENCES

Bargallo Font, E. 2014. *IFMIF accelerator facility RAMI analysis in the engineering design phase.* Barcelona: PhD thesis.

Distefano, S. & Xing, L. 2006. A new approach to modeling the system reliability: dynamic reliability block diagrams. In *Reliability and Maintainability Symposium, 2006. RAMS'06. Annual* (pp. 189–195). IEEE.

Kucherenko, S. & Albrecht, D. & Saltelli, A. 2015. Exploring multi-dimensional spaces: a Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques. *arXiv:1505.02350*.

Lie, C.H. & Hwang, C. & Tillman, F.A. 1977. Availability of Maintained Systems: A State-of-the-Art Survey. *AIIE Transactions*: 247–259.

Saltelli, A. & Ratto, M. & Andres, T. & Campolongo, F. & Cariboni, J. & Gatelli, D. & Saisana, M. & Tarantola, S. 2008. *Global Sensitivity Analysis: the primer*. John Wiley & Sons.

Schueller, G.L. 2009. Efficient Monte Carlo simulation procedures in structural uncertainty and reliability analysis—recent advances. *Structural Engineering & Mechanics 32*(1): 1–20.

Stamatis, D.H. 2003. *Failure mode and effect analysis: FMEA from theory to execution.* ASQ Quality Press.

Stano, P.M. 2018. Global Sensitivity Analysis for Event-driven Indicators Used in Probabilistic RAMI Analysis. In *Proceedings of 64th Annual Reliability & Maintainability Symposium, Reno, Nevada 22–25 January 2018*. IEEE.

Stapelberg, R. 2009. *Handbook of reliability, availability, maintainability and safety in engineering design.* Springer Science & Business Media.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A performance-margin-based belief reliability model considering parameter uncertainty

Qingyuan Zhang, Meilin Wen, Rui Kang & Tianpei Zu
School of Reliability Engineering and Systems Engineering, Beijing, China

ABSTRACT: Belief reliability is a newly developed reliability metric considering aleatory and epistemic uncertainty. Since Performance Margin (PM) is an important concept in belief reliability, there is a great need to develop PM-based belief reliability models. In this paper, we consider the parameter uncertainty of PM and propose a new belief reliability model. In this model, the performance parameter and its threshold, which determines the PM of a system, can be modeled as an uncertain variable or a random variable, and three belief reliability formulas are put forward based on different cases. To illustrate the model, a case study about belief reliability estimation of a contact recording head is performed.

1 INTRODUCTION

Nowadays, there is a growing interest in physical model-based reliability metric (cf. Physics-of-Failure (PoF) model based metric (Zeng et al. 2016), structural reliability metric (Choi et al. 2007), etc.), where reliability is predicted using deterministic physical models. The only uncertainty comes from the parameter variations, which are modeled as probability distributions. However, in real cases, besides the random variations (referred to as aleatory uncertainty (Kiureghian & Ditlevsen 2009)) affecting parameters, the physical model is also influenced by epistemic uncertainty caused by our lack of knowledge (Kiureghian & Ditlevsen 2009). For example, the model parameters may not be estimated accurately due to our limited knowledge about the system operation environment (Aven et al. 2014).

Considering the effect of epistemic uncertainty, many reliability metrics are proposed, such as evidence theory-based reliability metric (Bae et al. 2004), interval analysis-based reliability metric (Zhang et al. 2017), fuzzy interval analysis-based reliability metric (Flage et al. 2013) and posbist reliability metric (Cai 1996). The first three metrics are essentially reliability intervals which may cause interval extension problems, while the last one is a possibility measure which does not satisfy duality property (Kang et al. 2016). Recently, a new reliability metric called belief reliability is developed considering both aleatory and epistemic uncertainty Zeng et al. 2017b). To model uncertainty, belief reliability introduces uncertainty theory and chance theory to measure system reliability. Uncertainty theory is proposed by Liu (2007) as a new branch of axiomatic mathematic, founded on normality, duality, subadditivity and product axioms, and chance theory can be regarded as a mixture of uncertainty theory and probability theory. According to Kang et al. (2016), since belief reliability overcomes the shortcomings of other reliability metrics, it is regarded as a more appropriate metric in reliability engineering.

The concept of performance margin plays an important role in belief reliability (Zeng et al. 2017a). Performance Margin (PM), denoted as m, describes the distance between a critical performance parameter p to its associated threshold p_{th}. By analyzing the physical model of a system, we can obtain the PM model. In this process, there may be two types of uncertainties affecting the physical model: parameter uncertainty caused by indeterministic model parameters and model uncertainty due to the inaccuracy of the physical model. Till now, several research have considered the model uncertainty and applied the belief reliability model to real industrial systems (see Zeng et al. (2017a) and Yu et al. (2017)). However, there are still strong needs to develop PM-based belief reliability models considering parameter uncertainty. Therefore, in this paper, we managed to propose a simple model accounting for parameter uncertainty for future application. To represent aleatory and epistemic uncertainty, the PM of a system is modeled as an uncertain random variable. If the system is mainly influenced by aleatory uncertainty, PM will degenerate to a random variable; if the system is mainly affected by epistemic uncertainty, PM will degenerate to an uncertain variable. Since the parameter uncertainty will essentially affect the p and p_{th}, we discuss the model from three aspects, i.e., p and p_{th} are all modeled as uncertain variables, p is a random variable while p_{th} is an uncertain variable, and p is an uncertain variable while p_{th}

is a random variable. Some basic belief reliability formulas based on uncertainty theory and chance theory are then proposed based on the three cases.

The remainder of this paper are organized as follows. In section 2, we will introduce some basic concepts and results of the theory basis, i.e., uncertainty theory and chance theory. The developed PM-based belief reliability model will be introduced in section 3. We will first give the definition of belief reliability in the sense of PM, and then put forward three formulas. Finally, a real case study is performed in section 4 to illustrate the model.

2 PRELIMINARY

In this section, some basic concepts and results of uncertainty theory and chance theory are introduced.

2.1 *Uncertainty theory*

Uncertainty theory is a new branch of axiomatic mathematics built on four axioms, i.e., Normality, Duality, Subadditivity and Product Axioms. Founded by Liu (2007) in 2007 and refined by Liu (2010) in 2010, uncertainty theory has been widely applied as a new tool for modeling subjective (especially human) uncertainties. In uncertainty theory, belief degrees of events are quantified by defining uncertain measures:

Definition 2.1. Uncertain measure, Liu (2007)). Let Γ be a nonempty set, and $\mathcal{L}$ be a σ-algebra over Γ. A set function $\mathcal{M}$ is called an uncertain measure if it satisfies the following axioms,

Axiom 1. Normality Axiom $\mathcal{M}\{\Gamma\}=1$ *for the universal set* Γ.

Axiom 2. *Duality Axiom* $\mathcal{M}\{\Lambda\}+\mathcal{M}\{\Lambda^c\}=1$ *for any event* $\Lambda\in\mathcal{L}$.

Axiom 3. Subadditivity Axiom *For every countable sequence of events* $\Lambda_1,\Lambda_{2,\cdots}$, we have

$$\mathcal{M}\left\{\bigcup_{i=1}^{\infty}\Lambda_i\right\}\le\sum_{i=1}^{\infty}\mathcal{M}\{\Lambda_i\}.$$

Uncertain measures of product events are calculated following the product axiom (Liu 2009):

Axiom 4. Product Axiom. *Let* $(\Gamma_k,\mathcal{L}_k,\mathcal{M}_k)$ *be uncertainty spaces for* $k=1,2,\cdots$. *The product uncertain measure* $\mathcal{M}$ *is an uncertain measure satisfying*

$$\mathcal{M}\left\{\prod_{k=1}^{\infty}\Lambda_k\right\}=\bigwedge_{\infty}^{k=1}\mathcal{M}_k\{\Lambda_k\},$$

where $\mathcal{L}_k$ *are* σ*-algebras over* Γ_k, *and* Λ_k *are arbitrarily chosen events from* $\mathcal{L}_k$ *for* $k=1,2,\cdots$, *respectively.*

Definition 2.2. Uncertain variable, Liu (2007)). An uncertain variable is a function ξ from an uncertainty space $(\Lambda,\mathcal{L},\mathcal{M})$ to the set of real numbers such that $\{\xi\in\mathcal{B}\}$ is an event for any Borel set $\mathcal{B}$ of real numbers.

Definition 2.3. Uncertainty distribution, Liu (2007)). The uncertainty distribution Φ of an uncertain variable ξ is defined by $\Phi(x)=\mathcal{M}\{\xi\le x\}$ for any real number x.

For example, a linear uncertain variable $\xi\sim\mathcal{L}(a,b)$ has an uncertainty distribution

$$\Phi_1(x)=\begin{cases}0, & \text{if } x<a\\ \dfrac{x-a}{b-a}, & \text{if } a\le x\le b\\ 1, & \text{if } x>b\end{cases}\tag{2.1}$$

and a normal uncertain variable $\xi\sim\mathcal{N}(e,\sigma)$ has an uncertainty distribution

$$\Phi_2(x)=\left(1+\exp\left(\frac{\pi(e-x)}{\sqrt{3}\sigma}\right)\right)^{-1},\ x\in\Re\tag{2.2}$$

An uncertainty distribution Φ is said to be regular if it is a continuous and strictly increasing with respect to x, with $0<\Phi(x)<1$, and $\lim_{x\to-\infty}\Phi(x)=0,\lim_{x\to+\infty}\Phi(x)=1$. A regular uncertainty distribution has an inverse function, which is defined as the inverse uncertainty distribution, denoted by $\Phi^{-1}(\alpha),\alpha\in(0,1)$. Inverse uncertainty distributions play a central role in uncertainty theory, since the uncertainty distribution of a function of uncertain variables is calculated using the inverse uncertainty distributions:

Theorem 2.1. (Operational law 1, Liu (2010)). *Let* $\xi_1,\xi_2,\cdots,\xi_n$ *be independent uncertain variables with regular uncertainty distributions* $\Phi_1,\Phi_2,\cdots,\Phi_n$, *respectively. If* $f(\xi_1,\xi_2,\cdots,\xi_n)$ *is strictly increasing with respect to* $\xi_1,\xi_2,\cdots,\xi_m$ *and strictly decreasing with respect to* $\xi_{m+1},\xi_{m+2},\cdots,\xi_n$ *then* $\xi=f(\xi_1,\xi_2,\cdots,\xi_n)$ *has an inverse uncertainty distribution*

$$\psi^{-1}(\alpha)=f\left(\Phi_1^{-1}(\alpha),\cdots,\Phi_m^{-1}(\alpha),\Phi_{m+1}^{-1}(1-\alpha),\Phi_n^{-1}(1-\alpha)\right).$$

Theorem 2.2 (Operational law 2, Liu (2010)). *Let* $\xi_1,\xi_2,\cdots,\xi_n$ *be independent uncertain variables with continuous uncertainty distributions* $\Phi_1,\Phi_2,\cdots,\Phi_n$, *respectively. If* $f(\xi_1,\xi_2,\cdots,\xi_n)$ *is strictly increasing with respect to* $\xi_1,\xi_2,\cdots,\xi_m$ *and strictly decreasing with respect to* $\xi_{m+1},\xi_{m+2},\cdots,\xi_n$, *then* $\xi=f(\xi_1,\xi_2,\cdots,\xi_n)$ *has an uncertainty distribution*

$$\psi(x)=\sup_{f(x_1,x_2,\cdots,x_n)=x}\left(\min_{1\le i\le m}\Phi_i(x_i)\wedge\min_{m+1\le i\le n}\left(1-\Phi_i(x_i)\right)\right).$$

2.2 *Chance theory*

Chance theory is founded by Liu (2013b) as a mixture of uncertainty theory and probability theory, to deal with problems affected by both aleatory uncertainty (randomness) and epistemic uncertainty. The basic concept in chance theory is the chance measure of an event in a chance space.

Let $(\Gamma,\mathcal{L},\mathcal{M})$ be an uncertainty space, and $(\Omega,\mathcal{A},\Pr)$ be a probability space. Then $(\Gamma,\mathcal{L},\mathcal{M})\times(\Omega,\mathcal{A},\Pr)$ is called a chance space.

Definition 2.4. (chance measure, Liu (2013b)). Let $(\Gamma,\mathcal{L},\mathcal{M})\times(\Omega,\mathcal{A},\Pr)$ be a chance space, and let $\Theta\in\mathcal{L}\times\mathcal{A}$ be an event. Then the chance measure of Θ is defined as

$$\mathrm{Ch}\{\Theta\}=\int_0^1\Pr\{\omega\in\Omega\mid\mathcal{M}\{\gamma\in\Gamma\mid(\gamma,\omega)\in\Theta\}\ge x\}\,dx.$$

Definition 2.5 (Uncertain random variable, Liu (2013b)). An uncertain random variable is a function ξ from a chance space $(\Gamma,\mathcal{L},\mathcal{M})\times(\Omega,\mathcal{A},\Pr)$ to the set of real numbers such that $\{\xi\in B\}$ is an event in $\mathcal{L}\cdot\mathcal{A}$ for any Borel set B of real numbers.

Random variables and uncertain variables are two special cases of uncertain random variables. If an uncertain random variable $\xi(\gamma,\omega)$ does not vary with γ, it degenerates to a random variable. If an uncertain random variable $\xi(\gamma,\omega)$ does not vary with ω, it degenerates to an uncertain variable.

Definition 2.6. Let ξ be an uncertain random variable. Then its chance distribution is defined by $\Phi(x)=\mathrm{Ch}\{\xi\le x\}$ for any $x\in\Re$.

Theorem 2.3. (Liu 2013a) *Let $\eta_1,\eta_2,\cdots,\eta_m$ be independent random variables with probability distributions $\Psi_1,\Psi_2,\cdots,\Psi_m$, respectively, and let $\tau_1,\tau_2,\cdots,\tau_n$ be uncertain variables. Assume f is a measurable function. Then the uncertain random variable $\xi=f(\eta_1,\eta_2,\cdots,\eta_m,\tau_1,\tau_2,\cdots,\tau_n)$ has a chance distribution*

$$\Phi(x)=\int_{\Re^m}F(x;y_1,y_2,\cdots,y_m)\,d\Psi_1(y_1)\,d\Psi_2(y_2)\cdots d\Psi_m(y_m),$$

where $F(x;y_1,y_2,\cdots,y_m)$ is the uncertainty distribution of the uncertain variable $f(y_1,y_2,\cdots,y_m,\tau_1,\tau_2,\cdots,\tau_n)$.

3 THE BELIEF RELIABILITY MODEL

This section will develop the belief reliability model based on performance margin (PM). In subsection 3.1, we first give some basic definitions about PM and belief reliability, and discuss the source of uncertainty in PM. Then, three theorems are developed as reliability formulas in subsection 3.2 to calculate belief reliability indexes based on PM.

3.1 *Performance-margin-based belief reliability*

In an industrial system, there is usually a critical performance parameter with a failure threshold, describing the functional behavior and requirement of the system. In most cases, two categories of performance parameters exist:

1. Smaller-the-better (STB) parameters: Failure occurs when $p\ge p_{th}$.
2. Greater-the-better (GTB) parameters: Failure occurs when $p\le p_{th}$.

Definition 3.1. (Performance margin Let p be the critical performance parameter of a system, and p_{th} be its associated failure threshold. Then the performance margin m related to p is defined as:

$$m=\begin{cases}p_{th}-p, & \text{if } p \text{ is } STB,\\ p-p_{th}, & \text{if } p \text{ is GTB.}\end{cases}\tag{3.1}$$

It is easy to find that PM describe the distance between the performance parameter and its threshold, and failure occurs whenever $m\le 0$. In real cases, the PM is usually affected by both aleatory and epistemic uncertainties. Therefore, in this paper, we model the PM as an uncertain random variable and we have the following definition of belief reliability in the sense of PM.

Definition 3.2. (Belief reliability.) Let the system performance margin m be an uncertain random variable, then the system belief reliability is defined as the chance that m is greater than 0, i.e.,

$$R_B=\mathrm{Ch}\{m>0\}.\tag{3.2}$$

Remark 3.1. If the PM is mainly affected by aleatory uncertainty, m will degenerate to a random variable, and the belief reliability becomes $R_B=\Pr\{m>0\}$.

Remark 3.2. If the PM is mainly affected by epistemic uncertainty, m will degenerate to an uncertain variable, and the belief reliability becomes $R_B=\mathcal{M}\{m>0\}$.

The uncertainty of m comes from p and p_{th}. For the performance parameter, in practice, it is usually described by a physical model:

$$p = g(x_1, x_2, \cdots, x_n), \tag{3.3}$$

where $g(\cdot)$ denotes the deterministic model predicting p and $x_i, i = 1,2,\cdots,n$ denote input parameters. Due to the effect of aleatory and epistemic uncertainty, the input parameters are not crisp values anymore. To deal with the parameter uncertainty, the input parameters are usually modelled as random variables or uncertain variables. Therefore, p can be a random variable, an uncertain variable, or an uncertain random variable. For the failure threshold associated with p, it is usually regarded to be constant in traditional analysis. However, in some cases, p_{th} is also affected by uncertainty. For example, if p_{th} is estimated by experts, we model it as an uncertain variable; if p_{th} is estimated by experimental data, we tend to regard it as a random variable.

According to the above discussions, there may be four common conditions about p and p_{th}: (1) p and p_{th} are all random variables; (2) p and p_{th} are all uncertain variables; (3) p is random while p_{th} is uncertain; (4) p is uncertain while p_{th} is random. Since the first condition is similar to the traditional stress-strength interference reliability model, we only discuss the latter three conditions and develop some reliability formulas in next subsection.

3.2 *Reliability formula*

3.2.1 *Case I: p and p_{th} are all uncertain*

Theorem 3.1. *Suppose the system performance parameter p* and its associated threshold p_{th} are uncertain variables with uncertainty distributions $\Phi(x)$ *and* $\Psi(x)$, *respectively. Then the PM-based belief reliability of the system can be calculated by:*

$$R_B = \begin{cases} \sup\limits_{y \in \Re} \left(\Phi(y) \wedge \left(1 - \Psi(y)\right)\right), & \text{if } p \text{ is STB}, \\ \sup\limits_{y \in \Re} \left(\left(1 - \Psi(y)\right) \wedge \Psi(y)\right), & \text{if } p \text{ is GTB}. \end{cases}$$

Specially, if p_{th} is constant, the belief reliability will be

$$R_B = \begin{cases} \Phi(p_{th}), & \text{if } p \text{ is STB}, \\ 1 - \Phi(p_{th}), & \text{if } p \text{ is GTB}. \end{cases}$$

Proof Let $\xi = p - p_{th}$. Since p and p_{th} are all uncertain variables, ξ is also an uncertain variable. Here we assume the uncertainty distribution of ξ is $\Upsilon(x)$. According to Theorem 2.2, we have

$$\Upsilon(x) = \sup_{x_1 - x_2 = x} \left(\Phi(x_1) \wedge \left(1 - \Psi(x_2)\right)\right).$$

If p is a STB parameter, the belief reliability can be calculated as

$$\begin{aligned} R_B &= \mathcal{M}\{m > 0\} = \mathcal{M}\{p - p_{th} < 0\} = \Upsilon(0) \\ &= \sup_{y \in \Re} \left(\Phi(y) \wedge \left(1 - \Psi(y)\right)\right). \end{aligned}$$

When p_{th} is constant, it is easy to find

$$R_B = \mathcal{M}\{p < p_{th}\} = \Psi(p_{th}).$$

If p is a GTB parameter, the belief reliability can be calculated as

$$\begin{aligned} R_B &= \mathcal{M}\{m > 0\} = \mathcal{M}\{p - p_{th} > 0\} = 1 - \Upsilon(0) \\ &= 1 - \sup_{y \in \Re} \left(\Phi(y) \wedge \left(1 - \Psi(y)\right)\right) \\ &= \sup_{y \in \Re} \left(\left(1 - \Phi(y)\right) \wedge \Psi(y)\right). \end{aligned}$$

When p_{th} is constant, we have

$$R_B = \mathcal{M}\{p > p_{th}\} = 1 - \Phi(p_{th}).$$

3.2.2 *Case II: p is random while p_{th} is uncertain*

Theorem 3.2. *Suppose the system performance parameter p is a random variable with a probability distribution* $\Phi(x)$ *and the associated threshold p_{th} is an uncertain variables with an uncertainty distribution* $\Psi(x)$. *Then the PM-based belief reliability of the system can be calculated by:*

$$R_B = \begin{cases} \int_{-\infty}^{+\infty} 1 - \Psi(y) d\Phi(y), & \text{if } p \text{ is STB}, \\ \int_{-\infty}^{+\infty} \Psi(y) d\Phi(y), & \text{if } p \text{ is GTB}. \end{cases}$$

Proof The proof will be given from two aspects:

1. p is a STB parameter

Let $\xi = p - p_{th}$. Since p is a random variable while p_{th} is an uncertain variable, ξ is an uncertain random variable. Here we assume the chance distribution of ξ is $\Upsilon_1(x)$. According to Theorem 2.3, we have

$$\Upsilon_1(x) = \int_{-\infty}^{+\infty} F_1(x; y) d\Phi(y), \tag{3.4}$$

where $F_1(x;y)$ is the uncertainty distribution of an uncertain variable $q = y - p_{th}$. Based on Theorem 2.1, we have $F_1^{-1}(\alpha) = y - \Psi^{-1}(1 - \alpha)$, so $F_1(x;y)$ is $F_1(x; y) = 1 - \Psi(y - x)$. Then, (3.4) will be

$$\Upsilon_1(x) = \int_{-\infty}^{+\infty} 1 - \Psi(y - x) d\Phi(y).$$

Therefore, the belief reliability can be calculated by

$$\begin{aligned} &R_B = Ch\{p - p_{th} < 0\} = \Upsilon(0) \\ &= \int_{-\infty}^{+\infty} 1 - \Psi(y) d\Phi(y) \end{aligned}$$

2. p is a GTB parameter

Let $\eta = p_{th} - p$. It is easy to find η is an uncertain random variable. Assume its chance distribution if $\Upsilon_2(x)$, then we have

$$\Upsilon_2(x) = \int_{-\infty}^{+\infty} F_2(x;y)\, d\Phi(y), \tag{3.5}$$

where $F_2(x;y)$ is the uncertainty distribution of an uncertain variable $s = p_{th} - y$. Similarly, we have $F_2(x;y) = \Psi(x+y)$. Then, (3.5) will be

$$\Upsilon_2(x) = \int_{-\infty}^{+\infty} \Psi(x+y)\, d\Phi(y)$$

Therefore, the belief reliability can be calculated by

$$\begin{aligned} R_B &= Ch\{p_{th} - p < 0\} = \Upsilon(0) \\ &= \int_{-\infty}^{+\infty} \Psi(y)\, d\Phi(y) \end{aligned}$$

3.2.3 *Case III: p is uncertain while p_{th} is random*

Theorem 3.3. *Suppose the system performance parameter p is an uncertain variable with an uncertainty distribution $\Phi(x)$ and the associated threshold p_{th} is a random variables with a probability distribution $\psi(x)$. Then the PM-based belief reliability of the system can be calculated by:*

$$R_B = \begin{cases} \int_{-\infty}^{+\infty} \Phi(y)\, d\Psi(y), & \text{if } p \text{ is STB,} \\ \int_{-\infty}^{+\infty} 1-\Phi(y)\, d\Psi(y), & \text{if } p \text{ is GTB.} \end{cases}$$

Proof Since the proof of this theorem is similar to the previous one, it will not be repeated here.

4 CASE STUDY

In this section, we try to use the proposed model to analyze the static belief reliability of a contact recording head (Kawakubo, Miyazawa, Nagata, & Kobatake 2003). The contact recording head is a vital component of a contact recording system, which is used to achieve a recording density over 100 Gbits/in^2. Wear is the main failure mechanism of the recording head. To satisfy the recording density, the wear depth of the head cannot exceed 2 nm (determined by the overcoat thickness of the head) after 100h of using. Since a track seek also reciprocates during the disk rotation, the worn area is a donut.

Because the wear rate of the recording head decreases with increasing sliding distance in wear tests, to quantify the wear volume, an extended Archard's wear equation developed by Kawakubo, Miyazawa, Nagata, & Kobatake (2003) is utilized:

$$V = k_s \cdot L_s \cdot W \cdot \left(\frac{L}{L_s}\right)^{1-a} \left(\frac{B}{b}\right)^{a}, \tag{4.1}$$

where V denotes the wear volume, k_s denotes specific wear amount (SWA) at a standard sliding distance L_s, W denotes the sliding load, L denotes the total sliding distance, a denotes running-in coefficient ranging from 0 to 1, B denotes the sliding width, and b denotes the head contact width.

The uncertainty of this physical model may come from the input parameters. Among them, the coefficient a is relatively precise obtained by real experiment, k_s is calculated based on a and the material properties, L_s is determined by k_s, L is easily controlled to be a crisp value (total function time is 100h and the sliding speed is controlled to be 10m/s), and B and b are controlled during design and manufacturing phase. The only uncertain parameter is the W, which may be affected by many factors in real cases and we don't have enough data to estimate an accurate value. Therefore, in this paper, we model W as an uncertain variable following normal distribution (the form is shown in 2.2). Through uncertainty propagation, the uncertainty distribution of V can be easily calculated. In addition, since the threshold of V is given by experts according to the overcoat thickness of the head, it may be affected by epistemic uncertainty. To be more precise, we describe V_{th} as an uncertain variable as well. The values or the distributions of critical parameters are summarized in Table 1.

Table 1. Values or distributions of parameters.

Parameter	Value or distribution
SWA	$k_s = 2.55 \times 10^{-20}\,(m^2/N)$
Standard sliding distance	$L_s = 1000(m)$
Running-in coefficient	$a = 0.39$
Total sliding distance	$L = 3.6 \times 10^6\,(m)$
Sliding width	$B = 0.015(m)$
Head width	$b = 10^{-4}\,(m)$
Head contact area	$10^{-8}\,(m^2)$
Contact load	$W \sim \mathcal{N}(\mu = 0.7, \sigma = 0.03)\,(mN)$
Wear volume threshold	$V_{th} \sim \mathrm{L}(a = 2, b = 2.5)\,(10^{-17} m^3)$

1. $\mathcal{N}(\mu,\sigma)$ and $\mathcal{L}(a,b)$ are normal and linear uncertainty distributions in the form of (2.2) and (2.1), respectively.

Based on the operational laws of uncertainty theory, the wear volume of the head after 100h of using also follows a normal uncertainty distribution, i.e.,

$$V \sim \mathcal{N}\left(\mu_V = 1.8606, \sigma_V = 0.07974\right)\left(10^{-17} m^3\right),$$

Assume the distribution function of V and V_{th} are $\Phi_V(x)$ and $\Phi_{V_{th}}$, respectively. Since the wear volume V is a STB parameter, by using Theorem 3.1, the static belief reliability can be calculated to be

$$R_B = \sup_{x \in \Re}\left(\Phi_V(x) \wedge \left(1 - \Phi_{V_{th}}(x)\right)\right) = 0.97078.$$

5 CONCLUSIONS

In this paper, we develop some performance-margin-based belief reliability models using uncertainty theory and chance theory. Considering the uncertainty of performance parameter and its threshold, we discuss the model from three aspects: p and p_{th} are all uncertain variables, p is a random variable while p_{th} is an uncertain variable, and p is an uncertain variable while p_{th} is a random variable. Three theorems as belief reliability formulas are given. Finally, a case study about the static belief reliability evaluation of a contact recording head is performed to illustrate the reliability formulas.

REFERENCES

Aven, T., P. Baraldi, R. Flage, & E. Zio (2014). Uncertainty in Risk Assessment: *The Representation and Treatment of Uncertainties by Probabilistic and Non-Probabilistic Methods.* John Wiley & Sons.

Bae, H.R., R.V. Grandhi, & R.A. Canfield (2004). An approximation approach for uncertainty quantification using evidence theory. *Reliability Engineering & System Safety 86*(3), 215–225.

Cai, K. (1996). *Introduction of fuzzy reliability.* Norwell: Kluwer Academic Publishers.

Choi, S.K., R.A. Canfield, & R.V. Grandhi (2007). *Reliability based Structural Design.* Springer London.

Flage, R., P. Baraldi, E. Zio, & T. Aven (2013). Probability and possibility-based representations of uncertainty in fault tree analysis. *Risk Analysis An Official Publication of the Society for Risk Analysis 33*(1), 121.

Kang, R., Q. Zhang, Z. Zeng, E. Zio, & X. Li (2016). Measuring reliability under epistemic uncertainty: Review on nonprobabilistic reliability metrics. *Chinese Journal of Aeronautics 29*(3), 571–579.

Kawakubo, Y., S. Miyazawa, K. Nagata, & S. Kobatake (2003). Wear life prediction of contact recording head. *Magnetics IEEE Transactions on 39*(2), 888–892.

Kiureghian, A.D. & O. Ditlevsen (2009). Aleatory or epistemic? does it matter? *Structural Safety 31*(2), 105–112.

Liu, B. (2007). *Uncertainty Theory* (2nd ed.). Berlin Heidelberg: Springer Berlin Heidelberg.

Liu, B. (2009). Some research problems in uncertainty theory. *Journal of Uncertain Systems 3*(1), 3–10.

Liu, B. (2010). *Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty.* Springer Berlin Heidelberg.

Liu, Y. (2013a). Uncertain random programming with applications. *Fuzzy Optimization & Decision Making 12*(2), 153–169.

Liu, Y. (2013b). Uncertain random variables: a mixture of uncertainty and randomness. *Soft Computing 17*(4), 625–634.

Yu, S., Q. Zhang, M. Wen, & R. Kang (2017). Belief reliability evaluation of a quad redundant servo system: A case study. In *International Conference on Reliability Systems Engineering,* pp. 1–6.

Zeng, Z., M. Wen, & R. Kang (2013). Belief reliability: a new metrics for products reliability. *Fuzzy Optimization & Decision Making 12*(1), 15–27.

Zeng, Z., R. Kang, & Y. Chen (2016). Using pof models to predict system reliability considering failure collaboration. *Chinese Journal of Aeronautics 29*(5), 1294–1301.

Zeng, Z., R. Kang, M. Wen, & E. Zio (2017a). A model-based reliability metric considering aleatory and epistemic uncertainty. *IEEE Access 5*(99), 15505–15515.

Zeng, Z., R. Kang, M.Wen, & E. Zio (2017b). uncertainty theory as a basis for belief reliability. *Information Sciences 429*(1), 26–36.

Zhang, Q., Z. Zeng, E. Zio, & R. Kang (2017). Probability box as a tool to model and control the effect of epistemic uncertainty in multiple dependent competing failure processes. *Applied Soft Computing 56,* 570–579.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Bayesian updating with time dependent models

P. Beaurepaire
CNRS, SIGMA Clermont, Institut Pascal, Université Clermont Auvergne, ClermontFerrand, France

ABSTRACT: Bayesian updating is increasingly used in structural engineering; it is applicable as an inverse method to identify the model of uncertainty which best matches some available experimental data. This paper introduces a novel method for the definition of the likelihood function in case the numerical model is a time dependent function. The set of time instants which best describes the experimental data is identified using the maximum entropy principle. The marginal distributions of the model response are identified as well using the maximum of entropy. The dependence between the responses of the model for the different time instants is implemented using the linear coefficients of correlation obtained after a mapping into standard normal distributions. The joint probability density function is subsequently used in the formulation of the likelihood function. The relevance of the method is demonstrated through an application example.

1 INTRODUCTION

Inverse methods are widely used in science and engineering; they consist of identifying the input parameter of a numerical model leading to an adequate match with available experimental data. Model updating techniques provide an appropriate framework and received considerable attention from structural engineers during the past decades (Friswell & Mottershead 1995, Imregun & Visser 1991, Arora 2011). They are applied in case a forward numerical model is available but it is not possible or numerically too demanding to evaluate the inverse model. Such methods are used for instance in case sensors collect the vibration data, which are subsequently used to identify modal information (amplitudes, modes, damping, etc). A finite element model is then implemented and model updating is used to identify the input parameters leading to the best fit with the experimental data.

Bayesian updating methods allow to identify the optimal parameter values and as well the probability density function associated with them (Beck & Katafygiotis 1998, Katafygiotis & Beck 1998). Considering a model $\mathcal{M}$, $\boldsymbol{\theta}$ the set of parameters to be updated from the experimental data $\mathcal{D}$ using Bayes' theorem, which expresses as:

$$p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}) = \frac{p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})\,p(\boldsymbol{\theta}|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})} \tag{1}$$

where $p(\boldsymbol{\theta}|\mathcal{M})$ is the *prior distribution*, which gathers the initial knowledge on the parameters; $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})$ is the *likelihood function*, which quantifies the match between the experimental data and the outcome of the numerical model; $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ is the *posterior distribution*, a probability density function associated with the model parameters which considersthe information provided by the experimental data $\mathcal{D}$ and $p(\mathcal{D}|\mathcal{M})$ is the *evidence*, a constant guaranteeing that Equation (1) integrates to one.

Much research efforts on Bayesian updating are geared towards the implementation of computationally efficient numerical methods. The most frequently used approach consists of generating realizations of the posterior distribution and subsequently assuming that this set of realizations fully describes the posterior distribution. Markov chain Monte Carlo is frequently used in this context, as it can be applied to any arbitrarily given distribution. For instance, Ching and Chen (2007) proposed an efficient algorithm based on a sequence of intermediate distributions with a gradual convergence from the prior distribution to the posterior distribution, combined with an appropriate selection of the first state of the Markov chains; Beck and Zuev (2013) implemented a method based on importance sampling, Markov chain Monte Carlo and simulated annealing; Straub and Papaioannou (2015) introduced Bayesian Updating with Structural reliability methods (BUS), a procedure used to transform the updating problem into a reliability problem, which is subsequently solved with reduced efforts as multiple efficient algorithms are available for reliability analysis.

The focus of this paper is on the formulation of the likelihood involved in Equation (1). The formulation of this function is not difficult in case the

numerical model is a scalar function, it is possible to use for instance kernel density estimation (Goller 2011, Nagel & Sudret 2016). Specific formulations are available in case the problem involves modal data (Vanik, Beck, & Au 2000, Ching, Muto, & Beck 2006). In this paper, a novel implementation of the likelihood function is proposed in case the response of the numerical model is a time dependent function of the form:

$$y = f(t, \boldsymbol{\theta}) \tag{2}$$

where $\boldsymbol{\theta}$ denotes the uncertain parameters and t denotes time.

This manuscript is organized as follows: the methods of analysis are described in Section 2; Section 3 presents application examples and the paper closes with conclusions and perspectives in Section 4.

2 METHODS OF ANALYSIS

2.1 *Distribution at a given time instant*

In case Bayesian updating is applied, a set of experimental realizations of the numerical model is available and used to define the likelihood function. It is assumed here that it consists of a set of curves of the response of the model expressed in termsof time of the form $\left(y^{(1)}(t), y^{(2)}(t), ..., y^{(N_R)}(t)\right)$. The method developed here is applicable only in case the response of the model greater than zero for all the time instants.

The first step of the procedure is the discretization of the time dependent response of the model (as shown in Figure 1). A finite set of time instants $\boldsymbol{t} = \left[t_1, t_2, ..., t_{N_d}\right]$ is defined and the same discretization points are applied to all the experimental responses.

The response of the numerical model at the time t_i is a function involving uncertain parameters, it can therefore by modeled as a random variable Y. The probability density function associated with the model response needs to be identified for any predefined time instant. It is not possible to assign *a priori* a probability distribution (e.g. Gaussian, lognormal, uniform) with Y, as the most suitable distribution is problem dependent. Novi Inverardi and Tagliani (2003) proposed a flexible strategy to model the probability density function of an arbitrarily given distribution from a set of experimental data, it is expressed as:

$$p_Y(x \mid t_i) = \exp\left(-\sum_{j=0}^{M_i} \lambda_{ij} x^{\alpha_{ij}} \right) \tag{3}$$

where $p_Y(\cdot \mid t_i)$ denotes the probability density associated with the model response at the time t_i, $\lambda_i = \left(\lambda_{i1}, ..., \lambda_{iM_i}\right)$ and $\alpha_i = \left(\alpha_{i1}, ..., \alpha_{iM_i}\right)$ denote a set of parameters identified using the maximum entropy principle.

The distribution described in Equation (3) may be characterized by its fractional moments at the time instant t_i; which are expressed as:

$$\mu_{ij} = \mathrm{E}\left[y(t_i)^{\alpha_{ij}} \right] = \int_0^{\infty} x^{\alpha_{ij}} p_Y(x \mid t_i) \mathrm{d}x \tag{4}$$

These fractional moments can be estimated from the experimental data:

$$\mu_{ij} \approx \hat{\mu}_{ij} = \frac{1}{N_R} \sum_{j=1}^{N_R} \left(y^{(j)}(t_i) \right)^{\alpha_{ij}} \tag{5}$$

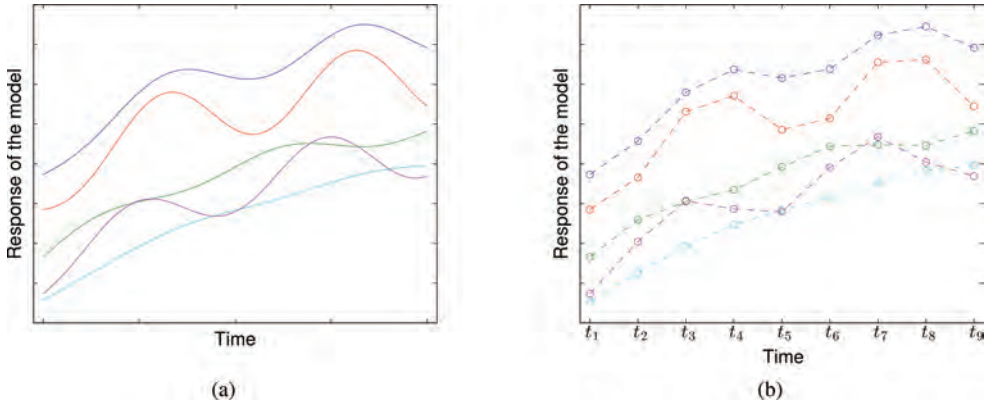


Figure 1. Schematic aspect of a time dependent response. Each curve can be associated with a realizations of the available data. (a) Original functions. (b) Discretization of the functions.

The identification of the maximum entropy distribution may be simplified using equispaced moments (see e.g. (Taufer, Bose, & Tagliani 2009)), i.e. $\alpha_{i0} = 0$ and $\alpha_{ij} = j\alpha_{i1}$ for $j > 1$. The coefficient λ_{i0} is used to guarantee that Equation (3) integrates to one; it is expressed as:

$$\lambda_{i0} = \ln\left[\int_0^{\infty} \exp\left(-\sum_{j=1}^{M_i} \lambda_{ij} x^{\alpha_{ij}} \right) \mathrm{d}x \right] \tag{6}$$

The integral involved in Equation (6) is estimated using Gauss-Laguerre quadrature.

The entropy associated with the distribution is expressed as:

$$\Gamma_i\left(\boldsymbol{\alpha}_{i1}, \ddot{\boldsymbol{e}}_i \right) = \lambda_{i0} + \sum_{j=1}^{M_i} \lambda_{ij} \hat{\mu}_{ij} \tag{7}$$

The maximum entropy distribution is subsequently identified using an optimization procedure to maximize Equation (7). The design variables of the optimization (i.e. the adjustable parameters) are α_{i1} and λ_i. The values of these coefficients obtained through the optimization procedure define a probability density function in good agreement with the experimental data.

The probability density function $p_Y(x \mid t_i)$ and its coefficients $\boldsymbol{\alpha}_i$ and $\boldsymbol{\lambda}_i$ should be expressed in terms of $\mathscr{D}$, as the experimental data is used to identify the coefficient value. However, this dependence is omitted in the notation to simplify the formulas.

It is referred to (Novi Inverardi & Tagliani 2003) for a detailed derivation of the equations described above, an in-depth discussion of the method and a procedure applicable to set the total number of fractional moments M_i to be used.

2.2 *Distribution associated with a set of time instants*

The experimental values of the numerical model for a set of time instants $\boldsymbol{t}$ are characterized by the marginal distributions and by the linear correlation coefficient. The procedure described in Section 2.1 is applied independently for all the time instants to identify the marginal distributions. The joint probability density function is subsequently defined using the Nataf' model (Nataf 1962, Liu & Der Kiureghian 1986), which is equivalent to using a Gaussian copula (see e.g. (Mai & Scherer 2012, Schölzel & Friederichs 2008)). Theiso-probabilistic transformation is used and auxiliary random variables with a standard normal distributions are introduced, they are defined as:

$$Z_i = \Phi^{-1}\left(P_{Y|_{t_i}}\left(Y\left(t_i \right) \right) \right) \tag{8}$$

where $P_{Y|_{t_i}}$ denotes the cumulative distribution function associated with the response of numerical model at the time instance t_i and Φ^{-1} denotes the inverse of the cumulative distribution function of the standard normal distribution. In the general case, the responses of the numerical model at the time instants t_i and t_j are correlated. Therefore, the random variables Z_i and Z_j are correlated as well. Their correlation coefficient is determined by applying the iso-probabilistic transformation to the experimental data:

$$\begin{aligned} z_i^{(k)} &= \Phi^{-1}\left(P_{Y|_{t_i}}\left(y^{(k)}\left(t_i \right) \right) \right) \\ &= \Phi^{-1}\left(1 - \int_{y^{(k)}(t_i)}^{\infty} p_Y(x \mid t_i) \mathrm{d}x \right) \end{aligned} \tag{9}$$

Gauss-Laguerre quadrature is used for the integral involved in Equation (9).

The coefficients of correlation are directly determined once all the experimental data are mapped into standard normal distributions. The joint probability density function of the numerical model's response is:

$$\begin{aligned} p_Y(\boldsymbol{x} \mid \boldsymbol{t}) = {} & p_Y(x_1 \mid t_1) p_Y(x_2 \mid t_2) \ldots p_Y(x_{N_d} \mid t_{N_d}) \\ & \cdot \frac{\varphi_{N_d}\left(\boldsymbol{z}, \boldsymbol{R} \right)}{\varphi\left(z_1 \right) \varphi\left(z_2 \right) \ldots \varphi\left(z_{N_d} \right)} \end{aligned} \tag{10}$$

where φ denotes the probability density function of a standard normal distribution, φ_{N_d} denotes the joint probability density function associated with a multivariate standard normal distribution, $\boldsymbol{R}$ being the matrix containing thecoefficients of correlation.

Alternative strategies may be considered to account for the correlation between the responses of the numerical model at different time instants. For instance, copulas offer a rational framework to considered alternative dependence structures. However, thejoint probability density function described in Equation (10) is involved in the definition of the objective function of an optimization problem (as discussed in the next subsection). Therefore, a large number of density functions need to be defined in an iterative way, which may become challenging in case multiple structures of dependence need to be considered. Only the Nataf model is implemented in this work for the sake of simplicity.

2.3 *Identification of a representative set of time instants*

A reduced set of time instants $\boldsymbol{t}$ needs to be identified to define the likelihood function and the maximum entropy principle is used as the quantity of information extracted from the experimental data can

be maximized using this technique (Jaynes 1957a, Jaynes 1957b). The entropy associated with $\boldsymbol{t}$ is:

$$\Gamma(\boldsymbol{t}) = \int_0^\infty \ldots \int_0^\infty p_Y(\boldsymbol{x}\,|\,\boldsymbol{t}) \ln\left(p_Y(\boldsymbol{x}\,|\,\boldsymbol{t})\right) \mathrm{d}\boldsymbol{x} \tag{11}$$

The expression of the joint probability density function expressed in Equation (10) can be injected in Equation (11); this formula can be expanded and simplified. This procedure is not derived in details herein and the simplified formulaof the entropy is:

$$\Gamma(\boldsymbol{t}) = \sum_{j=0}^{M} \Gamma_i + \frac{1}{2} \ln \det\left(2\pi e \boldsymbol{R}\right) \tag{12}$$

where Γ_i is the entropy defined in Equation (7). The term on the left-hand side of Equation (12) is the contribution of the marginal distributions to the total entropy and the term on the right-hand side is the contribution of the correlation. The set of time instants which best represents the experimental data is obtained by maximizing Equation (12):

$$\hat{\boldsymbol{t}} = \boldsymbol{t} \arg\max \Gamma(\boldsymbol{t}) \tag{13}$$

$\hat{\boldsymbol{t}}$ may be interpreted as the set of N_d instants which contains the maximum of information from the experimental data. It is therefore subsequently used in the definition of the likelihood function.

2.4 *Bayesian updating*

The likelihood function involves the numerical model y, which needs to be evaluated for the time instants maximizing Equation (12). It involves as an input the values of the uncertain parameters $\boldsymbol{\theta}$. The formulation of thelikelihood function is well known and widely applied in statistics in case realizations of a random variable and its probability density function are available. In the present problem, the joint probability density function described in the previous subsequently is used and only one realization is considered; it is the response of the numerical model associated with $\boldsymbol{\theta}$. The likelihood function is:

$$p(\mathscr{D}\,|\,\hat{\boldsymbol{e}}, \mathscr{M}) = p_Y(\boldsymbol{y}(\hat{\boldsymbol{t}}, \hat{\boldsymbol{e}})\,|\,\hat{\boldsymbol{t}}) \tag{14}$$

with $\boldsymbol{y}(\hat{\boldsymbol{t}}, \boldsymbol{\theta}) = \left[y(\hat{t}_1, \boldsymbol{\theta}), y(\hat{t}_2, \boldsymbol{\theta}), \ldots, y(\hat{t}_{N_d}, \boldsymbol{\theta}) \right]$. The dependence with respect to the experimental data $\mathscr{D}$ is not explicitly stated in the right hand side of refeqeq:likelihood. Nevertheless, this dependence is carried by the correlation matrix and by the coefficient α_{i1} and ν_i involved in the definition of the marginal distributions. Therefore, the value of the likelihood functions associated with $\boldsymbol{\theta}$ is changed in case the experimental data is modified or expanded. Samples of the posterior distribution expressed in Equation (1) are generated using crude Monte Carlo simulation as the application examples are numerically inexpensive. Advanced methods may be applied as well as the choice of the updating algorithm does not affect the relevance of the proposed formulation of the likelihood function.

3 APPLICATION EXAMPLES

3.1 *Damped oscillator*

The methods described in the previous section are applied to an example involving a mass spring damper system as shown in Figure 2. A force F is applied at the time $t = 0$ and the quantity of interest is the displacement of the mass during 25 seconds. This problem has an analytical solutions which is used as the numerical model. The applied force and the mass are deterministic and arbitrarily set to 1 N and 1 kg, respectively, and the same parameter value is used during the updating.12 samples of the damping and of the stiffness are selected arbitrarily and the corresponding displacement of the mass is determined. These trajectories of the displacement of the mass are used as the experimental data and Bayesian updating is subsequently used to identify the posterior distribution of the uncertain parameters. Therefore, this example does not consider model uncertainty, as the same model is used to generate the experimental data and during the updating procedure.

The prior distributions are uniform between 0.5 and 2 for both the stiffness of the spring and the damping coefficient, which corresponds to a *non-informative* prior distribution between the predefined ranges. It is initially assumed that all the regions of the domain are equally likely to contain samples of the uncertain parameters. Two fractional moments are used to define the marginal probability density functions associated with the experimental data expressed in Equation (3), i.e. $M_i = 2$ for all the time instants.

The procedure described in Section 2 does not provide any strategy to set N_d, the total number

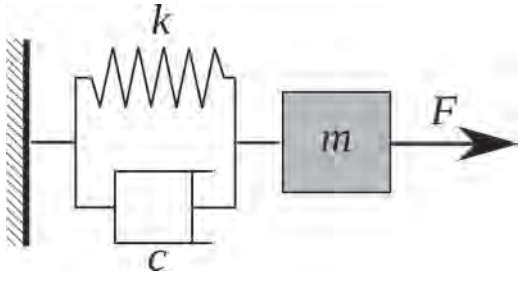


Figure 2. Mass spring damper system.

of representative time instants to be used for the updating. It is arbitrarily set to 1, 2, 3 and 5 and the results are shown in Figure 3.

3.2 *Over-constrained beam model*

The second example considered here is an over-constrained beam model as shown in Figure 4. The left hand side of the beam is clamped, the right hand side is simply supported and two forces are applied. All the units are expressed using the metricsystem (and the forces are therefore defined in Newtons). The Young's modulus of the material is equal to 200GPa, the total length of the beam is L = 1m and the moment of inertia is equal to 10^{-4}m^4. The random variables involved in the model are the forces f_1 and f_2. 20 samples are arbitrarily generated, the corresponding deflection of the beam is computed for all the coordinates of the beam and subsequently used during the Bayesian updating procedure. Therefore, the quantity of interest (i.e. the beam deflection) is expressed with respect to a space parameter. Even though a time parameter is considered in the description of the methods in Section 2; these procedures remain applicable without loss of generality to any parameterdependent problem and can therefore be applied to this example.

For both forces, the prior distribution is uniform between 4 and 16kN. Three fractional moments are used in the definition of the marginal distributions which define the likelihood function. As in the previous example, the parameter N_d (i.e. the dimensionality of the set of random variables used in the definition of the likelihood function) is arbitrarily set to 1, 2, 3 and 5. Figure 5 shows the results obtained from the Bayesian updating procedure.

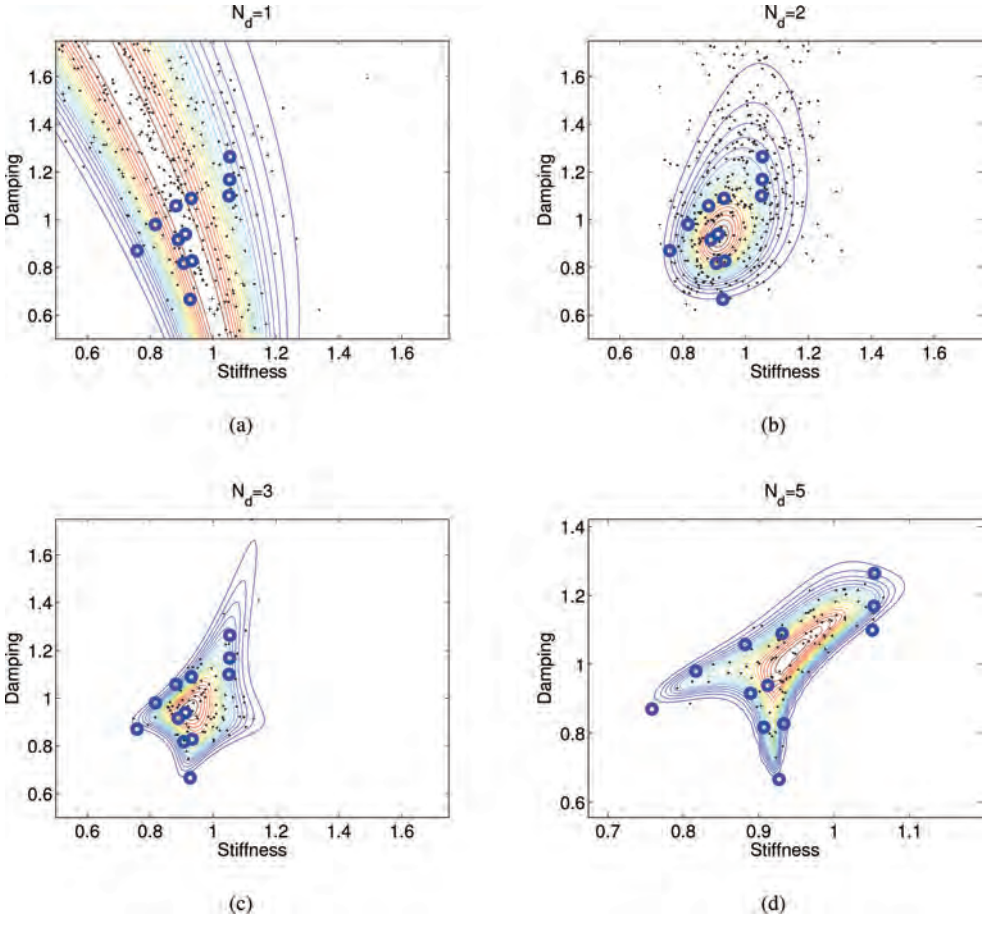


Figure 3. Results of the Bayesian updating procedure for the damped oscillator. The contour lines represent the iso-values of the posterior distribution, the dots represent realizations of this distribution and the circles are the samples used to generate the experimental data. (a) $N_d = 1$. (b) $N_d = 2$. (c) $N_d = 3$. (d) $N_d = 5$.

3.3 *Discussion*

For the two considered examples, the results are strongly influenced by N_d, the dimensionality of the set defining the likelihood function.

In case $N_d = 1$, a poor match between the posterior distributions and the samples used to generate the experimental data is observed. It is indeed not possible to identify the bivariate distribution function associated with the input parameters using a single response of the model, as an infinite number of values of the input parameters can be associated with the response.

The posterior distribution is smooth and regular for $N_d = 2$. In this case, the results are also in good agreement with the dependence in the experimental data. Moreover, it is graphically observed in the first example that the experimental input values are slightly positively correlated and negatively correlated in the second example. The same trend can be seen in the posterior distribution for both examples.

In case $N_d = 3$ and $N_d = 5$, the posterior distribution becomes irregular; it seems that the experimental data are over-fitted. The surface included in a contour line of the posterior distribution is non-convex and the regions of the parameter domain inthe vicinity of the samples used to generate the experimental data are favored (i.e. the posterior distribution has higher values in these regions).

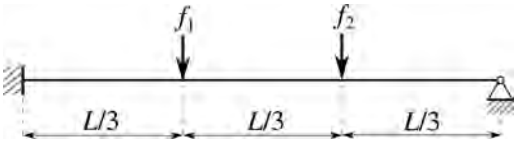


Figure 4. Over-constrained beam model.

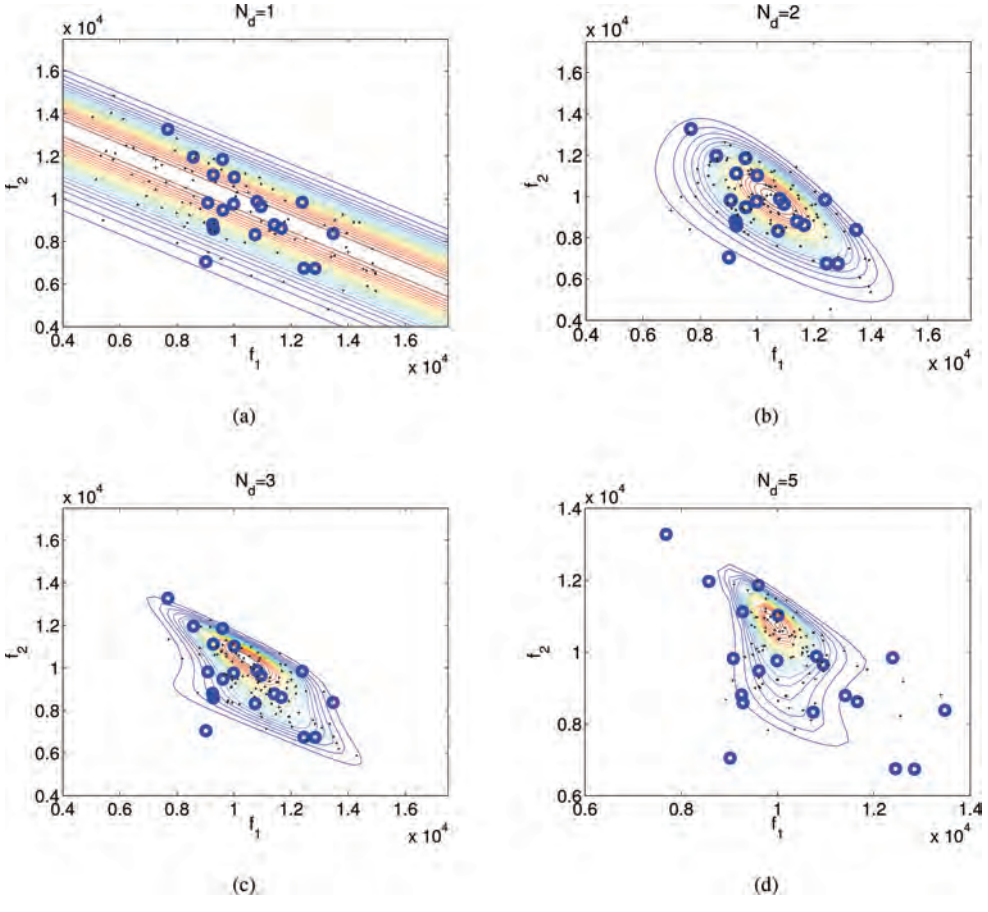


Figure 5. Results of the Bayesian updating procedure for the beam model. The contour lines represent the iso-values of the posterior distribution, the dots represent realizations of this distribution and the circles are the samples used to generate the experimental data. (a) $N_d = 1$. (b) $N_d = 2$. (c) $N_d = 3$. (d) $N_d = 5$.

In case the Bayesian updating problem does not include model uncertainties, N_d should not be greater than the dimensionality of the problem. For both examples, all the samples of the models' response used to define the likelihood function lay on a bidimensional manifold (included in a space of dimension three or five). The strategy used to define the joint probability density function associated with the response of the model becomes inappropriate, leading to the poor fit between the posterior distribution and the samples used to define the experimental data. example. As the the linear correlation coefficients are used to model the dependence between the model's response for the various time instants, the first two eigenvalues of the correlation matrix are non-negligible and all the other eigenvalues are verysmall

It seems to be empirically observed from these examples that the N_d should be equal to the total number of parameters involved in the model of uncertainty in case the problem does not involve model uncertainties.

4 CONCLUSIONS

This paper describes a method for the formulation of the likelihood function used for Bayesian updating of time dependent models. The maximum entropy principle is used to (I) define the marginal distributions for all the time instants; (II) identify the set of time instants which best represent the response of the model. The joint probability density function is defined by its marginals and by the matrix of the linear coefficients of correlation obtained after the Nataf transformation. This probability density function is subsequently used in the formulation of the likelihood, the definition of this function generally used in statistics is applied here.

The method is applied with success to a problem involving the response in the time domain of a spring damper system with two uncertain parameters: the stiffness of the spring and the damping coefficient.

The example heuristically suggests to use a parameter N_d equal to the total number of uncertain parameters involved in the numerical model (at least in case there are no model uncertainties).

Future work is geared towards the application of this method to examples involving model uncertainties and a larger number of random variables.

REFERENCES

Arora, V. (2011). Comparative study of finite element model updating methods. *Journal of Vibration and Control 17*(13), 2023–2039.

Beck, J. & K. Zuev (2013). Asymptotically independent Markov sampling: A new MCMC scheme for Bayesian inference. *International Journal for Uncertainty Quantification 3*(2), 445–474.

Beck, J. & L. Katafygiotis (1998). Updating models and their uncertainties. i: Bayesian statistical framework. *Journal of Engineering Mechanics 128*(4), 380–391.

Ching, J. & Y.-C. Chen (2007). Transitional markov chain monte carlo method for Bayesian model updating, model class selection, and model averaging. *Journal of Engineering Mechanics 133*(7), 816–832.

Ching, J., M. Muto, & J.L. Beck (2006). Structural model updating and health monitoring with incomplete modal data using Gibbs sampler. *Computer-Aided Civil and Infrastructure Engineering 21*, 242257.

Friswell, M. & J.E. Mottershead (1995). *Finite element model updating in structural dynamics*. Springer Science & Business Media.

Goller, B. (2011, June). *Stochastic Model Validation of Structural Systems*. Ph. D. thesis, University of Innsbruck.

Imregun, M. & W.J. Visser (1991). A review of model updating techniques. *Shock and Vibration Digest 23*(1), 9–20.

Jaynes, E.T. (1957a). Information theory and statistical mechanics. *Physical Review. Series II 106*(4), 620–630.

Jaynes, E.T. (1957b). Information theory and statistical mechanics ii. *Physical Review. Series II 108*(2), 171–190.

Katafygiotis, L. & J. Beck (1998). Updating models and their uncertainties. ii: Model identifiability. *Journal of Engineering Mechanics 128*(4), 463–467.

Liu, P.-L. & A. Der Kiureghian (1986). Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics 1*(2), 105–112.

Mai, J.-F. & M. Scherer (2012). *Simulating Copulas (Stochastic Models, Sampling Algorithms and Applications)*, Volume 4 of *Series in Quantitative Finance*. World Scientific.

Nagel, J. & B. Sudret (2016). A unified framework for multilevel uncertainty quantification in Bayesian inverse problems. *Probabilistic Engineering Mechanics 43*(Supplement C), 68–84.

Nataf, A. (1962). D´etermination des distributions de probabilit´es dont les marges sont données (in French). Technical report, Acad´emie des Sciences, Paris. X.

Novi Inverardi, P.L. & A. Tagliani (2003). Maximum entropy density estimation from fractional moments. *Communications in Statistics - Theory and Methods 32*(2), 327–345.

Sch¨olzel, C. & P. Friederichs (2008). Multivariate non-normally distributed random variables in climate research – introduction to the copula approach. *Nonlinear Processes in Geophysics 15*(5), 761–772.

Straub, D. & I. Papaioannou (2015). Bayesian updating with structural reliability methods. *Journal of Engineering Mechanics 141*(3), 04014134.

Taufer, E., S. Bose, & A. Tagliani (2009). Optimal predictive densities and fractional moments. *Applied Stochastic Models in Business and Industry 25*(1), 57–71.

Vanik, M.V., J.L. Beck, & S.-K. Au (2000). Bayesian probabilistic approach to structural health monitoring. *Journal of Engineering Mechanics*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Advanced methodology for uncertainty propagation in computer experiments with large number of inputs: Application to accidental scenario in a pressurized water reactor

A. Marrel
CEA, DEN, DER, SESI, LEMS, Saint-Paul-lez-Durance, France

B. Iooss
EDF R&D, Chatou, France
Institut de Mathématiques de Toulouse, Toulouse, France

ABSTRACT: Complex computer codes are often too time expensive to be directly used to perform uncertainty propagation or sensitivity analysis. A solution to cope with this problem consists in replacing the cpu-time expensive computer model by a cpu inexpensive mathematical function, called metamodel. Among the metamodels classically used in computer experiments, the Gaussian process (Gp) model has shown strong capabilities to solve practical problems. However, in case of high dimensional experiments (with typically several tens of inputs), the Gp metamodel building process remains difficult. To face this limitation, we propose a general methodology which combines several advanced statistical tools. First, an initial space-filling design is performed providing a full coverage of the high-dimensional input space (Latin hypercube sampling with optimal discrepancy property). From this, a screening based on dependence measures is performed. More specifically, the Hilbert-Schmidt independence criterion which builds upon kernel-based approaches for detecting dependence is used. It allows ordering the inputs by decreasing primary influence, for the purpose of the metamodeling. Furthermore, significance tests based either on asymptotic theory or permutation technique are performed to identify a group of potentially non-influential inputs. Then, a joint Gp metamodel is sequentially built with the group of influential inputs as explanatory variables. The residual effect of the group of non-influential inputs is captured by the dispersion part of the joint metamodel. Then, a sensitivity analysis based on variance decomposition can be performed through the joint Gp metamodel. The efficiency of the methodology is illustrated on a thermal-hydraulic calculation case simulating accidental scenario in a Pressurized Water Reactor.

1 INTRODUCTION

Quantitative assessment of the uncertainties tainting the results of computer simulations is nowadays a major topic of interest in both industrial and scientific communities. One of the key issues in such studies is to get information about the output when the numerical simulations are expensive to run. For example, in nuclear engineering problems, one often faces up with cpu time consuming numerical models and, in such cases, uncertainty propagation, sensitivity analysis, optimization processing and system robustness analysis become difficult tasks using such models. In order to circumvent this problem, a widely accepted method consists in replacing cpu-time expensive computer models by cpu inexpensive mathematical functions, called metamodels (Fang et al. 2006). This solution has been applied extensively and has shown its relevance especially when simulated phenomena are related to a small number of random input variables (see Forrester et al. (2008) for example).

However, in case of high dimensional numerical experiments (with typically several tens of inputs), depending on the complexity of the underlying numerical model, the metamodel building process remains difficult, even unfeasible. For example, the Gaussian process (Gp) model (Santner et al. 2003) which has shown strong capabilities to solve practical problems, has some caveats when dealing with high dimensional problems. The main difficulty relies on the estimation of Gp hyperparameters. Manipulating pre-defined or well-adapted Gp kernels (as in Muehlenstaedt et al. (2012), Durrande et al. (2013)) is a current research way, while coupling the estimation procedure with variable selection techniques has been proposed by several authors (Welch et al. 1992, Marrel et al. 2008, Woods and Lewis 2017).

In this paper, following the latter technique, we propose a rigorous and robust method for building a Gp metamodel with a high-dimensional vector

of inputs before using it to perform variance-based sensitivity analysis.

To build this metamodel, we use a sequential methodology where the technical core are updated with more relevant statistical techniques. For example, the screening step is raised by the use of recent and powerful techniques in terms of variable selection using a small number of model runs. Second, contrary to the previous works, we do not remove the non-selected inputs from the Gp model, keeping the uncertainty caused by the dimension reduction by using the joint metamodel technique (Marrel et al. 2012). The integration of this residual uncertainty is important in terms of robustness of subsequent safety studies and sensitivity analysis. Finally, a sensitivity analysis based on variance decomposition is performed through the joint Gp metamodel, yielding both the estimation of the influence of each selected inputs and the total effect of the group of non-selected inputs.

Each step of our methodology is detailed in a dedicated section and illustrated on a guideline application, namely a thermal-hydraulic calculation case simulating accidental scenario in a nuclear reactor. This use-case is first described in the following section.

2 THERMAL-HYDRAULIC TEST-CASE

Our use-case consists in thermal-hydraulic computer experiments, typically used in support of regulatory work and nuclear power plant design and operation. Indeed, some safety analysis considers the so-called "Loss Of Coolant Accident" (LOCA), which takes into account a double-ended guillotine break with a specific size piping rupture. It is modeled with code CATHARE 2.V2.5 which simulated the thermalhydraulic responses during a LOCA in a Pressurized Water Reactor (Mazgaj et al. 2016).

In this use-case, $d = 27$ scalar input variables of CATHARE are uncertain. They correspond to various system parameters as initial conditions, boundary conditions, some critical flowrates, interfacial friction coefficients, condensation coefficients, … The output variable of interest is a single scalar which is the maximal peak cladding temperature during the accident transient.

In our problem, minimal and maximal values are defined for each uncertain input and, in the framework of probabilistic approach, their uncertainties are modeled by probability laws defined on the domain of variation (uniform, log-uniform, truncated normal and truncated log-normal laws). Moreover, the d inputs are supposed independent. Our first objective with this use-case is to provide a good metamodel for sensitivity analysis, uncertainty propagation and, more generally, safety studies. Indeed, the cpu-time cost of this computer code is too important to develop all the statistical analysis required in a safety study only using direct calculations of the computer code. A metamodel would allow to develop more complete and robust demonstration.

In what follows, the system under study is generically denoted

$$Y = g\left(X_1, \quad , X_d\right) \tag{1}$$

where $g(\cdot)$ is the numerical model (also called the computer code), whose output Y and input parameters $X_1, \ , X_d$ belong to some measurable spaces $\mathcal{Y}$ and $\mathcal{X}_1, \ , \mathcal{X}_d$ respectively. $\mathbf{X} = \left(X_1, \quad , X_d\right)$ is the input vector and we suppose that $\mathcal{X} = \prod_{k=1}^{d} \mathcal{X}_k \subset \mathbb{R}^d$ and $\mathcal{Y} \subset \mathbb{R}$. For a given value of the vector of inputs $\mathbf{x} = \left(x_1, \ldots, x_d\right) \in \mathbb{R}^d$, a simulation run of the code yields an observed value $y = g(\mathbf{x})$.

3 STEP 1: INITIAL DESIGN OF EXPERIMENTS

The objective of the initial sampling step is to investigate the whole variation domain of the uncertain parameters in order to fit a predictive metamodel which approximates as accurately as possible the code in the whole domain of variation of the uncertain parameter. For this, we use a space-filling design (SFD) of a certain number n of experiments, providing a full coverage of the high-dimensional input space (Fang et al. 2006). This design enables to investigate the domain of variation of the uncertain parameters and provides a learning sample.

For the SFD type, a Latin Hypercube Sample (LHS) with optimal space-filling and good projection properties (Woods and Lewis 2017) would be well adapted. In particular, Fang et al. (2006) and then Damblin et al. (2013) have shown the importance of ensuring good low-order sub-projection properties. Maximum projection designs (Joseph et al. 2015) or low-centered L^2 discrepancy LHS (Jin et al. 2005) are then particularly well-suited.

Mathematically, this corresponds to the sample $\left\{\mathbf{x}^{(1)}, \ , \mathbf{x}^{(n)}\right\}$ which is performed on the model g. This yields n model output values denoted $\left\{y^{(1)}, \ , y^{(n)}\right\}$ with $y^{(i)} = g\left(\mathbf{x}^{(i)}\right)$. The obtained learning sample is denoted $\left(X_s, Y_s\right)$ with $X_s = \left[\mathbf{x}^{(1)T}, \ldots, \mathbf{x}^{(n)T}\right]^T$ and $Y_s = \left[y^{(1)}, \ldots, y^{(n)}\right]^T$. The goal is to build an approximating model of g using the n-sample $\left(X_s, Y_s\right)$.

The number n of simulations is a compromise between the CPU time required for each simula-

tion and the number of input parameters. Some thumb rules propose to choose n at least as large as 10 times the dimension d of the input vector (Loeppky et al. 2009, Marrel et al. 2008).

To build the metamodel for the LOCA test case, $N = 500$ CATHARE simulations of this test case are performed following a space-filling LHS with good projection properties as the design of experiments. The obtained inputs-output sample constitutes the learning sample.

Remark 3.1 Note that the input values are sampled following their prior distributions defined on their variation ranges. Indeed, as we are not ensured to be able to build a sufficiently accurate metamodel, we prefer sample the inputs following the probabilistic distributions in order to have at least a probabilized sample of the uncertain output, on which statistical characteristics could be estimated. Moreover, as explained in the next section, dependence measures can be directly estimated on this sample, providing first usable results of sensitivity analysis.

4 STEP 2: INITIAL SCREENING BASED ON DEPENDENCE MEASURE

From the learning sample, a screening technique is performed in order to identify the primary influential inputs (PII) on the model output variability. It has been recently shown that screening based on dependence measures (Da Veiga 2015, De Lozzo and Marrel 2016, Raguet and Marrel 2018) or on derivative-based global sensitivity measures (Kucherenko and Iooss 2017, Roustant et al. 2017) are very efficient methods which can be directly applied on a SFD. One of their great interest is that, additionally to their screening job, the sensitivity indices that they provide can be quantitatively interpreted and used to order the PII by decreasing influence, paving the way for a sequential building of metamodel. Note that Mara et al. (2017) recently compared the efficiency of several sensitivity measures to address the issue of factors fixing setting: Sobol'indices estimated with sparse polynomial chaos expansion method, density-based dependence measure and derivative-based global sensitivity measures.

In the considered LOCA test case, the adjoint model is not available and the derivatives of the model output are therefore not computed because of their costs. This considerably limits the interest of using derivative-based sensitivity measures. Moreover, as the number of uncertain inputs is large and HSIC dependence measures has showed good convergence properties (De Lozzo and Marrel 2016), we choose to use the latter for the screening step, directly estimated from the inputs-output sample (metamodel-free estimation).

4.1 *Screening based on HSIC dependence measure*

Da Veiga (2015) and more recently De Lozzo and Marrel (2016) have proposed to use dependence measures for screening purpose, by applying them directly on a SFD. These sensitivity indices are not the classical ones variance-based measures (see Iooss and Lemaître 2015 or Borgonovo and Plischke 2016, for a review on global sensitivity analysis methods). They consider higher order information about the output behavior in order to provide more detailed information. Among them, the Hilbert-Schmidt independence criterion (HSIC) introduced by Gretton et al. (2005) builds upon kernel-based approaches for detecting dependence, and more particularly on cross-covariance operators in reproducing kernel Hilbert spaces (RKHS).

If we consider two RKHS $\mathcal{F}_k$ and $\mathcal{G}$ of functions $\mathcal{X}_k \to \mathbb{R}$ and $\mathcal{Y} \to \mathbb{R}$ respectively, the crossed-covariance $C_{X_k,Y}$ operator associated to the joint distribution of $t(X_k, Y$ is the linear operator defined for every $f_{X_k} \in \mathcal{F}_k$ and $g_Y \in \mathcal{G}$ by:

$$< f_{X_k}, C_{X_k,Y} g_Y >_{\mathcal{F}_k} = \mathrm{Cov}\left(f_{X_k}, g_Y\right). \tag{2}$$

$C_{X_k,Y}$ generalizes the covariance matrix by representing higher order correlations between X_k and Y through nonlinear kernels. The HSIC criterion is then defined by the Hilbert-Schmidt norm of the cross-covariance operator:

$$\mathrm{HSIC}(X_k, Y)_{\mathcal{F}_k,\mathcal{G}} = \| C_k \|^2_{HS} . \tag{3}$$

From this, Da Veiga (2015) introduces a normalized version of the HSIC which provides a sensitivity index of X_k:

$$R^2_{\mathrm{HSIC},k} = \frac{HSIC\left(X_k, Y\right)}{\sqrt{HSIC\left(X_k, X_k\right) HSIC\left(Y, Y\right)}}. \tag{4}$$

Gretton et al. (2005) also propose a Monte Carlo estimator of $\mathrm{HSIC}\left(X_k, Y\right)$ and a plug-in estimator can be deduced for $R^2_{\mathrm{HSIC},k}$. Note that Gaussian kernel functions with empirical estimations of the variance parameter are used in our application (see Gretton et al. 2005 for details).

Then, from the estimated R^2_{HSIC}, independence tests can be performed for a screening purpose. The objective is to separate the inputs into two sub-groups, the significant ones and the non-significant ones. For a given input X_k, it aims at testing the null hypothesis "$\mathcal{H}_0^{(k)}$: X_k and Y are

independent", against its alternative "$\mathcal{H}_1^{(k)}$: X_k and Y are dependent". The significance level[1] of these tests is hereinafter noted α. Several statistical hypothesis tests are available: asymptotic versions, spectral extensions and bootstrap versions for non-asymptotic case. All these tests are described and compared in De Lozzo and Marrel (2016); a guidance to use them for a screening purpose is also proposed. At the end of the screening step, the inputs selected as significant are also ordered by decreasing R^2_{HSIC}. This order will be used for the sequential metamodel building in step 3.

4.2 *Application on LOCA test case*

From the learning sample of $N = 500$ simulations, R^2_{HSIC} dependence measures are estimated and bootstrap tests with $\alpha = 0.1$ are performed. Eleven inputs are selected as significantly influential. Ordering them by decreasing R^2_{HSIC} reveals the predominance influence of X_{10} ($R^2_{\text{HSIC}} \approx 0.39$), followed by X_2, X_{12} and X_{22}, ($R^2_{\text{HSIC}} \approx 0.04$, 0.02 and 0.02 respectively). X_{15}, X_{13}, X_9, X_5, X_{14}, X_{26} and X_{27} have a lower influence (R^2_{HSIC} around 0.01)) and the others variables are considered as negligible by statistical tests.

Note that the estimated HSIC and the results of significant tests are relatively stable when the learning sample size varies from $N = 300$ to $N = 500$. Only two or three selected variables with a very low HSIC (R^2_{HSIC} around 0.01) can differ. This confirms the robustness of the HSIC indices and the associated significance tests for qualitative sorting and screening purpose.

In the next steps, the eleven significant inputs are considered as the explanatory variables, denoted PII, in the joint metamodel and will be successively included in the building process. The other sixteen variables will be joined in a so-called *uncontrollable* parameter.

5 STEP 3: JOINT GP METAMODEL WITH SEQUENTIAL BUILDING PROCESS

Among all the metamodel-based solutions (polynomials, splines, neural networks, etc.), we focus our attention on the Gaussian process (Gp) regression, which extends the kriging principles of geostatistics to computer experiments by considering the correlation between two responses of a computer code depending on the distance between input variables. The Gp-based metamodel presents some real advantages compared to other metamodels: exact interpolation property, simple analytical formulations of the predictor, availability of the mean squared error of the predictions and the proved efficiency of the model (Santner et al. 2003).

However, for its application to complex industrial problems, developing a robust implementation methodology is required. Indeed, fitting a Gp model implies the estimation of several hyperparameters involved in the covariance function. In complex situations (e.g. large number of inputs), some difficulties can arise from the parameter estimation procedure (instability, high number of hyperparameters, see Marrel et al. 2008 for example). To tackle this issue, we propose a progressive estimation procedure which combines the result of the previous screening step and a joint Gp approach (Marrel et al. 2012).

5.1 *Sequential building process based on successive inclusion of explanatory variables*

At the end of the screening step, the inputs selected as significant (group of PII) are ordered by decreasing influence. The sorted PII are successively included in the metamodel explanatory inputs while the other inputs (remaining PII and the sixteen non-selected inputs) are joined in a single macro-parameter which is considered as an uncontrollable parameter (i.e. a stochastic parameter, notion detailed in section 5.2). Thus, at the j^{th} iteration, a joint Gp metamodel is built with, as explanatory inputs, the j sorted PII. The definition and building procedure of a joint Gp is fully described in Marrel et al. (2012) and summarized in the section 5.2.

However, building a Gp or a joint Gp involves to perform a numerical optimization in order to estimate all the parameters of the metamodel (covariance hyperparameters and variance parameter). As we usually consider in computer experiments anisotropic (stationary) covariance, the number of hyperparameters linearly increases with the number of inputs. In order to improve the robustness of the optimization process and deal with a large number of inputs, the estimated hyperparameters obtained at the $(j-1)^{th}$ iteration are used, as starting points for the optimization algorithm. This procedure is repeated until the inclusion of all the PII. Note that this sequential estimation process is directly adapted from the one proposed by Marrel et al. (2008).

5.2 *Joint Gp metamodel*

In the framework of stochastic computer codes, Zabalza et al. (1998) proposed to model the mean and dispersion of the code output by two interlinked Generalized Linear Models (GLM), called "joint GLM". Marrel et al. (2012) extends this approach to several nonparametric models and

[1]The significance level of a statistical hypothesis test is the rate of the type I error which corresponds to the rejection of the null hypothesis $\mathcal{H}_0$ when it is true.

obtains the best results with two interlinked Gp models, called "joint Gp". In this case, the stochastic input is considered as an uncontrollable parameter denoted $\mathbf{X}_\varepsilon$ (i.e. governed by a seed variable).

We extend this approach to a group of non-explanatory variables. More precisely, the input variables $\mathbf{X} = (X_1, \quad , X_d)$ are divided in two subgroups: the explanatory ones denoted $\mathbf{X}_{\mathbf{exp}}$ and the others denoted $\mathbf{X}_\varepsilon$. The output is thus defined by $y = g(\mathbf{X}_{\mathbf{exp}}, \mathbf{X}_\varepsilon)$. Under this hypothesis, the joint metamodeling approach yields building two metamodels, one for the mean Y_m and another for the dispersion component Y_d:

$$Y_m(\mathbf{X}_{\mathbf{exp}}) = \mathbb{E}(Y \mid \mathbf{X}_{\mathbf{exp}}) \tag{5}$$

$$Y_d(\mathbf{X}_{\mathbf{exp}}) = \mathrm{Var}(Y \mid \mathbf{X}_{\mathbf{exp}}) = \mathbb{E}\left[\left(Y - Y_m(\mathbf{X}_{\mathbf{exp}}) \right)^2 \mid \mathbf{X}_{\mathbf{exp}} \right]. \tag{6}$$

To fit these mean and dispersion components, we propose to use the methodology proposed by Marrel et al. (2012). First, an initial Gp denoted $Gp_{m,1}$ is estimated for the mean component with homoscedastic nugget effect. A nugget effect is required to relax the interpolation property of the Gp metamodel, which would yield zero residuals for the whole learning sample. Then, a second Gp, denoted $Gp_{v,1}$, is built for the dispersion component with, here also, an homoscedastic nugget effect. $Gp_{v,1}$ is fitted on the squared residuals from the predictor of $Gp_{m,1}$. Its predictor is considered as an estimator of the dispersion component. The predictor of $Gp_{v,1}$ provides an estimation of the dispersion at each point, which is considered as the value of the heteroscedastic nugget effect. The homoscedastic hypothesis is so removed and a new Gp, denoted $Gp_{m,2}$, is fitted on data, with the estimated heteroscedastic nugget. Finally, the Gp on the dispersion component is updated from $Gp_{m,2}$ following the same methodology as for $Gp_{v,1}$.

Remark 5.1 Note that some parametric choices are made for all the Gp metamodels: a constant trend and a Matérn stationary anisotropic covariance are chosen. All the hyperparameters (covariance parameters) and the nugget effect (when homoscedastic hypothesis is done) are estimated by maximum likelihood optimization process.

5.3 *Assessment of metamodel accuracy*

To evaluate the accuracy of the metamodel, we use the predictivity coefficient Q^2:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_{test}} \left(y^{(i)} - \hat{y}^{(i)} \right)^2}{\sum_{i=1}^{n_{test}} \left(y^{(i)} - \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} y^{(i)} \right.} \tag{7}$$

Table 1. Evolution of Gp_m metamodel predictivity during the sequential process building, for each new additional PII.

Additional PII	X_{10}	X_2	X_{12}	X_{22}	X_{15}	X_{13}
Q^2	0.60	0.64	0.70	0.79	0.81	0.83
Additional PII	X_9	X_5	X_{14}	X_{26}	X_{27}	
Q^2	0.85	0.85	0.87	0.87	0.87	

where $(x^{(i)})_{1 \le i \le n_{\text{test}}}$ is a test sample, $(y^{(i)})_{1 \le i \le n_{\text{test}}}$ are the corresponding observed outputs and $(\hat{y}^{(i)})_{1 \le i \le n_{\text{test}}}$ are the metamodel predictions. Q^2 corresponds to the coefficient of determination in prediction and can be computed on a test sample independent from the learning sample or by cross-validation on the learning sample. The closer to one the Q^2, the better the accuracy of the metamodel.

5.4 *Application on LOCA test case*

The joint Gp metamodel is built from the learning sample of $N = 500$: the eleven PII identified at the end of the the screening step are considered as the explanatory variables while the sixteen others are considered as the uncontrollable parameter. Gps on mean and dispersion components are built using the sequential building process described in section 5.1 where PII ordered by decreasing R^2_{HSIC} are successively included in Gp. Q^2 coefficient of mean component Gp_m is computed by cross validation at each iteration of the sequential building process. The results which are given by Table 1 show an increasing predictivity until its stabilization around 0.87, which illustrates the robustness of building process. The first four PII make the major contribution yielding a Q^2 around 0.8, the four following ones yield minor improvements (increase of 0.02 on average for each input) while the three last PII does not improve the Gp predictivity.

Thus, only 13% of the output variability remains not explained by Gp_m, this includes both the inaccuracy of the Gp_m (part of Y_m not fitted by Gp) and the total effect of the uncontrollable parameter, i.e. the group of non-selected inputs.

6 STEP 4: VARIANCE-BASED SENSITIVITY ANALYSIS

Sensitivity Analysis (SA) methods allow to answer the question "How do the input parameters variations contribute, qualitatively or quantitatively, to the variation of the output?" (Saltelli et al. 2008). These tools can detect non-significant input parameters in a screening context, determinate the most significant ones, measure their respective contributions to the output or identify an interaction between several inputs which impacts strongly

the model output. From this, engineers can guide the characterization of the model by reducing the output uncertainty: for instance, they can calibrate the most influential inputs and fix the non-influential ones to nominal values. Many surveys on SA exist in the literature, such as Kleijnen (1997), Frey and Patil (2002) or Helton et al. (2006). SA can be divided into two sub-domains: the Local SA (LSA) and the Global SA (GSA). The first one studies the effects of small input perturbations around nominal values on the model output (Cacuci 1981) while the second one considers the impact of the input uncertainty on the output over the whole variation domain of uncertain inputs (Saltelli et al. 2008). We focus here on one of the most widely used GSA indices, namely Sobol' indices which are based on output variance decomposition.

6.1 *Sobol' indices*

A classical approach in GSA consists of computing the first-order and total Sobol' indices which are based on the output variance decomposition (Sobol 1993, Homma and Saltelli 1996). If the variables $X_1, \ldots, X_d$ are independent and if $\mathbb{E}\left[g^2(X)\right] < +\infty$, we can apply the Hoeffding decomposition to the random variable $g(X)$ (Hoeffding 1948):

$$g(X) = \sum_{u \subset \{1,\ldots,d\}} g_u(X_u) \tag{8}$$

where $g_\phi = \mathbb{E}\left[g(X) \right]$, $g_i(X_i) = \mathbb{E}[g(X) \mid X_i] - g_\phi$ and $g_u(X_u) = \mathbb{E}[g(X) \mid X_u] - \sum_{v \subset u} g_v(X_v)$, with $X_u = (X_i)_{i \in u}$, for all $u \subset \{1,\ldots,d\}$. All the 2^d terms in (8) have zero mean and are mutually uncorrelated with each other. This decomposition is unique and leads to the Sobol' indices. These are the elements of the $g(X)$ variance decomposition according to the different groups of input parameter interactions in (8). More precisely, for each $u \subset \{1,\ldots,d\}$, the first-order and total Sobol' sensitivity indices of X_u are defined by:

$$S_u = \frac{Var\left[g_u(X_u)\right]}{Var\left[g(X)\right]} \text{ and } S_u^T = \sum_{v \supset u} S_v.$$

S_u represents the part of the output variance explained by X_u, independently from the other inputs, and S_u^T is the part of the output variance explained by X_u considered separately and in interaction with the other input parameters.

In practice, we are usually interested in the first-order sensitivity indices $S_1, \ldots, S_d$, the total ones $S_1^T, \ldots, S_d^T$ and sometimes in the second-order ones S_{ij}, $1 \le i < j \le d$. The model g is devoid of interactions if $\sum_{i=1}^{d} S_i \approx 1$.

Sobol' indices are widely used in GSA because they are easy to interpret and directly usable in a dimension reduction approach. However, their estimation (based on Monte-Carlo methods for example) requires a large number of model evaluations, which is intractable for time expensive computer codes. An common solution consists in using a metamodel to compute these indices. Note that, when the Q^2 of the metamodel is estimated on a probabilized sample of the inputs, it provides an estimation of the part of variance unexplained by the metamodel. This can be kept in mind when interpreting the Sobol' indices estimated with the metamodel.

6.2 *Sobol' indices with a joint Gp metamodel*

In the case where a joint Gp metamodel is used to take into account an uncontrollable input X_ε, we have shown in Marrel et al. (2012) how to deduce Sobol' sensitivity indices from this joint metamodel. Indeed, the variance of the output variable $Y\left(\mathbf{X_{exp}}, X_\varepsilon\right)$ can be rewritten and deduced from the two metamodels:

$$\mathrm{Var}\, Y\left(\mathbf{X_{exp}}, X_\varepsilon\right) \mathrm{Var}_{\mathbf{X_{exp}}}\left[Y_m\left(\mathbf{X_{exp}}\right)\right] + \mathbb{E}_{\mathbf{X_{exp}}}\left[Y_d\left(\mathbf{X_{exp}}\right)\right] \tag{9}$$

where $\mathbb{E}_X$ (resp. Var_X) denotes the mean (resp. variance) operator with respect to the pdf of X. Furthermore, the variance of Y is the sum of the contributions of all the d controllable inputs $\mathbf{X_{exp}} = (X_1, \ldots, X_d)$ and the uncontrollable one X_ε:

$$\mathrm{Var}(Y) = V_\varepsilon(Y) + \sum_{i=1}^{d} \sum_{|J|=i} [V_J(Y) + V_{J\varepsilon}(Y)] \tag{10}$$

where $V_\varepsilon(Y) = \mathrm{Var}_{X_\varepsilon}[\mathbb{E}_{\mathbf{X_{exp}}}(Y \mid X_\varepsilon)]$, $V_i(Y) = \mathrm{Var}_{\mathbf{X_i}}[\mathbb{E}_{\mathbf{X_{-i}}}(Y \mid X_i)]$, $V_{i\varepsilon}(Y) = \mathrm{Var}_{X_i X_\varepsilon}[\mathbb{E}_{\mathbf{X_{exp,-i}}}(Y \mid X_i X_\varepsilon)] - V_i(Y) - V_\varepsilon(Y)$, $V_{ij}(Y) = \mathrm{Var}_{X_i X_j}[\mathbb{E}_{\mathbf{X_{-i,-j}}}(Y \mid X_i X_j)] - V_i(Y) - V_j(Y)$...

Variance of the mean component $Y_m(\mathbf{X})$ denoted hereafter Y_m can be also decomposed:

$$\mathrm{Var}(Y_m) = \sum_{i=1}^{d} \sum_{|J|=i} V_J(Y_m). \tag{11}$$

As $V_i(Y_m) = \mathrm{Var}_{X_i}\mathbb{E}_{\mathbf{X_{exp,-i}}}[\mathbb{E}_{X_\varepsilon}(Y \mid \mathbf{X_{exp}}) \mid X_i] = V_i(Y)$, Sobol' indices according to input variables $\mathbf{X}_{exp} = (X_i)_{i=1\ldots d}$ can be derived and estimated from Y_m:

$$S_J = \frac{V_J(Y_m)}{Var(Y)} \text{ for any } J \subset \mathbf{X}_{exp}. \tag{12}$$

Similarly, the total sensitivity index of X_ε is given by:

$$S_{\varepsilon}^{tot} = \frac{V_{\varepsilon}(Y) + \sum_{i=1}^{d}\sum_{|J|=i} V_{J\varepsilon}(Y)}{Var(Y)} = \frac{\mathbb{E}_{\mathbf{X}_{exp}}\left[Y_d\left(\mathbf{X}_{exp}\right)\right]}{Var(Y)}. \quad (13)$$

Note that, as $Y_d\left(\mathbf{X_{exp}}\right)$ is a positive random variable, positivity of S_{ε}^{tot} is guaranteed. In practice, Var(Y) can be estimated from the data or from simulations of the fitted joint model, using equation (9).

S_{ε}^{tot} is interpreted as the total sensitivity index of the uncontrollable process. The limitation of this approach is that only the total part of uncertainty related to X_{ε} is estimated; its individual effect is not distinguished from its interaction with the other parameters. However, these potential interactions could be pointed out, considering all the primary and total effects of all the other parameters. The SA of Y_d can also be a relevant indicator: if an input variable X_i is not influential on Y_d, we can deduce that $S_{i\varepsilon}$ is equal to zero.

6.3 *Results on LOCA test case*

From the joint Gp built in section 5.4, Sobol' indices of PII are estimated from Gp_m metamodel using equation (12), Var(Y) being estimated with Gp_m and Gp_d using equation (9). For this, intensive Monte Carlo methods are used (see e.g. pick-and-freeze estimator of Gamboa et al. 2016). The first Sobol' indices of PII are given by Table 2 and represent 85% of the total variance of the output. X_{10} remains the major influential input with 59% of explained variance, followed to a lesser extend by X_{12} and X_{22} with for each of them 8% of variance. The *partial* total Sobol' indices involving only PII and derived from Gp_m show that additional 4% of variance is due to interaction between X_{10}, X_{12} and X_{22}. The other PII have negligible influence. Lastly, all PII explain around 89% of the output variance, of which 79% is only due to X_{10}, X_{12} and X_{22}. From Gp_d metamodel and using equation (13),the total effect of the uncontrollable parameter, i.e. the group of the sixteen not-explanatory inputs, is estimated to 9.7%. This includes the effect of the uncontrollable parameter alone and in interaction with the PII. To further investigate these interactions, Sobol'-based SA and HSIC-based statistical dependence tests are applied on Y_d and reveal that only X_{10}, X_{14}, X_2, X_{22} and X_4 potentially interact with the uncontrollable parameter.

Table 2. First Sobol' indices of PII (in %), estimated with Gp_m metamodel.

Input	X_{10}	X_2	X_{12}	X_{22}	X_{15}	X_{13}
1st Sobol' index	59	3	8	8	2	1
Input	X_9	X_5	X_{14}	X_{26}	X_{27}	
1st Sobol' index	2	0	2	0	0	

7 CONCLUSION AND PROSPECTS

Using an efficient sequential building process, we fitted a predictive joint Gp metamodel on a high dimensional thermal-hydraulic test case simulating accidental scenario in a Pressurized Water Reactor (LOCA test case). An initial screening step based on advanced dependence measures and associated statistical tests enabled to identify a group of significant inputs, allowing dimension reduction. The efforts of optimization when fitting the metamodel fitting can be concentrated on the main influential inputs and the robustness of metamodeling is thus increased. Moreover, thanks to the joint metamodel approach, the non-selected inputs are not completely removed: the residual uncertainty due to dimension reduction is integrated in the metamodel and the global influence of non-selected inputs is so controlled.

From this joint Gp metamodel, several statistical analyses, not feasible with the numerical model due to its computational cost, become accessible. Thus, on LOCA application, a sensitivity analysis based on variance decomposition is performed using the joint Gp: Sobol' indices are computed and reveal that the output is mainly explained by four uncertain inputs: one input is strongly influential with around 60% of output variance explained, the three others being of minor influence. The quite less influence of all the other inputs is also confirmed.

The next step is to use the joint Gp metamodel to perform uncertainty propagation for the estimation of failure probabilities and quantiles. In the LOCA test case, we are particularly interested by the estimation of high quantile (at the order of 95% to 99%) of the model output temperature. In nuclear safety, methods of conservative computation of quantiles (Nutt and Wallis 2004) have been largely studied. However, several complementary information are often useful and are not accessible in a high-dimensional context. Then, we expect that the joint Gp metamodel could help to access this information: the uncertainty of the influential inputs will be directly and accurately propagated through the mean component of the joint metamodel while a confidence bound could be derived from the dispersion component in order to take into account the residual uncertainty of the other inputs. On this last point, the interest of heteroscedastic approach in joint Gp could also be illustrated and compared with its homoscedastic version.

ACKNOWLEDGMENTS

We are grateful to Henri Geiser and Thibault Delage who performed the computations of the CATHARE code.

REFERENCES

Borgonovo, E. & E. Plischke (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research 248*(3), 869–887.

Cacuci, D. (1981). Sensitivity theory for nonlinear systems. I. Nonlinear functional analysis approach. *Journal of Mathematical Physics 22*, 2794.

Damblin, G., M. Couplet, & B. Iooss (2013). Numerical studies of space filling designs: Optimization of Latin hypercube samples and subprojection properties. *Journal of Simulation 7*, 276–289.

Da Veiga, S. (2015). Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation 85*, 1283–1305.

De Lozzo, M. & A. Marrel (2016). New improvements in the use of dependence measures for sensitivity analysis and screening. *Journal of Statistical Computation and Simulation 86*, 3038–3058.

Durrande, N., D.G.O., Roustant, & L. Carraro (2013). ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis. *Journal of Multivariate Analysis 155*, 57–67.

Fang, K.-T., R. Li, & A. Sudjianto (2006). *Design and modeling for computer experiments*. Chapman & Hall/CRC.

Forrester, A., A. Sobester, & A. Keane (Eds.) (2008). *Engineering design via surrogate modelling: a practical guide*. Wiley.

Frey, H.& S. Patil (2002). Identification and review of sensitivity analysis methods. *Risk Analysis 22*, 553–578.

Gamboa, F., A. Janon, T. Klein, A. Lagnoux, & C. Prieur (2016). Statistical inference for sobol pick freeze Monte Carlo methods. *Statistics 50*, 881–902.

Gretton, G., O. Bousquet, A. Smola, & B. Schölkopf (2005). Measuring statistical dependence with hilbert-schmidt norms. In *Proceedings Algorithmic Learning Theory*, pp. 63–77. Springer-Verlag.

Helton, J., J. Johnson, C. Salaberry, & C. Storlie (2006). Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering and System Safety 91*, 1175–1209.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics 19*, 293–325.

Homma, T. & A. Saltelli (1996). Importance measures in global sensitivity analysis of non linear models. *Reliability Engineering and System Safety 52*, 1–17.

Iooss, B. & P. Lemaˆıtre (2015). A review on global sensitivity analysis methods. In C. Meloni and G. Dellino (Eds.), *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications*. Springer.

Jin, R., W. Chen, & A. Sudjianto (2005). An efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference 134*, 268–287.

Joseph, V., E. Gul, & S. Ba (2015). Maximum projection designs for computer experiments. *Biometrika 102*, 371–380.

Kleijnen, J. (1997). Sensitivity analysis and related analyses: a review of some statistical techniques. *Journal of Statistical Computation and Simulation 57*, 111–142.

Kucherenko, S. & B. Iooss (2017). Derivative-based global sensitivity measures. In R. Ghanem, D. Higdon, and H. Owhadi (Eds.), *Springer Handbook on Uncertainty Quantification*. Springer.

Loeppky, J., J. Sacks, & W. Welch (2009). Choosing the sample size of a computer experiment: A practical guide. *Technometrics 51*, 366–376.

Mara, T., B. Belfort, V. Fontaine, & A. Younes (2017). Addressing factors fixing setting from given data: A comparison of different methods. *Environmental Modelling and Software 87*, 29–38.

Marrel, A., B. Iooss, F. Van Dorpe, & E. Volkova (2008). An efficient methodology for modeling complex computer codes with Gaussian processes. *Computational Statistics and Data Analysis 52*, 4731–4744.

Marrel, A., B. Iooss, S. Da Veiga, & M. Ribatet (2012). Global sensitivity analysis of stochastic computer models with joint metamodels. *Statistics and Computing 22*, 833–847.

Mazgaj, P., J.-L. Vacher, & S. Carnevali (2016). Comparison of CATHARE results with the experimental results of cold leg intermediate break LOCA obtained during ROSA-2/LSTF test 7. *EPJ Nuclear Sciences & Technology 2*(1).

Muehlenstaedt, T., O. Roustant, L. Carraro, & S. Kuhnt (2012). Data-driven Kriging models based on FANOVAdecomposition. *Statistics & Computing 22*, 723–738.

Nutt, W. & G. Wallis (2004). Evaluation of nuclear safety from the outputs of computer codes in the presence of uncertainties. *Reliability Engineering and System Safety 83*, 57–77.

Raguet, H. & A. Marrel (2018). Target and conditional sensitivity analysis with emphasis on dependence measures. *ArXiv e-prints*.

Roustant, O., F. Barthe, & B. Iooss (2017). Poincaré inequalities on intervals - application to sensitivity analysis. *Submitted https://hal.archives-ouvertes.fr/hal-01388758*.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Salsana, & S. Tarantola (2008). *Global sensitivity analysis - The primer*. Wiley.

Santner, T., B. Williams, & W. Notz (2003). *The design and analysis of computer experiments*. Springer.

Sobol, I. (1993). Sensitivity estimates for non linear mathematical models. *Mathematical Modelling and Computational Experiments 1*, 407–414.

Welch, W., R. Buck, J. Sacks, H. Wynn, T. Mitchell, & M. Morris (1992). Screening, predicting, and computer experiments. *Technometrics 34*(1), 15–25.

Woods, D. & S. Lewis (2017). Design of experiments for screening. In R. Ghanem, D. Higdon, and H. Owhadi (Eds.), *Springer Handbook on Uncertainty Quantification*. Springer.

Zabalza, I., J. Dejean, & D. Collombier (1998, september). Prediction and density estimation of a horizontal well productivity index using generalized linear models. In *ECMOR VI, Peebles*.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Accelerated degradation model based on geometric Liu process

Ji-Peng Wu, Xiao-Yang Li & Rui Kang
School of Reliability and Systems Engineering, Beihang University, Beijing, China
Science and Technology on Reliability and Environmental Engineering Laboratory, Beijing, China

ABSTRACT: Through evaluating stress levels, Accelerated Degradation Testing (ADT) can obtain degradation data in a limited period of time and then use these data for reliability evaluations. However, because of the high price of the test items or the test equipment, the sample size used in ADT is usually small, which causes a lack of knowledge on recognizing the population and then lead to the epistemic uncertainty. The small sample problem makes the probability theory based models, which need large samples, not appropriate any more. To address this problem, based on the uncertain theory, this paper uses the general geometric Liu process to construct an uncertain acceleration degradation model, and gives the corresponding statistical analysis method with objective measures. A carbon-film resistors case is used to illustrate the proposed methodology, and discussions are conducted on the sensitivity analysis of the proposed methodology to the sample sizes. Results show that the proposed methodology is a suitable choice for the reliability evaluations of ADT data under small sample situations.

1 INTRODUCTION

Products' reliability and lifetime are usually assessed by the life testing that uses time-to-failure data. But for highly reliable products, there are usually few or even no failure during the life testing, which makes it unappropriated to use life testing to assess these products' reliability and lifetime. Therefore, the accelerated reliability testing (ADT) has attracted much attention and been widely applied. ADT can obtain degradation data in a limited period of time by evaluating stress levels, and then uses these data for the reliability and lifetime assessments.

In a standard ADT analysis, there is a degradation model and an acceleration model. The degradation model describes the degradation paths under each stress level. Some of the parameters are assumed to be functions of stress levels, i.e., the acceleration model. In general, there are two broad categories of degradation models based on the probability theory, which are the degradation path models (Meeker et al., 1998) and the stochastic process models (Ye and Xie, 2015). These models are suitable for the situation where there are large samples. But in practical applications, the sample size in ADT is usually small due to the high price of the test items or the test equipment, which will cause a lack of knowledge on recognizing the population, and then lead to the epistemic uncertainty. Therefore, the probability theory based models are not appropriate for the small sample situation.

To quantify the epistemic uncertainty, various methods have been applied by utilizing subjective information such as belief degrees, including the Bayesian method (Li and Meeker, 2014), the interval analysis (Moore et al., 2009), and the fuzzy probability theory (Beer et al., 2013). The prior distributions, the intervals or the fuzzy variables are used respectively in these methods to utilize subjective information to quantify the epistemic uncertainty (Kang et al., 2016).

However, there still exit some problems. On the one hand, these methods quantify the epistemic uncertainty by subjective measures, which could result in different results from different researchers. On the other hand, these methods origin from the probability theory, which makes them unsuitable for the ADT data with small sample size.

Motivated by these problems, the uncertainty theory proposed by Liu (Liu, 2015) is introduced to the field of ADT modeling. The uncertainty theory is a branch of mathematics for modeling belief degrees and is used for the small sample (or even no sample) situations (Liu, 2012). It has been widely used in many fields such as risk assessment (Liu, 2010), reliability analysis (Zeng et al., 2013), supply chain (Huang et al., 2016), and so on.

In this paper, based on the uncertainty theory, a positive uncertain process named the general geometric Liu process is proposed to construct an uncertain accelerated degradation model, and the statistical analysis method for parameter estimations is proposed correspondingly. The proposed methodology quantifies the epistemic uncertainty by objective measures. The rest of the paper is organized as follows. Section 2 introduces preliminaries about the

uncertainty theory. Section 3 presents the uncertain accelerated degradation model, derives the reliability and lifetime distributions and gives the corresponding statistical analysis method. Section 4 conducts the case study and the sensitivity analysis. Section 5 concludes the paper.

2 PRELIMINARIES

In this section, we introduce some preliminaries about uncertain measure, uncertain variable, and uncertain process that will be used in the subsequent sections.

Definition 1 (Liu, 2015): Let Γ be a nonempty set, and $\mathcal{L}$ be a σ-algebra over Γ. Each element Λ in $\mathcal{L}$ is called a measurable set. A set function M from $\mathcal{L}$ to [0, 1] is called an uncertain measure if it satisfies the following axioms:

1. **Normality axiom:** $\mathrm{M}\{\Gamma\}=1$ for the universal set Γ.
2. **Duality axiom:** $\mathrm{M}\{\Lambda\}+\mathrm{M}\{\Lambda^c\}=1$ for any event Λ.
3. **Subadditivity axiom:** For every countable sequence of events $\Lambda_1,\Lambda_2,\ldots$, we have

$$\mathrm{M}\left\{\bigcup_{i=1}^{\infty}\Lambda_i\right\}\le\sum_{i=1}^{\infty}\mathrm{M}\{\Lambda_i\}. \tag{1}$$

4. **Product axiom** (Liu, 2009): Let $(\Gamma_k, \Gamma_k, \mathrm{M}_k)$ be uncertainty spaces for $k = 1, 2, \ldots$, then the product uncertain measure M is an uncertain measure satisfying

$$\mathrm{M}\left\{\prod_{k=1}^{\infty}\Lambda_k\right\}=\bigwedge_{k=1}^{\infty}\mathrm{M}_k\{\Lambda_k\}, \tag{2}$$

where Λ_k are arbitrarily chosen events from $\mathcal{L}_k$ for $k = 1,2,\ldots$, respectively.

Definition 2 (Liu, 2015): Liu introduced the concept of uncertainty distribution to describe uncertain variables. The uncertainty distribution Φ of an uncertain variable ξ is defined by

$$\Phi(x)=\mathcal{M}\{\xi\le x\},\quad\forall x\in\mathcal{R}. \tag{3}$$

Let be ξ an uncertain variable with regular uncertainty distribution $\Phi(x)$. Then the inverse function $\Phi^{-1}(\alpha)$ is called the inverse uncertainty distribution of ξ.

Definition 3 (Liu, 2008): Let T be a totally ordered set (that is usually "time"), and let $(\Gamma,\mathcal{L},\mathrm{M})$ be an uncertainty space. An uncertain process is defined as a measurable function from $T\times(\Gamma,\mathcal{L},\mathrm{M})$ to the set of real numbers, i.e., for each $t\in T$ and any Borel set B of real numbers, the set

$$\{X_t\in B\}=\{\gamma\in\Gamma\mid X_t(\gamma)\in\mathrm{B}\}\} \tag{4}$$

is an event. In other words, an uncertain process is a sequence of uncertain variables indexed by time.

Definition 4 (Liu, 2014): An uncertain process X_t is said to have an uncertainty distribution $\Phi_t(x)$ if at each time t, the uncertain variable X_t has the uncertainty distribution $\Phi_t(x)$.

Theorem 1 (Liu, 2014): (Sufficient and Necessary Condition) A function $\Phi_t^{-1}(\alpha):T\times(0,1)\to\Re$ is an inverse uncertainty distribution of independent uncertain process if and only if 1) at each time t, $\Phi_t^{-1}(\alpha)$ is a continuous and strictly increasing function; and 2) for any times $t_2<t_1$, $\Phi_{t_1}^{-1}(\alpha)-\Phi_{t_2}^{-1}(\alpha)$ is a monotone increasing function with respect to α.

3 METHODOLOGY

In this section, we use an uncertain process called the general geometric Liu process (Liu, 2015) to for ADT modeling, derive the reliability and lifetime distributions, and gives the corresponding statistical analysis method.

3.1 *Accelerated degradation modeling*

In practical applications, the degradation process is usually positive. To describe the positive degradation process under the small sample situation, we consider the following uncertain process

$$X(t)=\exp(e\cdot t+\sigma/\sqrt{t}\,C(t)), \tag{5}$$

where e is the log-drift parameter, also known as the degradation rate. $\sigma/\sqrt{t}$ is the log-diffusion parameter. $C(t)$ is the Liu process that follows a normal uncertainty distribution (N_u) with mean 0 and variance t^2, i.e. $C(t)\sim N_u(0,t)$. Eq.(5) is called the general geometric Liu process. Note that $X(t)$ in Eq.(5) follows a lognormal uncertainty distribution (Log_u), i.e., $X(t)\sim Log_u(et,\sigma\sqrt{t})$. Its uncertainty distribution can be expressed as

$$\Phi_t(x)=\left(1+\exp\left(\frac{\pi(e\cdot t-\ln x)}{\sqrt{3}\sigma\sqrt{t}}\right)\right)^{-1}. \tag{6}$$

In ADT modeling, acceleration models are usually utilized to describe the relationship between the degradation rate and the accelerated stress level, which can be expressed as

$$\ln e(s_i)=a+b\cdot s_i, \tag{7}$$

where a and b are unknown parameters. s_i is the normalized stress level that can be expressed as

$$s_i = \begin{cases} \dfrac{1/S_0 - 1/S_i}{1/S_0 - 1/S_H} & \text{Arrhenius model,} \\ \dfrac{\ln S_i - \ln S_0}{\ln S_H - \ln S_0} & \text{Power law model,} \\ \dfrac{S_i - S_0}{S_H - S_0} & \text{Exponential model.} \end{cases} \quad (8)$$

where S_0 is the normal stress level, S_i is the i^{th} accelerated stress level, and S_H is the highest accelerated stress level. From Eq.(8), it is easy to know that $s_0 = 0$, and $s_H = 1$.

For simplicity, our proposed uncertain accelerated degradation model in Eq.(5) and Eq.(7) is denoted by M_1. The unknown parameters in model M_1 are summarized as $\Omega = (a, b, \sigma)$, in which a and b are shown in Eq.(7), σ is shown in Eq.(5).

3.2 *First hitting time and reliability distributions of the proposed model*

After getting the proposed model M_1, we need to derive the reliability and lifetime distributions correspondingly.

Define ω as the failure threshold of the degradation process, then the lifetime T can be defined as the first hitting time (FHT) when the degradation process $X(t)$ reaches ω. Liu (2013) defined the FHT of the uncertain process as follows:

$$t_\omega = \inf\{t_\omega \geq 0 \mid X(t) = \omega\}. \quad (9)$$

According to Theorem 3 in (Liu, 2013), the FHT of an independent increment process with a continuous uncertainty distribution at each time can be expressed as follows,

$$\Upsilon(z) = \mathcal{M}\left\{\sup_{0\leq t\leq z} X(t) \geq \omega\right\} = 1 - \inf_{0\leq t\leq z} \Phi_t(\omega). \quad (10)$$

Therefore, before deriving the reliability and lifetime distributions, we firstly need to prove that $X(t)$ in Eq.(5) is an independent increment process with a continuous uncertainty distribution at each time.

Proof:

1. From Eq.(6), it is easy to know that $X(t)$ has a continuous uncertainty distribution at each time t.
2. For each $\gamma \in \Gamma$, the uncertainty distribution of $X(t \mid \gamma)$ can be expressed as

$$\Phi_t(\gamma) = \left(1 + \exp\left(\frac{\pi(e(s)\cdot t - \ln\gamma)}{\sqrt{3}\sigma\sqrt{t}}\right)\right)^{-1}. \quad (11)$$

Eq.(11) is obvious a continuous function with respect to time t.

3. From Eq.(6), we can get the inverse uncertainty distribution of $X(t)$ as follows,

$$\Phi_t^{-1}(\alpha) = \exp(e(s)\cdot t + \frac{\sigma\sqrt{3t}}{\pi}\ln\frac{\alpha}{1-\alpha}), \ \alpha \in (0,1). \quad (12)$$

At each time t, the derivative of Eq.(12) with respect to α is

$$\left(\Phi_t^{-1}(\alpha)\right)' = \frac{\sigma\sqrt{3t}}{\pi}\exp\left(e(s)t + \frac{\sigma\sqrt{3t}}{\pi}\ln\frac{\alpha}{1-\alpha}\right) \times \frac{1}{\alpha(1-\alpha)}. \quad (13)$$

Since $\alpha \in (0,1)$, we can get that $\alpha(1-\alpha) > 0$. It is easy to prove that $\Phi_t^{-1}(\alpha)$ is a continuous and strictly increasing function with respect to α.

4. For any times $0 < t_2 < t_1$, to prove $\Phi_{t_1}^{-1}(\alpha) - \Phi_{t_2}^{-1}(\alpha)$ is a monotone increasing function with respect to α, we need to prove the following condition:

$$\Phi_{t_1}^{-1}(\alpha) - \Phi_{t_2}^{-1}(\alpha) = \exp\left(e(s)\cdot t_1 + \frac{\sigma\sqrt{3t_1}}{\pi}\ln\frac{\alpha}{1-\alpha}\right) - \exp\left(e(s)\cdot t_2 + \frac{\sigma\sqrt{3t_2}}{\pi}\ln\frac{\alpha}{1-\alpha}\right) > 0. \quad (14)$$

Since $\exp(\bullet)$ is a monotone increasing function, Eq.(14) is equivalent to the following condition:

$$e(s)\cdot(t_1 - t_2) + \frac{\sigma\sqrt{3}}{\pi}\ln\frac{\alpha}{1-\alpha}\left(\sqrt{t_1} - \sqrt{t_2}\right) > 0. \quad (15)$$

According to the given information, it is easy to prove that the condition in Eq.(15) holds. Thus, $\Phi_{t_1}^{-1}(\alpha) - \Phi_{t_2}^{-1}(\alpha)$ is a monotone increasing function with respect to α. Based on **Theorem 1** in section 2, we can prove that $X(t)$ in Eq.(5) is an independent increment process.

From the above analyses, we can get that the uncertain process $X(t)$ in Eq.(5) is an independent increment process with a continuous uncertainty distribution at each time t. Thus, the uncertainty distribution of the FHT of $X(t)$ can be expressed as follows,

$$\begin{aligned} \Upsilon(z) &= 1 - \inf_{0\leq t\leq z}\Phi_t(\omega) \\ &= 1 - \inf_{0\leq t\leq z}\left(1 + \exp\left(\frac{\pi(e(s)\cdot t - \ln\omega)}{\sqrt{3t}\sigma}\right)\right)^{-1} \\ &= \left(1 + \exp\left(\frac{\pi(\ln\omega - e(s)\cdot z)}{\sqrt{3z}\sigma}\right)\right)^{-1}. \end{aligned} \quad (16)$$

The corresponding reliability distribution is

$$R_B(t) = \mathrm{M}\{t_\omega > t\} = \left(1 + \exp\left(\frac{\pi(e(s)\cdot t - \ln\omega)}{\sqrt{3}t\sigma}\right)\right)^{-1} \quad (17)$$

where $R_B(t)$ is known as the "belief reliability" with an uncertain measure (Zeng et al., 2013).

Meanwhile, the belief reliable life, i.e. $BL(\alpha)$, can also be derived and expressed as follows

$$BL(\alpha) = \sup_{0\le t\le z} R_B(t) \ge \alpha = \Upsilon^{-1}(1-\alpha). \quad (18)$$

3.3 *Statistical analysis method for parameter estimations*

With different loading profiles, there are different kinds of ADT plans, including the constant stress accelerated degradation testing (CSADT), the step stress accelerated degradation testing (SSADT), and the progressive stress accelerated degradation testing (PSADT). Here, we provide the statistical analysis method for parameter estimations in CSADT.

Liu (2015) employed the principle of least squares for parameter estimations of uncertain variables. In this section, we also use this method to estimate the unknown parameters of the proposed model M_1.

Suppose that x_{ijk} is the k^{th} observed degradation value for the j^{th} sample under the i^{th} stress level, and t_{ijk} is the corresponding measurement time, $i = 1, 2, \ldots, K$, $j = 1, 2, \ldots, n_i$, $k = 1, 2, \ldots, m_i$, where K is the number of accelerated stress levels, n_i is the sample size under the i^{th} stress level, and m_i is the number of measurements for the j^{th} sample under the i^{th} stress level.

The unknown parameters of the proposed model M_1 is $\boldsymbol{\Omega} = \mathrm{a}, \mathrm{b}, \sigma)$. We proposed a two-step statistical analysis method for the parameter estimations: 1) Collecting belief degrees; 2) Estimating unknown parameters. The procedure of the method is shown in Figure 1 and details are shown as follows:

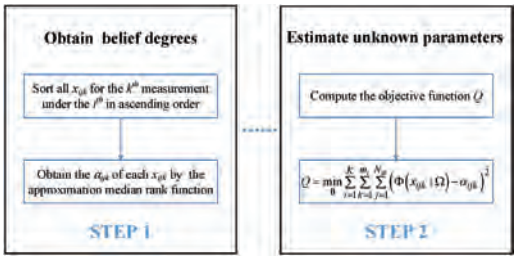


Figure 1. The two-step statistical analysis method for parameter estimations of the proposed model.

1. Obtain belief degrees

From Section 3.2, it is known that x_{ik} is an uncertain variable. All the degradation data x_{ijk} of the k^{th} measurement under the i^{th} stress level are the observations of x_{ik}, i.e., $x_{ik} = \{x_{i1k}, x_{i2k} \ldots x_{ijk} \ldots\}$, ($j = 1, 2 \ldots N_i$, and n_i is the upper boundary of N_i). Each of the element has a belief degree α_{ijk}. In this section, we use the approximate median rank functions to obtain belief degrees, which is expressed as follows,

$$\alpha_{ijk} = (j - 0.3)/(N_{ik} + 0.4),\ j = 1,2,\ldots,N_{ik}\ . \quad (19)$$

For all the degradation data of the k^{th} measurement under the i^{th} stress level, if there exist degradation data that are the same, then their belief degrees are also the same.

2. Estimate unknown parameters

According to the principle of least squares, the parameters estimations of the proposed model can be obtained by the principle of least squares, which is

$$Q = \min_{\boldsymbol{\theta}} \sum_{i=1}^{K}\sum_{k=1}^{m_i}\sum_{j=1}^{N_i} \left(\Phi\left(x_{ijk}\,|\,\Omega\right) - \alpha_{ijk}\right)^2 \quad (20)$$

4 CASE STUDY

In this section, the carbon-film resistors CSADT dataset (Meeker and Escobar, 1998) is used to illustrate the proposed methodology, and discussions are conducted for the sensitivity analysis of the proposed methodology to the sample sizes.

4.1 *The carbon-film resistors CSADT dataset*

In the carbon-film resistors CSADT dataset, there are 9, 10, 10 samples under each accelerated stress levels, which belongs to the small sample situation. Therefore, the proposed methodology can be used to this case for the reliability and lifetime evaluations under normal conditions. Details about this case is shown in Table 1.

Table 1. Basic information about the carbon-film resistors CSADT dataset.

Test information	Contents
Stress levels (temperature/°C)	83, 133, and 173
Normal conditions (°C)	50
Sample size	9, 10, and 10
Measurement times	4, 4, and 4
Failure threshold ω (%)	12

4.2 *Reliability and lifetime evaluations under normal conditions*

Since the accelerated stress is temperature, the Arrhenius model is selected as the acceleration model. Based on the proposed methodology, the parameter estimations are obtained as follows:

Taking the parameter estimation results in Table 2 into Eq.(17) and Eq.(18), the reliability and lifetime evaluations under normal conditions can be obtained. Results are showed in Figure 2.

From Figure 2 (a), it can be seen that the belief reliability changes from the initial value 1 and decreases gradually with the increasing time, which agrees with the intuitive cognition of human beings. If decision makers are interested at belief reliability $R_B = 0.9$, the corresponding belief reliable lifetime $BL(0.9) = 10181$ hours. It means that the belief degree that the products will survived at the normal conditions after 10818 hours is 0.9.

4.3 *Discussions*

For the sensitivity analysis of the proposed methodology to the sample sizes, we simulate several different situations that has different sample size, and remark each situation as ST_r (n_{r1}, n_{r2}, n_{r3}), $r = 1, 2,\ldots, 8$. ST_r represents the r^{th} situation. n_{r1}, n_{r2}, and n_{r3} represents the chosen sample size under each stress level. Details are shown in Table 3.

Table 2. Parameter estimation results.

Parameters	a	b	σ
Values	–15.18	3.975	9.060e-04

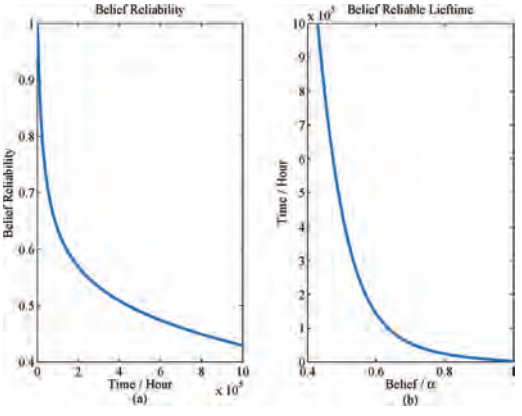


Figure 2. Reliability and lifetime evaluations of the carbon-film resistors ADT dataset under normal conditions.

As shown in Table 3, under each situation, there are many different combinations of samples, which will lead to many different reliability evaluations. To present the range of reliability evaluations under different situations, under each monitoring time t under the situation ST_r, we choose the minimum and maximum reliability evaluation results as the lower and upper boundaries of the reliability evaluations. So the lower and upper boundaries under each situation can be obtained, and results are shown in Figure 3.

Figure 3 show that with the increasing sample size, the lower and upper boundary are approaching gradually. It indicate that when there are more samples that can provide more information, the epistemic uncertainty in ADT data decreases, which will lead to more stable reliability evaluation results. In addition, the reliability evaluations

Table 3. Different situations for discussions.

Situations | Sample sizes	n_{r1}	n_{r2}	n_{r3}
ST_1	2	3	3
ST_2	3	4	4
ST_3	4	5	5
ST_4	5	6	6
ST_5	6	7	7
ST_6	7	8	8
ST_7	8	9	9
ST_8	9	10	10

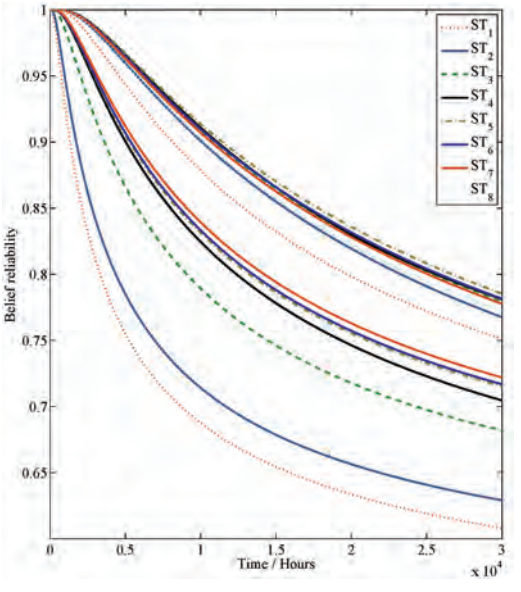


Figure 3. Lower and upper boundaries of the reliability evaluation results under different sample sizes.

results under ST_8 is included in the lower and upper boundaries under most situations (ST_3 to ST_7). As for ST_1 and ST_2, the sample size is very small, which makes the provided information too scarce to get stable reliability evaluation results.

The above analysis results show that the proposed methodology is a suitable choice for the small sample situation and can furtherly provide support for the subsequent decision making.

5 CONCLUSIONS

This paper deals with the positive degradation process with small samples in ADT, and concludes as follows,

1. Based on the uncertainty theory, the general geometric Liu process is used to conduct an uncertain accelerated degradation model, which takes the epistemic uncertainty due to small samples in ADT data into consideration.
2. The corresponding statistical analysis method with objective measures is provided for the unknown parameter estimations.
3. The application results show that the reliability evaluation results of the proposed methodology agrees with agrees with the intuitive cognition of human beings, and the discussion results show that the proposed methodology provides stable reliability evaluation results under small samples, which makes it a suitable choice for the small sample situation and can provide support for the subsequent decision making.

In addition to the work of this paper, there are other issues that are worthwhile for future researches. The proposed model is built up on the general geometric Liu process, which is a positive uncertain process. But in practical applications, there are degradation processes which are not only always positive but also strictly monotonic. It is necessary to apply other uncertain processes in ADT to model such degradation processes.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China [grant numbers 51775020, 61573043, 71671009 and 61104182].

REFERENCES

Beer, M., Ferson, S. & Kreinovich, V. 2013. Imprecise probabilities in engineering analyses. *Mechanical Systems & Signal Processing,* 37, 4–29.

Huang, M., Ren, L., Lee, L.H., Wang, X., Kuang, H. & Shi, H. 2016. Model and algorithm for 4PLRP with uncertain delivery time. *Information Sciences,* 330, 211–225.

Kang, R., Zhang, Q., Zeng, Z., Zio, E. & Li, X. 2016. Measuring reliability under epistemic uncertainty: Review on non-probabilistic reliability metrics. *Chinese Journal of Aeronautics,* 29, 571–579.

Li, M. & Meeker, W.Q. 2014. Application of Bayesian methods in reliability data analyses. *Journal of Quality Technology,* 46, 1.

Liu, B. 2008. Fuzzy process, hybrid process and uncertain process. *Journal of Uncertain systems,* 2, 3–16.

Liu, B. 2009. Some research problems in uncertainty theory. *Journal of Uncertain Systems,* 3, 3–10.

Liu, B. 2010. Uncertain risk analysis and uncertain reliability analysis. *Journal of Uncertain Systems,* 4, 163–170.

Liu, B. 2012. Why is there a need for uncertainty theory. *Journal of Uncertain Systems,* 6, 3–10.

Liu, B. 2013. Extreme value theorems of uncertain process with application to insurance risk model. *Soft Computing,* 17, 549–556.

Liu, B. 2014. Uncertainty distribution and independence of uncertain processes. *Fuzzy Optimization and Decision Making,* 13, 259–271.

Liu, B. 2015. *Uncertainty theory*, Springer Berlin Heidelberg.

Meeker, W.Q. & Escobar, L.A. 1998. *Statistical methods for reliability data*, John Wiley & Sons.

Meeker, w.Q., Escobar, l.A. & Lu, c.J. 1998. Accelerated degradation tests: modeling and analysis. *Technometrics,* 40, 89–99.

Moore, R.E., Kearfott, R.B. & Cloud, M.J. 2009. *Introduction to interval analysis*, SIAM.

Ye, Z.S. & Xie, M. 2015. Stochastic modelling and analysis of degradation for highly reliable products. *Applied Stochastic Models in Business and Industry,* 31, 16–32.

Zeng, Z., Wen, M. & Kang, R. 2013. Belief reliability: a new metrics for products' reliability. *Fuzzy Optimization and Decision Making,* 12, 15–27.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Reliability assessment for solid state drive based on measurement errors and fuzzy failure threshold

Peng Li
Technology and Engineering Centre for Space Utilization, Chinese Academy of Science, Beijing, China
University of Chinese Academy of Sciences, Beijing, China

Jiaxin Yuan
Beihang University, Beijing, China

Wei Dang
Technology and Engineering Centre for Space Utilization, Chinese Academy of Science, Beijing, China

ABSTRACT: For solid state drive (SSD) with high reliability and long life, it is nearly impossible to obtain sufficient amount of time-to-failure data within acceptable testing time by testing such products under normal operating environments. Thus, accelerated degradation testing (ADT) are introduced to solve reliability modeling problems based on the products' degradation information obtained from accelerated tests. The intimate link between performance degradation data and product failures can be obtained according to the degradation threshold failure mechanism. Through the hypothetical degradation process and failure time, we can estimate the failure time with a given threshold value. However, due to diverse users, and uncertainty of what explicit level of degradation will cause a failure, a probabilistic, rather than a deterministic threshold value should be taken into account. On the other hand, complex operation environment would result in inevitable noise, so it is necessary to consider the detecting error in the reliability analysis. This paper propose a reliability assessment method based on fuzzy failure threshold and measurement errors, aiming at improving the assessment precision. We establish the degradation modeling with fuzzy failure threshold and measurement errors, and the maximum likelihood estimation method is adopted to estimate the failure time distribution parameters. Then the reliability model can be used for subsequent forecasting and decision-making. A commercial off-the-shelf SSD is shown as an example to illustrate the procedure that how to predict time to failure, of which writing current is used as a precursor parameter and directly monitored. Finally the results show the superior performance of the proposed method over traditional methods.

1 INTRODUCTION

Solid state drives (SSDs) are the most important and widely-used data storage device in space application due to their high performance, low energy consumption, small size, and shock resistance compared with traditional hard disk drives (HDDs). As the amount of data stored in SSDs keeps increasing, it is important to understand the reliability characteristics of SSD under field conditions. However, for SSD with high reliability and long life, it is nearly impossible to obtain sufficient amount of time-to-failure data within acceptable testing time by testing such products under normal operating environments and sometimes even under harsher conditions (Nelson 2008).

The accelerated degradation testing (ADT) with higher stress levels provides a method to carry out the life test within a reasonable time frame based on the products' degradation information obtained from historical data or degradation tests (YaoHsu, Chih-YenSu et al. 2014). Generally, a failure occurs when the performance degradation data exceeds a specified threshold. The reliability can be predicted by extrapolating the degradation trend to its threshold after tracking the degradation path for a while (Ye and Xie 2015). Thus, numerous failure and reliability information can be obtained in the accelerated degradation process.

Great attention have been caught in this field. Ren proposed an original approach combining ADT and physics of failure (PoF) modeling for effectively assessing connector reliability (Ren, Feng et al. 2015). González presented a complete reliability analysis including the determination of the reliability functions and parameters obtained for concentrator mutijunction solar cells (Espinet-González, Algora et al. 2015). Chang proposed a reliability qualification test method for pneumatic cylinders based on performance degradation data (Chang, Kwon et al. 2014).

The actual failure point is usually unpredictable or random in the practical engineering application because of the lack of enough empirical information and experimental data, leading to the ambiguity in failure threshold definition. Therefore, it's significant to consider the potential uncertainties in the reliability analysis.

Chen presented the reliability analysis for system with random failure threshold based on the degradation model and failure threshold probability distribution (Chen, Ma et al. 2014). Hua proposed a novel performance degradation reliability based on an adaptive failure threshold (Hua, Zhang et al. 2013).

In this paper, a reliability assessment method based on measurement errors and fuzzy failure threshold is proposed. First, the life-stress model and lifetime distribution model are established. By combining these two models, we adopt the maximum likelihood estimation method to estimate the failure time distribution parameters, and then the reliability based on fixed failure threshold can be estimated. Subsequently, we present a reliability assessment method based on fuzzy failure threshold by applying the probability formula of the fuzzy event and the convolution formula. Then the reliability assessment method is validated by a type of commercial off-the-shelf SSD to be used in the Chinese space station. In the end, we take a comparative analysis on the reliability assessment curves with fixed failure threshold, fuzzy threshold without measurement errors, and fuzzy failure threshold and measurement errors. The result shows that the proposed method has a higher assessment accuracy, and it is more flexible.

2 RELIABILITY MODELING AND ANALYSIS

2.1 *Drift degradation data analysis and model modification*

For electronic products in ADT, performance degradation is accompanied with performance drift created by accelerated stress (Li et al. 2017). The degradation path model can be described by:

$$m_i(t) = x_i(t) + \Delta x_i + \varepsilon_i \tag{1}$$

where $m_i(t)$ is the measured value; $x_i(t)$ is the true degradation value; Δx is the performance drift; ε_i the measurement error which is independent of $x(t)$, and $\varepsilon_i \sim N(0, \sigma^2)$.

2.2 *Lifetime distribution and parameter estimation with degradation data*

2.2.1 *Life-stress model*

As temperature is the acceleration variable, the well-known Arrhenius model is selected:

$$\xi(T) = Ae^{-\frac{Ea}{kT}} \tag{2}$$

where $\xi(T)$ is the performance characteristic related to life; A is a constant; E_a is the activation energy (eV); k is Boltzmann constant (8.617×10^{-5} eV/K); and T is the absolute temperature.

2.2.2 *Lifetime distribution model*

In many standard statistical distributions used to model the various reliability parameters, Weibull distribution is the most common lifetime distribution (Bertsche 2010). Because it applies well to various failure modes by adjustment of the distribution parameters. In this paper, a two-parameter Weibull distribution is selected for the degradation life distribution, whose cumulative distribution function (cdf) has the expression:

$$F(x;t,T) = 1 - e^{-\left(\frac{x}{\eta}\right)^{\beta}} \tag{3}$$

where x is the true degradation, t is the measurement time, T is the accelerated temperature stress level, β and η are shape parameter and scale parameter, respectively.

Under assumption of constant temperature, (2) can be reduced to the following power formal function related with test time and stress level (Ren, Feng et al. 2015):

$$\eta(t,T) = a \cdot e^{\frac{b}{T}} \cdot t^c \tag{4}$$

where a, b, c are the parameters to be estimated.

The scale parameter of the Weibull model (η) depends on the temperature in the Arrhenius model, and the shape parameter (β) has been assumed constant for different temperatures. Combining the life-stress model and lifetime distribution model, we can obtain the combined Arrhenius-Weibull model:

$$F(x;t,T) = 1 - e^{-\left(\frac{x}{a \cdot e^{b/T} \cdot t^c}\right)^{\beta}} \tag{5}$$

where β, a, b, c are the parameters to be estimated.

The failure probability density function (pdf) can be expressed as follows:

$$f(x;t,T) = \frac{\beta}{\left(a \cdot e^{b/T} \cdot t^c\right)^{\beta}} \cdot x^{\beta-1} \cdot e^{-\left(\frac{x}{a \cdot e^{b/T} \cdot t^c}\right)^{\beta}} \tag{6}$$

2.2.3 *Parameters estimation*

In order to estimate the best-suited parameters of the Arrhenius-Weibull model (5), the maximum

likelihood estimation method (MLE) can be utilized. The likelihood function is (Ren, Feng et al. 2015):

$$L(\beta,a,b,c)=c\cdot\prod_{k=1}^{n}\prod_{i=1}^{q}\prod_{j=1}^{m}\frac{\beta}{\left(a\cdot e^{\frac{b}{T}}\cdot t^{c}\right)^{\beta}}\cdot x^{\beta-1}\cdot e^{-\left(\frac{x}{a\cdot e^{\frac{b}{T}}\cdot t^{c}}\right)^{\beta}} \tag{7}$$

Substituting the performance degradation data into (7) and solving the MLE equation $\partial L/\partial\beta=0$, $\partial L/\partial a=0$, $\partial L/\partial b=0$, $\partial L/\partial c=0$, we can get the estimates $\hat{\beta}\ \hat{a}\ \hat{b}\ \hat{c}$. The estimated true degradation distribution is expressed as:

$$F(x;t,T)=1-e^{-\left(\frac{x}{\hat{a}\cdot e^{\frac{\hat{b}}{T}}\cdot t^{\hat{c}}}\right)^{\hat{\beta}}} \tag{8}$$

The product will not work properly when the degradation value $x(t)$ reaches the threshold level D. For a fixed failure threshold, the reliability $R(t)$ is the probability that the true degradation value $x(t)$ is larger than the fixed threshold level D. Therefore, the reliability function based on fixed failure threshold is represented by the cumulative distribution function of the true degradation value distribution:

$$R(t,T)=P(x(t)\le D)=1-e^{-\left(\frac{D}{\hat{a}\cdot e^{\frac{\hat{b}}{T}}\cdot t^{\hat{c}}}\right)^{\hat{\beta}}} \tag{9}$$

2.3 *Reliability assessment method based on fuzzy failure threshold*

The membership function is used to describe the membership between an element u and a fuzzy set on the domain U. It is assumed that the performance degradation data has an increasing trend over time in this paper. In the beginning, the membership function is equal to 1. When the performance degradation data $x(t)$ increases to the lower limit of fuzzy failure threshold ($D_{\min}$), the membership function is $\mu_A(x)$, and the performance of the product continues to degenerate gradually; When the performance degradation data $x(t)$ is larger than the upper limit of fuzzy failure threshold ($D_{\max}$), the membership function is equal to 0 (Yan 2014).

Thus, the lower semi-trapezoid distribution membership function is:

$$\mu_A(x)=\begin{cases}1 & x<D_{\min}\\ \dfrac{D_{\max}-x}{D_{\max}-D_{\min}} & D_{\min}\le x\le D_{\max}\\ 0 & x>D_{\max}\end{cases} \tag{10}$$

2.4 *Establishment and assessment of reliability function based on measurement errors and fuzzy failure threshold*

Consider the measurement uncertainty of sensors, we assume the measurement errors ε_i obey a normal distribution ($\varepsilon_i \sim N(0,\sigma^2)$). So its probability density function is

$$f_\varepsilon(\varepsilon)=\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(\varepsilon-\mu)^2}{2\sigma^2}} \tag{11}$$

A fuzzy subset showing the good state of the product is expressed as follows:

$$A\overset{\Delta}{=}\{D\ge X(t)\}=\{D\ge Y(t)-E\} \tag{12}$$

Since the true degradation value $x(t)$ has a different distribution form with the measurement errors ε_i, the reliability function based on measurement errors and fuzzy failure threshold can be obtained by applying the probability formula of the fuzzy event and the convolution formula:

$$\begin{aligned}R_A(t)&=P(A)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mu_A(t)f(y-\varepsilon,t)f_\varepsilon(\varepsilon)dyd\varepsilon\\&=\int_{-\infty}^{\infty}\int_{0}^{D_{\min}}f(y-\varepsilon,t)f_\varepsilon(\varepsilon)dyd\varepsilon+\int_{-\infty}^{\infty}\int_{D_{\min}}^{D_{\max}}\mu_A(t)f(y-\varepsilon,t)f_\varepsilon(\varepsilon)dyd\varepsilon\\&=\int_{-\infty}^{\infty}\int_{0}^{D_{\min}}\frac{\beta}{\left(a\cdot e^{\frac{b}{T}}\cdot t^{c}\right)^{\beta}}\cdot(y-\varepsilon)^{\beta-1}\cdot e^{-\left(\frac{y-\varepsilon}{a\cdot e^{\frac{b}{T}}\cdot t^{c}}\right)^{\beta}}\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(\varepsilon-\mu)^2}{2\sigma^2}}dyd\varepsilon\\&+\int_{-\infty}^{\infty}\int_{D_{\min}}^{D_{\max}}\frac{D_{\max}-y}{D_{\max}-D_{\min}}\frac{\beta}{\left(a\cdot e^{\frac{b}{T}}\cdot t^{c}\right)^{\beta}}\cdot(y-\varepsilon)^{\beta-1}\cdot e^{-\left(\frac{y-\varepsilon}{a\cdot e^{\frac{b}{T}}\cdot t^{c}}\right)^{\beta}}\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(\varepsilon-\mu)^2}{2\sigma^2}}dyd\varepsilon\end{aligned} \tag{13}$$

Substituting the estimators $\hat{\beta}$ $\hat{a}$ $\hat{b}$ $\hat{c}$ in (13), then we acquire the reliability assessment result. If we ignore the measurement errors, the function reduce to be:

$$R_A(t) = P(A) = \int_{-\infty}^{\infty} \mu_A(t) f(\mathrm{x},t)\,dx$$
$$= \int_0^{D_{\min}} f(\mathrm{x},t)\,dx + \int_{D_{\min}}^{D_{\max}} \mu_A(t) f(\mathrm{x},t)\,dx$$
$$= \int_0^{D_{\min}} \frac{\beta}{\left(a \cdot e^{b/T} \cdot t^c\right)^{\beta}} \cdot y^{\beta-1} \cdot e^{-\left(\frac{y}{a \cdot e^{b/T} \cdot t^c}\right)^{\beta}} dy$$
$$+ \int_{D_{\min}}^{D_{\max}} \frac{D_{\max} - y}{D_{\max} - D_{\min}} \frac{\beta}{\left(a \cdot e^{b/T} \cdot t^c\right)^{\beta}} \cdot y^{\beta-1} \cdot e^{-\left(\frac{y}{a \cdot e^{b/T} \cdot t^c}\right)^{\beta}} dy \tag{14}$$

3 EXAMPLE OF SSD APPLICATION

3.1 *SSD test plan*

SSD consists of main control unit, power supply, connector, cache, and NAND flash. Non-volatile NAND flash memory is the core component. According to Fowler-Nordheim (FN) tunneling, electrons in memory cells write in or read out data in the NAND flash memory by passing through tunnel oxide repeatedly. This process reduces the reliability of memory cells, leading to the gradual degradation of the tunnel oxide.

As the environment temperature rises, the thermal motion energy of electrons increases, accelerating the destruction of tunnel oxide in tunneling. Therefore, temperature is the sensitive stress that causes SSD's performance degradation(Li, P., et al. (2017).

According to the reliability enhancement testing, the tested SSD's temperature design limit is proved to be –40 ~ 85°C, and operating limit is proved to be –80 ~ 125°C. Thus, a step-stress temperature preliminary test has been taken firstly by testing all the 9 samples from 25°C to 125°C with the same pre-fixed stress-change time. The degradation effect can be neglected because test duration is very short.

Results of the step-stress temperature preliminary test show that the majority of performance indexes nearly remain unchanged except the read/write current and quiescent current, which rise with the increasing of temperature. Therefore, the relationship between current drift Δx and temperature T can be modeled by regression analysis:

$$\Delta x = f(T) \tag{15}$$

Nine SSD samples are randomly divided into three groups. ADT is carried out in each group for each accelerated stress levels with 80°C, 90°C and 104°C. The test duration is set to 40 days and measurements are conducted every 96 hours. During the test process, performance characteristics (read/write current, quiescent current, read/write speed, read/write response time and bad block increment) are real-time monitored. Samples continue working all along at the same time. Results of the test indicate that write current is the most suitable index to track SSD degradation.

3.2 *SSD test results and reliability analysis*

After the determination of the relationship between write current drift and temperature, the degradation path model can eliminate the drift impact.

Assume the write current at 25°C to be the initial value, and apply the least square method to fit the average measured drift data. Then the fitting curve is illustrated in Figure.1.

The fitting model is expressed as:

$$\Delta x = a \cdot \exp(b \cdot \Delta T) \tag{16}$$

where Δx is the current drift; and ΔT is the temperature rise over initial temperature T_0 = 25°C. The parameters are estimated to be $\hat{a} = 2.088$, $\hat{b} = 0.03446$.

Samples tested under 80°C are named N-0, N-1, N-2, and N-3, N-4, N-5 under 90°C, and N-6, N-7, N-8 under 104°C. After eliminating current drift impact, the true degradation path of write current can be calculated under 80°C, 90°C and 104°C, which are shown in Figure. 2.

Figure. 3 shows the variation over time of the write current degradation distribution in terms of Weibull probability plots, for different levels of temperatures. The data is proved to follow the Weibull distribution as all the scatter points are close to the reference line.

Substituting the test observations by MLE with the Arrhenius-Weibull model presented above, the estimates $\hat{\beta}$ and $\hat{\eta}$ can be obtained. Further, with

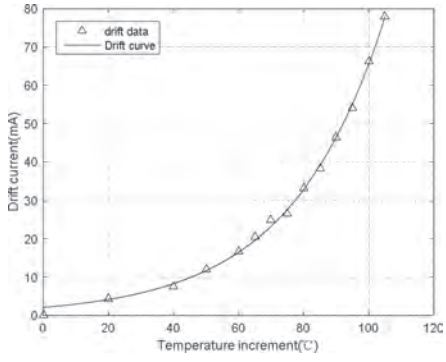


Figure 1. Regression analysis of current drift & temperature increment.

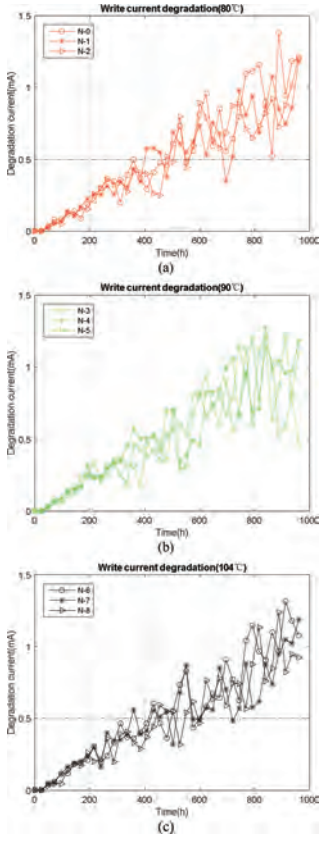


Figure 2. True degradation path of write current. (a) 80°C. (b) 90°C. (c)104°C

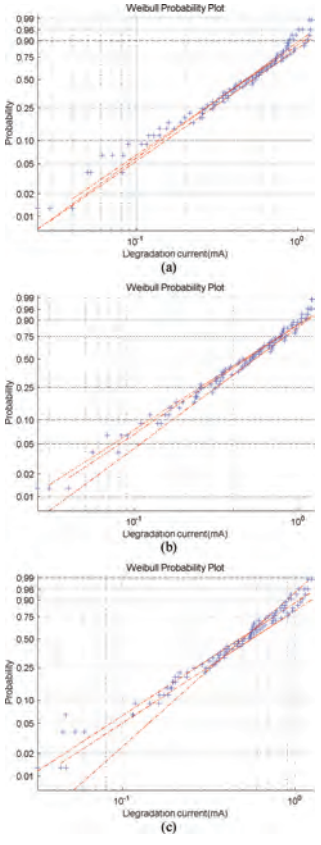


Figure 3. True degradation path of write current at different measuring time and temperature levels. (a) 80°C (N-0, N-1, N-2). (b) 90°C (N-3, N-4, N-5). (c) 104°C (N-6, N-7, N-8)

the estimates of $\hat{\eta}$ at different time points and stress levels, the parameters a, b, c can be estimated and the write current degradation distribution model is specified as:

$$F(x;t,T)=1-e^{-\left(\frac{x}{0.005e^{-350/T}\cdot t^{0.93}}\right)^5} \tag{17}$$

3.3 *Reliability extrapolation at normal working conditions*

Afterwards the Arrhenius-Weibull model is used to extrapolate the performance of SSD at normal temperature. According to the NAND flash memory manual, the failure threshold at normal temperature (25°C) is 125 mA (Isobe 2010). Then, the reliability function based on fixed failure threshold for SSD is:

$$R(t,T)=P(x(t)\le D)=1-e^{-\left(\frac{125}{0.005e^{-350/T}\cdot t^{0.93}}\right)^5} \tag{18}$$

Assuming the lower limit of fuzzy failure threshold ($D_{\min}$) to be 100 mA, and the upper limit of fuzzy failure threshold ($D_{\max}$) to be 150 mA. Besides, the variance of measurement error distribution is 0.8, namely $\varepsilon \sim N(0,0.8)$. As the integration in (13) and (14) have no explicit expressions, numerical integration method are used to solve the problems. Thus, the reliability curves based on fixed failure threshold, fuzzy failure threshold without measurement errors, and fuzzy threshold with measurement errors under normal temperature (25°C) are illustrated in Figure 4 respectively.

It can be observed that the three reliability curves are close to each other, but if we consider fuzzy threshold situation, the reliability is slightly smaller than the deterministic threshold. Moreover, the reliability with measurement errors and fuzzy threshold is the lowest. Thus, when dealing with SSD's reliability prediction, it's necessary to consider the influences of measurement errors and fuzzy threshold, which can result in a more conservative and practical value.

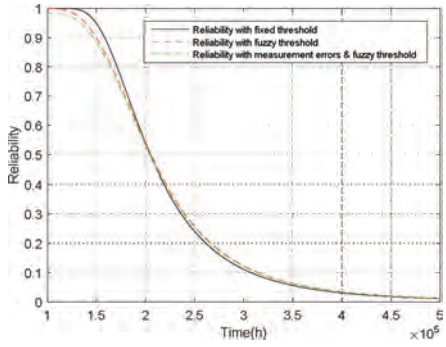


Figure 4. Reliability curve based on different assessment methods.

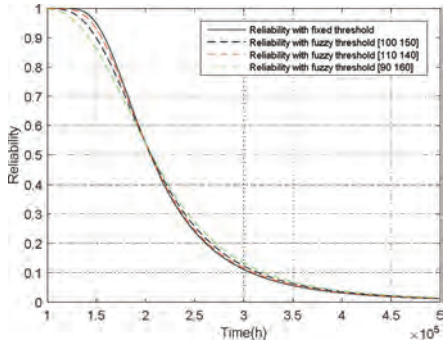


Figure 5. Reliability curve under different failure thresholds.

For further analysis, Figure 5 represents the reliability curves under different failure thresholds by changing the range of failure threshold.

It can be found that the smaller the fuzzy threshold range is, the closer the reliability assessment curve is to the reliability curve based on fixed failure threshold. The reliability comparison results show that the reliability assessment method based on measurement errors and fuzzy failure threshold is more accurate and flexible.

4 SUMMARY AND CONCLUSIONS

In this paper, a reliability assessment method based on fuzzy failure threshold and measurement errors is proposed. Firstly, we establish the life-stress model and lifetime distribution model, and then model parameters are estimated by the maximum likelihood estimation method. Secondly, we derive the reliability model based on fixed failure threshold by substituting the estimators into the combined Arrhenius-Weibull model. Thirdly, we present a reliability assessment method based on fuzzy failure threshold and measurement errors by applying the probability formula of the fuzzy event and the convolution formula. Subsequently, we take a type of SSD as an example to demonstrate the reliability assessment method. In the end, we compare the reliability assessment curve based on different situations. The comparison results show that the proposed method has a higher assessment accuracy, and it is more flexible.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (Grant No. 61703391) and Technology and Engineering Center for Space Utilization (Grant No. CSU-QZKT201714).

REFERENCES

Bertsche, B. (2010). "Reliability in Automotive and Mechanical Engineering." *VDI-Buch*.

Chang, M.S., et al. (2014). "Design of reliability qualification test for pneumatic cylinders based on performance degradation data." *Journal of Mechanical Science & Technology* **28**(12): 4939–4945.

Chen, J., et al. (2014). Reliability Analysis for System with Random Failure Threshold, Springer Berlin Heidelberg.

Espinet-González, P., et al. (2015). "Temperature accelerated life test on commercial concentrator III–V triple-junction solar cells and reliability analysis as a function of the operating temperature." *Progress in Photovoltaics Research & Applications* **23**(5): 559–569.

Hua, C., et al. (2013). "Performance reliability estimation method based on adaptive failure threshold." *Mechanical Systems & Signal Processing* **36**(2): 505–519.

Isobe, K. (2010). NAND flash memory, US.

Li, P., et al. (2017). "Statistical Analysis of Step-stress Accelerated Degradation Testing based on New and Used Samples." *Reliability & Maintainability Symposium, 2017*.

Li, P., et al. (2017). " Reliability Assessment of NAND SSD Based on Acceleration Degradation Test." *2017 IEEE International Conference on Industrial Engineering and Engineering Management*.

Nelson, W. (2008). Accelerated Testing: Statistical Models, Test Plans, and Data Analysis.

Ren, Y., et al. (2015). "A Novel Model of Reliability Assessment for Circular Electrical Connectors." *IEEE Transactions on Components Packaging & Manufacturing Technology* **5**(6): 755–761.

Yan,W.A. (2014). "Research on the method of strorage reliability for torpedo." *Northwestern Polytechnical University*, China.

YaoHsu, et al. (2014). "An analytical procedure for estimating field lifetime and failure rate of electronic packages." *Journal of the Chinese Institute of Engineers* **37**(1): 36–43.

Ye, Z.S. and M. Xie (2015). "Stochastic modelling and analysis of degradation for highly reliable products." *Applied Stochastic Models in Business & Industry* **31**(1): 16–32.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Effect of load-generation variability on power grid cascading failures

R. Rocchetta & E. Patelli
Institute for Risk and Uncertainty, Liverpool University, Liverpool, UK

L. Bing & G. Sansavini
Reliability and Risk Engineering Laboratory, Department of Mechanical and Process Engineering Institute of Energy Technology, ETH Zurich, Zurich, Switzerland

ABSTRACT: Cascading failures events are major concerns for future power grids and are generally not treatable analytically. For realistic analysis of the cascading sequence, dedicated models for the numerical simulation are often required. These are generally computationally costly and involve many parameters and variables. Due to uncertainty associated with the cascading failures and limited or unavailable historical data on large size cascading events, several factors turn out to be poorly estimated or subjectively defined. In order to improve confidence in the model, sensitivity analysis is applied to reveal which among the uncertain factors have the highest influence on a realistic DC overload cascading model. The 95*th* percentile of the demand not served, the estimated mean number of line failures and the frequency of line failure are the considered outputs. Those are obtained by evaluating random contingency and load scenarios for the network. The approach allows to reduce the dimensionality of the model input space and to identifying inputs interactions which are affecting the most statistical indicators of the demand not supplied.

1 INTRODUCTION

Assure high-reliability of electric power supply is a major concern for next-generation power grid. Power grid should have the ability to withstand know threats, such as N-1 and N-2 contingencies, but also poorly understood low-probability-high-consequence events such as N-k contingencies leading to cascading sequences. Due to the inherent complexity of cascading failure events, associated mathematical models are, generally, analytically not solvable. This is mainly due to the high dimensionality of the problem and to the complex, non-linear and dynamic behaviour characterizing domino failures.

Computational models for the simulation of the cascading sequences are used to provide a solution to the cascade problem. A wide variety of models have been proposed in the past, aiming at analysing different system behaviours and with several different objectives. For instance, models employing the AC power flow (PF) equations, such as the Manchester model (Nedic et al. 2006) or the linearized AC PF model (Li et al. 2016), the ORNL-PSerc-Alaska (OPA) model (Dobson, Carreras, Lynch, & Newman 2001) and DC PF-based models have been developed to simulate realistically cascading failures sequences.

Numerical models for cascading simulation have to be adequately designed, calibrated and validated (Bialek et al. 2016). Calibration and validation should use available historical cascading data, which is (in particular for large size cascade events) quite limited (Rocchetta et al. 2018) or affected by imprecision (Rocchetta et al. 2018). Consequently, the resulting model verification and calibration is very challenging and affected by high level of uncertainty. Uncertainty will result particularly prominent when the model is used to simulate rare events leading to very severe consequences.

To increase confidence in the cascading model results and better understand the relation between its inputs and outputs, all the relevant sources of uncertainty affecting the analysis should be quantified. Dimensionality and complexity issues are often involved in cascades analysis problems and the numerical simulators generally reflect these problems. In fact, the simulators often are time costly and involve a large number of uncertain variable and parameters.

Sensitivity analysis methods are useful to deal with both dimensionality and uncertainty issues. These methods can be used to reveal which sources of uncertainty are affecting the most the model output and can be used to reduce the dimensionality of the aleatory space by prioritizing only the most important factors. This is indeed a useful information, necessary to better comprehend inputs-outputs relations otherwise hidden within the complexity of the model.

Global sensitivity analysis methods are often employed by uncertainty analysts to sharpen the

view of the problem. Sensitivity analysis is sometimes regarded as a fundamental part of works that involves the assessment and propagation of uncertainty (Borgonovo and Plischke 2016). Applying global sensitivity analysis methods, insights can be gained regarding the input-output mapping and the key drivers of uncertainty can be clearly revealed (Borgonovo and Plischke 2016).

In this paper, an integrated framework for sensitivity analysis and power grids cascading analysis is proposed. The framework can be used to identify and prioritize the most relevant uncertain input factors by revealing their effect on different cascading failures indicators. Both system-level indicators, describing the overall impact of cascading failures, and component-level indicator, focusing on a single component performance, are considered. One of the aims of this work is to provide some guidance for the application of given data sensitivity analysis and screening methods to engineering practitioners, promoting their potential.

The framework is tested on a modified version of the RTS96 IEEE system. Two uncertainty cases are analysed, first accounting for only the uncertainty in the load demand. Then, a more complex and realistic case has been considered by accounting for randomness in the generators costs, thus inflating the dimensionality of the input space, i.e. more flexibility for the generators outputs. The analysis allows to point out which among loads and generator costs uncertainties is affecting the most the outputs of cascading failures model and for a modest computational effort.

The rest of the paper is organized as follows: Section 2 introduces global sensitivity analysis and screening methods. In Section 3 the algorithm for cascading failure simulation and the performance indicators are introduced. A benchmark case study, the RTS96 system, tests the framework in Section 4, 2 uncertainty cases are analysed. Section 5 closes the paper with a discussion on the results and conclusions.

2 SENSITIVITY ANALYSIS AND SCREENING

This section proposes a concise introduction to uncertainty quantification and methods for global sensitivity analysis. Traditionally, uncertainty quantification and analysis consist in the assignment of probability distributions to the model input factors (variables and parameters). Once the uncertainty has been characterized, it is propagated into the simulation code via Monte Carlo method. First, uncertain factors are characterised by assigning probability distributions. This is an important step which has to be performed adequately to assure high quality and consistency of results (Patelli, Pradlwarter, & Schuller 2010). Then, samples are obtained from the joint probability distribution of the input factors, e.g. by Latin hypercube sampling, quasi-random sequences or crude Monte Carlo inverse transform sampling (Patelli, Broggi, Angelis, & Beer 2014). Once the i^{th} input realisation is obtained $\boldsymbol{X}_i = \left[X_i(0),..,X_i(m)\right]$, the sample is forwarded to the computational model $M(X)$. This allows obtaining information about the input-output mapping defined by the computational model as follows:

$$M : \mathcal{X} \to \mathcal{Y},\ \boldsymbol{X} \to Y = M(\boldsymbol{X}) \tag{1}$$

where Y is the model output, for simplicity assumed 1-dimensional and without loss of generality.

Global sensitivity analysis methods have been developed to identify the most and the least relevant factors and gain additional insight on the input-output mapping defined in equation 1. Several global methods have been developed in the last decades. Screening methods, such as the one-at-a-time design of Morris (Morris 1991), variance-based methods, density-based methods (Borgonovo & Plischke 2016) are some of the most intensively applied methods.

2.1 *Given data Sobol's indices*

A variance-based statistic, commonly referred as the first order sensitivity coefficient, quantifies the (additive) effect of each input factor on the model output as follows (M. Sobol 1990):

$$S_i = \frac{V_{X_i}\left[\mathbb{E}_{\mathbf{X}_{\sim i}}[Y \mid X_i]\right]}{V[Y]} \tag{2}$$

where $V[Y]$ is the total variance of the output Y, X_i is the i^{th} uncertain input factor, $\mathbf{X}_{\sim i}$ is the matrix of all uncertain factors but X_i, $\mathbb{E}_{\mathbf{X}_{\sim i}}[Y \mid X_i]$, is the expectation of the model output Y taken over all possible values of $\mathbf{X}_{\sim i}$ while removing the X_i uncertainty (i.e. keeping X^i fixed) and $V_{X_i}[]$ is the variance taken over all possible values of X_i. The indices S_i can be used to reveal the importance of the input factor X_i on the variance of the output and it is a normalized index, that is $\sum_i S_i = 1$

The main effect index reveals what is the importance of each uncertain factor on the uncertainty in the model output. It relatively cheap to obtain as it can be efficiently computed using given data methods or from a single Monte Carlo run (Plischke et al. 2013). The main drawback of the index is that interactions between input factors are not accounted for with this sensitivity measure. Higher order Sobol's effects (second and higher order interactions) compose the so-called total effect index S_{Ti}. This is a variance-based measure of the influence of an input i accounting for all the inter-

actions with other uncertain factors. It is defined as follows:

$$S_{Ti} = \frac{\mathbb{E}_{\mathbf{X}_{-i}}\left[V_{X_i}[Y \mid \mathbf{X}_{-i}]\right]}{V[Y]} = 1 - \frac{V_{\mathbf{X}_{-i}}\left[\mathbb{E}_{X_i}[Y \mid \mathbf{X}_{-i}]\right]}{V[Y]} \tag{3}$$

where S_{Ti} account for all the contribution to the total variance of the output $V[Y]$ when the first order effect of $\mathbf{X}_{-i}$ is removed.

2.2 *Elementary effects and Morris diagram*

The Elementary Effects (EEs) is a screening method used identify the effect of input factors $X(i)$ with $i = 1,2,\ldots,m$ on the output Y of a mathematical or computational model $M(\mathbf{X})$. The method consists in the calculation of m incremental ratios, also called Elementary Effects, which are used to assess the influence of the input variables and parameters. The i^{th} elementary effect of the m-dimensional input vector $\mathbf{X}_0$ is defined as follows:

$$\delta_i(\mathbf{X}_0) = \frac{Y\left(X_0(1),..,X_0(i)+\Delta,..,X_0(m)\right) - Y(\mathbf{X}_0)}{\Delta} \tag{4}$$

where the quantity Δ is a given variation in the input factor whose effect has to be evaluated. Intuitively speaking, the input factors leading to the higher incremental ratios $\delta_i(\mathbf{X}_0)$ have to be considered as the most relevant for the output quantity Y. Of curse, this relevance metric is valid only locally, in $\mathbf{X}_0$, where Y has been evaluated. Repeated One-At-a-Time (OAT) evaluations of random vector configurations provide the elementary effect method with global sensitivity analysis features (Turati et al. 2017). The mean and standard deviation of the EEs, resulting from random input vector configurations, can be plotted in the well-known $\mu(\delta) - \sigma(\delta)$ plot proposed by Morris (Morris 1991). If a factor X_i results in a small absolute value of the mean and small variance, it should be considered less relevant for the model. On the other hand, a factor X_i resulting in a high $|\mu(\delta_i)|$ has to be considered highly relevant for the model, i.e. it leads to the average higher variation in the output. Similarly, a factor X_i resulting in a high $\sigma(\delta_i)$ is also of interest for the model output. In fact, high $\sigma(\delta_i)$ probably indicate a non-linear relation between the factor i and the output and/or a relevant interaction with other factors. An example of Morris plot is presented in Figure 1 where the standard error of the mean (SEM) is used to decompose the plot in different areas.

The method has some points of strength, worth highlighting: 1) It is relatively easy to implement; 2) Computationally cheap compared to other global sensitivity methods also for high number of factors; 3) It uses a sensitivity measure which is simple to communicate (similarity between incremental rations and partial derivatives) to non-experts; 4) Compared to variance/based measures, shows if the input factors are (in average) positively or negatively correlated to the output.

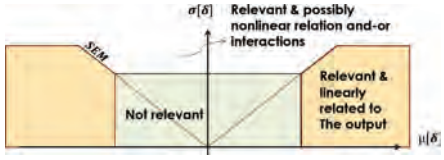


Figure 1. An example of Morris diagram and how to discern between important and non important factors.

3 THE CASCADING MODEL

A model for the simulation of steady-state operations of electric networks has been developed and calibrated in (Bing Li and Sansavini 2017). It can be used to simulate the initial contingencies that trigger the cascading events and estimate the post-contingency system states. The initial generation dispatch for each load demand is computed with a Security Constrained Optimal Power Flow (SCOPF), which takes into account the generators constraints, line flow constraints, voltage angles constraints and, optionally, the N-1 security constraints. After line tripping, DC power flow is used to evaluate the post-contingency power flow. The failures propagate in the grid through line over loading. Frequency control and protections, voltage protections and a variety of other automatic and realistic regulations and remedial actions are also included in the model.

A simplified flow chart of the cascading failures analysis is presented in Figure 2 adapted from (Bing Li and Sansavini 2017). The algorithm starts by loading power grid data, selecting the steady-state solver (e.g. DC-SCOPF) and a list of N-k contingencies. Then, for each contingency N-k, islands are identified, frequency deviation assessed and under frequency load shedding performed if necessary. Once power balance is restored, line flows are evaluated using the power flow solver and the lines exceeding their flow limit are removed from the grid topology. This process is repeated until grid stability is reached. The considered outputs are the total Demand-Not-Served (DNS) due to contingency N-k and lines failure indicator functions indicating if a line tripped during the simulation of the N-k contingency.

For simplicity, the contingency list has been obtained by random sampling N-1, N-2 and N-k line contingencies. To better identify and select

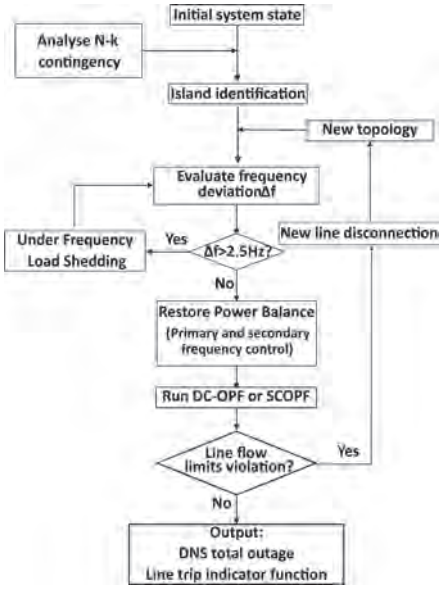


Figure 2. The flow chart of the algorithm for cascading failures analysis.

critical failure scenarios, methods such as the N-2 contingency screening, eg. the method presented in (Kaplunovich and Turitsyn 2016), could have been employed. However, a smart exploration of the contingency space was not the main aim of this work. Once the list is obtained, repeated N-k contingency analysis are performed as presented in Algorithm (Bing Li and Sansavini 2017).

3.1 *System and components performance indicators*

Several output measures can be obtained from the cascades model. In this work, we focus on 2 system-level indicators, which provide insights on the grid performance as a whole, and on N_l components performance indicators, one for each line in the system.

The indicators are the 95*th* percentile of the DNS cumulative distribution function $p_{95}(DNS)$, the average total number of lines tripped $\mu(N_f)$ and the line outage frequency $P_{f,l}$, defined as follows:

$$\mu(N_f) = \frac{\sum_{c=1}^{N_c}\sum_{l=1}^{N_l} I_{l,c}}{N_c}; \quad P_{f,l} = \sum_{c=1}^{N_c} \frac{I_{l,c}}{N_c};$$

where N_c is the total number of contingencies listed, N_l is the total number of lines in the system and $I_{l,c}$ is the indicator function for line l and contingency c. The indicator function will result 0 if the line survived the cascading propagation initiate by contingency c or 1 if the line failed, e.g. due to flows redistribution leading to an overload.

4 A CASE STUDY

The IEEE RTS96 power grid is used to test the methods and the cascading model and Figure 3 displays the grid layout. The power grid data can be found in (Grigg et al. 1999) and are not reported here for sake of synthesis. In this analysis, two representative uncertainty cases, named Case A and B, are considered. In Case A, the uncertainty associated with the load demand is explicitly modelled. In the second case, CASE B, also random generation costs are accounted for, thus introducing uncertainty in the power dispatch and increasing the dimensionality of the random input space. The DC cascading model presented in section 3, is employed for the solution of the cascading problem. A predefined contingency list is selected and includes 2444 line contingencies. The list counts the full set of N-1 and N-2 line failures and a set of 1000 random N-3 line failures. To simplify comparison between uncertainty cases and the different sensitivity analysis methods, the contingency list has been kept the same throughout all the analysis (i.e. the random set of N-3 contingencies has been sampled just once).

4.1 *CASE A: Random loads*

The first uncertainty case A assumes that uncertainty affects the 17 loads in the system due to inherent variability. The analysts lack better information regarding the variability affecting the load at each node, thus, the uncertainty in *Li* is simply modelled by assuming uniform distributions. The

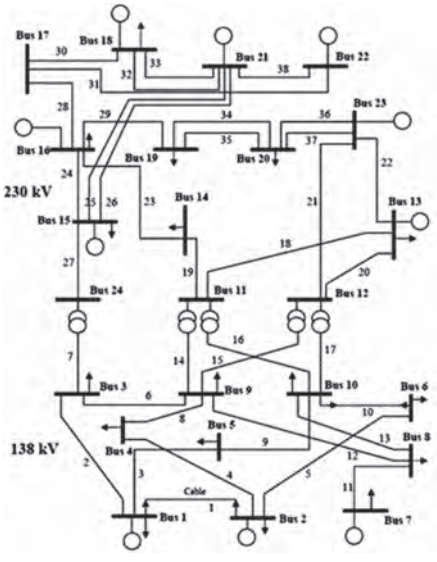


Figure 3. The IEEE RTS96 system, the connections between the 24 nodes, the lumped generators (32 generators) and the location of the aggregated loads (17 arrows).

distribution parameters have been selected to cover a range of values around the design loads and based on experts opinion:

$$L_i \sim U\left(0.5L_{d,i}, 1.2L_{d,i}\right) \quad i = 1,..,N_l$$

where $L_{d,i}$ is the design load of node i as presented in (Grigg, Wong, Albrecht, Allan, Bhavaraju, Billinton, Chen, Fong, Haddad, Kuruganty, Li, Mukerji, Pat ton, Rau, Reppen, Schneider, Shahidehpour, & Singh (1999) and the number of lines is $N_l = 17$.

Once the uncertainty sources are characterized, a preliminary uncertainty analysis is performed. Monte Carlo method is used to propagate 5e4 samples of the load profile. For each load sample, the cascading failure model is solved 2444 times, one for each contingency listed. The percentile of the demand not served, the average number of failed lines and the line outage frequencies are computed for each load sample as described in Section 3.1. The p_{95} (*DNS*) results are summarised in Figure 4. This figure presents a so-called cobweb plot, also known as parallel coordinates plot. It is a simple and effective way of visualising random input and output spaces in high dimensions. The X-axis reports the inputs loads and the percentile of the DNS (on the far right). The Y-axis reports the normalized inputs and output realisations of the Monte Carlo method. Each one of the dark dashed line in the background corresponds to one load profile realisation and corresponding The p_{95} (*DNS*) obtained through N_c model evaluations. Red solid lines are conditional samples, which highlight only the load combinations leading to the highest p_{95} (*DNS*). It can be observed, later confirmed by Morris' and Sobols' analysis, that there is a strong influence of some of the loads (e.g. in nodes 15 and 18) on the extremes of the DNS. In particular, when the power demanded in nodes 15 and 18 is small, the risk of facing severe DNS scenarios increases.

Morris and Sobol's indices have been computed aiming at better investigating which among the uncertain factors are key drivers for the output uncertainty. The Morris indices are obtained by selecting 250 random input vector realisations

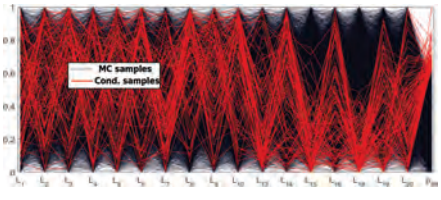


Figure 4. The parallel plot of the Monte Carlo loads and p_{95}(DNS) realizations. In red solid line the conditional samples which lead to the highest p_{95} and in the background (black dashed lines) all the MC realisations.

(saved from the MC) and computing incremental ratios δ as described in section 2.2. The Sobol's first order coefficents are obtained using given data sensitivity approaches, see ref. (Plischke, Borgonovo, & Smith 2013) for further details. This is a very convenient approach as for calculations, as the data from the MC run can be used for this and with essentially no-extra computational cost. On the other hand, total Sobol's indices require higher computational cost and in this work the Liu and Owen method (R. Liu 2006) is used for their computation.

The result relative to the DNS percentile and the average total number of line failed are presented and compared in Table 1. The Morris statistics and Sobol's main and total effect indices are also graphically presented in the $\mu - \sigma$ plot in Figure 4 and in Figure 5, respectively. Both methods identify L_{18} and L_{15} as the most influencing factors for the DNS and average number of line failures. Less relevant but, not to be neglected, is the effect of loads in nodes 8, 19 and 16. Morris analysis has the advantage of revealing an inverse relation between L_{18}, L_{15}, L_{19} and the outputs (see figure 4) which could not be revealed only using Sobol's indices. On the other hand, an increment in load 8 lead to higher risk of extreme DNS.

This result can be explained looking at the generators production profile, which is obtained solving the pre-contingency DC-SCOPF with objective

Table 1. Sobol's main and total effect mean and standard deviation for the elementary effects for the uncertainty case A for the DNS percentile and average total failed lines outputs.

	p_{95} (DNS)				$\mu(N_f)$			
	Sobol		Morris		Sobol		Morris	
	S_i	S_{Ti}	$\mu(\delta_i)$	$\sigma(\delta_i)$	S_i	S_{Ti}	$\mu(\delta_i)$	$\sigma(\delta_i)$
L_1	0.01	0.00	0.01	0.03	0.01	0.00	–0.1	0.4
L_2	0.01	0.00	0.01	0.03	0.00	0.00	–0.1	0.4
L_3	0.02	0.02	0.02	0.08	0.02	0.01	–0.5	0.7
L_4	0.01	0.00	0.01	0.03	0.01	0.00	–0.1	0.3
L_5	0.01	0.00	0.02	0.03	0.01	0.00	0.0	0.3
L_6	0.01	0.02	0.03	0.05	0.01	0.00	0.0	0.4
L_7	0.01	0.03	–0.01	0.06	0.01	0.01	0.0	0.7
L_8	0.04	0.03	0.06	0.08	0.03	0.05	0.4	0.7
L_9	0.01	0.02	0.02	0.05	0.01	0.01	–0.2	0.7
L_{10}	0.02	0.03	0.04	0.06	0.01	0.01	0.0	0.5
L_{13}	0.01	0.04	–0.01	0.09	0.01	0.03	–0.4	0.9
L_{14}	0.02	0.02	0.02	0.09	0.01	0.01	0.0	0.7
L_{15}	0.29	0.40	–0.18	0.20	0.33	0.33	–2.1	1.7
L_{16}	0.04	0.06	–0.05	0.09	0.03	0.02	–0.5	0.6
L_{18}	0.39	0.44	–0.20	0.20	0.47	0.54	–2.7	1.9
L_{19}	0.06	0.12	–0.08	0.13	0.03	0.05	–0.6	0.8
L_{20}	0.03	0.06	–0.04	0.08	0.02	0.02	–0.5	0.7

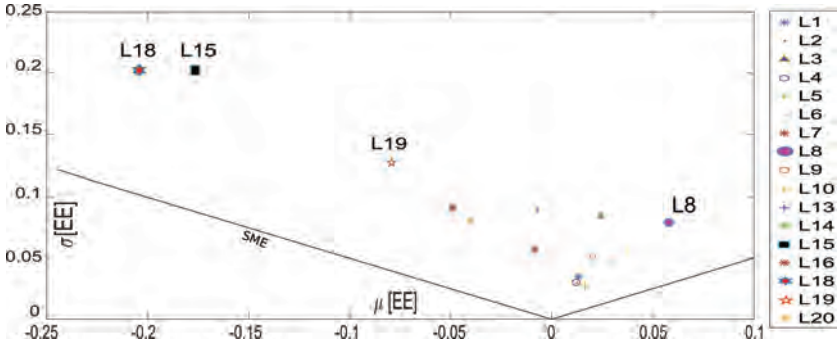


Figure 5. The Morris diagram for uncertainty case A and for the DNS percentile output. The mean and standard deviation of the EEs are reported on the X and Y axis, respectively.

of minimizing generation costs. The generators in nodes 18, 22 are associated with lower generation costs. This lead to the maximum exploitation of their production capacity, independently from the load profile realisation. Consequently, when electrical power is consumed in loco (e.g. the loads close to these generators as in 15 and 18), less power will be flowing from the 'northern' area to the 'southern' area of the network. On the other hand, if less power is demanded in, for instance, nodes 18 and 15 (or more power in 8), this increases the risk of higher loads on line such as 24, 25 and 26 which connecting the upper part of the grid with the lower part, and with it the risk of facing more severe post-contingency scenarios.

4.2 *CASE B: Random loads and generator costs*

The second uncertainty case B extends case A by accounting for generators costs uncertainties. The generation cost variability is characterised by uniform probability distributions as follows:

$$C_{g,i} \sim U(0.9,1.1) \quad i = 1,..,N_g$$

where $C_{g,i}$ is the cost of the generating unit i and the number of generators N_g is equal to 32. By assuming costs $C_{g,i}$ distributed uniformly between 0.9 and 1.1, the economic viability of the generators drastically changes if compared to case A. This lead to a higher variability in the economic dispatch, i.e. generators in nodes from 18 to 22 will sometime produce less than their maximum capacity. This case study shows the applicability of the method to larger input spaces and larger power grids. Furthermore it shows the impact of different generation profiles, in combination with load demands, on the cascading failures.

Similarly to the uncertainty case A, a Monte Carlo uncertainty propagation is performed and the Sobol's S_i indices and Morris $\mu(\delta)$ and $\sigma(\delta)$ have been calculated. The 5 most influencing factors

Table 2. Comparison between the top 5 most influencing factors according to the Sobol's main index and Morris mean and standard deviation. The output considered is the DNS percentile.

rank	S_i	$\lvert\mu(\delta)\rvert$	$\sigma(\delta)$
1	L_8	L_8	$G_{18}(1)$
2	L_3	$G_{18}(1)$	$G_{13}(2)$
3	$G_{18}(1)$	L_3	L_8
4	$G_{21}(1)$	L_6	$G_7(1)$
5	L_{18}	L_{18}	L_7

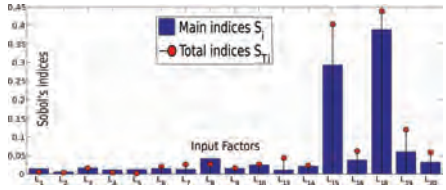


Figure 6. The Sobol s main and total effects obtained for the uncertainty case A and for the DNS percentile output.

(among the 17 loads and 32 generator costs) affecting the 95*th* percentile of the DNS are reported in Table 2. Multiple generators can be found in the same bus and to simplify the notation, the relevant costs are presented using the symbol $G_k(j)$, where j is the machine reference number within the bus k where the generator is installed. Differently from case A, load in node 8 emerged as the most relevant factor for the DNS percentile.

Uncertainty in the loads and generator costs has been propagated to the line outage frequency indicator $P_{f,l}$. The resulting MC realisations are displayed using a box plot in Figure 6. The X-axes shows the lines identification number and the Y-axes presents the $P_{f,l}$ values (red markers). Each box indicates the median (the central mark) and the bottom and top edges of the box indicate the

25th and 75th percentiles, respectively. It can be observed that the line connecting node 7 to node 8 results in the higher failure frequency and that lines in the lower voltage area of the grid (ID from 1 to 13) are more prone to failure. This result is probably due to the lower thermal limit (175 MW) and to the specific combination of grid topology, design load demanded in node 7 and 8 (125 and 171 MW, respectively) and generators in node 7 maximum lumped capacity (300 MW). Thanks to the sensitivity analysis, it has been possible to clarify which are the factors responsible for this peculiar behaviour, i.e. better understanding which are the variables which are contributing the most to $P_{f,7\text{–}8}$.

Main effect sensitivity indices have been computed for each line $P_{f,l}$ to reveal which of the input factors is affecting the most their variability. Results are presented graphically with a bar plot in Figure 8 and reported in Table 3. Table 3 presents only the factors leading to relatively high S_i, i.e.

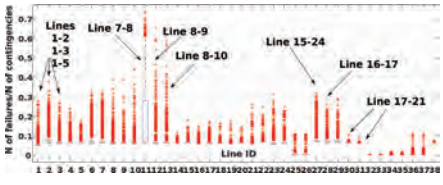


Figure 7. The box plot of the $\mathbf{P}_{f,l}$ realisations corresponding to different load and generation cost samples.

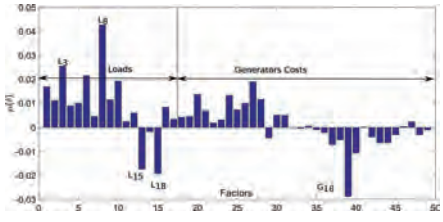


Figure 8. The tornado diagram presenting the mean o the elementary effects for the uncertain factors considered in case B.

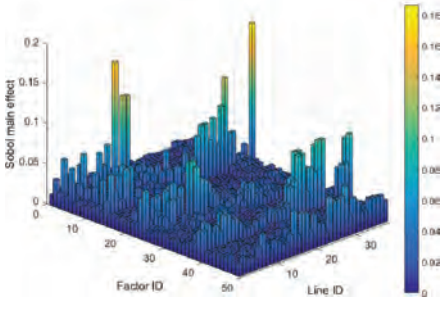


Figure 9. The $\boldsymbol{S}_i$ indices calculated for the 49 input factors and for the $\boldsymbol{P}_{f,l}$ outputs. The factors from 1 to 17 are loads at different locations and last 32 are the generator costs.

Table 3. The most influencing factors for the line failure probability. Factors leading to a $\boldsymbol{S}_i > 0.08$.

From	To node	Factors
7	8	L_7, $G_7(1)$, $G_7(2)$
8	9	L_8
8	10	L_8
15	21	L_{18}, $G_{18}(1)$, $G_{21}(1)$
15	21	L_{18}, $G_{18}(1)$, $G_{21}(1)$,
16	17	L_{18}
17	18	L_{18}, $G_{18,}(1)$, G_{21} (1)
17	22	L_{18}, $G_{18,}(1)$, G_{21} (1)
21	22	L_{18}, $G_{18,}(1)$, G_{21} (1)

greater than 0.08, and the corresponding components. It can be observed that the variability in the line 7–8 outage frequency is mainly affected by uncertainty in node 7 (generators and load). On the other hand, uncertainty in L_8 is not affecting much the variance of $P_{f,7\text{–}8}$ but it is the most relevant factor for $P_{f,8\text{–}9}$, $P_{f,8\text{–}10}$.

5 DISCUSSION AND CONCLUSIONS

In this paper, the sensitivity of a cascading failures model for power grids has been analysed. Variance-based global sensitivity analysis indices, i.e. Sobol's indices, have been computed to reveal which among the uncertainty sources is affecting the most the variances of the cascading failure model output. The Morris screening indices are also obtained and compared to variance based indices to improve confidence in the results and better understand dependencies between output and factors.

Different system-level and component-level indicators have been evaluated using the cascading model. The selected metrics were the 95*th* percentile of the DNS, the average total number of line failed and the frequency of line failure for each line. The IEEE RTS96 power grid has been selected as a representative case study and used to test the applicability of the methods to a real-world system. Two uncertainty cases (Case A and Case B) have been investigated, which were characterised by an increasing dimensionality of the aleatory space.

In the Case A, only load variability has been accounted for and the result suggested that two uncertainties in the loads in node 15 and 18 are the major contributors to the extremes of the demand not served. A similar result is obtained for the average total number of line failed. Morris had the advantage of showing a negative relationship between the DNS and loads in nodes 15 and 18. In reality prices are indeed affected by uncertainty, so a sensitivity analysis that assumes fixed prices (and therefore fixed generator dispatch) might be misleading in identifying critical components in the

power grid. Thus, in the uncertainty Case B, the variability of the generator costs and loads variability are both considered. The new economic setting changed the underling behaviour of the network and, consequently, of the cascading evaluation process. The Sobol's and Morris' analysis are fairly consistent in pointing out which among the load and generators costs are the most relevant for the system output. The results are quite different compared to case A, due to the difference in the economic setting of the generators. In addition, the sensitivity of the lines outage frequency has been computed.

This analysis was performed to investigate more in detail some cascading-relevant relationships between input loads, generators costs and line failures. The results are very interesting from an engineering perspective and at least 2 results can be highlighted which are helpful in a practical context:

- The vulnerable lines (i.e. prone to failure) and the most relevant factors affecting $P_{f,l}$ are identified (using sensitivity indices). This information can be helpful to support reliability-related decision, for instance, in deciding on weather it is better to replace the line with one having higher capacity (i.e. if $P_{f,l}$ high and is similarly affected by all the input factor), or if it may be more useful to intervene on the factors affecting $P_{f,l}$ (i.e. if $P_{f,l}$ high and sensitive to just few factors);
- When the uncertainty in the loads is identified as highly relevant for a system-level indicator, it is advisable to consider actions such as allocation of distributed generators or adopt peak-shaving (load variance reduction) control methods. This can be beneficial to reduce the uncertainty in the reliability performance of the network (reducing its variance).

The framework proved to be flexible and computationally quite cheap which is a requirement for its application to more realistic large size power networks. This will be the focus of future analysis.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the gracious support of this work through the EPSRC and ESRC Centre for Doctoral Training on Quantification and Management of Risk & Uncertainty in Complex Systems & Environments Grant number (EP/L015927/1).

REFERENCES

Bialek, J., E. Ciapessoni, D. Cirio, E. Cotilla-Sanchez, C. Dent, I. Dobson, P. Henneaux, P. Hines, J. Jardim, S. Miller, M. Panteli, M. Papic, A. Pitto, J. Quiros-Tortos, & D. Wu (2016, Nov). Benchmarking and validation of cascading failure analysis tools. *IEEE Transactions on Power Systems 31*(6), 4887–4900.

Bing Li, B.G. & G. Sansavini (2017, July). A genetic algorithm based calibration approach on validating cascading failure analysis. In *IEEE PES general meeting*.

Borgonovo, E. & E. Plischke (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research 248*(3), 869–887.

Dobson, I., B.A. Carreras, V.E. Lynch, & D.E. Newman (2001, Jan). An initial model fo complex dynamics in electric power system blackouts. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, pp. 710–718.

Grigg, C., P. Wong, P. Albrecht, R. Allan, M. Bhavaraju, R. Billinton, Q. Chen, C. Fong, S. Haddad, S. Kuruganty, W. Li, R. Mukerji, D. Patton, N. Rau, D. Reppen, A. Schneider, M. Shahidehpour, & C. Singh (1999, Aug). The ieee reliability test system-1996. a report prepared by the reliability test system task force of the application of probability methods subcommittee. *IEEE Transactions on Power Systems 14*(3), 1010–1020.

Kaplunovich, P. & K. Turitsyn (2016, Nov). Fast and reliable screening of n-2 contingencies. *IEEE Transactions on Power Systems 31*(6), 4243–4252.

Li, B., G. Sansavini, S. Bolognani, & F. Drfler (2016, July). Linear implicit ac pf cascading failure analysis with power system operations and automation. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1–5.

M. Sobol, I. (1990, 01). Sensitivity estimates for nonlinear mathematical models. 2.

Morris, M.D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics 33*(2), 161–174.

Nedic, D.P., I. Dobson, D.S. Kirschen, B.A. Carreras, & V.E. Lynch (2006). Criticality in a cascading failure blackout model. *International Journal of Electrical Power & Energy Systems 28*(9), 627–633. Selection of Papers from 15th Power Systems Computation Conference, 2005.

Patelli, E., H.J. Pradlwarter, & G.I. Schuller (2010). Global sensitivity of structural variability by random sampling. *Computer Physics Communications 181*(12), 2072–2081.

Patelli, E., M. Broggi, M. Angelis, & M. Beer (2014, June). Opencossan: An efficient open tool for dealing with epistemic and aleatory uncertainties. In *Vulnerability, Uncertainty, and Risk*, pp. 2564–2573–. American Society of Civil Engineers.

Plischke, E., E. Borgonovo, & C.L. Smith (2013). Global sensitivity measures from given data. *European Journal of Operational Research 226*(3), 536–550.

R. Liu, A.B.O. (2006). Estimating mean dimensionality of analysis of variance decompositions. *JASA 101*(474), 712721.

Rocchetta, R., E. Zio, & E. Patelli (2018). A power-flow emulator approach for resilience assessment of repairable power grids subject to weather-induced failures and data deficiency. *Applied Energy 210*(Supplement C), 339–350.

Rocchetta, R., M. Broggi, & E. Patelli (2018). Do we have enough data? robust reliability via uncertainty quantification. *Applied Mathematical Modelling 54*(Supplement C), 710–721.

Turati, P., N. Pedroni, & E. Zio (2017). Dimensionality reduction of the resilience model of a critical infrastructure network by means of elementary effects sensitivity analysis. pp. 457.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Managing interdependencies in critical infrastructures—a cornerstone for system resilience

P. Ferreira
Lusófona University—DAT, Lisbon, Portugal
University of Lisbon—IST-CENTEC, Lisbon, Portugal

E. Bellini
University of Florence, Florence, Italy

ABSTRACT: The management of critical infrastructures is increasingly challenged by the high complex and interdependent nature of their operations. The need for methods and tools that better address these challenges has been often argued in literature. An improved access and accuracy of information on local operational conditions is referred to as one of the assets to be pursued, towards better coping with complexity. Technology currently available facilitates the access to a wide range and amount of data and information. However, putting such data and information to use as an effective support to decision making remains poorly addressed. Project RESOLUTE proposes an approach to the management, aggregation and processing of "big data" towards an enhanced adaptive capacity of stakeholders within critical infrastructures. The approach is based on the understanding of functional interdependencies between stakeholders and the use of that understanding for the tailoring of existing data to various operational contexts and conditions. A modelling of critical infrastructures was developed using FRAM. This model is then used as a basis for the development of an IT platform that supports coordination and cooperation between stakeholders.

1 INTRODUCTION

The increasing interdependent nature that spans across all industry sectors poses major challenges. While most management practices remain grounded on the control of internal processes, an increasing number of threats emanate from beyond the formal boundaries of organisations, where control based approaches reveal many shortfalls. Safety barriers that are continuously raised against threats identified based on hindsight, often show little effectiveness in the face of the variability and uncertainty that emerge from a growingly interdependent world.

Risk management remains strongly attached to the assumption that phenomena can be sufficiently known and described, so as to fully master the likelihood and means through which undesired events may occur. However, there is a growing awareness that managing and coping with uncertainty is an unavoidable consequence of the interdependent world. Singly investing on the elimination of uncertainty through hindsight-based statistical and predictive approaches is no longer sufficient. The maturity of methods and tools to effectively manage uncertainty remains unsatisfactory in view of industry needs and desires of communities at large.

This paper presents an approach to the management of uncertainty within critical infrastructures based on the experience of project RESOLUTE. Rather than attempting to eliminate uncertainty through knowledge on how and when "things might happen", the focus is set on understanding their sources as a result of highly interdependent and tight coupled operations. Project methodology is outlined and the key aspects of interdependency and uncertainty within critical infrastructures are highlighted. Main results are then presented and, given that it is an ongoing project, the paper concludes with a discussion of foreseeable achievements in terms of potential enhanced adaptive capacities and ability to cope with uncertainty for critical infrastructure stakeholders. As argued by Woods (2015), enhancing such adaptive capacities can be placed at the core of systems resilience, in particular within the scope of the resilience engineering approach.

2 COMPLEXITY, UNCERTAINTY AND INTERDEPENDENCY

The notion of complexity is widely used and yet remains difficult to define with precision. Many definitions of the concept underline that

something complex challenges the ability to describe and also to predict in terms of its states or behaviours, and of outcome of any actions taken or changes introduced to it (Hollnagel, 2012a). This feature of complexity is particularly relevant for the scope of this discussion, as it can be placed at the source of uncertainty. According to Grote (2009), uncertainty may concern the probability of an event (state uncertainty), a lack of information on the outcomes of an event and the underlying cause-effect relationships (effect uncertainty), or a lack of information about response options and their consequences (response uncertainty). RESOLUTE proposes to address the need for understanding uncertainty through the modelling of functional interdependencies. This is addressed late in this paper.

Another important aspect of complexity is the lack of linear behaviours. As pointed by Flach (2012), things like a mechanical clock can exhibit a wide range of possible states and yet, because the relations between its parts assume a linear nature, these possible states can be predicted. This however, would fit the definition of something that is complicated rather than something complex. Complexity is related to non-linearity. Describing and predicting possible states and outcomes cannot be achieved by the knowledge of individual variables or components, and of their relations. Within complex environments, a given action may lead to many different outcomes and a given state can be achieved through many different combinations of variables or parts (Flach, 2012).

Interdependency if often at the source of non-linear behaviours. Variables or components can interact and combine in many different ways, much beyond simple cause-effect relations. Interdependency has been studied for a number of decades. James Thompson (Thompson, 1967) had anticipated in 1967 the critical role that this phenomenon would play across most social endeavours.

The pursuit of multiple goals and the need to secure the access to a wider diversity of resources to respond to this pursuit generate self-reinforced cycles. The more interdependencies are generated through coalitions and exchanges of resources, the higher the number and diversity of goals that are put into play. The conflicting priorities that must be negotiated, lead to the continuous pursuit of more convenient options, which in return can be placed at the source of a search for coalitions that may potentially offer higher benefits. Thus, multiple conflicting goals produce continuous adjustments in existing interdependent relations, and also generate the potential for the expansion towards new interdependencies. As interdependencies increase in number, diversity, dynamics and complexity emerges and with it, uncertainty increases (Flach, 2012).

3 THE RESOLUTE APPROACH

Complexity and its inherent interdependent nature, poses a challenge to the management of critical infrastructures. As stated by Flach (2012), "classical hierarchical or servomechanism-type control systems are inadequate as a basis for dealing with the unanticipated variability endemic to complex work domains".

3.1 *The context of critical infrastructures*

The European Directive 2008-114-CE from the European Commission defines critical infrastructures as an "asset, system or part of which is essential for the maintenance of vital societal functions, health, safety, security, economic or social well-being of people, and the disruption or destruction of which would have a significant impact".

Critical infrastructures are particularly prone to the emergence of complexity and uncertainty related issues. On the one hand, they operate at the intersection of a wide range of stakeholders, with diverse organizational and legal status (i.e. public service, non-profit and for profit…), and with multiple operational purposes (i.e. maintenance, oversight and control, support and service delivery…). On the other hand, critical infrastructures tend to operate within many different scales, both geographically and temporally.

In addition, the critical nature of the service they provide exposes these infrastructures and their operation to a particularly strong public and political scrutiny. This kind of exposure often generates important pressures towards heightened efficiency, reliability and safety of operations. This means that critical infrastructures are expected to generate the capacity to adapt to continuously changing environments under highly complex and uncertain operational conditions. Generating such adaptive capacities is at the core of resilience (Woods, 2015) and to which project RESOLUTE aims to contribute to. The challenge then becomes how to integrate flexibility as a requirement for adaptive capacities, whilst ensuring the degrees of coordination and alignment between multiple goals necessary to respond to efficiency and safety demands.

3.2 *RESOLUTE methodology*

Project methodology was based on the functional modelling of critical infrastructures as sociotechnical systems. The modelling activities were carried out with the FRAM – Functional Resonance Analysis Method (Hollnagel, 2012b). FRAM enables modelling activities through the description of human, technical and organisational elements as functions within an interdependent system. FRAM

is essentially a tool intended for the modelling of "real work". It has been mainly applied to the modelling of operations and work on what is often considered the "sharp end" where human activity and processes tend to be more tangible. Within the RESOLUTE approach, the aim was to develop a generic functional model at critical infrastructure level. The contrast with what is normally considered the "sharp end" is that, at critical infrastructure level, interactions between human, organisational and technical elements tend to be less direct and may become more difficult to perceive and describe. Interviews with subject matter experts were used to develop the necessary insight on operations that mainly consisted of decision-making processes distributed across multiple infrastructure stakeholders. Given the broad system scale that was needed in this case (a critical infrastructure), the functions described are often imprecise in terms of their nature (human, technical or organisational). The insight obtained through the contact with experts was used to ascertain if each function was described with an acceptable level of clarity and precision.

Functions were identified on the basis of "what (actions or processes) should be carried out in order for a given critical infrastructure to achieve its operational purpose (the delivery of a given service)". Functions are then described according to six different aspects: input, output, resources, preconditions, time and controls. In order to systemise the gathering of data and information for the modelling activity, a set of triggering questions were used. These are shown in Table 1.

Table 1. Questions supporting the description of functions.

	Triggering questions
Input	• What should start the function? • What should the function act on or change?
Output	• What should be the output or results of the function? • Do you should to inform anyone? • Do you have to collect or record/report anything? If so, where? • Who needs the output? Who will use what is produced?
Precondition	• What should be in place so that you can complete the function normally?
Resource	• What resources do you need to perform the function, such as people, equipment, IT, power, buildings, etc.?
Control	• Should be any formal procedures or instructions controlling the function? • Should be people, such as supervisors, controlling the function? • Should be there any priorities? • Should be there specific constraints?
Time	• Should be there any time related to the function? • Should be a certain time where you have to perform the function?

Through the identification of these six aspects for each of the functions described, the potential couplings between functions can be identified. Except for the function output, the remaining five aspects can be considered as inputs to the function that rely on couplings with "upstream" functions. These potential couplings may then become effective, as system operation is instantiated in the model. These couplings represent the "functional points" at which critical infrastructure stakeholders generate interdependencies to ensure process flows and overall conditions necessary to service delivery. This provided an important understanding of various critical operational aspects at system level, namely how key resources are exchanged amongst stakeholders, or operational control is carried out, among others.

The main focus of RESOLUTE was urban transport systems as a critical infrastructure. The generic FRAM model developed was interpreted in the context of urban transport systems, in order to produce a more concrete understanding of the different stakeholders and their interdependencies.

RESOLUTE proposes to urban transport stakeholders a cooperation and coordination platform, aiming at generating adaptive capacities that address the specific needs to cope with complexity and the uncertainty that it generates. The platform was designated as CRAMSS—Collaborative Resilience Assessment and Management Support System, and consisted mainly on the usage of "big data" to enhance the ability of sense-making at various decision-making processes within various management and operational levels of critical infrastructures.

3.3 *From FRAM to CRAMSS*

The CRAMSS integrates real-time and historical data retrieved from a wide range of operational and environmental aspects, namely traffic monitoring, weather conditions, public transport services, and water levels in river banks, among many others.

Information is vital for every kind of decision making process. The lack of information is one of the main sources of uncertainty (Grote, 2009). Through the powerful technology currently available in most decision scenarios, access to information and data can be relatively easy and fast. However, knowing what information and where to look for it within vast arrays of big data often poses many challenges. Determining what is relevant and needed at a given level and scenario of decision requires understanding its sources of uncertainty,

based on which available data and information may be processed and/or filtered.

The FRAM model previously developed was used as guidance for the tailoring of the data gathered and fed to the CRAMSS to the potential needs of different decision makers and actors, and across the different critical infrastructure stakeholders. In a first step, the functions described in the model were associated with different roles played by stakeholders in a given critical infrastructure. For instance, the function "monitor operations" was associated with the role and responsibilities attributed to operations control rooms such as those that be found at an urban transport network. The rational of this approach is illustrated in Figure 1.

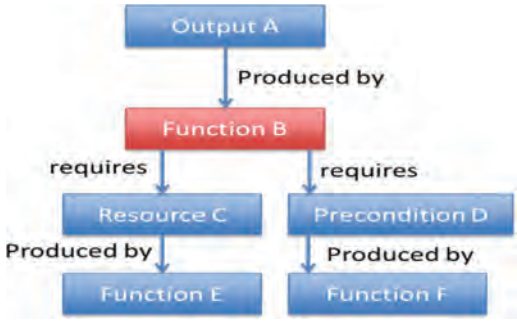


Figure 1. Structure of the analysis of function couplings.

Table 2. Functions in the FRAM model.

Function	Description
Develop Strategic Plan	Define the long-term objectives and identify critical resource needs and allocation strategy. It also involves the definition of policies, according to which all stakeholders should be strategically aligned. This is expected to take place by policy makers, regulators and with the participation of key stakeholders.
Manage financial affairs	Develop financial control and plan financial assets in accordance to financial needs of the operation and financial obligations. Often maintenance or renewal investments required by critical infrastructures greatly exceed the scope of legal ownership or responsibility of a given stakeholder. Managing such large scale projects requires detailed coordination amongst stakeholders and frequently the oversight of regulators, in particular for the oversight of financial responsibilities.
Perform Risk Assessment	Organisations carry out multiple risk assessment activities. Such activities tend to be developed within relatively limited scopes (i.e. specific tasks or projects, specific equipment…) and limited to a given domain of risk (i.e. safety, security, financial, environmental…). Assessment tools also tend to undermine risk factors that are not formally recognised and described in particular those emanating from beyond the formal boundaries of an organisation. This function should recognise the added value of integrating risk factors of diverse nature and of coordinating with multiple stakeholders, in particular along the supply chain of the service supplied by the critical infrastructure.
Coordinate Service delivery	The delivery of critical infrastructure services requires a thorough coordination amongst multiple stakeholders. Coordination activities should be carried out at various planning and operational stages of service delivery. This function contemplates operation related decision-making and activities that directly aim at keeping service delivery aligned with the strategic plan and the overall level of service in terms of quality and safety.
Manage awareness & user behaviour	As providers of fundamental public services, critical infrastructures tend to be significantly exposed to individual and collective behaviours, in many cases not just of the service end-users, but also of the wider public. Recent technological developments, in particular in relation to ICTs, offer a great potential for the enhancement of interactions with the public and the use of this potential towards an increased effectiveness in managing and deploying operational adjustments to various relevant events and circumstances.
Develop/ update procedures	The complete set of procedures forms a body of formal knowledge regarding management and operation requirements. They tend to reflect the structure of decision-making and production processes of a given organisation, so as to ensure coordination and shared understanding of operations and their goals. While this may be relatively well achieved at organisational level, amongst stakeholders of complex sociotechnical systems such as critical infrastructures, this is often very challenging. Procedures are essentially tools internal to organisations but within the scope of the function here described, to the extent possible and in addition to safety and efficiency requirements, procedures should also reflect the need for synchronisation and coordination amongst stakeholders at various CI process stages and supply chain levels. This should follow from the scope of a regulator's initiative, down to an active cooperation amongst stakeholders.
Manage human resources	Managing human resources within an organisation involves dealing with multiple relations between in-house and sub-contracting staff. The contractual boundaries may often be misaligned with real operational demands, where tight and dynamic cooperation amongst team members is required, regardless of the fact that various stakeholders are likely to be formally involved. Beyond the management of staff contractual relations, rosters and other human related operational needs, this function takes into account the need to manage the dynamics of close operational cooperation amongst multiple stakeholders and the need to align such dynamic relations with the formally established and recognised responsibilities and accountability.

(*Continued*)

Table 2. (*Continued*).

Function	Description
Training staff	In line with the principles previously outlined in relation the management of human resources, in addition to employee training needs, and their quality control, this function must also account for the need to provide and control the quality of training of staff that while working under other stakeholders, may operate on a more or less continuous basis with the premises of a given organisation such an infrastructure owner or manager. This relates to initiatives such as cross training and shared expertise programmes.
Manage ICT resources	Provide/maintain/update/develop/repair information and communications services to support critical infrastructure operation and management. Information systems may be owned and managed by a given organisation but may be strongly reliant on the operation and input from multiple stakeholders. ICT often gives shape to many interdependencies and the management of such resources should recognise this critical system role, namely by providing an overall system understanding of how these resources and made available and used by stakeholders in view of the overall service delivery (system operational purposes).
Maintain physical/cyber infrastructure	Maintenance activities require increasingly skilled and specialised staff and technical resources. Because their nature, maintenance services are often sub-contracted and providers become stakeholders with tight couplings with operational requirements. In addition to the planning, delivery and testing of maintenance activities, this function also incorporates the need to continuously assess the integration between in-house and sub-contracted maintenance resources, in view of process and technology changes, and overall operational environment demands.
Monitor Safety and Security	Integrated risk management has been recognised as a potentially valuable approach but has proven to present many managerial and operational challenges. As two of the fundamental risk domains for the operation of all critical infrastructures, safety and security should be managed in the scope of an integrating function, aiming to maximise efficiency and effectiveness of assessment and control measures and to integrate multiple interdependent risk factors that emanate from both within and beyond organisational boundaries.
Monitor Operations	The monitoring of service delivery performance is often singly based on lagging indicators and stakeholders tend to each assess their performance in reference to internal targets to be met, which may not necessarily reflect overall needs of the service delivered at critical infrastructure level. This function envisages the development of shared performance assessment practices amongst critical infrastructure stakeholders, mainly by integrating stakeholders' targets with overall service delivery needs. This becomes fundamental to generate overall system performance understanding.
Monitor Resource availability	Complex sociotechnical systems such as critical infrastructures rely on increasingly diversified and dynamic supply chains. This function focuses on generating an overall coordination of resource planning and deployment, taking into account the need to align multiple stakeholder needs with CI service delivery. This requires an understanding of resource flows and their main variability trends.
Monitor user generated feedback	Current technology provides the means to monitor service usage on a wide range of parameters and produce in real time, fundamental support to the deployment of operational adjustments. This function deals with the need for an integrated approach to the assessment of user generated feedback, mainly by placing this data and information in the context of operational monitoring. This requires the coordinated action amongst multiple stakeholders under a shared framework.
Coordinate emergency actions	The operation of critical infrastructures relies on the close cooperation amongst multiple stakeholders. Emergency response scenarios pose additional challenges, mainly by adding significant time pressure and high uncertainty (and therefore heightened risk) to this already complex operational environment. Coping with such challenges places even greater emphasis on the coordination needs and increased pressure on already limited resources. This function deals mainly with the requirements of an efficient distributed decision-making process under emergency response scenarios, namely the availability of accurate and timely information and data, and coordinated action of multiple and diverse stakeholders, often under unplanned and unforeseen circumstances.
Restore/Repair operations	Restoring operational capacities after significant damages requires much more than the re-allocation of system resources, namely those foreseen under maintenance and renewals projects. Dedicated teams are normally put in place to design, plan and execute specific projects, which in the case of critical infrastructures, in addition to the need to maintain minimum operation capabilities, is also likely to require the containment of impacts on other interdependent infrastructures.
Provide adaptation & improvement insights	With the scope of resilience, the operation of complex sociotechnical systems is challenged by two opposing needs: sustaining adaptive capacities to continuously changing operational conditions (flexibility) and the continued and coherent pursuit of goals within their own timescales (rigidity/robustness). For instance, Operation and production goals may be reassessed on an annual basis, while strategic goals may be addressed on a five year basis. Without compromising the consistency and feasibility of planned operations within each of those timescales, a given degree of flexibility must be ensured, in order to sustain the ability to respond to unforeseeable operational changes. This relates to aspects such as learning from ex-post event analysis, de-briefing, daily operations and providing insights for system capacities adaptation, keeping operations record, examining good practices, performing impact analysis of suggested actions, among others.
Collect event information	Collecting in house and external event data as good practices and/or historical data (archiving). From the perspective of resilience, this should not only address the occurrence of undesired events but most of all the understanding of factors that under highly variable circumstances become critical for achieving successful performance.

4 PROJECT ACHIEVEMENTS

The couplings described in the FRAM model were used to relate types of data to different user profiles for the CRAMSS. These user profiles were defined according to different roles and activities carried out by key stakeholders within the delivery of urban transport services. Emphasis was placed on conveying to users across multiple stakeholders, information on the conditions or operational status of other stakeholders with which relevant functional couplings had been identified through the FRAM model. The purpose was to create conditions for an improved cooperation and coordination between interdependent stakeholders. Providing information and data with a higher context and time dependent relevancy is expected to reduce uncertainty associated to decision-making and enhance the capacity to adapt to the continuously changing conditions under which such decisions are made, and their resulting actions taken.

For reasons of practicality, the complete FRAM model cannot be reproduced here in such a way that functions and potential couplings could be clearly presented. To illustrate the project outputs achieved so far based on the approach previously described, Table 2 provides a brief description of all the functions identified.

5 CONCLUSIONS

Beyond the provision of "big data" to support decision-making, the approach here outlined addresses the need to tailor a wide range of available data to different scenario and operation needs, taking into account interdependencies that are critical for process flows and the delivery of services.

The control of operations and processes remains strongly grounded on centralised mechanisms and on "top-to-bottom" procedural structures. One of the shortfalls of such approaches is the rigidity that it imposes on processes and operations, which often fails to account for the need for local adaptive capacities to uncertainty and continuously changing operational conditions (Woods, 2015). The conflicts generated by different goals and local priorities, and across multiple organisational boundaries frequently escape the prescribed work of procedures. The CRAMSS and the underlying FRAM model offers to each stakeholder an enhanced adaptive capacity. In some cases, this capacity means the ability to adjust operations and conditions in anticipation of critical events. In others, it may only provide the means for the detection of events in hindsight, but nevertheless, with an early warning and with increased precision and detail on the actual occurrence. Overall, these adaptive capacities and the ability to cope with complexity are at the core of system resilience.

Throughout the remaining duration of project RESOLUTE, the CRAMSS and the various tools on which it is build will be tested under real scenarios (a first one in the city of Florence and a second one in the city of Athens). The outcome of these testing activities will be fed back to the FRAM model, aiming to improve the understanding of stakeholder interdependencies that it provides.

REFERENCES

Flach J. (2012) Complexity: learning to muddle through. Cognition Technology and Work, vol. 14 (pp. 187–197).

Grote, G. (2009). Management of Uncertainty: Theory and Application in the Design of Systems and Organisations. Cranfield, UK: Springer.

Hollnagel E. (2012a) Coping with complexity: past, present and future. Cognition Technology and Work, vol. 14 (pp. 199–205).

Hollnagel E. (2012b) FRAM, the Functional Resonance Analysis Method: Modelling Complex Socio-technical Systems. Aldershot, UK: Ashgate.

Thompson J. (1967) Organizations in action. Social Science bases of Administrative Theory. McGraw-Hill, New York.

Woods D. (2015) Four concepts for resilience and the implications for the future of Resilience Engineering. Reliability Engineering and System Safety.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Application of PCE sensitivity analysis method to gas transmission network

V. Kopustinskas & P. Praks
European Commission, Joint Research Centre (JRC), Directorate for Energy, Transport and Climate Energy Security, Distribution and Market Unit, Ispra (VA), Italy

T. Mara & R. Rossati
European Commission, Joint Research Centre (JRC), Competence Centre on Modelling, Indicators and Impact Evaluation Unit, Ispra (VA), Italy

ABSTRACT: The study presents an uncertainty and sensitivity analysis exercise of the security of gas supply model implemented in the probabilistic gas network simulator ProGasNet. The study showed the potential and usefulness of sensitivity study applied to the ProGasNet gas network model. It has not only identified the most important model parameters for which more attention should be paid during the estimation process of the input parameter values, but also provided useful insights into the simulation process by confirming, e.g., the heterogeneity of the network from a sensitivity analysis perspective. The model was run for four different scenarios representing different disruption situations. The study confirms the results already observed in other studies that some disruption scenarios affect only part of the network (e.g. specific countries) while other parts of the network are not affected. This clearly indicates heterogeneity of the network and the need for further infrastructure development. The most important parameters for each country are identified, and peak demand value is the main parameter for the three countries.

1 INTRODUCTION

Natural gas networks can be viewed as complex technical systems, which are exposed to various threats, for example technical failures, natural disasters and human/political uncertainties. As a consequence of these threats, a subset of network components might fail. In order to simulate security of gas supply due to component failures/attacks, the probabilistic gas network simulator ProGasNet (Probabilistic Gas Network Simulator) is being developed.

ProGasNet is able to model, in a single computer model, capacity and reliability constraints of a natural gas network. The physical model is based on graph theory (maximum flow algorithm), whereas component failures are simulated by the Monte-Carlo method. The ProGasNet simulator has been developed with the primary purpose to quantify the security of gas supply situation in probabilistic metrics (V.Kopustinskas et al., 2012). The simulator can be used to perform risk assessment of the gas transmission network as required by the EC Reg 2017/1938 (EU Regulation, 2017), formerly known as 994/2010 (EU Regulation, 2010). In addition, the simulator can evaluate infrastructure development plans and the proposed PCI projects. ProGasNet was used in vulnerability assessment which could be of interest to critical infrastructure protection policy makers (Praks et al., 2017).

The ProGasNet has been applied to gas transmission networks of several EU countries, however geographical information cannot be disclosed. Various types of analysis have been performed: reliability, vulnerability, security of supply analysis and the results have reported reliability of supply estimates under different disruption scenarios (Praks et al., 2015). The ProGasNet model also provides an indication of the worst networks nodes in terms of security of supply and their numerical ranking. The model is very powerful to compare and evaluate different supply options, new network development plans and analyse potential crisis situations.

Despite of the insightful information that can be obtained regarding the security of gas supply within European regions with this simulator, still ProGasNet remains a strong approximation of the reality of gas supply in Europe (simplified assumptions, quasi steady-state modelling, lack of knowledge about some parameter values, etc). Therefore, it is important to assess the impact of epistemic uncertainties in gas network models in order to point out those sources of uncertainty that have an impact on the model responses of interest. Identifying important sources of (epistemic) uncertainty allows to guide further investigation for possible improvements of gas network models.

In this study we analyse an EU regional network already used for other purposes and bottleneck

analysis in particular (Kopustinskas et al, 2015). The focus of this study is to perform sensitivity analysis by using Polynomial Chaos Expansion (PCE) method (Sudret, 2008).

2 PROGASNET SIMULATOR

The ProGasNet (Probabilistic Gas Network) simulator is the JRC in-house developed software tool currently in use at the Joint Research Centre. ProGasNet is used for experimental simulation-based security of supply analyses of selected European gas transmission networks. Usually, one million Monte-Carlo simulations are automatically solved within one hour on a single-core processor computer. The software tool can make use of a multi-core computer, as multiple simulations (Monte Carlo runs) can be evaluated independently. The gas transmission network model implemented in ProGasNet is based on Maximum Flow (MF) algorithm well known in graph theory (Deo, 2008). The algorithm was modified to reflect gas transport property to have lower pressure at the points far from the supply source. This means that in case of disruption and lack of gas in the network, consumers that are closer to the source have better chances to be served rather than those that are located far away. Therefore, the maximum flow algorithm was modified to distribute gas first to the consuming nodes close to the supply source and distance from the nearest source was used as a priority criteria. Although not being a model requirement, this assumption was chosen as if network operator decision on how to distribute available gas among the users is not known. No doubt that in a real crisis management situation, the network operator has more options on how to supply and distribute the available volumes of natural gas.

The model is not running physical gas pressure and flow computations, but uses available results of physical models as a set of rules to define flow limitations. The ProGasNet simulates network facility failures (pipeline ruptures, failures of compressor stations, unavailability of LNG terminals and storages) by Monte Carlo method and each different network configuration is evaluated by modified maximum flow algorithm to evaluate available gas for each network consuming node. The statistical results are obtained from 1 million of Monte-Carlo runs.

3 STUDY CASE AND NETWORK DATA

Figure 1 shows topology of the study case gas transmission network. It is based on a real regional network topology and data, however geographical

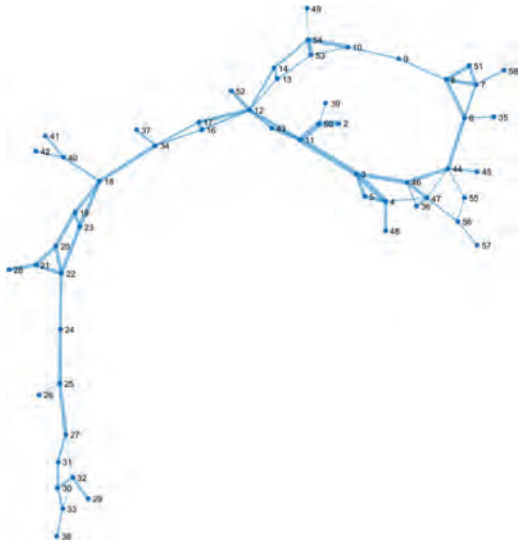


Figure 1. Topological layout of the gas network.

location is not displayed. The transmission network GIS data are converted to a graph by creating nodes and links (edges). The nodes are:

- Demand nodes (consumers connected to pressure reduction stations of the transmission network;
- Compressor stations;
- Supply nodes (storages, LNG terminals, import points at cross-borders).

The network links (edges) are typically pipelines. The model explicitly considers two parallel pipelines as two components (double links between nodes).

The basic network data are the same as already reported (Kopustinskas et al., 2015). Here we replicate only the most important network data.

The demand nodes are determined by daily demand value (Table 1). This value is a peak demand value, but it could be also average winter or summer consumption value depending on the purpose of the study.

Table 2 shows maximum capacities and type (pipeline, UGS or LNG) of input supply nodes. In case of Underground Gas Storages (UGS), also the output values of not fully loaded storages can be used.

The total maximum supply capacity is 77.5 mcm per day. The total network peak demand is 45.8 mcm/d, so the network has certain degree of spare capacity to compensate supply disruptions. The experience of different analysis already performed shows that depending on where the disrup-

Table 1. Network demand nodes, in millions of cubic meters per day (mcm/d).

Node	Demand	Node	Demand
4	0.1	34	0.5
5	3.2	35	0.1
6	0.1	36	4.2
7	0.3	37	1.3
9	0.1	39	0.3
10	1	41	0.6
13	0.5	42	0.6
17	0.1	43	0.2
18	8.5	44	0.7
20	0.6	45	1.3
25	0.5	47	0.5
26	0.8	48	1.8
27	3	49	0.2
28	6	51	7
30	0.5	52	0.6
33	0.5	53	0.1

Table 2. Maximum capacity of supply sources.

Node	Type	Capacity, mcm/day
2	Pipeline	31
10	LNG	10.2
11	Pipeline	7
19	UGS	30
29	Pipeline	4
38	Pipeline	1.2

tion happens, internal bottlenecks in the network prevent from full usage of this spare capacity.

For each network component, failure data must be provided. The following components (nodes) are considered for failures:

- Compressor Station (CS) failure: 2.5E-01/yr;
- Underground storage failure: 1.0E-01/yr
- LNG terminal failure: 1.5E-01/yr
- Pipeline failure: 3.5E-05 /km/yr.

The model uses annual failure data (probability of failure per year), however when simulations are performed, one month interval is considered. It is assumed that the same peak consumption in the network is constant during this one month period.

The compressor station node normally is modelled as working or failed, for each state determining the corresponding capacity of the outgoing pipelines. The capacity reduction due to compressor station failure is normally estimated by hydraulic model computations or expert evaluation. As a consequence due to a CS failure, capacity reduction by 20% of the inlet pipelines and also the outlet pipelines until the next connection node is assumed. This assumption is based on physical flow models, however is not accurate in all cases and also multiple CS failures will have more severe effects on the network operation.

The pipeline import points are not considered as failure-prone elements due to lack of upstream network model, however they are modelled as on/off elements by scenario analysis.

4 DISRUPTION SCENARIOS

In total four different gas disruption scenarios were analysed. The first scenario is the reference scenario during which the system operates under normal conditions without predefined disruptions.

Scenario 1: All available sources operate. Scenario 1 represents a basic scenario when all sources can be used for supply and the network components can fail randomly according to their reliability parameters.

Scenario 2: Node 2 disruption. In this scenario, supply node 2 (the largest gas source in capacity) is not available. This scenario can test the system for the largest source disruption which can be classified as N-1 situation looking at the network globally.

Scenario 3: Node 19 disruption. Scenario 3 runs the model with disconnected node 19, the second largest gas source.

Scenario 4: Loss of two largest gas sources (Nodes 2 & 19). The underground gas storage can be unavailable due to technical problems, failures or inability to fill it up during summer period. The Scenario 4 simulates a more challenging crisis in which the both sources of the highest capacity are unavailable. The scenario 4 is used to demonstrate vulnerability of the network, when the largest and the second largest gas sources are lost simultaneously. The network can be supplied only with source nodes 10 and 11 and small sources 29 and 38.

5 SENSITIVITY ANALYSIS METHODOLOGY

Uncertainty in model predictions stems from the lack of knowledge about the process of interest (in the present case, about the gas transport), model simplification and parameters' value. For some of the model inputs, it is possible to refine our knowledge about their probable value (that is, decreasing their uncertainty) but such a task is time consuming. Our strategy is i) to assign large but likely uncertainty ranges to the inputs of the

gas network model, ii) to check whether this yields large uncertainty in the model predictions of interest and iii) to identify eventually those inputs that are mostly responsible for the predicted uncertainty. In step iii), it is expected that only a few of the uncertain inputs are identified as influential in order to reduce the effort to pay during the subsequent input uncertainty refinement.

Twenty model inputs of the gas transmission network model have been deemed as uncertain (Table 3). It can be noted that some are assigned uniform distributions within plausible ranges while some others are assigned normal distribution. It is assumed that the model input values are independent of each other. These uncertainties reflect the experts' belief before further investigation. The prior uncertainties being large reflect the fact that the experts (here the modellers) have a vague knowledge about model inputs uncertainty.

This study has applied Polynomial Chaos Expansion (PCE) method to estimate sensitivity indices. The idea of PCE is to approximate variance decomposition terms (Sobol, 1993) by multidimensional orthonormal polynomials. The rate of convergence of such an expansion of course depends on the regularity properties of $g(\boldsymbol{x})$, the model response of interest. By exploiting the Parseval-Plancherel relationship, one obtains a standard variance decomposition equation. Therefore, we get (Wiener, 1938),

$$g(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in N^n} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \tag{1}$$

where $\boldsymbol{\alpha} = \alpha_1 \quad \alpha_n$, with $\alpha_i \in N$, is a multi-index indicating whether $\psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ depends on x_i ($\alpha_i > 0$) or not ($\alpha_i = 0$), a_α is the polynomial coefficient associated with $\psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ which is the so-called multivariate orthonormal polynomial chaos that is written

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \psi_{\alpha_1}(x_1) \cdot \quad \cdot \ \psi_{\alpha_n}(x_n)$$

$\psi_{\alpha_i}(x_i)$ being the α_i-th degree univariate polynomial basis element ($\psi_0 = 1$).

The expression of the univariate polynomial basis elements depends on the PDF assigned to the input variables. If $x_i \sim U(-1,1)$, $\psi_{\alpha_i}(x_i)$ is the Legendre polynomial of degree α_i, while if $x_i \sim N(0,1)$, $\psi_{\alpha_i}(x_i)$ is the Hermite polynomial of degree α_i. One can rely on the Wiener-Askey scheme to choose the appropriate polynomial family (Xiu & Karniadakis, 2002).

Once a PCE expansion such as Eq.(1) is obtained, it is straightforward to prove that the total variance of $g(\boldsymbol{x})$ is,

Table 3. Input uncertainty distributions of the gas network model.

N	Parameter	Baseline value	Distribution(1)	Type(2)	Accuracy(3)
X_1	Capacity of gas source N2	31.2	U(16,31.2)	A	L
X_2	Capacity of gas source N19	30.0	U(15,30)	A	L
X_3	Capacity of gas source N10	10.2	U(5,10.2)	A	L
X_4	Capacity of gas source N11	7.0	U(3.5,7)	A	M
X_5	Compressor station capacity reduction factor	0.2	$N(0.2,0.05^2)$	E	M
X_6	Peak demand of Country 1	15.5	$N(15.5,0.75^2)$	E	L
X_7	Peak demand of Country 2	12.1	$N(12.1,0.6^2)$	E	L
X_8	Peak demand of Country 3	5.3	$N(5.3,0.4^2)$	E	L
X_9	Failure frequency of LNG	0.15	$N(0.15,0.015^2)$	E	H
X_{10}	Failure frequency of storage facility	0.1	$N(0.1,0.01^2)$	E	M
X_{11}	Failure of compressor station	0.25	$N(0.25,0.025^2)$	E	M
X_{12}	Failure frequency of a pipeline	3.5E-05	$N(3.5 (10^{-5},(3.5 (10^{-6})^2)$	E	M
X_{13}	Capacity of DN1000 pipeline	30.6	$N(30.6,1.5^2)$	E	H
X_{14}	Capacity of DN800 pipeline	17.1	$N(17.1,0.86^2)$	E	H
X_{15}	Capacity of DN700 pipeline	12.1	$N(12.1,0.6^2)$	E	H
X_{16}	Capacity of DN600 pipeline	8.1	$N(8.1,0.40^2)$	E	H
X_{17}	Capacity of DN500 pipeline	5.1	$N(5.1,0.25^2)$	E	H
X_{18}	Capacity of DN400 pipeline	2.8	$N(2.8,0.14^2)$	E	H
X_{19}	Capacity of DN350 pipeline	2.0	$N(2.0,0.1^2)$	E	H
X_{20}	Capacity of DN300 pipeline	1.3	$N(1.30,0.065^2)$	E	H

(1) U = Uniform distribution, $N(\mu,\sigma^2)$ = Normal distribution of mean μ and variance σ^2.
(2) E stands for epistemic uncertainty as opposed to A-aleatory uncertainty.
(3) The modellers belief regarding the assigned prior uncertainty: L = Low, M = Medium and H = High.

$$V\left(g(\boldsymbol{x})\right) = D_y = \sum_{\boldsymbol{\alpha}\in N^n} a_{\boldsymbol{\alpha}}^2 - a_{0\ldots0}^2$$

by exploiting the orthonormality property of the polynomial basis elements, that is,

$$E\left(\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \cdot \psi_{\boldsymbol{\beta}}(\boldsymbol{x})\right) = \delta_{\boldsymbol{\alpha}\boldsymbol{\beta}}$$

where $\delta_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ is the symbol of Kronecker.

Therefore, it is possible to estimate the Sobol' indices from the PCE coefficients as follows,

$$S_{i_1} = \frac{V\left(E\left(g(\boldsymbol{x})\middle|x_{i_1}\right)\right)}{V\left(g(\boldsymbol{x})\right)} = \frac{\sum_{\alpha_{i_1}\in N} a_{0\ldots0\alpha_{i_1}0\ldots0}^2}{\sum_{\boldsymbol{\alpha}\in N^n} a_{\boldsymbol{\alpha}}^2 - a_{0\ldots0}^2}$$

$$ST_{i_1} = \frac{E\left(V\left(g(\boldsymbol{x})\middle|\boldsymbol{x}_{-i_1}\right)\right)}{V\left(g(\boldsymbol{x})\right)} = \frac{\sum_{-\in N^n:\alpha_{i_1}>0} a_{\boldsymbol{\alpha}}^2}{\sum_{\boldsymbol{\alpha}\in N^n} a_{\boldsymbol{\alpha}}^2 - a_{0\ldots0}^2}$$

Hence, the issue with the PCE approach for variance-based sensitivity analysis is to assess the PCE coefficients. In this work this is achieved with the Bayesian sparse PCE developed in (Shao et al, 2017). With this approach, the variance decomposition is obtained from one single Monte Carlo sample of size N. This is very computationally cheap compared with other classical approaches (Saltelli, 2002; Saltelli et al, 1999). The cost of the analysis is a criterion to keep in mind with a long-time run model like the ProGasNet.

6 UNCERTAINTY & SENSITIVITY ANALYSIS: RESULTS AND DISCUSSION

6.1 *Model output selection*

Given the computational time required to run ProGasNet, a sample of size 512 was considered. The input sample was generated according to the probability density function of each input variable (see Table 3). The low-discrepancy LPτ sequences of (Sobol' et al. 1992) were employed. It required four days of calculation with ProGasNet to propagate the input uncertainty into the model responses for the four different scenarios under analysis.

It is possible to analyse the impact of the model input uncertainty onto different model responses. For this purpose, while propagating the input uncertainty with a Monte Carlo sample, one has to save the different model responses of interest after each model run. In the present work, we have considered 20 different model responses, namely: $\bar{S}$ the mean volume of gas supply, $P(S=0)$ the probability of none gas supply, $P(S<0.2D)$ the probability of supplyinless than 20% of the demand, $P(S<0.5D)$ the probability of supplying less than 50% of the demand, $P(S<0.8D)$ the probability of supplying less than 80% of the demand and $P(S<D)$ the probability of supplying less than the demand. Doing so for each country provided a set of $3\times6=18$ different model responses.

The different model responses are more or less sensitive to the uncertainty in the model inputs. Figure 2 shows the Monte Carlo simulation results

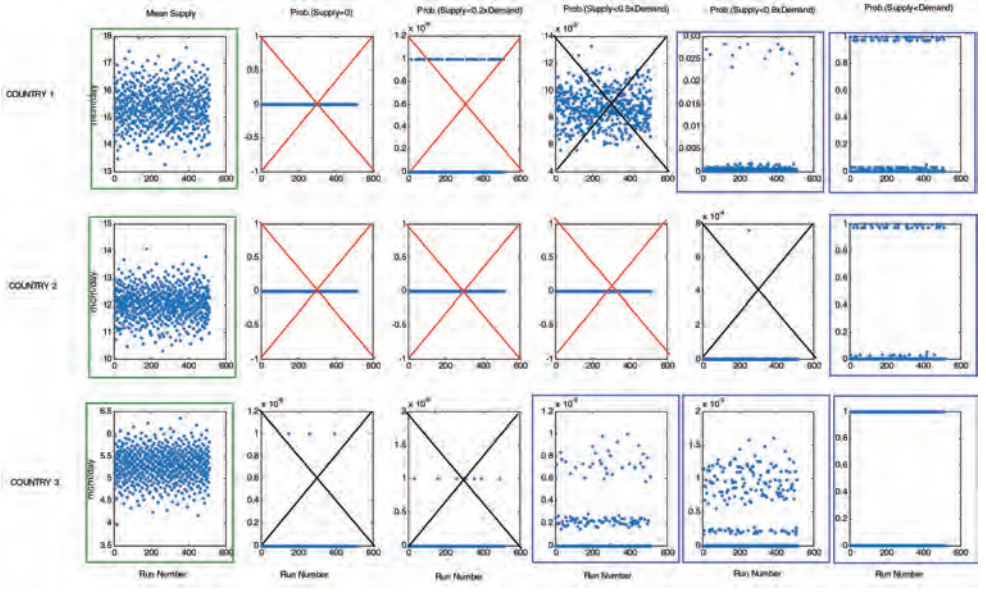


Figure 2. Monte Carlo predictions of the 18 model responses.

obtained for Scenario 2. In Figure 2 outputs with a cross do not change significantly and are not analysed. Outputs with a green frame are analysed with the PCE approach. It can be noted that some output variables do not vary at all (likewise $P(S=0)$ for country 1 & 2), some do not vary significantly (values less than 10^{-3}), some others only take discrete values (e.g. $P(S<D)$) while the averages volume of gas supply $\bar{S}$ vary continuously. Consequently, only those model responses that are significantly impacted by the input uncertainty are analysed in the sequel. Only the model responses that take continuous values are analysed with the polynomial chaos expansion.

6.2 *Results for Country 1*

In the following sections, we analyse the predictive uncertainty on the mean volume of gas $\bar{S}$ supplied by the network to the different countries. The estimated probability density functions of the mean volume of gas supply (in millions of cube meters/day) for the four different scenarios are depicted in Figure 3. We note that the PDF's are the same for scenario 1 (normal operation) and scenario 3 (Source 19 off). This means that the system is completely resilient to the failure of this important source regarding the gas supply in Country 1. When the main source of gas is off (scenario 2), the network remains quite resilient. However, as far as scenario 4 is concerned, the system completely fails at satisfying the gas demand (vertical dashed-line) although some quantity of gas is supplied.

Notably, in the first three scenarios, according to the model, the ability of the network to supply sufficient gas volume depends on the true value of some of the uncertain input variables. Indeed, Figure 3 indicates that the system might fail at satisfying the gas demand under certain uncertain conditions, linked to capacity of sources which is subject to strong aleatory uncertainty.

To guess which are the critical uncertain inputs responsible for the variability of this model response we have carried out a global sensitivity analysis for each scenario. Given that this model response takes continuous values, the PCE method

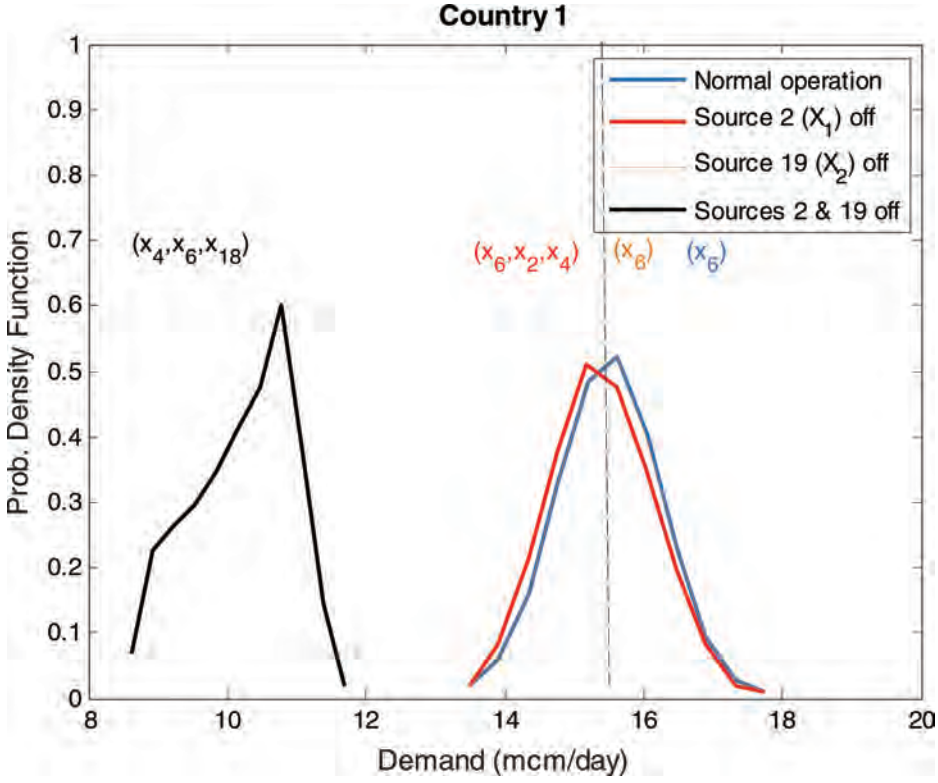


Figure 3. Predicted uncertainty of mean gas supply for country 1 with respect to the different disruption scenarios.

Table 4. Relevant inputs for the predicted mean gas supply for country 1.

Scenario	Relevant inputs	S_i%	ST_i%
1 (Normal)	X_6 = Peak demand of Country 1	100	100
2 (X_1 off)	X_6 = Peak demand of Country 1	80	85
	X_2 = Capacity of Source N19	7	16
	X_4 = Capacity of Source N11	1	9
3 (X_2 off)	X_6 = Peak demand of Country 1	100	100
4 (X_1, X_2 off)	X_4 = Capacity of Source N11	88	88
	X_6 = Peak demand of Country 1	9	9
	X_{18} = Capacity of DN400	3	3

is employed. The first-order (S_i) and total-order (ST_i) Sobol' indices are shown in Table 4. We recall that S_i represents the amount of variance of the predicted mean gas supply explained by the input variable while the difference ($ST_i - S_i$) represents the amount due to the interaction of the variable with the other ones. We note that only scenario 2 is subject to interactions.

The results indicate that the accuracy of the model to predict the mean gas supply to Country 1 heavily depends on the knowledge of the peak demand in that country. This is particularly crucial for scenarios 1 & 3. When source N2 is off, then the model response becomes more complicated involving sources N10 and N11. When both sources N2 and N19 are off (scenario 4), then the capacity of source N11 prevails for the mean gas volume supplied to country 1, the country peak demand (X_6) becoming less relevant. In summary, to predict accurately the mean gas volume supplied to Country 1 it is crucial to know accurately the value of the following inputs: X_6, X_4, X_2 (by order of importance) and in a less extent X_{18}.

6.3 *Results for Country 2*

The estimated PDF's of the mean volume of gas supply for the four different scenarios are not shown due to lack of space. Note that when source N19 is unavailable (scenarios 3 & 4), the predicted mean value of gas supply is much less than the demand of the country. This indicates that the source N19 is very important for the Country 2. However, the system is resilient to the failure of gas source N2 (recall that it is the highest capacity source in the region). This is due to the location of the Country 2 with respect to the sources and potential bottlenecks in certain connections. Another study of bottleneck analysis of this network has identified several bottlenecks in this network (Kopustinskas et al., 2015).

The most important input variable is the peak demand of the country 2 when the system operates normally and when Source N2 is off. Surprisingly, we note that when Source N19 fails, the capacity of the system to provide gas to Country 2 also depends on the peak demand of Country 3 and the peak demand of Country 1 (when both N2 & N19 are unavailable). This result is also a consequence of the priority algorithm implemented in the ProGasNet. Indeed, the latter supplies the closer demand nodes first before serving the other more distant nodes. Hence, if a country is far from the main sources of gas, it might be more subject to failure of gas supply.

6.4 *Results for Country 3*

For Country 3, the system behavior is simpler. The system is resilient to the failure of one of the main sources but it is incapable to provide gas to Country 3 when both main sources collapse (scenario 4). This is an important result, revealing the country vulnerability to crisis. However, the crisis simulated by scenario 4 is unlikely to happen. The prediction of volume of gas supply heavily depends on the knowledge of the value of the gas demand in the country.

7 CONCLUSIONS

The study presents an uncertainty and sensitivity exercise of the security of gas supply model implemented in the probabilistic gas network simulator ProGasNet. The selected network was an EU gas transmission network of several member states. It is based on realistic network topology and data, but due to sensitivity of information, the network is anonymised.

The study aims to identify and rank the model parameters that most significantly affect the security of supply. The main finding is that the risk of gas supply shortage in the three countries heavily depends on the accurate knowledge of different model inputs. In particular, the peak demand in each country is the key parameter whose precise estimation is very important for accurate model results.

The model was run for four different scenarios representing different disruption situations. The study confirms, the results already observed in other studies, that some disruption scenarios affect only parts of the network (e.g. specific countries) while other parts of the network are not affected. Even more, due to existing bottlenecks in the system, supply disruptions in one part of the network cannot be restored from the other part even if there is enough gas available. This clearly indicates heterogeneity of the network and the need for further infrastructure development.

Looking at each country, for Country 1, the peak demand is the most important parameter, followed by capacity of sources 19 and 11 and capacity of DN400 pipeline. This is important information for further development of the model and indicates where to target further research efforts.

For Country 2, again the peak demand is the most important parameter, followed by capacity of source 19, DN500 and DN400 pipelines. Interestingly, for some scenarios, peak demands of Countries 1 and 3 are identified as important. This can be well explained by the fact that in certain disruption scenarios, lower demand in neighbouring countries allows better supply to the other countries.

For Country 3, the supply is secured in case of scenarios 2 and 3, but in the case of unlikely scenario 4, the security of supply is threatened.

The study showed the potential and usefulness of sensitivity study applied to the ProGasNet gas network model. It has not only identified the most important parameters of the model for which more attention should be paid during the estimation process, but also provided useful insights into the simulation process by confirming, for instance, the heterogeneity of the network from the sensitivity analysis perspective.

REFERENCES

Deo N., 2008. Graph theory with applications to engineering with computer science. Prentice Hall.

EU Regulation, 2010. Regulation No. 994/2010 of the European Parliament and of the Council of 20 October 2010 concerning measures to safeguard security of gas supply and repealing Council Directive 2004/67/EC. Official Journal of the European Union, L295. 53:1–22.

EU Regulation, 2017. Regulation No. 2017/1938 of the European Parliament and of the Council of 25 October 2017 concerning measures to safeguard security of gas supply and repealing Regulation (EU) 994/2010. Official Journal of the European Union, L280, 28/10/2017, 1–56.

Kopustinskas, V. & Praks, P. 2012. Development of gas network reliability model, JRC technical report JRC78151, European Commission, Luxembourg.

Kopustinskas, V. & Praks, P. 2015. Bottleneck analysis of the gas transmission network using ProGasNet simulator. In Luca Podofillini et al. (ed.), Safety and Reliability of Complex Engineered systems; *Proc. of ESREL'2015, Zurich 7–10 September 2015.* Leiden, CRC Press/Balkema.

Praks, P., Kopustinskas, V. & Masera M. 2015. Probabilistic modelling of security of supply in gas networks and evaluation of new infrastructure. *Reliability Engineering and System Safety* 144:254–264.

Praks, P., Kopustinskas, V. & Masera M. 2017. Monte-Carlo based reliability and vulnerability assessment of a natural gas transmission system due to random network component failures. *Sustainable and Resilient Infrastructure* 2(3):97–107.

Saltelli, A. 2002. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145:280–297.

Saltelli, A., Tarantola, S., & Chan, K. 1999. A quantitative model independent method for global sensitivity analysis of model output. *Technometrics* 41(1):39–56.

Shao, Q., Younes, A., Fahs, M., & Mara, T.A. 2017. Bayesian sparse polynomial chaos expansion for global sensitivity analysis. *Computer Methods in Applied Mechanics and Engineering* 318:474–496.

Sobol', I.M. 1993. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling & Computer Experiments* 1(4):407–414.

Sobol', I.M., Turchaninov, V.I., Levitan, L.Y., & Shukman, V.B. 1992. Quasi-random Sequence Generator (Routine LPTAU51). Moscow: Keldysh Institue of Applied Mathematics, Russian Academy of Science.

Sudret, B. 2008. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety* 93(7):964–979.

Wiener, N. 1938. The homogeneous chaos. *American J. of Mathematics* 897–936.

Xiu, D., & Karniadakis, E.G. 2002. The Wiener-Askey polynomial chaos for stochastic differential equation. *SIAM J. Science of Computing* 24(2):619–644.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Application of fuzzy finite element method in addressing the presence of uncertainties

A.Y.N. Yusmye
Institute of Engineering Mathematics, Universiti Malaysia Perlis, Kampus Pauh Putra, Arau, Perlis, Malaysia

A.K. Ariffin, S. Abdullah & S.S.K. Singh
Department of Mechanical and Materials Engineering, Universiti Kebangsaan Malaysia, UKM Bangi, Selangor, Malaysia

M. Beer
Institute for Computer Science in Civil Engineering, Leibniz University, Hannover, Germany

ABSTRACT: This research works is focused on the analysis of Fuzzy Finite-Element Method (FFEM) with the present of uncertainties. In considering a major engineering science problems, like damage processes or loading in consequence of real incident, uncertainty are present. Uncertainty is due to lack of data, an abundance of information, conflicting information and subjective beliefs. With that reason, the present of uncertainties is needed to avoid for prevent the failure of the material in engineering. The goals of this study are to analyzed and determine the application of FFEM by taking into consideration of the epistemic uncertainties involved toward the single edge crack plate and beam. Since it is crucial to develop an effective approach to model the epistemic uncertainties, the fuzzy system is proposed to deal with the selected problem. Fuzzy system theory is a non-probabilistic method, and this method is most appropriate to interpret the uncertainty compared to statistical approach when the deal with the lack of data. Fuzzy system theory contains a number of processes started from converting the crisp input to fuzzy input through fuzzification process and followed by the main process known as mapping process. In mapping process stage, the combination of fuzzy system and finite element method are proposed. In this study, the fuzzy inputs are numerically integrated based on extension principle method. Obtained solutions are depicted in terms of figures and tables to show the efficiency and reliability of the present analysis.

1 INTRODUCTION

An uncertainty is define as a gradual assessment of the truth content of a proposition, doubt arises as to whether the truth content may be stated with sufficient accuracy using each of the data models in all cases. Also, uncertainties are defined as the vagueness and lack of the information or data, Farkas (2010). The element of uncertainty is one of the biggest challenges in the field of engineering. He (2007) mentioned that, in general, uncertainty can divided into three types, which are stochastic uncertainty, epistemic uncertainty and error. Stochastic uncertainty is due to variations in the system. For the epistemic uncertainty, it exists as a result of incomplete information, ignorance and lack of knowledge caused by the lack of experimental data. When compare to the error, this uncertainty is the uncertainty that can be identified due to the imperfections in the modelling and simulation.

For a several decades ago, uncertainty is modelled according to the theory of probability. Probability method is very effective in solving the problem of stochastic uncertainty, but this method is not suitable to be used to solve problem involving the lack of data. Some scholars hold that the use of non-probability methods are most appropriate to interpret the uncertainty compared to statistical approach when deal with the lack of data. The interval analysis (Zimmermann 2012), convex modelling (Tapaswini et al. 2012) and fuzzy set theory (Ozkoka & Cebi 2014) are the main categories of non-probabilistic methods. Specifically, fuzzy system is a system to be precisely defined and it applied all the theories that use the basic concept of fuzzy set theory. According to Savoia (2002), the main advantages of using the fuzzy set theory than the fuzzy probability theory is able to maintain the intrinsic random nature of the physical variables without the need for modeling the probability density function. The simple justification for fuzzy system theory is the real world is too complicated for precise explanation and description to be obtained. Fuzzy system theories are knowledge based or rule based systems. Fuzzy systems have been applied to a wide field.

Finite Element Method (FEM) has become a powerful tool in solving numerous complex scientific and engineering problems. By using FEM, the complicated structured of any materials can be discretized into a small finite element. All the elements are assembling then applying the respective requirement to obtain the output. System parameters such as geometry, material properties, external load or boundary conditions are considered as crisp value or can be defined exactly in the convectional FEM. However, rather than the particular value, it may have only the vague, imprecise and incomplete information about the variables being a result. The present of uncertainty is the biggest challenging in engineering in order to deal with uncertainty in designing the material and cannot be avoided.

Fuzzy Finite Element Method approach (FFEM) is present to deal with the uncertainty and it is the merger method of fuzzy approach with the conventional Finite Element Method (FEM). In FFEM approach, conventional FEM is used as a parent in order to obscure the data input mapping to the output data (Behera, 2012). In this paper, there are two illustrative examples which applied the fuzzy finite element method with considering the presence of uncertainties. The first illustrative example is the application of FFEM in single edge crack plate by considering a fuzzy variable of crack length. And the second one is to analyze the structural reliability on beam.

2 PRELIMINARIES

In the following paragraphs, some of the notation, definition and preliminaries which are used further in this paper discuss deeply.

Definition 1:

A fuzzy number is convex normalized fuzzy set of the real line such that

$$\mu_A = X \to [0,1] \tag{1}$$

Definition 2:

We can defined an arbitrary triangular fuzzy number as $\tilde{A} = [a_L, a_N, a_R]$ and trapezoidal fuzzy number as $\tilde{B} = [b_L, b_{NL}, b_{NR}, \mathrm{b}_R]$. The fuzzy number $\tilde{A}$ is said to be triangular fuzzy number when the membership function is given by

$$\mu_{\tilde{A}}(x) = \begin{cases} 0 & x \le a_L \\ \dfrac{x - a_L}{a_N - a_L} & a_L \le x \le a_N \\ \dfrac{a_R - x}{a_R - a_N} & a_N \le x \le a_R \\ 0 & x \ge a_R \end{cases} \tag{2}$$

The triangular fuzzy number $\tilde{A} = [a_L, a_N, a_R]$ can be transformed into interval form by using α-cut as Equation (3).

$$\begin{aligned} \tilde{A} &= [a_L, a_N, a_R] \\ &= \left[a_L + (a_N - a_L)\alpha, a_R - (a_R - a_N)\alpha\right] \end{aligned} \tag{3}$$

The fuzzy number $\tilde{B}$ is said to be trapezoidal fuzzy number when the membership function is given by

$$\mu_{\tilde{B}}(x) = \begin{cases} 0 & x \le b_L \\ \dfrac{x - b_L}{b_{NL} - b_L} & \mathrm{b}_L \le x \le b_{NL} \\ \dfrac{b_R - x}{b_R - b_{NR}} & \mathrm{b}_{NR} \le x \le b_R \\ 0 & x \ge b_R \end{cases} \tag{4}$$

The trapezoidal fuzzy number $\tilde{B} = [a_L, a_N, a_R]$ can be transformed into interval form by using α-cut as Equation (5).

$$\begin{aligned} \tilde{B} &= [b_L, b_{NL}, b_{NR}, b_R] \\ &= \left[b_L + (b_{NL} - b_L)\alpha, b_R - (b_R - b_{NR})\alpha\right] \end{aligned} \tag{5}$$

3 FUZZY FINITE ELEMENT METHOD

The fuzzy finite element method is a combination of fuzzy approach and conventional finite element method. Figure 1 shows a flow chart for a fuzzy finite element process which included the fuzzification process, mapping process, α-cut level process and ending with defuzzification process.

The procedure for the analysis by using the fuzzy finite element methods is started by transform the crisp input or real value for input into the fuzzy input uncertainty through the fuzzification process. The variable that have fuzzy uncertainty is probably the material properties, geometry of the materials, boundary condition or loading (Akapan 2001). Triangular fuzzy numbers are used for understanding the function of all parameters of fuzzy membership. This FFEM approach continues with the mapping process, where the fuzzy input will be used in the FEM model by using α -cut level. Vertex method is common numerical methods used in the implementation of the extension principle. Vertex method is a combination of interval arithmetic methods with α-cut method. The purpose of this mapping is to map two or more fuzzy input to the fuzzy output with binary combinations of as many as where is the number of fuzzy input parameters. Figure 2 showed the example of α-cut method for fourth level of alpha for each variable available.

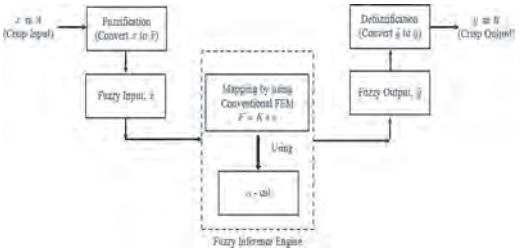


Figure 1. Flow chart for FFEM.

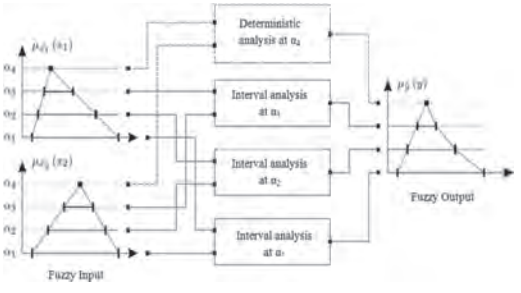


Figure 2. α- cut method for each alpha level Source: Farkas et al. (2010).

After the mapping process, the new probability distributions of the fuzzy output are created. The final stage in FFEM is defuzzification process in which stress intensity factor is the final output for this study. In defuzzification process, there are many different methods available which is center of gravity, center of area, mean of maximum, middle of maximum, fuzzy mean, fuzzy clustering defuzzification and so many. The center of gravity or centroid (COG) is the familiar defuzzification method used by many researchers and in this study. Defuzzification process is the process that transforms the fuzzy output into crisp output. This technique determines the point at which it will distribute one area into two parts which have the same value.

4 RESULT AND DISCUSSION ON ILLUSTRATIVE EXAMPLE

4.1 *A single edge crack plate by considering the fuzzy variable of crack length*

The model used in this study is Aluminium Alloy 2024-T351, since this material widely used in structural engineering such as parts of the hydraulic valve and piston, gear and shafts. In this study, three parameters that are Young's modulus, E, Poisson ratio, v and Density, ρ of Aluminium 2014-T3, are used as non-fuzzy parameters. Compare to the crack length, a is treated as a fuzzy parameter. The geometry used in this study is based on the stress analysis of crack handbook by Tada et. al (2000) as shown in Figure 3.

From analytical aspect, the stress intensity factor under mode I can be obtained by using the Equation (6) as below:

$$K_I = \sigma\sqrt{\pi a}F\left(a/b\right) \tag{6}$$

The geometry function $F\left(a/b\right)$ in equation (6) can be calculated by using the equation (7).

$$F\left(a/b\right) = 1.122 - 0.231\left(a/b\right) + 10.55\left(a/b\right)^2 - 21.7\left(a/b\right)^3 + 30.382\left(a/b\right) \tag{7}$$

where a is the crack length, b is width of the geometry, π is the constant with the value 3.1415927 and σ is the applied stress. The ratio of a/b for the geometry in Figure 3 is 0.1. The fuzzy value of crack length, a is represent in a trapezoidal fuzzy

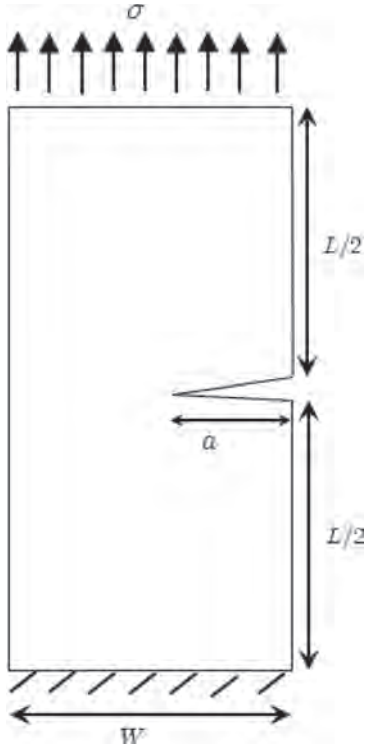


Figure 3. Two-dimensional loaded plate.

number in form as $a=(a_1,a_2,a_3,a_4)$ and written in α-cut form and finally the interval fuzzy number can written as Equation 8.

$$[a_L,a_R]=[a_1+(a_2-a_1)\alpha,a_4-(a_4-a_3)\alpha] \quad (8)$$

As mentioned early, this study only considered the fuzzy variable of crack length whereas the Young's Modulus and Poisson Ratio are considered as crisp variables.

Using fuzzy values are present the result in interval form and sketch it in trapezoidal membership function graph. The fuzzy number for crack length are $a=(0.1,0.32,0.41,0.49)$. By using Equation 6 and 7, the fuzzy intervals in term of α-cut are obtained as shown in Table 1. These fuzzy intervals are substitute into SIF formula in term of interval as

$$K_{I(R)}=\sigma\sqrt{\pi a_{(R)}}F\left(a/b\right) \quad (9)$$

$$K_{I(L)}=\sigma\sqrt{\pi a_{(L)}}F\left(a/b\right) \quad (10)$$

and continue till all SIF are obtained. Where $K_{I(L)}$) and $K_{I(R)}$) is a left and right value of interval respectively for each α-value. The output for SIF with the fuzzy crack length is shown in Table 2. Beside the result is shown in Figure 4. The cases of crack length are treating as interval at a width again increases as we increase the number of elements. In fuzzy, the large width of interval for membership functions is giving more accurate result.

The result is shown in Figure 4 depicts the SIF plot the single edge crack plate. It may be seen from the above numerical result that the natural stress (crisp values) are constant with increase in number of elements as it should be even though only two elements consider here. However it differs for values got in each element by using fuzzy method. The cases of crack length are treating as interval at a width again increases as we increase the number of elements. In fuzzy, the large width of interval for membership functions is giving more accurate result.

Table 1. Fuzzy value of crack length for each α-level.

Beta (α-Level)	a_L	a_R
0	0.1	0.49
0.1	0.122	0.482
0.2	0.144	0.474
0.3	0.166	0.466
0.4	0.188	0.458
0.5	0.21	0.45
0.6	0.232	0.442
0.7	0.254	0.434
0.8	0.276	0.426
0.9	0.298	0.418
1	0.32	0.41

Table 2. Interval value of SIF for each α-level.

Beta (α-Level)	a_L	a_R
0	6.40	14.16
0.1	7.06	14.04
0.2	7.67	13.92
0.3	8.24	13.81
0.4	8.77	13.69
0.5	9.27	13.57
0.6	9.74	13.45
0.7	10.19	13.32
0.8	10.63	13.20
0.9	11.04	13.08
1	11.44	12.95

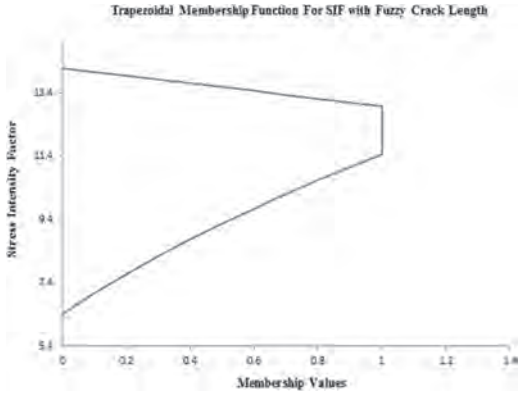


Figure 4. Trapezoidal membership function for SIF.

4.2 *Analysis on beam structure*

The analysis on structural reliability in the presence of uncertainties is performed in beam structure. In the analysis of beam structure in Figure 5, the moment of inertia and modulus of elasticity of beam are considered as fuzzy input parameters with incomplete data.

The existing data shows the moment of inertia is a normal distribution with mean value $6.58\times10^{-5}m^4$ and constant variance, COV of 0.1, while the elastic modulus has a normal distribution with mean and COV value of 0.1 and 73.1 GPa respectively. The uniformly distributed load applied to the beam is considered as an input that has no data. Thus, the opinion of an experienced

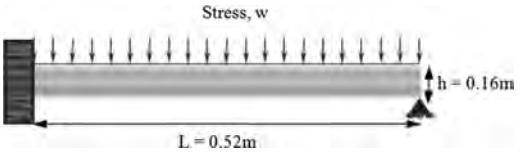


Figure 5. The geometry of the beam structure.

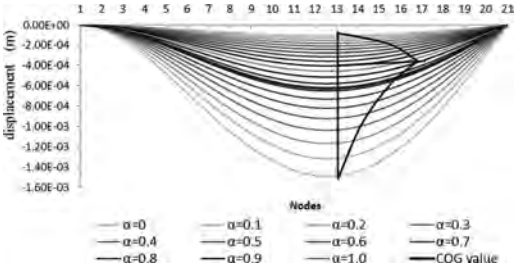


Figure 6. Fuzzy outputs of displacement using COG method.

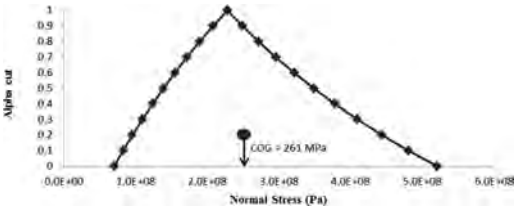


Figure 7. Fuzzy outputs of normal stress using COG method.

specialist should be considered to determine the likely distribution of that load. In this study, the distributed load is considered as a normal distribution with mean value 5.55 MN/m and COV value of 0.2. The base area of membership function for the three fuzzy input parameters considered to have the interval of 6 standard deviation and σ normal distribution, where 99% of the distribution included in the analysis.

In general, the deformation and stress in the beam plays an important role in determining the reliability of the structure. The output at each node in the FFEM is in membership function form. A COG value at each node is calculated with the centroid techniques used in the defuzzification technique. The maximum displacement is 0.00062 m at node 13. COG values in Figure 6 and Figure 7 represent the FFEM results considering uncertainties in the input parameters. In the reliability analysis of beam structures, the critical parameter is the stress. The COG value for the maximum normal stress is at node 1 with the value 261 MPa and this value is less than the yield strength of a material. The results showed that the reliability of the structure under this illustrative example is 0.9733, which mean it is close to 1. Therefore, the beam structure is still in a safe region of the elastic deformation range. Besides that, the maximum normal stress value for conventional FEM is 228 MPa. From this, it shown that the FFEM method was produced more conservative value of reliability compare to conventional FEM method.

5 CONCLUSION

Obtained solutions are depicted in terms of figures and tables to show the efficiency and reliability of the present analysis. This study shows that, the FFEM approach is conservative method to solve the uncertainties problems. One way to reduce the uncertainty in the data is by experiment. FFEM method developed does not require a large amount of data. FFEM method only required data to determine the profile or shape of membership function. The data can usually be obtained from the opinion of expert knowledgeable in the analysis associated with inductive reasoning or a genetic algorithm. Modeling input in the form of membership function, effectively involving the epistemic uncertainty in the analysis. Although both these illustrative example are based on fuzzy approach, but the presence of a finite element method in FFEM approach allows us to use easily analysis toward the complex structure or component. In addition, the factor that affects the developed simulation of FFEM is the number of fuzzy parameters. The more obscure parameters involved in the simulation, the most conservative result of the analysis. The important decision by using fuzzy approach is the large width of interval for membership functions will give more accurate in result. The trapezoidal and triangle membership function are considered for these two illustrative examples. The result is compare and it worth mentioning that by using fuzzy value given better result in term of interval width of the membership function and also reliability values.

REFERENCES

Akpan, U.O., Koko, T.S., Orisamolu, I.R. & Gallant, B.K. 2001. Practical fuzzy finite element analysis of structures. *Finite Elements in Analysis and* Design 38(2):93–111.

Behera, D. & Chakraverty, S. 2012. A new method for solving real and complex fuzzy system of linear equations. Computational Mathematics and Modelling (23), 507–518.

Farkas, L., Moens, D., Vandepitte, D. & Desmet, W. 2010. Fuzzy finite element analysis based on reanalysis technique, *Structural Safety*, 32(6): 442–448.

He, L.P., Huang, H.Z., Du, L., Zhang, X.D. & Miao, Q. 2007. A review of possibilistic approaches to reliability analysis and optimization in engineering design, *Human Computer Interaction* 4553: 1075–1084.

Ozkoka, M. & Cebi, S. 2014. A fuzzy based assessment method for comparison of ship launching methods. *Journal of Intelligent and Fuzzy Systems* 26: 781–791,

Savoia, M. 2002. Structural reliability analysis through fuzzy number approach, with application to stability. *Computers and Structures* 80(12): 1087–1102.

Tapaswini, S. & Chakraverty, S. 2014. Non-probabilistic uncertainty analysis of forest fire model by solving hyperbolic reaction-diffusion equation. *Fire safety journal*, 66: 8–14.

Zimmermann, H.J. 2010. An application-oriented view of modeling uncertainty, European journal of operational research 122(2): 190–198.

Zimmermann, H.J. 2001. Fuzzy set theory and its application, Kluwer academic publisher.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Application of evidential network to model uncertainty in quantitative risk assessment of Natech accidents

N. Khakzad & P.H.A.J.M. van Gelder
Safety and Security Science Group, Faculty of Technology, Policy, Management, TU Delft, The Netherlands

ABSTRACT: Natech is a technological accident which is triggered by a natural disaster. Increasing frequency of natural disasters along with an increasing growth of industrial plants are bound to increase the risk of Natechs in the future. Due to a lack of accurate field observations and empirical data, risk assessment of Natechs has largely been reliant on experts opinion and thus prone to epistemic uncertainty in addition to aleatory uncertainty originating from randomness of natural disasters. Evidential Network (EN) is a directed acyclic graph based on Dempster-Shafer Theory to explicitly model the propagation of epistemic uncertainty in system safety and reliability assessment. In the present study, we have illustrated an application of EN to handling epistemic uncertainty in risk assessment of flood-induced floatation of storage tanks.

1 INTRODUCTION

Technological accidents which are triggered by natural disasters such as earthquakes, lightning, storms, wildfires, tsunamis, and floods are known as Natechs. Natural disasters have reportedly led to the release of significant amounts of oil, chemicals, and radiological substances (Showalter & Myre 1994, Rasmussen 1995, Young et al. 2004).

The occurrence of Natechs in industrial plants, particularly oil terminals, can result in catastrophic consequences in terms of large spillage of petroleum products. In 2005, the floods triggered by the Hurricane Katrina in the U.S. caused a spillage of ~ 8 million gallons of oil into the ground and waterways. In August 2017, the Hurricane Harvey in the U.S. caused damage to storage tanks in refineries and petrochemical plants, leading to a substantial release of pollutants. The structural damage caused by natural events, however, does not compare with the environmental damage and revenue losses due to interruption in production and supply chain: the Hurricane Harvey made oil refineries shut down as for safety precautions, leading to at least a loss of more than 1 million barrels of oil per day in refining capacity (CNBC, 2017).

Natechs has been recognized in quantitative risk assessment of industrial plants by many researchers (Young et al. 2004, Godoy 2007, Cruz & Okada 2008, Antonioni et al. 2009, Haptmanns 2010, Krausmann et al. 2011, Landucci et al. 2012, Necci et al. 2013, Marzo et al. 2015, Mebarki et al. 2016, Khakzad & van Gelder 2017, 2018, Kameshwar & Padgett 2018). The scarcity of historical data, especially data with sufficient resolution and accuracy, has made the majority of previous studies reliant on analytical or simulative techniques (e.g., finite element modeling) in modeling and calculating the probability of failure modes. This mostly has been carried out based on modeling the envisaged failure mechanisms as a function of loads exerted by natural disasters (e.g., impact of tsunami wave) and the resistance of impacted vessels.

The stochastic features of natural disasters as well as randomness of failure mechanisms are naturally modeled via probability density functions. Either the types or the parameters of such density functions are usually estimated based on insufficient (either amount or accuracy) objective data. This lack of objective data is usually tried to be compensated for by experts opinion based upon their experience, knowledge, and even intuition, inevitably introducing degrees of epistemic uncertainty into the analysis.

The Evidence Theory (Dempster-Shafer Theory, DST), originally initiated by Dempster (1967) and further developed by Shafer (1976), is an effective tool to handle imprecise probabilities and reasoning under epistemic uncertainty. According to DST, all the possible states (mutually exclusive and collectively exhaustive) of a system is presented in a set known as *the frame of discernment* $\boldsymbol{\Omega}$. To each subset of Ω such as $\mathbf{A}$, an evidential weight $\mathbf{m(A)}$ can be assigned to indicate the degree of evidence (based on objective data or subjective opinion) in favor of the claim that a specific state in Ω belongs to A (Rakowsky 2007). Having m(A), which is also known as belief mass, the amounts of *belief* and

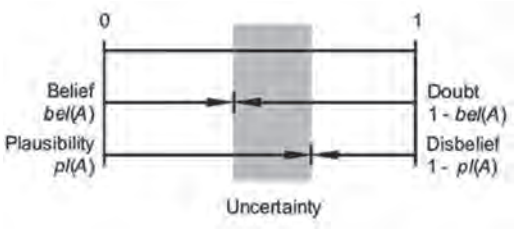


Figure 1. Quantification of epistemic uncertainty through Bel and Pls functions (Rakowsky, 2007).

plausibility of A, equivalent to lower and upper probability bounds of A, respectively, can be determined (Shafer 1976). The difference between plausibility **Pls(A)** and the belief **Bel(A)** represent the epistemic uncertainty of A (Fig. 1).

An Evidential Network (EN) is a directed acyclic graph to propagate uncertainty based on conditional belief functions (Xu & Smets 1996). Simon and Weber (2009) combined DST with Bayesian Network (Pearl, 1988) to take advantage of the junction tree algorithm developed by Jensen (1996) in propagating and computing the marginal belief functions of child nodes based upon those of their parent nodes.

The present study is an attempt to illustrate the potentiality of EN in system safety where due to lack of sufficient accurate data the analysis would be subject to epistemic uncertainty embedded in expert judgement. The application of EN will be demonstrated via safety assessment of oil storage tanks impacted by floods, with the floatation of tanks as the most common failure mode (Cozzani et al. 2010).

2 REASONING UNDER EPISTEMIC UNCERTAINTY

2.1 *Dempster-Shafer theory*

Assume that all the states of a system can be presented in a frame of discernment as Ω = {S1, S2, S3}. Accordingly, the set of all the subsets of Ω can be shown as:

$$A_i : \{\{\varnothing\},\{S1\},\{S2\},\{S3\},\{S1,S2\},\{S1,S3\},\{S2,S3\},\Omega\} \quad (1)$$

According to the available evidence (either objective or subjective), an expert may assign a belief mass to each A_i as $0 \le m(A_i) \le 1$. Each A_i for which $m(A_i) > 0$ is called a *focal set*. If all the states of the system are known, then $m(\varnothing) = 0$. Further, it must always hold that:

$$\sum_{A_i} m(A_i) = 1 \quad (2)$$

Having the belief masses determined, the belief and plausibility measures of each focal set can be defined:

$$Bel(A_i) = \sum_{B|B\subseteq A_i} m(B) \quad (3)$$

$$Pls(A_i) = \sum_{B|B\cap A_i \neq \varnothing} m(B) \quad (4)$$

$Bel(A_i)$ and $Pls(A_i)$, which are non-additive, can be taken as lower and upper probability bounds, respectively, of A_i (Simon & Weber 2009):

$$Bel(A_i) \le P(A_i) \le Pls(A_i) \quad (5)$$

$$Bel(A_i^c) = 1 - Pls(A_i) \quad (6)$$

$$Pls(A_i^c) = 1 - Bel(A_i) \quad (7)$$

where A_i^c is the complement of A_i. Having the Bel and Pls functions, the belief mass of a focal set can be determined using the möbius transformation as (Smets 2002):

$$m(A_i) = \sum_{B|B\subseteq A_i} (-1)^{|A_i - B|} Bel(B) \quad (8)$$

where $|A_i - B|$ refers to the difference between the number of elements of A_i and B.

2.2 *Evidential network*

Simon & Weber (2009) used a Bayesian network (BN) formalism to propagate imprecise probabilities using the belief mass functions assigned to the focal sets. Since the belief masses allocated to the focal sets of each component of the system add up to unity, they can be used as marginal probabilities of the nodes in the BN.

Combination of the mass belief functions of components (nodes) can readily be carried out by means of Boolean algebra. For the sake of exemplification, consider a system Z comprising two components X and Y as shown in Fig. 2.

In Fig. 2, the components and the systems are considered as binary nodes, i.e., being in one of *up* or *down* states. Thus, for instance, the frame of discernment of X and its focal sets can be presented as Ω_X = {up, down} and A_X = {{up}, {down}, {up,down}}, where {up,down} = {up}⊕{down}, respectively. Among the focal sets of X, {up,down} models the uncertainty, indicating that X can be in either up or down states. Now consider a case where

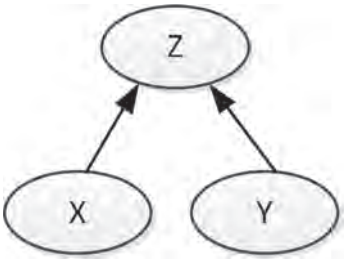


Figure 2. BN for reliability assessment of a two-component system.

Table 1. Truth table used to combine the focal sets of components X and Y via AND and OR gates (Simon & Weber, 2009).

		Z	
X	Y	AND	OR
{up}	{up}	{up}	{up}
{up}	{down}	{down}	{up}
{up}	{up,down}	{up,down}	{up}
{down}	{up}	{down}	{up}
{down}	{down}	{down}	{down}
{down}	{up,down}	{down}	{up,down}
{up,down}	{up}	{up,down}	{up}
{up,down}	{down}	{down}	{up,down}
{up,down}	{up,down}	{up,down}	{up,down}

X = {up} and Y = {up,down} are connected to Z by an AND gate; using Boolean algebra, the state of Z can be identified as {up}∩{up,down} = {up}∩{up} ⊕ {up}∩{down} = {up} ⊕ {down} = {up,down}. Likewise, in case of an OR gate, the state of Z can be identified as {up}∪{up,down} = {up}∪{up} ⊕ {up}∪{down} = {up} ⊕ {up} = {up}. The results of AND and OR gates in the form of a truth table have been presented in Table 1.

For the system shown in Fig. 2, assume that the analyst, based on his degree of belief, has assigned the marginal belief mass distributions to the focal sets of components X and Y as $m(A_X) = \{0.5, 0.4, 0.1\}$ and $m(A_Y) = \{0.4, 0.4, 0.2\}$. We in the next section will demonstrate using a case study how the belief mass distributions can be determined using Equations (2)–(8). Fig. 3 displays the resulting EN in which X and Y are connected to Z via an AND gate.

As can be seen in Fig. 3, the inference algorithm of BN can be used to calculate marginal belief mass distribution of Z based on the marginal mass distributions of X and Y and the truth table (see Table 1) as $m(A_Z) = \{0.2, 0.64, 0.16\}$. Having the belief mass distribution of Z, the belief of Z = {up} can be calculated using Equation (3):

$$Bel(\{up\}) = \sum_{B|B\subseteq up} m(B) = m(\{up\}) = 0.2.$$

This is because among the focal sets of Z, i.e., $A_Z = \{\{up\}, \{down\}, \{up,down\}\}$, only the focal set B = {up} is the subset of {up}. Likewise, the plausibility of Z = {up} can be calculated using Equation (4):

$$Pls(\{up\}) = \sum_{B|B\cap up\neq\varnothing} m(B) =$$
$$m(\{up\}) + m(\{up,down\}) = 0.2 + 0.16 = 0.36.$$

This is because among the focal sets of Z, only the intersections of focal sets {up} and {up,down} with {up} are not null. As a result: $0.2 \le P(Z = up) \le 0.36$.

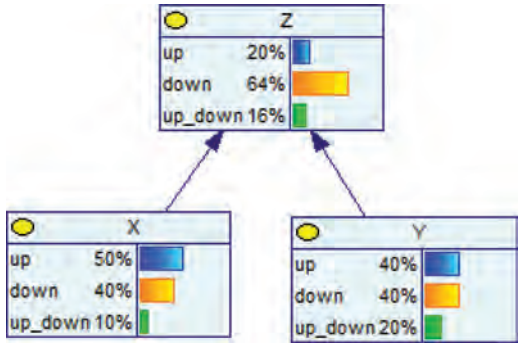


Figure 3. EN for reliability assessment of a two-component system using belief mass distributions.

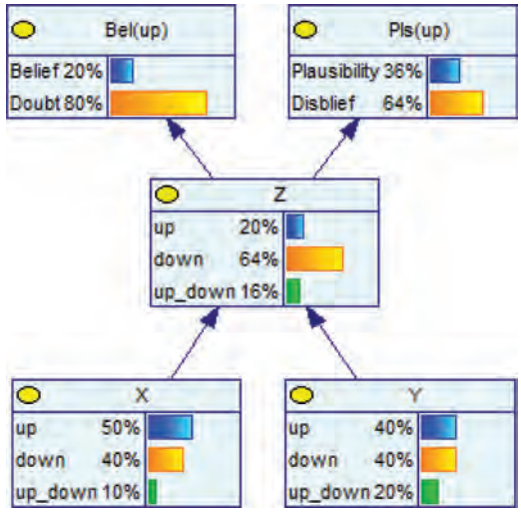


Figure 4. Adding belief (Bel) and plausibility (Pls) nodes in order to calculate epistemic uncertainty of Z = {up}.

Table 2. Conditional belief table used to calculate Bel(up) and Pls(up) of Z in Fig. 4.

Z	Bel(up)	Pls(up)
{up}	1	1
{down}	0	0
{up,down}	0	1

The procedure of calculating belief and plausibility can be carried out using the developed BN (which in fact is an EN) by adding the nodes Bel({up}) and Pls({up}) to the network (Fig. 4). The conditional belief table used to connect these two nodes to node Z is presented in Table 2. It should be noted that since Bel and Pls are non-additive (see Equations (6) & (7)), they have been presented as two separate nodes in the EN.

3 SAFETY ASSESSMENT OF STORAGE TANKS IN CASE OF FLOOD

3.1 *Floatation of storage tanks*

Floatation of storage tanks has reportedly been the most frequent failure mode during floods (Cozzani et al. 2010). Floatation of storage tanks occurs if the upthrust force of flood exceeds the bulk weight of the storage tank (weight of the tank plus the weight of its liquid containment). Fig. 5 presents the loading force (buoyancy) and resisting forces (bulk weight of the tank) contributing to the flotation of the storage tank.

When there is a lack of field or experimental data to relate the characteristics of the natural disaster to the failure modes and failure probabilities of an impacted equipment, one may choose to develop Limit-State Equations (LSE) based on influential loading and resisting forces. Development of LSEs helps the analyst combine his knowledge (though incomplete) of the influential parameters with available objective data to compensate for the inadequacy of objective data required for estimation of failure probabilities.

As for the floatation of storage tanks, the relevant LSE should take into account the weight of the tank W_T, the weight of the contained liquid W_L, and the buoyant force F_B. As can be seen from Fig. 5, we have considered a self-anchored storage tank (not bolted to the foundation) which is a common practice in case of atmospheric storage tanks. As such, the only resisting forces against the tank's floatation comprise the bulk weight of the tank. Considering the direction of the forces in Fig. 5, the LSE can be developed as:

$$LSE = F_B - W_T - W_L \tag{9}$$

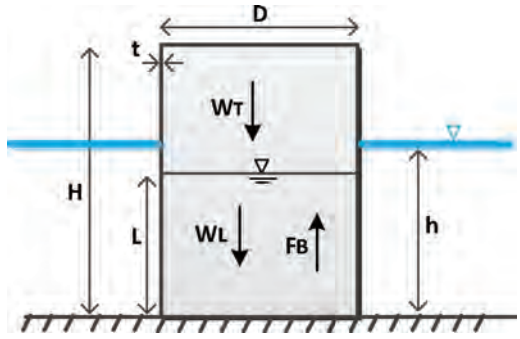


Figure 5. Schematic of the load-resistance forces considered for tank floatation.

$$F_B = \rho_w g \frac{\pi D^2}{4} h \tag{10}$$

$$W_T = \rho_s g \left(\pi DH + 2\frac{\pi D^2}{4} \right) t \tag{11}$$

$$W_L = \rho_l g \frac{\pi D^2}{4} L \tag{12}$$

where D: tank's diameter, H: tank's height, t: tank shell's thickness, L: height of liquid inside the tank, h: height of flood's inundation, ρ_w: flood water density, ρ_s: tank shell's density, ρ_l: liquid's density, and g: gravitational acceleration. Accordingly, the floatation probability of the tank can be presented as P(LSE > 0).

3.2 *Failure analysis*

For the sake of exemplification, assume that the analyst, based on objective data, would know the amounts of the tank's and flood's parameters as listed in Table 3, except the initial amount of chemical liquid (gasoline in this example).

Floatation of a storage tank due to slow submersion can result in an instantaneous release of liquid should the tank collapse, continuous release of the entire containment in a limited time in case of full disconnection of large pipelines, or a minor release in case of partial disconnection of flanges and pipelines (Cozzani et al 2010). In any of these release scenarios, if the initial inventory of the tank was not known before the floatation, the estimation of a priori inventory of the tank would be subject to uncertainty (epistemic).

Since the analyst would have doubts about the initial inventory of the tank before the flood impacted the plant, he decides to seek the opinion of two experts (e.g., operators working at the storage tank area). The first expert comes up with

Table 3. Parameters used for risk assessment of floatation.

Parameter	Value
H (m)	6
D (m)	10
t (m)	0.01
h (m)[¥]	N ($\mu = 1, \sigma = 0.2$)
ρ_s (kg/m^3)	7900
ρ_w (kg/m^3)	1024
ρ_l (kg/m^3)[§]	850

[¥] due to aleatory uncertainty inherent in flood's forecast.
[§] gasoline has been considered as the chemical liquid.

P(L = 0.5 m, L = 1.0 m, L = 1.5 m) = (0.2, 0.5, 0.3) whereas the second expert with P(L = 0.5 m, L = 1.0 m, L = 1.5 m) = (0.5, 0.3, 0.2). As such, the experts' uncertainty about the initial inventory of the storage tank can be expressed using imprecise probabilities as:

$$\begin{cases} 0.2 \le P(L = 0.5) \le 0.5 \\ 0.3 \le P(L = 1.0) \le 0.5 \\ 0.2 \le P(L = 1.5) \le 0.3 \end{cases} \quad (13)$$

According to the parameters in Table 3, the tank's weight, the weight of liquid containment, and the buoyancy force can be calculated as $W_T = 219$ (KN), $W_L = 655$ L (KN), and $F_B = 789$ h (KN), respectively. The floatation probability can thus be calculated as:

$$P(LSE > 0) = P(F_B > W_T + W_L) = P(789h > 219 + 655L) = P\left(h > \frac{219 + 655L}{789}\right) \quad (14)$$

3.3 *Uncertainty modeling*

Considering L as an uncertain variable with three states as L1 = 0.5 m, L2 = 1.0 m, and L3 = 1.5 m, its frame of discernment would be:

Ω_L = {L1, L2, L3}.

Consequently, the set of its focal sets would be:

A_L: {{L1}, {L2}, {L3}, {L1, L2}, {L1, L3}, {L2, L3}, {L1, L2, L3}}.

Using the equations in Section 2.1, the belief mass of each focal set can be determined. For example, consider the first focal set, {L1} with the lower and upper bound probabilities as shown in Equation (13). Based on Equation (5), Bel({L1}) = 0.2 and Pls({L1}) = 0.5. Since {L1} is a singleton, using Equation (8), m({L1}) = Bel({L1}) = 0.2. Similarly, m({L2}) = 0.3, and m({L3}) = 0.2.

As another example, consider the focal set {L1, L2}. Since {L1}, {L2}, and {L1, L2} are all the subsets of {L1, L2}, using Equation (8), we will have m({L1, L2}) = Bel({L1, L2}) – Bel({L1}) –Bel({L2}).

Further, based on Equation (6), Bel({L1, L2}) = 1 – Pls({L3}) = 1 – 0.3 = 0.7.

As a result, m({L1, L2}) = 0.7 – 0.2 – 0.3 = 0.2. Following the same procedure, m{{L1}, {L2}, {L3}, {L1, L2}, {L1, L3}, {L2, L3}, {L1, L2, L3}} = (0.2, 0.3, 0.2, 0.2, 0.1, 0, 0). Since m({L1, L3}) = m({L1, L2, L3}) = 0, they would not be considered as focal sets any more.

3.4 *Probability of floatation*

As can be seen from Equation (14), the only influential parameters in estimating the probability of floatation are the flood inundation height h and the liquid containment height L. To facilitate the propagation of uncertainty – aleatory uncertainty in h and epistemic uncertainty in L – the EN in Fig. 6 can be developed. It is worth noting that compared to the EN proposed by Simon & Weber (2009), in our EN both the belief and plausibility of *Floatation* have been modeled using a single node, considering the fact that:

$$Bel(A_i) + Unc(A_i) + Dis(A_i) = 1.0 \quad (15)$$

$$Unc(A_i) = Pls(A_i) - Bel(A_i) \quad (16)$$

$$Dis(A_i) = 1 - Pls(A_i) \quad (17)$$

where $Unc(A_i)$ and $Dis(A_i)$, respectively, refer to the uncertainty and disbelief about the focal set A_i (see Fig. 1).

In the EN shown in Fig. 6, the states of the node L have been represented by its focal sets with the respective belief masses as marginal probabilities (although belief masses are not probabilities, as discussed in Rakowsky (2007)). As opposed to the node L, the states of the node h are the discretized intervals of h with their (real) marginal probabilities calculated based on the normal distribution presented in Table 3. In this regard:

$$h = \begin{cases} h1 & if \quad 0 \le h < 0.8 \\ h2 & if \quad 0.8 \le h < 1.2 \\ h3 & if \quad 1.2 \le h < 1.6 \\ h4 & if \quad 1.6 \le h < 2.0 \end{cases}$$

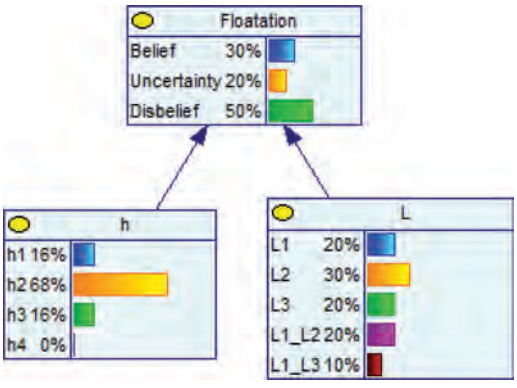


Figure 6. Evidential network to estimate the probability of floatation under aleatory and epistemic uncertainty.

The conditional belief table used to calculate the marginal masses of the node Floatation has been shown in Table 4. The conditional masses can readily be calculated using Equation (14). As an example, consider the combination of h2: $0.8 \le h < 1.2$ with the states (focal sets) of L:

- Case 1

{L1}: L = 0.5

$$P\left(h2 > \frac{219 + 655L1}{789}\right) = P\left(h2 > 0.69\right) = 1.0$$

Since h is always greater than 0.69 (note $0.8 \le h < 1.2$), the belief and plausibility of the floatation, as the lower and upper bounds of probability, are both equal to 1.0. This, in turn, yields a zero uncertainty (see Equation (16)) and a zero disbelief (see Equation (17)). See the 6th row in Table 4.

- Case 2

{L2}: L = 1.0

$$P\left(h2 > \frac{219 + 655L2}{789}\right) = P\left(h2 > 1.11\right) = P\left(1.11 < h < 1.2\right) = 0.133$$

As a result, Bel = Pls = 0.133, Unc = 0.0, and Dis = 0.867 (7th row in Table 4).

- Case 3

{L3}: L = 1.5

$$P\left(h2 > \frac{219 + 655L3}{789}\right) = P\left(h2 > 1.52\right) = 0.0$$

Since h is always smaller than 1.2 (note $0.8 \le h < 1.2$), the belief and plausibility of the floatation, as the lower and upper bounds of probability, are both equal to 0.0. This, in turn, yields a zero uncertainty and a disbelief of unity (8th row in Table 4).

- Case 4

{L1, L2}: L = 0.5 or 1.0

From Case 1 (L = 0.5) and Case 2 (L = 1.0), the probabilities of floatation were calculated as 1.0 and 0.133, respectively. Accordingly, the lower probability can be taken as Bel = 0.133 whereas the upper probability as Pls = 1.0. This in turn will result in Unc = 0.867 and Dis = 0.0 (9th row in Table 4).

- Case 5

{L1, L3}: L = 0.5 or 1.5

From Case 1 (L = 0.5) and Case 3 (L = 1.5), the probabilities of floatation were calculated as 1.0 and 0.0, respectively. Accordingly, the lower probability can be taken as Bel = 0.0 whereas the upper probability as Pls = 1.0. This in turn will result in Unc = 1.0 and Dis = 0.0 (10th row in Table 4).

As can be seen from Fig. 6, given the marginal and conditional probabilities and belief masses, the lower bound probability of floatation has been calculated as Bel = 0.3. Similarly, the amount of uncertainty has been calculated as Unc = 0.2, which together with the amount of belief results in an upper bound probability of floatation as Pls = 0.3 + 0.2 = 0.5.

Modeling the uncertainty of the floatation in a single node instead of two nodes (cf Fig. 4) comes

Table 4. Conditional belief mass distribution for the node Floatation in Fig. 6.

Index	h	L	Bel	Unc	Dis
1	h1	{L1}	0.098	0	0.902
2	h1	{L2}	0	0	1
3	h1	{L3}	0	0	1
4	h1	{L1,L2}	0.098	0	0.902
5	h1	{L1,L3}	0.098	0	0.902
6	h2	{L1}	1	0	0
7	h2	{L2}	0.133	0	0.867
8	h2	{L3}	0	0	1
9	h2	{L1,L2}	0.133	0.867	0
10	h2	{L1,L3}	0	1	0
11	h3	{L1}	1	0	0
12	h3	{L2}	1	0	0
13	h3	{L3}	0.003	0	0.997
14	h3	{L1,L2}	1	0	0
15	h3	{L1,L3}	0.003	0.997	0
16	h4	{L1}	1	0	0
17	h4	{L2}	1	0	0
18	h4	{L3}	1	0	0
19	h4	{L1,L2}	1	0	0
20	h4	{L1,L3}	1	0	0

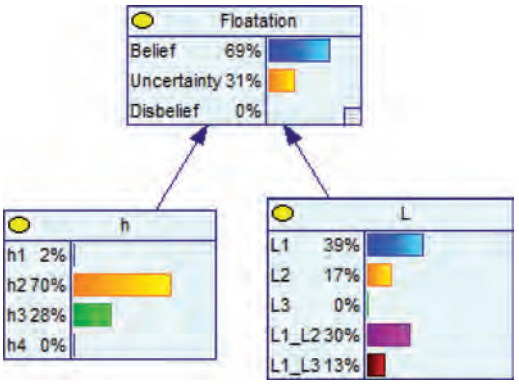


Figure 7. Updated belief masses by means of Disbelief = 0 as soft evidence.

in handy in reasoning about a priori inventory of the storage tank via belief updating. For instance, in case the storage tank is believed to have lost some of its containment as a matter of floatation, the initial belief masses assigned to L can be updated by instantiating the amount of Disbelief to zero (Fig. 7).

The updated belief masses have been depicted in Fig. 7, where L1 (i.e., L = 0.5 m) is believed to be the likeliest amount of initial inventory before the floatation.

4 CONCLUSIONS

In the present study we examined the applicability of evidential networks to system safety under both aleatory and epistemic uncertainties.

Modeling epistemic uncertainty of a parameter in a single node as an aggregation of the degrees of belief, uncertainty, and disbelief, makes it possible to perform belief updating by using a variety hard and soft evidence. We demonstrated the application of evidential networks to assess the vulnerability of storage tanks against flood-induced submersion. However, the methodology, without a loss of generality, can be applied to system safety and reliability assessment in a wide variety of domains.

REFERENCES

Antonioni, G., Bonvicini, S., Spadoni, G., Cozzani, V. 2009. Development of a frame work for the risk assessment of Na-Tech accidental events. *Reliability Engineering and System Safety* 94: 1442–1450.

CNBC (2017). Major refineries are shutting down in the wake of Harvey flooding. 27 Aug 2017. Available at https://www.cnbc.com/2017/08/27/hurricane-harvey-refineries-shutting-down.html. Last checked Dec. 14, 2017.

Cozzani, V., Campedel, M., Renni, E., Krausmann, E. 2010. Industrial accidents triggered by flood events: analysis of past accidents. *Journal of Hazardous Materials* 175: 501–509.

Cruz, A.M., Okada, N. 2008. Consideration of natural hazards in the design and risk management of industrial facilities. *Natural Hazards* 44: 213–227.

Dempster, A.P. 1967. Upper and lower probabilities induced by a multivalued mapping. The Annals of Mathematical Statistics 38: 325–339.

Godoy, LA. 2007. Performance of storage tanks in oil facilities damaged by Hurricanes Katrina and Rita. *Journal of Performance of Constructed Facilities* 21(6): 441–449.

Hauptmanns, U. 2010. A decision-making framework for protecting process plants from flooding based on fault tree analysis. *Reliability Engineering and System Safety* 95: 970–980.

Jensen, F. 1996. *An Introduction to Bayesian Networks*. UCL Press.

Kameshwar, S., Padgett, JE. 20158. Storm surge fragility assessment of above ground storage tanks. *Structural Safety* 70: 48–58.

Khakzad, N., van Gelder, P. 2017. Fragility assessment of chemical storage tanks subject to floods. *Process Safety and Environmental Protection* 111: 75–84.

Khakzad, N., van Gelder, P. 2018. Vulnerability of industrial plants to flood-induced natechs: A Bayesian network approach. *Reliability Engineering and System Safety* 169: 403–411.

Krausmann, E., Renni, E., Campedel, M., Cozzani, V. 2011. Industrial accidents triggered by earthquakes, floods and lightning: lessons learned from a database analysis. *Natural Hazards* 59: 285–300.

Landucci, G., Antonioni, G., Tugnoli, A., Cozzani, V. 2012. Release of hazardous substances in flood events: damage model for atmospheric storage tanks. *Reliability Engineering and System Safety* 106: 200–216.

Marzo, E., Busini, V., Rota, R. 2015. Definition of a short-cut methodology for assessing the vulnerability of a territory in natural–technological risk estimation. *Reliability Engineering and System Safety* 134: 92–97.

Mebarki, A., Jerez, S., Prodhomme, G., Reimeringer, M. 2016. Natural hazards, vulnerability and structural resilience: tsunamis and industrial tanks. Geomatics, *Natural Hazards and Risk* 7(S1): 5–17.

Necci, A., Antonioni, G., Cozzani, V., Krausmann, E., Borghetti, A., Nucci, C.A. 2013. A model for process equipment damage probability assessment due to lightning. *Reliability Engineering and System Safety* 115: 91–99.

Pearl, J. 1988. Probabilistic reasoning in intelligent systems. San Francisco, CA: Morgan Kaufmann.

Rakowsky, U.K. 2007. Fundamentals of the Dempster-Shafer theory and its applications to system safety and reliability modelling. *In Proc. of the ESRA Summer Safety and Reliability Seminar – SSARS 2007*, Sopot, Poland, July 2007.

Rasmussen, K. 1995. Natural events and accidents with hazardous materials. *Journal of Hazardous Materials* 40: 43–54.

Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.

Showalter, P.S., Myers, M.F. 1994. Natural disasters in the United-States as release agents of oil, chemicals, or radiological materials between 1980 and 1989. *Risk Analysis* 14: 169–181.

Simon, C., Weber, P. 2009. Evidential networks for reliability analysis and performance evaluation of systems with imprecise knowledge. *IEEE Transactions on Reliability* 58(1): 69–87.

Smets, P. 2002. The application of matrix calculus to belief functions. *International Journal of Approximate Reasoning* 31: 1–30.

Young, S., Balluz, L., Malilay, J. 2004. Natural and technologic hazardous material releases during and after natural disasters: a review. *Science of the Total Environment* 322: 3–20.

Dynamic risk and barrier management

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Towards an online risk model for dynamic positioning operations

Anna Yining Dong, Jan Erik Vinnem & Ingrid Bouwer Utne
Department of Marine Technology, NTNU, Norway

ABSTRACT: Automation and increasing complexity mean that operators have to handle data and alarms and emergent decisions under the pressure of unexpected and rapidly changing hazardous situations. Position loss during marine operations may lead to serious accidents, such as collision, loss of well integrity, etc. An online risk model aims at assisting operators in dynamic positioning operations to successfully recover the vessel's position in a good timing. The objective of this paper is to identify generic scenarios of position loss during operational phase and the information that is needed for successful recovery action. The results show that position loss normally involves of complex human machine interactions, generally in two patterns. Based on the findings, it has been recognized that risk model considering time aspect is of vital importance to develop an online risk model for DP operations.

1 INTRODUCTION

A dynamically positioned (DP) vessel is by the International Maritime Organization (IMO) defined as a vessel that maintains its position and heading (fixed location or pre-determined track) exclusively by means of active thrusters (1994). A DP system generally consists of three main subsystems, i.e., the power system, thruster system and DP control system. To further measure the designed equipment redundancy of the DP system, the IMO MSC Circ. 645 (1994) defines three classes, i.e., DP 1, DP 2 and DP 3. For DP class 1, position loss may occur given a single failure event of an active component. For DP class 2, position loss should not occur given a single failure, and for DP class 3, position loss should not occur given a single failure, including fire and flooding of watertight compartment or fire subdivision.

Based on more than two-decades of experience with safety management of DP marine operations, it has been shown that risk of position loss is intrinsic to all DP vessels (Chen and Nygård 2016). A position loss may happen on DP 1 vessels, as well as on DP 2 and DP 3 vessels. Meanwhile, offshore exploration and exploitation of hydrocarbons have opened up an era of DP vessels. There are wide applications of DP vessels in the offshore oil and gas industry, e.g., diving support vessels, pipe-layers, heavy lifting vessels, drilling rigs, subsea construction vessels, platform support vessels, shuttle tankers, etc. The focus of this paper is on offshore loading operations using DP shuttle tankers.

Nowadays, most liquid products (i.e., stabilised crude oil, condensate, liquefied petroleum gas, liquefied natural gas, etc.) from oil and gas fields in the North Sea are transported to refineries and terminals by DP shuttle tankers (ST). These large vessels with high thrust and power capacity may pose a significant collision risk to an adjacent offshore installation in case of position loss. Since 2000, there have been two collisions between shuttle tankers and facilities on the Norwegian Continental Shelf (NCS). In addition, there have been four near misses (collision events) and seven incidents related to loss of position, with varying degrees of severity.

There are two generic failure modes of position loss, i.e., drive-off and drift-off. The primary concern in this paper is on the drive-off scenario. The term *drive-off* is defined as a tanker moving away from its own target/desired position by its own power in off-loading operations. It might occur in different phases during offloading operations, i.e., approach, connection, loading and disconnection. Excessive relative motions between the FPSO (floating production storage and offloading) and tanker, categorized in surging and yawing modes, have been identified as the failure prone situations (drive-off) in tandem offloading (Chen 2003). Several recommendations have been given by (2003) regarding how to reduce the occurrence of excessive surging and yawing events. For instance, the coordination of mean heading control between the FPSO and tanker is important to minimize the probability of excessive yawing.

A recent review of DP accidents and incidents by the Petroleum Safety Authority in Norway (PSA) has shown that there is an increasing tendency in the number of DP drive-off accidents and incidents during offshore loading with shuttle tankers on the NCS in the past fifteen years (Kvitrud, Kleppestø

et al. 2012). Vinnem et al. (2015) indicate the need for new risk reduction measure and outline an overall concept for online risk management. A research project has been initiated by the Department of Marine Technology in the Nowegian University of Science and Technology (NTNU). The main goal of the project is to develop an online risk monitoring and decision support system.

An analysis was performed on DP accidents and incidents with emphasis on root causes and barrier failures (Dong, Rokseth et al. 2017). One of the major conclusions was that the most recently drive-off accidents and incidents on NCS involve both technical and human/operational failures. The development of DP operator (DPO) decision support should focus on reducing the combination of causes. Five design principles for the online risk model, including complementarity, integration, early detection, early warning, and transparency were proposed by Hogenboom et al. (2017). Moreover, it is likely that an online risk management system may reduce the risk due to human machine interface (HMI) failures. Automation and increasing complexity mean that DPOs have to handle data and alarms and take safety-critical decisions under the pressure of unexpected and rapidly changing hazardous situations. A previous study shows that human error is the most complex and least understood factor in the failures of complex systems, accounting for as much as 60% to 80% of complex system failures (Sudano 1994). As an additional barrier function (Vinnem, Utne et al. 2015), the new decision support system should aim at supporting information to operators, reducing the potential for catastrophes induced or exacerbated by human errors. The complex HMI determines the importance of information support for operators' decision-making.

The objective of this paper is to identify the human action and types of technical failures in the initiating event of drive-off. The purpose is to find out the challenges and problems for early detection during DP operations, which online risk model and decision support system can provide the information to contribute to improve DPOs' situation awarness and decision making, and understanding of system performance as well.

The paper is structured as follows: the challenges of DPO decision-making are stated in Section 2, where classifications of decision and risk information are also introduced. In Section 3, human actions in the initiating event of drive-offs and classifications of failures are presented, following by the description of analysis and result in Section 4. Based on the analysis and result, discussions are given in Section 5. Lastly, conclusions are summarized in Section 6.

2 DP OPERATOR DECISION-MAKING, CHALLENEGES, TYPES OF DECISION AND RISK INFORMATION

An information-decision-execution model (Figure 1) for DPO reaction in a drive-off scenario was introduced by Chen (2003). One important factor and dimension that needs to be under control is the *time*. As illustrated in Figure 1, the model is presented with the time reference. It is worth noting that the three stages (Ta; Td; T1) do not happen in a purely linear, sequential manner. The estimation of the DPO action initiation time (T1) is, accordingly, based on an estimation of the following three characteristic time interval values, as shown in Figure 1:

- Information time: 0-Ta
- Decision time: Ta-Td
- Execution time: Td-T1

Chen (2003) states that the challenge of human intervention is that the DPO needs to make a decision within typically 45 seconds to avoid a collision in the case of drive-off, given that the typical distance between vessels is 75 m. Some previous studies show that DPOs when collisions occurred, used about 3 minutes before taking manual evasive action.

Loss of situation awareness has been recognized as the main reason for no early detection (Chen 2014). Situation awareness (SA) is defined as "the perception of the elements in the environment within volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley 1995). Moreover, Endsley (1995) developed a three-level model to describe the different levels involved in the formation of SA. Level 1, perception, refers to the perception of attributes and dynamics of elements in the environment. Level 2, comprehension, refers to the integration and understanding of the information, i.e. it involves the human operator's sense-making to establish what is happening in the situation. Level 3, projection, refers

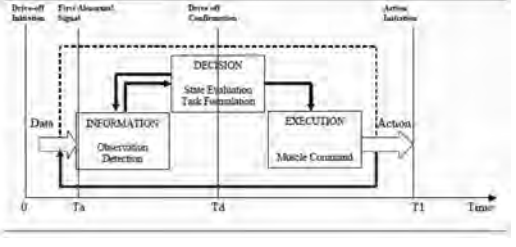


Figure 1. Information-Decision-Execution model for DP operator reaction in drive-off scenarios, adopted from (Chen 2003).

to the operator's estimation of future states of the system. The results of the assessment of the current situation can be utilized to determine future courses of action, thus supporting decision-making. However, Kjell et al. (2015) argued that the process of gaining SA does not follow sequentially from level 1 SA to level 2 SA, as set out by Endsley's model, but rather the build-up seemed to be adaptive and related to the work system's higher level goals, such as to avoid collision. He also found that in a majority of DP accidents and incidents, DPOs didn't expect the occurrence of an accident or incident. Some of the DPOs were not able to identify the relevant initiating events (lack of level 1 SA), or to understand the relevance of the initiating events (lack of level 2 SA) in DP accidents and incidents (Øvergård 2015). *Initiating event* is an identified event that upsets the normal operations of the system and may require a response to avoid undesirable outcomes (Rausand 2011). The initiating events that were used in the study include environmental impact, DP reference, human error, component failure, power management system, DP software failure (Kjell I. Øvergård 2015).

To avoid a collision and mitigate the consequence of position loss, successful human intervention has been considered as the main risk reduction measure, while efforts should be made on bridge ergonomics, HMI, alarm system, procedures and training. Meanwhile, it also needs to improve DPO's decision-making for gaining and maintaining situation awareness.

From a risk assessment perspective, decisions can be classified into planning decisions and execution decision. *Planning decision* is the decision made by blunt-end decision maker and middle level decision makers, such as operational managers. The time lag between decision and action is relative long. Enough time systematically identify and evaluate different alternatives. *Execution decision* is made by sharp-end personnel, who monitor or control ongoing operation and emergency response teams. The time lag between decision and action is much less. When DPO is in charge of making execution decision (illustrated in Figure 1), it can be further divided into *instantaneous decisions* and *emergency decisions*. Instantaneous decisions are taken spontaneously by sharp-end operators, e.g. to follow or deviate from procedures; ignore or react upon deviations in normal working conditions. The decision-making emphasizes situation assessment and pattern matching. This type of decision is normally taken quickly, although not necessarily. Emergency decisions are taken in emergencies to avoid or adapt to hazardous situations. Time dynamic is often so fast that pattern matching may not match the development of the situation. Decisions have to be made fast.

Furthermore, Yang and Haugen (2015) identify six different risk types to make the different operational decisions. To support execution (instantaneous and emergency) decisions, time-dependent action risk information is proposed. The time-dependent action risk can be estimated or predicted based on the margin between the performance of parameters in the current situation and operational limits.

In terms of early detection, focus should be on signals of deterioration of position to strengthen situation awareness during the monitoring (boredom) phases. Early warning including indicators derived from operating parameters against operating limits should faciliate early detection and reflect the operating limits and capabilities.

2.1 *Human actions in initiating event of drive-off*

To study the HMI in initiating event of drive-off, human action is used instead of human error. *Human error* is defined by Reason (1990) as all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, and when these failures cannot be attributed to the intervention of some agency. Reason (1990) emphasizes that the notion of intention and error are inseparable. Human action can be categorized as intentional action and nonintentional action. Reasons argues that human error is only associated with the intentional action, and it has no psychological meaning in relation to nonintentional behaviour. This view is also accepted in paper, although nonintentional human behaviour may contribute to system failure from safety point of view.

The human actions and their interations with technical failure events have been categorized into initiating action, response action and latent action:

- *Initiating action* is an action that initiates a failure event in the system.
- *Response action* is an action that responds to the system demands, typically under technial failure events or special external situations. Chen (2003) points out that the response action may save or worsen the situation or cause a transition to another event.
- *Latent action* is an action that influences (but does not directly initiate) the technical failure, e.g. maintenance action, and/or the above two types of human actions.

2.2 *Types of failures*

With respect to the performance of an item, it is necessary to explain the difference between failure, fault and error, a relationship between failure, fault

and error is given as follows (Rausand og Høyland 2004):

- A failure is an event that occurs at a specific point in time.
- A fault is the state of an item characterized by inability to perform a required function. While a failure is an event that occurs at a specific point in time, a fault is a state that will last for a shorter or longer period.
- The error is a discrepancy between a computed, observed, or measured value or condition and the true, specified, or theoretically correct value of condition. An error is present when the performance of a function deviates from the target performance (i.e., the theoretically correct performance), but still satisfies the performance requirement. An error will often, but not always, develop into a failure.

An illustration showing the relationship can be found in Figure 2. A failure may originate from an error. When the failure occurs, the item enters a fault state. A failure mode is always related to a required function and the associated performance requirement. A failure mode is a description of a fault (i.e., a state) and not of a failure (i.e., an event). A correct term would, therefore, be a fault mode.

In addition, failures may be classified according to their causes, effects, detectability and several other criteria. It is worth to mention a special category of cause is common-cause failures (CCFs). According to the effect of the failure, IEC 61508 (2010) classifies failures as follows:

- *Safe failure:* failure which does not have the potential to put the safety-related system in a hazardous or fail-to function state.
- *Dangerous failure:* failures has the potential to put the safety-related system in a hazardous or fail-to function state.
- *Non-critical failure:* failures where the main functions of the item are not affected.

Safe and dangerous failures may be classified further as either detected (by diagnostics) or undetected (not detected by diagnostics).

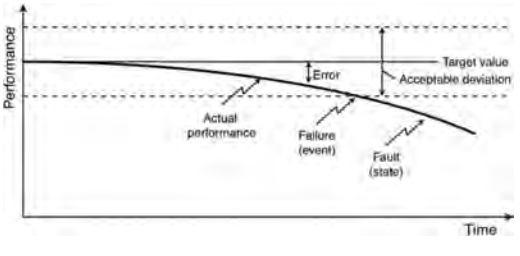


Figure 2. Difference between failure, fault and error (Rausand og Høyland 2004).

3 ANALYSIS AND RESULTS

The analysis is performed by reviewing the incident investigation reports of recently occurred DP accidents and incidents (a detailed overview can be found in (Dong, Rokseth et al. 2017)). According to the classifications of human action and failures that are stated above, the following keywords are used to analyse the accidents and incidents:

- Performance deviation
- Initiating event
- Failure (event), particularly technical failures
- Human initiating action
- Human response action

The result shows two identified situations regarding performance deviations.

Situation 1: Deviation is observed. Normal operational activity is required to perform.

Situation 2: Deviation is observed. Deviation represents the abnormal performance of the technical system.

For each situation, operator tasks, HMI, type of human action and typical technical failures are summarised in Table 1.

For the first situation (shown in Table 1), the operator needs to interact with the techncial system to perform operation activity when they observed deviations in normal working conditions. When performing the task, this situation mostly involved human initiating action. In addition, DP control logical failure has been identified as a typical technical failure that is initiated by the human action. For instance, the DPO might need to adjust ST heading to return backloading hose during disconnection. When giving the new setpoints, DPO initiated the software logic failure. However, it is challenging for the DPOs to identify the DP logic failure, since it is a type of undetected dangerous failure. Technical failures can be classified into (dangerous or safe) undetected and (dangerous or safe) detected failures. Dangerous Undetected (DU) failures are preventing activation on demand and are revealed only by testing or when a demand occurs. Sometimes, it is also called dormant failures (61508 2010).

The drive-off involving human initiating action also shows DPO lacks information for performing their task. A task analysis is conducted based on a case study of adjusting shuttle tanker position to return loading hose during disconnection. It is illustrated in Figure 3. The task analysis is made according to a six-step decision-making process (D. Husjord 2015) using in navigational training and practice. As shown in Figure 3, the main task is to adjust ST position to return loading hose

Table 1. Identification of situation that contributed to human action in DP accidents and incidents.

Situation	Description	Operator task	HMI	Type of human action	Identified technical failures	Type of technical failures
1	Deviation is observed. Normal operational activity is required to perform.	Interact with technical system to perform operation activity. i.e., adjust ST heading to return backloading hose during disconnection.	– Select an operation mode – Give new setpoints in user menu – Select DP reference orgin	Initiating action, i.e., DPO gives new setpoints.	– DP control logic failure. – Sensor failure.	Dangerous undetected
2	Deviations represent abnormal performance of technical system.	Interact with technical system to keep ST position.	Interaction depending on the type of technical failures (i.e., safe failures or dangerous failures) and causes of the technical failures.	Response action, i.e., DPO performs position drop-out to calibrate PRS	Inaccurate DP offset (s) for PRS (s) and/ or gyros deviating from true north	Safe detected

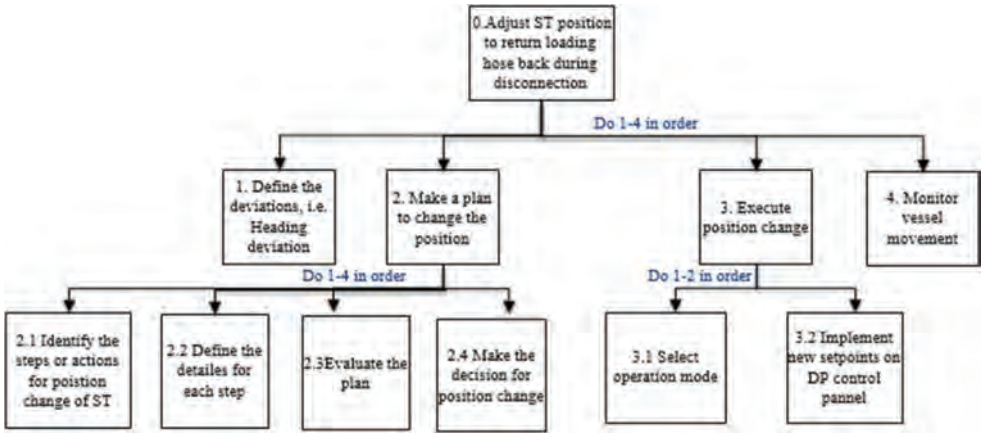


Figure 3. Task analysis of adjusting ST position to return backloading hose during disconnection.

during disconnection. To achieve it, there are a couple of steps, which are listed as follows:

Step 1: Define the heading deviations;
Step 2: Make a plan for changing the position;
Step 3: Execute the action for a position change;
Step 4: Monitor the movement of the vessel.

The operator should follow the Step 1-4 in order. However, this is not just a human task. It demands a human-machine collaboration to change the heading. While the operator decides whether the heading should be changed or not, execution of the position change needs thrust allocation, which is implemented by the DP control system. The problem is the DP control system might have limitations that are not stated in the user manual or hidden failures. Therefore, the operators need information about the real-time performance of the DP control system to avoid undesired outcome from the command they give to the control system. By referring to real-time, it means the weather conditions at the moment is also taken into account.

Regarding the second situation in Table 1, it is normally associated with technical failures. While

technical failures appear, the DPO identifies the deviations. To ensure safe operation, the DPO further identifies the risk that can be caused by the technical failures so that they can take action to respond to the failure. Nevertheless, they might misunderstand the technical failures, which means they should be able to distinguish between dangerous and safe detected failures.

- Dangerous detected failures (DD): dangerous failures that are detected immediately when they occur, for instance by an automatic built-in self-test.
- Safe detected failures (SD): Dangerous failures that are detected (normally by automatic self-testing).

The detection of technical failures and identification of dangerous or safe failures have been mainly given as the task of the automation, which is performed by *diagnostic self-testing* (Rausand og Høyland 2004). Therefore, the increasing trust of the reliability of the automatic function (i.e., automatic self-testing) may result in loss of skill of human operators. The operator needs information support for being aware that the actual performance of DP system is within acceptable deviation even though a deviation is observed.

4 DISCUSSION

Based on the classifications of human actions, initiating action and response action are identified from the recently occurred DP accidents and incidents. First, it shows that the initiating event of drive-off does have to be a technical failure. It can be a human initiating action that triggers a failure in technical system with the purpose to perform a normal operational task. While lack of information support during performing the task is identified, it also represents the deficiency of proof testing to detect dangerous undetected failure (i.e., DP control logic failure). DPO needs information for evaluation before executing the decision when time is available. Lack of information may result in loss of situation awareness and *overconfidence* of the DP system performance. Sometimes, overconfidence is referred to as *complacency*, and can have severe negative consequence if the automation is less than fully reliable (Wickens, Gordon og Liu 1997). The cause of complacency is probably an inevitable consequence of the human tendency to let experience guide our expectancies. When DP systems are marketed as quite reliable, we should avoid that the DPOs perceive the device to be of "perfect reliability". Otherwise, it becomes a natural tendency for the operator to cease monitoring its operation or at least to monitor it far less vigilantly than is appropriate. One implication of automation for human intervention related to situational awareness is that people are better aware of the state of processes in which they are actively participating in than when they are passive monitors of someone (or something) else. If they are carrying out those processes to detect a failure in an automated system, they will be less likely to intervene correctly and appropriately if they are out of the loop and do not fully understand the system's momentary state. All of this information will be essential in order to develop the risk model, which the on-line monitoring will be based on.

In addition, a drive-off event can also be triggered by an observed technical failure with interaction of human response action. Indeed, it requires the DPO to be able to analyse if it is a safe or dangerous failure. The operator needs information support for being aware whether the actual performance of DP system is within acceptable deviation or not.

To avoid DPO overconfidence in the automation and imporve their understanding of actual system performance, it is necessary with an additional supervisory system to assist operators in detection, situational awareness and skill loss.

One of the purposes for such a system is to support the DPO in two situations described in Section 4 to avoid initiating action and response action during DP operations. The information support should help DPOs to have the reference for the following questions:

- Will the action initiate a dangerous undetected failure?
- Is it a dangeorus technical failure? Will the action worsen the situation?

For the first question, the system aims to support the operator in the situation that they face deviations in normal working condition and need to perform an operational action about the deviations. For instance, if the DPO needs to adjust ST position to return loading hose during disconnection if heading deviation has been observed. This information will be used in the forthcoming development of a risk model which will be the main basis of the online risk modeling tool.

Initiation of action is a decision-making process. Therefore, we can call the new system an online decision support system. It will support operator aiming to reduce initiating action and response action. Based upon the findings, the online decision support system should address two types of information support. An illustration is given in Figure 4 to demonstrate the online decision support system will support the two types of information.

1. Support information for DPO in the task planning in the situation that DPO encounters

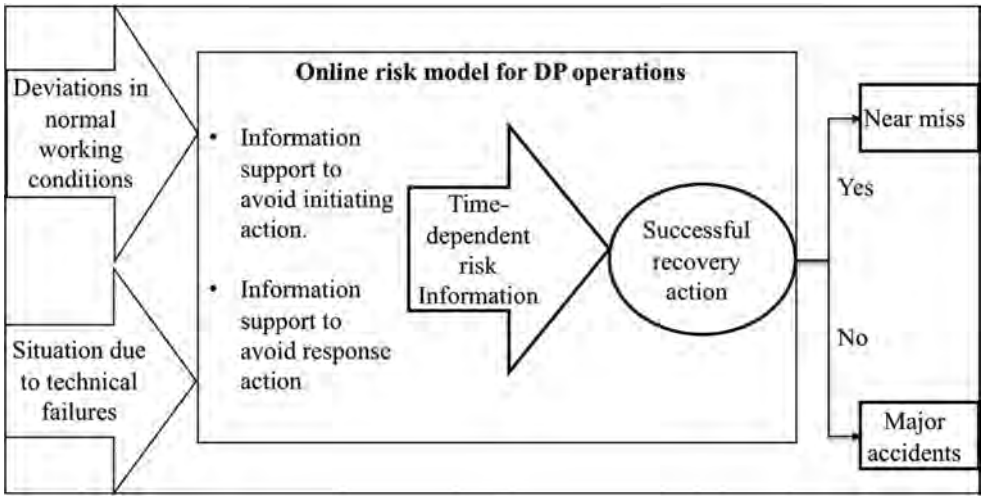


Figure 4. Two types of information support in an online decision support system.

deviations in normal working condition. The operator needs to be aware of whether their planned action will initiate a dangeorus undetected failure. This information support should avoid initiation of drive-off involving human initiating action.

2. Support information for DPO to analyse the actual system performance. The information is to help operator to be aware of whether their response action will make worse the situation leading to a drive-off.

Meanwhile, the importance of the time aspect should be emphasized, since responses may develop so fast in a drive-off situation that it might develop to a severe consequence, such as collision, within a very short time. Operator has to maintain awareness of the situation and catch the development of the situation.

5 CONCLUSIONS

Due to the nature of DP operations and human machine dynamics, loss of situation awareness has been recognized as the main reason for no early detection. One of the reasons is that the initiating event of position loss involves a complex HMI. Tanker drive-off potentially involves not only DP hardware and software, position reference systems, and vessel sensors and local thruster control system, but also the DP operator.

To improve DP operator's situational awareness and understanding of system performance, it is necessary to study the human machine interaction based on classification of human action and types of failures. It has been found that many DP accidents and incidents are involved human initiating action and response action.

Two situations are identified from the DP accidents and incidents involving initiating action and response action. First, DPO encounters the situation that is associated with deviations in normal working conditions (ST keeps its position within operating limits). Drive-off is initiated due to the interaction between human initiating action and dangerous undetected failures. Second, DPO faces a situation given by technical failures. The operator needs information support for being aware whether the actual performance of DP system is within acceptable deviation or not.

Based on the situations, some challenges are pointed out:

Challenge 1: In the first situation, the DPO should evaluate if their action will initiate a dangerous failure which has not been detected. It indicates the need for information support concerning the deficiency of proof testing, which should be supported operator' evaluation of planning a decision for the normal operational task.

Challenge 2: In the second situation, the DPO should analyze the detected failure if it is a safe failure or dangerous failure and if their response action will worsen the situation. Therefore, it is of vital importance for the operator to be aware of the actual system performance is acceptable deviations.

An online decision support system will be an advisory tool concerning the listed challenges to support the DPO to improve situation awarness and decision

support. A benefit of introducing the system is to avoid DPO overconfidence in the DP system.

REFERENCES

61508, IEC. 2010. *IEC 61508*: *Functional Safety of E/E/PE Safety-related Systems*. International Eletrotechnical Commission.

Chen, Haibo. 2014. May 20. Accessed December 2017. https://www.ntnu.edu/documents/10392/881108733/4+Haibo+Chen.pdf/56fe1148-8a72-4466-8763-2c3384aa29f0.

—. 2003. *Probabilistic evaluation of FPSO-tanker collision in tandem offloading operation*. Trondheim: NTNU.

Chen, Haibo, and Bjørn Nygård. 2016. "Quantified Risk Analysis of DP Operations—Principles and Challenges." *SPE International Conference and Exhibition on Health, Safety, Security, Environmental and Social Responsibility.* Stavanger, Norway.

Chen, Haibo, Torgeir Moan, and Harry Verhoeven. 2008. "Safety of dynamic positioning operations on mobile offshore drilling units." *Reliability Engineering and System Safety* 1072–1090.

Dong, Yining, Børge Rokseth, Jan Erik Vinnem, and Ingrid Bouwer Utne. 2017. "Analysis of dynamic positioning system accidents and incidents with emphasis on root causes and barrier failures." *Risk, Reliability and Safety: Innovation Theory and Pratice.* Glasgow.

Eltervåg, Aina, Tommy B. Hansen, Elisabeth Lootz, Else Rasmussen, Eigil Sørensen, Bård Johnsen, Jon Erling Heggland, Øyvind Lauridsen, and Gerhard Ersdal. 2017. *Barrierenotat 2017: Prinsipper for barriererestyring i petroleumsvirksomheten.* Oslo: PSA.

Endsley, M.R. 1995. "Toward a Theory of Situation Awareness in Dynamic Systems." *Human Factors* 37(1): 32–64.

Hogenboom, Sandra, Jan Erik Vinnem, and Ingrid Bouwer Utne. 2017. "Towards an online risk model for DP operations: decison-making and risk information."

Husjord, D., and E. Pedersen. 2009. "Operational Aspects on Decision-making in STS Lightering." *Proceedings of the Nineteenth (2009) International Offshore and Polar Engineering Conference.* Psaka, Japan.

Husjord, Dagfinn. 2015. *Guidance and decison-support system for safe navigation of ships operating in close proximity*. Trondheim: NTNU.

IMO MSC Circ.645. 1994. "Guideline for vessels with dynamic positioning systems."

Kjell I. Øvergård, Linda J. Sorensen, Salman Nazir & Tone J. Martinsen. 2015. "Critical incidents during dynamic positioning: operators' situation awareness and decisionmaking." *Theoretical Issues in Ergonomics Science* 16(4): 366–387.

Kvitrud, Arne, Harald Kleppestø, and Odd Rune Skilbrei. 2012. "Position Incidents during Offshore Loading with Shuttle Tankers on the Norwegian Continental Shelf 2000–2011." *Proceedings of the twenty-second international offshore and polar engineering conference.* Rhodes, Greece.

Øvergård, Kjell I., Linda J. Sorensen, Salman Nazir, and Tone J. Martinsen. 2015. "Critical incidents during dynamic positioning: operators' situation awareness and decsion-making in maritime operations." *Theoretical Issues in Ergonomics Science* 366–387.

Pan, Yushan, Sathiya Kumar Renganayagalu, and Sashidharan Komarndur. n.d. "Tacticle cues for ship bridge operations." *Proceedings 27th European Conference on Modelling and Simulation.*

PSA. 2013. "Principles for barrier management in the petroleum industry."

Rausand, Marvin. 2011. *Risk Assessment Theory, Methods and Applications.* Hoboken, New Jersery: John Wiley & Sons.

Rausand, Marvin, and Arnljot Høyland. 2004. *System Relaibility Theory: Models, Statistical Methods, and Applications.* Hoboken, New Jersey: John Wiley.

Reason, J. 1990. *Human Error.* Cambridge University Press.

Sorensen, Linda J., Kjell I. Øvergåard, and Tone J.S. Maritinsen. n.d. "Understanding human decision making during critical incidents in dynamic positioning."

Sudano, J.J. 1994. "Minimizing human-machine interface failures in high risk systems." *Aerospace and Electronics Conference*. Dayton.

Vinnem, Jan Erik, Ingrid Bouwer Utne, and Ingrid Schjølberg. 2015. "On the need for online decision support in FPSO-shuttle tanker collision risk reduction." *Ocean Engineering* 101: 109–117.

Vinnem, Jan Erik, P. Hokstad, T. Dammen, H. Saele, H. Chen, and S. Haver. 2003. "Operational safety analysis of FPSO-ST collision risk reveals areas of improvement." *Proceedings of the OTC conference.* Houston, USA.

Wickens, Christopher D., Sallie E. Gordon, and Yiliu Liu. 1997. "Chapter 16 Automation." In *An Introduction to Human Factors Engineering*, 493–512. New York: Addison Wesley Longman.

Yang, Xue, and Stein Haugen. 2015. "Classification of risk to support decision-making in hazardous process." *Safety Science* 115–126.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Development of dynamic safety envelopes for autonomous remotely operated underwater vehicles

J. Hegde, E.H. Henriksen, I.B. Utne & I. Schjølberg
NTNU Center of Autonomous Marine Operations and Systems, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

ABSTRACT: This paper describes the implementation of dynamic safety envelopes for Autonomous Remotely Operated Vehicles (AROVs). A safety envelope is defined as a three-dimensional spatial area around the AROV, which forms a virtual protective barrier against collision with known and unknown obstacles in the subsea environment. The Octree method is used to setup the cuboidal shape of the proposed safety envelope. A Fuzzy Inference System (FIS) is modeled to derive the size of the dynamic safety envelope. The three inputs of the proposed FIS are vehicle velocity, probability of acoustic sensor failure and time to collision risk indicator. A user interface allows for verification and visualization of the resulting dynamic safety envelope during live laboratory tests. The results show that similar to vehicular envelopes in other industries, dynamic safety envelopes can be implemented on AROVs. The proposed dynamic safety envelope may be used to model the behavior of AROVs when confronted with different collision scenarios.

1 INTRODUCTION

Globally, numerous research initiatives are investigating the use of autonomous remotely operated underwater vehicles to perform subsea inspection, maintenance, and repair (IMR) operations (Jamieson et al. 2012, Furuholmen et al. 2013, Mai et al. 2016, Gancet et al. 2016, Schjølberg et al. 2016). Autonomous remotely operatedc underwater (AROVs) are tethered/untethered underwater vehicles, which can independently control manipulator functions, permit shared control between the vehicle and the human operator. AROVs can navigate autonomously, perform self-diagnostics, and be equipped with remotely operated tool systems requiring limited operator control (Hegde et al. 2015). However, the introduction of autonomy in subsea IMR operations may also result in emerging risk factors. One such risk factor is the risk of collision posed by the use of AROVs (Hegde et al. 2016, Utne and Schjølberg 2014). Delayed IMR operations, loss of vehicle, loss of structural integrity may be some of the severe consequences of AROV collisions with the subsea structures, other AROVs and the seabed. Safeguarding the functions of subsea infrastructure and the AROVs is vital to ensure safe and cost efficient autonomous subsea IMR operations.

Studies to identify, assess, and avoid collision risk of vehicles are paramount for all vehicular systems. In the early 1970s, an increase in maritime traffic and need for safe envelopes around the marine vessel was highlighted by Fujii and Tanaka (1971). Influenced by the collision avoidance procedures in the aviation industry, Goodwin (1975) coined the term "*ship domain*". Goodwin (1975) defined "*ship domain*" as the "*sea around the ship, which the navigator would like to keep free, with respect to other ships and fixed objects*". Over the years, the size, shape, and the area covered by the ship domain has evolved continuously (Pietrzykowski and Uriasz 2009, Tam et al. 2009, Lewison 1978, Davis et al. 1980). Currently, in the automotive, maritime (surface vehicles), aviation and space industries, different forms of vehicular safety envelopes are utilized during operations. The primary aim of these vehicular envelopes is to suggest or autonomously modify the behavior of the vehicle when obstacles are detected inside the vehicular safety envelope.

Hegde et al. (2017) utilize the Octree method to design a static safety envelope for AROVs. The term safety envelope can be defined as *a 3D spatial area around the underwater vehicle forming a virtual protective barrier (in space and time) against collision with known and unknown obstacles in the subsea environment, influencing the behavior of the AROV* (Hegde et al. 2017). In a static safety envelope, the size of the envelope is constant and does not change during live IMR operations. This approach is valid when the AROV is in close proximity to the subsea equipment. However, when the AROV is moving from one location to another, a dynamic safety envelope may assist the AROV

and the human operator to adapt and react to different collision scenarios. In addition, a dynamic envelope can reduce the need to detect obstacles by decreasing the area of the envelope. This can result in decreased data processing requirements for the on-board collision detection module. At this time, such dynamic vehicular envelopes do not exist for AROVs (Hegde et al. 2015).

The objective of this paper is to develop dynamic safety envelopes for AROVs, i.e., the size of the safety envelope changes depending on operational parameters of the AROV.

This paper is organized as follows: Section 2 presents the design of safety envelope. The elements of the proposed fuzzy inference system is described in Section 3. Section 4 describes the laboratory setup used to test the dynamic safety envelope. The results from the laboratory tests are presented in Section 5. The findings are discussed in Section 6. Section 7 presents the conclusions and future work possibilities.

2 DESIGN OF DYNAMIC SAFETY ENVELOPES

An Octree is used to generate the dynamic safety envelopes. Octree is a recursive tree data structure, which consists of spatial cubes named Octants. Each Octant can further be divided into eight child Octants. Figure 1 illustrates the Level 1 and the Level 2 Octree rendering with the AROV in the center of the Octree. In the Level 1 Octree, eight cubes surround the AROV and in the Level 2 Octree sixty four cubes surround the AROV. Each of the cubes are allocated an unique identifier and linked to a safe subsea traffic rule. The subsea traffic rule aims to maximize the horizontal and vertical seperation from the identified obstacle. If an obstacle is detected in one or more Octants, a suitable subsea traffic rule is suggested to the AROV or the human operator.

According to Hornung et al. (2013), there are four main reasons to use the Octree method for robot applications.

1. Octrees can establish virtual spatial grids around the robot, which can be used to check for collisions with the obstacles in all three axis.
2. The resolution of Octrees can be increased or decresed, which can result in detailed obstacle tracking, if required.
3. Measurement data from multiple sensors can be probabilistically represented using Octrees.
4. Both active and passive sensors can be used to check for collisions in known and unknown environments.

AROVs are also exposed to collisions with obstacles. Data from multiple active and passive sensors can be used to detect obstacles in the subsea enviornment. Therefore, use of the Octree method as highligted by Hornung et al. (2013) can also be extended to underwater vehicle applications.

In addition, the size of the of the Octree can increase or decrease the computational load in detecting the obstacle. If the safety envelope is static, the constant computations required may consume battery power by the AROV even when there are no obstacles in the vicinity. A dynamic safety envelope may lead to decrease in the computations required by limiting the collision detection module to an optimized Octree area. The next section explores the use of fuzzy logic to derive the size of the dynamic safety envelope.

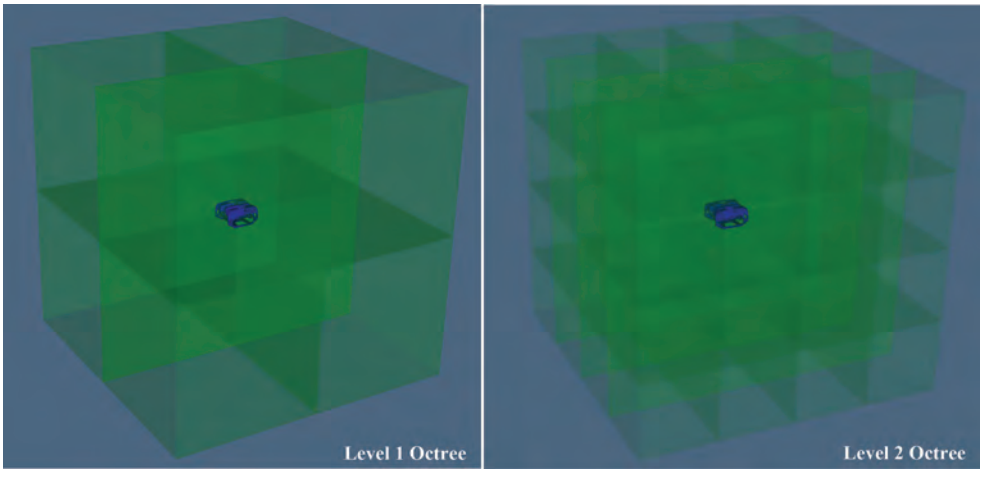


Figure 1. Rendering of static safety envelope for AROVs as proposed by Hegde et al. (2017).

3 FUZZY INFERENCE SYSTEM

Fuzzy logic delivers precise outputs from imprecise inputs. Input values are assumed to vary within a given range of values, which resembles real-life scenarios. Figure 2 is adapted from (Zadeh 2002, Zadeh 1996) and describes the overall methodology of a fuzzy inference system (FIS). In a FIS, input and output variables can contain n number of fuzzy sets with shared memberships among other fuzzy sets. This process of converting the crisp input to range values is known as fuzzification. A fuzzy operator is used to connect the antecedent (fuzzy inputs) to a consequent (crisp output) through an if-then logic. Defuzzification is achieved by calculating the membership of input variable fuzzy sets against the output variable fuzzy sets. Defuzzification results in a crisp value that can further be used as input to make decisions. Fuzzy inference systems are useful in two main use cases: first, to model systems that are highly complex and when the systems behavior is vaguely understood; and second, where an approximate, but quicker solution is acceptable.

Scikit Fuzzy, a fuzzy logic module in Python programming language is utilized to set up the proposed FIS (Warner et al. 2017). Figure 2 provides an overview of the proposed FIS, which has three input variables and one output variable.

3.1 *Fuzzification of variables*

Three input variables (operational performance indicators of the vehicle) are identified to influence the output variable i.e., the size of the safety envelope. This means that it is assumed that the vehicle performance influence the safe operation. This subsection describes the fuzzification of the input and output variables.

3.1.1 *Vehicle velocity*

Considering the kinetic energy of a moving vehicle, the velocity at which the vehicle moves can influence collision detection and avoidance ability of the underwater vehicle. An AROV traveling at high velocity can result in a faster approach to a potential obstacle. This means that the time needed to detect and avoid the collision scenario is inversely related to the velocity of the underwater vehicle. Therefore, vehicle velocity (vv) is used as one of the inputs in the proposed FIS. Three membership functions for the vehicle velocity input are assumed, namely low, medium and high.

The FIS was modeled according to the technical specifications of the Blue Robotics BlueROV2. The maximum achievable velocity of the BlueROV2 vehicle is 1 m/s (Blue Robotics 2017). Figure 3 illustrates the resulting membership functions (MFs) for vehicle velocity input. The three MFs are low, medium and high. The low velocity MF ranges from 0 to 0.4 *m/s*. The medium velocity MF ranges from 0.2 to 0.8 *m/s* and high velocity MF ranges from 0.8 to 1 *m/s*.

3.1.2 *Probability of acoustic sensor failure*

Stovner et al. (2017) demonstrate use of underwater acoustic sensor grid to aid localization capabilities of the AROVs. A grid of acoustic sensors is used to communicate with and track the position of the AROV during IMR operations. However, failure of one or more subsea acoustic sensors may result in inaccurate position, orientation and velocity estimates. Reliable measurements of vehicle position, orientation, and velocity are important to safely navigate in the subsea enviornment. Failure of acoustic sensors may lead to increased risk of collision or loss of the AROV. It is therefore, important that AROVs use the available acoustic sensor information to make informed decisions.

In this paper, the acoustic grid consists of four acoustic transducers on the bed of the pool and two acoustic transducers on the AROV as illustrated in Figure 7. Acoustic transducers placed on bed of the pool can communicate with the transducers

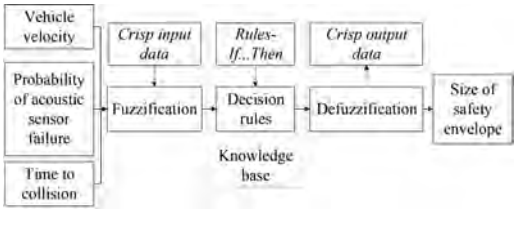


Figure 2. Overview of the proposed fuzzy inference system.

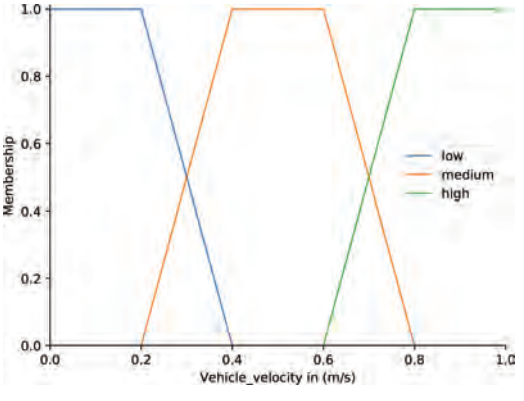


Figure 3. Membership functions for vehicle velocity (m/s).

on the AROV. This results in an acoustic network with eight possible range (distance) measurements. A minimum of four range values are needed for the acoustic positioning system to be classified as reliable. If there are fewer than four range measurements, the resulting estimates (position, orientation and velocity) are assumed to be unreliable. In such scenarios, the acoustic localization system is categorized as failed i.e., the probability of acoustic sensor failure is 1. Failure is defined as the termination of the ability of a functional unit to provide a required function or operation of a functional unit in any way other than as required (IEC 61508 2009). Therefore, the acoustic sensor voting scheme is 4*oo*8 (four out of eight). In a failed state, it is vital that the safety enelope around the AROV increases in size.

Figure 4 illustrates the membership functions of the probability of acoustic sensor failure. A trapezoidal membership function (TRAMF) is used to signify that at certain range of input values the membership is unity (1). As shown in Figure 4, the probability of the acoustic sensor failure variable consists of three MFs, namely low, medium and high. The low MF ranges from probability 0 to 0.3. The medium MF ranges from probability 0.2 to 0.7 and the high MF ranges from probability 0.6 to 1.

3.1.3 *Time to collision*

In the automotive and aviation industry, the time remaining for the vehicle to collide with an obstacle is used to suggest collision avoidance maneuvers. The term time to collision (TTC) is used to convey the criticality of the collision scenario. The lower TTC, the greater the risk of collision with the obstacle. In the Traffic Collision Avoidance System (TCAS), the value of TTC is utilized to calculate the criticality of the obstacle (US Department of Transportation and Federal Aviation Administration 2011).

Hegde et al. (2016) apply the TTC as a risk indicator that can indicate risk of collision in a given AROV path. The TTC can be classified as an operational parameter, in that it can change as the vehicle velocity and the distance to the obstacle varies. As AROVs are required to navigate through the subsea infrastructure, they may face many obstacles in their path.The TTC risk indicator can aid in classifying critical obstacles by monitoring continously. TTC is calculated by using Equation 1

$$Time\ to\ collision = \frac{Distance\ to\ obstacle}{Resultant\ velocity\ of\ \ AROV} \tag{1}$$

A recommended standard for autonomous subsea IMR also highlights the need for monitoring all existing obstacles in the vicinity of the subsea production system (Germanischer Lloyd Aktiengesellschaft 2009). Therefore, the inclusion of the TTC indicator in the proposed FIS allows the AROV to not only monitor the obstacles, but also devise collision avoidance behavior if they are under a threshold value. In the proposed FIS, the threshold values relate to the MFs. Three MFs are determined for the TTC input variable, namely low, medium and high. The low MF of TTC ranges from 0 to 3 *s*. The medium and high MF range from 2 to 5 and 5 to 10 *s* respectively.

3.1.4 *Size of safety envelope*

The output variable in the proposed FIS is the size of the safety envelope. The safety envelope is realized in form of a cuboid. Therefore, the FIS output: size of the safety envelope will increase or decrease uniformly along all three axes *North, East and Down*. The size of the safety envelope is proportional to the velocity of the vehicle and

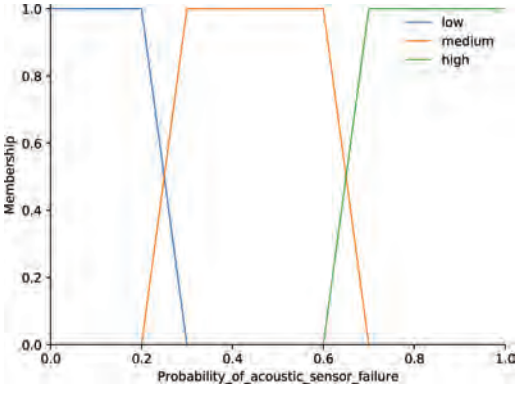


Figure 4. Membership functions for probability of acoustics sensor failure.

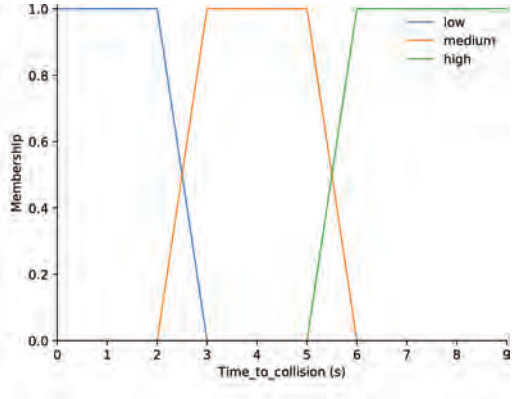


Figure 5. Membership functions for time to collision (s).

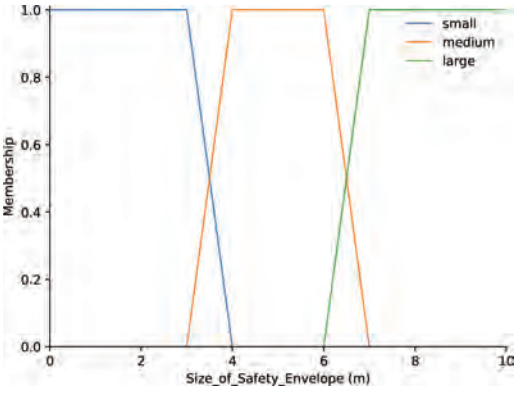


Figure 6. Membership functions for size of safety envelope (m).

probability of acoustic sensor failure (PASF) and inversely proportional to the time to collision input. In short, the safety envelope size increases when vehicle velocity and PASF increase and decreases when TTC value increases. A large safety envelope reflects a low safety margin whereas a small safety envelope reflects high safety.

Figure 6 illustrates the MFs for the FIS output variable. The size of the safety envelope is classified into three MFs, namely small, medium and large. The MF for the small safety envelope ranges from 0 to 4 *m*. The MF for the medium safety envelope ranges from 3 to 7 *m* and the MF for the large safety envelope ranges from 6 to 10 *m*.

3.2 *Fuzzy rule set*

Once the fuzzy sets of input and output variables are determined, the next step is to define fuzzy rules by combining the input and output variables using logic statements. Table 2 lists the twenty seven fuzzy rules resulting from the three input variables. The fuzzy logic operator AN D is used to derive the inference from input variables.

It has to be noted that the input variables and their influence on the output variable is different. For example, a low TTC is not favorable as this would mean that the obstacle is in close proximity to the AROV. On the other hand, low vehicle velocity and PASF are favorable. Relative importance of inputs are not considered in this paper. Weights are not allotted to the input variables and therefore the fuzzy rule set do not favour certain rules over others. All rules are given equal importance.

3.3 *Deffuzification*

The process of obtaining crisp values from fuzzy inputs is known as defuzzification. The Scikit Fuzzy library supports numerous defuzzification methods, such as centroid, bisector, mean of maximum (mom), min of maximum (som) and max of maximum (lom) (Warner et al. 2017). Centroid defuzzification method is used in this paper because it provides conistent crisp output values when compared to other deffuziffication methods within the uncertainty constraints. The centroid defuzzification method aggregates the total area under the membership functions of the input variables and calculates the centroid of the combined area (Sivanandam et al. 2007).

Table 1. Rule sets in the fuzzy inference system. vehicle velocity (VV), probability of acoustic sensor failure (PASF), time to collision (TTC).

Rule number	Antecedent: VV & PASF & TTC	Consequent size of safety envelope
1	Low & Low & Low	Large
2	Low & Low & Medium	Medium
3	Low & Low & High	Small
4	Low & Medium & Low	Large
5	Low & Medium & Medium	Medium
6	Low & Medium & High	Medium
7	Low & High & Low	Large
8	Low & High & Medium	Large
9	Low & High & High	Large
10	Medium & Low & Low	Large
11	Medium & Low & Medium	Medium
12	Medium & Low & High	Medium
13	Medium & Medium & Low	Large
14	Medium & Medium & Medium	Medium
15	Medium & Medium & High	Medium
16	Medium & High & Low	Large
17	Medium & High & Medium	Large
18	Medium & High & High	Large
19	High & Low & Low	Large
20	High & Low & Medium	Large
21	High & Low & High	Medium
22	High & Medium & Low	Large
23	High & Medium & Medium	Large
24	High & Medium & High	Large
25	High & High & Low	Large
26	High & High & Medium	Large
27	High & High & High	Large

4 LABORATORY SETUP AND TESTING

Figure 7 illustrates the laboratory setup to test the dynamic safety envelopes. The Mission Orientated Operating Suite (MOOS) middle-ware (Newman 2006) stores and retrieves information from the AROV. Four acoustic sensors are installed on the bed of the pool and two on the AROV. The acoustic

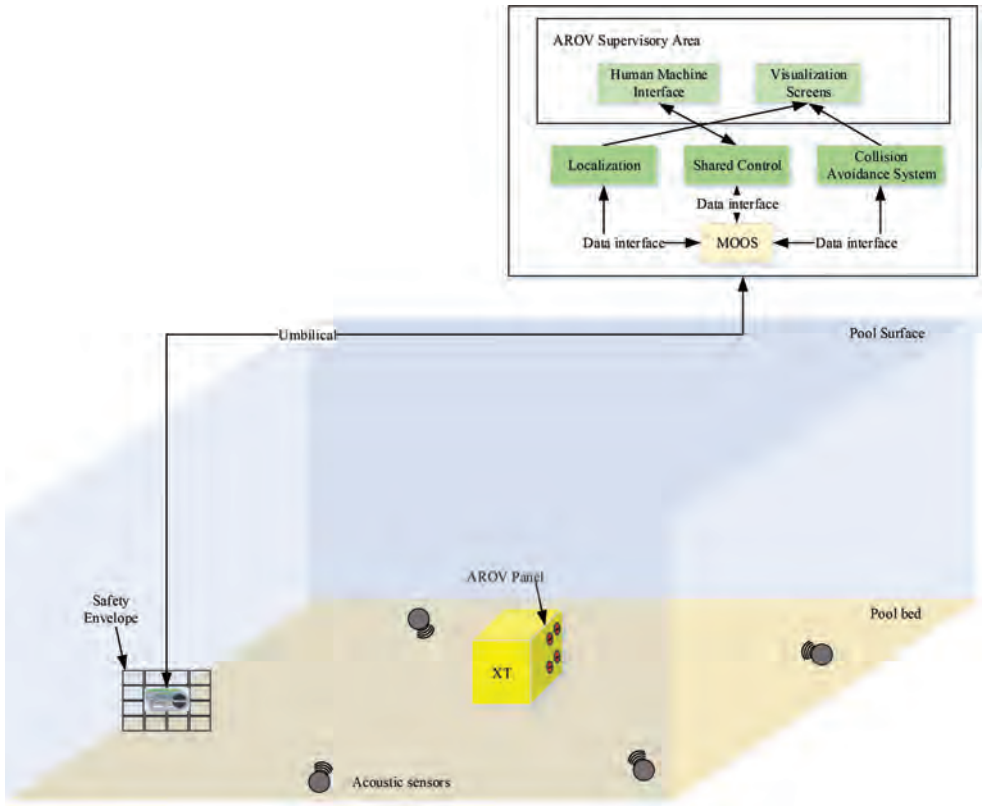


Figure 7. Laboratory setup to test feasibility of dynamic safety envelopes.

sensors provide the localization measurements, such as position of obstacle and AROV, velocity and orientation of the AROV. The localization module shares the data with MOOS. The shared control module retrieves data from the localization module via MOOS and utilizes it to control the AROV either autonomously or via shared control between the AROV and the AROV supervisor. The communication to the AROV is established through an umbilical.

The collision avoidance module posts and retrieves collision data to and from MOOS. The two parts of the collision avoidance module are the dynamic safety envelope and the subsea traffic rules. The subsea traffic rules are set of assigned safe navigation maneuvers that can be performed by the AROV to increase the vertical and/or horizontal separation (distance) from the obstacle (see Section V of Candeloro et al. (2016)). The subsea traffic rules are developed based on the rules from collision regulations (COLREGs) in the maritime and from the TCAS in the aviation industry. Each Octant in the Level 2 Octree in Figure 1 is assigned a subsea traffic rule to maximize vertical and/or horizontal separation from the obtascle.

The pseduocode implemented to derive the dynamic envelopes from the proposed FIS is as listed in Listing 1. The first step is to retrieve the velocity and position variables from MOOS followed by the distance to the potential obstacle. Then the available acoustic ranges are counted and a sensor voting scheme of 4oo8 is used to derive the PASF. The velocity of the AROV, distance to the obstacle are used to calculate the TTC. When the input data are collected, they are routed to the FIS as described in Figure 2.

The FIS computes the new size of the safety envelope and publishes the new size of the safety envelope to MOOS. The 3D renderer updates safety envelopes to the new size and the detection algorithm updates the potential detection volume.

Listing 1: Pseudocode of FIS implementation in the underwater collision avoidance system

```
# Initialization
Get position, velocity of AROV and obstacle
Get count of acoustic sensor ranges
Get envelope size
```

```
#Dynamic safety envelope
Set MFs for VV, PASF, TTC
Set fuzzy rule set
Compute PASF and TTC
Compute new size of safety envelope(FIS)
Update envelope size to new size
Update detection volume to new size

#Collision detection
Make empty list collisions
  If new envelope collides with world:
    for each octant in new envelope:
      If octant collides with world:
        append octant to collisions
```

5 RESULTS

In Figure 8, the Vehicle velocity is 0.02 m/s (MF = low) and PSAF is 0 (MF = low) and TTC is 391.01 s (MF = high). These inputs result in application of Rule 6. The resulting size of safety envelope is 1.76 *m* (MF = small). In Figure 9, the Vehicle velocity is 0.03 *m/s* (MF = low) and PSAF is 0.38 (MF = medium) and TTC is 391.01 *s* (MF = high). These inputs result in application of Rule 6. The resulting size of safety envelope is 5.00 *m* (MF = medium). Observations from the laboratory tests are as listed in Table 2.

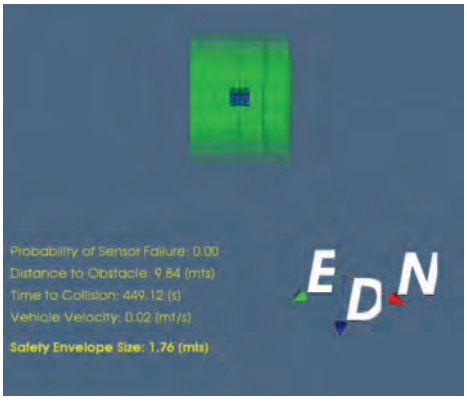


Figure 8. Rendering of data point 1.

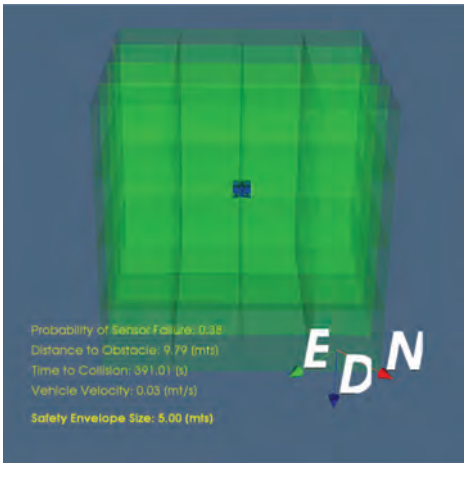


Figure 9. Rendering of data point 4.

Table 2. Observations from laboratory tests.

Data point	VV (m/s)	PASF	TTC (s)	Size of safety envelope (m)
1	0.02	0	449.12	1.76
2	0.24	0	22.12	2.49
3	0.63	0	16.52	5.00
4	0.03	0.38	391.01	5.00
5	0.25	0	18.52	2.58
6	0.20	0.12	12.69	1.83
7	0.33	0.25	5.68	3.68
8	0.36	0.25	10.56	4.11

6 DISCUSSION

This section discusses the learnings from the process and the impact to industrial applications through the development of the proposed dynamic envelopes.

The key drivers in development of safety envelopes in the aviation and maritime industries (TCAS, Ship Domain) are asset/personnel safety and ability to design an intelligent collision avoidance systems. By fusion of both active and passive sensor technologies, the safety envelopes in the aviation, maritime (surface vehicles) and automotive industries currently utilize dynamic safety envelopes. In comparision, current remotely operated vehicles are controlled by human operators. In the future, AROVs will also need to be able to make decisions both in presence or in absence of the human operators. Development of dynamic safety envelopes can be seen as the first step towards ensuring asset safety of AROVs by identifying, assessing, and mitigating the risk of underwater collisions.

As applications of AROVs to inspect and repair subsea production systems, offshore aquaculture systems, offshore wind turbines and facilitate subsea mining, asset safety of AROVs is vital. The proposed process to build dynamic safety envelopes using fuzzy logic allows the system developers to tweak the membership functions and fuzzy rule sets according to their respective industrial requirements. This allows for application specific dynamic safety envelopes. For example, requirements for dynamic safety envelope for subsea

IMR operations and subsea mining operation may vary as the later is more vulnerable to seabed collisions than collisions with the man-made subsea structures.

Use of fuzzy logic (expert-based systems) ensures that the system developers can understand the inherent behaviour of the system under different input conditions. However, two limitations of fuzzy logic in engineering applications can be highlighted. First, fuzzy logic is a form of deductive reasoning i.e., to conclude on a specific truth by using generic inputs (Ross 2009). An example for deductive reasoning is the ground is wet (input) therefore, it must be raining (truth). Second, the subjective nature of defining the membership functions of the fuzzy variables and deriving fuzzy rule set can be challenging. This is true for any technical system where experts are needed to provide input and they can disagree with each other's judgment. For example, in the proposed fuzzy rule set, weightage to different inputs are not implemented. All rules and input values are given the same importance. In the future, a modification may be necessary to include the relative importance of input variable. Is the TTC more important than PASF and VV or vice-versa?

The proposed dynamic safety envelopes are highly dependent on availability of reliable sensor measurements. The laboratory setup used in this paper consists of a grid of six acoustic sensors providing eight range measurements. Measurements from the acoustic sensors were used as the primary input to calculate the orientation, velocity and position of the AROV and the obstacle (passive sensor grid). However, the advantage of the proposed dynamic envelopes is that it can be easily modified to include input from either passive or active sensors. For example, active sensors, such as sonar and LiDAR can detect both known and unknown obstacles. This is possible because of the underlying architecture of the implemented 3D rendering program, which allows for scalability. In addition, if the sensor module comprises of redundant sensors, failure of one sensor type can be tolerated by the overall collision avoidance system. For example, if the acoustic position sensor fails to measure the depth of the AROV, measurements from a dedicated depth sensor can still provide a reliable source to the proposed FIS.

7 CONCLUSIONS

This paper proposes a novel approach to developing dynamic safety envelopes for autonomous remotely operated vehicles (AROVs). A proof-of-concept of the dynamic safety envelope is presented in this paper.

The proposed dynamic safety envelope was developed by using a fuzzy inference system (FIS) to adapt the size of the safety envelope. Three fuzzy input variables were used in the FIS, namely vehicle velocity, probability of acoustic sensor failure and time to collision. A FIS was implemented in an existing underwater collision avoidance system. Observations from the laboratory tests performed to verify the feasibility of dynamic safety envelopes are presented. Results show that the AROV safety envelope can increase or decrease in size depending on the three input variables. This allows the AROV to decrease or increase the obstacle detection area in a highly uncertain and sensitive subsea environment.

In presence of uncertainty, visualizations of obstacles that pose the risk of collision to the AROV may aid situation awareness of human operators. The size of the safety envelope can be used to make decisions related to maneuvering of the AROV either autonomously by the AROV or remotely by the human operator. To safely maneuver the AROVs during collision scenarios, further development and testing is required to implement dynamic safety envelopes together with the subsea traffic rules.

ACKNOWLEDGMENT

This work is supported by the Research Council of Norway, Statoil and TechnipFMC through the research project Next Generation Subsea Inspection, Maintenance and Repair Operations, 234108/E30.

REFERENCES

Blue Robotics (2017). BlueROV2 Datasheet.

Candeloro, M., A. Lekkas, J. Hegde, & A.J. Sørensen (2016). A 3D Dynamic Voronoi Diagram-Based Path-Planning System for UUVs. In *OCEANS'16 MTS/IEEE Monterey*, Monterey, US.

Davis, P.V., M.J. Dove, & C.T. Stockel (1980). A computer simulation of marine traffic using domains and arenas. *The journal of Navigation 33*(2), 215–222.

Fujii, Y. & K. Tanaka (1971). Traffic Capacity. *Journal of Navigation 24*(4), 543–552.

Furuholmen, M., A. Hanssen, R. Carter, K. Hatlen, & J. Siesjo (2013). Resident Autonomous Underwater Vehicle Systems A Review of Drivers, Applications, and Integration Options for the Subsea Oil and Gas Market. In *Offshore Mediterranean Conference*. Offshore Mediterranean Conference.

Gancet, J., P. Weiss, G. Antonelli, M.F. Pfingsthorn, S. Calinon, A.Turetta, C. Walen, D. Urbina, S. Govindaraj, P. Letier, X. Martinez, J. Salini, B. Chemisky, G. Indiveri, G. Casalino, P. Di Lillo, E. Simetti, D. De Palma, A. Birk, T. Fromm, C. Muel-

ler, A. Tanwani, I. Havoutis, A. Caffaz, & L. Guilpain (2016). Dexterous Undersea Interventions with Far Distance Onshore Supervision: the DexROV Project. *IFAC-PapersOnLine 49*(23), 414–419.

Germanischer Lloyd Aktiengesellschaft (2009, nov). Rules for Classification and Construction Ship Technology- Underwa- ter Technology - Unmanned Submersibles (ROV, AUV) and Underwater Working Machines.

Goodwin, E.M. (1975). A Statistical Study of Ship Domains. *Journal of Navigation 28*(3), 328–344.

Hegde, J., E.H. Henriksen, I.B. Utne, & I. Schjølberg (2017). Development of safety envelopes and subsea traffic rules for autonomous remotely operated vehicles. *Journal of Safety, MDPI* (Submitted Under Review).

Hegde, J., I. Utne, I. Schjølberg, & B. Thorkildsen (2015). Application of fuzzy logic for safe autonomous subsea IMR operations. In *Safety and Reliability of Complex Engineered Systems - Proceedings of the 25th European Safety and Reliability Conference, ESREL 2015*, pp. 415–422.

Hegde, J., I.B. Utne, & I. Schjølberg (2015, May). Applicability of Current Remotely Operated Vehicle Standards and Guidelines to Autonomous Subsea IMR Operations. In *Volume 7: Ocean Engineering*, pp. V007T06 A026. ASME.

Hegde, J., I.B. Utne, & I. Schjølberg (2016). Development of collision risk indicators for autonomous subsea inspection maintenance and repair. *Journal of Loss Prevention in the Process Industries 44*, 440–452.

Hornung, A., K.M. Wurm, M. Bennewitz, C. Stachniss, & W. Burgard (2013, apr). OctoMap: an efficient probabilistic 3D mapping framework based on octrees. *Autonomous Robots 34*(3), 189–206.

IEC 61508 (2009). Functional Safety of electrical/electronic/programmable electronic safety-related systems.

Jamieson, J., L. Wilson, M. Arredondo, J. Evans, K. Hamilton, & C. Sotzing (2012, apr). Autonomous Inspection Vehicle: A New Dimension in Life of Field Operations. In *OTC-23365- MS*, pp. 8. Offshore Technology Conference.

Lewison, G.R.G. (1978). The Risk of a Ship Encounter Leading to a Collision. *Journal of Navigation 31*(3), 384–407.

Mai, C., S. Pedersen, L. Hansen, K.L. Jepsen, & Zhenyu Yang (2016). Subsea infrastructure inspection: A review study. In *2016 IEEE International Conference on Underwater System Technology: Theory and Applications (USYS)*, pp. 71–76. IEEE.

Newman, P.M. (2006). MOOS -Mission Orientated Operating Suite. Tech. Rep. OE2003–07 (MIT Department of Ocean Engineering, Cambridge 2003). Technical report, Massachusetts Institute of Technology.

Pietrzykowski, Z. & J. Uriasz (2009, jan). The Ship Domain A Criterion of Navigational Safety Assessment in an Open Sea Area. *Journal of Navigation 62*(01), 93.

Ross, T.J. (2009). *Fuzzy logic with engineering applications*. John Wiley & Sons.

Schjølberg, I., T.B. Gjersvik, A.A. Transeth, & I.B. Utne (2016). Next Generation Subsea Inspection, Maintenance and Repair Operations. *IFAC-PapersOnLine 49*(23), 434–439.

Sivanandam, S., S. Sumathi, & S. Deepa (2007). *Introduction to fuzzy logic using MATLAB*, Volume 1. Springer.

Stovner, B.B., T.A. Johansen, & I. Schjølberg (2017). Globally exponentially stable aided inertial navigation with hydroacoustic measurements from a single transponder. In *American Control Conference (ACC), 2017*, pp. 1219–1226. IEEE.

Tam, C., R. Bucknall, & A. Greig (2009). Review of Collision Avoidance and Path Planning Methods for Ships in Close Range Encounters. *Journal of Navigation 62*(3), 455–476.

US Department of Transportation and Federal Aviation Administration (2011). Introduction to TCAS II – Version 7.1. Technical report.

Utne, I.B. & I. Schjølberg (2014). A systematic approach to risk assessment - Focusing on autonomous underwater vehicles and operations in Arctic areas. In *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering – OMAE*, Volume 10.

Warner, J., J. Sexauer, scikit fuzzy, twmeggs, A.M.S., A. Unnikrishnan, G. Castelo, F. Batista, T.G. Badger, & H. Mishra (2017, October). Jdwarner/scikit-fuzzy: Scikit-fuzzy 0.3.1.

Zadeh, L. (2002). From computing with numbers to computing with words - From manipulation of measurements to manipulation of perceptions. *International Journal of Applied Mathematics and Computer Science 12*(3).

Zadeh, L.A. (1996). Fuzzy logic = computing with words. *Fuzzy Systems, IEEE Transactions on 4*(2), 103–111.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Dynamic risk assessment during eco-driving behaviors for conventionally fueled vehicles

G.L. Mauri, E. Bressan & F.C. Velardo
KITE Solutions s.r.l., Varese, Italy

P.C. Cacciabue
KITE Albion Consulting Ltd., Kingston Upon Thames, UK

ABSTRACT: The urgency of applying a more ethical approaches to get to a sustainable low-carbon economy by 2050 requires to implement a model of sustainable mobility, which embraces social, economical and industrial activities. In this horizon, behavioral change can produce considerable benefits whose effects are distributed on wider period. In this direction, the GamECAR project aims at provoking a change in driver style towards a more sustainable and efficient use of private cars by applying gamification. However, safety must not be affected by the introduction of new technologies, tools or systems, which are not essential for the driving task. This paper presents a methodology for the control of the inference of a smart-phone application suggesting eco-driving hints to the driver, on the basis of the dynamic assessment of the risk exposure embedded in the current situation. Real-time measurements of physiological, behavioural and car performance parameters are combined with data-driven driver models to determine the safe communication of eco-driving suggestions to the driver. The methodology builds upon the structured approach to operational safety initially applied in aviation and its adaption to the road environment during the XCYCLE Project (Funded by the Horizon 2020 Framework Programme of the European Union – Grant n° 635975).

1 INTRODUCTION

Road transport is troubled by two major issues: environmental pollution induced by cars and serious injuries to people from traffic crashes. The common denominator between these problems can be found in the aggressive driving style. Studies found that fuel consumption can be affected by the driver's style up to 35%. Hard acceleration and braking, excessive speed, open windows results in higher emission rates from a vehicle compared with a more calm driving style.

We are now in a transitional phase, where electric and hybrid vehicles are becoming more popular, but they are still not sufficiently distributed to generate a significant effect. Eco-driving can become an intermediate tool for inducing a more sustainable and effective use of current private vehicles, carbon fuelled, but it can also inspire a behavioural change, that will be beneficial across generations.

Eco-driving has been defined as a decision making process that significantly affects the fuel economy and emission intensity of a vehicle, reducing its environmental impact. Ecological, economic, but also road safety and social benefits can be derived from adopting eco-driving conduct. However, changing the behaviour of a driver seems to be a challenging task. For example, the social context in which the driver is operating can result a mediating factors of behaviour.

In the context of the GamECAR Project (Funded by the Horizon 2020 Framework Programme of the European Union—Grant n°732068) a Decision Support System module embedding personalized driver models will combine the calculus of an eco-driving score with the dynamic evaluation of the risk associated to the context where the action is conducted. The most appropriate eco-related, personalized suggestion will be communicated to the driver if the operational risk results acceptable. This paper will present the methods adopted to combine real-time measurements of physiological, attitudinal and car performances parameters with data-driven driver models, to determine the safe communication of eco-driving suggestions to the driver. The methodology builds upon the structured approach to operational safety initially applied in aviation and its adaptation to the road environment during the XCYCLE Project (Funded by the Horizon 2020 Framework Programme of the European Union—Grant n° 635975).

2 THE FOUNDING CONCEPTS

Some keywords inspired the creation of the model.

Dynamicity: the road is, by nature, an ever-changing, dynamic environment. In such contexts, the discrepancy between the time available and time required to make a decision generates time pressure. Unfortunately, cognitive models of driving behaviour still fail to take into account the dynamic context that the road environment offers. GamECAR focuses very much on real-time and context dependent information, through which it can suggest new modelling parameters for a more realistic scheme.

Change: dynamicity causes revolution. Managing dynamic traffic scenarios is a complex exercise and the actual task of management is dynamic itself, according to the increase or decrease of the situation. Driver's capability also varies and affects the driver state. For example, the variety of stimuli could affect the driver's attention by causing distraction or by numbing his/her reaction times. Moreover, an impaired or fatigued driver may not be sufficiently responsive to the changing surrounding conditions.

Change is also an innate element of the social context in which the driver operates and can represent a mediating factor of behaviour.

Change is a key word for the GamECAR project itself, as one of its most important objectives is to provoke a behavioural revolution towards more sustainable driving style and general attitude.

Interactive: the traffic system is composed of three main actors, namely road users, vehicles and road environment. This frequent and heterogeneous interactions introduced by traffic impose demands on drivers, increasing the complexity of the activity. Considering road transport, at least three elements compose the scenario: different road users, their transport means and the environment in which the driving task is conducted.

Safety & Ecology: previous researches demonstrated that adopting eco-driving behaviours not only induces economic paybacks, but also contributes to road safety. Among these, it is essential to ensure that any just-in-time feedback must not compromise the safety of its user, by increasing his/her cognitive workload and shifting attention from the actual driving act.

3 A NEW USE FOR DYNAMIC RISK ASSESSMENT OUTCOMES

An adequate, modern and complete risk model for the road transport must include all the above elements as well as their mutual conditionings.

Beyond facing the challenge of designing a risk model capable of representing all members and their interactions through a sufficiently lean structure, thanks to the GamECAR project's requirements, the proposed model adds a significant novelty in the exploitation of risk assessments. In fact, traditionally, results from risk analysis, either dynamic or not, either proactive or retrospective, served to inform the user about the undesirable situation just experienced or possibly to encounter, in order to highlight the need for mitigation actions. In GamECAR, the output of the risk analysis will feed information to the platform, so as to decide whether it is appropriate or not to present the eco-driving information to the driver. The statement depends on the context in which the driving action is conducted, the driver's current state and personal traits, and the driving performances expressed by vehicles parameters. The results of the risk analyses aim at avoiding that new tools, such as the GamECAR display, despite their noble intentions, could overload the driver with unnecessary information and distract him/her from the primary task. The user will not be warned about the fact that s/he is going to experience a hazardous situation (e.g., a risk indicator, whose colour communicates the level of risk), neither about what this circumstance is (e.g., hazard's name and description), nor about the consequences s/he might face (e.g., icon depicting the possible negative outcome). S/he will not be conscious that a risk analysis is performed, but the effects of these background operations will be essential to maintain his/her focus on the primary task and to induce his/her behavioural change through a safe process.

4 STARTING POINTS

The GamECAR risk model builds upon the research conducted in previous EU projects, namely A-PiMod (FP7/2007–2013, contract number: 605141) and XCYCLE (H2020/2015–2018, contract number 635975). The former, through the design of a new adaptive automation concept based on a hybrid of three elements: a multi-modal pilot interaction, an adaptive distribution of tasks between flight crew and automation and real-time risk assessment, aimed at supporting flight management when hazards and unexpected conditions are encountered. The latter focuses on the development of technologies to improve active and passive detection of cyclists, as well as systems informing both drivers and cyclists of a hazard at junctions.

From the progressive experiences and breakthroughs in the aviation domain, in which a structured approach to safety has been applied for decades, and from the adaptation of such approach

in a different sector, such as the road transport one, the GamECAR model is sketched.

5 RISK MODEL COMPONENTS

The GamECAR model is constructed by three components: a LUT (Look Up Table), a calculus engine and a risk matrix (Figure 1).

The LUT represents a database containing static data, such as driver's name, age, residence, driving experience and eco-attitude, but also the list of hazards characterizing road types, such as straight lines or T-junctions.

Personal information can be inserted by the driver in the Profile section of the GamECAR system, either via the application or via the web platform. His/her eco-attitude is assessed through specifically designed questionnaires, which attempt to deduce also other personality traits, such as driver aggressiveness. The driver can decide to refill a questionnaire, updating his/her profile, while the GamECAR developers could add new questions to refine the user modelling.

The list of hazards is generated by GamECAR researchers on the basis of retrospective analysis of incidents and accidents related to specific road layouts as well as through expert judgment.

For example, some inputs may derive from the study of pre-fatal crash manoeuvres identified in the early stages of the XCYCLE project (Fruhen, L., et al. (2015). Threats and consequences are the other elements attached to the hazards and they contribute to improving the description of the situation (event) under exam with potential harm.

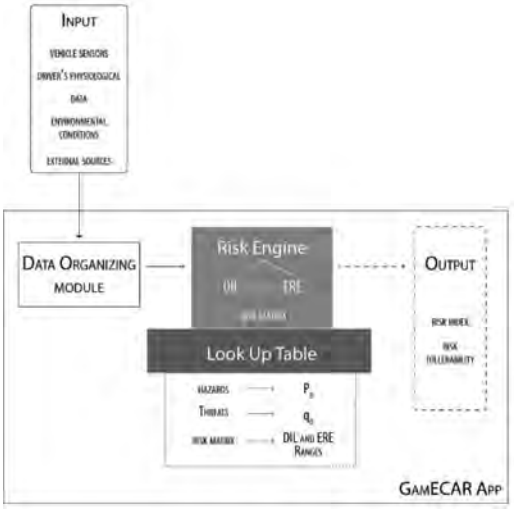


Figure 1. GamECAR risk module's component.

A threat is an external element that may generate or increase the effects of hazards, thus affecting the margins of safety and impacting on the development of incidents and accidents. A consequence is any possible measurable outcome, whether adverse or beneficial, resulting from the evolution of threats and hazards. Each element is represented in the LUT by variables, which are quantifiable by measurable parameters, in order to calculate the probability of occurrence or the weight of the influencing factors in the current situation. Every combination of threat(s), hazard and consequences generates a single incident sequence, which is expressed through its probability of occurrence, of risk.

Even if it the LUT contains only static data, its content can be updated periodically thanks to new research findings, newly discovered hazards or edited personal data.

The assessment engine runs two distinct functions, named DIL (Driver Impairment Level) and ERE (Environmental Risk Exposure). These functions are embedded in the GamECAR application. Therefore, every time a driver activates the GamECAR app and monitors h/her eco-driving performances, these dynamically evaluate the risk associated to the driving task. Their inputs derive both from the LUT as well as from the data measured in real time through the GamECAR sensors, such as a wearable device for the monitoring of physiological parameters and the OBD-II (On-Board Diagnostics) device reporting on vehicle performances.

By being two independent functions, modelling two different aspects of the overall driving activity, in order to obtain a single outcome, they must be combined into a specific matrix where the riskiness of distracting the driver by adding information emerges, together with its tolerability level.

The following three sections will provide more insight into: 1) the human factors aspects, through the evaluation of the Driver Impairment Level; 2) the driving context, i.e., the mean of transport and the surrounding environment; and 3) the matrix, discussing the meaning of its cells and how the tolerability level could be used to tailor the eco-message to display to the driver.

5.1 *The DIL function*

As stated before, the major peculiarities of the road environment are represented by three specific elements, namely: the driver in control of the action, the context in which the driving action is conducted and the transport mean. The DIL function deals with the first one, i.e., the person controlling (driving) a technological tool (i.e., the vehicle). According to Carsten (2007), two broad

types of models of behavior of the human-in-control can be distinguished in the literature. The first type is defined as "descriptive model" and it attempts to describe parts or the whole of a task in terms of what the operator has to do. The second one is named "motivational model" and it aims to describe how keen the operator is in managing risk or abnormal situations.

In GamECAR we concentrate on the second type, by developing a predictive and context-aware model, able to reproduce the effect of a dynamic environment on the driver's performance capabilities. In fact, if task demand exceeds the driver's capabilities, task overload is experienced, resulting in high workload (e.g., De Waard, 1996). Given that driving task difficulty fluctuates in a dynamic environment, instantaneous driver workload also fluctuates. As driving demand changes, driver capabilities also varies with driver state (for example, impaired, fatigued or distracted).

The model is based on five categories of driver capability, performance and behavior, which are related to safety. The outcome is the probability of error making generated by human related factors and possible operational conditions.

The Driver Impairment Level (DIL) focuses on the human related aspects and environmental dynamic circumstances that favour the possibility of error making. Five parameters have been selected as the most essential quantities that influence people's motivational aspects. They are not claimed as representing neither an exhaustive nor an independent set of quantities that characterizes human behaviour. However, they are a first instantiation of several theoretical and applied research studies and the main characteristics of this modelling approach will not be altered by the implementation of different sets of parameters.

The mechanism to describe DIL attempts to combine dynamic changing conditions expressed via Driver State (DS), Situational Awareness (SA) and Environmental Conditions (EC) and more static quantities that reveal the driver Experience/competence (EXP) and Attitudes/personality (ATT) (Cacciabue, P.C., 2010).

More in details:

- Experience/competence (EXP). This parameter enables to consider age and the number of years of driving license. Skills can be developed through practice; when a skill becomes automatic it is consciously controlled and more cognitive resources may be devoted for managing unanticipated events (such as emergency events) (static parameter);
- Attitudes/personality (ATT). Attitude and personality are typical individual traits that result in interpersonal differences and specific behavior. In the context of GamECAR it refers to the ecological and sustainable approach the driver applies not only in driving, but in his/her everyday life. The value is derived from a specifically designed questionnaire and it is updated through the averaged eco-score calculated by GamECAR over the different trips the driver has played. The driver would be asked to fill the form at his/her first access to the application and s/he will have the opportunity to edit the responses anytime. His/Her eco-score will be measured every time s/he performs a trip with the GamECAR App on and it is stored in his/her personal section of the Virtual Community Platform (quasi-static parameter);
- Situation awareness (SA). SA is a very important parameter, as it represents "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley 1994) (dynamic parameter);
- Environment Conditions (EC). Since the surrounding conditions might affect efficiency also at cockpit level, environmental characteristics should not be entirely disregarded, but included as describers of the operational contest. Inputs can come either from sources external from the GamECAR system (e.g., weather applications) or from GamECAR sensors, such as the vehicle OBD-II (dynamic parameter);
- Driver state (DS). The Driver State refers to those psycho-physiological variables that may affect task performance either permanently (e.g. physical/cognitive impairments) or temporarily (e.g. impaired performance due to fatigue or drowsiness) (dynamic parameter).

Driver's bio signals are monitored via wearables devices that provides measurements for heart rate, respiration rate and muscle activity, and transmit them to the GamECAR App via Bluetooth technology. The physiological data are correlated to the driver's psychological parameters, such as SA and DS. For example, by the detection of the heart rate and the setting of proper thresholds, we can derive the driver's level of anxiety or calm. Such data source and quantification process are some of GamECAR's main innovation.

Apart from the Environmental Conditions parameter, all others reflect individual traits.

All parameters vary between 0 and 1 with the meaning expressed as follows:

From a purely logical reasoning, the DIL is assumed to depend these five parameters as follows:

$$DIL_{(Driver,t)} = 1 - e^{-[\varphi+(2-\lambda)/4]}$$

Table 1. DIL parameters and their ranges.

	Novice	Low	Medium	High
EXP	0	0.25	0.5	1
	Bad	Poor	Medium	Good
ATT	0	0.25	0.5	1
	Full	Slightly degr.	Degraded	Loss
SA	0	0.25	0.5	1
	Good	Regular	Poor	Bad
EC	0	0.25	0.5	1
	Very good	Slightly degr.	Degraded	Bad
DS	0	0.25	0.5	1

where:

$$\varphi = \Sigma SA, DS, EC$$

$$\lambda = \Sigma ATT, EXP$$

Note that:

- best driver dynamic conditions = > φ = 0
- best driver static conditions = > λ = 2
- if ATT = EXP = 1 (best static pilots condition), and SA = DS = TD = 0 (best dynamic conditions) then DIL = 0, i.e., Driver Impairment Level is null.

The simple, basic hypothesis is that at the increase of a parameter corresponds an increase in the DIL value.

5.2 *The ERE evaluation*

The Environmental Risk Exposure (ERE) analysis mainly takes into account the other two elements of the road environment, i.e., the scenario and the transport mean. For each one of them, it retrieves inputs from external sources and merges them into a safety assessment of the situation. This data extraction is fulfilled by the sensors utilized in the GamECAR project, which are on-vehicle, in the smartphone, and worn by the driver. With regard to the vehicle, its data are provided via an OBD-II device connected directly to the CAN bus of the car. Real time information includes: active transmission gear, fuel type, consumption and tank level, throttle position, engine RPM, vehicle speed, and other data.

Moreover, the sensors integrated inside any modern smartphone, such as the linear accelerometer and the GPS receiver, enables to collect essential information on the location of the vehicle (i.e., longitude and latitude) and also about its acceleration.

Finally, other meaningful information about the environment are gathered from cloud services, e.g., streets maps, traffic, and climatic conditions.

The ERE builds upon an internal database, i.e. the LUT, a Data Organizer module that retrieves and manages the raw external inputs, and a Risk Engine, which performs the dynamic calculus for evaluating the exposure to hazards in a particular, forthcoming moment of time.

Standard incidental sequences, based on previous retrospective analysis are stored in the LUT. They are described by the well-known Bow-Tie logic, where to a central hazard the threats (or factors) leading to it are linked to its left inside, and the consequences representing the final result of the possible incident sequence are connected from its right inside. At the beginning, an absolute probability value is given to each hazard; moreover, static default values are allocated to each factor representing the baseline for the successive calculation steps. Due to the scarcity of recorded data about road transport, compared to the potential availability, these values have been assigned on the basis of the literature available and, mostly, through expert judgment.

In order to find a suitable trade-off between the theoretical framework, which sustains that more consequences can derive from the same hazard, and the practical real life, where possible incidental sequences share many commonalities and it is complicated to differentiate among them with a sufficiently accurate proves, we decided to link to each hazard only one direct consequence.

Another important database section includes a list of the road types, and a more detailed mapping of the roads sections and junction types, exploiting open source maps. To each element, a value expressing its harmfulness potential is assigned, determined from retrospective evaluations of incidents occurred in that particular spot. Hazards are linked to specific crossroads or streets on the basis of the possible occurrence of the associated direct consequence.

In case the user sets in advance the itinerary of the trip s/he is going to take, the Risk Engine would be able to analyse the specific road path with higher discretization, thanks to the pre-load of the needed information due to a more intense usage of the map info, thus providing more frequent and more refined outcomes.

Data in the LUT are static, but their quality can be refined and the dataset could be enriched with the scope of covering previously unforeseen scenarios.

The preparation of the database is executed offline and when completed becomes the baseline for the real time calculus. The Data Organizer module collects as many inputs possible among those previously described from the on-vehicle sensors and the external sources, and it serves them to the Risk Engine. The online risk calculus

dynamically modifies the values in the LUT, transforming the basic probability of each hazard $P_{0,i}$ into an adapted value $P_{1,i}$, which represents the exposure to risk the user suffers while s/he is driving. The Risk Engine is supported by a function eta (η) that is responsible to update the probability in accordance with the characteristics of the current action scenario.

$$\eta_i(\gamma) = \frac{c}{1 + b * e^{-\gamma}}$$

where:

$$\gamma = \frac{\sum_{i=1}^{n_f} \left(q_{j,i} * l_{j,i}\right) * \mu}{nn_f}$$

c and b are two constants that keep the function from exceeding the range's limit when multiplying it for the hazard probability value;

- $l_{j,i}$ is a vector containing the dynamic values calculated by exploiting the data received from the Data Organizer;
- $q_{j,i}$ is a matrix containing the weighting values for each specific factor$_j$ linked to the hazard$_i$;
- n_f is the number of evaluated factors, while nn_f is the number of not null factors calculated during the real time assessment;
- μ considers risk of each specific hazard, and to personalize the weight of the factor for different road configuration.

In order to transform the basic probability of the hazard$_i$ into a contextualized one, the following formula is applied:

$$P_{1,i} = P_{0,i} * \eta_i(\gamma)$$

This final outcome is a value included between 0 and 1. This is combined with the DIL value into a Risk Matrix designed for the purpose.

6 RISK MATRIX

Given that DIL and ERE are two independent functions, they must be combined into a matrix in order to obtained a single, merged "risk" value. The basis is the well-known risk matrix proposed by ICAO Doc. 9859 (Third Ed.), improved by assigning to each cell a numerical index, enabling to compare among hazards associated to the same location and to range them.

The tolerability range is also assigned by a five-colour code, in order to avoid the risk index to fall too often into a critical area, which might induce misunderstanding or overestimation of the situation. This reasoning arises from the experience of the use of risk matrices in dynamic and proactive assessments, in contrast to retrospective analyses, where a three-colours code could be sufficient.

The major difference relies in the meaning of the Severity. In GamECAR, the Severity is not represented by the potential harm of the consequences, but by the level of impairment of the driver, thus by his/her capacity to promptly react to abnormal situation. Instead of an esteem on the effects of a consequence, the attention is shifted towards the measurement of human performances.

7 CONCLUSIONS

The GamECAR model is still in its designing version since the project is currently at its mid-term. A dedicated software solution is under development and it will be integrated in the project App, linked to the shared database and providing inputs to the display function. This activity is scheduled for Spring 2018 and the so called Evaluation Phase will run in Summer 2018. For this reason, this paper could not present a practical case study where the full methodology has been experimented. Tests will be very significant in relation to the fine tuning of threats' weighting factors. At first, the travelling path will be pre-defined, thus allowing the researchers to fill the LUT with information, such as hazards, road types and harmfulness potential, etc. with more peculiar and accurate data. In the meantime, thanks to the experiences lived during the test-drives, the hazards list, as well as the threats list, can be increased, refined and improved, in order to better represent the complexity of the driving reality.

The possible exploitation opportunities of such model are various. In the GamECAR project framework, it determines which visualization option is more appropriate for the scenario in which the driving task is conducted, in order not to jeopardize safety in benefit of ecology. The same scope could be exploited by other functionalities, technologies, or tools the vehicle is equipped with. Generally, these functions support or customize the driving action, but they are not essential for its core fulfilment.

The main advantage of this model, and the main difference with the ones usually applied in other domains with high automation, such as aviation, is that the information is not presented to the final users, but it becomes a controlling input for another background service(s). This implies that the user does not have to be trained on risk principles, risk variables, index and colour meaning and that the system is suitable for drivers of any experience levels.

ACKNOWLEDGMENT

The outcomes described in this document arise from the activities conducted in the GamECAR project. The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°732068.

REFERENCES

Bressan, E. et al. (2017). Dynamic communication of hazards to cyclist by merging risk assessment and risk exposure. *Safety and Reliability – Theory and Applications.* ESREL 2017.

Cacciabue, P.C. & Carsten, O., 2010. A simple model of driver behaviour to sustain design and safety assessment of automated systems in automotive environments. *Applied Ergonomics* 41: 187–197.

Carsten, O. 2007. From driver models to modelling the driver: what do we really need to know about the driver? In P.C. Cacciabue (Ed.), Modelling Driver Behaviour in Automotive Environments, Springer, London.

De Waard, 1996 The Measurement of Drivers' Mental Workload.

Endsley, M.R., 1994. Individual differences in pilot situation awareness. The international journal of aviation psychology, 4, 241–264.

Fruhen, L. & Flin, R. (2015). Car driver attitudes, perceptions of social norms and aggressive driving behaviour towards cyclists. *Accident Analysis and Prevention*, 83, 162–170.

International Civil Aviation Organization (ICAO) (2013). Doc. 9859 Safety Management Manual (SMM). Third Edition.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

What could adaptive risk management look like in practice?

J.M. Nisula
Risk in Motion and ICS-IRIT, Université de Toulouse III, France

ABSTRACT: Complex adaptive systems pose challenges for risk management, e.g. due to the difficulty to forecast evolutions in the system and to predict system responses to different interventions. Due to the complex cause-and-effect relationships, straightforward actions would often fail to produce the desired effects in the system. A more adaptive approach is required. This paper focuses on the challenges of adaptive risk management and especially adaptive risk treatment. Key features of complex adaptive systems are described, as well as the consequences of these features for risk management. Different types of possible interventions are introduced, including simple actions, adaptive policies and portfolios of experiments. Advantages and challenges of adaptive risk treatment strategies are discussed and some examples are provided. It is argued that the scientific community should promote the use of adaptive approaches compared to the older less dynamic approaches which are not well-adapted for complex systems.

1 INTRODUCTION

Examination of the features of a complex adaptive system leaves no doubt that in our everyday lives we are more and more exposed to—or rather agents within—complex systems. These would include the various social networks, so much accelerated by social media; logistics systems like the transport system in a city; political systems such as the system of governance for a country or a company; and so on. Complexity makes risk management more difficult, especially when a modern risk perspective is adopted, putting a lot of emphasis on the knowledge dimension behind the assessments (see e.g. Cox 2012, Aven 2013, Bjerga et al. 2016).

It is argued that the very dynamic nature of complex systems requires dynamic risk management techniques. Adaptive risk management has been introduced in literature, see e.g. Cox (2012) and Bjerga & Aven (2015). The current paper tries to take a rather practical perspective to how adaptive risk management techniques could be used. Both risk assessment and risk treatment are discussed but the main focus will be on risk treatment.

Examples are picked from the challenging domain of managing transport safety risks.

2 COMPLEX ADAPTIVE SYSTEMS AND RISK MANAGEMENT

2.1 *Features of complex adaptive systems*

Cilliers (1998) presents the following 10 characteristics of complex systems:

1. Complex systems consist of a large number of elements.
2. The elements interact dynamically.
3. The interaction is fairly rich, i.e. any element in the system influences, and is influenced by, quite a few other ones (but there can be large differences between elements in this respect).
4. The interactions are often nonlinear. This means that small inputs in one part of the system can have large results in other parts of the system.
5. Each element is mainly dealing with its immediate neighbors in its local context. However, due to the interactions, influences can propagate through the system, while possibly getting modulated (enhanced, suppressed or altered).
6. There are feedback loops. Any feedback can be positive (enhancing, stimulating) or negative (inhibiting, dampening).
7. Complex systems are usually open systems, having live interactions with the environment and there is a multitude of influences between the system and its environment. The system borders could be defined in many ways depending on the purpose and the position of the observer.
8. Complex systems operate under conditions far from equilibrium being highly dynamic and constantly evolving.
9. Complex systems have a history and their past is co-responsible for their present behavior.
10. Each element in the system is ignorant of the behavior of the system as a whole, responding only to information that is available to it locally.

Reiman et al. (2015) summarize the key features of a Complex Adaptive System in the following way:

> "A complex adaptive system is a collection of individual agents with freedom to act in ways that are not always predictable, and whose actions are interconnected so that one agent's actions change the context for other agents. These agents interact in a non-linear way creating system-wide patterns and higher and higher levels of complexity. The agents differ from each other and none understands the system in its entirety. This diversity is a source of invention and improvisation. As the agents are interdependent of each other, relationships among agents can be considered to be the essence of a complex adaptive system. Understanding a complex adaptive system requires understanding of patterns of relationships among agents."

Complex systems can also be associated with the following concepts (Reiman et al. 2015):

- *Emergence*: properties and behaviors which cannot be deducted from the individual system components appear spontaneously in the system. Consequently, the properties of the whole can be significantly different from the properties of the parts, and the outcomes are not predictable, even when the initial conditions are known.
- *Self-organization*: agent interaction and connections lead to new structures, patterns and new forms of behaviors. Emergence, feedback and non-linearity contribute to self organization.
- *Nested systems*: complex systems are typically part of larger complex systems, thus the expression *system of systems*.

Meadows (2008, p. 11) argues that a system consists of *elements*, *interconnections* and a *function* or *purpose*. To these can be added the more visible aspects: *events* (Senge 2006 p. 21) and system *behavior* (Meadows 2008, p. 88). Senge points out that people are naturally drawn to events and their alleged causes, failing to see the longer-term behavior of the system. The system behavior and the resulting events are visible but from the system point of view they are consequences of the purpose, interconnections and elements. Consequently, the following hierarchy can be established:

- Purpose
- Interactions
- Elements
- Behaviors
- Events

This hierarchy is helpful both for understanding behaviors in complex systems and for guiding interventions. As Dekker (2011, pp. 130–133) points out, understanding a complex system requires going "up and out" (synthesis) rather than following the typical "down and in" (analysis) approach of the Newtonian-Cartesian worldview.

2.2 *Challenges for risk management in complex adaptive systems*

A complex adaptive system introduces many challenges for risk management, affecting both risk assessment and risk treatment.

Knowledge is a key aspect of new risk perspectives. Obtaining precise and comprehensive knowledge in a complex system is particularly challenging. On one hand, knowledge is spread all over the system and cannot be centralized in one place (cf. point 10 in Cilliers' list above). This makes gathering comprehensive knowledge about the system and its behavior virtually impossible. On the other hand, due to the *unpredictability* of the system, many things are simply unknowable in advance. The nonlinearity of the system also reserves potential *surprises* in terms of emerging phenomena and their magnitude. These challenges related to knowledge affect the capability for risk assessment.

Risk treatment suffers from the same unpredictability that affects risk assessment. One may have a good idea of the desired change in the system, however, it may be very difficult to establish an action plan which would achieve the desired outcomes with some level of certainty or repeatability. Methods, tools and techniques which work in simple and complicated systems are not adapted for complex systems (Kurtz & Snowden 2003). Modeling the system would almost certainly cover only some parts of the system behavior and any models would by definition not reproduce the full complexity of the real system. Controlled experiments are not possible because the system is open and keeps changing all the time whether or not this is desirable (Sterman 1994). When a specific intervention is introduced, nobody can predict with certainty what the various consequences within the system will be. Not only can an intervention fail to reach its desired outcomes—it could have different unwanted consequences, some of which may remain difficult to detect.

For example, a new strict rule introduced by the regulator in order to eradicate common shortcuts in aviation maintenance procedures may seem like the typical easy solution from the point of view of the regulator. However, if it misses the true reasons for the shortcuts it may be fairly inefficient. Moreover, the additional negative consequence could be loss of respect and faith in the regulator by maintenance professionals when they perceive that the regulator has not addressed the real problem but only opted for an apparent solution which, to make things worse, threatens them with punitive actions.

3 ADAPTIVE RISK MANAGEMENT

The classical way to act in trying to change a system is often a specific action at a specific point in time. This could typically be a new working method, new procedure, a new piece of regulation, or a reorganization. Such interactions are not adaptive, i.e. they cannot adapt to changing conditions after the implementation.

The main focus in this paper is on adaptive interventions in the context of risk treatment. Especially two types of adaptive interactions are discussed: adaptive policies and experiments.

3.1 *Adaptive policies*

The main idea behind adaptive policies is to develop policies that are not targeted to be optimal for a best estimate future, but which could be robust across a range of futures. Swanson et al. (2010) argue that policies designed for a certain range of conditions often face unexpected challenges outside of that range and this leads to unintended impacts and failures in accomplishing the original goals. Instead of being forced to change policies on an ad hoc basis repeatedly, adaptive policies have a built-in capability to adapt to new conditions. Such an approach is particularly suited for highly complex, dynamic and uncertain settings—so typically for complex adaptive systems.

Another key feature of adaptive policy making is that all major uncertainties do not need to be tackled prior to the implementation phase. As new knowledge is gained during the implementation, the policy is gradually adapted accordingly (Marchau et al. 2010).

Swanson et al. (2010) propose "seven tools for adaptive policymaker", which can be summarized as:

1. *Integrated and forward-looking analysis*. The idea is to try to identify the key success factors for the policy in advance, anticipate problems and to mitigate the foreseeable unintended impacts in advance.
2. *Built-in policy adjustment*. If the future needs for policy adjustment can be anticipated, fully or semi-automatic adjustment mechanisms can be built in the policy.
3. *Formal review and continuous learning*. Ideally a policy is introduced in a phased manner, using so-called policy pilots allowing early testing and adjustments.
4. *Multi-stakeholder deliberation*. Volunteers are used in a very structured manner to get valuable feedback from stakeholders.
5. *Enabling self organization and social networking*. Policies should promote self organization and networking so that local solutions can be discovered without external input.
6. *Decentralization of decision-making*. Both formal and informal feedback gets faster and better when the decision-makers are closer to the people affected.
7. *Promoting variation*. This can be achieved either by several parallel experiments, facilitating an environment where variation can occur, or by using feedback to create variation.

Going from simple one-shot actions to adaptive policies can already be a significant step in embracing the complexity of the surrounding system and creating more suitable solutions for that environment.

Adaptive policy making could very naturally be adopted by regulators and other administrative agencies. However, the same philosophy can also be applied in private companies (and various other types of organizations) in trying to replace company policies, rules and procedures by more adaptive alternatives.

3.2 *Experiments*

As the heart of the problem with complex systems is that any intervention could have unpredictable consequences, the most natural course of action is to organize probes, i.e. set up experiments and use the results to improve the next round of experiments. As Sterman (1994, p. 310) states, effective learning in the world of dynamic complexity and imperfect information can be based on continuous experimentation and feedback. The final aim is to gradually find the most effective solutions through trial and error.

By definition many experiments will fail. It is therefore important that the various cost dimensions (e.g. time, financial cost, potential damages) of the experiments are very limited.

Other key requirements for experiments are:

- Clear criteria for success and failure are necessary; otherwise the iteration towards better experiments will be too slow.
- It must be possible to stop the failing experiments rapidly.
- There must be a way to scale up the successful experiments.

The whole point of experiments is that many diverse strategies can be experimented with. Therefore, a thorough pre-screening of proposed experiments is not necessarily recommendable. A lot of different experiments may be ran, as long as they are safe-to-fail. Experiment proposals should not need to be convincing, they just need to have a certain level of coherence, showing that they *could* possibly work.

In designing experiments, it is good to keep in mind the hierarchy of elements within complex adaptive systems (see above, section 2.1), i.e. modifying the *purpose* or the *interconnections* within a system will probably bring a more tangible and durable change than working at the level of *elements* and *behaviors* (Meadows 2011, pp. 16–17).

3.3 *Example: Adaptive risk treatment based on experiments*

Let's use as an example the problem of high road accident rates of young drivers. The challenge from a transport safety agency's point of view is to find effective interventions to reduce this particular risk.

Instead of trying to figure out a single preferred solution, the solution could be built through experimentation with different types of interventions. It would be important to involve different types of people in creating the experiments, including people outside the typical group of transport safety experts. In particular, it would be valuable to involve some young drivers who are themselves part of the risk group. In this way, different perspectives to the problem can be included and relevant knowledge from the real-life context can be made available.

The working group creating the experiments might come up, for example, with the following experiments:

1. Adding a specific short training which highlights the limits of the driving skills of the young drivers. Typically, the drivers could learn about the speeds and conditions at which they no longer can control the car in a curve or stop the car by breaking. Rationale: reduce the overconfidence of young drivers by illustrating the limits of their skills, thereby achieving a more cautious driving style.
2. Introduce specific driving limitations for young drivers: forbid driving 10 pm-6 am on Fridays and Saturdays and forbid transporting more than one young person in the car, unless if at least one adult is present. Rationale: reduce exposure to dangerous driving situations at nighttime and under negative influence of peers.
3. Introduce specific driving limitations II: require that no weekend driving can take place without an adult present in the car. Rationale: eliminate risky driving during weekends.
4. Expose young drivers to their peers who have suffered a serious accident. Young drivers would be meeting face-to-face one or more people who have had serious consequences after a road accident as a young driver. Rationale: the encounter would give an emotional lesson that carelessness in traffic could have very serious and long-lasting consequences, and the encounter would contribute to a better driving style.
5. Oblige young drivers to take care of potential victims of their careless driving style. Typically, during several weeks the young drivers would be taking young children to school both by using their own car and also as pedestrians. Rationale: young drivers will experience the vulnerability of these people in a very tangible way and also create an emotional link with such potential victims. Being a pedestrian may create situations where the young drivers can observe other drivers' dangerous behavior threatening the safety of the school children that they are now responsible for.

The first experiment is close to what a regulatory action could be even without experimentation. The next two experiments are a little bit more creative but still within the regulatory domain. The last two experiments propose approaches that regulators would typically not use as a part of their interventions. All rationales can be considered coherent and the costs (in the large sense) are small, thus all five experiments can be considered valid.

From the practical point of view, each experiment could be running in a different region or city within the state in question. Probably the most challenging aspect would be to measure the impact of these experiments. Ideally, if one wants to get results in a short time it is not enough to wait for accident data. One should be able to capture feedback for all kinds of safety related problems and also comments from the drivers themselves as well as from any other key stakeholders depending on the experiment. It should be possible to adapt the course of action after a few months of experimentation.

It could happen that one or two approaches turn out to be the most effective ones, perhaps depending on certain conditions (e.g. big city vs. small town). Additionally, it could be that some experiments could be combined or their lessons learned could be fed into the winning strategies as additional elements. For example, it could be decided that all young drivers should "meet" previous young drivers who have had serious accidents, but that this could be made virtually through a film.

Once the winning strategies have been implemented it is important to keep in mind that some fresh experiments should be run in the not-too-distant future.

3.4 *Resilience as a risk treatment strategy*

It is difficult to talk about risk management within complex systems without touching the concept of resilience. Woods (2015) gives four separate interpretations for the concept of resilience:

1. *Rebound*: recovering the functional state after a major disturbance.
2. *Robustness*: increased ability to absorb perturbations.
3. *Graceful extensibility*: the capability to extend adaptive capacity in the face of surprise.
4. *Sustained adaptability*: being able to sustain adaptability over longer periods of time.

Resilience is a valid concept at different scales: e.g. one could talk about the resilience of a *crew* or an *organization* or even the resilience of a specific *system*.

From risk management point of view, a highly resilient organization could cope with virtually any kind of surprising disturbance without losing the operational capabilities. Therefore, building high resilience in an organization could be a preferred strategy compared to having to predict all possible failure scenarios and getting prepared for them. Moreover, for some types of risks—e.g. low-probability scenarios, including black swans—working at the level of individual scenarios is not even feasible due to their almost infinite number.

Achieving higher levels of resilience could also be taken as an objective within an adaptive risk treatment approach, i.e. one of the key goals for various experiments could be to increase the organizational resilience level. An innovative approach would be to use an inverse logic for such experiments: take the increased resilience as a given based on a robust scientific reasoning and use the experiments rather to test how well the organization can cope with the new constraints which come as the downside of the resilience-increasing measures.

3.5 *Dynamic risk assessment & treatment*

In the previous example, the experimental iteration took place within the risk treatment domain. However, it is possible to stretch the loop back to risk assessment and create a risk management process where both risk assessment and risk treatment are part of the adaptive approach.

Risk assessment is based on assumptions. Experiments can produce valuable feedback on the validity of the assumptions and lead to iterative corrections to the risk assessment. In this case, is not only the risk treatment for a given set of risks which may change—the priority to treat various risks may also change.

A very dynamic risk management response can be achieved, if the operational people are fully involved in the assessment and management of risks, and are empowered to take real-time decisions based on the understanding of risks, assumptions and the alternatives for action. This type of dynamic risk management requires an excellent cooperation between the risk management experts and the operational people.

For example, imagine that an airline has made a risk assessment but about its operation to a specific airport. It may have deemed the operational risks acceptable based on several assumptions, one of which could be certain acceptable weather conditions. If the pilots flying to the destination are aware of the risk assessment and the assumptions behind it, they are able to compare the real-world conditions of the day to the assumed ones and detect any mismatches. If the crew one day is faced with weather conditions different from the assumed ones, it would immediately know that the risk assessment is no longer valid. Different alternative courses of action could have been prepared in advance, and the crew could now take the most suitable one of them, for example divert to a close-by airport from which ground transportation to the final destination could be arranged.

Ideally, instead of having a one-time risk assessment produced before the operation, the dynamic approach creates an active risk management tool/process, engaging the expertise both from the risk management experts and the operational people.

3.6 *Macro-view to adaptive risk management*

It is easy to see that adopting an adaptive risk management approach replaces a one-time exercise by a dynamic activity where interventions are iterated continually based on feedback from the system.

Instead of having to make one or two big decisions, smaller decisions are taken, and they are taken more frequently.

Applying adaptive risk treatment techniques also means that at any point in time there will be several portfolios of experiments ongoing. One needs to develop skills to manage such portfolios. Attention may also have to be paid on adaptive policies if they are used.

Based on the results of the individual experiments, some would be expanded, others would be terminated, and some experiments could be merged together.

Like explained in the previous section, the process can be made even more dynamic if risk assessment becomes part of the adaptive loop and if operational people are engaged in a shared risk management and decision-making process together with the risk analysts.

4 ADVANTAGES OF ADAPTIVE TECHNIQUES

The main advantage of the described adaptive techniques should be better effectiveness due to the use of techniques which are better adapted to complex systems than less dynamic alternatives: as

it is very difficult to predict future evolutions and system responses to interventions, instead of trying to model the system it is better to use probes. Also, in a system which is in constant change, it is better to base decisions on fresh feedback from a recent experiment rather than follow a plan made several months or years ago.

As a positive spin-off from frequent experimentation, there is an opportunity to start gradually learning interesting aspects of the system behavior. This type of learning can be helpful in designing future experiments and interventions.

Finally, big decisions may be risky and frightening because of the potential losses if the intervention fails. In addition to the lost improvement opportunity and incurred costs, the failure may be damaging to the image and credibility of the organizations and people involved. The smaller decisions within the adaptive approach are much less frightening as the stakes will be smaller each time, and as the whole context is explicitly a context of experimentation where *graceful failure* is an acceptable option.

5 CHALLENGES OF ADAPTIVE TECHNIQUES

There are systems and phenomena which do not easily lend themselves to experimentation. First, by their nature some systems do not produce enough feedback so that the success of the experiment can be assessed. For example, a very safe system does not produce much feedback in the form of events where safety was compromised. Trying to test interventions aiming at further safety improvement will therefore suffer from the fact that there is not going to be enough feedback within a reasonable time period.

Secondly, one also has to accept that running a specific experiment successfully in a specific place at a specific time does not necessarily mean that scaling up the same idea in the same or in a different place, will again be a success. One has to remember that these experiments are not controlled experiments where various parameters can be frozen—because in a complex system, key parameters cannot be frozen.

Third, there may be man-made restrictions for experimentation. For example, it may be very difficult to set up experiments in a tightly regulated system. Work may be highly proceduralized, and testing new procedures in real working conditions may be judged too risky.

Finally, moving from traditional working methods to adaptive risk management is a significant change and brings with it several psychological barriers. Accepting the features of complex systems means also accepting the associated unpredictability and uncertainty. Both the people carrying out the risk management and the public might perceive admitting the uncertainty negatively. As Tickner & Kriebel (2006) point out, acknowledging uncertainty can weaken agency authority by creating an image of the agency as unknowledgeable and by threatening the objectivity of science-based standards.

The experiment-based working method itself might not look very controlled or scientific either, and could further deteriorate the credibility of the organization implementing it.

The challenge grows even bigger if one wants to explore unconventional interventions. This would typically require testing some very surprising and "crazy" ideas. It would also require involving people who are not experts in the subject matter or at least not perceived as experts. All such factors raise the bar higher for organizations to start applying modern adaptive approaches.

6 CONCLUSIONS

This paper has described typical features of complex adaptive systems and discussed the challenges that such systems pose to risk assessment and risk treatment.

It is argued that adaptive risk management is better adapted to complex systems than other classic approaches. The principal aim of this paper has been to illustrate how adaptive risk management and especially adaptive risk treatment could be used in practice. Three different approaches have been promoted: adaptive policies, experiments, and enlarging the adaptive iteration to cover also the risk assessment part of the process. The theoretical content has been complemented with a concrete example from the domain of road safety.

The key argument for promoting the use of adaptive risk management is the expectation of reaching better results in managing risks in the complex system.

However, several important challenges in implementing adaptive approaches have also been identified.

It is proposed that the scientific community should help promote adaptive risk management as a respected method which could be implemented in different organizations—including governmental agencies—without the fear of losing credibility in the eyes of peers or the public.

REFERENCES

Aven, T. 2013. Practical implications of the new risk perspectives. *Reliability engineering and system safety* 115: 136–145.

Bjerga, T. & Aven, T. 2015. Adaptive risk management using new risk perspectives—an example from the oil and gas industry. *Reliability engineering and system safety* 134: 75–82.

Bjerga, T., Aven, T., Zio, E. 2016. Uncertainty treatment in risk analysis of complex systems: The cases of STAMP and FRAM. *Reliability engineering and system safety* 156: 203–209.

Cilliers, P. 1998. *Complexity and postmodernism. Understanding complex systems*. London & New York: Routledge.

Cox, L.A. 2012. Confronting deep uncertainties in risk analysis. *Risk Analysis* 32(10):1607–1629.

Dekker, S.W.A. 2011. *Drift into failure. From hunting broken components to understanding complex systems*. Farnham: Ashgate.

Kurtz, C.F., Snowden, D.J. 2003. The new dynamics of strategy: sense-making in a complex and complicated world. *IBM systems journal* 42 (3): 462–483.

Marchau, V.A.W.J., Walker, W.E., van Wee, G.P. 2010. Dynamic adaptive transport policies for handling deep uncertainty. *Technological forecasting & social change* 77: 940–950.

Meadows, D.H. 2008. *Thinking in systems*. White river junction (VT): Chelsea green publishing.

Reiman, T., Rollenhagen, C., Pietikäinen, E., Heikkilä, J. 2015. Principles of adaptive management in complex safety critical organizations. *Safety Science* 71: 80–92.

Sterman, J.D. 1994. Learning in and about complex systems. *System dynamics review* 10(2–3): 291–330.

Senge, P.M. 2006. *The fifth discipline. The art & practice of the learning organization*. Revised edition of 2006. London: Random house group.

Swanson, D., Barg, S., Tyler, S., Venema, H., Tomar, S., Bhadwal, S., Nair, S., Roy, D., Drexhage, J. 2010. Seven tools for creating adaptive policies. *Technological forecasting & social change* 77: 924–939.

Tickner, J., Kriebel, D. 2006. The role of science and precaution in environmental and public health policy. In Fisher, E., Jones, J., von Schomberg, R. (eds.) *Implementing the precautionary principle.* Northhampton, MA, USA: Edward Elgar publishing.

Woods, D.D. 2015. Four concepts for resilience and the implications for the future of resilience engineering. *Reliability engineering & system safety* 141: 5–9.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Risk indicators for safety performance assessment of crane-operations in the chemical industry

G. Ancione & M.F. Milazzo
University of Messina, Messina, Italy

N. Paltrinieri
Norwegian University of Science and Technology NTNU, Trondheim, Norway

ABSTRACT: In recent years, new risk analysis approaches have been developed to be applied in major hazard establishments. Amongst these, approaches associated with the overall concept of dynamic risk are the most relevant. The main reason for the diffusion of Dynamics Risk Analysis (DRA) is the need to integrate new notions, accounting for dynamics of phenomena and information on system changes over the time and/or the impact of innovative technologies, within risk assessment procedures. DRA approaches are particularly useful when assessing the risk due to the lifting of load, especially when hazardous substances are handled or in major hazard industries. The reason is the potential for loss of containment, which could be followed by catastrophic scenarios (such as fires, explosions and toxic dispersions). When carrying out a crane operation, a common hazardous situation, leading to accidents, is associated with the hindered view of the workspace for the crane-operator. At this regard, a recently developed real-time monitoring tool, named *Visual Guidance System* (VGS), can be used to assist the worker during the load lifting. It gives back an alarm, when a potential impact between the handled load and any obstacle in the workspace is occurring. Data's collection, through the feedback provided by the VGS, allows deriving appropriate indicators correlated to the safety of crane operations in chemical industry. These indicators are continuously updated; therefore, they are useful parameters to be integrated within DRAs. This work aims at the definition of *risk indicators* for lifting and handling operations; the methodology to derive such indicators is described and an application is also given.

1 INTRODUCTION

Safety in crane-related operations is a complex issue, especially in the chemical industry, where an accident, due to a wrong load lifting or handling, could lead to severe incidental scenario (Milazzo et al., 2016). In this context, Dynamics Risk Analysis (DRA) approaches can be particularly useful in managing the risk. Crane accidents can be the cause of losses of containment, which can be followed by events, such as fires, explosions and toxic dispersions. More generally, approaches for dynamic risk assessment are based on the use of models integrating parameters that change over the time. Dynamic factors impact on both frequencies and consequences of incidents and, thus, on final risk results. Moreover, it is well-known that the integration of real-time monitoring data offers the opportunity to achieve a more effective control of activities, carried out in the workplace in view of worker safety, by allowing the prevention of accidents and the timely implementation of protective actions. Currently the use of DRA is becoming more widespread, some examples from the literature are given in the following: Eide and co-authors carried out dynamic environmental risk assessment for oil tankers, sailing along the North Norwegian coast (Eide et al, 2006; Eide et al. 2007); Milazzo et al (2009) made use of the concept of *dynamic geoevent* to represent a dynamic evolution of toxic dispersions. As pointed by Paltrinieri and Reniers (2017) and Paltrinieri and Khan (2016), DRA allows improving decision-making and supporting critical risk communication; it can also be used to describe the impact of innovative technologies on the overall safety.

In the chemical process industry, risk arises from complex systems and their management requires a large number of control measures (De Rademaeker et al., 2014). In this context, a common practise is to track performance of activities by using *indicators*, in order to continuously improve the safety and the operability. As defined by Øien (2001a), an *indicator* is a measurable/operational variable that can be used to concisely describe a broader phenomenon occurring when a plant is operating.

A small number of *key indicators* can monitor the status of whole systems. In the chemical industry, the most relevant *indicators* are those used to appraise safety or risk performance of systems. The terms *safety indicator* and *risk indicator* are distinguished by Øien et al. (2011). A *risk indicator* is a risk influencing factor, i.e. an event/condition that affects the risk level of a system/activity; whereas a *safety indicator* is a factor that has an effect on safety as it is related to some measures, different than risk metrics (as number of accidents or incidents or other). Thus, *risk indicators* are derived from a risk-based approach (Øien, 2001b), whereas *safety indicators* may be developed from a safety performance-based approach. *Indicators* are also distinguished as *leading* and *lagging indicators* (HSE, 2006): *leading indicators* represent a form of proactive monitoring of the effectiveness of a Risk Control System (RCS), by providing feedback about safety outcomes before an incident occurs; whereas, *lagging indicators* represent a form of a reactive monitoring of the effectiveness of a RCS, given that they provide feedback after the occurrence of a negative event.

Approaches to develop *safety* and *risk indicators* are grouped in two perspective typologies by Øien et al. (2011), i.e. *technical-human-organisational perspective* and *predictive-versus-retrospective perspective*. The first perspective allows developing *safety indicators* as it searches for causes of accidents occurred in the past, starting from technical to human and further to organisational causes (Leveson, 2004). The second one gives *risk indicators* and aims at predicting potential accidents by including all possible causes or by trying to establish the causes after the event (according to a retrospective point of view); this approach requires the use of quantitative risk models.

The above brief review underlines that the prevention of major accidents can benefit from both the use of dynamic risk analysis techniques and *safety/risk indicators*. The application of dynamics risk assessment techniques based on proactive indicators is suggested by Paltrinieri et al. (2016), it brings additional benefits, since the risk analysis is supplemented by information related to the early warning, which supports to manage in advance unwanted events. The integration of a set of collected indicators provided the risk assessment with dynamic and proactive features. Data collection and processing, for the purpose of such a DRA, take advantage of information technology supporting real-time data collection, sharing, processing, visualization, etc. According to Paltrinieri et al. (2016), dynamics risk assessment techniques based on proactive indicators can be classified in four levels by referring to the basic theory and provided results. The first level concerns to the use of *safety indicators*, it takes into account the effect of technical, human and organization factors; the second one is related to the use of *risk indicators;* thus, the application of risk models is needed; the third level refers to the application of techniques for frequency updating; finally, the forth level concerns the use of techniques for the aggregation of information, which are provided by indicators. This aggregation allows an accurate assessment of the variation of overall risk, also based on real-time data.

This work aims at the derivation of *risk indicators* for the load lifting and handling in chemical industry, to be integrated into the dynamic risk analysis procedure. The paper is structured as follows. Section 2 shows the methodology for the derivation of these indicators based on the approach proposed by HSE (2006); some indications are given on how *indicators* could be integrated into the dynamic risk analysis. Section 3 shows the description of a case study in which the application and validation of the indicators is carried out. Section 4 comments about the results obtained from this study. Finally, in Section 5 the conclusions of the work are presented.

2 METHODOLOGY

To derive *risk indicators* for the purpose of this study, the HSE approach (HSE, 2006) has been used. Figure 1 shows a simple scheme of the main steps of this general procedure, which is usually used to develop indicators related to processes. It is based on the development of *lagging* and *leading indicators* for the facility, related to each Risk Control System (RCS) and associated with a previously identified hazardous event, the aim is to prevent, control or mitigate major accidents.

2.1 *Identify the scope of indicators*

According to this methodology, to develop key indicators of a facility, preliminarily, it is fundamental to define the scope of such indicators. This allows selecting the proper information about the adequacy of safety controls. It is important to focus on a few critical indicators, giving a sufficient overview of the performance of systems and/or a more detailed picture in a hierarchical form.

To develop facility level indicators, major accidents that could involve the equipment have to be identified. Each indicator should refer to a RCS of the facility, then, it can be weighted to reflect its importance in guaranteeing the safety at that facility. Different weights will reflect criticalities of such RCS on that equipment, e.g. plant design might be the most critical RCS on Facility 1 in

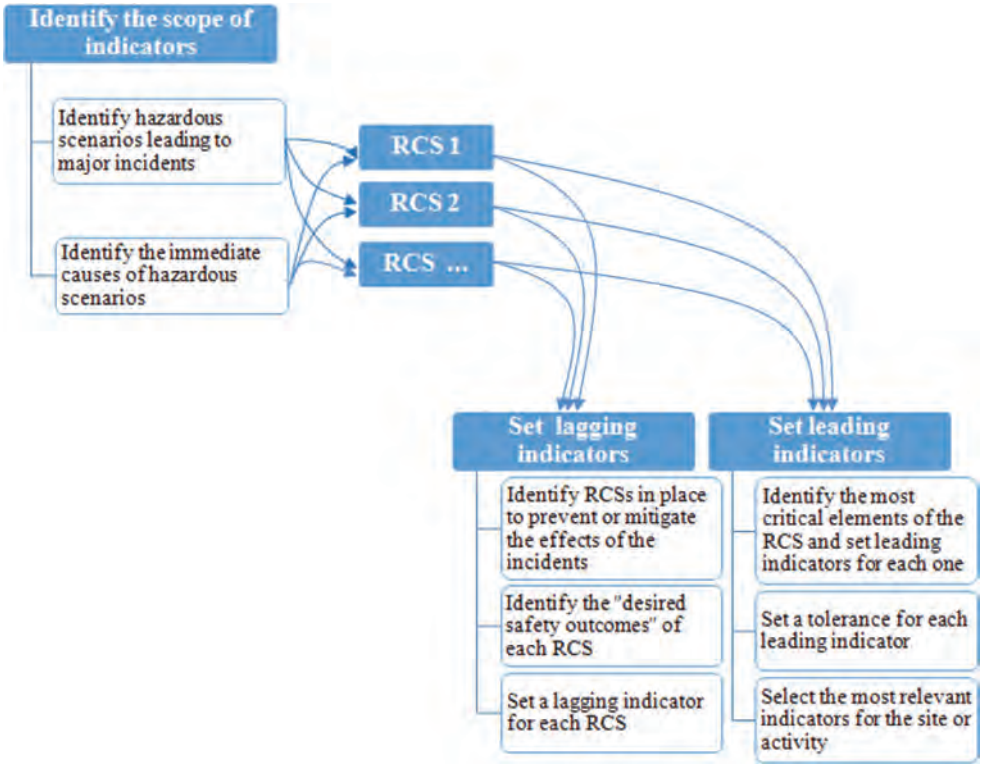


Figure 1. Main steps of the HSE approach for the development of risk indicators.

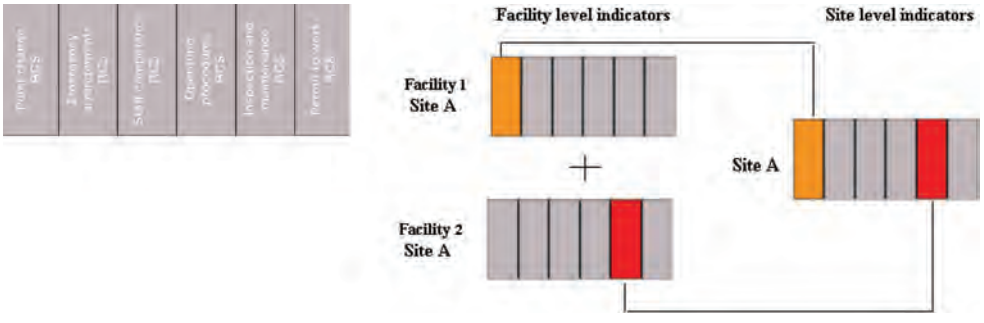


Figure 2. Development of facility level indicators.

Figure 2, whereas inspection and maintenance could be the most important RCSs at Facility 2.

2.2 *Set lagging indicators*

The identification of the hazard scenarios and the description of how these events can be generated help the analyst to focus on the most relevant activities and to identify which indicators should be set. The HSE methodology suggests considering the primary failure that gives an incident. Furthermore, attention must be given to areas where potential criticalities are present, i.e. where near-misses or past incidents have already occurred and where information from audits and inspections have collected.

Next step is to provide a list of RCSs, needed to prevent or mitigate the consequences of each hazardous scenario; the starting point is to identify their primary cause. Therefore, related safety

outcomes should be defined by means of the description of these hazard scenarios. The desired safety outcome represents the success for each RCS. For this reason, the safety outcomes must be clearly identified in term of "success/failure", otherwise it will be impossible to classify properly related indicators. At this moment, a *lagging indicator* for each RCS should be defined to highlight whether the desired goal is actually achieved.

2.3 *Set leading indicators*

After the identification of proper *lagging indicators*, the HSE approach suggests setting *leading indicators* for each critical element of all risk control systems (i.e. those actions or processes that must function correctly to deliver the outcomes). These indicate whether the RCSs provide the designed safety outcomes. The following factors must be considered to identify what are the most important critical aspects that the RCSs should cover to deliver the desired outcome:

- activities that must always be performed correctly;
- elements of the system that are susceptible of deterioration over time
- activities that are most frequently performed.

A range of tolerance for each *leading indicator* should be set. This is important, to permit to capture the analyst attention if deviations in performance are flagged up. Relevant information from *indicators* must be readily obtained, as well as the presentation of data collected should be as simple as possible to permit a prompt comprehension of them.

Finally, a review of all *indicators* implemented should also be executed, with the aim to highlight poor performance of them or just of a part of them. This means that if *leading indicators* of a RCS show poor performance, whereas *lagging indicators* give back satisfactory results for the same system, there is clearly a discrepancy. In this case, a system reviews are recommended to understand the reason of such discrepancy, because to improve the performance of the facility, or activity and so on, each deviation from the intended outcome or failure of a critical part of a RCS must be followed up. Each occasion to review provides an opportunity to consider whether improvements should be made.

2.4 *Indicators measuring the safety in load lifting and handling in chemical industry*

This paper focuses on major accidents due to use of equipment for the lifting and handling of loads in chemical industry, which involves the load and objects located in the workspace, i.e. workers or other equipment (Cheng and Teizer, 2014, Spasojević Brkić et al., 2015). In conducting such operations, there could be situations in which the operator does not have an entire visibility of the workspace. In these cases, a widespread practice is to be supported by an intermediary, which usually gives a guide to the crane-operator to correctly navigate the load. Under this condition, the crane-operator is subject to a high level of stress, as he/she has to trust in the judgment of others in order to carry out an operation for which he/she is absolutely responsible for. Indeed, one of the most important causes of accidents is the human error, which is underlined under one of the following forms distraction, wrong communication, inexperience, etc. For this reason, this contribution includes planned operator activities to control risk within the group of RCSs.

To avoid the cause of accidents associated with a partial or a total hindered view of the workspace, a real-time computer-aided visual guidance system has been recently developed, which is simply indicated as VGS (Ancione et al. 2017). This safety device is composed by hardware and

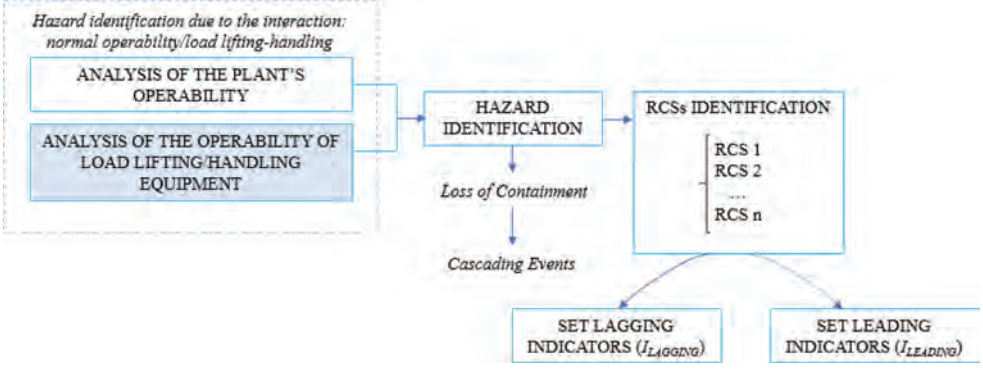


Figure 3. Methodology for the development of indicator for load lifting/handling in chemical industry.

software; it was designed to provide support to crane-operators in navigating the load. The use of the VGS for the execution of a load lifting or handling represents a supplementary RCS, aiming at the prevention of the release of hazardous substances, caused by collisions between the load and the equipment containing them. To identify the safety-related parameters for the activities associated with the crane operability, the HSE method has been adapted as shown in Figure 3 by including the investigation of the hazards due to the interference between the plant normal activity and the operations carried out with the crane.

3 CASE STUDY

The case study, analysed in this paper, is a reconstruction of an alkylation unit of a refinery, where a crane accident occurred. This event happened in 1987 in Texas (USEPA, 1993); it caused the release into the atmosphere of hydrofluoric acid, an extremely corrosive substance, forcing the authorities to evacuate thousands of people from their homes in the surrounding the refining plant. Due to the toxic dispersion more than 800 people were sent to the hospital.

The crane was lifting a multi-ton heater, which was accidentally dropped onto the top of a hydrofluoric acid storage vessel. This facility was being moved for repair and maintenance during a general plant turnaround. The dropped heater severed a 4-inch acid loading line and an inch pressure relief line causing the hydrofluoric acid to be released. Figure 4 shows the layout of the hydrofluoric acid service in the alkylation unit, where the crane that was temporary installed, and the trajectory for the load handling. All major equipment is labeled. The bulk of the acid was located in the storage drum, the settler, the acid cooler and the piping between the settler and cooler. Smaller volumes were contained in the fractionator, the fractionator accumulator, the splitter column and the acid rerun column. The acid unloading equipment was in southern boundary of the overall unit.

The event was investigated by OSHA (USEPA, 1993), which concluded that various causes contributed to the accident:

- not instituting accepted engineering control measures to prevent the release (i.e. emptying hydrofluoric acid vessel before hoisting a heavy load over it and not hoisting a heavy load over a hydrofluoric acid tank);
- the crane was not properly blocked (wooden blocks supporting crane outriggers were crushed);
- crane inspection documentation was not prepared;
- the crane safety devices were not inspected prior to use and a malfunction occurred.

It is also possible that, due to the evidence of an anomaly, which forewarned the accident, the crane operator decided to drop part of the moving system, in order to avoid a major accident due to impact with the vessel and, therefore due to the limited visibility and the complexity of the unit, he dropped it in the wrong place.

In general, the risk control system adopted by company to prevent a loss of containment and/or mitigate its risk, are: operational procedures, emergency procedures, training of workers, periodic inspection and scheduled manutention. The presence of the VGS, in that context, would have assisted the crane-operator in the navigation of the load by indicating where he would have dropped it, thus avoiding the chain of events that led to the release of the toxic cloud.

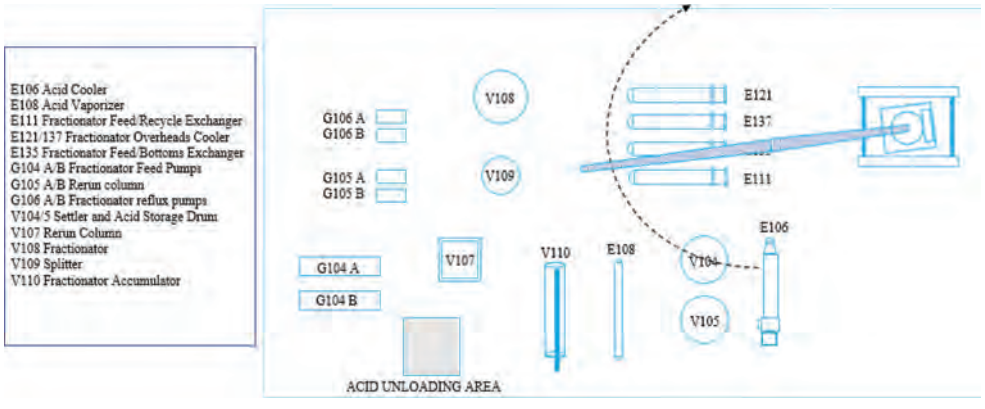


Figure 4. Layout for the hydrofluoric acid service in the alkylation unit.

4 RESULTS

The methodology described in the Section 2.4 has been applied to the case study. Initially, hazard scenarios and initial causes, which can lead to a major accident, have been defined. Then, all RCSs, that are in place to prevent or mitigate the effects of major accidents, have been identified.

Figure 5 shows a logic scheme for the identification of the incidental sequence that could occur during the crane activity. On the left side the initial cause for the hazard scenario is given, this represents what can go wrong when the crane is operating. Amongst various causes of failure for this activity, one appears associated with the hindered view and potentially leads to a major accident. It is represented by a dropped load and could be the cause of a loss of containment (hazard scenario), from which a toxic dispersion of hydrofluoric acid is originated.

The risk control systems, adopted by the Company, for the prevention and mitigation of losses of containment due to crane-related operations, are:

- Plant operating procedures (if these are not correctly followed, facilities age and, due to impacts, LOCs are more likely)
- Crane operating procedures
- Inspection and Maintenance procedure
- Work permit procedure
- Emergency procedure
- VGS

Compared to the use of traditional RCSs (considered in the case of the accident described in Section 3), the installation of the VGS has been included in the assessment.

Table 1 lists all relevant traditional RCSs that usually are adopted by the Company to prevent, control and mitigate the loss of containment. For each of them the desired outcomes are indicated, as well as the *lagging indicator* controlling the achievement of the related desired outcome. Then, critical elements of each RCS have also been

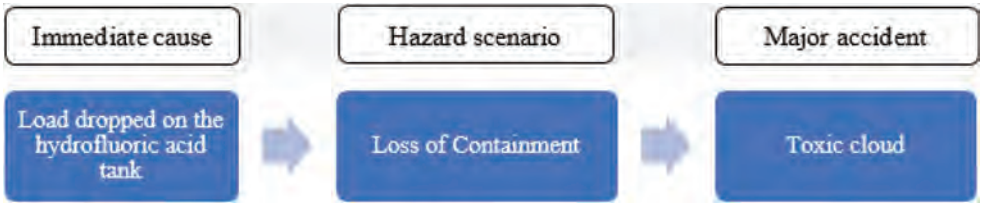


Figure 5. Schematic view of the identified hazard scenario for the case study.

Table 1. Lagging indicators.

RCS	Desired outcome	Selected lagging indicator
RCS1: Operating procedures	– Correct tank selection and operating the substance transfer – Correct cleaning, isolation and emptying	– No. of times the substance transfer does not proceed as planned
RCS2: Inspection and maintenance procedures	– No unexpected loss of containment due to failures of the crane or a not correct handling of the load or failure of the control instrumentation. – No fires or explosions caused by faulty or damaged electrical elements.	– No. of unexpected LOCs due to failures of the crane or a not correct load handling or failure of the control instrumentation.
RCS3: Crane operating procedures	– Correct execution of the load lifting/handling – Identification of lifting restriction areas – No involvement of escape route in the load trajectory – No interference with other equipment operations	– No. of times the load lifting/handling does not proceed as planned
RCS4: Work permit procedure	– High-risk maintenance activities are undertaken in a way that will not cause damage/injury	– No. of incidents due to error during maintenance
RCS5: Emergency procedures	– Minimum consequence in case of LOC	– No. of elements of the emergency procedures that fail
RCS6: VGS	– Full view of the working area and warning in case of approaching collision between load/crane and any obstacle in the workspace	– No impacts load/crane-obstacles

identified in Table 2 with the aim to define the potential *leading indicators*.

The most relevant indicators (a *lagging indicator* and one or two *leading indicators*) for the activity under analysis have been selected; this selection has been made by referring to the criterion that the most relevant criticality per RCS has to be represented.

Table 2. Leading indicators.

RCS	Critical elements	Selected leading indicator
RCS1: Operating procedures	– Procedures contain all elements (key actions, tasks including emergency actions – Procedures are clearly written and easy to be understood – Procedures are kept up to date	– Percentage of procedures reviewed and revised within the reference period
	– Information and training covering: hazardous properties of products, communication systems pre-transfer checks, load transfer controls and monitoring, post-transfer checks and emergency actions	– Percentage of staff attending safety courses within the reference period.
RCS2: Inspection and maintenance procedures	– Scope and frequency of the inspection and maintenance – Failures of critical elements of the crane and identified malfunctions	– Percentage of critical elements of the crane inspected and repaired
RCS3: Crane operating procedures	– Procedures contain all elements (key actions, tasks including emergency actions) – Procedures are clearly written and easy to be understood – Procedures are kept up to date – Execution of risk analysis for crane-operation	– Percentage of procedures reviewed and revised within the reference period – Percentage of activities covered by a preliminary risk assessment
	– Training covering: hazardous-properties of products handled, communication systems, load transfer controls and monitoring, and emergency actions	– Percentage of staff trained within the reference period.
RCS4: Work permit procedure	– Scope of activities, covered by the permit-to-work is clearly identified – Permits specify hazards, risks and control measures – Permits are only issued according to proper authorisation procedures – Duration of the permit – Work is conducted as per permit conditions, including demonstration of satisfactory completion of work	– Percentage of permits to work issued where the hazards, risks and control measures are adequately specified – Percentage of work conducted in accordance with permit conditions
RCS5: Emergency procedures	– Emergency plan covers all relevant elements (testing of emergency plan, raising alarm, shutdown/isolation procedures, firefighting, communication, evacuation	– Percentage of elements that have not failed – Percentage of staff/contractors who take correctly emergency actions
RCS6: VGS	– Device correctly shows the workspace and including elements – Alarm activated at the desired set points – Knowledge of tasks and relevant experience about substances, work processes, hazards and emergency actions	– Percentage of correct indication of the view – Percentage of warning at the set point – Percentage of staff having at least 10 years of experience

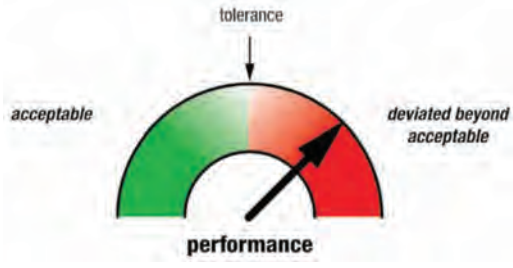


Figure 6. Example of tolerance set (source: HSE, 2006).

The HSE approach suggests that, after the definition of indicators, a tolerance value for each *leading indicator* has to be assigned with the aim to observe any deviation in performance, with respect the related RCS. This allows the plant management deciding the actions to be taken to restore the system from its deviation. For instance, for the indicator "*Percentage of elements that have not failed during simulation*", selected for the RCS "*Emergency procedures*", the tolerance may be set as zero, which means that 100% of actions must be correctly performed. Alternatively, it could be accepted a certain degree of failure if it has previously been highlighted to by the management team, this means that the tolerance of the indicator could be set below 100%.

Figure 6 shows an example of tolerance set for a *leading indicator*. However, the aim of this paper is the definition of *risk indicators* for lifting and handling operations and then, the assignment of tolerance levels for each leading indicator is objective of future study.

As underlined above, the choice of *indicators* that provide a warning when collisions could lead to LOCs, is based on the ability to monitor if RCSs are operating as desired. Moreover, the use of a proper set of *risk indicators*, which have to be collected, processed and evaluated on a regular basis, represents the fundamental for the integration of dynamic features in risk assessment and carrying out a Dynamic Risk Assessment.

5 CONCLUSION

Risk indicators for the load lifting and handling in chemical industry have been developed, according to the HSE methodology, which has been properly adapted with the aim to be integrated into the dynamic risk assessment procedure.

The application and validation of such *indicators* has been made by referring to a case study of an alkylation unit of a refinery. Firstly, the hazardous scenario (loss of containment from a hydrofluoric acid tanks), that can lead to a major accident (toxic dispersion), and its initial cause (dropped load) have been identified. This preliminary analysis allowed identifying RCSs in place to prevent or mitigate the effects of the dispersion. For each RCS, *lagging* and *leading indicators* have been defined.

Resulting indicators highlighted criteria for optimize the process by accounting for the proper safety levels. The use of a safety device, such as the VGS, could be an effective solution to control crane-operations, to guarantee the safety with respect to LOC initiated by a hindered view of the workspace and, finally, to integrate dynamic parameters within risk models in view of the execution of a DRA.

ACKNOWLEDGMENT

This work is partly supported by INAIL (Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro) within SmartBench project grant BRIC 2016 ID 15. Dr. Giuseppa Ancione thanks the NTNU and Prof. Nicola Paltrinieri for support given in the period spent in Trondheim as visiting researcher.

REFERENCES

Ancione G., Kavasidis, I. & Milazzo, M.F. (2017). Improving safety of crane-related operations in chemical industry by the support of a real-time computer-aided visual guidance system. *Safety and Reliability – Theory and Applications*. Proceeding of ESREL2017, 1787–1792.

Cheng, T. & Teizer, J., (2014). Modelling tower crane operator visibility to minimize the risk of limited situational awareness. *ASCE Journal of Computing in Civil Engineering*, 28(3).

De Rademaeker, E., Suter, G., Pasman, H.J., Fabiano, B. (2014). A review of the past, present and future of the European loss prevention and safety promotion in the process industries. *Process Safety and Environmental Protection*, 92(4): 280–291.

Eide, M., Endresen, Ø., Ervik, J.L., Brett, P.O. & Røang, K. (2006). Intelligent ship traffic monitoring for oil spill prevention: risk based decision support building on AIS. Marine Pollution Bulletin 54(10): 145–148.

Eide, M.S., Endresen, Ø., Breivik, Ø., Brude, O.W., Ellingsen, I.H., Røang, K., Hauge, J. & Brett,P.O. (2007). Prevention of oil spill from shipping by modelling of dynamic risk. Marine Pollution Bulletin 54(10): 1619–1633.

Health and Safety Executive, HSE (2006). Developing process safety indicators. HSE book.

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science*, 42: 237–270.

Milazzo M.F. Ancione, G., Lisi, R., Vianello, C. & Maschio, G., (2009). Risk management of terrorist

attacks in the transport of hazardous materials using dynamic. *Journal of loss prevention in the process industries* 22(5): 625–633.

Milazzo, M.F., Ancione, G., Spasojevic Brkic, V. & Vališ, D., (2016). Investigation of crane operation safety by analysing main accident causes. Risk, Reliability and Safety: Innovating Theory and Practice: Proceeding of ESREL2016, 74–80.

Øien K. (2001a). *A framework for the establishment of organizational risk indicators.* Reliability Engineering and System Safety 74: 147–167.

Øien K., (2001b). Risk Control of Offshore Installations. A Framework for the Establishment of Risk Indicators. Department of Production and Quality Engineering, PhD thesis. Norwegian University of Science and Technology NTNU, Trondheim, Norway.

Øien K., Utne, I.B. & Herrera, I.A., (2011). Building Safety indicators: Part 1 – Theoretical foundation. *Safety Science,* 49:148–161.

Paltrinieri, N. & Khan, F. (2016). Dynamic Risk Analysis in the Chemical and Petroleum Industry: Evolution and Interaction with Parallel Disciplines in the Perspective of Industrial Application, Butterworth-Heinemann.

Paltrinieri, N., Landucci, G., Nelson, W.R. & Hauge, S. (2016). Proactive Approaches of Dynamic Risk Assessment Based on Indicators. In: *Dynamic risk analysis in the chemical and Petroleum industry*, Chapter 6, pp. 63–73. Ed. Paltrinieri & Khan 2016.

Paltrinieri N. & Reniers, G. (2017). Dynamic risk analysis for Seveso sites, Journal of Loss Prevention in the Process Industries. 49: 111–119.

Spasojević Brkić, V., Milazzo, M.F., Brkić, A. & Maneski T. (2015). Emerging risks in smart process industry cranes survey: SAF€RA research project SPRINCE. *Serbian Journal of Management* 10(2): 247–254.

U.S. Environmental Protection Agency (USEPA), Chemical Emergency Preparedness and Prevention Office, (1993). Hydrogen Fluoride Study: Report to Congress, Section 112(n)(6) Clean Air Act as Amended: Final Report.

Natural hazards

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

A multidimensional risk evaluation framework for managing floods in urban areas

L.B.L. da Silva, R.P. Palha, M.H. Alencar & A.T. de Almeida
Center for Decision Systems and Information Development (CDSID)—Universidade Federal de Pernambuco (UFPE), Recife-PE, Brazil

ABSTRACT: The increasing migration of people to urban areas around the world contributes to increasing man's adverse impact on the environment and therefore on climate change which is an issue that presents one of the most important challenges of modern civilization. In urban areas, this impact leads to the occurrence of many extreme events, which means multidimensional losses. Over the decades, hydrological catastrophes, especially floods, have been studied so as to control and monitor risks. However, this type of risk assessment has to meet multiple objectives that often conflict with each other. This paper proposes a multidimensional framework which allows researchers to use approaches that support decision makers in the modelling and analysis of the risk of flooding in urban areas. This includes rating risk while considering multiple factors and to responding to risk using actions, such as mitigation measures, which prioritize preventive actions to combat disasters throughout a cyclical process. Some potential benefits of using this approach are discussed.

1 INTRODUCTION

Ever since human beings first formed communities, they have felt the need to interact with each other and to be in harmony with the environment. In fact, living in isolation is not conducive to creating a common culture, to using language or to exercising the human mind (Chatfield 2016). This means not only that people are interdependent—which they rarely admit—but also that it is interdependence which underpins the progressive development of technology which is used to adapt this environment so as to bring comfort to a social group. A look at several points in history confirms this behavior, thereby revealing why human beings have needed to harness and exploit the natural resources at their disposal and how they have done this.

Thus, society has a cyclical co-dependence on technology; in other words, they cannot be separated. This reality in post-modern society has become much more apparent, for example, in terms of communications, forms of transport, modes of learning and doing business or of the artefacts of comfort. Relevant tools or practices have been developed which simplify the way humans do things.

In spite of technology in itself not being harmful to society, its use to achieve specific goals can give rise to negative impacts on human life.

As to the environmental dimension, some features of technology are, in fact, designed to control and monitor natural phenomena. However, the growing dependence on technology in all spheres of life has nevertheless led to global pollution levels increasing because of human behavior. This has had an impact on climate change which coupled with other negative human behavior has caused natural disasters (Merchant 2014).

When the focus of attention turns to urban areas, it is apparent that the unrestrained urbanization of cities has created fragmented spaces into which people have been segregated, thus making cities more vulnerable not only to social inequality but also to the consequences of environmental degradation. As a result, several natural problems adversely affect, in particular, people who suffer from social and economic disadvantages. Natural catastrophes which are considered to be geophysical, meteorological, hydrological and climatological events—such as landslides, hurricanes, floods, and fires—have affected more than one billion people, in this century alone, and caused US$ 2,512 billion in economic losses worldwide (MunichRe 2017).

With this in mind, hydro-meteorological phenomena deserve special attention, since they reinforce the impact of climate changes brought about by human activities. For example, these have caused increases beyond tolerable proportions in the emission of carbon dioxide, deforestation, the growth of metropolises and the manipulation of river basins, all of which have played a part in there being more flood events in urban areas.

In this context some studies have been developed to reach a more specific understanding of the relationship between the risk of floods—which have occurred more and more frequently over the decades—and changes in the climate. In the urban context, some research studies seek to improve prevention and emergency plans in large cities in order to strengthen urban resilience and reduce vulnerability (Ke et al. 2012, Kourgialas & Karatzas 2016, Piloneet al. 2017, van Wesenbeeck et al. 2016).

Thus, by considering climate change issues that affect the environment of urban cities, this article sets out to analyze how to control and mitigate floods by implementing strategic projects.

Assessing the risk of floods, which are extreme events, and the policies used to mitigate them have to take account of multiple strategic objectives that are often conflicting and integrate three dimensions: social, economic and environmental. That is why some researchers, on seeking approaches to support decision makers (DMs) in their decisions, have applied multi-criteria methods.

This study proposes a multi-criteria framework to model and analyze the risks of flooding in urban areas that take human, social and environmental factors into account. In addition, this framework seeks to make a contribution towards aiding public policy to prioritize preventive actions so as to combat disasters. It does so by setting out a detailed classification of risk.

The present paper is structured as follows: Section 2 presents a literature review on the impacts of climate change on urbanization, in order to justify what prompted this paper. Sections 3 and 4 present research that addresses mitigating the risk of disaster, a review of the literature on risk and multidimensional analysis for decision support. Section 5 describes the proposed framework for flood risk management in urban areas, while Section 6 makes some final remarks and suggestions for future studies on the subject.

2 IMPACTS OF CHANGES IN THE CLIMATE ON URBANIZATION

Although the issue of climate is one of the most important challenges of modern civilization, studies undertaken in the twentieth century pointed out that urbanization was one of the main factors driving global growth and this had the potential to cause natural disasters (Mitchell 1993).

In fact, the ever greater migration of people to urban areas around the world changed the interaction between man, natural resources and the environment as a whole—the Earth and the atmosphere. People's behavior led to their actions having an increasing impact on the environment and thereby contributed to climate change.

Chen & Frauenfeld (2016) investigated what the impact on the climate of urbanization in China might be in the future. They used a model to project what China's climate could be by 2050 and found it would be drier and warmer, thus indicating that human activity will have catastrophic effects.

It needs to be borne in mind that there is considerable uncertainty associated with climate changes, and therefore it is a challenge to establish secure projections that can be used to inform the management of risk. This is because projections about extreme weather events and changes in the climate are forecasts of their frequency, intensity, duration, and month or season of occurrence, which, by definition, cannot be estimated with certainty.

Studies published by IPCC—about global temperature increase—suggest that, on analyzing different scenarios, global temperatures will increase by about 1 degree Celsius between 2000 and 2040 (IPCC 2012).

O'Brien & O'Keefe (2014) assert that while such studies indicate that there is enough time to develop mitigation actions, this may be misleading because there is little "wiggle room" and decisions on mitigation trajectories and deadlines need to be made soon.

In addition, the interference of political factors contributes to the complexity of managing risks. Giddens (2009) affirms through his paradox that governments will not act until something really goes wrong by which time it is too late to take corrective actions. The explanation for this is that politicians keep postponing making difficult decisions if they believe this could be to their electoral disadvantage and therefore would lead to their not staying in power. Examples include the need to increase taxes significantly to pay for major projects such as protecting coastal cities; causing large-scale unemployment as a result of introducing stringent pollution controls or running down extractive industries; and building new nuclear power plants near heavily-populated areas.

Given this set of circumstances, little progress has been made in international climate negotiations and planning preventative and mitigating actions for both long-term and extreme events are needed urgently.

Therefore, effective management strategies for managing risk are needed to counter-act the increasing occurrence of natural disasters in the urban environment.

As an example, NATECH events have, to date, had huge impacts worldwide, and consequently, modern society has started to pay great attention to them and to call for appropriate actions.

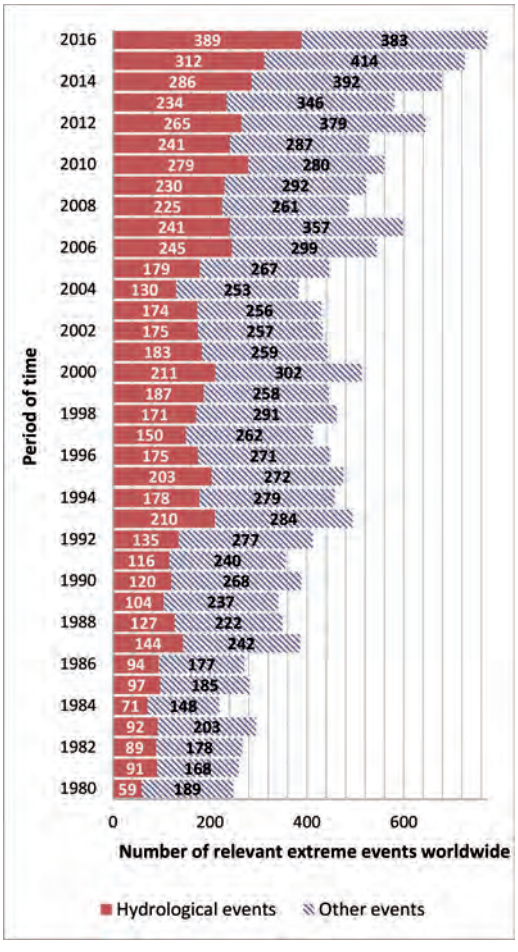


Figure 1. Number of relevant extreme events worldwide from 1980 to 2016. Source: MunichRe (2017).

Thus, Nascimento & Alencar (2016) made a systematic review of the literature on these events because of their great impact and due to their high level of complexity in terms of risk management. Their study revealed the surge in seeking improved scientific knowledge on how to mitigate extreme events, due to their occurring more and more frequently, year after year.

However, the nature and severity of the extreme impacts on climate cannot be assessed only by their intensity, but must also include assessing the exposure and vulnerability of the areas most likely to be affected (IPCC 2012).

Therefore, there is an urgent need to understand how natural disasters, which are increasing in intensity, impact the critical infrastructures of urban centers. These infrastructures have very often been integrated and this is essential for the maintenance of community life so if any break down, the impact on the rest of the system is all the more severe and potentially catastrophic.

Due to this need, the literature now contains a growing number of studies on how to tackle this problem, an example of which is et al. 2017). They propose that urban infrastructure systems be improved by structuring the problem in detail, and do so by involving a wide range of specialists including those from the areas of energy, drainage, transportation, and sanitation. Their model sets out how to conceptualize projects, actions and policies in order to reduce the exposure of areas in urban centers to extreme events.

Figure 1 illustrates the occurrence of relevant events, which have been officially registered since 1980, worldwide. The statistics show that the frequency of hydrological events, of which floods are predominant, has grown dramatically (MunichRe 2017).

The proportion of hydrological events as measured against all other extreme events has grown from 50% in the 80s, to about 80% in the 2010s and, moreover, in 2016 they were nearly equal in number to all other events. In view of this, the authors were prompted to undertake the research that led to producing this article.

3 FLOOD RISK MANAGEMENT AND THE DECISION MAKING PROCESS

Adverse impacts are considered disasters when they produce widespread damage and cause serious changes in the normal functioning of urban societies. Such events may occur because of extremes of climate and the exposure and vulnerability of places to these. Changes in climate can arise from a wide range of factors, including anthropogenic climate change, natural climate variability, and socioeconomic development (IPCC 2012).

That is why research studies seek to investigate how risk management can be used to reduce exposure and vulnerability and increase resilience to the potential adverse impacts of extreme climates, even if risks cannot be completely eliminated. As a result, ways to adapt to and mitigate these events can complement each other and together can dramatically reduce the risks that arise from climate change.

As to the context of floods, however, the qualitative and quantitative evaluations of these factors are subject to various uncertainties (Aven 2012). Therefore, it must be understood that the context of risk and hydrological uncertainty are important concepts that need to be addressed to support decision making in many situations.

The challenge is to know how to describe, measure and communicate risk and uncertainty.

This involves economic factors, including potential financial losses, and environmental and social impacts, such as estimating the number of fatalities and the impact on health services and the sanitation systems of large cities.

Several papers in the literature seek to quantify and analyze risks for which various tools are used, such as decision analysis (Cuellar & McKinney 2017) and Hydro-Meteorological analysis (Patra et al. 2015). However, this paper tackles the multi-objective point of view as a motivator to analyze in an integrated manner the dimensions involved, should an extreme event occur. This is set out in the following section.

4 MULTIDIMENSIONAL RISK EVALUATION

The multi-objective feature of a problem that a risk mitigation policy should reveal can be used in multicriteria analysis.

The application of Multiple Criteria Decision Making/ Aiding (MCDM/A) methods seeks to establish preference relations to evaluate several alternatives against various previously determined criteria during the decision process.

Due to the plurality of points of view that directly influence the decision, de Almeida et al. (2015) pointed out that the methods developed seek to translate the way people have always made their decisions. Despite the diversity of existing methods, the basic elements are simple: a set of undominated actions evaluated by considering at least two criteria that cannot be translated into economical evaluation and a decision maker.

Thus, several studies over the decades have allowed great advances in modelling problems, using multicriteria aggregation methods even more closer to the reality of their applications. Koksalan et al. (2011) and Edwards et al. (2007), for example, analyzed the evolution, history and perspectives of MCDM/A area over time and emphasized this fact.

Such advances leaded researchers to apply MCDM/A methods in many areas: water distribution networks (Fontana & Morais 2017), research and development (R&D) projects (Karasakal & Aker 2017), cleaner energy and electricity market (Cucchiella et al. 2017), civil construction (Miniotaite 2017), financial sector (Ferreira et al. 2018), among other applications.

In the risk management context, there are studies in the literature that integrate multicriteria analysis with environmental risk. Medeiros, et at. (2017) enhance previous suggestions for a multicriteria decision model that evaluates the multidimensional risks of gas transportation by pipeline.

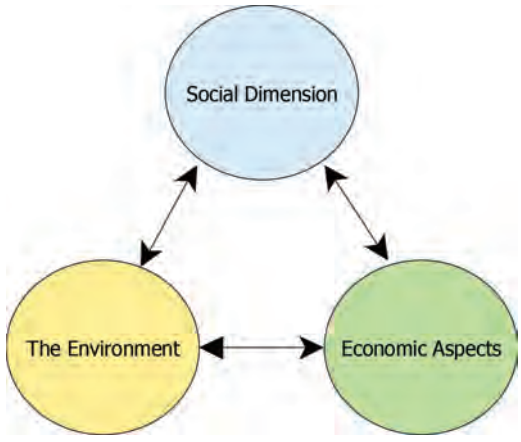


Figure 2. Integration of dimensions for evaluation. Source: This paper (2017).

Their model offers tools that help to prioritize maintenance efforts, and therefore to optimize the use of human, financial and other resources.

In fact, MCDM/A methods have been used in many risk management contexts, and a systematic literature review made by de Almeida et al. (2017) identified research trends in dealing with multicriteria models. These include that they offer a multidimensional view of problems and take into account a DM's preference structures.

By choosing, ranking or sorting actions, these methods can make recommendations on how a DM should analyze results. Thus, multi-dimensional methods admit subjectivity as part of the decision process, and support DMs by using algorithms and methodologies that aid them to build their preference structure. This way, the DMs are able to obtain more robust and reliable results, which might be submitted to sensitivity analysis.

Thus, a multidimensional risk assessment can be carried out in order to incorporate subjective and elements in the social environment that have hitherto been less exploited as presented in Figure 2. This leads to putting forward a framework for flood risk management in urban areas, in Section 5.

5 A FRAMEWORK PROPOSAL FOR MANAGING RISK FROM FLOODS

In this research study, a decision support framework was developed not only so as to understand the risks from floods and to integrate the dimensions of risk, but also so as to use this information to improve how projects are selected and portfolios of projects are managed, where a DM faces uncertainty.

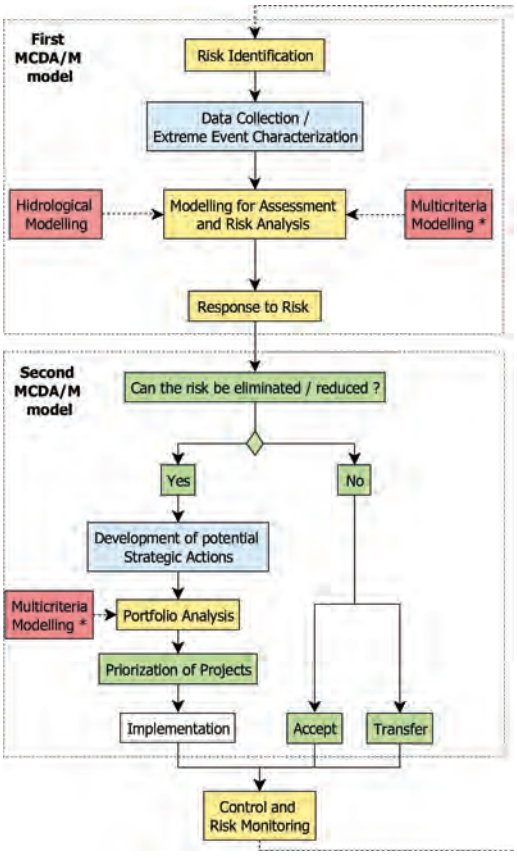


Figure 3. A framework proposal for managing flood risk. Source: This paper (2017).

The framework presented in Figure 3 is divided into 5 steps. It starts by collecting data to provide the model with enough parameters to model the problem.

Broadly speaking, the framework proposes using a multidimensional evaluation for rating risk, which has already been developed, to define possible actions to avoid risk. In this context, when this risk can be eliminated or mitigated, the framework then leads to a way to model multicriteria that will be used to manage projects that combat risk. These projects are held in a portfolio and include alternatives that perform effectively with regard to avoiding the impacts of an extreme event.

The framework proposal conducts the multidimensional flood risk analysis step by step, as described below:

5.1 *Risk identification*

This step corresponds to the initial stage of understanding the problem. A detailed characterization is made, in addition to which data is collected on important parameters which will be used as input for the risk modeling (considering the inherent probabilistic character).

Flood risk identification covers (Rausand 2011):

- Identification of all flood events that are relevant throughout the area to be studied;
- Describe characteristics such as the way, the type, the volume, when and where the flood was present in the system under study;
- Identify possible related triggering events;
- Assess if flood hazards could cause potential hazardous events; and
- Collect hydrological data, such as historical series, return time and characteristics of the hydrographic basin under analysis, and climatological/natural data, such as climate, humidity of the air, wind and soil characteristics.

If social and economic data, such as fatalities, economic losses, devastated areas, are available, it is important to take them into account and this could be done in this step (Steijn et al. 2017).

5.2 *Modelling for assessment and risk analysis*

This stage comprises the process of modeling the problem using the data collected in order to quantify the flood risk in the area of study by considering the influence and interaction of natural, social and economic aspects of floods. Thus, a multicriteria model is used along with the hydrological model to analyze and determine this measure of risk, using methods that are appropriate to the problem of risk rating in the area of study.

Thus, several procedures have been presented in the literature to support DMs in constructing a model. For example, de Almeida et al. (2015) divide the procedure into 3 phases—in successive refinements:

a. A preliminary phase, in which the decision makers, the objectives, the criteria used to model these objectives, the space of actions and the problematic are defined. In addition, uncontrolled factors are identified.
b. The modelling of preferences phase is developed by choosing what MCDM/A method will be used in addition to modelling the DMs' preferences; and
c. The finalization phase, during which the evaluation of alternatives and recommendation are presented, and the decision is implemented.

This way, building MCDM/A models to represent real problems can be faced as a creative process, which intellectual and cultural background of DMs are important for understanding its complexity.

However, an analyst help decision makers in all phases mentioned previously, giving factual information about the problem. Through the interaction of all actors of this process, DMs increase their perceptions about objectives, criteria, space of actions, what indeed enrich the model to the evaluation and implementation of the decision.

In the context of risk identification, in general, this step

- estimates unknown parameters of the model;
- uses an MCDM/A approach to generate an estimate of the risk indices and considers a broad range of different aspects.

It should be noted that the elements of a decision model, such as objectives, criteria and preferences, are constructed in order to promote a realistic assessment of the flood risk. This includes making recommendations that arise from rating the risks and establishing what the response to the risks should be.

5.3 *Response to risk*

As to disaster risk management, several procedures, which were divided into stages, have been developed in the literature, among which O'Brien & O'Keefe (2014) divides it into 4 phases in a cyclical process.

Figure 4 shows that the disaster management function is focused on preparedness for an event and then responding to that.

The authors commented that the recovery stage is often treated as the responsibility of other bodies. Mitigation efforts are dealt with similarly, but typically there will be linked with the disaster management function about the type of mitigation measures; for example, the scale and location of flood defenses.

Figure 4. Disaster management cycle. Source: O'Brien & O'Keefe (2014).

In the context of MCDM/A methods (de Almeida et al. 2015), categorizing risk based on disaster management cycle allows DMs to respond to risk in three ways:

- Accept: there is no action to be implemented, since it takes a lot of time to prepare a strategy to manage risk or high-cost actions will be needed to deal with it;
- Transfer: outsource or share risk to a third party or parties that can manage the outcome;
- Eliminate/mitigate: action is taken to eliminate the causes of threats wherever possible or reduce the likelihood of occurrence of the risk or resulting consequences.

Therefore, if the assessed risk can be eliminated / mitigated, then potential flood control alternatives can be established, while taking the critical infrastructures that are affected by it into account. This creates a discussion among scientific communities, companies, local population and public administration to help DMs to generate a portfolio to be analyzed in a later stage.

5.4 *Portfolio analysis*

As a set of projects or programs and other activities, gathered for effective management, the portfolio constructed from the risk analysis of the delimited area should be managed in order to select the set of projects that maximize satisfaction in social (such as risk by population), economic (loss reduction), and natural (environmental preservation) ways.

However, it is known that they often conflict with each other, which prompts the application of multicriteria decision support methods in this problem.

Since DMs seek to prioritize their strategic goals, which demonstrates what really matters to the organization, a portfolio needs financial and human resources but, generally speaking, the amounts required exceed the limits available (Larson & Gray 2011), and the support of MCDM/A methods is necessary for its resolution.

In addition, the interaction process of all actors is similar to the first MCDM/A model (second step), and which contributes to better results.

Broadly speaking, the portfolio problem consists of choosing, within a set of actions, a subset that meets the objectives and constraints. This results in prioritizing projects and then implementing them.

5.5 *Control and monitoring*

This is the last step of the framework and it allows a monitoring plan to be implemented. It might not only record the risks to be addressed, but also the approval of required projects and the results of the actions developed to mitigate floods. Thus, the step promotes a continuous improvement:

- to design and execute projects;
- to assess the risk and its implications;
- to understand the problem and to model the DM's preferences more coherently;
- to learn and exploit opportunities to benefit affected populations.

This stage is required to:

- Ensure the implementation of risk plans and assess their effectiveness in reducing risk.
- Follow the identified risks, including the watch list.
- Monitor residual risks and identify new risks arising from project implementation.
- Assess the perception of risk by the population;

Thereafter, a cyclical process is implemented for continuous risk assessment, which involves updating the framework and generating new input data.

It needs to be pointed out, however, that this paper seeks to propose a framework, the consolidation of which will be based on a careful analysis, step by step, of the modeling to be done, in order to correctly evaluate the problem, according to the simplifications, specifications and constraints of the model.

6 FINAL REMARKS AND FUTURE STUDIES

Combating and preventing flood risks requires a broad and integrated view of the social, economic and environmental dimensions involved. The development of risk responses in urban areas requires a multi-dimensional analysis of flood risk. Thus, the framework presented seeks to include multivariate elements to control risk in a more balanced and coherent way, thereby making cities less vulnerable and less exposed to extreme events, while their resilience grows due to efficient urban planning.

Some final considerations can be made:

- When implementing this framework in practice, it may contribute to diagnose the resilience of urban areas;
- The multicriteria analysis can give balanced and coherent recommendations, in order to increase the perception of the risk and reduce exposure and vulnerability;
- The cyclical analysis can be understood as a learning process and new information of parameters and modelling can be inserted as procedures. In addition, elements of the decision model—objectives, criteria and evaluation—can be better estimated. Thus, it can promote better portfolio management.
- The modeling of preferences becomes a challenge for DMs. Thus, prior support is required from analysts to understand not only the problem but also the MCDM/A method used. Thus, knowledge of the multidimensional problem makes the process more reliable when implementing the decision.

As future studies, we intended to apply this framework to cities around the world where floods have a considerable impact on critical infrastructures as well as on citizens. Therefore, accurate data need to be collected so that there will be enough parameters to guarantee the benefits mentioned previously, on applying this approach.

Once the decision environment of this problem is dynamic, the framework can be extended for group decision analysis. The aim is to translate strategic objectives of public power into effective actions in the fight against urban floods in order to reduce the damage caused by this natural disaster worldwide.

ACKNOWLEDGEMENTS

This paper is part of a research study funded by the Brazilian National Research Council (CNPq).

REFERENCES

Aven, T. (2012). *Foundations of risk analysis*. John Wiley and Sons, Ltd.

Chatfield, T. (2016). What does it mean to be human in the age of technology? Access: 10 november 2017, from https://www.theguardian.com/technology/2016/jan/20/humans-machines-technology-digital-age

Chen, L., & Frauenfeld, O.W. (2016). Impacts of urbanization on future climate in China. *Climate Dynamics*, *47*(1), 345–357. https://doi.org/10.1007/s00382–015–2840–6

Cucchiella, F., Gastaldi, M., & Trosini, M. (2017). Investments and cleaner energy production: A portfolio analysis in the Italian electricity market. *Journal of Cleaner Production*, *142*, 121–132. https://doi.org/10.1016/j.jclepro.2016.07.190

Cuellar, A.D., & McKinney, D.C. (2017). Decision-making methodology for risk management applied to Imja Lake in Nepal. *Water (Switzerland)*, *9*(8), 14–16. https://doi.org/10.3390/w9080591

de Almeida, A.T., Alencar, M.H., Garcez, T.V., & Ferreira, R.J.P. (2016). A systematic literature review

of multicriteria and multi-objective models applied in risk management. *IMA Journal of Management Mathematics*, *28*, 153–184. https://doi.org/10.1093/imaman/dpw021

de Almeida, A.T., Cavalcante, C.A.V., Alencar, M.H., Ferreira, R.J.P., de Almeida-Filho, A.T., & Garcez, T.V. (2015). *Multicriteria and Multiobjective Models for Risk, Reliability and Maintenance Decision Analysis*. *International Series in Operations Research and Management Science* (Vol. 231). https://doi.org/10.1007/978–3-319–17969–8

Edwards, W., Jr., R.F.M., & Winterfeldt, D. von (Orgs.). (2007). *Advances in Decision Analysis: From Foundations to Applications* (1st ed). Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511611308

Ferreira, L., Borenstein, D., Righi, M.B., & de Almeida Filho, A.T. (2018). A fuzzy hybrid integrated framework for portfolio optimization in private banking. *Expert Systems with Applications*, *92*, 350–362. https://doi.org/10.1016/j.eswa.2017.09.055

Fontana, M.E., & Morais, D.C. (2017). Water distribution network segmentation based on group multi-criteria decision approach. *Producao*, *27*, 1–13. https://doi.org/10.1590/0103–6513.208316

Giddens, A. (2009). Politics of Climate Change.

IPCC. (2012). Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. https://doi.org/10.1017/CBO9781139177245

Karasakal, E., & Aker, P. (2017). A multicriteria sorting approach based on data envelopment analysis for R&D project selection problem. *Omega (United Kingdom)*, *73*, 79–92. https://doi.org/10.1016/j.omega.2016.12.006

Ke, Q., van Gelder, P.H.A.J.M., Jonkman, S.N., & Rijcken, T. (2012). An explorative analysis of the potential flood risk in downtown Shanghai city along the Huangpu River.

Koksalan, M., Wallenius, J., & Zionts, S. (2011). Multiple criteria decision making: From early history to the 21st century.

Kourgialas, N.N., & Karatzas, G.P. (2016). A flood risk decision making approach for Mediterranean tree crops using GIS; climate change effects and flood-tolerant species. *Environmental Science & Policy*, *63*, 132–142. https://doi.org/10.1016/j.envsci.2016.05.020

Larson, E.W., & Gray, C.F. (2011). Project managment: the managerial process.

Medeiros, C.P.P., Alencar, M.H.H., & de Almeida, A.T.A.T. (2017). Multidimensional risk evaluation of natural gas pipelines based on a multicriteria decision model using visualization tools and statistical tests for global sensitivity analysis. *Reliability Engineering & System Safety*, *165*(March), 268–276. https://doi.org/10.1016/j.ress.2017.04.002

Merchant, B. (2014). "Natural Disasters Caused by Human Activity Have Increased." Are Natural Disasters Increasing? *Greenhaven Press*, (Opposing Viewpoints in Context). Access 10 november 2017, from link.galegroup.com/apps/doc/EJ3010598222/OVIC?

Miniotaite, R. (2017). Multicriteria Analysis of Assembling Buildings from Steel Frame Structures. *IOP Conference Series: Materials Science and Engineering*, *245*(2). https://doi.org/10.1088/1757–899X/245/2/022077

Mitchell, J.K. (1993). Natural Hazard Predictions and Responses in Very Large Cities. In J. Nemec, J.M. Nigg, & F. Siccardi (Orgs.), *Prediction and Perception of Natural Hazards: Proceedings Symposium, 22--26 October 1990, Perugia, Italy* (p. 29–37). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978–94–015–8190–5_4

MunichRe. (2017). NatCatService. Access: 10 march 2017, from http://natcatservice.munichre.com/

Nascimento, K.R.D.S., & Alencar, M.H. (2016). Management of risks in natural disasters: A systematic review of the literature on NATECH events. *Journal of Loss Prevention in the Process Industries*, *44*, 347–359. https://doi.org/10.1016/j.jlp.2016.10.003

O'Brien, G., & O'Keefe, P. (2014). *Managing Adaptation*. (Routledge, Org.).

Patra, J.P., Kumar, R., Mani, P., Cuellar, A.D., McKinney, D.C., Yu, D., … Shimatani, Y. (2015). Evaluating the importance of catchment hydrological parameters for urban surface water flood modelling using a simple hydro-inundation model. *Procedia Engineering*, *118*(8), 1096–1103. https://doi.org/10.3390/w9080591

Pilone, E., Mussini, P., Demichela, M., & Camuncoli, G. (2017). Municipal Emergency Plans in Italy: Requirements and drawbacks. *Safety Science*, *85*, 163–170. https://doi.org/10.1016/j.ssci.2015.12.029

Priori, L., Alencar, M.H., & de Almeida, A.T. (2017). Adaptations to Possible Climate Change Impacts: Problem Structuring Based on VFT Methodology. In W. Leal Filho (Org.), *Innovation in Climate Change Adaptation* (p. 145–157). Cham: Springer International Publishing. https://doi.org/10.1007/978–3-319–25814–0_11

Rausand, M. (2011). *Risk Assessment: theory, methods, and applications*. Hoboken, N.J. : J. Wiley & Sons, [2011] ©2011.

Steijn, W.M.P., Groeneweg, J., Van Der Beek, F.A., Van Kampen, J., & Van Gelder, P.H.A.J.M. (2017). An integration of human factors into quantitative risk analysis: A proof of principle, (2011), 321–328.

van Wesenbeeck, B.K., de Boer, W., Narayan, S., van der Star, W.R.L., & de Vries, M.B. (2016). Coastal and riverine ecosystems as adaptive flood defenses under a changing climate. *Mitigation and Adaptation Strategies for Global Change*, 1–8. https://doi.org/10.1007/s11027–016–9714-z

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Impacts of climate change on rail systems: A new climate risk analysis model

T. Wang, Z. Qu & T. Nichol
Liverpool Business School, Liverpool John Moores University, UK

Z. Yang
Liverpool Logistics, Offshore and Marine Research Institute, Liverpool John Moores University, UK

D. Dimitriu
Centre for Aviation, Transport, and the Environment, Manchester Metropolitan University, UK

G. Clarke & D. Bowden
Logistics Department, AECOM (UK) Ltd, UK

ABSTRACT: Risk analysis has been widely used in climate adaptation practice. However, traditional probabilistic risk analysis methods are not capable of tackling the unavailability or incompleteness of climate risk data. To deal with such challenges, this paper further applies an advanced Fuzzy Bayesian Reasoning (FBR) model for climate risk analysis of railways system in the UK. Its novelty lies in the realisation of climate risk ranking under high uncertainty in data and its practical contribution on the risk perception of stakeholders in the UK railway systems. To test the feasibility of the developed model in the transport industry, a large scale of surveys are conducted to collect data, regarding the timeframe of climate hazards, likelihood of occurrence, severity of consequences, and infrastructure resilience for the analysis of climate risks threatening British rail systems. The findings will provide transport planners with useful insights on the identification of climate hazards of high risks to facilitate the development of cost-effective climate adaptation strategies.

1 INTRODUCTION

Current variability in climate poses a challenge for rail infrastructure and operation. In the majority of countries, the related activities of transport systems are sensitive to different weather extremes, which include but are not limited to, changes in temperature, precipitation, thunderstorms, winds, visibility and sea level (e.g., Love et al. 2010).

Risk analysis as a critical element in climate change adaptation has been widely utilised through a variety of approaches and techniques. The selection of cost-effective climate adaptation measures requires systematically analysing risk reduction combining with the associated costs due to the implementation of these measures. A variety of methods in risk quantification have been proposed (Wilby et al. 2009). However, their availability and effectiveness of common-used methods are challenged in existing climate risk studies (Yang et al. 2015). One of the main challenges is that the unavailable or incomplete objective data fails to precisely evaluate the risk reduction and costs on many occasions, resulting in difficulties to assess risk and costs. Owing to the high level of uncertainties in this data, many conventional risk assessment approaches (e.g., Quantitative Risk Assessment (QRA)) which have been widely applied to conduct risk analysis in many sectors become unsuitable (UNCTAD 2012).

Fuzzy set methods have been applied to climate risk assessment on ports in several pioneering studies to deal with this challenge (i.e., Ng et al. 2013, Yang et al. 2015, 2016, 2017). In assessing climate change risks on ports, linguistic terms were regarded as variables, namely, the stakeholders' perceptions of the impacts of climate change (Yang et al. 2015). Through modelling subjective input data (linguistic terms) on climate risk evaluation based on the stakeholders' perceptions collected from their current interpretations of the frequencies, severity of consequences and timeframes of climate risks where they occur. By these studies, this paper develops a new Fuzzy Bayesian model by dividing the risk parameters into two levels. The junior-level parameters are "*Timeframe (T)", "likelihood*

(L)", "Severity of Consequences (C)" and add the fourth one "*Climate resilience (S)*". In particular, three sub-parameters of *"Severity of Consequences (C)"*, namely, "*Damage to Infrastructure (INF)*", "*Injures and/or Loss of Lives (INJ)*" and "*Damage to Environment (ENV)*" are added to extend and specialise the risks parameters. As a result, this Fuzzy Bayesian Reasoning (FBR) approach in which subjective evaluations by domain experts enables to complement the unavailable objective data to realise climate risk ranking more precisely under high uncertainty in data.

This paper aims to achieve two primary objectives. Firstly, it develops a new FBR model by subdividing and adding new parameters to quantify the risk perception of the stakeholders towards climate threats s. This new development will offer a notable foundation for further adapting to climate risks and tackling the uncertainties in climate risks. Most importantly, this FBR model is exemplified by the British rail system through conducting a nation-wide survey amongst 21 major rail stakeholders in the UK. This applicability will contribute to reveal the real climate risks on UK rail from the perspective of both industry and academia. We believe that the outcomes of this study will be of considerable value to rail planners, decision makers and industrial practitioners, helping them to create and implement adaptation plans, strategies and practices.

The rest of the paper is structured as follows. The critical review of the risk analysis for climate change is presented in section 2. The FBR methodology including a step-by-step risk analysis framework is given in section 3. In section 4, climate risks and adaptation on UK rail systems, as well as the data collection through a survey amongst the 21 rail stakeholders across the UK, are described. Finally, the discussion and conclusion implying the contribution and revelation for further research can be found in section 5.

2 CRITICAL REVIEW: RISK ANALYSIS FOR CLIMATE CHANGE

Current research on climate-related risk analysis have commons on interpreting and identifying the existing and future risks, estimating the level of risk as well as determining the level of uncertainties (Yang et al. 2015). A variety of methods in risk quantification have been proposed which can be categorised into three methods: methods requiring limited resources (i.e., sensitivity analysis), methods with modest resource needs (i.e., empirical downscaling) and methods with high resources needs (i.e., dynamical downscaling) (Wilby et al. 2009). However, traditional Probabilistic Risk Analysis (PRA) methods sometimes are unable to tackle the unavailability or incompleteness of climate risk data (i.e., Yang et al. 2015). Fuzzy set and Bayesian Networks (BNs) have been applied to climate risk assessment on ports in several pioneering studies to deal with this challenge. In this section, we review the development of fuzzy theory and BNs, as well as their applications in risk analysis.

2.1 *Fuzzy set and fuzzy logic*

Mathematically, a fuzzy set can be defined by assigning to each possible individual in the universe of discourse a value on behalf of its grade of membership in the fuzzy set (Klir & Yuan, 1995). One of main approaches to reasoning under uncertainty is possibility theory, which is called fuzzy logic (Jenso, 2001). It is the logic of classes with unsharp boundaries and has been widely utilised to cope with the problem of computer understanding of natural language (i.e., Zadeh, 1992; 2010).

In recent decades, fuzzy set and fuzzy logic have been widely accepted tools in risk assessment, providing a theoretical framework of expert decision making under uncertainty (i.e., Sii et al., 2002; Andrews & Moss, 2002). For instance, Singh & Benyoucef (2011) utilised F-TOPSIS to the solution of MCDM problems in selection of supply chain coordination. Yang et al. (2011) employed F-TOPSIS for vessel selection under uncertain environment and Yazdani-Chamzini (2014) applied F-AHP and F-TOPSIS for selection and evaluation of available handling equipment.

Fuzzy set methods had made successful attempts in climate risk assessment for climate change in recent years (Yang et al. 2015). For example, in assessing the climate risks on ports, linguistic terms were regarded as variables, namely, the stakeholders' perceptions of climate impacts (Ng et al. 2013, Yang et al. 2015, 2016, 2017). On the basis of the stakeholders' perceptions collected from their interpretations of climate risks' frequencies, consequences and timeframes in which they occur, these subjective input data were modelled to estimate climate risks. The fuzzy risk score R was defined by using a discrete fuzzy set, $R = C \circ (L \times T)$ (Ng et al. 2013) or a fuzzy set manipulation approach, $R = T \otimes C \otimes L$ (Yang et al. 2015). Nevertheless, these methods could neither avoid the loss of useful information in fuzzy operations nor deal with a huge amount of risk input data (Yang et al. 2016). A novel fuzzy-Bayesian model proposed by Yang et al (2017) overcame these issues by taking advantages of fuzzy rule bases (i.e., IF-THEN rules) in modelling non-linear relation between risk parameters and output and BN in realising fuzzy rule integration and risk inference. Employing fuzzy IF-THEN rules in Fuzzy logic theory allows the

antecedent and conclusion parts containing linguistic variables model the qualitative features of experts' knowledge and reasoning process when there is lack of precise quantitative analysis.

2.2 *Bayesian networks*

When multiple sets of data (from different experts) are employed, it is difficult to use normal fuzzy rule inference mechanisms as the calculation could take a long time. Bayesian Networks (BNs) is a sound mathematical method in minimising the uncertainties and increasing knowledge. These are achieved by combining probability distributions or functions of different parameters and updating their probabilities when new information emerges (Wang, 2003).

Bayesian modelling is a proven inter-disciplinary tool (Tebaldi et al. 2005). There are a variety of benefits with BNs including integrating different types of variables and data within a framework, and efficiently being updated when new information and knowledge become available (Castelletti & Soncini-Sessa, 2007, Cinar & Kayakutlu, 2010). In particular, BNs are capable of compensating the absence of historical statistics and handling incomplete uncertainty through combining various pieces of information and making use of expert judgments (i.e., Tighe et al., 2007). BNs has achieved a wide range of application in multiple fields, such as safety assessment, intelligent decision making and computer network diagnosis, which involve uncertainty due to incomplete data and limited cognitive capacity (Zhang et al., 2013). It is employed for the estimation of future climate change conditions, such as precipitation mean state and seasonal cycle in South Africa (Boulanger et al., 2007) and quantitative prediction and assessment of long-term shoreline change associated with sea level rise and its uncertainty (Gutierrez et al., 2011). Bertone et al. (2015) integrated BNs into participatory modelling to develop a risk assessment tool for managing water-related health risks associated with extreme events by combining System Dynamics.

Nevertheless, a major disadvantage of BNs in incorporating expert judgements is that it may fail to probabilistically forecast subjective fuzziness in a precise way as it lacks understanding of probability theory. Bayesian approaches were criticised as they often requires too much information regarding prior probabilities which are hard to receive in risk analysis. Also, this information mainly relies on subjective judgments such as provided by experts' knowledge, which might result in unexpected bias into BNs (i.e., Tversky & Kahneman, 1975; 1990). BNs might become computationally inefficient when models have a large number of variables (Liang & Lee, 2008). It cannot process time-series data and address feedback regulations (Ristevski, 2013).

To compensate these disadvantages of BNs, some theoretical research and applications have identified the benefits to combine fuzzy logic and Bayesian reasoning especially in the applications of Fuzzy-Bayesian approaches to safety and reliability (Bott & Eisenhawer, 2002). However, more advanced and capable methods are required to enhance structure learning algorithms, allowing for constraints based on expert knowledge, with precise rules for interventional risk management and decision making (i.e., Zhou et al. 2014, Constantinou et al. 2016).

Overall, the evaluation of climate change risks on the rail systems in this paper may contain various types of uncertainty. Similarly, due to the scarcity of historical/statistical data (UNCTAD, 2012), is carried out by subjective judgments. Thus, combining fuzzy set theory and BNs is appropriate to model subjective linguistic variables and cope with the discrete problem, handling incomplete uncertainty and complicated calculation.

3 A NEW FBR RISK ANALYSIS FRAMEWORK

In Section 3, it indicates a step-by-step risk analysis framework by utilising FBR approach. The proposed framework for modelling climate change risks consists of four main steps, which outlines each step required for risk estimation and assessment.

3.1 *Identify environmental drivers*

Based on previous literature review, we investigated four primary environmental drivers due to climate change affecting on the British railways: 1) temperature increase, 2) intensive rainfall/flooding, 3) increased intensity and/or frequency of high wind and/or storms, and 4) sea level rise. Hence, this risk analysis is made for each of these environmental drivers to evaluate the risk level of their corresponding potential climate threats.

3.2 *Identify fuzzy input and output variables*

To define the subjective risk estimates, five threat-based risk permeants are identified which include at both the senior and junior levels respectively. The senior-level parameter is "Risk Level (RL)". Refer to the definition of RL used in the FBR climate risk analysis on port systems (i.e., Ng et al. 2013, Yang et al. 2015, 2016), it expresses such linguistic variables as "Very High", "High", "Medium", "Low" and "Very Low".

Three junior parameters closely associated with climate change risks have been identified in previous studies, namely "Timeframe (T)", "likelihood (L)" and "Severity of Consequences (C)" based on the FMEA approach (Yang et al. 2008, 2009, Ng et al. 2013, Yang et al. 2016). Especially, in this paper, "Severity of Consequences (C)" is divided into three subcategories, namely, "Damage to Infrastructure (INF)", "Injures and/or Loss of Lives (INJ)" and "Damage to Environment (ENV)". At the meantime, "Climate resilience (S)", designed as the fourth important junior parameter in assessing climate risks in railways, is characterised in Table 1.

Table 1. Climate resilience.

Grade	Linguistic terms	Description	Fuzzy memberships
1	Very Weak (VW)	Very weak (0–20%) capacity of the transportation system to anticipate, absorb, accommodate, or recover from the effects of a climate event and requiring a very long period (a year) and very high cost of recovery (£10 million above)	(0, 0, 0.1, 0.3)
2	Weak (W)	Weak (20–39%) capacity of the transportation system to anticipate, absorb, accommodate, or recover from the effects of a climate event and requiring a long period (a month) and high cost of recovery (£1 million above)	(0.1, 0.3, 0.5)
3	Average (A)	Average (40–59%) capacity of the transportation system to anticipate, absorb, accommodate, or recover from the effects of a climate event and requiring certain length of time (a week) and cost of recovery (£100, 000–£1 million)	(0.3, 0.5, 0.7)
4	Strong (S)	Strong (60–80%) capacity of the transportation system to anticipate, absorb, accommodate, or recover from the effects of a climate event in a relatively timely and efficient manner (a day) and requiring some cost of recovery (£10,000–£100,000)	(0.5, 0.7, 0.9)
5	Very Strong (VS)	Very strong (80% above) capacity of the transportation system to anticipate, absorb, accommodate, or recover from the effects of a climate event in a very timely and efficient manner (12 hrs) and requiring slight cost of recovery (0–£1,000)	(0.7, 0.9, 1, 1)

3.3 *Establish a fuzzy rule-base network with the belief structure*

Fuzzy system theory, through collecting fuzzy IF-THEN rules from experts or domain knowledge and combining the rules into a single system, offers a systematic procedure for transforming knowledge bases to non-linear mappings (Sii & Wang 2002, Yang 2006).

While in some real cases, subjective degrees of belief (DoBs) are assigned to the linguistic variables used to express the conclusion attribute (RL) for modelling the incompleteness of expert judgments. Refer to the climate risk models on ports by Yang et al. (2016, 2017), in the fuzzy rule-base network of this study, four junior-level fuzzy input parameters including 20 (5+5+5+5) linguistic variables are assembled to generate 625 (5*5*5*5) antecedents with a rational degree of belief (DoB) distribution. Simultaneously, we constructed a secondary level network between the three subcategory parameters (INF, INJ and ENV) and junior-level parameter C, containing 15 (5+5+5) linguistic variables assembling to create 125 (5*5*5) antecedents.

3.4 *Conduct risk inference by BN techniques*

The constructed belief structures can be used to conduct risk inference using BN techniques. First of all, the rule base with belief structures is expressed in the form of conditional probabilities. Through utilising a BN technique, the FBR constructed can be modelled and converted into a five-node converging connection. It includes four parent nodes, NT, NL, NC and Ns (Nodes *T, L, C* and *S*); and one child node NRL (Node *RL*).

The prior probabilities of NT, NL, NC and Ns can be achieved through questionnaire surveys (i.e., on rails). The prior profanities of NT, $p(Li)$, for example, can be obtained through asking the question, "using the defined linguistic variables (i.e. VH, H, A, L and VL), how likely the effect will occur when you expect first to see this climate threat poses impacts to the rail your organisation associated with?", to collect the data from domain experts' risk perception. Through analysing all the prior probabilities of the four nodes, the marginal probability of NRL can be computed as (Jensen, 2001):

$$p(RLh) = \sum_{i=1}^{5}\sum_{j=1}^{5}\sum_{k=1}^{5}\sum_{l=1}^{5} p(RLh \mid Ti, Lj, Ck, Sl) \; p(Ti)p(Lj)p(Ck)p(Sl)$$
$$(h = 1, \ldots, 5) \quad (1)$$

To prioritise the climate risks, RLh (h = 1, …, 5) requires the assignment of appropriate utility values URLh. The linguistic description can then be converted into a crisp value using a centroid defuzzifcation method (Yang et al. 2009). Accordingly, RLh (h = 1, …, 5) = {0.11, 0.3, 0.5, 0.7, 0.89}. Hence, a new risk criticality ranking index can be developed as follows:

$$RI = \sum_{h=1}^{5} p(RLh)U_{RLh} \quad (2)$$

where the smaller the value of RI is, the higher the risk level of potential climate threats.

4 CASE STUDY: RISKS ANALYSIS OF CLIMATE CHANGE ON UK RAILWAYS

4.1 *Climate risks and adaptation on UK rail systems*

In the UK, the transport sector has been recognised as one of six key sectors which will be most vulnerable to the impact of climate change (McKenzie Hedger et al. 2000). The predicted climate impacts on rail transport in the UK include an increased number of hot days, a decreased number of cold days, increased heavy precipitation, drought, sea level change, seasonal change, extreme events and wind (i.e., Jaroszweski et al. 2010, Peterson et al. 2008, Hooper & Chapman. 2012).

The extreme events posed the most devastating impacts (e.g., heat waves and storms) on rail transport. Higher temperatures in summer will cause rail buckling as well as decreased thermal comfort; more intense precipitation in winter will result in flooding, landslips and bridge scour. However, this might be beneficial for lower winter maintenance. Flooding was regarded as one of the significant impacts on the rail network (EPA 2009). The damage caused by climate change on railway networks took into account approximately 29% to 71% of the total (Chatterton et al. 2010). Dora (2012) investigated the leading climate change impacts to infrastructure operations on UK rail transport systems as a result of the projected changes in temperature and precipitation. This report stressed the effects, including the increases in track buckling, days of track maintenance and exposure of staff to hear the stress and overhead power cables sagging in hot weather. Some issues included air quality in urban areas and remarkable differences between the North and South in the UK owing to the growth temperature, the increased possibility of track inundation and of scouring affecting river bridges' stability and incidence of landslips due to heavy and extreme precipitation. Overall, although there have been widespread effects on diverse transport modes, it is only recently that more attention has been given to the impacts of climate change on British railways (Hooper & Chapman 2012).

The lack of data on the impacts of climate change, as well as cost-benefit analysis for climate change pose a significant challenge for transportation planners, which also results in the failure of adaptation strategies in the transport sector (i.e., Koetse & Rietveld 2012). Owing to high uncertainties related to the future climate, adaptation measures should be robust to retain the option value of the portfolio of measures. Hence, through conducting a nation-wide survey of UK rail systems, this paper examines the new risk analysis model in Section 3 which overcomes the shortage of data and the uncertainty of climate risks to reveal the real climate risks for British rail.

4.2 *A nation-wide online survey on climate risks assessment*

To validate the efficiency of the FBR model, a nation-wide survey was conducted to collect the first-hand data through examining the perceptions of rail planners and stakeholders on the impacts of climate change within the rail systems.

This survey aims to illustrate the general situation of climate risks in UK rail systems and to justify the necessity of adaptation planning. Four main environmental drivers due to climate change were identified: temperature increase, intensive rainfall/flooding, increased intensity and/or frequency of high wind and/or storms and sea level rise. The specific potential climate threats and corresponding adaptation measures were summarised according to the Network Rail's adaptation framework (Network Rail 2015). To guarantee the validity and feasibility of questionnaire designing, a pilot study was conducted in April 2017 by inviting ten professional rail experts and academics in the UK. From May to September 2017, a nation-wide online survey was conducted amongst the 21 British rail stakeholders to assess their perception of climate change risks, including specific impacts on their operation, performance and infrastructure.

The population of this survey is all the rail stake-holders in the UK who are from rail companies and authorities, governmental departments, academics and NGOs etc. The databases of the national rail networks were chosen from the national maps to be used to select the transportation entities (Network Rail 2016). The participants in the railway survey were mainly chosen from

members of the Railway Industry Association (RIA) and the Rail Freight Group (RFG) representing major UK-based suppliers of the world's railways and the leading body for rail freight in the UK (RIA n.d., RFG n.d.). Over 200 member companies crossing the whole range of railway supply with diverse skills and resources are the typical rail entities of UK national railway.

Considering the uniqueness and complexity of climate change issues (i.e., the characteristics, geographic distribution, types and levels of risks posed by climate change on rails), non-probability sampling, including a combination method of judgment sampling and snowball sampling, was utilised in this survey. Some small entities in remote regions might lack necessary knowledge or experience of climate change issues and the representativeness of the samples is more critical than its generalisability in judgment sampling (Vogt et al. 2012). Consequently, a sample of 30 administrators representing the most critical transport institutions in different regions of the UK (e.g., Network Rail, Transport for Greater Manchester, AECOM UK, etc.) was selected to assess their perceptions of climate change risks. Snowballing was utilised via one or two key informants at each entity from the targeted population.

The 30 questionnaires were distributed online through BOS Online Survey (BOS 2017). E-mails and phone calls were used to contact all the respondents during business days. In the end, 21out of 30 effective responses were received with a high response rate 70%. The questions were divided into two types: closed-end questions, which utilise multiple choices and a linguistic evaluation approach to quantify responses; open-ended questions, which provide more freedom to respondents and produced more precise data.

Data screening was conducted to eliminate missing and ineffective data such as incomplete input information and insane responses. Accordingly, 4 out of 21 feedbacks became invalid after the screening process. The consistency of the remaining 17 sets of data was addressed through the comparative climate risk analysis. Finally, associated data from the first 11 questions from the questionnaire were inputted in this FBR model to rank and analyse the top potential risks posed by climate change.

Table 2. Questionnaire results of climate risk analysis on UK railways.

Environmental driver due to climate change	Potential climate threat on the railway	Result of risk level	Utility value	Ranking
Temperature increase	A1. Track buckling causing derailment risks & reducing opportunities for track maintenance	{0.1154, 0.1808, 0.3003, 0.1922, 0.2116}	0.54	6
	A2. Unreliable signalling, power line side systems, failure of temperature controls and overheating of electronic equipment	{0.8580, 0.2016, 0.3228, 0.2116, 0.1783}	0.54	6
Intensive rainfall/flooding	B1. Bridge foundations damaged leading to bridge collapse and derailment risk	{0.1083, 0.3299, 0.2613, 0.2109, 0.0896}	0.47	1
	B2. Landslips caused obstruction in increasing derailment risk	{0.1590, 0.2313, 0.2406, 0.2700, 0.0991}	0.48	2
	B3. Heavy rain affect visibility, scheduled work may have to be rescheduled for safety and welfare reasons	{0.0621, 0.1617, 0.2115, 0.3192, 0.2455}	0.6	7
	B4. Track drainage overloaded leading to flooding of the track	{0.0927, 0.2874, 0.2502, 0.2716, 0.0981}	0.5	3
Increased intensity and/or frequency of high wind and/or storms	C1. Trees falling onto the line	{0.1138, 0.2045, 0.2863, 0.2944, 0.1010}	0.51	4
	C2. High winds affect visibility and scheduled work may have to be rescheduled for safety and welfare reasons	{0.0997, 0.2689, 0.2229, 0.1833, 0.2252}	0.53	5
	C3. Instability of structures	{0.0500, 0.1482, 0.4197, 0.1832, 0.1899}	0.56	8
Sea level rise	D1. Breach of sea wall, flooding and derailment risk	{0.0830, 0.2974, 0.3390, 0.2142, 0.0663}	0.48	2
	D2. Reduced maintenance opportunities, bridges/sea walls may not be safely inspected	{0.0156, 0.2488, 0.3305, 0.2458, 0.1594}	0.56	8

Based on the fuzzy Bayesian approach in Section 3, the climate risk results of each potential climate threat of environmental driver related to UK rails were calculated and elaborated in Table 2. For instance, the impacts of temperature increase were divided into two potential threats, namely, "A1. Rack buckling is causing derailment risks & reducing opportunities for track maintenance" and "A2. Unreliable signalling, power line side systems, failure of temperature controls and overheating of electronic equipment". Then, the evaluations of the two threats were based on the four aforementioned risk parameters: Timeframe (T), Likelihood (L), Severity of occurrence (C) and Climate Resilience (S).

Utilising Equation (6) and Hugin software (HUGIN v.8.5 2017), the risk results of "A1. Rack buckling causing derailment risks & reducing opportunities for track maintenance" and "A2. Unreliable signalling, power line side systems, failure of temperature controls and overheating of electronic equipment" can be calculated as {11.54% RL1, 18.08% RL2, 30.03% RL3, 19.22% RL4, 21.16% RL5} and {8.58% RL1, 20.16% RL2, 32.28% RL3, 21.16% RL4, 17.83% RL5}, respectively. According to Equation (7) and Hugin, their risk index values are calculated both as 0.54.

Based on the ranking in Table 2, the highest potential climate threats to British rail are "B1. Bridge foundations damaged leading to bridge collapse and derailment risk", "B2.Landslips caused obstruction in increasing derailment risk" and "B4.Track drain-age overloaded leading to flooding of the track" due to the intensive rainfall/flooding, as well as"D1.Breach of seawall, flooding and derailment risk" due to sea level rise. Interestingly, all the top potential climate threats are attributed to the intense rainfall/flooding. However, the lowest threats are "C3.Instability of structures" owing to the increased intensity and/or frequency of high wind and/or storms, and "D2.Reduced maintenance opportunities, bridges/sea walls may not be safely inspected" posed by sea level rise.

5 DISCUSSION & CONCLUSION

This paper presents an innovative mathematical FBR model to quantify the risks posed by climate change and applied to the British rail system through conducting a nation-wide survey. Fuzzy Bayesian Reasoning (FBR) approach, in which subjective evaluations by domain experts, enables to complement the unavailable objective data to realise climate risk ranking under high uncertainty in data. Based on the previous modelling research of climate adaptation in ports (e.g., Ng et al. 2013, Yang et al. 2015, 2016, 2017), this Fuzzy-Bayesian Network is innovated by employing the climate resilience (S) as the fourth most important junior parameter in assessing climate risks taking, as well as dividing the Severity of consequences (C) into three subcategories ("Damage to Infrastructure (INF)", "Injures and/or Loss of Lives (INJ)" and "Damage to Environment (ENV)"). We believe that this new risk analysis model will offer a more precise and effective tool for researchers and rail stakeholders to tackle the uncertainties in responding to climate risks.

To test the feasibility of the model and provide the empirical evidence on its applicability in the transport industry, a large scale of surveys were conducted to collect data for the analysis of climate risks threatening the rail systems in the UK. Through previous literature reviews, we identified four primary environmental drivers due to climate change. The risk level for each potential climate threat was evaluated by the timeframe of climate hazards, likelihood of occurrence, severity of consequences, and infrastructure resilience. Unsurprisingly, the top potential climate threats to British rail are highly related to the occurrence of intensive rainfall/flooding, which is also consistent with the current priorities of adapting to flooding issues in climate change adaptation planning in the UK. It is noticeable that the research presented is part of an ongoing project and further activities and data collection (i.e., a second-round survey and in-depth interviews) are expected. Nevertheless, the outcomes of this study will provide policy makers and transport planners with useful insights for the identification of climate hazards of high risks to facilitate the development of cost-effective climate adaptation strategies before the final implementation of the relevant infrastructure.

One of the current dilemmas in climate risk assessment is that traditional probabilistic risk analysis (PRA) methods usually pay insufficient attention to a particular type of climate change event or transportation assets. This study focuses on primary impacts of climate change on British rail systems, to provide region-specific customisation and ongoing trend observation. Further research on other transport modes (e.g., road and air) and multiple regions (i.e., developing countries) is vital to establish a practical and robust adaptation framework on climate change. Hence, the present preliminary analysis that this study makes will be incorporated into a comparative study of climate risks and adaptation in the rail and road systems on the next stage. Also, it is suggested that there are more quantitative estimates, cost-effectiveness evaluations and more complex decision models to increase the robustness of the risk model (Adger et al. 2007, Walker et al. 2011, Wu et al. 2013). Accordingly, the next step of this

research is to investigate the relationship between climate change risk reduction and the costs of adaptation measures. A new economic model will be constructed complementary with the current FBR model and evidential reasoning techniques to enhance data consistency and credibility to the overall modelling.

REFERENCES

Adger, W.N., Agrawala, S., Mirza, M.M.Q., Conde, C., o'Brien, K., Pulhin, J., Pulwarty, R., Smit, B. & Takahashi, K. 2007. Assessment of adaptation practices, options, constraints and capacity. *Climate change*: 717–743.

Bertone, E., Sahin, O., Richards, R. & Roiko, R.A. 2015. Bayesian Network and system thinking modelling to manage water-related health risks from extreme events. In *Industrial Engineering and Engineering Management (IEEM), 2015 IEEE International Conference*: 1272–1276.

BOS, 2017. BOS Online Survey Tool. Retried on 7 May 2017, from https://www.onlinesurveys.ac.uk/.

Bott, T.F. & Eisenhawer, S.W. 2002. Risk analysis using a hybrid Bayesian-approximate reasoning methodology. In *Reliability and Maintainability Symposium, 2002. Proceedings. Annual*: 127–133. IEEE.

Boulanger, J.P., Martinez, F. & Segura, E.C. 2007. Projection of future climate change conditions using IPCC simulations, neural networks and Bayesian statistics. Part 2: precipitation mean state and seasonal cycle in South America. *Climate Dynamics*, *28*(2–3): 255–271.

Castelletti, A. & Soncini-Sessa, R. 2007. Bayesian Networks and participatory modelling in water resource management. *Environmental Modelling & Software*, *22*(8): 1075–1088.

Cinar, D. & Kayakutlu, G. 2010. Scenario analysis using Bayesian networks: A case study in energy sector. *Knowledge-Based Systems*, *23*(3): 267–276.

Constantinou, A.C., Fenton, N., Marsh, W. & Radlinski, L. 2016. From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support. *Artificial intelligence in medicine*, *67*: 75–93.

Dora. J., 2012. A Climate Change Report Card for Infrastructure Working Technical Paper Transport: Rail.

EPA 2009. *Climate Change Health and Environmental Effects: Adaptation*. United States Environmental Protection Agency. Retrieved on 10 June, 2015 from https://www.epa.gov/sites/production/files/signpost/cc.html.

Gutierrez, B.T., Plant, N.G. & Thieler, E.R. 2011. A Bayesian network to predict coastal vulnerability to sea level rise. *Journal of Geophysical Research: Earth Surface*: *116*(F2).

Hooper, E. & Chapman, L., 2012. Chapter 5 The Impacts of Climate Change on National Road and Rail Networks. In *Transport and Climate Change*: 105–136. Emerald Group Publishing Limited.

HUGIN v. 8.5 2017. *Hugin Expert*. Retrieved on 10 May, 2017 from http://www.hugin.com/index.php/resources/.

Jaroszweski, D., Chapman, L. & Petts, J. 2010. Assessing the potential impact of climate change on transportation: the need for an interdisciplinary approach. *Journal of Transport Geography*, *18*(2): 331–335.

Jensen, F.V. 2001. *Bayesian Networks and Decision Graphs. Series for Statistics for Engineering and Information Science*. Springer-Verlag, NY, USA.

Klir, G., & Yuan, B. 1995. *Fuzzy sets and fuzzy logic.* Vol. 4. New Jersey: Prentice hall.

Koetse, M.J. & Rietveld, P. 2012. Adaptation to climate change in the transport sector. *Transport Reviews*, *32*(3): 267–286.

Liang, Y. & Lee, J.D. 2008. Comparing Support Vector Machines (SVMs) and Bayesian Networks (BNs) in detecting driver cognitive distraction using eye movements. *Passive Eye Monitoring: Algorithms, Applications and Experiments; Springer: Berlin/Heidelberg, Germany*: 285–300.

Love, G., Soares, A. & Püempel, H., 2010. Climate change, climate variability and transportation. *Procedia Environmental Sciences*, *1*: 130–145.

McKenzie Hedger, M., Gawith, M., Brown, I., Connell, R. & Downing, T.E. 2000. Climate Change: assessing the impacts—identifying responses. The first three years of the UK Climate Impacts Programme. *UKCIP and DETR, Oxford*.

Network Rail 2015. *Climate Change Adaptation Report 2015.* Retrieved on 7 April 2017, from http://16cbgt3sbwr8204sf92da3xxc5m-wpengine.netdna-ssl.com/wp-content/uploads/2016/11/Climate-Change-Adaptation-Report-2015_FINAL.pdf.

Network Rail 2016. *Great Britain National Rail Network Diagram.* Retrieved on 7 April 2017, from http://www.nationalrail.co.uk/static/documents/content/routemaps/nationalrailnetworkmap.pdf.

Ng, A.K.Y., Chen, S.L., Cahoon, S., Brooks, B. & Yang, Z.L. 2013. Climate change and the adaptation strategies of ports: the Australian experiences. *Research in Transportation Business and Management, 8*: 86–194.

Peterson, T.C., McGuirk, M., Houston, T.G., Horvitz, A.H. & Wehner, M.F. 2008. Climate variability and change with implications for transportation. *Transportation Research Board*.

RFG n.d.. *Rail Fright Group.* Retrieved on 7 April, 2017, from http://www.rfg.org.uk/about-rfg.

RIA n.d.. *Rail Industry Association*. Retrieved on 7 April, 2017, from http://www.riagb.org.uk/about-ria/introduction-to-ria/.

Ristevski, B. 2013. A survey of models for inference of gene regulatory networks. *Nonlinear Anal Model Control*, *18*(4): 444–465.

Sii, H.S. & Wang, J. 2002. Safety assessment of FPSOs—The process of modelling system safety and case studies. *Report of the Project—The Application of Approximate Reasoning Methodologies to Offshore Engineering Design~EPSRC GR/R30624 and GR: 32413*.

Singh, R.K. & Benyoucef, L. 2011. A fuzzy TOPSIS based approach for e-sourcing. *Engineering Applications of Artificial Intelligence*, *24*(3): 437–448.

Tebaldi, C., Smith, R.L., Nychka, D. & Mearns, L.O. 2005. Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the

analysis of multimodel ensembles. *Journal of Climate*, *18*(10): 1524–1540.

Tighe, M., Pollino, C.A., Cuddy, S.M. & Whitfield, S., 2007. A Bayesian approach to assessing regional climate change pressures on natural resource conditions in the central west of NSW, Australia. In *International Congress on Modelling and Simulation (MODSIM2007), December*: 10–13.

Tversky, A. & Kahneman, D. 1975. Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making*: 141–162. Springer Netherlands.

Tversky, A., Kahneman, D. & Moser, P. 1990. Judgment under uncertainty: Heuristics and biases. *In Rationality in action: Contemporary approaches*: 171–188.

UNCTAD 2012. Ad Hoc Expert Meeting on Climate Change Impacts and Adaptation: A Challenge for Global Ports. *Geneva, Palais des Nations, 29–30 September 2011.* Main Outcomes and Summary of Discussions. UNCTAD, Geneva, Switzerland.

Vogt, W.P., Gardner, D.C. & Haeffele, L.M. 2012. *Sampling for Surveys. In When to use what research design:* 128–129. Eds W. Paul Vogt, Dianne C. Gardner and Lynne M. Haeffelethe, Guilford Press, New York.

Walker, L., Figliozzi, M., Haire, A. & MacArthur, J. 2011. Identifying Surface Transportation Vulnerabilities and Risk Assessment Opportunities under Climate Change: Case Study in Portland, Oregon. *Transportation Research Record: Journal of the Transportation Research Board, (2244):* 41–49.

Wang, J. 2003. Technology and safety of marine systems, Elsevier.

Wilby, R.L., Troni, J., Biot, Y., Tedd, L., Hewitson, B.C., Smith, D.M. & Sutton, R.T. 2009. A review of climate risk information for adaptation and development planning. *International journal of climatology*, *29*(9): 1193–1215.

Wu, Y.J., Hayat, T., Clarens, A. & Smith, B. 2013. Climate change effects on transportation infrastructure: scenario-based risk analysis using geographic information systems. *Transportation Research Record: Journal of the Transportation Research Board*, (2375): 71–81.

Yang, Z.L. 2006. *Risk assessment and decision making of container supply chains* (Doctoral dissertation, Liverpool John Moores University).

Yang, Z.L., Bonsall, S. & Wang, J. 2008. Fuzzy rule-based Bayesian reasoning approach for prioritization of failures in FMEA. *IEEE Transactions on Reliability*, *57*(3): 517–528.

Yang, Z.L., Bonsall, S. and Wang, J., 2011. Approximate TOPSIS for vessel selection under uncertain environment. *Expert Systems with Applications*, *38*(12):14523–14534.

Yang, Z.L., Cahoon, S., Chen S.L., Ng A.K.Y. & Becker, A. 2016. In: *Ng, AKY, Cahoon S, Chen SL, Austin B (Eds.) Analyzing Risks Posed by Climate Change on Ports: A Fuzzy Approach:* 24–44. Climate Chang and Adaptation Planning for Ports, Routledge Publishing, UK.

Yang, Z.L., Ng, A.K.Y., Lee, P.T.W., Wang, T., Qu, Z., Rodrigues, V.S., & Lau, Y.Y. 2017. Risk and cost evaluation of port adaptation measures to climate change impacts. *Transportation Research Part D: Transport and Environment.*

Yang, Z.L., Ng, A.K.Y., Lee, P.T.W., Wang, T., Rodrigues, VS., Pettit, S. & Harris, I. 2015. Modelling risk based cost analysis of port adaptation measures to climate change. *Proceedings of the International Conference of Asian Logistics Round Table (ALRT) 2015, Taipei, Taiwan, 31 August–1 September.*

Yang, Z.L., Wang, J., Bonsall, S. & Fang, Q.G. 2009. Use of fuzzy evidential reasoning in maritime security assessment. *Risk Analysis*, *29*(1): 95–120.

Yazdani-Chamzini, A. 2014. An integrated fuzzy multi criteria group decision making model for handling equipment selection. *Journal of Civil Engineering and Management, 20:* 660–673.

Zadeh, L.A. 1992. Fuzzy logic and the calculus of fuzzy if-then rules. In *Multiple-Valued Logic, 1992. Proceedings, Twenty-Second International Symposium:* 480. IEEE.

Zadeh, L.A. 2010. A summary and update of "fuzzy logic". In *Granular Computing (GrC), 2010 IEEE International Conference:* 42–44. IEEE.

Zhang, D., Yan, X.P., Yang, Z.L., Wall, A. & Wang, J. 2013. Incorporation of formal safety assessment and Bayesian network in navigational risk estimation of the Yangtze River. *Reliability Engineering & System Safety*, *118*: 93–105.

Zhou, Y., Fenton, N. & Neil, M. 2014. Bayesian network approach to multinomial parameter learning using data and expert judgments. *International Journal of Approximate Reasoning*, *55*(5): 1252–1268.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Data management for the development of a flood vulnerability model

J.-P. Pinelli & D. Rodriguez
Florida Tech, Melbourne, Florida, USA

D. Roueche
Auburn University, Auburn, Alabama, USA

K. Gurley & M. Baradaranshoraka
University of Florida, Gainesville, Florida, USA

S. Cocke & D.-W. Shin
Florida State University, Tallhassee, Florida, USA

L. Lapaiche & R. Gay
Ecole Nationale d'Ingenieurs de Saint Etienne, France

ABSTRACT: The Florida Public Hurricane Loss Model (FPHLM) is a catastrophe model, which estimates hurricane damage. The FPHLM team has access to three main sources of exposure and claims data: county tax appraiser databases, National Flood Insurance Program (NFIP) portfolios, and private wind insurance portfolios. These databases were processed and cross-referenced at the county level. The FPHLM hazard team assigned estimates of surge and wave height, or fresh water inundation height, to each NFIP claim, as well as estimates of wind speed to each wind claim. The paper shows how combinations of more accurate building descriptions and cause of loss facilitates FPHLM flood model development, calibration and validation. The reliability evaluation, cleaning, geocoding, alignment, and integration of the different datasets, as well as the methods for the development, calibration and validation of the model are described. This is a work in progress and the paper presents some preliminary vulnerability curves.

1 INTRODUCTION

The purpose of catastrophe models such as the Florida Public Hurricane Loss Model (FPHLM) is to estimate the potential damages caused by hurricane events, including coastal flood (storm surge), inland flood, and wind. Cat models have three main components: a hazard component, which models the hurricane hazards; a vulnerability component, which model the effect of the hazard on the buildings; and an actuarial component, which translates the damage into insured losses.

Calibration and validation of the FPHLM wind and flood models outputs are conducted in part by comparisons against historical National Flood Insurance Program (NFIP) and wind insurance claims data. These comparisons require that different kinds of information are available including: 1) actuarial claim data, like values, and cause of loss; 2) building claim data, like location, elevation, age, and building characteristics; and, 3) hazard data, like date and type of event, and hazard intensity.

Once this information has been gathered for all claims in a portfolio, the teams can use the data for model development, validation and calibration. In the first case, vulnerability curves can be derived based on direct or indirect regression analysis techniques on the enhanced claim data. In the second case, empirical vulnerability curves, obtained through regression analysis of the claim data vs. hazard intensity, are compared to the corresponding model vulnerability curves. The empirical and modeled vulnerability curves are then compared to validate the FPHLM flood and/or wind models.

A challenging aspect of this methodology is that, very often, the building data in the insurance portfolios are incomplete, missing, or erroneous, and the hazard intensity data is not provided. The purpose of this paper is to describe the process for gathering missing information, and the methodology for associating that information with a specific claim. Aspects of the subsequent model development, validation and calibration processes are then described.

2 FLORIDA PUBLIC HURRICANE LOSS MODEL

The Florida Office of Insurance Regulation (FLOIR) sponsored the development of the Florida Public Hurricane Loss Model (FPHLM) as a tool for insurance regulation in the state of Florida (FPHLM, 2016; Hamid et al., 2011). The FPHLM originally analyzed portfolios of single family homes, including manufactured homes, subject to wind hazard only. Currently, the scope includes commercial residential buildings (either apartment or condominium buildings), as well as inland and coastal flooding. The purpose of the model is to predict aggregated insured losses for residential properties in the form of annual expected losses (AEL) and probable maximum losses (PML). Such loss estimates are used by insurance companies and state regulators to help evaluate rate filings. The model can also be used to conduct scenario analyses to estimate losses for hypothetical events and historical storms.

Under the sponsorship of FLOIR, the FPHLM team has recently expanded the previously wind-only scope of the FPHLM to include coastal and inland flood hazards (Baradaranshoraka et al., 2017). Their strategy was to adapt the large body of tsunami related building fragility curves, especially the work of Suppasri et al. (2013), to coastal flood, and to adapt the work of the USA Corp of Engineers (USACE, 2006, 2015) for inland flood. In all the models, building vulnerability curves provide estimates of mean building damage ratio as a function of hazard magnitude (wind speed in the case of wind, and inundation height in the case of coastal or inland flood) (Pinelli et al., 2011). The damage ratio is the cost of repair of a damaged building divided by the replacement value of the building. Paramount to that effort is the validation and calibration of the building vulnerability curves.

In addition to building damage, which includes the damage to both the exterior and interior of the building, insurance companies cover the contents damage. Contents is anything inside the building but not attached to the building (e.g. furniture, rugs, appliances, etc.). These contents vulnerability curves are being developed based on regression techniques using the flood claim data, as described in this paper.

3 DATABASES

The FPHLM teams use three primary sources of exposure and claim data

- The National Flood Insurance Program (NFIP) database: exposure files, and a claim files
- Twenty three wind insurance companies: exposure files and claim files
- County tax appraiser databases: building descriptors

Exposure files include all the policies insured by a given company, while the claim files include only the policies that suffered a loss for a particular event. The datasets provided by these sources are briefly summarized below.

3.1 *NFIP database*

The NFIP claims and exposure portfolios were provided to the FPHLM by FLOIR. The claims database contains 153,751 claims between July 1975 and January 2014 for 126 different flood events. The exposure portfolios were provided by year from 1992 to 2012. A combined portfolio was created containing all policies between 1992 and 2012. The hazard team analyzed the claims data locations and dates to associate a specific hazard to a given claim. From these datasets, a trial dataset of 43,552 claims from the eight storms with the most claims was chosen for validating the FPHLM. The eight storms are listed in Table 1.

The NFIP claim files contain information such as the date of loss, policy number, physical address, cause of damage, total property value, financial damage to building and contents, and replacement cost. Fields are present in the files for structural information such as exterior wall type and foundation type, but do not contain values for 97% of the claims. The exposure files contain policy number, flood zone, address, original construction date, base flood elevation, and more, but no structural building information.

3.2 *Tax appraiser databases*

Tax appraiser (TA) datasets have been collected for a total of 51 counties and comprising approximately 97% of Florida's population. However, the Miami-Dade TA database contains virtually no information useful for the purposes of FPHLM calibration and validation.

An ideal TA dataset would include all building properties and components that are damageable by a hurricane: interior and exterior wall composition, roof shape, roof covering, floor covering,

Table 1. Hurricane events currently included in the claims trial dataset from NFIP and wind portfolio databases.

Storm ID	Storm name	Storm year
AL041992	Andrew	1992
AL071998	Georges	1998
AL032004	Charley	2004
AL062004	Frances	2004
AL092004	Ivan	2004
AL112004	Jeanne	2004
AL042005	Dennis	2005
AL252005	Wilma	2005

year built, and number of stories. In addition, the building location, elevation, footprint area, and the number of apartment units per building is desired. However, there is no uniform format or content standard across the different county TA databases, and the amount and quality of data collected by each tax appraiser differs from county to county.

The TA databases typically consist of data tables, which contain the building attribute information, and GIS shapefiles, which typically contain the polygons defining the geographic boundaries of each parcel, and the unique parcel identification number in a linked database file. Some counties did not provide GIS shapefiles, or the shapefiles did not contain parcel identification numbers that matched any unique identifiers in the data tables. When that occurred, GIS shapefiles were sourced from the Florida Department of Revenue which contained parcel identification numbers matching those provided by the county.

3.3 *Wind insurance portfolios*

The wind insurance claim datasets represent 23 different wind insurance companies and nine hurricane events. There are 667,573 claim records among all the insurance portfolios and hurricane events. The data contained in the claim files varies by company and event, but generally include the policy ID, zipcode, county, year built, construction type, and loss to structure, contents, appurtenant structures, and time related losses (additional living expenses). The exposure portfolios also vary in details provided, but generally include policy ID, zip code, county and construction type (frame or masonry). A few companies provided more detailed exposure files that included roof shape, number of stories, roof cover and opening protection. In all, the exposure files contain 13.5 million policies. The latitude and longitude values and/or addresses are provided in the 2012 exposure files for the location of the individual building, and in the claim data files for some of the 2004 and 2005 hurricanes. The precise locations of the buildings in the 2012 exposure portfolios makes it possible to relate the exposure information to NFIP or tax appraiser databases, which is the grand challenge for integrating these datasets. It makes it also possible to identify the location of some claim records, based on the policy ID.

4 PROCESSING OF THE DATABASES

4.1 *Reformatting*

The first step in processing the databases was to reformat the building attribute data contained within the exposure, claims and TA databases into a consistent nomenclature. The standard nomenclature for the main building attributes are provided in Table 2. For each attribute within each database, a link table was developed to relate the unique values for a given attribute within the database to a value matching the common nomenclature provided in Table 2. This process is illustrated in Table 3, where unique values of exterior wall type from the Martin County TA database are recorded to match the standard nomenclature. Generating the link tables was intuitive for most counties and attributes, but documentation on the exact definition of attribute values were generally not available from the counties, which introduces uncertainty in the link tables. For example, in Table 3 it is unclear what type of structural system "Fire Resistant" actually refers to, and so without any other supporting documentation, it is coded as "other" .

Table 2. Common nomenclature for primary building attributes.

Building use	Building category	Roof shape	Roof cover	Exterior wall
RES	PR	Gable	Shingle	Timber
MANUF	LR-CR	Hip	Tile	Masonry-unreinforced
RENTAL	MHR-CR	Other	Metal	
CONDO			Other	Masonry-reinforced
COM				
Other				Other

Table 3. Example link table to reformat attributes into common nomenclature.

TA-Assigned exterior wall	Common nomenclature
Wood frame	Timber
Concrete block	Masonry
Fire resistant	Other
Conc block 12	Masonry
Concrete	Masonry
Reinforced concrete	MasonryR
Fireproof steel	Other
Metal	Other
6 Concrete block	Masonry
Stone local cut	Other
8 concrete block	Masonry

4.2 *Geocoding and integration of the databases*

For a given building with a claim due to flood-induced and/or wind-induced damage during a hurricane, the following data sources can be joined: 1) standardized building attribute infor-

mation from the county TA database, 2) the wind insurance exposure portfolios, 3) the NFIP exposure portfolio, 4) the NFIP claims portfolio, 5) the wind claims portfolios, and 6) the hazard model output. Combined, this will allow for the classification of the claims by building type and hydrological states in the case of flood hazard, where 4 different flood condition states were defined by via the relationship between the wave height to inundation depth. This will facilitate the development of semi-empirical vulnerability curves, as explained in the next section, as well as the comparison of empirical vulnerability curves against the engineering model curves, to facilitate validation and calibration.

The various databases are joined by specific links as illustrated in Figure 1. The links can broadly be classified in two ways—spatial joins and table joins. In table joins, a common field between two fields was used to link different databases together. Table joins based on policy number were used to join the NFIP claims and exposure portfolios, and table joins based on a unique parcel identification number were used to join the tax assessor data tables and shapefiles containing the individual parcel polygons. Two methods of spatial joins were used. For generating the hybrid database linking the hazard model outputs with the NFIP claim portfolio, first the physical address contained in the NFIP exposure database was geocoded to obtain a GPS coordinate. Then the hazard model generated coastal flood heights and/or inland flood heights at the geocoded location for each claim, and the outputs were joined by matching the GPS coordinates (match-spatial-join). Finally, for combining the TA database with the hybrid hazard-NFIP database and, if necessary, the wind exposure or claim

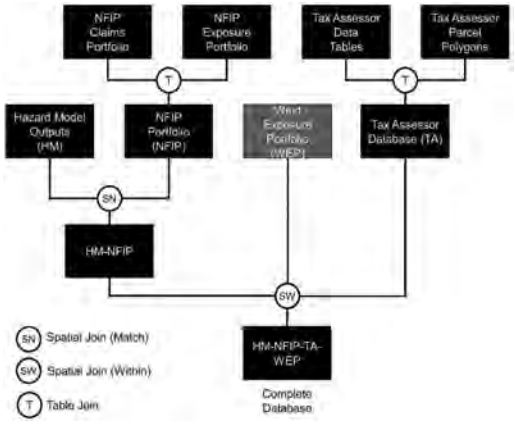


Figure 1. Process for linking hazard model output, NFIP claims and building attributes.

portfolios, a within-spatial-join was used. This was accomplished by coding a graphical user interface (GUI) in Matlab (Mathworks, 2016a) that reads in the county shapefiles and finds any elements (using the GPS coordinates obtained from geocoding the physical address) of the hybrid hazard-NFIP database or wind exposure or claim portfolios that fell within a parcel polygon. For each claim that fell within a parcel polygon, the associated building attributes from the TA database and the wind exposure portfolio were appended to the hybrid hazard-NFIP database, providing a combined dataset linking claims, hazard intensities and building attributes.

4.3 *Hazard information*

The NFIP claims data contains several attributes which can link the loss to a hazard event for each claim. Among these attributes are the "catastrophe number", "cause of loss", "date of loss" and property location. However, upon close examination of the data the FPHLM team found that the information provided by these attributes are not always complete or accurate, and do not provide sufficient detail pertaining to the hazard event. For example, the catastrophe number is often missing, or multiple hazard events may be assigned the same number if they occur around the same time period. The cause of loss often appears to be incorrect. The team found cases where the loss was listed as "Tidal Water Overflow", even though the property was too far from the coast for such an event to occur. In addition, the claims data do not provide important hazard information, such as the flood elevation or wave conditions.

In order to remedy the above issues, algorithms based on observed hazard information from a variety of sources associated hazard event information to each claim. New attributes appended to the claims data provide this information. When no observations are available, such as the case for flood elevation or wave conditions, models provided these estimates. The paragraphs below describe the new attributes.

Distance to coast is the distance of each property to the nearest coast. This information is useful to determine if coastal flood was likely to occur. The team computed the distance using the 2011 Multi-Resolution Land Characteristics Consortium (MRLC) National Land Cover Database (NLCD) (Homer et al., 2015). The NLCD has approximately 30 meter resolution and has a classification for water body. A heuristic algorithm distinguishes coastal waters from inland bodies of water.

Distance to nearest body of water. This information was computed similarly to the distance to

coast, but computes the distance to nearest body of water, regardless of type. This can help determine if the flooding could be due to river, stream or lake overflow.

Precipitation maximum. This data was derived from the high resolution (4 km) PRISM daily rainfall database (Daly et al, 2008). The team selected the largest precipitation amount that occurred with +/– 1 day and within 10 km of the property location. This information helps determine if the cause of loss could be due to accumulation of rainfall.

Storm identification. Many of the claims are due to tropical storms or hurricanes. The HURDAT2 database, from the National Hurricane Center, determines if a storm was in the vicinity of the property during the claimed date of loss. An additional attribute is also provided to include an indication of the storm intensity level (depression, tropical storm or hurricane).

Flood elevation and wave height. Since observed flood elevation data is generally not available, a coastal flood model provided an estimate of the flood elevation due to surge, and an inland flood model in the case of flood due to the accumulation of rainfall. The coastal flood model is based on the CEST model (Zhang et al, 2008), and was driven by estimated observed winds from the H*Wind analyses (Powell et al., 1998). The inland flood model is based on the SWMM model and was driven by observed NEXRAD rainfall data. A simplified wave model developed at the University of Notre Dame was used to provide an indication of wave conditions in the case of coastal flood.

The above information from these new attributes lead to a revision of the cause of loss. If the original cause of loss was tidal water overflow, but the distance to coast was greater than 20 km, the team then checked the precipitation maximum. If it is greater than 1 inch, the cause was revised to be accumulation of rainfall, otherwise it is designated as questionable. If the original cause of loss was accumulation of rainfall, but the precipitation maximum was less than 0.2 inch, then the cause of loss is marked as questionable. For hurricane events, if the original cause of loss is unknown for a property that is less than 20 km to the coast and the precipitation maximum is greater than 1 inch, then the cause of loss is marked as undetermined but is either tidal water overflow or accumulation of rainfall, and further review is needed. Otherwise the cause of loss is marked as tidal water overflow.

The end result is that each claim in the NFIP portfolio can be assigned to one of four hydrological states: inland flood with no waves; coastal flood with minor waves; coastal flood with moderate waves; and, coastal flood with severe waves.

A similar process assigns wind speed to each claims in both the NFIP and wind insurance portfolios but is not reported here.

5 DEVELOPMENT, VALIDATION AND CALIBRATION OF FLOOD VULNERABILITY CURVES

5.1 *Introduction*

The processing of the data is still in progress and will result in enhanced NFIP and wind claims subset that contains loss, value, hazard and hazard intensity, and structural characteristics. When complete, the enhanced NFIP claims data will be used for the development of the flood content vulnerability curves, and the validation and calibration of the FPHLM flood building vulnerability curves.

The model building or content vulnerability curves output are the expected damage ratios (mean damage over replacement value), where the replacement value is the cost to replace the property with a new item of like kind and quality. In the claim data, building and content coverage are used as proxies for the respective building and content replacement values. In the case of the NFIP data, there is no reliable replacement value data provided, and the coverage limits in the NFIP policies are not a true measure of the value of a building or its content. In the future, this issue will be solved through a triangulation between the TA databases, the NFIP exposure files, and the wind insurance exposure files. The idea is to use the coverage limits of the wind policies (which is a better measure of the true value of building or contents) as a proxy for the replacement values of the building and contents in the NFIP claims and exposure portfolios.

Another issue is that extreme hazard events have very large return periods and are underrepresented or altogether absent from the historical claim data. For example, the number of NFIP claims within the State of Florida with a hazard intensity of more than 1.5 m (5 ft) is not significant, and this problem becomes more pronounced as the flood height increases. Hurricanes such as Katrina produced storm surges more than 20 feet (The maximum high water mark observation was 27.8 feet at Pass Christian, MS) (NOAA, 2006), however the storm surge height for the State of Florida only reached up to 5 feet and the FPHLM is constrained to claims in the State of Florida. The result is that any validation and calibration of a vulnerability model based on claim data is more applicable to lower hazard intensities.

This last issue also complicates the development of vulnerability models based on simple regression over the claim data, since not only the assignment of hazard intensity to the claims is subject

to caution, but there might be no data to regress upon for higher hazard intensities. To resolve this problem, the FPHLM team developed the method for the creation of contents vulnerability curves.

5.2 *Development and calibration of contents vulnerability curves for coastal and inland flood*

The FPHLM team used the NFIP claim data in the development of the coastal and inland flood model content damage component. The building vulnerability curve are converted into a content vulnerability curve using a relationship derived from the NFIP claim data. To derive this relationship, the building damage ratio (building damage to building coverage) and content damage ratio (content damage to content coverage) are calculated for each claim.

Figure 2 shows a plot of content damage ratios vs. building damage ratios for a subset of the NFIP claim data of personal residential buildings that have a masonry structure, are not elevated, and are one or two stories. The data is from 16 counties, where the structural information and number of stories have been added (around 6000 of the 43,500 paid claim data for the eight hurricane events stated in Table 1). The 16 counties are spread throughout the State of Florida and contain more than half of the NFIP claim data for the entire state. The NFIP claim data does not provide information regarding the structure type or the number of stories and these were added by matching the NFIP claim data with the tax appraiser databases.

A statistical analysis using a two-way histogram defines the relationship between building damage and content damage ratios. Figure 3 shows a plot of the content damage ratio vs. the building damage ratio relationship derived using statistical analysis with the corresponding curve fit.

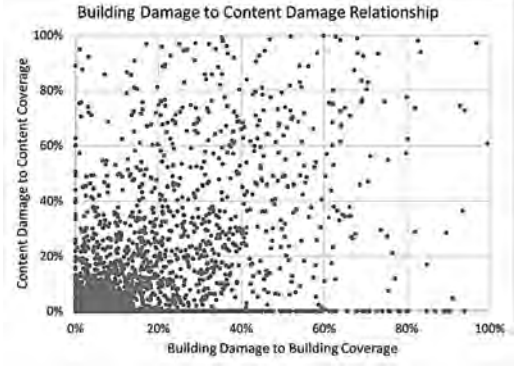


Figure 2. Example of building damage ratio to content damage ratio relationship.

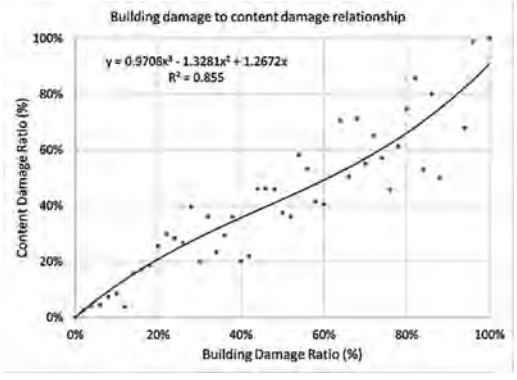


Figure 3. Example of the plot of the building vs. content damage ratios and the polynomial line fitted to the mean values.

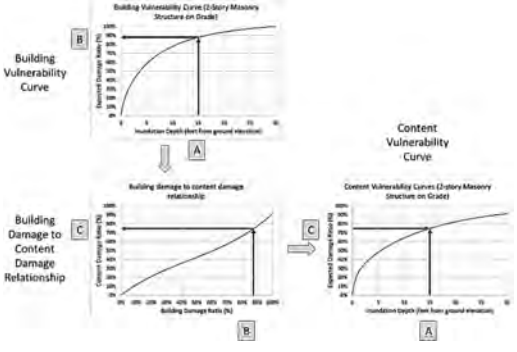


Figure 4. Converting the building vulnerability curves into content vulnerability curves using the building and content damage ratios relationship.

Figure 4 illustrates the process of converting the building vulnerability curves into content vulnerability curves for a two-story on grade masonry structure susceptible to coastal flood with *severe* waves.

The process starts with the building vulnerability curve. For specific hazard intensity ("A" box on the top left graph of Figure 4) we identify the building damage ratio ("B" box on the top left graph). Then, we move to the building and content damage ratios' relationship curve. For the specific building damage ratio found in the first curve ("B" box on the lower left graph), the content damage is identified ("C" box on the lower left graph). Now, for the specific hazard intensity ("A" box on the lower right graph) we know what the equivalent content damage is ("C" box on the lower right

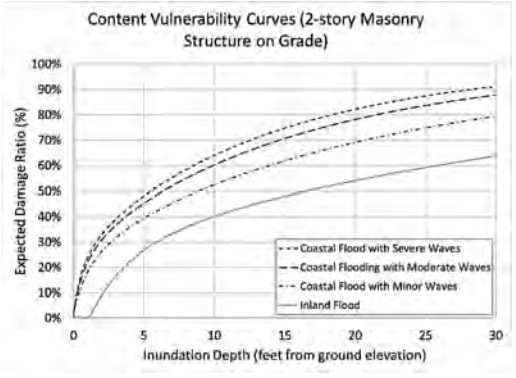


Figure 5. Example of content vulnerability curves. (1 ft = 0.305 m).

graph). This will give a point in the content vulnerability curve. Repeating this process for different hazard intensities will result in the content vulnerability curve.

This methodology resolves the issue of the uncertainty attached to the assignment of a specific hazard intensity to each claim. Instead the claims are simply grouped, according to their appurtenance, to one of the 4 hydrological states. It also resolves to a certain extent the lack of claim data for higher hazard intensities, by combining the claim data with the building vulnerability model, which extends to the whole range of hazard intensity. Finally, it ensures the compatibility between both the building vulnerability model, the claim data, and the resulting content vulnerability model.

Figure 5 shows an example of content vulnerability curves for different hydrological states of a two-story on grade masonry structure. These results are preliminary and intended for the development of the methodology. Significant refinements to the building vulnerability curves are in progress, and the resultant content vulnerability will differ from that shown in Figure 5.

6 CONCLUSIONS AND RECOMMENDATIONS

The analysis of insurance claims data provides a way for developing, validating, and calibrating hurricane risk model outputs. However, before they can be used they need to go through extensive processing and interpretation. This paper describes some of the challenges risk modelers face while handling insurance claims data.

The paper presents a methodology to convert building vulnerability curves into content vulnerability curves using the building damage to content damage relationship derived using NFIP insurance claims data. This method produces contents vulnerability curves compatible with both the claim data and the building vulnerability models.

This is a work in progress. Additional work includes: the validation of the building vulnerability curves, within the limitations of the data; the validation of the wind model; and, the validation of the combined wind and flood model

In addition, the merging of tax appraiser databases with the wind and flood insurance portfolio has the potential of increasing the accuracy of the portfolio analyses, since more data will be available during the analyses.

ACKNOWLEDGEMENT

The Florida Office of Insurance Regulation provided financial support for this work. The opinions, findings, and conclusions presented in this article are those of the authors alone, and do not necessarily represent the views of the FLOIR. Special thanks to Yuepeng Li from Florida International University, and Andrew Kennedy from University of Notre Dame, who provided hazard information

REFERENCES

Baradaranshoraka, M. & Pinelli, J.-P. & Gurley, K., Peng, X. & Zhao, M. 2017. Hurricane wind versus storm surge damage in the context of a risk prediction model. *ASCE Journal of Structural Engineering Multi-hazard special issue* 143(7)

Daly, C. et. al. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol* 28: 2031–2064.

Federal Emergency Management Agency (FEMA) 2015. HAZUS-MH 2.1 hurricane model technical manual. *Federal Emergency Management Agency, Mitigation Division* Washington, DC. Last updated January 29, 2015.

FPHLM–6.2 (2016). Florida public hurricane loss model 6.2. *Florida public hurricane loss projection model (FPHLPM), Laboratory for insurance, financial, and economic research.* International Hurricane Research Center (IHRC), Miami, FL.

Hamid, S. & Pinelli, J.-P. & Cheng, S.-C. & Gurley, K. 2011. Catastrophe Model Based Assessment of Hurricane Risk and Estimates of Potential Insured Losses for the State of Florida. *Natural Hazard Review* 12(4): 171–183.

Homer, C.G. & Dewitz, J.A. & Yang, L. & Jin, S. & Danielson, P. & Xian, G. & Coulston, J. & Herold, N.D. & Wickham, J.D. & Megown, K. 2015. Completion of the 2011 National Land Cover Database for the

conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing* 81(5): 345–354.

National Oceanic and Atmospheric Administration (NOAA). 2006. Hurricane Katrina, A Climatological Perspective. Preliminary Report. *NOAA' National Climatic Data Center*. Asheville, NC. Last updated August 2006.

Pinelli, J.- P. & Pita, G. & Gurley, K. & Torkian, B. & Hamid, S. & Subramanian, C. 2011. Damage Characterization: Application to Florida Public Hurricane Loss Model. *Natural Hazard Review* 12(4): 190–195.

Powell, M.D. & Houston, S.H. & Amat, L.R. & Morisseau-Leroy, N. 1998. The HRD real-time hurricane wind analysis system. *Journal of Wind Engineering and Industrial Aerodynamics* 77 & 78: 53–64.

Suppasri, A. & Mas, E. & Charvet, I. & Gunasekera, R. & Imai, K. & Fukutani, Y. & Abe, Y. & Imamura, F. 2013. Building damage characteristics based on surveyed data and fragility curves of the 2011 Great East Japan tsunami. *Natural Hazards*, 66(2): 319–341.

United States Army Corps of Engineers (USACE) (2015). "US North Atlantic Coast Comprehensive Study: Resilient Adaptation to Increasing Risk." Physical Damage Function Summary Report, US Army Corps of Engineers.

United States Army Corps of Engineers (USACE) 2006. Depth-damage relationships for structures, contents, and vehicles and content-to-structure value ratios (CSVR) in support of the Donaldsonville to the Gulf, Louisiana, Feasibility Study. *US Army Corps of Engineers (USACE)*, New Orleans District, Louisiana.

Zhang, K. & Xiao, C. & Shen, J. 2008. Comparison of the CEST and SLOSH Models for Storm Surge Flooding. *Journal of Coastal Research* 24(2): 489–499.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Optimizing warnings for slippery runways based on weather data

Arne B. Huseby
University of Oslo, Oslo, Norway

Marit Rabbe
Air Navigation Services, Avinor, Norway

ABSTRACT: Slippery runways represent a significant risk to aircrafts especially during the winter season. In order to apply the appropriate braking action, the pilots need reliable information about the runway conditions. Unfortunately the accuracy of runway reports can sometimes be unsatisfactory. In order to obtain more precise and up-to-date information about the current conditions, a warning system based on various types of weather data was suggested by Huseby & Rabbe (2012). See also Huseby & Rabbe (2008) and Huseby et al. (2010). The system is based on a set of scenarios known to cause slippery conditions. By monitoring meteorological parameters like air and ground temperature, humidity, visibility and precipitation, and comparing these to the given scenarios, the system can issue warnings to the ground personnel. This system is currently being used on 16 Norwegian airports. In the present paper this warning system is reviewed. Ideally, the warning system should issue warnings whenever the estimated runway conditions are medium or worse. At the same time the system should not issue warnings when the runway conditions are good. Thus, there are two types of errors we need to take into consideration. Type 1 errors occur when the system does not issue a warning even though the conditions are medium or worse, while Type 2 errors occur if a warning is issued when the conditions are good. When designing the system, we need to find the optimal balance between these types of errors taking into account that a Type 1 error to a certain degree is considered to be worse than a Type 2 error. The paper describes how the system can be optimized using a combination of weather data and flight data.

1 INTRODUCTION

Slippery runways represent a significant risk to aircrafts especially during the winter season. Accidents, such as the Southwest Airlines jet skidding off a runway at Chicago Midway Airport in December 2005, as well as the similar accident with the Delta Connection flight at the Cleveland Hopkins International Airport in Ohio in February 2007, show that this is indeed a serious problem. More recently, an aircraft skidded off the runway at Oslo Airport in May 2015 due to wet conditions. Fortunately this accident only caused minor damages.

In order to apply the appropriate braking action, the pilots need reliable information about the runway conditions. Unfortunately the accuracy of runway reports can sometimes be unsatisfactory. During the Southwest Airlines accident the pilots based their landing on an assumption that conditions were fair. However, computer calculations after the crash showed that the actual conditions were in fact worse than poor. Given the correct information about the landing conditions, including a significant tailwind of 8–9 kt, the flight should in fact have been diverted. For more details about this accident see Rosenker et al. (2007). For a discussion of the effect of contaminated runways on aircraft braking performance see Giesman (2005).

Having reliable methods for identifying slippery runway conditions is very important. However, measuring the runway friction with a satisfactory precision is very difficult. There are two main reasons for these problems. Firstly, measuring the runway friction with a satisfactory precision is very difficult. While many different measurement devices have been developed, it is hard to find equipment that produces stable and consistent results. The second problem is that in order to measure friction, the runway needs to be closed for traffic. Thus, in order to avoid severe delays, such measurements cannot be carried out too frequently. As a result the runway reports are not as useful as one could hope. This is especially true during heavy snowfalls, or when the temperature suddenly drops below the freezing point, where the conditions change very rapidly.

Rosenker et al. (2007) discusses the difficulties with assessing runway condition, and notes that

no standardized and universally accepted correlation exists to define the relationship between the runway surface condition, using any of the available runway surface assessment methods, and an airplane's braking ability.

In Haugen et al. (2002) an alternative approach to this problem was developed. Contaminated runways were characterised in terms of a function of local weather parameters. The main idea was that this function would be easier to update compared to friction measurements which are based on the last runway report. Thus, by using weather data one could bridge the gaps between the runway reports.

Based on a large-scale study of runway conditions carried out during two winter seasons at two Norwegian airports the ideas suggested in Haugen et al. (2002) were developed further. A complete report from this project, referred to as the SWOP-project, is given in Aarrestad et al. (2007). The study was carried out by *Avinor*[1] with contributions from the three airlines SAS, Norwegian and Widerøe. As in Haugen et al. (2002) the main goal was developing methodology for predicting runway conditions utilizing weather data in addition to runway reports. Throughout the two seasons various kinds of weather data were collected, such as air and surface temperature, humidity, precipitation, visibility and wind. Using these data a scenario based weather model for slippery conditions was developed. At a given point of time, the weather model compares the current conditions to a set of different scenarios. If the conditions match any of these scenarios, the model classifies the runway conditions as *potentially slippery*.

Using the experiences from the SWOP-project, an integrated runway information system, called IRIS, has been developed. This system is now implemented on 16 Norwegian airports. For a description of this system see Søderholm et al. (2009). This system consists of three parts: a *weather* model, a *runway* model and a *development* model. The weather model uses a scenario approach to identify slippery conditions. A description of an early version of this model can be found in Huseby & Rabbe (2008), while a revised version is presented Huseby & Rabbe (2012). The weather model currently in use is yet another revision based on more recent data. The runway model uses mainly runway report data and assesses runway conditions on a five level scale ranging from *poor* to *good*. This model is discussed in Huseby et al. (2010). See also Klein-Paste et al. (2012). The development model combines runway report data and precipitation and temperature data in order to issue warnings when the runway conditions appear to be deteriorating.

In the present paper we focus on the weather model and show how this model can be optimized. In order to analyze and optimize the model, a large data set including weather data, runway report data and flight data has been collected. The full data set consists of data from 16 Norwegian airports. In the present paper, however, we only consider data from the airports at Oslo and Tromsø. For both airports we have observations from *9 winter seasons* starting at the winter 2008/2009. The weather observations are sampled every minute starting at midnight November 1 and ending at midnight April 30. The flight data sets are provided by Scandinavian Airlines Services (SAS) and Norwegian Air Shuttle AS. In this paper, however, only the flight data sets from SAS are used. At Oslo airport there are two runways, *West* and *East*. In the analysis these runways are treated separately.

2 FRICTION LIMITED LANDINGS

In order to optimize the weather model, flight data was obtained from the Quick Access Recorder of Boeing 737-600/700/800 NG airplanes. The data was provided with approval of the Pilots Associations of the cooperating airlines. Starting at the time of touchdown, a 60 seconds record was taken including among others the following main parameters:

- Airplane weight
- Longitudinal acceleration
- Airspeed
- Ground speed
- Flaps settings
- Spoiler settings
- Engine rotational speed
- Brake pressures
- Auto brake settings
- Longitude and latitude positions

The flight data was analyzed using the *Boeing Airplane Performance model* which is based on general equations of motion for airplane, along the length direction of the runway. The model gives an *airplane braking coefficient*, denoted by μ_B. This parameter is used to represent the contribution of the wheel brakes to stopping the airplane. The coefficient μ_B is the ratio of the stopping force contribution of the wheel brakes to the average airplane weight on wheels. In general μ_B will include both the wheel braking and the effect of contaminant drag force.

A key concept when analysing flight data is the notion of *friction limited landings*. Unless the pilot challenges the runway friction during the

1. Avinor is a state owned limited company that operates most of the civil airports in Norway.

landing, the maximum friction available will not be utilized. In this case μ_B reflects the amount of tire-pavement friction that was used. When wheel brakes are applied fully or to a high degree on a slippery runway, the maximum attainable friction from the runway is used during the stop. In this case the airplanes deceleration and stopping capability is limited by the friction available from the runway. The obtained μ_B will then reflect the amount of tire-pavement friction that was available. It is therefore crucial to determine whether or not the stop was limited by the friction available from the runway. A landing where this is the case, is said to be *friction limited*. For more details about the Boeing Airplane Performance model see Klein-Paste et al. (2012).

Given the airplane braking coefficient μ_B it is possible to obtain the socalled *runway braking action* (BA) associated with the landing. This is a simplified measure given according to a five level scale randing from *poor* to *good*. The relation between μ_B and BA is given Table 1.

In the validation of the weather model only the friction limited landings are used. In Table 2, Table 3 and Table 4 the total number landings as well as the number of friction limited landings for the three runways are listed. In the same tables we have also included the number of landings with braking action level less than or equal to 3 (i.e., *medium*).

Table 1. The relation between μ_B and runway Braking Action (BA).

μ_B	BA-level	BA
[0.000, 0.050]	0	NIL
(0.050, 0.075]	1	Poor
(0.075, 0.100]	2	Medium to poor
(0.100, 0.150]	3	Medium
(0.150, 0.200]	4	Medium to good
(0.200, ·]	5	Good

Table 2. Number of friction limited landings at Oslo West.

	Count	Percentage
Total number of landings	57324	100.0%
Friction limited landings	997	1.7%
Landings with BA ≤ 3	845	1.5%

Table 3. Number of friction limited landings at Oslo East.

	Count	Percentage
Total number of landings	54794	100.0%
Friction limited landings	981	1.8%
Landings with BA ≤ 3	764	1.4%

Table 4. Number of friction limited landings at Tromsø.

	Count	Percentage
Total number of landings	12362	100.0%
Friction limited landings	2214	17.9%
Landings with BA ≤ 3	1866	15.1%

3 THE WEATHER MODEL

In this section we review briefly the scenario based weather model which is a central part of the runway condition prediction methodology. This model is based on the work in Rabbe (1974), and includes eight different scenarios which are known to cause slippery runway conditions. In this context we will not describe all these scenarios in detail. Instead we refer to Huseby & Rabbe (2012) for a more complete description.

All scenario evaluations are typically done relative to a given point of time T representing the touchdown point of time for a given flight. In order to describe the weather conditions and check whether or not any of the scenarios has occurred at T, weather data from two time intervals $[T - S_2, T - S_1]$ and $[T - S_1, T]$ is used. In our study $S_1 =$ 1 hour, while $S_2 = 4$ hours. The length of the first interval is $(T - S_1) - (T - S_2) = S_2 - S_1 = 3$ hours, while the length of the second and most recent interval is $T - (T - S_1) = S_1 = 1$ hour. Thus, the two intervals represent a total of 4 hours of observations. Throughout this period the different weather parameters are ideally sampled once every minute, so given the four hours of observations, each weather parameter is sampled 180 times during the first interval and 60 times during the last interval. In real life, the number of observations is typically slightly less than this, but still there is more than enough of data for the model calculations. We assume that all the data are indexed, and let I_1 and I_2 denote the index sets corresponding to the first and second interval respectively. Moreover, for $i \in I_1 \cup I_2$ we introduce:

- p_i = the ith precipitation type,
- t_i^a = the ith air temperature,
- t_i^r = the ith runway temperature,
- h_i = the ith relative humidity,
- v_i = the ith horizontal visibility,

The scenarios are divided into two groups, where the first group includes three scenarios with precipitation, while the second contains the remaining three scenarios with no precipitation.

3.1 *Scenarios with precipitation*

The following five scenarios are all characterized by the presence of some sort of precipitation.

Scenario 1. *Dry snow*
Dry snow is usually not so dangerous as wet snow. However, if this condition persists over some time, the runway may become polished by the snowflakes, which can result in slippery conditions.

Scenario 2. *Snow*
Snowfalls can have dramatic effects on braking performance. Falling snow is typically a mixture of ice crystals and water with a temperature close to 0°C. Dry snow containing less water is less slippery than snow with a higher water content. This scenario is defined to occur when the air temperature is between –6°C and +2°C. Severe conditions occur with heavy snowfalls, temperatures close to the freezing point, and relative humidity close to 100%.

Scenario 3. *Freezing rain/drizzle*
Freezing rain or freezing drizzle occurs when warm air tries to replace cold air from above. If rain or drizzle falls from the warm layer through the cold layer, it will be supercooled on its way down. When these supercooled drops hit the frozen ground, they will freeze immediately, and as a result the runway becomes extremely slippery. The weather sensor can sometimes identify this condition as a specific type of precipitation. However, in our scenario definition, we also include another set of conditions based on the precipitation type *rain*.

Scenario 4. *Freezing fog*
When temperatures at ground level drop to or below freezing, the water droplets making up fog often freeze on contact. This condition is called *freezing fog*. The result can be black ice, which makes the runway very slippery and dangerous.

Scenario 5. *Rain or drizzle on ice-coated or supercooled runway*
When rain or drizzle falls on an ice-coated or supercooled runway, some of it will start freezing. As a result the braking actions will be reduced to a minimum.

3.2 *Scenarios without precipitation*

The scenarios included in this subsection does not involve precipitation, at least not during the most recent interval, $[T - S_1, T]$.

Scenario 6. *Wet runway, clearing sky*
This scenario occurs when the weather is clearing after overcast and rain. Due to the outgoing radiation, the temperature both in the air and in the runway surface gradually drop below zero. The water left on the runway, will then start to freeze, and the friction coefficient may drop quickly.

Scenario 7. *Stratus/fog, air temperature below* 0°C
When a low stratus cloud or a fog layer at freezing temperatures flows into the airport, small drops will settle on the cold runway surface and partially freeze. As a result the runway becomes slippery.

Scenario 8. *Rime, sublimation, ice crystals*
In clear autumn and winter nights the temperature in the air and on the ground can fall below zero. At the same time the humidity increases towards 100%. Sometimes this results in fog, while other times moisture may settle on the runway as rime. When water vapor sublimates on solid objects, the result is ice crystals. Some of this may melt under the wheels of the aircrafts, and create a slippery ice-coat on the runway.

In the analysis the weather model is calculated each minute throughout the 9 winter seasons based on the data from these seasons. As as result we have as many as 2348640 realizations of this model for the three runways in our data set. Figure 1, Figure 2 and Figure 3 show the rates of occurences of the eight scenarios (denoted $S1$, …, $S8$ in the diagrams). We have also included the null-scenario (denoted $S0$ in the diagrams) representing the fraction of the realizations with no identified scenarios. Note that the scenario definitions allow several scenarios to occur at the same time. Thus, the sum of the rates of occurences (including the null-scenario) is greater than 100%.

We observe that the two dominating scenarios are $S2$ (*snow*) and $S8$ (*Rime, sublimation, ice crystals*). Scenario $S2$ is relatively easy to identify. Thus, we will not discuss this scenario further here. Scenario $S8$, on the other hand, is much more difficult to identify. The scenario definition depends

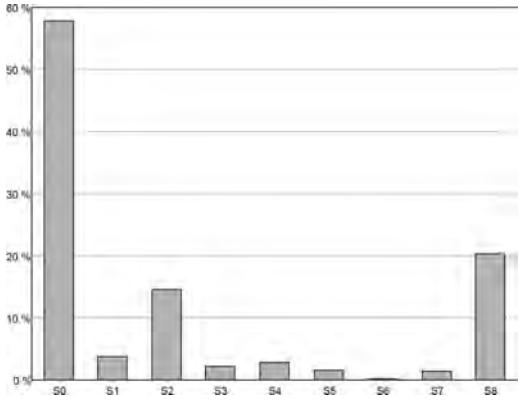


Figure 1. Rates of occurences of the weather scenarios at Oslo West.

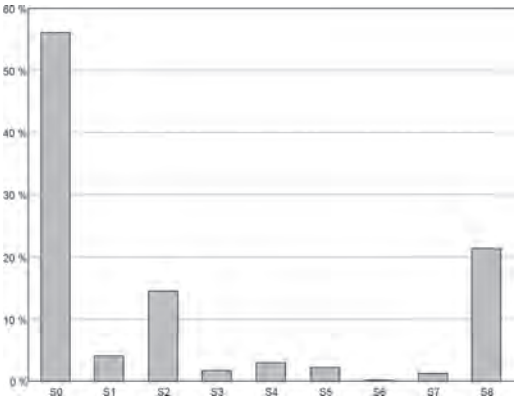


Figure 2. Rates of occurences of the weather scenarios at Oslo East.

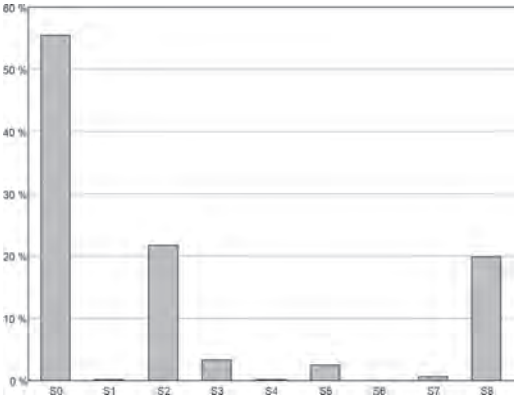


Figure 3. Rates of occurences of the weather scenarios at Tromsø.

on several parameters which need to be finetuned in order to optimize the results. In the next section we show how this problem can be attacked.

4 OPTIMIZING THE RIME SCENARIO

In this section we show how to optimize the rime scenario, i.e. scenario $S8$. We say that this scenario has occurred if the following conditions hold:

- The number of minutes with precipitation[2] during the last 4 hours should not exceed 10.
- The air temperature is decreasing during the last 4 hours, and last value is less than or equal to α°C,

  OR

  The air temperature is less than or equal to 0°C during the last hour
- The runway temperature is less than or equal to 0°C during the last hour
- The relative humidity is in the interval [β%, 100%] at least half of the time during the last four hours

The quantities α and β mentioned in the conditions, are parameters which will be subject to simultaneous optimization. Based on meteorological insight, however, it was decided that the parameters should be chosen within the following intervals: $\alpha \in [0, 3]$ and $\beta \in [70, 85]$. As a *base case* we let $\alpha = 2$ and $\beta = 75$. In the optimization only the parameters α and β will be considered here. However, as scenario $S8$ is only one out of eight scenarios, the contributions of the other scenarios need to be taken into account as well.

One of the difficulties with optimizing the weather model is the lack of precision with respect to the response. Even when the runway is very slippery, this does not need to be reflected in the flight data if the pilot does not challenge the runway friction during the landing. On the other hand even when no weather scenario is identified, the flight data might indicate slippery conditions if this is due to the presence of older contamination on the runway. Thus, when the weather model is applied, a substantial number of both Type 1 and 2 errors are to be expected. Still, finding the best balance between the two types of errors is important.

We now consider an arbitrary point of time t, and introduce the following events:

A = A scenario is identified at time t

B = The true BA-value is 3 or less at time t

A Type 1 error corresponds to the event $A^c \cap B$, while a Type 2 error corresponds to the event $A \cap B^c$. In order to balance the two types of errors, we introduce a loss function L defined as follows:

$$L = \begin{cases} K_1 & \text{if } A^c \cap B \text{ occurs} \\ K_2 & \text{if } A \cap B^c \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

The expected loss is then given by:

$$\mathrm{E}[L] = K_1 \mathrm{P}(A^c \cap B) + K_2 \mathrm{P}(A \cap B^c)$$

The constants K_1 and K_2 are relative numbers chosen in order to reflect that a Type 1 error is usually much worse than a Type 2 error. In the analysis we have chosen to let $K_1 = 25$ and $K_2 = 1$.

2. In this context *mist* is not considered as precipitation.

The probabilites $P(A^c \cap B)$ and $P(A \cap B^c)$ are easily estimated based on the available data, and depend on the values of the parameters α and β. In particular the event A is identified using the weather data, while the event B is identified using flight data.

In particular, by using the data in Table 2, Table 3 and Table 4 we get the following estimated probabilities for the event B:

$$P(B) = \begin{cases} \frac{845}{57324} & = 1.47\% \text{ at Oslo West} \\ \frac{764}{54794} & = 1.39\% \text{ at Oslo East} \\ \frac{1866}{12362} & = 15.09\% \text{ at Tromsø} \end{cases}$$

It is easy to see that $P(A)$ is increasing in α and decreasing in β. Hence, the Type 1 error probability, $P(A^c \cap B)$ is decreasing in α and increasing in β, while the Type 2 error probability, $P(A \cap B^c)$ is increasing in α and decreasing in β. Moreover, for the base case where $\alpha = 2$ and $\beta = 75$, we get the error probabilities and expected losses listed in Table 5.

In order to optimize the weather model we run it for all the relevant combinations of the parameters α and β. The resulting expected losses are shown in Table 6, Table 7 and Table 8. The parameter combinations yielding the minimum losses are indicated in bold face.

For Oslo West and East we observe that the optimal values are $\alpha = 0$ and $\beta = 85$. This is the most restrictive combination which implies that the probability of a Type 1 error is at its maximum while the probability of a Type 2 error is at its minimum. From a meteorological perspective, these parameter values are hardly realistic. The reason why this combination still comes out as the best, is due to the fact that Oslo Airport has a fairly proactive runway maintenance strategy which prevents the different scenarios from causing problems. As a result $P(B)$ is relatively small for this airport. Thus, if we use $\alpha = 0$ and $\beta = 85$ instead of the base case values $\alpha = 2$ and $\beta = 75$, the Type 1 error probability increases slightly from 0.23% to 0.24%, while the Type 2 error probability decreases significantly from 31.01% to 26.06%.

For Tromsø we have the opposite situation where the optimal values are $\alpha = 3$ and $\beta = 70$. This is the least restrictive combination which implies that the probability of a Type 1 error is at its minimum while the probability of a Type 2 error is at its maximum. At this airport it is more common to have a contaminated runway during parts of the winter season. As a result avoiding Type 1 errors are more important.

Table 5. Type 1 and Type 2 error probabilities and expected losses.

Runway	$P(A^c \cap B)$	$P(A \cap B^c)$	$E[L]$
Oslo West	0.23%	31.01%	0.367
Oslo East	0.29%	44.70%	0.520
Tromsø	3.37%	28.87%	1.132

Table 6. Expected losses for various combinations of α and β at Oslo West. Optimal combination is indicated in bold face.

	$\beta = 70$	$\beta = 75$	$\beta = 80$	$\beta = 85$
$\alpha = 0$	0.3685	0.3586	0.3431	**0.3204**
$\alpha = 1$	0.3741	0.3634	0.3477	0.3245
$\alpha = 2$	0.3782	0.3668	0.3508	0.3265
$\alpha = 3$	0.3792	0.3674	0.3513	0.3268

Table 7. Expected losses for various combinations of α and β at Oslo East. Optimal combination is indicated in bold face.

	$\beta = 70$	$\beta = 75$	$\beta = 80$	$\beta = 85$
$\alpha = 0$	0.5310	0.5155	0.4973	**0.4716**
$\alpha = 1$	0.5351	0.5188	0.5004	0.4742
$\alpha = 2$	0.5362	0.5195	0.5007	0.4748
$\alpha = 3$	0.5370	0.5203	0.5012	0.4752

Table 8. Expected losses for various combinations of α and β at Tromsø. Optimal combination is indicated in bold face.

	$\beta = 70$	$\beta = 75$	$\beta = 80$	$\beta = 85$
$\alpha = 0$	1.1044	1.1446	1.2399	1.4198
$\alpha = 1$	1.0980	1.1370	1.2369	1.4185
$\alpha = 2$	1.0929	1.1320	1.2369	1.4175
$\alpha = 3$	**1.0857**	1.1264	1.2330	1.4175

Table 9. Expected losses for various combinations of α and β at Tromsø given that $K_1 = 10$. Optimal combination is indicated in bold face.

	$\beta = 70$	$\beta = 75$	$\beta = 80$	$\beta = 85$
$\alpha = 0$	0.6312	0.6265	0.6405	0.6906
$\alpha = 1$	0.6320	0.6262	0.6412	0.6916
$\alpha = 2$	0.6330	0.6260	0.6423	0.6919
$\alpha = 3$	0.6306	**0.6240**	0.6408	0.6919

It should be noted that the optimal parameter values depend on the choice of the loss factors K_1 and K_2 also. If more less weight is put on Type 1 errors, i.e., if K_1 is reduced, the optimal parameter values will change accordingly. As an illustration we have computed the expected losses for Tromsø airport given that K_1 is reduced to 10. The results are shown in Table 9. In this case the optimal parameter combination is $\alpha = 3$ and $\beta = 75$.

5 CONCLUSIONS AND FUTURE WORK

In the present paper we have reviewed the weather model used in the integrated runway information system IRIS. We have shown how the model can be optimized by using a combination of weather data and flight data. The methodology is demonstrated on a simplified problem with only two parameters. In a full scale analysis, a multidimensional optimization must be carried out. Furthermore, the model needs to be finetuned in order to work in combination with the other models in the system. In order to carry out such a complex optimization it is important to screen the parameters and choose the relevant ranges carefully. Moreover, in order to handle the enormous amount of data, an efficient database structure must be applied.

This work is just a small part of a much larger study which includes all the weather scenarios, the use of all available flight data as well as separate analysis for all 16 airports where the IRIS system is installed. A more extensive report from this analysis will be available later.

ACKNOWLEDGEMENTS

The authors are grateful to Avinor for funding this study.

REFERENCES

Aarrestad, O., A. Norheim, A.B. Huseby, & M. Rabbe (2007). Safe winter operation project (SWOP). *Project Report, Avinor, Norway*. In Norwegian.

Giesman, P. (2005). Wet runway, physics, certification, application. *Boeing Performance and Flight Operations Engineering Conference* (8), 1–24.

Haugen, Ø., T. Kirkhus, M. Carlin, & O. Løvhaugen (2002). Empirical prediction of stop distances for aircrafts at contaminated runways. *Technical Report, SINTEF, Norway* (STF72 F02607).

Huseby, A.B., A. Klein-Paste, & H.J. Bugge (2010). Assessing airport runway conditions—a bayesian approach. In A.B, P.I, and Z.E. (Eds.), *Reliability, Risk and Safety Back to the Future*, pp. 2024–2032. London CRC Press.

Huseby, A.B. & M. Rabbe (2008). Predicting airport runway conditions based on weather data. In M.S, G.S.C., and B.J. (Eds.), *Safety, Reliability and Risk Analysis: Theory, Methods and Applications*, pp. 2199–2206. London CRC Press.

Huseby, A.B. & M. Rabbe (2012). A scenario based model for assessing runway conditions using weather data. In P. ESREL (Ed.), *11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference*, pp. 5092–5101. Curran Associates, Inc.

Klein-Paste, A., A.B. Huseby, J. Anderson, P. Giesman, H.J. Bugge, & T.B. Langedahl (2012). Braking performance of commercial airplanes during operation on winter contaminated runways. *Cold Regions Science and Technology* (79–80), 29–37.

Rabbe, Å. (1974). Slippery runways caused by meteorological factors. *Technical Report, Meteorological Institute, Norway* (20).

Rosenker, M.V., R.L. Sumwalt, D.A.P. Hersman, K.O. Higgins, & S.R. Chealander (2007). Runway overrun and collision southwest airlines flight 1248. *Aircraft Accident Report, National Transportation Safety Board* (NTSB/AAR-07/06 PB2007-910407).

Søderholm, B., H.J. Bugge, A.B. Huseby, M. Rabbe, A. Klein-Paste, E. Bergersen, & P. Skjøndal (2009). Integrated runway information system (IRIS). *Project Report, Avinor, Norway*. In Norwegian.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Risk management for natural hazards based on reliability analysis: A case study of landslides

Jongook Lee
Interdisciplinary Program in Landscape Architecture, Seoul National University, Seoul, South Korea

Dong Kun Lee
Research Institute of Agriculture Life Science, Seoul National University, Seoul, Republic of Korea

ABSTRACT: To cope with natural disasters, it is necessary to establish a strategy for adaptation after analyzing the risk originating from natural hazards. When conducting risk analysis, the first step is estimating the frequency of occurrence, followed by calculating the probability of occurrence. This study aimed to estimate the frequency of landslide occurrence, using the correct units, and to present how the probability of occurrence can be calculated utilizing the concept of reliability, which corresponds well to the outcomes obtained when a Poisson distribution is employed. For the analysis, a pixel unit of GIS-based spatial information was considered as a component to denote the unit for the frequency of landside occurrence, and the equation for the probability of occurrence was derived from the definition of reliability. As a result, the frequency of landslide occurrence in the study region was demonstrated, and a sample calculation of the probability of landslide occurrence is presented.

1 INTRODUCTION

Disasters caused by natural hazards are especially dangerous because they affect larger areas, with greater intensity, than disasters caused by human activities. In addition, the frequency of natural disasters has increased in recent years due to climate change (IPCC, 2013). In order to cope with natural disasters, it is necessary to establish a strategy for adaptation after analyzing the risk originating from natural hazard, to mitigate the risk below acceptable criteria, and to manage it constantly to prevent unwanted loss of life, property, or environment. Landslides are a type of natural disaster that frequently have caused significant damage, especially near mountainous regions (Guzzetti, 2000). Thus, risk analysis for landslides was conducted in this study, utilizing the concept of reliability.

The first step in conducting risk analysis is to estimate the frequency of occurrence (Corominas et al., 2014), followed by calculating the probability of occurrence. However, estimating the frequency can be challenging because of the long interval between relevant events over time, and the lack of accumulated inventory information available in a study region (van Westen et al., 2006, Jaiswal et al., 2010). In addition, frequency units considering spatial and temporal probability are often applied in misinformed ways in risk analysis. Regarding the probability of natural dis-aster occurrence, calculations are often based on a Poisson distribution, sometimes without a clear explanation.

The objective of this study was to estimate the frequency of landslide occurrence, using the correct units, and to present how the probability of occurrence can be calculated utilizing the concept of reliability. For the analysis, a pixel unit of GIS-based spatial information was considered as a component to denote the unit for the frequency of landside occurrence, and the equation for the probability of occurrence was derived from the definition of reliability (Lee et al., 2017).

To estimate the frequency of landslide occurrence, the point locations where landslide events have occurred were considered as failed components, and the other areas, in which no landslide has occurred, were regarded as surviving components. To find the specified time between landslide events, a rainfall threshold was established to estimate the frequency of landslide occurrence due to the limitation of landslide inventory data. A number of methodologies for setting the rainfall threshold have been examined (Frattini et al., 2009, Polemio and Sdao, 1999, Caine, 1980, Aleotti, 2004); however, the validity of those methods are pertinent only with the local geo-spatial properties (Martelloni et al., 2012, Jakob et al., 2006). Therefore, the rainfall threshold value for this study was deter-mined based on local research.

To express the equation of the probability of landslide occurrence, the reliability function, which was derived from the definition of reliability, was used. The equation derived from the concept of reliability has the same outcomes as those reported by Crovelli (2000) based on a Poisson distribution model, which is commonly employed to model the random disastrous events caused by natural hazards over time.

As a result, the frequency of landslide occurrence was determined, and a sample calculation for the probability of landslide occurrence is presented for the study region, Gangwon Province in South Korea, from which landslide inventory data and a landslide hazard map were available for the analysis. The resulting model can be used for risk management of landslides, and also can be extended to assess various mitigation measures to handle the risk from natural hazards.

2 METHOD

2.1 *Study site and inventory data*

Gangwon Province is well known as the region where Pyeongchang County, the host of the 2018 Olympic Winter Games, is located. The province includes high mountainous regions in its boundaries, as shown in Figure 4. The elevation of its northeast area is higher than other parts of the country, and the area has steeper terrain. Due to its topographical features, Gangwon Province is prone to landslides. The population of Gangwon Province is currently growing, reflecting the high demand for leisure activities and property investment. Consequently, housing is expanding, both for permanent residences and for tourism, into mountainous areas that are susceptible to landslides.

The landslide inventory data were provided by the local government of Gangwon Province. The inventory was conducted when devastating damage was reported after Typhoon Ewiniar brought heavy rainfall to the region in July 2006. The damage and losses caused by the typhoon were recorded as a historic natural disaster, as it caused 62 casualties in total, and the loss of property was estimated to be over 1.5 billion USD (N.E.M.A., 2007). The landslide inventory data used for the study were part of the second and third rounds of data collection after the field survey that covered the overall region of Gangwon Province, and the inventory data contained a summary of observed landslide events, with information about the location and damaged area. The survey results were recorded in a polygon-type file in a GIS system and plotted on a 1:25000 scale, including the GPS location data and the areas affected by landslides.

2.2 *Landslide hazard map*

In order to differentiate the frequency of landslide occurrence, we adapted the landslide hazard map produced by the Korea Forest Service in 2012, which was made based on logistic regression analysis considering the following nine factors of mountain properties: slope inclination, slope orientation, slope length, slope curvature, topographic wetness index, the type of forest, the age of the forest, soil depth, and bedrock (Korea Forest Service, 2012). The map classified the analyzed area into five hazard classes, from grade 1 (the highest) to grade 5 (the lowest), with pixel units sized 10 m × 10 m, and the map was projected on a 1:25000 scale. Grade 5 was excluded from our analysis because those pixel data were marked with a null value, including areas of water and flat land with no hazard of landslides.

2.3 *Frequency of landslide occurrence*

Landslide risk can be briefly expressed by an equation in which the probability of a hazardous event is multiplied by the probability of loss of life or property (A.G.S., 2000). For risk analysis, the frequency of landslide occurrence must be identified prior to calculating the probability of landslide occurrence. To estimate the frequency of landslide occurrence, a unit pixel of GIS-based spatial information was considered as a component item to denote the unit for the frequency of landside occurrence. The frequency of landside occurrence can be interpreted as the instantaneous failure rate, presented in terms of the number of failures per unit time, and it is based on measurements of the quantity of components exposed to a stressful environment (Goble and Cheddie, 2005).

Based on the assumption above, the frequency of landslide occurrence can be derived from the concept of the failure rate as below:

$$\lambda = (N - Ns)/(Ns \times \Delta t) \quad (1)$$

where λ = the failure rate of the pixel components, which corresponds to the frequency of landslide occurrence; N = the number of total items; Ns = the number of surviving items; and Δt = the time between landslide occurrences.

Given that Ns is defined as the number of pixel components with no landslide initiation, subtracting Ns from N in Formula 1 can be considered as a measure of the number of landslide events because the point locations of landslide initiated were regarded as the failed components. Time to fail, which refers to the time to landslide reoccurrence, is given by Δt in Formula 1, and it expresses the probabilistic period between landslides. It can be estimated by establishing the rainfall threshold.

More information is provided below regarding the rainfall threshold for landslide initiation.

As per Formula 1, the units for the frequency of landslides can be expressed quantitatively as below, provided that the unit pixel (10 m × 10 m) is obtained from the data of a landslide hazard map:

$$F_L = \text{landslide event} \times \text{pixel}^{-1} \times \text{year}^{-1} \quad (2)$$

where F_L is the frequency of landslide occurrence and the area of a unit pixel corresponds to 100 m^2.

Meanwhile, Corominas and Moya (2008) described two separate approaches to the spatial probability and temporal probability of landslide occurrence. The relative frequency is the ratio of the number of landslides recorded to the unit area, which allows multiple regional landslide events to be described. The units for the relative temporal frequency are the same as those given in Formula 2. The relative temporal frequency of landslides could be identified in this study because the landslide inventory data included multiple landslide events in the province region that were triggered by a heavy rain event.

The overall procedure for analyzing the frequency of landslide occurrence based on the concept of the failure rate in the reliability study and the differentiated frequency according to landslide hazard grades is shown in Figure 1.

Notably, when the number of landslide events is classified by the landslide hazard grade, the maximum landslide hazard grade was assigned to each landslide event with an overlay of the inventory polygon data and landslide hazard map on the GIS platform. Moreover, in a conservative approach to risk analysis, the area where landslides did not occur was measured when determining the total area corresponding to each landslide hazard grade. It should also be noted that the estimated value of the frequency varies depending on the size of the area. The frequency units for this study are only representative of Gangwon Province.

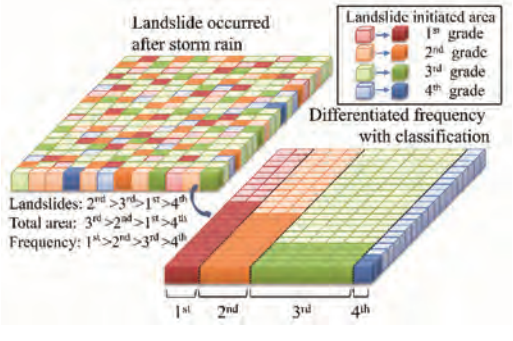


Figure 1. The frequency of landslide occurrence based on the concept of the failure rate.

2.4 *Probability of landslide occurrence*

To estimate the probability of random disastrous events caused by natural hazards over time, a Poisson distribution is often employed. Crovelli (2000) presented a Poisson distribution model to express the probability of landslide occurrence in continuous time in natural environments, as below:

$$P\{N(t) = n\} = e^{-\lambda t} \frac{(-\lambda t)^n}{n!} \quad (3)$$

where n = 1, 2, 3…; λ = the rate of occurrence of landslides; t = the specified time; and N(t) = the number of landslides that occurred during time t.

The probability of one or more landslides occurring in time t, which is referred to as the exceedance probability, is expressed as below, when λ is much less than one (λ<<1):

$$P\{N(t) \geq 1\} = 1 - e^{-\lambda t} \quad (4)$$

This model of the probability of landslide occurrence can also be presented using the definition of reliability. The definition of reliability usually contains four basic elements: probability, adequate performance, time, and operating conditions (Billinton and Allan, 1992), and one of the definitions in general terms can be introduced as follows: the probability that an item will perform a required function without failure under stated conditions for a stated period of time (O'Connor and Kleyner, 2012). The reliability function R(t) in mathematical terms is expressed as follows (Kapur and Pecht, 2014):

$$R(t) = \frac{Ns(t)}{N} \quad (5)$$

where Ns = the number of surviving items; and N = the number of total items.

Unreliability, F(t) is given as:

$$F(t) = 1 - R(t) = \frac{N - Ns(t)}{N} \quad (6)$$

and,

$$f(t) = \frac{dF(t)}{dt} = -\frac{1}{N}\frac{dNs(t)}{dt} \quad (7)$$

When the hazard rate h(t) is normalized with its surviving items Ns(t) instead of the total number

of items N from the unreliability rate f(t) equation, we obtain the hazard rate h(t), as shown below, with a more conservative meaning:

$$h(t) = \frac{f(t)}{R(t)} \tag{8}$$

The integral of the hazard rate h(t) over the time from 0 to t is:

$$\int_0^t h(\tau)\, d\tau = -\ln R(t) \tag{9}$$

Then, R(t) is:

$$R(t) = e^{-H(t)} \tag{10}$$

where H(t) is the number of hazards in time t and can be expressed as $H(t) = \lambda t$.

Finally, we obtain F(t) as:

$$F(t) = 1 - R(t) = 1 - e^{-H(t)} = 1 - e^{-\lambda t} \tag{11}$$

By using Formula 11 above, the probability of landslide occurrence can be calculated using the frequency of landslide occurrence estimated from the concept of the failure rate, as expressed in Formula 2.

2.5 *Time period estimation by Rainfall threshold*

In order to estimate the time period between landslide occurrences, a rainfall threshold was established, since the mechanism of landslide occurrence is triggered by the increase in pore water pressure and rain water seepage forces (Cullen et al., 2016). Since Caine's research (1980) examined the relationship between the minimum rainfall duration and intensity required to cause a landslide, a number of methodologies to identify the rainfall threshold have been examined, with the goal of finding the most suitable correlation with landslide initiation (Aleotti, 2004).

However, the examined proposals are valid only with the local geo-spatial properties (Martelloni et al., 2012, Jakob et al., 2006). Thus, domestic research results were applied to reflect the features of local geology, vegetation, and topography. Kim and Chae (2009) reported that landslides tend to occur in South Korea when the consecutive rainfall is over 200 mm for 48 hours. Based on their results, cumulative precipitation of more than 200 mm for 48 hours was adopted as a criterion for the rainfall threshold.

To determine the daily rainfall intensity, which was another factor used to determine the threshold, daily precipitation records were reviewed from when Typhoon Ewiniar caused heavy rainfall in July 2006. This decision was based on the assumption that landslides are likely to occur in the future in similar environmental conditions. Despite the lack of continuous landslide inventory data, this method provides a basis for estimating the frequency of landslide occurrence. Rainfall data from the Automatic Weather Station (AWS) located in the center of Gangwon Province were adopted as a representative sample to estimate landslide frequency, considering the geographic location and the high severity of damage caused by the typhoon.

The daily rainfall records in the region for the last 10 years are plotted in Figure 2. The graph shows that daily precipitation exceeded 150 mm on both July 15 and July 16, 2006, and it is possible to assume that landslides occurred after the rainstorms on those dates.

Therefore, a daily rainfall of more than 150 mm was set as another factor contributing to the rainfall threshold. As a result, the rainfall threshold was established as including both 48-hour cumulative precipitation over 200 mm and daily precipitation over 150 mm.

By screening using the rainfall threshold, as shown in Figure 3, the average landslide occurrence interval was estimated by counting the dates

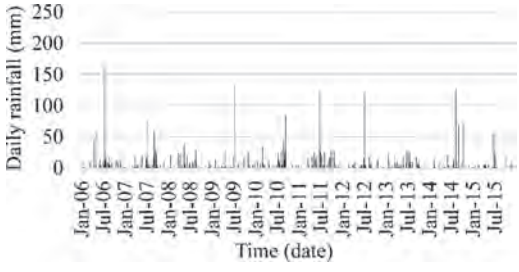


Figure 2. Sampled daily rainfall in Gangwon Province for 10 years (2006–2015).

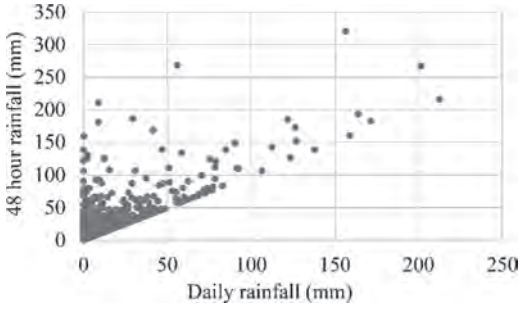


Figure 3. Scatter diagram of daily rainfall and 48-hour cumulative rainfall showing that 3 events exceeded the rainfall threshold in 10 years (2006–2015).

that satisfied these criteria, which resulted in 3 events during the reviewed 10-year period. Thus, the probabilistic period of landslide occurrence was estimated as 3.3 years for the analysis of the probability of landslide occurrence in this study.

3 RESULT

3.1 *Estimation of landslide frequency*

The locations in the inventory data where landslides have occurred in Gangwon Province are presented in Figure 4.

An analysis of the landslide hazard map of Gangwon Province shows that grades 2 and 3 predominated throughout the study region, while grade 1 areas were sparsely scattered near mountainous areas. The resulting estimation of the frequency of landslide occurrence is summarized in Table 1. The total areas of the each landslide hazard grade are shown, except for grade 5, which had null data, and it is shown that grade 3 occupied the largest area of 5815.2 km^2, followed by grade 2 (4611.6 km^2), grade 4 (250.0 km^2), and grade 1 (186.5 km^2).

When the number of landslide events was counted and classified along with the landslide hazard grades, a total of 72 landslide events were found in grade 1 areas, followed by 700 landslides in grade 2 areas, 433 in grade 3 areas, and 8 landslides in grade 4 areas.

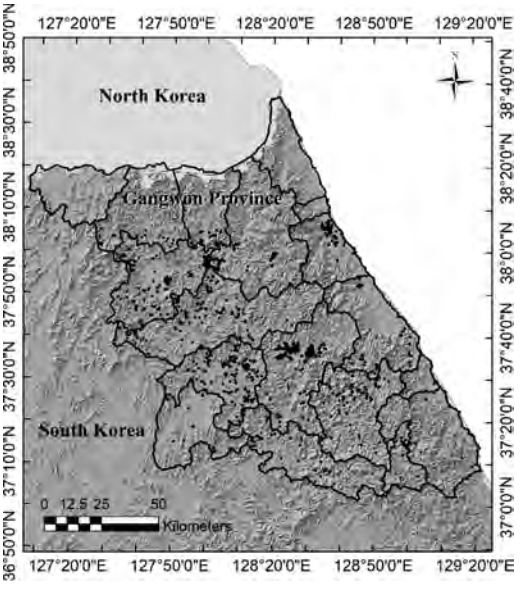


Figure 4. Locations of landslide occurrence in the study area.

Table 1. The analyzed frequency of landslide occurrence.

Landslide hazard grade	No. of Landslides	Total area * (km^2)	Landslide occurrence frequency (λ)**
1	72	186.5	1.17E-05
2	700	4611.6	4.60E-06
3	433	5815.2	2.27E-06
4	8	250.0	9.70E-07

*The unit pixel (10 m × 10 m) is used for frequency estimation.
**The unit of landslide occurrence rate (λ) is landslide event × pixel −1 × year −1.

The most landslides occurred in grade 2 areas because they accounted for the most area. However, if we examine the occurrence ratio, which is defined as the number of landslide events divided by the total area, it can be seen that the highest number of landslide events per area occurred in grade 1 areas.

Thus, our results indicate that areas with a landslide hazard of grade 1 had the highest value of landslide occurrence frequency (1.17E-05 landslide events × pixel^{-1} × year^{-1}). The frequency decreased from grade 2 to grade 4, which had the lowest value of landslide occurrence frequency (9.70E-07 landslide events × pixel^{-1} × year^{-1}). It should, however, be kept in mind that the estimated landslide occurrence frequencies are only valid for Gangwon Province area.

3.2 *The probability of landslide occurrence*

Given the estimated frequency of landslide occurrence along with the landslide hazard grades, the probability of landslide occurrence was calculated following Formula 11 and plotted with a logarithmic scale on the Y-axis. The graph in Figure 5 shows an increase in the overall probability of landslide occurrence over time, as well as presenting discrete curves according to the landslide hazard grade. The resulting graph indicates that grade 1 areas had the highest value of probability of landslide occurrence, followed by grade 2 areas, grade 3 areas, and grade 4 areas, sequentially.

Our results indicate that locations with different grades of landslide hazard are exposed to different risk levels, which can be analyzed by calculating the probability of landslide occurrence. The probability of a landslide, which is a disaster caused by a natural hazard, can be estimated based on the concept of reliability. The calculated probability value can be used as a basis for landslide risk management.

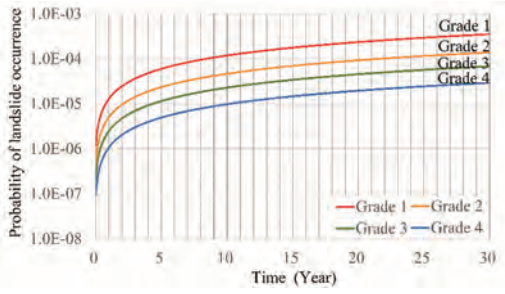


Figure 5. Increases in the probability of landslide occurrence depending on the landslide hazard grade.

4 DISCUSSION

In this study, it was shown the frequency of landslide occurrence can be estimated and the probability of landslide occurrence can be calculated using the concept of reliability analysis, which is commonly used in mechanical studies and other engineering fields. Since the occurrence of disasters due to natural hazards is affected by numerous environmental variables, estimating their frequency and calculating the probability of their occurrence are usually difficult tasks. However, when we consider a pixel unit of spatial information in a GIS platform as the component for reliability analysis, it becomes possible to calculate the frequency of occurrence with correct units and to calculate the probability of occurrence.

A set of cross-sectional inventory data obtained after the destructive damage caused by a typhoon was used due to the lack of landslide inventory data in this study. Therefore, continuous data collection in accordance with appropriate landslide inventory frameworks should be implemented, and updating landslide event observation data is important to maintain valid data regarding the frequency of occurrence.

To cope with natural disasters, it is necessary to establish a strategy for adaptation after an analysis of the risk originating from natural hazards. The method presented in this paper for estimating the frequency and the resulting probability value can be used as the basis for landslide risk analysis and management. After baseline risk analysis is conducted, various safety barriers for risk reduction can be applied to manage the risk quantitatively within acceptable criteria. Monitoring of potentially unstable slopes is a potential safety barrier serving as a preventive measure to reduce the likelihood of landslide occurrence (Uhlemann et al., 2016). Adopting an early warning system is a safety barrier to prevent unwanted loss by timely detection (Sättele et al., 2015). Major engineering work to reinforce an unstable slope with a retaining wall or to install a fence or net can be a mitigation measure to reduce the severity of the consequences of a landslide (Dai et al., 2002). Safety barriers are associated with specific magnitudes of risk reduction, and the appropriate level of risk reduction through the use of safety barriers can be decided by quantitatively considering the exposed risk, which is estimated from the probability of landslide occurrence.

Analysis of the risk originating from natural hazards through the concept of reliability can be considered a convincing approach to manage risk. Landslide risk management based on reliability analysis can be also actively applied with local frequency data that are suitable to other regions in order to prevent unwanted loss of life, property, or environment.

5 CONCLUSION

In this study, the frequency of landslide occurrence was estimated by reliability analysis and a sample calculation of the probability of landslide occurrence was presented for Gangwon Province in South Korea, from which landslide inventory data and a landslide hazard map were available. For the analysis, a pixel unit of GIS-based spatial information was considered as a component to denote the unit for the frequency of landside occurrence.

It was found that more landslide events were initiated in areas with a hazard grade of 2 or 3 than in grade 1 areas because of their greater total areas; however, the most landslide events per area occurred in grade 1 areas. It was shown that areas with a greater landslide hazard had higher values of the landslide occurrence frequency in the study region, and the estimated frequency data were used to calculate the probability of landslide occurrence in an analysis based on the concept of reliability.

By examining natural hazards through reliability analysis, we have ascertained that the estimated frequency and the resulting probability value can be used as the basis of landslide risk analysis and management. This technique can also be extended to assess various mitigation measures to handle the risks that stem from natural hazards.

REFERENCES

Australian Geomechanics Society 2000. Landslide risk management concepts and guidelines. *Australian Geomechanics Journal,* 37, 1–44.

Aleotti, P. 2004. A warning system for rainfall-induced shallow failures. *Engineering Geology,* 73, 247–265.

Billinton, R. & Allan, R.N. 1992. *Reliability Evaluation of Engineering Systems: Concepts and Techniques.*

Caine, N. 1980. The rainfall intensity-duration control of shallow landslides and debris flows. *Geografiska Annaler Series A,* 62, 23–27.

Corominas, J. & Moya, J. 2008. A review of assessing landslide frequency for hazard zoning purposes. *Engineering Geology,* 102, 193–213.

Corominas, J., Van Westen, C., Frattini, P., Cascini, L., Malet, J.P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M.G., Pastor, M., Ferlisi, S., Tofani, V., Hervãs, J. & Smith, J.T. 2014. Recommendations for the quantitative analysis of landslide risk. *Bulletin of Engineering Geology and the Environment,* 73, 209–263.

Crovelli, R.A. 2000. Probability models for estimation of number and costs of landslides. *United States Geological Survey open file report 00 249*.

Cullen, C.A., Al-Suhili, R. & Khanbilvardi, R. 2016. Guidance Index for Shallow Landslide Hazard Analysis. *Remote Sensing,* 8, 17.

Dai, F.C., Lee, C.F. & Ngai, Y.Y. 2002. Landslide risk assessment and management: An overview. *Engineering Geology,* 64, 65–87.

Frattini, P., Crosta, G. & Sosio, R. 2009. Approaches for defining thresholds and return periods for rainfall-triggered shallow landslides. *Hydrological Processes,* 23, 1444–1460.

Goble, W., Cheddie H. 2005. *Safety Instrumented Systems Verification: Practical Probabilistic Calculations*. U.S.A.: The Instrumentation, Systems and Automation Society

Guzzetti, F. 2000. Landslide fatalities and the evaluation of landslide risk in Italy. *Engineering Geology,* 58, 89–107.

Ipcc, S., T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex And P.M. Midgley (ED.) 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Jaiswal, P., Van Westen, C.J. & Jetten, V. 2010. Quantitative landslide hazard assessment along a transportation corridor in southern India. *Engineering Geology,* 116, 236–250.

Jakob, M., Holm, K., Lange, O. & Schwab, J.W. 2006. Hydrometeorological thresholds for landslide initiation and forest operation shutdowns on the north coast of British Columbia. *Landslides,* 3, 228–238.

Kapur, K.C. & Pecht, M. 2014. Reliability Engineering.

Kim, W.-Y. & Chae, B.-G. 2009. Characteristics of Rainfall, Geology and failure Geometry of the Landslide Areas on Natural Terrains in Korea. *The Journal of Engineering Geology* Vol 19, 331–344.

Korea Forest Service 2012. Landslide Hazard Map, available at: http://www.forest.go.kr/newkfsweb/html/HtmlPage.do?pg=/fgis/UI_KFS_5002_020600.html&mn=KFS_02_04_03_04_06&orgId=fgis (in Korean)

Lee, J., Lee, D.K., Kil, S.H. & Kim, H.G. (2017) Risk-based analysis of monitoring time intervals for landslide prevention. Nat. Hazards Earth Syst. Sci. Discuss., 2017, 1–22.

Martelloni, G., Segoni, S., Fanti, R. & Catani, F. 2012. Rainfall thresholds for the forecasting of landslide occurrence at regional scale. *Landslides,* 9, 485–495.

National Emergency Management Agency 2007. National Emergency Management Agency Key Statistics and Data in 2007.

O'connor, P.D.T. & Kleyner, A. 2012. Practical Reliability Engineering.

Polemio, M. & Sdao, F. 1999. The role of rainfall in the landslide hazard: The case of the Avigliano urban area (Southern Apennines, Italy). *Engineering Geology,* 53, 297–309.

Sættele, M., Brþndl, M. & Straub, D. 2015. Reliability and effectiveness of early warning systems for natural hazards: Concept and application to debris flow warning. *Reliability Engineering & System Safety,* 142, 192–202.

Uhlemann, S., Smith, A., Chambers, J., Dixon, N., Dijkstra, T., Haslam, E., Meldrum, P., Merritt, A., Gunn, D. & Mackay, J. 2016. Assessment of ground-based monitoring techniques applied to landslide investigations. *Geomorphology,* 253, 438–451.

Van Westen, C.J., Van Asch, T.W.J. & Soeters, R. 2006. Landslide hazard and risk zonation—Why is it still so difficult? *Bulletin of Engineering Geology and the Environment,* 65, 167–184.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Probabilistic seismic hazard assessment for offshore structures in Andaman Sea

T. Ornthammarath
Department of Civil and Environmental Engineering, Faculty of Engineering, Mahidol University, Thailand

ABSTRACT: A set of probabilistic seismic hazard maps for Offshore structures in Andaman sea has been derived using procedures developed for the latest US National Seismic Hazard Maps. In contrast to earlier hazard maps for this region, which are mostly computed using delineated seismic source zone, the presented maps are based on the combination of smoothed gridded seismicity, crustal-fault, and subduction source models. The ground motion hazard map is presented over a 10 km grid in terms of peak ground acceleration and spectral acceleration at 1.0 undamped natural periods and a 5% critical damping ratio for 10 and 2% probabilities of exceedance in 50 years, which have generally been used for Seismic Analysis and Design of Offshore Structures.

1 INTRODUCTION

The Andaman Sea is situated in an active back-arc basin lying above and behind the Sunda subduction zone, which is the Indo-Australian and Eurasian boundary zone comprise the convergent margins, including the Burma oblique subduction zone, Andaman thrust and Sunda arc, to the North West, west and south, respectively. Within Andaman Sea, several earthquakes with magnitude greater than 4 have been observed during the years 1964 to 2017. However, the earthquake activity rate is much lower than those occurred near plate boundary. For this study, the southern part of Andaman Sea (Blue dash line and the study area is bounded latitude 5° to 10° and longitude 94° to 98°, Figure 1) is of special interested due to ongoing human activities for gas pipeline and offshore facilities, several platforms and subsea gas pipelines have been developed offshore. In addition, onshore supports such as control and maintenance centre along with gas metering stations have been developed in this region. The largest instrumental earthquake within this zone is Mw 6.3 on 16 May 1933 at 10 km depth. In addition, this region is situated about 200 and 300 km from the Sumatran faults and Sumatra subduction zone, respectively. And these seismic active structures have historically produced moderate to large earthquakes with long-period ground motions that were felt in high-rise buildings in Singapore and Malaysia (Pan and Sun, 1996; Pan, 1997; Pan et al., 2001). A probabilistic seismic hazard assessment for offshore structures located in this area has been carried out.

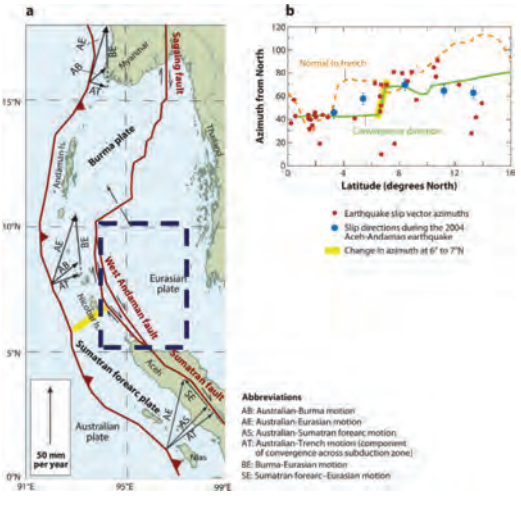


Figure 1. (a) Map showing possible relative motions of the Sumatran and Burma forearc plates relative to adjacent Indian-Australian and Eurasian (Sunda) plates. Adapted with permission from Curray (2005). The vector diagrams show the partitioning of the total convergence of Australia relative to Eurasia (vector AE) into components of subduction (vectors AS and AB) and strike-slip (vectors SE and BE). (b) Plot of earthquake slip vector azimuths (red dots). Blue dash line showing the area of current study.

The main objective of this work is to determine appropriate earthquake ground motion parameters for the seismic design of offshore structures based on available scientific information. These

ground motion parameters include: horizontal Peak Ground Acceleration (PGA) and Spectral Acceleration (SA) values at the project site with the return periods of 475 and 2475 years.

2 SEISMOTECTONIC SETTINGS

The Indo-Australian and Eurasian boundary zone comprise the convergent margins, including the Burma oblique subduction zone, Andaman thrust and Sunda arc, to the North West, west and south, respectively. The plate kinematics of the Sumatran region is, in a broad sense, the simple interaction of the Indian-Australian and Eurasian plates (Figure 1). However, in detail it is much more complex than that. Deformation rates across these plate boundaries are variable. The observed seismicity and seismotectonic settings of these plate boundaries clearly indicate the capability of producing large events, where the 26 December 2004 earthquake occurred. A convergence rate of 65–70 mm/year as a result of Australia moving toward South East Asia is reported by McCaffrey (1996).

Deformation of the overriding plate leads to larger complexities in plate motions. Sumatra sits at the southwestern edge of the Sunda plate (Bird 2003), which moves at a few millimeters per year to a centimeter per year eastward relative to Eurasia (Chamot-Rooke & Le Pichon 1999, Bock et al. 2003) (Figure 1). The resulting convergence between the Sunda plate and the oceanic plates to the southwest is somewhat slower than it would be relative to Eurasia. The rate and direction of subduction of the lithosphere under the Sunda forearc, however, are further modified by the independent motion of the forearc. Fitch (1972) explained the presence of the Sumatran fault and other similar faults inboard subduction zones by the process now known as slip partitioning. That is, in some cases of oblique subduction where the two plates do not converge at a right angle to the strike of the trench, it requires smaller overall shear force to share the shearing (trench-parallel) component of the relative motion between two separate faults instead of on one fault. In the case of partitioning, one fault is the subduction thrust, which takes up all of the trench-normal slip (the dip-slip component) and some fraction of the trench-parallel slip (the strike-slip component). A second fault, within the overriding plate and commonly strike-slip in nature, takes up a portion of the trench-parallel motion. The subduction thrust and strike-slip fault isolate a wedge of forearc called the sliver plate. The slip rates on the separate faults can be inferred from their geometries and knowledge of the overall convergence.

The motion of the Sunda forearc (sliver plate) is not known well, particularly in the Andaman section, and hence the subduction vector is highly uncertain. Earlier estimates of the relative motions assuming a rigid forearc sliver plate failed to predict convergence in the Andaman section of the trench, which probably indicates, as is now accepted, that there is extensive internal deformation within the forearc. One possibility, evident in the change in earthquake slip directions along the margin, is that the Andaman section (called the Burma plate) and Sumatran sections of the forearc move independently with a break near 6° to 7°N (Subarya et al. 2006).

3 EARTHQUAKE CATALOGUE

The earthquake catalogue for current study is composed of instrumental earthquake records from different international earthquake observatories including:

1. International Seismological Centre-Global Earthquake Model (ISC-GEM) (1907–2009),
2. Engdahl (EHB)'s earthquake catalog (1960–2007),
3. USGS/NEIC preliminary earthquake database (1900–2013), and
4. Global centroid moment tensor (GCMT) (1976 – December 2010),

The compiled instrumental earthquake catalog is covered from 1900 to 2013 (Figure 1). All reported event magnitudes have been converted to moment magnitude by using Scordilis (2006) relationship for mb-Mw and Ms-Mw. Subsequently, all duplicated events have been removed, and earthquake magnitude and location correction have been performed manually to remove any obvious errors. The largest earthquake magnitude 9.2, Great Sumatra-Andaman earthquake in northern Sumatra on 26 December 2004 at 0 7.58 local time.

3.1 *Magnitude conversion*

In the final updated earthquake catalogue, several different magnitude scales are used to define the earthquake magnitude. For example, the 20-s surface-wave magnitude (M_S) and the short-period P-wave magnitude (mb) are commonly used in the data from USGS, ISC, and other international database sources, and the moment magnitude (M_W) is reported in the Global Centroid Moment Tensor catalogue. It is necessary to convert all these different magnitude scales into a single magnitude scale. In this study, the moment magnitude scale is chosen as the single representative scale. Since the accuracy of reported magnitudes is dependent on magnitude definitions, the more reliable magnitude

is then preferred for using in magnitude conversion as follows: M_w, M_S,mb, and M_L Conversions between magnitude scales are made using the equation provided in Table 2 in Ornthammarath et al. (2011). After the magnitude conversion, we merged duplicate entries (from different data sources) into a single entry for each earthquake event.

3.2 *Declustering*

One basic assumption of the adopted seismic hazard assessment methodology is that earthquake occurrences are statistically independent (the Poisson assumption). Therefore, the earthquake catalogue to be used for seismic hazard assessment must be free of dependent events, such as foreshocks and aftershocks. The process to eliminate dependent events from earthquake catalogues is called "declustering". Gardner and Knopoff(1974) declustering algorithm, is chosen for the present study. This approach states that foreshocks and aftershocks are dependent (a non-Poissonian process) on the size of the main event, and these earthquake events need to be removed in accordance with space- and time- windows. Normally, a large main earthquake event leads to larger aftershocks over a larger area and for a longer time. Therefore the time- and distance-window parameters for larger main events are greater than those for smaller events. Declustering eliminates about 60% of the 64,866 events in the catalogue. The final declustered catalogue includes 25,654 earthquake events (4218, 7585, and 13851 events for shallow, intermediate, and deep earthquake, respectively) with M_W greater than or equal to 3.0 in the study region from 1900 to 2013.

3.3 *Catalogue completeness*

It is recognized that earthquake data in the catalogue are not complete, and that failure to correct for the data incompleteness may lead to underestimation of the mean rates of earthquake occurrence. The correction can be made by identifying the time period of complete data for prescribed earthquake magnitude ranges. Reliable mean rates of earthquake occurrence for the given magnitude ranges can then be computed from the complete data.

Two methods were employed for completeness analysis of the catalogue: (a) the Visual Cumulative method (CUVI) and (b) Stepp's method. Both algorithms provided a similar result; hence, the former technique was adopted. We divide the study region into three zones, i.e., shallow, intermediate, and deep earthquakes (BG-I, BG-II, BG-III, respectively). The data completeness analysis is carried out separately for each of these zones, and the results are presented in Table 2 in Ornthammarath et al. (2011).

4 MODELING OF EARTHQUAKE SOURCES

To properly describe the complex earthquake environments in the region, they are modeled as a mixture of background seismicity, subduction area sources, and crustal faults. These are described in more detail below.

4.1 *Background seismicity model*

The background seismicity model represents random earthquakes in the whole study region except the subduction zones. The model accounts for all earthquakes in areas with no mapped seismic faults and for smaller earthquakes in areas with mapped faults. In this approach, it is not necessary to divide the region into many small areas. One large area may be used, but the rate of seismicity is assumed (or allowed) to vary from place-to-place within the area. The rate of seismicity is determined by first overlaying a grid with a given spacing, in the current case 0.10° in latitude and 0.10° in longitude, approximately 10 by 10 km, onto the study region, and counting the number of earthquakes with magnitude greater than a reference value (M_{ref}) in each grid cell. The rate of seismicity is computed by dividing the number of earthquakes by the time period of earthquake data. The rate is then smoothed spatially by a Gaussian-function moving average and comparing with the observed seismicity. By this approach, the spatially-varied seismicity can be modeled with confidence relating to source uncertainty.

In hazard calculations, earthquakes smaller than magnitude 6.0 are characterized as point sources at the centre of each grid cell, whereas earthquakes larger than magnitude 6.0 are assumed to be hypothetical finite vertical or dipping faults centered on the source grid cell. Lengths of finite faults are determined using the Well and Coppersmith (1994) relations. Consecutively, the pre-calculated average source-to-site distance from virtual faults with strike directions uniformly distributed is employed (Petersen et al. 2008).

The whole study region is divided into 6 source zones: BG-I, BG-II, BG-III, SD-A, SD-B, SD-C, (see Fig. 3). The zones SD-A, SD-B, and SD-C are subduction zones, which will be described in detail below. The zone BG-I is a background seismicity zone covering the whole study area excepting subduction zone, and the zone BG-II and BG-III are background seismicity zone for intermediate and deep earthquake except subduction

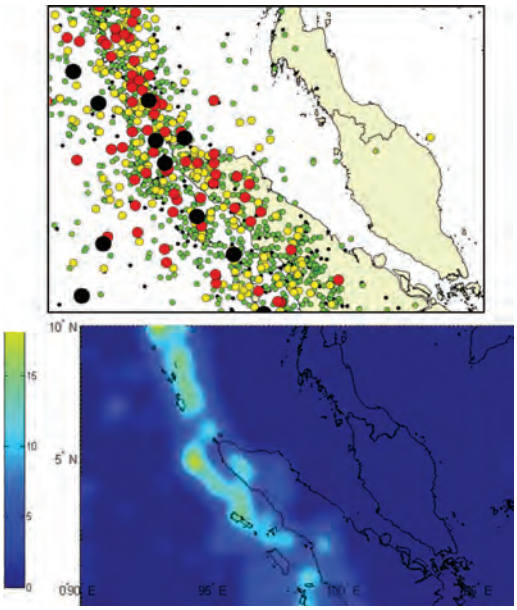


Figure 2. Observed shallow seismicity Black circle represent magnitude greater than 7.0, Red circle represent magnitude between 6.0–7.0; Yellow circle represent magnitude between 5.0–6.0; Green circle represent magnitude between 4.0–5.0 (Upper) and the smoothed activity rate 10^a value of BG-I for shallow earthquake (Lower).

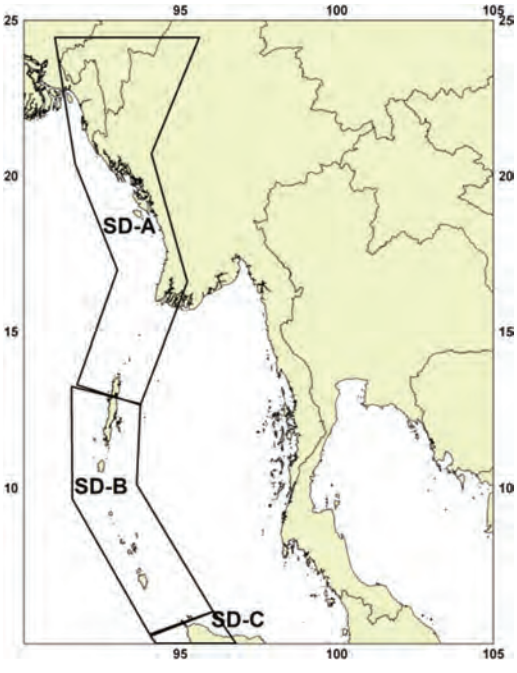


Figure 3. Subduction zones SD-A, SD-B, and SD-C considered in this study.

zones. Earthquake data, particularly greater than magnitude 3.0 earthquakes, in BG-I, BG-II, and BG-III could be reliably recorded since 2004 due to the high earthquake detection capability of a fairly dense seismograph network in Malaysia and surrounding region. Hence, the accuracy of the estimated seismicity rate in BG-I can be significantly improved by including small earthquakes ($3.0<M_W<5.0$) in the seismicity rate calculation

Furthermore, this activity rate computation is also based on the observation that moderate earthquakes generally occur in areas where there have been a significant number of magnitudes 3 events. On the other hand, in BG-II and BG-III earthquake data with $M_W > 5.0$ can be used for computing the seismicity rate due to the incompleteness of small earthquake data. Nevertheless, a lack of small earthquake data is not a major problem because the seismicity rate in these zones are relatively high; thus, the rate can be reliably estimated from moderate-sized earthquakes. In addition, the overall influence of BG-II and BG-III on the seismic hazard in Andaman sea is lower than that of BG-I. We model the magnitude-dependent characteristic of the seismicity rate in each background seismicity zone by a truncated exponential model (Gutenberg-Richter model):

$$\text{Log10}\,(N(M_W)) = a - bM_W \qquad (1)$$

where $N(M_W)$ is the annual occurrence rate of earthquakes with magnitude greater than or equal to M_W, and a and b are the Gutenberg-Richter model parameters. The b-value is assumed to be uniform throughout the whole background region. Hence, we used complete earthquake data with magnitude greater than 4.0 in current study area to compute a single regional b-value. The obtained regional b-value is 0.90, and this value is used for BG-I BG-II and BG-III. The a-value varies from place to place within each zone. It is computed by using a grid with spacing of 0.10° in latitude and longitude and is spatially smoothed using a two-dimensional Gaussian moving average operator with a correlation distance parameter C (Frankel 1995). Earthquake data with $M_W > 3.0$ and C = 50 km are used for BG-I, while earthquake data with $M_W > 5.0$ and C = 50 km are used for BG-II and BG-III. The correlation distance is chosen based on Frankel (1995) and it is comparable to earthquake location error. Note that at present there are no fixed rules or guidelines to determine an appropriate C value. If the value of C is too small, the resulting smoothed seismicity will be concentrated around the epicenters of past recorded earthquakes. On the other hand, if the value of C is too large, the resulting smooth seismicity will be blurred and will not reflect the true spatial variation pattern of seismicity. The chosen C values are believed to suitable as the computed smoothed rate 10a values

(presented in Fig. 2) are in agreement to observed spatial pattern of seismicity.

In the truncated Gutenberg-Richter models of BG-I, BG-II, and BG-III, the minimum earthquake magnitude is set equal to 4.5 because earthquakes with smaller magnitude than this are judged not to cause damage to buildings and structures. The maximum (upper bound) magnitude is set to 7.0 for BG-I and 8.0 for BG-II and BG-III to account for the largest earthquake magnitude that have been observed in these zones, as shown in Fig. 2. The average depth used in the model for BG-I, BG-II, and BG-III are 20, 75, and 120 km, respectively.

4.2 *Subduction zone model*

As explained earlier, the megathrust Sunda subduction zone is divided into seven sub-zones based on seismicity characteristics: the Burma zone (SD-A), the Northern Sumatra-Andaman zone (SD-B), and the Southern Sumatra zone (SD-C). Each sub-zone is modelled as a seismic area source with a uniform rate of seismicity (the traditional area source model), and the magnitude-dependent seismicity rate is modeled by a truncated Gutenberg-Richter relation. The geometry and recurrence times of large earthquakes associated with these active tectonic structures are largely based on available paleotsunami and seismic history studies (Jankaew et al., 2011), summarized geodetic data reported in, Berryman et al., (2013).

The calculated Gutenberg-Richter model parameters (a and b values) are shown in Table 3 in Ornthammarath et al. (2011). The minimum earthquake magnitude in the Gutenberg-Richter model is set to 6.5 as the subduction zones are very near to studied area and hence large earthquakes are also important for long period structures in southern Andaman Sea. The maximum magnitude for each zone is set to the maximum observed magnitude plus 0.5 magnitude units. The maximum magnitude for zone SD-B and SD-C is set to 9.2 as the 2004 Sumatra earthquake and the 2005 Nias earthquake

4.3 *Crustal fault source model*

Two different approaches are employed to model the magnitude-dependent characteristic of the seismicity rate of these crustal faults: the Gutenberg-Richter model and Characteristic Earthquake (CE) model. In the Gutenberg-Richter model, a magnitude-frequency distribution for crustal fault model is assumed from the minimum magnitude of 6.5 to the upper-bound magnitude (Mmax). To account for the uncertainty in estimating Mmax, we consider three different cases with Mmax set equal to MC – 0.2, MC, and MC + 0.2. The probabilistic weights of 0.2, 0.6, and 0.2 are assigned to these cases, respectively. In each case, theb-value is set equal to the regional b-value of 0.90, and the a-value is determined from the seismic moment rate, which is computed from the fault slip rate.

In the characteristic earthquake model, three characteristic earthquake magnitudes (M_C) are also considered: M_C – 0.2, M_C, and M_C + 0.2. The probabilistic weights of 0.2, 0.6, and 0.2 are assigned to these cases, respectively. In each case, the earthquake occurrence rate is computed from the characteristic magnitude and the fault slip rate (to match with the seismic moment rate of the fault). The recurrence interval for the characteristic model is determined from:

$$\text{Recurrence Interval} = \mu u LW/M_{0C} \quad (2)$$

where is μ shear modulus, 3.0×10^{11} dyne/cm^2, L is rupture length, and W is rupture width, u is the fault slip rate, M_0C is the characteristic earthquake moment, which is calculated from:

$$\log(M_{0C}) = 1.5M_C + 16.05 \quad (3)$$

and the magnitude is assumed to be normally distributed around the characteristic value with a standard deviation of 0.12. The properties and parameters of Sumatra faults considered in this study are shown in Table 2 in Petersen et al. (2007).

5 GROUND MOTION PREDICTION EQUATIONS (GMPES)

In this study, three Next Generation Attenuation (NGA) models were developed for shallow crustal earthquakes in the Western United States and similar active tectonic regions were applied for background earthquakes in BG-I and for earthquakes from crustal faults in the study region. These NGA models were developed by Boore and Atkinson (2008), Campbell and Bozorgnia (2008), and Chiou and Youngs (2008) during the NGA project.

In addition, based on comparison of several different Ground Motion Prediction Equations (GMPEs) with recorded ground motion from interface subduction earthquakes by Ornthammarath et al. (2014), it has been decided that three subduction GMPEs could be used for probabilistic seismic hazard analysis in this region, and probabilistic weights assigned to these models are 0.25, 0.25, and 0.50 for Atkinson and Boore (2003; 2008), Youngs et al. (1997), and Zhao et al. (2006), respectively. These weights are relatively consistent with residual of observed recorded data and estimated values of three selected GMPEs. To calculate ground motion for intermediate-and deep-depth earthquakes, the equations of Young et al. (1997) and Atkinson and Boore (2003) are considered with equal weights.

6 PROBABILISTIC SEISMIC HAZARD ANALYSIS

The PSHA results for southern Andaman Sea, Figure 4 and 5, are presented in terms of seismic hazard maps at 475- and 2475-year return periods at bedrock condition. For southern Andaman Sea, the observed seismicity in and around Sumatra subduction zone and Sumatra faults control the hazard for most considered structural periods. Estimated bedrock PGA near subduction zone at 2475-year return period range between 0.6g to 1.0g; however, for area near coast of Thailand and Myanmar, estimated bedrock PGA at 475-year return period are relatively less intense varied from 0.05g to 0.15g. This is primarily due to its location which is far removed from any major active structure in augmented with low observed seismicity rate of background seismicity.

For long structural period (Figure 5), large part of southern Andaman Sea is subjected by moderated long period ground motion due to lower attenuation rate of long periods. Subduction zone earthquakes contribute high seismic hazard to long period offshore structure in southern Andaman Sea. Estimated bedrock SA (T = 1s) near subduction zone at 2475-year return period range between 0.4g to 1.0g; however, for area near coast of Thailand and Myanmar, estimated bedrock PGA at 2475-year return period are relatively less intense varied from 0.10g to 0.20g. The suitable ground motion records for performing time history analysis of long period structures in southern Andaman Sea should be selected for large earthquake at long distance. In addition, the computed ground motions are comparable to those in Petersen et al. (2007) with minor different in short period ground motion near subduction zones.

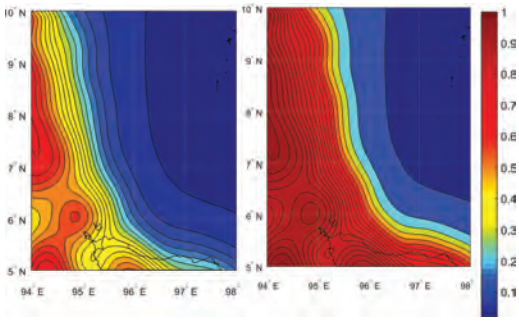


Figure 4. PGA (g) map for southern Andaman Sea at 475 - (Left) and 2475 - (Right) year return period.

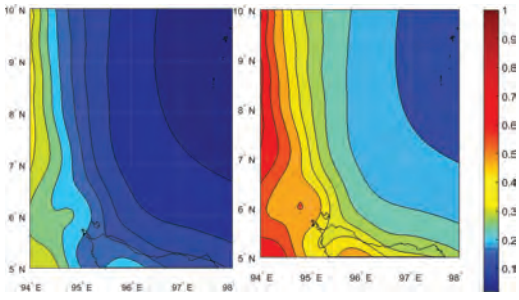


Figure 5. SA (T = 1s) (g) map for southern Andaman Sea at 475 - (Left) and 2475 - (Right) year return period.

REFERENCES

Atkinson and Boore. 2003 Empirical ground-motion relations for Subduction-zone earthquakes and their application to Cascadia and other regions. *Bull Seism Soc Am* 93(4):1703–1729.

Atkinson and Boore. 2008 Erratum to empirical ground-motion relations for subduction zone earthquakes and their application to Cascadia and other regions. *Bull Seism Soc Am* 98(5):2567–2569.

Berryman et al.2013, The GEM Faulted Earth Subduction Characterisation Project, pp. 43.

Boore DM, Atkinson GM. 2008 Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01s and 10.0s. *Earthq Spectra* 24(1):99–138.

Campbell KW, Bozorgnia Y. 2008 Ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10.0s. *Earthq Spectra* 24(1):139–171.

Chiou B, Youngs R. 2008 A NGA model for the average horizontal component of peak ground motion and response spectra. *Earthq Spectra* 24(1):173–215.

Jankaew, Kruawun, Maria E. Martin, Yuki Sawai and Amy L. Prendergast. 2011. Sand Sheets on a Beach-Ridge Plain in Thailand: Identification and Dating of Tsunami Deposits in a Far-Field Tropical Setting, The Tsunami Threat - Research and Technology, Nils-Axel MÃ¶rner (Ed.), ISBN: 978-953-307-552-5, InTech, DOI: 10.5772/14010.

Ornthammarath T, Warnitchai P, Worakanchana K, Zaman S, Sigbjörnsson R, Lai CG. 2011 Probabilistic seismic hazard assessment for Thailand. *Bull Earthquake Eng* 9(2):367–394.

Petersen M, Harmsen S, Mueller C, Haller K, Dewey J, Luco N, Crone A, Lidke D, Rukstales K. 2007 Documentation for the Southeast Asia Seismic Hazard Maps, Administrative Report, U.S. Geological Survey, pp. 65.

Youngs RR, Chiou SJ, Silva WJ, Humphrey JR. 1997 Strong ground motion attenuation relationships for subduction zone earthquakes. *Seismol Res Lett* 68(1):58–73.

Zhao JX, Zhang J, Asano A, Ohno Y, Oouchi T, Takahashi T, Ogawa H, Irikura K, Thio H, Somerville P, Fukushima Y, Fukushima Y. 2006 Attenuation relations of strong ground motion in Japan using site classification based on predominant period. *Bull Seism Soc Am* 96(3):898–913.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Power outage forecasting: Methods, results, and uncertainty

S.D. Guikema
University of Michigan, Ann Arbor, Michigan, USA

ABSTRACT: Power outage forecasting for severe events such as hurricanes provides valuable information to those managing power systems and those dependent on electric power such as other utilities and individuals in society. A number of different approaches to power outage forecasting have been developed. This paper provides an overview of different approaches with a focus on how these approaches work in practice. It then focuses on the role and representation of uncertainty in power outage forecasts. What are the sources of uncertainty? How do different outage forecasting approaches handle these sources of uncertainty? How is uncertainty represented in the resulting forecasts?

1 INTRODUCTION

Sever weather is a major cause of power outages and expensive restorations associated with these outages. A critical aspect of managing weather-induced power outages for a utility is being prepared to restore power quickly and cost-effectively. Power outage forecasting models are an important part of this process, providing estimates of both total outages and the spatial distribution of these outages in advance of a storm. An increasing number of utilities are reportedly using power outage forecasting models. For example, approximately 85% of utilities responding to a recent benchmarking survey reported that they have some sort of storm power outage prediction model (Guikema et al. 2017).

A number of different approaches for predicting weather-induced power outages have been developed. The two primary approaches are fragility-based models and statistical models. Fragility-based models (e.g., Winkler et al. 2010) take as input estimates of key aspects of the weather hazard, such as gust wind speeds, and they then estimate the probability of damage at the level of individual assets via a fragility function, a mathematical function that translates the value of a hazard parameter (e.g., gust wind speed) into a damage probability. These asset level damage probabilities are then used to simulate a number of replications of sets of damaged assets. For each simulated set of damaged assets, a power flow or connectivity model is used to estimate which customers have power and which do not. This provides the overall estimate of power outages together with its spatial distribution.

A statistical power outage forecasting model (e.g., Han et al. 2009a, 2009b, Guikema et al. 2010, Nateghi et al. 2011, Guikema and Quiring 2012, Guikema et al. 2014, Nateghi et al. 2014a, 2014b, Quiring et al. 2014, McRoberts et al. 2017, He et al. 2017) uses data about the performance of power systems during past storms such as the number of meter outages in defined geographic areas together with data about the utility system, environmental conditions, and the weather conditions. These data are used to train and validate a model that can predict the impact of future weather events. The types of statistical models used vary widely, from simple linear regression models to more advanced ensemble data mining methods.

Statistical power outage forecasting models are in much wider use within electric power utilities. Based on conversations with utility personnel, many U.S. utilities have at least an Excel-based linear regression model that they use in house. It should also be noted that to date, statistical outage forecasting models have shown substantially better predictive accuracy that fragility-based models. Because statistical outage forecasting models are both more widespread and, to date, more accurate, they will be the focus of this paper.

In this paper I focus not on the technical details of statistical outage forecasting models. Rather, I take a perspective grounded in Bayesian probability and risk analysis and provide a critical assessment of the foundations of these models. To the degree possible, I use models that I have developed as the basis for the critiques. The goal here is to suggests ways in which the approaches underlying outage forecasting can advance. It is important to note that the focus here is not the meteorological side of the problem. That is, I focus on the modeling approach, not on which additional weather parameters might be helpful to include, though there is potentially substantial merit to exploring a wider array of weather variables.

This paper is structured as followed. The second section summarizes some of the practical limitations of the current approaches, drawing on the discussion in Guikema et al. (2017), laying the foundation for the following sections. The third section then focuses on discussing how these models conceptualize risk and how this is reflected in the estimates. The fourth section focuses in particularly on how uncertainty is treated in outage forecasting models. The fifth concluding section presents some suggestions for paths forward.

2 PRACTICAL LIMITATIONS

Note, this section is a summary of the discussion in Guikema et al. (2017).

The first limitation is that nearly all of the existing models in the academic literature are for a single hazard. That is, they were developed and validated for a single type of weather event, such as a hurricane (e.g., Han et al. 2009a, 2009b, Guikema et al. 2014, Nateghi et al. 2014a). The two exceptions to this that I am aware of is the model presented in Guikema et al. (2017), where an "all weather" model (wind storms, rain storms, lightening events, heat events, and to a limited extent mountain snow events) is presented and the model presented in He et al. (2017) where a model for hurricanes, thunderstorms, and blizzards is presented.

A second limitation is that the data used as input to the models can vary significantly, with some models leaving out potentially important explanatory factors. For example, our previous work suggests that variables such as the duration of strong wind, soil moisture levels, and soil type play an important role in predicting the impact of high wind events (e.g., Quiring et al. 2011, Nateghi et al. 2014a, 2014b,). For high-temperature events, the variable set becomes challenging in other ways because the duration of high temperatures and nighttime lows may be particularly important. Predictive accuracy is limited if an incomplete set of explanatory variables are used.

A third limitation is that many of the models we have seen in use within utilities vary greatly in the rigor with which they were developed and validated. It is a possible to train a statistical model that fits an outage data set extremely well, but offers poor predictions for new events. Careful validation testing, using out-of-sample holdout testing is critical for properly balancing the trade-off between bias and variance, maximizing prediction accuracy for future storms.

A fourth limitation is that most of the available models, with a few notable exceptions, provide point estimates of impacts and do not represent the uncertainty inherent in any prediction of the impacts of hazard weather events on power systems. This point will be discussed further below.

3 CONCEPTUALIZATION OF RISK AND TREATMENT OF UNCERTAINTY IN POWER OUTAGE FORECASTING MODELS

3.1 *Background*

Risk has been conceptualized and measured in many different ways. Here I draw on Aven (2012) and Guikema and Aven (2015) and focus on three conceptualizations of risk: (1) risk is an expected value, (2) risk is a function of probabilities, consequences, and outcome scenarios, and (3) risk is a function of uncertainties, consequences, and outcome scenarios. These conceptualizations can be summarizes as: (1) $R = E$, (2) $R = f(S,P,C)$, and $R = f(S,U,C)$.

The first approach to risk assumes that risk can be summarized by an expectation. This could be expected outcome, with outcome measured in financial terms or some other measure appropriate for the situation, or it could be expected utility, converting the outcomes of the different scenarios into utility. The different possible outcome scenarios are accounted for, but the risk measure is condensed down to a single measure by summing probability multiplied by the consequence measure over all possible outcome scenarios. In this sense uncertainty is included in the measure, but only to the degree the probabilities over the scenarios are included in calculating the expectation. This approach implicitly assumes a fixed value function, either through assuming risk neutrality (for an expected outcome) or a fixed risk attitude and utility functional form. This approach also assumes that uncertainty can be fully represented by probability, with a Bayesian perspective typically adopted.

The second approach is different in that rather than summarizing the probabilities and consequences by a single expected outcome measure, the full probability distribution is presented. Kaplan and Garrick (1981) is arguably the paper that originated this approach, though it was used in the earlier WASH 1400 nuclear reactor study as well. It is widely used within many areas of risk analysis. The results are often given in the form of a F-N (Frequency-Number) curve, a special case of an inverse cumulative distribution. This approach allows decision-makers to consider more than just a point estimate given that the full probability distribution is available to them. It also does not assume a fixed value function. This means that the decision maker(s) are able to obtain the probability-consequence pairs for all scenarios and

using them in their own utility function should they wish to. This approach does still assume that uncertainty can be adequately represented by probabilities. Typically, a Bayesian perspective is taken with regard to the probabilities.

The third approach is similar to the second, except in how uncertainty is treated. Probabilities are often still provided, again sometimes in the form of an F-N curve. However, additional qualitative descriptions of the information underlying those probabilities is provided. Flage and Aven (2009) provide approaches for doing this in an organized, grounded manner. This approach is fully Bayesian and focuses on describing the background state of knowledge underlying any probability assignment made within a Bayesian framework. As such, this approach provides substantially more information to decision-makers, particularly information that would allow them to judge how much confidence to place in the assessed probabilities.

3.2 *How do outage models conceptualize risk and treat uncertainty?*

Power outage forecasting models deal with an inherently spatial process. That is, they provide spatial estimates of power outages. For example, Figure 1 below shows examples of predicted and actual outages for the hurricane power outage forecasting model of Nateghi et al. [2014a] for Hurricane Ivan.

In these types of forecasts, predictions are given as spatial point estimates. For example, in the model output shown in Figure 1, a prediction consists of an estimate of the number of power outages in each 12,000 ft. by 8,000 ft. (3.66km by 2.43 km) grid cell.

How does such a model conceptualize risk? The predictions are not probabilistic in any sense. One could argue here that the predictions are intended to represent a "best estimate" with one interpretation of this being that point estimate is intended to be a represent the mean, mode, or some other measure of central tendency of the unknown, unestimated underlying distribution. This, however, is reading more into these models than is actually there. Many of the current generation of outage models (e.g., Nateghi et al. 2014a, 2014b, Guikema et al. 2014, McRoberts, 2017) use distribution-free ensemble data mining methods such as Random Forest. These types of models do not directly estimate distributions, and they lack asymptotic approximations of distributions as one would have with generalized linear models. It is thus problematic to say exactly what the point estimates represent. In a sense, these models could be thought of as a R = E conceptualization, but the details are unclear.

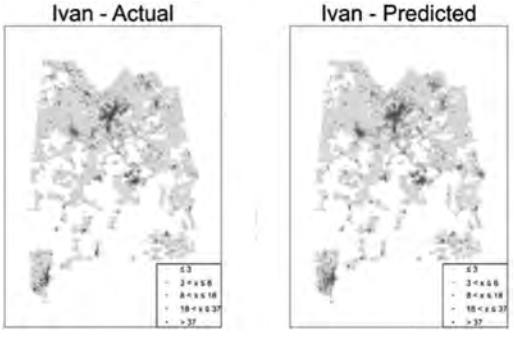


Figure 1. Example power outage forecasts from Nateghi et al. (2014a).

Even in these point-estimate models, some sources of uncertainty have been represented. For example, the models are usually in a forecasting mode, i.e., predicting outages based on a weather forecast, and arguably the largest source of uncertainty in the outage predictions is uncertainty in the weather forecast. The one paper I am aware of that explicitly included weather forecast uncertainty in outage forecasting is Quiring et al. (2014). In this approach, a deterministic model, the one underlying the results in Figure 1, was coupled with 1,000 replications of simulated hurricane forecasts. The deterministic outage forecasting model was run for each replication of the weather forecast, yielding an empirical probability density function for the outage forecast. An example, taken from Quiring et al. (2014) is shown in Figure 2.

In Figure 2, the red bar shows what the result would be if the deterministic model was run for the single best estimate weather forecast. The star represents the actual number of outages, and the green bar represents the mean of the ensemble. We see that there is a high degree of uncertainty in the outage forecasts due to the uncertainty in the weather forecast, uncertainty that is ignored in the deterministic forecasts shown in Figure 1. This approach can be conceptualized as a R = f(S,P,C) framework, though only a portion of the uncertainty is accounted for.

As a third approach, consider a model that is explicitly trained to estimate quantiles of the distribution for outages conditional on the input parameter values. This is the approach developed by both Guikema et al. (2017) and He et al. (2017). Both of these papers used a Quantile Regression Forest (QRF), a variant of the popular random forest algorithm that is designed to estimate specified quantiles of the distribution of the response variable. For concreteness, I will use the model developed by Guikema et al. (2017) as an example, though He et al. (2017) is similar.

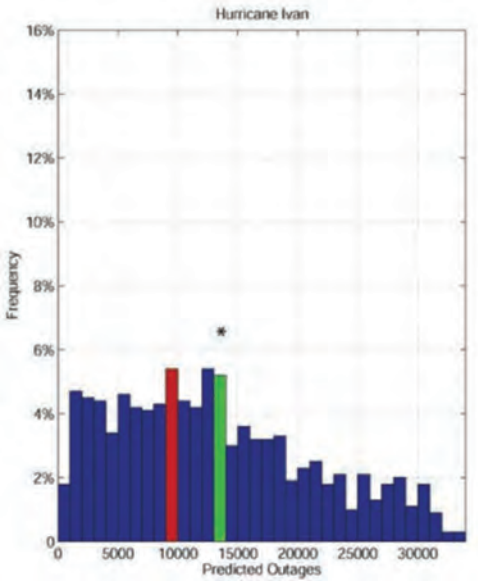


Figure 2. Example of including weather forecast uncertainty in an outage forecasting model. The figure if from Quiring et al. (2014).

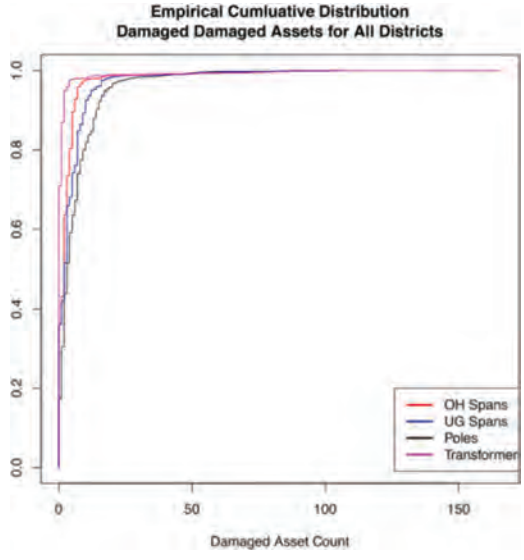


Figure 3. Example of the probabilistic predictions of asset damage from Guikema et al. (2017).

The Guikema et al. (2017) model was developed for estimating damage to power systems due to adverse weather events. Damage is estimated separately for four classes of assets—poles, transformers, overhead line spans, and underground cable runs—where the quantity of interest for each is the number of damaged items in each asset class. The model is a two-stage model. The first stage is a classification model the estimates the probability of their being damage in any one (or more) of the asset classes on a given day due to the weather. The second stage consists of four QRF models, one per asset class, that estimates quantiles (in 0.01 increments) of the distribution of the number of damaged assets in that asset class. There is a separate model that converts this into a probabilistic estimate of the person-hours needed for restoration, but I will focus on the damage model here.

To make a prediction for a given day with the Guikema et al. (2016) model, N (typically N = 10,000) replications are simulated from the two-step process, leading to an estimated probability distribution. Figure 3 shows an example of this type of prediction, presented as a cumulative density function. We see from this figure for that this particular day, there is a high probability of very little damage but that there is a long tail to the distribution. That is, there is a small chance of substantial damage on this particular day given the weather forecast. This approach directly captures the uncertainty in the forecast conditional on the weather forecast. It does not, however, explicitly model and propagate the weather forecast uncertainty through the predictive model as Quiring et al. (2014) does. This approach can be thought of as a R = f(S,P,C) approach, though it captures different aspects of the uncertainty differently than Quiring et al. (2014).

4 DISCUSSION AND RECOMMENDATIONS FOR A PATH FOREWARD

4.1 *Summary of the state of the literature in representing uncertainty in power outage forecasting*

From the discussion above, it is clear that power outage forecasting models have evolved over time in how they handle uncertainty and how they conceptualize risk. Early models were deterministic with at best an asymptotic approximation of the predictive distribution. One model that I am aware of has explicitly included weather forecast uncertainty, the largest source of uncertainty, in the forecasting. Two other models have used explicitly probabilistic prediction models to directly estimate probability distributions for outcomes, but they did not explicitly include weather forecast uncertainty. None of the models to date have moved beyond probabilistic representations of uncertainty to include descriptions of the strength of the knowledge underlying the predictions. This is potentially

quite important because there is substantially more knowledge and data about the performance of power systems under some storm conditions than others.

4.2 *Suggested path forward*

Much has been achieved in the past 12 years during which power outage forecasting has been an active area of research. Model accuracy has improved substantially, and an increasing number of explanatory factors are being included in outage forecasting models. Progress on incorporating uncertainty into the forecasts has been made as well, though this has been more limited as discussed above. Where should the field move next? What types of developments are needed? Here I discuss three key developments that are needed, focused on the treatment of uncertainty in outage forecasting models.

First, models should be developed that more completely characterize uncertainty probabilistically. I am aware of only one paper that has propagated weather forecast uncertainty through an outage forecasting model, and the model in this paper has not been used operationally due to computational limitations. The two probabilistic outage forecasting papers (i.e., Guikema et al. 2017 and He et al. 2017) only handle weather forecast uncertainty indirectly, to the extent that it is represented in the uncertainty in the conditional predictions given the outage forecast. Models are needed that combine these probabilistic conditional forecasts with an explicit modeling of weather forecast uncertainty. While this would be technically challenging it would give a much more complete picture of the uncertainty in a given forecast.

Second, approaches should be developed that move beyond probabilities to provide a clear description of the strength of the information underlying the prediction. Consider for example two situations. In the first, an outage model is to be run for weather event that is similar to a large number of past events that have impacted that utility system, and there is strong confidence in that weather forecast for that day. The model was presented with a large set of highly relevant data during the training process. In the second, the weather forecast is much more uncertain, and the forecast conditions have only been experienced by that power system a small number of times in the past. The model may give similar numerical estimates in these two cases, but many would argue that the utility should have much more confidence in the model in the first case. Qualitative descriptors of the strength of knowledge and the similarity of the forecast conditions to past events, if properly developed and described clearly in practice, could help utility decision makers better judge the degree of confidence they should have in the model in a given situation.

Third, better methods for communicating uncertainty in predictions to utility personnel need to be developed. This is related to the second suggestion, but the focus here is on how the probabilistic forecasts themselves are communicated to utility personnel. In my experience, utility personnel do not understand probability distributions as predictions or, in some cases, even discrete quantiles from the distribution particularly well. They are also making expensive decisions under intense time pressure after a major storm. If they are confused by model output, the model is not helpful to them. If probabilistic predictions are to be used in utilities in practice, new methods for communicating these predictions much be found.

5 CONCLUDING THOUGHTS

Much progress has been made in power outage forecasting for storms. Outage forecasting models are in increasingly wide use in practice, and the available models are increasing in both accuracy and sophistication. However, challenges remain in how these models conceptualize risk and uncertainty and how they model and present uncertainty in predictions to model users. This paper suggests some specific steps that should be taken as this field moves forward, steps that would help to improve how power outage forecasting models deal with the sometimes-substantial uncertainty in their predictions.

REFERENCES

Aven, Terje. *Foundations of risk analysis*. John Wiley & Sons, 2012.

Aven, T., and Guikema, S. 2015. On the concept and definition of terrorism risk. *Risk Analysis*, *35*(12), 2162–2171.

Flage, R., and Aven, T. 2009. Expressing and communicating uncertainty in relation to quantitative risk analysis. *Reliability & Risk Analysis: Theory & Application*, *2*(13), 9–18.

Guikema, S.D. and S.M. Quiring. 2012. "Hybrid Data Mining-Regression for Infrastructure Risk Assessment Based on Zero-Inflated Data," *Reliability Engineering & System Safety*, Vol. 99, pp. 178–182.

Guikema, S.D., S.R. Han, S.M. Quiring. 2010. "Pre-Storm Estimation of Hurricane Damage to Electric Power Distribution Systems," *Risk Analysis*, Vol. 30, No. 12, pp. 1744–1752.

Guikema, S.D., R. Nateghi, S.M. Quiring, A. Reilly, M. Gao. 2014. "Predicting Hurricane Power Outages to Support Storm Response Planning," *IEEE Access*, Vol. 2, OPEN ACCESS. DOI: 10.1109/ACCESS.2014.2365716

Gukema, S.D., S.M. Quiring, K. Buckstaff, M. Beck, B. McRoberts, R. Nateghi, and T. Logan. 2017. Storm Damage and Restoration Labor Estimation: An All-Weather Model, Working paper, University of Michigan (to be submitted to *IEEE Access*).

Han, S.-R., S.D. Guikema, and S.M. Quiring. 2009a. "Improving the Predictive Accuracy of Hurricane Power Outage Forecasts using Generalized Additive Models," *Risk Analysis,* Vol. 29, No. 10, pp. 1443–1453.

Han, S., S.D. Guikema, S.M. Quiring, K. Lee, D. Rosowsky, and R.A. Davidson. 2009b. "Estimating the Spatial Distribution of Power Outages during Hurricanes in the Gulf Coast Region," *Reliability Engineering & System Safety*, Vol. 94, No. 2, pp. 199–210.

He, J., Wanik, D.W., Hartman, B.M., Anagnostou, E.N., Astitha, M., & Frediani, M.E. (2017). Nonparametric Tree-Based Predictive Modeling of Storm Outages on an Electric Distribution Network. *Risk Analysis*, *37*(3), 441–458.

Kaplan, S., & Garrick, B.J. (1981). On the quantitative definition of risk. *Risk analysis*, *1*(1), 11–27.

McRoberts, D.B., S.M. Quiring, and S.D. Guikema. 2016. Improving Hurricane Power Outage Prediction Models Through the Inclusion of Local Environmental Factors, *Risk Analysis*, published online October 2016. DOI: 10.1111/risa.12728 [in press for print version].

Nateghi, R., S.D. Guikema, and S.M. Quiring. 2011. "Comparison and Validation of Statistical Methods for Predicting Power Outage Durations During Hurricanes," *Risk Analysis*, Vol. 31, No. 12, pp. 1897–1906.

Nateghi, R., S.D. Guikema, S.M. Quiring. 2014a. "Forecasting Hurricane-Induced Power Outage Durations," *Natural Hazards*, Vol. 74, No. 3, p. 1795–1811.

Nateghi, R., S.D. Guikema and S.M. Quiring, 2014b. "Power Outage Estimation for Tropical Cyclones: Improved Accuracy with Simpler Models," *Risk Analysis,* Vol. 34, No. 6, p. 1069–1078. DOI 10.1111/risa.12131.

Quiring, S.M., L. Zhu, and S.D. Guikema. 2011. "Importance of soil and elevation characteristics for modeling hurricane-induced power outages" *Natural Hazards*, Vol. 58, No. 1, pp. 365–390.

Quiring. S.M., A. Schumacher, and S.D. Guikema. 2014. "Incorporating Hurricane Forecast Uncertainty into Decision Support Applications," *Bulletin of the American Metrological Society*, Vol. 95, pp. 47–58.

Winkler, J., Duenas-Osorio, L., Stein, R., & Subramanian, D. 2010. Performance assessment of topologically diverse power systems subjected to hurricane events. *Reliability Engineering & System Safety*, *95*(4), 323–336.

Occupational safety

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Probabilities in safety of machinery: Sample space of yearly accident data

Heinrich Mödden
German Machine Tool Builders' Association (VDW), Frankfurt am Main, Germany

ABSTRACT: The issues "risk assessment" and "tolerable risk" are causing conflicting reactions not only among Health and Safety experts. Experienced designers in the machinery sector are sometimes unsettled, too. The controversies are mainly about numerical probabilistic representations. These are new in the field of general machinery safety, and the key term "probability" turned out to be ambiguous. Recently introduced probabilistic methods encounter a well-proven practical state-of-the-art, which is merely based on qualitatively defined requirements.

If objective findings are to be taken as a basis, it is obvious that the sequence of the total annual figures for reportable accidents can be considered as random independent events in a population of comparable elements. Mathematical concepts seem to make sense if the annually recorded (machine-specific) accident data are interpreted as the overall result of a huge "random experiment" in a relevant observation framework (e. g. DGUV statistics in Germany).

1 PROBABILITIES IN SAFETY OF MACHINERY

The issues "risk assessment" and "tolerable risk" are causing conflicting reactions not only among Health and Safety experts. Experienced designers are sometimes unsettled, too. The controversies are mainly about numerical probabilistic representations. These were recently introduced in the field of general machinery safety, when the revised European Machinery Directive 2006/42/EG (2006) extended the "hazard analysis" of former versions to a "risk analysis" by introducing the term "probability" in the expression: "*estimate the risks, taking into account the severity of the possible injury or damage to health and the probability of its occurrence*". Since this alteration in a legal text, simplified probabilistic methods as in ISO 13849-1 (2008) are being developed. They encounter a well-proven practical state-of-the-art, which is merely based on qualitatively defined requirements. They were mainly focussing on hazards as such (and their countermeasures) rather than "risks, severities of injuries and their probability of occurrence". Nevertheless, on the background of customer demands for very high availabilities (which is equivalent to high inherent safety), it brought about a well-tried state-of-the-art following the three-step-reduction method of ISO12100 (2010).

It is hard to believe, but for the time being, the normative frame for machinery safety does not contain a numerical risk model, which summarizes the single factors plausibly to an overall risk, in order to e.g. compare theoretical results with empirical field data. As a consequence, fictitious hazards and actual risks are being mashed up again and again, so that in controversial discussions in the world of safety standardization, mostly the strongest gut feeling overtrumps logical considerations.

The state-of-the-art is defined on a non-quantitative descriptive background in harmonized safety standards in the Official Journal of the European Commission (2017) since more than 20 years. However, the advantage of the quantitative probabilistic theory is clearly visible: the entire risk reduction process can be detailed in single quantitative factors showing the proportions to enable a more effective engineering. Surprisingly, the transition from "qualitative" to "quantitative" requirements was not so easy, since the key term "probability" turned out to be ambiguous: subjective probabilities are disturbing the discussions, since many traditional experts claim that objective probabilities were missing or difficult to derive. Is this really so?

2 REFERENCE TO ACCIDENT RECORDS

The accident numbers of machine tools (for metal working) in Germany are decreasing since the introduction of a new statistical framework of the European Union in the year 2004, and also before. This pleasant trend of decreasing overall numbers

of reportable accidents is the result of a considerable effort among all stakeholders, on the manufacturer side and on the user side as well (BGHM).

For instance, German manufacturers of machine tools are closely working together in a specific working group of VDW (German Machine Tool Builders Association, Frankfurt) on safety issues in order to exchange their experiences. This is necessary, because the commenting phases of product safety standards, which are being repeated every 5 years, urge them to review existing standard provisions.

The predominant question is then, whether a state-of-the-art can be considered as "proven-in-use" to be safe, or if there are reports of accidents (or incidents) indicating the need for certain upgrades in the safety design.

Answering this question is only possible, if the relevant safety experts of BGHM (German Occupational Safety Organisation for metal-working machines) are involved, see Kesselkaul and Meyer (2016), Adler (2015), Platz (2016) and Preusse (2005).

2.1 *Record low for machine tools was in 2014 and it was missed in 2015*

Most recently, the journal "BGHM-Aktuell" reports in June 2016 a significant decrease of summarized accident data for the branches wood- and metal-working, see Platz (2016). For the separated numbers of metal-working machine tools (i.e. without wood-working), the situation of Figure 1 remains. The decreasing trend in the overall numbers of reportable accidents is also visible here. Noteworthy is that in 2015 "only" 17.235 reportable accidents are allocated to metal-working machine tools (out of the machine perspective). The delightful results in the overall numbers of wood and metal of are unfortunately different when focussing closer on the subsets of metal. For instance, in Fig. 2 the subset "new pention payments" is displayed for (metal-working) machine tools, and Fig. 3 "fatal accidents".

Obviously, since 2008 the numbers are oscillating up and down, e.g. 281 in 2014 and 350 in 2015; a stable decrease since 2008 is not visible. Thus, the above mentioned record low in 2015 does not exist

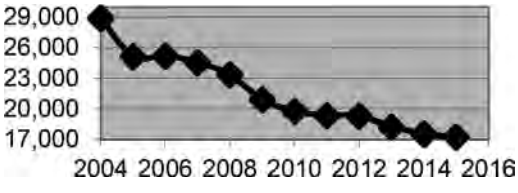


Figure 1. Reportable accidents of German machine tools from 2004 to 2015 (Source: DGUV, VDW (2015).

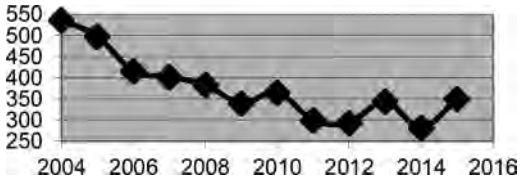


Figure 2. New pention payments for machine tools from 2004 to 2015 (Source: DGUV and VDW (2015)).

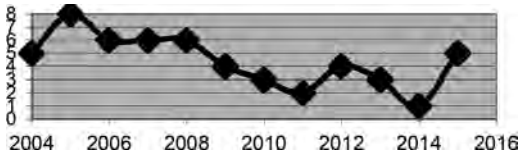


Figure 3. Fatal accidents for machine tools from 2004 to 2015 (Source: DGUV and VDW (2015)).

for machine tools, because it happened in the year before: so there is still a need for action.

Remark: the interpretation limits of statistical findings of DGUV are considered in an extended version of this paper.

3 CONNECTING THE ACCIDENT STATISTICS TO A PROBABILITY SPACE

Obviously, the sequence of yearly overall numbers for reportable accidents can be regarded as random independent events, since each year different random effects are causing the reportable accidents (different machines and operators are affected in different situations).

3.1 *Random experiments and accident modelling*

Mathematical concepts as in Barner et al (1983), Bauer (1990,1991), Bertsche et al (2004) and Feller (1967) seem to be useful for the reduction of real accident risks, if the yearly recorded accident data are interpreted as the overall outcome of a huge "random experiment" in a relevant observation frame (which e.g. is given by Occupational Safety (DGUV) Statistics in Germany):

a. the operation of an average number of α comparable machines during the actual record year;
b. an average number of β operators,
c. who are during an average number of γ hours busy in operating them.

Consequently, the annual accident statistics must be linked to a clearly defined "probability space" (also known as "sample space"). As a result, Health and Safety experts can monitor yearly accident

records on a probabilistic basis. This is recommendable, because the law in Europe Machinery Directive 2006/42/EC refers to the term "probability".

Regrettably, the probability theory does usually not belong to the education of engineers. Neither is the measure theory being lectured among mechanical engineers. Since the probability theory is based on the measure theory, some notations shall be explained graphically and vividly.

3.1.1 *Basic terms*

The term "probability" can be expressed in different ways, see Moedden (2015):

a. as a real number between 0 and 1, where 0 means impossible and 1 means a certain event; also
b. the expected frequency is used, which is the fraction of the number of relevant events over all possible events (e.g. expressing a statistical finding, empirical or theoretical); if this fraction is connected to a given period of time, it means the probability of an event per time;
c. where no statistical data exist, "probability" can express a degree of belief about facts (e.g. due to subjective "experience").

It is obvious that the latter might become a mere "gut feeling". This is the worst kind of a probability: maybe that is the reason, why Feller (1967) emphasizes in the preface: "it is the purpose of this book to treat probability theory as a self-contained mathematical subject rigorously avoiding nonmathematical concepts".

Also the expected frequency is often misunderstood. For instance, if there is a lack of experience, e.g. when empirical data or theoretical findings are not available, it amazingly often happens that the *probability* of an event is incorrectly mistaken for a *mere possibility* of occurrence. In doing so, the probability is not calculated with respect to the correct reference frame as a fraction of numerator/ denominator; this error is called "denominator neglect", see Blastland (2014), Spiegelhalter et al (2014). I.e. the reference frame for the calculation of probability is omitted or ignored.

As an important logical fundament for quantitative probabilities, Kolmogoroff's axioms are based on an event space of elementary singletons. The entire probability theory can be developed from them. Extract from Feller (1967) for the axiomatic probability definition (quote):

The "probability" is not defined as such. The modern theory rather uses the "probability" as a term, which has to fulfill certain axioms. The axioms established by Kolmogoroff are:

1. To every random event A a real number $P(A)$ can be assigned such that $0 \le P(A) \le 1$, which is called the probability of A. This leans on the properties of relative frequencies.
2. The probability of the sure event E is $P(E) = 1$ (scaling axiom).
3. If $\{A_i, i \ge 1\}$ are random events, which are pairwise disjunct, i.e. $A_i \cap A_j = 0$ for $i \ne j$, then follows with the limes $n \to \infty$ (countable additivity axiom):

$$\mathbf{P}\left(\bigcup_{i=1}^{n} \mathbf{A}_i\right) = \sum_{i=1}^{n} \mathbf{P}(\mathbf{A}_i) \tag{1}$$

3.1.2 *Four steps to build a probability space*

In order to take the bearings of an effective risk reduction, several steps are necessary to apply the probability theory to the "random experiment" of yearly accidents. To start with, the real accident records can be connected to some notations of the mathematical measure theory, see Bauer (1990).

Obviously, the number of accidents are "measures" in a sample space. Then, these measures need to be calibrated such that a probability space is built.

Step 1: Define the sample space Ω of this random experiment with all possible outcomes.

The entirety of all possible elementary outcomes ω_i, $(\omega_i \in \Omega)$ for every single machine and every single operator has a variety of numerous causes-to-effect relations. It spreads from technical failures over human errors to forces majeure risk (e.g. lightning stroke). So as regards the cause-to-effect relation, the elementary outcomes ω can be thought of as a countably infinite set of singletons $\omega_i = \{\omega_i, i \ge 1\}$. Nevertheless, the entirety of all possible elementary singletons, which are to be assigned to severities of injuries, can be divided in five subsets (i.e. resulting events in a record year, here sorted by their expected frequency of occurrence from high to low), as shown in Table 1.

In addition to this sample space Ω, there are some related events (e.g. further dimensions due to possible repetition): At the end of a record year in the above mentioned observation frame, every single operator in the basic population of β operators is going to be either in the fortunate subset A_1, or he/she was (once or more) lucky in subset A_2 or (once or more) unlucky in subsets A_3, A_4. Or it could even happen that he/she were in the mortal subset A_5. Accordingly for a record period, every single machine of the basic population of α comparable machines can be connected to subset A_1 or to one or more of the subsets A_2, A_3, A_4 or A_5. And finally, every single hour of the on average $\alpha \cdot \gamma$ hours can be allocated at least to event A_1, or to one of the other four subsets.

Table 1. Sample space Ω with all possible outcomes, examples given for singletons.

Subset of Ω	Meaning	Example of one singleton $\omega_i \in A_j \subset \Omega$
Event A_1	No hazardous event, i.e. undisturbed machine operation (or standing idle), no relevant hazards	Five days of automatic production from Monday to Friday without interrupts.
Event A_2	A hazardous event without causing an injury (i.e. a near accident occurred)	One Monday morning, a drilling tool broke. It was set free, but retained in the machine enclosure.
Event A_3	A hazardous event causing an accident with slight or severe reversible injury	During maintenance work, an operator crushed his finger. After months of medical treatment it healed.
Event A_4	A hazardous event causing an accident with a severe irreversible injury (pension payment)	After having defeated the safeguards of a turning machine, an operator's eye was hit by ejected swarf so heavily, that he was handicapped afterwards.
Event A_5	A hazardous event causing an accident with a fatal injury (pension payment to the relatives)	On a Friday afternoon, an operator was entangled by a CNC-lathe, when he was manually polishing the surface of the turning workpiece.

It has to be remarked that this sample space Ω is not absolutely "mathematically correct", because the fact is ignored that the basic populations of machines and operators presumably are slightly different from one year to the next. However, this "transition error" alters the findings below only gradually, but not substantially, because the conclusions remain the same.

As regards real design practice, Heisenberg (2013) illustrates an (implicitly described) sample space for a certain range of products and the according event records, which fits exactly to the explicit one in Table 1.

Step 2: A measure space is required, which contains the measurable sets (called events).

The measure theory illustrates that with a collection of measurable events $\mathcal{F} = \{A_i, i \geq 1\}$, we get the measurable space $(\Omega, \mathcal{F})$. A measure on $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0, \infty]$ such that:

i. $\mu(\varnothing) = 0$
ii. If $\{A_i,\ i \geq 1\}$ is a sequence of disjoint sets in $\mathcal{F}$, then the measure of the union (of countably infinite disjoint sets) is equal to the sum of measures of individual sets, i.e.

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \tag{2}$$

The second property stated above is known as the *countable additivity* property of measures. From the definition, it is clear that a measure can only be assigned to elements of $\mathcal{F}$. The triplet $(\Omega, \mathcal{F}, \mu)$ is called a *measure space*. μ is said to be a *finite measure* if $\mu(\Omega) < \infty$; otherwise, μ is said to be an *infinite measure*. In particular, if $\mu(\Omega) = 1$, μ is said to be a *probability measure*.

Since an event is a subset (of comparable singletons) of the sample space Ω, to which a probability will be assigned, for the measurable subsets of the accident sample space follows: The disjoint sets in $\mathcal{F}$ are here $\{A_i, i = 3,5\} = \{A_3, A_4, A_5\}$ with the property:

$$\mu\left(\bigcup_{i=3}^{5} A_i\right) = \sum_{i=3}^{5} \mu(A_i) = \mu \cdot (A_3 + A_4 + A_5) \tag{3}$$

Step 3: Assign measures to the events of $\mathcal{F}$.

The measure μ can be understood as the overall number of all accidents of the type reportable (R) in the relevant observation frame above, since the subsets of severe accidents with pension payments (PP) and mortal accidents (M) are contained therein.

$$\mu(A_3 + A_4 + A_5) = \mu R \tag{4}$$

However, a closer look to different severities requires a distinction of the numbers of the subsets so that they are disjoint:

M =: number of all mortal accidents (remains): this number corresponds to event A_5 above.

PP^* =: number of severe accidents with pension payments (PP) minus fatal accidents: $PP^* = PP - M$: this number corresponds to event A_4 above.

R^* =: number of all reportable accidents minus the number of severe accidents with pension payments (PP) and minus the number of all mortal accidents (M): $R^* = R - PP^* - M$: this corresponds to event A_3 above.

Now, the values R^*, PP^*, M can be derived from the yearly accident records. They can be assigned to the $\mathcal{F} = \{A_3, A_4, A_5\}$ following these principles:

$$\mu = R = R^* + PP^* + M \tag{5}$$

Remark: Since the measurable space $\mathcal{F}$ is simple with only three $\mathcal{F}-$ measurable sets $\{A_3, A_4, A_5\}$, it fulfills the conditions of a Algebra, a σ–Algebra and a Borel Algebra (see Feller (1967) and Jaganathan (2015)).

Step 4: Assign probabilities to the events of $\mathcal{F}$. The triplet $(\Omega, \mathcal{F}, \mathrm{P})$ is called a probability space, if the following three properties (sometimes referred to as the axioms of probability) are fulfilled: a probability measure is a function $\mathrm{P}: \mathcal{F} \to 0,1$ such that:

i. $\mathrm{P}(\varnothing) = 0$
ii. $\mathrm{P}(\Omega) = 1$.
iii. (Countable additivity:) If $\{A_i, i \geq 1\}$ is a sequence of disjoint sets in $\mathcal{F}$, then

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(A_i) \tag{6}$$

Here, we have the single probability measures $\mathrm{P}(A_1)$, $\mathrm{P}(A_2)$, $\mathrm{P}(A_3)$, $\mathrm{P}(A_4)$, $\mathrm{P}(A_5)$, which are strictly positive. Because of the limitations of the above defined random experiment of yearly accident records, we know that for every combination of α, β, γ the following equation holds: $\mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) + \mathrm{P}(A_4) + \mathrm{P}(A_5) = 1$. But, we don't know the actual numbers α, β, γ. Obviously, these probabilities are not easy to determine, because the frame parameters α, β, γ of the basic population are not known. Thus, the measures R^*, PP^*, F of the events A_3, A_4, A_5 are only rough reference numbers, which can be plotted of the years.

3.2 *Available data put into an observation frame*

So, for the sake of showing the proportions of $\mathrm{P}(A_1)$, $\mathrm{P}(A_2)$, $\mathrm{P}(A_3)$, $\mathrm{P}(A_4)$, $\mathrm{P}(A_5)$, some practical probability measures out of the record numbers R^*, PP^*, M, the values α, β, γ shall be estimated for above defined observation frame, to start with (see remark below):

$$\alpha \cong 500.000, \beta \cong 50.000, \gamma \cong 4 \tag{7}$$

The average hours per day $ah_d = \alpha \cdot \gamma$ are then: $ah = 500.000 \cdot 4 = 2{,}000.000$. And per year with on average about 220 working days: $ah_y = 2{,}000.000 \cdot 220 = 4{,}4 \cdot 10^8$.

The probability of an event A_5 can then be expressed as a relative frequency (rf):

$$\mathrm{P}(A_5) = \mathrm{rf}(A_5) = {}^{M}\!/_{ah_y} \tag{8}$$

Accordingly, the probabilities of the events A_3 and A_4 are:

$$\begin{aligned} \mathrm{P}(A_3) &= \mathrm{rf}(A_3) = {}^{R^*}\!/_{ah_y} \quad \text{and} \\ \mathrm{P}(A_4) &= \mathrm{rf}(A_4) = {}^{P^*}\!/_{ah_y} \end{aligned} \tag{9}$$

The residual probabilities $\mathrm{P}(A_1) + \mathrm{P}(A_2)$ are then:

$$\mathrm{P}(A_1) + \mathrm{P}(A_2) = 1 - \mathrm{P}(A_3) + \mathrm{P}(A_4) + \mathrm{P}(A_5) \tag{10}$$

The accident record for machine tools of 2014 shall serve as an example (see DGUV):

$M = 1$, $PP^* = 281 - 1 = 280$, $R^* = 17.563 - 280 - 1 = 17.282$.

$$\begin{aligned} \mathrm{P}(A_5) &= \mathrm{rf}(A_3) = {}^{1}\!/_{4{,}4 \cdot 10^8} = 2{,}3 \cdot 10^{-9} \\ \mathrm{P}(A_4) &= \mathrm{rf}(A_4) = {}^{280}\!/_{4{,}4 \cdot 10^8} = 6{,}4 \cdot 10^{-7} \\ \mathrm{P}(A_3) &= \mathrm{rf}(A_3) = {}^{17.282}\!/_{4{,}4 \cdot 10^8} = 3{,}9 \cdot 10^{-5} \\ &\to \mathrm{P}(A_1) + \mathrm{P}(A_2) \\ &= 1 - 3{,}9 \cdot 10^{-5} - 6{,}4 \cdot 10^{-7} - 2{,}3 \cdot 10^{-9} \\ &= 0.99996. \end{aligned}$$

In Figure 4 the proportions are illustrated according to the assumption of eq. (7). The light grey right column represents the probabilities of the subsets A_1 and A_2, where no injuries happen. The stepwise darker grey columns indicate the probabilities of the subsets A_3, A_4 and A_5, where accidents are rerted as slight or severe reversible injuries, severe irreversible injuries or even fatal injuries. Please take note of the fact that only the absolute values of eq. (10) should be questioned, since their relative proportions are largely independent of the presumptions made above in eq. (7). If absolute values were to be acquired, the assumptions in eq. (7) could be spreaded in terms of empirical worst case, average, best case as a histogram.

Because $\{A_i, i = 3,5\}$ is a sequence of three disjoint sets, eq. (6) is fulfilled:

$$\mathrm{P}\left(\bigcup_{i=3}^{5} A_i\right) = \sum_{i=3}^{5} \mathrm{P}(A_i) = P(A_3) + P(A_4) + P(A_5) \tag{11}$$

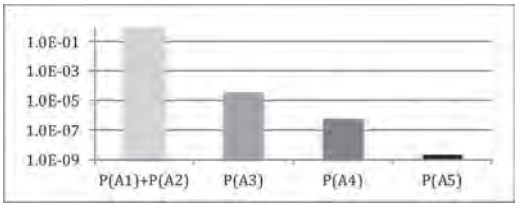


Figure 4. Relative proportions in event probabilities are largely independent of assumptions in eq. (7).

3.3 *Fascinating vision: no accidents ("zero risk")*

Now, as we have built a probability space based on a "σ–Algebra", we can have a closer look to the fascinating vision "zero risk". Based on the definitions 1. to 4. above, the zero-one-laws of the probability theory can be applied. They indicate that the probability of events of a certain type (here a set of the "limes superior") is either 0 or 1. That is to say: Those events happen either almost surely or they are almost impossible.

In particular the Borel-Cantelli Lemmas seem to be suitable to check the decreasing trend of accident numbers, whether possibly the fascination "zero risk" can be realized in the future (see Moedden (2017)). Thus, the sum of all single probabilities (of all subsequent events) is the crucial quantity: if the sum is increasing limitless, the repeated occurrence of events (here accidents) is virtually certain. As a consequence of this comprehension, safety experts cannot be content with current safety standards and their (implicitly/explicitly) defined tolerable risk.

Obviously, the tolerance level has to be continuously reduced for reaching a "zero risk". To start with the foundation of the Borel-Cantelli Lemmas in Moedden (2017), two for convergence and divergence relevant event types (or subsets) of a σ – algebra $\mathcal{F}$ shall be explained.

3.3.1 *"Limes inferior"*

The sequence of subsets $\{A_i, i = 1,n\}$ in a σ – algebra $\mathcal{F}$ shall be given, e.g. A_i could be a single record of reportable accidents of year "*i*" in an observation frame. Now, we form a specific subset of elements ω, which are contained in (only) almost all A_i, i.e. these elements are contained (at the utmost) in only a finite number of A_i:

The example in Figure 5 indicates that for a singleton ω, there are two subsets: i. $\{\omega \notin A_1, A_4, A_7\}$ and ii. $\{\omega \in A_2, A_3, A_5, A_6, A_8, A_9, \ldots\}$. The sequence of both subsets is called the "limes inferior", which is a kind of lower accumulation point of a sequence of random events. It is defined as of the union $(n = 1,\infty)$ of the intersections of all partial sequences $(i = n,\infty)$:

$$A_{\inf} = \lim\inf A_n = \bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} A_i \qquad (12)$$

The subset A_{inf} is the set of all elements $\omega \in \Omega$, which are contained in almost all (finitely many) of the subsets A_n. Let us have a closer look to a single element ω, which could e.g. mean that ω is the event "no fatal accide at grinding machines in an entire record year". As shown in Figure 5, the singleton ω is only contained in finitely many A_i. That is to say, there is a time point n^* such that for all $n \geq n^*$: $\omega \in A_i$. Then ω is in the inter-section: $\omega \in \bigcap_{i=n^*}^{\infty} A_i$, which means (in this example) no more fatal accidents would occur at grinding machines in the relevant observation frame beyond $n \geq n^*$ ("zero risk" accomplished). If all time points are considered, we get: $A_{inf} = lim\inf A_n = \bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} A_i$. The reverse argument needs to be proven. Let ω be one singleton with $\omega \in \bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} A_i$, then there is a number n^* with $\omega \in \bigcap_{i=n^*}^{\infty} A_i$. From this time point on, $\omega \in \bigcap_{i=n^*}^{\infty} A_i$ holds. For $1 \leq i < n^*$ (finitely many), we don't know how often ω is contained. Consequently, the element ω is contained in (only) almost all A_i.

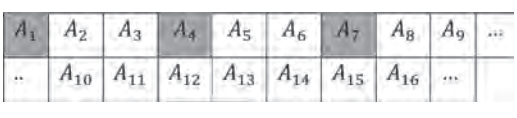


Figure 5. Example for the lower accumulation point "limes inferior".

3.3.2 *"Limes superior"*

The sequence of subsets $\{A_i, i = 1,n\}$ in a σ–algebra $\mathcal{F}$ shall be given. We form another subset of elements, which are contained in infinitely many A_i:

The example in Figure 6 indicates that for a singleton ω, there are two subsets:

i. $\{\omega \notin A_1, A_2, A_3, A_5, \ldots\}$ and
ii. $\{\omega \in A_4, A_6, A_8, A_9, \ldots\}$. So ω is contained in the non-prime number subsets, but not in the prime number subsets. This sequence of subsets is called the "limes superior", which is a kind of upper accumulation point of a sequence of random events. It is defined as the intersection $(n = 1,\infty)$ of the union of all partial sequences $(i = n,\infty)$:

$$A_{sup} = A_n \; i.o. = lim\sup A_n = \bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} A_i \qquad (13)$$

The subset A_{sup} is the set of all elements $\omega \in \Omega$, which are contained in infinitely many of the subsets A_n, and obviously $\omega \in A_n$ applies for infinitely many $n \in \mathbb{N}$, i.e. "A_n i.o." with "i.o. = infinitely often". A_{sup} is a nested intersection of unions. The union $\bigcup_{i=n}^{\infty} A_i$ can be thought of as a sequence $(B_n)_{n\in N}$ with $B_n \supset B_{n+1} \supset B_{n+2}$ … The (B_n) are

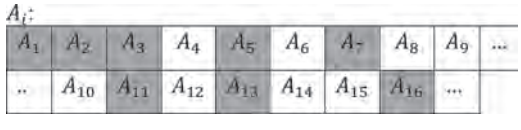


Figure 6. Example for the upper accumulation point "limes superior".

Figure 7. Matryoshka nested doll (Source: Elisabeth Mödden).

nested decreasing subsets. This resembles the Matryoshka nested doll in Fig. 7, where the smallest innermost doll is the "intersection" of all dolls.

Formally, let $(B_n)_{n\in N} = \bigcup_{i=n}^{\infty} A_i$, then $(A_n\ i.o.) = \bigcap_{i=1}^{\infty} B_n$. In doing so, B_n is the event that at least one event $A_n + A_{n+1} + \ldots + A_\infty$ occurs (that is a bigger set than at least one event $A_{n+1} + A_{n+2} + \ldots + A_\infty$ occurs). It is sometimes referred to as the "end-tail" event, which means that from n onwards at least one event $A_n + A_{n+1} + \ldots + A_\infty$. occurs. Then $(A_n\ i.o.)$ is the intersection of the "end-tail" events, it means that for n an event B_n occurs. That is to say, no matter how big n is, at least one of the A_i will occur.

$$P(A_n\ i.o.) = P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathrm{P}(B_n) = \lim_{n\to\infty} \mathrm{P}\left(\bigcup_{i=n}^{\infty} A_i\right) \tag{14}$$

It is obvious that $A_{inf} \subseteq A_{sup}$, because if one singleton ω is contained only in finitely many A_n, it is also contained in infinitely many A_n. Let one singleton ω be contained in infinitely many A_i. How can this be expressed without using the term "infinitely"? For every n there is one k \$ n: $\omega \in A_k$. For every n, $\omega \in \bigcup_{k=n}^{\infty} A_k$ holds, and therefore $\omega \in \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$. The reversion is obvious.

As regards the application of the zero-one-laws on the accident records, it is fortunate that they consist of scalar values summarizing comparable singletons. So they resemble a sequence of values (a_n) which sum up a collection of $\omega_i \in \Omega$ that are assigned to a selected kind of events in the subset sequence A_n. For a sequence (a_n), the necessary and sufficient condition for convergence of sequences can be applied (see Barner (1983): (a_n) is convergent if and only if (a_n) is bounded and:

$$\lim_{n\to\infty} a_n = A_{inf} = A_{sup}$$
This is equivalent to: $A_{inf} = A_{sup}$ (15)

3.3.3 *Interpretation for available accident records*
The probability space from 3.1.2 is the platform for the interpretation: For the sake of simplicity, the accident records in Figure 1 can be interpreted as a finite sequence (i.e. as a partial sequence of an infinite sequence in the future), and the available records contain twenty-two random events $A_1, A_2, \ldots A_{22}$ until the year 2015 (i.e. the available "measures" of reportable accidents in the observation frame: injuries caused by machine tools in Germany). Since the sequence is monotonously decreasing (at least during the last ten years for the overall numbers), the "limes superior" of the only 22 (instead of ∞) events can be thought of as the minimum number of accidents, which happened in (almost) every of these twenty two years. Apparently, this is the number of accidents in the last recorded year A_{22} = 17.563. So, the "limes superior" for the partial sequence $A_1, A_2, \ldots A_{22}$ is the value of the lowest number of accident records (the "smallest doll" contained in all of the Matryoshka nested dolls from above, see Jaganathan (2015), which is here equal to the last record:

$$A_{sup} = A_n\ i.o. = lim\ sup\ A_n = \bigcap_{n=1}^{22} \bigcup_{i=n}^{22} A_i = A_{22} = 17.563 \tag{16}$$

It has to be remarked that this simplification is only "almost mathematically correct", because the intersection of the union of twenty-two different subsets ignores the fact that the basic populations of machines and operators presumably have been slightly different from one year to the next etc. However, this reservation does not alter the fact that the Borel-Cantelli lemmas can illustrate the vision "zero risk" mathematically.

4 SUMMARY AND OUTLOOK

Admittedly, theoretical risk assessment starts in the hypothetical "what if" domain, where theoretical risk can be estimated only logically in cause and effect (at the most), but not in an *absolute* scale. However, actual *relative* risk reduction effects between different yearly accident records can be calculated and compared quite exactly, because the real risk actually can be "measured" precisely as by the DGUV records, as well as in individual records of machine tool builders. Therefore, this paper tries to support plausible risk considerations connecting theory and reality. Hopefully, it also helps to improve a common understanding of the term "probability". Only then can the 1. Borel-Cantelli lemma be met, which leads the way to the fascinating vision "zero risk", at least approaching it yearly step by step.

Of course it is indispensable that the a.m. accident investigation of BGHM and DGUV is going to be continued meticulously. VDW is offering

support with this paper. This is a precondition such that a more complete Pareto diagram (than the one in Figure 12 of Moedden (2017)) can be brought about, showing potential safety gaps in current design and operational practices, which need to mended first. Subsequently, possible upgrades in the safety design against not significant hazards can be looked at: doing the first thing and not leaving the other option out. In addition, individual manufacturers of specific product ranges can monitor their own yearly records separately, such as convincingly demonstrated in Heisenberg (2013). Moreover, VDW is supporting the member companies accordingly in research projects (see Nowizki et al (2016)) in order to make German machine tools as safe as reasonably possible.

SYMBOLS AND OTHER DEFINITIONS

Symbol	Dim.	Meaning
α, β, γ	–	Parameters of the observation frame of acc. statistics
μ	–	Measure of a subset in a sample space Ω
$\sigma, \mathcal{F}$	–	Indicator for certain algebra
k_1, k_2	–	Adjustment parameter
Ω	–	Sample space
ω_i	–	Singleton of the sample space Ω
A_1 *to* A_5	–	Accident specific subsets of the sample space Ω
A_i, $P(A_i)$	–	General events in the sample space Ω and their probabilities P
$A_{inf} = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i$	–	Limes inferior of a sequence A_i also indicated as: A_i *f.o.* = *lim inf* A_i
$A_{sup} = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i$	–	Limes superior of a sequence A_i also indicated as: A_i *i.o.* = *lim sup* A_i
R, PP, M	–	Number of accidents: reportable, pension pay., mortal
$\in, \notin, \forall$	–	Element of, not an element of, holds for all

ACKNOWLEDGEMENT

Gratitude shall be expressed to Christoph Thomann (DGUV) for giving insight into the statistical framework of accident reports in Germany. Also to Ralf Kesselkaul, Christoph Meyer and Christian Adler (BGHM, 2015, 2016) for the very useful preliminary conclusions from their accident investigation reports, which presumably entailed a lot of depressing details of human misfortune.

REFERENCES

Adler, Christian 2015: *Unfallanzeigen für Schleifmaschinen*, BGHM Hannover.

Barner, M.; Flohr, F 1983: *Analysis I,* (2. Auflage), de Gruyter Lehrbuch, Freiburg.

Bauer, H. 1991: *Wahrscheinlichkeitstheorie* (4. Auflage), de Gruyter Lehrbuch, Erlangen.

Bauer, H. 1990: *Maß- und Integrationstheorie* (1. Auflage), de Gruyter Lehrbuch, Erlangen.

Bertsche, B.; et al 2004: Zuverlässigkeitstechnik im Fahrzeug-und Maschinenbau. Springer, Stuttgart.

Blastland, M. 2014: The Tiger That Isn't – Seeing Through a World of, Cambridge, profilebooks.

EN ISO 12100, 2010. Safety of machinery – General principles for design – Risk assessment and risk reduction. Germany; Beuth Verlag GmbH, Berlin.

EN ISO 13849-1, 2008. Safety of machinery – Safety-related parts of control systems – Part 1: General principles for design. Germany; Beuth Verlag GmbH, Berlin.

Feller, William 1967: *An Introduction to Probability Theory and Ist Applications*, Volume I, Third Edition, John Wiley & Sons, Inc., Princeton.

Heisenberg, David 2013: *Produktsicherheit systematisch integrieren*, VDMA Nachrichten 04/2013, Frankfurt

Jaganathan, K. 2015: Probability Foundations for Engineers, manuscript, University Madras.

Kesselkaul, R., Meyer, C. 2016. Priorisierung des Handlungsbedarfs (Prioritizing the Need for Action - Focus of Accident Occurrence). VDW-Technologietag 2016. Dusseldorf.

Moedden, H. 2015: Probabilities in Safety of Machinery – Elements of a Risk Model and Comparison with Field Data, VDW, Frankfurt a.M., Germany, ESREL Zurich Switzerland.

Moedden, H. 2017, Probabilities in Safety of Machinery: Borel-Cantelli Lemmas lead to a Prevention Dogma based on the Pareto principle, 15th International Probabilistic Workshop, Dresden.

Nowizki, N., P. Zeiler, B. Bertsche, University of Stuttgart, Germany, H. Moedden, VDW Germany 2016: Statistical Analysis of Field Data for a Proven-In-Use Assessment in the Machine Tool Industry according to ISO 13849, ESREL Glasgow.

Platz, Albert; et al. 2016: *BGHM-Aktuell, Magazin für sicheres & gesundes Arbeiten,* 5/2016, Berufsgenossenschaft Holz und Metall Mainz.

Preusse, C. 2005. Manipulation an Schutzeinrichtungen an Maschinen, St. Augustin: HVBG.

Spiegelhalter, D. 2014: *Norm Chronicles – Stories and Numbers about Risk*, Cambridge, profilebooks.

The European parliament: Machinery Directive 2006/42/EC. And its list pf harmonized safety standards in the Official Journal of the European Commission (Dec. 2017), Brussels.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Probabilities in safety of machinery—how fixed and movable guards bring about a significant risk reduction

Heinrich Mödden
VDW, Frankfurt am Main, Germany

Eckart Uhlmann, Lukas Prasol & Simon Thom
Institute for Machine Tools and Factory Management, Technische Universität Berlin, Germany

Bernd Duchstein
Schiess Tech GmbH, Aschersleben, Germany

ABSTRACT: If hazards arising from machine tools cannot be completely eliminated by design, protective devices must be provided. Separating guards prevent people from accessing or entering the danger area; in addition, they retain any parts that may have been released in the work area. An attempt shall be made here to supplement the currently purely intuitive (qualitative) consideration of the protection effects by a probabilistic scaling of the risk reduction effects. By scaling these effects, the Pareto principle can be applied: achieving the best possible benefit with minimal effort. This procedure is indispensable for machine tool manufacturers to master economic risks in global competition. Since there is no plausible risk model in current safety standards for machine safety with which risk reduction effects can be scaled, a simplified quantitative risk model is presented in this paper for this purpose (and two further papers are presented at ESREL 2018).

1 INTRODUCTION

In accordance with the guiding standard ISO 12100 (2010) risk is a function of possible damage severity and probability of occurrence, whereby the latter can be represented as the frequency of the damage. Because, over a large basic population of comparable situations, the mathematic-theoretical "probability" of an event becomes an empirical (countable) "relative frequency", which enables the verifiability of this event in the totality of all possible events, see Haigh (2012), Taschner (2013).

As a result, risk can be reduced by avoiding or reducing the extent of damage (effect related measure) and/or by reducing the probability of occurrence or frequency of the occurrence of damage or the associated hazard (cause-related measures). Separating protective devices (guards) plays an elementary role in both measures and also offer a high cost-benefit advantage. These have the effect of significant risk reduction both for protection against accidental access and for protection against flying parts, substances and, in particular, fire protection.

As for conventional machine tools with drive powers in the range of 100 kW and powerful hydraulic clamping devices, it cannot be denied that (very low) failure probabilities of technologically highly reliable machine functions or safety functions include also a basic possibility of serious damages/injuries. This is demonstrated, e. g. by the depressing fact that in Germany, machine tool operators suffer around 100 life-changing accidents and even fatal accidents in the single-digit range per year, see Kesselkaul & Meyer (2016). The typical distribution of severity is also known as the "accident pyramid" and can be displayed in a histogram, see Moedden (2017).

How can a designer live with this frustrating insight into the world of real probabilities, i. e. the annually recurring serious or even fatal accidents? One might even ask oneself the question: Can a machine tool be built safely at all?

The answer is "yes", because if the extent of injury cannot be reduced in itself, then the reduction of the relative frequency of injuries remains as a means of reduction, i. e. how often damage repeats itself in a given period of time. With reference to a basic population with comparable elements, this is equivalent to their probability.

The risk reduction of machine tools is therefore all about reducing the "probability of hazardous

situations", so that damage is avoided as far as practicable. The likelihood of hazards occurrence (see also ISO 12100 (2010), Section 5.5.2.3.2 "Occurrence of hazard events") can be reduced in several respects and is scalable as parameter O, e. g. with $0 \leq O \leq 1$ (see Fig. 2 in ISO 12100 "Risk reduction process from the viewpoint of the designer").

The starting point is always that technological machine functions (which can also become safety functions when the safety doors are open) already have very low failure probabilities. How the possible resulting hazards can be massively reduced even further with the help of separating protective devices shall be considered here and scaled plausibly. It is not the accuracy of the post-decimal value that matters, but rather the power of 10, as is usual in risk representations, see Baruch et al (2011).

2 SAFETY MEASURES IN PRODUCT STANDARDS

Separating guards for risk reduction on machine tools should be designed in such a way that they are sufficiently capable of holding back normally expected off-flying parts, see ISO 14120 (2002). The retention capability of a guard describes the resistance against the penetration of flying elements. Machine-specific design conventions are anchored in standards such as ISO 23125 (2010). In the case of impact energies that increase beyond the design conventions, e. g. due to the release of large workpieces or clamping jaws, the probability of failure increases, Meister (2017).

2.1 *Failure hypotheses*

Basic risk assessments have been carried out in product standards and corresponding safety measures have been defined for some typical machine configurations, such as machine tools, woodworking machines and balancing machines. There detailed dimensioning recommendations in the harmonized standards are available for separating protective devices in order to ensure the required retention capability. The starting point is a typical failure hypothesis for released elements that leads to defined masses, velocities and the energies derived from them, as in ISO 23125 (2010). Fragment energies that may exceed these hypotheses are therefore further reduced in terms of control technology, e. g. by limiting the number of r. p. m. of spindles or instructively by means of detailed instructions in the operating instructions, e. g. for the safe clamping of workpieces. This ought to reduce the risk of operator injury to a level that is "As Low As Reasonably Practicable" ("ALARP"), UK-HSE (1974).

The following constructional distinction is applied to protect against released elements in order to minimize the residual risk of injury:

i. Closed door (automatic operation): The safety guard must have sufficient impact resistance against ejected parts.
ii. Open door (setting mode, special or service mode): The rotational speeds of rotating parts are safely reduced, and unnecessary movements are safely stopped, so that the risk of ejected elements is reduced.

The application duration of operating modes according to ii.) should therefore be minimized as far as possible in order to minimize the risk. In the machining process, on the other hand, the observability of a process is sometimes indispensable.

2.2 *Guiding standard for safety guards ISO 14120*

For machine tools, the full enclosure as in Figure. 1 in particular is a very effective means for risk reduction (see below transition in probabilities from PFH_d to PHE; this is assigned to step two of the three-stage method in Figure. 2). For this purpose, important aspects are regulated in a Guiding Standard for separating protective devices, *e. g. access to hazardous areas (EN ISO 14120 (2002), Sec. 5.1.2), retention of ejected parts and other impacts (EN ISO 14120, Sec. 5.1.3), retention of hazardous substances (EN ISO 14120, Section 5.1.4), retention capacity (EN ISO 14120, Section 5.5) and corrosion resistance (EN ISO 14120, Section 5.6).*

In particular, the type of construction according to Section 3.5.2 is widespread in machine tools:

Figure 1. Automatic vertical turning centre acc. to ISO 23125 (2010) (Source: INDEX).

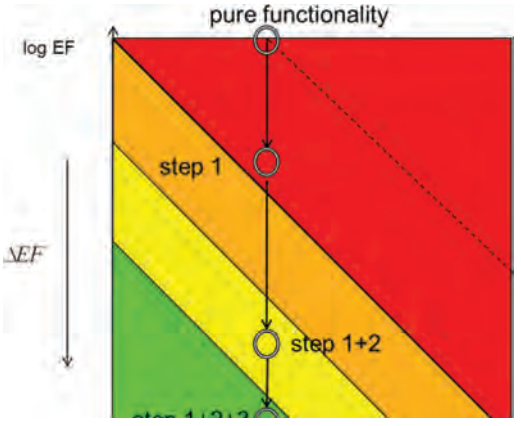


Figure 2. Risk reduction according to the three-step method in the Farmer Diagram, Baruch et al. (2011).

Interlocking guard with guard locking
Guard associated with an interlocking device and a guard locking device so that, together with the control system of the machine, the following functions are performed:

- *the hazardous machine functions "covered" by the guard cannot operate until the guard is closed and locked,*
- *the guard remains closed and locked until the risk due to the hazardous machine functions "covered" by the guard has disappeared and*
- *when the guard is closed and locked, the hazardous machine functions "covered" by the guard can operate (the closure and locking of the guard do not by themselves start the hazardous machine functions.)*

3 SEPARATING PROTECTIVE DEVICES (GUARDS) AND PROBABILISTIC SAFETY

The Machinery Directive 2006/42/EC requires the preparation of risk assessments for all life cycle phases of a machine as proof of compliance with safety and health requirements in accordance with Annex I. In point 1, it deals directly with the risk elements "severity of possible injuries" and "probability of their occurrence". Since 2012, ISO 13849 (2008) has been the new guiding standard for control safety in mechanical engineering. It determines the safety reliability of a control chain via the "Performance Level" (PL). A new feature is the quantitative treatment of safety: The PL is a theoretical characteristic value as "average probability of a dangerous failure per hour" and it is given as PFH_d-value. This concept is not yet fully compliant with the Machinery Directive 2006/42/EC and the Guiding Standard for Risk Assessment, as explained by Steiger (2014). This paper contributes to a better understanding.

3.1 *Intuition and better approaches*

It is hard to believe, but for the time being, the normative frame for machinery safety does not contain a numerical risk model, which summarizes the single factors plausibly to an overall risk, in order to e. g. compare theoretical results with empirical field data. As a consequence, fictitious hazards and actual risks are being mashed up again and again, so that in controversial discussions in the world of safety standardization, mostly the strongest gut feeling (subjective) overtrumps logical (objective) considerations.

The controversies are mainly about numerical probabilistic representations. These were recently introduced in the field of general machinery safety, when the revised European Machinery Directive 2006/42/EC extended the hazard analysis of former versions to a risk analysis by introducing the term "probability" in the expression: "*estimate the risks, taking into account the severity of the possible injury or damage to health and the probability of its occurrence*".

Intuitively, it is immediately obvious that a separating protective device (e. g. a full enclosure, as shown in Figure. 1) is an effective means of risk reduction—both with regard to the probability of occurrence and, in particular, with regard to the extent of damage. In addition, this applies to technical causes of failure on the one hand and to human error on the other. It is also clear that separating protective devices cause a considerable risk reduction because they significantly reduce the duration of the operator's exposure to hazards, for example, not only in the case of protection against unintentional access, but also for protection against ejected parts, substances and especially in the case of fire protection. This can significantly reduce both the expected severity of an injury and, in particular, its expected frequency.

An attempt shall be made at ESREL 2018 to supplement the intuitive consideration with an additional scaling of the risk reduction effects. Therefore, the key term "probability of hazardous situations" is explained in Moedden (2018). Therein, the scaling of risk reduction is explained by means of a detailed probabilistic risk model, see Moedden (2016).

For full enclosures, a high probability can be assumed that risks from the workspace are completely controlled in cases of failure that lie below the design convention (e. g. chips from machining).

For the range of energies up to the design concept (e. g. the chuck jaw hypothesis according to

ISO 23125 (2010), see also Mewes (1999), Ising (2001)), the standard-compliant safety guards contribute considerably every day to protect the operators. Beyond the design convention, however, the probability decreases significantly, see Meister et al (2017). As a separating protective device must be opened regularly, e. g. when changing workpieces manually, during troubleshooting or when setting up the machining process, the protective effect of the separating protective device is sometimes (partly) lost, too. The consequences of technical failure and human error can then be dramatic. This applies in particular if the automatic processing is carried out under "defeating" of the protective devices (manipulation), as explained in Kesselkaul & Meyer (2016), Preusse (2005) and Meyer (2017).

3.2 *Inherent safety and diagnostic functions*

As far as the reduction of technical failures is concerned, a reliable technological function is basically the starting point (step one in Figure 2), because the more reliable a function is, the less frequent are failures and even rarer failures critical to safety. This is because the latter have all failure events to the upper set; in any case, safety-relevant failures cannot be more frequent than all failure events as a whole. Obviously, non-safety-relevant and safety-relevant failures are in a complementary relationship to each other and form the entirety of all failures, see Bornemann et al (2015). Diagnostic functions can be used to anticipate some safety-relevant failure events, so that safety-related reactions can be carried out which lead to a (safe) standstill of the machine in the event of an imminent critical failure, see Moedden (2016). If this is done in time, a potentially safety-relevant failure becomes a non-safety-relevant failure with machine downtime as a result without any hazard from the motion control of the machine tool. This applies in particular if a full enclosure is closed when a safety-relevant failure is detected in order to prevent access to the working area (e. g. in case of fire extinguishing with CO_2 flooding after detection of a fire in the working area, ISO23125(2010)). After this step two in Figure 2, only the always possible stumbling, falling etc. remain as a hazard.

The entirety of all failures can be represented as a repetition rate in defined time periods, e. g. as "probability of failure per hour", PFH. In principle, the PFH value of a technological function does not increase due to an additional diagnosis, but rather if the diagnostic function is hypersensitive and fails to perform correctly (e. g. often in case of smoke detectors in living quarters).

A high reliability of a technological function (i. e. with low PFH-values) thus contributes to the inherent safety required by the Machinery Directive 2006/42/EC and ISO 12100 (2010) as a first step towards the three-step risk reduction method, see Figure 2 (Legend: EF: Expected frequency, S: Degree of severity). The diagnostic function itself is not an inherent safety element, but an additional safety measure according to step two of the three-step method. ISO 13849-1 (2008) uses the term "safety function" for this purpose.

3.3 *Target "zero risk" is missed*

With the measures described so far, an expected frequency of safety-relevant technical failures remains, because not all of them can be shut down safely with diagnostic functions. Therefore, if a "zero-risk" acc. to Platz (2016) cannot be achieved, a certain repetition rate of safety critical failures (which can cause hazards) must be expected, e. g. as "probability of dangerous failure per hour", defined as PFH_d in ISO 13849-1(2008).

A full enclosure primarily reduces exposure to hazards. However, this is not the only risk reduction achieved by safety guards. The controllability of a hazardous situation can also be increased by separating protective devices, e. g. for partial claddings around the primary flight circle of chips, so that the operator can either choose a safe location himself, or is compelled to take up a safe position by means of location binding of the control panel.

A further protective effect of a full enclosure results from the ratio of partial load operation to full load operation. In a paper at ESREL 2016, this important "secret of success in machine tool safety" is explained, the implicit error detection in the process on the basis of a full enclosure, Moedden (2016). The probability theory indicates that most of the failures are to be expected in automatic operation, i. e. when the safety doors are closed, i. e. then they occur (nearly) without hazard.

However, there is still a probability of occurrence of a hazard event with a full enclosure in automatic mode (index "1"): $PHE_1 > 0$ because a full enclosure cannot safely hold back all conceivable failures but only up to the respective design convention, as already mentioned, see Ising (2001).

With correct dimensioning, however, the far most common of the conceivable hazards are mastered, i. e. on this side of the dimensioning convention and partly beyond it. For example, the chuck jaw hypothesis of ISO 23125 (2010) partially covers the workpieces rotating around the main axis if a chuck loss leads to symmetrical release of the workpiece, Ising (2001).

In automatic mode, the numerical value for PHE_1 is close to $PHE_1 \approx 0$ because, according to accident records, damage caused by parts released from the working area actually occurs in the basic

population of all machines acc. to Kesselkaul & Meyer (2016), but it is extremely rare in standard-compliant machines, see Moedden (2017). Meister et al (2017) made a first attempt to describe the withstand probability and the probability of failure as a probability function depending on the translational energy of fragment impacts.

In the setting mode (index "2"), a numerical value $PHE_2 > 0$ must be assumed for the risk PHE_2, since, according to accident statistics, most of the damage occurs in the basic population of all machine tools when the safety doors are open. The same also applies to all other operating modes with open safety doors, in particular for "manipulation", Kesselkaul & Meyer (2016), Preusse (2005).

3.4 *The "accident pyramid" at the boundary of a "zero risk"*

According to Moedden (2018), the equation of a technical risk reduction for a. m. cases one and two is in the interval $t_a \le t \le t_b$, (eq. 1).

$$PHE''_{1,2} = \left(1 - C_{1,2}\right) \cdot t_{\text{exp},1,2} \cdot f_{\text{exp},1,2} \cdot O_{1,2} \times \int_{t_a}^{t_b} \sum PFH_{d,\,tech,1,2} \cdot dt$$

Eq. (1) contains all risk reduction principles from the three-stage method as illustrated in Figure 2.

Nevertheless, PHE > 0 remains as a finding, i.e. even with very reliable controls ($\sum$PFHd $\approx$ 0) and standard-compliant fully enclosed enclosures, residual hazards remain. But does a PHE > 0 really mean that there is an according percentage in the basic population of severe damages? The answer is "no", because the probability of a hazard must still be associated with the severity degree of an injury.

Furthermore, in the operational environment, there is a supplementing indirect, intuitive way of avoiding injuries. It is a kind of "statistical warning principle", based on the so-called "accident pyramid", see Manuele (2011). Considered as relative frequencies in probability theory on the background of a clear cause-to-effect relation, this means that—before a serious accident occurs—a lot of "near-accidents" occur first, then the less frequent accidents with real injuries, from slight reversible over severe to fatal. The severity of an injury needs to be connected to the equation of risk reduction eq. (1), it can be roughly estimated with the "accident pyramid".

As the "near-accidents" happen comparatively more often than real accidents, the critical situations in the operational area should be largely known. The warning effect takes place only on a averaged statistical background, because a "near-accident" takes place many times, before it comes to an injury, and even more times before it comes to a serious or fatal injury. This means that experienced operators can also protect themselves intuitively by being and remaining particularly vigilant in critical areas. Inexperienced operators should therefore pay close attention to the warnings of their experienced colleagues.

3.5 *Comparison of machine tool and robot*

In order to illustrate the probabilistic risk model in Eq. (1), a calculation example shall be used to compare an automatic machine tool with full enclosure, as shown in Fig. 1, with a collaborative robot without protective housing (instead using op-tical and/or tactile proximity sensors), as in Fig. 3.

In order to make the comparison easier to understand, the estimated average values in Table 1 are compiled on the basis of "risk snapshot parameters" of typical activities; these can, of course, be adjusted on a case-by-case basis, e. g. out of evaluated video records. For statistical purposes, they can be collected in a histogram, showing the bandwidth and share of the parameters.

A manual workpiece change on the machine tool, which should last 3 minutes and be repeated 11 times per hour, thus means 33 minutes of possible hazard exposure (per hour). The robot is a support device for collaborative manual activities as shown in Figure 3, which should last 1.8 minutes and repeat 25 times p.h, i. e. 45 min. hazard exposure (p.h.).

Using an 8-hour shift, the risk reduction of a machine tool (MT) is compared with that of a robot (ROB). The safety design of the latter depend almost entirely on reliable control technology, e. g. in the case of a safety function "safe operational stop"; three superimposed movements are assumed to be stopped safely. The safety concept of an automatic machine tool, on the other hand, depends heavily on a standard-compliant,

Figure 3. Coll. robot without protective enclosure, see Kentsch (2016).

Table 1. "Risk snapshots" of a) Machine tool with full enclosure, and b) collaborating robots without protective housing.

Object/ Action/ Safety function	$\Sigma PFH_{d,tech}$ [h^{-1}]	$t_b - t_a$ [h]	O [–]	$t_{exp,1}$ [h]	$f_{exp,1}$ [h]	C [–]	PHE [–]
a) MT/ manual workpiece change/ SCW	$6.5 \cdot 10^{-6}$	8	0.5	0.05	11	0.5	$7.5 \cdot 10^{-6}$
b) ROB/ manual support/ SOS	3-times $5 \cdot 10^{-7}$	8	1	0.03	25	0.1	$8.1 \cdot 10^{-6}$

fully enclosed work area and on the presumption that it is used as intended. For the sake of simplicity, one safety function is considered at a time. The above mentioned "Safe Operation Stop" (SOS) for the robot, whereby the associated hazard can constitute an impact to the operator's head, with potentially serious injuries and for the machine tool, a "Safe Clamping of the Workpiece" (SCW). The associated hazard during manual activities with open guards could be a loss of workpiece due to gravity, with a hazard of crushing the operator's hands in the work area and possibly serious injuries.

In the right-hand column of Table 1, the probability of occurrence of a corresponding hazard based on the input parameters acc. to Eq. (1) is estimated.

The simple example in Table 1 shows that:

a. A safety function (SF) with a PL = b (i. e. PFH_d in the range of $3 \cdot 10^{-6}$ $h^{-}1$ to $1 \cdot 10^{-5}$ $h^{-}1$, here averaged: $6.5 \cdot 10^{-6}$ $h^{-}1$). This is active as a technological work function within a full enclosure of a machine tool, but it becomes a safety function, when the safety doors are open. If the SF fails, the occurrence probability is assumed to be O = 0.5 and the avoidability to C = 0.5. Concerning the risk, it can be equivalent to:
b. A safety function with PL = d (i. e. PFH_d in the range of $1 \cdot 10^{-6}$ h^{-1} to $1 \cdot 10^{-7}$ h^{-1}, here averaged: $5 \cdot 10^{-7}$ h^{-1}, for three superimposed movements) of a robot without a full enclosure. Such safety functions are used without full enclosure for collaborating robots according to ISO 10218 [27]. If the SF fails, the occurrence probability is assumed to be O = 1 and the avoidability to C = 0.1.

Corresponding to the risk model in eq. (1), for the first case in Table 1 an occurrence probability of a hazard PHEMT can be derived to $7.15 \cdot 10^{-6}$ per 8-hour shift. And for the second case and occurrence probability of a hazard PHEROB is estimated to $8.1 \cdot 10^{-6}$.

For the robot, this means that the probability of a hazard occurrence can largely be compared with that of a fully enclosed machine tool, despite the higher reliability of the safety function (PL = d). The reason why the machine tool achieves (despite a comparatively low PL = b) a safety level comparable to a robot is that in the event of a control failure, the safety guard on the machine tool can still protect the operator here (but not for the robot). *This comparison demonstrates the significant risk reduction effect of a full enclosure!*

This protective effect of a full enclosure is not sufficiently taken into account in ISO 13849-1 (2008), especially the derivation of performance level recommended (PL_{rec}). This is the current reason why the technological clamping functions of machine tools, which can also be used as safety functions when safety guards are open, are underestimated (allegedly max. PL = b, c according to ISO 13849-1 (2008)). As illustrated by Steiger (2014), this standard is not correctly based on the three-step method of risk reduction in ISO 12100 (2010). The greatly simplified approaches of ISO 13849-1 (2008) may be suitable for the situation on robots, but not for the complex interactions involved in risk reduction of safety guards of machine tools.

3.6 *Best practice at the EMO safety day 2017*

The actual implementation of the three-step method described above depends on the type of machine. However, there are also cross-cutting issues, such as the operating modes of a machine, see Steger (2017), and the equipment for fire protection, see ISO 19353 (2016). Therefore, the machine-specific measures are summarized in product safety standards (here the type C standard EN ISO 23125 (2010)).

At EMO Safety Day 2017, field data evaluations of more than 93,000,000 operating hours of turning machines (Figure. 1) without accident were presented. These results and the subsequent discussion presented by Nowizki (2016) proved that the provisions in type C standards can be applied successfully, when a sufficiently dimensioned full enclosure is available.

The "secret of success" in reducing accidents lies, as explained above, to a large extent in the massive risk reduction effects of combining a) fixed and moveable guards for the full enclosure of stationary machines (comparable to a driver's cab in mobile work machines), and b) fault detection in the process, explicitly and implic-

itly, Moedden (2016). These two factors together result in a kind of "systematic safety integrity" at the machine level. In doing so, the very high availability of machine tools should never be ignored because this contributes to a high level of inherent safety,

The proven-in-use studies of the VDW and Nowizki (2016) thus confirm that the theoretically possible probabilities of hazards, as estimated in Table 1, can actually be significantly reduced by suitable measures, e. g. the protective effect of a properly designed full enclosure. The full enclosure as shown in Figure. 1 provides a practically proven state of the art which has been defined in product standards for more than ten years on a kind of "macroscopic" level (i. e. thicknesses of protective doors and windows are actually visible and measurable). The safety measures in it follow the three-step reduction method of ISO12100 (2010) and create a significant probabilistic distance between the failure probability (e. g. of a component) and the occurrence probability of according hazards.

On the other hand, the discussion about control failures takes place merely on a kind of "microscopic" level where the PFHd-values are invisible, not measurable. However, they must be brought to the "macroscopic" level, as suggested in eq. (1), in order to plausibly connect to the probability of occurring hazards.

4 LACK OF MAINTENANCE

Another important aspect for the operational safety is the maintenance in the plants. The ageing of sight protection windows has been known for more than ten years, see Würz et al (2002), Duchstein (2010), Mewes et al (1999), Wahrlich et al (1999), Mewes (2003) and Uhlmann et al (2012), and the countermeasures are anchored in standards such as ISO 23125 (2010), see also Adler (2013), Mewes (2001). Nevertheless, the criteria for replacing embrittled or damaged protection windows—the most important component here is polycarbonate—are still being ignored by some person in charge. This can pose a considerable risk to the machine tool operators who may not be protected sufficiently against fragments of material released from the work area.

This item also includes preventive maintenance which is intensively marketed for components that are relevant to machine availability and machining quality. In doing so, a great deal of effort to ensure that replacement / maintenance is carried out within the specified time intervals so that availability and quality do not suffer.

Unfortunately, in contrast to this, the component "safety window" which directly affects the safety of the operator is often seen as a "cost driver". Can availability and quality be more important than the safety of the operators who ultimately make availability and quality possible by exposing themselves to the above restrictions?

5 SUMMARY AND OUTLOOK

Since 1996, the Institute for Machine Tools and Factory Management (IWF) at the Technische Universität of Berlin in cooperation with VDW has used impact and ageing tests to define design recommendations for separating protective devices (i. e. fixed and moveable guards with vision panels) and to supplement existing knowledge. Both, this database and the experimental investigation possibilities in Berlin are used in many other industries for the dimensioning of guards in order to close gaps in the existing standardization. In addition, current research projects have considerably expanded the experimental equipment at the IWF in Berlin, see Prasol et al (2017), Meister et al (2014). And close scientific relations exist with comparable research institutes in Italy, see Landi et al (2017), and Japan, see Yui (2017). The longstanding cooperation of VDW and IWF has led to the fact that the German machine tool industry has a high level of safety ass proven by Nowizki et al (2016). This can be seen before all other factors due to the low number of accidents resulting from machine-related faults, see Kesselkaul & Meyer (2016), Moedden (2017).

Nevertheless, there is still a need for further research and knowledge transfer. On the one hand, this is aimed at the machine tools users, since these are the innovation drivers and define the requirements for the manufacturers. On the other hand, machine tool manufacturers should also be mentioned as a further target group, as they should also be made more aware of this issue. In particular, it is of interest to convey that safety does not in principle go hand in hand with an increase in costs. As an example, it is referred to the dimensioning of separating guards on grinding machines, see Adler et al (2017). In the area of marketing of safe machine tools, it would be possible to use safety as a "Unique Selling Proposition (USP)" similar to the "Blue Competence" label, in order to distinguish itself more strongly from competitors. This has been successfully practiced in the automobile industry for a long time with reference to the different consumer motivations and also of people who purchase industrial goods, see Duchstein (2011).

This composition is also an appeal to a professional use of the term "probability". An unprofessional application of the term leads to arbitrary interpretation which is not an option in effective safety engineering. The aim is always to reduce the number of accidents that are recorded annually by the strong effect of the "Pareto principle". These represent the sum of all probabilities of occurrence of hazards, with reference to a yearly period for the respective basic population. In Moedden (2017) the situation for machine tools in Germany is explained in detail; a transfer to other machine types and other countries is possible.

6 CONCLUSION

The current safety standards are difficult to understand and partly contradictory, especially when it comes to the use of the term "probability". Although, taken together, they contain a vast number of obligations, there is no common risk model in them. For effective risk reduction, however, a logically structured risk model is indispensable, which should also be probabilistically scalable. For this purpose, further proposals are presented here, in particular for a quantification of the significant risk reduction effect of full enclosures. Also, in order to put the following two previously not logically connected representations in relation to each other: a) a dimensionless probability (e. g. of the Machinery Directive and ISO 12100 (2010)), and b) a probability per time unit, in the context of control functions often indicated as PFH_d-value (e. g. in ISO 13849-1 (2008)).

Theoretically and empirically, it could be demonstrated that fixed and moveable guards make a significant contribution to reducing the probability of occurrence of hazard events. However, they cannot completely reduce the probability of hazard occurrence to zero for two reasons:

1. Fixed and moveable guards need to be opened for certain activities, see Moedden (2018).
2. In the closed state, they have only a defined retention capacity against released parts in the working area, see Meister (2017).

In order not to lose the protective effect of separating protective devices, two appeals to the operators remain important:

1. To keep the boundaries the intended use of the machines, such that manipulation of full enclosures does not happen.
2. Preventive maintenance of embrittled polycarbonate vision panels must be carried out carefully.

SYMBOLS AND OTHER DEFINITIONS

Symbol	Dim.	Meaning
PFH_{tech}, ($PFH_{d,tech}$)	h^{-1}	Acronym stands for "probability of (dangerous) technical failure per hour". It is shown in [11] that this is an average density function.
PHE	h^{-1}	Acronym stands for "probability of a hazardous event (technical and human sources)". In Eq. (1) the association with PFHd, tech value is shown.
Parameter O	-	Probability of occurrence of hazard events
t_{exp}	sec	Exposure time, expressed also as an intervall: $t_a \leq t \leq t_b$
f_{exp}	h^{-1}	Repetition rate of an exposure time
$F_{rel} = t_{exp} \cdot f_{exp}$	–	Relative share of exposure
A_v, C	–	Avoidability, controllability

REFERENCES

Adler, C.; Mewes, D.; Herbst, P. (2013): *Dimensionierung von trennenden Schutzeinrichtungen an Schleifmaschinen*. In: 4th European Conference on Grinding, 27.-28.11.2013, Bremen.

Baruch; Fischhoff; Kadvany, (2011): *"Risk – A Very Short Introduction"*, Paperback, Oxford University Press 2011.

Bornemann, A., Y. Froese, L. Landi, H. Mödden, (2014): *Probabilities in safety of machinery-Part 1: Risk profiling and farmer matrix*, Safety and reliability Methodology and Application CRC Press/Balkema, 1933–1942, European Safety and Reliability Conference, ESREL 2014, Wroclaw; Poland; 14–18 September 2014.

DIN EN ISO 19353 (2016), Anforderungen zum Brandschutz an Maschinen, Beuth Verlag.

Duchstein, B. (2010): *Aufprallprüfungen an definiert gealterten Polycarbonat-Sichtscheiben,* Mitteilungen aus dem Produktionstechnischen Zentrum Berlin, FUTUR 01/2010, Aktuelles aus Forschung und Entwicklung.

Duchstein, B. (2011): Wahrnehmung der Sicherheit als "Unique Selling Propostion (USP) im Vergleich zwischen Automobil- und Werkzeugmaschinenbau, Studienarbeit von Nadine Poblenz of Technische Universität Berlin.

EN ISO 12100, (2010). Safety of machinery – General principles for design – Risk assessment and risk reduction. Berlin, Germany; Beuth Verlag GmbH.

EN ISO 13849-1, (2008). Safety of machinery – Safety-related parts of control systems – Part 1: General principles for design. Germany; Beuth Verlag GmbH.

EN ISO 14120, (2002) (Revision in ISO/FDIS 14120:2015) Sicherheit von Maschinen - Trennende Schutzeinrichtungen - Allgemeine Anforderungen an Gestaltung, Bau und Auswahl von feststehenden und beweglichen trennenden Schutzeinrichtungen, Berlin.

EN ISO 23125, (2010), *Machine tools safety – Turning machines* Berlin, Germany; Beuth.

Haigh, J. (2012). Probability – A Very Short Introduction. Oxford: Oxford University Press.

Industrieroboter—Sicherheitsanforderungen—Teil 1: Roboter (ISO 10218-1:2011); Deutsche Fassung EN ISO 10218-1:2011, Berlin, Beuth Verlag.

Ising, M. (2001): *System zur sicherheitsgerechten Konstruktion von Werkzeugmaschinen.* Dissertation. Institut für Werkzeugmaschinen und Fabrikbetrieb, Technische Universität Berlin: IWF TUB, 2001.

Kentsch, S. (2016): Bosch zeigt berührungslos kollaborierende Roboter für die flexible Fertigung, 12.09.2016 Pressemeldung Industrie 4.0.

Kesselkaul, R., Meyer, C. (2016): Priorisierung des Handlungsbedarfs (Prioritizing the Need for Action - Focus of Accident Occ.). VDW-Fair 2016. Düsseldorf.

Landi, L.; Uhlmann, E.; Meister, F.; Pera, F.; Mödden, H. (2017): Probabilities in Safety of Machinery - Risk reduction through fixed and moveable guards by standardized impact tests, part 2: Possible improvements with FE impact simulations, European Safety and Reliability Conference (ESREL 2017), Portorož, Slovenia.

Landi, L.; Uhlmann, E.; Meister, F.; Pera, F.; Mödden, H. (2017): Probabilities in Safety of Machinery - Risk reduction through fixed and moveable guards by standardized impact tests, part 1: Applications and consideration of random effects, European Safety and Reliability Conference (ESREL 2017), Portorož, Slovenia.

Manuele, F.A. (2011): *Reviewing Heinrich Dislodging Two Myths From the Practice of Safety*, Professional Safety OCTOBER 2011, www.asse.org, New York.

Meister, F. et al (2017): Probabilities in Safety of Machinery - Hidden Random Effects for the Dimensioning of Fixed and Moveable Guards, 15th International Probabilistic Workshop 2017 in Dresden.

Mewes, D. (2001): *Aufprallfestigkeit von Werkstoffen für trennende Schutzeinrichtungen an Fräsmaschinen und Bearbeitungszentren.* Sicherheitstechnisches Informations- und Arbeitsblatt 330 620, Hrsg.: Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung IFA, Sankt Augustin.

Mewes, D. (2011) *Alterung von Polycarbonat-Sichtscheiben an Werkzeugmaschinen.* Kennzahl 330 630. In: IFA-Handbuch Sicherheit und Gesundheitsschutz am Arbeitsplatz, Lfg. 1/11, V/2011. Hrsg.: Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung IFA, Sankt Augustin. Erich Schmidt, Bielefeld 2003 Losebl. Ausg.

Mewes, D. et al (1999): Rückhaltefähigkeit von Polycarbonatscheiben nach betrieblichem Einsatz, BIA, Nr.:1999 23452.

Mewes, D. et al. (1999): Trennende Schutzeinrichtungen an Werkzeugmaschinen, „Werkstatt und Betrieb" (Heft Nr. 9/97); in der „WT Werkstattstechnik" (10/99).

Mewes, D.; Trapp, R.-P.; Warlich, H.-J. (1996) Trennende Schutzeinrichtungen Die Aufprallfestigkeit von Werkstoffen prüfen und beurteilen. Materialprüfung 38 (1996) Nr. 9, S. 368/372.

Mewes, D.; Trapp, R.-P.; Warlich, H.-J. (1998): *Festigkeit von Werkstoffen bei Aufprallbeanspruchungen.* Materialwissenschaft und Werkstofftechnik 29 (1998) Nr. 9, S. 258/262.

Meyer, C. (2017): *Process observation operating mode: blessing or curse?* KANMAIL 2/2017, Kommission Arbeitsschutz und Normung.

Mödden, H. (2017): Probabilities in Safety of Machinery: Borel-Cantelli Lemmas lead to a Prevention Dogma based on the Pareto principle, 15th International Probabilistic Workshop 2017 in Dresden.

Moedden, H. (2015): Probabilities in Safety of Machinery – Elements of a Risk Model and Comparison with Field Data, VDW, Frankfurt a. M., Germany, ESREL 2015, Zurich Switzerland.

Moedden, H. (2016): Probabilities in Safety of Machinery – Risk Reduction Effects by Combination of Full Enclosure and Fault Detection in the Process, ESREL 2016, Glasgow, Scotland.

Moedden, H. (2018): Probabilities in safety of machinery - Markov model for the scaling of risk reduction effects due to limited exposure. 2018 Trondheim, Norway

Nowizki, N.; P. Zeiler; B. Bertsche, (2016), Institute of Machine Components, Univ. of Stuttgart, Germany, H. Moedden, VDW Germany: *Statistical Analysis of Field Data for a Proven-In-Use Assessment in the Machine Tool Industry according to ISO 13849*, ESREL 2016.

Parliament of the United Kingdom. Health and Safety at Work etc. Act (1974); 194. London.

Platz, A. et al. (2016): *BGHM-Aktuell, Magazin für sicheres & gesundes Arbeiten,* 5/2016, Berufsgenossenschaft Holz und Metall Mainz, ISSN 1612-5428.

Preusse, C. (2005). Manipulation an Schutzeinrichtungen an Maschinen, St. Augustin: HVBG.

Steger, P. (2017): *Bericht zu Betriebsarten,* Aufsatz der Fa. Grob im EMO 2017 Vorlauf, Mindelheim.

Steiger G. (2014), Funktionale Sicherheit aus der Sicht der Normanwender im Maschinenbau, *Maschinenrichtlinie aktuell Heft 4 / 2014*.

Taschner, R. (2013): Die Zahl, die aus der Kälte kam: Wenn Mathematik zum Abenteuer wird, Hanser, Wien.

The European parliament: Machinery Directive 2006/42/EC, (2006), Brussels.

Uhlmann, E., Meister, F. (2014): Untersuchung vom Alterung von Maschinenschutzscheiben – Die Alterung aufhalten, Werkstatt und Betrieb Juli-August 2014.

Uhlmann, E., Prasol, L.; Meister, F.; Adler, C. (2017): *Safety of Grinding Machines: Secret of the Success of ISO 16089 and Residual Risk.* Presentation at EMO Safety Day 2017 Hannover, Germany.

Uhlmann, E.; Bell, T.; Duchstein, B.; Meister, F.; Mewis, J.; Mödden, H. (2012): *Sicherheit an Werkzeugmaschinen - Die Bedeutung der trennenden Schutzeinrichtung für die Risikoreduktion an Werkzeugmaschinen.* Mitteilungen aus dem Produktionstechnischen Zentrum Berlin, ZWF Jahrgang 107/2012, Aktuelles aus Forschung und Entwicklung.

Warlich, H.-J. et al. (1999) Position Einzelfallbeurteilung von der BG, Titel: Alterung von Polycarbonatscheiben (Eingang beim VDW am 9. Dez. 99), BG Mainz, Fachauschuss Eisen und Metall II.

Würz, T.; Kuhnmünch, K.-P.; Mödden, H. (2002): *Polycarbonat-Sichtscheiben in Werkzeugmaschinen,* Verein Deutscher Werkzeugmaschinenfabriken e.V. (VDW).

Yui, A.; Fukui, T.; Kitajima, T. (2017): Study on protection performance of grinding wheel safety guard against the soft and brittle abrasive projectile. Euspen's 17th International Conference & Exhibition, 19.06.2017, Hannover.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Analysis of fatal fires in Norway over a decade, – a retrospective observational study

C. Sesseng, K. Storesund & A. Steen-Hansen
RISE Fire Research AS, Trondheim, Norway

ABSTRACT: Five-hundred-and-seventy-one fatalities were registered in the official fire statistics in Norway between 2005–2014. However, little is known about the victims. This study collected information from several sources to build a holistic database and gain more knowledge about the technical and social aspects of the incidents, forming a basis for more targeted measures. Human behaviour greatly affects the risk of fire, which supports why social aspects of incidents should be considered when identifying risk factors associated with the victims. The results showed a clear distinction between victims above and below the age of 67 with respect to risk factors. For the elderly, reduced mobility, impaired cognitive ability, mental disorders and smoking were observed risk factors. For the younger victims known substance abuse, mental illness, alcoholic influence and smoking were observed, mostly in combination. This shows that fire is a social problem, and should be prevented by initiating customised measures.

1 INTRODUCTION

In this study, information from fire statistics of the Norwegian Directorate of Civil Protection (DSB) and other sources has been analysed to gain more detailed knowledge than before about who dies in fires and why. This will help to implement more targeted measures in order to reduce the number of people perishing in fires. Results from previous studies in Norway and other countries have provided important data in designing the study, in the interpretation of results, and comparisons of the evolution of fatal fires over time. Demographic and cultural differences over time and between countries will amongst others impact on attitudes and risk behaviour as concern fires. Variations in the employment of various preventive measures will also be reflected in statistics on fatal fires. Factors that may impact on the proportion of registered victims in fatal fires are the methods of data collection, different definitions of fire fatalities and to what extent fires are investigated, plus the way in which fires are registered, and the thoroughness in which the consequences of a fire are followed up afterwards.

2 METHOD AND ANALYSES

2.1 *Data collection*

2.1.1 *Sample*

This study surveys persons having died in fires in Norway during the time period of 2005–2014. According to DSB's fire statistics 517 fatal fires with 571 fatalities occurred during this period.

2.1.2 *Sources*

Our data material consisted of DSB's fire statistics, police investigation reports, the fire victim's medical records, and the Norwegian Cause of Death Registry (NCoDR) from the Norwegian Institute for Public Health. Generally, the police investigation reports also include post-mortem reports.

In order to obtain access to these sources, authorisations were applied for and granted from several authorities.

DSB's statistics provided the basis for which police investigation reports we requested access to. Hence, cases not listed in these statistics are not included in the study.

By means of the police reports the fatalities were identified by name and national identity number. This formed the basis for which medical records we asked to access, and for which persons we requested data from NCoDR. The medical records are not accessible through a common register, but each medical record resides with the doctor who was general practitioner/family doctor at the time of death. This meant that the Norwegian Directorate of Health (NDE) needed to provide us with a key to connect each fatality to the GP/family doctor at the time of death.

2.1.3 *Data registration and categorization*

To register data from the various sources, a database was set up to handle the fire data and to

link data on fire victims. DSB's fire statistics were imported directly into the database, and were used as a basis for the requests to access investigation reports sent to the police districts. Additionally, variables to handle relevant data were added.

To ensure consistent extraction and storage of data from the various cases between project collaborators, an electronic form was prepared which was completed for each incident. Some of the input was pre-defined multi-choice categories, other was free text input. This made interpretations more objective, and it became easier to quantify qualitative data.

Extraction of data from NCoDR, which contained information regarding cause of death, was also made. Of all the identified persons there were 29 with an incomplete ID-number (lack of information in the police reports). Sixteen of these were found in NCoDR by searching their name. The remaining 13 persons were foreigners without a Norwegian ID-number or D-number (ID number for temporary residents in Norway) who had died in Norway. The extraction from NCoDR was also imported into our database and linked to each separate person. Of the NCoDR data we got access to, the categories *underlying cause of death* and *injury code* were of largest interest.

2.2 *Statistical analysis*

All data registered in the database was exported to the statistics program Statistica, version 12 (Dell Inc 2015).

Statistical tests were conducted to test and examine apparent differences between sub-groups in the population. In all essentials non-parametric tests were employed, such as Mann-Whitney U-test, Fisher exact-test, chi-square test and regression analysis. For all analyses a significance limit of $p \leq 0.05$ was employed. A p-value between $0.10 \geq p > 0.05$ is considered as a trend.

Multiple Correspondence Analysis was carried out in order to identify whether the various characteristics of the victims frequently occurred simultaneously. This includes e.g. examining whether victims with an established substance abuse also tend to be smokers. This method cannot quantify any similarities or dissimilarities, but it may provide a qualitative impression and give a basis for further analyses.

3 RESULTS

3.1 *Registration of fatalities*

DSB's fire statistics contain some data on the fire itself, as well as the number of injured persons and the gender and age of the fire causalities. Requests for access to police reports were made for all these fires, and we received 347 police reports (68% response) that were reviewed. Data from these reports were combined with data from DSB's statistics. The cases were geographically dispersed across the entire country, although we did not receive any police reports for three out of 19 counties in Norway.

From the received police reports 387 fatalities were registered. The vast majority of these were identified through their ID-number, and a request for access to the medical records was sent to the respective GP/family doctor at the time of death. A small number lacked ID-number data and were consequently excluded from the analysis of medical records. There were also a number of cases where only the name and date of birth were stated, which often sufficed to find the medical record of the person in question. We received 248 medical records, which represent 64% of the identified victims.

3.2 *The fatalities*

3.2.1 *Sample description*

Table 1 shows the distribution of various characteristics of the fatalities in fires during the 2005–2014 period.

As concerns the category «non-native speaker», it is meant to include persons who are assumed not to communicate adequately in a Nordic language or English, and with whom it will be difficult to communicate relevant information about fire safety. According to a rough estimate based on Norwegian statistics on immigration, there were 5–6% non-native speakers (to a varying extent) in Norway in 2014 (Sesseng et al., 2017).

3.2.2 *Age and gender*

The persons who perished were in all essentials well into their adulthood. Half of the fatalities had an age between 44–78 years, see Figure 1 showing the age distribution of the fatalities.

When taking the number of people in each age group into consideration, one sees there is almost an exponential connection between age and the number of fire fatalities, see Figure 2.

Table 1. Sample description.

Gender	Male	Female		n
	56.1%	43.9%		387
Age	Median	Interquartile range	Min–Max	n
[years]	59	44–78	1–97	386
Non-native speaker	No	Yes	Unknown	n
	88.6%	7.8%	3.6%	387

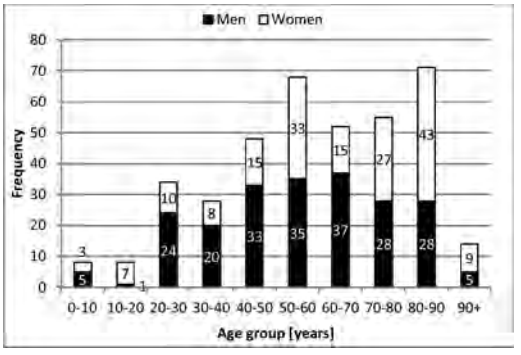


Figure 1. Age distributed on gender for fatalities, n = 386.

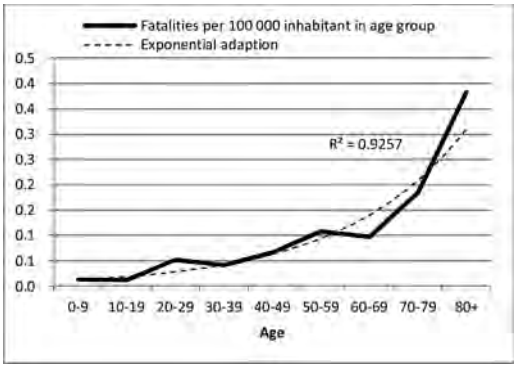


Figure 2. The graph shows the ratio between number of fatalities and number of inhabitants in each age group for fatal fires during the 2005–2014 period.

During the 2005–2014 period 56% of fatalities were men. Men are overrepresented in all age groups under 70 years, also taken into account the gender distribution in these age groups.

According to Statistics Norway, in 2007 there were about as many men as women around the age of 63 in Norway. After this age the proportion of women increased, and with age the surplus of women continued to rise. Among persons over the age of 80 around two thirds were women, and among those over 90 years three fourths were women (Falnes-Dalheim & Slaastad 2007).

After the age of 80, more women than men perish in fires (approx. 60% of the fatalities). However, taken the gender distribution in the age groups into consideration there is an equal risk for both genders (Statistics Norway 2017).

3.2.3 *Risk factors*

Table 2 shows the registered risk factors linked to the fire fatalities in our selection. The results show

Table 2. Registered risk factors related to the fatalities. The category «unknown» means that the medical records hold no data. Light grey cells mark a high proportion of observations of risk factor for persons above the age of 67, while dark grey cells mark the same for persons below the age of 67.

Vision	Normal	Visually impaired	Blind	n
All	86.8%	13.2%	0.0%	257
<67 years	92.4%	7.6%	0.0%	145
≥67 years	79.5%	20.5%	0.0%	112
Hearing	**Normal**	**Hearing impaired**	**Deaf**	**n**
All	89.9%	10.1%	0.0%	257
<67 years	95.2%	4.8%	0.0%	145
≥67 years	83.0%	17.0%	0.0%	112
Reduced mobility	**Normal**	**Reduced**	**Immobile**	**n**
All	69.2%	27.4%	3.4%	266
<67 years	84.6%	12.1%	3.4%	149
≥67 years	49.6%	47.0%	3.4%	117
Impaired cognitive abilities	**No**	**Yes**	**Unknown**	**n**
All	16.0%	18.7%	65.3%	262
<67 years	20.7%	7.6%	71.7%	145
≥67 years	10.3%	32.5%	57.3%	117
Known substance abuse	**No**	**Yes**	**Unknown**	**n**
All	9.1%	36.5%	54.4%	263
<67 years	9.2%	54.0%	46.1%	152
≥67 years	9.0%	25.2%	64.8%	111
Mental illness	**No**	**Yes**	**Unknown**	**n**
All	6.5%	44.3%	49.2%	262
<67 years	6.6%	51.7%	41.7%	151
≥67 years	6.3%	34.2%	59.5%	111
Alcoholic influence	**No**	**Yes**	**Unknown**	**n**
All	38.9%	41.2%	19.9%	386
<67 years	28.4%	59.0%	12.7%	229
≥67 years	54.1%	15.3%	30.6%	157
Women	50.0%	27.6%	22.4%	170
Men	30.1%	51.9%	18.1%	216
Smoker	**No**	**Yes**	**Unknown**	**n**
All	9.3%	34.6%	56.1%	387
<67 years	6.6%	35.8%	57.6%	229
≥67 years	13.4%	32.5%	54.1%	157

that the majority of the victims had normal vision as well as hearing. On the other hand, we see that half of the victims at pension age had reduced mobility. Further, one third of the same age group had impaired cognitive abilities, and an equally large share suffered from mental illness. For the younger age group we see that half of the victims had a reputation for substance abuse. Equally many suffered from mental illness, and an equal proportion was under the influence of alcohol during the fire.

When examining the fatalities confirmed being under the influence of alcohol compared to those not being under the influence at time of death, regardless of age, we find no difference as to the time of day they perished when dividing the day into four 6-hour periods (00–06, 06–12, 12–18 and 18–24) ($p = 0.147$). Even though there is no statistically significant difference between morning/day and night (06–18 and 18–06), there is a trend suggesting that more fatalities were intoxicated at night time ($p = 0.052$).

There is a significant difference between women and men when it comes to being under the influence of alcohol during the fire ($p = 0.000$). When disregarding the cases where it was not possible to neither prove nor disprove that there was alcohol in body fluids, one sees that around two thirds of all male fatalities were under alcoholic influence. The case is the opposite for women, where one third was under the influence of alcohol.

There is no significant difference in the distribution of the cause of fires in fires where the victim was under alcoholic influence compared with fires where the victims were not under alcoholic influence ($p = 0.433$, the rarest categories were excluded from the analysis).

To conduct a Multiple Correspondence Analysis the population was divided into two sub-groups; persons with age < 67 years, and age ≥ 67 years (Norwegian retirement age). Further, analyses were made of factors representing somatic ailments and factors relating to intoxication, psychiatry and lifestyle. The reason why psychiatry is placed in the same category as substance abuse, smoking and alcoholic influence during the fire, is because psychiatric cases may be triggered by drug abuse, and therefore drug abuse and psychiatry occur simultaneously in many cases.

For the group with age ≥ 67 years the material shows that persons with mental illness also frequently are smokers. Moreover, we have seen that being under the influence of alcohol during a fire does not appear systematically in combination with other factors (no pattern). As regards somatic ailments we do not see any evident patterns. This may signify that *combinations* of e.g. reduced mobility and vision impairment are not overrepresented in the fatal fire statistics for this group. Nevertheless, 61% of this group have *either* impaired vision, hearing, or mobility, which may have contributed to the death.

For the group of fatalities below the age of 67 our data material shows that factors «known substance abuse», «alcoholic influence during fire», «mental illness» and «smoking» often occur together. Only 13% of the fatalities in this age group had none of the mentioned factors, while the majority (66%) had various combinations of two or more factors.

As concerns somatic functional impairment no pattern is identified, except from the fact that factors normal mobility, normal vision, and non-reduced cognitive ability often occur together.

3.2.4 *Cause of death*

Figure 3 shows the distribution of cause of death for the fatalities in the sample. Asphyxiation is the chief cause (57%), followed by burns (15%). In addition to that, a combination of asphyxiation and burns was concluded in 10% of cases. In 13% of the cases the cause of death is unknown, which may be attributed to the fact that the victim was so heavily burnt that it was difficult to carry out an examination or draw some conclusions.

Information about the underlying incident and causes of death was collected from NCoDR. The underlying incident says something about the incident leading to the person's death (e.g. exposure to open flame), while the causes say something about the actual cause of death (e.g. burn).

The data showed that exposure to open flame stands for 88.4% of all incidents, and that 5.8%

Figure 3. Cause of death for fatalities in fires during the 2005–2014 period, N = 387.

of incidents involved intentional self-harm (suicide). Correspondingly, it was seen that one form of asphyxiation or other stands for 74.7% of the causes of death, while some form of burn stands for 21.6% of the causes of death. The asphyxiation rate in NCoDR (74.7%) is somewhat higher than what was registered from post-mortem reports in our selection (57%). By adding the fatalities where it was concluded that the cause of death was a combination of fire injury and asphyxiation, one gets a rate of (67%), which is still somewhat lower than in NCoDR. It therefore seems to be an inconsistency between the post-mortem reports and what is being reported to NCoDR.

4 DISCUSSION

4.1 *Cause of death*

The distribution of causes of death for fire fatalities in Norway are in stark contrast to the corresponding US numbers for 2012–2014 where e.g. asphyxiation alone stood for 37% of the causes of death (DHS 2016a). We are unaware of the reason for these differences, but an assumption is that the causes of death are defined and registered differently in the US compared to Norway. This emphasizes the uncertainty when making such comparisons between countries.

The content in fire smoke is complex, consisting of a number of components which have various effects on humans. Asphyxiation may occur at an early stage in the course of a fire, even before the fire has developed enough to have a visible flame. In order to prevent deaths caused by asphyxiation it is therefore important to focus on preventing fires from arising in the first place, and to ensure that outbreaks of fire are detected and extinguished as early as possible. Materials with a high combustibility should be avoided, and an adequate control of potential ignition sources should be in place.

4.2 *Number of persons involved and fire outcome*

From the police reports we retrieved data on the number of people present in the building or living unit when the fire started and the number of persons perished or injured. The results, presented in Table 3, show that in the majority of the fires only one person was present at start of fire, and that in over nine of ten there was only one fatality, and in an almost equal proportion there were no injuries in addition to the fatality.

4.3 *Risk factors*

For the purpose of the study the population is divided into different sub-populations, e.g. victims below the age of 67 and victims of 67 years or more. This makes it possible to identify the factors that often recur for these groups. These results should be used by home care services, the GP/family doctor and other agencies who are in contact with the inhabitants of the municipality, and who provide care or fire prevention measures, in order to implement the right actions for the individual. The subsequent sub-chapters deal with the various risk factors.

Table 3. Distribution of number of persons present at fire start, number of fatalities and injured.

Present at start	1 pers.	2 pers.	3 pers.	≥4 persons	n
Fatalities	71.1%	15.2%	5.0%	8.7%	342
	1	2	3	≥4.	n
	92.6%	5.7%	1.2%	0.5%	513
Number of injured	0 pers.	1 pers.	2 pers.	≥3 pers.	n
	86.1%	8.1%	3.5%	2.3%	347
Where victims were found	Point of origin	Neighbouring room of origin	Other room in living unit	Outside living unit	n
	40.1%	11,3%	42.7%	5.9%	354

4.3.1 *Age and health*

During the 2005–2014 period seven children aged 0–7 years died in a fire, which constitutes 2.2% of the fatalities. Six of the children died in fires occurring during the night while the family was asleep. In comparison 15.1% of all fatalities during the 1970–1979 period were children aged 0–7 years, and for the 1990 to 1992 period the rate was 4.5% (Storesund 2013). This means it can be confirmed that the number of children perishing in fires is comparatively low and that it has decreased since the 1970s. Childcare in Norway has undergone major changes since the 1970s. Children spend less time on their own and more time in buildings with high fire safety standards (schools and kindergartens), which altogether probably has contributed to improving fire safety, thereby reducing the risk of children perishing in a fire.

Studies establish that elderly people are more exposed to perishing in a fire than younger persons, and we have found the same in our study (DHS 2016a, 2016b, Xiong et al. 2015). Half of the fatalities are 59 years old or older. When taking into account the number of individuals within the various age groups, one finds an exponential increase in the number of fatalities with increasing

age. The increase is particularly noticeable for age groups 70–79 years and 80+. E.g., the latter group has a 9.3 times higher probability of perishing in a fire than persons between 30–39 years according to our material. DSB's study of fatal fires during the 1986–2009 period showed that elderly people over the age of 70 had a probability of perishing in a fire that was around 4 times higher than the population as a whole (DSB 2010). Figures from the US show that elderly people (age 85+) had a 4.1 times higher probability of perishing in a fire than the population at large. The corresponding figure from our study is 4.9 times, which must be said to be comparable.

With an anticipated population growth as well as a growth in the proportion of elderly people (67+), it will be important to target preventive measures toward this group in society, in particular in view of the fact that more of them will live longer in their own homes than before, and often also alone. Until date the number of places in Norwegian nursing homes and similar in recent years have remained rather stable (Kristiansen 2016). Nursing and care services have changed from being provided in nursing homes and similar to being provided to a larger degree at home, while places in nursing homes mainly are being offered to the most sickly of elderly people. There is reason to believe that this trend will continue (Steen-Hansen et al., 2010).

For those who have reached retirement age we principally see four recurring risk factors: reduced mobility (47%), impaired cognitive ability (the elderly people who fell into this category often had Alzheimer's disease or other forms of dementia) (33%), mental illness (34%) and smoking (33%). However, no recurring patterns of combinations of various risk factors were identified, except mental illness and smoking. The report entitled *Correct measures in the right place* (English translation) does not define age as a risk factor in itself, but argues that age may involve physical and cognitive challenges (Storesund et al. 2015, Gjøsund et al. 2016, Storesund et al. 2016). The thought behind such an approach is that one should not see elderly people as a homogeneous group facing the same challenges, but rather implement safety measures meeting individual challenges. The results of this study of fatal fires support this perspective. The fact that these analyses group persons having reached pensionable age in a separate group does not signify that one must or should introduce measures targeting all elderly people, but one should rather be particularly attentive to the risk factors that stand out for this age group.

For those below retirement age the most conspicuous risk factors are known substance abuse (54%), mental illness (52%), alcoholic influence (59%) and smoking (36%) It was also found that these risk factors often occur in combination with each other. This shows that some of those who perish in a fire in this age group carry several risk factors, which increases the risk of perishing in a fire. Actually, 87% of the fatalities in this group have one or more of the mentioned risk factors. Without having any basis for asserting it in this study, it cannot be excluded that these factors may affect their independent living skills, and consequently also the risk of a fire occurring in their living unit.

4.3.2 *Gender*

More men than women perish in fires. However, the gender distribution varies for the different age groups. In age groups 20–49 and 60–69 there are around twice as many men perishing than women, while there is an overweight of women in age group 80+ and between 10 and 19 years. The variations between women and men for some age groups may be explained by population distribution and risk behaviour.

The variations also show us that being male is not a risk factor. There is still a high proportion of women who perish in fires, and prevention measures must be directed toward this group on an equal footing with men.

Also for the 80+ age group the picture is more nuanced when taking the population into consideration. Sixty-two percent of the fatalities in age group 80+ are women. Figures from Statistics Norway show that the proportion of women in this age group is 62.3% (Statistics Norway 2017). However, these figures are from 2017, which is a few years after the focus period of this study. However, from what we know there is nothing to suggest that the gender composition has changed considerably during these years. Thus one may conclude that women in this age group probably do not carry a heightened risk of dying in a fire than men.

4.3.3 *Smoking*

Thirty-five percent of the fatalities were confirmed smokers (56% unknown), where 57% of the smokers were men and 43% women. Approximately 13% of fires were caused by smoking. In comparison, figures from Statistics Norway show that the rate of daily smokers in Norway has been reduced from around 25% to 13% during the period we studied (Kristiansen 2016). Since the share of smokers in fatal fires is a lot higher than in society in general, it appears that smoking is a risk factor.

4.3.4 *Influence of alcohol and intoxication*

The fact that known substance abuse and mental illness are relatively common occurring factors in fatal fires is not surprising, but coincides with previous studies made in Norway and abroad. Studies show that fatal fires more often than other fires are caused by the victim himself/herself; in our selection the fire was directly caused by the

victim in almost 40% of the fires. This points to the importance of being able to prevent and handle a fire, and it shows that the probability of fatal fire increases when these abilities are limited. Substance abuse and some mental illnesses may lead to impaired judgment, both as concerns the incidents leading up to fire start, and in connection with the fire itself.

In around 40% of the fatalities the post-mortem examination identified traces of alcohol in the body fluids, while no alcohol was found in an equally large proportion. As concerns the last 20% this is uncertain, as it was not reported in the post-mortem reports, or the post-mortem report was unavailable.

The figures also show that there are variations between women and men. When considering men only, one finds that the figures are not dissimilar to previously reported findings (Skaar 2013). There is, however, a larger discrepancy where women are concerned compared to the same study. It was reported that only 20% of women were under influence of alcohol during the fire, while our figures show a proportion almost twice as big (36%).

An explanation to this discrepancy may be differences in focus periods between the two studies, which were 1993–2008 and 2005–2014, respectively. If this is the explanation, it may point to a change in the drinking pattern of women (frequency and volume) over the period. However, we did not examine this aspect any further.

4.3.5 *Single persons*

In over 70% of the fires in our data basis there was only one person present at fire start. Forty percent of the victims were found in the room of origin, which suggests that the victims had few opportunities of handling the fire and escaping at an early stage of the fire.

These figures show that there is an inherent risk associated to *being alone*. The likelihood of the fire being detected in time to survive is reduced, and it is harder to escape if one is alone at the start of fire. Persons living alone are probably more often alone in the living unit than persons living with others. This means that there is probably an indirect, increased risk associated with living alone.

4.3.6 *Culture, attitudes and language*

Attitudes to fire safety impact on the likelihood of an occurrence of fire. Attitudes may be related to risk behaviour, the willingness to use preventive equipment, and tidiness and maintenance. In our study we registered whether the victim was a non-native speaker, which was not common. The background was amongst other to examine whether the ability to read and communicate in the Norwegian language had any large impact on the risk of perishing in fire. We did not register any detectable connection between being a non-native speaker and the risk of perishing in a fire.

Examination of socio-economic factors (income, education, ethnicity, profession, etc.) was not a part of this study.

4.4 *Social circumstances*

Social problems are described as challenges in the relationship between the individual and the society. The term embraces the individual's problems caused by societal circumstances and the individual's challenges in adapting to societal structures and norms. The causes of social problems are many and often interdependent, but factors such as poor working conditions, lack of employment and poor housing conditions are mentioned as examples. In medical context, it is said that social problems comprise a bodily, a mental and a social component, which all interact (Bruusgaard 2017).

The dominating risk factors in our material all fall in under social problems in the medical context, as they either affect the individual's bodily, mental or social functions, or a combination of these.

As discussed and exemplified by (Storesund et al. 2015), it is therefore of key importance to not only focus on the physical circumstances of the dwelling when initiating fire preventive measures, but also consider the individual's specific challenges and needs (bodily and mental) and the individual's social circumstances. E.g. a person with impaired hearing will have other challenges regarding preventing, perceiving and escaping a fire than a person with impaired vision. Further, a person with reduced cognitive abilities would have other challenges.

By conducting a thorough risk analysis, the individual's needs could be identified and suitable organisational and/or technical measures could be initiated.

4.5 *Critique of methodology*

The police reports vary when it comes to richness in detail, which may impact on our analyses. In some cases we see that investigation reports are highly inadequate, while in other cases they are exhaustive. This makes it difficult to draw categorical conclusions as regards some factors. We have taken care not to colour the information with our own interpretations, but have in cases where some factors are not mentioned labelled them «unknown» to underline the lack of data. In a few number of cases where the police concluded with an unknown cause of fire, but where we, based on a professional assessment, believe that one cause of fire is overwhelmingly likely, we have stated this as the cause. This was a conscious choice which we believe gives a more correct picture of the actual causes, even

though it may give an imprecise picture of the police's clearance rate as concerns fatal fires.

The medical records were also very thin in some cases. In these cases the majority of categories were therefore marked as unknown. Similarly, cases potentially relating to cognitive ability, substance abuse and mental illness were labelled as unknown, as these conditions most often are not evaluated in cases where the doctor does not have any suspicion. It is therefore reasonable to assume that a large percentage of cases where these conditions are labelled unknown that entails no existence of impaired cognitive ability, known substance abuse or mental illness. In cases where the victim rarely had been to see the doctor, the medical records tended to be old with scarce updated information. The relevance of data has therefore been evaluated from one case to the next.

To which extent one is affected by a certain blood-alcohol level depends on a variety of factors, e.g. weight and the person's alcohol tolerance. It can therefore not be stated with any certainty to which extent alcohol actually affected the victim and the outcome of the fire.

Based on information in the investigation reports and the post-mortem reports we therefore registered categorically whether or not the fire victims were under the influence of alcohol at the time of death.

5 CONCLUSIONS

Individuals who have died in fire cannot be divided into groups of common denominators, but there are some combinations of factors that we have seen repeatedly:

For those who have reached retirement age, we mainly see four risk factors: *reduced mobility, impaired cognitive ability, mental disorders* and *smoking*.

For those under retirement age, the risk factors are *known substance abuse*, *mental illness*, *alcoholic influence* and *smoking* that appear, either alone or in combination with each other.

There is an increasing risk of dying in a fire with increasing age. Generally speaking, men do not have a higher risk than women, but in some age groups, the risk of fatality is greater for men. Alcohol constitutes a greater risk factor for men than for women.

Fatal fires in Norway are a social problem, which calls for measures individually adapted to the persons with the identified risk factors.

The current study has generated extensive data which can serve as a baseline to track whether the fatality rate for different groups increases or decreases over time. This could also be valuable when evaluating the effect of future fire preventive measures.

REFERENCES

Bruusgaard, D., 2017. Sosialt problem. Store medisinske leksikon.

Dell Inc, 2015. Dell Statistica (data analysis software system). Dell Inc.

DHS, 2016a. Civilian fire fatalities in residential buildings (2012–2014) (Topical report No. Volume 17, Issue 4). U.S. Department of Homeland Security, U.S. Fire Administration, Emmetsburg, Maryland, USA.

DHS, 2016b. Fire Risk in 2014 (Topical report No. Volume 17, Issue 7). U.S. Department of Homeland Security, U.S. Fire Administration, Emmetsburg, Maryland, USA.

DSB, 2010. Kjennetegn og utviklingstrekk ved dødsbranner og omkomne i brann. En gjennomgang av DSBs statistikk over omkomne i brann 1986–2009. Direktoratet for samfunnssikkerhet og beredskap.

Falnes-Dalheim, A., Slaastad, T.I., 2007. Færre unge—flere eldre.

Gjøsund, G., Almklov, P.G., Halvorsen, K., Storesund, K., 2016. Vulnerability and prevention of fatal fires, in: In: L. Walls, M. Revie and T. Bedford, '*Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016. Presented at the ESREL 2016,* Taylor & Francis Group, CRC Press, Glasgow, Scotland.

Kristiansen, J.E., 2016. Dette er Norge 2016 - Hva tallene forteller. Statistisk Sentralbyrå, Oslo/Kongsvinger.

Sesseng, C., Storesund, K., Steen-Hansen, A., 2017. Analysis of fatal fires in Norway in the 2005–2014 period (RISE-report No. A17 20176:2). RISE Fire Research, Trondheim, Norway.

Skaar, T.E., 2013. Alkohol og brann. Rapport fra kartlegging og sammenhenger mellom alkoholbruk og dødsfall i boliger. Norsk brannvernforening.

Statistics Norway, 2017. Tabell—Folkemengd, etter kjønn, alder og sivilstand. 1. januar 2017 (SÅ 60) [WWW Document]. Statistisk sentralbyrå. URL http://www.ssb.no/300113/folkemengd-etter-kjonn-alder-og-sivilstand.1.januar-2017-sa-60 (accessed 6.21.17).

Steen-Hansen, A., Heskestad, A.W., Mostue, B.A., Stensaas, J.P., 2010. Brannsikkerhetsnivået i sykehjem og pleieinstitusjoner for eldre (SINTEF NBL No. NBL A09130 rev 1). Trondheim, Norway.

Storesund, K., 2013. Forprosjekt; Dødsbranner i Norge (SINTEF-rapport No. NBL A13113). SINTEF NBL as, Trondheim.

Storesund, K., Sesseng, C., Steen-Hansen, A., Bøe, A.G., Stølen, R., Gjøsund, G., Halvorsen, K., Almklov, P.G., 2015. Rett tiltak på rett sted—Forebyggende og målrettede tekniske og organisatoriske tiltak mot dødsbranner i risikogrupper (SPFR report No. A15 20075:1). SP Fire Research AS, Trondheim, Norway.

Storesund, K., Steen Hansen, A., Sesseng, C., Gjøsund, G., Halvorsen, K., Almklov, P.G., 2016. How can fatal fires involving vulnerable people be avoided? Presented at the *Interflam Fire Science and Engineering Conference,* Interscience Communications Limited, Windsor, UK.

Xiong, L., Bruck, D., Ball, M., 2015. Comparative investigation of 'survival'and fatality factors in accidental residential fires. *Fire Safety Journal* 73, 37–47.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Probabilities in safety of machinery—Markov model for the scaling of risk reduction effects due to limiting the hazard exposure

H. Mödden
German Machine Tool Builders' Association (VDW), Frankfurt am Main, Germany

ABSTRACT: The most recent product safety standards for machine tools such as ISO 16090 for the safety of milling machines follow the three-step-method of risk reduction, as it is explained in ISO 12100. Step 1 is always the design of inherent safety of the machine. Step 2 follows with the design of additional safety measures, and step 3 focusses on instruction for use (including warning signs at the machine) and training. A very effective risk reduction can be achieved, if the work area of the machine is fully enclosed (this belongs to step 2). In doing so, the automatic machining processes can take place without exposing the operator to the hazard. However, operator access to the work zone is sometimes necessary for manual intervention, such as setting actions inside the work area or workpiece change. Then careful instructions are necessary, which shall enable the operator to fully control the situation and protect himself by awareness of the respective hazards.

The many different kinds of operator activities, which vary with the selected machining process, can be allocated to specific modes of operation. In ISO 16090, a modes of operation concept comprises five selectable options: 0: manual mode/1: automatic machining/2: setting mode/3: special mode with limited manual intervention/and separately for selected operators only: service mode. The purpose of such a concept is to reduce the relative exposure of the operator to occurring hazards as far as for the intended use possible.

The combination of full enclosure and modes of operation concept brings about a significant risk reduction, however the effect is being discussed only qualitatively on an intuitive basis. Unfortunately, a scalable model is still missing, which could quantify the risk reduction effects, e.g. for the purpose of parameter optimization.

In order to improve the engineering abilities, a very first simplified Markov model is presented here, which is founded on a probabilistic concept for the description of the operator activities at a machine. As a result, the a.m. risk reduction effects can be scaled probabilistically. This is of advantage, if it comes to the quantitative reliability requirements of safety functions according to ISO 13849-1 such as e.g. a safe standstill of gravity loaded vertical axes. Because of the potentially severe hazards of those axes in setting mode or during manual intervention in the work area, the required reliability is high, e.g. performance level PL = d according to ISO 13849-1. This paper explains for typical manual interventions, how this requirement and supplementary safety provisions can be justified probabilistically.

1 PROBABILITIES IN SAFETY OF MACHINERY

Numerical probabilistic representations were recently introduced in the field of general machinery safety, when the revised European Machinery Directive 2006/42/EG extended the "hazard analysis" of former versions to a "risk analysis" by introducing the term "probability" in the expression: "*estimate the risks, taking into account the severity of the possible injury or damage to health and the probability of its occurrence*". Since this alteration in a legal text, simplified probabilistic methods as in ISO 13849-1 (2015) are being developed. They encounter a well-proven practical state-of-the-art, which is merely based on qualitatively defined requirements. They were mainly focussing on hazards as such (and their countermeasures) rather than "risks, severities of injuries and their probability of occurrence". Nevertheless, on the background of customer demands for very high availabilities (which is equivalent to high inherent safety), it brought about a well-tried state-of-the-art following the three-step-reduction method of ISO12100 (2010). The state-of-the-art is defined on a non-quantitative descriptive background in harmonized safety standards since more than 20 years.

2 CONDITIONS OF A MARKOV MODEL IN SHORT FORM

Markov chains are suitable for describing stochastic processes that change between finitely many states, as it is the case during machine operation. Markov's modelling is illustrated in Feller (1967), for example, by constant transition probabilities. Discrete system states are defined with according probabilities of a) either keeping a state, or b) the transition to another state. The representation with time-dependent transition probabilities on the basis of constant failure rates, as with an exponential distribution, is also possible. The relative frequencies of the system states (as an approximation of the according probabilities) could be derived, for example, from a video evaluation of typical activities. Because, probability values can be represented as relative frequencies, if a large basic population or a large number of repetitions of the observed random events are given. The latter is the case over the service life of a machine with regard to the system states defined under section 3 below. Accordingly, a Markov model shall be created here for the manual intervention in the workspace of a machine tool. In doing so, certain conditions need to be fulfilled.

Check the conditions:

Condition 1: The probability that the state Z_i, which occurs at an i-th time step, only depends on the state Z_{i-1}, which was present at the (i-1)-th time step, (Markov property).

Reflection: The probability of the state Z_i "Manual intervention" at the i-th time step depends only on the state Z_{i-1}, which was taken to the (i-1)-th state: namely, if the (i-1)-th state was "automatic processing", there is a certain (average) probability to change into the state "manual intervention".

The Markov property may not be fully attainable with very detailed state models, for example, if certain sequential patterns are passed through. In this case, the transition probability of the next state may not only depend on the last state, but also on the previous one. For a first simplified model with three states in Figure 1, however, the Markov property can be considered fulfilled.

Condition 2: In a homogeneous Markov process with a discrete range of parameters and finitely many states, the transition matrix B(i,i+1) must not depend on i. At discrete time intervals $t_i = \mathrm{i} \cdot t_{ref}$ this means that the transition matrix is independent of time, so it must be stationary.

Reflection: The use of a machine leads to changing operating modes. Their occurrence probabilities can be approximated as averaged constants at all transitions (i, i+1). The requirement for a transition matrix that is independent of "i" might not be fully satisfied in the short term, because, for example, certain sequences of machining steps that differ from the sequences averaged over long periods of time occur on a machine during a day shift. For long periods of time however, the requirement for a transition matrix that it is independent of "i" (i. e. indep. of time) is approximately met because of the average effect. But, only for the "inner states" of the model, because with the "absorbing states" a temporal increase relative to global time comes into play (see below). Unforeseeable failures during the repeated change of operating modes above, either of technical or of human cause, require "absorbing states".

Then, a homogeneous Markov model needs to be expanded to an inhomogeneous Markov process. The reason is that technical faults have a time-dependent probability of occurrence, e. g. exponentially distributed. Even an inhomogeneous Markov model still fulfils the Markov condition that the further course of the process is clearly defined at all times.

As a result, both conditions above are fulfilled and a Markov model can be created for manual intervention in the workspace of a machine tool.

3 SYSTEM STATES OF THE MODEL

In accordance with the normatively anchored operating modes, such as in ISO 16090-1 (2016), only 3 states of the machine are considered here at first:

System State 1: Production in automatic mode with closed prot. doors (with Z1 as state prob.)

System State 2: Setting mode to prepare automatic op. with opened prot. doors (with Z2 as state prob.)

System State 3: Manual intervention with opened protection doors, e. g. to carry out a workpiece change or to correct a fault; either from the previous state 1 or from state 2 (with Z3 as state prob.)

Figure 1 shows a possible sequence of changes of states: To prepare automatic operation from time 1 on, setting mode is started at time 0. In event

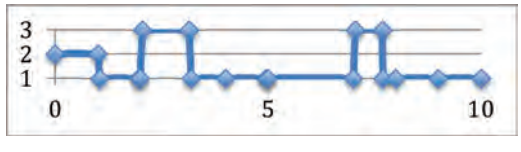


Figure 1. Example of state changes in automatic mode (x-axis: time in minutes, y-axis: indicator for state).

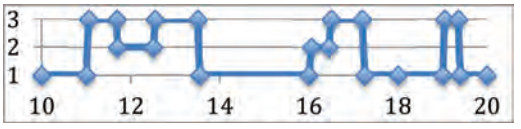


Figure 2. Example of state changes in setting mode (x-axis: time in minutes, y-axis: indicator for state).

2, a malfunction occurs (transition from state 1 to 3) that is rectified manually up to time 3.15. From then on, automatic operation continues undisturbed. At point 7, the workpiece is finished and is being replaced by an unmachined part, which lasts until point 7.7, after which the automatic machining continues. Fig. 2 shows another possible continuation of state changes. At time step 10, there is still automatic processing of Fig. 1.

The workpiece is finished at time point 11 and is being removed until point 11.7. Afterwards, a new setup operation takes place in which up to 12.6 manual operations are necessary. And from time point 13.5 on, a new workpiece is clamped.

Afterwards an automatic processing is carried out up to time point 16, where a malfunction occurs. This leads to a switch-over into setting mode, whereby manual intervention is necessary from point 16.5 to 17.3. From time point 17.3 the automatic processing can be continued. And with time point 19, another workpiece change takes place, which is finished at time point 19.4. The new workpiece is then to be processed aomatically.

The activities in Figs. 1 and 2 can thus be grouped into four types of human-machine interaction:

Activity A: Manual workpiece change in automatic mode (Z1è Z3)

Activity B: Manual troubleshooting in automatic mode (Z1è Z3)

Activity C: General setting mode (Z2) or manual workpiece change during setting (Z2è Z3)

Activity D: Manual troubleshooting in setting mode (Z2è Z3).

In the following section, another activity occurring in real practice shall also be introduced for the preparation of a follow-up paper:

Action E: Manual troubleshooting in setting mode (Z2è Z3), which could not be resolved.

4 PRESENTATION OF THE STATE AND TRANSITION PROBABILITIES

The real key question in this paper is: How likely or relatively frequent is it that the operator is exposed to hazards, which could possibly occur during manual work in the work area?

In order to answer the question quantitatively with scalable probabilities, a homogeneous

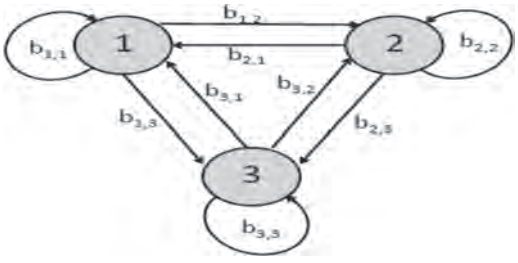


Figure 3. State and transition prob. of hom. Markov model.

Markov model is suitable as a quantitative reference frame. This should be adapted to the above three system states. Fig. 3 shows a graphical representation. The state probabilities at subsequent "i" time steps P(i) result from the so-called "B-matrix of state and transition probabilities" of a homogeneous Markov process, see Eq. (2).

The states 1,2 and 3 are called "inner states" because they can merge into each other.

For time step i = 0, an initial start vector P(0) from state probabilities is suitable as a representation:

$$\underline{P}(0) = \begin{pmatrix} Z1(0) \\ Z2(0) \\ Z3(0) \end{pmatrix} \tag{1}$$

$$\underline{B} = \begin{pmatrix} b_{1,1} & b_{1,2} & b_{1,3} \\ b_{2,1} & b_{2,2} & b_{2,3} \\ b_{3,1} & b_{3,1} & b_{3,3} \end{pmatrix} \tag{2}$$

The progressively developing connection probabilities are described in Feller (1967):

$$\begin{aligned} &\underline{P}(i+k) = \underline{P}(i) \cdot \underline{B}^k \\ &or\ \underline{P}(k) = \underline{P}(0) \cdot \underline{B}^k \end{aligned} \tag{3}$$

Thus, forecasts of repetitions of an assumed reference period can be deduced at time point $k \cdot t_{ref}$.

4.1 *Derivation of the transition probabilities*

As explained in section 2, probabilities over the lifetime of a machine and for a basic population of comparable machines can be presented as relative frequencies.

For the duration and repetition rates of the five a.m. activities, an averaging over several 8 hour shifts, shall be assumed. This serves as a frame of reference for the hypothetical estimation of the needed transition probabilities. In this paper, the values listed in Table 1 are assumed, e.g. a manual workpiece change on the machine tool takes

Table 1. Five typical activities on a machine tool.

Action	$t_{exp,1}$ [h]	$f_{exp,1}$ [h^{-1}]	F_{rel} [–]	*Fraction* [–]	*Coeff.* ($b_{1,j}$)	*Connecting Coeff.*
A	0.02	6	0.12	$(b_{1,3})_A = 0.12$	$b_{1,3} = (b_{1,3})_A + (b_{1,3})_B = 0.1325$	$b_{1,1} = 1 - b_{1,2} - b_{1,3} = 0.6675$ with: $b_{2,2} = 0$ and $b_{2,3} = 0.008$ follows:
B	0.05	0.25	0.0125	$(b_{1,3})_B = 0.0125$		
C	8	0.025	0.2	$(b_{1,2})_C = 0.2$	$b_{1,2} = 0.2$	
D	0.1	0.05	0.005	$(b_{2,3})_D = 0.005$	$b_{2,3} = (b_{2,3})_D + (b_{2,3})_E = 0.008$	$b_{2,1} = 1 - b_{2,3} = 0.992$ with: $b_{3,3} = (b_{2,3})_E$ and
E	0.5	0.006	0.003	$(b_{2,3})_E = 0.003$		$b_{3,2} = (b_{2,3})_D$ follows: $b_{3,1} = 1 - b_{3,2} - b_{3,3} = 0.992$

1.2 min. (= 0.02 hours) and repeats itself 6 times per hour, this means 7.2 min. of non-productive time p.h. The rel. exposure is 0.12.

From Table 1, the state and transition probabilities can be analytically derived from state 1 (with probability Z1) with $b_{1,i}$ (i = 1,3) to:

$$b_{1,1} = 0.6675;\ b_{1,2} = 0.2;\ b_{1,3} = 0 \tag{4}$$

Idition, it is clear that the setting mode (Z2) must always switch to automatic mode (Z1), either directly or via a troubleshooting of a fault (Z3) i.e.:

$$b_{2,2} = 0.0. \tag{5}$$

With the additional assumptions in Table 1, the remaining coefficients can be determined analytically as follows: 99.2% of the general setting mode (Z2) is transferred to automatic operation (Z1) without manual intervention (Z3). From the counterconclusion, a 0.8% usage share of the setting mode for troubleshooting can be deduced:

$$b_{2,1} = 0.992,\ b_{2,3} = 0.008 \tag{6}$$

In addition, it shall be assumed that in automatic mode (Z1), 100% of the faults are either eliminated or assigned to setting mode (Z2). For a follow-up paper, it could be further assumed that only 5/8 of all faults in the setting mode can be rectified and 3/8 cannot.

Assuming that the unrecoverable faults (E) occur in state (Z3) via setting mode (Z2), the result would be: $b_{3,3} = (b_{2,3})_E$. And because the remediable faults (D) are dealt with in setting mode, the result would be: $b_{3,2} = (b_{2,3})_D$. It follows:

$$b_{3,1} = 0.992;\ b_{3,2} = 0.005;\ b_{3,3} = 0.003. \tag{7}$$

Therefore, the state and transition probabilities required for the matrix representation of a Markov model are available as coefficients. The so-called "B-matrix of the state and transition probabilities" of a homogeneous Markov process results to:

$$\underline{B} = \begin{pmatrix} 0.6675 & 0.2 & 0.1325 \\ 0.992 & 0 & 0.008 \\ 0.992 & 0.005 & 0.003 \end{pmatrix} \tag{8}$$

Plausibility check: the row sum of $\underline{B}$. must be 1.

4.2 *State vector after repetitions*

For the i-th repetition of a state with a Markov model, the following vector of the state probabilities can be defined:

$$\underline{P}(i) = \begin{pmatrix} Z1(i) \\ Z2(i) \\ Z3(i) \end{pmatrix} \tag{9}$$

For example, to initialize the model with a completed setting operation to be followed by an automatic process, an initial condition (i = 0) can be:

$$\underline{P}(0) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \tag{10}$$

4.3 *State vector after one week/one year*

As assumed above, the reference frame t_{ref} for estimating the transition probabilities here is an 8-hour shift. The coefficients of the B-matrix are derived from Table 1 above in Eq. 3. In this way, the values of the time proportions of the states for a reference time period can be concluded (e.g., here a 40-hour week with 5 times 8-hour days):

Z1(i) Proportion of total usage time in state 1 (automatic operation),
Z2(i) Proportion of total usage time in state 2 (setting mode),
Z3(i) Prop. total usage time in condition 3 (fault).

As a representation in vector form:

$$\underline{P}(i)^T = \left(Z1(i), Z2(i), Z3(i)\right)^T$$

Table 2. Short oscillation of the Markov model.

Dur.	*Repet. i*	$\underline{P}(i)^T$
1 day	1	(0.6675, 0.20, 0.1325)
1 week	5	(0.748, 0.151, 0.101)
1 year	220	(0.789, 0.150, 0.100)

Table 2 shows the rapid transient oscillation of the Markov model, such that a quasi-stationary vector of the state probabilities already appears after 1 week. For the steady state after 1 week in table 2 a proportional undisturbed production of: Z1(i = 5) = 0.784 = 78.4% results.

The interruption of production results proportionately to: 1- 0.784 = 0.216 = 21.6%. And from this the system state 3 „manual intervention" has a share of Z3 = 0.1 = 10%, based othe assumptions above. This provides a comete a scalable framework for plausibly estimating the probability of a hazard or injury during manual iervention.

5 HAZARDS OF MANUAL ACTIVITIES

A whole range of hazards are conceivable during manual intervention in the work area: crushing of hands or arms, shearing of fingers, entanglement of clothing etc. are basically possible. Risk assessment is in the first step aimed at the completeness of all possible hazards. This was formerly known as a hazard analysis. However, in order to become a risk analysis, a continuation sp needs to follow: every "possibility of a hazard" is to be converted into a "probability of this hazard".

Here, state 3 is "Manual intervention" to carry out a workpiece change or to correct a malfunction; either from the previous state 1 or from state 2. The subsequent state after 3 could be a hazard with an injury. It shall be combined in a further state:

System State 4: Hazards can occur that could possibly cause injuries (with Z4 as the probability of their occurrence). Because these injuries could be severe, a standstill of the machine for accident investigation can be assumed such that (Z4 = 1) would apply. Thus, state 4 is an "absorbing state".

It is supplemented in Fig. 4. If an absorbing state is taken once, it will not change ($b_{4,4}$ = 1). Absorbing states represent critical states of a system. Therefore, one is interested in the probability with which such critical states occur during the process.

The extension of the model with state 4 turns the above *homogeneous* Markov model into an *inhomogeneous* Markov model. Because state 4 is not time-independent, i.e. the transition probability $b_{3,4}$ is not constant, but depends on the time: $b_{3,4} = f(t)$.

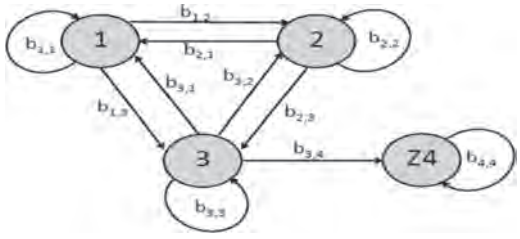


Figure 4. Extension to an inhomogeneous Markov model: state and transition probabilities and an absorbent state 4.

In simplified terms, the exponential distribution over the PFH_d-values assumed in ISO 13849-1 (2015) can be regarded as the trickle rate of an "hourglass" and thus a time proportionality can be applied, see Moedden (2014). The following hypothetically derived values for the above activities A, B, C, D and E are intended to provide the connection for an answer to the key question above.

It is a very first proposal of "risk snapshots". The purpose is merely to estimate the risk reduction of guards (e.g. full enclosure) in the context of an exemplary modes of operation concept.

A calculation example shall be used to compare the five activities A, B, C, D and E as "risk snapshots" in a scaled form, as assumed above, see Table 3. An 8-hour shift is used again. When manually intervening in the working area, the operator has to completely trust in reliable control technology, e. g. for the safety functions "Safe operation stop" and "Safe reduced speed".

The relevant failure probabilities can be assigned as $\sum PFH_{d,tech}$-values in a superimposed form. The typical hazards can be superimposed, too. The probability of occurrence of relevant hazards (risk element O according to equation 11) and their controllability (or avoidability, risk element C according to equation 15) are simply estimated on the safe side, just for illustration purposes.

Activity A: Manual workpiece change in automatic mode (Z1è Z3), with:

- Safety function "Safe operational stop (SOS):
- 2 rotary axes $\left(\sum PFH_{d,tech} = 2 \cdot \left(6.5 \cdot 10^{-6}\right) \frac{1}{h}\right)$, which can cause an entanglement hazard (EN),
- and 3 translational axes $\left(\Sigma PFH_{d,tech} = 3 \cdot (4.5 \cdot 10^{-6}) \frac{1}{h}\right)$, which can cause crushing (CR),

Activity B: Manual fault rectification in automatic operation (Z1 → Z3) with the same safety function and hazards as for activity A,

Activity C: General setting mode (Z2) or manual workpiece change (Z2 → Z3) with the same safety function as for activity A,

Activity D: Manual troubleshooting in setting mode (Z2 → Z3) with:

Table 3. "Risk snapshots" of fiveypical activities on a fully enclosed machine tool.

Activity/Safety function/Hazards	$\Sigma PFH_{d,tech}$ [η^{-1}]	*O* [–]	F_{rel} [–]	*C* [–]
A: /SOS/EN + CR	$2\cdot(6.5\cdot10^{-6})$ $+3\cdot(4.5\cdot10^{-6})$	1	0.12	0
B: /SOS/EN + CR	$2\cdot(6.5\cdot10^{-6})$ $+3\cdot(4.5\cdot10^{-6})$	1	0.0125	0
C: /SOS/EN + CR	$2\cdot(6.5\cdot10^{-6})$ $+3\cdot(4.5\cdot10^{-6})$	1	0.2	0
D: /RSP/EN + CR	$1\cdot(6.5\cdot10^{-6})$ $+1\cdot(4.5\cdot10^{-6})$	0.2	0.005	0.5
E: /RSP/EN+CR	$1\cdot(6.5\cdot10^{-6})$ $+1\cdot(4.5\cdot10^{-6})$	1	0.003	0

– Safety funct. "Safely reduced speed (RSP) with:
– 1 rotary axis $\left(\sum PFH_{d,tech} = 1\cdot(6.5\cdot10^{-6})\frac{1}{h}\right)$, which can cause an entanglement hazard (EN),
– and 1 translational axis ($\Sigma PFH_{d,tech}$ = 1 · (4.5 · $10^{-6})\frac{1}{h}$), which can cause crushing (CR),

Activity E (for a follow-up paper): Manual troubleshooting in setting mode whout success with the same safety function and hazards ain activity D; but without controllability: C = 0 and with very high probability of occurrence of a hazard: O = 1.

The likelihood of a hazard occurring after a control failure (e. g. causing an unexpected movement) has been assumed for activities A, B, C with O = 1. This means that a failure is very likely to lead to a hazard in the respective situation. For activity D, only O = 0.2 is assumed. For the first three activities a controllability is not possible: C = 0, and for the fourth activity it shall be estimated with C = 0.5.

The superimposed failure probabilities are compared in Figure 5 as $\sum PFH_{d,tech}$ – values.

These base values can be used to carry out a risk assessment for manual interventions A, B, C, D and E. For this purpose, the risk elements probability of occurrence of a hazard (after a control failure or after a human error), relative exposure and controllability ought to be reasonsmated for Eq. (14). In addition, the terms used should be clear in order to obtain a logical risk model.

5.1 *Key term "Probability of hazardous events"*

The likelihood of hazards can be caused by technical or human factors or force majeure. The designer cannot influence the latter, but he/she must concentrate on mastering the causes of technical failures (index "tech"), also on foreseeable human errors. For quantitative aspects, an objectively scalable probability plays a special role.

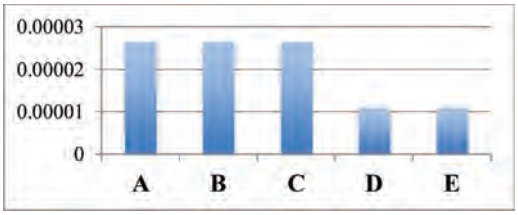


Figure 5. Bar chart of the $\sum PFH_{d,tech}$ – values of "Risk snapshots" (x-axis: indicator for activity, y-axis: values in h^{-1}.)

For this purpose, a correctly defined probability space is indispensable, see Moedden (2017). Two quantitative representations of the term "probability" are established: a) generally, in most cases the dimensionless probability such as illustrated in Feller (1967); and in the context of control chains for safety functions, often b) a probability per time unit (PFH_d-value, see ISO 13849-1 (2015)) is used. The first can be derived from the latter via a time integral, as explained by means of a "hourglass analogy" in Moedden (2014).

> In this context, the interpretation of the so-called "Murphy's Law" must be reasonably restricted, it reads: "Anything that can go wrong will go wrong". Murphy's law is an alleged wisdom of life, which makes a statement about human failure or sources of error in complex systems. In the short form with the emphasis on "…will go wrong" there is presumably an assumption contained, which must not go unmentioned here, since the key term "probability" has to be distinguished from the term "possibility". Presumably, the precondition of "Murpy's law" is: either an infinite number of repetitions, or an infinite basic population. But, "Murpy's law" doesn't work for finite repetitions and/or a finite basic population to the conclusion "…will go wrong".

5.2 *Scaling of risk reduction measures*

Practical machine design methods, as collected for a milling machine in ISO 16090-1 (2016), apply a whole range of measures that bring about a considerable distance in the probabilities of the following events:

Event I: a failure in a mechatronic control chain of a certain safety function, for instance due to the failure of a mechanical brake actuator in the safety function "SOS" (see abovtivity A) and

Event II: hazards occurring from this failure (here: crushing and entanglemen

The likelihood of event I. is being expressed as a PFH_d-value of the safety fction. The unit is a number per hour. The probability oevent II. in a certain time frame $t_a \le t \le t_b$, i.e. the occurrence

probability of according hazards, is given as a dimensionless number: it represents a probability and the parameter PHE is used here.

If the a.m. over-interpretation of "Murphy's Law" would be made, the conclusion were: $PHE = PFH_d \cdot (t_b - t_a)$

> The reality is different: there is a well-tried art of safety design such that PHE is generally significantly lower $PHE \ll PFH_d \cdot (t_b - t_a)$, because additional risk reduction measures "surround" the mechatronic control chains in a cascaded form. For instance, protective guards such as a full enclosure of the work area are very effective in reducing the probability of hazardous events, which might be caused by failures inside the work area. Thus, within in a cascaded safety design, not every failure of a safety function leads to a hazard. Only if an operator is exposed to potentially hazardous control chains, a hazard can occur. The PFH_d-values of control chains represent the "threatening" failure prob.'s. That is why a differentiation has to be made between an open or closed protective guard (see ste definitions above), if it comes to deriving PHE-values. Obviously, hazardous events can occur, when the protection grds are open. With closed guards however, an operator is n exposed to possible hazardous events inside the full enclosure.

The risk reduction effect by limiting the relative exposure is explained in section 6.1 below as a product of time duration and repetition rate.

A dimensionless probability wh a scaling factor "O" can be defined as follows for the time period $t_a \le t \le t_b$. It corresponds to the risk element "probability of occurrence of a hazard" (ISO 12100, section 5.5.2.3.2):

$$\begin{aligned} PHE &= O \cdot \int_{t_a}^{t_b} \sum PFH_d \cdot dt \\ &= O \cdot \sum PFH_d \cdot [t_b - t_a] \text{ with } 0 \le O \le 1 \end{aligned} \quad (11)$$

6 PROBABILISTIC RISK MODEL

In automatic operation (index 1 below), closed guards of fully enclosed work areas hold back any released tool and/or workpiece fragments up to the withstand capabilities of the respective dimensioning regulations as in ISO 16090-1 (2016). Even with open protection guards as in setting mode (index 2 below), a full enclosure can also considerably reduce the failure probability of a technological function (index "tech") during setting mode indirectly, as explained in the "implicit fault detection in the process", see Moedden (2016).

As regards released parts of tools or/and workpieces after a failure in a control chain, the withstand capability of the enclosure against causes a strongly reduced probability of occurrence of a technically caused hazardous event, i. e. with $O_{1a} << 1$. For other non-ejection hazards even $O_{1b} = 0$ holds (with $O_1 <= O_{1a} + O_{1b}$). Therefore, PHE_1 during automatic mode in the period $t_a \le t \le t_b$ is:

$$PHE_1 = O_1 \cdot \int_{t_a}^{t_b} \sum PFH_{d,tech,1} \cdot dt \quad (12)$$

In setting mode, a numerical value $PHE_2 >> 0$ ought to be assumed for hazardous events, because, according to accidents in the entire population of all machines, injuries occur more often, when the protection doors / guards are open than in closed condition, see Moedden (2017).

6.1 *Exposure reduction*

In the case of hazard exposure, reducing the duration of exposure time t_{exp} and reducing the repetition rate f_{exp} act as a product to reduce the relative exposure $F_{rel} = t_{exp} \cdot f_{exp}$. As there is a kind of double proportionality, this can lead to a considerable risk reduction. For example, if workpiece and tool changes are mainly carried out automatically (instead of manually) and the product $t_{exp} \cdot f_{exp} << 1$ reaches very small values. However, some recurring hazard exposure events in setting mode (e. g. to prepare for automatic operation) are always left over. This results for a.m. setting operation in the period $t_a \le t \le t_b$ to the following expression for a probability of a hazardous event:

$$PHE'_2 = t_{exp,2} \cdot f_{exp,2} \cdot O_2 \cdot \int_{t_b}^{t_b} \sum PFH_{d,tech,2} \cdot dt \quad (13)$$

6.2 *Increasing controllability/avoidability*

If the above risk reduction measures have been implemented as far as reasonably possible, a "zero risk" can practically almost be reached acc. to Moedden (2017). However, even then (theoretically hypothetically) a very small technical residual risk remains. For this purpose, step 3 of the three-step method of risk reduction is being applied, the so-called instructive safety: in risk models, mostly described as avoidability or controllability (parameter C) in ISO 12100 (2010). In the above case of setting mode (indicated as "2"), and considering the avoidability effect with $C_2 > 0$, the remaining event rate of hazards in the period is calculated as follows $t_a \le t \le t_b$ in (Eq. 14):

$$PHE''_2 = (1 - C_2) \cdot t_{exp,2} \cdot f_{exp,2} \cdot O_2 \cdot \int_{t_a}^{t_b} \sum PFH_{d,tech,2} \cdot dt \quad (14)$$

PHE''_2 can be equated with the probability of injury, i. e. it corresponds to injury risk during setting.

6.3 *Comparison of different activities*

In accordance with the risk model in Eq. (14), the probability of occurrence of a hazard event is shown in the right-hand column of Table 4; and a corresponding bar chart is shown in Figure 6. The proportions of activities A, B, C, D and E in system state 3 are taken from Table 3 with $F_{rel.}$. The simple comparison in Table 4 obviously proves that not only the PFH_d-values in the range of some $10^{-6}h^{-1}$ are decisive for the probability of occurrence of a hazard event PHE. The other risk elements O, F_{rel}, C also have a considerable influence. This can be seen, because the bar chart of PHE values in Fig. 6 differs from the bar chart in Fig. 5, which merely shows the distribution of the summed PFH_d-values. Fig. 6, however, is more important for the risk of injury than Fig. 5. And Fig. 6 shows that the highest probability of occurrence of hazardous events are in activities C and A of this parameter study. Activities B, D and E are far less hazardous.

The coefficient $b_{3,4}$ in Figure 4 and Eq. (8*) is time-dependent: $b_{3,4} = f(t) = f(i \cdot t_{ref})$. And $b_{3,4}$ results from the sum of the PHE-values in the right-hand column of Table 4 to $b_{3,4} = 0.00007$; it has to be multiplied by as much 8-hour-shifts, as shall be considered. And because state 4 is an absorbing state: $b_{4,4} = 1$. Based on an eight-hour shift, the B-matrix in Eq. (2) and Eq. (8) can be extended from 3*3 to 4*4. This means that the probability of a hazard event can also be estimated.

The new B-matrix would look like this, Eq. (8*):

$$\underline{B^*} = \begin{pmatrix} 0.6675 & 0.2 & 0.1325 & 0 \\ 0.992 & 0 & 0.008 & 0 \\ 0.992 & 0.005 & 0.003 & b_{3,4} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Table 4. "Risk snapshots" and hazard probabilities.

Activity/ SF/HAZ.	$\Sigma PFH_{d,tech}$ [h^{-1}]	$t_b - t_a$ [h]	O [–]	F_{rel} [–]	C [–]	PHE [–]
A: /SOS/ EN + CR	$2\cdot(6.5\cdot10^{-6}) + 3\cdot(4.5\cdot10^{-6})$	8	1	0.12	0	2,5
B: /SOS/ EN + CR	$2\cdot(6.5\cdot10^{-6}) + 3\cdot(4.5\cdot10^{-6})$	8	1	0.0125	0	2,6
C: /SOS/ EN + CR	$2\cdot(6.5\cdot10^{-6}) + 3\cdot(4.5\cdot10^{-6})$	8	1	0.2	0	$4{,}2\cdot10^{-5}$
D: /RSP/ EN + CR	$1\cdot(6.5\cdot10^{-6}) + 1\cdot(4.5\cdot10^{-6})$	8	0.2	0.005	0.5	$4{,}4\cdot10^{-8}$
E: /RSP/ EN + CR	$1\cdot(6.5\cdot10^{-6}) + 1\cdot(4.5\cdot10^{-6})$	8	1	0.003	0	$2{,}6\cdot10^{-7}$

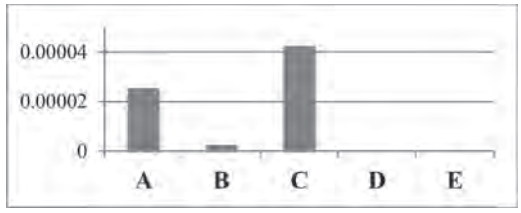


Figure 6. Bar chart of the PHE-values for the "Risk Snapshots" above (x-axis: indicator for activity, y-axis: values in [–]).

Table 5. Result for inhomogeneous Markov model.

Dur.	i	$\underline{P}(i)^T$
1 day	1	(0.6675,0.20, 0.1325, 0.00007)
1 week	5	(0.748, 0.151, 0.101, 0.00035)
1 year	220	(0.789, 0.150, 0.1000, 0.0154)

The state vector in vector form is:

$$\underline{P}(i)^T = \left(Z1(i), Z2(i), Z3(i), Z4(i)\right)^T$$

From Table 5, it can be seen that the values Z1, Z2 and Z3 are convergent, whereas Z4 is divergent, i.e. steadily increasing.

It can also be derived that the probability of the system state 4 in a basic population of n comparable machines is $Z4 \approx 1$ after 1 year, if $n = 1/0.01538 \approx 65$. This means that hazardous events, which can cause injuries, are very probable with above assumptions.

7 SUMMARY AND OUTLOOK

The system states of a simplified Markov model of man-machine interaction in operating machine tools have been defined. A diagram for the state and transition probabilities was developed for parameter studies.

The Markov model presented here has proven that reducing operator exposure is as important a means of reducing risk as the high reliability of control functions. Consequently, the model can also be used to compare safety measures with each other in a scaled form. In doing so, the designer has the opportunity to optimise effort and benefit. In addition, the model can also weigh different effects against each other in order to compensate, for example, i) an increased exposure due to extending the operating modes concept by ii) reducing the probability of failure of relevant control chains that can generate hazards, if they fail. This approach

has proven to be effective over more than a decade. Also a consideration of technical reliability versus human reliability can be performed in a scalable frame. For example, when clamping workpieces for machines equipped with the relatively new technology of vertical turning on milling machines, see Wittstock (2017).

Five typical activities have already been examined here in more detail to answer the key question above. Further operating modes are to be examined, i.e. mode 3: Special mode and Service mode.

NOMENCLATURE

Symbol	Dimension	Meaning
t_{ref}	h	Reference time step for the Markov model
t_{exp}	h	Time duration of a manual intervention
f_{exp}	$\frac{1}{h}$	Relative frequency of a manual intervention
O, F_{rel}, C	Probability	risk elements acc. to the parameters of ISO 12100
PFH_d	$\frac{1}{h}$	Average prob. of dangerous failure per hour. It is only for single i/o-channels equivalent to the failure rate (i.e. the frequency with which an engineered system fails, often denoted by λ)
$t_b - t_a$	h	Time period for integration
PHE	Probability	Prob. of hazardous event
$\underline{P}(n)^T$	Probability	State vector with prob.
$\underline{B}$	Probability	Transition matrix with prob. and coefficients $b_{i,j}$

REFERENCES

Feller, William, 1967: *An Introduction to Probability Theory and Ist Applications*, Volume I, Third Edition, John Wiley & Sons, Inc.

ISO 12100, 2010. *Safety of machinery—General principles for design—Risk assessment and risk reduction.* Berlin, Germany; Beuth Verlag GmbH.

ISO 13849-1, 2015. *Safety of machinery—Safety-related parts of control systems—Part 1: General principles for design.* Berlin, Germany.

ISO 16090-1: 2016, *Machine tools safety—Machining centres, Milling machines, Transfer machines—Part 1: Safety requirements,* Berlin, Germany.

Mödden, H. 2014. *Probabilities in Safety of Machinery—Part 1: Risk Profiling and Farmer Matrix,* ESREL Symposium Wroclaw.

Moedden, H., 2015: *Probabilities in Safety of Machinery—Elements of a Risk Model and Comparison with Field Data,* ESREL Zurich Switzerland.

Moedden, H., 2016: *Probabilities in Safety of Machinery—Risk Reduction Effects by Combination of Full Enclosure and Fault Detection in the Process,* ESREL 2016 Glasgow, Scotland.

Moedden, H., 2017: *Probabilities in Safety of Machinery: Borel-Cantelli Lemmas lead to a Prevention Dogma based on the Pareto principle,* 15th International Probabilistic Workshop, Dresden.

Spiegelhalter, D. et al, 2013: *Norm Chronicles—Stories and Numbers about Risk,* Cambridge.

The European parliament & the council of the European Union 2006: *Machinery Directive 2006/42/EC.* - Linked to Official Journal of EC: Summary list of titles and references of harmonised standards under Direct. 2006/42/EC for Machinery.

Wittstock, V., 2017: *VDW Nr. 020 Gefährdungsrisiko freigesetzter Werkstücke bei.. Fräsmaschinen mit tels probab. Berechnungsansätze,* Chemnitz.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Information flow and knowledge transfer of accident investigation results in the Norwegian construction industry

K. Wasilkiewicz
Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: This paper examines the information flow of accident investigation results in the construction industry, and how this affects learning processes. The aim is to document the state-of-the art in the industry on information flow that follow accident investigations, and to find how accident investigation results better can be learned from and used as input for proactive safety management. A literature review was undertaken on the relation between accidents and learning. An interview study was undertaken with different actors (clients, contractors and consulting engineers) in the construction industry on accident investigations and information flow of accident investigation results. The preliminary results from the interviews are presented. Mostly, results after accident investigations are shared within a company, and there is no systematic sharing of information between companies, other than occasional sharing. Further research needs on information sharing after accident investigations are discussed.

1 INTRODUCTION

The construction industry is a complex industry, with constantly changing processes and activities, several actors involved that depend on each other, and external factors, such as state of the market, that affect construction projects.

Just as the industry varies a lot in terms of company sizes, sizes of construction sites, resources available, and competence and experience of managers and workers, so are accident investigations and the results of them influenced by these characteristics and conditions.

The aim of this research is to look closer at safety in the construction industry, by studying the processes after accident investigations, which are meant to accommodate for learning processes and prevention of future accidents.

This paper presents preliminary results of research concerning the results of and the information flow after accident investigations, as it is a prerequisite for the learning processes.

2 BACKGROUND

2.1 *The complexity of the construction industry*

A construction project can be many things; a small cottage to be built, a highway, a tunnel or a skyscraper. "The construction of a building can be regarded as a complex, information dependent, prototype production process were conception, design and production phases are compressed, concurrent and highly interdepended in an environment where there exists a usually large number of internal and external uncertainties" (Pryke, 2012, p.64). This shows how the construction industry varies with regards to project size and durability, company size, contract models and so on. According to Lingard and Rowlinson (2005, p. 3) the project structure which companies in the construction industry operate in, is an important characteristic that challenges the safety work in the industry. Each project site is different, forming a new, temporary organisation. There are large circulations of personnel; some come in for as shorts as a day, others stay for the whole project period. Further, construction projects are operating in an ever-changing environment largely influenced by the state of the marked. All these variations are influencing how safety work is being implemented and executed in the daily life of construction projects.

The socio-technical system involved in risk management by Rasmussen (1997), illustrates the different levels that will influence the safety in the sharp end. The model includes environmental stressors that can influence the different levels that again can influence the overall safety.

Further early project phases can affect the next phases, i.e. decisions in early phases of projects can influence the safety during construction. Therefore, it is important to look at the whole range of actors involved in the project not only during the construction phase, but also in earlier phases.

2.2 *The construction industry in Norway*

The Norwegian construction industry was employing almost 235 000 people in 2016, and the industry comprised of more than 57 000 enterprises, of which approximately 90% had 0–9 employees, and around 1% had more than 50 employees (SSB, 2017a).

The industry is one of the industries with the highest number of work related fatalities and injuries on mainland Norway. Between 2010–2015, there was in total 69 fatalities related to construction work, i.e. close to 12 fatalities per year (NLIA, 2016). Most of the fatal accidents involve falls, followed by collisions, being hit by an object or being crushed or trapped.

In 2016, there were 9 fatalities and more than 2700 reported injuries in the construction industry (SSB, 2017b). More than half of these resulted in long term absence (more than three days' absence from work). The most frequent types of accidents resulting in serious injuries were: fall from roof/ floor/ platform, fall from scaffolding, contact with a falling object, contact with moving parts of machine, being hit by an object in a lifting operation, fall from ladder, and fall from height when unsecured (NLIA, 2017).

2.3 *Incident reporting and accident investigations*

Incident and accident reporting is an important tool for accident understanding, and thus learning and future accident prevention. Incidents and accidents should be reported and investigated internally to prevent them from reoccurring (Lingard & Rowlinson, 2005, p.163–164), and in some cases also investigated externally (e.g. incident with high or potentially high consequences).

According to the Working Environment Act §5–2 (2005) in Norway, employers are obligated to report accidents with fatal or serious outcome to the Norwegian Labour Inspection Authority (NLIA) or the Police. Nearly all fatalities are reported, however in the "serious injury" category there are still unrecorded numbers (NLIA, 2015a). NLIA conduct inspections after all fatal accidents (NLIA, 2015b), but less severe accidents are not examined as closely as accidents with severe outcome.

2.4 *Learning from accidents*

Learning from accidents is one of the goals with accident investigations, both to prevent similar accidents to reoccur, and to prevent other accidents.

Learning can refer to either a product or a process, respectively something learned and the activity of learning (Argyris and Schön, 1996, p. 3). In relation to construction safety, the product can be accident understanding, which is one of the products after an accident investigation. To improve the safety and avoid similar accidents from reoccurring, the knowledge obtained needs to be applied. However, taking actions might require major changes (e.g. cultural and behavioural) (Love et al., 2013), which in practice can be challenging and will require designated resources in a company. Correct actions taken after accidents, show the results of learning in practice, by applying the obtained knowledge. This takes learning one step further towards improvement.

Organisational learning can be divided in two types of learning. Single-loop learning refers to learning that results in changes of action so that the outcome is desired. It does not change the "theory of action" (Argyris and Schön, 1996, p.20), meaning that the focus is only on the symptoms of the problem, and not the underlying cause. Therefore, this is a lower level of learning. Double-loop learning on the other hand, focuses on changes "in values in theory in use" (Argyris and Schön, 1996, p.21), meaning that it goes deeper into the root causes of the problem.

Organisation size, complexity, and number of levels in the organisation are factors affecting what type of organisational learning (single or double-loop) learning results in (Argyris and Schön, 1996, p. 25–26).

3 METHODOLOGY

3.1 *Literature review*

A literature review was undertaken to get an overview of literature treating the relation between accidents and learning. The focus was on the construction industry, however papers discussing other industries were also looked at to find good examples and general knowledge about the topics of safety and learning.

Searches for literature were made in the following databases: Scopus, Googles Scholar and Oria. The search strings used were: accidents, learning, and construction. Searches were made both on two of the words at a time, and on all three at the same time.

In Scopus, the review was undertaken in a systematic way, which means it is a replicable, scientific and transparent process (Bryman, 2012, p.102). The search using "accidents" and "learning" as search string resulted in 4,983 results. From the 4,983 documents found from 1930–2017, the majority is published in the 2000s. After round of sorting out, first based on titles and years, then on abstracts, in total 34 articles were found as the most relevant concerning different fields, however not all were accessible.

The aim of the literature review was to get a background for the data collection.

3.2 *Interviews*

The first round of interviews was undertaken with actors in the Norwegian construction industry on information flow and knowledge transfer after accident investigations.

An interview guide was made with the following topics: general introduction, procedures for accident investigations, results of accident investigations, information flow, learning arenas, improvement potential and closing questions. The questions in the interview guide were adjusted to the three different actors (clients, contractors and consulting engineers).

In total 13 interviews with 19 persons responsible for Health, Safety and Environment (HSE) at clients, contractors and consulting engineers were undertaken. Table 1 presents an overview of the interviewees.

All the interviewed companies are large, professional companies that are well established in the Norwegian construction industry.

Interviewees were recruited through convenience selection, through contact persons in the industry. Further selection of interviewees will be done strategically to cover the construction industry widely.

The interviews were conducted between October 2017 and January 2018. Each interview took from 30–80 minutes. Most of the interviews took about an hour. Eight of the interviews were conducted in person, and five over phone. All the interviews except one, were recorded and transcribed. For the one that was not recorded, detailed notes were taken. The interviews were transcribed in NVivo, and coded according to the interview guide. Additionally, new codes were created while going through the data. A first, preliminary review of the data was done, resulting in main topics for discussion.

3.3 *Methodological considerations*

This research only comprises of a smaller sample of interviews and the preliminary results from these. To get a more general picture of the state-of-the art of the whole industry, more interviews will be undertaken, and preferably other methods should be used as well (e.g. questionnaire).

Table 1. Overview of interviewees.

Actors	Interviews	Documentation
Clients	2 interviews	1 company
Contractors	8 interviews^	3 companies
Consulting Eng.	3 interviews*	1 industry association

^one group interview with three interviewees.
*one group interview with representatives from five companies.

4 FINDINGS

4.1 *Accident investigations in practice*

The research shows that there are large variations when it comes to accident investigation practices between companies (within different actors) in the construction industry, and also between different projects within a company. The variations concern both resources available, investigation competence and investigations execution (i.e. methods used). Companies have their own criteria for when and how investigations should be performed, also these vary between the companies.

Mostly the accident investigations are undertaken separately by different actors, however interviews are conducted with persons in different companies during the investigations if it is relevant for the investigation. This results in separate investigations at clients, contractors and sub-contractor if more of the actors are deciding to undertake an investigation.

The consulting engineers reported that they are usually not involved in accident investigations, unless the unwanted event is directly caused by a calculation error performed by them.

Some companies use external parties to undertake investigations, others mainly perform internal investigations. It was also reported by one interviewee at a client, that they sometimes have to request contractors in order for the contractors to perform accident investigations.

Competence was seen as a success factor for performing good investigations However, the research shows that the knowledge and experience of HSE-managers about accident investigations varies. It was pointed out by the interviewees that methods and tools to be used for accident investigations should be pre-defined, and the methods and tools should be easy to use to ensure that the investigations can start quickly after the accident, and in order to conduct the investigations in a good way to obtain learning. Some of the HSE-mangers found this to be unclear in their company. One of the interviewees stated that it is important to go deeply into the causes to prevent future accidents:

> *"The most important thing is the learning one can get out of accident investigations. That must be the main goal. If you really manage to uncover the root causes, that is when you have the opportunity to prevent the same from happening again. That must be the foremost goal." (HSE-manager, contractor).*

Further, some of the HSE-managers reported that the roles and responsibilities in terms of who is responsible for the accident investigations and follow-up of it in the projects and in the companies, are sometimes unclear.

4.2 *Results of accident investigations*

The results of the accident investigations included investigation reports, learning sheets and changes of procedures. One of the companies had put together experiences from several accidents into a short film. Most companies finalise accident investigation with an investigation report. The reports vary between companies in size and content.

Learning sheets, which are short one page summaries of accidents, have started to become increasingly popular. The drawback that was pointed out with these by some of the interviewees, is that they do not go deeply into the causes and are more like event descriptions. Further, some remarked that the focus on these learning sheets as an answer to the challenge of learning and knowledge transfer is too large.

> *"Generally, I think that there is a large focus on learning sheets and sharing of learning sheets, as if they solve everything. I think perhaps it is somewhat too much focus on only this one solution" (HSE-manager, contractor).*

4.3 *Information flow of the accident investigation results*

The results of the investigations are mainly distributed within the company which undertakes the investigation. Some companies have systems for sharing results after accident investigations within the company, such as management systems, procedures, and best practice databases.

Information sharing across companies is even lesser systematised. There are no automatic mechanisms for sharing results between companies. In one contracting company, it was reported that if the unwanted event happened at a sub-contractor, and the main contractor or client investigated the event, the sub-contractor would have to request the report in order to get it. In the same way, the NLIA can request access to investigation reports from companies. It was also reported that the NLIA sometimes requests companies to make investigation reports. However, it was mentioned that this could affect what the companies put into the report, as they would not want to face additional consequences.

Further, "breakfast-meetings" that some companies hold after accidents, were perceived as very good knowledge sharing arenas across companies. At such breakfast-meetings, a company shares experiences from an accident they think the industry as a whole can learn from. These meetings are held rather seldom, and are suited only for certain types of incidents, e.g. general activities that resulted in an accident or near accident, and where good measures to prevent this type accidents are found.

Another arena for information sharing that was mentioned by several of the interviewees, were workshops held by the NLIA. These workshops were perceived as a good for knowledge sharing. Additionally, HSE-conferences (e.g. SHA-dagene, HMS-konferansen) were other examples of knowledge sharing arenas. These are large conferences that occur yearly, which mostly mangers with exceptions attend. However, not all the interviewees were aware of these arenas, and it was pointed out that the events are occasional.

How information is shared between companies, is in large degree steered by the systems within the companies, and the contracts between companies. It was also mentioned that a client or a main contractor can put requirements regarding incident and accident reporting into contracts to easier obtain safety information from projects.

4.4 *Utilisation of accident investigation results*

Experiences of the interviewees show that investigation competence in the investigation team is important for the outcome of accident investigation. Further, the team compositions regarding the members´ role in the event is also important, so the persons are not too closely related to events or persons affected in the event. The members of the team should not have a conflict of interest with the investigation.

Further, it was mentioned by the interviewees that certain events are better suited to learn and share knowledge from than others. The outcome or consequence of the event (e.g. fatality, serious injury etc.) in large degree influence how the results are used further. In cases where the events result in police investigations and legal proceedings, there can be resistance that will be of disadvantage for the results and for the learning process. Especially, near-misses and high potential incidents (HIPOs) which have not become police cases are good to learn from, as the question of guilt in larger degree is eliminated.

Several of the interviewees mentioned the question of guilt as a factor that impede knowledge sharing, as this concerns the reputation of the company, future projects as well as compensations for injuries.

Further, it was mentioned that it can be challenging to share information in cases of serious

injuries where police investigations are undertaken, as these often take long time. This leads to the company accident investigation report being held back and thus delayed, also delaying the learning processes.

5 DISCUSSION

5.1 *Deficits with accident investigations*

To be able to learn from something that has happened, information about what happened and why it happened is needed. Accident investigations are important to gain this information. Gibb et al. (2014) highlight the importance of going in depth into accidents and finding underlying causes of accidents for a good learning outcomes.

In the construction industry in Norway there are no standardised methods to investigate accidents. Accidents largely vary when it comes to type, size and severity, and different accidents may therefore require different types of investigations. One important issue to research in relation to accidents is how accidents are selected for investigation (Lindberg et al., 2010). The criteria for investigating accidents and the degree of investigations vary between companies, and even between projects within a company. This is a challenge when it comes to learning after accidents, as the investigations vary and thus give different foundations for further work with safety.

The quality of accident investigation results is in large degree dependent on the investigation team; their relation to the accident and to the company, knowledge about the industry, investigation knowledge and experience. The knowledge and the experience of the responsible persons in the companies varies as seen in the interviews undertaken, and this affects the outcomes of the investigations. Le Coze (2013) highlights the importance of expertise on accident models, to apply them in proper way. This was also states by some of the interviewees.

Further, as mentioned, the construction industry is characterised by having many actors, many phases, and constant progress and changes in the projects. The cooperation between levels of actors, between different phases of construction projects, between companies in the same phase performing different operations, and between operations within a company is important for good safety. From what is seen in the interviews, there is not much cooperation on accident investigation between companies that are involved in an unwanted event. Mostly, the investigations are performed separately between companies if more companies are undertaking investigations. The weakness with this is that important viewpoints and the causes behind the event can be overseen, due to lack of specialised knowledge (e.g. when consulting engineers are not involved), but also that other involved companies do not get access or ownership of the investigation results and measures suggested in the investigation report to prevent future similar accidents. Lundberg et al. (2009) write in their paper about WYLFIWYF (What You Look For Is What You Find), which shows the importance of using several perspectives in accident investigations, weather it is accident models, methodologies or specialists.

If the aim of the investigation is also to learn from what has happened, the learning perspective should also be integrated into the investigation, to provide for information that will lay the groundwork for learning.

5.2 *Knowledge transfer as a premise for learning*

Nonaka and Takeuchi (1995) describe knowledge as different from information as it is about beliefs, commitment and actions. Both have in common that they are about meaning. Simply said: *"Information is a flow of messages, while knowledge is created by that very flow of information, anchored in the beliefs and commitment of its holder"* (Nonaka and Takeuchi, 1995).

After accident investigations, the knowledge obtained needs to be shared if learning from previous accidents is the goal. The importance of how this knowledge is shared for learning is also highlighted by Lindberg et al. (2010). Drupsteen and Guldenmund (2014) point out that there often are limited processes to follow-up learning after accidents, and that such knowledge is often shared through one-way communication, which does not encourage interaction and thus learning processes. The findings of the current research are similar; uncertainties about who should follow-up the accidents were found, as well as examples of one way dissemination of accident investigation results (e.g. learning-sheets).

Further, the way information is shared is another challenge in the industry seen from the interviews. Internally, companies might have some systems or ways to share information, however they are not necessarily good enough to share information with all levels in the company. Within companies, results are often shared though learning sheets. These are meant as an information sharing arena for all levels in the company; from the top to the sharp end. However, different users require different degrees of details of the information. In example, for other HSE-managers, the information which is on the learning sheets might be too vague to be useful for safety work.

One further deficit as the research shows, is that this information and knowledge is not in large scale shared across companies. A good platform

for sharing information across the industry is missing, even though there are a few conferences and other smaller arenas where some experiences can be shared. A knowledge sharing platform can be one solution for sharing information and experiences across companies in the industry, e.g. a common database. An accident data base could be used to collect all severe accidents in the construction industry. Accidents with a potentially serious outcome also need to be registered. Having set criteria for systemising the accident types, causes and possible use would be useful for the user of such a database.

The knowledge from the investigations can serve several purposes such as input for risk assessments, decision making and to create awareness about important circumstances that can affect safety. To make use of such information companies and the industry need to have certain tools available. Information needs to be shared internally in the company, and externally for the whole industry to improve.

Lingard and Rowlinson (2005, p.366) write that learning from past accidents is important for safety management, and in an organisational context an incident information systems must be available to collect, analyse and create preventive measure. However, only having system is not enough according to them. It is also important to be aware of how the organisation is currently running, and having a vision for the desired safety work and performance, the management's safety focus and safety work being an integrated part of the operations is highlighted.

5.3 *Learning*

According to Nonaka and Takeuchi (1995) learning can be looked at as a dynamic spiral, between the two learning loops that Argyris and Schön (1978) describe. The spiral goes between tacit and explicit knowledge and from explicit to tacit through four phases. The spiral goes on as *"organisational knowledge creation is a continuous and dynamic interaction between tacit and explicit knowledge"* (Nonaka and Takeuchi, 1995, p. 70).

Lingard and Rowlinson (2005, p.365) highlight the need for collective learning in the construction industry and the current lack of this. They write that similar accidents reoccur in the industry across countries, and yet the industry does not manage to improve the occupational safety enough.

In relation to the construction industry and safe-ty, it is therefore important to acknowledge the individuals in the organisation when creating and implementing measures for accident prevention. In the same manner, during accident investigations, tacit knowledge should be a part of the information foundation in an investigation, as when it goes back as learning points.

Drupsteen and Guldenmund (2014) point out that it is hard to identify organisational factors and managerial weaknesses that are root causes of events, which limit the possibility of double-loop learning. Le Coze (2013) suggest more cross-disciplinary research on learning from accidents. This shows the need for more research on the topic, and combining different topics together.

5.4 *Input for safety management*

Which results that can be used from an accident investigations for proactive safety management, depend on the type of accident, the outcome of the investigation as well as the way the information it is shared. It is suggested to make specifications and criteria related to characteristics of accidents (e.g. types of accidents, causes, processes) in order to decide what learning purposes they can serve.

Results of accident investigations can in example be used as input for proactive safety management, e.g. in the Safety Management System (SMS) of a company, and as an input in building information models (BIM) which can include early phase actors (i.e. consulting engineers) in the learning loop. One of the challenges for consulting engineers when it comes to occupational safety during construction, is that they in small scale get feedback if their solutions could be executed safely in practice by construction workers. By using new solutions and tools (e.g. digital solutions), these actors could easier be involved in occupational safety work.

The results can also be adapted to serve safety purposes in different processes during a construction project. Procurement processes both in early phases of a project as well during construction put a foundation and boundaries for safety. It is important to transfer knowledge also to these processes from accident investigations, to reduce the safety risks during construction.

5.5 *Coping with the diversity of actors*

The industry is, as mentioned before, diverse and numerous actors are involved in construction projects. This diversity pose a challenge in relation to learning, as different actors have different needs and requirements. This means that adaptation is required when it comes to ways of sharing knowledge and learning. A flow of information is required both between the different levels, and at each of the levels.

To analyse this diversity of actors and activities Pryke (2012) suggests a graphical representation and a social network analysis of how the specific actors and activities are related. Such an analysis

could be linked to safety management. Mapping the relations and information flows after accident investigations might give a better understanding of deficits in communication and knowledge sharing. The different actors in the actor chain in the construction industry, introduce boundaries and conditions that affect safety. A social network analysis can also be used to map other factors such as frame conditions (e.g. contract conditions) and how they affect different actors (Pryke, 2012).

Different actors have different roles in safety management, and this apply also for learning. By mapping relations in the construction network, information flow, finding out who has which needs, and who should facilitate whom, can help in knowledge transfer and learning.

It is suggested to perform a similar analysis as the social network analysis on safety information flow after accident investigations.

6 CONCLUSIONS

Preliminary findings of the research show the importance of coordination and cooperation between actors of construction projects. Accident investigations are important to avoid that similar accident reoccur, however there are elements that hinder knowledge transfer and learning from earlier accidents. Accident investigations in large degree vary between projects, clients and contractors. Often each actor performs individual investigations with limited sharing of the results across companies. Consulting engineers are rarely involved in investigations, unless the problem has clearly been related to calculations.

Having a good knowledge foundation, based on facts including root causes, is crucial to ensure that correct measures are taken after accidents and to enable learning. For this it is important with competence and experience of the investigation team. It was found that experience and knowledge about accident investigations of the HSE responsible persons in companies varies a lot.

To obtain learning after accident investigations, information must be shared. Certain types of accidents are more suitable for sharing and learning purposes, e.g. near misses and high potential incidents. It is suggested to make specifications and criteria related to characteristics of accidents for specific learning purposes.

Information sharing after accident investigations mostly happen within companies. Between companies, knowledge sharing and learning is not systemised and occurs occasionally, and tools to share information between companies are lacking.

The large diversity of actors in the industry challenges practices, information sharing and learning processes. To enable learning in the industry both across organisations and within organisations, there is a need to understand the different relations between actors, processes and needs. A social network analysis of the information flow of the results after accident investigations in the construction industry might help to find the deficits as well as the centre points of communication and relations between actors, that might enable knowledge transfer and learning.

REFERENCES

Argyris, C. & Schön, D.A., 1978. Organizational learning: a theory of action perspective, Reading, Mass, Addison-Wesley.

Bryman, A., 2012. *Social research methods*, Oxford, Oxford University Press.

Drupsteen, L. & Guldenmund, F.W., 2014. What Is Learning? A Review of the Safety Literature to Define Learning from Incidents, Accidents and Disasters. *Journal of Contingencies and Crisis Management*, 22, 81–96.

Gibb, A., Lingard, H., Behm, M. & Cooke, T., 2014. Construction accident causality: Learning from different countries and differing consequences. *Construction Management and Economics,* 32, 446–459.

Le Coze, J.C., 2013. What have we learned about learning from accidents? Post-disasters reflections. *Safety Science*, 51, 441–453.

Lindberg, A.K., Hansson, S.O. & Rollenhagen, C., 2010. Learning from accidents—What more do we need to know? *Safety Science*, 48, 714–721.

Lingard, H. & Rowlinson, S., 2005. *Occupational health and safety in construction project management,* London, Spon Press.

Love, P.E.D., Lopez, R. & Edwards, D.J., 2013. Reviewing the past to learn in the future: Making sense of design errors and failures in construction. *Structure and Infrastructure Engineering,* 9, 675–688.

Lundberg, J., Rollenhagen, C. & Hollnagel, E., 2009. What-You-Look—For-Is-What-You-Find—The consequences of underlying accident models in eight accident investigation manuals. *Safety Science*, 47, 1297–1311.

NLIA, 2015b. KOMPASS Tema nr. 3 2015 Arbeidskadedødsfall i Norge. Utviklingstrekk 2009–2014 og analyse av årsakssammenhenger i fire næringer. [Work fatalities in Norway] Trondheim: Norwegian Labour Inspection Authority

NLIA, 2016. KOMPASS Tema nr. 8 2016 Ulykker i bygg og anlegg 2015 [Accidents in the construction industry 2015] Trondheim: Norwegian Labour Inspection Authority

NLIA. 2015a. KOMPASS Tema nr. 4 2015 Skader i bygg og anlegg: Utvikling av problemområder [Accidents in the construction industry: Development of problem areas] Trondheim: Norwegian Labour Inspection Authority

NLIA. 2017. KOMPASS Tema nr. 1 2017 Helseproblemer og ulykker i bygg og anlegg [Health problems and accidents in the construction industry] Trondheim: Norwegian Labour Inspection Authority

Nonaka, I. & Takeuchi, H., 1995. The knowledge-creating company: how Japanese companies create the dynamics of innovation, New York, Oxford University Press.
Pryke, S., 2012. *Social Network Analysis in Construction,* Oxford, UK, Oxford, UK: Wiley & Blackwell.
Rasmussen, J., 1997. Risk management in a dynamic society: a modelling problem. *Safety Science*, 27, 183–213.
SSB, 2017b. Accidents at work. [Internet]. Version corrected 03.10.2017. Accessed: 12.02.2018 Available from: https://www.ssb.no/en/helse/statistikker/arbulykker
SSB. 2017a. Construction, structural business statistics [Internet]. Updated: 02.06.2017. Accessed: 12.02.2018 Available from: https://www.ssb.no/en/bygg-bolig-og-eiendom/statistikker/stbygganl
Working Environment Act, 2005. Lov om arbeidsmiljø, arbeidstid og stillingsvern mv. nr. 62 av 17. juni 2005. Arbeids—og sosialdepartementet.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Personal protective equipment detection in industrial facilities using camera video streaming

C.B. Souto Maior, J.M. Santana, L.M. Nascimento, J.B. Macedo, M.C. Moura & D.L. Isis
CEERMA—Center for Risk Analysis, Reliability and Environmental Modeling, Recife, Brazil
Department of Production Engineering, Universidade Federal de Pernambuco, Recife, Brazil

E.L. Droguett
CEERMA—Center for Risk Analysis, Reliability and Environmental Modeling, Recife, Brazil
Department of Production Engineering, Universidade Federal de Pernambuco, Recife, Brazil

ABSTRACT: Organizations and industries must ensure safe operation of their facilities, employing rigorous risk management techniques for planning and executing their activities. The use of Personal Protective Equipment (PPE) represents the closest layer of protection to workers, and can considerably reduce the risk of exposure to hazards, being critical for safety in industrial environments. The hazards addressed by protective equipment include physical, acoustic, electrical, heat and chemicals. Despite the substantial efforts in increasing awareness about the benefits of PPE to strive towards zero accident philosophy, operators often neglect its use when not being supervised. However, organizations commonly have surveillance cameras installed which might provide useful visual information on correct usage of PPE. In this context, computer vision is an interdisciplinary field that seeks to automate tasks that the human visual system can do and includes domains of signal and image processing, pattern recognition and artificial intelligence. Moreover, object recognition is a prominent technology from computer vision for finding and identifying objects in an image or video sequence. Then, this work aims to create an automatic PPE detection from surveillance cameras and other video streams using computer vision and machine learning. Equipment such as helmets, safety glasses, earplugs and other garments are checked for whether they are being used by operators in a real-time monitoring, alerting supervisors to prevent accidents and ensure a safer environment.

1 INTRODUCTION

Even with the scientific and technological progress, statistics provided by the International Labour Organization (ILO) demonstrate that working conditions in many countries (e.g. European Union) have not changed to such a degree as to significantly reduce the problem of occupational injuries (Cavazza & Serpe, 2009). Therefore, every effort to decrease the number of accidents or, at least, maintain its rate at an acceptable range is highly important, and can be employed either by organizational actions, collective training or individual safeguard.

The traditional approach to avoid loss is the implementation of barriers, which plays a central role in the prevention of accidents. Sklet (2006) defines safety barriers as 'physical and/or non-physical means planned to prevent, control, or mitigate undesired events or accidents'.

Indeed, there are many opportunities to interrupt or change an accident sequence of events before it evolves into a loss. First, an answer is to change the preconditions for an accident to occur by eliminating the energy source or modifying the energy characteristics from the hazard. Second, barriers may interrupt, dilute, or redirect the energy flow during the latter part of the accident process (e.g. separating the victim from the energy flow). As last barrier, it is possible to improve the victim's ability to endure the energy flow (e.g. wearing some protective equipment), which is the ultimate protection to avoid damage (Kjellén and Albrechtsen, 2017).

In this context, Personal Protective Equipment (PPE) is usually adopted to protect the individual against health or safety risks at work. It includes items related with protection of head, face, eye, hand, arms, and legs (Health and Safety Executive, 2013). There are consolidated regulations for the usage of PPE in industries (Occupational Safety and Health Administration—OSHA 2004; U.S. Homeland Security 2002) that aim to decrease the frequency of misuse or absence of PPE. Also,

PPE's positive impacts are very significant (e.g. rate of eye injury and lost work time can be reduced by 50% or more when PPE is worn (Lipscomb, 2000).

Head, as a vital body part containing possibly the most important human organ, needs appropriate attention. Every year, approximately 1.7 million people are hospitalized or die as a result of a traumatic brain injury (TBI) only in the United States (McCrory et al., 2009). Protective headgear and helmets decrease the potential for severe TBI following a collision by reducing the acceleration of the head upon impact, thereby decreasing both the brain-skull collision, as well as the sudden deceleration induced axonal injury (Newman et al., 2005).

There are several types for head protection such as industrial safety helmets, bump caps and firefighters' helmets. The use of those equipment is necessary in activities like low-level fixed objects with risk of collision (e.g. pipework, machines, scaffolding) and transport activities involving the risk of falling material (e.g. hoists, lifting plant, conveyors) (Health and Safety Executive, 2015). The energy absorbing material of a helmet compresses itself to absorb force during the collision and slowly restores itself to its original shape. This compression and restoration has the effect of prolonging the duration of the collision, while reducing the total momentum transferred to the head (Pellman et al., 2006).

The problem relies on the fact that, even with understanding about the safety improvement that the usage of PPE leads, its usage is often neglected in industry. The report of the ILO estimates that 2.34 million people die every year in the world due to occupational accidents, some of these deaths caused by non-use of PPE (International Labour Office, 2011). A common approach is to impose fines and penalization to workers, who do not wear the required PPE when performing specific activities. However, supervision to guarantee its use is normally performed in person by a higher-level employee, which makes almost impossible to control all operators during the whole labor time.

Indeed, there is an extensive discussion concerning ethical issues in workplace surveillance, referring to management's ability to monitor, record and track employee performance, behaviors and personal characteristics in real time (Ball, 2010). Most of the discussion involves the so-called Electronic Performance Monitoring (EPM) about employee's control in social and technological forms (e.g. Internet and email monitoring, location tracking, biometrics) and the understanding of privacy boundaries surrounding employee information (Alder, 1998) (Allen et al., 2015). However, our discussion is to assure that proper safety protocol is followed, preventing injury to employees, as well as avoiding damage to the assets through a consistent and trustworthy model.

Therefore, an automatic method for monitoring PPE usage presumably is significant worthy for industrial safety, representing an impactful opportunity for the use of Computer Vision (CV). CV is an interdisciplinary field aiming to investigate and develop computers with high-level understanding from digital images or videos, describing the world that we see and to reconstruct its properties (Szeliski, 2010). From the perspective of engineering, it seeks to automate tasks that the human visual system can do (Sonka, Hlavac and Boyle, 2008). The development of high-powered computers, the availability of high quality and inexpensive video cameras, and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms in CV field (Yilmaz, Javed and Shah, 2006). Therefore, this paper aims to develop a model for automatic PPE detection from industrial video streams using computer vision and machine learning, employing modern technologies to create tools capable of modifying and innovating methods currently used in industries.

The rest of this paper is organized as follows: Section II introduces some ideas and concepts on CV, while Section III presents some works of object/person detection, describing the methodology applied in PPE detection. Section IV demonstrates the developed model. Section V provides possible usages as decision supporter and Section VI concludes remarks.

2 COMPUTER VISION

Computer Vision (CV) studies the automated extraction of information from images and videos. Information can mean anything from 3D models, camera position, object detection and recognition to grouping and searching image content (Jan Erik Solem, 2012). CV gathers knowledge from many fields, such as image processing, pattern recognition, mathematics and artificial intelligence. One of its main goal is to enable computers to reproduce core functions of human vision, such as motion perception and scene understanding.

Hence, visual object tracking have been constantly studied and presents three key steps for detection in video analysis: detection of movement of objects, tracking of such objects from frame to frame, and analysis of object tracks to recognize their behavior (Yilmaz et al., 2006). Essentially, the basis of visual object tracking is to robustly estimate the motion state (i.e., location, orientation, size, etc.) of a target object in each frame of an input image sequence (Li et al., 2013).

Specifically, intelligent visual surveillance systems deal with the real-time monitoring of persistent and transient objects within a specific environment (Valera & Velastin, 2005). The goal of these systems is not only to put cameras in the place of human eyes, but create an entire surveillance system as automatically as possible (Hu et al., 2004).

There exist some well-known visual surveillance systems such as W4 (Haritaoglu et al., 2000); Haar-wavelet Adaboost (Enzweiler & Gavrila, 2009) and ViBe (Barnich & Van Droogenbroeck, 2011), mainly developed to detect different vehicles types, groups of people, pedestrians, people access control. Every system is developed seeking to compensate the capability limitation of human operators in monitoring enormous number of cameras at the same time. Thus, exploring similar tools and challenges to detect usage of PPE in order to avoid accidents in industries represents an interesting case.

3 REAL TIME OBJECT DETECTION

Techniques from statistical pattern recognition have, since the revival of neural networks, obtained a widespread use in digital image processing (Egmont-Petersen et al., 2002). Due to the outstanding work of Krizhevsky et al. (2012) for image classification, Deep Neural Networks (DNN) have been successfully studied in different fields of application such as speech recognition (Hinton et al., 2012), vibration analysis (Guo et al., 2016), electronic nose data (Längkvist et al., 2013) and physiological data (Mirowski et al., 2008). However, surely, the most promising results are found in the field of computer vision, bringing impressive developing in tasks like automatic object and face recognition.

One promising, open and free project that uses DNN for object detection is You only look once (YOLO). YOLO is a system for detecting objects and was first created on the Pascal VOC 2012 dataset, detecting the 20 Pascal object classes, such as person, birds, dogs, car, bicycle, bottle, table and chair, as could be seen in Fig. 1 (Redmon et al., 2015).

The developers adopt a different approach than the standard object detection models that uses classifier based-systems applied at multiple locations and scales in an image, which typically considered as detections high scored regions of the image. In YOLO, a single DNN is executed to the full image. This network divides the image into regions and predicts bounding boxes and probabilities for each region, with these bounding boxes being weighted by the predicted probabilities. It has considerable

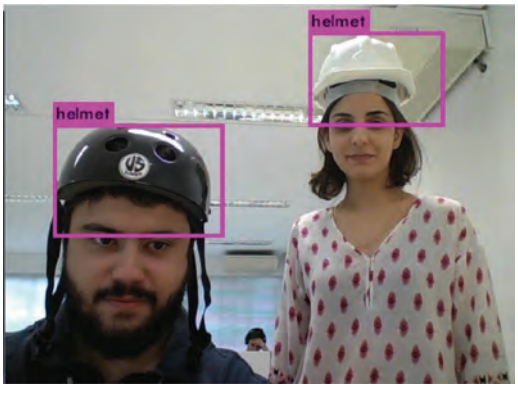


Figure 1. Example of object detection using YOLO. Adapted from Redmon et al. (2015).

advantages to other object detection models, once it looks at the whole image, and then its predictions are informed by global context in the image (Redmon et al., 2015). It also makes predictions with a single network evaluation, making YOLO extremely fast, allowing usage even for computers without a Graphics Processing Unit (GPU).

Still, an improved model, YOLOv2, has already been developed. More robust, detecting more than 9000 objects without losing real-time performance, YOLOv2 is a state-of-art object detection system with results comparable or with even better than many other systems (Redmon & Farhadi, 2017). Moreover, the YOLO project is open, well described, easily explained and user-friendly to anyone, who has some basis in computer programming. It even demonstrates how to include objects that were not on its detection basis, how to process and train a new model, allowing adaptation for different purposes.

Hence, using YOLOv2 as a key tool, we trained a new model to automatically detect PPE usage. Specifically, we were interested in identifying whether workers were wearing or not a safety helmet when performing some activities in which the protection was required.

4 HELMET DETECTION USING YOLO

YOLO project easily provides a pre-trained model, which could be used as a basis for detecting new types of objects. As any machine learning algorithm, YOLO requires a training dataset that will 'teach' the machine how an unknow object looks like. For our specific goal, 731 images containing helmets were used to give sufficient information about its appearance. All images were obtained from ImageNet (Jia Deng et al., 2009), and loca-

tion of helmets in images were annotated manually. ImageNet is an image database organized according to the WordNet (Fellbaum, 1998) hierarchy, in which each node of the hierarchy is depicted by hundreds and thousands of images. It presents useful resource for researchers that needs image data, containing innumerous classes of items.

The network was trained for about 8 hours, running in a Nvidia GeForce GTX 960 m GPU, with 4GB of video random access memory (VRAM). Once the algorithm finished its training, our helmet detection model could be applied to a specific image or to a video stream, such as a camera feed, processing every frame. Our model runs in real-time, maintaining the frame rate of the camera (30 frames per second—FPS). Fig. 2 depicts the model applied to a standard web camera video streaming.

Then, a script was created to alert surveillance operators whether an abnormal situation appears (i.e. helmets were not detected). Due to simplifications proposed, the model aims to be applied in a room containing specific number of employees that should be using helmets. For each frame, our algorithm detects the use of helmet and counts how many are present in the scene. If the number of detected helmets is different from number of previously defined people in the room for more than a brief period (e.g. 10 seconds, or 30 seconds), then an alert was emitted. Fig. 3 shows a computer screen when an alert was presented (i.e. one person is not using helmet in the image).

Moreover, if the number of detected helmets does not return to its normal value, alerts may continue to be emitted, but the period between alerts may be defined as desired (e. g. 30 seconds, or two minutes). Properly adjusting this period avoids unnecessary alerts that will continually distract the surveillance operator even after recognition of first anomaly situation, while still reminding that an abnormal situation is ongoing.

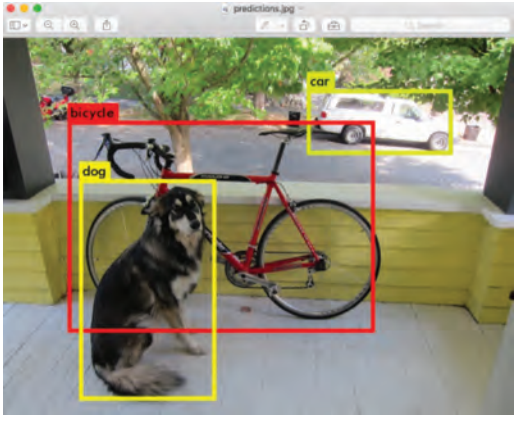


Figure 2. Helmet detection model applied for video streaming.

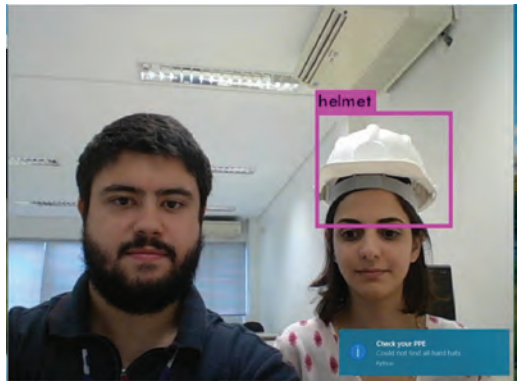


Figure 3. Alert emitted when model detect anomaly situation.

5 SUPPORTING DECISIONS

In practice, the presented model could be explored as a tool in different contexts, supporting decisions for the safety manager. The idea was to develop the model to be highly adaptable and manageable for various situations, providing specific information accordingly.

For example, as aforementioned, the alert period is easily adjustable to avoid unnecessary warnings. Still, other types of warnings (e.g. depending on how long operators remains without PPE; how many operators are not wearing the PPE) could be easily implemented and customable, providing information for the decision-maker to determine whether or not someone must be notified. Moreover, it is also possible to use information and statistics provided by the model (e.g. how many times alerts were displayed per day; who long operators had remained without PPE) as a safety indicator.

Still, implementation of a real time alert for the operators (e.g. a particular warning light is lit somewhere in the room) connected with the model would emphasize (or create) the sense of autoregulation among them, reducing (or sharing) the surveilling workload expected for the supervisor.

With further and wider adaptations, this surveillance technology could be also implemented to monitor other barriers than PPE. Related with the initial barriers where, usually, reliable sensors are already available, CV could act as a redundancy, rather than substituting existing technologies, aiming to improve detection effectiveness of hazards (e.g. fire, toxic gases) in order to interrupt the energy flow in case of accidents. For inner barriers, it is possible

to create alerts and warnings for whether an operator approaches a danger zone based on images of the area. In both previously mentioned barriers, CV would help to eliminate or reduce the consequences of unwanted energy flow, rather than dealing with the last safety impediment represented by the PPE.

6 CONCLUSIONS

This paper presented an approach for automatically detecting PPE usage in a controlled environment, using object detection with YOLO. By using YOLO, this method achieves a reasonable balance between speed and confidence, running in real-time, which results in relative low computational resource usage. Moreover, it is possible to adjust the model to different scenarios according to specific requirements. This could lead to beneficial results to safety engineering since the detection is performed automatically and does not require constant human attention.

As a matter of our current research, we aim at extending this model for application in a wider range of situations. For instance, YOLO can be trained to identify other types of PPEs, so that it could be used for simultaneously monitoring usage of different PPEs. Also, the script used for alerts could be improved, allowing this method to cover a wider range of scenarios. A further step is to use real surveillance videos as input, detecting the usage of PPE in realistic environment, preventing accidents and providing an improvement on the safety monitoring system of industries.

REFERENCES

Alder, G.S. (1998) 'Ethical Issues in Electronic Monitoring: A Consideration of Deontological Perspectives', *Journal of Business Ethics*, 17(7), pp. 729–743.

Allen, M.W., Coopman, S.J., Hart, J.L. and Walker, K.L. (2015) 'Workplace Surveillance and Managing Privacy Boundaries', *Labor History*, 51(1), pp. 172–200.

Ball, K. (2010) 'Workplace surveillance: an overview', *Labor History*, 51(1). doi: 10.1080/00236561003654776.

Barnich, O. and Van Droogenbroeck, M. (2011) 'ViBe: A universal background subtraction algorithm for video sequences', *IEEE Transactions on Image Processing*, 20(6), pp. 1709–1724. doi: 10.1109/TIP.2010.2101613.

Cavazza, N. and Serpe, A. (2009) 'Effects of safety climate on safety norm violations: exploring the mediating role of attitudinal ambivalence toward personal protective equipment', *Journal of Safety Research*, 40(4), pp. 277–283. doi: 10.1016/j.jsr.2009.06.002.

Egmont-Petersen, M., de Ridder, D. and Handels, H. (2002) 'Image processing with neural networks—a review', *Pattern Recognition*, 35(10), pp. 2279–2301. doi: 10.1016/S0031–3203(01)00178–9.

Enzweiler, M. and Gavrila, D.M. (2009) 'Monocular pedestrian detection: Survey and experiments', in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2179–2195. doi: 10.1109/TPAMI.2008.260.

Fellbaum, C. (1998) 'WordNet: An Electronic Lexical Database', *MIT Press, Cambridge, London, England*, p. 423. doi: 10.1139/h11–025.

Guo, X., Chen, L. and Shen, C. (2016) 'Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis', *Measurement: Journal of the International Measurement Confederation*, 93, pp. 490–502. doi: 10.1016/j.measurement.2016.07.054.

Haritaoglu, I., Harwood, D. and Davis, L.S. (2000) 'W4: Real-time surveillance of people and their activities', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), pp. 809–830. doi: 10.1109/34.868683.

Health and Safety Executive (2013) *Personal protective equipment (PPE) at work: A brief guide, HSE Books.*

Health and Safety Executive (2015) *Personal protective equipment at work.*

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. and Kingsbury, B. (2012) 'Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups', *IEEE Signal Processing Magazine*, 29(6), pp. 82–97. doi: 10.1109/MSP.2012.2205597.

Hu, W., Tan, T., Wang, L. and Maybank, S. (2004) 'A survey on visual surveillance of object motion and behaviors', *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 34(3), pp. 334–352. doi: 10.1109/TSMCC.2004.829274.

International Labour Office (2011) *ILO introductory report: global trends and challenges on occupational safety and health: XIX World Congress on Safety and Health at Work*.

Jan Erik Solem (2012) 'Programming Computer Vision with Python', *Programming Computer Vision with Python*, p. 264. doi: 10.1017/CBO9781107415324.004.

Jia Deng, Wei Dong, Socher, R., Li-Jia Li, Kai Li and Li Fei-Fei (2009) 'ImageNet: A large-scale hierarchical image database', *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. doi: 10.1109/CVPRW.2009.5206848.

Kjellén, U. and Albrechtsen, E. (2017) *Prevention of Accidents and Unwanted Occurrences*. CRC Press.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) 'ImageNet Classification with Deep Convolutional Neural Networks', *Advances In Neural Information Processing Systems*, pp. 1–9. doi: http://dx.doi.org/10.1016/j.protcy.2014.09.007.

Längkvist, M., Coradeschi, S., Loutfi, A. and Balaguru Rayappan, J.B. (2013) 'Fast classification of meat spoilage markers using nanostructured ZnO thin films and unsupervised feature learning', *Sensors (Switzerland)*, 13(2), pp. 1578–1592. doi: 10.3390/s130201578.

Li, X., Hu, W., Shen, C., Zhang, Z., Dick, A. and Hengel, A. Van Den (2013) 'A survey of appearance models in visual object tracking', *ACM Transactions on Intelligent Systems and Technology*, 4(4), pp. 1–48. doi: 10.1145/2508037.2508039.

Lipscomb, H.J. (2000) 'Effectiveness of interventions to prevent work-related eye injuries.', *American journal of preventive medicine*, 18(4 Suppl), pp. 27–32.

McCrory, P., Meeuwisse, W., Johnston, K., Dvorak, J., Aubry, M., Molloy, M. and Cantu, R. (2009) 'Consensus statement on concussion in sport: The 3rd International Conference on Concussion in Sport held in Zurich, November 2008', in *Journal of Athletic Training*, pp. 434–448. doi: 10.1016/j.pmrj.2009.03.010.

Mirowski, P.W., LeCun, Y., Madhavan, D. and Kuzniecky, R. (2008) 'Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG', in *Proceedings of the 2008 IEEE Workshop on Machine Learning for Signal Processing, MLSP 2008*, pp. 244–249. doi: 10.1109/MLSP.2008.4685487.

Newman, J.A., Beusenberg, M.C., Shewchenko, N., Withnall, C. and Fournier, E. (2005) 'Verification of biomechanical methods employed in a comprehensive study of mild traumatic brain injury and the effectiveness of American football helmets', *Journal of Biomechanics*, 38(7), pp. 1469–1481. doi: 10.1016/j.jbiomech.2004.06.025.

Occupational Safety and Health Administration (OSHA) (2004) *Personal Protective Equipment*. doi: 10.5923/j.safety.20160501.02.

Pellman, E.J., Viano, D.C., Withnall, C., Shewchenko, N., Bir, C.A. and Halstead, P.D. (2006) 'Concussion in professional football: Helmet testing to assess impact performance—Part 11', *Neurosurgery*, 58(1), pp. 78–95. doi: 10.1227/01.NEU.0000196265.35238.7C.

Redmon, J. and Farhadi, A. (2017) 'YOLO9000: Better, Faster, Stronger', *Conference on Computer Vision and Pattern Recognition*, 7(3). doi: 10.1142/9789812771728_0012.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2015) 'You Only Look Once: Unified, Real-Time Object Detection'. doi: 10.1109/CVPR.2016.91.

Sklet, S. (2006) 'Safety barriers: Definition, classification, and performance', *Journal of Loss Prevention in the Process Industries*, 19, pp. 494–506. doi: 10.1016/j.jlp.2005.12.004.

Sonka, M., Hlavac, V. and Boyle, R. (2008) *Image Processing, Analysis, and Machine Vision*, *Thomson Learning*. doi: 10.1007/978-1-4899-3216-7.

Szeliski, R. (2010) *Computer Vision : Algorithms and Applications*. Springer.

U.S. Department of Homeland Security (2002) 'Guide for the Selection of Personal Protective Equipment for Emergency First Responders', I(November), pp. 1–118.

Valera, M. and Velastin, S.A. (2005) 'Intelligent distributed surveillance systems: a review', *IEE Proceedings—Vision, Image, and Signal Processing*, 152(2), p. 192. doi: 10.1049/ip-vis:20041147.

Yilmaz, A., Javed, O. and Shah, M. (2006) 'Object tracking: A survey', *ACM Comput. Surv.*, 38(4), p. 13. doi: http://doi.acm.org/10.1145/1177352.1177355.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Norwegian police training in the use of force: A preparation for facing the realities of street challenges?

S. Vee Henriksen & A. Snortheimsmoen
Norwegian Police University College, Oslo, Norway

B.I. Kruke
Centre for Risk Management and Societal Safety, University of Stavanger, Stavanger, Norway

ABSTRACT: The Norwegian police have a tradition for reticent use of force. Norway is one of the very few countries in the world with unarmed police on daily duty. The aim of this paper is to study the degree to which the Norwegian Police University College (NPUC), and the police districts, training in the use of force reflects the need for reliable handling of the various weapons at the disposal of the police officers and how experiences on the use of force gained by police officers in their daily duty is made available for police training. However, there is no national strategy on collection, dissemination, interpretation and integration of experiences from the streets of Norway. Even though experienced police officers lecture and supervise students and fellow police officers, the lack of systematic collection of experiences hampers the quality of the training, especially at the NPUC, but also in the police districts.

1 INTRODUCTION

The Norwegian police have a tradition for reticent use of force, a tradition rooted in the Norwegian society (NOU 1981:35; Norwegian Parliamentary White Paper No. 42 (2004–2005)). However, from 2007–2016 the Norwegian police has experienced a sharp increase in armed assignments from 1507 in 2007 to 5816 by November 2016 (POD 2017a). It is mostly ordinary emergency response police personnel who handle the acute phases of armed assignments (Norwegian Parliamentary White Paper No. 21 (2012–2013)). Heightened terrorist threats and organised crime may expose the police to extreme situations characterized by rapid assessments and decisions, based on ambiguous information and the possibility of fatal consequences. Changes in street realities may challenge both operational training and practice in the use of force in the Norwegian police.

This paper aims to study the degree to which the Norwegian Police University College (NPUC), and the police districts, training in the use of force reflects the need for reliable handling of the various weapons at the disposal of the police officers facing street challenges, and how experiences on the use of force gained by police officers in their daily duty is collected, disseminated, interpretation and integrated in police training. The paper draws on observations made through many years of experience from operational service in the police. One of the authors is former head of the police counter terrorist unit (Delta), and another is a former lecturer and head of studies at NPUC. The paper starts with a historical introduction to the police use of force, followed by a presentation of the frameworks on crisis, situational awareness, decision-making, training, exercising and learning. We then present the results on operational use of force based on literature studies of the results of the Firearms Commission (NOU 2017:9) and of reports on the use of force. We also include findings based on participant observation in training sessions at the NPUC and in the police districts. These findings are then discussed together with the theoretical frameworks before we make some concluding remarks on police training on the use of force in relation to the operational needs on the street.

2 THE NORWEGIAN POLICE

A Norwegian national police force organized and employed by the state has existed since 1936. The Norwegian police is organized under the Ministry of Justice and Public Security, and managed by the Police Directorate (POD). All police services in the

12 police districts, and the five national specialized agencies[1] (POD 2017b), follow the same law, instructions and guidelines.

2.1 *The Norwegian Police and the Norwegian Police University College (NPUC)*

The NPUC is one of the national specialized agencies, and is responsible for all under graduate and post graduate education, including all police training in the use of force. The joint national basic education program secures the same basic training in the use of force and coercive means, including firearms. The police are divided into five competence levels for emergency response personnel (IP), according to the level of annual operational training (PBS 1:38):

IP 1: The police counter-terrorist personnel (Delta)
IP 2: The dignitary protection personnel
IP 3: The police special response unit personnel
IP 4: The police emergency response personnel (ordinary police personnel)
IP 5: The police emergency response personnel with limited operational training (not licensed to carry firearms).

To be authorized to carry firearms, emergency response personnel have to carry out at least 48 hours of operational training and pass a standardized annual firearms test to reach competence level 4 (IP 4). It is NPUC, along with POD, which determines the annual joint national operational training. The yearly training normally consists of firearms training, arrest techniques, first aid, tactics and situational training. In recent years training on how to handle an active shooter has been emphasized for all police emergency response personnel (Norwegian Parliamentary White Paper No. 10 2016–2017).

2.2 *Directives and instructions on the use of force*

There are no written directives on how the police should perform their operational training. NPUC develops the professional content on the basis of inputs from instructors, collaboration with POD and PJS (Police Joint Services), dialogue meetings with police districts, participation in national exercises and regular meetings with the National Reference Group for Police Operative Disciplines. The professional content of the training is detailed in relevant academic literature, in instructor manuals and in subject booklets, which ensures an equal approach in all police districts. Thus, at a strategic level, NPUC has a central role in developing and coordinating national training.

The specific instructions for use of force are provided in the Police Act, Police Instructions and Firearms Instructions (Lovdata.no 20.11.17). The use of physical force is authorized through law in the Police Act, is defined in the Police instructions and is essentially exercised on the basis of professional judgment and discretion. Use of coercive measures, such as pepper spray, baton and firearms, are strictly regulated in the Firearms instructions.

2.3 *Models for the force continuum*

The Norwegian police use a model, "Force pyramid" (see Figure 1), for visualizing coercive measures in a force continuum. The model shows how the use of force is increased the higher up in the pyramid one comes, based on the amount of physical and mental injury which may be expected from the respective coercive measure. The pyramid is not a complete and generally valid description of the use of force, but a ranking of different coercive measures (Lie & Lagestad 2011:10). However, the model is used in all training in the use of force at NPUC:

In this paper we limit the further discussion to *physical* force, shown as the steps above the dotted horizontal line in Figure 1. The first steps are use

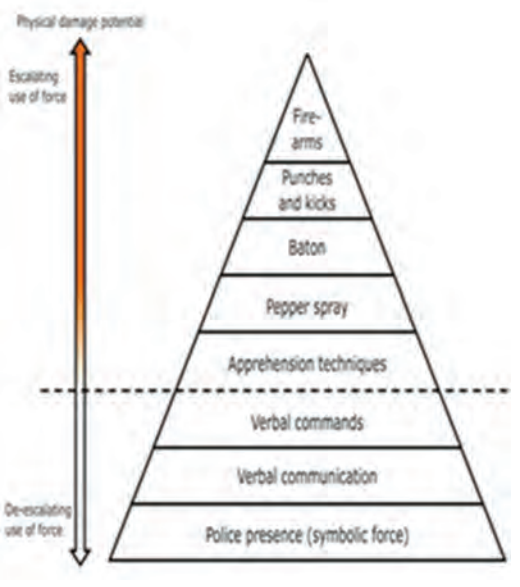


Figure 1. Force pyramid.

[1]Norwegian Police University College (NPUC), Central Mobile Police Service (CMPS), National Criminal Investigation Service (NCIS), Police Joint Services (PJS) and Norwegian National Authority for Investigation and Prosecution of Economic and Environmental Crime (NNAIPEE).

Figure 2. Police use of force model.

of arrest techniques in more controlled situations, the use of pepper spray, and the use of baton. The next step is the use of punches and kicks as self-defence in more uncontrolled situations. As a last resort, the police can use firearms. The police use of force may also be displayed in another model (see Figure 2).

This model, unlike the force pyramid model (Figure 1), also includes Electronic control devices (ECD, for example TASER) and the use of dogs, in addition to situational assessment. ECD is under review for possible acquisition. The police use of force model (Figure 2) is based on the New Zealand Police "Tactical options framework" model. This modified version was first presented at the annual research conference at NPUC in 2016 (Vee Henriksen 2016). The inner circle of the model describes the level of resistance faced by the police. The outer circle describes the coercive measures in relation to the level of resistance. Proportionate use of force is decided based on situational assessment, conducted by the police emergency response personnel.

It must be emphasized in both models that the police do not have to try every step before stronger force is used; this depends on the development of the situation in question. In daily service the Norwegian police will carry pepper spray and baton in the belt, but their firearms will be stored in locked boxes in the patrol cars.

3 CONCEPTUAL FRAMEWORKS

To be able to understand and discuss police use of force, we need to clarify our understanding of incidents, crisis, situational awareness and decision making. We then need to look at collection, dissemination, interpretation and integration of experience and finally learning about police use of force.

3.1 *Incident and crisis*

A crisis may be defined as "a serious threat to the basic structures or the fundamental values and norms of a system, which under time pressure and highly uncertain circumstances necessitates making critical decisions" (Rosenthal et al. 1989: 10). This definition points to critical decision-making in the midst of the uncertainty and time pressure of a dynamic incident or crisis. For the sake of this paper we will not distinguish between the terms incident and crisis. There is some research on differences between events such as incident, crisis and catastrophe (Engen et al. 2016), for instance due to their response needs. We will not go into this discussion in this paper, but use the terms incident and crisis independent of the size of the event in question.

Some crisis characteristics may be of particular importance in incident or crisis management. 't Hart and Boin (2001) distinguish crises based on their development and termination patterns. A fast burning crisis develops and terminates fast. A slow burning crisis has slow development and termination patterns. Gundel distinguishes crises based on their predictability and the degree to which they can be influenced (Gundel 2005). Unexpected crises are in his terminology difficult to predict, but fairly easy to influence. Intractable crises are easy to predict but difficult to influence. It is fair to assume that an unexpected crisis (Gundel 2005) with a fast developing pattern ('t Hart and Boin 2001) may be particularly challenging to manage for police officers.

3.2 *Situational awareness and decision-making*

Situational Awareness (SA) is crucial for decision-makers in dynamic incident management. Situational awareness is defined by Endsley as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley 1997:97). SA makes it possible to operate quickly and efficiently even during demanding operations (Lee and Kirlik 2013) necessitating swift decision-making, also on the use of force (including firearms). The quality of decision-making on modus operandi, and the ability to switch from one practice to another (Schakel et al. 2016), to manage the situation at hand, including the use of force, may have a crucial impact on the police officers and the public. Decisions may be based on analytic and/or intuitive reasoning

(Eid and Johnsen 2006, Helsloot and Ruitenberg 2004). The important issue initially is to reduce uncertainty, to find out what is going on. Boin and colleagues describe this initial phase as the sense-making phase, or the "what the hell is going on" phase (2005). In unexpected and fast developing situations, characterized by a high degree of uncertainty, it is fair to assume that intuitive decisions will be of particular relevance. Intuitive decisions are based on previous experience from similar situations (Eid and Johnsen 2006) and on improvisation (Weick 1993). Recognition of familiar patterns, déjà vu (Weick 1993), is a comforting thought for decision-makers in these situations, for recognition-primed decision-making (Klein 1989, 2011). Weick describes decision-making in this connection as «an act of interpretation rather than an act of choice» (1995:185). Furthermore, Weick (1995) points out that a good decision-maker is as good as his memory and the type of information that is stored there, and that good decisions are just as much based on a correct understanding of what has happened as on a correct understanding of what is going on. The level of expertise, based on experience, training and education, is therefore important for situational awareness and decision making.

3.3 *Training, exercises and learning*

Paoline and Terrill (2007) focus on experience as being the most important factor for learning within the police. Experiential learning is a form of learning that builds on practical situations and practical exercises that provide personal experience as well as learning (Kolb 1984), as learning by doing (Dewey 1938). Dreyfus and Dreyfus (1988) present five levels of proficiency; novice, advanced beginner, competent, proficient and expert. They focus on the development of occupational knowledge in an educational perspective: «If, and only if, experience is gained in this a theoretical manner, and intuitive behaviour replaces considered reactions, can mastery be regarded as developed» (Dreyfus and Dreyfus 1986:56). The transferring of experience and development of competence at the moment of learning however requires occupational experience to be passed on as knowledge-based learning. Thus, police officers experiential learning (Kolb 1984) need to by shared, disseminated, and collectively integrated and interpreted (Dixon 1999) in the police force and thereby made available for instructors and training sessions in the NPUC and the police districts.

4 EMPIRICAL FINDINGS

The empirical findings stem from participant observation in training sessions and operational use of force, and statistics from the Norwegian Police Directorate (POD), and the Firearms Commissions report (NOU 2017). Reference will also be made to earlier surveys relating to the subject.

4.1 *Use of force*

Even though risk assessments must not only take into account experiences but also cater for surprises and uncertainty (NOU 2012:14), proper data gathering is a vital foundation for these assessments. The Norwegian police do not collect data on the use of force in a structured way. Based on the lack of reporting systems in Norway, there is limited knowledge about to which extent and in which contexts the police exercise force. The only use of force that is being reported is in the form of statistics for armed assignments, when the police have threatened to use firearms and where they have actually used them. However, there is no national standard for reporting on the use of firearms. Thus, routines may vary amongst the police districts.

Figure 3 shows the development of armed assignments from 2007–2016. It is noted that the sharp decline in the period 2014–2015 is due to the temporary arming of the Norwegian police due to the increased threat of terrorism (NOU 2017:9). The temporary arming meant that all uniformed emergency response personnel approved to carry firearms, were required to carry a pistol during regular service. In most police districts, assignments that required armed response were no longer registered, as the police already had a general order for arming (therefore the decline in Figure 3 in this period). Some police districts, however, continued the normal practice of registration, despite the fact that the police already were armed (NOU 2017:9).

How many shots are fired by police officers each year? Figure 4 shows the numbers of shots fired at a person in the period 2005–2014.

Figure 4 shows an average of 2,5 shots fired at a person per year. In 2014, with just above 4000

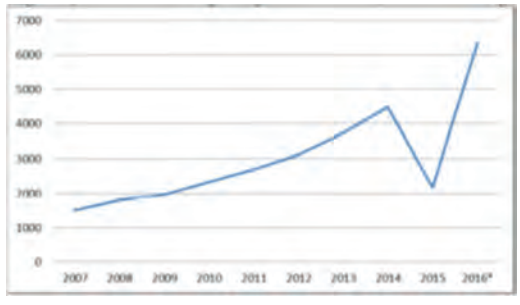


Figure 3. Annually reported armed assignments (POD 2017a:8).

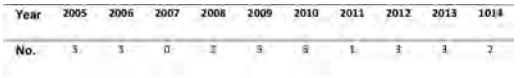


Figure 4. Annually reported shots fired at a person (POD 2015:3).

armed assignments (see Figure 3), 2 shots were fired at a person (Figure 4).

In situations where police officers have fired shots at a person, an investigation must be carried out according to the Norwegian Prosecution Instructions (Lovdata 22.08.16). The investigation will check the police officer's action only based on a *legal perspective*, and the incidents will not be evaluated for gathering knowledge in a *learning perspective*. All cases of police use of firearms in the period specified in Figure 4 have been filed without any form of legal consequences against the police. The number of shots fired at persons each year is low, and no substantial change was recorded in the number of situations where the police threatened to use firearms or actually fired shots during the period of temporary arming (NOU 2017:9:153).

Except for various standards of reporting related to the use of firearms, it is no national reporting system on the police use of force. Any documentation will only appear in the police local orderly book and/or administrative documents in the individual criminal cases. The lack of an overall police system for collection, dissemination, interpretation and integration of experience in the use of force has been pointed out in official reports (NOU 2009:12; NOU 2013:9). A comprehensive police analysis in 2013 pointed out that: «Despite the fact that expertise is a crucial input factor, the police today lack a comprehensive strategy and approach to the development and management of competence in the police. Competence and personnel development work in the police is to a lesser degree systematic» (NOU 2013:9, p. 209). In this context, it is reasonable to ask questions about the basis for Norwegian police training in the use of force. Is the training based on proven experience-based knowledge or based on assumed needs of the police? Very few surveys related to police use of force have been conducted in Norway. Lagestad (2008) conducted a study which included examining how often the police train and use arrest techniques. In this study the participating police officers assumed that they were using physical force approximately once a month. Lie (2010) prepared a report on mapping and testing of police knowledge and skills in arrest and control techniques. 90% responded using physical force to put someone on the ground once or twice the last two years, 16% had used pepper spray in the last year and 4.5 percent had used telescopic baton. According to Holmberg's (2013) report on the use of pepper spray in the Scandinavian countries, in Oslo Police District the use of pepper spray was reported approximately twenty-seven times a year on average in the period 2005–2011. It must be underlined that the Firearms Commission suggested that the Norwegian police should report on all use of force (NOU 2017:9). It is not known if and when this will be implemented.

4.2 *Temporary arming in 2014–2016*

Due to a heightened terrorist threat ordinary police officers were armed on daily duty in the period 25th November 2014 to 3rd February 2016. The Norwegian police officers were armed while on daily duty for the first time, with no time for preparations. In the Norwegian Official Report (NOU 2017:9) it is pointed out that the Police Directorate particularly noticed two main issues regarding the arming; (1) the risk for accidental discharges (AD), and (2) the possibility of the police being deprived of their own firearms. According to the Norwegian Official Report (NOU 2017: 9, 151) "it cannot be said that a central and systematic risk assessment of accidental discharges was made in advance of the arming."

4.3 *Accidental Discharges (AD)*

Prior to the period of temporary arming, there was no systematic registration of accidental discharges (AD) in the Norwegian Police Force, nor was there any clear specification of what should be considered as AD (NOU 2017:9). Guidelines for preventing ADs when the arming occurred were not in place. It was also not established a system for collection, dissemination, interpretation and integration of experiential learning from such incidents. However, AD became a very relevant issue during the temporary arming. In a letter of 10th August 2015 the Ministry of Justice and Public Security requested the registration of ADs during the temporary arming, and instructed POD to retrospectively map ADs in the period 2011–2015. Twenty-eight ADs were registered during the period of temporary arming, a sharp increase from earlier years, but earlier numbers are uncertain as they have been retrieved retrospectively (NOU 2017:9:147), and not based on a continuous systematic reporting. Figure 5 shows accidental discharges registered in the period 2011 – January 2016.

It must be underlined that the registered numbers of ADs do not include ADs during firearms training or when weapons have been emptied and checked in projectile collectors, unless resulting in injuries.

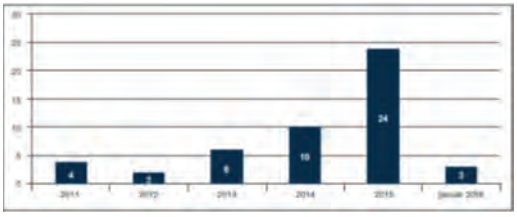


Figure 5. Shows AD registered in the period 2011-January 2016 (NOU 2017:9:148).

4.4 *Attempts to deprive the police of firearms*

During the period of temporary arming the Police Directorate received twenty-two reports of attempts to deprive the police of firearms. The numbers might be incorrect because the reporting was not mandatory until august 2015, nine months after the Police Directorate gave the order to arm all police emergency response personnel (NOU 2017:9). The attempts to deprive the police officers of their firearms occurred in different contexts. There were reports of several attempts to deprive the police of their firearms assisting medical care, include transportation of mentally ill persons. Similar incidents occurred when the police performed arrests. At least two such incidents resulted in injuries on police officers while defending their firearms.

4.5 *Police basic education on the use of force*

The Norwegian police basic education lasts for three years (a bachelor degree). The program consists of one year of theory at one of the NPUC`s three departments, a year of practice where each student is guided by an experienced police officer/tutor, and a final year of theory back at the NPUC. The operational education, in terms of training in the use of force, is primarily distributed throughout the second and third year. The operational training reflects that the Norwegian police officers still are unarmed in their daily duty. In the first year of studies there is no regular tactics or weapons training. However, the students receive basic training in arrest techniques, and training and approval to use OC pepper spray and the telescopic baton. In the second year the students undergo a two week course in basic tactics, situational training and basic training on Heckler & Koch MP5 (sub-machine gun). Tactical and situational training takes place without students carrying firearms during this course. The students also attend a basic course in the use of the Heckler & Koch P30 L pistol. Students will to varying degrees also participate in the operational training held in the police district in which they are assigned. This applies to both tactics training and weapons training. In addition, they continue to receive basic training in arrest techniques. During the third year the students have a number of exercises in situational training. In addition, they receive training in a simulator with exercises in situational awareness and decision-making connected to the use of force, including focus on the use of firearms. The simulator training focuses on the model for force continuum (see Figure 1) and the legal framework for police use of force. During this final year the students also continue training in arrest techniques, and have their final exam. During the last semester, students participate in a three-week course, which places special emphasis on firearms training and armed assignments. During this course, students also undergo the standardized annual firearms test for both firearms (pistol and sub-machine gun). When completed, they will be approved as police emergency response personnel (IP4), authorized to carry firearms in service.

4.6 *Use of experiential learning*

There is no position or function on a national level explicitly responsible for the collection of experiences from operational service, and making these experiences available for education and training in the police. Use of experiential learning is sporadic and dependent on individual initiatives. Experience is not collected in a systematic manner. We have also seen variations both between departments and police districts on collection, dissemination, interpretation and integration of experience-based learning. At the same time there are some good examples where experience-based operational training has been developed. The police counter terrorist unit (Delta) participated in the development of a concept for arresting a person with a knife based on experiences from such arrests. This concept is now mandatory training for all students at the NPUC.

The project manager for the development of the NPUC simulator training in the use of force, including firearms, used all available Decisions of Prosecution from the Norwegian Bureau for the Investigation of Police Affairs as a basis for the professional content of the simulations. All relevant legal assessments were reviewed and implemented in the training, and many of the scenarios that were recorded are based on real events. This training is also mandatory for all the students at the NPUC.

5 DISCUSSION

Is the police training in the use of force based on experiences from incident management on the streets, preparing students and police officers for facing the realities of street challenges? First of all we have to go back to our understanding of

incidents and incident management, and try to find out what characterizes assignments where coercive force may be used. As previously stated, there is a lack of reporting in connection with experiences gained by police officers in the use of physical force, except in the case of firearms. Reporting in the use of firearms is also fragmented and not based on a standardised format implemented by all police districts. There has been a sharp increase in the number of armed assignments in Norway. At the same time the number of fired shots from the police emergency response personnel stays at a very low level (se Figures 3 and 4). Armed assignments may suddenly arise, leaving limited time for preparation and decision making. Some assignments terminate without any kind of contact with a suspect, some result in arrests without use of force and a few results in confrontation with a suspect where various forcible measures are used. Armed assignments may escalate to a crisis as they represent a serious threat to the security of the people involved and the surroundings. Nevertheless, time pressure and highly uncertain circumstances may necessitate the need for critical decision-making (Rosenthal, Charles et al., 1989). Based on the numbers of armed assignments (see Figure 3) it is fair to assume that new incidents will arise, but it will be difficult both to depict the time and place of occurrence and therefore the ability to influence the development of the event in question. Unexpected crisis (Gundel 2005) with a fast developing pattern ('t Hart and Boin 2001) may therefore be difficult for police emergency response personnel to plan for and manage. The NPUC, and the rest of the police force, need to consider both expected and unexpected scenarios when planning and developing the proper competence level for police students and officers. It is difficult to predict the future challenges the police may face, fully implement measures related to preventing and preparing for worst case scenarios, and plan police preparedness in case of unlikely events. Risk assessments must not only take into account experiences but also cater for surprises and uncertainty (NOU 2012:14). Thus, a reliable modus operandi to manage the situation at hand may not only be based on "following the book". Reliable decision-making in a crisis situation may be the result of an ability to switch from one practice to another (Schakel et al. 2016), on intuitive decisions based on previous experience from similar situations (Eid and Johnsen 2006), on a feeling of déjà vu (Weick 1993), on improvisation (Weick 1993), and, finally, on recognition of familiar patterns (Klein 1989). These are capacities normally associated with highly experienced response personnel. However, mostly ordinary police emergency response personnel (IP: 4) manage the initial phase of an acute incident or crisis. They need to be trained and equipped to face such situations. How do the police use experience to manage these types of events in training and education? Besides the statistics on armed assignments and the use of firearms there is no national systematic collection of experience on the use of forcible measures for use in training and education, either at the NPUC or the police districts.

In armed assignments adequate situational awareness (SA) is crucial for decision-making undertaken by police officers, and thus the outcome of the situation. These assignments are often characterized by rapid development, uncertainty and lack of or ambiguous information. SA makes it possible to operate quickly and efficiently even during such demanding assignments (Lee and Kirlik 2013). We are well aware that SA is one of the most central elements in training on the use of force, both at NPUC and in the police districts. During simulator training at NPUC SA is constantly referred to in connection with the choice of coercive measures based on the force continuum models (Lie and Lagestad 2011, Vee Henriksen 2016). As a result of their SA the police emergency response personnel decide whether or not to use force, and the appropriate kind of force and forcible means to be used. In situations where the police are armed, decision making will be particularly important since the use of firearms may have fatal consequences. To make the right decisions Eid and Johnsen (2006) point out that intuitive decision making will be of particular relevance as it is based on experience from similar situations. Weick (1995) also underlines that a good decision maker is as good as his memory and the available information that is stored there. However, training in the use of force is primarily done in orderly terms with less influence of interfering factors that will be applicable to real situations. Paoline and Terril (2007) underlines that experience is the most important factor for learning within the police, and Kolb (1984) refers to experiential learning based on practical situations that provide experience and learning. As we have pointed out, there is a severe lack of structured approach to collection, dissemination, interpretation and integration of experience on the police use of force in Norway. There is no documentation for the extent of use of force, what kind of force is being used and the outcome in these situations. How can we then maintain that the training is experience-based? And how can we be sure that this training prepares police students and officers for street challenges? Our experience shows that training largely depends on the individual instructor's experience, and random work based on individual initiatives as for example the development of the simulator training at NPUC. This means that there will be varying quality in the professional experiences conveyed during training.

Norwegian official reports have several times pointed out a lack of systematic use of experiential

learning in the police (NOU 2009:12, NOU 2012:14, NOU 2013:9, NOU 2017:9). It is a fundamental principle in learning theory that training and education should be based on experience. You train as you fight. The Firearms Commission has suggested a systematic registration of police use of force and evaluation of incidents where the police have used firearms (NOU 2017:9). This would be an important contribution to the systematic use of experiences in the future, and support competence development within the police preparedness. The police response is not likely to be more efficient than the actual police preparedness when an incident occur (Norwegian Parliamentary White Paper 2012–2013). Thus, as also reflected in the report of the 22nd July commission (NOU 2012:14), the police emergency response personnel must have relevant competence and training in the use of force. The number of accidental discharges in the police (see Figure 4), lack of systematic reporting on the use of force, and the less degree of systematic competence and personnel development work in the police (NOU 2013:9, p. 209), indicate that experience is an underused asset in police training.

6 CONCLUSIONS

The aim of this paper has been to study the degree to which the NPUC and the police districts training in the use of force reflects the need for reliable handling of the various weapons at the disposal of the police officers facing street challenges, and how experiences on the use of force gained by police officers in their daily duty is collected, disseminated, interpretation and integrated in police training.

Even though experienced police officers lecture at the NPUC, and supervise police students during the three-year study program, it is no national overall strategy on collection, dissemination, interpretation and integration of experiences from the street in the Norwegian police force. It is fair to say that this is hampering the quality of the professional content of the training, especially at the NPUC, but also in the police districts. As long as there is no documented experience that can be conveyed in the form of experience-based knowledge, the quality of the professional content of the training and lecturing will depend on the individual instructor`s experience and competence, also related to the street challenges faced by police officers.

REFERENCES

Boin, A., et al. (2005). *The Politics of Crisis Management: Public Leadership under Pressure*. Cambridge, UK: Cambridge University Press.

Dewey, J. (1938). *Experience and education*. New York: Macmillian.

Dixon, N.M. (1999). *The Organizational Learning Cycle: How we can learn collectively.* Aldershot: Gower.

Dreyfus, H.L. and S.E. Dreyfus (1986). *Mind over machine. The Power of Human Intuition and Expertise in the Era of the Computer.* New York: Free Press.

Eid, J. and B.H. Johnsen (2006). *Operative psychology*. 2. ed. Bergen: Fagbokforlaget.

Engen, O.A., B.I. Kruke, P. Lindøe, K.H. Olsen, O.E. Olsen and K.A. Pettersen (2016). *Perspectives on societal safety and security* (own translation). Oslo: Cappelen Dam.

Endsley, M.R. (1988). *Design and evaluation for situational awareness enhancement.* Paper presented at The Human Factors Society 32nd Annual Meeting, Santa Monica, CA.

Gundel, S. (2005). Towards a New Typology of Crises. *Journal of Contingencies and Crisis Management 13(3)*: 106–115.

Helsloot, I. and A. Ruitenberg (2004). Citizen Response to Disasters: a Survey of Literature and Some Practical Implications. *Journal of Contingencies and Crisis Management, 12(3)*: 98–111.

Holmberg (2013). Police use of pepper spray—a Scandinavian comparison with focus on Denmark (own translation). *Nordisk Tidsskrift for Kriminalvidenskab.*

Klein, G.A. (1989). Recognition-primed decisions. In Rouse, W.B. (1989). *Advances in Man-Machine Systems Research.* Greenwich, CT, JAI: 47–92.

Klein, G.A. (2011). *Streetlights and Shadows: Searching for the Keys to Adaptive Decision Making.* Boston: MIT Press.

Kolb, D. (1984). *Experiential Learning: Experience as the Source of Learning and Development.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Lagestad (2008). Police use of physical force (own translation). *Nordisk tidsskrift for kriminalvidenskab, 95(3)*: 298–314.

Lee, J. D and A. Kirlik (2013). *The Oxford Handbook of Cognitive Engineering.* Oxford: Oxford University Press.

Lie, A.L. (2010). Police use of physical force (own translation). *PHS Forskning 2010:2*.

Lie, A.L. & P. Lagestad (2011). *Apprehension techniques* (own translation). Gyldendal Norsk Forlag AS.

Lovdata. no (2017). *Weapon instructions for the police.* Downloaded 5th December 2017: https://lovdata.no/dokument/INS/forskrift/2015-07-02-1088?q=våpeninstruks.

Norwegian Parliamentary White Paper No. 21 (2012–2013). *Terror preparedness*. Follow-up of NOU 2012: 14 Report from 22 July Comm. Min of Justice and Public Security.

Norwegian Parliamentary White Paper No. 10 (2016–2017). *Risk in a safe society. Societal safety* Ministry of Justice and Public Security.

Norwegian Parliamentary White Paper No. 42 (2004–2005). Police role and tasks (own translation). Oslo: Ministry of Justice and Public Security.

NOU 2009:12. Norwegian Official Report. A responsible police, openness, control and learning. Ministry of Justice and Public Security.

NOU 2012:14. Norwegian Official Report. Report from the 22nd of July Commission. Oslo Norway: DSS.
NOU 2013:9. One police—prepared to meet the challenges of the future (own translation). Oslo, Norway: DSS.
NOU 1981:35. The police role in society (own translation). Ministry of Justice and Public Security.
NOU 2017:9. Norwegian Official Report. Police and arming. Legality, necessity, proportionality and accountability. Oslo, Norway: DSS.
Paoline, E.A. & W. Terrill (2007). Police Education, Experience and the Use of Force. *Criminal Justice and Behavior*, Feb.
POD (2017a). Evaluation of the practice for arming based on the locally orderly book (own translation). Downloaded 5th December 2017: https://www.politiet.no/globalassets/05-aktuelt-tall-og-fakt/bevapning/bevapning.pdf.
POD (2017b). Police organization. Web page accessed 7th November 2017: https://politi.no/om/organisasjonen/.
POD (2011). Police Emergency Response System 1. Guidelines for police preparedness. Oslo: POD.
POD (2015). Police threat of the use of firearms or the use of firearms 2002–2014 (updated June 1 2015). Downloaded 5th December 2017: https://www.politiet.no/globalassets/05-aktuelt-tall-og-fakt/bevapning/bruk-av-skytevapen.pdf.
Rosenthal, U., M.T. Charles and P. t'Hart. (1989). *Coping With Crises: The Management of Disasters, Riots, and Terrorism*. Springfield, Illinois, U.S.A., Charles C. Thomas.
Schakel, J.-K., O.C. van Fenema & S. Faraj (2016). Shots fired! Switching between practices in police work. *Organization Science, 27*(2), 391–410.
't Hart, P., & A. Boin (2001). Between Crisis and Normalcy: The Long Shadow of Post-crisis Politics. In Rosenthal, Boin & Comfort (Eds.) (2001) *Managing Crises*. Springfield, Illinois: Charles C. Thomas. 28–46.
Vee Henriksen, S. (2016). *Police use of force. Forcible means, knowledge-based learning and development opportunities*. Police at new borders: Threats, force and protection. Oslo: NPUC.
Weick, K.E. (1993). The Collapse of Sensemaking in Organizations—the Mann Gulch Disaster. *Administrative Science Quarterly 38(4)*: 628–652.
Weick, K.E. (1995). *Sensemaking in Organazations*. London: Sage Publications.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

The role of employers, safety engineers and safety reps in the improvement of safety level at enterprises

G. Hrenov, K. Reinhold, M. Tint & P. Tint
Tallinn University of Technology, Estonia

ABSTRACT: Fifteen Estonian enterprises were investigated to determine the safety level, using Method for Industrial Safety and Health Activity Assessment (MISHA) method. Some of the firms have implemented OHSAS 18001 or belong to the foreign companies. One of the ideas to improve the safety level at the enterprise is learning through the interviews. The interview was based on MISHA method. The safety performance elements were divided into three parts: formal, real and combined ones. The statistics was carried out using factor analysis with Barlett's test, ANOVA and T-square test with Wilks' Lambda row. The main possibilities to influence on safety level in the firm to employ the Working Environment Specialists (WES), as they are more educated and supported by the law in the work safety area. In small- and medium-size enterprises, there are only few resources to hire the WES. If the OHSAS 18001 was implemented, then the assignment of tasks and responsibilities in Occupational Health and Safety (OHS) was committed to the top management ($p = 0.000$), the employer was revising the safety policy ($p = 0.000$), the personnel's responsibilities and authorities in OHS were clearly defined ($p = 0.013$). The role of the working environment representative was not particularly significant in the investigated enterprises ($p = 0.350$). At the same time, the firm's type was significant on the supervisor/employee communication ($p = 0.001$) and on general communication procedures ($p = 0.006$).

1 INTRODUCTION

The work environment is a multifunctional term and it occupies not only the physical work settings, but also the psychological and psychosocial features that are subject to the people personalities and approaches (Hrenov et al. 2017a). Treatment of risks at work is a key apprehension in today's working environment (Lafuente & Abad 2018). Researchers and general practitioners have perceived a remarkable modification in the role of safety supervision, which has developed from a slim view associated to an overpriced managerial problem do an functioning primacy with significant economic and social impression (Abad et al. 2013; Das et al. 2009). According to the European Agency for Safety and Health at Work (EU-OSHA 2013), the economic costs and functioning damages of work accidents to workers, productions, and public supervision signify 3% of the EU's gross domestic product.

The OHSAS 18001 certification is becoming the leading international safety system approved by the administrations to involve in processes to encourage constant developments of work safety conditions (Fernandez-Muniz et al. 2012; Lo et al. 2014). OHSAS 18001standard is the basis for the new ISO 45001 standard that would likely to be accessible in the nearby time.

There are many different safety management systems (Li & Guldemund 2018). A safety management system in industry (SMS) (Robson & Bigelow 2010) can be defined as a planned, documented safety program that incorporates certain basic management concepts and activating elements into a well-organized safety system. The safety activity areas and supporting elements that comprise this system act and interact on one another to help achieve the desired safety or risk level. A total safety management system (SMS) consists of the following objects: parametres such as input, process, output, and feedback control; attributes: properties of parameters such as the external manifestation of the way in which an object is known, observed, or introduced in a process; relationships: bonds that link objects and attributes in the system process.

SMS can be categorized as a set of institutionalized interconnected and interacting strategic elements designed to establish and attain occupational health and safety goals and objectives (Yurio et al. 2015; Kim et al. 2016).

There are different key persons in the enterprise who have to take care of occupational health and safety (OHS): the employer (EMP), the working environment specialist (in some countries safety engineer, WES) and working environment representatives (or reps, WER) (Hrenov et al. 2017a). The roles of these key-persons in dissimilar nation

states are different (Paas 2015a, b, c, d; Hrenov et al. 2016).

The previous studies for improvement of safety and health at workplace (Paas 2015a; Hrenov et al. 2016, 2017a, 2017b), the roles of the employer, working environment specialist and working environment representatives are given specifically.

The current study gives the comparison of the results of the interviews of these three parties (EMP, WES, and WER).

The aim of the study is to improve the safety level at small- and medium-sized enterprises in Estonia through the cooperation and close communication of the employer, working environment specialist and the working environment representative.

2 THEORETICAL PART

Safety is an intellectual conception. This state of freedom from something that could have undesirable consequences, such as damage to humans or nature, financial loss, or any other form of damage or defeat. For example, in a hospital, the safety of patients means holding patients in a steady state by avoiding the risk of adversarial occasions (Shojania et al., 2001). The current paper is concerned with industrial safety, the unexpected events and risks arise within the context of manufacturing activities. However, a zero-risk condition, does not occur. Although some companies achieve a zero accident record for a convinced period of time, it does not indicate they are risk-free. "Risk is a degree of the likelihood and consequence of undefined upcoming occasions; it is the casual of an undesirable endings" (Yoe, 2011), while safety is, according to IEC 61508, "freedom from unacceptable risk". We can consequently settle that the safety of an industry is judged by its acceptable risk.

Safety management (SM) means "a systematic control of worker performance, machine performance, and the physical environment" (Heinrich et al. 1980).

Over the latest past, the incidence of major accidents and crises have made it clear that organizations must still progress their competencies to discourse safety through the application of a systematic and proactive attitude. Safety management systems (SMS) are changing from a prescriptive style to a more "self-regulatory" and "performance oriented" model (Frick & Wren, 2000; Bluff 2003) that is more proactive, participative and better integrated with commerce activities.

Goetsch (1998) introduces the concept of total safety management (TSM) as a performance-oriented approach that gives organizations a supportable advantage in the market-place by creating a safe work environment that is conductive to ultimate performance and persistent improvement. In a total safety approach, business processes are combined with safety engineering methods within a nonstop upgrading culture that affects all levels in the organization. A work process is a complex network of interdependencies between physical items, information, communication and knowledge passages and decision-making activities (Zou & Sunindijo 2015; Kontogiannis et al. 2017).

2.1 *Participative risk assessment*

In furthermost cases, a risk assessment is performed on how jobs should be accomplished rather than on how they are actually performed in practice. As a result, critical alterations or destructions of events are missed in this analysis. To avoid this oversight, a participative risk assessment is required that would involve people at all organizational levels in certain stages of the analysis. This has the benefit of catching risks associated with how the work is done and also involves staff in SM. Finally, it would be easier to design safety measures and barriers that are compatible to the competencies and preferences of workers when they are part of that process, hence enabling a more efficient human-system interface (Kontogiannis et al. 2017).

2.2 *Processes for appreciative hazards and risks*

Risk assessment (RA) is significant because it helps create awareness of hazards and risks related to serious work activities. A participative method to safety analysis will allow workers and supervisors to update the risk analysis with everyday information about critical risks in the workplace and technical processes. Hence, a process of "safety preview and screening" would be required at the start of risk analysis. Some tests related to overwhelming weaknesses of RA (like residual risks) should be addressed through new changes in RA methods (Kontogiannis et al. 2017). Digital RA methods are very looked-for in the current area.

The research question in the current study is the following: which are the possibilities of the employer, working environment specialist (safety engineer) and working environment representative (rep) to improve the safety level at small and medium-sized enterprises (Figure 1).

In the previous study (Paas 2015a,b,c), the key-elements of safety management system have been divided into formal (like safety documents, content of the policy, correlation between the safety activities and the implementation or non-implementation of OHSAS 18001, revising the safety policy, written safety policy existence, assignment of tasks and responsibilities etc.), real (like top management commitment to the safety policy,

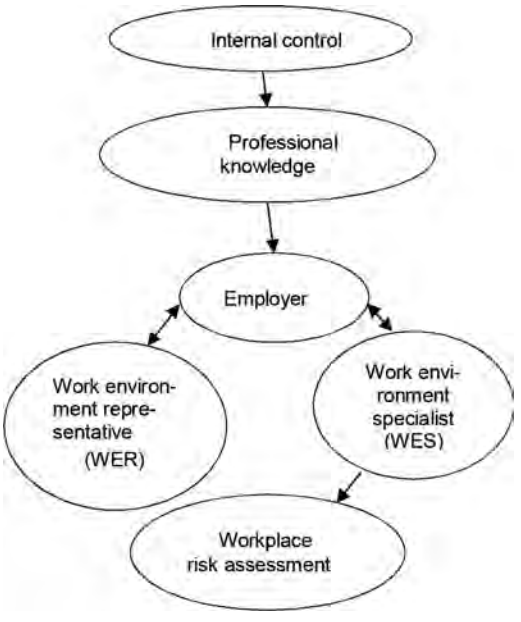


Figure 1. The arrangement of OHS command at workplace.

communication, participation in workplace design etc.) and combined (like participation in the preparation of the safety policy, workplace hazard analysis, the assessment of the work environment etc.) safety elements.

Three hypothesis were formulated and the area in which they are proved concerning employer's activities were as follows:

H1. Standard OHSAS 18001 has an impact on Formal safety performance in companies,
H2. Standard OHSAS 18001 has an impact on Real safety performance,
H3. Standard OHSAS 18001 has an impact on Combined safety performance at enterprises.

2.3 *Learning from experience*

Learning involves a monitoring process of near-misses, changes and success/failures of modifications as well as a reviewing process of strengths/weaknesses of risk analysis, unreliability risk assumptions and problematic risk acceptance criteria (Kontogiannis et al. 2017). This TSM mainstay also includes communication and training intended to provide safety information and skills to the employees to manage risks. Communication includes spread of information throughout the organizational hierarchy and feedback from the workforce. Safety training is very important. Safety training can take several forms such as classroom seminars, apprentice training, simulator training and computer-generated reality training. Although many small and medium-sized enterprises (SMEs) have understood the importance of the safety monitoring, they rarely know what data to collect, what safety parameters are important, how often to collect the data and what safety parameters are important, how to assure data reliability and confidentiality. For example, many workers may hold significant risk information but they do not know when and how to report it. Other cases contain the reporting of near misses in standard forms without the identification of causal factors (Kontogiannis et al. 2017).

One of the ideas to improve the safety level at the enterprise is learning through the interviews in the course of the questioning the employers, WES and WER (workers) using modified MISHA method (Paas 2015d). For example, the interview with the employer individually lasts 2–3 hours and is focused only of OHS matters. During this time, the person (employer or worker) learns as much as he will learn during the compulsory safety training course (24 hours) carried out in groups of people from different areas and interests and sometimes the lecturers on these safety training course (outside the house) are not giving not-interesting and old-fashioned information. This modified MISHA questionnaire is an educational tool for the WES: this is the mode of learning through interviews (Paas 2016d).

3 MATERIAL AND METHODS

Fifteen Estonian manufacturing enterprises (Table 1) were examined with modified MISHA method (Kuusisto 2000; Paas 2016d) to explain the role of EMP, WES and WER in OHS matters as well as for studying the perspectives to improve the safety level of the enterprise through their co-operation.

There are four areas in MISHA method: A) organization and administration; B) participation, communication, and training; C) work environment, D) follow-up (altogether 200 questions).

The MISHA questionnaire was modified taking into account some of the workplace hazards that were not included into the original MISHA questionnaire (Kuusisto 2000). For example, the influence of vibration and electromagnetic fields on the workers was asked in the course of the interview (Guldemund 2007).

The interviews with the learning aims consist of the questionnaire that includes "whether" and "how" questions. In the original questionnaires compiled for the assessment of safety, activities at

Table 1. The characterization investigated companies (N = 15).

Id. of the company	The activity area	Size, employees	OHSAS company/ corporated company
I	Plastic industry	50–249	+/
II	Electronics	>250	/+
III	Food industry	>250	/+
IV	Electronics	>250	+/
V	Textile industry	50–249	–/–
VI	Printing industry	<50	–/–
VII	Glass industry	<50	+/
VIII	Chemical industry	50–249	+/
IX	Chemical industry	50–249	–/–
X	Metal industry	50–249	–/+
XI	Metal industry	>250	–/+
XII	Agriculture farm (milk production)	<50	–/–
XIII	Agriculture farm (grain production)	<50	–/–
XIV	Construction	<50	–/–
XV	Transport	<50	–/–

Table 2. The characterization and results of quantitative study by the MISHA method in the investigated enterprises (N = 15).

Id.of the company	OHSAS company/ corporated company	Person interviewed; position, age	Total average score (100 max)
1	2	3	4
I	+/	Quality manager, 41	78 ± 3.0*
		Safety manager, 62	76 ± 2.5
		WER, 25	78 ± 2.5
II	/+	Quality manager, 35	84 ± 2.0
		WES, 42	90 ± 1.0
		WER, 53	80 ± 1.0
III	/+	WES, 62	75 ± 2.0
		WER 1, 34	80 ± 2.5
		WER 2, 39	58 ± 3.0
IV	+/	Quality manager, 59	92 ± 1.0
		WES, 39	88 ± 2.0
		WER, 39	78 ± 1.0
V	–/–	Quality manager, 38	47 ± 3.5
VI	–/–	Quality manager, 36	29 ± 4.0
VII	–/–	Quality manager, 41	41 ± 4.0
VIII	+/–	Manager, 55	88 ± 1.0
		WER, 62	85 ± 1.0
		External auditor, 34	78 ± 2.0
IX	+/–	Manager, 45	87 ± 1.0
		WER, 40	87 ± 1.0
		External auditor, 34	78 ± 1.0
X	–/–	Manager, 40	61 ± 1.5
		WER, 53	55 ± 2.0
		External auditor, 53	50 ± 2.0
XI	–/+	WES, 35	89 ± 1.0
		Trade union rep, 60	86 ± 1.0
XII	–/–	Employer, 50	46 ± 2.0
XIII	–/–	Employer, 56	60 ± 1.0
XIV	–/–	Active manager, 40	50 ± 2.0
XV	–/–	Personnel manager, 45	65 ± 2.0

*Mean difference in reviewers (4) assessment score.

enterprises can be used as a tool for learning and obtaining more information on safety in companies. Learning is likely to be more effective when participants are actively involved in dialogue in which they are co-constructors of the meaning (Kines et al. 2011).

In locally owned companies, where the safety level is rather low, the managers did not recommend to have interviews with WER as by their (employers') view, the knowledge of WER in OHS tends to be negligible.

In those companies where OHSAS 18001 was implemented, the work instructions, instructions for safety training and safety organization's activity programs existed. This may not be the case in companies where OHSAS 18001 was not implemented.

The statistics used in the paper involved IBM SPSS Statistics 22.0 and R.2.15.2. The following statistical methods were used: correlation, MANOVA, factor analysis, principal component method, independent T-test (Field, 2013).

4 RESULTS

The interviews with the enterprises' representatives (column 3, Table 2) were carried out and recorded; afterwards listened and analysed by the three authors of the paper and one expert independently. The total average score (column 4) is derived with MISHA method.

4.1 *The employers role in safety level formation*

From the MISHA questionnaire, the questions concerning the activities of the employer were selected. The comparison of the interviewer assessments' on OHSAS-implemented companies compared with non-implemented companies was assessed with the statistics (Tabel 3).

In these safety areas (Table 3), the implementation of OHSAS 18001 has been successful.

Table 3. SM areas where the employer's influence on the safety level is significant.

Safety key element	Sum of Squares (KMO and Barlett's test	p-value
Formal safety elements		
A1.4. Assignment of tasks and responsibilities to the top management	13.375	0.000
A1.8. Revising the safety policy: has the employer defined how often the policy is revised?	25.688	0.000
C2.3- Does the personnel's responsibilities and authorities are clearly defined?	4.576	0.013
Real safety elements		
A1.9. Dissemination of the policy: has the employer defined how the policy is made available to the personnel?	21.007	0.000
A2.8. Resources: does the company has the resources for OHS improvement?	22.688	0.000
B2.1. Does the manager arrange the information meetings to the employers on OHS?	2.896	0.006
D2.1. Does the company has the system for redesigning the work or workplaces of a person with disabilities?	0.047	0.013
Combined safety elements		
A1.6. Dissemination of the policy: has the employer defined how the policy is made available to the personnel?	13.375	0.001
A1.10. Informing external bodies about the company's safety policy	17.241	0.001
B3.1. Does the employer affords the safety training for the personnel in a regular basis?	2.854	0.004
D1.2. The reduction of accidents: has the plan elaborated and presented to the top manager?	4.125	0.007
D3.1. Does the company have a system for measuring social climate?	19.125	0.000

4.2 *The working environment specialist role in safety level formation*

Similarly to the chapter 4.1, the questions concerning the WES activities, where selected (Table 4).

Table 4. SM areas where WES has the strongest influence on the company's safety level.

Safety key element	Sum of Squares (KMO and Barlett's test	p-value
Formal safety elements		
A1.1. Does the company has the written policy?	22.250	0.000
A1.3, Comments of the policy: a description of the safety tasks?	19.285	0.000
A1.4. Are the tasks assigned to the safety and health personnel?	13.375	0.000
A1.8. Revising the safety policy, who are responsible?	25.688	0.000
C2.3. Definition of the personnel's responsibilities: are the persons responsible for health and safety trained for their responsibilities?	4.576	0.0013
D1.1. Does the company make statistics on accident rates and summaries on accident causes?	21.000	0.000
Real safety elements		
A1.9. DEssemination of the policy: is safety involved?	21.007	0.000
A2.8. Does the company seek advice in resources to health and safety from safety personnel?	22.688	0.000
B1. Does the safety manager instruct the personnel?	5.672	0.001
B2.1. Has the safety manager arranged the hazards management system in the workplace?	2.896	0.006
B2.4. Does the safety manager arrange the safety campaigns?	9.797	0.001
B3.4. Has the safety manager defined which work permits are necessary (e.g., permit do fire hazardous work?)	6.750	0.004
C1.8. Has safety manager involved in the cleaning of plant area?	4.500	0.002
Combined safety element		
A1.5. Participation in preparation of the safety policy	21.500	0.000
A1.6. Initial status review	13.375	0..001
B3.1. Safety training needs, are they determined to the personnel?	2.854	0.004
C3.1 Workplace hazard analysis: has safety manager carried out?	9.491	0.000
D1.2. Accident investigation: are the near accidents investigated?	4.125	0.007

The main possibilities to influence on the safety level in the company is to employ the working environment specialists, as they are more educated and supported by the law in the work safety and health (OHS) area. Nevertheless, in OHSAS 18001 implemented companies, the results are better than in non-implemented ones.

4.3 *The working environment representatives' role in safety level formation (all combined area)*

The questions were selected from the interviews, where WER activities could possibly improve the safety level at enterprises (Table 5). All these questions happened to be only in the combined safety elements' area.

4.4 *Results of the hypothesis*

Three hypothesis were formulated and the area in which they are proved concerning employer's activities were as follows:

H1. Standard *OHSAS 18001* has an impact on *Formal* safety performance in companies (*p value* < 0.013) – if OHSAS 18001 has been implemented, then the assignment of tasks and responsibilities in OHS is committed to the top management, the employer is revising the safety policy, and the personnel's responsibilities in OHS are clearly defined.

H2. Standard *OHSAS 18001* has an impact on *Rea*l safety performance. ($p < 0.013$) – if OHSAS 18001 is implemented, then the top manager promotes dissemination of safety policy: the policy is made available to all of the personnel; resources for improvement are arranged by the top management; the top manager arranges meetings in OHS; there is a system for redesigning the workplaces for the persons who have difficulties in coping with the work.

H3. Standard *OHSAS 18001* has an impact on *Combined* safety performance ($p < 0.007$) – if OHSAS 18001 implemented, then: the top management is participating in the preparation of safety policy, top manager is reviewing the safety policy and is informing the external bodies about the company's safety policy's effectiveness; the top manager arranges safety training for all of the personnel; there is a plan for reduction of accidents; it has been elaborated by the top manager; the company has a system for measuring the social climate in the company.

Table 5. SM areas where WER can influence on the safety level at enterprises.

Combined safety elements	Sum of Squares (KMO and Barlett's test	p-value
A1.5. Participation in the preparation of the policy	21.250	0.000
A1.6. Initial status review	13.375	0.001
A1.10. Informing external bodies about the policy	17.241	0.001
A3.3. Selection of the line management	3.063	0.017
B3.1. Safety training needs	8.491	0.000
D1.2. Acident investigation	4.125	0.007
D3.1. Assessment of the social environment	19.125	0.000

5 DISCUSSION

Our study revealed that management plays an essential role in OHS improvement in the company. By O'Toole (2002), it is also postulated that the leadership's position is influencing the employee's perceptions of the safety management systems. Those perceptions appear to influence on the employee's decisions that relate to at-risk behaviours and decisions on the job. Organizational commitment did affect the perceived safety at work, but not on work accidents.

In the current study, it was declared that the plan for reduction of accidents if it is worked out by the employer, has very strong influence on the combined safety at enterprises. If the safety standards (OHSAS 18001 etc.) are implemented then the organizational climate will also be in higher level (Neal et al. 2000).

Taking into account the results of the previous studies (Hrenov et al. 2016; Arghami et al. 2014), where the safety and health level on the enterprise measured with MISHA method was carried out from the viewpoint of the working environment representative (WER) (Hrenov et al. 2016) and the employers (Arghami et al 2014), it can be concluded that the safety engineers have the best overview and knowledge of the safety system.

The other key persons (WER, employer) are hesitant in some questions, concerning for example, the safety policy expanding to the workers in the firm. The working environment specialist often views the current situation realistically while WER and employer may overestimate the situation as their daily work is not connected with safety matters.

MISHA method is not the only method for assessment and showing the improvement's points in safety and health at enterprises (Kines et al 2011;

Guldemund 2007). By Arghami et al. (2014), the safety climate questionnaire is built up on another principle than in MISHA method. It contains seven (7) different factors: management commitment to safety and personnel collaboration (the influence of total safety level R = 0.954), safety communication (R = 0.830), supportive environment (R = 0.793), work environment (R = 0.803), formal training (R = 0.774), priority of safety (R = 0.740), personal priorities and need for safety (R = 0.547). These results are comparable with the results in the current paper: the safety policy might be worked out properly and on the high level, but the safety policy usually does not reach the personnel, from up to down. One of the lowest scores (R = 0.431) is given to the question: "my line manager/supervisor does not always inform me of current concern and issues".

Our study examined three different types of companies: OHSAS certified companies, corporated companies and small and medium-sized locally owned companies. It turned out that "small enterprises" may be diverse: the definition covers many types of work activities, which naturally lead to large differences in the work environment. Small enterprises are more susceptible to influence from various "external" sources e.g., though the ownership structure. It might be important whether the small enterprise is part of a larger organization and whether it is publicly or privately owned (Sorensen et al., 2007). This problem remains for the future research.

Compared to Estonian OHS system in companies, Nordic OHS regime contains three different collaborating structures within the company: 1) a work environment or safety committee with balanced representation from the parties; 2) safety representatives elected by the employees; 3) in-house or external health and safety experts employed by and representing the management (Lindoe et al., 2001). According to the OHS Act (1999), based on EU Framework Directive 89/91, the employer and employees have to co-operate and opportunities for both parties to consult on the relevant OHS matters should be available. The need to ensure worker participation is stated in mandatory forms of industrial health and safety national legislation and in the EU Framework Directive 89/391. In Estonia, WER has to be trained following the 24-h training programme provided in the regulation. In Norway, the social partners agree that a 40-h course covers the basic training necessary to function as a WER (Hovden et al., 2008).

From our interviews, we concluded that WERs assessed the time for dealing with OHS matters unsatisfactory. The results in Nordic countries (Hovden et al., 2008) show similar pattern – often WERs complained about lack of time. The examples of the best experiences of the Nordic countries should be used in order to increase workers' participation and representation in health and safety matters.

6 CONCLUSIONS

1. OHSAS 18001 implementation helps to improve the following *formal* safety elements where safety manager is involved: to write the safety policy, the description of tasks of the personnel in safety area, the responsibilities of the safety personnel are clearly determined.
2. OHSAS 18001 implementation helps to improve the following *real* safety elements: dissemination of the safety policy, the safety personnel is advising the top management in safety and health questions, the safety manager instructs thoroughly the personnel in safety matters, the safety personnel is advising the top management how to allocate the resources.
3. OHSAS 18001 implementation in the firm helps to improve the following *combined* safety elements: safety manager compiles the initial safety review, the safety training needs of the personnel are determined, workplace hazard analysis are carried out.
4. The position of safety representative has often a low status in the company. WERs do not have enough time to fulfil their safety functions to keep employees safe.
   There is a limited understanding among employers about the role of WER. The study showed that in small enterprises, the WER has a formal position, although required by the law. In that case, employers do not understand the need of the WER and while electing them only formally, there is no practical value and often, employees are unaware of the position. The interviews also revealed that it is complicated to find the candidates to the WER position even in larger companies, especially in locally owned companies as managers do not know how to motivate workers on taking an additional responsibility. Safety management system plays a role in effective work of WERs. If the management does not give enough priorities to OHS, the employees will follow the example of the employer. WER should be elected among the peers rather than using WERs from other departments.
5. The WER of the organization is not well known or acknowledged by all the employers and subcontractors. The subcontracting work may cause several accident and near-accident situations. The importance of the person (WER), who knows how to deal with the problems in OHS, becomes evident only after the accident

has occurred or some of the workers are already seriously ill with occupational disease, such as musculoskeletal disease (MSD). The MSD is, at the present time, the number one occupational illness in almost every European country (Kaergaard & Andersen, 2000).

6. To be successful in WER commitments may be complicated due to conflicting expectations from employer and colleagues. The interviews revealed that nobody in the enterprise wants to be the resolver of a risky situation or even accident. Therefore, it is particularly important to prevent these situations by increasing the knowledge on OHS. For this occasion, WER and his/her knowledge and activities are a very good solution. It is important to mention that he/she needs enough time to gather the information on OHS and his/her activity has to be acknowledged by the employer.

REFERENCES

Abad, J., Dalmau, I. & Vilajosana, J. 2014. Taxonomic proposal for integration levels of management systems based on empirical evidence and derived corporate benefits. *J Clean. Prod.* 78:164–173.

Arghami S. et al.. Reliability and validity of a safety climate questionnaire. 2014. *Journal of Research in Health Sciences*, 14:140–145.

Bluff, L. 2003. Systematic management of occupational health and safety. Working paper 20. National Research centre for Occupational health and safety regulation. Regulatory Institutions Network, Research School of Social Sciences, Australian national University.

Das, A. et al. 2008. Towards a theory of the linkages between safety and quality. *J. Oper. Manage.* 29:753–765.

European Agency for Safety and Health at Work. 2013. EU-OSHA Priorities for occupational safety and health research in Europe: 2013–2020. Luxembourg.

Fernandez-Muniz, B., Montes-Peon, J.M. & Vazquez-Ordas, C.J. 2012. Safety climate in OHSAS 18001-certified organizations: antecedents and consequences of safety behavior. *Accid. Anal. Prev*. 45:745–758.

Field, A. 2013. Discovering Statistics Using IBM Statistics. Fourth Edition. SAGE Publications Ltd., http://www.uk.sagepub.com/field4e/main.htm.

Frick, K. & Wren, J. 2000. Reviewing occupational safety and health management: multiple roots, diverse perspectives and ambiguous outcomes. In: Frick, K., Jensen, P., Quinlan, M. & Wilthagen, T. (Eds.). Systematic Occupational safety and health management: perspectives on an International development. Emerald Group Publishing Limited, Bingley.

Goetsch, D.L. 1998. Implementing total safety management: safety, health and competitiveness in the Global market. Prentice Hall, KJ.

Guldemund, F.W. 2007. The use of questionnaires in safety culture research- and evaluation. *Safety Science* 45:723–743.

Heinrich, H.W., Peterson, D. & Roos, N. 1980. *Industrial Accident Prevention: A Safety Management Approach*, fifth ed. McGraw-Hill, New York.

Hovden, J. et al. 2008. The safety representative under pressure. A study of occupational health and safety management in the Norwegian oil and gas industry. *Safety Science* 46, 493–509.

Hrenov, G., Paas, Õ., Tint, P.& Reinhold, K. 2016. Workers' representatives in OHS activities: example of Estonian industrial sector. *Agronomy Research* 14:377–391.

Hrenov, G., Reinhold, K. & Tint, P. 2017a. *Employers' role in the improvement of safety level in Estonian enterprises*. In Environment. Technology. *Resources. Proc. of the 11th Intern. Scientific and Practical Conference. 15–17 June, Rezekne Academy of Technologies*, 1:115–120. Rezekne, Latvia.

Hrenov, G. et al. 2017b. Working environment specialist's role in the improvement of safety level in Estonian enterprises. In: Proceedings of the 16th Int. Scien. Conf. Engineering for Rural Development. Jelgava, 24–26 May.

IEC 61508. 2015. *Functional Safety*. International Electrotechnical Commission. Geneva.

Kaergaard, N & Andersen, J.H. 2000. Musculoskeletal disorders of the neck and shoulders in female sewing machine operators: prevalence, incidence, and prognosis. *Occupational and Environmental Medicine* 57:528–534.

Kines P. et al. 2011. Nordic safety climate questionnaire (NOSACQ-50): a new tool for diagnosing occupational safety climate. *International Journal of Industrial Ergonomics*, 21:634–646.

Kim, Y., Park, J. & Park, M. 2016. Creating a culture of prevention in occupational safety and health practice. *Safety and Health* at Work 7:89–96.

Kontogiannis, T. et al.. 2017. Total safety management: principles, processes and methods. *Safety Science* 100:128–142.

Kuusisto, A. 2000. Safety management systems: audit tools and reliability of auditing [dissertation]. Tampere: Tampere University of Technology.

Kysor, H.D. 1973. Safety management system. Part 1: the design of a system. *Nat. Safety News* 108:98–102.

Lafuente, E. & Abad, J. 2018. Analysis of the relationship between the adoption of the OHSAS 18001 and business performance in different organizational contexts. *Safety Science* 103:12–22.

Li, Y. & Guldenmund, F.W. 2018. Safety management systems: a broad overview of the literature. *Safety Science* 103:94–123.

Lindoe, P.H. et al.. 2001. *A Nordic way of handling occupational health and safety regulations*. Nordic Council of Ministers, Copenhagen.

Lo, C.K.Y., Pagell, M., Fan, D., Wiengarten, F. & Yeung, A.C.L. 2014. OHSAS 18001 certification and operating performance: the role of complexity and coupling. *J. Oper. Manage.* 32:268–280.

Neal, A. et al. The impact of organizational climate on safety climate and individual behaviour. *Safety Science* 34, 99–109.

O'Toole, M. 2002. The relationship between employees' perceptions of safety and organizational culture. *Journal of Safety Research* 33:231–243.

Paas, Õ. 2015a. *Development of the safety management system at enterprises*. PhD thesis. Tallinn University of Technology, 164 pp.

Paas, Õ., Reinhold, K. & Tint, P. 2015b. Estimation of safety performance by MISHA method and the benefits of OHSAS 18001 implementation in Estonian manufacturing industry. *Agronomy Research* 13:792–809.

Paas, Õ., Reinhold, K. & Tint, P. 2015c. OHSAS 18001 contribution to real and formal safety elements of safety management in manufacturing companies: results of statistical analysis. *Agronomy Research* 13:1260–1274.

Paas, Õ., Reinhold, K. & Tint, P. 2015d. Learning through questioning in occupational health and safety. *Scientific Journals of Poznan University of Technology series of "Organization and Management"* 67:69–85.

Robson, L.S. & Bigelow, P.L. 2010. Measurement properties of occupational health and safety management audits: a systematic literature search and traditional literature synthesis. *Can/Public Health*, 101 (Suppl. 1): 534–540.

Shojania, K.G. et al 2001. *Making health care safer: A critical analysis of patient safety practices*. Agency for Healthcare Research and Quality Rockville, Maryland.

Sorensen, O.H., Hasle, P. & Bach, E. 2007. Working in small enterprises – Is there a special risk? *Safety Science* 45: 1044–1059.

Yoe, C. 2011. *Primer and risk analysis: decision making under uncertainty*. CRC Press, New York (NY).

Yurio, P.L., Willmer, D.R. & Moore, S.M. The previous studies for improvement the safety and health at workplaces 2015. Health and safety management systems through a multilevel and strategic management perspective: theoretical and empirical considerations. *Safety Science* 72:221–238.

Zou, P.X.W. & Sunindijo, R.Y. 2015. Strategic Safety management in Construction and Engineering. Wiley & Sons, NY.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Standardized risk assessment techniques: A review in the framework of occupational safety

F. Brocal
Department of Physics, Systems Engineering and Signal Theory, University of Alicante, Alicante, Spain

C. González & M.A. Sebastián
Department of Manufacturing Engineering, National Distance Education University (UNED), Madrid, Spain

G.L.L. Reniers
Faculty of Technology, Policy and Management, Safety and Security Science Group (S3G), TU Delft, The Netherlands
Faculty of Applied Economics, Antwerp Research Group on Safety and Security (ARGoSS), University Antwerp, Antwerp, Belgium

N. Paltrinieri
Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology—NTNU, Trondheim, Norway

ABSTRACT: The publication in 2018 of ISO 45001 is the first international ISO standard in the field of occupational health and safety management systems. ISO 45001 is the result of an international consensus on the subject and describes the best international preventive practices and incorporates the requirements of a management system aligned with the so-called High-Level Structure of the ISO standards of management systems. Simultaneously, ISO/IEC 31010:2009 provides for the selection and application of systematic techniques for risk assessment. However, this standard does not address safety in a specific way. It is a generic standard for risk management and any reference to safety is simply informative. Thus, the main objective of the present work is to identify and classify the main techniques included in the ISO/IEC 31010:2009 standard applicable in the field of occupational safety and in line with the requirements for the ISO Standard 45001. As a secondary objective, it is sought in the same context to identify the main non-standard techniques of new or emerging nature. This process of identification and classification has been carried out by means of a systematic review of the scientific literature specialized in the matter. The results have been classified according to bibliometric indicators in the following groups: (a) Main standard techniques for application to occupational safety; (b) Main techniques developed through specific standards (such as: IEC 61882:2016 – Hazard and operability studies [HAZOP studies] – Application guide; IEC 61025:2006 – Fault Tree Analysis [FTA]); (c) Main non-standard techniques of a new or emerging nature for application to occupational safety. Finally, the results obtained by the classification mentioned above have been analyzed in order to determine the degree of coverage and standardization of the main risk assessment techniques applied to occupational safety.

1 INTRODUCTION

The European Agency for Safety and Health at Work together with the International Labour Organization has estimated of the cost of poor occupational safety and health. Such estimation reveals (EU-OSHA, 2017): Worldwide work related injury and illness result in the loss of 3.9% of GDP, at an annual cost of roughly € 2680 billion; Work-related illnesses account for 86% of all deaths related to work worldwide, and 98% of those in the EU; 123.3 million DALY (disability-adjusted life years) are lost globally (7.1 million in the EU) as a result of work-related injury and illness. Of these, 67.8 million (3.4 million in the EU) are accounted for by fatalities and 55.5 million (3.7 million in the EU) by disability; and in most European countries, work-related cancer accounts for the majority of costs (€ 119.5 billion or 0.81% of the EU's GDP), with musculoskeletal disorders being the second largest contributor.

In order to combat the problem, ISO has developed a new standard, ISO 45001, Occupational health and safety management systems—Requirements, that will help organizations reduce this

burden by providing a framework to improve employee safety, reduce workplace risks and create better, safer working conditions, all over the world. This standard follows other generic management system approaches such as ISO 14001 and ISO 9001 (ISO, 2017).

The current version ISO/DIS 45001.2:2017, contemplates in its section 6.1.2.2 that the organization must establish, implement and maintain one or several processes to assess the risks for safety and health at work from the identified hazards, taking into account legal requirements and other requirements and the effectiveness of existing controls (the concept of organization, include among others, a company, firm, enterprise, etc.). Nevertheless, this standard does not specify the techniques or methodologies to assess the occupational safety and health risks.

Simultaneously, the standard ISO/IEC 31010:2009 provides for the selection and application of systematic techniques for risk assessment.

However, this standard does not address safety in a specific way. It is a generic standard for risk management and any reference to safety is simply informative.

Thus, the main objective of the present work is to identify and classify the main techniques included in the ISO/IEC 31010:2009 standard applicable in the field of occupational safety in a way in line with the requirements for the risk assessment included in the ISO Standard 45001 (Currently, ISO/IEC 31010 is under review. For the moment, ISO/IEC/DIS 31010: 2017 is published under development)

As a secondary objective, it is sought in the same context to identify the main non-standard techniques of new or emerging nature.

2 METHOD

We conducted a systematic search in the occupational safety literature. A systematic review of the literature is typically based on a detailed and comprehensive plan and search strategy derived a priori in order to reduce bias (Uman, 2011). We aim to present an overview of techniques addressed in both quantitative and qualitative research on occupational safety, and their general direction (e.g. Cornelissen et al., 2017). Below, we will elaborate on our systematic selection process and analysis.

2.1 *Literature search*

As indicated by Goerlandt et al. (2017), in traditional indexing systems such as Scopus and Web of Science, risk analysis is not considered as a separate category in the scientific research areas. Instead, contributions related to risk are typically listed under "mathematics", "social sciences" or "engineering". Hence, general searches in those systems on terms such as "risk analysis", "validation" and "QRA" results in many hits, with low relevance to the above stated aims.

Therefore, another review method has been applied (November, 2017). To do this, we chose a literature search using broad search terms as a starting point. To this end, we selected as search criteria the use of the keywords "risk" and "review" in the journal title. In addition, we delimited the search in "engineering" and "chemical engineering" fields using ScienceDirect databases. The results were the following: Engineering field: 90 records published all years (1995–2018); Chemical engineering field: 48 records published all years (1988–2017).

With these criteria, the journal principals identified in the work of Reniers and Anthone (2012) have been included, except Journal Risk Analysis. These authors found that the most well-respected journal by expert opinion was the Journal of Loss Prevention in the Process Industries. However, taking into consideration both the respondents' results and the citation-based results into consideration, the Journal of Hazardous Materials is the most influential journal, followed by Reliability Engineering and System Safety, Risk Analysis, Accident Analysis and Prevention and Safety Science.

2.2 *Article selection*

The further selection of articles was performed in steps, as depicted in Figure 1.

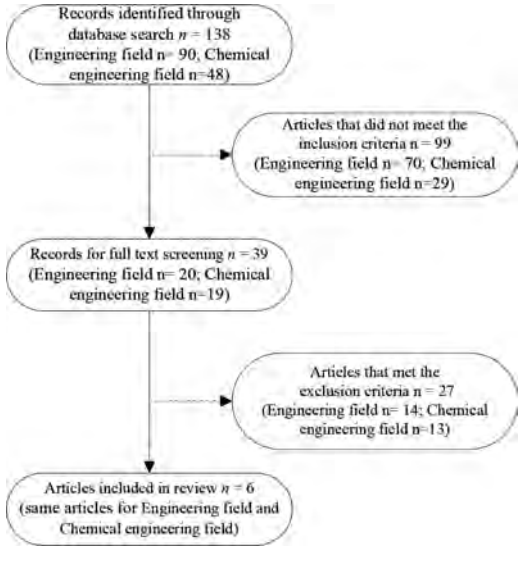


Figure 1. Flow diagram of study selection.

First, the titles were evaluated and articles that did not meet the following inclusion criteria were marked: (a) describe safety in an occupational setting; (b) focus on the application, study or analysis of techniques of identification, analysis or risk assessment applied to occupational safety; (c) conducted in the construction, manufacturing, offshore, or petrochemical sector; (d) published in a peer-reviewed journal; and (e) be written in English.

Second, the full text of the articles obtained after the application of the previous step, were analyzed. In this regard, the following exclusion criteria were applied: (a) only one specific technique is applied; (b) collect several techniques only by way of example; (c) conducted on driving or transportation; (c) conducted on risks in natural disasters.

Third and finally, the articles obtained were included in the review process. In this regard, both the Engineering field and the Chemical engineering field obtained the same 6 results (Table 1).

2.3 *Analysis*

The analysis of the articles was developed as follows.

a. Articles that contain a classification or set of techniques related to standard ISO/IEC 31010: 2009. First, each article has been analyzed with the aim of identifying those techniques that coincide with the list of techniques included in Annex A of ISO/IEC 31010: 2009. Second, those techniques that are developed through specific standards have been identified. To do this, the reference documents of the standard have been analyzed (ISO/IEC 31010: 2009 and ISO/IEC/DIS 31010:2017). In addition, each technique has been checked if it has been developed through any standard published by any of the following standardization bodies: International Organization for Standardization (ISO), European Committee for Standardization (CEN) and International Electrotechnical Commission (IEC).
b. Articles containing other techniques not collected to ISO/IEC 31010. The main techniques have been identified.

Table 1. Articles included in review (n = 6).

Nr	Paper citation	Field of application	Journal
1	Kjellén and Sklet (1995)	Offshore industry	SS
2	Tixer et al. (2002)	Risk analysis on an industrial plant	JLPPI
3	Marhavilas et al. (2011)	Work sites	JLPPI
4	Dallat et al. (2017)	Safety management, (risks that may lead to accidents)	SS
5	Villa et al. (2016)	Chemical and process Industries	SS
6	Yang et al. (2017a)	Petroleum activities	SS

SS: Safety Science; JLPPI: Journal of Loss Prevention in the Process Industries.

3 RESULTS

The results presented in this section follow the scheme accordingly for the analysis of the articles included in the review.

3.1 *Main standard techniques for application to occupational safety*

Table 2 lists the existing relations between the techniques identified in the reviewed articles (1–4) and the techniques included in Annex A of ISO/IEC 31010: 2009.

The following results can be observed in the Table 2: 18 techniques are collected in a single reference (Nr Paper citation); 7 techniques are collected in two references; 3 techniques are collected three references; and 3 techniques in four references.

3.2 *Main techniques developed through specific standards*

Those techniques that are developed through specific standards have been identified in the Table 3.

Thus, 10 techniques of the 31 collected by ISO/IEC 31010: 2009 are identified. Comparing these 10 techniques with the results of Table 2, it can be observed: 4 techniques are collected in a single reference (techniques Nr 11, 12, 24 and 25); 1 technique is collected in two references (13); 3 techniques are collected in three references (6, 20 and 22); and 2 techniques (14 and 15), in four references.

Of the 10 techniques listed in Table 3, all of them except Reliability centered maintenance (RCM), have also their correspondence with European Standards (ENs) published by the European Committee for Standardization (CEN, 2017).

3.3 *Main non-standard techniques of a new or emerging nature for application to occupational safety*

Tables 4 and 5 list the existing relations between the techniques identified in the reviewed articles (5–6).

Table 2. Relations between techniques identified in reviewed articles and Annex A of ISO/IEC 31010: 2009.

Nr	Annex A ISO/IEC 31010/2009	Nr Paper citation 1	2	3	4
1	Brainstorming	N	N	N	Y
2	Interviews	N	N	N	Y
3	Delphi technique	N	Y	N	Y
4	Check lists	N	Y	Y	Y
5	Preliminary Hazard Analysis (PHA)	N	Y	N	Y
6	Hazard and operability studies HAZOP	N	Y	Y	Y
7	Hazard analysis and critical control points HACCP	N	N	N	Y
8	Environmental Risk Assessment (ERA)	N	N	N	Y
9	Structured what if technique SWIFT	N	Y	Y	Y
10	Scenario analysis	N	N	N	Y
11	Business Impact Analysis (BIA)	N	N	N	Y
12	Root Cause Analysis (RCA)	N	N	N	Y
13	Failure Modes and Effect and Analysis (FMEA)	N	Y	N	Y
14	Fault Tree Analysis (FTA)	Y	Y	Y	Y
15	Event Tree Analysis (ETA)	Y	Y	Y	Y
16	Cause consequence analysis	N	N	N	Y
17	Cause and effect analysis	N	N	N	Y
18	Layers of Protection Analysis (LOPA)	N	N	N	Y
19	Decision tree	N	N	N	Y
20	Human Reliability Analysis (HRA)	N	Y	Y	Y
21	Bow tie analysis	N	N	N	Y
22	Reliability Centered Maintenance (RCM)	N	Y	Y	Y
23	Short cut risk analysis (SCRAM)	N	Y	N	Y
24	Markov analysis	N	N	N	Y
25	Monte Carlo analysis	N	N	N	Y
26	Bayes analysis and Bayesian networks	N	N	N	Y
27	F/N diagrams	N	N	Y	Y
28	Risk índices	N	N	N	Y
29	Consequence likehood matrix	N	N	Y	Y
30	Cost-benefit analysis	N	N	N	Y
31	Multi criteria analysis (MCDA)	N	N	N	Y

Y (YES): The technique (Nr) is collected by the Paper (Nr)
(NOT): The technique (Nr) is not collected by Paper (Nr).

Theoretical and practical limitations affecting results of hazard identification suggest the need for an improvement of current techniques (Paltrinieri et al., 2016). Yang et al. (2017a) indicated that several aspects of operational decision making creates a need for different risk analyses compared to the traditional Quantitative Risk Analysis (QRA) with principles and guidelines described in standard ISO 31000:2009, which are developed more for design purposes.

Yang et al. (2017a) reviewed 11 risk influence-frameworks that integrate organizational and human factors in a structured way (Table 4). The intention was to evaluate how these frameworks

Table 3. Techniques included in Annex A of ISO/IEC 31010: 2009 that are developed through specific standards.

Nr	Annex A ISO/IEC 31010/2009	Specific standards
6	Hazard and operability studies HAZOP	IEC 61882:2016
11	Business Impact Analysis (BIA)	ISO TS 22317:2015; ISO 22301:2012
12	Root Cause Analysis (RCA)	IEC 62740:2015
13	Failure modes and effect and analysis (FMEA)	IEC 60812:2006
14	Fault Tree Analysis (FTA)	IEC 61025:2006
15	Event Tree Analysis (ETA)	IEC 62502:2010
20	Human Reliability Analysis (HRA)	IEC 62508:2010
22	Reliability Centered Maintenance (RCM)	IEC 60300-3-11: 2003
24	Markov analysis	IEC 61078:2016; IEC 61165:2006; ISO/IEC 15909-1:2004
25	Monte Carlo analysis	IEC 62551:2012; ISO/IEC Guide 98-3-SP1:2008

Table 4. Risk influence frameworks (techniques) that integrate organizational and human factors in a structured way (adapted from Yang et al., 2017a).

Technique	Paper citation	Field of application
Model of Accident Causation using Hierarchical Influence Network (MACHINE)	Embrey (1992)	General
WPAM		Nuclear
System Action Management (SAM)	Paté-Cornell and Murphy (1996)	General
ω-factor model	Mosleh et al. (1997); Mosleh et al. (1999)	Nuclear
Integrated Risk (I-RISK)	Papazoglou et al. (2002)	Chemical
(Organizational Risk Influence Model (ORIM)	Øien (2001a, 2001b)	Oil & Gas
Barrier and Organizational Risk Analysis (BORA)	Aven et al. (2006)	Oil & Gas
RISK_OMT	Vinnem et al. (2012)	Oil & Gas
Hybrid Causal Logic (HCL)	Røed et al. (2009)	Oil & Gas
Socio-Technical Risk Analysis (SoTeRiA)	Mohaghegh et al. (2009); Mohaghegh and Mosleh (2009)	General
Phoenix	Ekanem et al. (2016).	General

and identified Risk Influencing Factors (RIFs) can be used for activity related risk analysis.

Real-time data and periodical risk evaluation may be considered as a key improvement to allow for effective decision-making support (Bucelli et al., 2017). However, being static QRA precludes any possible update and integration of the overall risk figures, due to the actual real world ever-changing environment or later improvements based on new risk notions. To overcome this limit, during the last decade several efforts have been devoted to the development of novel approaches to risk assessment and management, which can be considered the dynamic evolution of conditions, both internal and external to the system, affecting risk assessment (Paltrinieri and Scarponi, 2014). The main purpose of dynamic risk assessment is the development of an appropriate technique that allows for the effective aggregation of heterogeneous information and provides risk estimation over time reflecting the current condition of the system (Yang et al., 2017b).

Villa et al. (2016) reported a brief description of the most relevant methodologies and applications of dynamic approaches to risk analysis in the chemical process industry.

Table 5 shows a list of these dynamic methodologies. These methodologies (or techniques) have evolved over time according to Villa et al. (2016) describe in detail. In this way, the list of techniques collected in Table 5 are linked to the most recent citations according to Villa et al. (2016).

Table 5. List of the most relevant methodologies (techniques) of dynamic approaches to risk analysis in the chemical process industry (adapted from Villa et al., 2016).

Technique	Paper Citation
Dynamic Risk Assessment Methodology (DRA)	Kalantarnia et al. (Abimbola et al., 2014; Kalantarnia et al., 2010, 2009; Khakzad et al., 2013a
Dynamic Procedure for atypical scenario identification (DyPASI)	(Paltrinieri et al., 2011, 2013a, 2013b).
Coupling of DRA and DyPASI	(Paltrinieri et al., 2014b, 2014c, 2013a, 2013b).
Dynamic risk assessment with bayesian networks	Khakzad et al. (2013b, 2011)
Risk barometer	The Center for Integrated Operations in the Petroleum Industry (Hauge et al., 2015; Paltrinieri and Hokstad, 2015; Paltrinieri et al.,2015, 2014a)

4 ANALYSIS AND DISCUSSION

The results obtained can be grouped into two groups of techniques for application in the field of occupational safety: standardized and non-standard techniques of a new or emerging nature.

In relation to the standardized techniques, of the 31 techniques included in Annex A of ISO / IEC 31010, the following techniques are among the most used: HAZOP, SWIFT, HRA, FTA, ETA and RCM. In turn, these techniques are developed through specific standards (IEC-EN), except the SWIFT technique.

Regarding non-standard techniques of a new or emerging nature, new risk influence frameworks as well as new dynamic techniques are observed.

Among the first techniques, the following ones can be cited according to its closeness in time (last decade): HCL (2009), RISK_OMT (2012), SoTeRia (2009) and Phoenix (2016). In relation to dynamic techniques, stand out the following: DRA, DyPASI, DRA-DyPASI and Risk barometer.

However, the set of the foregoing results should be considered as an approximation in the framework of occupational safety. This approximation is due to the characteristics of the method followed, which are linked to three important limitations. The first limitation is due to the use of a single database; the second to the inclusion and exclusion criteria used; and the third limitation is due to the lack of differentiating and applicative criteria between the techniques used in the field of safety occupational and safety linked to major accidents.

In relation to the first limitation, it has allowed analyzing the main journals that emerge from the work of Reniers and Anthone (2012). However, these authors analyzed a total of 35 representative safety journals, of which with the present work a total of 11 journals have been analyzed.

Regarding the inclusion and exclusion criteria used, they prevent analyzing the evolution of standardized risk assessment techniques in a broader and deeper way through the scientific literature.

As for the third limitation, with this work the dividing line that exists between occupational safety and safety linked to major accidents has not been analyzed directly (for example, this important aspect can be observed in the Table 5). It is evident that both branches of the safety are interconnected and that therefore share risk assessment techniques with the aim of avoiding accidents in processes and industries.

However, such dividing line between these safety branches is diffuse, so it would be advisable to deepen the analysis of differentiating and

applicative criteria. The differentiating criteria could allow a structured and interconnected risk and loss analysis. These criteria could be aligned with the results of the EU-OSHA (2017). With the application criteria, the scope, use, complexity, strengths and limitations of each technique in the field of occupational safety could be defined.

5 CONCLUSIONS

With this work the main objective consisting in identifying and classifying the main techniques included in the ISO/IEC 31010:2009 standard applicable in the field of occupational safety, was achieved. These techniques could be compatible with the current version ISO/DIS 45001.2:2017 (section 6.1.2.2).

In addition, a secondary objective has also been achieved, that is, to identify the main non-standard techniques of new or emerging nature (especially dynamic characteristics).

However, three important limitations have been identified that can point the direction of future research. Such research could focus on two objectives. The first objective concerns pursuing analyzing in greater depth the impact and degree of development of standardized techniques and non-standard techniques of a new or emerging nature. To do this, the method used must be modified to obtain results with greater representativeness. This modification should include an update of the results obtained by Reniers and Anthone (2012). The second objective concerns the focus on the analysis of differentiating and applicative criteria between the techniques used in the field of safety occupational and safety linked to major accidents.

In any case, the final publications of ISO 45001 as well as ISO/IEC 31010 should be considered, both publications being foreseen for 2018.

ACKNOWLEDGEMENTS

This work was funded by the Ministry of Economy and Competitiveness of Spain, title: "Analysis and Assessment of technological requirements for the design of a New Emerging Risks standardized management SYStem (A2 NERSYS)" with reference DPI2016-79824-R.

REFERENCES

Abimbola, M., Khan, F., Khakzad, N. 2014. Dynamic safety risk analysis of offshore drilling. *J. Loss Prev. Process Ind. 30, 74–85.*

Aven, T., Sklet, S., Vinnem, J.E. 2006. Barrier and operational risk analysis of hydrocarbon releases (BORA-Release): Part I. Method description. *J. Hazard. Mater. 137, 681–691.*

Bucelli, M., Paltrinieri, N., Landucci, G. (in press). 2017. Integrated risk assessment for oil and gas installations in sensitive areas. *Ocean Eng.*

Cornelissen, P.A., Van Hoof, J.J., De Jong, M.D.T. 2017. Determinants of safety outcomes and performance: A systematic literature review of research in four high-risk industries. *J. Safety Res. 62, 127–141.*

Dallat, C., Salmon, P.M., Goode, N (in Press). 2017. Risky systems versus risky people: To what extent do risk assessment methods consider the systems approach to accident causation? A review of the literature, *Saf. Sci.*

Ekanem, N.J., Mosleh, A., Shen, S.-H. 2016. Phoenix – a model-based human reliability analysis methodology: qualitative analysis procedure. *Reliab. Eng.Syst. Saf. 145, 301–315.*

Embrey, D.E. 1992. Incorporating management and organisational factors into probabilistic safety assessment. *Reliab. Eng. Syst. Saf. 38, 199–208.*

European Agency for Safety and Health at Work (EU-OSHA) (December 2 st 2017). Consulted from: https://osha.europa.eu/en/about-eu-osha/press-room/eu-osha-presents-new-figures-costs-poor-workplace-safety-and-health-world

European Committee for Standardization (CEN, 2017). (December 4th, 2017). Consulted from: https://www.cen.eu

Hauge, S., Okstad, E., Paltrinieri, N., Edwin, N., Vatn, J., Bodsberg, L. 2015. Handbook for Monitoring of Barrier Status and Associated Risk in the operational Phase, the Risk Barometer Approach. SINTEF F27045. Trondheim, Norway.

IEC. Dependability management – Part 3–1: Application guide – Analysis techniques for dependability – Guide on methodology. IEC 60300-3-11: 2003. Geneva: IEC, 2003.

IEC. *Fault tree analysis (FTA).* IEC 61025:2006. Geneva: IEC, 2006.

IEC. Analysis techniques for system reliability – Procedure for failure mode and effects analysis (FMEA). IEC 60812:2006. Geneva: IEC, 2006.

IEC. *Application of Markov techniques.* IEC 61165: 2006. Geneva: IEC, 2006.

IEC. Analysis techniques for dependability – Event tree analysis (ETA). IEC 62502:2010. Geneva: IEC, 2010.

IEC. Guidance on human aspects of dependability. IEC 62508:2010. Geneva: IEC, 2010.

IEC. Analysis techniques for dependability – Petri net techniques. IEC 62551: 2012. Geneva: IEC, 2012.

IEC. *Root cause analysis (RCA).* IEC 62740:2015. Geneva: IEC, 2015.

IEC. Hazard and operability studies (HAZOP studies) – Application guide. 61882:2016. Geneva: IEC, 2016.

IEC. *Reliability block diagrams.* IEC 61078: 2016. Geneva: IEC, 2016.

ISO. Systems and software engineering – High-level Petri nets – Part 1: Concepts, definitions and graphical notation. ISO/IEC 15909-1:2004. Geneva: IEC, 2004.

ISO. Supplement 1 – Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM:1995) – Propagation of distributions using a Monte Carlo method. ISO/IEC Guide 98-3-SP: 2008. Geneva: IEC, 2008.

ISO. Risk Management – Principles and Guidelines. ISO 31000:2009. Geneva: ISO, 2009.

ISO. Risk management — Risk assessment techniques. ISO/IEC 31010:2009. Geneva: ISO, 2009.

ISO. Societal security – Business continuity management systems—Requirements ISO 22301:2012. Geneva: ISO, 2012.
ISO. Societal security – Business continuity management systems – Guidelines for business impact analysis (BIA). ISO TS 22317:2015. Geneva: ISO, 2016.
ISO. *Risk management—Risk assessment techniques.* Draft International Standard. IEC/DIS 31010:2017 (E). Geneva: ISO, 2017.
ISO. Occupational health and safety management systems – Requirements with guidance for use. ISO/FDIS 45001. Geneva: ISO, 2017.
ISO. (December 4th, 2017). Consulted from: https://www.iso.org/iso-45001-occupational-health-and-safety.html
Kalantarnia, M., Khan, F., Hawboldt, K. 2010. Modelling of BP Texas City refinery accident using dynamic risk assessment approach. *Process Saf. Environ. Prot. 88, 191–199.*
Kalantarnia, M., Khan, F., Hawboldt, K. 2009. Dynamic risk assessment using failure assessment and Bayesian theory. *J. Loss Prev. Process Ind. 22, 600–606.*
Khakzad, N., Khan, F., Amyotte, P. 2013a. Quantitative risk analysis of offshore drilling operations: a Bayesian approach. *Saf. Sci. 57, 108–117.*
Kjellén, U., Sklet, S. 1995. Integrating analyses of the risk of occupational accidents into the design process Part I: A review of types of acceptance criteria and risk analysis methods. *Saf. Sci. 18, 215-227.*
Marhavilas, P.K., Koulouriotis, D., Gemeni, V. 2011. Risk analysis and assessment methodologies in the work sites: On a review, classification and comparative study of the scientific literature of the period 2000–2009. *J. Loss Prev. Process Ind. 24, 477–523.*
Mohaghegh, Z., Kazemi, R., Mosleh, A. 2009. Incorporating organizational factors into Probabilistic Risk Assessment (PRA) of complex socio-technical systems: a hybrid technique formalization. *Reliab. Eng. Syst. Saf. 94, 1000–1018.*
Mohaghegh, Z., Mosleh, A. 2009. Incorporating organizational factors into probabilistic risk assessment of complex socio-technical systems: principles and theoretical foundations. *Saf. Sci. 47, 1139–1158.*
Mosleh, A., Goldfeiz, E., Shen, S. 1997. The X-factor Approach for Modeling the Influence of Organizational Factors in Probabilistic Safety Assessment. *IEEE Sixth Annual Human Factors Meeting, Orlando, Florida, pp. 9–18.*
Øien, K. 2001a. A framework for the establishment of organizational risk indicators. *Reliab. Eng. Syst. Saf. 74, 147–167.*
Øien, K. 2001b. Risk indicators as a tool for risk control. *Reliab. Eng. Syst. Saf. 74, 129–145.*
Paltrinieri, N., Tugnoli, A., Bonvicini, S., Cozzani, V. 2011. Atypical scenarios identification by the DyPASI procedure: application to LNG. *Chem. Eng. Trans. 24, 1171–1176.*
Paltrinieri, N., Tugnoli, A., Buston, J., Wardman, M., 2013a. DyPASI methodology: from information retrieval to integration of HAZID process. *Chem. Eng. Trans. 32, 433–438.*
Paltrinieri, N., Tugnoli, A., Buston, J., Wardman, M., Cozzani, V. 2013b. Dynamic procedure for atypical scenarios identification (DyPASI): a new systematic HAZID tool. *J. Loss Prev. Process Ind. 26, 683–695.*
Paltrinieri, N., Hauge, S., Dionisio, M., Nelson, W.R., 2014a. Towards a dynamic risk and barrier assessment in an IO context. In: Reliability and Risk Analysis: Beyond the Horizon – Proceedings of the European Safety and Reliability Conference, ESREL 2013. Amsterdam, Netherlands, pp. 1915–1923.
Paltrinieri, N., Khan, F., Amyotte, P., Cozzani, V. 2014b. Dynamic approach to risk management: application to the Hoeganaes metal dust accidents. *Process Saf. Environ. Prot. 92, 669–679.*
Paltrinieri, N., Hokstad, P., 2015. Dynamic risk assessment: development of a basic structure. In: Safety and Reliability: Methodology and Applications – Proceedings of the European Safety and Reliability Conference, ESREL 2014.Wroclaw, Poland, pp. 1385–1392.
Paltrinieri, N., Hauge, S., Nelson, W.R. 2015. Dynamic barrier management: a case of sand erosion integrity. In: Safety and Reliability of Complex Engineered Systems. Proceedings of the European Safety and Reliability Conference, ESREL 2015. Zurich, Switzerland, pp. 523–531.
Paltrinieri, N., Tugnoli, A., Cozzani, V. 2016. Chapter 3 – Advanced Technique for Dynamic Hazard Identification, In *Dynamic Risk Analysis in the Chemical and Petroleum Industry*: 27–36. Butterworth-Heinemann.
Papazoglou, I.A., Aneziris, O.N., Post, J.G., Ale, B.J.M. 2002. Technical modeling in integrated risk assessment of chemical installations. *J. Loss Prev. Process Ind. 15, 545–554.*
Paté-Cornell, M.E., Murphy, D.M. 1996. Human and management factors in probabilistic risk analysis: the SAM approach and observations from recent applications. *Reliab. Eng. Syst. Saf. 53, 115–126.*
Reniers, G., Anthone, Y. 2012. A ranking of safety journals using different measurement methods, *Saf. Sci.* 50 (7), 1445–1451.
Røed, W., Mosleh, A., Vinnem, J.E., Aven, T. 2009. On the use of the hybrid causal logic method in offshore risk analysis. *Reliab. Eng. Syst. Saf. 94, 445–455.*
Tixier, J., Dussuerre, G., Salvi, O., Gaston, D. 2002. Review of 62 risk analysis methodologies of industrial plants. *J. Loss Prev. Process Ind. 15, 291–303.*
Uman, L.S. 2011. Systematic reviews and meta-analysis. Journal of the Canadian Academy of Child and Adolescent Psychiatry, 20(1), 57–59.
UNE (Spanish Association for Standardization). 2017. ISO/DIS 45001.2:2017. Sistemas de gestión de la seguridad y salud en el trabajo. Requisitos con orientación para su uso. [Occupational health and safety management systems. Requirements with guidance for use]. UNE. Madrid.
Villa, V., Paltrinieri, N., Khan, F., Cozzani, V. 2016. Towards dynamic risk analysis: A review of the risk assessment approach and its limitations in the chemical process industry. *Saf. Sci. 89, 77–93.*
Vinnem, J.E., Bye, R., Gran, B.A., Kongsvik, T., Nyheim, O.M., Okstad, E.H., Seljelid, J., Vatn, J. 2012. Risk modelling of maintenance work on major process equipment on offshore petroleum installations. *J. Loss Prev. Process Ind. 25, 274–292.*
Yang, X., Haugen, S., Yuandan, L. 2017a. Risk influence frameworks for activity-related risk analysis during operation: A literature review. *Saf. Sci. 96, 102–116.*
Yang, X., Haugen, S., Paltrinieri, N. (in press) 2017b. Clarifying the concept of operational risk assessment in the oil and gas industry. *Saf. Sci.*

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Indicator on the performance of barriers against fatal accidents in construction

U. Kjellén
NTNU, Trondheim, Norway

ABSTRACT: The paper presents work to develop a safety performance indicator suitable for real-time management of major accident hazards in construction. Data on 60 fatal accidents in the period 2011–2016, resulting in 63 fatalities, have been analysed. About 70% of the accidents belonged to three main categories: fall from height, driver or person outside the cabin killed by moving construction machine/vehicle, and person killed by load or equipment during material handling. The three main categories have been further divided into subcategories (seven in all) and analysed to identify barriers to prevent adverse consequences. This analysis has resulted in checklists, one for each subcategory. They list observable conditions at a construction site that, if found substandard, will indicate that one or more of the important barriers are seriously deteriorated. The paper highlights the results of the accident concentration and barrier analyses. It also reviews remaining work to develop and test the performance indicator.

1 INTRODUCTION

Construction activities are characterized by the management of large amounts of energy such as in transportation, excavation, assembly, work at height etc. Loss of control of the energy flow may have major consequences. Natural hazards (rock fall, land slide etc.) represent significant additional risks. The statistics on severe accidents in construction reflect these conditions. According to ILO estimates, the fatal accident risk in construction is five times the average among workers worldwide (Murie 2007). Statistics from Norway for 2009–2014 show a fatal accident frequency rate of three times the general average for workers (Norwegian Labour Inspection Authority 2015).

Construction work is organised in projects with a limited duration. A project goes through different phases from site establishment and excavation to installation and completion, and the conditions at site and the activities change accordingly. Traditional safety management using performance measurements (such as the TRI rate) and feedback control is inadequate, due to lagging characteristics of most current safety performance indicators (Kjellén 2009, Lingard et al. 2017). There is a need for indicators that provide real-time data on safety performance to ensure timely feedback for control of safety performance (Kjellén 2018).

Behavioural sampling was developed in the 1950s to meet this requirement (Rockwell 1959). The method uses observations on deviations from safe work practices and conditions as data input. The "TR safety monitoring method" represents an application of behavioural sampling to the construction industry (Laitinen et al. 1999, Laitinen & Päivärinta 2010). Experiences show that the method produces reliable and valid results related to the prevention of ordinary occupational accidents.

There is a general lack of 'real-time' performance indicators suitable for the control of hazards in construction with fatal accident potential. The principles behind the "barrier performance indicators" developed by the process and oil and gas industries for the prevention of fires and explosions offer such as opportunity (Health and Safety Executive 2006, OGP 2011). The indicators measure the compliance of safety barriers to a standard.

Accident statistics indicate that this approach may be valid for the construction industry, despite the high variety of activities in the industry. A relatively small number of types of central events according to the bow-tie accident model, each representing the loss-of-control of significant amounts of energy, account for a substantial share of the fatal accidents in construction (Visser 1998, Swuste et al. 2012). By identifying barriers to prevent these central events and/or reduce their consequences, input may be provided to performance indicators on the risk of fatal accidents in construction projects.

Experiences from two case studies support the validity of this approach. An Indian hydropower project reported eight fatalities due to road departures and falling rocks in tunnels during 2,5 years

after start-up (Kjellén 2012). The project implemented improved safety routines directed at barriers to prevent these types of accidents. It was completed three years later with one additional fatality not related to these types of events.

A large international construction contractor identified six dominating concentrations of fatal accidents in their operations world-wide (A. Berglund, personal communication, Nov. 17, 2017). These included falls from height, conflict between human and machine, structural failure, lifting operations (falling objects), fire/explosions, and electric arcs. The contractor implemented lifesaving rules directed at barriers to prevent these types of accidents. They experienced a reduction in the frequency of the affected types of fatal accidents by 60% in a five-year period after the intervention, compared to the previous five-year period. The number of construction workers was about the same in the two periods.

1.1 *This paper*

The paper presents research that is part of an ongoing project for the construction industry. Its aim is to develop safety performance indicators that are better suited for the management of safety by client and contractor companies than the lagging safety performance indicators in use today.

The research presented here focuses on developing a performance indicator of the availability of barriers against hazards with fatal accident potential. The intention is to provide data in 'real time' and thereby allow the involved companies to accomplish effective feedback control of the fatal accident risk at construction sites.

The paper highlights the first part of the research, aiming at identifying dominating fatal accident concentrations in the Norwegian construction industry and critical barriers that will prevent the relevant types of accidents from occurring. In the subsequent steps, these results will be used as input to the development of indicators for barrier availability for use by safety practitioners in the industry. These indicators will be tested and evaluated in intervention studies in construction projects.

2 MATERIAL AND METHOD

2.1 *Sources of data*

The analysis consists of two parts, an accident concentration analysis and a barrier analysis. A set of 60 fatal accidents resulting in 63 fatalities in the Norwegian construction industry between 2011 and 2016 represents the source data for the accident concentration analysis. The Norwegian Labour Inspection Authority (NLIA) provided the data from their general register of fatal occupational accidents in Norway.

The accident data was documented in a spreadsheet with the following information on each accident: date of the accident, registration number, type of construction business (classification), type of activity (classification), number of people killed, type of injury (classification), accident type (classification), free text resumé of the sequence of events.

In the subsequent barrier analysis, data from NLIA was supplemented with data from other sources:

- Observations at three construction sites and interviews with senior construction managers
- The author's library of in-depth investigations into fatal accidents and high potential incidents that fall within any of the accident concentrations selected for further analysis
- Lifesaving rules developed by major construction contractors
- Relevant regulatory requirements in Norway

The results were reviewed in workshops by high-level safety experts from construction client companies, contractors and an organisation of regional safety representatives.

2.2 *Methods*

2.2.1 *Accident concentration analysis*

An accident concentration analysis aims at identifying clusters of 'accident repeaters' with common characteristics (Kjellén & Albrechtsen 2017). By directing preventive measures at a selection of dominating accident concentrations, a significant risk reduction is expected.

The analysis is carried out stepwise in several dimensions to identify accidents with common characteristics. A natural starting point for the analysis of fatal accidents was to group the accidents according to the main type of energy involved.

In the current study, the dataset was first analysed by accident type according to NLIA's classification. For each accident type, the free-text descriptions of the accidents were reviewed to identify common patterns. A new, composite classification of the accidents was developed, where each 'accident concentration' was made-up of at least three fatal accidents with common characteristics.

2.2.2 *Barrier analysis*

The barrier analysis is rooted in the principles of defence in depth and the energy model of loss causation (Rasmussen 1993, Haddon 1980).

A model of the accident sequence encompassing three successive phases is shown in Figure 1

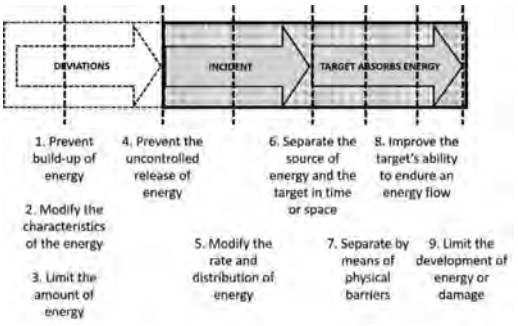


Figure 1. Barrier intervention in the accident sequence to eliminate or reduce loss (Kjellén & Albrechtsen 2017).

(Kjellén & Larsson 1981). Nine of Haddon's ten accident prevention strategies are introduced in the model as barrier functions. Each of these may, dependent on the type of accident, have the capacity to change the sequence of events and thereby eliminate or reduce loss.

The function of a barrier is realised through a barrier system. Whereas passive barriers consist of a physical element, active barriers are more complex with several elements including a control system. One or more human operators may be part of the control loop.

Input data to the barrier analysis consisted of the results of an analysis of accident concentrations among fatal accidents in construction in Norway between 2011 and 2016. For each accident concentration, barriers that are critical in preventing accidents in the accident concentration in question were identified. Next, the required performance of the elements that make up the barrier were identified for each identified barrier.

3 RESULTS

3.1 *Accident concentration analysis*

The distribution of the 60 fatal accidents in NLIA's database by accident type is shown in Table 1. This analysis represents the starting point for the identification of accident concentrations.

Falls represents the most common accident type in NLIA's database, followed by squeezed/caught and hit by object.

The analysis revealed some inconsistencies in the accident type classification by NLIA. A2 and A3 include accidents involving vehicles, but similar vehicle related accidents were classified as category A4.

The accident concentration analysis identified 12 categories of 'accident concentrations', covering 93% of the fatal accidents in the material,

Table 1. Distribution of accidents in NLIA's database on fatal accidents in construction in the period 2011–2016 (N = 60). The table also shows a summary description of accidents with common characteristics for each accident type.

Accident type (NLIA)	# of events	Coarse description of accidents
A1 Hit by object	9	Hit by load or machinery part during material handling, structural collapse, falling rock, flying object (bolt from pistol)
A2 Collision, hit by vehicle	11	Road/work area departure by vehicle (incl. mobile machine), collision, hit by vehicle
A3 Roll-over	3	Roll-over of vehicle
A4 Squeezed or caught	13	Hit by vehicle, road/work area departure by vehicle, roll-over of mobile machine, squeezed by lifting/transportation equipment during operation, rock fall, collapsing trench, structural collapse
A5 Fall	19*	Electric current through body (resulting in fall), fall from height (roof, deck, scaffold, ladder, machinery/equipment)
A7 Electric voltage	2	Electric current through the body from electric installation or hand tool
A10 Explosion, fire	3**	Accidental blast, explosion

* Two persons killed in one fall accident, ** Three persons killed in one blast accident.

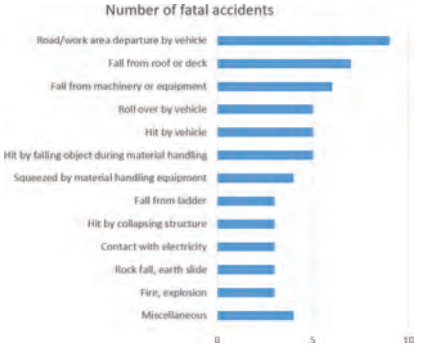


Figure 2. Distribution of fatal accidents by 'accident concentration' (N = 60).

Figure 2. The category 'miscellaneous' made up the remaining 7%.

28% of the accidents were related to fall from height. Similarly, 23% involved vehicles (including

mobile construction machinery). If we add material handling by crane, conveyor belt, truck etc., we cover in all two thirds of the fatal accidents.

Accident concentrations with four or more fatalities were selected for further analysis. These included:

- Road or work area departure by vehicle: The driver was killed due to loss-of-control of the vehicle followed by departure from a road or construction area.
- Fall from or through roof or deck: The person being killed fell either outside the edge of a roof or deck or through and opening or weak point of the roof/deck.
- Fall from machinery or equipment: The person being killed fell when moving or working on machinery or equipment. Falls from ladder was introduced as a special case.
- Roll over of vehicle (or machine during transfer): The driver/operator has been killed due to roll-over during unloading, during transfer (when a machine behaved as a vehicle), or when a parked vehicle has accidently started to move.
- Hit by vehicle: The person being killed was present in the danger zone of a vehicle in motion and the driver was either not aware of the person or lost control of the vehicle.
- Hit by falling object during material handling: The fatalities occurred during crane handling or unloading of truck or trailer. The accidents involved loss of control of load or sudden, uncontrolled movements of equipment.
- Squeezed by movements of personnel lifts or material handling equipment: The person being killed was either squeezed by uncontrolled machinery movements, missing machine guarding or because the operator was not aware of the person being in the danger zone.

3.2 *Barrier analysis*

A basic assumption in this research project is the considerations of barriers as an additional measure implemented in a production system to achieve a tolerable level of risk. It means that the basic design of the work system, where the fatal accidents occurred, has not been questioned. Surface transportation, for example, was not considered as a barrier function to avoid being hit by falling objects during material handling. In Table 2, barrier types 1 and 2 have been excluded for this reason.

The barrier analysis has been based on some additional assumptions:

3. Limit the amount of energy: Has been included in barrier function no. 8 for falls from height (use of fall arrest system).

Table 2. Results of the barrier analysis. A barrier type marked "X has a function that is considered essential to prevent fatalities in the accident concentration in question.

	Type of barrier (see Figure 1):						
Accident concentration	3	4	5	6	7	8	9
Road/work area departure by vehicle		X	X		X	X	X
Fall from or through roof or deck	X	X	X		X		
Fall from machinery or equipment	X	X	X		X		
Fall from ladder	X	X	X				
Roll over of vehicle (or machine during transfer)		X	X		X		
Hit by vehicle		X		X			
Hit by falling object during material handling		X		X			
Squeezed by movements of lifting and material handling equipment		X		X	X		

4. Prevent uncontrolled release of energy: This barrier function is relevant to all analysed main types of accident concentrations. The barrier function is not relevant to subsets of accidents, where a machine operator has not been aware of personnel being present in the danger zone.
5. Modify the rate and distribution of energy: This applies to the use of safety belt in case of road/work area departure or roll-over and use of fall arrest system.
6. Separate the source of energy and the target in time or space: This barrier function is relevant to vehicle and material handling related accidents. A possible exception is sudden machinery movements due to mechanical failure, which could not reasonably have been foreseen.
7. Separate by means of physical barriers: This barrier function is relevant to all types of accident concentrations. In practice, the implementing this type of barrier is restrained by feasibility considerations.
8. Improve the target's ability to endure an energy flow: This barrier function is implemented through personal protective equipment. It is especially relevant to falls, when collective measures (physical barriers) are not applied. Use of helmet has limited effect in case of falling objects with high energy.
9. Limit the development of energy or damage: The general organisation of emergency response at site falls outside the scope of this analysis. Preparedness directed at specific types of hazards has been included such as use of mobile machine close to water way. In this case, the barrier relates to preparedness to safe the driver from drowning in case of departure from road or work area.

Table 3. Requirements related to barrier elements for each of three barriers essential to prevent fall from ladder.

Limit the amount of energy	Prevent uncontrolled release of energy	Modify the rate and distribution of the energy transfer
• Ladder used as an exception in lack of feasible alternative?	• Ladder checked and found in adequate condition?	• Use of safety harness and life line hooked to a solid attachment point
• Ladder used for a height difference to solid ground <6 m?	• Ladder positioned on solid ground and secured from tilting? • Ladder raises > 1 m above roof or landing?	

Table 4. Requirements related to barrier elements for each of two barriers essential to prevent personnel being hit by vehicle.

Prevent uncontrolled release of energy	Separate the source of energy and the target in time or space
• Vehicle is certified and regularly checked and maintained?	• Areas for transportation and loading/unloading separated from work areas and pedestrian traffic?
• Driver instructions regarding controlled operation of the vehicle available?	• Driver has full overview from the cabin of the operating zone?
• Adequate training and certification of driver?	• Driver instructions to have control of the operating zone for other personnel? Adequate operator training and certification?
• Foundation of the operating area of the vehicle adequately stable, inclination and friction satisfactory to ensure full operator control?	• Site instructions and training regulating avoidance of the operating zone of mobile machines? Adequately supervised and respected? • Use of highly visible uniforms by all personnel at the site? • Adequately illuminated work areas and roads?

The next step in the analysis has aimed at identifying performance requirements of barrier elements necessary for the applicable barrier function to be effective. These requirements provide the basis for the barrier performance indicator, which will measure average percent compliance with the requirements. The results of this analysis are illustrated by two examples in Table 3 and Table 4.

4 FURTHER WORK

Work is under way to develop and operationalise the identified performance requirements in cooperation with Client and Contractor personnel of a large infrastructure construction site. The requirements for successful performance of the different barrier elements will be scrutinized by site personnel in cooperation with the researchers and further developed to meet certain quality criteria (Kjellén & Albrechtsen 2017, Laitinen et al. 1999):

– Flexible for use at different types of construction sites,
– Transparent and easily understood by site personnel,
– Address site conditions that are observable and quantifiable,
– Sensitive to changes in the safety standard at construction sites,
– Produce reliable results when applied by different observers,
– Robust against manipulation.

The task of the observers will be to identify work at the construction site involving any of the seven identified major accident hazards. Next, the observers will use the relevant checklist to review the status of the different barrier elements and classify them as either correct or incorrect or not relevant. Incorrect barrier elements will be qualified through a short description. The results will be summarized, showing% of the checked barrier elements that are correct. This can be shown in total and for each type of major accident hazard.

To produce reliable results, characteristics of the various barrier elements that separate correct from incorrect must be defined in the checklists to the extent needed when used by experienced personnel. In addition to physical observations of the work, the observers will have to consult available documentation and make interviews to check items such as operator training and qualifications and maintenance standard of vehicle. This makes the use of the method to resemble a system audit more than an inspection, but not fulfilling the requirement to independence.

The next step of the research project will include testing and evaluation of the barrier performance

indicator and underlying requirements to barrier elements in routine safety practice. A systematic approach will be applied in assessing the operational experience of the intervention itself, and various stages of outcome, including degree of implementation and immediate, intermediate and end results (Shannon et al. 1999). It will not be possible to monitor any effects on the fatal accident rate by this trial.

5 CONCLUSIONS

The work presented in this paper represents the first phase in developing a real-time performance indicator aiming at managing the risk of fatal accidents.

An important prerequisite has been the existence of dominating fatal accident concentrations in construction that may be prevented through a few well-defined barriers. This will allow the development of an indicator based on requirements for barrier availability that is adequately comprehensive still practicable. Results of the accident concentration analysis shows that this prerequisite generally is satisfied. There exists variation between individual accidents in the concentrations that are more complex. It is necessary in the further development and operationalisation of the requirements to check for these variations to ensure that the prerequisite is satisfied.

Another critical issue is the representativeness of the data used in this analysis for the fatal accident risk in construction in Norway. The period and number of events covered by the analysis, and the consistence of the results with findings in international research and safety practice referred to in this paper, call for a positive answer to this question.

The ultimate test of the viability of the proposed performance indicator will take place through integration is safety practice at selected construction sites. Here it will be possible to evaluate the indicator against predefined criteria. Only extensive and long-term use will validate the performance indicator as an efficient tool in preventing fatal accidents.

REFERENCES

Haddon, W. 1980. The basic strategies for reducing damage from hazards of all kinds. *Hazard Prevention* 16:8–12.

Kjellén 2018. Experience feedback. In: Niklas Möller, Sven Ove Hansson, Jan-Erik Holmberg & Carl Rollenhagen (eds), *Handbook of Safety Principles*. Hoboken, NJ: Wiley, Essentials in Operations Research and Management Science.

Kjellén, U. & Albrechtsen, E. 2017. *Prevention of accidents and unwanted occurrences—Theory, methods, and tools in safety management.* Boca Raton, FL: CRC Press.

Kjellén, U. & Larsson, T.J. 1981. Investigating accidents and reducing risk—a dynamic approach. *Journal of Occupational Accidents* 3:129–140.

Kjellén, U. 2009. The safety measurement problem revisited. *Safety Science* 47:486–489.

Kjellén, U. 2012. Managing safety in hydropower projects in emerging markets—Experiences in developing from a reactive to a proactive approach. *Safety Science* 50:1941–1951.

Laitinen, H. and Paivarinta, K. 2010. A new generation safety contest in the construction industry—a long-term evaluation of a real-time intervention. *Safety Science* 48: 680–686.

Laitinen, H., Marjamäki, M. and Päivärinta, K. 1999. The validity of the TR safety observation method on building construction. *Accident Analysis & Prevention* 31:463–472.

Lingard, H., Hallowell, M., Salas, R. and Pirzadeh, P. 2017. Leading or lagging? Temporal analysis of safety indicators at a large infrastructure construction project. *Safety Science* 91:206–220.

Norwegian Labour Inspection Authority 2015. Arbeidsskadedødsfall i Norge—Utviklingstrekk 2009–2014 og analyse av årsakssammenhenger i fire næringer (Occupational fatalities in Norway—Trends 2009–2014 and Analysis of causes). *Kompass tema* 3.

Rasmussen, J. 1993. Learning from experience? How? Some research issues in industrial risk management. In B. Wilpert and T. Qvale (eds.), *Reliability and safety in hazardous work systems*: 43–66. Hove, UK: Lawrence Erlbaum Associates.

Rockwell, T.H. 1959. Safety performance measurement. *Journal of Industrial Engineering* 10:12–16.

Shannon, H.S. and Manning, O.P. 1980. Differences between lost-time and non-lost time industrial accidents. *Journal of Occupational Accidents* 2:265–272.

Shannon, H.S., Robson, L.S., and Guastello, S.J. 1999. Methodological criteria for evaluating occupational safety intervention research. *Safety Science* 31:161–179.

Swuste, P., Frijters, A. and Guldenmund, F. 2012. Is it possible to influence safety in the building sector? A literature review extending from 1980 until the present. *Safety Science* 50:1333–1343.

Visser, J.P. 1998. Developments in HSE management in oil and gas exploration and production. In A.R. Hale and M.S. Baram (eds.), *Safety management—The challenge of change*: 43–66. Bingley, UK: Pergamon.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Maritime safety culture and safety behaviours in Greece and Norway: Comparing professional seafarers and private leisure boat users

T.O. Nævestad
Institute of Transport Economics, Norway

A. Laiou
Department of Transportation Planning and Engineering, National Technical University of Athens, Greece

K.V. Størkersen
NTNU Samfunnsforskning, Norway

R. Phillips
Institute of Transport Economics, Norway

G. Yannis
Department of Transportation Planning and Engineering, National Technical University of Athens, Greece

T. Bjørnskau & A. Amundsen
Institute of Transport Economics, Norway

ABSTRACT: The present study compares professional seafarers and private leisure boat users in Norway and Greece. The aims of the present study are to examine the safety behaviours related to personal injuries and accidents among these groups and to study the factors influencing these behaviours. This will serve as a backdrop to a general discussion of why the level of fatalities is higher among private boat users than among professional seafarers and what the former may learn from the latter. The study is based on surveys to crew members on Norwegian and Greek cargo and passenger vessels and leisure boat users in Norway and Greece. Our study indicates that while unsafe behaviours related to work pressure and risk taking are important among professional seafarers (i.e. risk acceptance and violations), unsafe behaviours related to the leisure/holiday situation was important for the leisure boat users (i.e. alcohol use while driving a boat). Additionally, we discuss how the situation of private leisure boat users is less regulated than that of professional seafarers. Our study indicates that both in the professional and the private setting, norms for interaction and conduct seem to be influenced by norms and expectations rooted in different socio-cultural groups, e.g. the national culture, the specific sector in question, the organisations and in peer groups.

1 INTRODUCTION

1.1 *Background*

Maritime transport is a substantial part of world trade, as approximately 90% of the goods traded worldwide are transported by sea. Although safety improvements have led to a significant decrease in the mortality rates of seafarers in recent decades, seafaring is still termed one of the most hazardous occupations (Oldenburg & Jensen 2012). In a questionnaire study including 6461 participants in 11 countries, Jensen et al. (2004) found that during the latest tour of duty, 9% of all seafarers were injured and 4% had an injury with at least 1 day of incapacity. According to Nævestad et al. (2015) there were on average 15 killed and 424 injured annually on Norwegian ships, i.e. Norwegian Ordinary Ship Register (NOR) and Norwegian International Ship Register (NIS) in the period 2004–2013. At EU level, in the period 2011–2016, there were on average 100 fatalities and 935 injuries annually reported in the European Marine Casualty Information Platform (EMCIP) (EMSA, 2017).

In Norway, as in many other countries, the number of persons killed using different transport modes has decreased since 1970. This is especially true for road transport, but it also applies to professional seafarers (Nævestad et al 2015). Leisure

boating on the other hand, have not had the same positive development, and in Norway the number of deaths using leisure boats have more or less stayed the same since 1995 (with some annual variations).

In 2006–2015, on average 33 people died each year in recreational boating accidents in Norway, which is about 0.65 persons per 100 000 inhabitants. Approximately 4,4 persons have died per 100 000 vessels (registered and not). More than 90% of these persons are men, and a majority of the victims were not wearing personal floating devices (PFD). Most of the accidents happen during the summer months, when the boats are in more use. Only 20–30% of the people found dead was wearing a lifejacket.

1.2 *Aims*

The present study compares professional seafarers and private leisure boat users in Norway and Greece.

The aims of the study are to examine the safety behaviours related to personal injuries and accidents among these groups and to study the factors influencing these behaviours. This will serve as a backdrop to a general discussion of why the level of fatalities is higher among private boat users than among professional seafarers and what the former may learn from the latter.

The data in this study have been collected as part of the SafeCulture project, which is funded by the Norwegian Research Council, and undertaken by the Institute of Transport Economics—TØI (Norway) and the National Technical University of Athens—NTUA (Greece).

The research on safety culture suggests that if we are to fully understand its effects on safety in transport, we should study not only safety culture particular to organisations, but that particular to peer-groups, sectors, regions and nations. We define transport safety culture (TSC) as shared norms prescribing certain transport safety behaviours, shared expectations regarding the behaviours of others and shared values signifying what's important (e.g. safety, mobility, respect, politeness) (Nævestad & Bjørnskau, 2012). An important aspect of our approach is that overall TSC is a composite of overlapping safety cultures associated with different types of sociocultural unit. Thus, we apply the safety culture concept to the national level, to organisations and to peer groups in the present study.

1.3 *Previous research*

There seem to be few studies examining the relationship between safety behaviours and work accidents in the maritime sector, although there are some exceptions (cf. Håvold and Nesset, 2009). The existing studies within this area do, however, indicate that demographic factors (age, nationality, position, line of work) influence work accident risk, and we should assume that this relationship is mediated by some kind of unsafe behaviour (e.g. risk taking, violations), resulting in injuries. Younger seafarers have a higher risk (Hansen et al 2002; Jensen et al 2004). Foreigners have a considerably lower accident risk than local (in the specific study, Danish) citizens (Hansen et al 2002). Previous research also indicates that alcohol consumption may be an important risk factor in the maritime sector (Akhtar & Bouwer Utne 2014, Hetherington et al 2014), and that alcohol and drug abuse are greater for seafarers compared to workers ashore (Nitka 1990; Kariris 2012 in Zhang & Zhao 2017), partly because of their working situation (e.g. social isolation). However, given the relatively unregulated character of private boat use, we may perhaps assume that alcohol consumption "boating while under the influence", is an even more important risk factor in this sector. Likewise, me may perhaps also assume that the other risky behaviours related to boating accidents (e.g. over speeding close to shore) are more prevalent among the less regulated private boat users.

Based on a review of previous foreign studies of recreational boating accidents in Norway, Amundsen (2016) asserts that questions about alcohol use and lifejacket use are common in almost all of the international surveys. We may infer from this that alcohol use and life jacket use are key safety behaviours influencing the risk of accidents among private leisure boat users. Amundsen (2016) reports that the questions used in the different countries are adapted to the specific use of leisure boat in that country, and the accident situation. Based on a review of studies relevant to Norway, Amundsen and Bjørnskau (2017) point to the following safety behaviours as likely to influence the safety of private boat users: Drive faster than the permitted speed close to shore, Carry more passengers than the boat is licensed for, Drink a beer or a glass of wine before going boating, Drive in the dark without using the lantern/lights, Wearing a life jacket, Carrying enough lifejackets for everybody onboard the boat. The questions are partly based on the findings from a review of the safety situation for the recreational boaters performed by the Norwegian Maritime Authority in 2012.

Moreover, it is also important to ask whether the difference between the two groups are due to differences in private and professional maritime safety culture in Norway and Greece. The professional maritime safety culture is closely related to the safety regulation (e.g. the ISM-code) in

professional maritime transport. The International Safety Management (ISM) code of the International Maritime Organisation requires shipping companies to implement Safety Managemen Systems (SMS) on board their vessels, including describing safety roles, goals, procedures, monitoring, reporting, follow up etc. (Thomas 2012). Studies indicate that the SMS requirements of the ISM code foster a positive safety culture on board vessels (Lappalainen et al 2014). Additionally, shipping companies also often work to implement a positive organizational safety culture, including policies for seafarer behavior. Based on previous research, we may hypothesize first that organizational safety culture influences safety behaviours among professional seafarers (cf. Håvold & Nesset 2009, Lu & Tsai 2010).

Also, professional seafarers have undergone an IMO approved training in their respective home countries. Thus, this training, the SMS and safety culture are elements which are likely to influence the professional maritime safety culture. Additionally, it is important to remember that professional seafarer culture also is likely to be influenced by the working conditions of professional seafaring, which may include a high work pressure, demanding working conditions, fatigue etc. (cf. Nævestad 2017). Størkersen et al (2011) found that a third of the respondents in the Norwegian coastal cargo sector reported that they put themselves in danger to get the job done, while about 40% violate procedures to get the job done, especially because of efficiency demands (Størkersen et al 2011).

Moreover, research has also highlighted the importance of national safety culture for the safety behaviours of professional seafarers (Håvold 2005). We compare two countries (Norway and Greece), and we therefore, also compare the influence of national safety culture. The theoretical link between safety culture and safety behaviours is often omitted in research (Ward et al 2010). In the present study, we conceptualise this relationship as both direct social pressures and more subtle social mechanisms, producing important normative influences on behaviour (Cialidini et al., 1990). Individuals' perceptions of peers' opinions about a given behaviour are often defined as injunctive norms, while individuals' perceptions of what peers actually do often are defined as descriptive norms (Ajzen 1991; Rivis & Sheeran 2003; Ward et al 2010). Since injunctive norms are normative they can be expected to directly influence peoples' behaviour (Cialidini et al. 1990). In the present study national culture is measured as descriptive norms. Descriptive norms may influence behaviour by providing information about what is normal, but they can also influence behaviour through the false consensus bias, in which individuals overestimate the prevalence of risky behaviour among their peers in order to justify their own behaviour. The focus on normative influences on behaviour is important in the theory of planned behaviour (TPB) (Ajzen, 1991, 2006), and in the critique of it (Rivis & Sheeran 2003). In short, TPB predicts that our behaviour is the result of our intention to carry out the behaviour, and that our intention to carry out a particular behaviour is influenced by our attitudes towards the behaviour, injunctive norms and our perceived control over our behaviour (Ajzen 1991, 2006).

Additionally, research on maritime safety has found that the framework conditions and safety level varies considerably between (sub)sectors (Størkersen 2017; Hansen et al 2002; Jensen et al 2004). The influence of sector and sector safety culture is examined for professional seafarers. Studies of private transport operators have found that other sociocultural groups, e.g. peer groups (Nævestad et al 2014) and region (e.g. urban vs. rural) (Rakauskas et al 2009) are important when it comes to influencing safety behaviours Thus, we also seek to examine the influence of peer-groups and regional maritime safety culture. Boat type and background variables are examined for leisure boat users.

2 METHOD

2.1 *Recruitment of respondents*

The Norwegian professional seafarers were recruited through the Norwegian researchers' contact with Norwegian shipping companies, i.e. shipping companies that are located in Norway, with mainly Norwegian crew members. Web links to the questionnaires were distributed by the shipping companies to all employees working on board vessels, along with an introductory text explaining the purpose of the survey, and stressing that the surveys were confidential. The Norwegian private boat users were recruited through a) the Norwegian researchers' contact with a boating association distributing survey links to members, and b) distribution on a member website for leisure boat owners, which in many years has been Scandinavia's largest boat forum (e.g. with 1.6 million posts submitted by members). The Greek professional seafarers were recruited through a marketing research company in Greece, which was under the scientific supervision of researchers from the NTUA. Seafarers working for Greek shipping companies, i.e. shipping companies that are located in Greece, with mainly Greek crew members, were approached. Private boat users in Greece were also recruited by the same marketing research company.

2.2 *Survey measures: Professional seafarers*

The present paper analysis of professional seafarers builds on and takes further the knowledge gained from two previous conference papers. The first (Nævestad et al 2017) compares organizational safety culture, working conditions and occupational injuries in Norwegian cargo and passenger transport. The second (Nævestad et al 2018), compares cultural influences on maritime cargo transport in Norway and Greece. The present paper takes the knowledge from these two papers further, as it compares professional seafarers with private leisure boat users.

1. **Background variables** (15 questions): e.g. gender, nationality, age group, seafarer experience, position/area of work, employment status, vessel type, vessel size, manning on board, ship register,
2. **Safety performance** (5 questions): respondents' occupational injuries on board, ship accidents, type of ship accidents, safety compromising fatigue and assessment of work place safety level (1–10).
3. **Safety behaviours** (7 questions): questions on safety behaviours. Respondents were asked: How often do you think the following events tend to occur for every 100 working days/nights on board?: 1) I accept small risks because the "situation demands it" (e.g. because of time pressure, bad weather), 2) I violate procedures to get the job done, 3) I work, even though I am so tired that safety may be compromised, 4) I refrain from using the required protection equipment in my work, 5) I work while being under the influence of alcohol (e.g. one beer or more), or while being hungover, (Answer alternatives: 1) Never, 2) 1–2 times, 3) 3–5 times, 4) 6–10 times, 5) 11–15 times, 6) 16–20 times 7) More than 20 times, 8) Do not know/not relevant)
4. **Working conditions** (3 questions): How often do you think the following events tend to occur for every 100 working days/nights on board: 1) Your shift change is delayed because of work operations, for instance port calls?, 2) You work more than 16 hours in the course of a 24-hour period?, 3) You are interrupted when you are off duty". (Answer alternatives: 1) Never, 2) 1–2 times, 3) 3–5 times, 4) 6–10 times, 5) 11–15 times, 6) 16–20 times 7) More than 20 times, 8) Do not know/not relevant).

We removed the eight answer alternative and made a "Demanding working conditions index" of these three questions (Cronbach's Alpha:.728). The survey also included a question on work pressure: "Sometimes I feel pressured to continue working, even if it is not perfectly safe"

5. **Organisational safety culture** (7 questions): We made an organisational culture index, consisting of questions from the GAIN-scale on organisational safety culture. We have used this scale in previous research from different transport sectors (Bjørnskau & Longva, 2009; Nævestad & Bjørnskau, 2014). The GAIN-scale is presented in the "Operator's Safety Handbook" (GAIN 2001). The GAIN-scale originally consists of 25 questions measuring five themes, but we have reduced the scale to 7 questions, e.g. 1) Ship management regards safety to be a very important part of all work activities, 2) The shipping company regards safety to be a very important part of all work activities, 3) Ship management detects crew members who work unsafely, 4) Ship management often praises crew members who work safely etc.
6. **National safety culture** (7 questions): In the present study we measure national safety culture as descriptive norms (Cialdini 1990) at the national level meaning "what respondents expect that other seafarers from their own country do" expressed through question"When working on vessels, I expect the following behaviours from other seafarers from my country:" 1) That they sometimes violate procedures to get the job done, 2) That they sometimes refrain from using the required protection equipment in their work, 3) That they sometimes work, even when they are so tired that safety may be compromised, 4) That they sometimes work being under the influence of alcohol (e.g. one beer or more), or while hungover, 5) That they sometimes take small risks if the "situation demands it" (e.g. because of time pressure, bad weather), 6) That they sometimes avoid telling colleagues taking risks to work safely, 7) That they sometimes refrain from reporting safety problems and unsafe situations that they experience in their work to the ship management. An exploratory factor analysis (EFA) was conducted to examine the underlying factor structure of the 7 national safety culture (descriptive norms) items.
7. **Sector safety focus** (2 questions): We measure sector safety focus by means of two questions that were selected after a "scale if items deleted" analysis (including five items): 1) Safety is more important than deadlines to our customers, 2) Safety is more important than price to our customers (CA = .875).

2.3 *Survey measures: Private boat users*

1. **Background variables** (12 questions): gender, nationality, age group, experience as a boat driver, participation in organized boat training/ educational programme, boat type, use of navigation equipment, boat length, engine capac-

ity, maximum boat speed, purpose of boat use, municipality of residence, education level.

2. **Safety performance** (4 questions): respondents' accidents/incidents, injuries in accidents, safety self-assessment as a boat driver (1–10), boat use duration.
3. **Safety behaviours:** (12 questions): Respondents were asked:
   A. For every ten times you are driving your boat, approximately how often do you do the following things, before you go out: 1) Tell someone where I will be going and when I will be back, 2) Check the weather forecast, 3) Check the fuel level, 4) Drink two units of alcohol (e.g. two beers, two gl. of wine)
   B. For every ten times you are driving your boat, approximately how often do you do the following things: 1) Personally wear a life jacket the entire trip, 2) Drink two units of alcohol (e.g. two beers, two glasses of wine), 3) Drive faster than the permitted speed close to shore, 4) Drive so fast or offensively that passengers or others (e.g. other boat drivers) express concern or react in other ways, 5) Look down at navigational equipment/GPS for so long that I have been surprised to see other boats, islands, skerries etc. when I look up, 6) Become angered by a certain type of boat driver and indicate your hostility by whatever means you can.
   C. For every ten times you are driving your boat with passengers, approximately how often do you do the following things: 1) Ensure that adult passengers on your boat wear a lifejacket, 2) Ensure that child passengers on your boat wear a lifejacket. (Answer alternatives for A, B and C: 1) Never, 2) 1–2 times, 3) 3–4 times, 4) 5–6 times, 5) 7–8 times, 6) more than 8 times but not always, 7) Always
4. **National safety culture** (3 questions): National safety culture is again measured as descriptive norms at the national level meaning "what respondents expect that other boat drivers from their own country do" expressed through question "Based on your experience, how many boat drivers in your country do you think do the following:" 1) Drink two units of alcohol (e.g. two beers, two glasses of wine) while driving the boat, 2) Drive faster than the permitted speed close to shore, 3) Drive so fast or offensively that passengers or others (e.g. other boat drivers) express concern or react in other ways. The questions were combined into an index. The survey included six additional questions about this that are not listed here.
5. **Peer group safety culture** (3 questions): The same principle and questions as for national safety culture are applied.
6. Safety culture at municipality level (3 questions):

The same principle and questions as for national safety culture are applied.

3 RESULTS

3.1 *Professional seafarers*

3.1.1 *Which behaviours influences personal injuries?*

A logistic regression analysis was conducted with personal injuries as dependent variable, to find the variables predicting personal injury among our respondents (Table 1). In this analysis, the injury variable, which originally had four answer alternatives, was dichotomized, 0 = no personal injury, 1 = personal injury. B values are presented and they indicate whether the risk of personal injuries is reduced (negative B values) or increased (positive B values), when the independent variables increase by one value.

Table 1 provides three main results. The first is that nationality influences respondents' work injuries in the last two years on board. This is the variable with the strongest contribution. The Norwegian seafarers reported to have been more involved in injuries than the Greek seafarers. The variable with the second strongest contribution is the Risk acceptance/violations index; indicating that the more violations and risk accepting behaviour you are involved in, the more likely it is that you are injured on board. The variable with the third strongest contribution is age group, indicating that controlled for the other variables, the youngest seafarers have a higher risk of being injured on board. In Table 1, the Nagelkerke R^2 is 0.188 which indicates that the independent variables explain 19% of the variance in the dependent variable.

3.1.2 *Which factors influence safety behaviours?*

In Table 2 we show results from a hierarchical, linear regression analysis, where independent variables are included in successive steps to examine the variables predicting respondents' scores on the Risk acceptance/violations index.

Table 2 provides five main results: first, the more demanding working conditions that the respondents experience, the more likely they are to be involved in Risk acceptance/violations. Second, we see that the national safety culture—descriptive norms index contributes positively, indicating that the more unsafe behaviours the respondents say that they expect from seafarers from their own country, the more likely they are to be involved in unsafe behaviours themselves. Third, the higher organizational safety culture scores the respondents report, the less unsafe are their behaviours.

Table 1. Logistic regression. Dependent variable: Personal injuries on board in the last two years (dichotomized: 0: no personal injury, 1 = personal injury). B values.

Variables	B value
Age group (26 years = 0, Other = 1)	0.373**
Nationality (Greek = 0, Norwegian = 1)	2.226**
Vessel type (Live fish carrier = 0, Other = 1)	0.888
Position/line of work (Deck crew = 0, Other = 1)	0.657
Risk acceptance/violations index	1.164***
Working under the influence of alcohol/ hungover	0.304
Non-reporting/non-intervention index	0.940
Sometimes I feel pressured to continue working even if it is not perfectly safe	1.224
Organisational safety culture index	1.025
Nagelkerke R^2	0.188

*P < 0.1, **p < 0.05, ***p < 0.01.

Table 2. Linear regression. Dependent variable: "Risk acceptance/violations Index". Standardized beta coefficients.

Variables	Beta coeff.
Age group (26 = 2)	0.003
Nationality (Greek = 2)	–0.030
Position (Apprentice = 2)	0.052
Vessel type (Tank = 2)	–0.031
Sometimes I feel pressured to continue working, even if it is not perfectly safe	0.167**
Demanding working conditions index	0.281**
Organisational safety culture index	–0.195**
Sector focus on safety	–0.144**
National safety culture: descriptive norms	0.206**
Adjusted R^2	0.453

*P < 0.1, **p < 0.05, ***p < 0.01.

Thus, a positive organisational safety culture may reduce the negative contribution of demanding working conditions and safety compromising work pressure. The same applies to the index "sector focus on safety". In Table 2, Adjusted R^2 is 0.453 which indicates that the independent variables explain about 45% of the variance in the dependent variable.

3.2 *Private boat users*

3.2.1 *Which behaviours influences personal injuries?*

A logistic regression analysis was conducted with boating incidents (grounding, collision, intake of water) as dependent variable, to find the variables predicting personal injury among our respondents (Table 3). Seven percent of the respondents had experienced this. The incident variable, was dichotomized, 1 = no personal injury, 0 = personal injury. B values are presented and they indicate whether the incident risk is reduced (negative B values) or increased (positive B values), when the independent variables increase by one value.

Table 3 provides two main results. The first is that nationality influences respondents' experiences with boating incidents in the last two years. This is the variable with the strongest contribution. The effect is negative, meaning that Greek boat users are involved in fewer incidents, controlled for the other relevant variables.

The second result is that alcohol use during trips as a boat driver contributes significantly and negatively, meaning that increased alcohol use increases the likelihood of boating incidents. In Table 3 the Nagelkerke R^2 is 0.222 which indicates that the independent variables explain 22% of the variance in the dependent variable.

3.2.2 *Which factors influence safety behaviours?*

In Table 4 we show results from a hierarchical, linear regression analysis, where independent variables are included in successive steps to examine the variables predicting respondents' scores on the variable "Alcohol use during boat trip as driver".

Table 4 provides four main results: first, national safety culture, specified as descriptive norms (boat users from your own country's alcohol use, over speeding close to shore and offensive driving) provides the strongest contribution to respondents' alcohol use while driving a boat. Respondents who report of unsafe behaviours among boat users in their country are more likely to drink alcohol while boating themselves. We made similar indexes for the peer group and the municipality level. The peer group level refers to "friends who own a boat".

Table 3. Logistic regression. Dependent variable: boating incidents in the last two years (dichotomized: 0: incident, 1 = no incident). B values.

Variables	B value
Age group (46–55 years = 0, Other = 1)	0.711
Exposure	0.002
Boat type (motor boat w/sleeping facilities = 0, other = 1)	0.328
Alcohol use during trip as boat driver	–0.359**
Nationality (Greek = 0, Norwegian = 1)	–1.683***
Education/training in boat use (No = 1)	–0.179
Navigational equipment on board	0.826
Nagelkerke R^2	0.222

*P < 0.1, **p < 0.05, ***p < 0.01.

Table 4. Linear regression. Dependent variable: "Alcohol use during boat trip as driver". Standardized beta coefficients.

Variables	Beta coeff.
Age group (Under 56 years = 1, over = 2)	–0.147**
Nationality (Norwegian = 1, Greek = 2)	–0.160**
Boat type (Other = 1, motor boat w/ sleep = 2)	0.171***
Purpose of trip (Other = 1, Leisure = 2)	0.119*
Perceived enforcement: police/coast guard	0.028
Peer group safety culture	0.151*
Municipal safety culture	–0.157*
National safety culture	0.218***
Adjusted R^2	0.167

*P < 0.1, **p < 0.05, ***p < 0.01.

We see that that peer group safety culture and municipality safety culture only contributes significantly at the 10% level. The contribution of peer group safety culture is positive, as the national culture variable, but the municipality contribution is negative. This is unexpected, and we return to it in the discussion section.

Second, we see that boat type (motor boat with sleeping facilities) contributes positively, indicating that using this boat type involves a higher incident risk in our sample. Third and fourth, we see that age (>56 years) and nationality (Greek) gives a lower risk of having experienced incidents.

Finally, we also see that purpose (i.e. leisure and holiday) contributes positively to incidents, but only at the 10% level. Thus, we see, not unexpectedly, that compared with other purposes (e.g. fishing, transport), boat drivers on leisure/trips are more likely to drink alcohol while driving. In Table 4, Adjusted R^2 is 0.167 which indicates that the independent variables explain about 17% of the variance in the dependent variable.

4 CONCLUDING DISCUSSION

The aims of the study were to examine the safety behaviours related to personal injuries and accidents among professional seafarers and private leisure boat users in Norway and Greece, and to study the factors influencing these behaviours.

4.1 *Factors predicting injuries/accidents*

Looking at the factors predicting injuries/accidents in the two groups, we saw that nationality (Norwegian), risk acceptance/violations and age group (<26 years) predicted professional seafarers' work injuries in the last two years on board. The contribution of age and nationality is in accordance with previous research on professional seafarers (Hansen et al 2002; Jensen et al 2004).

Looking at the private boat users, we also saw that nationality influenced respondents' risk of boating incidents in the last two years. Second, we found that alcohol use during trip as a boat driver increased the likelihood of boating incidents. Working under the influence did not contribute significantly to professional seafarers' risk of work accident.

This contrasting result is in line with the hypothesis we mentioned in the introduction; that private boat use seems to be a relatively unregulated behaviour compared with professional seafaring. Previous research indicates that alcohol consumption may be an important risk factor in the maritime sector (Akhtar & Bouwer Utne 2014, Hetherington et al 2014), and that alcohol and drug abuse are greater for seafarers compared to workers ashore (Nitka 1990; Kariris 2012 in Zhang & Zhao 2017), partly because of their working situation (e.g. social isolation). However, as private boat use is less regulated than professional boat use, we hypothesized that alcohol consumption "boating while under the influence", would be an even more important risk factor among private boat users. Results indicate that this is the case, at least based on our sample.

As noted, Amundsen (2016) also asserts that questions about alcohol use and lifejacket use are common in almost all of the international surveys, indicating the importance of these factors for boating safety. Moreover, Norwegian boating accident statistics report a number of death involving alcohol, and where the drowned person did not were a life jacket (Amundsen & Bjørnskau 2017).

4.2 *Factors predicting safety behaviours*

Analysing the factors influencing professional seafarers' risk acceptance and violations, we found that demanding working conditions and work pressure were important factors. This is in line with previous research (Størkersen et al 2011, Nævestad 2017). The former was the most important factor. We also found that a positive organisational safety culture may reduce the negative contribution of demanding working conditions and safety compromising work pressure. This has also been pointed out in a previous study (Nævestad 2017). We also found that "sector focus on safety" may reduce the negative influence of demanding working conditions on professional seafarers' safety behaviours.

Additionally, we, found that the national safety culture—descriptive norms index contributed

positively, indicating that the more unsafe behaviours the respondents say that they expect from seafarers from their own country, the more likely they are to be involved in unsafe behaviours themselves. We found the same in the analysis of the private boat users; in fact, this analysis showed that, national safety culture, specified as descriptive norms (boat users from your own country's alcohol use, over speeding close to shore and offensive driving) provided the strongest contribution to respondents' alcohol use while driving a boat. To our knowledge, there are few other studies that have examined the influence of national culture (specified as descriptive norms) on both professional seafarers and private boat users.

Examining the cultural influences on private boat users' safety behaviours, we also found that peer group and municipality safety culture contributed significantly at the 10% level. The contribution of peer group safety culture is positive, as the national culture variable, but the municipality contribution is negative. This is likely to be a result of a collinearity effect, indicating that these two variables are strongly related and measure "the same effect". In practice (and based on observing the means and the standard deviations on these two variables) it seems that respondents do not separate clearly between boat users in their own municipality and their peer group. This is understandable, given the memory, knowledge and analytical separation required to do this. Thus, we should exclude the municipality level from the analysis, as it is likely that this is the level (compare to peer group) that respondents know less about.

Analysing, the influences on private boat user behavior, we also found that boat type (motor boat with sleeping facilities) involves a higher incident risk in our sample. We also found that age (>56 years) and nationality (Greek) gives a lower risk of having experienced incidents. We also found that purpose (i.e. leisure and holiday) contributed positively to incidents, but only at the 10% level. This brings us to the important differences between the two groups that we study. Finally, previous research on private boat users has also found such background variables to be important for safety behaviours, e.g. type of boat used, gender, age, experience, what kind of activity they usually use the boat for (e.g. fishing, competition, holiday, recreation), type of location where they usually use the boat (cf., Amundsen 2016; Amundsen & Bjørnskau 2017).

4.3 *Why are the number of fatalities higher for leisure boat users than professional seafarers?*

An overarching purpose of our study was to discuss possible reasons to the higher number of fatalities for leisure boat users than professional seafarers. We wanted for instance to examine the kind of behaviours that are related to injuries/accidents in the two groups, and subsequently to examine the factors influencing these behaviours. We may of course only speculate based on our study, indicating hypotheses that should be examined further in future research, but our study indicates that the settings and purposes are important to understand this difference.

While unsafe behaviours related to work pressure and risk taking are important among professional seafarers (i.e. risk acceptance and violations), unsafe behaviours related to the leisure/holiday situation was important for the leisure boat users (i.e. alcohol use while driving a boat).

Additionally, it seems that the situation of private leisure boat users is less regulated than that of professional seafarers. The International Safety Management (ISM) code of the International Maritime Organisation requires shipping companies to implement Safety Management Systems (SMS) on board their vessels, including describing safety roles, goals, procedures, monitoring, reporting, follow up etc. (Thomas 2012). Additionally, shipping companies also often work to implement a positive organizational safety culture, including policies for seafarer behavior. Also, professional seafarers have undergone an IMO approved training in their respective home countries. Thus, this training, the SMS and safety culture are elements which are likely to influence the professional maritime safety culture.

Private boat users, on the other hand are not part of such a system of international and national regulation, involving education, inspections from port states, flag states, classification societies, transport buyers etc. Compared to the number of people who go boating in different countries, the risk of accidental death is quite high compared to that of other private transport modes. Despite of this, recreational boating is to a small extent being regulated and the level of enforcement is low (Amundsen & Bjørnskau 2017). Some countries seem to take safety for leisure boat users more seriously than others. In a few countries it is, for instance, now mandatory to report all incident you experience while boating, even if no persons were injured in the incident/accident. Finally, it is important to remember that the above mentioned view points merely are hypothesis that must be examined in future research.

4.4 *Cultural influences on maritime safety behaviours*

The theoretical link between safety culture and safety behaviours is often omitted in research

(Ward et al 2010). In the present study, we conceptualise this relationship as descriptive norms, that may influence behaviour by providing information about what is normal. In the professional (organisational) setting, managers are an important source of social pressure, as well as colleagues, and the interaction between people within the organisation is important for the creation and maintenance of a safety culture influencing behavior, as indicated by the effect of organizational safety culture on professional seafarers' safety behaviors in Table 2.

In the private setting, there will not be a similar strong link from managers to transport safety culture. Some peers are, however, likely to assert stronger social influence than others, and may be as important as managers in organizations in exerting social pressures that shape safety culture and influence behaviour. In our study (Table 4), we saw that peers are central as advocates of social norms related to safety (i.e. drinking alcohol while boating), but we also saw that the reference to other people in the boat users' country were even more important. In Table 2 we also saw the importance of sector for professional seafarer behavior.

To conclude, our study indicates that both in the professional and the private setting, norms for interaction and conduct seem to be influenced by norms and expectations rooted in different socio-cultural groups, e.g. the national culture, the specific sector in question, the organisations and in peer groups.

ACKNOWLEDGEMENTS

This research was funded by the Norwegian Research Council's Transport 2025 program.

REFERENCES

Akthar M.J. & I. Bouwer Utne (2014) Human fatigue's effect on the risk of maritime groundings—A Bayesian Network modeling approach, *Safety Science*, vol. 62. pp. 427–440.

Antonsen, S. (2009). The relationship between culture and safety on offshore supply vessels, Safety Science, Vol. 47. Issue 8, pp. 1118–1128.

Bjørnskau, T., F. Longva, Sikkerhetskultur i transport. TØI rapport 1012/2009. 2009: Transportøkonomisk institutt.

Cialdini, R.B., R.R. Reno and C.A. Kallgre (1990) A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places, *Journal of Personality and Social Psychology,* 58, pp. 1015–1026.

EMSA (2017). Annual overview of marine casualties and incidents 2017. European Maritime Safety Agency.

GAIN (Global Aviation Network) (2001). Operator's Flight Safety Handbook.

Hale, A. (2000). Editorial: Culture's Confusions, *Safety Science,* vol. 34, pp. 1–14.

Hansen, H.L et al. (2002). Occupational accidents aboard merchant ships, Occup Environ Med, 59, pp. 85–91.

Håvold, J.I. (2005). Safety-culture in a Norwegian shipping company. Journal of Safety Research, 36, pp. 441–458.

Hetherington, C., Flin, R., Mearns, K. (2006). Safety in shipping: The human element. *Journal of Safety Research*, 37(4), pp. 401–411.

Jensen, O.C. et al. (2004). Incidence of self-reported occupational injuries in seafaring—an international study, Occupational Medicine, doi:10.1093/occmed/kqh090.

Nævestad T-O. and T. Bjørnskau (2014), Kartlegging av sikkerhetskultur i tre godstransportbedrifter. TØI rapport 1300/2014. 2014: Transportøkonomisk institutt.

Nævestad, T.-O. (2017) "Safety culture, working conditions and personal injuries in Norwegian maritime transport", *Marine Policy,* Vol. 84. pp. 251–262.

Nævestad, T.O. et al. (2015). Work-related accidents in road sea and air transport: prevalence and risk factors, TØI report 1428/2015, Oslo: Transportøkonomisk institutt.

Nitka, J. (1990) Selected medical and social factors and alcohol drinking in Polish seafarers, Bull. Inst. Marit. Trop. Med. Gdynia 41 (1990) 53–57.

Oldenburg, M. & H.J. Jensen (2012). Merchant seafaring: A changing and hazardous occupation, Occupational and environmental medicine 69(9):685–8 · June 2012.

Rivis, A., P. Sheeran (2003). Descriptive norms as an additional predictor in the theory of planned behaviour: A meta-analysis. Current Psychology: Developmental, Learning, Personality, Social, 22, p. 218–233.

Størkersen K.V. (2017). Coastal cargo work: How can safety shout instead of whisper when money talks? Paper to be presented at the 2017 ESREL conference.

Størkersen, K.V.et al. (2011). Sikkerhet i fraktefarten. Analyse av drifts- og arbeidsmessige forhold på fraktefartøy, NTNU Samfunnsforskning AS, Studio Apertura, Trondheim: NTNU.

Ward, N.J. et al.(2010). White Paper on Traffic Safety Culture. White Papers for "Toward zero deaths: a national strategy for highway safety" Series—White Paper No.2, Montana State University.

Zhang, P., Zhao, M. (2017) Maritime health of Chinese seafarers (2017) *Marine Policy*, Vol. 83, pp. 259–267.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Accident and disease prevention in working life: Common grounds and areas for mutual learning

E. Albrechtsen, R.B. Jørgensen, T.Ø. Kongsvik & K.V.H. Svendsen
Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT: Globally, there are more than six times more fatalities caused by a poor working environment than due to occupational accidents. In this paper we compare the basic strategies involved in accident and disease prevention. We find that the basic thinking is the same. The preventive strategies involve control of hazards in a hierarchy from elimination to application of personal protective equipment. Also, the Norwegian regulation on internal control of HSE is applicable for both accident and disease prevention, involving the idea of continuous improvement. Still, the nature of the hazards differ, as well as the possible consequences. In the area of occupational safety, the hazards are usually visible and the consequences of accidents are immediate. In the area of occupational health, the hazards are often invisible and the consequences of exposure delayed. There is a potential for better integration of the two areas in practical management, and a potential for mutual learning from concepts and models in the two fields.

1 INTRODUCTION

Protection from harm in relation to work is a responsibility for politicians, authorities and employers. Still, work related diseases, accidents, injuries and fatalities continues to be a significant challenge.

According to ILO, globally 2.3 million deaths take place due to occupational injuries and diseases each year (Takala et al., 2014). Of these, over 2 millions of deaths are due to occupational diseases, among them 23% due to work related cancers.

In Norway, we have from 47 to 25 deaths each year from injuries in the period 2013–2016 (The Norwegian Labour Inspection Authority 2017), and 1% of the working population reported work related lung complaints in 2013 (NOA, 2017). In addition, it has been estimated that approximately 3000 new cases of work related Chronic Obstructive Pulmonary Disease (COPD) each year in Norway, and that 200 persons will die from this disease due to working conditions (Leira, 2011). Although these figures are high, there seems to be a positive trend both in work related accidents and diseases. The number of fatalities was more than 100 annually in the early 1970s, but has steadily declined. In 2016 the number of fatalities was 25 (The Norwegian Labour Inspection Authority 2017). The number of self-reported exposures for chemicals are declining, as well as reports of other work-related diseases (NOA, 2011, 2017)

Accidents in the construction industry still get high attention. However, some highlights the fact that the workers in this sector have increased risk for lung diseases (Bergdahl et al., 2004; Robinson, Petersen, Sieber, Palu, & Halperin, 1996; Vermeulen, Heederik, Kromhout, & Smit, 2002). Also in Norway the statistics from NOA shows that workers in the construction industry have more respiratory complaints, and more declared work related respiratory complaints than the national mean (Aagestad, 2015). Recent studies have also detected an association between exposure and an increased risk of COPD in the construction industry (Fell, Aasen, & Kongerud, 2014). In addition 59% of construction workers state that that they inhale smoke, dust or exhaust in their work situation (NOA, 2017).

The purpose of this paper is to compare the main strategies for the prevention of occupational diseases and occupational accidents.

2 PREVENTION OF OCCUPATIONAL DISEASES

2.1 *Exposure assessment*

The traditional approach for chemical risk assessment is to compare the exposure level of the chemical agent to their Occupational Exposure Limits (OELs). These OELs were established from late 19 century (Jayjock MA, 2000). Still this approach is regarded to be the best practice for risk assessment for chemical agents and noise. These limit values are given for occupational exposures during 8 hours shift, and there is guidelines on how

to assess risk of exposure to noxious agents (NS-ISO689) and sound (NS-4815–1).

The sampling and analyses of airborne contaminants and comparing results with the national OELs have been challenging for different reasons. Due to variability in exposure both within workers and between workers, several samples have to be taken for each group of workers regarded as homogenously exposed (Chen, Chuang, Wu, & Chan, 2009; Rappaport, Lyles, & Kupper, 1995). The cost for each analysis and the resources needed to perform the measurements have resulted in a limited number of measurements from the different parts of the Norwegian industry. Reported measurement results from the construction sector are rather sparse, except some from tunnel construction, cement work and Bricklayeres (Bakke, Ulvestad, Thomassen, Woldbæk, & Ellingsen, 2014; Beaudry et al., 2013).

A new approach to risk analyses on health effects have been widely used since the introduction of REACH (Money, 2003; Office, 2017; UK, 2017). This new approach take into consideration the health classification of chemicals or particles are use this together with an exposure assessment (Money, 2003). The classification of health effects are performed on the basis of the agent's inherent toxic property, often by use of the CLP classification (CLP-ref) but the risk assessment are also dependent on the exposure. This exposure assessment may be founded on subjective assessment and exposure toolkits (Office, 2017; UK, 2017). The subjective assessment method uses a structured approach, based on descriptive information about work activities and the work environment, and have been validated against exposure measurements (Cherrie & Schneider, 1999). This new approach makes it possible to do risk assessment of chemicals and particles which do not have an OEL and without measurements. However, when the risk analyses shows that there may be an unwanted risk; measures have to be taken to comply to the model, or measurements have to be performed in the old fashion way to document that there is an acceptable risk.

2.2 *Hierarchy of exposure controls*

Within occupational hygiene, control of exposure is a fundamental method for protection of workers. Traditionally ahierarchy of controls has been used as a mean of determining how to implement feasible and effective control solutions. The hierarchy is often illustrated as in Figure 1, where elimination is the most effective measure, and include physically removal of the hazard. Substitution is replacement of the hazard, engineering controls isolate people from the hazard, administrative

Figure 1. Control of exposure.

controls change the way people work and Personal Protective Equipment (PPE) is the weakest measure protecting the worker against the hazard.

The principle behind this hierarchy is that the control methods at the top of graphic are potentially more effective and protective than those at the bottom.

In practical use, however, it is often more difficult. Using a smelting plant as example, the production of metal or metal alloy like aluminium, silicon, silicon carbide or ferromanganese is the basic idea behind a production plant. The top two measures, elimination and substitution of the raw material is impossible; and the occupational hygienist only have the three lowest ranked measures available. The smelting industry is very energy-intensive and access to cheap energy often determines where the factories are located. The processes are often old and physically relatively simple, but produce large amounts of pollutants. In the silicon carbide industry the cancer risk have been documented since 2000 (Romundstad, Andersen, & Haldorsen, 2001), and the dust exposure is documented to contain both fibers (Bye, Eduard, Gjonnes, & Sorbroden, 1985) crystalline silica, silicon carbide (SiC) and sulphur dioxide (S. Føreland, Bye, Bakke, & Eduard, 2008). The workers in this industry is heavily loaded by personal protective equipment using dust mask, CO alarm, eye protection, hearing protection, safety helmet, gloves and safety clothes. The Silicon carbide industry is not the only industry using this as main principle, even that isolation of the process; local exhaust ventilation or general ventilation would have been more effective types of measures.

This was pointed out by Føreland (Solveig Føreland, Bakke, Vermeulen, Bye, & Eduard, 2013) as late as 2012stating that "recommendations for exposure reduction based on this study are (i) to separate the sorting area from the furnace hall,

(ii) minimize manual work on furnaces and in the sorting process, (iii) use remote controlled sanders/grinders with ventilated cabins, (iv) use closed systems for filling pallet boxes, and (v) improve cleaning procedures by using methods that minimize dust generation".

Following the hierarchy normally leads to the implementation of inherently safer systems, where the risk of illness or injury has been substantially reduced, but as the example shows, it is not always possible to use the most favorable measure.

The same kind of example could be used for the construction sector, as the work itself produces pollutants that is impossible to avoid as concrete, wood dust, stone dust and exhaust from vehicles. The only one that may be substituted is the exhaust when new vehicle technology is developed. This makes it necessary to use the last favorable protective measure; the personal protection devices.

3 PREVENTION OF OCCUPATIONAL ACCIDENTS

The barrier concept is a basic foundation in strategies for occupational accident prevention. An accident can be understood as caused by energy out of control (Gibson, 1961), where a hazard (source of energy) releases energy. This energy is then absorbed by a victim which leads to loss (health, life). Following this energy model, accidents can be prevented by barriers that stops or prevents a sequence of events that lead to loss of control of hazards (Kjellén and Albrechtsen, 2017). As shown in Figure 2. the barrier philosophy follows the same logic as control of exposure (Figure 1)

Seen in a functional perspective, safety barriers perform tasks, such as preventing falling objects from hitting people working below. Such functions or tasks are performed by different barrier elements that constitute a barrier system (Rosness et al. 2010). Physical devices, human actions and administrative procedures serve as barrier elements meant to protect vulnerable targets from harm (Sklet, 2006).

A 'defence in depth' strategy is commonly applied in the prevention of accidents. According to Reason (1997:12), major accidents occur as a result of failures in multiple layers of the defences separating potential hazards from people and assets. Accident trajectories pass through 'holes' in these defences, created by active failures—errors and violations—and/or latent conditions, such as lack of competence, design flaws and unrealistic procedures.

For example for a lifting operation, two barrier functions must be in place: 1) prevent sudden release of the gravity energy that the lift represent and 2) separate the gravity energy that the lift represent and workers (establish a safety zone). The first barrier function (prevent sudden release of energy) is realized by a set of barrier elements: sling mechanism, crane driver competence, slinger competence, signal man, control rope etc.

Haddon's (1980) defined ten generic principles for the prevention of harm (injury) from transfer of energy. These ten strategies are generic barrier functions to either control the hazard; separate the hazard and a victim; or make the offer more robust to harm. See Figure 3.

Haddon's (1980) strategies and the energy accident model (Gibson, 1961) have had significant influence on European legislation and standardisation work such as that related to hazardous chemicals (European Council 1998) and machinery safety (European Council 2006). The strategies are central components in accident investigation methods such as in the OARU, Management Oversight and Risk Tree MORT, Safety Management and Organisation Review Technique (SMORT) (Kjellén and Albrechtsen, 2017).

Figure 2. Control of hazards.

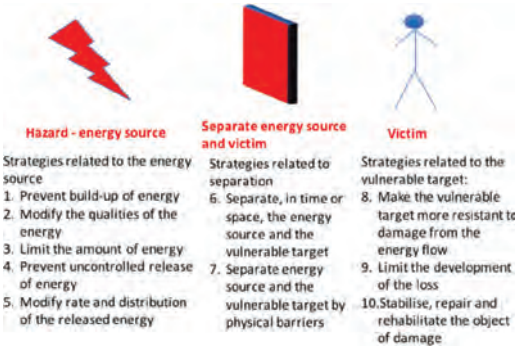


Figure 3. Haddon's principles for accident prevention.

Risk mitigation in the European Council (2006) Directives on Machinery is based on Haddon's strategies. ISO 12100:2010shows a strategy for selection of safety measures of machinery, see Figure 4. The strategy reflects that hazards should be prevented or limited at the design stage, i.e. designed out. If this is not possible, protective measures and safety controls should be established to separate victims and the hazards of the machinery. Residual risk after these measures can be accepted, but requires that the producer inform the user about these. The user of the machinery is responsible for training, and safe operation at work, including providing necessary personal protective equipment.

To implement the correct barrier system (functions and elements) risk assessments are essential. The results of risk assessment serve as decision-making support to implement adequate safety measures. ISO31000 Risk Management gives descriptions of the principles and steps in risk assessment and risk handling. Briefly, the steps are: identify hazards and incident scenarios; analyze causes; analyze frequency and consequences; analyze risk; evaluate risk according to risk acceptance criteria; and mitigate risk. The analysis of risk is made by systematically collect available knowledge about the analysis object and use this knowledge to express what can go wrong in the future; what the likelihood of it happening; and the consequences if it happens (Rausand, 2011). This risk picture is then evaluated to risk acceptance criteria to detriment whether the risk is acceptable or not.

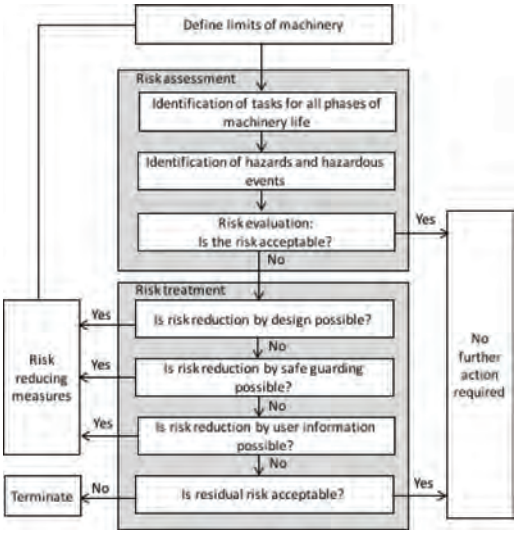


Figure 4. Risk treatment and risk assessment of machinery (based on ISO12100).

The Directives of Machinery shows how risk assessment is the input to risk mitigation, see Figure 3. ISO 12100:2010show the steps for risk assessment of machinery. It follow the same steps for risk assessment as described in ISO31000, but starts with an identification of the machinery and intended/unintended use of the machinery for all life phases of the machinery.

3.1 *Tunnel construction as example*

An analysis of fatal accident from hydropower instruction projects in Kjellén and Albrechtsen (2017) show that there are two types of fatal accidents in tunnel excavation: falling rock/rock burst and workers squeezed or driven over by vehicle. Barriers to prevent such fatal accidents would mainly be to separate in time/space the hazard and the victim. For blasting work workers are moved away from the blasting area.

During construction work other than blowing work, accident risks are in particular conflicts between workers and heavy machinery, falling objects (rock fall/rock burst) and getting squeezed during rigging.

One of the work activities in tunnel excavation are assembly of different materials and infrastructures in the roof. The accident risk for such operations are falling objects and squeezing. In Norway, there has been fatal accidents at scissor lifts, where the victim has been squeezed between the lift and the roof. Risk mitigation for this scenario is to design in pressure-released stop of the lifting mechanisms.

Work at height will also involve significant occupational hygiene risk. From an occupational hygiene perspective it is well known that tunnel workers are exposure to both particulate and gaseous air pollution and that tunnel workers are known to be at increased risk of long-term and short–term lung function decline and CORP (Ulvestad et al.). Different activities causes different exposures to these workers. Exposure to gases and particles from diesel emissions has been considered to be among the dominating burdens during tunnel construction due to weel-going diesel machines; also during drilling and blasting operations, workers are exposed to dust, with α-quartz as the most important agent. α-quartz in the dust from tunnels varies between <1% and more than 50% (Norwegian Tunnelling Society,Publication No.13) and α-quartz exposure may lead to COPD. Exposure to oil mist and oil vapour is another type of exposure that may also occur during drilling. Exposure to oil mist may cause occupational asthma and also pulmonary fibrosis (Robertson et al., 1988) Risk assessment hence have to be performed with focus on all these possible exposures.

If we look into Figure 1 in order to identify preventive measures it seems logic that elimination and substitution is difficult. None of these activities could be eliminated if the tunnel should be constructed. Substitution might be possible if we look into the type of explosive used for blasting. Ammonium nitrate fuel oil is used as explosive, and if this agent is substituted with size-sensitised emulsion, the worker exposure lower (Ulvestad). Apart from this example engineering controls often is the first possible preventive measure, where ventilation of the tunnel, pollution-abatement equipment for diesel vehicles is some of the measures often used. High frequency maintenance routines are an example of administrative controls, while personal protection equipment only should be the last solution, but in real life often is used on a daily basis.

Figure 5. Deming's circle.

4 DISCUSSION

4.1 *Common grounds*

The law/regulation on internal control makes it clear that all enterprises in Norway should have a system for safeguarding health, safety and environment. This system includes information on health and safety regarding issues in the working environment, requirements for establishing goals for the HSE-work, performing risk analyses for any hazard and establish routines for unveiling, correcting and preventing violations of the law and regulations.

The internal control regulation that is the source for management systems for HSE control is the same for both safety and workers health, and builds on quality principles. Theories on quality gained much attention in industry from the 1980s, and was first related to improvements in the production processes. Total quality management became an influential movement, spurred by the works of Deming (1986) and Juran. Later, the principles were used as a foundation for internal control for HSE in Norway (Saksvik & Nytrø, 1996).

The idea of continuous improvement is evident in both the prevention of occupational diseases and in the prevention of occupational accidents, illustrated in Deming's circle (Figure 5):

In the prevention of occupational diseases related to the exposure to chemical, quality principles are at work when exposure levels are compared to OELs, and when measures are taken to mitigate or eliminate the exposure, illustrated in the hierarchy of controls (Figure 1).

In the prevention of occupational accidents, the idea of continuous improvement is the foundation for safety information systems, and the experience feedback such systems entail. By means of safety indicators, safety audits, and accident investigations, information is applied to implement measures for accident prevention.

The barrier philosophy of both areas is the same. Both aim at first prioritizing efforts directed at the source of danger by elimination, modification and limitation. Second, both areas emphasize to avoid interface between risks and victims by substition and separation. Third, both areas emphasize engineering control by built-in solutions in design. Finally, the last measure in both areas is to approach the victim by information, training, procedures and lastly personal protective equipment.

Many of the occupational hygiene risk factors can contribute to higher accident risk by influencing human performance. The human operator is an essential barrier element to realize barrier functions. Stress (fatigue, time load and task load); situation/environment (physical and chemical work environment); and human-machine interactions are among performance shaping factors (Groth and Mosleh, 2012) that affect the quality of the human barrier element in accident prevention.

4.2 *Differences in the nature of hazards and consequences*

One obvious difference between the health and safety field, is the nature of the hazards that should be handled. Hazards in the field of safety are a form of energy that is not properly controlled. Further, hazards may cause immediate harm if safety barriers are not in place, or if they are not functioning as intended.

In the area of occupational health, hazards are not limited to energy sources, although vibration, radiation and noise is clearly within the energy perspective. But also toxic fumes, cancer inducing and poisonous agents represent hazards, not directly related to energy and more or less invisible in nature.

Further, hazards within the occupational health domain do in many instances not cause immediate harm, but the harm may be delayed. In some instances it may take several decades from exposure to the loss is evident. Although there are nuances in this, this can be summed up in a general manner as in Table 1:

Even if more than six times more occupational deaths are caused by diseases than accidents, there is a tendency that accidents get more attention. Accidents are concrete, often dramatic events, and with immediate consequences. This will naturally generate attention from employers, authorities and the general public, often followed by demands for measures that ensures that similar events will not take place in the future.

Occupational diseases are less dramatic, and as the consequences are often delayed, there will also be an issue of employer responsibility for the harm. The employee might have changed employer several times before the disease is evident. Diffused responsibility, coupled with the more invisible nature of the hazards, and the latency period from exposure to disease, might result in less attention from outside actors.

In the end it might also lead to less resources to the prevention of occupational diseases, relative to the needs. The regulatory authorities should be aware of such mechanisms, and implement requirements to ensure that the prevention of occupational diseases get proper attention and the resources.

4.3 *Cross-disciplinary learning?*

Although some of the preventative strategies related to accidents and diseases seems to be the same, the professional concepts that are applied differ. For example, control of exposure vs. control of hazards refers to the same line of thinking. The professional concepts in the two fields have developed over a long period, and are institutionalized and applied in research and theory building. The distinct repertoires of concepts ensures precision and are a foundation for theory development within the two fields. Thus, the development of a common professional language seems unrealistic and also undesirable.

Table 1. Differences in the nature of hazards and consequence in occupational safety and occupational health.

	Hazards	Consequences
Occupational safety	Visible	Immediate
Occupational health	Invisible	Delayed

Still, mutual awareness of the concepts, models and theories that have been developed in the respective fields can represent a fruitful cross-pollination, and instigate new ideas for prevention. For example, within the safety field, there exists many accidents models; domino models (Heinrich, 1931), information models (Turner & Pidgeon, 1997), and the swiss cheese model (Reason, 1997) to name a few. There are also perspectives related to what kind of organizational characteristics that may prevent accidents, including the theory of high-reliability organizations (Weick, Sutcliffe, & Obstfeld, 1999) and Resilience Engineering (Hollnagel, 2014). Researchers with occupational health might find such models to be inspiring and of relevance.

A basic concept related to the prevention of occupational diseases is OELs. Actual exposure levels of chemical agents are compared to OELs to determine whether the exposure might induce harm. Much research lies behind the definitions of OELs. This might be an interesting area for occupational safety. Although the hazards are of a different nature, setting clear thresholds for exposure might be an issue for further exploration, even if there exist similar lines of thinking (e.g. the notion of acceptance criteria, and the ALARP principle, As Low as Reasonably Practicable).

In many instances, HSE practitioners working within companies (e.g. HSE engineers, managers etc.) will have responsibilities related to both health and safety. From their professional background, they will have insight into the different research fields, and be in a position to integrate them, and treat HSE as a holistic concept. Academics tend to be more specialized within one of the fields of research. Thus, for researchers interested in cross-disciplinary learning, HSE practitioners might be a resource for practical knowledge integration. It is also a responsibility for academics and educators to prepare prospective HSE workers for the cross-diciplinary challenges they will meet in working life. Real working life problems do not necessarily follow professional boundaries, but require a cross-disciplinary approach. Thus, cross-disciplinary learning in the areas of occupational health and safety should be an important issue in university education.

5 CONCLUSIONS

The paper demonstrates that there are similarities between management of occupational health and

safety, but also that there are potential improvements for better integration of the two areas in practical management.

As indicated in the introduction of the paper there are far more death and personnel harm due to poor working environment than there are deaths and injuries due to occupational accident. However, accidents and accident prevention seems to get more attention by HSE practitioners and mass media. Possible explanation for this picture could be the different nature of the hazards and the lagging consequences of occupational exposure compared to the sudden consequences of an accident.

Common for both prevention of both occupational disease and occupational injury is that adequate planning would prevent many events by establishing adequate barriers

Topics for further research of the interaction between occupational health and safety can include:

How is HSE implemented in practice in the building and construction industry? Is safety more focused than work environment? And if any difference could be identified, what is the explanation behind this. Some hypotheses could be investigated: A) Simple measurable parameters within security, such as accidents per 1000 hours of work or absence per 1000 hours make safety easier to control. B) Control of exposure is not defined as a project-specific activity and hence an activity belonging to the internal control system of the sub contractors, which in practice means greater variation in how much focus it gets in practical work. C) Differences between safety culture and work environment culture exist.

REFERENCES

Aagestad, C., Tynes, T., Johannesen, H., Gravseth, H.M., Løvseth, E., Alfonso, J.H., Aasnæss, S., Sterud, T. 2015. faktabok om arbeidsmiljø og helse 2015. Oslo.

Bakke, B., Ulvestad, B., Thomassen, Y., Woldbæk, T., & Ellingsen, D.G. 2014. Characterization of Occupational Exposure to Air Contaminants in Modern Tunnelling Operations. The Annals of Occupational Hygiene, 58(7), 818–829. doi:10.1093/annhyg/meu034

Beaudry, C., Lavoué, J., Sauvé, J.-F., Bégin, D., Senhaji Rhazi, M., Perrault, G., Gérin, M. 2013. Occupational Exposure to Silica in Construction Workers: A Literature-Based Exposure Database. Journal of Occupational and Environmental Hygiene, 10(2), 71–77. doi:10.1080/15459624.2012.747399.

Bergdahl, I.A., Toren, K., Eriksson, K., Hedlund, U., Nilsson, T., Flodin, R., & Jarvholm, B. 2004. Increased mortality in COPD among construction workers exposed to inorganic dust. Eur Respir J, 23(3), 402–406.

Bye, E., Eduard, W., Gjonnes, J., & Sorbroden, E. 1985. Occurrence of airborne silicon carbide fibers during industrial production of silicon carbide. Scand J Work Environ Health, 11(2), 111–115.

Chen, C.-C., Chuang, C.-L., Wu, K.-Y., & Chan, C.-C. 2009. Sampling Strategies for Occupational Exposure Assessment under Generalized Linear Model. The Annals of Occupational Hygiene, 53(5), 509–521.

Cherrie, J.W., & Schneider, T. 1999. Validation of a New Method for Structured Subjective Assessment of Past Concentrations. The Annals of Occupational Hygiene, 43(4), 235–245. doi:10.1093/annhyg/43.4.235

Deming, W.E. 1986. Out of the crisis: quality, productivity and competitive position. Cambridge: Cambridge University Press.

European Council. 1998. The protection of the health and safety of workers from the risks related to chemical agents at work. Council Directive 98/24/EC. Brussels.

European Council. 2006. Machinery. Council Directives 2006/42/EC. Brussels.

Fell, A.K., Aasen, T.O., & Kongerud, J. 2014. [Work-related COPD]. Tidsskr Nor Laegeforen, 134(22), 2158–2163. doi:10.4045/tidsskr.14.0255

Føreland, S., Bakke, B., Vermeulen, R., Bye, E., & Eduard, W. 2013. Determinants of Exposure to Dust and Dust Constituents in the Norwegian Silicon Carbide Industry. Annals of Occupational Hygiene, 57(4), 417–431.

Føreland, S., Bye, E., Bakke, B., & Eduard, W. 2008. Exposure to Fibres, Crystalline Silica, Silicon Carbide and Sulphur Dioxide in the Norwegian Silicon Carbide Industry. The Annals of Occupational Hygiene, 52(5), 317–336. doi:10.1093/annhyg/men029

Gibson, J. 1961. The contribution of experimental psychology to the formulation of the problem of safety. In Behavioral approaches to accident research, eds. H.H. Jacobs, Association for the Aid of Crippled Children, New York.

Groth, K.M. and Mosleh, A. 2012. A data-informed PIF hierarchy for modelbased human reliability analysis. Reliability Engineering and Systems Safety 108:154–174.

Haddon, W. 1980. The basic strategies for reducing damage from hazards of all kinds. Hazard Prevention 16:8–12.

Heinrich, H.W. 1931. Industrial Accident Prevention: a Scientific Approach. New York: McGraw-Hill.

Hollnagel, E. 2014. A tale of two safeties. Nuclear Safety and Simulation, 4, 1–9.

ISO. 2010. Safety of machinery. General principles for design. Risk assessment and risk reduction. International standard ISO 12100:2010. International Organization for Standardization, Geneva.

Jayjock MA, L.J., Nelson DI. 2000. Risk assessment principles for the industrial hygienist. In AIHA (Ed.). Fairfax: AIHA press.

Kjellén, U. & Albrechtsen, E., 2017. Prevention of Accidents and Unwanted Occurences. CRC Press

Leira, H. 2011. Kols-ikke bare hos røykere. Tidsskriftet den norske legeforening, 18(131).

Money, C.D. 2003. European experiences in the development of approaches for the successful control of workplace health risks. Ann Occup Hyg, 47(7), 533–540.

NOA. 2017. Nasjonal overvåking av arbeidsmiljø; Astma og KOLS. Retrieved from http://noa.stami.no/arbeidsmiljoindikatorer/helseutfallarbeidsskader/sykdom/kols/

Norwegian Tunnelling Society,Publication No.13 Available from http://tunnel.no/wp-content/uploads/2014/01/Publication_13.pdf.

Office, I.L. 2017. International Chemical Control Toolkit. SafeWork. Retrieved from http://www.ilo.org/legacy/english/protection/safework/ctrl_banding/toolkit/icct/guide.pdf website:

Rappaport, S.M., Lyles, R.H., & Kupper, L.L. 1995. An exposure-assessment strategy accounting for within— and between—worker sources of variability. The Annals of Occupational Hygiene, 39(4), 469–495.

Rausand, M. 2011. Risk assessment. Theory, methods and application. Wiley, Hoboken, NJ.

Reason, J. 1997. Managing the risks of organizational accidents. Farnham: Ashgate.

Robertson AS, Weir DC, Sherwood Burge P. Occupational asthma due to oil mists. Thorax 1988;43:200–5.

Robinson, C.F., Petersen, M., Sieber, W.K., Palu, S., & Halperin, W.E. 1996. Mortality of Carpenters' Union members employed in the U.S. construction or wood products industries, 1987–1990. Am J Ind Med, 30(6), 674–694.

Romundstad, P., Andersen, A., & Haldorsen, T. 2001. Cancer incidence among workers in the Norwegian silicon carbide industry. Am J Epidemiol, 153(10), 978–986.

Saksvik, P.Ø., & Nytrø, K. 1996. Implementation of internal control (IC) of health, environment and safety (HES) in Norwegian enterprises. Safety Science, 23(1), 53–61.

Sklet, S. 2006. Safety barriers: Definition, classification, and performance. Journal of loss prevention in the process industries 19:494–506.

Takala, J., Hamalainen, P., Saarela, K.L., Yun, L.Y., Manickam, K., Jin, T.W.,. . Lin, G.S. 2014. Global Estimates of the Burden of Injury and Illness at Work in 2012. Journal of Occupational and Environmental Hygiene, 11(5), 326–337. doi:10.1080/15459624.2013.863131

Turner, B.A., & Pidgeon, N.F. 1997. Man-made disasters. Oxford: Butterworth-Heinemann.

UK, H. a. S.E. 2017. Control of Substances hazardous to health. www.hse.gov.uk/pubns/guidance/coshh-technical-basis.pdf.

Ulvestad B, Lund MB, Bakke B, et al. 2015. Short-term lung function decline in tunnel construction workers Occup Environ Med 72:108–113.

Vermeulen, R., Heederik, D., Kromhout, H., & Smit, H.A. 2002. Respiratory symptoms and occupation: a cross-sectional study of the general population. Environ Health, 1(1), 5.

Weick, K.E., Sutcliffe, K.M., & Obstfeld, D. 1999. Organizing for high reliability: Processes of collective mindfulness. In R.S. Sutton & B.M. Staw (Eds.), Research in Organizational Behavior (Vol. 1, pp. 81–123). Stanford: Stanford Jai Press.

Security

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Customs—a vital contributor to safe societies? A study of the Norwegian customs service

L.K. Stene & R. Folgerø
University of Stavanger, Norway

ABSTRACT: Prior to the 2011 Norway terrorist attacks, widely referred to as '22 July', the Norwegian Customs Service (Customs) had alerted the Norwegian Security Police Service to the importation of weapon-related products by a known terrorist. The purchases made by this terrorist initially led officials to place him on the watch list of the Norwegian Security Police Service, but no action was taken based on the information given. In the federal Report of the 22 July Commission, the poor coordination of the Norwegian Security Police with Customs was criticized. In this paper, we address whether Customs, as an important player in preventive safety and security, is actually being regarded as such. From a societal safety and risk-based regulatory perspective, we discuss the challenges of the state in recognizing and supporting the effectiveness of Customs as a vital contributor to a safe society.

1 INTRODUCTION

In today's struggle to ensure safe, secure and resilient societies, we strive to avoid industrial and technological accidents and hazards, man-made disasters and health-threatening hazards and pandemics. We take steps to prepare for natural disasters and climate change as well as to ensure the safety of critical infrastructures and supply chains. In this effort to ensure safe societies, federal customs services are increasingly important players, especially with respect to ensuring border security and fighting international crime and terrorism, as well as addressing the relatively recent European migration challenge.

A safe society must find a way to regulate and control the above-mentioned hazards and threats, which might initially seem to be an impossible task. Safety and security policy frameworks cover many societal levels and sectors. The implementation of such policy frameworks presupposes the interaction and close coordination between politicians, bureaucratic organizations, interest groups and other relevant players, as well as citizens in general. This again depends on the effective exchange of information and communication between the relevant players. This information may include, for example, policy documents, project or research results and field or operational experiences, all of which requires horizontal and vertical coordination. Security and risk management in the areas of border control and criminal networks is also influenced by international policies covering various sectors and operations. As such, many individuals at different levels in a range of organizations must be aware of each other. Are they? In this article, we discuss the extent to which the resources held by the Norwegian Customs Service (Customs) are recognized and utilized in Norway's effort to ensure a safe society. In our study, one of the authors investigated a total of 36 official strategic documents both from Customs and other significant organizations that had been generated over a 17-year period (2000 to 2017).

2 WHAT IS A SAFE SOCIETY?

What do we consider to be a safe society? The term safe society is related to the threats and dangers faced by that society. We can define a safe society as follows: *The ability of a society to uphold important social functions and safeguard its citizens' lives, health, and basic needs under different forms of strain* (Engen et.al. 2016:30).

According to Beck (1992), our perception of risk and the development of modern societies are closely connected. Beck argues that globalization and societal development brings new risks and threats that require new ways of thinking to effectively protect ourselves. The author divides modern development into three phases—pre-modern society, modern society and risk society. These phases or stages are essentially connected to the periodization of social change, namely pre-modernity, simple modernity and reflexive modernity.

In Beck's (1992) perspective, modern society is associated with how old and new risks are organized

and how a society develops its social systems, governance, and politics with respect to risk handling. The risk society developed as an outcome of the industrial society. The hallmarks of industrial and modern societies are technological and economic progress and the development of national democratic institutions. In this phase, the state is governed by regulations that guarantee economic growth and the distribution of wealth, risk, safety and security.

However, Beck (1992) also argued that modern societies' continuous production of more goods to an increasingly wealthy population has drawbacks. These include the side effects of chemical, biological, and/or environmental pollution, as well as complex technological developments that increasingly intervene into the social life of individuals, and which lead to new risks and insecurities. The risks associated with technological development and globalization are complex, are not always obvious or easy to detect and can strike arbitrarily. Individuals cannot protect themselves against these risks, in contrast to the risks associated with the industrial society, which are more often aligned with social class. The risk society is more individualized, but is also still an industrial society, due to its involvement in the creation of the technologies leading to the new risks.

Beck (1992) argues further that the transition from a modern society to a risk society can be described as a transition from the distribution of goods to the distribution of risk, which implies that new risks can strike at random and individually, and they can be global and impossible for national institutions to manage. This random aspect can be debated as, for example, climate change is likely to impact third-world countries harder than westernized countries.

An interesting point in Beck's (1992) description of a risk society is the question of how societies can meet the challenges of living in a world with so many uncontrolled risks. Can we fully depend on existing risk analyses and political strategies to effectively address these new risks?

The hallmark of a safe society is its population's trust in the state governance (Engen et. al. 2016). One way states gain this trust is to develop their abilities, to the extent possible, to appropriately regulate risks (Baldwin et. al. 2012).

1.1 *Societal safety and security*

According to Beck (1992), Kaldor (2007) and Giddens (1999), among others, we have moved from a threat perspective, in which state security was the main focus, to a perception in which the threats to the society have become our essential focus. Today, our vulnerability is more dependent on global events and processes, so customs services have become even more relevant as societal protectors.

Engen et al. (2016) argue that the terms risk, security and safety have politically explosive force. Questions concerning national security, sustainable development, accident prevention and human security are examples of this force and give us some clue to the most important aspects characterizing a safe society.

Also central to the concept of a safe society is the term insecurity. Risk, simply defined, is a product of probabilities and consequences that tell something about the future. Insecurity is a more interesting dimension of risk. Insecurity can sometimes be offset by broadened knowledge, but always involves the absence of future insight, which must be compensated for by political assumptions and estimations when making decisions (Engen et. al. 2016).

Threats and vulnerability have different time perspectives; some occur suddenly, some develop slowly, and all require good and flexible planning, regulation and handling (Rosenthal et.al.2001).

What is most important to protect? Some areas have already been mentioned, and critical infrastructure is certainly an important component. In this paper, however, we focus on the public institutions that safeguard the important functions that ensure the proper functioning and safety of societies. These formal organizations, for example, customs services, are fundamental and essential societal arrangements.

Our 'modern threats' challenge these institutions in that they often do not recognize national borders. Examples include international criminal networks, climate change, large migration populations, epidemics and economic/ financial systems, which all increase the vulnerability of our society.

In the struggle to ensure a safe society, Norway has highlighted the need for key resource organisations to identify and cooperate with each other in appropriate ways by implementing contingency preparedness principles that are applicable to public institutions. These principles, including responsibility, proximity, similarity and coordination (also known as risk management principles), underpin the support and regulation of governance in the crisis management practices adopted in ministries, departments and directorates. These kinds of regulations, regulations in general and especially risk regulation are needed to ensure a safe society. Risk regulation is a challenging task in modern societies, termed "risk societies" by Beck (1992), due to the pace of globalization and change.

2 RISK REGULATION IN A CHANGING SOCIETY

To build a safe society, the state establishes risk management practices through regulations, stand-

ardizations and requirements that stipulate the responsibilities of different actors in the marketplace (Lindøe et al. 2015). Risk management is mainly the balance between the ability to realise what is wanted and avoiding what is unwanted. According to Lindøe et al. (2015), the general economic liberalization characterizing the globalized marketplace has led to a focus on effective production and division of labour, as well as the removal of trade barriers and the development of harmonized regulations.

Risk management governance has been on the OECD agenda for some years. Regulation policy has evolved from a focus on negative effects and costs to a positive regulatory approach as an important premise for added value and democratic governance. Many policy documents, guidelines, reform programmes and procedures have been issued to ensure the use of regulation as a positive policy instrument (OECD 2010, OECD 2011). Risk analysis and risk considerations have also become more integrated in matters of common governance and inspections (Lindøe et. al. 2015).

What *is* risk regulation, why do we regulate and what is "good" regulation? These are vital questions for political debate and discussion regarding regulation to better understand the need for risk regulation in a changing society. Selznick (1985) defined regulation as "*a sustained and focused control exercised by a public agency over activities that are valued by a community.*"

Baldwin et al. (2012) described regulation as a specific set of commands involving a social group, such as Health, Safety and Environment (HSE) rules, a deliberate state influence (like revenue) and all forms of social and economic influence, whether they are state-based or have other sources such as the self-regulating marketplace. The latter is addressed by the theory of 'smart regulation,' which states that regulation may be carried out by a host of bodies other than the state. These might include corporations, self-regulators, voluntary organizations or professional or trade bodies.

Regulation represents a multidisciplinary approach whereby political, financial, juridical and social matters are taken into consideration. Regulation discussions are wide-ranging and can be based on assessments or official and political debates related to, e.g., social protection, future planning, the challenge to ensure human rights and the principle of division of powers, or to counter monopolization and ensure the fair allocation of scarce resources (Baldwin et.al. 2012).

A conglomeration of laws, regulations, standards, guidelines, rules and procedures are taken into consideration in discussions of how to regulate areas of interest. "Red-light" and "green-light" concepts are then used, which are based on different regulatory perceptions. The red-light concept is based on the idea of regulation as restricting behaviour and preventing the occurrence of undesirable activities. The green-light concept, in contrast, is based on a broader view of regulation as more enabling or facilitating of desired behaviours or activities.

With respect to risk regulation, there are particular areas of interest related to a safe society. These interests might be, for example, ensuring the provision of product information to citizens, establishing social solidarity (to support a welfare state), maintaining important basic state services, discouraging unwanted or criminal behaviour, and protecting vulnerable interests and citizens from harm or hazards in their work environment and in general.

Engen et al. (2016) argues that regulation is both a controlling mechanism and a judicial body that regulates a special area or field. For example, how should resources, safety and security be addressed in the oil and gas industry? Generally, there will be public interest in good governance and risk control regarding the threats and hazards associated with this industry. Risks that threat public infrastructure, public interests, human or technical security or the environment—both with respect to material and intangible values—as well as crime and terrorist attacks are often subject to risk assessment and regulation, which are in the public interest.

Nevertheless, legislation, regulation, governance and jurisdiction often lead to dilemmas, such as that referred to by Foucault (1999) in his term 'governmentality,' which pertains to how power (the state) establishes itself and constructs 'truth,' which becomes the basis for measures to facilitate regulation and control mechanisms.

According to Baldwin et al. (2012), a number of pitfalls must be avoided to ensure that affected citizens, interests or social groups will accept the given regulation. First, there must be a balance between efficiency and control—the regulator must be sensitive to democratic rights. This again presupposes regulatory legitimacy, which is a premise for success. The premise of legitimacy is that a regulator has the necessary competence and a relationship of trust with the affected group. Experience shows that, to succeed, the processes of implementation must be transparent and fair, as well as efficient and easy to implement.

Baldwin et al. (2010, 2012) argues that there are some well-recognized reasons for regulation, including social protection to ensure social solidarity, human rights to protect weaker citizens, rationing to ensure the allocation of scarce commodities or goods based on the public interest, and the avoidance of moral hazards in sharing costs and benefits.

However, many sectors and industry regulation efforts are based on a combination of rationales and have corresponding strengths and weaknesses related to their design and implementation, including the societal safety and preparedness principles operating in Norway, which we discuss in the following.

3 CUSTOMS' CONTRIBUTION TO A SAFE SOCIETY

Is Customs regarded as an important contributor to ensuring safe societies? Customs in Norway is organized under the Ministry of Finance, is divided into six regions and a directorate, and employs roughly 1,600 people.

Societal safety efforts to ensure public security and civil protection in Norway is based on four principles: responsibility, similarity, proximity and cooperation, as found in the abovementioned contingency preparedness principles applicable to public institutions. The first three principles were established in 2000 by the Norwegian Official Report NOU 2000:24 A vulnerable society. The principle of responsibility implies that an organization normally responsible for a task is also responsible for the necessary emergency preparedness and crisis management in that area. The principle of similarity means that the organization operating during a crisis should be as similar as possible to the regular organization. The principle of proximity states that a crisis should be handled at the lowest possible level.

The principle of cooperation was introduced after the terrorist attacks in 2011, which are referred to as 22 July. It requires authorities and agencies to have independent responsibility to ensure good cooperation with relevant parties in prevention-related issues and crisis management (White Paper: Meld. St. 10 (2016–2017) Risk in a Safe and Secure Society).

As a contributor to societal safety, Norwegian Customs seemed to have been given scope for a greater role following the establishment of the principle of cooperation. Customs' presence along the borders of Sweden, Finland and Russia, and at international airports and along the coast, points to its great potential for becoming a vital contributor to the protection of society.

The goal of this study is to determine the extent to which Customs is actually regarded as a contributor to Norway's societal safety, and if it is so regarded, whether or not this attitude has changed over the past 17 years.

The reason for the rather limited time span is that the term societal safety had not yet been established in Norway prior to this time. The term was officially defined in 2002 (White Paper (St.meld. nr. 17 (2001–2002)), although it had been in unofficial and limited use for a couple of years prior.

Nine of the documents we investigated in this study were strategic official documents from Norwegian Customs, 15 were from the Ministry of Finance, and 12 were strategic official documents related to societal safety published between the years 2000 and 2017. The latter were mainly Official Norwegian Reports, White Papers and draught resolutions to Parliament (Stortinget).

In this study, we examined strategic Norwegian Customs documents published in the past 17 years, since in 2001, Customs began to explicitly claim its role as an important protector of society. Customs has been increasingly conscious of this role in recent years, and its intention seems to have intensified after the Norwegian Tax Administration took over the responsibility of collecting taxes and duties in 2016. In the last document studied (2017), high ambitions are expressed in Custom's statement that it protects society through risk-based, customized and effective control measures.

The strategic documents produced by the Ministry of Finance that we investigated in this study do not explicitly acknowledge the role of Customs as a protector of society until 2016, but then do so emphatically.

The first Norwegian Official Report included in this study is the "NOU 2000:24 "A vulnerable society." Its main focus was to assess societal vulnerability to strengthen Norway's levels of security and preparedness. The role of the Ministry of Finance with respect to societal protection is limited to banking and financial issues, and Customs is not mentioned at all as a relevant contributor.

Only two of the 12 documents related to societal safety clearly recognize Customs as a protector of society. After the terror attack of 22 July 2011, a government-appointed commission authored a Norwegian Official Report (NOU 2012:14 22. juli-kommisjonens rapport) evaluating the incident and proposing answers as to how such a terrorist attack could happen. In that report, Customs is mentioned almost 150 times, and it is evident that prior to the terrorist attacks, Customs had alerted the Norwegian Security Police Service regarding the terrorist's importation of weapon-related products. The terrorist's purchases initially resulted in his being placed on the watch list of the Norwegian Security Police Service, but no further action was taken in response to the information given. The commission states in its report that Customs is an important contributor to societal protection, and it criticized the Norwegian Security Police's poor coordination with Customs.

The second document that acknowledges Customs as a protector of society is the Office of the

Auditor General of Norway, who audited the Customs' border control in 2014 (Document 3:7 (2013–2014).

The most recent document included in this study is the white paper titled: Risk in a Safe and Secure Society (Meld.St.10 (2016–2017). It defines and discusses eight core areas considered to be highly significant for public security, three of which are contagious diseases, hazardous substances and civil–military cooperation. With respect to contagious diseases and dangerous substances, the white paper refers to the national strategy for Chemical, Biological, Radiological, Nuclear, and Explosive (CBRNE) preparedness. It also addresses various kinds of heavy crime, including the financing of terrorism. In all these fields, Customs plays a natural role through its administration and control of the flow of goods over the border.

The white paper clearly states the legal authorities considered to be relevant contributors to a safe and secure society, and Customs is not among them. Despite the fact that Customs obviously protects society from, e.g., CBRNE-goods and the transportation of dangerous goods over the border, the agency is only marginally referenced and is not mentioned at all in relation to societal protection.

The term societal safety has also changed during the years included in this study. After the 22 July terrorist attacks, the concept was broadened, and today it also includes a preventive dimension, i.e., the protection of society. This change affects Customs as it now is aligned with its core responsibility. Through the administration and control of the flow of goods, Customs can contribute to the protection of society. Specifically, it does so by preventing the smuggling of dangerous goods and illegal imports/ exports that can destabilize the economy and/or may be directly hazardous to the population. One could perhaps expect this definition change to naturally move Customs closer to its role as societal protector. However, the results of this study reveal that there has been no change in the way Customs is officially regarded in the documents presented to Parliament in 2000 versus those in 2017.

4 DISCUSSION

Is Customs regarded as a vital contributor to societal safety?

Beck (1992) emphasized that political and public institutions lack the political expertise to handle risks and threats. With this in mind, we might ask why the administrative system in Norway did not follow up on the proposals concerning risk management coordination and cooperation introduced in the "NOU 2000:24 "A vulnerable society", and later in the 2006:6 white paper, and most recently in the 2012 22 of July report. By implementing the societal safety and preparedness principles of responsibility, similarity, proximity and cooperation, Norwegian authorities have made efforts to regulate and meet the requirements of societal protection. Societal protection, as Baldwin et al. (2012) argue, is a well-recognized justification for risk regulation. However, in reality, its implementation, especially with respect to organisational cooperation, has not been an easy task.

The Ministry of Justice is in charge of the overall horizontal governance between ministries and departments for the enforcement of the societal safety and preparedness principles. The review of the public documents in this research project spanning the past 17 years reveals that cooperation between governmental departments is fragmented. This situation might be related to the principle of responsibility, which is strong in Norwegian governance, thereby leading to ministries to leave this responsibility to the Ministry of Justice, and to forsake their own capacities to implement cooperative measures.

With the focus on enabling and facilitating desired behaviour and activity—the intention of the societal safety and preparedness principles—regulating public institutions according to "green-light" concepts is a challenging task. Public institutions are to some extent self-governing and represent complex organizations with many levels and factions.

The National Audit Office (Riksrevisjonen) has noted the difficulties of the Ministry of Justice in accomplishing organisational cooperation (Document 3:8, 2016–2017). This report found the mechanisms for cooperation to be weak or not functioning at all, as presupposed. When rules and regulations are to be enforced by various authorities under different departments, the inevitable result is weaker cooperation mechanisms.

Additionally, the fact that Customs is organized under the Ministry of Finance, whereas the Ministry of Justice has the responsibility for providing risk or societal safety regulations, highlights the fragmentation of the hierarchy in maintaining and enforcing laws and regulations. Vital societal safety actors, such as the Ministry of Justice, seem to lack sufficient knowledge about how Customs enforces numerous regulations related to various authorities. This knowledge gap regarding the opportunity set of Customs might be the main reason for the Ministry's failure to involve Customs in cooperation and coordination efforts to prepare for and manage the risks related to societal safety.

The result of this situation has been fragmentation, and the dilemma regarding the adoption of

centralized or decentralized management has led to something in between, or a vacillation between both. According to Fimreite et.al. (2011), an overall body is missing, one whose main tasks include the effective cooperation, harmonization, and integration of societal safety public authorities in Norway.

The specialization and resources of Customs and their broad control authority do not yet seem to have been taken into consideration with respect to ensuring societal safety. The question is whether the Customs' vague or less visible role as a vital societal safety actor can be regarded as an informal deregulation, which can only lead to increased risk in the form of undesired incidents or accidents that threaten societal safety.

4.1 *Risk management in public institutions*

When vertical cooperation in public institutions does not function optimally, neither do the communication and information processes that are vitally important in effective risk management. Some public institutions have developed a hierarchical steering model incorporating a hierarchical command and control system. In these systems, the horizontal dimension of cooperation and coordination can be challenging.

Risk regulation presupposes that governing units hold a combination of specializations, both vertically and horizontally, which is one reason why the task of ensuring overall cooperation, as led by the Ministry of Justice, is so important. Critics from the National Audit Office (Riksrevisjonen) (Document 3:8, 2016–2017) point especially to the need to strengthen cooperation between actors across sections or sectors during planned exercises and to then follow up. Customs' role in societal safety issues would benefit from better horizontal cooperation and coordination,

Despite the stated ambitions of both the Ministry of Finance and Customs itself regarding the important role played by Customs in societal protection, there remains a dilemma. This dilemma is between the sectorial decentralized responsibilities and follow up regarding societal safety issues on one side and the centralized decision-making authorities and resources on the other. This leads to weak governance and cooperation and, accordingly, the inevitable invisibility of the role of Customs to central decision-making authorities.

This contrasts with Baldwin's (2012) concept emphasizing that, to be efficient and easy, processes of implementation must be transparent and fair. How else to ensure legitimacy, which presupposes that the regulator has the necessary competence and trust of the populace?

4.2 *An appropriate regulation regime?*

If we consider Norway's public institutions such as Customs and other important cooperating authorities, can we conclude that Norway has an effective risk regulation regime?

The societal safety and preparedness principles of responsibility, proximity and similarity are based on an understanding that various specializations are needed in public institutions. The concept behind the cooperation principle is to ensure that the other principles are not developed using a sectorial approach that erects barriers between actors.

In white paper No. 17 (2001–2002), the cooperation responsibility of the Ministry of Justice concerning the preparedness and civil protection of society was underlined.

As Customs is subordinate to the Ministry of Finance and the Ministry of Justice has responsibility for societal safety, it is difficult to readily achieve the necessary horizontal cross-section overview and cooperation necessary for good societal safety governance.

Governance consists of planned and objective-oriented activities coordinated between relevant actors, which presupposes mutual dependence (Engen et. al. 2016). If the vertical coordination (between departments and directorates) is weak and controlling authorities lack knowledge about Customs' enforcement of regulations related to various authorities, the opportunity set for Customs to contribute to safe societies is reduced.

This highlights the importance of the participation of relevant actors in the decision-making process. If the Ministry of Finance does not participate in relevant assemblies in which societal safety issues are discussed and decided upon, there is no avenue for obtaining the support from those in charge of societal safety. The successful outcome of political decision-making processes depends on who participates and the responsibilities or mandates of the actors.

When these actors do not view Customs as an equal participant, the decision-making processes and negotiations concerning the role of Customs as a societal protector cannot be effective. In addition, Customs must be allowed to participate in these processes to demonstrate their societal protection resources and goals.

Since societal safety policy is organized piecemeal across sectors, it has become fragmented due to the spread of responsibilities among many actors. Fimreite et al. (2011) stated that with improved horizontal cooperation between the departments involved, the development of different policies that undermine each other can be avoided. Instead, Norway can make the most of

scarce resources by bringing together various policy interests to initiate combined action and synergy in the field of societal safety.

The Ministry of Finance may not be reaching the necessary political circles to share the goals and resources of Customs, since there is such a clear discrepancy between the Ministries of Finance and Justice regarding the role of Customs in matters of societal safety.

5 CONCLUSIONS

Is Customs recognized as an important contributor to ensuring a safe society in Norway? In this article, we discussed the extent to which the resources held by Customs are known and utilized in this effort to ensure such a safe society.

There will always be dualism in the Customs portfolio, i.e., despite its ambitions to do so, it can never fully control the flow of goods over its borders. Societal safety must be balanced against democratic values and the national obligation to maintain free movement of legal goods, as aligned with what Baldwin et al. (2012) refers to as regulator sensitivity to democratic rights. However, the documents we analysed reveal that, although Customs and the Ministry of Finance are eager to perceive Customs as having a vital role in protecting society, Customs is not sufficiently visible in the national preparedness overview. Our analysis indicates that the intensified focus of Customs itself as well as the Ministry of Finance with respect to its societal safety role has not been recognized by the Ministry of Justice, which is responsible for societal safety and preparedness in Norway. Based on the societal safety and preparedness principles of responsibility, proximity, similarity and cooperative governance, Customs has knowledge of, presence and wide statutory authority along Norway's borders, and can thereby make a vital contributions to ensuring a safe society. However, it seems that Customs has yet to be regarded by the state as an important contributor to safe societies; at least Customs is not *described* as such in the documents analysed. Furthermore, there has been no change in this perspective over the 17-year study period. This represents a paradox in light of Beck's (1992) description of a risk society and the increasing globalization in recent decades.

The results of this study shed light on the potential for better governance with respect to the cooperation principle related to risk management efforts. By focussing more on cooperative governance, Customs' importance as a societal protector in the nation's preparedness regime might become more visible, effective, and valued.

As yet, however, the complexity of the policy frameworks, lack of cooperation between vital actors and the resulting fragmentation of information can only lead to a continued lack of awareness and poor capacity building and training initiatives that are so vital for ensuring Customs' ability to be fully operationalised in ensuring a safe society.

REFERENCES

Aven, T., 2007. *Risikostyring* (English: Risk Management). Universitetsforlaget. Oslo.

Baldwin, R., Cave, M., and Lodge, M., 2010. *The Oxford Handbook of Regulation*. University Press. Oxford.

Baldwin, R., Cave, M., and Lodge, M., 2012. *Understanding Regulation: Theory, Strategy and Practice*. University Press. Oxford.

Beck, U., 1992. *Risk Society: Towards a New Modernity*. Sage. New Delhi.

Engen, O.A., Kruke, B.I., Lindøe, P., Olsen, K.H., Olsen, O.E. and Pettersen, K., 2016. *Perspektiver på samfunnssikkerhet*. Cappelen Damm Akademisk. Oslo.

Folgero, R., 2017. From Tax-Revenue Collector to Societal Protector. Master thesis. University of Stavanger.

Fimreite, A.L., Lægreid, P., Rykkja, L.H. 2011. *Organisering, samfunnssikkerhet og krisehåndtering (Organizing, Societal Safety and Crisis Mangement)*. Universitetsforlaget. Oslo.

Giddens, A., 1999. *Risk and Responsibility Modern Law Review* 62(1):1–10.

Lindøe, P., Kringen, J., Braut, G.S. (2015): *Risiko og tilsyn, risikostyring og rettslig regulering. (Risk Management and Judicial Regulation)*. Universitetsforlaget. Oslo.

OECD 2010. *Risk and Regulatory Policy. Improving the governance of Risk, Paris: OECD reviews of Regulatory Reform*. Paris: OECD publishing.

OECD 2011. *Regulatory Policy and Governance. Supporting Economic growth and serving the Public interest*. Paris: OECD Publishing.

Selznick, P. 1985. Focusing organizational Research on Regulation in R. Noll (ed.) *Regulatory Policy and the Social Sciences*. (Berkeley, CA, 1985), 363, quoted Opus, Regulation, 1.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Management of airport security screening system effectiveness

J. Skorupski
Faculty of Transport, Warsaw University of Technology, Warsaw, Poland

P. Uchroński
Upper Silesian Aviation Group, Poland

ABSTRACT: Airports are potential targets for terrorist attacks. High level of risk inclines airport managers to take preventive measures. The decision-making process is difficult, as there have been no methods for the quantitative assessment of changes in security levels as a result of the actions taken. The aim of this study is to develop a method that quantitatively evaluates the effectiveness of security screening, based on various operating parameters of the system, which may become decision variables in the safety management process. The method uses the theory of fuzzy inference for overall assessment of the effectiveness of three key ways to secure air transport: screening of passengers, hand and checked baggage. The method based on multi-criteria group decision making was adopted for validation of the method and the developed calculation tool FASAS. The simulation experiments that were carried out allowed a practical presentation of a method for increasing security levels in medium and long time horizon. The method gives also an answer to the question about the quantitative effects of the decisions made. Application of the fuzzy inference theory enables an effective calculation tool to be developed which is usable in managerial practice.

1 INTRODUCTION

Every air travel is preceded by passenger and baggage screening. Terrorist threats forces the airport management to take measures to ensure security to the passengers and personnel. These involve considerable expenses and pose a serious organizational challenge. Therefore, the security control becomes a significant part of airport budget and affects the functioning of the entire company. At the same time, process management is difficult due to the lack of proper supporting methods. This applies particularly to evaluating the effects of the measures taken in relation to the achievable security levels.

In this study a quantitative method for evaluating the effectiveness of the airport passenger and baggage security screening system is proposed. It is difficult to accurately describe this ill-defined problem, so it often comes down to an intuitive or a 'trial and error' approach. Our approach allows us to formalize the expert knowledge and achieve more objective results, and certainly makes it possible to carry out a comparative analysis.

The method is based on fuzzy logic, more precisely on the fuzzy inference systems. The computer-aided tool FASAS (Fuzzy Airport Security Assessment System) enables practical support of airport management in terms of security control.

1.1 *Managing the system of an airport passenger and baggage security screening*

The security checks of passengers and baggage in airports is regulated by extensive regulatory system (European Commission 2015). However, compliance with the regulatory requirements does not exclude the option of making individual managerial decisions that can significantly affect security, capacity or comfort of passengers. The legislation in force does not give an indication on how to practically organize the airport control system, which includes not only the physical activities visible to the passengers, but also a series of infrastructural, personal and procedural actions, requiring expenses relevant to the scale of passenger traffic.

There are usually several key issues found in the passenger and baggage security screening management.

1. Choosing the number of Security Control Areas (SCA).
2. Having a proper number of staff to perform their duties.
3. Selection of SCA equipment.
4. SCA organization.
5. Dynamic modification of system operating parameters.
6. Operational management of the system.

The airports mostly function as economic agents and attempt to achieve a positive financial result, which in turn is a determinant of the investments planned for airports. However, the managers responsible for planning investments (in security equipment or training, for example) would like to know the measurable effects of such actions. This also applies to the comparison of several alternative investment decisions. The lack of quantitative methods makes it impossible to evaluate their actual effects.

In the literature on airport security management an important issue is the scope of control operations and their effect on an airport capacity (Hainen et al. 2013, Butler & Poole 2002, Leone & Liu 2005, Van Boekhold et al. 2014, Kierzkowski & Kisiel 2015) and the passenger comfort and satisfaction (Alards-Tomalin et al. 2014, Benda 2015, Gkritza et al. 2006, Sakano et al. 2016). Increasing the scope of control operations requires obviously increased expenditures which are not always reasonable (Stewart 2010, Stewart & Mueller 2014 2015, Gerstenfeld & Berger 2011, Gillen & Morrison 2015, Prentice 2015).

1.2 *Concept of the study*

This study assumes the perspective of airport management and, more precisely, the person in charge of passenger and baggage screening, undertaking medium—and long-term upgrade actions. The method developed will meet the proactivity criterion, which means that it will allow planning of actions before security risks are present. We will seek an action strategy that will increase the effectiveness of control without compromising the capacity, assuming that funds to pay for the actions are available.

This study is a finalization of the previous works, in which we proposed methods for evaluation of hand baggage (Skorupski & Uchroński 2015a), checked baggage (Skorupski & Uchroński 2015b) and passenger (Skorupski & Uchroński 2016a) screening systems. All those methods are based on fuzzy inference systems. The objective of this paper is to integrate in hierarchical structure and demonstrate that the calculation tool FASAS created can be used for operational management of an airport security screening system.

The remaining part of the paper is organized as follows. Section 2 gives a brief overview of theoretical grounds and presents an integrated fuzzy inference system which is the result of the study. Section 3 provides an analysis of the passenger and baggage security screening system at Katowice International Airport (ICAO code: EPKT, IATA code: KTW). Section 3.2 analyses several scenarios and determines quantitatively the effectiveness of the control relative to the upgrade actions taken in the medium-time horizon. Section 4 contains the summary and conclusions.

2 FUZZY SETS IN ASSESSING THE EFFECTIVENESS OF SECURITY SCREENING IN AIR TRANSPORT

2.1 *Uncertainty and subjectivity in making decisions in aviation security screening systems*

In technical activities, including transport processes, the information available often tends to be inaccurate and incomplete. If this is the case, decisions are made in uncertainty conditions. There are many types of uncertainty and various mathematical methods and ways to reduce its adverse effects on the decisions being made.

In airport passenger and baggage screening systems, the knowledge of the effectiveness of screening cannot be obtained from measurements. It is necessary to acquire the knowledge from experts in the field. They usually use informal language and their knowledge is expressed inaccurately and in approximation. This is therefore a typical example of acting in linguistic uncertainty conditions. For that reason, fuzzy inference systems based on fuzzy logic were used in this study (Siler & Buckley 2005).

2.2 *General information on fuzzy inference systems*

As a fuzzy set we understand a set in a form

$$A = \left\{ \left(x, \mu_A(x) \right) : x \in X, \mu_A(x) \in [0,1] \right\}$$

where μ_A is a membership function of this set. Every object can belong to a certain degree to a fuzzy set. The element degree of membership in a fuzzy set is determined by the membership function. They can have various shapes as the case may be.

A linguistic variable refers to a variable whose values are words or sentences in a natural or artificial language. Such words or sentences are called the linguistic values of a linguistic variable. In formal terms, a linguistic variable can be defined as the five-tuple (Czogała & Pedrycz 1980):

$$L, T, X, G, M \tag{1}$$

where:
L – linguistic variable name,
T – a set of syntactically correct linguistic variable values L,
X – consideration space of a linguistic variable L,
G – syntactics of a linguistic variable,

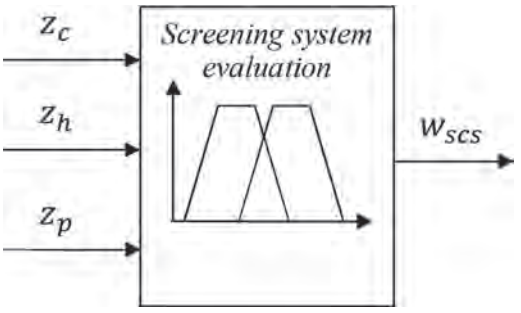


Figure 1. General scheme of the fuzzy model of an airport passenger and baggage screening system.

M – semantics of a linguistic variable, determined by a set of algorithms allowing the assignment of linguistic variable value of certain fuzzy set A.

2.3 *General structure of a fuzzy model Screening system evaluation*

The general structure of the fuzzy model for evaluating the effectiveness of the passenger and baggage screening system is shown in Fig. 1. The input variables of the *Screening system evaluation* model are:

Z_h – effectiveness of hand baggage screening (linguistic variable *Hand baggage*),
Z_c – effectiveness of checked baggage screening (linguistic variable *Checked baggage*),
Z_p – effectiveness of passenger screening (linguistic variable *Passenger screening*). All the input variables are the outputs of other local models. They are presented in detail in the following sections.

2.4 *Input variable hand baggage*

Hand baggage screening has two aspects. The first one of them is scanning luggage with X-rays. The second one is manual inspection carried out by the security screening operator (SSO). Three input variables correspond to those two aspects (Skorupski & Uchroński 2015a). Two of them—*Device evaluation* (y_d) and *Type A Errors* (x_{eA}) – are related to the X-ray scanning of cabin baggage.

The *Device evaluation* linguistic variable makes it possible to express the impact of the technical factor on the possibility of effectively detecting a prohibited article in baggage (Skorupski & Uchroński 2016b). That parameters taken into account are: detectability of different materials, presence and efficiency of Threat Image Projection (TIP), the number of detection lines used, and the age of the device.

The other linguistic variable—*Type A Errors*—describes the actual skills of SSO to use the X-ray device smoothly and efficiently. Type A error consists in failure to indicate the virtual prohibited item interposed on the image of the scanned baggage by the TIP system.

The third variable used in the model is *Manual Inspection* (y_m). It is used to describe the efficiency of manual inspection carried out for some cabin baggage. We assumed that the quality of inspection depends on linguistic variable *Employee Evaluation* (y_p). It is, in turn, dependent on the experience of SSO, the amount of time since the last comprehensive or current training they have undergone, and their overall attitude to work they perform (Skorupski & Uchroński 2015c). On the other hand, the quantity of baggage subjected to manual inspection also has an impact on the efficiency of such screening. We have broken down the concept of manual inspection into two linguistic variables *Number of type B manual inspections* (x_B) and *Number of type C manual inspections* (x_C).

2.5 *Input variable checked baggage*

Evaluation of the effectiveness of checked baggage screening depends on two factors: efficiency of X-ray equipment and efficiency of the checks performed at SCA, particularly with participation of SSOs (Skorupski & Uchroński 2015b). So the local model *Checked baggage* has two input variables—*Device's assessment* (y_d) explained in Section 2.4 and *SCO control* (y_o). The latter depends on the employee evaluation, type A errors (variables *Employee Evaluation* and *Type A Errors*—Section 2.4) and an important variable *Control organisation option* (x_o). This variable describes the checked baggage screening organisation.

2.6 *Input variable passenger screening*

The *Passenger screening* variable (z_p) depends on three input variables: *WTMD's evaluation* $\left(y_{WTMD}\right)$, *Frequency of manual control* $\left(x_{mp}\right)$ and *Manual Control* $\left(y_m\right)$. This fuzzy inference system creates a hierarchical structure as the *WTMD's evaluation* and *Manual control* are outputs of local fuzzy models (Skorupski & Uchroński 2016a).

An evaluation of the effectiveness of Walk-Through Metal Detector (WTMD) in the passenger security control depends on: number of detection areas, ability to set sensitivity in different detection areas, visualisation of the detection areas and the system for the support of manual control.

The input variable *Frequency of manual control* is functionally dependant on two decision variables: *WTMD's sensitivity* $\left(x_s\right)$ and *Frequency of additional controls* $\left(x_{am}\right)$. The former may be set

directly at the WTMD, while the latter is arbitrary based on the applicable, current regulations and the current threat level for the given airport.

The *Manual control* variable is in fact an output of the local fuzzy inference model. The inputs of the model are the linguistic variables: *Employee number* (x_w) and *Employee evaluation* (y_p), which is the output of local inference model with the appropriate inference system (Section 2.4).

2.7 *Output variable screening system evaluation*

A general evaluation of the effectiveness of a security screening system at an airport is described by the linguistic variable *Screening system evaluation*. Its membership function which is the same as the one presented in Fig. 2, is the output value.

A group of experts were asked to define the fuzzy inference rules. They are practitioners with extensive experience in airport security management. The following problem has been brought before them. If we assume that the purpose of the whole security system at the airport is a safe flight (in which there is no explosion, hijacking or an assault on another passenger), which of these three types of screening is the most important to achieve this objective? The knowledge base consists of 125 fuzzy rules. Some of them are presented in Table 1.

2.8 *Validation of the model Screening system evaluation*

To validate the rules of the model and also to remove any inconsistencies the method described in (Skorupski, 2015) was used. It consists in using expert opinions expressed in terms of multiple criteria in the form of both numerical and linguistic assessments. Experts define the conclusions of rules as so-called half-marks in order to increase the method's flexibility. Automatically generated rules are compared to the rules that were provided by experts in order to detect inconsistencies. As a result of using a new concept of half-marks, it is possible to perform not only the verification, but also automatic selection of the final form of a conclusion.

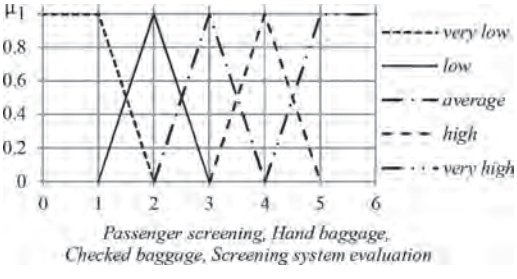


Figure 2. Membership functions of the linguistic variables: Passenger screening, Hand baggage, Checked baggage, Screening system evaluation.

Table 1. Fuzzy inference rules for the model *Screening system*.

Rule	Passenger screening	Hand baggage	Checked baggage	Screening system evaluation
6	Low	Low	Very low	Very low
17	Average	High	Very low	Low
54	Very high	Very low	Average	Average
93	High	High	High	High
120	Very high	High	Very high	Very high

3 EVALUATION OF THE EFFECTIVENESS OF AN AIRPORT PASSENGER AND BAGGAGE SECURITY SCREENING SYSTEM

This section describes the experiments with the use of the method. At first, the reference variant will be examined, and then the variants presenting the benefits of various actions: extending the scope of training, equipment change and organisational activities.

3.1 *Example of a security screening system effectiveness assessment in KTW airport*

This section will present the reference variant that corresponds approximately to the situation in KTW in 2014. The starting point in various airports may differ, however the solutions proposed constitute a kind of a standard that can be assumed as a typical situation. The basic scenario S_0 parameters required are compiled in Table 2.

Simulation analysis carried out for typical values of parameters listed in Table 2, using the FASAS tool, provided the results as given in Table 3.

The final, defuzzified rating value of the entire system is 2.95, which corresponds to the value *medium*. This is in line with the expectations and typical for most airports. Meeting the regulatory requirements and acting in typical threat levels causes the system to be configured so as to achieve the average effectiveness in detection of prohibited items.

3.2 *Analysing the possibilities of influencing the effectiveness of airport security screening*

This section analyses the effects of implementing some measures to improve the security screening effectiveness: shortening training intervals, partial or complete replacement of equipment, greater emphasis on the system condition monitoring.

Table 2. Input parameters of scenario S_0.

Parameter	Passenger screening	Hand baggage screening	Checked baggage screening
Security screening operators			
Experience [mth]	36	36	36
Comprehensive training [mth]	18	18	18
Ongoing training [mth]	2	2	2
Attitude	2 (average)	2 (average)	2 (average)
Type A errors	–	15%	20%
X-ray equipment			
Age [yrs]	–	9	5
Number of TIPs	–	6000	0
TIP frequency	–	2.8%	0
Number of lines	–	1	2
Detectability	–	6	6
Walk-Through Metal Detectors			
Visualisation	20	–	–
Detection areas	20	–	–
Sensitivity [g]	165	–	–
Support	1 (yes)	–	–
Other parameters			
Organisation variant	–	–	4
Number of employees	4	–	–
Additional checks	13%	–	–
Checks B	–	9%	–
Checks C	–	0%	–

Table 3. Evaluation of the passenger and baggage screening system in the scenario S_0.

System	Defuzzified rating	Linguistic rating
Passenger screening	3.16	> *medium*
Hand baggage screening	2.30	*low*/*medium*
Checked baggage screening	2.72	*low*/*medium*
Total	2.95	*medium*

3.2.1 *Greater emphasis on training and staff awareness*

As a standard, comprehensive training is given every 3 years which results in the average time from the last comprehensive training is 18 months. In the simulation experiment described below (scenario S_1) it was assumed that the training intervals were reduced to 6 months, which means that the mean time, and therefore the value of the *Comprehensive training* variable (for all three systems) is 3 months.

Shortening the training intervals has also positive effects on the number of type A errors made by the SSOs. Measurements using the software available in TIP system showed that a comprehensive training results in a decreased number of type A errors on average by 3 percentage points for hand baggage screening operators, and on average 8 percentage points for those screening the checked baggage.

The results of evaluation of the passenger and baggage screening system in the scenario S_1 are given in Table 4.

As can be seen, reducing training cycles do not give satisfactory improvement of an airport security screening system—the grade is 2.98 vs. 2.95 for a standard training system. It is basically understandable, as already for the initial parameters the ratings of staff in all subsystems are at the level *very high* and there is not much room for improvement.

3.2.2 *Equipment change*

Given in Table 2, the specifications of the equipment used at Katowice Airport indicate that it is not very advanced. The evaluation made with the use of FASAS system shows that they vary between *low* and *medium*.

The effects of the decision to replace the X-ray equipment with more advanced devices will be analysed. Let Scenario S_{2a} denote replacing half of the equipment and Scenario S_{2b} denote replacing all equipment.

A part of input data that has been changed in relation to the reference variant is shown in Table 5.

The simulation experiments carried out indicate a noticeable improvement in the assessments of those systems which had been provided with new equipment, and also the overall rating of the entire screening system. The results are listed in Table 6.

As it can be noted, the effect of an upgrade activity is considerable. This is particularly true for the replacement of all equipment. However,

Table 4. Evaluation of the passenger and baggage screening system in the scenario S_1.

System	Defuzzified rating	Linguistic rating
Passenger screening	3.24	>*medium*
Hand baggage screening	2.51	*low*/*medium*
Checked baggage screening	2.72	*low*/*medium*
Total	2.98	*medium*

Table 5. Input data for the equipment upgrade decision (Scenarios S_{2a} and S_{2b}).

Parameter	Passenger screening	Hand baggage screening	Checked baggage screening
Age [yrs]	–	0	0
Number of TIPs	–	6000	1000
TIP frequency	–	2.8%	2.8%
Number of lines	–	4	2
Detectability	–	10	10

Table 6. Evaluation of the passenger and baggage screening system in the scenarios S_{2a} i S_{2b}.

System	Scenario S_{2a}		Scenario S_{2b}	
	Defuzzified rating	Linguistic rating	Defuzzified rating	Linguistic rating
Passenger screening	3.16	>*medium*	3.16	>*medium*
Hand baggage screening	2.9	<*medium*	3.67	*medium*/*high*
Checked baggage screening	3.68	*medium*/*high*	4.0	*high*
Total	3.25	>*medium*	3.75	*medium*/*high*

attention should be drawn to quite a considerable cost of such project.

3.2.3 *Combined activity—equipment replacement and changing the training system*

The results of the experiment discussed in Section 3.2.1 indicate that shortening training intervals for employees working on low standard equipment does not give the desired outcome. Let us discuss Scenario S_3 involving this solution to be implemented as the second stage of the upgrade, after upgrading the equipment. Table 7 presents the experiment results.

In this case, further improvement of grades to *high* is noticeable. It is worth mentioning here that this does not prevent the possibility to improve the effects of security screening system management, as the changes in equipment did not affect the passenger screening system at all. There is further potential for growth in it.

3.2.4 *Emphasis on the diagnostics of personnel status*

While the role of the ongoing diagnostics of a facility condition is very important, when analysing

Table 7. Evaluation of the passenger and baggage screening system in the scenario S_3.

System	Defuzzified rating	Linguistic rating
Passenger screening	3.24	> *medium*
Hand baggage screening	4.0	*high*
Checked baggage screening	4.0	*high*
Total	4.0	*high*

Table 8. Evaluation of the passenger and baggage screening system in the scenario S_4.

System	Defuzzified rating	Linguistic rating
Passenger screening	3.24	> *medium*
Hand baggage screening	3.5	*medium*/*high*
Checked baggage screening	3.5	*medium*/*high*
Total	3.66	*medium*/*high*

reliability of socio-technical systems, it seems that it is slightly underestimated in diagnosing the personnel status.

The task of diagnosing personnel status in a security screening system is performed in many ways. One of them include the use of TIP system. The effects of resignation from using that system were analysed (Scenario S_4). This is practically reasonable in little used passages. Table 8 presents the results of Scenario S_4.

As can be seen, resignation from using TIP system results in lowering the grades to somewhere between *medium* and *high*. The extent of the reduction can be even greater since an SSO who is aware of not being evaluated may work carelessly, and take a very light-hearted approach by investigating minor doubtful issue less carefully or not at all. The above issues will be subject of further studies.

In contrast to the resignation from the TIP system, let us consider the effectiveness of the screening system change if more stringent standards are established, for example that an SSO is to make no more than 10% of type A errors (Scenario S_5).

The simulation experiment proves that imposing so high requirements does not result in further improvement in the effectiveness of the screening system. Reducing the type A error rates to 12% proves to be sufficient to achieve the best assessment of a worker. It is impossible to completely eliminate the errors in recognising images

of prohibited items, and moreover, the parameter discussed is not the only, but still very important, factor describing the quality of screening.

The last scenario (S_6) to be considered is the influence on the alertness (suspicion) of the SSOs in relation to the persons and baggage being checked. In the study (Skorupski & Uchroński 2015a) it is referred to as the B type checks.

Simulation experiment for Scenario S_6 was performed by modifying the input data for the Scenario S_3 in such a manner that we increased the number of B type checks from 9% (as observed and measured in real work conditions at KTW) to 15%.

The results of the experiment for the Scenario S_6 show that this results in an overall effectiveness grade of 4.41, which corresponds to somewhere between the *high* and *very high* rating. The result is so much interesting that it means a relatively high increase achieved by 'soft' actions involving influencing the SSO psychology.

4 SUMMARY AND CONCLUSIONS

In many areas the priority of security is declared as the governing rule. These areas include transport, particularly the air transport. Surprisingly, however, introducing innovations intended to increase security levels is often delayed in those areas and forced mainly by effective legislative procedures and enforcement methods. In our opinion, one of the major causes of that is the lack of tools providing an objective and quantitative assessment of the effects implementation of a given solution.

The method developed, together with the FASAS computer tool enables quantitative assessments of the effects of various managerial decisions on the effectiveness of an airport security screening system. It was discovered that, for the personnel training to bring a positive effect, the equipment needed an upgrade first. Of course, the equipment replacement alone contributes to the improved system effectiveness. A reverse approach proves to be less effective.

An important observation is the importance of personnel diagnostics. While monitoring of technical equipment condition is quite common, the ongoing diagnostics of the human factor is rare. Perhaps, it is because there are no proper methods and tools to do so. An example of such a tool in an airport security screening is the TIP system. It is not always possible to use similar solution in other areas, however, such attempts should be made and the role of diagnostics in management process should be increased.

Further works, including multi-criteria analyses, are required taking into account the overall security and capacity maximisation. In general, activities in average risk situation should focus on increasing security levels without compromising the capacity. Regrettably, for most of the solutions possible, this involves great cost and is time-consuming.

To conclude, simulation experiments indicate the importance of quantitative methods for evaluating the effectiveness of airport security control systems. The lack of such methods makes it difficult to provide reasoning for their implementation. A commonly observed attitude is that if ones activity complies with the standard and regulatory requirements, and possible innovation is expensive, there is not much sense in introducing it. A hypothetical quality improvement in the security level is difficult to prove. However, the quantitative methods, similar to that presented herein, make it possible to compare the results from various possible innovative activities what is critical for sound management.

REFERENCES

Alards-Tomalin, D., Ansons, T.L., Reich, T.C., Sakamoto, Y., Davie, R., Leboe-McGowan, J.P., Leboe-McGowan, L.C., 2014. Airport security measures and their influence on enplanement intentions: Responses from leisure travelers attending a Canadian University. *Journal of Air Transport Management* 37, 60–68.

Benda, P., 2015. Commentary: Harnessing advanced technology and process innovations to enhance aviation security. *Journal of Air Transport Management* 48, 23–25.

Butler, V., Poole, R.W., 2002. Rethinking Checked Baggage Screening. Policy Study*, Reason Public Policy Institute* 297.

Czogała, E., Pedrycz, W., 1980. Elements and methods of fuzzy sets theory, Silesian University of Technology, Gliwice (in Polish).

European Commission, 2015. Commission Implementing Regulation (EU) 2015/1998 of 5 November 2015 laying down detailed measures for the implementation of the common basic standards on aviation security.

Gerstenfeld, A., Berger, P.D., 2011. A decision-analysis approach for optimal airport security. *International Journal of Critical Infrastructure Protection* 4, 14–21.

Gillen, D., Morrison, W.G., 2015. Aviation security: Costing, pricing, finance and performance. *Journal of Air Transport Management* 48, 1–12.

Gkritza, K., Niemeier, D., Mannering, F., 2006. Airport security screening and changing passenger satisfaction: An exploratory assessment. *Journal of Air Transport Management* 12, 213–219.

Hainen, A.M., Remias, S.M., Bullock, D.M., Mannering, F.L., 2013. A hazard-based analysis of airport security transit times. *Journal of Air Transport Management* 32, 32–38.

Kierzkowski, A., Kisiel, T., 2015. An impact of the operators and passengers behavior on the airport's security screening reliability, w: Nowakowski, T., (red.), *Safety and Reliability: Methodology and Applications*: 2345–2354, CRC Press/Taylor & Francis/Balkema.

Leone. K., Liu, R., 2005. The key design parameters of checked baggage security screening systems in airports. *Journal of Air Transport Management* 11: 69–78.

Prentice, B.E., 2015. Canadian airport security: The privatization of a public good. *Aviation Security* 48, 52–59.

Sakano, R., Obeng, K., Fuller, K., 2016. Airport security and screening satisfaction: A case study of U.S. *Journal of Air Transport Management* 55, 129–138.

Skorupski, J., 2015. Automatic verification of a knowledge base by using a multi-criteria group evaluation with application to security screening at an airport. *Knowledge-Based Systems*, 85, 170–180.

Skorupski, J., Uchroński, P., 2015a. A fuzzy reasoning system for evaluating the efficiency of cabin baggage screening at airports. *Transportation Research Part C: Emerging Technologies*, 54, 157–175.

Skorupski, J., Uchroński, P., 2015b. Fuzzy inference system for the efficiency assessment of hold baggage security control at the airport. *Safety Science*, 79, 314–323.

Skorupski, J., Uchroński, P., 2015c. A fuzzy model for evaluating airport security screeners' work, *Journal of Air Transport Management* 48, 42–51.

Skorupski, J., Uchroński, P., 2016a. Managing the process of passenger security control at an airport using the fuzzy inference system, *Expert Systems with Applications*, 54, 284–293.

Skorupski, J., Uchroński, P., 2016b. A fuzzy system to support the configuration of baggage screening devices at an airport, *Expert Systems with Applications*, 44, 114–125.

Stewart, M.G., 2010. Risk-informed decision support for assessing the costs and benefits of counter-terrorism protective measures for infrastructure, *International Journal of Critical Infrastructure Protection* 3 (1), 29–40.

Stewart, M.G., Mueller, J., 2014. Cost-benefit analysis of airport security: Are airports too safe? *Journal of Air Transport Management* 35, 19–28.

Stewart, M.G., Mueller, J., 2015. Responsible policy analysis in aviation security with an evaluation of PreCheck. *Journal of Air Transport Management* 48, 13–22.

Van Boekhold, J., Faghri, A., Li, M., 2014. Evaluating security screening checkpoints for domestic flights using a general microscopic simulation model. *Journal of Transportation Security* 7, 45–67.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Information power supporting the rail systems safety

T. Kertis & D. Procházková
Czech Technical University in Prague, Praha, Czech Republic

ABSTRACT: Railway systems are complex systems interconnected with other concurrent or superordinate systems by physical, information, territorial and logical linkages, i.e. they are so-called system of systems. Introduced linkages are the causes of system of systems vulnerabilities, which directly limit the rate of safety of such system. The railway systems are also complex cyber-physical systems, because they use an information system (cyber space) that manages the physical system and they are also influenced by the physical system. The cyber (information) domain is very important because it ensures the semiautomatic and automatic control of system technology during the normal, abnormal and even critical conditions. Therefore, it is very important the care on the information power. The information power is defined as the rate of information effectiveness of systems. The information effectiveness is the key aspect for coping with risks connected with the interconnections in management of complex system, and it is main assumption for protection or ensuring the cyber security of such systems at build up their safety. The present work is focused on the information power, its definition and issues, how to manage and improve the system property. It also provides two examples of selected railway accidents, on which the information power issues are demonstrated. By multicriterial approach application the main criteria for information power size are proposed.

1 INTRODUCTION

Today's society is dependent on the technical and cybernetic systems that help to satisfy its basic needs. The systems, for their proper function, need correct and timely information from a real physical environment. For this purpose, information and communication systems and their technologies are used to create links between systems of different kinds and are able to process information according to the given rules faster than a person.

The stronger the human effort in improvement and streamline of processes towards their higher economic benefit is, the stronger dependency of human society on information technologies is, and therefore, the incessant need for their development arises. By improving and streamlining the processes in the direction to growth of economic benefits, we are introducing the new links, i.e. dependences, and by this we make up more complex systems, which are more vulnerable. This vulnerability leads to their failures in critical conditions, which have in many cases an impact on human security, provision of basic human needs and main functions of states. Therefore, also in the field of information technologies we are talking about critical information infrastructure, which, however, is rightfully interconnected with other technologies. The critical cyber-physical system arises by linking these elements.

To ensure the human security, we need to ensure the secured and, in many cases, the safe cyber-physical system. For formation of appropriate principles for ensuring the secure and safe cyber-physical systems, it is necessary to use the theory of information, and especially to determine parameters, which affect information power. Information power is just the quantity, which size determines the quality of decision; the higher information power is, the higher probability of quality decision is, and vice versa.

Transportation infrastructure and rail systems are not different; they are complex systems of systems and cyber-physical systems are dependent on information technologies, in which it is number of vulnerabilities that can lead to severe traffic accidents. Therefore, the work gives a selected part of theory of information related to the information power. It analyzes two examples of railway accidents and demonstrates the role of information performance on them. The result of work is the proposal of criteria determining the size of information power for different places of railway cyber-physical system.

2 THE INFORMATION TECHNOLOGIES INFLUENCE ON TRAIL SYSTEMS SAFETY

We are influencing the level of railway safety by improving the qualitative parameters, which are transport speed, capacity and the amount of goods and people transported, RAM parameters (Reliability, Availability, Maintainability), Life Cycle Costs, interoperability. In some cases, the qualitative parameters improve the safety, especially when safety is dependent on reliability and availability of performed function. In other cases, the qualitative parameters impair the safety, e.g., the increase of speed and coincident shortening the intervals between trains with a larger number of passengers, lead to higher danger for people.

Railway safety has a long-standing tradition in terms of technology, but there is still site for improvement in terms of human security. The level of safety is always limited by the operating conditions in which the system operates; if the conditions are very different from those for which the system is designed and when certain limits are exceeded, the system gets into a dangerous state, i.e. in state, in which it endangers itself and its surroundings, i.e. the surrounding systems, technology, the environment, economic links, the lives and health of people, and others (Kertis & Prochazkova 2017a, b, c).

Safety of the railway system means to ensure that the rail system is able to work perfectly in wide range of conditions. If range limits are exceeded, the system needs to recognize the change of conditions and to pass to another mode of safety management system, e.g. to apply measures and activities according to security plans, continuity plans, etc. At present, great emphasis is given on the development of well-secured rail systems (ARTEMIS 2014, CEN-CENELEC 2017, EU 2020a,) and therefore, they are focused mainly on safe technological platform, which is very important in terms of safety, but it does not solve complex issues, i.e. human security. Integral safety focusing on human security is still mostly ignored in practice. The size of investment in safety and security is also given by economic aspects (Prochazkova 2013, Kertis 2015, 2016, Kertis & Prochazkova 2016).

Information systems based on information technologies are implemented in all above mentioned areas, i.e. ensuring the quality of railway transportation, security and safety of railway systems. Information technologies interpret, help to manage or, in the case of automated operations, control all of these qualitative and safety parameters. Information systems and technologies are integral part of rail systems. Table 1 shows examples, how information systems are used in different areas of railway transportation. Correctly chosen information system parameters ensure the size of their information power, i.e. quality of information enabling the system to effectively react on possible unacceptable conditions. By this way they improve railway safety, not only at normal conditions, but also at abnormal and especially critical conditions.

Table 1. Examples of information systems used in railway domains.

Management and planning:
– evaluation of data from operation, timetabling,
– breakdown of staff services,
– decision-making, economic, accounting,
– communication with rescue forces and the police.
Management and control of operation:
– central supervision and management, dispatching,
– station and track technologies,
– data collection and processing on the train route,
– communication between stationary and train systems,
– protection devices.
Train operation:
– train control, vehicle control units,
– data transmission among train devices,
– tracking and control of train equipment (door, air conditioning, train radio, power equipment),
– human-machines interfaces (HMI, driver-technologies).
Passenger:
– information light boards,
– passenger check-in systems,
– entertainment systems, Wi-Fi,
– navigation systems—direction signs, for the disabled.

3 INFORMATION SYSTEMS AND TECHNOLOGIES, INFORMATION ORIGINATION PROCESS, INFORMATION POWER AND SECURITY

The theory introduced below is based on knowledge in domains of information systems, cyber-physical systems, complex systems and critical infrastructure (Moos & Malinovsky 2008, Prochazkova 2012, 2015, Novobilsky at. al. 2016, Kertis & Prochazkova 2017b, c).

Information, information systems and technologies include very wide domain that makes up the couplings among the systems. Information is today beside the material, energetic and financial resources ranked to main factors that determine the progress, not only in technology, but also in other domains of human activities. Information flows in systems make up the important linkages

and couplings the elements and whole systems in complex technological facilities. Without certain level of information, it is not possible to make up and to manage the processes of any nature, including the human society. The origin of information is conditioned by monitoring the certain properties of observed object or the common properties of group of objects. Each information system follows the entity properties using the particular language, which serves to creation of information on observed object. Following types of information systems are distinguished according to how the information is interpreted: syntactical information systems, which create the set of information images of state quantities of the object being observed; and process information systems, which represent the set of processes. Then an action information system affects the observed object by feedback, or it creates a model or a real object. In area of railway operation management, the action process information systems are mainly applied, and therefore, our work is focused on them. The process of origin of information, information system, new object or modify object, consists of sub-processes, or sets of links and their relationships, which are described in Table 2.

The qualitative characteristics of information systems and technologies can be influenced by appropriate settings of their parameters, such as: quality of process of producing information images (based on Frege's concept); information amount (equation 1); parameters of transmission matrix (equation 4). These parameters influence the size of information power (equations 5 and 6), and thus also the ability to deliver quick and correct decision of information system (7). Fregge functional concept of information image origination is composed from the sets: O_i is set of rated quantities on the object; P_i is set of states (observers); Φ_i is set of syntactic strings (data flow); I_i is set of information images of state quantities. Relations of the mentioned sets, which determine the quality of the

Table 2. Process of information origin.

No.	Process/set	Affected abstract nodes	Used information technologies	Process inputs	Process outputs
	Object identification	Object, observer	Physical receptors (sensors)	Observed condition (physical) quantities of object	Signals
2	Observation statement	Observer, language (syntax)	Sampling, quantization, coding/encoding	Signals	Data
3	Communication between the source and receiver of message	Language (of observer resp. System of data acquisition), message receiver	Telecommunication, transmission and communication systems	Data	Data
4	Interpretation set, information origination	Language (of observer or system of data acquisition, or receiver), information set (see no. 6)	Ontology, language	Data	Information
5	Relations hips of functions and structural arrangement of object, integrity verification	Information (see no. 6), the object	Actuator of system, action information system	Object, information	Information correctness, change of object
6	Information set in set of information systems	Information systems	Information systems	Information	Information
7	Interpretation process	Information (see no. 6), the new object	Signalizing and representation technology, artificial intelligence	Information	Image of object, new object

information image origination process according to are described by following parameters:

- a_{OP} identification,
- a_{PO} invasity (danger of breaking the integrity of state variables on the object being observed),
- $a_{P\Phi}$ projection in a set of symbols and syntactical strings,
- $a_{\Phi P}$ uncertainty correction and identification,
- $a_{\Phi I}$ interpretation, information origin,
- $a_{I\Phi}$ language constricts reflection,
- a_{IO} relation of functions and structural regularity,
- a_{OI} integrity verification.

The information measure is mostly characterized by Hartley tolerance rate, the information amount for the binary symbols' system (i.e. for the most present cyber-physical systems) it is expressed by equation:

$$I = \frac{1}{\ln(2)} \cdot \ln(N) \tag{1}$$

in which N is the number of possible reports (data):

$$N = S^n \tag{2}$$

where "S" means the number of characters in the alphabet A(A1, A2,…AS) and "n" is the number of elements in the character set.

The process information systems are characterized by graphs assigned to relationships:

$$I_i \approx F\left[P(t), \Phi(t)\right] \tag{3}$$

This assignment enables to perform the structural interpretations of complex information systems, evaluation of feedbacks, and the quality of transmission and information processing in partial information systems, and its information segment goes out from the matrix representation in following form equation:

$$\underbrace{\begin{pmatrix} I_2 \\ \Phi_2 \end{pmatrix} \approx \begin{pmatrix} t_a & t_b \\ t_c & t_d \end{pmatrix} \cdot \begin{pmatrix} I_1 \\ \Phi_1 \end{pmatrix}}_{[T_i]} \tag{4}$$

where T_i is the transmission matrix of i-th information segment (information systems segment). In real system from equation (3) it follows that information or set of information I_i is in relationship with set of system conditions and information flows in time. We can assign the information segment from equation (4) to data acquisition system, where I_1 are input (initial) information, Φ_1 is an input information flow and on output side I_2 are output information and Φ_2 is a transmitted information flow. Parameters $t_{a,b,c,d}$ can be obtained using both, the quantitative and the qualitative ways and in terms of case from they express:

- t_a – interpretation capability (for $t_a < 1$ the system has very small knowledge and interpretation capability, for $t_a = 1$ it has capability of interpretation of object properties in information system, for $t_a > 1$ it goes on the expert system with capability to create own information about the object on the basis of obtained data),
- t_b – filtration capability (for $t_b < 1$ the system on its output interprets lower amount of information than it obtained at its input information flow, for $t_b > 1$ vice versa),
- t_c – communicativeness (capability to provide output information flow on the basis of input information),
- t_d – system information throughput (i.e. the capability to transfer the input information flow to output information flow, in case of redundancy t_d is much higher than 1).

Qualitative parameters of systems of systems (including rail systems), which directly affect human security, such as safety, integrity, reliability, quality, availability, continuity, accuracy are directly dependent on the effectiveness of information systems. Information systems provide the required accuracy and timeliness of information and, in the case of action information systems, also the speed of the right decision. The efficiency level of the information system is expressed by quantity of the information power quantity:

$$P_i(t) = I_i(t) \cdot \Phi_i(t) \tag{5}$$

where Pi(t) is immediate information power (performance). The information power is also equal to the value of eliminated uncertainty quantity E per time unit t:

$$P = I \cdot \Phi = \frac{E}{t} \tag{6}$$

To ensure the safety of railway control systems, it is important to operate information systems, which provide quickly correct decision, which is closely related to information power. The probability of correct solution choice from set of alternatives and the probability of proper decision of control operation system, i.e. PCD (Probability of Proper Decision), are given by function that is dependent on level of knowledge of function "k" and information flow in time "Φ (t)":

$$PCD = F\left[\Phi(t), k\right] \tag{7}$$

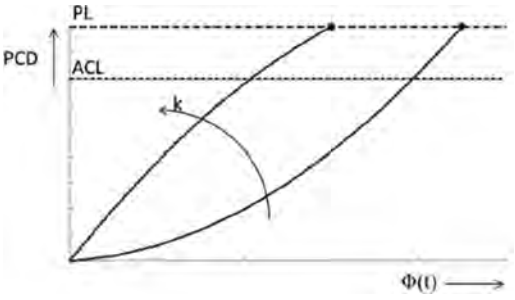


Figure 1. Probability of proper decision in systems control operation (Moos & Malinovsky 2008).

The relationship is depicted in Figure 1, where PL is the PCD maximum, and ACL is acceptable PCD level.

From these facts it follows that control systems, which rely on higher level of knowledge, are able to make quickly correct decision at lower load, i.e. they decide faster.

Each system works correctly only under certain conditions. Therefore, cyber-physical systems need to have: determined certain limits and conditions, which determine their qualitative parameters; and mechanisms, which react to surrounding system conditions. For very different conditions, they need to have plans for transition to other activities, i.e. operation rules for abnormal and critical conditions, which can happen at disaster origin. These issues of information systems are done by processes ensuring the information security, which are based on the protection of important cyber assets in a way that provides the required degree of confidentiality, integrity and availability for important information (CIA).

4 CASE STUDIES, ACCIDENTS COMMON CYBER CAUSES

For illustration of causes of traffic accidents caused by cyber network failures, we give two real cases; the first from the Czech Republic and the other from Spain.

4.1 *Moravany 2008*

On May 19, 2008, the serious railway accident at the rail station was in Moravany at 4h 48m; crash of freight train with passenger train followed by derailment. The accident caused one death, 4 slight injuries and direct financial damage were 12 643 092.- CZK (CR MD 2017).

The accident investigation report states the following direct causes: track circuit contact loss between train No. 5011 and track circuit 1K; reaction of interlocking system ESA 11 on unexpected change of station track No. 1 status of occupation. Next underlying causes: incompatibility of railway vehicles and track circuits as far as sanding (improving the friction between wheel and track) is concerned; internal logic of interlocking system, as far as processing of information about station track occupation received after track circuit reactivation is concerned. In terms of safety management system, the report emphasizes: operation of railway vehicles incompatible with track circuits without adequate safety measures.

This event is not unique; source (CR MD 2017) also states: "On August 29, 2008 at the Hulín railway station the event with the same background took place on an event as the event on May 19, 2008 at the Moravany railway station. After the same defect of same sending device of same railway vehicle of same series, the same procedure of the stationary protection device of the same type occurred at 17 h46 m55 s to change the status indication of the 3rd station track to "free", although it was still occupied by the stationary personal train Os 4256. This event took place without consequences only because of favourable circumstances and immediate proper response of the participating employees.

This event, in terms of common cyber causes (Kertis & Prochazkova 2017a), is combination of two types of causes, namely the false SW and/or inadequate HW: the failure of the sanding device that was still active after switched off by engine driver—insufficient maintenance (HW); use of sand with larger grain size—organizational error (HW); the sanding indication for the driver does not signify the actual state, only the intention—whether an electronic signal is given for sanding—insufficient design (SW/HW); signalling the unoccupied track circuit at the occupied track—insufficient design (HW/SW); automatic route setting to occupied track circuit—insufficient design (SW); absence of a remote train stop system—insufficient design (HW/SW); and insufficiency of communication equipment (freight train driver did not react to the stop call) - insufficient design, organizational error (HW).

The main cause of the accident is the fact that information was distorted due to both, the sanding and the bad response of rail protection system in the station, i.e. it is an error that occurred at interface of cybernetic and physical (cyber-physical) system.

4.2 *Santiago de compostela 2013*

The rail accident in 2013, which was a few kilometres from Santiago de Compostela in Spain, was

the worst rail accident in Spain in the last forty years. On July, 2013 at 20 h41 m, the derailment of high-speed passenger train at 179 km/h velocity originated on the Angrois curve (where the speed limit is 80 km/h). After the derailment, the most of coaches hit the concrete wall along the curve and the rear generator car caught fire. The result was 80 fatalities and 152 injured (almost all the passengers) (EU 2015).

Cause of the accident was the speed of train and the state working professional commission of inquiry ultimately arraigned the engine driver from failure and non-observance of rail regulations. These commission conclusions are not open for public, and they were questioned by the European Rail Agency (ERA). This agency described the root accident causes in its document for EU, i.e. it also revealed the weaknesses in the overall rail control system.

We briefly outline the ERA report conclusions (EU 2015) and we append the comments from our experiences from similar accident investigation (Prochazkova 2017):

1. The accident of the Alvia Class 730 passenger high-speed train 150/151, modification of Class 130; both train ends are equipped by generator car including the diesel tank. *Note:* just spilled diesel has being the main cause of huge fire in railway accidents.
2. Train 150/151 assembles from 13 vehicles: two power cars appended by two generator cars at both ends; eight passenger coaches; and restaurant car. The train weight was 382 tons. *Note:* just the train's parameters have a direct influence on impacts and severity of the accident.
3. The train was equipped by two security systems, the ASFA Digital and the ERTMS/ETCS. Due to problems in reliability and availability of ERTMS configuration in this railroad, the operator (Renfe Operadora) applied to operate under a trackside signal block (BSL) with the protection of ASFA Digital. Note: just incompatibility of multiple systems from different levels of automation cause disruption of safety integrity, and contribute to serious accidents.
4. The followed rail line is equipped with the BSL device, ERTMS/ETCS level 1, with exception of its beginning and end, and with the back-up ASFA Digital. *Note:* just the transition between various control systems leads often to confusion of staff and to human errors.
5. The low speed curve (max. 80 km/h) with design radius of 402 m is located in the end of rail line solely equipped with ASFA Digital. *Note:* at that time this system did not consider the speed limit and allowed to train to pass at with unacceptable speed, which contributes to the derailment.
6. Along the curve, the massive, concrete wall is. Maximal permitted speed is 80 km/h, which is given on table at railway. *Note*: such tables are not clearly visible in the speed and driver's occupancy, which contributed to the accident.
7. The signals and path for the train was set in position that indicated "track clear". ***Note:*** just the gaps in the automatic control system, which allow incorrect indication of freedom or limitations on the track, caused the accidents.
8. The speed change marker before the curve at kilometre point (PK) 84 + 273 did not warning. *Note:* this seemingly malignant fact has been often the cause of traffic accidents.
9. In the engine driver cabin there were several communications systems (e.g. the radiotelephones between trains; the mobile phones (GSM-R)) for corporate communication in train control system; they were in order. *Note:* service calls and communication with dispatching, along with other train control tasks, especially at changes of control systems, lead to heavy workload of engine drivers and contribute to mistakes.
10. The timetable book for driver showed the speed change: i.e. limitation of speed to 80 km/h at PK 84 + 230 (the Angrois curve). *Note:* this is possible driver's mistake due to oversight.
11. According to regulations, in this site the driver needs to start the slow down (form 200 km/h to 80 km/h), namely without technical supports of control. *Note:* this is a possible driver's mistake.
12. In a given case, the train was delayed, ca 2 up 3 minutes. *Note:* such delay often increases the engine driver stress because he/she needs to adhere the timetable.
13. The records showed that the engine driver reacted to service call of train manager ca 6000 meter before curve beginning and the call lasted 100 seconds. *Note:* this is possible cause, why driver did not react to sign dictating the speed reduction.
14. The technical analysis showed that the train brakes were not sufficiently activated for ensuring the required speed reduction. *Note:* it goes on late reaction of engine driver.

The ERA conclusion stated that the formal investigation of the Commission for investigating railway accidents in Spain (CIAF—Comisión de Investigación de Accidentes Ferroviarios) did not consider all facts at accident root causes determination, because it leaned on narrow view on matter, i.e. the driver always needs to respond well, and he / she cannot rely on notice from security systems, i.e. he / she promptly needs to adjust the train speed to required 80 km/h by braking (EU 2015).

The assessment of traffic accident made on the basis of the ERA conclusions and our notes to 14 domains, shows that the combination of number of mistakes have been occurred in this case, especially the rail traffic safety management system complexity contributed to accident.

From the viewpoint of cybernetic causes given in (Kertis & Prochazkova 2017a), it goes on: distortion of monitoring data—according to driver's testimony, it came to his confusion on current position due this fact; and inadequate HW—in dangerous part the security level was reduced owing to the ETCS, level 1 switching off, which leaded to loss remote control of permitted line speed in that section.

4.3 *Common cyber causes*

From the analysis of above-mentioned railway accidents, their common cybernetic causes are clear. They are mainly connected with problems on systems' interfaces, which are designed, implemented and operated by various subjects on the different level of their responsibilities. In goes on: problems at human-machine interfaces; problems at cyber-physical systems interfaces; problems at socio-technical systems interfaces; problems at determination of responsibilities, namely not only between subjects but also between system processes (mutual compatibility); problems at interfaces among systems with different criticalities (levels of automation, levels of security, safety integrity levels).

These problems correspond to the findings of the analysis of common causes of railway accidents in the Czech Republic (Kertis & Prochazkova 2017a), which are in terms of cybernetics: distortion of monitoring data; false software that does not consider all possible variants of operating conditions; insufficiently robust hardware that causes incorrect or slow processing and evaluation of data. In some cases, there were noticed also hackers' attacks on control centre of dispatcher's workplace. The underlying root cause of accidents is insufficient system decision-making validity, i.e. the low information power and low knowledge of information systems, i.e. their low information amount.

5 METHOD OF CONSTRUCTION OF MEASURES IMPROVING RAIL SAFETY

To reach the highest level of safety of rail control system, i.e. also the information power, all stakeholders need to introduce the TQM approaches (Zairi 1991) with taking into consideration of integral risks. The approaches need to be adapted to specific features of rail systems. Therefore, for work with risks, multi-criteria approaches need to be used. Because some criteria for reaching the safety integrity are conflicting, the optimal solution needs to be determined for certain sufficiently wide extent of conditions can be used to respect the limits (Prochazkova 2013, 2015, Kertis 2015).

6 MEASURES IMPROVING RAIL SAFETY

In the present case, we have adjusted the above-described approach for the rail information systems, which are characterized above. According to (Kertis & Prochazkova 2017b), Figures 2 to 4, with a certain degree of abstraction, show a gradual change of system during the implementation

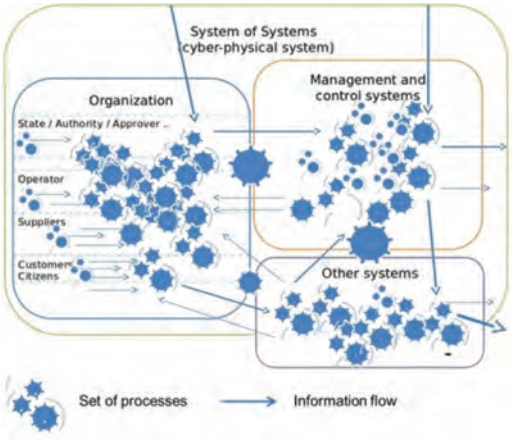


Figure 2. Level of CPS security (Common system of systems).

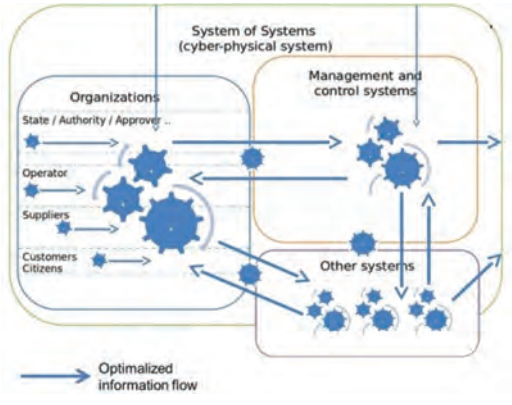


Figure 3. Level of CPS security (an optimized system with higher information power).

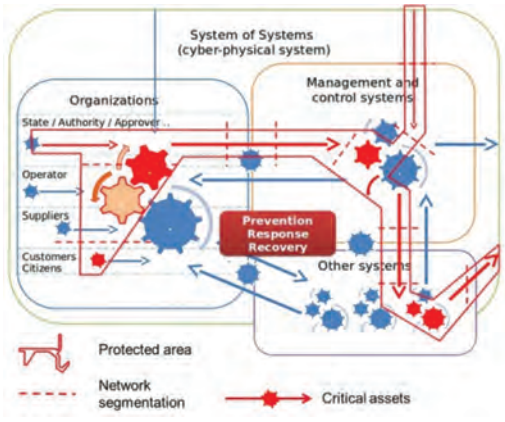


Figure 4. Level of CPS security (the secured system with higher information power).

of various techniques for increase of information power and system security.

Figure 2 shows the open system of systems, in which processes and linkages are implemented to perform defined functions. With consideration of large number of system linkages and interactions of large number of entities involved in the system, under normal conditions, the system performs the required functions, but it is prone to errors, especially in the case of higher operational deviations and imbalances in external system environment.

Figure 3 shows the system of systems with optimized information power that increases the system's durability and maintainability, and by this security. By optimization, i.e. streamlining the flows and improving the information power parameters, it reaches that the system is less prone to internal mistakes in the case of various known deviations in operation and in the surroundings.

For increase of information power and minimizing the resources, which are necessary for these systems formation, a number of methods can be used in practice, e.g.: COBIT for auditing the information systems from the top management point of view (Vitous 2000); ITIL for management of information systems and services, the parts of which are standardized (ISO 2011); refactoring, i.e. changes in software system, which do not influence the external behaviour of information system, but they improve its internal structure (Moos & Malinovsky 2008).

Introducing the management systems (ISO 2015, 2017) with the support of information systems designed using the methods listed above, greatly contributes to improving the information power and transmission matrix parameters according to the formula (4), but only in case if the context of management systems focuses on the rail system as a whole, with a uniform terminology and focus on interfaces of the systems across all stakeholders, i.e. concerned subjects.

Figure 4 shows a well-secure system with optimized information power that we get from the case, shown in Figure 3 by protecting it against significant external and internal influences, i.e. we introduce the preventive and mitigating measures, and prepare reactive measures for case of incidents, as well as also measures for rapid recovery by help verified continuity plans.

In management and governance of railway system and related organizations, the systems and methods according to (ISA 2007, ISO 2009, 20013) are gradually introduced. However, they need to be implemented by all stakeholders and, mainly, the problems associated with system interfaces need to be get over; i.e., it needs to consider the whole information origin process, used information technologies and the quality of their parameters according to Chapter 3.

From this it follows that for safe operation it is very important to deal not only with security of information and technological assets, but also to provide the required information power of the system across all levels and concerned subjects. Parameters of information power are influenced in all sub-processes of information origin process, i.e. the optimization needs to take into consideration each such sub-process and the appropriate information technology listed in Table 2.

Security of such system is then focused on identification and management of important assets. Since, everything cannot be ensured, we need to select critical assets, i.e. critical processes, information flows or other supporting information and physical assets. On the basis of their criticalities, we assess the primary risks and introduce appropriate preventive measures. In case of disaster occurrence (including the cyber-attack), we perform response and recovery according to established policy (Kertis & Prochazkova 2017b, c).

According to the systems of system safety principles, the whole rail system needs to be built according to the Defence-in-Depth approach (Prochazkova 2015) and to introduce different types of safety management to reflect the expected operating conditions of system, and eventually, for severe disasters, to have also way of management that protects also other assets than the assets of system under consideration, in which we monitor physical, organizational and cyber assets.

7 CONCLUSION

Examples of rail accidents presented in the paper point several common cybernetic causes, which

consist in combinations of several errors, mainly in design of safety management system and in its information systems, which should be increased not only the operation economic benefit, but also level of system safety regarding the public assets of European space. With regard to the fact that today's processes of railway management, i.e. the considered safety management and other parts of the railway system (control systems, infrastructure, vehicles and devices) cannot do without information systems, it is very important to find faults also in cyber domain.

Analysis of common cybernetic causes performed on basis of two case studies points to issues connected with systems' interfaces, which are in administration of different subjects, or have different physical nature. On the basis of above mentioned outcomes and knowledge, it is needed to use multi criterial approaches and propose measures for coping with found vulnerabilities, i.e. improving the parameters that heighten the information power and ensuring the information security at critical processes of railway management systems, namely in whole information origin process and at information processing. In the case of application of proposed principles, the work contributes to increasing the railway safety.

ACKNOWLEDGEMENT

Authors thanks to Czech Technical University in Prague for support (grant SGS2015-17).

REFERENCES

ARTEMIS 2014. *Project SESAMO: Security and Safety Modelling.*

CEN-CENELEC 2017. Rail Sector Forum: *Railway in Future*.

ČR MD 2017. *Zpráva o výsledcích šetření příčin a okolností vzniku mimořádné události: Moravany (trať Česká Třebová—Praha—Libeň)*. Praha: Drážní Inspekce (MD).

EU 2020a. *Project CertMILS: Compositional security certification for medium to high-assurance COTS-based systems in environments with emerging threats*.

EU 2020b. *Project CITADEL: Critical Infrastructure Protection Using Adaptive MILS*.

EU 2015. *Advice ERA/ADV/2015-6 OF THE EUROPEAN RAILWAY AGENCY FOR EUROPEAN COMMISSION REGARDING*. Brussels: European Rail Agency.

ISA 2007. *ANSI/ISA–62443-1-1 Security for industrial automation and control systems: concepts, terminology and models.* Washington, DC: ANSI.

ISO 2009. *ISO/IEC 15408-1: Information technology—Security techniques—Evaluation criteria for IT security—Part 1: Introduction and general model.* Geneva: IOS.

ISO 2011. *ISO/IEC 20000: Information technology—Service management.* Geneva: ISO.

ISO 2013. *ISO/IEC 27001: Information technology—Security techniques—Information security management systems—Requirements.* Geneva: ISO.

ISO 2015. *EN ISO 9001: Quality management systems. Requirements.* Geneva: ISO.

ISO 2017. ISO/TS 22163:2017 *Railway applications—Quality management system—Business management system requirements for rail organizations: ISO 9001:2015 and particular requirements for application in the rail sector*. Geneva: ISO.

Moos, P.& Malinovsky, V. 2008. *Information systems and technologies*. Praha: ČVUT.

Novobilsky, P., Kertis, T. & Prochazkova, D. 2016 Cyber security of metropolitan railway communication infrastructure. *Risk of business and territorial processes: 78–91*. Usti nad Labem: FVTM UJEP.

Kertis, T. 2015. *Bezpečnostní plán vybrané stanice pražského metra (Diploma Thesis)*, Praha: CTU, Faculty of transportation sciences.

Kertis, T. 2016. Porovnání přístupů pro řízení bezpečnosti v dopravě. *Rizika podnikových a územních procesů a poznatky pro krizové řízení: 34–59*. Praha: CTU.

Kertis, T.& Prochazkova, D. 2016. Risk management plan for metro station safe operation. *Risk, Reliability and Safety: Innovating Theory and Practice: 1306–1314*. London: CRC Press.

Kertis, T. & Prochazkova, D. 2017a. Railway accidents in the Czech Republic, causes of risks and their mitigation. *Safety and Reliability—Theory and Applications:* 1667–1673. London: Taylor & Francis Group.

Kertis, T., & Prochazkova, D. 2017b. Cyber security of underground railway system operation. *Smart City Symposium Prague*, Prague: CTU Faculty of transportation sciences.

Kertis, T., & Prochazkova, D. 2017c. Information power and cybernetic causes of rail accidents. *Řízení rizik procesů spojených s technickými díly (ŘRTD) 2017.* Praha: CTU Faculty of transportation sciences.

Prochazkova, D. 2012. *Bezpečnost kritické infrastruktury*. Praha: ČVUT.

Prochazkova, D. 2013. *Krizové řízení pro technické obory*. Praha: CTU.

Prochazkova, D. 2015. *Safety of complex technological facilities*. Saarbruecken: Lambert Academic Publishing.

Prochazkova, D. 2017. *Zásady řízení rizik složitých technologických zařízení*. Praha: ČVUT.

Vitous, M. 2000. Cobit 5 v malych a strednich firmach. *IT Systems: specializovany mesicnik o podnikove informatice*. Brno: CCB.

Wikimedia. 2017. https://commons.wikimedia.org/wiki/File: Tragedia_en_Santiago_de_Compostela_(g).jpg.

Zairi, M. 1991. *Total Quality Management for Engineers*. Cambridge: Woodhead Publishing.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Empirical studies of methods for safety and security co-analysis of autonomous boat

Erik Nilsen Torkildson, Jingyue Li & Stig Ole Johnsen
Norwegian University of Science and Technology, Norway

Jon Arne Glomsrud
DNVGL, Norway

ABSTRACT: Many autonomous systems are safety-critical, e.g., autonomous cars, boats, or aerial vehicles. Autonomous systems rely on software and communications. Security vulnerabilities of software and communication will give adversaries possibilities to attack and compromise security and safety. Therefore, when analysing safety, security should be co-analysed. In this study, we explored three safety and security co-analysis methods: Systems-Theoretic Process Analysis (STPA) plus STPA-Security Analysis (STPA-Sec), Failure Mode, Vulnerabilities and Effect Analysis (FMVEA), and Combined Harm Assessment of Safety and Security for Information Systems (CHASSIS). The purpose is to compare applicability, efficiency, and hazards identified by the different methods. An autonomous boat is used as the case study. Results of the study show that STPA plus STPA-Sec and CHASSIS can be more time consuming to use than FMVEA. However, STPA plus STPA-Sec and CHASSIS can help analysers identify more hazards of autonomous systems than FMVEA. Results of the study reveals weaknesses of each method to analyse autonomous systems with different levels of autonomy. We therefore propose possible improvements and combinations of the methods.

1 INTRODUCTION

Autonomous systems like drones, driverless cars, and autonomous boats are being developed. The key mechanism in an autonomous system is its ability to be independent of a human operator. The system manages to sustain situation awareness and decision-making capability, when an expected or unexpected event occurs. By shifting degrees of situation awareness and decision-making responsibilities from humans to the system, we can design autonomous systems with different levels of autonomy. As an example, the Society of Automotive Engineers have described 6 levels of autonomous driving (SAE, 2016) from no automation, driver assistance, partial automation, conditional automation, high automation, to full automation.

Without systematic safety/security analysis and design of autonomous systems, mishaps can happen and harm users and the environment. For example, on 24th July.2015, Fiat Chrysler Automobiles ordered recall of 1.4 million vehicles that was vulnerable to a threat of remote control and hijacking (Guzman, 2015). In 2013, Samy Kamkar demonstrated with the Parrot AR that it was possible to hijack other drones, with what he called SkyJack (Kamkar, 2013). Google self-driving cars had few accidents but was sometimes involved in a rear-end collision with human-driven cars, due to that human drivers did not anticipate actions from the autonomous system (Teoh and Kidd, 2017).

Traditionally, system safety analysis focuses on accidental component failures or software bugs. As industrial and autonomous control systems are increasingly interconnected through networks, system safety can also be compromised by security breaches. "*Although of great importance, it is not sufficient to address accidental threats (hazards) of such systems—also threats of intentional origin need to be covered* (Aven, 2007)." "*Security functions are not meant to cope with physical hazards and failures; likewise, safety functions might not detect and respond to attacks that target the digital components of the system. We infer that safety and security are complementary and should be treated jointly to improve risk management* (Kriaa et al., 2015)."

Several methods have been proposed to combine safety and security analysis of industrial control systems. Some studies have empirically compared different security and safety co-analysis methods using specific systems. For example, FMVEA (Failure Mode, Vulnerabilities and Effect Analysis) and CHASSIS (Combined Harm Assessment of Safety and Security for Information Systems) were compared using an automotive cyber-physical system (Schmittner et al., 2015). The comparison focused

on level of abstraction, comparability of repeated analysis, reusability of analysis artefacts, scope of analysis, suitability for risk rating, and adaptability to changing context. However, we believe that more empirical comparisons of different security and safety co-analysis methods are needed, because many methods are proposed but are not thoroughly evaluated. In addition, few studies have used autonomous systems as cases for evaluation. We have been interested in several methods including the STAMP method (STPA—System-Theoretic Process Analysis) since it has a modern system approach, looking at key control issues.

In this paper, we present an empirical study that compares three security and safety co-analysis methods using an autonomous boat that is under development as the case. The autonomous boat Revolt (www.dnvgl.com/technology-innovation/revolt/index.html) with the present design and sensor fitting is not pure autonomous yet, but a remotely operated dynamically positioned boat. This boat still misses sensors and functions for tracking other objects to be more autonomous. This study is just the first step of analysing safety and security of the autonomous boat. We will follow the development of Revolt and perform re-analysis when new functions are added. Such follow-up analyses will give us insights into different safety and security issues of autonomous systems with different levels of autonomy. Our key focus of our study was to compare *applicability, efficiency, and hazards identified by different methods.* The methods piloted and the sequence of the pilot are 1) FMVEA, 2) STPA plus STPA-Sec, and 3) CHASSIS.

Results of the study show that STPA plus STPA-Sec and CHASSIS are potentially more time consuming than FMVEA. Results also illustrate that different methods have different strengths and weaknesses for identifying different hazards. Based on results of this study, we propose to improve and combine the methods to meet the requirements of security and safety analysis of different autonomous systems.

The rest of this paper is organized as follows. Section 2 defines relevant terminologies. Section 3 introduces the state of the art of security and safety co-analysis, focusing on the three methods we evaluate. Section 4 presents our study design and results. Section 5 discusses evaluation results, and Section 6 concludes.

2 DEFINITIONS

There are many definitions of security and safety. Usually, safety is being used to describe accidental harm, while security is used to describe intentional harm. In (Firesmith, 2003), safety is defined as "*the degree to which accidental harm is prevented, reduced and properly reacted to*", and security is defined as "*the degree to which malicious harm is prevented, reduced and properly reacted to.*" In SEMA reference framework (Pietre-Cambacédès and Chaudet, 2010), safety and security are graphically mapped on a conceptual grid, which has two dimensions. The first dimension distinguishes between accidental and malicious threats. The second dimension differentiates safety and security based on origin and consequences. In the SEMA reference framework, the origin and consequence of safety is system and environment respectively. For security, the origin and consequence could be environment to system, system to environment, and system to system. In (Schmittner et al., 2016), the authors clarified the terminologies to be used for STPA plus STPA-sec analysis as follows. We follow the safety and security related definitions in (Schmittner et al., 2016) in our study.

- *Accident*: Event which causes undesired losses of life, asset damage, data, availability etc.
- *Hazard*: Dangerous system states which can lead to accidents.
- *Threat*: Potential cause of an unwanted incident, which may result in harm to a system and/or environment.
- *Vulnerability*: Weakness of an asset or control that can be exploited by one or more threats.
- *Attack*: Attempt to gain unauthorized access to or make unauthorized use of an asset.

3 STATE OF THE ART

3.1 *Security and safety co-analysis*

Many studies listed in (Kriaa et al., 2015) propose that it is necessary to consolidate the security and safety co-analysis, because security breaches can bring risks to system safety. However, the study (Eames and Moffett, 1999) identifies possible disadvantages of security and safety co-analysis "*We believe that consolidation of safety and security could reduce developers' understanding of the system being analysed, and prevent a thorough analysis of either property*." In addition, the study (Eames & Moffett, 1999) says that "*An additional danger is that a unified approach might actually hide the requirements conflicts that it aims to resolve.*" To address the possible disadvantages, it is critical to closely examine the various kinds of interdependencies between safety and security. Safety–security interactions can be classified into four categories (Pietre-Cambacédès, 2010).

- Conditional dependency: Satisfaction of safety requirements conditions security or vice-versa.

- Mutual reinforcement: Satisfaction of safety requirements or safety measures contributes to security, or vice-versa, thereby enabling resource optimization and cost reduction.
- Antagonism: When considered jointly, safety and security requirements or measures lead to conflicting situations.
- Independency: No interaction.

The security and safety co-analysis methods can generally be classified into three categories (Kriaa et al., 2015). One category is generic approach, such as FMVEA (Schmittner et al., 2014a) and Fault Tree Analysis (Kornecki and Liu, 2013). Another category is model-based graphical methods, such as CHASSIS (Raspotnig et al., 2012) and method using Bayesian Belief Networks (Kornecki et al., 2013). The third category is model-based non-graphic methods, such as STPA (Young and Leveson, 2013) and unified framework (Asare et al., 2013). Autonomous systems are often cyber-physical systems that integrate computation, networking, and physical processes. In addition, autonomous systems need to have proper situation awareness using various sensors, and need to make correct decisions based on the sensor information. Thus, we decided to evaluate one method that is relevant to cyber-physical system in each category mentioned in (Kriaa et al., 2015). We chose FMVEA, CHASSIS, and STPA plus STPA-Sec, because FMVEA and CHASSIS are shown to be applicable to automotive cyber-physical systems (Schmittner et al., 2015), and STPA plus STPA-Sec focuses strongly on software dependent systems.

3.2 *FMVEA*

FMVEA (Schmittner et al., 2014a, Schmittner et al., 2014b) is a FMEA (Failure Mode and Effect Analysis) analysis technique extended with security analysis. FMVEA is based on a three-level Data Flow Diagram (DFD). The first step of the method is to model the system and then to identify failure and threat modes of each component of the system. The failure mode covers the safety aspect, by describing the way the component could potentially fail. The threat mode covers the security aspect, describing the way the component could be potentially misused. The threat modes are based on the STRIDE model, developed by Microsoft (Microsoft, 2002). The STRIDE classification (spoofing/authentication, tampering/integrity, repudiation/non-repudiation, information disclosure/confidentiality, denial of service/availability, elevation of privilege/authorization) enables possible attacks on such components to be found. What is dependent on creating failure and threat modes is knowledge about the system. The potential risks and the effect they could have, are each related to a component (context level).

In addition to identifying vulnerability, threat modes, threat effects, and system effects, FMVEA also tries to quantify the attack probability by estimating system susceptibility and threat properties.

3.3 *CHASSIS*

CHASSIS (Raspotnig et al., 2012) defines a unified process for safety and security assessments. The process includes the use of Misuse Case (MUC) (Sindre and Opdahl, 2005) and Misuse Sequence Diagram (MUSD) (Katta et al., 2010) for visual modelling for security analysis. MUC is also used for safety assessment, but it is combined with Failure Sequence Diagram (FSD) instead of MUSD for detailed failure analysis (Raspotnig and Opdahl, 2012). As shown in Figure 1, there are three stages and 8 steps in CHASSIS. The first stage (steps 1–3) is to draw Use Case and Sequence Diagrams based on some operational and environmental descriptions of the system. In the second stage (steps 4–6), MUC diagrams are created by using a set of hazard and operability study (HAZOP) guidewords (Kletz, 1997) applied for the use cases. The MUC diagrams are then described in textual MUC templates (step 5). FSDs and MUSDs are used to refine the harm scenarios defined in the templates (step 6). When the textual Misuse cases are finished, HAZOP tables are prepared (step 7) and corresponding safety or security requirements are defined (step 8).

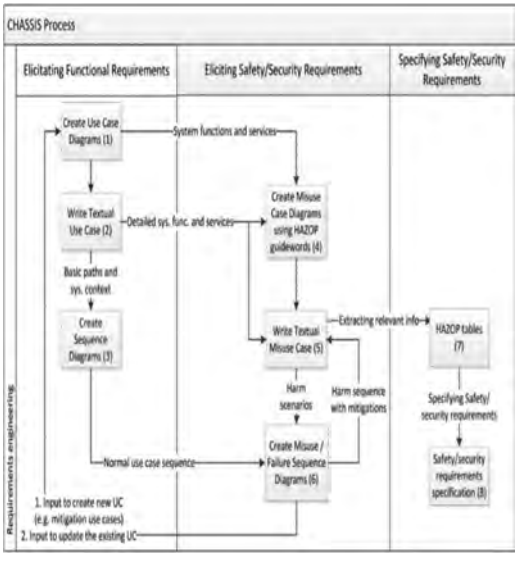


Figure 1. CHASSIS' unified process.

3.4 *STPA and STPA-Sec*

STPA-Sec (Young and Leveson, 2013, Young and Leveson, 2014) extends STPA, which is a safety analysis method (System-Theoretic Process Analysis) (John P., 2013, Leveson, 2012). The extension is to includes security analysis. STPA-Sec "*Shifts the focus of the security analysis away from threats as the proximate cause of losses and focuses instead on the broader system structure that allowed the system to enter a vulnerable system state that the threat exploits to produce the disruption leading to the loss (Young and Leveson, 2013)*."

The main steps of STPA plus STPA-Sec are:

- Identifying what essential services and functions must be protected or what represents an unacceptable loss.
- Identifying system hazards and constraints.
- Drawing the system control structure, physical hardware and network structure, and identifying unsafe control actions.
- Determining the potential causes of the unsafe control actions. The potential causes could be security vulnerability and threats. To facilitate the security analysis, some guide words like tampered feedback, injection of manipulated control algorithm, and intentional congestion of feedback path, are added (Schmittner et al., 2016).

Compared to other security analysis methods, STPA-Sec does not focus on countermeasures that should be taken. STPA-Sec focuses mainly on identifying those scenarios that could lead to losses.

4 STUDY DESIGN AND RESULTS

4.1 *Scope: Autonomous boat*

The autonomous boat Revolt shown in Figure 2 was made by Stadt Towing Tank (STT), on a mission from DNVGL in 2014. The model is a 1:20 scale model of the concept ship. The model ship has a length of 3 meters and weighs 257 kg.

Although Revolt is still under development and is not a fully autonomous boat, we still want to use it as a case since it gives us the opportunity to explore hazard and threats of two main issues i.e. 1) Safety and security of autonomous steering of the ship (i.e. losing control; ship damaged/ destroyed) and 2) security of data-communication between onshore and offshore (sensitive data compromised).

4.2 *Security and safety co-analysis using FMVEA*

The FMVEA analysis focuses on the embedded computer. The attack surface is the highest for

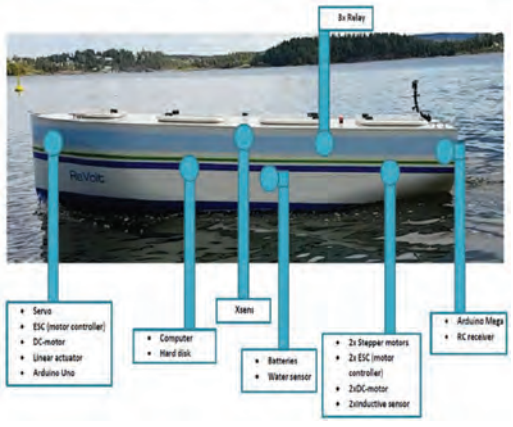


Figure 2. Overview of Revolt and its components.

the embedded computer in the Revolt, since other components in some way are connected to it. Microcontrollers are connected to (and controlled by) the embedded computer via USB. Analogue components (water sensor etc.) are connected to the microcontrollers.

To perform the FMVEA analysis, we fill in the table as proposed in (Schmittner et al., 2014a). The table includes columns for qualitative safety and security analysis, such as component, failure mode, threat mode, failure effect, threat effect, system status, system effect. The table also includes columns, such as severity, system susceptibility, treat properties, attack/failure probabilities, and risks, for quantitative analysis and for ranking the hazards.

4.3 *Security and safety co-analysis using STPA plus STPA-Sec*

When performing STPA plus STPA-Sec analysis, we start with the following unacceptable losses/ accidents and safety constraints.

- Collision with vessels, objects, humans/mammals, structures, grounding
- Fire or explosion
- Foundering (sinking, failing or plunging)
- Loss of cargo
- Loss of mission objectives
- Loss of information

Then, we read the network structure and the control structure documents of the boat to identify unsafe control actions. We follow the systematic method proposed in (John P., 2013) and enumerate full combinations of possible values of process variables and evaluate where control actions can be unsafe if the control action is given, is not given, is given too early or too late, too large or too

small value. The control actions (CAs) we analyse include:

- CA1: Control the position of the vessel
- CA2: Control the speed of the vessel
- CA3: Control the course of the vessel
- CA4: Control the access to the vessels system

After identifying the Unsafe Control Actions (UCA), the last step of the analysis is to identify possible causal factors of the UCA, including possible security breaches that can lead to the UCA. In this last step, STPA-Sec analysis is applied by using the guide words proposed in (Schmittner et al., 2016).

4.4 *Security and safety co-analysis using CHASSIS*

To perform CHASSIS analysis, we first identify use cases and draw use case diagrams. The use case we focus on is "operating and monitoring Revolt remotely through the Revolt Intelligent System (RIS)". Then we make security and safety misuse case through using the HAZOP keywords proposed in (Schmittner et al., 2015). Examples of the safety and security misuse cases are shown in Figure 3.

4.5 *Comparisons of effort spent on co-analysis*

The inputs to the methods are very different. FMVEA analysis focuses on components. STPA plus STPA-Sec analysis focuses on control actions. CHASSIS analysis focuses on use cases. Thus, it is difficult to have direct comparisons of the effort spent on applying the methods. However, by analysing the hours spent on each activity shown Table 1, we can still observe that STPA plus STPA-Sec and CHASSIS can be more time-consuming than FMVEA, because more activities are included and each activity requires more effort.

4.6 *Comparisons of safety hazards identified*

Like comparisons of effort, it is difficult to perform direct comparisons of safety issues identified by using different methods, because the methods have different inputs. However, through comparing safety issues identify by each method, we can observe strengths and weaknesses of each method. FMVEA helps us identify mostly the hazards that are related to single component failure, e.g., communication connection is lost or updates fails. The input of FMVEA does not require as many inputs as the two other methods. It requires only a list of components of the system and how they are connected. This is an advantage. However, it may also be a restriction for early analysis, because early system development might not have a system design.

Compared to FMVEA, STPA plus STPA-Sec method helps us identify more hazards that are related to interactions between different components or actors. STPA is a top down approach that

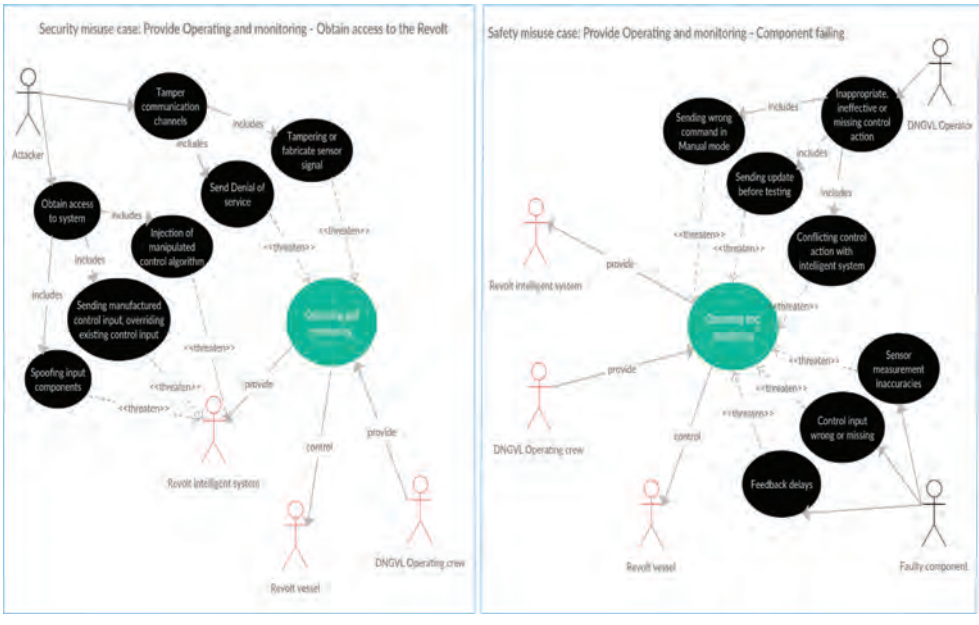


Figure 3. Examples of security and safety misuse cases.

Table 1. Effort spent on each method.

FMVEA		STPA and STPA-Sec		CHASSIS	
Activities	hr.	Activities	hr.	Activities	hr.
System level analysis	5	Define unacceptable losses	7	Elicitation functions and services	4
Selection of component	3	Identify hazards and safety constraints	5	Use case diagram	9
Identify functions of component	10	Create functional control structure	30	Safety misuse case diagram	9
Failure Mode, Vulnerabilities and Effect Analysis	27	Identify hazardous control actions	20	Security misuse case diagram	9
Risk assessment	10	Identify causal factors and scenarios	20	Final misuse case with mitigations	15
		Identify mitigations	23	Misuse Sequence Diagram	8
				Failure Sequence Diagram	8
				Fill in HAZOP table	20,5

looks at the operative picture and identifies unsafe system operation. STPA analysis covers not only the physical system, but also human operators and actors. One hazard example identified by STPA is "setting route for shipment and launch position when the shipping dock has not permitting the action, because other ships are dispatching at the same time". STPA can identify such hazards due to its use of process model variables to identify hazardous control actions, which generates scenarios that otherwise would be omitted. Another advantage of STPA is that it does not assume or need the fully designed observability in the system from the beginning, and this can possibly be achieved through several iterations. We usually find out some UCA based on the preliminary design of the system. When we explore casual factors of UCA, we may find new constraints or requirements for observability and control to handle the identified accident causes, or new need to obtain proof that the accident causes will not practically occur. However, the challenge of STPA plus STPA-Sec analysis (John P., 2013) is that it relies heavily on enumerating process control variables. If many process control variables are present, the analysis can be time consuming.

Compared to FMVEA and STPA, the strength of CHASSIS is that it helps us find hazards that are related to operation sequences. One example hazard identified by using CHASSIS is "the operator performs operations on the Revolt before having done security and safety procedures, and the Revolts components are having feedback delays and commands are executed too late". The weakness of CHASSIS is that it relies more on expert judgement than FMVEA and STPA. As observed in (Schmittner et al., 2015), the possible risk could be that "*if a CHASSIS analysis is repeated by a new group, due to the differences in the experts, new viewpoints can be introduced that change the results*." A restriction of CHASSIS we identify is that its starting point is the use case. If the use case is too broad, steps that follows in the process might be difficulty to perform.

4.7 *Comparisons of security issues identified*

FMVEA security analysis uses STRIDE classification. The identified security threats are limited to threat targeted at single component, e.g., wireless connection is targeted to jamming. When using FMVEA, the safety and security analysis can be done independently. Thus, safety–security interactions may be overlooked.

CHASSIS identifies threats and vulnerability using misuse cases. The hazards identified by CHASSIS are mostly related to operation and use of the system, e.g., the communication system might have vulnerabilities that could lead to modification of system files. By integrating security misuse cases and safety misuse cases, it is possible to analyse safety–security interactions. However, like the safety analysis, the possible weakness of CHASSIS is that it relies heavily on expert knowledge. Thus, the analysis results may not be replicable.

The security analysis of STPA-Sec focuses on identifying security vulnerabilities that may lead to unsafe control actions. For example, providing CA2 (control the speed of the vessel) too late from shore to the boat when the WIFI connection is jammed. Comparing to FMVEA, the strength of STPA plus STPA-Sec is that it focuses more on

safety–security interactions. However, the limitation of the STPA-Sec is that the security analysis focuses mainly on vulnerability that can be the casual factors for safety hazards. The security vulnerabilities, which may lead to information leakage or privacy issues, but will not lead to safety hazards, may be overlooked. The study (Schmittner et al., 2016) proposes to enhance STPA-Sec with more focus on losses related to confidentiality. In our study, we list "loss of information" as an accident and find out some threats that can lead to this loss. However, STPA plus STPA-Sec method use enumeration of process control variables to identify possible information loss. If certain security vulnerabilities, e.g. improper encryption of stored data, are not reflected directly in existing process control variables, the vulnerabilities may not be identified in the analysis. Thus, we believe that integrating STPA-Sec with more security oriented analysis methods, e.g., misuse cases or threat modelling, can be beneficial.

5 DISCUSSIONS

5.1 *Comparison with related studies*

The study (Schmittner et al., 2015) compared FMVEA and CHASSIS. Our study included STPA and STPA-Sec in the comparison.

- Level of abstraction: CHASSIS is a quite high-level approach. It can be applied in early requirement and concept phase, when a system is not clearly defined and little information is known. In contrast, FMVEA needs at least a list of system elements and connections between the elements to generate meaningful results (Schmittner et al., 2015). STPA plus STPA-Sec requires information of the hardware, the network nodes, the network input/output lists to identify process control variables and unsafe control actions.
- Replicable analysis results: CHASSIS depends more on expert knowledge. In contrast, FMVEA and STPA will more likely provide comparable results, even if the analysis is performed by different persons.
- Reusability of analysis artefacts: All three methods use guidewords. FMVEA uses failure modes and STRIDE classification. STPA plus STPA-Sec uses guide words proposed in (Schmittner et al., 2016). CHASSIS uses HAZOP keywords. In all three methods, the quality and completeness of the keywords will strongly influence the quality of the analysis.
- Scope of analysis: FMVEA and STPA plus STPA-Sec depend to a higher degree on the accuracy of the system model and control structure. For CHASSIS, "*it is possible to expand the consideration of risk scenarios which do not arise directly from the system model* (Schmittner et al., 2015)."
- Suitability for a risk rating: FMVEA targets at rating the risks. STPA plus STPA-Sec and CHASSIS focus mostly on generating a list of possible safety and security issues rather than rating them. A possible combination of the method is to perform STPA plus STPA-sec or CHASSIS analysis to identify hazards and then use FMVEA for quantitative comparisons of certain hazards.
- Adaptability to changing context: It is easier for CHASSIS to consider different usage scenarios and changing environment than FMVEA, because CHASSIS is less formal and focuses on high level analysis. STPA plus STPA-Sec and FEMVA analyses results need to be updated when the system design changes.

5.2 *Applicability of the methods for analysing autonomous systems*

Autonomous systems have different levels of autonomy. Based on our observations of strengths and weakness of the three methods, we propose applying different methods for analysing systems with different levels of autonomy.

- For systems with high automation, STPA plus STPA-Sec may be more applicable than FMVEA to analyse interactions between systems, and interactions between systems and environment.
- For systems with many sensors, STPA plus STPA-Sec may be more applicable than CHASSIS. CHASSIS focuses on sequential messages. In contrast, STPA deals with fusions of sensor messages that come at the same time better. However, STPA plus STPA-Sec also needs to be improved. The current STPA proposed in (John P., 2013) is limited to analyse single control action. For many cyber-physical systems and autonomous systems like autonomous boat, some control actions are mutually dependent and might be issued in pairs. For example, in emergency cases, the boat needs to change course and slow down at the same time to avoid collision. Our solution for analysis mutually dependent control actions is to add the control action as a process control variable of another control action, if another control action has dependency with it. For example, in the table to analyse control action CA3 (i.e. control the course), the CA2 (i.e. Control the speed) is added as process control variables with values "speed up" and "slow down".
- For autonomous system with high level intelligence and learning capability, none of the three methods will be sufficient. AI will make it harder to review the system due to its increasing "black

box" and "black code" nature and its learning capability. For those systems, STPA plus STPA-sec or CHASSIS analysis may outperform FMVEA, because the operational level is the same regardless of system implementation. STPA and CHASSIS are good at analysing the operational safety with the system interaction. For autonomous systems with learning capability, however, it is necessary to have continuous verification along with the learning.

5.3 *Limitations of the study*

One main limitation of this study is that the safety and security hazards identified by this study may not be complete. It is because the completeness relies much on the domain knowledge and the guide words. However, the purpose of the study is to compare the three methods rather than to identify all hazards of the system. We believe that, even if other researchers identify slightly more security and safety hazards than us or identify different hazards from Revolt, our observations of the main differences of the three methods are still valid.

6 CONCLUSIONS AND FUTURE WORK

Many security and safety co-analysis methods have been proposed from academia and industry. However, few empirical studies have been performed to compare and evaluate the methods. In this study, we have evaluated three methods using an autonomous boat, called Revolt, as a case study. Results of the study show advantages and disadvantages of each method. Our future study is to extend and strengthen existing methods to analyse safety and security issues of intelligent and complex control actions of autonomous systems. In addition, we need to check the validity of the method, based on observing performance and incidents of the Revolt system.

ACKNOWLEDGEMENT

This work is supported by the SAREPTA (Safety, autonomy, remote control and operations of industrial transport systems) project, which is financed by Norwegian Research Council with Grant No. 267860.

REFERENCES

Asare, et al.. (2013) FSTPA-I: a formal approach to hazard identification via system theoretic process analysis. P*roceedings of the ACM/IEEE 4th International Conference on Cyber-Physical Systems.* Philadelphia, Pennsylvania, ACM.

Aven, T. (2007) A unified framework for risk and vulnerability analysis covering both safety and security. *Reliability Engineering & System Safety,* 92, 745–754.

Eames, D.P. & Moffett, J. (1999) The Integration of Safety and Security Requirements. In Felici, M. & Kanoun, K. (Eds.) Computer Safety, Reliability and Security: 18th International Conference, SAFECOMP'99 Toulouse, France, September 27–29, 1999 Proceedings. Berlin, Heidelberg, Springer Berlin Heidelberg.

Firesmith, D. (2003) Common Concepts Underlying Safety, Security, and Survivability Engineering. Carnegie Mellon University.

Guzman, Z. (2015) Hackers remotely kill Jeep's engine on highway.

John P., I., Thomas (2013) Extending and automating a systems-theoretic hazard analysis for requirements generation and analysis. Massachusetts Institute of Technology.

Kamkar, S. (2013) SkyJack, http://samy.pl/skyjack/.

Katta, V., Karpati, P., Opdahl, A.L., Raspotnig, C. & Sindre, G. (2010) Comparing Two Techniques for Intrusion Visualization. In Van Bommel, P., Hoppenbrouwers, S., Overbeek, S., Proper, E. & Barjis, J. (Eds.) The Practice of Enterprise Modeling: Third IFIP WG 8.1 Working Conference, PoEM 2010, Delft, The Netherlands, November 9–10, 2010. Proceedings. Berlin, Heidelberg, Springer Berlin Heidelberg.

Kletz, T.A. (1997) Hazop—past and future. R*eliability Engineering & System Safety,* 55, 263–266.

Kornecki, A. & Liu, M. (2013) Fault Tree Analysis for Safety/Security Verification in Aviation Software. Electronics, 2, 41.

Kornecki, A.J. et al. (2013) Studying interrelationships of safety and security for software assurance in cyber-physical systems: Approach based on bayesian belief networks. 2013 Federated Conference on Computer Science and Information Systems.

Kriaa, S. et al. (2015) A survey of approaches combining safety and security for industrial control systems. *Reliability Engineering & System Safety*, 139, 156–178.

Leveson, N.G. (2012) *Engineering a Safer World: Systems Thinking Applied to Safety,* MIT Press.

MICROSOFT (2002) The STRIDE Threat Model.

Piètre-Cambacédès, L. & Chaudet, C. (2010) The SEMA referential framework: Avoiding ambiguities in the terms "security" and "safety". *International Journal of Critical Infrastructure Protection*, 3, 55–66.

Piètre-Cambacédès, L. (2010) Des relations entre sûreté et sécurité. (The relationships between safety and security).

Raspotnig, C. & Opdahl, A. (2012) Supporting Failure Mode and Effect Analysis: A Case Study with Failure Sequence Diagrams. In Regnell, B. & Damian, D. (Eds.) Requirements Engineering: Foundation for Software Quality: 18th International Working Conference, REFSQ 2012, Essen, Germany, March 19–22, 2012. Proceedings. Berlin, Heidelberg, Springer Berlin Heidelberg.

Raspotnig, C. et al. (2012) A Combined Process for Elicitation and Analysis of Safety and Security Requirements. In Bider, I., Halpin, T., Krogstie, J., Nurcan,

S., Proper, E., Schmidt, R., Soffer, P. & Wrycza, S. (Eds.) *Enterprise, Business-Process and Information Systems Modeling.* Berlin, Heidelberg, Springer Berlin Heidelberg.

SAE (2016) SAE International standard "J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems.".

Schmittner, C. et al. (2014a) Security Application of Failure Mode and Effect Analysis (FMEA). In Bondavalli, A. & Di Giandomenico, F. (Eds.) Computer Safety, Reliability, and Security: 33rd International Conference, SAFECOMP 2014, Florence, Italy, September 10–12, 2014. Proceedings. Cham, Springer International Publishing.

Schmittner, C. et al. (2014b) FMVEA for Safety and Security Analysis of Intelligent and Cooperative Vehicles. In Bondavalli, A., Ceccarelli, A. & Ortmeier, F. (Eds.) Computer Safety, Reliability, and Security: SAFECOMP 2014 Workshops. Florence, Italy, September 8–9, 2014. Proceedings. Cham, Springer International Publishing.

Schmittner, C. et al. (2015) A Case Study of FMVEA and CHASSIS as Safety and Security Co-Analysis Method for Automotive Cyber-physical Systems. Proceedings of the 1st ACM Workshop on Cyber-Physical System Security. Singapore, Republic of Singapore, ACM.

Schmittner, C. et al. (2016) Limitation and Improvement of STPA-Sec for Safety and Security Co-analysis.

Sindre, G. & Opdahl, A.L. (2005) Eliciting security requirements with misuse cases. *Requirements Engineering*, 10, 34–44.

Teoh, E.R. & Kidd, D.G. (2017) Rage against the machine? Google's self-driving cars versus human drivers. *Journal of Safety Research,* 63, 57–60.

Young, W. & Leveson, N. (2013) Systems thinking for safety and security. *Proceedings of the 29th Annual Computer Security Applications Conference.* New Orleans, Louisiana, USA, ACM.

Young, W. & Leveson, N.G. (2014) An integrated approach to safety and security based on systems theory. Commun. ACM, 57, 31–35.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

An overview on the obsolescence of physical assets for the defence facing the challenges of industry 4.0 and the new operating environments

V. Gonzalez-Prida
University of Seville, Seville, Spain
UNED, Madrid, Spain

J. Zamora
UNED, Madrid, Spain

A. Crespo Márquez
University of Seville, Seville, Spain

L. Villar-Fidalgo
UNED, Madrid, Spain

A. De la Fuente, P. Martínez-Galán & A. Guillén
University of Seville, Seville, Spain

ABSTRACT: This contribution is intended to observe special features presented in physical assets for defence. Particularly, the management of defence assets has to consider not only the reliability, availability, maintainability and other factors frequently used in asset management. On the contrary, such systems should also take into account their adaptation to changing operating environments as well as their capability to changes on the technological context. This study approaches to the current real situation where, due to the diversity of conflicts in our international context, the same type of defence systems must be able to provide services under different boundary conditions in different areas of the globe. At the same time, new concepts from the Industry 4.0 provide quick changes that should be considered along the life cycle of a defence asset. As a finding or consequence, these variations in operating conditions and in technology may accelerate asset degradation by modifying its reliability, its up-to-date status and, in general terms, its end-of-life estimation, depending of course on a diversity of factors. This accelerated deterioration of the asset is often known as "obsolescence" and its implications are often evaluated (when possible), in terms of costs from different natures. The originality of this contribution is the introduction of a discussion on how a proper analysis may help to reduce errors and mistakes in the decision-making process regarding the suitability or not of repairing, replacing, or modernizing the asset or system under study. In other words, the obsolescence analysis, from a reliability and technological point of view, could be used to determine the conservation or not of a specific asset fleet, in order to understand the effects of operational and technology factors variation over the functionality and life cycle cost of physical assets for defence.

1 INTRODUCTION

Throughout history, technological advances have always been applied in armed conflicts to allow certain superiority against the enemy. Such conflicts have served in many cases as fields of experimentation to validate progress that, subsequently, have had their application in the civilian world and vice versa. Identically, today's conflicts apply technology, where the digital transformation becomes more and more present. In fact, there is a historical constant where the same means used to create wealth, are those used for warfare, and vice versa [1]. I.e., there are dual-use technologies, so that the developments and commercial innovations take advantage to the military sector and vice versa. To this fact, it must be added that enemies can also have these high-tech devices easily.

Many functions from gadgets that emerged before, keep its purpose now, adapting them to the new circumstances, both operational and technological. In the case for example of an armoured vehicle, the basic characteristics that emerged during the First World War (protection, mobility and

firepower), are the same that are expected currently, though, with the own technological advances of today. These basic functions are seen nowadays in continuous development and improvement according to the new technologies that are emerging to meet specific operational needs. To such functions it is added now a fundamental property as the reliability of the systems themselves, as well as other concepts such as availability, maintainability and safety. This contribution deals with these technical characteristics from the standpoint of a military asset, introducing terms such as obsolescence and useful life of these systems, and also relating the effect on these assets and their characteristics of new technological tools included under the concept of industry 4.0.

2 CONCEPTS OF ASSET MANAGEMENT AND ITS ADAPTATION TO THE CASE OF DEFENSE

2.1 *RAMS parameters of an asset and its life cycle cost*

Reliability, availability and maintainability are parameters used in assets management whose analysis is usually known by the acronym in English of RAM. Frequently, it is added the S of security and/or safety, although some authors referenced it as sustainability, and hence the acronym results as "RAMS analysis".

- Reliability: capacity of the system not to fail, i.e., the probability that the system complies with what is expected of it
- Availability: proportion of time that the asset is useful to be used (in principle, it does not have to be operational, although it can be considered also an operational availability of the asset)
- Maintainability: ease of the asset to keep its normal operation, i.e. it would be inversely proportional effort in maintenance activities
- Safety/security: features set of measures that are taken to avoid and prevent accidents, or protect from illegal activities

Apart from the a. m. concepts, stricter definitions can be found in the references [2], [3] or [4]. These terms are closely linked to the concept of useful life, which is associated with the time during which the system continues fulfilling its functions [5]. Over the useful life of an asset, it must maintain and keep its value. Each of the stages of the life cycle of an asset will have some associated costs being finally life cycle cost the sum of all these costs. In other words, a life cycle cost analysis must take into account: costs initial acquisition of the physical asset (covering the costs of development and investment); operational costs; maintenance costs (planned, corrective as well as overhauls); and the costs associated with the divestiture of assets or dismantle the installation. If the asset is still used beyond than expected, this latter term, instead of a cost may be considered a Residual value (if for example is sold to a third party). All of the above are often treated from the standpoint of an industrial physical asset, which tend to have a certain operating profiles, to a greater or lesser extent, under controlled environments.

From the perspective of a military asset, they will find themselves under the paradigm of having to deploy to different environments, with changing mission profiles and where logistics is critical to maintain its performance. I.e. when a machine is acquired for a production process, this process can be more or less predictable or stable, while in the case of an asset for defence, it should be added the uncertainty of mission (surveillance, defence, humanitarian support… and other profiles), the conditions of operation that will be used (deserts, icy areas, forest, urban locations...), and of course technological change. Consequently, the estimation of lifespan for an industrial asset does not have to coincide in a military system. This end of life of an asset is known as obsolescence, which can slow down if redesigns and updates throughout the life of the asset are considered.

2.2 *Obsolescence and modernization of an asset*

The term "obsolescence" was used for the first time in relation to basic products [6], [7]. However, it has been employed in the industrial environment [8], [9], relating to functional factors (changes in use), economic (cost of continuing to use it, regard the cost of replacing it with an alternative), technological) efficiency of technology assets, in comparison with the new alternatives available), or social (trends of users, changes in legislation or regulations of health and safety...) [10], [11]. These asset shifts changed its durability and obsolescence [12]. Therefore, there are tangible factors, such as the functional and economic factors more related to the depreciation of the assets, and intangible factors such as technological and social, more subjective or prospects-related to the market [13].

Accordingly, maintenance activities are linked to the functional and economical obsolescence, preserving the value of the asset in a physical sense and combating the economic depreciation of own assets. In this sense, there is a time-dependent relationship between the obsolescence and the analysis based on reliability according to functional and economic factors [14] (life cycle costs). In other words, maintenance should be assessed by comparing the assets value regarding the cost involved

in repair or replacement. In this analysis, it may be the case that the decision to repair the asset results in disproportionate expense to preserve the value of the investment [15]. On the other hand, the modernization activities will be more closely linked to combat social and technological obsolescence.

In both cases, an improvement in maintenance and/or the possibilities of a modernization has to go exclusively through direct changes on the asset itself. It can be also related to assure the execution of tasks, training, spare parts and tools distribution, logistics required for the coordination of all aspects etc... allowing the system to be available and in operation as long as possible, maintaining and implementing its performance during its lifetime. I.e. new technologies not only can and must influence improvements on own assets, but also on the integrated logistics support provided to it.

3 INNOVATION IN MILITARY ASSETS

As seen in the previous section, with a view to slowing down the obsolescence of a military asset, it can be considered incorporating modernizations (on the assets and/or on its logistical support) taking advantage of the new technologies that now are at our disposal. It is important to emphasize that these advantages are not only applicable to the own assets, but they can facilitate and improve all those activities that surround system and needs that are essential to keep it in "life". With that intention, it is relevant to deal with the industry 4.0 concept, providing a panoramic view of possible applications in the military assets, focusing on the possibilities in terms of logistic support integrated with a view to improving precisely the reliability, availability, maintainability and safety of assets for the defence.

3.1 *Evolution to the "4.0"*

The origin of the term industry 4.0 refers to the fourth Industrial Revolution, understanding that the first arose when machine steam at the end of the XVIII century, the second when we implemented the use of electricity and manufacture in series in the last third of XIX century, and the third Industrial Revolution to that when automate factories began already in the s. XX century. Today, the fourth Industrial Revolution has to do with the digital transformation applied to the industry in the search for connectivity and operational excellence. It was named for the first time in a study conducted in Germany in 2011: "Smart Manufacturing for the Future", Germany Trade and Invest [16]. Associated with the term industry 4.0 are usually define 9 technologies such as the ones shown in the following table (Table 1).

Apart from these nine technologies, sometimes are added some others such as vehicles or autonomous or unmanned aircraft (drones), new materials (Graphene), artificial intelligence, digital platforms... having their place in the nine

Table 1. Technologies associated with the industry 4.0 concept.

Technology	Description
Robotics and computer vision	• Cooperative and autonomous robots. • Numerous integrated sensors and interfaces standardized.
Additive manufacturing and 3D scanner	• 3D printing, especially for prototypes and spare parts. • Decentralized 3D installations to reduce inventory and transportation distances. • Flexibility of forms, quickly not to use tools, cost savings
Augmented and virtual reality	• Augmented reality for maintenance, logistics and all kinds of operational procedures. • Supporting information display, for example, through smart glasses.
Simulation and modelling	• Simulation of value networks (flows). • Optimization based on data in real time from intelligent systems.
Horizontal and vertical integration	• Integration of data between companies based on standards of data transfer. • Precondition for a fully automated value chain (the company customer, the plant management)
Industrial Internet (IoT. Internet of things)	• Network of products and machines. • Multi-directional communication among objects in the network. • Cyber-physical Systems
Cloud computing	• Management of huge volumes of data in open systems. • Communication in real time for production systems.
Cybersecurity	• Operation in networks and open systems. • High level of networking between machines, products and intelligent systems.
Big data and data analysis	• Complete evaluation of the available data (for example, ERP, SCM, CRM, and data of machine). • Support and optimization in real time for decision-making

mentioned categories. All previous technologies have applications without a doubt in the military sphere. In this sense, one of the first objectives of digitization can be precisely:

- To support the concept of life cycle of military assets.
- Determine the needs of services and infrastructures in the evolution of the means for the defence.
- Study the incorporation of management software mass of data that allow the use of artificial intelligence methods.
- Evaluate trends.
- Get lessons learned.
- Control and manage the asset lifecycle and engineering.

In general, the inclusion of the concept of logistics support 4.0.

I.e. apart from the operational improvements that may benefit the assets themselves, first advantage is that assets management can be improved for the defence in maintenance (corrective action and scheduled inspections), in the management of spare parts and material (supply of spares, tools and consumables in time and place), in the adaptation of systems to operational requirements, control of assets configuration (design modifications) and, in general terms, the determination, evaluation and improvement of support along the life cycle through the support updating capabilities of the army.

3.2 *Technological trend of Defence 4.0*

The first half of the XXI century is characterized as a period of great political, social and ecological changes (translation of the geopolitical focusses, increase of the world population, consumption of large amounts of resources, effects on the biosphere...). All this happens in a highly globalized world and parallel to the highest scientific and technological transformation of history [17]. This context implies complex scenarios with increased uncertainty where the effect of technology cannot be ignored. Recent times correspond to the most technologically advanced in the history of mankind, driven by the digital revolution, but also by biology and nanotechnology [1]. In the digital realm, a growing convergence of cyber, physical and biological domains appear with more complex and higher value-added products and services. Therefore, at conventional operating environment: land, sea and air; it must be added now the virtual or cyber space (Figure 1).

In general terms, it is found an increasingly globalized combat environment more connected and with greater presence of digitized and automated means that (making a comparison with the industry) could be called "Defence 4.0". All this obliges the armies to a better adaptation to the changes because there are factors that radically transform the character of the war fighting. The Defence sector must therefore nourish of the digital revolution in aspects such as connectivity in real time (internet of things, "low cost" sensing), the collaborative robotic (drones, autonomous vehicles, 3D printing, exoskeletons), the Automation (augmented and virtual reality, artificial intelligence, Big Data, processing in the cloud...).

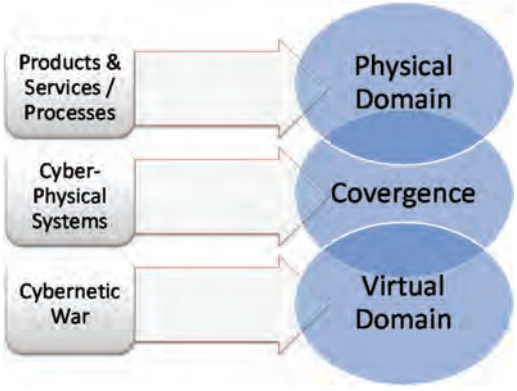


Figure 1. New operational environments.

The application of these technologies must take into account the platform on which it is incorporated (aircraft, vehicles, weapons, general equipment systems...), means for implementation (software, hardware...), as well as logistics and tactical elements (equipment of support, training and simulation systems, digital stands for technical documentation, mission planning systems...) [18].

4 CONCLUSIONS

Today, advances in technology and changes in operation environments oblige the designs of assets for the defence to be flexible and to evolve with the life of the system. In other words, the conservation of a fleet of military assets updated and the fulfilment of operational demands require redesigns and upgrades throughout the life of the items, as well as to streamline everything related to their logistical support. In order to emphasize those scientific issues addressed along this paper, it is relevant to explore the current willingness of industries, together with universities and research centers to collaborate and constitute a kind of community of interest in the area of maintenance [19], logistics support [20] and, in general terms, the integration of technological solutions [21], in

order to overcome the problem of obsolescence in the defense sector.

As it has been observed throughout this document, the application of technologies associated with the industry 4.0 concept, is without doubt a great help to increase the usefulness and efficiency of assets, improving both its operation and maintenance. For example, large data collection allows scheduled maintenance according to the different mission profiles, reducing costs by preventing unnecessary actions, and tasks etc. All intended to a result where it is increasing easiness, flexibility and immediacy of the spare parts, an improvement of maintenance policies, customizing them, as well as other aspects which, in summary, affect the parameters of assets reliability, availability, maintainability and safety.

As future lines of action and research, it is suggested the development of unified and simplified protocols for analysis and work processes. Depending on the case, it will be desirable to observe the applicability of methods of artificial intelligence (Machine Learning, Deep Learning), framed initially (for simplicity) in the concept of Integrated Logistic Support 4.0 for the establishment of decision-making processes, flow data and organizational structure that can certainly be implemented in the own assets ILS. As an academic added value, this paper stands aligned with the principles established by the current international defense sector, in the search for greater logistic efficiency [22] through an interconnected system among the members (Armies, Governmental Organizations and Companies from the allied countries), where innovation is encouraged and promoted [1], [16]. Finally, note that having operational armed forces requires today an advanced and competitive technological and industrial manufacturing. It should be a priority strategic objective for the industrialization and modernization of the defence sector.

REFERENCES

[1] Sanjurjo J.M. (2016). El futuro tejido industrial. XX Congreso Internacional de dirección e Ingeniería de Proyectos. Cartagena, julio 2016.

[2] UNE-EN 15341 (2007). Indicadores principales de desempeño de Mantenimiento. European Standard. CEN (European Committee for Standardization), AEN/CTN, INGEMAN, España.

[3] Jardine A. (1999). Measuring maintenance performance: a holistic approach. *International Journal of Operations and Production Management,* 19(7): 691–715.

[4] Parra Márquez C.A., Crespo Márquez A. (2015). Ingeniería de Mantenimiento y Fiabilidad aplicada en la Gestión de Activos. INGEMAN (Asociación Española para el Desarrollo de la Ingeniería de Mantenimiento). ISBN: 978-84-95499-67-7.

[5] Campbell JD, Jardine AKS. (2001). Maintenance excellence. New York: Marcel Dekker.

[6] Bulow, J.I. [1982]. Durable Goods Monopolists. *Journal of Political Economy*, 90: 314–332.

[7] Colwell, P.E. [1991]. Functional Obsolescence and an Extension of Hedonic Theory. Journal of Real Estate Finance and Economics, 4: 49–58.

[8] Handfield, R.B. & Pannesi, R.T. [1997]. Managing component life cycles in dynamic technological environments. *International Journal of Purchasing and Materials Management,* 30(2): 19–27.

[9] Bradley, M. & Dawson, R.J. [1998]. An analysis of obsolescence risk in IT systems. *Software Quality Journal* 7: 123–130.

[10] Iselin, D.G. & Lemer, A.C. [1993]. Fourth Dimension in Building: Strategies for Avoiding Obsolescence. National Research Council Staff, National Academies Press.

[11] Lemer, A.C. [1996]. Infrastructure obsolescence and design service life. *Journal of Infrastructure Systems* 2(4): 153–163.

[12] Trowbridge, C.R. [1964]. Deterioration. Appraisal Journal, January: 91–6.

[13] Lee, I.H. & Lee, J. [1998]. A Theory of Economic Obsolescence. *The Journal of Industrial Economics* 0022–1821 Volume XLVI September No. 3.

[14] Akgül, F. & Frangopol, D.M. [2004]. Time-dependent interaction between load rating and reliability of deteriorating bridges. *Engineering Structures* 26: 1751–1765.

[15] Mansfield, J.R. & Pinder, J.A. [2008]. Economic and functional obsolescence: Their characteristics and impacts on valuation practice. *Property Management* Vol. 26 No. 3: 191–206.

[16] GTAI (2014). Industrie 4.0–Smart Manufacturing for the Future. Germany Trade & Invest, Berlin, Germany (2014). www.gtai.de.

[17] Bonvillian W.B. (2013). Advanced Manufacturing Policies and Paradigms for Innovation. Science. Vol. 342, Issue 6163, pp. 1173–1175. DOI: 10.1126/science.1242210.

[18] Martorell J. (2017). Nueva organización del Mando del Apoyo Logístico del Ejército del Aire. El ciclo de vida y las nuevas tecnologías. Symdex 4ª Edición. Madrid, Junio 2017.

[19] Guillen A., Gonzalez-Prida V., Gomez J., Crespo A., Turconi G., Ventola G. (2017). Maintenance 4.0. Review of maintenance role in the industry 4.0 revolution. Ed. Taylor and Francis, CRC Press: Safety and Reliability—Theory and Application: ESREL 2017.

[20] González-Prida V., Crespo A. (2014). After-sales Service of Engineering Industrial Assets. A Reference Framework for Warranty Management. London: Springer-Verlag.

[21] Hernán A.B. (2017). Modelos predictivos en el MALE basados en RBS. El ciclo de vida y las nuevas tecnologías. Symdex 4ª Edición. Madrid, Junio 2017

[22] Lambert, K.R.. (2017). Supporting high-technology systems during periods of extended life-cycles by means of integrated logistics support. *South African Journal of Industrial Engineering,* 28(1), 125–132.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Security risk and vulnerability analysis in military operational planning: The why's and how's

S. Malerud & H. Fridheim
Norwegian Defence Research Establishment (FFI), Kjeller, Norway

ABSTRACT: A military operational plan helps prepare the armed forces for future security challenges. Planning necessitates making assumptions about the future, and these assumptions constitute operational risks; possible negative operational effects in case the assumptions fail. Hence, a thorough risk analysis should be part of the operational planning process. In this paper, we propose a method for analysing operational risks and vulnerabilities in support of operational planning. In particular, we focus on risks related to assumptions about availability of critical capabilities. Using a bow-tie model where the undesired event is a capability gap, we explore likely causes and consequences by combining gaps with identified vulnerabilities. The outcome of the analysis is an overview of operational risks. The method is applied on an example within a fictive planning scenario. While developed primarily for military planning, the method is also relevant for security risk analysis in the civilian domain.

1 INTRODUCTION

A military operational plan is developed to prepare the armed forces for future security challenges. The future is inherently uncertain; hence, we should aspire to develop plans that are robust and adaptable both in the short-term and long-term perspective. This also necessitates making planning assumptions that are assertions about the future that underlies the plan (Dewar et al. 1993). Uncertainty and pertaining assumptions entail operational risks, which can be expressed as the combination of the probability that something goes wrong and the related operational consequences; negative effects on the operational effectiveness.

In order to facilitate development of relevant plans, we need to balance their desired properties, such as robustness, adaptability and flexibility, with the planning assumptions. This could be assumptions about the future threat environment and the availability of critical capabilities and resources to handle different situations. These assumptions are vulnerable to future changes, and thus, should be monitored and followed-up in order to ensure relevant plans.

Most military plans are developed following the Operational Planning Process (OPP) described in the Comprehensive Operations Planning Directive (COPD) (NATO 2013). COPD focuses on what to do, i.e. which planning products to produce, but not so much on how to do it. The COPD planning process results in comprehensive plans that include planning assumptions and risk assessments. However, the process usually stops short of identifying adequate indicators which can help determine when assumptions are violated or failed, i.e. when it is time to revise the plans.

There are various sources of uncertainty to include in the planning process, e.g. adversaries' intentions and capabilities, own and allied capabilities and the properties of the operational environment. In order to develop relevant plans, all these sources need to be considered. Ignoring uncertainty and associated risks is not a viable option. According to Walker et al. (2013), ignorance may reduce the ability to take timely corrective actions and increases the risk of missing emerging opportunities.

In order to support development of relevant plans, we need to explore uncertainties, vulnerable planning assumptions and possible consequences if assumptions are violated. This information can be obtained by a risk and vulnerability analysis. Hence, in this paper we discuss how risk and vulnerability analysis can support development of relevant plans.

For this reason, we propose a method for risk and vulnerability analysis based on the well-known bow-tie model; see for instance Aven (2015) and Rausand & Utne (2009). The starting point is an initiating, undesired event, and the model is used to explore likely causes for and possible consequences of this event. The model is a multi-method approach, which means applying different combinations quantitative and qualitative methods that

work well together to solve the problem. This makes the method flexible and adaptable to user requirements.

The proposed method can be applied to support development of relevant plans by testing the validity of planning assumptions: What happens if planning assumptions fail or are violated? The output of this analysis supports development of more adaptable plans, and it enables decisions about the need to either revise an existing plan or develop a new one. Further, the method can be used to analyse risks and vulnerabilities in support of plan development following the Operational Planning Process (OPP). The result of the risk and vulnerability analysis can be summarized in a risk picture providing necessary information for risk management.

In particular, we focus on risks related to capability gaps, i.e. gaps in the ability to perform a certain task or function. We explore likely causes and consequences by combining capability gaps and identified vulnerabilities, where we also assess the impact of barriers and implemented measures. The outcome of the analysis is an overview of operational risks that can be used to validate planning assumptions.

2 BACKGROUND

A relevant plan is one that is useful for its purpose, i.e. prepare for future decision making. Developing a relevant plan is not trivial, and it involves recognizing uncertainties and finding the right balance between robustness, adaptability and planning assumptions. Planning assumptions are necessary due to the irreducible uncertainties related to the future state of the world. Making planning assumptions is always challenging, because they have a direct influence on the relevancy of the plan. Hence, it is important to ensure that the assumptions are robust and relevant. The validity of assumptions is not static; it will change over time and depend on the planning time-horizon. It is more likely that the assumptions are valid in the short term than in the long term.

Some assumptions are more critical and vulnerable than others, i.e. violation or failure may have huge impacts on the operations covered by the plan. Thus, it is important to identify assumptions, both the explicit and implicit, and to assess how critical these are with respect to the plan. According to Rosenhead & Mingers (2001), a robust plan will perform satisfactory on selected assessment criteria across a wide range of plausible futures, i.e. be less vulnerable to changes (anticipated variations) in factors like the behaviour of adversaries and the operational environment. An adaptive plan is a plan that facilitates updating when new information becomes available.

2.1 *Assumption-based planning*

Dewar et al. (1993) introduce "Assumption-Based Planning" (ABP) as a way to make more robust and adaptable plans. ABP is not a tool for developing plans, but for improving the robustness and adaptability of existing plans by identifying and examining underlying planning assumptions. Events that violate the assumptions can originate from different sources, such as potential opponents/adversaries, the capability of own and allied forces, and the operational environment. ABP is about identifying and applying relevant signpost or indicators to detect changes in the vulnerability of planning assumptions that may have severe consequences for the operation. However, ABP does not give a clear recipe for how best to identify these signposts or how to determine appropriate threshold values.

Risk and vulnerability analysis may be a suitable approach to identify vulnerable assumptions and possible consequences if these are violated or fail. The bow-tie model provides a generic approach to risk analysis using an undesired event, e.g. a violated or failed assumption, to initiate an analysis of possible causes for the event and likely consequences. A vulnerability analysis is an integrated part of the bow-tie approach, supporting both the cause and effect analysis. We believe that a thorough risk and vulnerability analysis can support identification of signposts with pertaining threshold values, and relate these to consequences for the operation.

In this paper, our primary focus is on assumptions related to the availability and performance of own capabilities required to accomplish a specific mission. A typical undesired event in this context is a capability gap, i.e. the discrepancy between the capability requirement and available resources. However, assumption about the availability of required capabilities is only one of many assumptions that are usually made in a military contingency plan. Other assumptions are made related to the behaviour of the opponent/adversary, own Courses of Action (CoA) and properties of the operational environment.

For more information about ABP and related planning methods, see for instance Walker et al. (2013) which provides a review of planning approaches for adaption under deep uncertainty.

2.2 *Risk analysis*

An overview of risks is as natural part of planning and decision processes. In our definition, risk is associated with negative consequences of uncertainty; however, uncertainty may also introduce

emerging opportunities. In this case, the challenge is to exploit the "window of opportunity" in order to gain an advantage. This underlines the importance of being aware of uncertainties, planning assumptions and risks.

The bow-tie model provides a basis for developing more advanced and detailed risk analysis methods. According to Aven (2015), there are three main categories of risk analysis methods: simplified risk analysis (qualitative), standard risk analysis (qualitative or quantitative) and model-based risk analysis (primarily quantitative). All these categories are relevant to support development of robust and adaptable plans.

Regardless of the choice of method, it is hard to identify all the vulnerable planning assumptions and to assess their associated risks. A plan may contain many underlying assumptions, both explicitly and implicitly given, which need to be exposed. Further, it is necessary to develop relevant and plausible scenarios that include violated or failed assumptions. Based on this, it should be possible to assess the likelihood of a violated or failed assumption constituting an undesired event for the risk analysis. The next step is to assess possible consequences of the undesired event and in particular reveal events that may have a huge impact on the planned operations, i.e. a high negative influence on the attainment of objectives and effects. Lastly, the results of the risk analysis are to be presented in a risk picture suitable for making decisions about the relevance of the plan and the need for revision.

Thus, some basic requirements to the risk and vulnerability analysis in support of plan development are:

- Provide a structured approach to identify vulnerable assumptions.
- Identify likely causes for violated/failed assumptions.
- Identify possible consequences of violated assumptions.
- Identify indicators to detect violated or failed assumptions.

In the next section, we present our suggested method.

3 A METHOD TO SUPPORT DEVELOPMENT OF RELEVANT PLANS

Our method aims to combine ABP (Dewar et al. 1993) with a standard risk analysis. We use the requirements stated in the previous section as guiding principles for the method development.

Although the method is developed for analysing risks related to capability gaps, the approach is generic and can thus be applied to analyse other planning assumptions as well.

3.1 *Risk analysis*

Figure 1 shows the main steps of a risk and vulnerability analysis based on the bow-tie model (Aven 2015).

The first step is to perform proper problem structuring and framing. In order to produce relevant results, it is crucial to understand what decisions the results of the analysis should support. In addition, it is necessary to limit the scope of the analysis and determine the analysis object (system). Decision makers and stakeholders should participate in formulating the problem and determining the scope of the analysis.

In our method, an undesired event is any event associated with violated or failed planning assumptions. Various events and vulnerabilities can be underlying causes for the occurrence of undesired events like capability gaps. A violated assumption can result in reduced performance of an activity/task, which further can affect the outcome of the operation.

In the cause analysis we aim to reveal the causes behind an undesired event. There can be several causes for a capability gap, e.g. lack of relevant resources to fill the capability requirement, lengthy reaction times due to low preparedness of critical resources, or insufficient sustainability due to lack of fuel, food/water or manpower. In order to avoid or reduce capability gaps, barriers and functions are implemented. A probability estimate of the occurrence of a gap should include assessments of the effect of both vulnerabilities and barriers. The severity of undesired events varies, depending

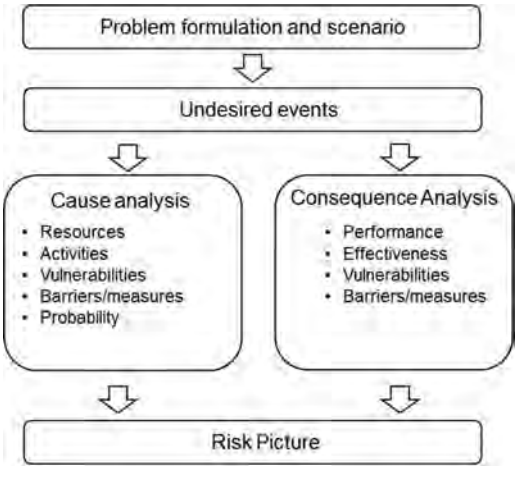


Figure 1. Risk analysis process.

on how critical the capability is for the operation. Thus, consequences depend on the scenario, the chosen Courses of Action (CoA) and the concept of operation.

The results of the cause analysis and the consequence analysis are combined in a risk picture that provides the required information to support the planning and decision process.

3.2 *Risk analysis of capability gaps*

Figure 2 gives an overview of the main components of the proposed method for analysing risks of capability gaps.

The method combines capability and risk analysis. First, we derive capability requirements and compare these to available resources. Identified capability gaps are further used as initiating events in a risk and vulnerability analysis, looking at the likelihood of occurrence of capability gaps and possible consequences for the outcome of the operation.

The different steps in Figure 2 are explained in more detail in the following. A possible application of the method is illustrated by an example.

3.2.1 *Problem structuring and scenario development*

Proper problem structuring and framing should be the first step of a risk and vulnerability analysis. Examples of questions we want to answer are: What is the purpose of the analysis? Who are the end-users? What are their expectations? Which resources are available for the analysis? Do relevant scenarios exist that can be adapted (Malerud & Fridheim 2013)?

Problem structuring can be done by informal discussions or by applying more formal methods and techniques, such as structured brainstorming, soft systems methodology or cognitive mapping.

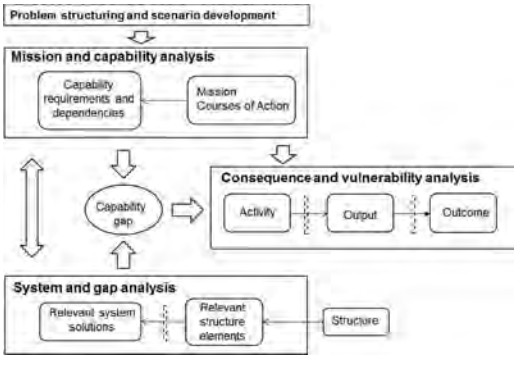


Figure 2. Method for analysing risks of capability gaps.

For an overview of problem structuring methods, see for instance (Rosenhead & Mingers 2001, Pidd 2003).

In this paper we use risk and vulnerability analysis to identify vulnerable planning assumptions and possible consequences if these are violated or fail. More specifically, we look at planning assumptions related to availability of critical capabilities. In Assumption-Based Planning (ABP), as presented in the background section, signposts are used as indicators for discovering when the vulnerability of an assumption is changed. It is particularly important to keep track of vulnerable assumptions that may have severe consequences for the planned operation.

In the example used to illustrate the method, we identify and analyse vulnerable assumptions about the availability of capabilities to handle small-scale operations as a part of daily, routine military operations. The actual plan is a contingency plan, which assumes access to capabilities necessary to perform specific tasks and functions in order to achieve the planned effects and objectives. Capability requirements are expressed using the following parameters: Capacity, reaction time, sustainability and interoperability. The plan covers small-scale crisis events at sea, and our task is to test and validate the plan in one particular scenario designed to challenge certain planning assumptions. Our focus is on capability gaps that may hamper or impede the planned operation.

Some relevant assumptions are:

- We are able to build a relevant situation picture to support situation awareness in a certain area of operation (AOO) at sea.
- We are able to react on, document (for forensic purposes) and handle an acute situation involving illegal fishing.

The scenario should in particular challenge the reaction time of critical capabilities and their ability to cooperate and exchange information.

Scenario development: Fit-for-purpose scenarios that challenge the planning assumptions are a premise for performing the risk and vulnerability analysis. Usually, a plan is developed for handling a scenario or a class of scenarios, thus there is a scenario base that can be adjusted to fit the purpose of the risk analysis. Otherwise, we need to develop new scenarios. We will not go into details with respect to scenario development here, but rather refer to Malerud & Fridheim (2016) which presents a method for developing fit-for-purpose scenarios. This method can be applied to develop new scenarios or adapt existing scenarios to fit the problems of the analysis.

Example scenario: Illegal fishing. Fishing vessels from a foreign shipping company X performs ille-

gal fishing in a Norwegian fisheries protection zone in the Barents Sea. In phase I, the illegal activity is discovered and triggers a response from Norwegian authorities. Surveillance units are allocated to the area to collect information and help build a relevant situation picture. In addition, units that are capable of ending the illegal activity are deployed.

In phase II, an inspection team from the Norwegian Coast Guard (CG) is restrained as hostages on a foreign fishing vessel. This vessel speeds up and heads for international waters.

This incident comes without advance warning, and the armed forces are performing daily, routine operations.

3.2.2 *Mission and capability analysis*

In the mission and capability analysis, information from a plan is used to develop a mission with effects, objectives and a desired end-state (what we want to achieve). A particular Course of Action (CoA) is developed comprising tasks and activities that needs to be performed in order to achieve the desired effects and objectives of the mission. Examples of effects and objectives for our scenario are:

- A relevant situation picture is established and maintained.
- The illegal activities are stopped and documented for forensic.
- The hostage situation is resolved and the hostage takers arrested.
- End state: Situation is back to normal.

An example of a CoA for handling this situation is to allocate necessary surveillance units to the area, allocate coast guard vessels (CG) to inspect and document illegal fishing activities, arrest vessels that don't comply, and finally to prepare for boarding of the vessel holding the hostages and to set them free.

Required capabilities are: In phase I: maritime surveillance, command and control (C2), action-team for boarding and inspection. In phase II: in addition to the above, action-team for hostage release.

In the following, we will focus on the maritime surveillance capability.

3.3 *System analysis*

As shown in Figure 2, the system analysis utilises a defined force structure as input, with Structure Elements (SE) such as CG vessels, surveillance aircrafts and maritime Task Forces (TF).

These SEs are combined into sub-systems to fulfil capability requirements. If a capability requirement is not entirely fulfilled, the result is a capability gap that might have impact on the operation. There are many sources that can cause capability gaps, e.g. lack of resources, loss/degradation or faults/defects of relevant SEs and low availability of SEs because they are busy doing other missions. Capability gaps arise when these sources are combined with vulnerabilities such as lack of redundancy, dependencies between capabilities, low preparedness and insufficient resilience. On the other side, barriers and mitigating actions may have been implemented to counter these deficiencies. Hence, both vulnerabilities and barriers should be considered in the gap analysis.

A part of the system analysis is to assess the performance of the SEs or sub-systems on the four performance parameters: capacity, reaction time, sustainability and interoperability. Input to these assessments can be expert judgements, historical data and simulations.

Gaps are identified by comparing system performance to capability requirements. Table 2 shows

Table 1. Capability requirements for maritime surveillance.

Task/activity	Capability	Requirement
Build and maintain a maritime situation picture	Maritime Surveillance	*Capacity*: surveillance aircraft + CG vessel *Reaction time*: Within 2 hours *Sustainability*: Duration of operation *Interoperability*: Ability to exchange information with other systems and to build a common operational picture

Table 2. Capability analysis of selected capabilities related to example scenario.

Capability requirement	SE/Sub system	Vulnerabilities	Gap
Maritime Surveillance	CG vessels Aircrafts Satellite	Dependency other capabilities Low redundancy Low preparedness	Reaction time Interoperability

an example of a gap analysis of the maritime surveillance capability.

A capability gap constitutes an undesired event that can have impact on the operations. It is possible to estimate the probability of capability gaps of varying severity, e.g. by using historical availability of resources and compare this to the capability requirements, or by assessments of Subject Matter Experts (SME). Probabilities can be expressed on a qualitative scale, e.g. high, medium and low. In this case the qualitative scale has to be defined, i.e. what is the meaning of a high probability. It is also possible to quantify the probabilities by applying methods to elicit probabilities, see for instance (O'Hagan et al. 2006).

The severity of consequences depends on the type and size of the gap. The gap may have huge or minor consequences dependent on the scenario and how critical the capability is.

In the next step, we assess possible consequences of capability gaps, with emphasis on gaps related to critical capabilities.

3.3.1 *Consequence analysis*

A capability gap results in reduced capability performance, and hence can have a negative effect on the output of an activity which further has negative impact on the achievement of effects and objectives of an operation.

Figure 3 shows an example of an event tree analysis of a gap in the maritime surveillance capability.

A moderate gap in the maritime surveillance capacity can result in reduced sensor coverage of the area of interest, delayed overview of the situation (slow reaction time), time periods with insufficient surveillance (sustainability) and inadequate common understanding of the situation due to insufficient interoperability.

The consequence analysis should comprise a thorough assessment of vulnerabilities that can lead to severe consequences, as well as associated barriers and mitigating actions that are implemented to avoid or reduce the consequences. Some examples of vulnerabilities are dependencies between capabilities, low level resilience, lack of alternatives and too high ambitions with respect to effects and timelines. At the branch points in the event tree model, identified vulnerabilities are compared to relevant barriers and mitigating measures in order to assess consequences for output and outcome of the activity/task. Such an event tree can be constructed for each capability gap.

Some relevant questions related to the event tree are:

- B: To what extent does the gap in maritime surveillance capacity cause a significant reduction in sensor coverage of AOI (vulnerabilities and barriers)?
- C: Does the reduced sensor coverage result in an insufficient situation picture?
- D: To what extent does an insufficient situation picture contribute to insufficient situation awareness?
- How severe is the gap with respect to achievement of planned objectives and effects of the operation (EH, H, M, L)? Insufficient situation awareness influences on the ability to make timely decisions.

The event tree is a suitable tool for assessing consequences. It can be supplemented by e.g. quantitative tools such as fault trees and Bayesian networks, see for instance Aven (2015).

3.3.2 *Risk picture*

Our main task is to identify and analyse vulnerable planning assumptions that can cause severe consequences if they are violated or fail. We have limited our scope to look at assumptions about the availability of relevant capabilities. The results of the risk analysis are combined and presented in a risk picture that is suitable for identifying and deciding on mitigating measures, in order to support decisions about the need for revising the plan. The risk picture contains a description of critical capability gaps, their probability of occurrence and possible consequences of the gaps for the operation (operational risks). Figure 4 shows how an operational risk picture can be presented for the example scenario.

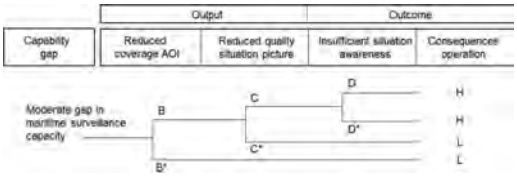


Figure 3. Example of an event tree of a gap in maritime surveillance capacity.

ID	Risk gap in maritime surveillance	Probability	Operational consequence	Mitigation
1	Capacity: Insufficient all-weather sensor coverage	L	L	More sensor units with relevant sensors. Prioritize sensor units
2	Reaction time: Sensor units insufficient reaction time	M	M	Increase preparedness of units
3	Sustainability: Sensor units don't have sufficient sustainability due to lack of logistics	L	L	N/A
4	Interoperability: Not sufficient interoperability between units	M	M	Invest in interoperable communication and information system. Collective training/exercises

Figure 4. Example of a risk picture.

4 DISCUSSION

In this paper we study the question: how can risk and vulnerability analysis support development of relevant military plans? What we want to achieve are robust and adaptable plans that help decision makers achieve overarching goals and objectives, as described in (Dewar et al. 1993 and Walker et al. 2013).

A plan will always, due to irreducible uncertainty, rely on a set of underlying assumptions. Making assumptions is necessary, but this should be done with great care because of the potentially huge consequences if the assumptions are violated or fail.

In Assumption-Based Planning (ABP) (Dewar et al. 1993), signposts are identified and applied to monitor if the vulnerability of a certain planning assumption is changed. ABP is one feasible approach to obtain more adaptive plans. It is, however not clear how signposts are identified.

Many military plans are developed using the COPD process (NATO 2013). COPD recognize the importance of adaptability and flexibility, however, it gives little guidance on how to achieve adaptable and flexible plans. In addition the COPD process lacks adequate measures for when and how a plan should be revised.

Hence, it is necessary to establish a method for detecting invalid assumptions that potentially can have huge impact on an operation. A proper risk and vulnerability analysis seems to be a promising line of approach.

This paper outlines a method for combining a capability and risk analysis to test the validity of planning assumptions, in particular assumptions related to the availability of critical capabilities for a military mission. Based on the bow-tie model, likely causes for a violated or failed assumption (capability gap) are explored. There are various sources that can cause a violated assumption, e.g. properties of the opponent/enemy, availability and performance of own and allied forces and properties of the operational environment. By identifying and assessing vulnerabilities, relevant barriers and measures implemented to reduce or avoid violation of assumptions, it is possible to obtain an overview of vulnerable assumptions and the likelihood that they are violated. This again will aid the development of signposts to monitor the vulnerabilities. This analysis relies on fit-for-purpose scenarios that challenge the planning assumptions. Ideally, one should test the planning assumptions in a wide variety of scenarios. However, in practice it is only feasible to include a few scenarios in the analysis. Thus, it is crucial to determine scenarios that cover different planning assumptions. To support this process, we suggest applying a scenario development method as described in Malerud & Fridheim (2016). Simulations may be a feasible approach to test assumptions in a larger set of scenarios and to identify the most challenging scenarios with respect to the assumptions.

In order to reveal which assumptions are most critical for achieving the objectives of the operation, it is necessary to perform a consequence analysis. We apply the proposed method to assess consequences of capability gaps by first considering how gaps influence on the output of the activity performed by the capability. Depending on the type and size of the gap, it will affect the output differently. This assessment also comprises the effects of identified vulnerabilities and barriers. In the next step, we assess consequences of reduced performance on the achievement of effects and objectives of the operation. Depending on how much the performance is reduced and the criticality of the capability, the gap will affect the outcome of the operation differently.

The proposed risk and vulnerability analysis is a multi-methodology that can adopt different combinations of methods covering the three main categories of risk analysis methods outlined in Aven (2015): simplified risk analysis which relies on qualitative methods, standard risk analysis which can be both qualitative and quantitative and model-based risk analysis which is mostly quantitative. The actual choice of method depends among others on availability of data and information, time and resources available for the analysis, and stakeholder expectations/requirements. This should be clarified in the initial problem structuring.

The example used to illustrate an application of the method is simplified and qualitative. However, it is possible to utilize more quantitative methods to enhance the quality of the risk analysis. Methods such as event trees, fault trees and Bayesian networks are promising candidates

Although the proposed method is only tested on simple cases, we believe risk and vulnerability analysis is promising with respect to support development of robust and adaptable plans. In order to refine and validate the method, we need to apply it on more cases involving real military contingency plans.

5 SOME FINAL REMARKS

We are currently not in the position to make any firm conclusions about the applicability of the method; the results are preliminary. However, we have gained some experiences that indicate that the method is appropriate for supporting development of relevant plans.

– It supports deriving signposts indicating when the vulnerability of an assumption changes.

– It can provide a risk picture comprising vulnerable planning assumptions, possible events/situations that can cause an undesired event (capability gap), and possible operational consequences if these assumptions are violated or fail (operational risk).
– It gives traceability between underlying causes for capability gaps and consequences for output and outcome of the operation.
– It provides a structured and traceable approach.

Although the method is developed to support military operational planning, we also believe it has a wider area of application, for instance to help develop robust and adaptable contingency plans in the civilian domain.

REFERENCES

Aven, T. 2015. *Risk analysis*. 2nd edition, John Wiley & Sons, Chichester.

Dewar, J.M., Builder, C.H., Hix, W.M. & Levin, M.H. 1993. Assumption-based planning—A planning tool for very uncertain times. Report MR-114-A, RAND, Santa Monica.

Malerud, S. & Fridheim, H. 2013. Metode for utvikling av scenarioer til spill og øvelser. FFI/Report 2013/00219 (in Norwegian).

Malerud, S. & Fridheim, H. 2016. A method for analysing security threats in operational risk analysis, in Walls, Bedford and Revie, *Risk, Reliability and Safety: Innovating Theory and Practice*, 2017 Taylor & Francis Group, London, p487–493.

NATO. 2013. Allied command operations comprehensive operations planning directive. COPD interim V2.0. 4 October 2013.

O'Hagen et al. 2006. *Uncertain judgements: Eliciting experts' probabilities.* John Wiley & Sons, Chichester.

Pidd, M. 2003. *Tools for thinking*. 2nd edition, John Wiley & Sons, Chichester.

Rausand, M. & Utne, I.B. 2009. *Risikoanalyse—teori og metoder*. Tapir Akademisk Forlag, Trondheim (in Norwegian).

Rosenhead, J. & Mingers, J. 2001. *Rational analysis for a problematic world revisited*. 2nd edition. John Wiley & Sons, Chichester.

Walker, W.E., Haasnoot, M. & Kwakkel, J.H. 2013. Adapt or perish: A review of planning approaches for adaption under deep uncertainty. *Sustainability*, 5, 955–979.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Security and availability on embedded systems

N. Burger, Y. Langeron, R. Cogranne & P. Lallement
University of Technology of Troyes, Troyes, France

ABSTRACT: With the fast-paced development of the Internet of Things and its applications within the emerging field of Industry 4.0 – decentralizing decisions by remotely monitoring data and automata – the issues of security and reliability of the whole communication pipeline between the connected devices taking part in this smart industry become crucial. In such context of embedded systems, microcontrollers are widely preferred over microprocessors as they are cheaper, smaller and less energy consuming. Unfortunately, the implementation of security features on microcontrollers, such as signing and ciphering functions, can largely reduce the availability of embedded systems because these functions are energy consuming and computationally complex. Thus, a trade-off has to be found between the prescribed level of availability and security. It is important to note that such a trade-off greatly depends on how the embedded systems will be used, how they are supposed to communicate between each other and if a central node with high computing resources is available. For instance, a common architecture typically consist of several embedded systems communicating up and down with a unique server. Indeed, this architecture is used in several areas where a monitor must supervise and treat data, which is the reason why this setup is chosen. The present paper aims at proposing a method to reach the right trade-off between security and availability, depending on the available resources. However, this problem is difficult to address because of the complexity to measure the security or the availability of a system. Solutions featuring those characteristics and a generic approach are presented to find the most suitable trade-off, in the use case of Industry 4.0.

1 INTRODUCTION

Embedded systems communicate between each others since the beginning of computer engineering. Today, these objects communicate on a network larger day after day: the Internet. For several years now the expression "the Internet of Things" is used. However, the expression "embedded systems" regroups many things: Smartphones, automobile electronics, sensors, and miniaturized computers like Raspberry. These objects have different functionalities, hence different computational power. For example, a communicating sensor embedded in a mechanical piece has physical constraints. Components such as microcontroller, battery, antenna, etc will be impacted by these constraints. If a Smartphone, because of these constraints, is less powerful than a computer, a communicating sensor will be much less powerful than a sensor. These constraints alone can define connected objects and their issues (Agrawal & Das 2011).

The eco-system of embedded systems changed drastically since their emergence. Embedded systems were *connected* to a physical interface, that was possible to physically secure. But with the more and more numerous communicating objects—with a server (for supervision purposes) or with other objects—physical security to access an object is not sufficient anymore (Sagstetter et al. 2013).

Whether the communication is radio-based, wired, or other, it is necessary to secure data exchanges. Security is a word that brings together many concepts. These concepts can be split in three categories:

- Confidentiality: The data are exchanged without anyone being able to understand it.
- Integrity: A modification of the exchanged data is detectable.
- Authentication: The data are signed by an emitter. The emitter can not repudiate the message. Another entity can not impersonate the emitter.

These three categories cause several needs that embedded systems must meet. Regardless of the system, the tasks answering these needs will take time and energy. Moreover, embedded systems will have different options to answer these needs, depending on their capabilities.

Comparing all the possible embedded systems outweighs the scope of this paper. That is why this study will focus on a specific restraint embedded system based on a small, procedural, real-time micro-controller with a transceiver. Possible applications, for a micro-controller like this, can be mass production of communicating sensors for the industry 4.0 or home automation.

2 PROBLEM AND CONTEXT

2.1 *Security definitions*

Authentication, integrity and confidentiality are based on mathematically complex functions. From an emitter point of view, authentication can be answered by a signature algorithm (S), integrity can be answered by a hash algorithm (H), and confidentiality can be answered by an encryption algorithm (E). From a receiver perspective, authentication can be answered by a verification algorithm (V), integrity can be answered by the same hash algorithm (H), and confidentiality can be answered by a decryption algorithm (D).

These functions take some time to execute:

- T_s: the time taken by S to sign a message
- T_h: the time taken by H to hash a message
- T_e: the time taken by E to encrypt a message
- T_v: the time taken by V to verify a signature
- T_d: the time taken by D to decrypt a message

Moreover, security also influences autonomy because these functions use computational power, hence energy, and reduce the autonomy of an embedded system. A balance must be found between security and energy consumption, reliability and real time constraints (Jiang et al. 2017). Nevertheless this study will focus on the availability of the receiver from a real-time situation perspective. Autonomy, influenced by the energy consumption of the cryptographic functions, is discarded in this paper.

2.2 *Embedded systems secured communications*

Let's take two embedded systems as in Figure 1: an emitter and a receiver. The emitter must:

- 1. Hash its message
- 2. Sign the hash of its message
- 3. Encrypt its message

The computing time for the emitter will be

$$T_{em} = T_h + T_s + T_e \tag{1}$$

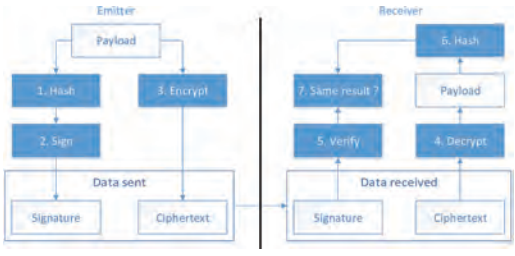


Figure 1. RSA signature.

And the receiver must:

- 4. Decrypt the message
- 5. Hash the message decrypted
- 6. and 7. Verify the signature and compare with the has previously calculated.

The computing time for the receiver will be

$$T_r = T_d + T_h + T_v \tag{2}$$

The signature (emitter side) and decryption (receiver side) have to be done with a private key. The receiver must be able to communicate with any emitter, and because of memory limitations, it is not possible to store the keys of all the emitters on the receiver memory.

A private key, k_{1d}, is used on the receiver for decryption: only the receiver can decrypt the messages from the emitter. And these messages were encrypted with the public key k_{1e} on the emitter.

Another private key, k_{2s} is used on the emitter for signing: only the emitter can sign its own messages. These signatures will be verified by the receiver with the public key k_{2v}. Two couples of keys are then needed: $K_1 = (k_{1e}, k_{1d})$ and $K_2 = (k_{2s}, k_{2v})$.

The receiver potentially manages many emitters. For example, a receiver in a nuclear power plant receives messages from hundreds of emitters. Once the messages received, they must be proceeded quickly so the data are supervised in real-time.

When the receiver receives too many messages at the same time, it will lead to message losses because the receiver will be busy to decrypt and verify the messages.

This study offers some methods to approach and quantify these *too many messages* and *at the same time* variables.

3 APPROACH AND METHODOLOGY

3.1 *Security level and algorithms*

First, some variables need to be fixed: the security level chosen is the most secured in the sense it manages the confidentiality, the integrity and the authentication of the data.

Not all messages in real life need to be so secured, but for the purpose of this article, the problem is simplified to focus on the workload of the receiver. Indeed, the receiver will be busy a certain amount of time to check these three security parameters. First to calculation of this duration is first needed.

RSA-1024 bits with SHA1 is chosen, mainly for the simplicity of the implementation, thanks to a C Microchip library. Each algorithm is set to its specific functions:

- *S*: RSA-1024 bits sign
- *H*: SHA1
- *E*: RSA-1024 bits encrypt
- *V*: RSA-1024 bits verify
- *D*: RSA-1024 bits decrypt

The RSA functions of encryption-decryption can be considered the same than the RSA signature-verification ones since the mathematical operations are the sames (Pawar & Ghumbre 2016). Indeed, in our case, the only difference is the key used for each operation:

- Encryption and decryption:

$$m^{k_{1e}} = c(modn) \tag{3}$$

$$c^{k_{1d}} = m(modn) \tag{4}$$

- Signature and verification:

$$m^{k_{2s}} = s(modn) \tag{5}$$

$$s^{k_{2v}} = m(modn) \tag{6}$$

The PKCS standard is used for the decryption and verification (on the receiver). So the function is mathematically optimized. Moreover, to balance the workload on the receiver, smaller exponent are used on its side, so the workload will be more important on the emitter for one ciphering or signature. Then, the workload per message is smaller on the receiver than the emitter, but the receiver manages several messages.

That is why equations 4 and 6 take a smaller amount of time than equations 3 and 5.

Using a smaller exponent as a private key on the receiver side for encryption may be a security issue: attackers can guess more easily the private key if they suppose it is a small exponent.

3.2 *Embedded systems definition*

The embedded system is an important variable on this problem. A modest micro-controller has been chosen to amplify the workload of both the receiver and emitter. Choosing a micro-controller means fixing the CPU speed, the Program Memory and the RAM. Each of these three variables can influence the security in some ways, because of memory limitation for example.

The measurements were conducted on a dsPIC30F3014 from Microchip with the specifications described in Table 1.

An output of the micro-controller was set to 1 during the operation, so the amount of time spent by the micro-controller to do each operation can be checked.

Table 1. dsPIC30F3014 specifications.

Parameter	Value
Architecture	16-bit
CPU Speed (MIPS)	30
Memory Type	Flash
Program Memory (KB)	24
RAM (KB)	2

3.3 *Variables*

Let's note T_r the total duration needed for the receiver to verify the security (see 2), T_{em} the duration between two emissions of the same emitter, and N the number of emitters.

Other duration, like time to send data to a central server, time to read some input on the receiver, time to read the message from the antenna can influence the availability of the system. However, these functions were voluntarily omitted to simplify the problem and to expose how security specifically influences the availability (Jiang et al. 2012).

From these data, the number of messages a receiver can treat without being overload can be determined, depending of the security (more specifically the time used for it):

$$N = \lfloor T_{em}/T_r \rfloor \tag{7}$$

4 RESULTS AND DISCUSSIONS

4.1 *Measurements*

In the Figure 2, it is shown that with a specific T_r and T_{em}, the receiver is totally busy. Graphically, it is visible that after 10 emitters, the receiver will miss some messages.

The assumption that messages follow each other was done to simplify the problem. Real life emitter can send messages at the same time, overlapping each other time frame. A CSMA/CA-like can be implemented by introducing an alea between each messages but itwould eventually return to this simple case of receiver overload.

The execution times of the different operations are referred in Table 2 and were given by an oscilloscope.

The total duration of unavailability for the receiver when receiving a message is

$$T_r = 6.8 + 6.8 = 13.6ms$$

The minimum time frame for an emitter between two message emissions is:

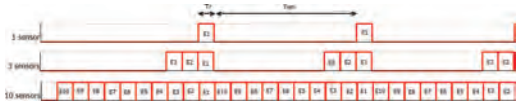


Figure 2. Receiver process time occupation.

Table 2. Operations and execution times.

Operation	Execution time
RSA 1024 signature + SHA1	158.4 ms
RSA 1024 verification + SHA1	6.8 ms
RSA 1024 encryption	159.2 ms
RSA 1024 decryption	6.8 ms

$$T_{em} = 158.4 + 159.2 = 317.6ms$$

If emitters never sleep and continuously send messages to the receiver, a message will be received each 317.6 ms. It can now be determined that the number of emitter a receiver can manage in optimal scenario where messages are sent one after another is:

$$N = \lfloor T_{em} / T_r \rfloor = \lfloor 317.6 / 13.6 \rfloor = 23 \quad (8)$$

Thus, 23 emitters can be associated to a receiver.

4.2 *Comparisons and improvements*

This result can be adapted to other scenario. For example, time measurements with RSA-2048 bits will probably take more time for both emitters and receivers and thus, change the number of emitters N.

Totally different encryption/decryption and signing/verifying algorithms can be used and compared on this embedded system. For example, a combination of AES and RSA will probably give a different N.

This comparison tool can be used to compare embedded systems instead of algorithms. For example, a measure done with AES128-CCM (symmetric solution of encryption + integrity + authentication) on a CC2538SF53 microcontroller supporting of Hardware acceleration for cryptography gave us a computing time of 210 μs (encryption) and 120 μs (decryption).

Symmetric algorithms, however, give a hard time for key management. This paper doesn't take into account this complexity, but some engineers could use the asymmetric algorithms to exchange a symmetric key.

This means that several type of messages will then be exchange: some encrypted with RSA, some with AES for example. That will imply different duration for these messages but it can be easily implemented on this method.

5 CONCLUSIONS

These results can be used to determine how many sensors can be attributed to a receiver on a specific area. Methods such as a header containing the ID of the sensor can help the receiver to know if it must verify the message or ignore it and thus, save time.

If an industry needs to deploy 30 sensors for instance, in the case studied previously, two receivers will be used.

Moreover, this method can be used to compare micro-controllers, and it can help embedded systems designers to better choose their components, regarding their needs.

More research can add other constraints on the equation, like the impact of security on the battery of the emitter (receivers are supposed to be plugged on a power source). This way, embedded systems designers will be able to choose their level of security depending on the real time and autonomy requirements.

REFERENCES

Agrawal, S. & M.L. Das (2011, December). Internet of Things – A paradigm shift of future Internet applications. In *Nirma University International Conference on Engineering*, pp. 1–7.

Jiang, W., Z. Guo, Y. Ma, & N. Sang (2012, June). Research on Cryptographic Algorithms for Embedded Real-time Systems: A Perspective of Measurementbased Analysis. In *2012 IEEE 14th International Conference on High Performance Computing and Communication 2012 IEEE 9th International Conference on Embedded Software and Systems*, pp. 1495–1501.

Jiang, W., P. Pop, & K. Jiang (2017, July). Design optimization for security- and safety-critical distributed real-time applications. *Microprocessors and Microsystems 52*(Supplement C), 401–415.

Pawar, A.B. & S. Ghumbre (2016, December). A survey on IoT applications, security challenges and counter measures. In *2016 International Conference on Computing, Analytics and Security Trends (CAST)*, pp. 294–299.

Sagstetter, F., M. Lukasiewycz, S. Steinhorst, M. Wolf, A. Bouard, W.R. Harris, S. Jha, T. Peyrin, A. Poschmann, & S. Chakraborty (2013, March). Security challenges in automotive hardware/software architecture design. In *2013 Design, Automation Test in Europe Conference Exhibition (DATE)*, pp. 458–463.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Constructing a method for classification of complex infrastructures for security threats: A case study of Norwegian ISPS port facilities

K. Brattekås, J.A. Bruvoll & M. Maal
Norwegian Defence Research Establishment (FFI), Norway

J.F. Aae & A. Breivik
Norwegian Coastal Administration, Norway

ABSTRACT: When an unspecified terrorist threat against Norway was identified in the summer of 2014, the security level was heightened at all of Norway's 600 plus International Ship & Port facility Security code (ISPS) ports. A type of classification system was necessary in order to identify which ports required security measures dependent on the different types of situations which could potentially arise. However, classifying infrastructures by criticality is a complex task with little guidance available. The main challenge was how to decide which level of protection the facilities would require. We evaluated whether the ad hoc approach, used to return security levels to baseline, was a good approach, and if there were other best practices available. Furthermore, we asked how the more than 600 different facilities could apply the same classification system. This paper proposes an approach for how to arrange different port facilities into "security profiles". A general threat assessment was made, based on a literature review and contact with security authorities, in order to determine what kind of scenarios would be relevant for the ports. We then continued to map the different types of ports, and common denominators for security issues. Through the literature review, workshops and input from the Norwegian Coastal Administration, we developed value-based categories for criticality. The findings in this paper can help clarify and present solutions which might help practitioners overcome challenges related to assessing security and threats to their facilities. The approach presented in this paper may also be a useful framework for other critical infrastructures to help select categories for ranking or classification.

1 INTRODUCTION

Knowing which infrastructure is the most critical or most exposed in security incidents is a great challenge, and there are few systematic methods available. Several approaches have been developed to identify and assess critical infrastructures over the years, yet classification and prioritisation of asset or value protection remains a difficult task given the occurrence of security threats or risks. In this paper, we present a method developed to classify port facilities into different security profiles.

This method may also be a useful framework for other types of infrastructure when it comes to assessing criticality. Even though the method presented in this paper is customised for the nature, rules and regulations of the maritime sector, the broader framework and the mindset may be applicable for further use in other infrastructure classifications. This paper aims to present the background for – and development of – this methodological framework.

Port facilities play a key role in the global maritime transportation, as over 90 percent of the world's trade is freighted by sea (IMO, 2017). Ships, engaged in international voyages, which hold an International Ship Security Certificate (ISSC) can only be served by ISPS-approved port facilities creating a security network for global maritime transportation.

The Norwegian Defence Research Establishment (FFI) and the Norwegian Defence Estates Agency (NDEA) conducted a study in order to develop a user-oriented method for classifying Norwegian ports and port facilities. The study provided guidance in relation to the risk-based supervision conducted by the Norwegian Coastal Administration (KYV), and on how to determine the maritime security level in Norwegian ports and port facilities, particularly if the general threat level was increased. The focus for this article is the latter, where we have developed a method for classifying port facilities based on their criticality.

Various categories for criticality were developed in this study, which included ranking tables. Examples of categories developed include the number of annual passengers, the type of goods being stored and transported, and the port facility's strategic importance. Eight ranking tables were used for each

port facility to determine the overall score, and thus determine which security profile the port facility belonged to (security profile 1 is low criticality, 2 is medium criticality, and 3 has high criticality).

It was also recommended that KYV base their risk-based supervision on the ports' security profile as well as the Port Facility Security Assessments (PFSA).

2 BACKGROUND

In the summer of 2014, the general terror threat level in Norway was raised by the Police Security Service (PST) (PST 2014). The terror threat was high, but unspecified, and due to the raised threat level, KYV decided to raise the security level for all Norwegian port facilities to security level 2 from 24th July 2014. This resulted in increased costs and workloads at each port, and when the threat level continued to stay high, KYV started to lower the security level to normal (security level 1) by the end of July, depending on the features of each facility.

Following this situation, KYV wanted a more substantiated method for determining the ports' inherent security classification in order to better differentiate the security levels for incidences when the general threat level is raised, or where certain threats are specified. With this background, FFI and NDEA worked in 2015 and 2016, in close collaboration with KYV, to develop such a method.

2.1 *The ISPS code and the SOLAS convention*

Several international and national laws and regulations apply to ports and ships in Norway. The ISPS Code was created as a result of the terrorist events in the USA September 11 2001, and was adopted by the International Maritime Organization (IMO) in the Safety of Life at Sea (SOLAS) Amendments in 2002. The objective of the ISPS Code and amendments was to enhance international maritime security.

The ISPS Code Part A, and some of Part B, is applicable for Norwegian regulations. The Code applies to port facilities that operate the following ships in international traffic: passenger ships, cargo vessels over 500 BT, moveable drilling rigs and special ships (SPS). All ships with ISSC are considered to be in international traffic. The ISPS framework is not applicable to war ships, military auxiliary vessels, ships in state-run, non-commercial operation, fishing vessels, ships without propulsion machinery, primitive wood ships and pleasure vessels (IMO 2003 and IMO 2012).

As described, the ISPS Code describes three security levels for port facilities (IMO 2012: 34–36). KYV is responsible for changes in security level, and notification of the Port Security Officers and Port Facility Security Officers (PSO/PFSO) about changes in their operation areas. All port facilities are required to have a Port Facility Security assessment (PFSA) and, in some cases, a Port Facility Security Plan (PFSP).

2.2 *General threat assessment*

The authors conducted a literature study and a general threat assessment where trends in the maritime sector were described. The analysis was based on trend reports from the NCIS (Kripos), the PST, and the Norwegian Intelligence Service as well as FFIs research reports and international databases of terrorist incidents. We also had a conversation with PST on general threat trends in Norway, and the KYV headquarters presented their incident log with various types of threats and security incidents in their sector.

The literature study consisted of reports concerning Norwegian sea transport risk assessment (Eggereide et al 2007; Rutledal 2002a and Rutledal 2002b), as well as a report describing maritime threat trends (Tønnessen 2007) and a threat assessment done by PST in 2013. International sources and databases (RAND 2006) were used to identify international incidents relevant for the general threat assessment. The threat assessment includes considerations about terrorism, intelligence/espionage, sabotage and other crime. Piracy has been left out due to the scope of this study focusing on Norwegian ports.

It is worth noting that there is wide academic agreement regarding the high uncertainty with security related risks, and, in particular, terrorism risk (Aven 2015; Bruvoll, 2017; Fischhoff 2002; Renn 2008; Jore and Njå 2012; Weiss 2007; DSB 2014; NSM 2016; Busmundrud et al. 2015; Maal et al. 2016). This will of course impact risk analyses and assessments where these risks are addressed. The traditional way of assessing risk, where the parameters likelihood and consequence result in an estimated, quantitative risk score, can be less expedient for security related events (Jore and Njå 2012; Aven et al. 2004; Renn 2008, and Pettersen and Engen 2010). Maritime terrorism has been a declining trend in the Northern hemisphere in the last years, and one of the main recommendations from the general threat assessment was to also focus risk assessments on other security threats and not just terrorism.

3 SCOPE

The general threat assessment that was conducted in the original body of work will not be further described in this article. This article focuses on the

method developed, and the preconditions for the choice of approach.

3.1 *Definitions*

Value assessment has the purpose of mapping the organisation's values (sometimes also referred to as "assets"), and considers which of these are most important for the organisation's mission and deliveries. This requires systematic review of which consequences would happen if the values are affected (NSM 2016:11).

According to Norwegian Standard 5830 the term *value* is defined as "a resource that if affected by unwanted influence will cause a negative consequence for the ones owning, managing or benefiting from the resource" (NS 5830:2012). A *threat* is defined as a "possible unwanted action that can have a negative consequence for a port facility's security" (ibid.). *Vulnerability* is defined as "lacking ability to resist an unwanted event or restore a new stable condition if a value is affected by unwanted influence" (ibid.).

3.2 *Port facilities and ports*

A port facility is described as land, buildings, facilities and other infrastructures used in port operations, including quays, terminal buildings, loading, unloading and transhipment facilities, and storage and administration buildings (KYV, 2011).

A port is defined as areas that are for use by vessels that load or unload goods or transport passengers as part of maritime transport or other business activities (ibid.). A port may contain several port facilities within its perimeters.

4 METHODS

The methods were based on reviews of primary and secondary literature. Furthermore, the authors held several workshops with relevant national and regional stakeholders in the Norwegian Coastal Administration.

4.1 *Literature review*

The primary literature reviewed consisted of 24 risk assessments from selected port facilities from all regions in Norway as well as a classification guide for ISPS facilities developed in Denmark (Danish Transport Authority, 2015). This primary material gave unique insight into the main threats, as well as the diversity in Norwegian port facilities. Furthermore, the Danish example gave a baseline for a best practice method. The secondary literature studied ranged from relevant rules and regulations, to academic theory about risk, risk based supervision and vulnerability, as well as previous reports and threat assessments about maritime security.

4.2 *Workshops and interviews*

Several workshops were conducted with relevant stakeholders and users from all the regions where KYV are situated, as well as more limited workshops with selected stakeholders. This was to gain an overview of the diverse ports and port facilities in Norway, as well as ensure the user perspective throughout the study. The representatives from KYV also provided important expertise for the study.

Semi-structured interviews were conducted with experts outside of KYV for the threat assessment and risk based supervision.

4.3 *Validation*

When the first draft of the methodology for classifying port facilities was done, KYV were able to test the framework at several facilities, and give feedback on whether the framework would be applicable or required further adjustment. This was an iterative process where every draft of the classification system was circulated and considered for improvements in order to ensure sufficient validation of the system.

4.4 *Limitations*

The number of port facilities regulated by the ISPS code in Norway is more than 600. Thus, assessing all of them individually was out of scope for this study, and an approach based on "types" of ports was selected and done on a general basis in order to include all ISPS port facilities.

The analysis has included risk for intentional acts (security). Considerations regarding ICT-security or personnel security were not included.

5 CLASSIFICATION OF CRITICALITY

Categorising all Norwegian port facilities is highly demanding and very complex, as the facilities are very different. It is a balance between too broad categories and too narrow categories. After several takes, it was recommended that the port facilities themselves should determine which security profile they are, based on certain criteria. It is recommended that the PSO/PFSO is involved in the classification, which can then be quality assured by KYV. Based on this classification, KYV can identify some common values linked to the total score and security profiles of each port and its facilities.

The classification method consists of eight different assessment categories which all constitute a point score. The categories are all described quantitatively (by numbers and points) and by a qualitative description that may influence which profile the port facility belongs to in order to classify criticality.

5.1 *Applied research and best practices*

As previously described, the original framework that was used to lower the security level of the port facilities to security level 1 in 2014 was inadequate. The new framework needed to consider and include features for all the port facilities in Norway, and also adhere to the ISPS Code and existing best practices. Hence, the starting point included the rules and regulations, combined with regional experience from Norway and the best practice framework from Denmark.

5.2 *Categorising port facilities*

According to the ISPS Code part B, 15.5-10 the identification and consideration of important values and critical infrastructure is a process that needs to consider the potential for loss of life, the port's economic significance, symbolic value and presence of government installations. This is reflected in the damage assessment form, where the consequence classes are (i) down time/operative ability, (ii) environment, (iii) life and health and (iv) reputation. From these parent categories, we developed some sub categories to assess criticality/importance, as presented in Figure 1. This was necessary as the users needed more specific reference points to assess the different port facilities.

5.2.1 *Environment, life and health*

The ISPS Code reads that the main focus should be to "avoid death or damage". According to the Norwegian Security Act (Sikkerhetsloven 1998), one should consider if there is a potential hazard for the environment or the population's life and health. Therefore, we have used the category "number of yearly passengers" to capture how many passengers could be affected. The number of employees in a port facility was also included to capture how many people are present daily.

It was also important to see how events in a port facility could affect the life and health of the population situated around the area. Therefore, we added the category "port facility proximity to populated places". Populous areas can be places where many people are convened, such as industry areas, infrastructure such as train stations, residential areas etc. Another element is the presence of dangerous goods and hazardous materials at the port facility, which could have consequences for environment, life and health. The category for hazardous goods and substances captures this aspect. These categories are visualised on the left side of Figure 1.

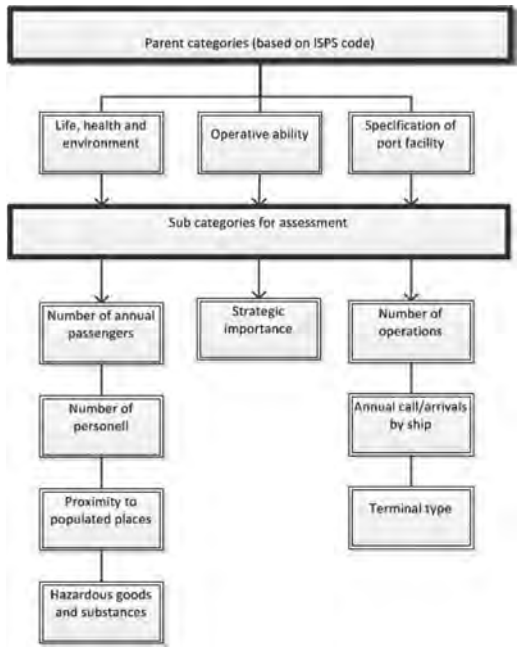


Figure 1. Categories for assessing criticality.

5.2.2 *Operative ability*

In the centre of Figure 1, the category "strategic importance" was included. According to the ISPS Code part B/15.6 it is important to consider if the port facility can still function without the given asset (or value), and in which degree it is possible to restore normal operations quickly. Downtime for the entire operation or delays in important deliveries can result in substantial costs. Addressing how some ports/port facilities can have such critical deliveries to society, is essential since any downtime will have impacts on society. This is why we added a category for strategic importance and redundancy. This is connected to national security and sovereignty, as well as critical infrastructure.

Today, there is no agreed method to assess whether the organisation is a critical infrastructure, and qualitative, knowledge based considerations are necessary. According to the Security Act, one needs to consider if downtime can have consequences for national security and defending the country, as well as critical societal functions for civilians. The assessment questions that addressed this aspect were: (i) Does the port facility have

import and export goods of strategic importance? (ii) Is your port facility the only one that has specialised operations or has special tools and facilities (e.g. a type of cables that are only produced/delivered for a certain port or goods for supply safety)? (iii) Does the port facility have strategic importance for the Armed Forces and proximity to defence installations (such as critical equipment like RoRo and LoLo (roll-on-roll-off and load-on-load-off))? (iv) Is the port facility a previously "national designated port"?

5.2.3 *Specifications for port facilities*

From a societal perspective, it is important to identify ports that are bigger and have more traffic than others. The sub categories "number of operations in the port facility" and "expected annual calls/arrivals by ship" cover this dimension. The thought was to also capture the complexity including actors, operations etc. that is ongoing in the port. We have also included "terminal type" with categories for different types of cargo. The categories were developed together with KYV, and include a ranking of different goods: (i) dry bulk, one-terminal facilities for timber, stone, gravel, asphalt, scrap iron etc., (ii) bigger goods and bulk facilities and groupage, (iii) containers and LPG, (iv) supply bases, oil- and gas production and cruise, (v) RO/PAX (roll-on/roll-off passenger) for international traffic, ferry terminals for international shipments. The reasoning for this weighting is based on classified threat assessments and what could be seen as attractive targets for a threat actor. These categories can be seen on the right side of Figure 1.

5.3 *Assessing and ranking criticality*

After workshops with KYV, we developed a general model to assess and rank criticality. The ranking tables associated with each category collect all the quantitative information. Several of the threshold values in the tables are based on Danish best practice adjusted to Norwegian conditions. KYV has tested the model on port facilities they have great knowledge of, and have returned with input regarding threshold values. However, we also included a field for each ranking table where the assessments should be described.

KYV also specified two conditions for the classification. If a port facility has more than 200 000 yearly passengers (score 5, red colour), the port facility will automatically end up in security profile 3 "high criticality". The other condition is tied to the category "terminal type". If the port facility has RO/PAX (score 5, red colour), the port facility will automatically end up in security profile 3 "high criticality".

As mentioned, the method for classification is based on both quantitative and qualitative assessments. The quantitative data should be relatively easy to collect through the PFSAs, as well as the KYV regional personnel and PFSOs' local knowledge.

The quantitative assessment will lead to a total score, which implies the criticality of the port and port facilities. Risk assessments for security related risks are difficult to manage solely based on quantitative approaches, so it was deemed necessary to supplement these with discretionary assessments. The qualitative assessments can both increase and lower the total score based on the quantitative assessments. They are meant as a supplement where e.g. quantitative numbers cannot describe the complexity or simplicity of the systems. Under each ranking table, there is a field where the qualitative assessments should be described. Including this will achieve a more holistic overview of the different ports and port facilities. The combined quantitative and qualitative data provides grounds for a nuanced overview in order to make an informed decision about the final security profile. Table 1 is one example of a ranking table including quantitative and qualitative assessment.

The security classification should make the process for raising and lowering the security level for different port facilities and ports more effective during and after changes in threat level. However, it is preferable to have continuous, updated information about the threat level and types of threats if available. The responsible stakeholders, such as KYV and PFSO/PSOs, should stay updated on openly accessible threat assessments from the relevant authorities.

The score for the port facility must be considered against the scores of the other ISPS port facilities in the port. The highest score in a port will affect the overall score for the port. This is also in terms with EU Port Security Directive 2005/65/EC.

The following figure shows the different categories explained in this chapter, and is a visualisation that will be further described in chapter 6.

Table 1. Example of a ranking table.

Ranking table Number of personnel in the port facility					
Number	<50	51–150	151–300	301–500	<500
Score	1	2	3	4	5
Describe the assessment	*Ex: There are normally around 20 persons in the port facility on a daily basis. However, during July this number will be 60 as there are more tourists and boats visiting. We have assessed that the port should still score 1.*				

6 USING THE METHOD

The first six ranking tables that should be completed use a score from 1 to 5, and the last two use a tripartite score, 0-3-5. However, the assessment criteria for each table differ. The quantification of the categories was done in close collaboration with KYV, who has extensive knowledge about the different types of facilities.

To better describe the step-by-step method for the users, we created a fictional port facility, "Hjertvik port facility", based on several real PFSAs. This example will also be described here in order to explain the approach.

The starting point for using the method is to count the "number of operations" in the port facility, which entails number of operations, people associated with the various operations etc., ranging from one to more than five. In the Hjertvik example, there were four operations in the port facility, which gave a score of 4 points.

The second step of the classification is to determine the "port facility proximity to populated places". This is described as proximity to populations or crowded places (cities, villages, industries, bases, installations with symbolic value and other infrastructure). This is measured in kilometres/metres ranging from less than 300 metres to more than 3 kilometres. Hjertvik port facility had more than 3 kilometres to the nearest densely populated area, which gave a score of 1 point.

As the third step, the "number of annual passengers" should be accounted for. The expected numbers should be filled in on a scale from less than 1000 to more than 200 001, and if the port facility does not handle passengers, the score is 1 (green). In this category, it should also be considered whether this is seasonal. The Hjertvik port facility did not handle passengers, and hence got a score of 1 point.

Fourth, the "number of personnel working daily at the port facility" should be counted based on the expected number on a scale from 0 to more than 200. Hjertvik had 21–50 people working on a daily basis, and got a score of 3 points in this category.

The fifth step is to assess the "expected annual calls/arrivals by ship" on a scale from less than 50 to more than 500. At Hjertvik, the annual arrivals were 51–150, and they got 2 points in this category.

Sixth, the "terminal type" should be described. This was done by examples for different scores: Does the terminal handle wet bulk, dry bulk, containers, general cargo, gas, building, RO/PAX and maintenance, as described in chapter 5.3.2. If the port facility handles different types of goods, the starting point is the type of cargo that gives the highest score. Here, Hjertvik had containers and LPG (liquefied petroleum gas-wet bulk), and ended up with a score of 3 points.

The seventh step includes "hazardous goods and substances". The term "hazardous substances" is used for any substance that may constitute an unreasonable risk to the health and safety of the operators, personnel or the environment if it is not handled and processed properly by storage, production, processing, packing, use, destruction or transportation. When assessing the dangers that may be potentially harmful to humans, the HazMat "diamond" is often used (United Nations 2011 and 2015). The scores were determined by consequence scores from the HazMat diamond, and whether the hazardous substances are stored in, or in proximity to, the port facility. Hjertvik had several types of dangerous goods and got a score of 5 points.

The eighth, and last, step is to determine "strategic importance". In this category it should be assessed whether the port facility has strategic significance for national security and sovereignty (e.g. territorial sovereignty and integrity, national freedom of action, relations with other states, democratic governance). It is also important to consider the importance of the port facility in the market and if it is of national symbolic value. In addition to the description, we prepared several guiding questions to help with this category, as described in chapter 5.2.2. The three scores were: no impact on national security and sovereignty, strategic import/export, uniqueness in operations, and importance for the armed forces, or a previous "national designated harbour". Hjertvik port facility did not currently have strategic importance, and hence got a score of 0 points.

In order to summarise the port facility's overall security profile, the point scores are added together and placed in the security profiles presented in Table 2.

Hjertvik port facility got a total score of 19 points, and ended up in "Security profile 2: Medium criticality", which is visualised in yellow in Table 2.

6.1 *End user experience*

During the spring of 2017, KYV completed the classification of all ISPS-approved Norwegian port facilities by applying the described classifica-

Table 2. The port facility security profiles.

The port facility security profile	
Security classes	Score
Security profile 1: Low criticality	7–16
Security profile 2: Medium criticality	17–24
Security profile 3: High criticality	25–40

tion method. KYV employees who have their main occupation in port security conducted the work, in association with PFSO when needed. Even though the aim was to make the PSO/PFSO determine what security profile they are, only to be quality assured by KYV, KYV ended up doing most of the classifications due to the familiarity with the method.

The method has a step-by-step approach making it intuitive to conduct, and does not require comprehensive knowledge about security risk assessments. Yet, it requires a joint interpretation of the ranking tables and their initial questions. During the process, KYV experienced that the questions were perceived differently among those who conducted the classification. This entailed that more or less similar port facilities could be assigned unequal security profiles.

As a response to this, KYV had to make some further refinements for how to understand the different ranking tables. For example, in the ranking table showing the number of operations in a port facility, KYV had to specify that operations in this context are related to the main ship-port activities. This means that a port handling container ships and dry bulk ships has two operations. For future use, a more accurate definition of each ranking table would improve the reliability of the method. Another remark is that classification must be considered as a continuous process. KYV has completed classifications of all Norwegian port facilities, but a change in a port facility's operational pattern can also entail change in the security profile.

Overall, the method provides a time efficient framework that applies to all Norwegian port facilities. Decision-makers in KYV are now provided with a simple, yet efficient tool when a situation calls for an increased security level in port facilities.

7 CONCLUSION AND RECOMMENDATIONS

In this paper, we have described a method developed to classify Norwegian port facilities and ports into different security profiles. This paper describes our approach and limitations to the study, as well as the data that has been analysed.

In the study, categories to assess criticality have been developed, with appurtenant ranking tables. Eight ranking tables are used for each port facility in order to determine what score they get, and thereby which security profile they belong to (security profile 1 is low criticality, 2 is medium criticality, and 3 has high criticality). Following this, the port facility's score must be compared with other ISPS port facilities in the port. All ISPS port facilities in a port must be dimensioned after the highest score in a threat situation.

We also described general threat trends in the maritime sector, focusing on the categories terrorism, intelligence/espionage, sabotage and other crime, and recommend that KYV stay updated on the available threat information in order to determine relevant security levels.

It is recommended that the security profile classification is used as one of the baseline information sources upon which KYV carry out their risk based supervisions. The frequency of supervisions should be considered with regard to both the criticality of the port facility and experiences from previous supervisions. The basis for the supervision should be well embedded, and the port facilities (e.g. PFSO/PSO) should be involved in the process of classification and risk assessments.

The approach presented here is a case that can be used as an example for other infrastructures as well. The background for the classification is rooted in a more general way of assessing consequences of security incidents, and may also be applicable to different kinds of critical infrastructures. Lessons from this case can also be relevant to consider for critical infrastructure owners or operators who are unsure of how they should address classification.

ACKNOWLEDGEMENT

The original study was funded by the Norwegian Coastal Administration. The authors would like to thank the co-authors of our original report from the Norwegian Defence Estates Agency (Siw Tynes Johnsen and Leif Riis).

REFERENCES

Aven, T. (2015). *Risikostyring*. Oslo: Universitetsforlaget AS.

Aven, T., Boyesen, M., Njå, O., Olsen, H.O. og Sandve, K. (2004). *Samfunnssikkerhet*. Oslo: Universitetsforlaget AS.

Bruvoll, J. (2017). Hvordan kan vi kommunisere det vi ikke vet? En kvalitativ studie om risikoforståelse og risikokommunikasjon i en terrorismekontekst. FFI-report 17/00182. From: http://www.ffi.no/no/Rapporter/17-00182.pdf.

Busmundrud, O., Maal, M., Hagness Kiran, J. and Endregard, M. (2015). *Tilnærminger til risikovurderinger for tilsiktede uønskede handlinger*. FFI-report 2015/00923. From: https://www.ffi.no/no/Rapporter/15–00923.pdf.

Danish Transport Authority (2015). Tilsynsplan for maritim sikring 2015. Risikobaseret tildeling af konfrontationsdage for perioden 2015–2017. 7th May 2015.

DSB (Directorate for civil protection) (2014). *Nasjonalt risikobilde 2014*. From: http://www.dsb.no/Global/Publikasjoner/2014/Tema/NRB_2014.pdf.

Eggereide, B. Kråkenes, T. Fridheim, H. (2007). *Innenriks sjøfart som mål for terror – en risikovurdering.* FFI/REPORT-2007/00004 (Restricted).

EU (2005). *Legislation of Maritime Security*. From: https://ec.europa.eu/transport/sites/transport/files/modes/maritime/security/doc/legislation_maritime_security.pdf.

Fischhoff, B. (2002). *Assessing and Communicating the Risks of Terrorism.* From: http://www.orau.gov/cdcynergy/erc/content/activeinformation/resources/FischhoffAAAS.pdf.

NSM (2016). *Håndbok for risikovurdering for sikring.* Nasjonal sikkerhetsmyndighet, mars 2016.

NS (2012). Samfunnssikkerhet – Beskyttelse mot tilsiktede uønskede handlinger – Terminologi. Norsk Standard NS 5830:2012.

Maal, M., Busmundrud, O. and Endregard, M. (2016). *Methodology for security risk assessments – Is there a best practice?* In Walls, L., Revie, M. & Bedford, T. (eds.). Risk, Reliability and Safety: Innovating Theory and Practice. Proceedings of the European Safety and Reliability Conference (ESREL) 2016. 25.09–29.09.2016, Glasgow. CRC Press. Taylor & Francis Group. ISBN 978-1-138-02997-2.

Maal, M., Brattekås, K., Johnsen, S.T., Bruvoll, J., and Riis, L.D. (2016). *Metode for klassifisering av havneanlegg og en overordnet trusselvurdering.* FFI-report 16/02319. From: http://www.ffi.no/no/Rapporter/16–02319.pdf.

International Maritime Organization (2003). *ISPS Code*. 2003 Edition. International Ship & Port Facility Security Code and SOLAS Amendments 2002. London.

IMO (2012). *Guide to Maritime Security and the ISPS Code*, 2012 Edition, IMO Publication, London.

International Maritime Organization (2017). Our Work. From: http://www.imo.org/en/OurWork/Environment/Pages/Default.aspx.

Jore, S.H. and Njå, O. (2010). Risk of Terrorism: A Scientifically Valid Phenomenon or a Wild Guess? The Impact of Different Approaches to Risk Assessment. Universitetet i Stavanger.

Kystverket (2011). *Veiledning om havne- og farvannsloven*. From: http://www.kystverket.no/Regelverk/Havne--og-farvannsloven/.

Pettersen, K. and Engen, O.A. (2010). Rethinking risk theory: a critical realist approach to aviation security. Universitetet i Stavanger.

PST, press release, 24th July 2014, *Mulig terrortrussel mot Norge.* From: http://www.pst.no/media/pressemeldinger/mulig-terrortrussel-mot-norge/.

PST (2013). Periodisk gjennomgang av obligatorisk sårbarhetsvurdering for innenriks sjøtransport og av havneanlegg som betjener slik trafikk. (Restricted).

RAND (2006). *Maritime Terrorism – Risk and Liability.* From: http://www.rand.org/content/dam/rand/pubs/monographs/2006/RAND_MG520.pdf.

Renn, O. (2008). Risk Governance. Coping with Uncertainty in a Complex World. Earthscan: London.

Rutledal, F. (2002a). *Systembeskrivelse av norsk sjøtransport.* FFI/RAPPORT-2002/01363.

Rutledal, F. (2002b). *Vurdering av sårbarheten i norsk sjøtransport.* FFI-RAPPORT-2002/04551 (Confidential).

Sikkerhetsloven (The Security Act). (1998). From: https://lovdata.no/dokument/NL/lov/1998-03-20-10.

Tønnesen, T. (2007). Maritim terrorisme – nye trender. FFI-report 2007/00015. From: https://www.ffi.no/no/Rapporter/07-00015.pdf.

United Nations (2011). *Globally Harmonized System of Classification and Labelling of Chemicals*. Fourth revised edition. ISBNN 978-92-1-117042-9.

United Nations (2015). *European Agreement concerning the International Carriage of Dangerous Goods by Road*. From: https://www.unece.org/trans/danger/publi/adr/adr_e.html.

Weiss, C. (2007). Communicating Uncertainty in Intelligence and Other Professions. *International Journal of Intelligence and Counter Intelligence*, 21: 57–85, 2008.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Mobile data interception in 4G via diameter interconnection

Silke Holtmanns
Nokia Bell Labs, Espoo, Finland

Jani Ekman
Nokia Oy, Espoo, Finland

Cathal McDaid
Adaptive Mobile, Dublin, Ireland

ABSTRACT: The IP Exchange (IPX) connects telecommunication networks with each other. It enables features like roaming and data access while traveling. Designed as a closed network it is now opening up and unauthorized entities now misuse the IPX network for their purposes. The interconnection networks suffer from many Signaling System No 7 (SS7) protocol attacks. Advanced operators now use Diameter based LTE roaming. We will illustrate how under certain network configurations an attacker can collect sufficient amount of insider information and then modify the subscriber profile to change the access point configuration in different core network nodes and by that place himself in the middle of the data traffic path for the user. We will close with recommendations on how to prevent such an attack.

1 INTRODUCTION

It is taken for granted that we can use our phone for data and calls when abroad. We rarely consider what happens in the background when we switch on our phone after our arrival in another country. You connect to a network that knows at that point of time nearly nothing about you, still in the end you can make calls, access your webmail and twitter and are being charged on your home-network bill. This is all possible because operator networks communicate through private signalling networks, i.e. an Interconnection Network. All network operators in the world are connected through it with each other, sometimes directly, sometimes indirectly via service providers.

The first Interconnection Network was the so called Nordic Mobile Telephone Network between Norway, Finland, Sweden and Denmark [1] in 1981. At that time most network operators were state-owned and there was trust between the partners. The main goal was to enable services for their users. They designed protocols and messages to serve that goal. The Signalling System No. 7 (SS7) is a network signalling protocol stack used worldwide between network elements and between different types of operator networks, service providers on the interconnection and within operator networks. It was standardised by the International Telecommunication Union, Telecommunication Standardisation Sector (ITU-T) more than 35 years ago [2].

At that point of time, security was not the main de-sign concern, as the usage of SS7 was considered to be only in a closed network between trusted partners. Later with VoIP and other IP services being handled by mobile devices it was necessary to use an IP based Interconnection network for mobile operators, this lead to the IP Exchange network (IPX) and it has been sup-ported and partially standardized by the GSMA Association since 2004.

2 DIAMETER BACKGROUND

Diameter is the evolution of the SS7 and its Message Application Protocol (MAP) [3] protocol that is used within and between the 4G Long Term Evolution (LTE) networks. LTE uses the Diameter protocol and the Session Initiation Protocol (SIP) for communication between the network elements inside a network and between networks. We will focus on the diameter part of the network in this article. In a Diameter based network architecture all elements are connected via an IP interface. The network nodes all support the Diameter base protocol specified in IETF RFC 6733 [4]. In Diameter each interface has its own application interface specification which is defined separately

in a different specification document and specifies application specific additions to the base protocol.

To connect two LTE diameter based operators together, often an IPX or interconnection service provider is used. Smaller operators utilize IPX service providers and aggregators to be able to offer their customers a large roaming selection. Operator networks usually deploy a Diameter Edge Agent (DEA) that resides on the border of the network as the first contact point for messages coming over the interconnection link. The most important nodes from the security point of view are the Home Subscriber Server (HSS) which holds the subscriber profile information and the Mobility Management Entity (MME) which takes care of the user's mobility (often combined with a Visitor Location Register VLR).

Diameter based communication can be secured using Network Domain Security NDS/IP as specified in 3GPP TS 33.210 [5] i.e. IPSec (Figure 1). However, even if IPSec is implemented in many core network nodes and separate gateways (e.g. SeGWs) it is commonly not used i.e. the Figure 1 is wishful thinking in most cases. The reason for that are manifold. Since this is an international network, the question of the trustworthy root certificate authorities, revocation list, certificate chains, trustworthiness of self-signed certificates, key generation etc. becomes a political one. In addition, there are Interconnection Service Providers i.e. messages often traverse several "hops" between the operators. The routes between operators are chosen based on costs, therefore, the realm based routing to and from an operator might be different. The support of security would require the creation of a large Public Key Infrastructure, including dynamic certificate status validation, which goes along with some substantial investments also for player who did not invest in security in the past (as the IPX network was closed). And some operators just don't have the financial resources or expertise to secure their network communications to their partners. We will focus on the usage of Diameter on the S6a/S6d between HSS and MME as specified in 3GPP TS 29.272 [6] and the Sh interface between the HSS and an IMS application server as specified in TS 29.329 [7] and TS 29.328 [8].

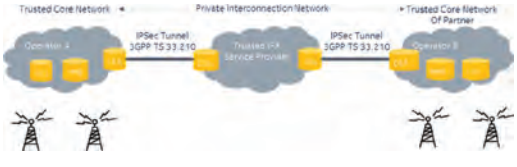


Figure 1. Interconnection between LTE operators using Diameter with NDS/IP security.

Diameter is constantly extended also for 5G and IoT (Internet of Things) connected devices. Even though Diameter is a different protocol that what is used within SS7, the underlying functional requirements e.g. authenticating the user to enable a IMS-SIP based session setup, transferring user information etc. remain much the same so there are many similarities in the messages used for Diameter and the SS7 MAP protocol messages. However, there isn't an exact one-to-one mapping for each MAP message to each Diameter command and vice versa.

3 RECENT SECURITY BREACHES

For SS7 many attacks via the Interconnection link are known and observed e.g. location tracking, eavesdropping, SMS interception, fraud, DoS, One-time password theft, credential theft, unblocking of stolen phones etc.

The first interconnection vulnerabilities were published for the 4G Diameter protocol:

- Location tracking [9], [12].
- Denial of Service/Fraud [10], [12].
- SMS interception [11].

Recently, an interesting GPRS data attack was observed during a live assessment, where a modification of the CAMEL profile of the user was manipulated [13] (Figure 2). We are doing a different type of profile modification, but it was inspired by the approach of [13].

In this attack, an incoming SS7-MAP Insert Subscriber Data (ISD) packet was detected being sent to the SGSN that a subscriber was currently attached to. Within the ISD packet the subscriber profile was modified with these specific CAMEL settings:

1. A GPRS-CSI Trigger Point was set, this set GPRS-TriggerDetectionPoint to a value of 11 (pdp-ContextEstablishment)

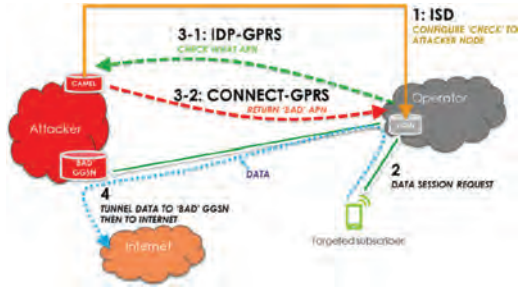


Figure 2. Data interception for GPRS [13].

2. A gsmSCF-Address was set, this is the SS7 Node address that the SGSN was to send any resultant CAMEL packet to. This SS7 address was the same as the source of the ISD packet (a SS7 Node based in another con-tinent)
3. DefaultGPRS-Handling was set to 0 (continueTransation)

The overall effect of this trigger point was to instruct the SGSN to inform the attacker node of any PDP context establishment procedure started by the sub-scriber, i.e. if they tried to use the Internet or set up a data connection. At this point the attacker node, having received the IDP-GPRS would override the APN indicated by the user (i.e. the one that is stored in his phone), with its own APN in its CONNECT-GPRS response.

The purpose of setting DefaultGPRS-Handling field by the attacker is if the attacker node did not respond, then the SGSN would continue to use the original APN, this removes the need to generate an explicit response each time, and reduces the attacker traffic.

The main obstacle for an attacker is to gain access to the private Interconnection network. In theory, this should not be possible for a private individual or an attacker, however the legal rules for network operators for renting out access to the interconnection to service providers differ between countries, and some IP-using nodes are attached and visible on the Internet (e.g. GGSN nodes via shodan.io [18]). Therefore, attackers with sufficient technical skills or financial resources have found ways to breach the privacy of the network. The GSMA Association has provided their members with a set of protection measures for SS7 and Diameter interconnection security. That those attack vectors are also very likely exploited in practice can be seen from the information contained in the leaked e-mails of HackingTeam [14] from 2013, which state that they are developing at that time already exploits for LTE/Diameter.

4 ATTACK DESCRIPTION

The attack we will describe has three major phases, which are independent of each other. Each of those phases has some assumptions on the configurations and also some variants and alternatives which an attacker may apply to reach the final goal of data interception.

We assume that the network does not have a signaling aware filtering software deployed at the edge of the network, typically represented by a Diameter Edge Agent (DEA) or a Diameter Routing Agent. Our assumption on the Diameter edge node is that it was placed there by the operator under the presumption, that it runs as a sort of router between trusted entities in a private and closed network.

The attacker is in possession of the phone number of the victim (MSISDN) and has access to the Interconnection network.

For attack preparation, the attacker acquires information about the network operators he wants to attack and about the network node he pretends to be. Operators that offer roaming have a so-called IR.21 document, this document describes various details of their architecture to enable the configuration for the roaming connections with their partners. Those IR.21 documents are not public documents and should be only used by the roaming partners for proper configuration of the roaming interface, nevertheless, some of them can be found on the Internet. In addition, operators also often use blocks of addresses for their nodes. Both IR.21 leakage and block usage make it easier for an attacker to brute force e.g. SGSN attacks.

The attacks are based on results obtained from specifications defining the behavior of the nodes and the test network that is usually used to verify correctness and security of new software code / updates for real operator networks before the new software is rolled-out.

4.1 *Subscriber data harvesting*

4.1.1 *Getting the IMSI*

The first step for an attacker is gather information about the user. He starts by obtaining the user's International Mobile Subscriber Identity (IMSI). This user identifier is needed for core network communication, but is also needed as critical information to base subsequent attacks on. There are several ways of obtaining IMSIs. One can set up a false base station and just call all devices in the area to send them their IMSI. Alternatively, a WIFI access point which is able to issue a EAP-SIM call to the device [15] or using the SS7 MAP SRI_SM (Send Routing Info for Short Message) command [16].

We will focus on how to obtain the IMSI via Diameter based Interconnection, as we assume that the attacker does not want to travel to his victim. The attacker impersonates a Short Message Center (SMSC) i.e. he claims to have a SMS for a user and he wants to deliver it and needs therefore the "contact details" from the users Home Subscriber Server (HSS), which contains all the important subscriber information elements (Figure 3). This is a common and valid roaming scenario.

For that purpose, the attacker sends a Diameter Send Routing Info for (Short Message) SM Request to the HSS of the user's operator.

This message contains the MSISDN (phone number) of the user. In response, the HSS will provide a Send Routing Information for SM Answer

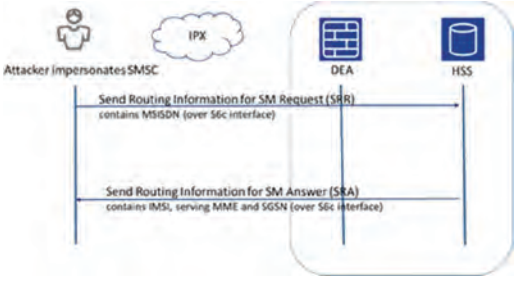


Figure 3. IMSI Retrieval using SMSC impersonation.

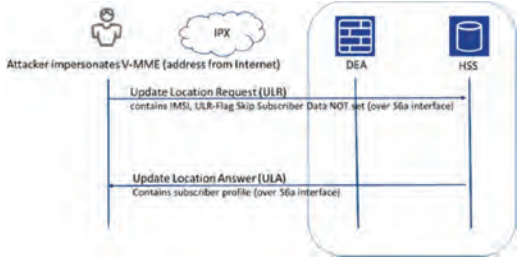


Figure 4. Profile extraction using ULR.

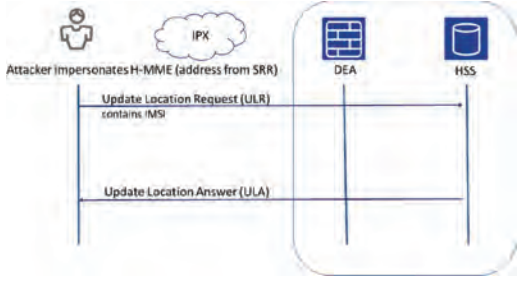


Figure 5. Setting back the MME to "old home MME".

with the IMSI and the serving nodes for the user i.e. serving MME and Serving GPRS Support Node (SGSN). The information will then later be re-used in the 2nd step of the attack.

It should be noted, that between the "harvesting of the IMSI" and the actual usage of the IMSI in attack years may go by. For some cases e.g. in IP Multimedia Subsystem IMS the IMSI is not even needed at all e.g. in the User Data Request (UDR) message.

The second piece of information the attacker craves is the subscriber profile. The subscriber profile contains detailed technical information and key attributes about the subscription the user holds.

4.1.2 *Subscriber profile retrieval*

4.1.2.1 From MME

The attacker is now in possession of the IMSI of the user, the serving Mobility Management Entity (MME). The next step is the attacker can perform a location update i.e. the attacker claims, that this user has "landed" in his network. For this he makes a Diameter Update Location Request (ULR) over the S6a interface according to 3GPP TS 29.272 [6], using the information obtained in the previous IMSI retrieval attack (Figure 4).

In this location update request he does NOT set the ULR-Flag "Skip subscriber data", this indicates to the HSS that the MME requests a fresh copy of the subscriber profile for synchronization purposes, which is a common roaming scenario. The HSS then sends back an Update Location Answer (ULA). This answer contains the requested subscriber profile.

We assume that once an attacker holds a complete subscriber profile of a user of one operator, he can deduct the operator specific details of the structure of the profile, and then figure out how to make a fake entry look real for this or another subscription. In Figure 5 a subtle attacker would also reset the MME back (assuming again that the DEA does not properly differentiate between internal and external originated traffic, in a really fully trusted and closed network this kind of differentiation would not be needed).

If the network has a very basic filtering in place i.e. H-MME messages should not come via interconnection link, then this setting back does not work, but for the main attack itself this 'extra' setting back of the MME is not needed, it is only a way for an attacker to reduce the risk of being noticed.

4.1.2.2 From HSS

An alternative method is to request the information from the HSS, this though is more difficult for Diameter. It may be possible as IMS networks usually deploy many Application Servers (AS) and not all of them reside directly in the core network. The Sh interface is between an Application Server and the HSS and is specified in TS 29.329 and TS 29.328 [7], [8] and should be used as a network internal interface only.

But, an operator has usually many application servers and not all AS services might be provided by that operator himself. They might be provided by the mother company of that operator or some other service provider, then the Sh interface becomes an interface that is open on the interconnection link (might not be visible, but still processes Sh messages).

In the case that the DEA passes Sh messages to the HSS (e.g. if the operators uses an application

server that does not reside directly in their network), the attacker can send a User Data Request (UDR) request message, which then returns the user profile. In IMS, this message can be populated with the MSISDN only as a user identity and the returned user profile would contain also the IMSI (Figure 6).

Below in Figure 7 a snapshot of such an update request and the corresponding answer in Figure 8:

As the Application Server would be on the other side of the DEA, the HSS would see only the IP address of the DEA and any IP address access control list for IMS Application Servers would not work, as that IP address control mechanism was designed for AS that reside in the same core network.

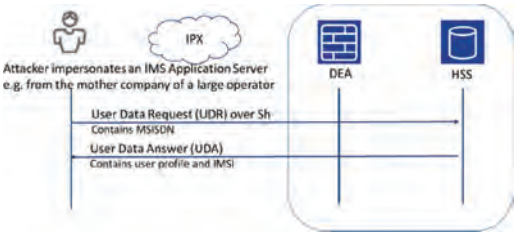


Figure 6. Subscriber profile and IMSI retrieval using UDR over Sh.

Figure 7. A UDR request containing a full profile.

Figure 8. UDA user data answer “ok”.

4.2 *Inserting the false access point node*

Depending on the different possible network configuration and on the target data communication, the attack discussed in this paper has several flavors. We will now outline how to place modified APN settings in various nodes.

4.2.1 *GPRS—profile modification in serving GPRS Support Node (SGSN) and MME*

The attacker aims to modify the subscriber data stored in the Serving GPRS Support Node (SGSN). From here on, the attack follows the discovery of [13] and evolves it to a Diameter attack. The attacker pretends to be the users home HSS issuing an Insert Subscriber Data Request (IDR) to the serving SGSN. The serving SGSN information has been obtained as outlined above or an attacker just brute forces all SGSNs of that operator or by extracting the subscriber profile.

The serving node information is part of the subscriber profile i.e. MME/VLR and SGSN. This malicious IDR message contains the subscription data AVP (attribute value pair) of type grouped. In particular, the subscription data contains:

– GPRS Subscription Data
– Access Point Network (APN) Configuration Profile (with APN con-figuration details)

In [13] based on SS7, the CAMEL settings were modified directly. In the LTE case this approach does not work, as the S6d interface does not support SGSN CAMEL subscription data. But this obstacle can be circumvented, by directly modifying the APN-Configuration-Profile AVP in the SGSN, which contains the APN-Configuration AVP (TS 29.272 [6], section 7.3.35) in the subscriber profile in the IDR diameter message.

APN-Configuration: = <AVP header: 1430 10415>
{Context-Identifier}
{PDN-Type}
{Service-Selection}

The Context-Identifier in the APN-Configuration AVP uniquely identifies the 4G Evolved Packet System (EPS) APN configuration per subscription. Each APN profile configuration can contain several APN configurations, therefore an attacker can set specific fake APNs for GPRS, EPS etc. TS 29.272 (section 7.3.34 and 7.3.35). Service Selection contains the name of the subscribed APN in use (Figure 9).

Depending on the deployed filtering the placement of a false APN in the SGSN is harder, while the user is in his home network. There the attacker would need to know the address of the home-HSS (potentially again from leaked IR.21 documents),

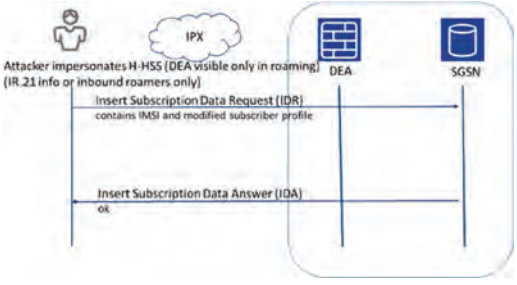


Figure 9. Placing of modified profile in the SGSN.

but that address is "less public" then for example a DEA address.

There are two "flavors" to this IDR attack. If the user is roaming and the serving SGSN resides in the visited network, then the attack has high chances of succeeding. The attacker, when sending the IDR with the malicious APN, would impersonate the Home-HSS, but due to roaming the visited network would only see the DEA address of the home network (which can be spoofed by setting the origin realm and origin host), as the message answer to the IDR does not really need to go through it is no issue to spoof the origin. The DEA address could be found from various IR.21 documents on the Internet or brute forcing the operator ranges.

Also, the network would not even have the lowest of all checks i.e. it does not check, if a message arrives on the interconnection edge which claims to come from an internal node. Therefore, the attack is considered harder, when the target user is not roaming. The modified profile would stay active until the SGSN synchronizes again with the HSS and indicates that it would need a fresh profile.

4.2.2 *Profile modification in HSS Using Sh*

An alternative method is to do subscriber profile modification in the HSS via the Sh interface. It is for an attacker very interesting to change the "master copy" of the subscriber profile which resides in the HSS, we will describe how in a not very well secure network that might be achieved. An attacker can pose as an Application Server and send a Profile Update Request (PUR) message to the HSS (Figure 10). This "synchronization message" can contain a modified subscriber profile i.e. it would have a modified APN settings. Actually, an AS is not supposed to touch that part of the subscriber profile, but the specification does not mandate fine grained profile processing on the HSS side. So depending on the implementation, the HSS will just process the incoming profile data, regardless if it is "transparent" profile data or not.

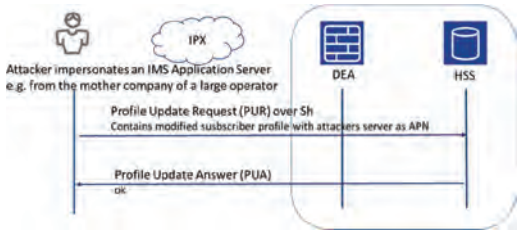


Figure 10. Update of profile in HSS using Sh interface and PUR message.

4.3 *Modification of APN settings in HSS using S6a*

The main focus of this paper is to highlight the risk that comes by an unprotected IMS Sh interface. But it should be noted, that parts of the attack can also be performed using other interfaces, like the S6a interface.

The update location message over the S6a interface allows the MME or SGSN to include a dynamic APN information to restore the Packet Data Network Gateway (PDN GW) data in the HSS e.g. for the case that a reset of the HSS has occurred and the APN information need to be restored (for detail see TS 29.272 [6] 5.2.1.1.2). The active APN AVP configuration element, contains the list of active APNs stored by the MME or SGSN, including the identity of the PDN GW assigned to each APN. The attacker would fill then in his APN information (Figure 11). The following information is required to be present:

- *Context-Identifier: context id of subscribed APN in use*
- *Service-Selection: name of subscribed APN in use*
- *MIP6-Agent-Info: including PDN GW identity in use for subscribed APN*
- *Visited-Network-Identifier: identifies the PLMN where the PDN GW was allocated (which would be the network the attacker has rented its access from)*

For the case that the above approach is not working the second alternative is to use a Wildcard APN by inserting the following information:

- *Context-Identifier: context id of the Wildcard APN- Specific-APN-Info: list of APN-in use and related PDN GW identity when the subscribed APN is the wildcard APN*

4.4 *Redirection of user*

When the user now requests data services, he would first use the APN configuration in the terminal. This APN configuration would not match the one

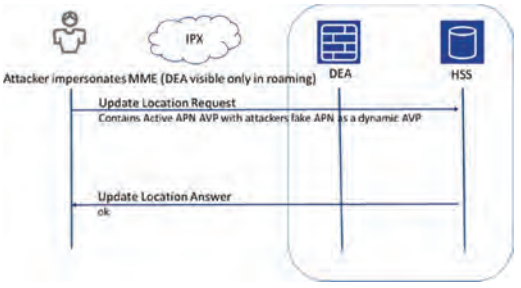


Figure 11. APN insertion with ULR over S6a.

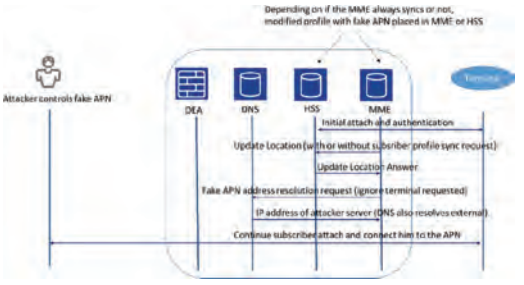


Figure 12. DNS resolution message flow.

subscriber profile in the SGSN/MME. There is an error process for this, as the basic assumption of the net-work would be that there is a misconfiguration in the user terminal. This assumption comes from the times, when the user had to manually configure their terminals. Therefore, the SGSN (GPRS) / MME (EPS) contacts the operators DNS using a NAPTR query to resolve the corresponding APN configuration to the attacker's node and the GTP (GPRS Tunneling Protocol) is established to the attackers GGSN.

Any profile modification in the SGSN and also in the MME are of "temporary nature", as those nodes synchronize with their Home-HSS at some point of time. Those synchronization configurations are operator dependent and depend on how much load usually a HSS has. If an operator has a of load on his HSS, those synchronization will be rarer. The attacker has to do some try-and-error to see, if it is sufficient to change only the MME subscriber profile or if he needs to change it in the HSS.

In some products exist the feature for such an error case to configure a default APN in the APN remap tables, but those remapping approaches can potentially be "tricked" by modifying the entry in the HSS/HLR. Also, there exist SGSN which attach automatically a "mncXXX.mccXXX.gprs" before sending out the request (similar for EPS), which would make such an attack difficult, but this kind of automatic attachment is not part of the mandatory standard processing TS 29.303 [17] and behavior and is a product specific extension.

It should be noted, that an operator should deploy different DNS for internal and external resolutions or at least different DNS views based on the requestor, but for efficiency reasons this is not necessarily always done. Note however, that even if different DNS are in place for both internal and external, and the internal DNS only contains entries of home and roaming partner APNs, it is possible that the malicious APN resolved could be resolved from the domain of a trusted roaming partner that has provisioned the malicious APN (to the attacked network). After all, the attacker needs to have some entry vector into the interconnection vector and that means he is "in some operator" already. This though, is a more complex and invasive method.

With those different approaches to modify and update the APN_Configuration_Profile AVP in the subscriber profile and setting it to the APN an attacker server, the attacker can vary his attack strategy according to configuration of the victim. The attacker does not know, if the MME would synchronize or not, so he would try and error, if he needs to modify the APN setting in the MME or in the HSS. When then a normal flow takes place and the DNS is serving internal and external requests, then the MME would redirect the terminal to the attacker's server (Figure 12).

In general, profile modification and data interception attacks pose a real threat and were observed in the wild for GPRS and therefore, we expect that with the advances in technology towards LTE, the attackers will extend their attack portfolio to cover also those attacks.

5 PROTECTION MEASURES

This article outlines, that an attacker is not bound to one technology or protocol choice when attempting to intercept data and modify the user profile via the interconnection link.

Attacks can be modular in structure and the attacker would have a toolbox depending on the encountered actual configuration of the victim network. From this toolbox he can use a mix-and-match approach to achieve his goals by combining several potential typical deployment weaknesses in Diameter networks. We validated in our test network specification conformant messages and nodes and their behavior under the different configurations. The main point of this article is to show that an information leakage and "harmless looking" configuration weaknesses in one interface combined with typical roaming messages can lead to a data breach.

The countermeasures are dependent on the actual deployments, but the following general recommendations would make this kind of attack harder:

- Interface checking at DEA i.e. the DEA should not answer Sh messages and related messages need to be blocked and logged.
- If the network relies use an Application Server outside of the core network security zone, then this should be secured with a direct certificate secure tunnel and explicit certificate based access control and authentication.
- Usage of whitelists and location-distance checks at the DEA (e.g. for S6a interface) and deploying a dedicated protocol aware signaling firewall for it.
- Validation if a reset procedure was done before executing a dynamic APN in an ULR message. In addition validation of the source of the update location i.e. if the this update location for this particular user is feasible (location-distance-travel-time-checking).
- Deployment of a dedicated DNS for internal queries only and put it behind the firewall e.g. multiple DNS servers or DNS views for each domain; Alternatively, no shared DNS configurations/physical DNS servers for trusted and untrusted domains.
- DNS security i.e. protection against cache poisoning and internal DNS should not be in front of the firewall.
- IMSI based address resolution (to avoid automatic resolution of attacker address).

These methods would make the described attacks at least much harder, if not impossible.

ACKNOWLEDGMENTS

This work was partially funded by the SCOTT project. The SCOTT (www.scott-project.eu) has received funding from the Electronic Component Systems for European Leadership Joint Undertaking under grant agreement No 737422. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and Austria, Spain, Finland, Ireland, Sweden, Germany, Poland, Portugal, Netherlands, Belgium, Norway.We would also like to thank the security group in GSMA which drive the improvement of the interconnection security. In addition, the authors would like to thank Tobias Engel from GSMK for his valuable feedback.

REFERENCES

[1] International Telecommunication Union (ITU) – T, Signalling System No.7 Specifications (T-REC-Q).
[2] Nordsveen Arve M., Norsk Telemuseum, 'Mobiltelefonens historie i Norge' 2005.
[3] 3rd Generation Partnership Project (3GPP), TS 29.002, 'Mobile Application Part (MAP) specification,' v14.3.0, Release 14, (2017).
[4] Internet Engineering Task Force, IETF RFC 6733 'Diameter Base Protocol', October 2012.
[5] 3rd Generation Partnership Project (3GPP), TS 33.210, '3G Security, Network Domain Security (NDS), IP Network Layer Security" v14.0.0 Release 14 (2016).
[6] 3rd Generation Partnership Project (3GPP), TS 29.272, 'Evolved Packet System (EPS); Mobility Management Entity (MME) and Serving GPRS Support Node (SGSN) related interfaces based on Diameter protocol', v14.3.0, Release 14 (2017).
[7] 3rd Generation Partnership Project (3GPP), TS 29.329, 'Sh interface based on the Diameter protocol; Protocol details' v15.0.0, Release 15 (2017).
[8] 3rd Generation Partnership Project (3GPP), TS 29.328, 'IP Multimedia (IM) Subsystem Sh interface; Signalling flows and message contents', v15.1.0 (2017).
[9] Rao S., Holtmanns S., Oliver I., Aura T., 'We know where you are', IEEE NATO Cy-Con, 8th International Conference on Cyber Conflict (2016), pp 277–294.
[10] Kotte B., Holtmanns S., Rao S., 'Detach me not - DoS attacks against 4G cellular users worldwide from your desk', Blackhat Europe 2016.
[11] Holtmanns S., Oliver I., 'SMS and One-Time-Password Interception in LTE Networks', IEEE ICC Conference, Paris (2017).
[12] Schmidt H., Mende D., 'Attacking NextGen Roaming Networks', Blackhat Europe 2017.
[13] McDaid C., Adaptive Mobile Blog, https://www.adaptivemobile.com/blog/malicious-data-interception-via-ss7
[14] Wikileaks, HackingTeam e-mails, https://wikileaks.org/hackingteam/emails/emailid/20790
[15] O'Hanlon P., Borgaonkar R., 'WiFi - Based IMSI Catcher', Blackhat Europe (2016).
[16] Engel T., 'Locating Mobile Phones using Signaling System 7', 25th Chaos Communica-tion Congress 25C3 (2008).
[17] 3rd Generation Partnership Project (3GPP), TS 29.303, 'Domain Name System Procedures; Stage 3', v.14.3.0 (2017).
[18] Internet of Things Search Engine, https://www.shodan.io/.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Finding your aim—choosing your game

T. Grunnan & H. Fridheim
Norwegian Defence Research Establishment (FFI), Kjeller, Norway

ABSTRACT: There are many ways to conduct crisis management games and exercises, with many formats to choose from. A recurring challenge is to find the most useful format for reaching the specific objectives of a given game or exercise. We have conducted a Limited Objective Experiment (LOE) to study organizational uncertainty related to this and to look at ways to support game and exercise planners. The experiment shows that precise aims and objectives are important for choosing suitable formats for games and exercises. Additionally, it shows that planners need guidance and support tools to make the right choices, even when the aim is specified. This is likely related to uncertainty associated with terminology, methodological biases and the fact that there are numerous available games and exercise formats. The findings are relevant to practitioners within both the fields of security and safety.

1 INTRODUCTION

There are many ways to conduct crisis management games and exercises and many types and formats to choose from. Relevant formats include table tops, seminar games, wargames and full-scale exercises. A recurring challenge is to find the most useful format for reaching the specific objectives of a given game or exercise. In our experience, there are often uncertainties within organizations on how best to do this.

Handbooks on wargaming and exercise planning are often descriptive, in the sense that they explain how to set up and conduct various types of games and exercises (DHS 2013, MSB 2014, Burns et al. 2015, NVE 2015, DSB 2016, MOD 2017). The handbooks often take the form of instruction manuals, teaching the reader to plan and conduct a game or an exercise. However, Grunnan & Fridheim (2017) shows that planners and decision-makers often are unaware of how choices in both design and execution affect the results from games and exercises. Handbooks often describe what to do, but to a lesser extent why a particular format or type is the best for a given purpose, what works and what will not.

In this paper, we will expand on our previous research (Grunnan & Fridheim 2016, Grunnan & Fridheim 2017, Fridheim et al. 2017) and argue that finding the aim of an activity is essential and a pre-requisite for choosing the right type of game or exercise to support it. However, we often notice that game/exercise planners are uncertain of what the most suitable format is, even when aims and objectives are given. The uncertainty is partly due to the bewildering array of overlapping and contradicting terms related to gaming and exercises, partly due to methodological biases of planners. Thus, we have started work on how to manage this uncertainty, using our own organization as a case. We have conducted a Limited Objective Experiment (LOE) to study whether the perceived uncertainty is real or not. We have also looked at possible tools to support game or exercise planners in choosing the right formats. This paper describes our initial findings.

The target audience for the paper is anyone involved in game/exercise planning, for example contingency planners, exercise planners, military planners, crisis managers, and consultants. In addition, the findings are relevant for decision-makers, researchers and analysts who use games and exercises as a method to investigate a problem and collect data.

2 BACKGROUND AND THEORY

2.1 *The organizational challenge*

The Norwegian Defence Research Establishment (FFI) has a long history of planning and conducting games and exercises, both generally to support analysis in running projects and specifically for various military and civilian customers. There is a wide range of activities that fall within the terms "games" and "exercises", but their purpose can roughly be divided in two:

1. Learning: Activities where the main purpose is to provide learning and training for the participants, so they are better prepared to make decisions in the future.

2. Analysis: Activities where the main purpose is to collect or develop data and information which can support future decisions.

While there are many activities that can support these purposes, they have several common features, similar to the elements of a wargame as listed in (MOD 2017):

- Aims and objectives for the activity
- Players/participants and their decisions
- A scenario and a setting that provides an immersive environment for the participants
- A simulation of the real world
- Rules, procedures and adjudication of decisions
- Data to create the scenario and the simulation
- Supporting personnel and subject matter experts
- Analysis and data collection

Although customer feedback on the conducted games and exercises at FFI is generally good, there are still areas for improvement. We have run a simple in-house problem-structuring session to help identify the challenges that must be addressed to develop more relevant and useful games and exercises in the future. Here we observed an uncertainty within the organization on which game and exercise types are best suited to support different purposes. Given the huge variety of possible formats, which should be chosen to give the most relevant answers and outputs for any given aim and objective?

Some of the causes for this uncertainty were identified to be:

- There is no common understanding of the terms related to "games" and "exercises".
- There is too little collaboration in the planning and conduct of games and exercises. Instead of sharing knowledge with each other, different project groups have established local practices and approaches.
- There is a practical approach to planning and conduct of games and exercises, instead of an academic one. The main goal is often to find a format that works given available resources, without considering options to find the best approach.

These causes are not uncommon for FFI. For instance, the lack of collaboration in exercise planning and conduct is also identified in OECD (2014), where local practices are described to be "…knowledge as a canon best passed on through mentor-mentee communication and informal apprenticeship or their designs as proprietary trade secrets best left undocumented and un-diffused".

2.2 *Games or exercises?*

The terminology related to games and exercises has long been subject to debate, and there are several definitions of both terms, with none being universally accepted. How you define the terms also depends on your area of work. Of interest to FFI is the difference between military and civilian approaches.

The military often uses the term wargaming or war gaming (tellingly, there is no consensus on whether the words should be combined or separate). One widely used definition of wargaming is from Perla (1990): "…a warfare model or simulation that does not involve the operations of actual forces, in which the flow of events affects and is affected by decisions made during the course of those events by players representing the opposing sides". The basic components of this definition are identified by the US Naval War College (USNWC 2017): "War-games involve PEOPLE making DECISIONS in a context of competition or CONFLICT (with themselves, other people, or their environment)". Exercises may differ from wargames in the sense that they use actual military forces instead of simulated ones (Simpson Jr 2015).

On the civilian side, one definition of exercises is made by the Department of Homeland Security (DHS 2013): "An exercise is an instrument to train for, assess, practice, and improve performance in prevention, protection, mitigation, response, and recovery capabilities in a risk-free environment." The same publication defines a game as "… a simulation of operations that often involves two or more teams, usually in a competitive environment, using rules, data, and procedures designed to depict an actual or hypothetic situation. Games explore the consequences of player decisions and actions". While this is in line with the definition of wargaming above, the military separation between the real (exercises) and the simulated (games) is muddled when DHS classifies games as a "discussion-based exercise technique". A similar approach is made by the Norwegian Directorate for Civil Protection, who defines "game exercise" as one of four exercise formats in their guidelines for planning, conduct and evaluation of exercises (DSB 2016).

At FFI, one result of these terminology uncertainties is that different terms are used about similar activities, and vice versa. On the military side, a tabletop discussion could be labelled as a seminar game. A similar activity on the civilian side could be classified as a crisis management exercise. However, for all practical purposes, these are the same activities, with the same aims and objectives, using similar formats. This confusion adds to the uncertainty of choosing the right game or exercise format for any given purpose.

2.3 *Choosing your game or exercise style*

There are many different styles of games and exercises available, and naturally, there have been

many attempts at categorizing them. Related to wargaming, a recent and widely used categorization is made by Pournelle (2017). He identifies six wargame categories, depending on the purpose of the games (creating knowledge, conveying knowledge or entertainment) and whether they address structured or unstructured problems, see Table 1.

Within the six categories, Pournelle (2017) identifies four major styles of operational wargames: Seminar games, matrix games, free kriegsspiel and rigid kriegsspiel. These styles get less creative, more predictable and provide more analytic rigour from left to right. In general, the effort necessary to arrange the games also increase from left to right.

Within these general styles, you will find various types of wargames. Mouat (2017) lists several types, mostly depending on how the operational environment is represented: Computer wargame, map wargame, board wargame, seminar wargame, sand table wargame and "soft issues" wargame (matrix game).

Related to exercises, DHS (2013) categorizes exercises as either:

- Discussion-based, Seminars, workshops, tabletop exercises and games.
- Operations-based, Drills, functional exercises and full-scale exercises.

Similar to Pournelle's game styles, the operations-based exercise categories are more effort-intensive than the discussion-based.

Within the range of styles, opinions naturally differ on which is the most suited game/exercise format for a given aim or objective. In many cases, the choice of format is subject to less thought than it should be. As discussed in (Grunnan & Fridheim 2016), customers often decide on a particular exercise format without finalizing the exercise goals. Similarly, game and exercise planners are very much subject to the "law of the instrument", as expressed by Maslow (1966): "I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail." The result is that old habits die hard when game and exercise planners re-use the methods they are familiar with, without considering new or improved approaches for the given problem.

Table 1. Six wargame categories (Pournelle 2017).

Problem	Creating knowledge	Conveying knowledge	Entertainment
Unstructured	Discovery games	Education games	Roleplaying
Structured	Analytic games	Training games	Commercial kriegsspiel

Similar effects can be noticed at FFI, where those working on the strategic and operational level tend to go for simpler discussion-based formats, while those working on the tactical level usually reach for more rigid formats with support of modelling and simulation. In many cases, this is a natural consequence of the problems that are to be studied, but not always.

In summary, it may be emphasized that there is need for more clarity on terms, as well as a need for tools or frameworks that enable us to choose the right game or exercise format for a given purpose.

3 METHOD

In this chapter, we explain the methods we have used in order to identify and manage uncertainty related to matching aims and objectives with suitable game and exercise formats. We have combined initial problem structuring with a simple Limited Objective Experiment.

3.1 *Problem structuring*

In order to 1) identify relevant challenges, and 2) demonstrate and discuss which game/exercise format is suitable and applicable for given purposes, we conducted two problem-structuring sessions in-house. The first session was a brainstorming session where the aim was to identify different formats/types of games and exercises, as well as to identify critical factors and criteria for matching formats and objectives.

After the initial problem structuring session (described in chapter 2.1), we used a creative technique called «Starbursting» (NATO 2017). The technique is used to generate questions about a problem, and it provides a useful structure for getting started with a research topic. We drew a star with six points and wrote who, what, why, when, where and how at each point. In the middle we put the topic of discussion, how to find the best format of a game or exercise in order to reach specific objectives. We systematically went through the six points, brainstorming possible questions. Afterwards, we organized and prioritized the questions.

The final step was to formulate a plan on how to proceed, based on an analysis of the findings. The questions laid the foundation for further literature studies and the organization of a seminar with a simple experiment. Starbursting is an effective method, especially in the early phases of a project, since many people find it easier to raise questions than to find answers (NATO 2017).

3.2 *Limited objective experiment*

The next stage was to conduct a Limited Objective Experiment (LOE). A LOE is a narrowly scoped, analytically focused assessment or validation event. We invited 24 researchers and military officers at FFI to take part in the experiment in the context of a half-day seminar. The participants' experience and expertise related to games and exercises ranged from no experience to much experience, see Table 2 in Chapter 4.2.

The seminar had three objectives. First, we wanted to strengthen internal competence and awareness of what games and exercises are and how they can be used at FFI. The authors presented a range of formats for games and exercises, using real examples. Second, we wanted to facilitate more collaboration on games and exercises across the organization, to enable learning and strengthen the quality of future games and exercises. Finally, and most importantly, we conducted the LOE with the participants, to study the perceived uncertainty related to game and exercise formats. In addition to being a competence-enhancing measure, the presentations and discussions were mood setters intended to prepare the participants for the experiment.

The purpose of the LOE was two-fold: First, we wanted to study whether the uncertainty we thought was present in the organization actually was there. Second, we wanted to see if a simple supporting form would help participants select the format of a game or exercise, when the aims and objectives were given.

The participants were given a form with the title: "How to choose the format of a game/exercise". On one page we described three cases. They were inspired by actual assignments FFI had received in the past. On the flip page, there was a form with boxes to fill out, as shown in Figure 1. The figure is a matrix with two dimensions, operational environment and adjudication. Operational environment is defined by the authors as "how the operational environment is represented in the game or exercise", i.e. how reality is simulated. Adjudication is "the process of judging how the game progresses and develops when the participants make decisions". The figure had these two dimensions represented on each axis. On each of the axes there were a number of values associated with the current dimension, 8 for adjudication and 6 for operational environment. Combinations of one value from each axis would constitute a format for a game or exercise.

Table 2. Participants' experience with games/exercises.

None	Little	Some	Much	Total
2	9	8	3	22

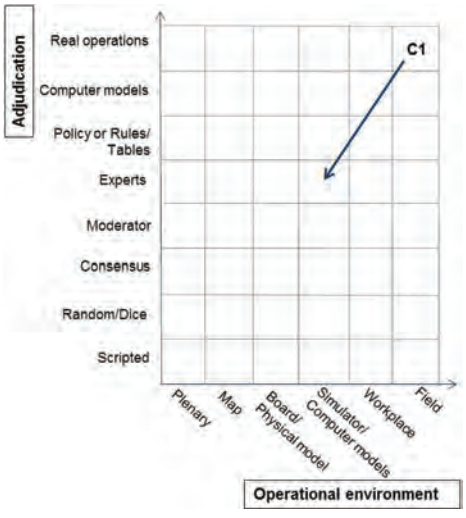


Figure 1. Example of form used in experiment.

The participants were asked to write C1, C2 or C3 in the figure where they would intuitively place cases 1, 2 and 3. When everyone had filled out the form, there was a short break to mingle and reflect, discuss and interact with the other participants. After the break, participants were allowed to make changes with an arrow to a new format they considered more appropriate, see example in Figure 1. Finally, the form was submitted to the organizers.

4 FINDINGS

In this chapter, we present the three cases we used in the experiment, before we describe and discuss the findings. The results from each case are presented in figures. The participants had the opportunity to modify the answers they made initially, and these changes are described in writing. We comment on the overall results, and we discuss how various factors may have impacted on the results from the experiment.

4.1 *Three cases*

The three cases chosen for the experiment were inspired by assignments given to FFI in the past. They have been conducted as games or exercises by researchers at the institute, including the authors, but not by the participants in the experiment.

The participants were asked to read the description of the cases and note in the form which game/exercise formats they intuitively would choose related to the two dimensions, operational environment and adjudication.

4.1.1 *Case 1 – Bilateral ministerial game*

As part of strengthening their bilateral cooperation, ministries in two countries want to carry out a wargame with a crisis in the Arctic region. The purposes of the game are:

1. Gain insight into the countries' strategic thinking, threat assessment and action in a potential crisis.
2. Help improve understanding of situational awareness and explore opportunities and limitations for cooperation in crisis and war.

4.1.2 *Case 2 – Internal wargame*

As part of internal learning processes, a research organization wants to conduct a wargame for its employees. The scenario is a bilateral conflict between two countries, which ends in a military attack on one of the countries.

The purpose of the game is to assess the two countries' practices and courses of action in different phases of the conflict, based on a given Order of Battle (i.e. available forces).

4.1.3 *Case 3 – Educational wargame*

A ministry is to participate in an annual international crisis management exercise. In order to prepare for the exercise, the ministry wants to conduct an internal game. The purposes of the game are:

1. Develop work methodology and procedures internally, within and between the various elements of the crisis management organization.
2. Prepare the players for the scenario and key issues related to the international exercise.
3. Exercise the ministry's crisis management procedures in relation to processes in international crisis management organizations.

4.2 *Experience with games and exercises*

24 employees took part in the experiment. However, the sample for "experience in games or exercises" is 22, as one participant did not fill out the table, and another marked both little experience with exercises and much experience with games. 9 participants had little experience with games or exercises, 8 participants had some experience, three had much experience, while two had no experience at all. This is shown in Table 2.

In the forms, a few answers differed considerably from the others. We examined these outliers and found that they were made by participants with no experience.

4.3 *Results*

The results in Figures 2–4 show the intuitive responses from the participants on which combination of adjudication and operational environment they would choose for the three cases. The results are their immediate reflections, based on their own skills and knowledge and on the information that was given to them in a presentation before the experiment. The number of respondents varies from case to case, due to incorrect use of the form.

4.3.1 *Results – Case 1*

For Case 1 – Bilateral Ministerial Game, 10 of 24 participants chose a *plenary* format for the operational environment, as shown in Figure 2. *Map* was the second choice, with 8 respondents finding this environment the most suitable for the given case.

Adjudication based on *consensus* between players was the preferred format for both *plenary* and *map* operational environment, with a total of 10 responses. Another preferred adjudication method was *moderator,* with 7 responses. No respondents chose *simulator/data models* as the operational environment or *scripted* and *policy or rules/tables* as means of adjudication for case 1.

Four participants made changes for Case 1 after the break. One participant changed the adjudication method from *consensus* to *scripted*, but kept *plenary* as the operational environment. One changed the operational environment from *workplace* to *plenary*, but kept *moderator* as a mean of adjudication, and another changed the operational environment from *board game* to *simulator*, but kept *computer models* for adjudication. The last modification was a change made on both axes, from *real operations—field* to *policy or rules/tables—workplace*.

4.3.2 *Results – Case 2*

Map was clearly the most preferred format for operational environment for Case 2 – Internal Wargame,

Figure 2. Results from Case 1.

Figure 3. Results – Case 2.

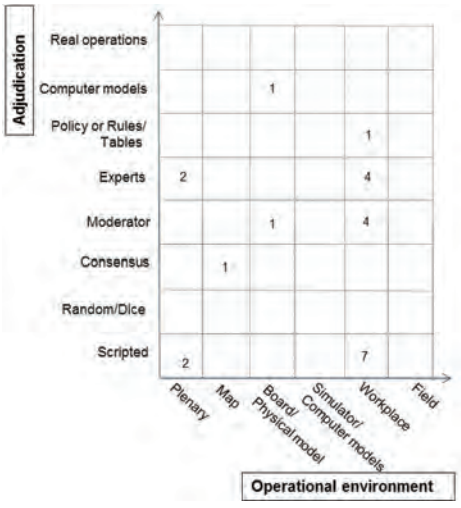


Figure 4. Results – Case 3.

with 13 of 24 respondents choosing it, see Figure 3. *Simulator*/*Data model* followed with 6 participants finding this environment the most suitable. Nobody chose *plenary* or *field*, the two "extremes" on the operational environment axis, and nobody chose *real operations* as a possible adjudication method.

Adjudication by *experts* was the preferred method with 9 responses, while *computer models* followed with 7 responses. *Experts* were chosen as the most preferred adjudication method in combination with a *map* operational environment, among four other choices. Nobody chose *experts* in combination with *simulator*/*data model*, which, evidently, was related to *computer models*.

After the break, five participants modified their responses for Case 2. Four of the changes were made from the combination *experts* as mean of adjudication and *map* as operational environment. Two of these kept *experts*, but changed *map* to *simulator*. Two others kept *map*, but adjusted the method of adjudication to *consensus* and *computer model*. One participant kept *computer model* for adjudication, but changed operational environment from *simulator* to *board*.

4.3.3 *Results – Case 3*

For Case 3 – Educational Wargame, 16 of 23 respondents chose *workplace* as the ideal operational environment, see Figure 4. No one preferred *simulator* or *field* for this case. *Scripted* adjudication was preferred with 9 answers, followed by *experts* which got 6 answers.

Four participants made changes after the break. All of them kept their initial choice of operational environment, but changed the method of adjudication. Two participants changed adjudication from *expert* to *scripted* and one participant changed from *moderator* to *scripted*. These participants' preferred operational environment was *workplace*. The last participant had *map* as operational environment, but modified the method of adjudication from *consensus* to *scripted*. The modifications strengthened *scripted* as the favorite mean of adjudication for case 3.

4.4 *Summary of results*

When combining the results from all three cases, represented in Figures 2–4, we find that the most popular choices for operational environment overall are *map*, *workplace* and *plenary*. *Field* is not considered particularly suitable for the cases. The most preferred adjudication method overall is the use of *experts*, followed by *moderator,* and *consensus* and *scripted* share the third option. The choices on the adjudication axis are more spread out than the ones on the operational environment axis.

The preferred choices in the experiment matched well with the formats used in the real games that the three cases were inspired by, see Table 3. The three means of adjudication used in the real games were *moderator*, *experts* and a *scripted* design, whereas the operational environment used were *plenary* sessions, *map* and *workplace*. The participants were presented the actual formats used after the experiment was completed, and they were able to comment on the differences.

In the experiment, the preferred combination in Case 1 was adjudication by *consensus* and a *plenary* representation of the operational environment. As shown in Table 3, the format for adjudication used in the real game was *moderator.* This was the second most popular choice in the experiment. For Case 2 and Case 3, the preferred formats in the experiment were similar to those used in the actual games.

Table 3. Actual formats used in real games.

	Case 1	Case 2	Case 3
Adjudication	Moderator	Experts	Scripted
Operational environment	Plenary	Map	Workplace

Overall, the experiment results correspond very well with real choices. However, Figures 2–4 show that the results were largely spread out, which indicate some uncertainty in the answers and that the participants intuitively thought differently.

4.5 *Discussion of results*

Various factors may have impacted on the results from the experiment. They are discussed in the following.

4.5.1 *Selection of cases*

There were similarities between the real games that the three cases were based on, in the sense that they were all discussion-based and tabletop-oriented. More rigorous games supported by models and simulations were not included, and having these as cases in the experiment could have provided clearer trends. However, we decided to choose cases that are representative for the assignments that the authors have worked on, where we knew the rationale behind the real choices of formats. Thus, it was interesting to see how a larger group of participants would respond to three tabletop-oriented cases. While the overall results from the experiment matched well with the real formats, individual responses were spread out across possible formats.

A major challenge with the cases is that they include several different objectives in the same assignment. This makes choosing just one format, or combination of adjudication and operational environment, difficult. This is especially relevant in Case 3. These challenges were present when FFI received these tasks from customers, and the participants also found this challenging when filling out the form. They gave feedback on this matter in the break between the two sessions. In reality, game and exercise planners often receive assignments with different objectives. Unlike the experiment, real assignments are often accompanied by an order or expressed wish for the type or format the customer wants. That brings us back to the initial scope of the paper: We argue that finding and defining the purpose is crucial for the choice of game or exercise, but that there is a need for supplementary guidance to find the most suitable format.

The cases could also have been presented with more detail, and the participants had little time to reflect on the cases and their choices. However, we were looking for the initial reactions from a group of researchers/officers interested in games and exercises. The results provide good indications that there is uncertainty within the organization about how best to match aims and goals with the myriad of available formats for games and exercises.

4.5.2 *Selection of dimensions*

In the experiment, we used two dimensions in the form to help choose game formats, adjudication and operational environment. The two dimensions illustrate the tradeoff between feasibility and robustness/complexity in game or exercise planning, where effort and necessary resources generally increase along the axes. It is more work involved to run a simulator than a plenary session, and data models are more resource-intensive to use for adjudication than a moderator. At the same time, the analytical rigor of the results will likely improve along the axes. This is in line with the categorizations and classifications referred to in Chapter 2, and therefore, it was natural to use these factors in the experiment.

In the experiment cases, we did not include how much time and resources were available for planning and conducting the games. In reality, time and resources are important entry values (Grunnan & Fridheim 2016), and the experiment participants commented on the lack of this information when discussing their selections. In other words, our form does not cover all relevant factors for choosing game formats. However, it gives the opportunity to quickly assess alternatives, preferably in dialogue with the client (Grunnan & Fridheim 2017).

4.5.3 *Selection of values*

The selection of values for the two dimensions in the form, presented on the axes in Figures 1–4, is influenced by the fact that both games and exercises are covered in the same form. Due to the uncertainties regarding terminology as discussed in Chapter 2, we have added values related to both games and exercises. Terms such as rules, tables, models, dices, maps, boards and simulators are often related to (war) games, while real operations, experts, moderators, workplace and field are often included in exercise vocabularies. There is a possibility that merging games and exercises in the same set-up may have confused the respondents and made them more uncertain.

4.5.4 *Number of participants*

It would have been beneficial to have more participants attending the experiment. Still, 24 participants is a fairly large number when reflecting on the number of experts working on this topic within the organization. Table 2 in chapter 4.2 shows an even distribution of experienced and less experienced participants.

5 CONCLUSION

The LOE shows that purpose should guide the choice of format for a game or an exercise, and it is therefore important to be precise when defining aims and objectives. A clear and precise task or assignment will make choosing an appropriate format easier (Grunnan & Fridheim 2017). However, our experiment indicates that even though you know the purpose, different people will still instinctively choose different formats. This is likely related to the uncertainty related to terminology, methodological biases and the fact that there are numerous games and exercise formats to choose from. Therefore, it is useful to have a quick support tool to assist the planners, like the form used in our experiment and shown in Figures 1–4. Although our findings demonstrate that there was a good match between the formats chosen in the experiment and the actual choices used in real-life, there was also a large spread of answers for each case.

The LOE is a pilot study, as it is the first time we have tested our ideas on this topic on a group of researchers. We carried out a first impressions session at the end of the experiment. The form (or matrix, as it was referred to by some participants) was especially considered useful for expanding the range of possibilities in the planning phase. The challenge is to choose formats when factors such as time and resources are included. For instance, which formats are best for a one-day or a week-long event respectively? Also, the participants found it difficult to choose only one format, especially in the cases which had several objectives in the assignment text. Often there is a need for different formats to cover all objectives, and this is not easy to visualize in one form.

There are many possibilities for future research on this topic. In coming experiments, it is possible to make adjustments such as adding more cases, choosing other dimensions in the form, and analyzing results from experienced and unexperienced participants respectively. Furthermore, for internal organizational use, we recognize that there may be a need for an internal classification of different game/exercise formats and their overall strengths and weaknesses for different purposes.

While the exercises and games covered by the paper are mainly related to security challenges, our findings are relevant also for more safety related issues. The form used in the experiment provides an opportunity to evaluate and discuss options and formats independently of the subject area.

ACKNOWLEDGEMENTS

The authors would like to thank our colleague Alf Christian Hennum for his review of the paper.

REFERENCES

Burns, S., Della Volpe, D., Babb, R., Miller, N. & Muir, G. 2015. *War Gamers Handbook – A Guide for Professional War Gamers.* US Naval War College Technical Report, November 2015.

Department of Homeland Security (DHS). 2013. *Homeland Security Exercise and Evaluation Program (HSEEP)*. US Department of Homeland Security, April 2013.

Directorate for Civil Protection (DSB). 2016. *Veileder i planlegging, gjennomføring og evaluering av øvelser. Grunnbok: Introduksjon og prinsipper*. Direktoratet for samfunnssikkerhet og beredskap, October 2016 (in Norwegian).

Fridheim, H., Grunnan, T. & Malerud, S. 2017. How to develop fit for purpose scenarios for crisis management exercises. In Cepin & Bris (Eds.), *Safety and Reliability – Theory and Applications*. London: Taylor & Francis Group.

Grunnan, T. & Fridheim, H. 2017. Planning and conducting crisis management exercises for decision-making: the do's and don'ts. In *EURO Journal on Decision Processes*: 1–17. DOI: 10.1007/s40070–017–0065–0.

Grunnan, T. & Fridheim, H. 2016. Planning and conducting crisis management exercises — what works and what does not? In Walls, Bedford & Revie (Eds.), *Risk, Reliability and Safety: Innovating Theory and Practice.* London: Taylor & Francis Group: p443–450.

Maslow, A. 1966. *The Psychology of Science.* Joanna Cotler Books, First edition (September 1966), p 15.

Ministry of Defence (MOD). 2017. *Wargaming handbook*. Ministry of Defence, Development, Concepts and Doctrine Centre, UK, released August 2017.

Mouat, T. 2017. *Wargaming*. Presentation at Connections UK, London, 5 September 2017.

NATO, 2017. *The NATO Alternative Analysis Handbook*. Second Edition, October 2017.

Norwegian Water Resources and Energy Directorate (NVE). 2015. *Øvelser. En veiledning i hvordan planlegge og gjen-nomføre øvelser innen energiforsyningen*. Rapport 39:15. Oslo: NVE (in Norwegian).

OECD. 2014. *Strategic Crisis Management Exercises: Challenges and Design Tools*. 4th Meeting of the OECD High Level Risk Forum, 10–12 December 2014, Paris, France. GOV/PGC/HLRF (2014) 9.

Perla, P. 1990. *The Art of Wargaming. A Guide for Professionals and hobbyists.* Annapolis, Maryland: United States Naval Institute. p165.

Pournelle, P. 2017. Designing Wargames for the Analytic Purpose. *Phalanx - The Magazine of National Security Analysis*. Volume 50, Number 2, June 2017, pp 48–53. Military Operations Research Society.

Swedish Civil Contingencies Agency (MSB). 2014. Övningsvägledning. Grundbok – Introduktion till och grunder i övningsplanering. Stockholm: MSB (in Swedish).

Simpson Jr, W.L. 2015. *A Compendium of War Gaming Terms*, US Naval War College, 7 July 2015.

US Naval War College (USNWC). 2017. *About Wargaming. What is a War Game?* Available at https://usnwc.edu/Research-and-Wargaming/Wargaming/About-Wargaming [Accessed 14 December 2017).

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Optimizing security patrolling scheduling in chemical industrial parks by using game theory

L. Zhang & G.L.L. Reniers
Faculty of Technology, Policy and Management, Safety and Security Science Group (S3G), TU Delft, Delft, The Netherlands

ABSTRACT: Protecting chemical clusters from intentional attacks has been a hot topic during the last decade. Besides intrusion security countermeasures such as cameras, entrances control etc., patrolling also fulfils an important role in the security of chemical facilities and industrial parks. Current patrolling strategies in industry are mainly single-plant driven and purely randomized or based on the patroller's preference. Such an approach in a chemical industrial park is on the one hand not able to cover the more hazardous facilities more than the less hazardous plants within the park, and on the other hand is not able to deal with strategic (intelligent) human adversaries w.r.t. terrorism. This paper therefore investigates a game theoretic model for optimizing the schedule of patrolling in chemical clusters. The industrial defender and the intelligent/adaptive attackers are modelled as two players in the game. The defender aims at increasing the probability of detecting the attacker, by randomly but strategically scheduling her patrolling route. The attacker aims at causing maximal consequences with highest success probabilities, by choosing a proper attack time and a proper target. The model is further illustrated by a case study.

1 INTRODUCTION

Since the unfortunate 9/11 attack, the protection of critical infrastructures has been an urgent topic, both in academia and in practice. Chemical industries have an important role in modern society. They provide materials for human being's daily necessities such as clothes, food, energy, etc. However, chemical facilities may also pose huge threat to modern society. The sadly Bhopal disaster (toxic gas leakage), among others, caused more than 3000 deaths and life-long suffering for over 300,000 [1]. In the security aspect, the malicious attack to a chemical plant in France reminded people that there is a possibility of a successful attack to chemical facilities. An investigation carried out by Orum and Rushing [2] concluded that a successful attack to a top 101 dangerous chemical plant in U.S. may result in more than one million casualties.

Furthermore, due to economic and management reasons, chemical plants are nowadays geographically clustered, forming chemical clusters, e.g., the Antwerp port chemical cluster, the Rotterdam port chemical cluster etc. Besides intrusion security countermeasures within each plant, patrolling is also scheduled, for securing these chemical facilities. The patrolling can be either single plant oriented, which is scheduled by the plant itself, or multiple plants oriented, which should be scheduled by a multiple plant council (MPC) [3]. Both types of patrolling have a drawback of not being able to deal with intelligent attackers. Some patrollers follow a fixed patrolling route, and the attacker thus can predict the patroller's position at a certain time. Other patrollers purely randomize their patrolling, without taking into consideration the hazardous level that each facility/plant holds, and the attacker can attack more dangerous facilities/plants since all the facilities/plants are equally patrolled.

Game theory has been introduced to the security domain to optimally allocate security resources. In a security problem, the attacker (human beings) is able to plan his attack according to the defender's defence, while the defender knows the fact and thus she can also defend accordingly. This procedure is called the 'intelligent interactions' between the defender and the attacker. Game theory was invented to model strategic decision making in multiple actor systems, thus it perfectly fits the necessity of modelling the 'intelligent interactions' in the security domain. Tambe and his co-authors [4] employed game theory for optimizing patrolling of protecting ferries, of protecting wild animals etc. Alpern and his co-authors [5] theoretically studied the optimization problem of patrolling in a graph. Amirali et al. [6] introduced a game theoretic model for optimally scheduling pipeline patrolling. No literature has investigated the use of game theory to optimize patrolling in chemical clusters, neither for the single plant patrolling nor for the multiple plants patrolling.

This paper proposes a Chemical Cluster Patrolling (CCP) game, which answers the question how to optimally randomize the patrolling, to better secure a chemical cluster, by using a game theory model. The reminder of this paper is organized as follows: Section 2 briefly demonstrates the CCP game. A case study is introduced in Section 3 and the results of the case study are given in Section 4. Conclusions are drawn in Section 5.

2 THE CHEMICAL CLUSTER PATROLLING (CCP) GAME

2.1 *Graphic modelling*

A chemical cluster can be descript as a graph $G(V, E)$. The vehicle entrances of each plant and the cross points of the vehicle road form the nodes of the graph. The vehicle roads between different plants (to be more specified, they should be "between different entrances") are modelled as edges of the graph. Furthermore, all entrance nodes which belong to the same plant are modelled to be full connected, which means edges also exist between every two nodes in these cases.

Based on the graphic model, the chemical cluster patrolling can be descript as a graphic patrolling problem: 1) a patroller (team) starts her (In this paper, we denote the patroller/defender as she/her/her, and denote the attacker as he/him/his.) patrolling from a node (the base camp); 2) she moves in the graph; 3) when arriving a node, she may decide whether to stay at the node for a specific period of time t_k^p (i.e., patrol the plant) or not (i.e., move to another plant without patrolling the current plant); 4) after a period T, the patroller terminates the patrolling.

A directed patrolling graph pG(pV, pE) is defined based on the graphic model of the chemical cluster. A node of pG is defined as a tuple of (t, i), in which $t \in$ denotes time dimension and $i \in \{1,2,\ldots,|V|\}$ denotes a node in graph G(V, E) (i.e., a plant (entrance) in the chemical cluster). Node (t, i) means that at time t the patroller arrives or leaves node i. A directed edge of pG from node (t_1, i_1) to node (t_2, i_2) therefore means that the patroller moves from node i_1 at time t_1 to node i_2, and arrives at t_2. Figure 3 shows the patrolling graph of the case study.

2.2 *Game theoretic modelling*

A game theoretic model consists of players, strategies, and payoffs.

Players

Players of the chemical cluster patrolling (CCP) game are the patroller team and the potential attackers. The CCP game is a two players game and both players are assumed with perfect rationality.

Strategies

An attacker's strategy consists of three parts: i) which plant to attack; ii) when to attack; and iii) what attack scenario to use, thus can be expressed as:

$$s_a = (t, i, k_i) \tag{1}$$

In which t denotes the attack start time, i represents the target plant, k_i is the attack period (e.g., 7 minutes) which should be determined by both the attack scenario and the target plant.

A mathematic formulation of the defender's strategy is shown in Formula (2).

$$s_d = \prod_{(s,e)\in pE} c_{s-e} \tag{2}$$

In which c_{s-e} denotes the probabilistic number assigned to the edge (of pG) from node s to node e, $\prod$ denotes the Cartesian product of all edges in pG (i.e., all $(s,e) \in pE$).

An important property of these probabilities is that, for each node (of pG), the sum of all the income probabilities must equal the sum of all the outcome probabilities. Formula (3) illustrates the abovementioned property.

$$\begin{aligned} sP_{pv} &= \sum_{in\in\{s\in pV|(s,pv)\in pE\}} c_{in-pv} \\ &= \sum_{out\in\{e\in pV|(pv,e)\in pE\}} c_{pv-out} \end{aligned} \tag{3}$$

Payoffs

Formulas (4) and (5) define the patroller and the attacker's payoff, in which f is the probability that the attacker would fail, and if the attacker failed, the patroller gets a reward R^d (e.g., obtaining bonus) and the attacker suffers a penalty P^a (e.g., being sent to prison). If the attacker succeeds, the patroller suffers a loss L^d and the attacker obtains a gain G^a.

$$u_d = R^d \cdot f - L^d \cdot (1-f) \tag{4}$$

$$u_a = G^a \cdot (1-f) - P^a \cdot f \tag{5}$$

Computing the f

The probability that the attacker would be detected can be calculated by Formula (6), in which f_{cpp} denotes the probability that the intrusion detection systems (IDS) in the target plant would detect the attacker, f_p is the probability that the patroller would detect the attacker

$$f = 1 - (1 - f_{cpp}) \cdot (1 - f_p) \tag{6}$$

Note that f_{cpp} is a plant-specific parameter (a number belongs to [0,1]). While f_p can be calculated by Formula (7), in which r denotes the overlap situ-

ation of that the patroller's staying in the plant and the attacker's intrusion and attack procedure, σ_r is the detection probability of situation r. Furthermore, the probability that the patroller would be in situation r is denoted as τ_r.

$$f_p = \sum_r \sigma_r \cdot \tau_r. \tag{7}$$

Denote the defender's strategy in a vector form as $\vec{c}$. It is worth noting that τ_r would be a linear polynomial of $\vec{c}$, and f_{cpp} and σ_r are user provided parameters. Therefore, f is a linear polynomial of $\vec{c}$ as well.

Stackelberg equilibrium

In the CCP game, the attacker is assumed to be able to collect information of the patroller's patrolling route. A Stackelberg equilibrium $(s_d^*, s_a^*) = (\vec{c^*}, (t^*, i^*, k_i^*))$ for the CCP game is a defender-attacker strategy pair that satisfies the following condition:

$$(t^*, i^*, k_i^*) = argmax\{u_a(\vec{c}, (t, i, k_i))\} \tag{8}$$

$$\vec{c^*} = argmax\{u_d(\vec{c}, (t^*, i^*, k_i^*))\} \tag{9}$$

Formula (8) reflects that observing the defender's strategy $\vec{c}$, the attacker would play a strategy which will maximize his own payoff (i.e., a best response). Formula (9) represents that the defender can also work out the attacker's best response to her strategy, thus she plays accordingly.

3 ILLUSTRATIVE CASE STUDY

Figure 1 provides the layout of a chemical cluster from the Antwerp port (data source: Google map). There are 5 plants in this cluster, indexed as plant 'A', plant 'B', and so forth. The yellow dot lines demonstrate the vehicle routes, and the patroller only drives on the vehicle route. Figure 2 shows the graph model of the cluster shown in Figure 1.

Figure 1. Layout of a chemical park in Antwerp port.

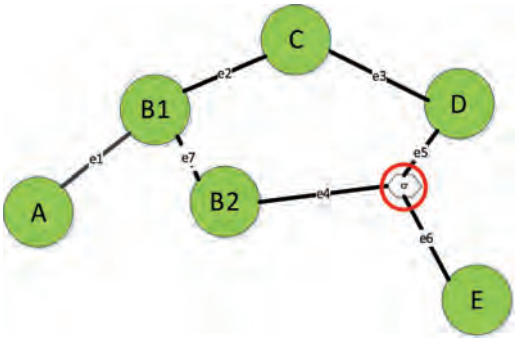


Figure 2. Graphic modelling of the chemical park.

As we may see, each plant (i.e., 'A', 'C', 'D', 'E') in Figure 1 is modelled as a node (with the same name) in Figure 2. The cross point of the vehicle road between plant 'D' and 'E' in Figure 1 is also denoted as a node in Figure 2 (i.e., node 'cr'). Moreover, plant 'B' has two vehicle entrances, and two nodes (i.e., nodes 'B1' and 'B2') are used in Figure 2 to denote the two different entrances of plant 'B'. Edges 'e1' to 'e6' reflect the vehicle roads between different plants, while edge e7 is added between node 'B1' and 'B2' because these 2 nodes belong to the same plant and hence should be full connected.

We set: $t_1^d = 2, t_2^d = 3, t_3^d = 4, t_4^d = 3, t_5^d = 2, t_6^d = 2$, and further set $t^p('A','B','C','D','E') = [9,7,6,5,7]$. t_i^d represents the driving time of edge 'ei' in Figure 2. For instance, t_1^d is the driving time from node 'A' to 'B1'. $t^p('X')$ denotes the time needed to patrol plant 'X'. If the patroller may have multiple patrolling intensity in a plant, then the t^p should not only be a number, but be a set of numbers. In this paper, we only consider one patrolling intensity in each plant and all the temporal data are unified in minutes.

Table 1 shows the time of moving from one node to another node. For instance, from node 'A' to node 'B1' needs $t_1^d = 2$ minutes. It is worth noting that i) numbers in the diagonal denote the time needed to patrol the plant, e.g., patrolling plant 'A' needs $t^p('A') = 9$ minutes; ii) the number from one entrance node to another entrance node of the same plant (e.g., from node 'B1' to 'B2') also represents the time needed to patrol the plant. Case (ii) means that the patroller comes into and leaves the plant from different entrances.

Figure 3 shows the patrolling graph pG for the chemical cluster shown in Figure 1, with the data in Table 1 and further assume a patrolling time T = 30. Patroller's base camp is assumed close to the cross road node, thus 'cr' is chosen as the patroller's base camp.

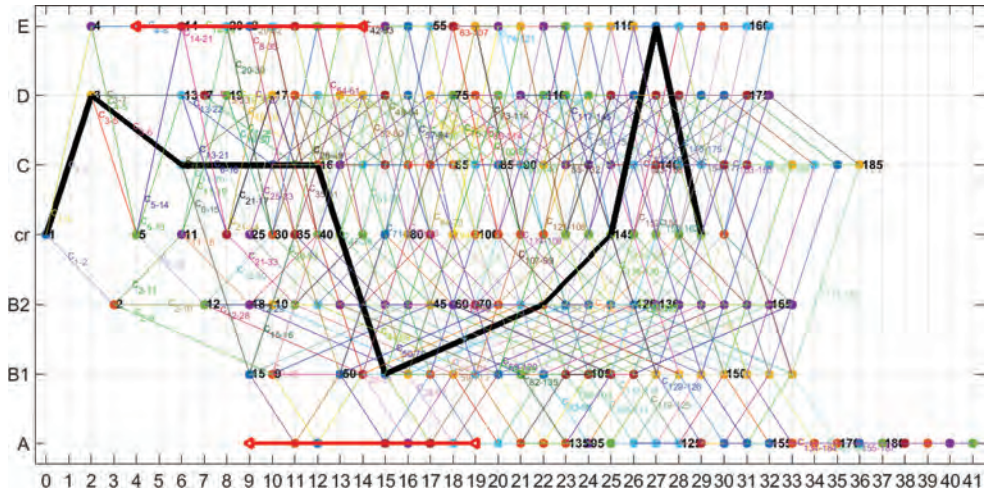


Figure 3. Patrolling graph of the illustrative example.

Table 1. Superior connection matrix for Figure 2 with the illustrative numbers.

	A	B1	B2	cr	C	D	E
A	9	2	∞	∞	∞	∞	∞
B1	2	7	7	∞	3	∞	∞
B2	∞	7	7	3	∞	∞	∞
cr	∞	∞	3	∞	∞	2	2
C	∞	3	∞	∞	6	4	∞
D	∞	∞	∞	2	4	5	∞
E	∞	∞	∞	2	∞	∞	7

In Figure 3, the x axis denotes the time dimension, while the y axis represents the different nodes in Figure 2. Therefore, any coordinates in Figure 3 can be a possible node for pG. As we may see, node 1 (at the left hand side of the figure) in Figure 3 is (0, *cr*), and it means that at time 0, the patroller starts from her base camp (i.e., 'cr'). Thereafter she has 3 choices: i) to come to plant 'B' (more accurately, entrance 'B2') with a driving time t_4^d, and reaches node 2; ii) to come to plant 'D' with a driving time t_5^d, and reaches node 3; and iii) to come to plant 'E' with a driving time t_6^d, and reaches node 4. Subsequently, at new nodes (e.g., 2, 3, or 4), the patroller has the same choice problem, that is, to patrol the current plant or to come to another plant. In Figure 3, the indexes of some nodes and the weight of some edges are not shown, for the clarity of the figure.

Table 2. Model inputs.

	R^d	L^d	G^a	P^a	f_{cpp}
'A'	1	16	10	3	0.45
'B'	1	11.2	6	3	0.3
'C'	1	14	8.3	3	0.42
'D'	1	12	7.1	3	0.45
'E'	1	15	10	3	0.5

For example, the bold (and black) line in the figure, denotes a patrolling route as: '*cr*' → '*C*' → *C*' → patrol plant '*C*' → '*B1*' → patrol plant *B*' → leave plant '*B*' from '*B2*' → '*cr*' → '*E*' → '*cr*'.

Finally, when time comes to the end of the patrol, the patroller terminates the patrolling and comes back to her base camp. In this research, to keep the continuity of coverage of each plant, the patroller is required to prolong their patrolling in the plant until that the next patroller team might be able to arrive the plant. For instance, in Figure 3, though the patrolling time is set as $T = 30$, however, the patrolling in plant 'A' is not stopped until $t = 41$. The reason is that, the shortest time that the next patrolling team can arrive plant 'A' (from 'cr') is 11 (By following a path '*cr*' → '*B2*' → '*B*1' → '*A*'). If the current patroller team does not prolong their patrolling, and the next patroller team starts at time 30 and stars from their base camp (i.e., 'cr'), then plant 'A' will definitely not be covered during time (30, 41). This approach may increase the patroller's workload. However, if we set T slightly smaller than the patroller's real workload, the problem will be solved. For example, if a patroller team's workload is 240 minutes per day, and we may set $T = 220$.

For the sake of clarity, only one type of attacker and only one attack scenario is considered. Further assume that the intrusion and attack procedure of the employed scenario would last for 10 minutes. For instance, the two horizontal bold dot red lines in Figure 3 represent attack strategies that attack plant 'A' start at time 9 (the line at below) and attack plant 'E' start at time 4 (the line at above), with an intrusion and attack period of ten time units, respectively.

Table 2 gives the model inputs, i.e., the defender's reward (loss) of (not) detecting an attacker; the attacker's gain (penalty) from a (not) successful attack; the probability that the intrusion detection system (IDS) can detect the attacker. The probability that the patroller can detect the attacker (i.e., σ_v) should also be provided by security experts. However, in this paper, we simply assume that in each time unit, if the attacker and the patroller stay in the same plant (i.e., overlap), there is a probability of 0.05 that the attacker would be detected by the patroller.

4 RESULTS

4.1 *Stackelberg equilibrium*

Figure 4 shows the Stackelberg equilibrium (SE) of the case study. The black (and narrow) lines demonstrate the patroller's optimal patrolling strategy. The associated numbers on the line denotes the probability that the defender will take this action. For instance, $c1 = 0.22747$ means that at time 0, the patroller should drive to node 'B2' at probability 0.22747. Furthermore, in patrolling practice, if the patroller arrives at a node in the figure, the conditional probabilities of following actions can be calculated as $cP = c/sP$, in which c denotes the probability assigned to the edge, sP_v denotes the probability that the patroller would be at the node. For instance, the probability that the patroller would arrive at the red node $(6,'C')$ in Figure 4 is $sP_v = 0.41734$, and the conditional probabilities that she should take the 2 actions are $cP_1 = 0.4979, cP_2 = 0.5021$.

Table 3. The patroller's actions that may detect the attacker.

Edge	τ	Overlap	σ
25	0.00220	[9,13]	0.20
41	0.09935	[9,16]	0.35
85	0.11143	[11,18]	0.35
159	0.09935	[16,19]	0.15
186	0.00220	[17,19]	0.10
206	0.11143	[18,19]	0.05

Table 4. Comparison of the CCP strategy and the purely randomized strategy.

Edge	Overlap	τ_c	τ_{rc}	σ
82	[11,19]	0.1926	0.0046	0.4
98	[12,19]	0.1942	0.0139	0.35
156	[15,19]	0	0.0019	0.2
176	[16,19]	0	0.0071	0.15
196	[17,19]	0	0.0024	0.1
216	[18,19]	0	0.0039	0.05
425	[9,10]	0	0.0100	0.05
430	[9,11]	0.3358	0.0274	0.1

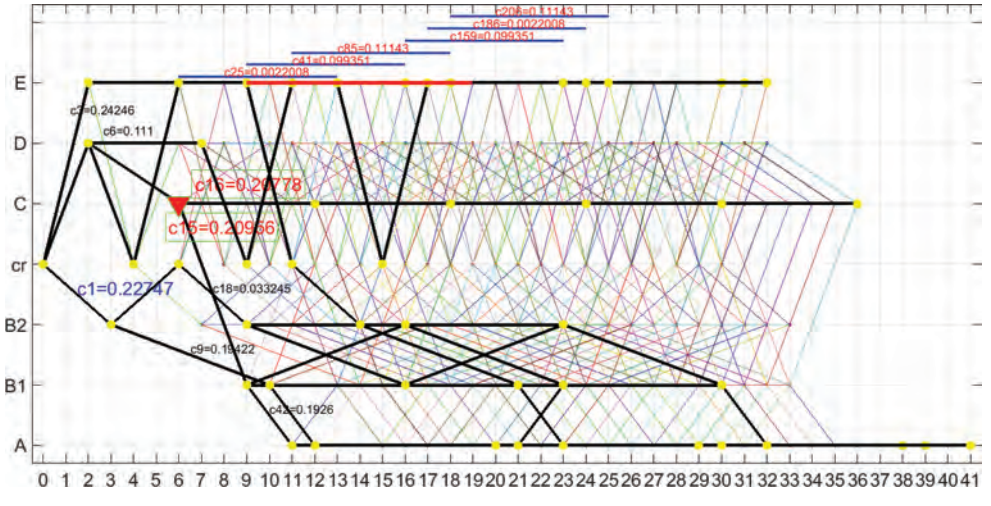


Figure 4. The optimal patrolling strategy and the attacker's best response (Stackelberg equilibrium).

The attacker's best response in the SE is to attack plant 'E' at time 9, as shown in the figure as a red bold line. The short lines above the attacker's best response line represent the defender's patrolling actions which would have overlap with the attacker's strategy. Table 3 shows detail information of these patrolling actions.

Based on the result in Table 3, we have that: $f_p = \sum_r \tau_r \cdot \sigma_r = 0.089101$, $f = 1-(1-0.5)*(1-f_p) = 0.09491$, $u_a = 2.88311$, $u_d = -6.24074$.

4.2 *Comparing to random patrolling*

In the current patrolling practise, patrollers may randomly schedule their patrol route. This situation, one looks at Figure 3, is simply assigning the same probabilities to edges that start from the same node. For instance, at the starting node (i.e., (0,'*cr*')), the patroller would come to plant (entrance) 'B2', 'D', and 'E' at the same probability, and the probability is 1/3.

In the case study, if the defender would purely randomize her patrolling, then the attacker's best response would be attacking plant 'A' at time 9. The attacker and the defender would obtain a payoff of 4.0653 and –8.2393, respectively. Comparing to the result of the CCP game, the defender suffers a higher lose.

Table 4 illustrates the differences between the CCP strategy and the purely randomized strategy. The 'Edge' column in Table 4 shows the edges in the patrolling graph that have an overlap with the attacker's strategy (i.e., attack plant 'A' at time 9). The overlap column illustrates which period of the attack procedure is overlapped by the edge. The 'c' and 'rc' column show the probability that the patroller will go the edge, resulting from the CCP game and from the randomized strategy respectively. The sigma column shows the probability that the attacker will be detected by the patroller by this edge and it is simply calculated as 0.05 multiplied by the overlapped time units. According to the result in Table 4, we can calculate the probability that the attacker would be detected by the patroller (see Formula 7), and the results are: $f_p^c = 0.17860$, $f_p^{rc} = 0.01183$. These results reveal that the CCP strategy has a higher probability of detecting the attacker at plant 'A', and thus transfers the attacker's best response target from plant 'A' to plant 'E'.

5 CONCLUSION

Terrorism has been a global problem. The chemical industry can be an attractive target for terrorists, due to the existence of hazardous materials. A chemical cluster is formed by multiple chemical plants, and can be of extra interest for attackers.

Besides intrusion security countermeasures of each plant, security countermeasures at the cluster level are also recommended. The current patrolling in chemical clusters are either single plant based or purely randomized, being economically not efficient and theoretically not optimized. The security adversaries are human beings, and they may learn the patroller's daily patrolling routes and plan their attack accordingly.

This paper therefore proposes a chemical cluster patrolling (CCP) game. The CCP game generates randomized but strategic patrolling routes for the cluster patrolling team. The intelligent interactions between the patroller and the potential attackers are modelled in the CCP game. An illustrative case study shows that the patrolling routes generated by our CCP game outperforms the purely randomized patrolling strategy.

REFERENCES

[1] Gupta J. The Bhopal gas tragedy: could it have happened in a developed country? *Journal of Loss Prevention in the Process Industries*. 2002;15(1):1–4.

[2] Orum P, Rushing R. Chemical Security 101, What You Don't Have Can't Leak, or Be Blown Up by Terrorists. 2008.

[3] Reniers G, Pavlova Y. Using game theory to improve safety within chemical industrial parks: Springer; 2013.

[4] Tambe M. Security and game theory: algorithms, deployed systems, lessons learned: Cambridge University Press; 2011.

[5] Alpern S, Morton A, Papadaki K. Patrolling games. *Operations Research. 2*011;59(5):1246–1257.

[6] Rezazadeh A, Zhang L, Reniers G, Khakzad N, Cozzani V. Optimal patrol scheduling of hazardous pipelines using game theory. *Process Safety and Environmental Protection*. 2017;109:242–256.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Perception of security and use of public travel modes in an urban Norwegian public

T. Rundmo, A.-M. Kummeneje & T. Nordfjærn
Norwegian University of Science and Technology (NTNU), Trondheim, Norway

ABSTRACT: The aim of the study was to examine judgement of security problems when using public transportation. A self-completion questionnaire survey was carried out among representative samples of residents in six Norwegian urban areas (n = 1047). Respondents who most frequently use public travel modes assessed the security problems to be larger compared to less frequent users. Frequency of use as well as past personal experience of a security problem enhanced the assessment of future probability of experiencing such an event. Perceived risk consistency was also measured and a median split on both these variables was carried out and four groups emerged. The first group was a group of risk insensitive respondents, the second group consisted of risk inconsistent respondents, the third consisted of risk consistent respondents and the fourth was risk sensitive respondents. Travel mode use and personal experience of a security-related problem were positively associated with risk sensitivity.

1 INTRODUCTION

More knowledge is needed to understand how future choices could be moved into a more pro-environmental and safe direction by use of public transport as well as walking and cycling. The current study focuses on the role of perception of security risks by use of public travel modes. Several studies carried out previously have examined perceived risk related to security (probability of violence, acts of terror, etc.) in public transport. Roche-Cerasi, Rundmo, Sigurdson and Moe (2014) showed that there were differences in overall perceived risk evaluations between frequent users of private and public travel mode users in representative samples in Paris and Oslo. The group including those who most frequently used public travel modes perceived the probability of experiencing violence and acts of terror on public travel modes to be larger compared to those who most frequently used car. Frequent public travel mode users were also found to be more worried about security issues on public transportation. In a representative sample of the Norwegian public, Rundmo, Nordfjærn, Iversen et al. (2011) also reported differences in perceived risk evaluations and worry with regard to criminality and acts of terror on public travel modes when comparing a group of respondents who used public travel modes most frequently with a group of respondents who most frequently used private motorised travel modes. Frequent users of public transport perceived the security problems to be more probable compared to frequent users of own car. Nordfjærn, Şimşekoğlu, Lind et al. (2014) showed that perceived risk evaluations and worry related to terrorism, sabotage, theft, harassment and other uncomfortable episodes, as well as violence significantly predicted travel mode use.

The studies presented above primarily examined the role of perceived risk and worry in travel mode use. Accordingly, the current study also hypothesises perceived risk evaluations and worry to be significant predictor variables of travel mode use, i.e. use of public travel modes versus car. In addition, perceived risk evaluations have been distinguished from risk sensitivity (Sjøberg, 1996). Risk sensitivity is the general tendency to perceive all risks as large or small. Consequently, risk sensitivity is linked to perceived risk. The current study aims to examine coincidence between perception and sensitivity in risk judgements.

Rundmo and Moen (2006) showed that perceived risk evaluations in various types of transport were significantly associated. Those who perceived the risk to be large concerning one type of transport also perceived the risk to be large in other areas, and vice versa. Sjöberg (1996, 2004) found that risk amplification-attenuation was a significant predictor variable of personal as well as general perceived risk evaluations. The risk amplification-attenuation framework attempts to explain the process by which risks are amplified, receiving public attention, or attenuated, receiving less public attention. In a random sample of the Swedish public, Sjøberg (1996) showed that risk sensitivity was a significant predictor vari-

able of general perceived risk evaluations among other factors, e.g. trust in authorities, attitudes, and the pooled original psychometric dimensions (see Fischhoff, Slovic, Lichtenstein et al., 1978). In these studies perceived risk evaluations and risk sensitivity were distinguished and measured as two different constructs. However, if risk is rated to be high in one domain it is more likely to be rated as high in another domain, which implies risk sensitivity. Thus, it could be that risk sensitivity is identical to perceived risk evaluations.

In the current study risk sensitivity is conceived to be pooled perceived risk evaluations. While risk evaluations have been found to be equal to hazard perception (Rundmo & Nordfjærn, 2017), risk sensitivity adds to the understanding of the concept by introducing an element of perceived risk that is a general tendency in risk perception which is independent of the object that is perceived. Thus, perceived risk evaluations consist of two elements. The first is linked to the characteristics of the object that is perceived (i.e. subjective assessment of the probability of experiencing a negative event and judgement of severity of consequences if it should occur) and a general characteristic reflected in the sensitivity of risk in general. Risk sensitivity should also be distinguished from perceived risk consistency. Respondents may vary in risk sensitivity, i.e. the level of perceived risk when judging a number of risk sources, as well as in risk consistency, i.e. how consistent or stable the risk level of all the sources is perceived. The current study hypothesises that risk sensitivity and risk stability significantly predicts travel mode use. The current study aims to investigate the role of priority of security and risk sensitivity in use of public travel modes versus car among an urban public. An additional objective is to discuss the relationship between risk perception and risk sensitivity.

2 METHODS

The results of the current study are based on a self-completion questionnaire survey carried out among residents above 18 years of age in the six most urbanised areas in Norway (n = 1043). The response rate was 18% A total of 1043 respondents (18%) replied to the questionnaire. The respondents were asked to assess the occurrence probability of a security problem on a seven-point evaluation scale ranging from 'not at all probable' to 'very probable' (see also Rundmo et al., 2011). Transport mode use was measured by asking the respondents about their weekly ordinary use of transport modes. The questionnaire also contained questions about the demographic characteristics of the respondents, including gender, age, level of education and car access. They were also asked about whether they themselves had been victimised due to a security-related problem when using public transportation and how they prioritised security measures to prevent theft, harassment and acts of terror.

3 RESULTS

The respondents assessed the probability of "theft" and "harassment" to be larger compared to the other types of security problems that were measured. The probability of "sexual assault" and "terrorism" was perceived to be low. A MANCOVA aimed to examine differences in the respondents' assessment of probability for experiencing security problems in such transport was carried out (significant differences shown in bold). Those who most often used public travel modes were compared to those who most frequently used car. There was a significant overall difference in judgement of security problems (Wilks' $\lambda = .95$, $p < .001$). Gender, age group and educational level were covariates in the analysis. The frequent public travel mode users perceived the probability of security problems in general to be larger compared to the group of frequent users of private motorised travel modes. This was the case for assessment of the probability for "theft" ($F = 9.17$, $p < .01$), "sexual assault" ($F = 7.68$, $p < .01$), "harassment" ($F = 7.20$, $p < .01$), and "terrorism" ($F = 9.37$, $p < .01$). There were also tendencies, however not significant, in the same direction for assessments of "blind violence" and "sabotage". Thus, respondents who most frequently used public travel modes assessed the security problems related to use of such modes to be larger compared to less frequent users.

The respondents were also asked to assess the probability of "being too late at work" due to public travel mode delay. In this case the group difference was in the opposite direction. Those who most seldom used public transportation assessed the probability of delay to be larger when using public travel modes compared to the cluster group of frequent users of such modes ($F = 12.33$, $p < .001$).

It could be argued that "being too late at work" is not a security problem in line with the other security problems in public transport. While the other items have to be conceived as the probability of outside inflicted damages, the probability of "being too late at work" is either caused by the traveller or by the operating travel company. In this case it is not a problem inflicted to a victim during the travel. Most often the consequences are more trivial compared to other security problems concerning public transportation. The internal consistency of the judgements of risk sensitivity measured

on a single dimension was satisfactory ($\alpha = .839$). When excluding the item "being too late at work", the reliability improved marginally ($\alpha = .859$). There were large and significant positive associations between the probability assessments of the security problems, i.e. those who assessed one security problem to be large also tended to assess the other security problems in the same way and vice versa, indicating the presence of risk sensitivity. The coefficients varied between .15 and .65.

In addition to probability assessments the respondents were also asked whether or not they had been exposed to one or more security risks when using public transportation during the last five years. A total of 14.5 per cent reported that they had experienced themselves either "theft", "sexual assault", "harassment", "terrorism", or "sabotage" during this time period. Thus, in addition to frequency of use of public travel modes, the respondents' previous experience of public travel mode security hazards have to be taken into consideration. A MANCOVA was carried out to examine the differences in probability assessments of public travel security problems due to risk exposure (frequent users of public versus private transportation) and previous personal experience with security problems when using public transportation. As shown there were significant differences in the judgement of probability due to previous hazard experience (Wilks' $\lambda = 0.87$, $p < .001$).

A median split on risk sensitivity as well as risk stability was carried out and four groups emerged. The first group consisted of respondents who scored low on risk sensitivity as well as on risk consistency ($n = 342$). In addition to rating the probability as low they were characterised by a low Sd, indicating that they judge the overall probability to be low, i.e. consistent low score. This group could be defined as consisting of risk ignorant respondents. The next group consisted of respondents who rated some of the risks to be large and at the same time their evaluations were characterised by a low Sd ($n = 152$). This group was characterised of risk instability and consisted of risk steady respondents. The third group of people who rated most of the probabilities to be low, however, was characterised by a high Sd, indicating a variety in judgements ($n = 188$). This group was entitled risk fluctuating respondents. Finally, there were respondents who rated the probability of all risks to be high, and consequently the group was characterised by a low Sd ($n = 351$). This group was entitled risk sensitive respondents. A total of 52.2 per cent of the respondents scored high on risk stability and 48.7 per cent on risk sensitivity.

The number of risk sensitive female respondents were larger than expected by statistical inference compared to male respondents $\chi 2 = 30.06$, $p < .001$. There were also more risk sensitive respondents than expected among those who had a vocational practical education and a university education less than 3 years. The number of risk insensitive respondents were larger than expected by statistical inference among those who had a university education lasting for more than three years compared to the other groups, $\chi 2 = 23.71$, $p < .05$. The age groups were also compared and risk sensitivity decreased by age, $\chi 2 = 28.71$, $p < .01$.

The respondents were also asked to rate how they prioritized security in transport, i.e. how important it was for themselves. This evaluation included the importance of implementing countermeasures to prevent theft, harassment and terror. The results showed a significant overall difference in priority of security due to risk sensitivity, Wilks' $\lambda = .96$, $p < .001$, gender, Wilks' $\lambda = .7$, $p < .001$, educational level, Wilks' $\lambda = .97$, $p < .001$, and age group, Wilks' $\lambda = .97$, $p < .01$. However, there were no significant differences in priority of security due to past security risk experience, Wilks' $\lambda = .99$, NS. Risk ignorant and risk steady respondents prioritised security measures to a lower extent compared with risk fluctuating and risk sensitive respondents.

The Post Hoc tests showed that there were no statistically significant differences between the security priorities of the groups of risk insensitive and risk stable respondents. The tests showed an identical pattern of differences for all the three assessed factors. The priority of security measures in the risk sensitive group differed significantly from all the other three groups. However, both the two last groups (the groups of risk steady and risk sensitive respondents) differed significantly from risk fluctuating and risk sensitive respondents, $p < .001$. This was the case for priority of countermeasures to reduce the risk of theft, harassment as well as terror. Of note, there were no significant differences between the security priorities of the risk fluctuating and risk sensitive respondents. The pattern was the same for all the three prioritised areas. Thus, risk fluctuation and risk sensitivity seems to enhance the priority of security measures to reduce the risk of theft, harassment as well as terror. Risk perception varied due to previous risk security experience. However, there were no significant differences in priority of risk reduction measures due to past experience of security problems in public transport.

The next step was to examine how priority of security and risk sensitivity predicted travel mode use. Hierarchical and k-means cluster analysis was carried out separately for leisure and work travels to identify mode user groups. The first group consisted of those who most frequently used public transport and the second group consisted of respondents who mainly used car. The same cluster groups emerged for leisure as well as work travels.

The first analysis concerned leisure travelling. As expected the first block consisting of demographic variables, distance from home to the nearest public transport point, and car access significantly predicted leisure travel mode use ($\chi 2 = 213.74$, $p < .001$). Adding priority of security significantly improved the model ($\chi 2 = 16.05$, $p < .05$) and adding risk sensitivity as the final block added further to the explained variance ($\chi 2 = 4.03$, $p < .05$).

When all the blocks were adjusted for in the model, level of education (OR = 1.36, $p < .001$), annual income (OR = 0.67, $p < .01$), and car access (OR = 0.02, $p < .001$) significantly predicted travel mode use. Annual income and access to car were negatively associated with use of public transportation. Gender, minutes to walk from home to the nearest public transport point and previous personal experience with security hazards was non-significant predictor variables. In the second block the significant predictor variables were priority of security against theft (OR = 1.80, $p < .05$) and harassment (OR = 1.41, $p < .001$). Priority of countermeasures to reduce theft was largest among frequent car users and countermeasures for reducing harassment was prioritised among public travel mode users. Finally, the probability of belonging to the group of frequent public travel mode users were larger when risk sensitivity increased (OR = 1.24, $p < .05$).

Concerning work travels, the predictor variables of the first block also significantly predicted work travel mode use ($\chi 2 = 57.62$, $p < .001$). This was also the case for the third block ($\chi 2 = 4.37$, $p < .05$). However, the predictors of the second block seemed not to be equally important for mode use on work travels ($\chi 2 = 1.53$, NS) compared to leisure travels. The significant predictor variables in the final analysis were annual income (OR = 0.69, $p < .05$) and access to car (OR = 0.25, $p < .001$) from the first block. Risk sensitivity also significantly predicted mode use (OR = 1.28, $p < .05$). The prediction of leisure travel mode use (Cox & Snell's $R2 = 0.35$, Negelkerke's $R2 = 0.49$) was more successful than predicting mode use on work travels (Cox & Snell's $R2 = 0.11$, Negelkerke's $R2 = 0.20$).

4 DISCUSSION

The current study showed that priority of security and risk sensitivity was significant predictors of travel mode use among an urban public when demographic factors were controlled for. In studies carried out previously (e.g. Sjøberg, 1994, 2004) risk sensitivity was conceived to be a predictor of "risk perception". However, risk sensitivity is the tendency to perceive all risks to be high and risk insensitivity the opposite. Explaining a large percentage of the variance in "perceived risk" (Sjøberg, 1996, 2004) could partly have been caused by the use of risk sensitivity as a predictor variable which is coincident with the criterion variable. Rundmo and Nordfjærn (2017) found no support of fit of the data to a model where risk perception was conceived to be a formative construct of subjective assessments of probability and judgement of severity of consequences. Subjective judgements of risk should be conceptualised as perceived risk assessments when the judgements are not considered to be a formative construct.

The current study was restricted to examining the probability component of the perceived risk evaluations. In future research the role of severity of consequences should also be included. This could add to explained variance in prediction of travel mode use. Sjøberg (1999) as well as Rundmo and Moen (2006) showed that the subjective judgement severity of consequences if a negative event should occur was a more significant predictor of demand for risk mitigation compared to the probability assessment. Studies carried out previously have shown that subjective assessments of risk and judgement of severity of consequences may predict worry and worry has been found to be associated with demand for risk mitigation in transport (Rundmo & Moen, 2007) as well as mode use preferences (Nordfjærn et al., 2014; Rundmo et al., 2011). Probability assessment was found to be rather insignificant for such demands. The current research did not aim to re-examine the role of worry in travel mode use. It is interesting to note that the assessment of the probability-component of risk sensitivity alone contributed significantly to prediction of travel mode use (public transportation versus use of car).

The results of the current study showed that the same set of predictor variables explained a significantly larger proportion of explained variance in leisure travel mode use compared to work travel mode use. There could be several explanations for this. Most obvious is that the freedom to choose travel modes could be different on the two types of travels. It may be that the freedom of choice is larger for leisure travels compared to work travels. As expected access of private travel modes was an important predictor of travel mode use. In addition, the power of this predictor variable was significantly larger for travel mode use on work travels compared to leisure travels, indicating that it may be easier to choose other travel modes for leisure travels. Also demographic factors seemed to be of less importance for travel mode use in leisure time than for work travels. Other possible explanations, which could be further investigated, include possible differences in the role of habits. Because work travels could have a more repetitive nature than

many travels conducted during the leisure time, habits could play a larger role in mode use on these travels (see also Bamberg et al., 2003).

In this study, perceived risk evaluations and risk sensitivity have been considered to contain the same data of evaluations of probability assessments and judgement of severity of consequences. Consequently, the concepts of perceived risk evaluations and risk sensitivity are unquestionably woven together. They are two parts of the same intuitive evaluation of risk; a direct assessment of probability of an event with negative consequences and the stability or consistency in the evaluation of various risk sources. The first element may vary because it relates closely to the hazard or object of evaluation. The second element is not primarily related to the object, but is a general tendency influencing on the direct evaluation of risk. This element consists of two elements. The first is risk sensitivity, i.e. judging a set of hazards or risk sources on a continuum varying from high (indicating risk sensitivity) to low (indicating risk ignorance). The second element is risk stability, which varies from high (indicating risk stability) to low (indicating risk flexibility). Perceived risk evaluation is only the "basis material" for calculating risk sensitivity. Therefore, risk sensitivity is the main concept, covering the perceived risk evaluations, including intuitive judgments of probability as well as severity of consequences across a set of risk sources.

Further research should examine predictors of risk sensitivity as well as risk stability in further detail. It could be interesting to investigate how aspects of subjective risk judgements not directly associated with characteristics of the risk source influence judgement. Therefore, research on risk sensitivity should be given priority, not the mere analysis of single hazard or risk source evaluation. Further investigations could focus on associations between personality variables and risk sensitivity, which in previous research have been found to be associated with risk perception as well as risk-taking behaviour. Such behaviour has been connected to personality factors e.g. sensation seeking. Another hypothesis is that attitudinal factors, e.g. attitudes towards risk-taking and risky behaviour, may stabilise risk sensitivity on a high or low level, working more or less independently in the judgement of single hazards or risks. The current study showed that priority of security also was associated with travel mode use. The relations between priority of security, personality factors and risk sensitivity should be investigated more thoroughly in future research.

REFERENCES

Bamberg, S., Rölle, D., Weber, C. (2003). Does habitual car use not lead to more resistance to change of travel mode? Transportation, 30, 97–108.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). How safe is safe enough? A psychometric study of attitudes towards technological Read. Policy Studies, 9, 127–152.

Lind, H.B., Nordfjærn, T., Jørgensen, S.H. & Rundmo, T. (2015). Using the Value-Belief-Norm theory to explain personal norms and pro-environmental travel mode use in urban areas. Journal of Environmental Psychology, 44, 119–125.

Nordfjæern, T., Şimşekoğlu, Ö., Lind, H.B., Jørgensen, S.H. & Rundmo, T. (2014). Transport priorities, risk perception and worry associated with mode use and preferences among Norwegian commuters. Accident Analysis and Prevention, 72, 391–400.

Roche-Cerasi, I, Rundmo, T., Sigurdson, J.F. & Moe, D. (2014). Transport mode preferences, risk perception and worry in a Norwegian urban population. Accident Analysis and Prevention, 50, 698–704.

Rundmo, T. & Moen, B.E. (2006). Risk perception and demand for risk mitigation among experts, politicians and lay people in Norway. Journal of Risk Research, 9, 623–640.

Rundmo, T. & Nordfjæern, T. (2017). Does risk perception really exist? Safety Science, 93, 230–240.

Rundmo, T., Nordfjærn, T., Iversen, H.H., Oltedal, S. & Jørgensen, S.H. (2011). The role of risk perception in transportation mode use. Safety Science. 49, 226–235.

Sjøberg, L. (1996). A discussion of the limitations of the psychometric and cultural theory approaches to risk perception. Radiation Protection Dosimetry, 68, 219–225.

Sjøberg, L. (1999). Consequences of perceived risk: Demand for risk mitigation. Journal of Risk Research, 2, 129–149.

Sjøberg, L. (2004). Explaining individual risk perception: The case of nuclear waste. Risk Management: An International Journal, 6, 51–64.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

A systematic classification scheme for cyber-attack taxonomy

S. Kim, J. Shin & G. Heo
Department of Nuclear Engineering, Kyung Hee University, Gyunggi, Korea

J.G. Song
Nuclear ICT Research Division, Korea Atomic Energy Research Institute, Daejeon, Korea

ABSTRACT: Development of digital and network technology has led to a big change in the industry, especially in ICS (Industrial Control System) and SCADA (Supervisory Control And Data Acquisition) system. The components of analogue in facilities are changed into digital components and new facilities of ICS and SCADA systems are composed of various digital instrumentation and control systems. Because of these changes, the security of ICS and SCADA system became an important factor in the industry. Nevertheless, there are few researches on cyber-attack taxonomy for ICS and SCADA systems. Even though some papers and researches suggested a cyber-attack taxonomy, it was not enough or comprehensive for industrial oriented ICS and SCADA systems. Therefore, in this paper, the classification scheme is proposed to classify the cyber-attack taxonomy of PLC (Programmable Logic Controller), DCS (Distributed Control System), and network equipment, which are core components of ICS and SCADA systems. In this paper, cyber-attack is subdivided into foot printing/scanning, password cracking, spoofing, sniffing, hijacking, MITM (Man In The Middle), virus, DoS (Denial of Service), backdoor installation, and hiding files. These ten cyber-attack categories are related with cyber-attack scenario, which is composed of prior preparation, gaining access, maintaining access, and clearing tracks. The grouped categories mentioned above were subdivided according to the characteristics and principles of cyber-attack. The subdivided cyber-attack access path is classified into physical/network, internal/external, and accidental/intentional access. After that, the detailed method of access is investigated. The consequences that can be caused by cyber-attacks are defined as disclosure, modification, destruction, and interruption. Finally, the prevention and mitigation of cyber-attack suggested the specific ways to reduce or escape the damage of cyber-attack consequence.

1 INTRODUCTION

In the ICS (Industrial Control System) and SCADA (Supervisory Control And Data Acquisition) system, the analog instrumentation & control device is being replaced by the digital instrumentation & control device because the digital control system provides better control accuracy and ease of maintenance. However, digital control systems have problems that are subject to cyber-attack. An example of the seriousness of the situation is Stuxnet, which was discovered in 2011, which has caused great damage to the Iranian nuclear infrastructure and major industrial infrastructure in China [1]. This example implies that ICS & SCADA systems, which have closed networks with air gap, are no longer safe from cyber-attack.

Therefore, a new digital instrumentation & control device with cyber security function is required to protect from malicious cyber-attack. In order to apply a new measurement control device with cyber security function to ICS & SCADA system such as nuclear power plants, a compatibility test for cyber security must be performed along with the development of design requirements. In order to conduct a conformance test for cyber security, a systematic classification of attack types and a cyber-attack taxonomy investigation including the latest attack types should be preceded. Most of the preceding cyber-attack taxonomy and researches focused on IT (Information Technology). However, this classification scheme is not suitable for ICS & SCADA systems that are composed of closed networks with air gap and take into account the characteristics of the equipment such as DCS (Distributed Control System) and PLC (Programmable Logic Controller). Therefore, this paper suggested a systematic classification scheme for cyber-attack taxonomy that selectively considered the characteristics of ICS & SCADA systems. The paper classified the cyber-attacks as foot printing & scanning, password cracking, spoofing, sniffing, hijacking, MITM (Man in the middle), virus, DoS (Denial of Service), backdoor installation, and hiding files according to the general cyber-attack scenarios. Then, each category of attacks is subdivided into detailed attacks, access means, result of cyber-attack, vulnerability,

and defense of cyber-attack. In addition, the classification of cyber-attack taxonomy can be applied to identify the attackable digital devices such as DCS, PLC, and network device and the interval of cyber-attack test such as DCS to network device, DCS to PLC, and so on.

2 BACKGROUND

2.1 *General cyber-attack process*

As mentioned above, this paper classified ten categories of cyber-attack. Before the systematic classification scheme for cyber-attack taxonomy, a brief description of the ten categories of cyber-attack is first presented.

Foot printing & scanning is the preliminary task of gathering information about the system to be attacked. Password cracking is the attack that extracts passwords through various methods and means, and can gain full access rights to the system through password cracking. Spoofing is the word meaning cheat, which allows spoofing on any connection that exists on the Internet or locally. Sniffing is an act of peeping packets exchanged by other parties on the network, similar to the dictionary meaning of 'sniff'. Session hijacking is a cyber-attack technique that steals and accesses another person's session state. MITM is the way of intercepting and exchanging information between two parties communicating with each other. A computer virus is a type of malicious software program that can replicate itself by modifying other computer programs and inserting its own code.

A DoS is an attack that causes an attacker to deplete the resource of target so that the user is no longer able to receive services for the resource used by the attacker. Backdoor installation is the method of passing through authentication, ensuring remote access. Hiding files is a process that clears the trace of the attack and deletes the related log to avoid tracing and also hides key files or information for later use or for confidential use of information stored in the victim system.

2.2 *Digital assets for ICS*

Analyzing the digital devices is necessary to make relationship between cyber-attack type and a digital asset. This paper analyzed PLC, DCS, and network device, which are core components of ICS & SCADA systems.

2.2.1 *PLC*

The PLC is a digital control device that can store commands and perform control algorithms that can perform various functions such as logic, operation, counting, and sequential processing to control various types of machines or processes. The PLC is basically the central processing unit that manages and controls all the functionality, a process input / output for exchanging signals, a memory for storing programs, a power unit for supplying electricity to the PLC, and other peripherals [2,16]. The differences between PLC and analog type relays are shown in the Table 1.

2.2.2 *DCS*

The DCS divides one central processing unit to distribute various functions. The overall system configuration is connected to each computer with a small number of central processing units via a communication network. DCS is developed to replace the PID (Proportional Integral Derivative) controller and generally controlled the complex process. The basic feature of DCS is to distribute the process control functions to several computers to improve the reliability and minimize the ripple effect in the event of an error. In addition, it facilitates data processing and operation management by concentrating information, driving operation, and management functions of distributed computers on the Distributed Operate Console. The basic components of the DCS are a CPU (Central Processing Unit) that performs data in real time, a data highway, LAN (Local Area Network) connecting a communication line, a signal input/output unit, an interface for power supply, and a power supply for supplying power [2]. The comparison between PLC and DCS is shown in the following Table 2.

2.2.3 *Network*

The network has a fieldbus, which is a type of digital network system. The fieldbus has been developed to digitally replace analog signals, providing high accuracy and reliability. The fieldbus is a path for transferring network data from the field where PLC-based control equipment exists to the network system between equipment. In addition, there are fieldbus as a controller subnetwork that digitizes/network connection between analog type control devices and field devices, and sensor bus which is a network that transmits sensor signals at high speed. The fieldbus as a controller subnetwork is also classified as a FA (Factory Automation) system and a PA (Process Automation) system. The fieldbus requires only one signal line to transmit signals to the main cable to which a plurality of cells can connect, thereby reducing the wiring cost. It is not only makes it easy for the operator to identify all the devices included in the system, but also facilitates mutual operation between individual devices, thereby lowering the maintenance cost and ensuring reliability. In addition, it has an advantage that system performance

Table 1. Differences between Relay and PLC.

	Relay	PLC
Control method	Hard Logic	Soft Logic
Control function	• Relay (AND, OR by serial/parallel) • Timer • Counter (The function is limited and enlarged according to size)	• Relays (AND, OR, NOT, etc.) • Up/down counter • Arithmetic operation • Logical operation • Transmission (Function is limited and can be enlarged)
Control element	• Reed switch (Limited lifetime, low speed control)	• Solid contact (High reliability, long lifetime, high speed control)
Change content	Disconnection and rewiring of wire	Change through program change
Construction period	• Control panel production after specification • Prolonged inspection and commissioning period	• Specification can be made in parallel with assembly of hard decision • Reduced inspection and commissioning period
Integrity	Difficult to repair	High reliability and easy maintenance
Scalability	Difficulty in expanding the system	• Easy to expand system • Transmission of work information is possible by connecting with computer
Size	Difficult to miniaturize	Possible to miniaturize

Table 2. Differences between PLC and DCS.

	PLC	DCS
Control target	Unit control	Process system
Constitution	PLC	• PLC • Distributed I/O • Touch panel
Response time	Fast (relatively) Suitable for situations requiring real-time action	Late (relatively)
Scalability	Manage thousands of I/O	• Manage hundreds of thousands of I/O • Suitable for advanced process control
Complexity	Impossible to request PLC based control system	Better than PLC
Process change	Suitable for processes that are not frequently changed	Ideal for complex or coordinated analysis that combines large amounts of data

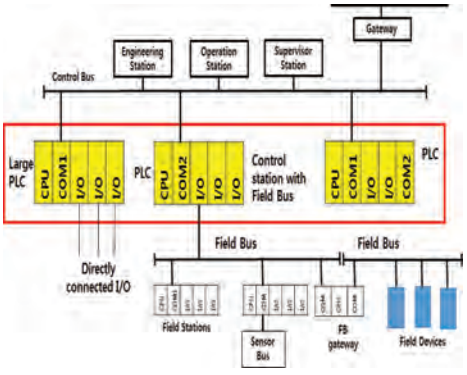


Figure 1. Typical structure of PLC.

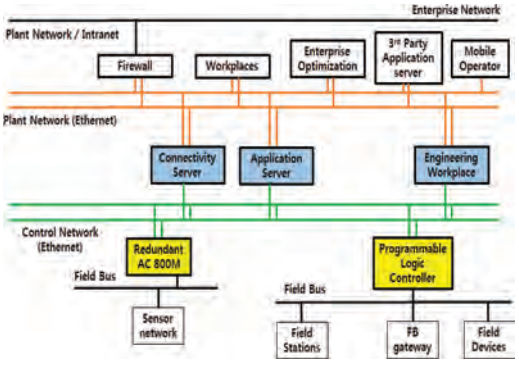


Figure 2. Structure of DCS.

can be improved by facilitating the information collection of field devices. Typical types of these field buses are Modbus, Interbus, Profibus, RS232C/RS485, and Ethernet TCP (Transmission Control Protocol)/IP (Internet Protocol) [9].

3 CLASSIFICATION SCHEME FOR CYBER-ATTACK

3.1 *Subdivision of attack categories*

This paper classified cyber-attack scenario as four stages: prior preparation, gaining access, maintaining access, and clearing tracks [18].

Prior preparation for cyber-attack includes information that target IP range, namespace, vulnerability, and protocol. Gaining access is to bypass access controls to gain access to the system. Maintaining access is retaining ownership of the system. Clearing track is the activities carried out by an attacker to hide malicious acts.

The step after classifying the cyber-attack in terms of cyber-attack scenario is to further subdivide the cyber-attack. The reason for subdivision of cyber-attacks is that even if the basic principles of cyber – attacks are the same, the conditions and vulnerabilities for the cyber-attacks are different, and the attack results and countermeasures are not the same accordingly. Also, if the cyber-attack is further subdivided and complementary measures are taken, it can be a more stable system. As a typical example, password cracking can be subdivided into keylogger attack, dictionary attack, hybrid attack, brute-force attack, and precomputed Hashes attack.

With these definitions, ten types of attacks were matched to the cyber-attack scenario and Table 3 shows their relationship.

Table 3. Cyber-attack scenario with corresponding attack.

Cyber-attack scenario	Corresponding Attack
Prior preparation Gathering information for attack	Foot printing & Scanning Spoofing, sniffing, MITM
Gaining access Bypass access controls to gain access to the system.	Password cracking Hijacking, DoS
Maintaining access Retain ownership of the system	Virus, Backdoors
Clearing tracks The activities carried out by an attacker to hide malicious acts	Hiding files

3.2 *Classification scheme for cyber-attack*

Generally, the cyber-attack scenario proceeds as follows. In the prior preparation stage, the attacker will identify the attack target information, attack through the designated access, and achieve the desired result. After that, security vendors will develop a defense. This paper focused on the general cyber-attack scenarios and selected items to be considered in the attack classification system as method of access, outcome, vulnerability, and defense. The Figure 3 shows the overall flow chart of cyber-attack taxonomy.

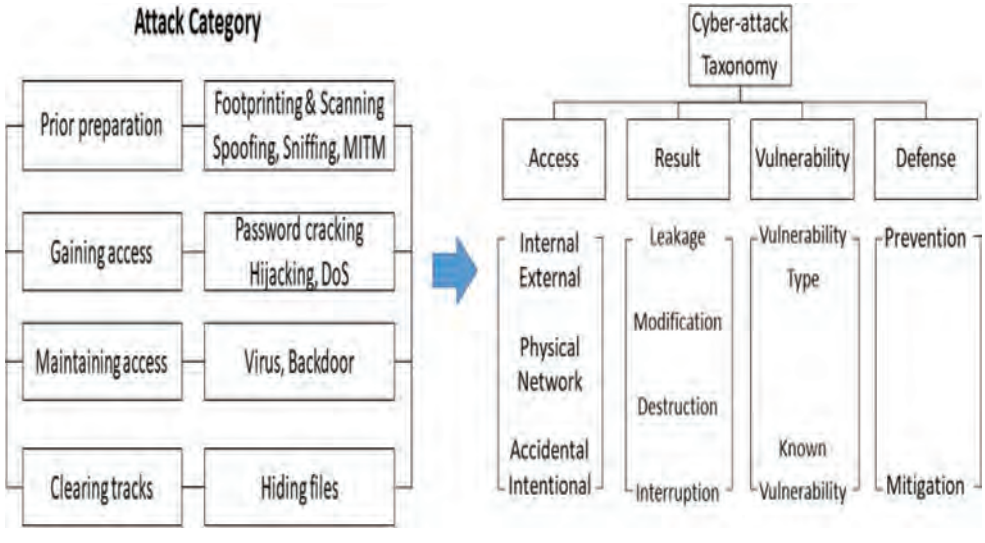


Figure 3. Flowchart of cyber-attack taxonomy.

3.2.1 *Access means of cyber-attack*

The first of the cyber-attack classification is the access of cyber-attack. The ICS & SCADA system is disconnected with the external network, so usually attack of internal is mainly considered. However, in consideration of recent ICS & SCADA system accidents such as Stuxnet, external attacks are also importantly considered as the accessible path of cyber-attack. After distinguishing internal/external attack, considered access of cyber-attack is physical access which include USB or CD and network access which using communication network [4]. This can be used to determine whether physical security is required or network and software security is required to increase the security of system. The last one considered is whether an attack is intentional or accidental. Considering only intentional attacks, there is a limit to the availability guarantee. Therefore, in order to provide a high level of availability and security for ICS operation, accidental access such as misconfiguration of ICS components and ICS equipment failure are considered [2,7].

3.2.2 *Attack result from cyber-attack*

The next classification is based on attack result. In order to derive the attack result, ten cyber-attack categories are considered which were earlier mentioned. For example, in the case of foot printing & scanning, the purpose is to leak data by secretly viewing the data. In the case of DoS, the goal is to destroy the system by putting the system and server in a saturation state. In the case of MITM, the purpose is to modify the data between the client and the server. Considering purpose of the remaining cyber-attacks, the results of the attack are classified as leakage, modification, destruction, and interruption [7].

3.2.3 *Vulnerability with cyber-attack*

The next classification is based on the vulnerability. With the development of information technology, the complexity of software has increased and vulnerabilities have also increased. Vulnerability is a flaw that can be security threat in hardware or software and can be exploited directly by a hacker or as a means of spreading a virus or malicious program [15]. As a result, some countries have databased vulnerabilities and have developed countermeasures against them. This enables early response to specific cyber-attack. This paper also considered vulnerability as one of the classification items, so that it can be used for early response by matching with vulnerability in case of specific cyber-attack. List the vulnerabilities of the software used in ICS and SCADA systems and investigate where the software is used to know anticipated cyber-attacks in the system and use them for countermeasures and mitigations in case of cyber-attack. In addition, the newly discovered vulnerabilities can be continuously matched and added to the cyber-attack, which can effectively upgrade and supplement the system. Vulnerability databases include USA's NVD (National Vulnerability Database), Japan's JVN (Japan Vulnerability Notes) and CNVD (China National Vulnerability Database). In this paper, NVD based on NIST (National Institute of Standards and Technology), which is widely used, is selected and shown in Table 4 [15].

3.2.4 *Defense from cyber-attack*

The last classification is the defense from cyber-attack. There are two kinds of method to reduce the damage from cyber-attack. One is the prevention of cyber-attack and the other is the mitigation of the cyber-attack [17].

Considering typical vulnerabilities of ICS is a way to prevent cyber-attack. Typical vulnerabilities of ICS are unauthorized protocols, aged hardware, vulnerable user authentication, vulnerable file integrity checks, vulnerable Windows operating systems, and undocumented third party relationships [13].

Table 4. Example for usage of cyber-attack taxonomy.

Attack Category	Subdivided Attack	Definition of Subdivided Cyber-attack				Defense		
		RST hijacking is a kind of TCP/IP hijacking in which RST packets are injected. ... (Omitted below)				The countermeasures of RST Hijacking are to reduce the SYN-received latency, or to block RST packets at the router or firewall. … (Omitted below)		
Session Hijacking	RST Hijacking	Access Method (1)	Access Method (2)	Attack Method & Condition	Attackable OS, CPU, S/W	Attack Result	Known Vulnerability	Vulnerability Type
		Internal External	NA: TCP/IP	TCP vulnerability (3-handshake)	OS: Window ver, Linux 3.2.24	Interruption	CVE-2002–1778	Other

The prevention of the cyber-attack which derived from ICS vulnerabilities are antivirus, firewall, one-way gate, and media control. Antivirus is the use of signature-based engines with solutions to detect and treat malware. Firewall is used to allow and block information transmitted during data transmission between ICS networks. One-way gate is a method of establishing a physical unidirectional environment to intercept the intrusion when linking infrastructure network and external network data. Media control is the disconnection of an interface control device such as a USB connected to an ICS network [15]. There is mitigation as a way to reduce the damage after the cyber-attack. Treatment and mitigation is the method needed to normalize the system after a cyber-attack. Treatment and mitigation are various like as vulnerability. As mentioned above, the mitigation is linked with vulnerability and can be used for early response after cyber-attack. With these classifications, Table 4 shows the example of systematic classification scheme for cyber-attack taxonomy.

3.3 *Example of cyber-attack taxonomy utilization*

The surveyed cyber-attack taxonomy can be used to determine attack target, attack area, and test method considering the principles and characteristics of PLC, DCS, and network device of ICS/SCADA system proposed in 2.2. PLC, DCS, and network devices used in the actual industrial field vary, and security performance and vulnerabilities are also different. The PLC, DCS, and network devices mentioned in this paper are assumed to have general performance and characteristics.

An example of cyber-attack, which shows the usage of cyber-attack taxonomy, is the RST hijacking. RST hijacking is a kind of TCP/IP hijacking in which RST packets are injected into the network approach. RST Hijacking can observe the communication between the client and the server as well as TCP session hijacking. Also, it is possible to extort the session using the trust and the session using the TCP. This attack exploits the weaknesses of the TCP three handshake method, which can cause the device to overload and stop functioning similar to a DoS attack. By applying the information of the cyber-attack taxonomy investigated to the ICS & SCADA system, the attackable device can be selected by the DCS and the network because RST hijacking is intended for communication between client and server like DCS. It is expected that the test interval in the cyber security test can be between DCS and network. It means that knowing of the particular cyber-attack, which occurs in a particular section of the ICS and SCADA system, can reduce the effort and cost of defending and mitigating from cyber-attack. Also, if many cyber-attacks are systematically classified and accumulated like RST-hijacking mentioned above, the user or operator will be able to quickly recognize the type of attack when an arbitrary cyber-attack is applied. It means that the user or operator can take action quickly.

4 CONCLUSION & FUTURE WORK

This paper classified cyber-attack scenario as four stages such as prior preparation, gaining access, maintaining access, and clearing tracks matched with the ten categories of cyber-attack such as footprinting, scanning, password attack, spoofing, sniffing, hijacking, MITM, virus, DoS, backdoor installation and hiding files. Then, ten kinds of cyber-attacks are subdivided and classified into access means, result of cyber-attack, vulnerability, and defense according to the classification scheme for cyber-attack. Access of cyber-attack can be classified with internal/external, physical/network, and intentional/accidental. Result of cyber-attack can be classified with leakage, modification, destruction, and interruption. Vulnerability of cyber-attack is based on NVD and it can be used for early response from cyber-attack when it combined with defense of cyber-attack. Defense of cyber-attack can be classified with prevention and mitigation. Through this classification system, the cyber-attacks and the possible intervals in ICS are exemplified. Future works include further subdivision of the ten categories of cyber-attack, and using it to prove possible attacks and attack intervals through penetration testing in a cyber-attackable experimental environment.

ACKNOWLEDGEMENT

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20161510101830).

REFERENCES

[1] Bonnie Zhu, Anthony Joseph, and Shankar Sastry (2011). A Taxonomy of Cyber Attacks on SCADA Systems. 2011 International Conference on and 4th International Conference on Cyber, 380–388.

[2] Jung-Chan Na, and Hyun-Sook Cho (2013). Classification of ICS abnormal behavior in terms of security. Electronics and Communications Research Institute, Vol. 23 No.2.

[3] Carl E. Landwehr, Alan R. Bull, John P. Mcdermott, and William S. Choi (1993). A Taxonomy of Computer Program Security Flaws, with Examples. NRL/FR/5542–93–9591.

[4] C. Meyers, S. Powers, and D. Faissol (2009). Taxonomies of Cyber Adversaries and Attacks: A Survey of Incidents and Approaches. Lawrence Livermore National Lab.

[5] Crispin Cowan, Perry Wagle, Calton Pu, Steve Beattie, and Jonathan Walpole (2000). Buffer Overflows: Attacks and Defenses for the Vulnerability of the Decade. DARPA Information Survivability Conference and Exposition.

[6] Hossain Shahriar, and Mohammad Zulkernine (2011). Taxonomy and classification of automatic monitoring of program security vulnerability exploitations. *Journal of Systems and Software*, Vol. 84, 250–269.

[7] James J. cebula, and Lisa R. Young (2010). A taxonomy of operational cyber security risks. CERT Carnegie Mellon University.

[8] Jelena Mirkovic and Peter Reiher (2004). A Taxonomy of DDoS Attack and DDoS Defense Mechanisms. SIGCOMM Computer Communication Review. 39–53.

[9] Justin King, Kiran Lakkaraju, and Adam Slagell (2009). A taxonomy and adversarial model for attacks against network log anonymization. ACM Symposium on Applied Computing.

[10] Maria Kjaerland (2005). A taxonomy and comparison of computer security incidents from the commercial and government sectors. *Computers and Security*. Vol.25, 522–538.

[11] Simon Hansman and Ray Hunt (2005). A Taxonomy of Network and Computer Attacks. Computer and Security, Vol.24, 31–43.

[12] Suhair Hafez Amer and John A. Hamilton Jr. (2010). Intrusion Detection Systems (IDS) Taxonomy—A Short Review. Defense Cyber Security: Policies and Procedures.

[13] Lee Neitzel and Bob Huba (2014). Top ten differences between ICS and IT cyber security, InTech.

[14] NIST (2015). Guide to Industrial Control Systems (ICS) Security. NIST 800–82(Rev.2).

[15] Du-Sun Yoon (2014). Quantitative Analysis of Vulnerabilities in Database Management System Using Vulnerability Standards. Chungbuk National University.

[16] PLC code vulnerabilities through SCADA systems (2013). Sidney E Valentine. University of South Carolina.

[17] AVOIDIT: A Cyber Attack Taxonomy (2009). Chris Simmons, Charles Ellis, Sajjan Shiva, Dipankar Dasgupta, and Qishi Wu. Department of Computer Science University of Memphis Memphis.

[18] CEH (Certified Ethical Hacker) (2014). EC-Council official curriculum. Official courseware Vol 1.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Organizational information security culture in critical infrastructure: Developing and testing a scale and its relationships to other measures of information security

T.O. Nævestad & S. Frislid Meyer
Institute of Transport Economics, Oslo, Norway

J. Hovland Honerud
University College of Southeast Norway, Norway

ABSTRACT: The aims of the present study are to 1) develop and test a scale measuring organizational information security culture, and 2) examine its relationships to other aspects of information security. The study focuses on an organization providing critical infrastructure. We developed the scale by conducting qualitative interviews (N = 22) and three focus groups (N = 15) in an organization providing critical infrastructure, and by reviewing previous research on culture in organisations. Based on our literature review and the interviews, we chose to measure organizational information security culture by reformulating one of the few existing general organizational safety culture questionnaires. We first tested the questionnaire in a small pilot survey, and then conducted a questionnaire survey (N = 323) including all departments in the organization. Our examination of the factor structure of the scale indicated two factors. Regression analyses indicate that our adapted GAIN-scale, measuring organizational information security culture is the most important variable influencing information security behavior in the model.

1 INTRODUCTION

1.1 *Background*

Information security is often defined as protection against breaches of confidentiality, integrity and accessibility. This applies to information that is oral, written or electronic. Confidentiality refers to ensuring that only those who are authorised to access information, accesses it. Integrity refers to protecting the accuracy and entirety of information and processing methods. Accessibility refers to ensuring that authorised users have access to the information and associated equipments when necessary (Report to the Storting 29. 2011–2012).

Ruighaver et al (2007) assert that it was not until the start of the century that scholars began to recognise the importance of organizational information security culture for information systems security in organisations. The importance of culture for security and safety has also gained recognition in the Norwegian society in recent years. One of the most important conclusions of the report of the investigation commission following the terrorist attack in Oslo and Utøya, July 22. 2011 was that future efforts to secure sensitive objects (e.g. people and critical infrastructure) and information should focus on culture, focusing especially on the acknowledgement of risk and leadership.

The study organization is a provider of critical infrastructure in Norway. As a provider of critical infrastructure, the study organization is obliged to follow the requirements of the Security Act ("Sikkerhetsloven") when it comes to preventive safety work, which includes safety analyses, securing objects, information security and safety drill. Based on these requirements, the study organization decided to map and analyse their own organizational security culture. Critical infrastructure means the facilities and systems that are completely necessary to maintain society's critical functions, which in turn meet society's basic needs and respond to the population's need for a perception of safety (NOU 2006).

1.2 *Aims*

The aims of the present study are to 1) develop and test a scale measuring organizational information security culture and 2) examine its relationships to other aspects of information security.

1.3 *Research on culture in organisations*

1.3.1 *Organisational information security culture*

Although Ruighaver et al. (2007) note that the organisational security culture concept has gained recognition, they also underline that there is

lacking consensus when it comes to how the concept should be defined and conceptualized (cf. Chia et al., 2002). Additionally, they also assert that in spite a large amount of research on organisational security and how it should be improved, this research only focus on certain aspects of security and not how these aspects can be analysed as part of a larger organisational culture.

Based on this understanding, Ruighaver et al. (2007) choose to draw on organisational culture research in their analysis of security culture. This approach is similar to that applied by scholars studying organisational safety culture, who analyse safety culture as a focused and safety relevant aspect of the larger organisational culture (e.g. Hale, 2000, Haukelid, 2008, Antonsen, 2009). Based on this, we may also analyse security culture as "security relevant" aspects of the larger organisational culture, define and conceptualising using models of organisational culture (e.g. Schein, 2004). In this paper, we suggest that the research on information security culture could learn from the research on safety culture. Nosworthy (2000) asserts for instance that one of the key challenges of information security culture implementation is how to educate the people of the organization to successfully implement the requirements of the information security policy. A lot of effort has been put in to understand this in safety culture research, discerning between formal (structure; safety management system; procedures, training, routines etc.) aspects of safety and informal aspects (culture) (Antonsen, 2009). Additionally, Knapp et al. (2006), depict the top management support as a significant predictor of an organization's security culture and level of policy enforcement. This also reflects a key finding in organizational safety culture and safety culture research, and thus it is relevant to also draw on the knowledge gained in these research fields.

1.3.2 *Organisational culture*

The influential scholar Schein defines organizational culture as: "(...) a pattern of shared basic assumptions that the group learned as it solved its problems of external adaptation and internal integration that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think and feel in relation to those problems." (Schein 1992: 12). According to Guldenmund (2000: 222–225), organizational culture has the following characteristics: 1. It is a construct in the sense that it is an abstract, not a concrete concept, 2. It is relatively stable, 3. It has multiple dimensionality, in the sense that it can be described in many different ways, 4. It is shared by groups of people, 5. It consists of various aspects, which means that several different cultures can be identified within organizations, depending on the issue at hand, 6. It constitutes practices, 7. It is functional. Guldenmund describes organizational culture in the following manner: "Overall, organisational culture is a relatively stable, multidimensional, holistic construct shared by (groups of) organisational members that supplies a frame of reference and which gives meaning to and/or is typically revealed in certain practices." (Guldenmund 2000: 225).

As the research on organizational safety culture seems to have been through many of the challenges that the organizational security culture research now is facing, we draw on the experiences of the former, e.g. when it comes to analyzing security culture as a focused aspect of organizational culture.

1.3.3 *Organizational safety culture*

Even though the concept of safety culture has become popular since it first was introduced in the wake of the Chernobyl accident in 1986, it is not well understood (Reason, 1997). Safety culture scholars may disagree on a range of different issues, but they seem to agree that the research on safety culture and its relationship with safety is fragmented and unsystematic (e.g., Cox & Flin, 199, Pidgeon, 1998, Hopkins, 2006, Guldenmund, 2007; Choudry et al., 2007; Glendon, 2008). In spite of this disagreement, most scholars seem to agree that safety culture refers to shared and safety relevant ways of thinking or acting that are (re) created through the joint negotiation of people in social settings (Nævestad, 2010a), and as noted as safety-relevant aspects of organisational culture (Hale 2000). The element of safety culture that can be measured is often referred to as safety climate. Thus, safety climate can be conceived of as "snapshots", or manifestations of safety culture (Cox & Flin, 1998). Quantitative measurements of safety culture can provide leading indicators of safety and consequently offer predictive assessments that enable safety improvements without having to wait for accidents or incidents to happen (Antonsen, 2009). Senior management commitment to safety is the most studied and best-documented characteristic of a good safety culture, independent of sector (Flin et al. 2000; Guldenmund 2000).

2 METHOD

Our methodological approach is based on a literature review conducted in 2012, interviews (N = 22) and focus groups (N = 15) in 2014 and survey in 2014 (N = 323).

2.1 *Interviews and focus groups*

We started with the qualitative part of the study before conducting the quantitative survey, so that

we could form a picture of key issues concerning safety and security work at the study organisation. This is important because we had to adapt the security culture questionnaire, and because it gave us the opportunity to add questions that are central to safety and security work at the study organisation to the questionnaire. We have conducted 22 in-depth interviews, primarily managers, and one group interview with 3 respondents. Focus groups, primarily employees: 2 focus groups with a total of 12 persons. This makes a total of 37 in-depth interviews.

We used a semi-structured and relatively open interview guide based on the safety and security culture topics from the safety culture index. Our point of departure was topics related to information security and the protection of critical infrastructure. The interview guide had to be open so that we could depend on the interviewees' understanding of how different features of the organisation culture in the study organisation have had and can have consequences for safety and security.

The interviews were built up around the following main topics: 1) In general about the department's and respondent's responsibility and roles, 2) Security focus, and relation to safety in the information security, HSE safety, deliverance reliability, 3) Organizational framework, management lines and communication, 4) Safety culture issues; safety, training, expertise, procedures, etc.

2.2 *Literature review*

The literature review was originally conducted as part of another project (cf. Nævestad & Bjørnskau 2012), but we nevertheless draw on it in the present study, as it also was relevant to the present study, and as our choice of safety culture scale for the present study was based on it. This is based on our mentioned ambition to learn from the safety culture literature when measuring and understanding organisational information security culture.

In this review, we conducted literature searches for articles and reports that document experiences with different safety culture measurement tools. We conducted searches through two key scientific databases, "Science direct" and the ISI web of science. A search for "Safety climate" in "abstract/title/key words", "safety climate scale" and "safety climate questionnaire" in scientific publications (primarily referenced journal articles, but also some books) for all years, gave everything in all 249 results. The next search we made from the scientific database "ISI-Web of Knowledge". Here we searched for articles with "safety climate" in title or subject, for all years, and received 458 hits.

The scales were reviewed according to the following criteria, whether: 1) they are based on a solid scientific approach (e.g. based on previous research and existing theory, have been validated in several studies), 2) they are universal, 3) they are) user-friendly; do not include too many themes and questions, which are understandable for people who are not researchers and 4) A key criterion has been that the themes and the items in the scales are in accordance with the key results of the interviews and focus groups. Our review resulted in 11 scales that we perceived as relevant enough to be evaluated systematically against these criteria.

In the present study, we choose to reformulate one of the few existing universal organizational safety culture scales, the GAIN-scale for safety culture, into an organizational security culture scale. The GAIN scale was chosen first, as our previous literature review, conclude that this is one of very few universal safety culture surveys (Nævestad & Bjørnskau, 2012). Thus, the wording of each item can be adapted to different sectors (and presumably also to security) without obviously altering the particular aspect which that item measures. Thus, the scale has the potential to be developed as a generic measure.

Second, the scale was chosen, as it is founded on a relatively solid scientific foundation. It must be noted that we ended up recommending another scale in the above mentioned 2012 review. In this review, we chose the NOSAQ-50 scale (Kines et al., 2011), over the GAIN-scales (GAIN, 2001), as this had been subjected to a more systematic literature review. We have however conducted several studies using the GAIN scale since 2012 (e.g. Nævestad & Bjørnskau, 2014, Nævestad et al., 2017, Nævestad, 2017), subjecting it to exploratory and confirmatory factor analyses, and we have also analyzed the relationship between the scale and safety outcomes (e.g. Nævestad, 2017). The scale has also been used to study and compare safety culture in different transport sectors like road, rail, helicopter and aviation (Bjørnskau & Longva, 2009).

Third, the scale was chosen, as it is relatively easy to use. The GAIN-scale has for instance considerably shorter than the NOSAQ 50; it has only half the items. Additionally, the questions are relatively short, and it is relatively easy to change and adapt the wording to information security culture.

2.3 *Survey*

2.3.1 *Sample characteristics*

A total of 323 individuals responded to the survey, from 11 different departments, giving a response rate of 56%. More than 90 per cent of the respondents are permanent employees and seven per cent are hired consultants. We also see that seven per cent are section or department managers. It should also be mentioned that more than 50% those who responded to the survey had been employed by the

study organisation for five years or less. This is a relatively high percentage. It explains that more than 40 per cent of the respondents have been employed by 3–5 other business in their working life before the study organisation. Almost a quarter of those who responded have however been employed for more than 20 years. This is an approximate reflection of the actual distribution in the study organisation, but respondents with the shortest seniority are overrepresented. 56 per cent of the respondents are above the age of 46. This is interesting, considering that around half have seniority of five years or less. 61 per cent of the respondents are men and 66 per cent have graduated from university/university college.

2.3.2 *Pilot survey*

As we developed several new questions in the survey which had never been tested before, we conducted a small pilot study (N = 12) directed at personnel in the study organisation to obtain feedback and assess how the questions worked. In the pilot study we received some useful feedback, including that we should use the term "my immediate supervisor" rather than "my department manager" in the survey on safety culture and information security culture

2.3.3 *Survey topics*

The survey contains mainly questions about ten topics. It first contains a set of background questions (e.g. gender, age, education, experience, level) that were sent to all respondents. In addition, three short indexes follow with questions about three different types of security related to information security, HSE safety and deliverance reliability. The questions are identical and have the same scale so that we can directly compare the meanings of the three forms of safety and security in the study organisation. Furthermore, the questionnaire 13 contains questions about attitudes and behaviour regarding information security culture.

2.3.3.1 Background variables

The survey also includes questions on demographic background variables and various performance targets related to safety. The background variables include information on: 1) gender, 2) age, 3) education, 4) seniority, 5) employment in other businesses, 6) level in the organization and 7) employment status in the organization (permanent, hired). These background variables are only presented at the enterprise level.

2.3.3.2 The GAIN scale

Global Aviation Information Network (GAIN) is a voluntary association of airlines, manufacturers, trade unions, governments and other organisations in aviation. The GAIN questionnaire contains 24 questions concerning (we excluded one of the original questions, because of the wording):

1. Management's attitude to and focus on safety:

Man 1: My immediate supervisor discovers employees who fail to take sufficient considerations to information security in their work
Man 2: My immediate supervisor often praises employees for maintaining information security
Man 3: My immediate supervisor is aware of the most important information security issues in the company
Man 4: My immediate supervisor often discusses information security issues with the employees
Man 5: My immediate supervisor is personally involved in activities to improve information security
Man 6: My immediate supervisor postpones tasks/activities if information security is not sufficiently ensured
Man 7: My immediate supervisor considers information security to be very important in all tasks and activities
Man 8: My immediate supervisor does everything he/she can do to avoid breaches of information security

2. Employees' attitudes to and focus on safety:

Emp 1; My colleagues do everything they can to avoid breaches of information security
Emp 2: Employees encourage one another to safeguard information security
Emp 3: Employees usually report all breaches and irregularities related to information security that they experience at work

3. Reporting culture and reactions to incident reporting:

Rep 1: Those who pursue breaches of information security in the business attempt to find the real causes rather than just blaming the employees
Rep 2: There are routines and procedures at my workplace so that I may report information security-related breaches or irregularities
Rep 3: After a breach of information security, measures are implemented to prevent this from happening again
Rep 4: All irregularities and information security issues that are reported are remedied in a short time
Rep 5: Everyone has plenty of opportunities to forward suggestions related to information security

4. Safety training and education:

Tra 1: Employees in my company are provided with adequate training in the secure use of ICT systems (e.g. e-mail, storage, encryption)
Tra 2: All new employees are provided with adequate training for tasks and the secure use of ICT systems (e.g. e-mail, storage, encryption)
Tra 3: Everyone is provided with sufficient feedback on how the enterprise is performing with regard to information security

Tra 4: Everyone is informed of any changes that may impact information security

5. General questions concerning safety in the organization in question:

Gen 1: There are procedures that must be followed in the event of an emergency situation in my workplace
Gen 2: Information security in my business is better that in other businesses
Gen 3: Regular security audits are carried out
Gen 4:_Information security is generally well taken care of at my workplace

Respondents can rate the questions from 1 (totally disagree) to 5 (totally agree). Thus, a safety culture index with a minimum value of 24 (1 × 24) and a maximum value of 120 (5 × 24) can be compared across companies and sectors. According to GAIN (2001), organizations with a score of 93–125 points on the safety culture index have a positive safety culture, 59–92 indicates a bureaucratic safety culture and 25–58 indicates a poor safety culture.

2.3.3.3 Questions about information security
Based on the interviews (and literature review of organizational security culture scales, e.g. SjekkIT developed by NTNU and Sintef for the Norwegian National Security Authority (NSM), we also included 22 additional questions about information security in the organisation. These were themes representing special information security challenges in the organization.

3 RESULTS FROM INTERVIEWS AND FOCUS GROUPS

In the interviews, we discussed organisational security management with the key managers, and based on the interviews, we found that they perceived the five GAIN themes as important and relevant. Management and employee commitment for safety was perceived as key. The organisation had also developed a reporting system covering information security, and they were also engaged in several initiatives to educate employees in information security issues.

Based on the interviews and focus groups, we also developed 22 survey questions, reflecting the most important information security challenges in the organisaion. We included several untested questions among the 22, and we experienced that nine of these did not work because some of them had relatively large shares of "neither/nor" responses. The 13 questions on information security we ended up with after removing the 9 that did not work (of 22 questions in total) may be divided into the following topics. We unfortunately don't have the opportunity to present all here due to space considerations, but will nevertheless reproduce the topics and questions, as they are an important result of our qualitative surveys:

1. Knowledge/attitudes—information security:

We constructed an index for knowledge of and attitudes to information security with five questions. All of the questions have five values, such that the minimum value for the index is 5 and the maximum value is 25 (Cronbach's Alpha = 0.740).

In my work, I have a clear understanding of what the term information security means.
I have a clear understanding of what entails a breach of information security in my business.
I feel that I have adequate knowledge on the secure use of ICT systems (e.g. e-mail, storage, encryption)
All unfamiliar persons at the workplace are noticed, and one investigates what they are doing there.
When I am asked for information, I always think carefully about whether the information can be used for other purposes than originally intended.

2. Security assessment—PC and cell phone:

My cell phone contains sensitive information.
If I'm working on a PC from home, information security is just as high as it is at work

3. Classified information and accessibility.

We created an index of the following three questions on classified information (Cronbach's Alpha was 0.721):

I am well aware of which type of information that is sensitive and classified
I am well aware of who has access to various types of classified information
I take precautions when I come into contact with sensitive and classified information

The respondents were also asked to consider the following statement: "Considerations to information security (for example passwords to log on) impede my work".

4 RESULTS FROM THE QUANTITATIVE SURVEY

4.1 *Clear understanding of information security?*

The questionnaire that measures information security culture is based on the GAIN safety culture index, where the word "safety" is replaced by "information security." The questionnaire opened with definitions of information security and related sub-concepts.

4.2 *Organisational information security culture index*

We have combined the 24 statements with five response options on the five different aspects of information security in an information security culture index. The indexes for the departments correspond to the average scores for the respondents. Since we have removed a statement from the GAIN index, the minimum score will be 24 (24×1) and the maximum score will be 120 (24×5). Cronbach's Alpha for the 24 questions in the index is 0.913, which means very good agreement between the questions and that the index is very good.

Factor 1: Information security management and commitment	**Factor loadings**
Man 4: My immediate supervisor often discusses information security issues with the employees	0.774
Tra 4: Information security is generally well taken care of at my workplace	0.745
Man 7: My immediate supervisor considers information security to be very important in all tasks and activities	0.743
Man 6: My immediate supervisor postpones tasks/activities if information security is not sufficiently ensured	0.735
Man 8: My immediate supervisor does everything he/she can do to avoid breaches of information security	0.733
Rep 3: After a breach of information security, measures are implemented to prevent this from happening again	0.729
Man 3: My immediate supervisor is aware of the most important information security issues in the company	0.691
Man 5: My immediate supervisor is personally involved in activities to improve information security	0.678
Rep 4: All irregularities and information security issues that are reported are remedied in a short time	0.654
Emp 2: Employees encourage one another to safeguard information security	0.644
Man 2: My immediate supervisor often praises employees for maintaining information security	0.639
Emp 3: Employees usually report all breaches and irregularities related to information security that they experience at work	0.634
Gen 3: Regular security audits are carried out	0.615
Rep1: Those who pursue breaches of information security in the business attempt to find the real causes rather than just blaming the employees	0.598
Man 1: My immediate supervisor discovers employees who fail to take sufficient considerations to information security in their work	0.593
Emp 1: My colleagues do everything they can to avoid breaches of information security	0.591
Rep 5: Everyone has plenty of opportunities to forward suggestions related to information security	0.572
Gen 2: Information security in my business is better that in other businesses	0.548
Rep 2: There are routines and procedures at my workplace so that I may report information security-related breaches or irregularities	0.534
Tra 1: There are procedures that must be followed in the event of an emergency situation in my workplace	0.508

Factor 2: Information security training	**Loadings**
Tra 2: All new employees are provided with adequate training for tasks and the secure use of ICT systems (e.g. e-mail, storage, encryption)	0.691
Tra 1: Employees in my company are provided with adequate training in the secure use of ICT systems (e.g. e-mail, storage, encryption)	0.629
Tra 3: Everyone is provided with sufficient feedback on how the enterprise is performing with regard to information security	0.562
Tra 4: Everyone is informed of any changes that may impact information security	0.550

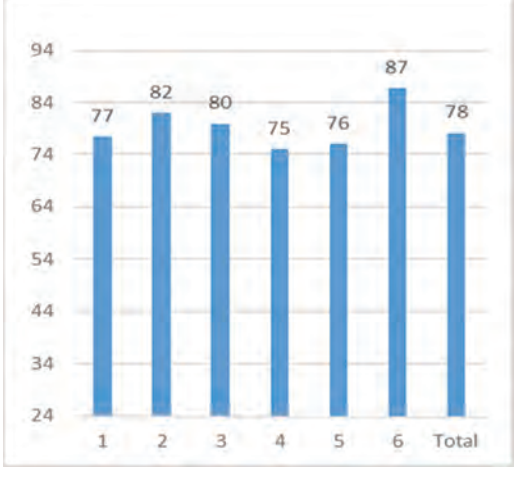


Figure 1. Mean scores on the GAIN index applied to organsiational information security culture. Departments in the study organisation. Minimum 24, maximum 120. (N = 249).

The figure shows that DEP 6 has the highest score and that DEP 4 has the lowest score. The difference between the highest and lowest score is more than 11 points. The differences between the departments are significant at the 1% level. We are careful of comparing the departments and in relation to the information security culture questions. The number of "neither/nor" responses indicate that the respondents are unwilling or incapable of considering the statements on the area, which may make it difficult to interpret the numbers. It also entails that respondents who have actually had an opinion are a minority for some of the questions. All in all, we therefore consider this index to be less robust than the others. This applies to all departments, except DEP 6.

The first topic in the index was "Immediate supervisor's attitude to and focus on information security." Here DEP 6 had the highest score, DEP 4 the lowest.

The second topic in the index is "Employees' attitude to and focus on information security." Once again DEP 6 had the highest score while DEP 1 had the lowest. The differences are significant at the 5% level. This should be interpreted in the light of DEP 1 having responsibility for following up such work, and is probably more critical in its assessment.

The third topic in the index is "Reporting culture and reactions to incident reporting." DEP 6 and DEP 2 had the highest scores, while DEP 4 and DEP 5 had the lowest. Differences were significant, of more than two points.

The fourth topic in the index is "Training in information security thinking." DEP 2 and DEP 6 had the highest scores, while DEP 1 had the lowest. The fifth topic in the index is "General information security issues." DEP 6 had the highest score, while DEP 1 had the lowest. The differences are significant at the 1% level.

4.2.1 *Exploratory factor analysis*

An Exploratory Factor Analysis (EFA) was conducted to examine the underlying factor structure of the 24 items in the sample. Although the original GAIN-scale for safety culture is comprised of five themes, we chose EFA as we apply it to a new topic; information security culture. Tests indicated that the items and the data were suitable for factor analysis. Bartlett's test of sphericity (approx. Chi-square) was 3380,834 ($p < 0.001$). The Kaiser–Meyer–Olkin's measure of sampling adequacy showed a value of 0.909. An unrotated principal component analysis (PCA) was used. We sat the cut off value of factor loadings equal to or above 0.40, as Matsunaga (2010) suggests that this perhaps is the lowest acceptable threshold on a conventional liberal-to-conservative continuum. Results showed five components with initial Eigenvalues higher than 1, which explained a total of 64.9% of the variance. The choice of the number of factors to retain was based on a combination of a) Eigenvalues, b) inspecting the scree plot for a bending point, c) inspecting the factor loadings in the component matrix, and d) conceptual and theoretical consideration. By inspecting the scree plot, a bend was relatively clearly identified between factor 5 and 6, indicating a five-factor solution. This is in line with the Eigenvalues. However, when looking at the factor loadings, we saw that all items loaded on the first component, while there were seven cross-loading. Four of these had lower factor loadings on the other factors than the first factor, and they were distributed on different factors. They were therefore kept in the first factor, with one exception. Three of the cross-loading items were all in the second factor, they had higher factor loadings in the second factor and they all concern security training. They were therefore attributed to a second factor. Additionally, one of the first mentioned cross-loading items had quite similar loadings in both factors (0.567 vs. 0.562), but as it matched the second conceptually, we attributed it to this factor.

Thus, based on our analysis of the factor loadings and a conceptual and theoretical consideration (the four latter items all concern information security training), we chose a two-factor solution, which explained a total of 50% of the variance, i.e. about 15% less than the three-factor solution.

4.2.2 *Regression analysis: What influences organisational information security scores?*

The information security culture scores vary according to conditions such as age, education and seniority, but we do not know which conditions that are most important to explain the variation in information security culture, or whether the effect we see from education is actually due to age or vice versa. We have conducted regression analyses to assess which conditions explain variation in the information security culture index.

We have used linear regression as the dependent variable is continuous. We add various independent variables in steps, so that we can assess their isolated effect on the dependent variables, i.e. when the values of the other variables remain unchanged. In this manner we can examine the effect of education controlled for age, for example.

We add the gender, age, education and seniority variables in the study organisation and department. We have converted the department variable to a dichotomous variable, i.e. with two values. The reason is that in regression analyses one cannot have independent variables that are at the nominal level, i.e. with values that are mutually exclusive, but which can't be ranked. The two values of the department

variable are 1) all other departments, 2) DEP 6. We have done this because DEP 6 had the highest score on the information security culture index.

We see that age contributes significantly and positively in 2; the older the respondents are, the better their information security culture score. In model 3 however, the age variable stops being significant, and that indicates that the age effect is actually due to it correlating with education. This means that younger respondents have higher education and a lower information security culture score and vice versa. Seniority does not make a significant contribution in any of the models.

Finally, we see that the department variable makes the strongest contribution in the regression analysis in Table 1. Belonging to DEP 6 predicts a positive score on the information security culture index. We already knew this, but in the regression analyses in Table 1 also show that this also applies when controlling for gender, age, education and seniority. We may therefore conclude that DEP 6's high score on the information security culture index is not due to underlying variables such as gender, age, education and seniority.

We see that the adjusted R^2 value, which indicates which proportion of the variation in the dependent variable that is explained by the independent variables significantly increasing in model 3 when education is included in the analyses, and that it increases by more than twice as much when department is included in model 5. The independent variables education and department explain 9.7% of the variation in the information security culture index.

4.3 *Regression analysis: What influences information security behaviour?*

We have conducted regression analyses to assess which conditions explain variation in the variable "I have never caused a breach of information security." This is a variable with five options from 1 (strongly disagree) to 5 (strongly agree). The overall "neither/nor" share is 26.3% for this question. What one answers here is probably to a certain extent dependent on whether one has a clear understanding of what information security is, or what it means for day to day work. The answer will also depend on how many opportunities one has to breach information security in one's work. We have used linear regression as the dependent variable is continuous. We add four independent variables in steps, so that we can assess their isolated effect on the dependent variables, i.e. when the values of the other variables remain unchanged. We add the gender, age and seniority variables in the study organisation and information security culture.

The table shows that seniority and information security culture contribute significantly to explain the variation in the variable "I have never caused a breach of information security." Both effects are positive. The positive effect of seniority means that

Table 1. Linear regression. Dependent variable: Organisational information security culture standardised beta coefficients.

Variable	1	2	3	4	5
Gender	–0,003	0,028	0,021	0,020	0,003
Age		**0,158****	0,105	0,076	0,034
Edu (Uni = 2)			**–0,185*****	**–0,157****	**–0,138****
Seniority				0,078	0,059
Department (DEP 6 = 2)					**0,244*****
Adj. R^2	–0.004	0.016	0.044	0.045	0,097

*p < 0,1; **p < 0,05; ***p < 0,01.

Table 2. Linear regression. Dependent variable: "I have never caused a breach of information security." Standardised beta coefficients.

Variable	Mod. 1	Mod. 2	Mod. 3	Mod. 4
Gender	–0,043	–0,042	–0,046	–0,051
Age		0,005	–0,070	–0,087
Seniority			0,159**	0,132*
Information security culture				0,192***
Adjusted R^2	–0.002	–0.006	0.010	0.041

*p < 0,1; **p < 0,05; ***p < 0,01.

the longer one has been employed by the study organisation, the higher the likelihood that one has not caused a breach of information security. This is perhaps somewhat unexpected, as one would assume that the longer one has been employed, the more opportunities (in terms of time) there have been to breach information security. It further means that there is reason to assume that the study organisation had scored somewhat higher on security culture if the distribution of respondents had corresponded to that in the organization: In the survey, 51 per cent of respondents had 0–5 years seniority, while in reality there are 38 per cent who have 0–5 years seniority in the study organisation. There is nevertheless no reason to believe that this possible skewness alters fundamental conclusions, as the difference is too small.

The effect of information security culture is however strongest, and this is the most important variable to explain variations in breaches of information security in the analyses. The higher the information security culture score is, the higher the likelihood that one has not caused a breach of information security.

We see that the adjusted R2 value, which indicates which proportion of the variation in the dependent variable that is explained by the independent variables, is negative in the two first models, but that it is at 1 and 4.1% in the last two. This happened when we included seniority and information security culture. These variables explain 4.1% of the variation in the variable "I have never caused a breach of information security".

5 CONCLUDING DISCUSSION

Learning from research on organizational culture and safety culture, we have adapted an organizational safety culture scale to measure organizational information security culture. Our examination of the factor structure of the scale indicated two factors. Regression analyses indicate that our adapted GAIN-scale, measuring organizational information security culture is the most important variable influencing information security behavior in the model.

REFERENCES

Antonsen, S. 2009. "The relationship between culture and safety on offshore supply vessels", *Safety Science*, Vol. 47. Issue 8, pp. 1118–1128.

Bjørnskau, Torkel og Frode Longva 2009. Sikkerhetskultur i transport. TØI rapport 1012/2009 Oslo: Transportøkonomisk institutt.

Chia P, Maynard S, Ruighaver AB. Understanding organizational security culture. In: *Sixth pacific Asia conference on information systems,* Tokyo, Japan; 2–3 September 2002.

Cox, S.J. & R. Flin (1998): "Safety Culture: Philopher's Stone or a Man of Straw?", *Work & Stress,* Vol 12, No 3 189.

Flin, R., K. Mearns, P. O'Connor & R. Bryden (2000): "Measuring safety climate: identifying the common features", *Safety Science,* Vol.34, 177–192.

GAIN (Global Aviation Network) 2001. Operator's Flight Safety Handbook, http://flightsafety.org/files/OFSH_english.pdf.

Guldenmund, F.W. (2000): "The Nature of Safety Culture: a Review of Theory and Research", *Safety Science*, vol. 34, 1–14.

Hale, A.(2000): "Editorial: Culture's Confusions", *Safety Science,* vol. 34, 1–14.

Haukelid, K. (2008): "Theories of (safety) culture revisited—An anthropological approach", *Safety Science*, Vol. 46/3, 413–426.

Kines, P.J. Lappalainen, K. Lyngby Mikkelsen, E. Olsen, A. Pousette, J. Tharaldsen, K. Tómasson & M. Törner (2011): Nordic safety climate questionnaire (NOSACQ-50): A new tool for diagnosing occupational safety climate, *International Journal of Industrial Ergonomics,* Vol. 41, pp. 634–646.

Knapp KJ, Marshall TE, Rainer RK, Ford FN. (2006) Information security: management's effect on culture and policy. *Information Management & Computer Security* 2006;14(1):24–36.

Nosworthy J. (2000) Implementing information security in the 21st Century—do you have the balancing factors? *Computers and Security;*19(4):337–47.

NOU (2006). Når sikkerhet er viktigst, Beskyttelse av landets kritiske infrastrukturer og kritiske samfunnsfunksjoner.

Nævestad, T.-O. (2010a): *"Cultures, crises and campaigns: examining the role of safety culture in the management of hazards in a high-risk industry"*, Ph.D. dissertation, Centre for Technology, Innovation and Culture, Faculty of Social Sciences, University of Oslo.

Reason, J. (1997): *Managing the Risk of Organisational accidents,* Aldershot: Ashgate.

Ruighaver, A.B.S.B. Maynard, S. Chang (2007) Organisational security culture: Extending the end-user perspective, computers & security 26 (2007) 56–62.

Schein, E.H. (2004): *Organizational Culture and Leadership,* Third Edition, San Francisco: Jossey-Bass.

SjekkIT. Verktøy for å måle informasjonssikkerhetskultur utviklet av NTNU og Sintef for NSM.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

How can we explain improvements in organizational information security culture in an organization providing critical infrastructure?

T.O. Nævestad
Institute of Transport Economics, Oslo, Norway

J. Hovland Honerud
University College of Southeast Norway, Norway

S. Frislid Meyer
Institute of Transport Economics, Oslo, Norway

ABSTRACT: The aims of the present study are to 1) Compare results from a study conducted before and a study conducted after efforts to improve organizational information security culture in an organization providing critical infrastructure, and 2) discuss the results of the comparison. In this study, we compare the results of two surveys done over a period of just over two years; the first early in 2014 (N = 323) and the second late in 2016 (N = 446). Organizational information security culture was measured with an index which was made by adding the scores of respondents' answers to 24 items with answer alternatives ranging from 1 to 5. Thus, the minimum score of the information security index was 24 and the maximum score was 120. We found a statistically significant improvement in the index, comparing 2014 to 2016. Changes are discussed considering respondents' experiences with the implemented measures, sample characteristics and other methodological factors. We conclude that it seems that management implementation of measures aimed at improving organizational security culture has led to improvements.

1 INTRODUCTION

1.1 *Background*

Ruighaver et al (2007) assert that it was not until the start of the century that scholars began to recognise the importance of organisational security culture for information systems security in organisations. Information security is often defined as protection against breaches of confidentiality, integrity and accessibility. This applies to information that is oral, written or electronic. Confidentiality refers to ensuring that only those who are authorised to access information, accesses it. Integrity refers to protecting the accuracy and entirety of information and processing methods. Accessibility refers to ensuring that authorised users have access to the information and associated equipments when necessary (Report to the Storting 29. 2011–2012).

Throughout the report of the 22 July Commission, following the Oslo and Utøya terror attacks in 2011, leadership, culture and attitudes are emphasized as key challenges, both back in time and for the future. In the 22 July Commission's report, part VI "Learning, 19.9" (page 458) the Commission's main conclusion and recommendations are presented: The Commission's most important recommendation is that leaders at all levels of the administration work systematically to strengthen their own and their organisations' fundamental attitudes and culture in respect of: e.g. the acknowledgement of risk and leadership. Critical infrastructure means the facilities and systems that are completely necessary to maintain society's critical functions, which in turn meet society's basic needs and respond to the population's need for a perception of safety (NOU 2006).

1.2 *Aim*

The aims of the present study are to 1) Compare results from a study conducted before and a study conducted after efforts to improve organizational information security culture in an organization providing critical infra-structure, and 2) discuss the results of the comparison.

The study organization is a provider of critical infrastructure in Norway. As a provider of critical infrastructure, the study organization is obliged to follow the requirements of the Safety Act ("Sikkerhetsloven") when it comes to preventive safety work, which includes safety analyses, securing objects, information security and safety drill. Based on these requirements, the study organiza-

tion decided to map and analyse their own organizational security culture.

1.3 *Previous research on culture in organisations*

Although Ruighaver et al (2007) note that the organisational security culture concept has gained recognition, they also underline that there is lacking consensus when it comes to how the concept should be defined and conceptualized (cf. Chia et al 2002). Additionally, they also assert that in spite a large amount of research on organisational security and how it should be improved, this research only focuses on certain aspects of security, and not how these aspects can be analysed as part of a larger organisational culture (Ruighaver et al 2007). Based on this understanding, they choose to draw on organisational culture research in their analysis of security culture. This approach is similar to that applied by scholars studying organisational safety culture, who analyse safety culture as a focused and safety relevant aspect of the larger organisational culture (e.g. Hale 2000; Haukelid 2008 Antonsen 2009). Based on this, we may also analyse security culture as "security relevant" aspects of the larger organisational culture, define and conceptualising using models of organisational culture (e.g. Schein 2004).

The influential scholar Schein defines organizational culture as: "(...) a pattern of shared basic assumptions that the group learned as it solved its problems of external adaptation and internal integration that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think and feel in relation to those problems." (Schein 1992: 12). According to Guldenmund (2000: 222–225), organizational culture has the following characteristics: 1. It is a construct in the sense that it is an abstract, not a concrete concept, 2. It is relatively stable, 3. It has multiple dimensionality, in the sense that it can be described in many different ways, 4. It is shared by groups of people, 5. It consists of various aspects, which means that several different cultures can be identified within organizations, depending on the issue at hand, 6. It constitutes practices, 7. It is functional. Guldenmund describes organizational culture in the following manner: "Overall, organisational culture is a relatively stable, multidimensional, holistic construct shared by (groups of) organisational members that supplies a frame of reference and which gives meaning to and/or is typically revealed in certain practices." (Guldenmund 2000: 225).

As the research on organizational safety culture seems to have been through many of the challenges that the organizational security culture research now is facing, we draw on the experiences of the former, e.g. when it comes to analyzing security culture as a focused aspect of organizational culture. This also applies to management of culture. Knapp et al (2006) asserts for instance that top management support is a significant predictor of an organization's security culture and level of policy enforcement. Learning from the organizational culture literature, we may conceptualise this in terms of Schein's (2004) "six primary embedding mechanisms" that managers can use to shape culture is typical of the integration view:

1) What managers pay attention to, measure and control on a regular basis, 2) How managers react to critical incidents and organizational crises, 3) How managers allocate resources, 4) Deliberate role modelling, teaching and coaching, 5) How managers allocate rewards and status, and 6) How managers recruit, select, promote and excommunicate. It should however be noted that although organisational culture is a concept that often is examined with the intention to influence, organizational culture scholars give different answers to whether, to what extent and how safety culture can be managed, depending on their conceptualization of culture (cf. Nævestad 2010a). To conclude, the research on culture interventions in organisations generally indicate that safety culture change come about in the dynamic between "top-down" processes inititated from the management and "bottom-up" processes based in sub-groups (Olsen et al 2009).

2 METHOD

2.1 *Interviews*

We used a semi-structured and relatively open interview guide. Our starting point was topics related to information security, follow-up of safety and security work since 2014. The interview guide had to be open so that we could depend on the interviewees' understanding of how different organisational and cultural features of the study organisation have had and can have consequences for security. The interviews were built up around the following main topics: a) Reviews and measures related to security in the study organisation since 2014, b) The organisation's and employees' relationship to security. Security in everyday life, and as a topic of conversation, c) Clarity, management and rules related to security, d) Training, reporting/notification and risk assessments.

2.2 *Survey items*

The survey mainly contains questions about ten topics. It first contains a set of background questions, (e.g. gender, age, experience, education). that were sent to all respondents. In addition, the questionnaire contains 13 questions about attitudes and behaviour concerning information security culture.

We have chosen to not report results for these items here, as we wanted to focus on the development in the organizational security culture scale in 2014 and 2016, also comparing departments and measures implemented after the first measurement.

2.2.1 *The GAIN scale*

In this study, we choose to reformulate one of the few existing universal organizational safety culture scales, the GAIN-scale for safety culture, into an organizational security culture scale. Global Aviation Information Network (GAIN) is a voluntary association of airlines, manufacturers, trade unions, governments and other organisations in aviation. The GAIN questionnaire contains 25 questions concerning: 1) management's attitude to and focus on safety, 2) employees' attitudes to and focus on safety, 3) reporting culture and reactions to incident reporting, 4) safety training and education, and 5) general questions concerning safety in the organization in question. Respondents can rate the questions from 1 (totally disagree) to 5 (totally agree). Thus, a safety culture index with a minimum value of 25 ($1 \cdot 25$) and a maximum value of 125 ($5 \cdot 25$) can be compared across companies and sectors. According to GAIN (2001), organizations with a score of 93–125 points on the safety culture index have a positive safety culture, 59–92 indicates a bureaucratic safety culture and 25–58 indicates a poor safety culture.

The scale was chosen first, as the research on organizational safety culture seems to have been through many of the challenges that the organizational security culture research now is facing. Although Ruighaver et al (2007) note that the organisational security culture concept has gained recognition, they also underline that there is lacking consensus when it comes to how the concept should be defined and conceptualized (cf. Chia et al 2002). Additionally, they also assert that in spite a large amount of research on organisational security and how it should be improved, this research only focus on certain aspects of security and not how these aspects can be analysed as part of a larger organisational culture. Based on this understanding, Ruighaver et al (2007) choose to draw on organisational culture research in their analysis of security culture. We choose the same strategy in the present paper, drawing on the experiences of the safety culture literature, and choosing the GAIN-scale. The safety culture literature seems to have matured a bit more conceptually and methodologically, as it has employed the culture perspective in a few more years than the field of security research.

2.2.2 *Three types of safety/security/reliability*

In the 2014 interviews related to the first survey, we were given indications that many different types or nuances of safety and security are key in the study organisation. To assess the importance of, or the focus on these types of safety and security, we created four indexes with similar formulations of statements: one about information security, one about HSE, one about deliverance reliability (and one about security in connection with sensitive objects which is not reported here). The indexes consist of three statements:

I. My head of department considers ... to be very important
II. The study organisation's senior management considers ... to be very important
III. My colleagues consider ... to be very important

In the four indexes we replace ... with 1) information security, 2) protection of sensitive objects, 3) HSE and 4) deliverance reliability. All of the questions have five values, so the minimum value is 1 (totally disagree) and the maximum value is 5 (totally agree). The four indexes therefore have the same minimum and maximum values so that they can be compared to each other. In this way we can see what types(s) of safety and security managers and employees focus most on in the study organisation as a whole; in the various departments.

2.3 *Samples*

In this study, we compare the results of two surveys done over a period of just over two years; the first in the spring of 2014 and the second in the autumn of 2016. While we have used the same questionnaire, we added some questions to the last form. A total of 323 respondents responded to the survey in 2014, while 446 responded in 2016. The survey respondents are shown in Table 1.

2.4 *Analysis*

We calculate the significance of the differences in scores for all surveys. In this way, we examine the probabilities that the differences are due to statistical coincidences. This is done by calculating the

Table 1. Responses, number of employed and response rate per department in 2016, compared with response rates in 2014.

Dep.	Res.	Empl	2016	2014
1	84	112	75%	79%
2	26	31	84%	–
2	62	85	73%	69%
4	38	54	70%	70%
4	115	162	71%	46%
6	84	109	77%	24%
7	13	17	76%	92%
8	24	28	86%	113%
Total	446	600	74%	56%

average scores' confidence intervals. These indicate the error margins of the average scores, i.e. the range which, with a given probability, contains the true number measured. When comparing average scores, we can generally say that the differences between them are statistically significant if they do not lie within each other's confidence intervals.

Probability is given as a percentage. This is also often given as a so-called P value. In choosing the confidence interval, you choose how much uncertainty you will accept. A 90% confidence interval means that you have decided on a 90% probability level, indicating that, on average, an error will be concluded in one of ten cases. A 95% confidence interval means that there is a 95% chance that the "true" risk number is within this range. We use confidence intervals of 90%, 95% and 99%, and we say that the differences are statistically significant at the 10%, 5% and 1% levels, respectively.

3 RESULTS

The questionnaire that measures information security culture is based on the GAIN organisational safety culture index, where the word "safety" is replaced by "information security." The questionnaire opened with definitions of information security and related sub-concepts. In order to map the respondents' relationship with information security in their work, we asked them to take a position on the statement: "In my work, I have a clear understanding of what the term information security means." The results show that all of the departments have an average score above 4 (= agree somewhat).

3.1 *Management commitment*

The immediate supervisor's attitude to and focus on information security is an index with eight questions, with a minimum value of 8 and a maximum value of 40. Cronbach's Alpha for this index in the data from 2014 was 0.911, which means very good agreement between the questions and that the index is good. The index is based on the following questions:

- My immediate supervisor discovers employees who fail to take sufficient considerations to information security in their work
- My immediate supervisor often praises employees for maintaining information security
- My immediate supervisor is aware of the most important information security issues in the company
- My immediate supervisor often discusses information security issues with the employees
- My immediate supervisor is personally involved in activities to improve information security
- My immediate supervisor postpones tasks/activities if information security is not sufficiently ensured
- My immediate supervisor considers information security to be very important in all tasks and activities
- My immediate supervisor does everything he/she can do to avoid breaches of information security

The score in 2014 was 26 points, while it was 29 in 2016. When we test the significance of the difference between the total scores for 2014 and 2016, we see that the differences are significant at the 1% level. The differences between the departments are significant at the 1% level in 2014 and 2016. This indicates that there are significant differences between the departments as to whether employees feel that their immediate supervisors focus on information security. DEP 6, DEP 2, DEP 1 and DEP 5 had the highest scores in 2016. In addition, we also see that all of the departments saw improvement in this index in 2016, especially DEP 1, DEP 2, DEP 6 and DEP 4.

3.2 *Employees' attitude to and focus on information security*

Employees' attitude to, and focus on information security is an index with three questions, with a minimum value of 3 and a maximum value of 15. In 2014, we measured a Cronbach's Alpha at 0.754, a high value considering that the index consists of three questions.

- My colleagues do everything they can to avoid breaches of information security
- Employees encourage one another to safeguard information security
- Employees usually report all breaches and irregularities related to information security that they experience at work

All of the departments saw an improvement on this index in 2016, especially DEP 1 and DEP 5. DEP 6 had the highest score again in 2016. The general improvement went from 10 to 11 points on the index. When we test the significance of the difference between the total scores for 2014 and 2016, we see that the differences are significant at the 1% level.

3.3 *Reporting culture and reactions to incident reporting*

Reporting culture and reactions to incident reporting is an index with five questions, with a minimum value of 5 and a maximum value of 25 (Cronbach's Alpha = 0.785 in 2014).

Those who pursue breaches of information security in the business attempt to find the real causes rather than just blaming the employees
There are routines and procedures at my workplace so that I may report information security-related breaches or irregularities
After a breach of information security, measures are implemented to prevent this from happening again
All irregularities and information security issues that are reported are remedied in a short time
Everyone has plenty of opportunities to forward suggestions related to information security

Again DEP 6 had the highest score in 2014 and 2016, followed by DEP 2. All of the departments regressed on this index in 2016, especially DEP 1, which noted a decline of more than 3 points on the index. The average decline for all of the departments from 2014 to 2016 was 2.1 points. This could be due to a learning effect and/or increased focus on what reporting means in the organisation (i.e. increased expectations), and/or an actual decrease. This change is statistically significant at the 1% level.

3.4 *Training/instruction in information security thinking*

Training/instruction in information security thinking is an index with four questions, with a minimum value of 4 and a maximum value of 20 (Cronbach's Alpha = 0.866 in 2014).

Employees in my company are provided with adequate training in the secure use of ICT systems (e.g. e-mail, storage, encryption)
All new employees are provided with adequate training for tasks and the secure use of ICT systems (e.g. e-mail, storage, encryption)
Everyone is provided with sufficient feedback on how the enterprise is performing with regard to information security
Everyone is informed of any changes that may impact information security

Again DEP 6 department had the highest score in 2016 (and 2014), followed by DEP 4 and DEP 1. All of the departments saw an improvement on this index in 2016, especially DEP 1, which increased by more than 4 points on the index. The average improvement for all of the departments from 2014 to 2016 was 2.4 points. This change is statistically significant at the 1% level.

3.5 *General information security issues*

General information security issues is an index with four questions, with a minimum value of 4 and a maximum value of 20 (Cronbach's Alpha = 0.720 in 2014).

There are procedures that must be followed in the event of an emergency situation in my workplace
Information security in my business is better that in other businesses
Regular security audits are carried out
Information security is generally well taken care of at my workplace

DEP 6 department again had the highest score in 2016 (and 2014), followed by DEP 1. All of the departments saw an improvement on this index in 2016, especially DEP 1, which increased by 3 points on the index. The average improvement for all of the departments from 2014 to 2016 was 1.4 points. This change is statistically significant at the 1% level.

3.6 *Index for information security culture*

We have combined the 24 statements with five response options on the five different aspects of information security in an information security culture index. The indexes for the departments correspond to the average scores for the respondents. Since we have removed a statement from the GAIN index (as the wording was unsuitable for the context), the minimum score will be 24 (24 · 1) and the maximum score will be 120 (24 · 5). Cronbach's Alpha for the 24 questions in the index was 0.913 in 2014, which means very good agreement between the questions and that the index is very good. Table 2.

We see that the DEP 6 department again had the highest score in 2016 and 2014, followed by DEP 2 and DEP 1. All of the departments saw an improvement on this index in 2016, especially DEP 1, which increased by 12 points on the index.

The average improvement for all of the departments from 2014 to 2016 was 9 points. This change is statistically significant at the 1% level. The differences between the departments are significant at the 1% level, both in 2014 and 2016.

If we are to transfer the GAIN scale values from organisational safety culture to organisational

Table 2. Scores on the index for information security culture per department in 2014 and 2016 (min = 24 points, max = 120 points).

Department	2014	2016
1	77	89
2	82	89
3	80	82
4	75	84
5	76	87
6	87	95
Total	78	87

information security culture, we see that none of the departments had a positive information security culture in 2014. However, in 2016 we find that DEP 6, DEP 2 and DEP 1 were within the part of the scale that we refer to as a positive culture. The limits for "positive organisational culture" range from 88 points to 120 points on the GAIN index. The moderate culture scale goes from 47 points to a maximum of 87 points, and scores below 46 points correspond to a poor culture.

3.7 *Questions about the department's follow-up of the survey in 2014*

We asked the respondents three questions about what actions their immediate supervisor has taken in his/her own department/section after the evaluation of the information security culture in 2014. We opened the questions with the text: "We want to know a little about what actions your immediate supervisor has taken in your department/section after the evaluation of the information security culture in 2014".

We made an index based on the three questions we have reviewed. All of the questions include six response options: 1 = totally disagree and 5 = totally agree and 6 = Have no knowledge about this. When we made the index, we removed the sixth response option. The index is based on the following questions:

- The head of my department/section has gone through the results of the evaluation with us
- My department/section has taken steps to improve the information security culture (e.g. focus on passwords and security-critical information) based on the results of the evaluation
- I am satisfied with how my department/section has worked on information security over the past two years

Table 3 shows the average index score for actions the immediate supervisor has initiated in his/her own department/section after the evaluation of the information security culture in 2014.

3.8 *Questions about training*

We asked the respondents three questions related to whether they have received useful information over the past two years that has increased their knowledge and awareness of information security:

- During the past two years I have received useful information (e.g. from security coordinator, manager, intranet) about what a secure password is
- During the past two years, I have received information (e.g. from security coordinator, manager, intranet) that made me more aware of strangers in our premises
- During the past two years, I have received information (e.g. from the security coordinator, manager, intranet) that has given me more insight into what security-critical information is

Since we, to some extent, assume that this information comes from the security coordinator, manager and intranet, we also refer to these questions as training in information security 2014–2016.

We made an index based on the three questions. All of the questions include six response options: 1 = totally disagree and 5 = totally agree and 6 = Have no knowledge about this. When we created the index, we removed the sixth response option.

Table shows the average index score for information security training in 2014–2016 in combination with scores for the action index in each department.

3.9 *Three types of safety/security/reliability*

Through initial interviews in the first survey we found that it was appropriate to distinguish between different types or nuances of security, safety and reliability in the study organisation. To assess the importance of, or the focus on these

Table 3. Index for actions the immediate supervisor has initiated in his/her own department/section after the evaluation of the information security culture in 2014. Minimum value: 3, maximum value: 15. (N=313).

Department	Measure index
1	12
2	11
3	11
4	11
5	12
6	14
Total	12

Table 4. The average index score for information security training in 2014-2016 in combination with scores for the action index in each department.

Department	Training index
1	13
2	12
3	12
4	13
5	13
6	14
Total	13

types of safety and security in the study organisation as a whole, and in different departments, we created indexes with identical questions: one about information security, one about HSE, one about deliverance reliability and one about security in connection with sensitive objects. The index about sensitive objects was not included in 2016.

In line with what we have commented on, we see in Figure 1 that all of the departments did better on the indexes in 2016 compared with 2014. However, the improvement is greatest on the information security index.

This is interesting because in 2014 we concluded that the results indicated that little attention was paid to information security compared with the "primary task" of deliverance reliability and the more traditional focus on HSE. The fact that the respondents in 2016 deemed that the focus on information security is equally strong as for the other types of security can indicate that the study organisation's efforts to put this topic on the agenda has worked as intended.

3.10 *Interview results*

The organisation has carried out extensive changes in its organisation and practices during the period. The scope and significance of the changes are emphasised through interviews. Information security and HSE-related support and correction from colleagues seems to be a fairly consistent practice. They also seem to be topics that are the subject of conversations. This is well in line with the general increase in all types of security culture in the agency. There has been a strong focus on information security both internally in the organisation and through auditing, supervision and ownership requirements. Since 2014, a more clear "security organisation" has been established, to use the organisation's own term. Security organisation means roles, ownership, and processes related to security.

The security culture now seems more strongly rooted the higher one gets in the organisation. This is partly based on interviews, which quite uniformly point to a clear focus on security in the agency's top management, and more varied further follow-up. This also correlates with the fact that managers state to a greater degree than staff that security is being followed up.

4 CONCLUDING DISCUSSION

4.1 *Improvements and measures*

We saw above that DEP 6 scored the highest on the information security culture index, and in chapter we also saw that DEP 6 scored the highest on the index for managers' measures aimed at information security culture. This indicates a correlation between immediate supervisors' measures since 2014, and the information security culture score.

Figure 2 confirms the impression that good work on measures produces results in the form of high scores on the information security culture

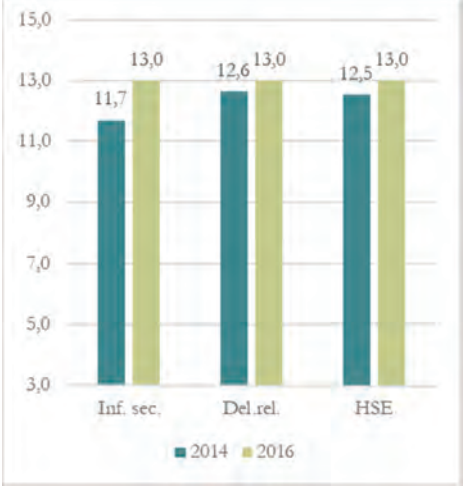


Figure 1. Total scores for the indexes for indexes on: Information security (N=323 and 446), Deliverance reliability (N=323 and 446), HSE (N=85 and 169).

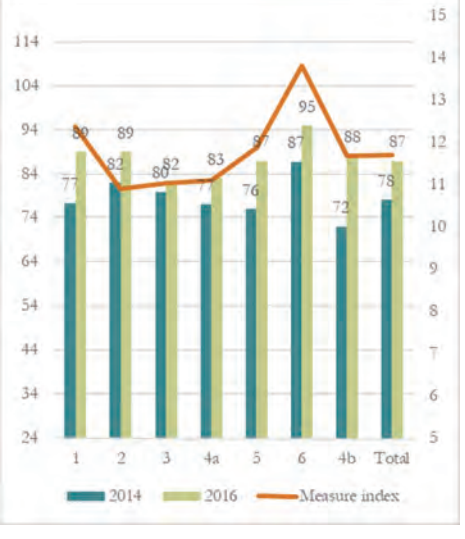


Figure 2. Information security culture scores in 2014 and 2016, for each department, including a measure index.

index. On the other hand, the impact of measures can also be assessed on the basis of which department has made the greatest improvement on the index from 2014 to 2016.

This result is in line with both the assertion of Knapp et al (2006), depicting the top management support as a significant predictor of an organization's security culture and level of policy enforcement. It is also in line with Schein's (2004) view on the possibilities to manage culture in organisations. As noted, he has coined "six primary embedding mechanisms" that managers can use to shape culture: e.g. what managers pay attention to, measure and control on a regular basis, how managers allocate resources. Thus, our study may indicate how organisations may improve their organizational security culture through cultural management. Moreover, the training measures also seem to be related to high security culture scores. Thus, this organization may provide an example of a possible solution to the key challenge addressed by Nosworthy (2000): how to we educate the people of the organization to successfully implement the requirements of the information security policy?

4.2 *Is the improvement of information security culture real?*

Hawthorne effect? We have seen an average increase of 9 points from 2014 to 2016 in the study organisation's information security culture index. This is a significant improvement that can be said to indicate that the study organisation's information security measures from 2014 to 2016 have produced good results. On the other hand, the improvement may be interpreted as a "measurement effect". This means that the improvement has more to do with the repeated survey than a real effect.

Such measurement effects can be due to numerous mechanisms. On a general basis, one can point to the so-called "Hawthorne effect," which indicates that being measured repeatedly produces improvement in the surveys because the researchers' focus per se causes the studied elements to do better in measurements. This is the "placebo effect" of intervention research, i.e., one can expect a certain amount of improvement in a repeated survey, only because the survey is conducted.

Learning effect? Another measurement effect that occurs in some cases is that the respondents learn from the first questionnaire and respond differently to the second questionnaire for that reason. However, this can yield both positive and negative outcomes; the respondents may think that they assessed their workplace too "harshly" when they think more about the topics from the survey and respond more positively the second time. However, they may also think that they assessed their workplace too "leniently" when they think more about the topics from the survey and respond more negatively the second time. This can result in lower scores in a repeated survey.

Noise. A third effect that can impact the second survey is that reorganisation, noise and the like make the respondents more negatively disposed to their workplace and the topics that are being measured and therefore show their dissatisfaction by giving negative answers. This would involve lower scores, which only applies to reporting culture.

We can perhaps conclude to some extent that the comparisons seem realistic, first since the results for the short indexes for safety, security and reliability also indicate a clear, increased focus on information security in the study organisation. In 2014, we saw a greater focus among managers and staff in the study organisation on deliverance reliability and HSE than on information security. However, in 2016 we see that the score for information security is at the same high level as deliverance reliability and HSE. Second, we saw a relationship between measures, training and security culture scores. Third, interviews indicate increased focus on information security in 2016 compared with 2014.

4.2.1 *Can the improvement be explained by sample differences?*

Changed scores should also be discussed in the light of possible sample differences in the two surveys. The respondent rate in 2016 was higher than it was in 2014. The fact that there were more respondents in 2016 may have influenced the answers given and the scores on the indexes we looked at. In 2016, the percentage of respondents who had been employed for 16 or more years was down five points compared with 2014. This should have indicated lower scores on the information security culture index, since we know that higher age and seniority yield higher scores. However, it should be pointed out that the increased shares without higher education in the 2016sample probably did not respond to the questions about information security culture, since this higher share was probably due to higher response rates from DEP 3.

It should also be mentioned that over half of those who responded to the survey had been employed by the study organisation for five years or less in both 2014 and 2016. This is a relatively large share. However, we do not know how many of the employees have been employed for less than two years.

Finally, it should be noted that an important difference between the departments' samples in 2014 and 2016 is that all of the departments in 2016 were given a code for "staff" that counts in the department's average. This is a small number of persons in each department, which usually includes

one supervisor and the rest staff. We considered removing staff from the comparisons between the departments in 2014 and 2016, but decided not to after a systematic analysis showed that the staff groups did not have higher scores than the sections in the departments.

Average scores for GAIN information security culture increase with the age of the respondents and decline with education. The difference is the same between the age groups in 2014 and 2016, and about the same for the highest and lowest level of education (14 in 2014 and 13 in 2016). There is no significant difference between the scores for men and women. Finally, we see that those with 11 and more years of seniority have significantly higher scores on the index than those with 10 or fewer years of seniority. The difference is 6 points in 2014 and 5 points in 2016.

4.2.2 *Information security culture scores with and without managers*

Since we also include managers in the survey, we will examine whether managers' responses pull the scores up or down in the various departments. The differences between the information security culture scores of staff, other managers and department and section managers were insignificant in 2014. In 2016, however, the differences were greater. Section or department managers consider the information security culture to be 6 points higher than staff do. This is interesting, but perhaps not unexpected, because section or department managers answer on the basis of their immediate supervisor, who is either the department manager or the study organisation's director. The results can therefore indicate that efforts and measures aimed at information security culture in 2016 are etter rooted at the management level than at the staff level. Similarly, this may indicate that efforts and measures are better rooted at higher management levels in the organisation, which correlates with statements from the interviews. However, the differences between the three groups within each measurement point are not statistically significant, so these results must not be accorded too much weight ($P = 0.14$).

5 CONCLUSION

We conclude that it seems that management implementation of measures aimed at improving organizational information security culture has led to improvements. Thus, our study may indicate how organisations may improve their organizational security culture through cultural management. It is important, however, to note that the study lacks a control group, and that it is impossible to rule out the influence of a possible Hawthorne effect. Additionally, we discuss whether sample characteristics may explain the observed improvements.

REFERENCES

Antonsen, S. 2009. "The relationship between culture and safety on offshore supply vessels", Safety Science, Vol. 47. Issue 8, pp. 1118–1128.

Chia P, Maynard S, Ruighaver AB. Understanding organizational security culture. In: Sixth pacific Asia conference on information systems, Tokyo, Japan; 2–3 September 2002.

Cox, S.J. & R. Flin (1998): "Safety Culture: Philopher's Stone or a Man of Straw?", Work & Stress, Vol 12, No 3 189.

Flin, R., K. Mearns, P. O'Connor & R. Bryden (2000): "Measuring safety climate: identifying the common features", Safety Science, Vol.34, 177–192.

GAIN (Global Aviation Network) 2001. Operator's Flight Safety Handbook, http://flightsafety.org/files/OFSH_english.pdf.

Guldenmund, F.W. (2000): "The Nature of Safety Culture: a Review of Theory and Research", Safety Science, vol. 34, 1–14.

Hale, A.(2000): "Editorial: Culture's Confusions", Safety Science, vol. 34, 1–14.

Haukelid, K. (2008): "Theories of (safety) culture revisited—An anthropological approach", Safety Science, Vol. 46/3, 413–426.

Knapp KJ, Marshall TE, Rainer RK, Ford FN. (2006) Information security: management's effect on culture and policy. Information Management & Computer Security 2006;14(1):24–36.

Nosworthy J. (2000) Implementing information security in the 21st Century—do you have the balancing factors? Computers and Security;19(4):337–47.

NOU (2006). Når sikkerhet er viktigst, Beskyttelse av landets kritiske infrastrukturer og kritiske samfunnsfunksjoner.

Nævestad, T.-O. (2010a): "Cultures, crises and campaigns: examining the role of safety culture in the management of hazards in a high risk industry", Ph.D. dissertation, Centre for Technology, Innovation and Culture, Faculty of Social Sciences, University of Oslo.

Reason, J. (1997): Managing the Risk of Organisational accidents, Aldershot: Ashgate.

Ruighaver, A.B.S.B. Maynard, S. Chang (2007) Organisational security culture: Extending the end-user perspective, compute r s & s e c u r i t y 2 6 (2 0 0 7) 5 6–6 2.

Schein, E.H. (2004): Organizational Culture and Leadership, Third Edition, San Francisco: Jossey-Bass.

SjekkIT. Verktøy for å måle informasjonssikkerhetskultur utviklet av NTNU og Sintef for NSM.

Digitalization and big data

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Pitfalls of machine learning for tail events in high risk environments

C. Agrell, S. Eldevik, A. Hafver, F.B. Pedersen & E. Stensrud
DNV GL AS, Norway

A. Huseby
University of Oslo, Norway

ABSTRACT: Most of today's Machine Learning (ML) methods and implementations are based on correlations, in the sense of a statistical relationship between a set of inputs and the output(s) under investigation. The relationship might be obscure to the human mind, but through the use of ML, mathematics and statistics makes it seemingly apparent. However, to base safety critical decisions on such methods suffer from the same pitfalls as decisions based on any other correlation metric that disregards causality. Causality is key to ensure that applied mitigation tactics will actually affect the outcome in the desired way. This paper reviews the current situation and challenges of applying ML in high risk environments. It further outlines how phenomenological knowledge, together with an uncertainty-based risk perspective can be incorporated to alleviate the missing causality considerations in current practice.

1 INTRODUCTION

In this paper we highlight some pitfalls to avoid in the design of Machine Learning (ML) applications for high risk engineering applications. We also propose some recommendations to consider in model development, and emphasize the introduction of causality constraints based on phenomenological knowledge.

1.1 *Background*

ML is recognized as one of the key enablers for the fourth industrial revolution[1]. In this setting it is often communicated as a tool that, together with increased access to data and computational power, can unlock a huge potential for increased efficiency, new insights and ultimately new revenue streams. Success stories of businesses that have disrupted entire industries are often shared to inspire investment in similar technologies. However, for operation of complex engineering systems in high risk environments, the new challenges that appear are often not clearly expressed.

Concerns have been raised regarding the reliability and trustworthiness of systems relying on Artificial Intelligence (AI) in general, and specifically related to the current main strategy of implementation. Knight (2017) emphasizes that "not knowing how the most advanced algorithms do what they do might become a serious problem as computers become more responsible for making important decisions". The problem of accidents in ML systems is discussed in detail in a paper led by Google Brain researchers Amodei, Olah, Steinhardt, Christiano, Schulman, & Mané (2016). Here the authors motivate the increasing need to address these safety problems by some general trends, one of which relates to the increasing autonomy in AI systems. "Systems that simply output a recommendation to human users, such as speech systems, typically have relatively limited potential to cause harm. By contrast, systems that exert direct control over the world, such as machines controlling industrial processes, can cause harms in a way that humans cannot necessarily correct or oversee."

In this paper we turn our attention towards the introduction of ML in design and operation of complex engineering systems in high risk environments. These are systems where today's methods for assessing risk relies heavily on understanding the underlying physical processes and our ability to model these. This is in contrast to e.g. mass production of components where more data-driven, statistically founded methods can be applied to estimate rates of failure.

We acknowledge the need for ML technologies to address increasing complexity of engineering systems, and the challenges that follow in quantifying uncertainty and risk based on detailed numerical simulation. However, this type of application reduces the tolerance for erroneous model behavior. Most of the methods applied in ML are based

[1]For a broader discussion see e.g. Lu (2017) and Dopico, Gomez, De la Fuente, García, Rosillo, & Puche (2016).

on historic statistical models, enhanced by recent breakthroughs in computer science. The assumptions, limitations and practical challenges of these statistical models still remain, and serve as recurring pitfalls in the digital era.

For a given application both the statistical and ML mindsets may have its advantages and drawbacks, but we argue that for the high risk engineering problems discussed in this paper, uncritical application of ML is unwarranted. We believe that experienced ML practitioners know this, and the main problem lies in how ML is communicated in the current digitalization boom, and how it is perceived by the increasing number of new practitioners in the field that may also be strongly incentivized to develop solutions for cutting costs through automation.

1.2 *Correlation and causation*

Most of today's ML methods and implementations are based on correlations, which we define in the general sense as *any statistical relationship, whether causal or not, between random variables*, i.e. the degree of which two or more variables tend to vary together.

"Correlation does not imply causation" is a well used phrase within statistics, and describes what still remains as one of the major pitfalls in the analysis of data. The importance of this distinction depends on the degree to which one intends to intervene, and the consequence of erroneous intervention. See e.g (Pearl 2010). For many ML applications this may not be significant. But for the use cases considered in this paper it plays an important role, and we will argue that causality constraints from phenomenological knowledge should be incorporated in ML models—to strengthen model performance for tail events, to increase model transparency, and to make it easier to falsify models that do not comply with observations.

1.3 *Structure of this paper*

Section 2 is included for readers that are unfamiliar or new to the field of ML, and gives a brief introduction to some core concepts and their relation to classical statistics. The experienced ML practitioner may jump directly to Section 3, where we propose a set of model properties that should be accommodated for ML applications in high risk, low probability scenarios. Going beyond model selection, some general pitfalls are highlighted in Section 4 which gives an example illustrating the use of ML for anomaly detection and space exploration in Structural Reliability Analysis (SRA). Finally, our conclusions and some final remarks are summarized in Section 5.

2 A BIT OF HISTORY

Often times discussions regarding ML and AI can become fairly opaque, colored by data science jargon, cognitive analogies and marketing buzzwords. This section will clarify the relation between ML and statistical methods by comparing the classes of ML problems with their statistical counterparts, and tracing their origins. This section will only briefly present the statistical background that underpin the main classes of ML methods. See e.g. (Hastie, Tibshirani, & Friedman 2001) for a thorough introduction, or (Domingos 2012) for a more informal description on how they are used in practice.

2.1 *A formal definition of machine learning*

Although the field of ML is often defined in cognitive terms, as giving computers the ability to learn without being programmed, a more formal definition often cited is the one by Mitchell (1997). "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E." Further, ML is generally classified into two categories called supervised and unsupervised learning.

Supervised learning: For some unknown relationship between variables $\mathbf{x} \rightarrow \mathbf{y}(\mathbf{x})$ where N pairs of data are observed $\left\{\left(\mathbf{x}_i, \mathbf{y}_i\right)\right\}_{i=1}^{N}$, the goal is to estimate $\mathbf{y}^*$ for unobserved $\mathbf{x}^*$. Here, the experience E is the observed data, the task T is to predict $\mathbf{y}^*$ and P is usually a measure of the difference between the predicted $\mathbf{y}^*$ and the true value $\mathbf{y}(\mathbf{x}^*)$.

Unsupervised learning: The goal is to discover patterns in unlabeled data. For instance, based only on the x-values in the above example, $\left\{\mathbf{x}_i\right\}_{i=1}^{N}$, estimate the distribution over the data or identify clusters, etc. With the task T of clustering in mind, the experience E is still the observed data and P might be a measure on cluster compactness or separability.

2.2 *Machine learning and statistical modelling*

The increasing popularity of ML today may be credited to recent advancements within computer science, although most of supervised and unsupervised learning have roots in traditional statistical methods. An overview is illustrated in Figure 1. *Supervised learning* is based on *prior* knowledge, and covers regression and classification (discriminant analysis).

Regression considers a continuous dependent variable represented by a model function fitted to data. Assumptions are generally made about the data generation process, e.g. homoscedasticity (equal finite

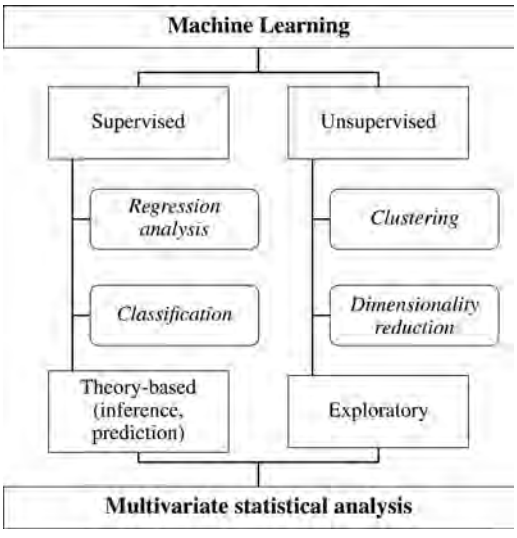


Figure 1. Machine learning vs. multivariate statistics.

variance), independence and normality. The earliest form of regression was the method of least squares, published by Legendre (1805) and Gauss (1809).

Classification or discriminant analysis has the same objective as regression, but where the dependent variable is discrete, and typically comes in a categorical form as labels from a finite set. Some of the earliest work was by Fisher (1936) leading to Fisher's linear discriminant function. Other popular methods are Logistic regression that dates back to Verhulst (1838), decision trees (Morgan & A. Sonquist 1963) and k-nearest neighbors (Cover & Hart 1967).

Note that due to the close link between these two categories of supervised learning problems, regression models may be altered to work for discriminant analysis and vice versa.

Unsupervised learning is related to data exploration problems. The main exploratory methods are often classified as clustering or dimensionality reduction.

Clustering analysis considers the task of grouping objects into sets (clusters) such that objects in the same set are more similar to each other than to those in other sets. No precise definition of a cluster exists, and this is one of the reasons why there are so many different clustering algorithms (Estivill-Castro 2002). Some popular alternatives representing different approaches to the problem of clustering are k-means (MacQueen 1967), hierarchical clustering (Sibson 1973), Gaussian mixture models using expectation-maximization (Dempster, Laird, & Rubin 1977) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester, Kriegel, Sander, & Xu 1996).

Dimensionality reduction is the procedure of reducing the number of input variables to a smaller set of principal variables. One fundamental approaches is principal component analysis (Pearson 1901), developed by Karl Pearson, considered the father of modern statistics.

The methods mentioned above is by no means a complete list, and the references cited are meant to give indications on early work on the different subjects as true origins are often debatable and outside the scope of this paper. A lot of research has gone into extending or improving on these methods since first invented, which has spawned the large variety of models and algorithms used in ML today. There are also other popular methods and widely used techniques within ML that can be placed in more than one category in Figure 1, e.g. artificial neural networks dating back to Hebb (1950) and using the backpropagation algorithm developed by Werbos in 1974.

2.3 *Model design vs. trial-and-error*

Many of the similarities between statistics and ML are just hidden behind different terminology. For instance, where we in ML refer to *features* in a model and that model *weights* are *learned* from data, a statistician would refer to *variables* in a model where data is used to *estimate* model *parameters*. But there are also some important differences in how the (same) models are used, and whether the main emphasis is on designing a good model or obtaining good prediction by trial-and-error.

In classical statistics, the focus is often on testing hypothesis of causes and effects and interpretability of the models applied. A common aphorism in statistics is that "All models are wrong, but some are useful". The analyst should know when the model will break, how it breaks, and if one can still use it anyways. The main goal is understanding the underlying mechanisms that drive the things we observe.

On the contrary, ML has more focus on predictive accuracy of models, with less attention towards model interpretation. Model selection is often based on trial-and-error through cross validation to evaluate goodness-of-fit criteria, where prediction accuracy is evaluated on data that was excluded in the learning process. Although the main goal is to obtain a good prediction, special considerations are made based on what the prediction is used for in a given application. For high risk applications for instance, false positives may be far worse than false negatives (or vice versa), and the model optimization can be weighted accordingly. Still, it is based on observing the desired behavior in future or excluded data.

Although science always has been concerned with both prediction and explanation, the different

philosophies have resulted in much debate between the statistics and machine learning communities. See for instance the discussion in Breiman (2001).

3 ML FOR HIGH RISK ENGINEERING APPLICATIONS

This section highlights some of the main challenges when using ML for high-risk and low probability scenarios. We continue by proposing recommendations to consider in ML model development to address these challenges.

3.1 *Tail events in high risk environments*

Three of the main challenges with ML applied to high risk and low probability scenarios are related to the following:

- **High risk reduces the tolerance for wrong predictions**. The consequence might be catastrophic.
- **Critical consequences often relate to tail events—for which data is naturally scarce**. This increases uncertainty and reduces the accuracy of predictions.
- **The ML models that are able to fit the data well are often opaque.** This makes the model less falsifiable, increasing uncertainty and reducing decision makers' ability to trust the model.

The first point relates to the decisions that are made based on model predictions. When the high risk is associated with a catastrophic consequence, the tolerance for a wrong prediction is clearly low. Moreover, severe consequences are often related to a rare event. This introduces additional uncertainty as data is scarce[2]. These first two points are in direct contrast to the typical ML applications today (e.g. non-consequential recommendation engines). In this respect, assurance of safety critical systems relying on ML is also recieving increasing attention, see e.g. Brandsæter & Knutsen (In press).

The last point on model opaqueness (lack of transparency) relates to quantification of model discrepancy and the ability to falsify models that do not comply with observations. The added uncertainty from model opaqueness may at first seem purely subjective. However, as we will see in the following section, addressing this by increasing model transparency may relax requirements on accuracy and vice versa.

[2]Scarce in the sense that the size of data is small compared to the number of relevant dimensions. This is because the event under consideration, e.g. structural failure, is rare and expensive to approximate by experiments.

3.2 *Recommendations on model development*

The challenges stated above will make any ML or statistics based model prone to erroneous, and potentially dangerous, use. However, many of the most common pitfalls in using such correlation based or data-driven methods may be avoided by proper model selection and design. To increase confidence in the safe use of ML models we propose some recommendations in this section.

For supervised learning, and when we want to use ML for prediction in general engineering applications, one should seek to adopt or develop models with the following attributes:

1. **Flexibility:** The models ability to represent a large class of functions.
2. **Constrainability:** The ability to impose model constraints based on phenomenological knowledge.
3. **Probabilistic inference:** Probabilistic representation of model output defined on the entire model range (i.e. the model output is represented by a distribution).

The first two concepts relate to the problem of underfitting and overfitting respectively. The third is important when the model is used in assessment of risk and reliability.

From an engineering perspective, **flexibility** is needed to capture the behavior of complex physical systems and their response. Most non-parametric models fulfill this requirement. The definition of "a large class of functions" in this context may not be precise, as it certainly depends on the problem at hand. A typical example may be all continuous or differentiable functions with a finite number of discontinuities.

Constrainability is beneficial for two main reasons. First, it reduces the possibility of overfitting the data, thus increasing the robustness and the performance for application on future data. This is particularly important for tail behavior problems where data is scarce. The second reason is that imposing constraints based on phenomenological knowledge reduces model opaqueness, i.e. increases transparency.

By a *transparent model* we mean that the relationship between model inputs and outputs can be meaningfully understood by humans, and that model characteristic properties and limitations of the model can be understood without explicit numerical computation. The most straightforward example is simple linear regression with linear basis, i.e. fitting a line. In contrast, for an opaque model, the model behavior can only be investigated through computation. These models are often referred to as black boxes, where the only way to fully characterize the model's properties is through

exhaustive computation; evaluating the model for *all* possible inputs.

The scientific method is based on the principle that any model, or hypothesis, should be falsifiable. All models, including ML models, are based on a set of assumptions. For a model or an assumption to be falsifiable, it must, in principle, be possible to make an observation that would show the assumption to be false. Thus, model transparency is important as it enables one way of falsification. ML models are typically falsified by observing poor accuracy. That is, observations are made that differ significantly from model predictions. However, understanding why such discrepancies occur is often difficult in an opaque model. If we assume that the observation is not erroneous (observed discrepancy is not related to noise), then the root cause might either be lack of relevant data, i.e. the model is built on too few datapoints close to the observed discrepancy—in which case a larger degree of uncertainty is expected. Or the root cause is related to violation of the underlying model assumptions—and the model must be changed.

In order to develop useful ML models for high risk applications, a compromise usually has to be made between model transparency and flexibility. To counteract the negative tradeoff of an initially opaque, but flexible model, we might impose constraints based on phenomenological knowledge related to causality. This can be thought of as "putting the black box inside a white box", i.e. enabling deduction of bounding model properties through the imposed constraints.

Figure 2 illustrates constraints in the form of boundedness, monotonicity and convexity. Three datapoints have been observed, and for simplicity we assume the data does not contain noise. This means that we are looking for an interpolation model, a function passing through all three points. The shaded areas show where the function cannot enter due to the imposed constraints. Hence, starting with the space of all functions passing through the three points, the constraints reduce the space of possible interpolation functions. Assuming that the constraints are based on phenomenological knowledge that hold in reality, the constrained models are less prone to overfitting (more robust). In the case where an observation is made within the shaded area, the model is falsified immediately as the assumptions behind the constraints do not conform with an observed outcome. Hence, by imposing constraints based on phenomenological knowledge, either a) performance is increased, or b) the model is falsified and the modeler learn something fundamentally new about the phenomenon studied, which can be applied in future modelling.

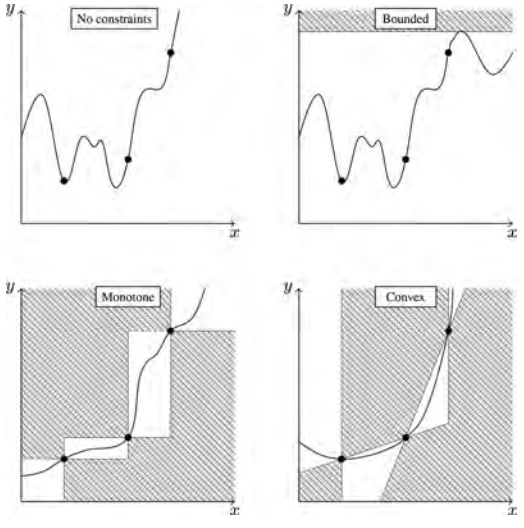


Figure 2. Effect of some different constraints for interpolation in data without noise. The interpolation function cannot enter the shaded areas.

The constrained models shown in Figure 2 may be restricted further by imposing multiple constraints, e.g boundedness and monotonicity or monotonicity and convexity. Note also that for noisy data, this means interpreting constraints in terms of probabilities using the assumed distribution of noise. The example is motivated by the more general class of constraints in the form of partial differential inequalities, for which phenomenological knowledge related to causal effects in various physical phenomena may often be available.

Practically, imposing constraints such as the ones illustrated in Figure 2 means translating the phenomenological constraints to constraints in the ML optimization algorithm. Many techniques exist for including constraints in ML through optimization, usually in order to obtain regularization effects, but we emphasize that developing the necessary links between these constraints and phenomenological knowledge will be highly beneficial. See for instance (Yu 2007) or (Maatouk & Bay 2017) for some examples and further discussion.

Probabilistic inference on the ML model output is needed for risk and reliability analysis applications. This means that the model output should optimally be in the form of a distribution. Model predictions in the form of fixed values and best estimates are not applicable. The dangers of expressing risk through expected values is well known, and modern definitions of risk usually relate to the distribution over possible outcomes. Hence, some quantification of prediction uncertainty is essential. It should be noted that this goes beyond the probabilities often used to report model accuracy for ML classifiers. The fact that an object is correctly classified 99% of the time might be insignificant if the outcomes of the remaining 1% is associated with severe consequences.

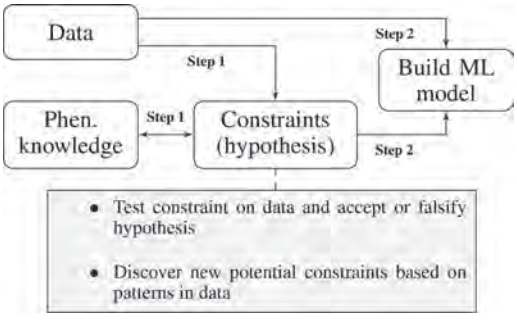


Figure 3. ML workflow with emphasis on constraints.

There is a traditional approach to this problem, where ML has played a role in mathematical models for calculation of risk and reliability. First, the models are scrutinized through human quality control to ensure that the model accuracy is sufficient or at least on the conservative side. This quickly becomes infeasible for higher dimensional models. Further, a single distribution representing the model uncertainty is often established from statistical analysis of prediction accuracy alone, assuming uniformly distributed data. This is no longer feasible for higher dimensional models or when the input data is far from uniformly distributed.

3.3 *Working with constraints*

Figure 3 illustrates of the workflow for building ML models with emphasis on constraints. For simplicity we ignore work on data cleaning and feature selection that naturally comes prior to model development.

Any hypothesis on model constraints coming from assumed causalities in the phenomenon under consideration must be tested to identify to what degree they hold in the observed data. Hence, the task of hypothesis testing is emphasized. There might also be valid constraints that are not immediately identifiable from phenomenological knowledge. It could therefore be valuable to search for possible constraints by unsupervised learning, to serve as hints to the modeler, and help identifying additional constraints before further testing and possible inclusion in the ML model.

This type of workflow will move the typical approach for building ML models today closer to traditional statistical modelling.

4 RELEVANT AREA OF APPLICATION – SRA

In this section we give a concrete example of an application area where ML is linked with engineering risk analysis – Structural Reliability Analysis (SRA), where the recommendations given Section 3 are highly relevant. In addition, we highlight two general pitfalls related to a common, but possibly misconceived, idea on how ML may be used in this context.

4.1 *Structural reliability analysis*

Structural reliability analysis, or SRA for short, is the fundamental building block of modern risk-based engineering methodologies. For a thorough introduction reference is made to Madsen, Krenk, & Lind (2006). The underlying theory combines structural analysis with statistics and probabilistic modelling to assess uncertainties of information that contributes to the probability of structural failure.

SRA may generally be described as the problem of establishing the probability

$$P\left(G(\mathbf{x}) \leq 0\right) \tag{1}$$

where $\mathbf{x}$ is a vector of stochastic variables, e.g. structural dimensions, material properties, loads and model uncertainties. The function $G(\mathbf{x})$ is referred to as the limit state, and is defined such that $G(\mathbf{x}) \leq 0$ if and only if the scenario represented by $\mathbf{x}$ results in structural failure, see Figure 4. In the literature the limit state is often presented as

$$G(\mathbf{x}) = R(\mathbf{x}) - L(\mathbf{x}) \tag{2}$$

where $R(\mathbf{x})$ and $L(\mathbf{x})$ represent the structural resistance, or capacity, and load effect respectively, although these are often not separable in practice.

The main tasks in a structural reliability analysis is to establish a suitable limit state function and distributions of all the input parameters, so that one may estimate the failure probability given by Eq. 1 and analyse the sensitivity of parameters and

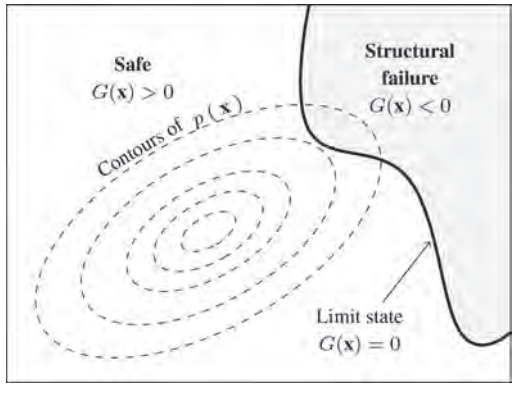


Figure 4. Illustration of SRA in two dimensions.

decisions that will affect the system. Informally, one might say that we use data and domain knowledge to establish the input distributions describing the current state of the system, and extrapolate to states where the system has failed using very limited data related to system failure in combination with advanced phenomenological simulations based on first principle physics.

4.2 *Machine learning in SRA*

Some of the key challenges in SRA today relate to the rapid increase in structural complexity of engineering systems, including more automation and software intensive control systems, together with a demand for higher system utilization and the need for more accurate and reliable models to support decision making under uncertainty. At the same time, ubiquitous sensor data provides new information that could potentially reduce the uncertainty if the information could be incorporated into the models.

In practice, this means that the function $G(\mathbf{x})$ in Eq. 1 and the distribution over the input space $\mathbf{x}$ will take a more complicated form. To cope with this, the technologies we now label ML can be useful, e.g. to address the following problems:

- Find $\mathbf{x}$: Establish distribution over $\mathbf{x}$ using all relevant data.
- Find G: Through data related to structural behavior, combined with data from past experience, experiments and simulations of structural failure, establish the limit state function that classifies all $\mathbf{x}$'s as *Safe* or *Failure*.

Note that in practice these two problems are generally intertwined, in the sense that inference about a model parameter (the $\mathbf{x}$'es) may only be observable through its effect on the system response. E.g. some $y(\mathbf{x})$ is observed, where the mapping $y(\cdot)$ is in our representation baked into the general function G.

The use of ML in SRA has traditionally been confined to smaller subcomponents where human quality control is possible, but for more complex systems this quickly becomes infeasible. The more general task of approximating functions like the limit state $G(\mathbf{x})$ using ML together with a limited number of realisations (experiments or numerical simulations) has received increasing attention over the last decades, within the field of Uncertainty Quantification (UQ). See for instance (Sullivan 2015). UQ aim to quantify the ML uncertainty introduced through approximation, as well as how uncertainty in input parameters propagates through such models.

4.3 *General ML pitfalls in SRA*

As for all ML applications, there are some general pitfalls to look out for. This section highlights some of the challenges related to how introduction of ML in reliability analysis is often depicted. This relates to a growing appetite for ideas like the following

- Due to the increasing instrumentation of systems, more data is available about the loads and structural behavior of systems at any time. By combining this with historical data from many other similar systems where the structural integrity is known (we know whether or not they have failed, and how), we could detect any abnormal behavior. With this information we could create warnings before potential failure occurs, and possibly also help the system back into normal operation.

Anomaly detection will probably play a larger role in risk assessment in the future, but there are some pitfalls that the industry needs to be aware of. The first relates to the quantification of the safety margin, which is often represented as a probability of failure or through some other metric relating the current physical condition with conditions corresponding to structural failure.

- For complex engineering systems, quantifying the margin of safety based on operational data alone is unlikely.

This statement might be obvious, from the many different ways a system may fail in practice and the assumption that these systems are designed not to fail. The next argument however is a bit more subtle

- From a data exploration perspective, when observing system states outside normal operation one might unknowingly have transitioned away from the default system behavior, leaving all previous observations biased, and possibly irrelevant.

This statement impacts the basic assumption in ML that future data will come from the same distribution as the data the model was trained on. This is illustrated by an example in Figure 5.

Following the SRA setting illustrated in Figure 4, we assume that the limit state is defined in terms of material over-utilization. Often the criteria for when ultimate failure occurs is difficult to express mathematically, and conservative approaches are applied by defining failure as some identifiable prior event. One such limit from material science is the yielding criteria of ductile materials such as steel. The stress-strain relationship of a material under some loading is linear up until the yielding point, and the material behaves elastic in this region. I.e. unloading the material will bring it back to its original unharmed state[3]. For continued loading beyond the yield point, the material will exhibit plastic behavior until rupture.

[3]Ignoring other failure modes such as fatigue.

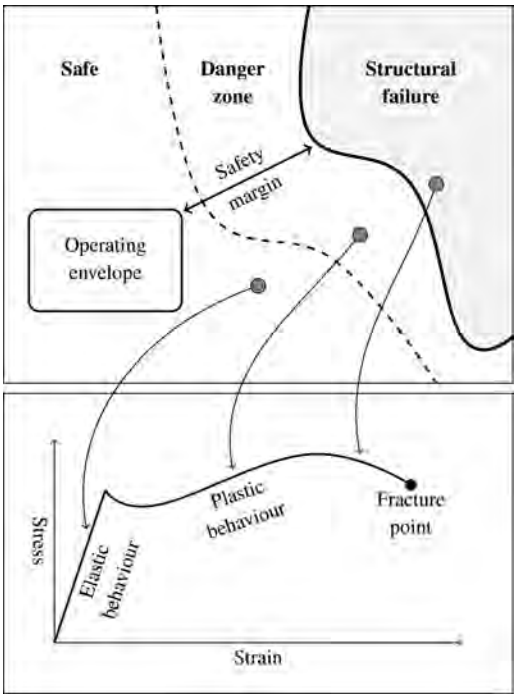


Figure 5. Illustration of structural response in different operating environments.

A limit state defined from the yield criterion can be illustrated as the dashed line in Figure 5, whereas the boundary of the structural failure set (black line) represents material fracture[4]. Imagine an anomaly detection agent that warns whenever the system behaves outside the normal operating envelope, and estimates procedures for moving the system back into normal operation. For elastic material response (leftmost point), the data used to train the model is still valid. But when materials are pushed closer to their limits, some properties are fundamentally changed. The elastic limit will change due to work hardening, and the stress-strain curve is no longer valid. Furthermore, when the material is over-utilized over a certain threshold, the reduction of load may initiate failure modes previously non-relevant, and uninformed decisions may be catastrophic.

This example is an oversimplification, but illustrates some challenges with introducing purely data driven agents. Due to the increased computational capacities and scientific models available today there is an increased push to utilise systems closer to their limits. In the above example this means allowing operation closer to the true failure limit, and compensating by increased control, uncertainty reduction, and more detailed understanding of the failure modes.

[4]For load controlled scenarios the top of the stress-strain curve represents maximum capacity. Unless the load is decreased the material will eventually fracture

5 CONCLUDING REMARKS

The field of ML is largely based on statistical methods, but with a focus that is shifted more towards predictive accuracy and with limited attention towards model interpretation and testing hypotheses on causes and effects.

For tail events in high risk environments the modeler is faced with additional challenges, as the tolerance for error is reduced and accuracy is needed in distribution tails rather than where the main bulk of data is. Because of this, opaque black-box type models are difficult to work with as the means for falsification may be limited to observations of future performance.

Therefore, research and development of ML models for such applications should be guided towards enabling incorporation of causality constraints reflecting the modeler's phenomenological knowledge.

REFERENCES

Amodei, D., C. Olah, J. Steinhardt, P. Christiano, J. Schulman, & D. Mané (2016). Concrete problems in AI safety. *CoRR abs/1606.06565*.

Brandsæter, A. & K.E. Knutsen (in press). Towards a framework for assurance of autonomous navigation systems in the maritime industry. European Safety and Reliability Conference, ESREL 2018.

Breiman, L. (2001, 08). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *16*.

Cover, T. & P. Hart (1967, January). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory 13*(1), 21–27.

Dempster, A.P., N.M. Laird, & D.B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological) 39*(1), 1–38.

Domingos, P. (2012, October). A few useful things to know about machine learning. *Commun. ACM 55*(10), 78–87.

Dopico, M., A. Gomez, D. De la Fuente, N. García, R. Rosillo, & J. Puche (2016). A vision of industry 4.0 from an artificial intelligence point of view. In *Proceedings on the International Conference on Artificial Intelligence (ICAI)*, pp. 407. WorldComp.

Ester, M., H.-P. Kriegel, J. Sander, & X. Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. pp. 226–231. AAAI Press.

Estivill-Castro, V. (2002, June). Why so many clustering algorithms: A position paper. *SIGKDD Explor. Newsl. 4*(1), 65–75.

Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics 7*(2), 179–188.

Gauss, C. (1809). Theoria motus corporum coelestium in sectionibus conicis solem ambientum.

Hastie, T., R. Tibshirani, & J. Friedman (2001). *The Elements of Statistical Learning*. Springer.

Hebb, D.O. (1950). The organization of behavior: A neuropsychological theory. *Science Education 34*(5), 336–337.

Knight, W. (2017). The dark secret at the heart of AI. *MIT Technology Review 120*(3), 55–63.

Legendre, A. (1805). *Nouvelles méthodes pour la détermination des orbites des comètes*. Nineteenth Century Collections Online (NCCO): Science, Technology, and Medicine: 1780–1925. F. Didot.

Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. *Journal of Industrial Information Integration 6*(Supplement C), 1–10.

Maatouk, H. & X. Bay (2017, Jul). Gaussian process emulators for computer experiments with inequality constraints. *Mathematical Geosciences 49*(5), 557–582.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, Berkeley, Calif., pp. 281–297. University of California Press.

Madsen, H., S. Krenk, & N. Lind (2006). *Methods of Structural Safety*. Dover Civil and Mechanical Engineering Series. Dover Publications.

Mitchell, T.M. (1997). *Machine Learning* (1 ed.). New York, NY, USA: McGraw-Hill, Inc.

Morgan, J. & J.A. Sonquist (1963, 06). Problems in the analysis of survey data and a proposal. *58*, 415–434.

Pearl, J. (2010, Sep). The mathematics of causal relations. *In P.E. Shrout, K. Keyes, and K. Ornstein (Eds.), Causality and Psychopathology: Finding the Determinants of Disorders and their Cures*, 47–65.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2*(11), 559–572.

Sibson, R. (1973, 01). Slink: An optimally efficient algorithm for the single-link cluster method. *16*.

Sullivan, T. (2015). *Introduction to Uncertainty Quantification*. Texts in Applied Mathematics. Springer International Publishing.

Verhulst, P.F. (1838). Notice sur la loi que la population suit dans son accroissement. *Curr. Math. Phys 10*, 113–120.

Werbos, P. (1974, 01). Beyond regression: new tools for prediction and analysis in the behavioral sciences.

Yu, T. (2007). *Incorporating prior domain knowledge into inductive machine learning: its implementation in contemporary capital markets*. Ph. D. thesis, University of Tech., Sydney. Faculty of Information Technology.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data

I. Abdallah, V. Dertimanis, H. Mylonas, K. Tatsis & E. Chatzi
Department of Civil, Environmental and Geomatic Engineering, ETH Zürich, Zürich, Switzerland

N. Dervilis & K. Worden
Department of Mechanical Engineering, The University of Sheffield, Sheffield, UK

E. Maguire
Vattenfall AB, Edinburgh, UK

ABSTRACT: In the context of Operation and Maintenance of wind energy infrastructure, it is important to develop decision support tools, able to guide engineers in the management of these assets. This task is particularly challenging given the multiplicity of uncertainties involved, from the point of view of the aggregated data, the available knowledge with respect to the wind turbine structures, as well as the varying operational and environmental loads. We propose to propagate wind turbine telemetry through a decision tree learning algorithm to detect faults, damage, and abnormal operations. The use of decision trees is motivated by the fact that they tend to be easier to implement and interpret than other quantitative data-driven methods. Furthermore, the telemetry consists of data from condition and structural health monitoring systems, which lends itself nicely in the field of Big Data as large amounts are continuously sampled at high rate from thousands of wind turbines. In this paper, we review several decision tree algorithms, we then train an ensemble Bagged decision tree classifier on a dataset from an offshore wind farm comprising 48 wind turbines, and use it to automatically extract paths linking excessive vibrations faults to their possible root causes. We finally give an outlook of a cloud computing based architecture to implement decision tree learning involving Apache Hadoop and Spark.

1 INTRODUCTION

Wind Turbines (WTs) maintenance and inspection relies on conventional techniques (Yang & Sun 2013), such as visual inspection, non-destructive evaluation and standard signal processing, trend analysis and statistics of data streamed from the Supervisory Control And Data Acquisition (SCADA). Specialized Condition Monitoring (CM) systems are only available on specific components such as the gearbox and main bearing (Hameed et al. 2009), while far forming part of the actual engineering practice (Grasse et al. 2011) Structural Health Monitoring (SHM) systems are deployed mostly for research purposes, or temporarily during the certification stage. In fact, there exists a dislocation between (i) data derived from CM systems (e.g. power output, rotor RPM), (ii) data obtained from specialized SHM (e.g. tower acceleration, strain on blade root), and (iii) specialized maintenance strategies of individual WT components. As a result, there are currently no holistic approach, and systematic, quantitative and automated tools for monitoring and diagnostics of WTs, for operation, maintenance (O&M) and decision making within their life-cycle. Towards this end, we propose to perform automated fault diagnostics and root cause analysis of faults on wind turbines (WTs) on the basis of decision tree classifiers. A decision tree is a predictive model that maps observations to their target values or labels. The key concept lies in running WT telemetry data through a decision tree learning algorithm for detecting faults, errors, damage patterns, anomalies and abnormal operation (i.e., end states). A decision tree essentially comprises a machine learning tool for classification of event outcomes. For a given initiating event, multiple end states are possible, linking each event to an associated probability of occurrence. Once built and trained, and given a new set of measurements, the decision tree may be used to predict end states and classify (discover) previously unknown end states. By examining the paths that lead to failure-predicting leaf nodes, one may distinguish the possible sources (root causes) of error. The remainder of this article is organized as follows. In section 2 we revisit the decision tree learning theory. In section 3 we

train an ensemble of bagged decision tree classifiers with the standard CART algorithm on a condition monitoring data set from the Lillgrund offshore wind farm comprising 48 wind turbines and use it to perform a diagnostics to elucidate the root cause of excessive vibrations. Finally, in section 4 we further our discussion to show how decision tree learning can be expanded to big data based applications for monitoring and diagnostics for wind turbines using the object-oriented based decision tree concept cite (Abdallah 2017).

2 DECISION TREES

A Decision Tree (also called Classification or Prediction Tree) is designed to classify or predict a discrete category from the data. Decision Trees (DTs) are a non-parametric supervised learning method used for classification (and regression). In the machine learning sense, the goal is to create a classification model (classification tree) that predicts the value of a target variable (also known as label or class) by learning simple decision rules inferred from the data features (also known as attributes or predictors). From Figure 1 an internal node *N* denotes a test on an attribute, an edge *E* represents an outcome of the test, and the Leaf nodes *C* represent class labels or class distribution.

Four reasons motivated us to work with decision tree classifiers. First, they can be learned and updated from data relatively fast compared to other methods. Second, they are visually more intuitive, simpler and easier to assimilate and interpret by humans/engineers. Third, unlike other classification methods, with decision tree classifiers one is able to perform data-driven root cause analysis of faults; one can trace a path from the end state (e.g. blade damage) to the initiating event (e.g. wrong parameters in control system), a way that follows the sequence and chronology of how events are interlinked. Last, it has been shown that

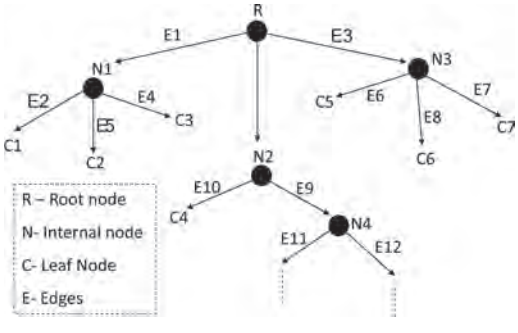


Figure 1. Graphical representation of a decision tree (DT) classifier. DT terminologies are also shown.

the accuracy of decision tree classifiers is comparable or superior (especially ensemble descision tree classifiers) to other models and in fact display the best combination of error rate and speed (Lim, Loh, & Shih 1997, Hand 1997, Lim, Loh, & Shih 2000, Caruana & Niculescu-Mizil 2006).

A decision tree is is a tree-structured classifier built by starting with a single node that encompasses the entire data and recursively splitting the data within a node, generally into two branches (some algorithms can perform *multiway* splits) by selecting the variable (dimension) that best classifies the samples according to a split criterion, i.e. the one that maximizes the information gain (Equation 1) among the random subsample of dimensions obtained at every point. The splitting continues until a terminal leaf is created that meets a stopping criterion such as a minimum leaf size or a variance threshold. Each terminal leaf contains data that belongs to one or more classes. Within this leaf, a model is applied that provides a fairly comprehensible prediction, especially in situations where many variables may exist that interact in a nonlinear manner as is often the case on wind turbines (Carrasco Kind & Brunner 2013). Algorithm 1 shows pseudocode of a generic decision tree learning algorithm.

Formally, the splitting is done by choosing the attribute that maximizes the Information Gain (I_G), which is defined in terms of the impurity degree index:

$$I_G(T,M) = I_d(T) - \sum_{m \in M} \frac{|T_m|}{|T|} I_d(T_m) \tag{1}$$

where T is the training data in a given node, M is one of the possible dimensions along which the node may be split, m are the possible values of M, $|T|$ is the size of the training data, $|T_m|$ is the number of objects for a given subset m within the current node, and I_d is the function that represents the degree of impurity of the information. There are three standard methods to compute the impurity index (I_d). The first method is by using the information entropy (H), which is defined by:

$$I_d(T) \equiv H(T) = -\sum_{i=1}^{n} f_i log_2(f_i) \tag{2}$$

where i is the class to be predicted, n is all possible classes, and f_i is the fraction of the training data belonging to class i. The second option, is to measure the Gini impurity (G). In this case, a leaf is considered pure if all the data contained within it have the same class. The Gini impurity can be computed inside each node using the following simplified equation:

$$I_d(T) \equiv G(T) = 1 - \sum_{i=1}^{n} f_i^2 \quad (3)$$

The third method is to use the classification error (C_E):

$$I_d(T) \equiv C_E(T) = 1 - \max f_i \quad (4)$$

where the maximum values are taken among the fractions f_i within the data T that have class i. Figure 2 shows the three impurity indices, for a node with data that are categorized into two classes, as a function of the fraction of the data having a specific class. If all of the data belong to one class, the impurity is zero. On the other hand, if half of the data have one class and the remaining data belong to the other class, the impurity is at its maximum (Carrasco Kind and Brunner 2013).

There exist several algorithms for training decision trees from data including *ID3, C4.5, C5.0, J48, SPRINT, FACT, FIRM, SLIQ, CHAID, QUEST, CRUISE, PUBLIC, BOAT, RAINFOREST, MARS, RIPPER and CART*. In the following we will briefly mention some ofthe more common algorithms. Ross Quinlan developed ID3 (Quinlan 1986) which stands for Iterative Dichotomiser 3, and its later iterations include C4.5 and C5.0. ID3 attempts to generate the smallest multiway tree. If the problem involves real-valued variables, they are first binned into intervals, each interval being treated as an unordered nominal attribute. C4.5 converts the trained trees into sets of if-then rules (Quinlan 1994) and improves on ID3 by allowing both discrete and continuous attributes, missing attribute values, attributes with differing costs, and pruning trees to avoid over-fitting are usually applied to improve the ability of the tree to generalise to unseen data. The accuracy of each rule is then evaluated to determine the order in which they should be applied. C5.0 is essentially the same as C4.5 but uses less memory and builds smaller rule sets while being more accurate.

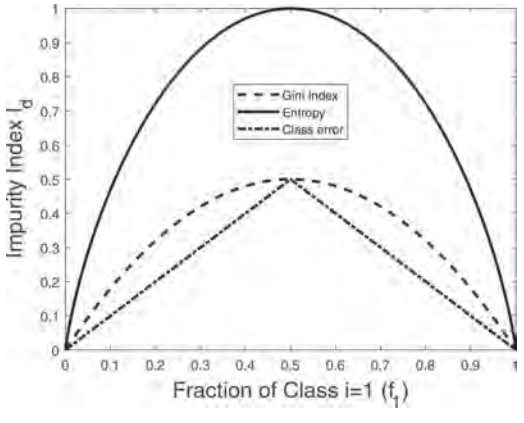


Figure 2. Impurity index I_d for a two-class example as a function of the probability of one of the classes f_1 using the information entropy, Gini impurity and classification error. In all cases, the impurity is at its maximum when the fraction of data within a node with class 1 is 0.5, and zero when all data are in the same category.

CART (Breiman, Friedman, Olshen, & Stone 1984) which stands for Classification and Regression Trees is very similar to C4.5, but it differs in that it supports numerical target variables (regression) and constructs binary tree based on a numerical splitting criterion recursively applied to the data instead of constructing rule sets.

According to (Lim, Loh, & Shih 2000) *C4.5* and *QUEST* have the best combinations of error rate and speed, but C4.5 tends to produce trees with twice as many leaves as those from Quest. QUEST is a binary-split decision tree algorithm for classification and regression (Loh & Shih 1997). It uses a significance test to select variables for splitting. An advantage of the QUEST tree algorithm is that it is not biased in split-variable selection, unlike CART which is biased towards selecting split-variables which allow more splits, and those which have more missing values.

CRUISE is one of the most accurate decision tree classifiers (Loh 2011) that is also efficiently capable of performing multiway splits. (Kim & Loh 2001) proposed CRUISE which stands for Classification Rule with Unbiased Interaction Selection and Estimation that splits each node into as many as subnodes, which precludes the use of greedy search methods. CRUISE is practically free of selection bias (Kim & Loh 2001) and is capable of integrating interactions between variables when growing the tree. CRUISE borrows and improves upon ideas from many sources, but especially from FACT, QUEST, and GUIDE for split selection and CART for pruning.

Finally a word about *RainForest* which is not a decision tree classifier per se but rather a framework. *RainForest* was proposed to make decision tree construction more scalable (same as *BOAT* which is in fact faster than *RainForest* by a factor of three while constructing exactly the same decision tree, and can handle a wide range of splitting criteria (Gehrke et al. 1999)). According to (Gehrke et al. 2000), a thorough examination of the algorithms in the literature (including C4.5, CART, CHAID, FACT, ID3 and extensions, SLIQ, Sprint and QUEST) shows that the greedy schema described in Algorithm 4 can be refined to a generic RainForest tree induction schema. In fact, most decision tree algorithms consider every attribute individually and need the distribution of class labels for each distinct value of an attribute

to decide on the splitting criteria. Rainforest is a comprehensive approach for decision tree classifiers that separates the scalability aspects of algorithms for constructing a decision tree from the central features that determines the quality of the tree. Rainforest concentrates on the growth phase of a decision tree due to the time consuming nature of this step RainForest, closes the gap between the limitations to main memory datasets of algorithms in the machine learning and statistics literature and the scalability requirements of a data mining environment (Singh & Sulekh 2017).

Next we present a demonstration of decision tree classifier learning based on a real-world SCADA data set from the Lillgrund offshore wind farm.

Algorithm 1: Pseudocode of a generic recursive decision tree learning algorithm.

```
DTClassifier (TR, Target, Attr);
  Input  : TR: training examples, Target:
           target label, Attr: set of descriptive
           attributes
  Output: DT: decision tree classifier
Create a Root node for the tree;
if TR have all the same label t_i then
    return a single-node tree, corresponding to
     leaf node with that label;
else if the set of attributes Att is empty then
    Return the single-node tree, i.e. Root, with
     most common value of Target in TR;
else
    pick an attribute A from Att (such that A
     maximizes I_G) and create a node R for it;
    for each possible value v_i of A do
        Let TR_{v_i} be the subset of TR that have
         value v_i for A;
        Add an out-going edge E to node R
         labeled with the value v_i;
        if TR_{v_i} is empty then
            attach a leaf node to edge E labeled
             with the target value = most
             common value of Target in TR;
        else
            call
             DTClassifier(TR_{v_i}, Target, Attr −
             A) and attach the resulting tree as
             the subtree under edge E;
        end
    end
    Return the subtree rooted at R;
end
```

3 DEMONSTRATION

Condition monitoring data from 48 wind turbines was collected over a period of 12 months and sampled every 10 minutes, across 64 channels. In total, more than 2.5 million records were available, of which 980 excessive vibration error events are recorded across all wind turbines. The error event of interest in this demonstration is *excessive structural vibrations*.

Data Pre-processing The first step prepares the data for the construction of the prediction tree classifier. Knowledge of the process is helpful in the elimination of parameters (features) that are not significant. The SCADA system recorded parameters can be grouped into the following categories: (i) system related data, e.g., turbine number and index, time stamp, are turbine specific and, therefore, can be excluded from the decision tree classifier training, (ii) operating performance parameters such as power output, pitch and rotor speed, (iii) environmental parameters such as wind speed and wind direction, (iv) temperature measurements for various components such as gearbox, bearings and generator, and finally (4) dynamics parameters such as tower top accelerations. The attributes chosen in this demonstration are shown in Table 1. The table includes both original sensor attributes such as maximum generator rotational speed over a 10min period (*GRpm_max*) and additional derived parameters such as the difference between max and min wind speed over a 10min period (*DMaxMinV*). We open a small parenthesis here. As with most pattern recognition methods, tree-based classification methods work best if the proper features are selected to start with; preprocessing by a data dimensionality reduction techniques such as principal component analysis (PCA) or independent component analysis (ICA) or optimal feature selection approaches such as the wrapper approach integrated with the genetic or the best-first search algorithms can be effective because they find important axes to be used as guideline for the selection of the features upon which a decision tree is trained. However,

Table 1. The attributes (features) used as input to train the decision tree classifier.

Attributes	Description
HSGenTmp	Mean temperature of gear bearing on generator side
Po_max	Max value of active power
DMaxMeanPow	Difference between max and mean active power
GRpm_max	Max generator RPM
DMaxMeanGRpm	Difference between max and mean generator RPM
Pi_min	Min collective pitch
DMaxMinV	Difference between max and min wind speed

in this demonstration we chose not to in order to test the limit of a decision tree classifier for fault diagnostics; for instance wind speed, power and RPM are strongly inter-dependent attributes, but for the same wind speed a wind turbine could be found operating at two different power output levels (during distinct time periods) or two different generator RPM levels, and both cases are considered normal operating modes. How so? Indeed this happens very often when a wind turbine has to de-rate the power output following a demand from the grid side or reduce the generator RPM under a specific noise or load control mode. The target variable *TurbineState* for classification is shown in Table 2. It can take two labels, *NoFault* indicating that the wind turbine is producing electric power under normal operating conditions, and *Vibr* indicating an excessive vibration fault resulting in the wind turbine shutting down, furthermore triggering a message to be send to the vibration support technical team (in order for some corrective action to take place).

Bagged decision tree construction The second step is the construction of the Bagged decision tree classifier. An ensemble of decision trees is often more accurate than any single tree classifier (Bauer & Kohavi 1999, Dietterich 1996). Bagging (Breiman 1996), boosting (Schapire 1990) and random forest are three popular methods of creating accurate ensembles. Previous research indicates that boosting is more prone to overfitting the training data (Freund & Schapire 1996, Opitz & Maclin 1999). Consequently, the presence of noise causes a greater decrease in the performance of boosting. Therefore, this study uses bagging to create an ensemble of bagged decision tree classifiers (using the standard CART algorithm) to better address the noise in the condition monitoring data. Other decision tree algorithms and ensemble approaches will be investigated in future work. This technique can be summarized as, (i) take a bootstrap sample from the data set, (ii) train an un-pruned classification tree and (iii) aggregate the trained tree classifiers. In more detail, Bagging predictors comprise a method for generating multiple versions of a predictor, each on random subsets of the original dataset, and fusing these into a unique final aggregated predictor. This aggregated predictor can typically be used for reducing the variance of a black-box estimator, by introducing randomization into the construction procedure and forming an ensemble (for proof refer to (Breiman 1996, Bühlmann 2012)). The bagging algorithm consists in (1) constructing a bootstrap sample $\left(X_*^{(1)},Y_*^{(1)}\right),\ldots,\left(X_*^{(n)},Y_*^{(n)}\right)$ by randomly drawing n times with replacement from the original data $\left(X^{(1)},Y^{(1)}\right),\ldots,\left(X^{(n)},Y^{(n)}\right)$ (2) computing the bootstrapped estimator (i.e. tree classifier) $\hat{g}_* = h_n\left(X_*^{(1)},Y_*^{(1)}\right),\ldots,\left(X_*^{(n)},Y_*^{(n)}\right)$ where the function $h_n(.)$ defines the estimator as a function of the data, and (3) repeating steps 1 and 2 M times, where M is often chosen as 50 or 100, yielding $\left\{\hat{g}_*^k, k=1,\ldots,M\right\}$ and the bagged estimator is $\hat{g}_{bagg} = \sum_{k=1}^{M} \hat{g}_*^k / M$. In theory, $M \to \infty$ if the bagged estimator is

$$\hat{g}_{bagg} = \mathrm{E}\left[\hat{g}_*^k\right] \tag{5}$$

In the machine learning and statistics literature, the two main performance measures for a classification tree algorithm are its predictive quality and construction time of the resulting tree. In this paper the predictive quality is given by the misclassification rate on the validation data set. As shown in Figure 4, the misclassification rate of the trained Bagged tree is less than 1% when the bag size is more than 10.

Fault diagnostics (offline) The final step uses the newly generated decision tree classifier (e.g. Figure 3) to create diagnostics for individual target fault classes, i.e. *Vibr*. One aspect of fault diagnostics deals with offline root cause analysis, which we will demonstrate here. When diagnosing faults, we are interested in identifying the root causes (or sequence of events) that lead to a large portion of the overall abnormal behavior, where the decision tree edges leading to faults become root cause candidates. In the literature, one can find a limited number of proposed programatic algorithms by which decision tree classifiers can be scanned/probed for root cause analysis (Solé, Muntés-Mulero, Rana, & Estrada 2017). One implementation is as follows (Zheng, Lloyd, & Brewer 2004):

1. Ignore the leaf nodes that correspond to normal operations (i.e. *NoFault*) as they will not be useful in diagnosing faults.
2. Identify, in the decision tree, all leaf nodes with the target fault class of interest, i.e. *Vibr*
3. Ranking: list the leaf nodes with the target fault class ranked by failure count to prioritize their importance.

Table 2. Target variable: *Turbine State*. Turbine state is defined in this demonstration according to two labels.

Class label	Description
NoFault	Normal operation, turbine is producing power, no faults
Vibr	Structural vibrations error resulting in: "Inform Vibration Support"

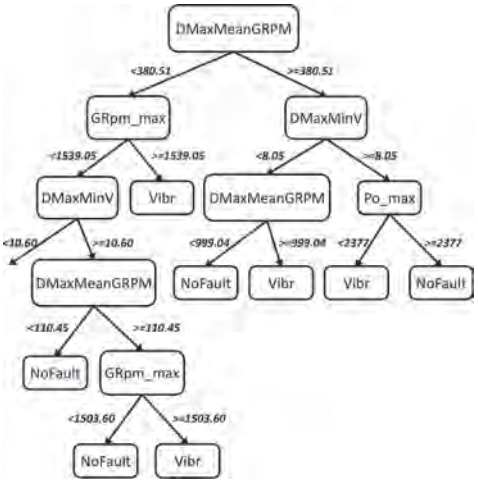


Figure 3. Part of the decision tree classifier based on the SCADA data from 48 wind turbines.

4. Noise Filtering: in diagnostics, we are interested in identifying the root causes that result in a large proportion of the overall faults. Thus we retain the leaf nodes accounting for more than a certain threshold of the total number of faults (e.g.) $threshold = 5\%$)
5. Extract traces (paths) containing the target fault leaf node (C), and all edges (E) and internal nodes (N) up to the root node (R).
6. Extract rules at each internal node of the traces
7. Node Merging: we merge nodes on a path by eliminating ancestor nodes that are logically subsumed by successor nodes.

Below is an example of an automated extraction of a sequential trace of events (root causes) from the trained decision tree classifier (see Figure 3), leading from the classified fault to the root of the tree:

$$Vibr \leftarrow 1503.6 <= GRpm\_mar < 1539.05 \leftarrow 110.455 <= DMaxMeanGRpm < 380.51 \leftarrow DMaxMinV >= 10.65$$

This sequential trace of events indicate that the fault *Vibr* can possibly occur when, over a period of 10 minutes, the maximum generator speed is in the range 1503.6–1549.05 *RPM*, the difference between max and mean generator speed is in the range of 110.45–380.51 *RPM* and the difference between the max and min wind speed exceeds 10.65 *m*/*s*. Note that it is not possible to infer from the data the time interval during which the large change in wind speed occured but it could well be an indication of agust or excessive turbulence over a short period of time resulting in the excessive vibration fault. Another example of an automated extraction of a sequential trace of events (root causes) leading from the classified fault to the root of the tree is as follows:

$$Vibr \leftarrow Po\_max < 2377 \leftarrow DMaxMinV >= 8.05 \leftarrow DMaxMeanGRpm >= 380.51$$

This sequential trace of events indicate that the fault *Vibr* can possibly occur when, over a period of 10 minutes, the maximum electric power output does not exceed, 2377 *kW*, the difference between max and mean generator speed exceeds 380.51 *RPM* and the difference between the max and min wind speed exceeds 8.05 *m*/*s*.

This approach to data-driven root cause analysis elegantly elucidates the traces of events that lead to a fault. These traces can subsequently be used by an engineer to design simulation scenarios to try and replicate the faults and to propose mitigating actions.

4 OUTLOOK: BIG DATA BASED MONITORING AND DIAGNOSTICS FRAMEWORK

So far we demonstrated how offline root cause analysis can be conducted via an ensemble based Bagged decision tree learning from SCADA data of 48 wind turbines based on the CART algorithm. In this section we summarize an outlook for a monitoring and diagnostics framework that would perform on big data and in real-time. The intent is for this framework to be deployed on a cloud such as Azure and scale as the volume of streamed data increases. Figure 5 shows an overview of the architecture of the framework. The main features of the framework include:

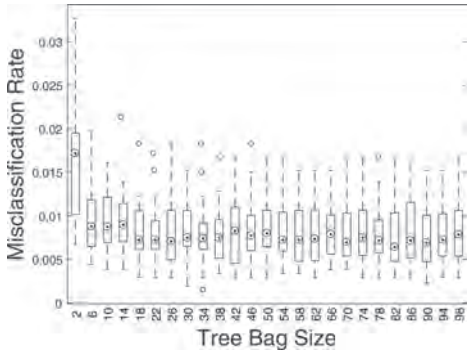


Figure 4. Misclassification rate of the validation set as a function of tree bag size.

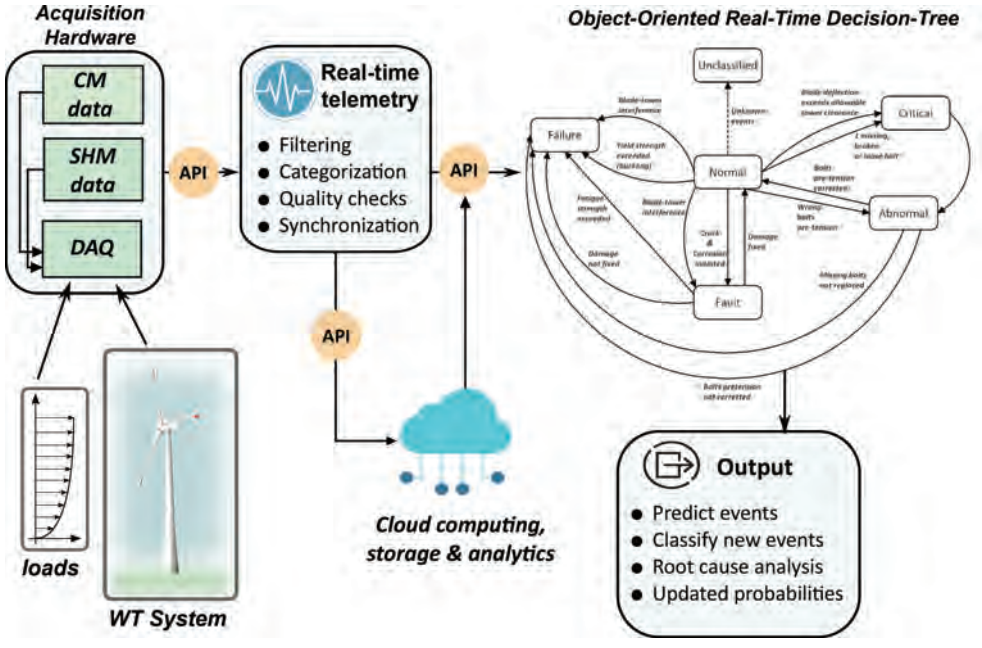


Figure 5. Graphical summary of the proposed framework decision tree learning with big data for fault diagnostics.

- Hardware-software package, able to acquire and fuse in real-time both condition monitoring (CM) (e.g. electric power output, rotor speed) and specialized structural health monitoring (SHM) data (e.g. tower accelerations, strains on blade) from WT components (e.g. blades, gearbox, tower, etc.).
- Computational core: object-oriented decision tree learning algorithm, Bayesian network (BN) based computation of probabilities and root-cause discovery algorithm (as demonstrated in the previous section).
- Distributed cloud based data storage and analytics of the high rate of real-time data streaming from the measuring unit on the WT, which interfaces with the real-time decision-tree software unit. Hence, the toolkit does not need to save data and do heavy computations on remote WT.
- Online user interface to visualize the output from the decision tree (decision support tool).

In the following we elaborate a bit more on: (i) the cloud computing, storage and analytics, and (ii) object-oriented decision tree learning aspects.

i. For efficient fusion of the large bulk of information made available from the condition and structural health monitoring systems constitutes a Big Data problem, as large amounts of data are continuously sampled at high and diverse rates from the WTs within a wind farm. Thus, the integration of a distributed cloud based data storage and analytics is a natural fit with consequent reduction in the cost of handling and manipulating of said data (see Figure 6). In the proposed framework, large scale storage of historical data from WTs is done via the Hadoop Data File System (HDFS). HDFS is a distributed file system that is fault tolerant and can load large volumes of data, in a distributed manner, even if any errors occurs. Telemetry and data acquisition is supported by Apache Kafka which is a distributed streaming platform that aims to provide a unified, high-throughput, low-latency methodology for handling real-time data feeds. Kafka is a fault-tolerant system, meaning that if any error occurs, published data would still be available in case the application or Kafka is stopped, is offline and then restarted. Thus, Kafka can be guaranteed not to miss any data that is issued from the wind turbines. Finally, for data processing purposes, Apache Spark is used. Spark is an in-memory fast and general engine for large-scale data processing. In addition, Spark can scale to thousands of machines in a fault-tolerant manner. These characteristics make Apache Spark very suitable to process and analyze the large volumes of data that are gathered from the wind turbines (Canizo, Onieva, Conde, Charramendieta, & Trujillo 2017).

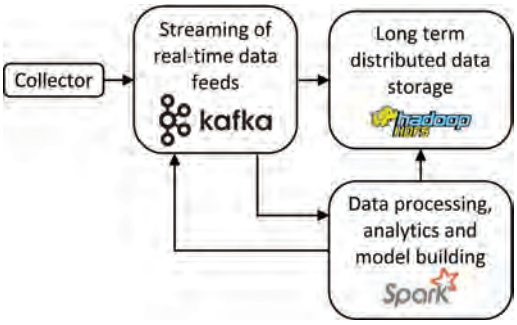


Figure 6. Cloud computing and storage.

ii. A priori Decision Trees are built first to serve the fundamental diagnostic tasks by combining engineering knowledge of the system, failure modes and domain knowledge. The dynamic decision tree classifiers are trained over time from wind turbine telemetry (CM & SHM data). The limitation in traditional decision tree classifiers appears when the component displays a behaviour with feedback (i.e., after a repair/update in the system) or for evolving systems (e.g. when new sensors are integrated or aging of the system), which implies a need to establish several decision trees based on the possible ordering of the events or based on new initiating events. One way around this is an innovation that we introduced in this framework in the form of object-oriented decision tree learning (Wyss & Durán 2001). To this end, a WT is viewed as a multi-layered system of objects (e.g. structure, controller, actuator, etc.) that are defined on the basis of abstract super-classes, attributed with specific properties and methods. Decision trees classifier are further mapped into Bayesian Networks for further assessment of the conditional probabilities (Bearfield & Marsh 2005, Jassens, Wets, Brijs, Van Vanhoof, Arentz, & Timmermans 2006). The integration of the object-oriented decision tree learning and BN delivers updated probability of occurrence associated with each event and end state, which would form a solid indicator of the risk of future failure of any given component.

5 CONCLUSIONS

We presented a review of several decision tree classification algorithms. We then demonstrated how data-driven and automated root cause analysis can be conducted via an ensemble based Bagged decision tree learning from SCADA data of 48 wind turbines based on the CART algorithm. Root cause analysis is here cast in the sense of programmatically discovering the sequence of events (paths) leading from a classified fault at the leaf, all the way to the root of the decision tree classifier. These traces can subsequently be used by an engineer to design simulation scenarios to try and replicate the faults and to propose mitigating actions. Finally, we briefly presented an architecture for a monitoring and diagnostics framework that would perform on big data. In particular we highlighted the need for cloud based storage and computing, and an innovative approach based on object-oriented decision tree learning that extends the traditional decision tree classifier concept. In the future, more concrete implementations and results of the proposed framework will be disseminated.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the European Research Council via the ERC Starting Grant WINDMIL (ERC-2015-StG #679843) on the topic of Smart Monitoring, Inspection and Life-Cycle Assessment of Wind Turbines. We would like to thank Vattenfall AB for providing access to the SCADA data from the Lillgrund offshore wind farm.

REFERENCES

Abdallah, I., V. Dertimanis, & E. Chatzi (2017). Data-driven identification, classification and update of decision tress for monitoring and diagnostics of wind turbines. In *2nd ECCOMAS Thematic Conference, UNCECOMP*, Rhodes Island, Greece.

Bauer, E. & R. Kohavi (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Mach. Learn. 36*, 105–139.

Bearfield, G. & W. Marsh (2005). Generalising event trees using bayesian networks with a case study of train derailment. In *International Conf. on Comp. Saf., Reli., and Sec., SAFECOMP*, Fredrikstad, Norway, pp. 52–66.

Breiman, L. (1996). Bagging predictors. *Mach. Learn. 24,* 123–140.

Breiman, L., J. Friedman, R. Olshen, & C. Stone (1984). *Classification and Regression Trees*. Wadsworth Inc.

Bühlmann, P. (2012). Bagging, boosting and ensemble methods. In J. Gentle and Y. Mori (Eds.), *Handbook of comp. stat*., Chapter 33, pp. 985–1022. Springer.

Canizo, M., E. Onieva, A. Conde, S. Charramendieta, & S. Trujillo (2017). Real-time predictive maintenance for wind turbines using big data frameworks. In *Proc. on IEEE Int. Conf. on Prognostics and Health Management*, Allen, TX, USA, pp. 70–77.

Carrasco Kind, M. & R. Brunner (2013). Tpz: Photometric redshift pdfs and ancillary information by using prediction trees and random forests. *Monthly Notices of the Royal Astronomical Society 432*, 1483–1501.

Caruana, R. & A. Niculescu-Mizil (2006). An empirical comparison of supervised learning algorithms. In *Proc. of the 23rd inter. conf. on Mach. lear.*, Pittsburgh, Pennsylvania, USA, pp. 161–168.

Dietterich, T. (1996). Applying the weak learning framework to understand and improve c4.5. In *Proc. 13th International Conference on Machine Learning*, pp. 96–104.

Freund, Y. & R. Schapire (1996). Experiments with a new boosting algorithm. In *Proc. 13th International Conference on Machine Learning*, pp. 148–156.

Gehrke, J., R. Ramakrishnan, & V. Ganti (2000). Rainforest —A framework for fast decision tree construction of large datasets. *Data Min. Knowl. Discov. 4*(2/3), 127–162.

Gehrke, J., V. Ganti, R. Ramakrishnany, & W. Lohz (1999). BOAToptimistic decision tree construction. In *Proc. of the 1999 ACM SIGMOD intern. conf. on Manag. of data*, Stockholm, Sweden, pp. 169–180.

Grasse, F., V. Trappe, S. Thoens, & S. Said (2011). Structural health monitoring of wind turbine blades by strain measurement and vibration analysis. In *Proc. of the 8th International Conference on Struct. Dyn.(EURODYN)*, Leuven, Belgium.

Hameed, Z., Y.S. Hong, S.H. Ahn, & C.K. Song (2009). Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renewable and Sustainable Energy Reviews 13*, 1–39.

Hand, D. (1997). *Construction and Assessment of Classification Rules*. Chichester: John Wiley and Sons.

Jassens, D., G. Wets, T. Brijs, K. Vanhoof, T. Arentz, & H. Timmermans (2006). Integrating Bayesian Networks and decision trees in a sequential rule-based transportation model. *Euro. J. Oper. Res. 1*, 16–34.

Kim, H. & W. Loh (2001). Classification trees with unbiased multiway splits. *J. Amer. Statist. Assoc. 96*, 598–604.

Lim, T., W. Loh, & Y. Shih (1997). An empirical comparison of decision trees and other classification methods. Technical Report TR 979, Department of Statistics, UW Madison.

Lim, T., W. Loh, & Y. Shih (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning Journal. 4*, 203–228.

Loh, W. & Y. Shih (1997). Split selection methods for classification trees. *Statistica Sinica 7*, 815–840.

Loh, W.Y. (2011). Classification and regression trees. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 1*, 14–23.

Opitz, D. & R. Maclin (1999). The strength of weak learn ability. *Popular ensemble methods: an empirical study 1*, 169–198.

Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning 1*, 81–106.

Quinlan, J.R. (1994). *Programs for Machine Learning.* Morgan Kaufmann Publishers.

Schapire, R. (1990). The strength of weak learn ability. *Mach. Learn. 5*, 197–227.

Singh, K. & R. Sulekh (2017). The comparison of various decision tree algorithms for data analysis. *International Journal Of Engineering And Computer Science 6*, 21557–21562.

Solé, M., V. Muntés-Mulero, A.I. Rana, & G. Estrada (2017). Survey on Models and Techniques for Root-Cause Analysis. *ArXiv e-prints.*

Wyss, G.D. & F.A. Duŕan (2001). OBEST: The object based event scenario tree methodology. Technical Report SAND2001-0828, Sandia National Laboratories, Albuquerque, CA, USA.

Yang, B. & D. Sun (2013). Testing, inspecting and monitoring technologies for wind turbine blades: A survey. *Renewable and Sustainable Energy Reviews 22*, 515–526.

Zheng, A.X., J. Lloyd, & E. Brewer (2004). Failure diagnosis using decision trees. In *Proc. of the 1st Intern. Conf. on Autonomic Comp., ICAC*, Washington, DC, USA, pp. 36–43.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Cyber physical systems implementation for asset management improvement: A framework for the transition

L. Villar-Fidalgo
Universidad Nacional de Educación a Distancia (UNED), Spain

A. Crespo Márquez, V. González Prida, A. De la Fuente, P. Martínez-Galán & A. Guillén
Universidad de Sevilla, Spain

ABSTRACT: The transformation of the industry due to recent technologies introduction is an evolving process whose engines are competitiveness and sustainability, understood in its broadest sense (environmental, economic and social). This process is facing, due to the current state of scientific and technological development, a new challenge yet even more important: the transition from discrete technological solutions that respond to isolated problems, to a global conception where the assets, plant, processes and engineering systems are conceived, designed and operated as an integrated complex unit. This vision is evolving besides a set of concepts that are, in some way, to guide this development: Smart Factories, Cyber-Physical Systems, Factory of the Future or Industry 4.0, are examples. The full integration of the operation and maintenance (O&M) processes in the production systems is a key topic within this new paradigm. Not only that, this evolution necessarily results in the emergence of new processes and needs of O&M, i.e. also, the O&M will undergo a profound transformation. The transition from actual isolated production assets to such Industry 4.0 with CPS is far from easy. This document presents a proposal to develop such transition adapting one iteration of the Model of Maintenance Management (MMM) integrated into ISO 55000 to the complexity of incorporating "System of Systems" CPSs maintenance. It involves several stages: identification, prioritization, risk management, planning, scheduling, execution, control, and improvement supported by system engineering techniques and agile/concurrent project management.

1 INTRODUCTION

Cyber-Physical Systems (CPS) is a widespread concept with several meanings, usually linked with embedded and connected systems.

The term cyber-physical systems was coined by Helen Gill in 2006 at the National Science Foundation in the U.S. to refer to the integration of computation with physical processes.

One of the earliest and referred papers describes them as "… are integrations of computation with physical processes. Embedded computers and networks monitor and control the physical processes, usually with feedback loops where physical processes affect computations and vice versa." (E.A. Lee 2008). In the early 2010's was defined as "transformative technologies for managing interconnected systems between its physical assets and computational capabilities" (Baheti & Gill 2011). It was a wider definition and avoids any technical description. The most recent release of Framework for CPS released by American National Institute of Standards and Technology (NIST CPS Public Working Group 2017) opens even more the definition: "... are smart systems that include engineered interacting networks of physical and computational components". But simultaneously the technical description needs a complete section –2.1 – of the document, and details in other section –1.1.2 – thirteen main differences with conventional product, system, and application design. This process of increasing the technical description extension while broaden the scope by shrinking the term definition and using vague words, reveals the complexity of these systems.

Despite this complexity, Industry will have to deal with the integration of CPS into their asset portfolio and their maintenance framework. It is a straightforward way to improve their performance in the global and competitive market that they face nowadays. The disruptive aspect of CPS in Operation & Maintenance of assets are two. The first one is their ability to share information & self-compare their behavior in an autonomous way as a community of equipment. For example: a network of CPS pumps will be able to predict the failure of one of their members based on shared information, without the intervention of higher level supervisors. The prognostics and health management has an open road ahead. The second disruptive aspect is

their ability to identify a misuse or improper use by human operators: two conveyor belts could compare themselves and rise a warning if one of them is overloaded while the other is underutilized. Unfortunate ly for maintenance personnel, most of their interventions will be over legacy and operating Systems of Systems interacting with human beings: plant operators, maintenance technicians, etcetera. Their intervention will have also to deal with the seven samurais (Martin 2004) systems: context, intervention, realization, deployed, collaborating, sustainment and competing.

This paper will propose a framework to introduce CPS assets and their philosophy for Operations & Maintenance in those real environments.

2 METHOD

Assuming a company that follows the Maintenance Management Model (Márquez 2007) as a framework for maintenance that fulfills the ISO 55000standard for asset management (Crespo & Parra 2018). The starting point for CPS integration in the asset portfolio should appear as a decision of phase 8 of an iteration (Continuous improvement and new techniques utilization): The company decision is to adopt the industry 4.0 concepts. In doing so it will implement Cyber-Physical Systems (CPS) as future assets and uplift the existent ones to this concept. Therefore, a new iteration in the complete Maintenance Management Model is proposed[1]. This iteration will also follow the Cyber Physical Framework. The actual target is to achieve the Function III (Cyber level) of the 5C level architecture for implementation of CPS (J. Lee, Bagheri, & Kao 2015), as a previous step to the full completion (Configuration Level) in future iterations.

It will be analyzed in the next paragraphs the eight phases of Maintenance Management Model as decision areas to implement our strategy regarding the inclusion of CPS in our asset portfolio. There will not be further references to "business as usual" activities of above mentioned framework related to maintenance process of the company.

3 PHASE 1: DEFINITION OF THE MAINTENANCE OBJECTIVES AND KPI'S

Along this phase there is a conceptualization facet: the main effort is to obtain the model that will satisfy our requirements under the distinct aspects of CPS framework: Functional, Business, Human, Trustworthiness, Timing, Data, Boundaries, Composition and Lifecycle. It is one of the phases deeply impacted by CPS implementation. During this phase, the decisions taken will determine the level I of the 5C architecture: The Smart Connection Level function with Plug & play, tether-free communication and sensor network as attributes.

Basic tool for this phase is Balanced Scorecard integrating not only economic performance and technical indicators of operation and maintenance but also project execution and Human factors of iteration to integrate CPS in the assets portfolio.

3.1 *Objectives to add*

1. CPS implementation policy. This policy should consider the importance of achieving tangible objectives of CPS as soon as possible, to engage the main stakeholders. So, assets to upgrade/ substitute must be carefully chosen; on later iterations under MMM it can be implemented the fully transformation of portfolio.
2. Maintainability, risk reduction, reliability and availability improvement to achieve by new/ uplifted CPS assets. Obviously, these parameters should be increased above the average.
3. Economic impact of overall operation. Capital expenditure should be considered carefully: In the future, the balance between investment and return obtained will be scrutinized for such technology change and can jeopardize further deployment.
4. Domains of CPS framework implementation should be identified; these are the areas of deployment in which stakeholders may have domain-specific (manufacturing) and cross-domain concerns (Energy, transportation). Groups of conceptually equivalent or related concerns will became the Aspects of the CPS framework of our interest.

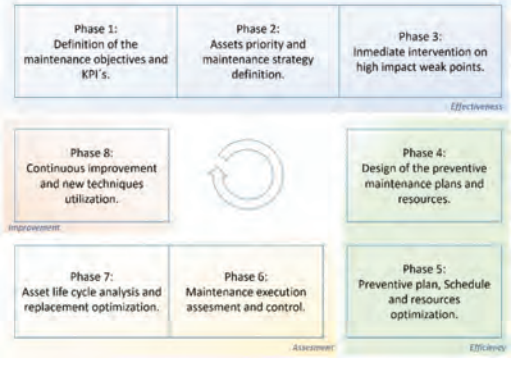


Figure 1. MMM schema.

[1]Only the new activities due to CPS are mentioned in the paper. The *business as usual* activities of the cycle are not included: refer to the original

Level	Function	Techniques	Attributes	System of systems action level
L5	**Configuration**	Resilient Control System (RCS):Action to avoid	• Self configure for resilience • Self adjust for variation • Self optimize for disturbance	DECISION
L4	**Cognition**	Decision support system (DSS): Prioritize and Optimize Decisions	• Integrated simulation and synthesis • Remote visualization for human • Collaborative diagnostics and decision making	
L3	**Cyber Physical Systems** (CPS)	Self-Compare	• Twin model for components and machines • Time machine for variation identification and memory • Clustering for similarity in data mining	DATA COLLECT
L2	**Data to information conversion**	Prognosis and Health Management: Self-Aware	• Smart analytics for component machine health. • Multi dimensional data correlation • Degradation and performance prediction	
L1	**Smart connection**	Condition Based Maintenance: Condition Monitoring	• Plug & play • Tether-free communication • Sensor network	

Figure 2. 5C level architecture (Lee et al.).

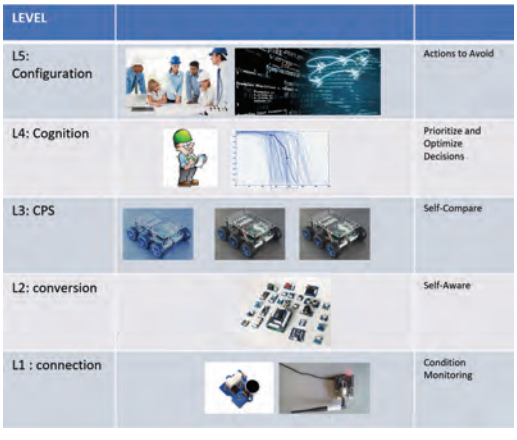


Figure 3. 5c architecture overview examples (Lee et al.).

3.2 *Identify the critical questions to be answered and key decisions to be taken*

1. New Stakeholders identification and documentation of their expectations. Consultants and technology providers are new stakeholders, but it should be also reexamined the old ones with new concerns: maintenance personnel and operators should change their approach to these systems improving their technical skills in software/hardware; it will imply new training. Report processes, command chain and IT systems will be also impacted.
2. Legal & safety concerns: Due to the recent development of CPS, there is a lack of regulation, but it will have a deep impact in the future for safety or environmental risky systems. The legislation will affect them and their crypto conditions to avoid any unintended access. The CPS framework has a full detailed coverage of this concerns and the assurance facet is described in detail.
3. Technological state of the art. Due to the lack of maturity, it must be evaluated which technologies will survive in a few years. Open standards will help to survive or, at least, to ease the transition to new ones in the future.
4. Identify leaders (internal and external) for new know-how transference and acquisition.
5. Other Standard, Policies, Directives and Procedures to be adapted. It has been mentioned the technological ones, but other business domains should be also considered. As an example: CPS themselves could inform of earned value to the project management software (units produced, finish of testing phase, finish of startup process…)
6. Cyber-level infrastructure and Machine-cyber interface.
7. ERP and Enterprise Asset Management software data flow impact.
8. Big data analysis systems and AI deployment.
9. Training of personnel.
10. Infrastructures, Special tools and test equipment needed or affected.

IT infrastructures play a key role in CPS systems. So, critical questions must be faced in this

stage as a decision-making breakdown structure for IT to be implemented in the CPS:

In the software level, the continuous evolution of information technologies is a risk factor and the use of open and standard methodologies for communications, like REST APIs with HATEOAS to exchange data, based on JSON or XML should be a priority. A wrong step in this direction could endanger the rest of the project by making CPS unable to communicate among them.

Regarding physical layers of communications (wired or wireless) it should be followed the same principle: looking for standard and open technologies. Of course, wired TCP/IP should be there, but wireless technologies are here to stay, and it is far more confusing the election:

Low-rate wireless personal area networks under IEEE 802.15.x or WIFI could be a reasonable decision for locally deployed systems.

For wide areas, it could be chosen between licensed LPWA technologies standardized by 3GPP (EC-GSM-IoT, LTE MTC Cat M1 and NB-IoT) or other commercial LPWA solutions: LoRAWAN or SigFox. These are closed technologies.

The future would pass through commercial 5G bands categorized into three generic services, namely, extreme mobile broadband, massive machine-type communications, and ultra-reliable machine-type communications.

3.3 *KPIs*

1. Human factors: strong leadership is needed, to overcome resistance and barriers, to change mindsets, to push through organizational change, to sustain investment, and to keep the team involved, specifically during the transition. Indicators over these soft factors will help to measure the pulse of the organization. It is important because the benefits of the CPS will appear after the integration of some of them: at the beginning of the process, only problems will arrive without any apparent benefit.
2. Speed of update in infrastructures, assets, etcetera: transition should be implemented fast enough to achieve evident benefits during the first stages, but avoiding over stress the organization (shutdown and start up production, new training, new processes, etcetera)
3. AI transition and assets peer to peer comparison. As one of the theoretically most disruptive changes that brings CPS, we should monitor the efficiency of this behavior through the economic impact in our organization. Installing only a modern gadget will fail as objective.
4. Earned Value management indicators (EV, CPI, SPI) are a good reference for any project to deploy CPS. It should also use the American Defense Contract Management Agency 14 points of Baseline Execution Index for planning and schedule. These are objective indicators of project's financial status, as one of the main concerns of the enterprise.

3.4 *Audits*

The use of audits will have two variants: the ones for control and continuous improvement according to MMM (MES, QMEM, etcetera) and specific ones to check the efficiency and effectivity of CPS implementation. These audits should focus on the incremental evolution of CPS upgrade: The added value of firsts units implemented should be audited against their objectives.

4 PHASE 2: ASSETS PRIORITY AND MAINTENANCE STRATEGY DEFINITION

Basic tool is Criticality Analysis, upgraded to include the risk of project—failure during implementation of CPS: It is needed to prioritize those assets with higher Return on Investment and lower technical risks. During the first iteration one of the highest risks is disaffection of CPS by main stakeholders. So, the asset to be upgraded at this early stage should be with high improvement-visibility, affordable capital expenditures effort, lower technological risk, easy to O&M in the near future.

4.1 *Determination of actions on actual assets based on risk factor analysis to:*

1. Replace/substitute for a new asset.
2. Uplift and improve the existent ones.

It will be scored using the concerns of CPS-FWK tailored accordingly. Without been exhaustive, in this first iteration it should be paid special attention to several concerns enumerated in the framework:

1. Functional: monitorability and communication.
2. Business: cost, time to market, utility and interoperability.
3. Timing: awareness and resilient time.
4. Boundaries: networkability, responsibility.
5. Composition: adaptability, constructivity and discoverability.
6. Lifecycle: procureability, deployability and mantenability.
7. Human: usability.
8. Trustworthiness: safety, reliability and cybersecurity.

These are, under author considerations, the key aspects to assure stakeholders engagement during the first iteration.

4.2 *Planning and scheduling to retrofit the assets. Based on the CPS 5C level architecture, with target of cyber level*

It will be used turnaround, shutdown and Outage operations to implement the new CPS assets or transform the legacy ones in CPS systems.

The authors recommend using Kanban agile practices embedded in a general schedule (Villar-Fidalgo, Espinosa-Escudero & Domínguez-Somonte 2016 & 2017), according to PMI® Agile Practice Guide (2017). The "product owner" value team might produce a roadmap to show the anticipated sequence of CPS to add or legacy systems to upgrade over the time. This planning will include conceptual, preliminary and detailed design, development, construction, M&O and disposal phases. This team re-plans the roadmap based on what the results are.

Due to the complexity of the CPS concept and lack of maturity, incremental and iterative changes during several iterations are preferable than a complete transformation in a single endeavor with waterfall planning and scheduling. Besides the intrinsic resilience and flexibility of a well-designed CPS will benefit this approach.

Agile-lean Kanban Method fits perfectly in particular steps of the proposal: Starts with current state, it is incremental, respects the current process, roles, responsibilities and titles. It will help dealing with overcoming requirements

The main adversary is that enterprise culture not always embrace leadership attitude at all levels.

4.3 *EV baseline evaluation*

This baseline will help to measure the performance of transition. The variances over baseline estimations will identify the need of upgrade the plan or maintain the initial target. Deviations off the baseline will not only affect the financial performance of the transition project but also stakeholder's engagement with CPS vision.

5 PHASE 3: IMMEDIATE INTERVENTION ON HIGH IMPACT WEAK POINTS

Basic tool is Failure Root Cause Analysis (FRCA), to look for high impact reliability enhancements and ensure a very effective definition of subsequent maintenance plan activities. Adding a new cause of failure is necessary, as it is recognized in CPS framework aspect of trustworthiness, based on security, privacy, safety, reliability and resilience of future CPS. The high complexity of network based software is a new factor to consider in the FRCA analysis over traditional physical, human and latent causes inherent to any kind of systems.

Nevertheless, it should be also intervened in high impact favorable assets to CPS transformation: those items which could lead to enhance the visibility of CPS advantages. A key reactor of high economical value deems for sure to be upgraded to CPS level, but their actual monitorization as a single asset with conventional SCADA and PLCs plugged to the network will dilute the improvement of such upgrade. On the other hand, a set of water flow pumps will not be as spectacular as the reactor, but their number and failure rate could make them ideal candidates to be upgraded if it is possible to achieve some good failure prognostics based on their shared data, including condition based on sensors and usage rate.

5.1 *Retrofit of highest priority assets to CPS level in first incremental iterations of schedule*

The priority will be based on:

1. Affordability of implementation and economic benefit for asset exploitation.

   Feasibility of full CPS application.
   Trustworthiness of implementation.
   Visibility of CPS enhancements.

5.2 *Data capture and data mining to extract the information, probably under "big data" considerations*

The first data extracted and initial behavior of CPS itself should follow the planned strategy. Otherwise will be necessary to adapt, preferably through agile methodologies, the next steps.

It should be achieved the maximum net benefits per system replaced/upgraded during the earliest stages of implementation, because of the priority criteria used. If the results are not the expected, it is an alarm signal that should draw attention to strategy or, at least, re-plan the implantation priorities. During this phase, it should be achieved the 5C architecture function 2: Data—to information Conversion Level. The attributes are smart analytics for component machine health with multi-dimensional data correlation and finally the degradation and performance prediction.

6 PHASE 4: DESIGN OF THE PREVENTIVE MAINTENANCE PLANS AND RESOURCES

Basic tool is Reliability Centered Maintenance: where operations and maintenance start to be influenced by CPS, depending on the level assigned in the risk plan and within the operational mode. It is far from the function 5 of 5C architecture: Configuration Level, with attributes of self-X of the

CPS, but it has information that, supported by new plans and their optimization, will take it to desired Function III of 5C architecture: the cyber level.

After implementation, the maintenance plan should be oriented to Condition Based Maintenance/Prognostic Health Management to exploit all the advantages of CPS. The CPS will probably rise overcoming requirements due to higher precision in sensors and data analysis (big data, M2M peer review), that will fine tune the detection of catastrophic failures. But this fine tune detection will increase the need of fast actuation: it will not only detect the normal worn of a friction bushing but the fast failure of an axle due to unexpected fatigue because of micro cracks (that CPS could infer due to misuse of the asset). The maintenance plans should also embrace the Agile philosophy.

7 PHASE 5: PREVENTIVE PLAN, SCHEDULES AND RESOURCES OPTIMIZATION

Basic tool is Risk-Cost Optimization considering the new opportunities that CPS brings. Use of self-compare behavior and machine to machine data exchange (peer monitoring) to identify the state of the asset portfolio and accuracy prognosis will allow to take advantage of dynamic planning and scheduling forecasting with optimal allocation of resources for maintenance and operation. The capacity of CPS to identify misuse or workloads unbalanced will also affect the production plans or operators training plans.

Again, the agile methodologies imbricated in high-level integrated master plan is a must: it should be ready to deal with overcoming requirements that need to be addressed on line. Baseline updates with a transition from original contour conditions (or "samurais") to updated ones will allow to measure performance without losing contact with the new reality. This way audits will continue delivering value.

8 PHASE 6: MAINTENANCE EXECUTION ASSESSMENT AND CONTROL

Basic tool is Operational Reliability Analysis, but it should not be forgotten specific aspects of transition phase: under a project environment with high-risk technological transition.

All the KPIs defined in phase 1 will help to control the evolution of the project and the impact in O&M operations:

1. Measurement, analysis and evaluation of earned value indicators for asset retrofit and start-up: cost and schedule through CPI, SPI, S-curves and Integrated Program Management Report.
2. Evaluation of indicators of actual performance in reliability, maintainability and future improvement based not only on probabilistic assessments, but also on truthful information and prognostics delivered by new CPS.
3. Evaluation of Risk with cost/benefit evaluation of mitigation plans.
4. CPS transition-speed and acceptance among stakeholders.

Finally, it will be needed to control the improvement achieved with CPS already installed. If we are not able to follow the baseline technical-plan, it should be reconsidered the overall project.

9 PHASE 7: ASSET LIFE CYCLE ANALYSIS AND REPLACEMENT OPTIMIZATION

This phase, as the first one, is deeply impacted by CPS. Basic tool is Life Cycle Cost Analysis, but it must include all the facets that new systems will bring to the asset portfolio.

It will face new cost-categories: software upgrades, trustworthiness analysis and developments, data networks deployment and maintenance, etcetera. Nevertheless, it can also achieve important savings: better use of systems by operators, higher accuracy in prognosis maintenance, lower supervision costs due to the "self-awareness" of CPS groups. Unfortunately, this kind of improvement will become very often incomputable because the traditional life cycle analysis has not included these factors. The maintenance team will have to struggle against skepticism to show the advantages of the CPS. This is one of the reasons to make key decisions and answer critical questions during the first phase: once the iteration is at this point, it should have evidences from CPS deployed to support the benefits to Life Cycle Cost, otherwise there will be only important capital expenditures and intangible benefits that could lead to disaffection of critical stakeholders like CFO's.

Another key aspect to highlight is the improvement in risk management and "probability/risk number" of CPS systems, due to a better knowledge of their health as a system.

10 PHASE 8: CONTINUOUS IMPROVEMENT AND NEW TECHNIQUES UTILIZATION

It will be necessary to analyze the targets from phase 1 achieved in the concluded iteration. If the result is satisfactory there will have two roads ahead: spread the CPS architecture to more assets or take another step with actual CPS towards

Function IV of CPS architecture: The Cognition Level. Probably the wisest decision in these times of technological immaturity is spread the CPS concept and study lessons learnt during the first iteration to smooth the continuous uplift / renovation of our asset portfolio.

The Function IV (Cognition Level) and V (Configuration Level) of 5C architecture have an intrinsic high degree of technological uncertainty that make them too risky under the engagement of main stakeholders. They should go through a System Integration Laboratory or prototype phase before the full integration in production assets.

11 DISCUSION

The architecture developed by Lee et al. (J. Lee et al. 2015) or the Framework released by NIST (NIST CPS Public Working Group 2017) are clear starting points for deployment of CPS in manufacturing industries. Nevertheless, these documents are focused on the CPS itself.

Our approach is a complementary and holistic view of the CPS implementation in a dynamic environment like the active enterprise. The Human Factor is a cornerstone during business transformation and should be included in the equation.

Another intended contribution is the necessary link between the sequential workflow of construction of a CPS with the iteration phases of the maintenance model and business itself.

12 CONCLUSIONS AND FURTHER DEVELOPMENT

This paper presents a framework to incorporate CPS, up to Function 3 of 5C architecture, in our asset portfolio with a holistic view and under a consolidated maintenance management model. The objective is to avoid early failures that discourage stakeholders from supporting this technology after the first iteration.

This work will continue with use cases evaluation and further iterations to achieve the Function V (Configuration Level) where all the advantages of CPS could be exploited in benefit of maintenance of full assets portfolio, and therefore the business objectives.

REFERENCES

*Agile Practice Guide.*2017. Global Standard ed. 14 Campus Boulevard. Newton Square, Penssylvania 19073–3299 USA: Project Management Institute.

Baheti, Radhakisan and Helen Gill. 2011. "Cyber-Physical Systems." *The Impact of Control Technology* 12: 161–166.

Crespo Márquez, Adolfo and C. Parra Márquez. 2018. "On the Family of Standards UNE-ISO 55000 and how to Effectively Manage Assets." In *Advanced Maintenance Modelling for Asset Management*, edited by A.Crespo Márquez et al., 1–16: Springer International Publishing AG.

Lee, Edward A. 2008. "Cyber Physical Systems: Design Challenges.", Object oriented real-time distributed computing (isorc), 2008 11th IEEE international symposium on.

Lee, Jay, Behrad Bagheri, and Hung-An Kao. 2015. "A Cyber-Physical Systems Architecture for Industry 4.0-Based Manufacturing Systems." *Manufacturing Letters* 3 (Supplement C): 18–23. doi:https://doi.org/10.1016/j.mfglet.2014.12.001.

Martin, James N. 2004. "The Seven Samurai of Systems Engineering: Dealing with the Complexity of 7 Interrelated Systems. Proceedings of 14th Annual Symposium of the International Council on Systems Engineering." Toulouse, France.

Márquez, Adolfo Crespo. 2007. *The Maintenance Management Framework: Models and Methods for Complex Systems Maintenance* Springer Science & Business Media.

NIST CPS Public Working Group. "CPS PWG Cyber-Physical Systems (CPS) Framework Release 1.0.", https://pages.nist.gov/cpspwg/.

Villar-Fidalgo, Luis, M. Mar Espinosa-Escudero, and Manuel Domínguez-Somonte. 2016. "Cronogramas Para Toma De Decisiones Ágiles En Entornos Concurrentes Con Incertidumbre." *Dyna Management* Vol 2017.

Villar-Fidalgo, Luis, Manuel Domínguez Somonte, and María Espinosa Escudero. 2017. "La Gestión Ágil Y Concurrente De Proyectos Con Incertidumbre." *Dyna Ingenieria E Industria* 92: 16–17.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Automated train driver competency performance indicators using real train driving data

R.A.H. El Rashidy, P. Hughes, M. Figueres-Esteban & C. van Gulijk
Institute of Railway Research, University of Huddersfield, Huddersfield, UK

ABSTRACT: On Train Data Recorders (OTDR) are used within the GB Railways to collect data relating to train operations and the state of various train systems throughout a journey. These data include power and brake controller position and driver acknowledgement of signaling system warnings. This data could be used to assess driver competency but an assessment framework is required to extract the data sensibly. This paper proposes a train driver competency framework to define aspects that are related to train driver functions based on documents analysis, cab-rides and informal interviews. It also explores the utilization of OTDR in the quantification of the train driver competency framework by introducing a number of indicators under each aspect covered by the framework. The proposed indicators demonstrate to how OTDR data can be useful in routine systematic checks and pre-incident investigation, for example, identification of the deviation from recommended rules that may have safety implications. Furthermore, the data may allow for improved understanding of driver performance that in turn could allow the development of more effective safety management strategies. A number of numerical example presented to illustrate applicability of developed algorithm.

1 INTRODUCTION

On train data recorder (OTDR) offers an opportunity to understand the driver use of power, brake and safety systems during a journey. Despite the potential of OTDR data, it is not widely used to facilitate the automatic analysis of driver performance.

This paper presents a train driver competency framework based on official documents and reflecting the professional driving policies. It will also explore the use of OTDR data to quantify different areas covered by the proposed train driver competency framework to assess train driver performance aspects such as drivers' use of safety systems and braking, in addition, to derive indicators to measure the vigilance level of a driver.

2 AUTOMATED SAFETY ASSESSMENT

In the UK, current practice for assessing driver competence performance is in-cab riding by driver managers. A number of train operating companies use in-cab assessment to monitor drivers' operational usage of Driver's Reminder Appliance (DRA) (McCorquodale et al., 2002). In some cases, digital cameras implemented to record driver's action but they tend to be unpopular (RSSB, 2004). These techniques have their merits in assessing the driver performance as they supplied comprehensive details about the driver's behavior. However, drivers may behave differently under observation, limiting the potential for independent driver assessment. Add to that, the time and cost traditional methods hinder their use for continuous monitoring.

A number of research studies (Balfe, 2016; El Rashidy & Van Gulijk, 2016; Walker & Strathie, 2014; Green et al., 2011) explored the advantage of using OTDR source in different areas such as station duties, driver assessment and interaction with warning systems. Green et al. (2011) introduced a number of indicators to assess driver performance. They are:

- The speed at which power Notch 4 (out of a total of 4 notches) is selected when accelerating;
- The percentage of time in a braking sequence that the driver selects brake step 3 (out of a total of 4 steps);
- The of the train as it traverses a Train Protection and Warning System grid (TPWS) approaching a Permanent Speed Restriction (PSR);
- The speed through a PSR as a percentage of the maximum speed and the mean speed when the warning system (AWS) horn is received.
- Erroneous events such as wrong-side door release and system trips such as TPWS brake demand.

These indicators are compared with the average performance of the whole population of train drivers

to assess an individual's driving performance in relation to the cohort of drivers. The study introduced initial learning OTDR analysis but did not make use of all the available OTDR that related to the driver performance.

The aim of this paper is similar to the papers above: to assess driver performance in relation to safe driving of a train but we propose a more comprehensive framework.

3 METHOD

The method proposed in this work comprises of two steps, *viz.* developing a train driver competency framework and introducing a number of performance indicators to quantify elements of the framework using OTDR data.

3.1 *Train driver competency framework*

The framework is based on documents analysis, cab-rides and interviews.

Document analysis clarified the driver function and best practices in relation safe professional driving. For example, the professional driving policy (e.g. SWT, 2012; LM, 2009) was used to identify the recommended travel speed when approaching a red aspect and braking rules. The Rule Book – Train Driver Manual (GE/RM8000/ train driver) was also used to identify rules that the driver should comply with such as the use of safety systems.

To gain more knowledge about the driver environment and driver reaction under different situations, cab-rides were carried out. In addition, consultations, in the form of informal interviews, with a driver and a driver manager were conducted to discuss some operational issues and clarify some technical points.

Based on above processes the following aspects were identified:

- The driver handling of trains;
- The driver's compliance with rules;
- The driver's vigilance;

Under each aspect, a number of indicators were introduced, as presented in Figure 1, to facilitate the conversion of the conceptual framework to operational indicators that can be used to assess each aspect. Train handling aspect covers how the driver uses the brake system and the power system whereas compliance deals with rules in relation to driver's handling safety systems. The vigilance level of the driver has been assessed by a number of TPWS brake demands, wrong-side door release and percentage of instant cancellations of AWS horn.

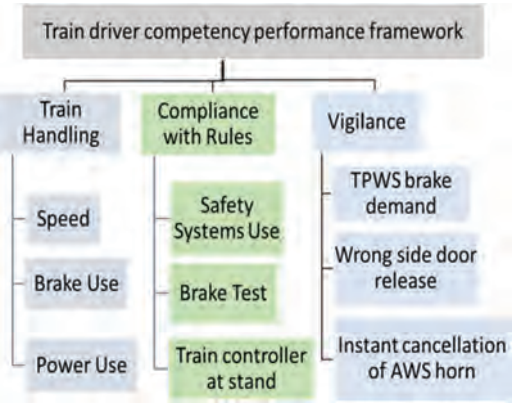


Figure 1. Train driver competency framework.

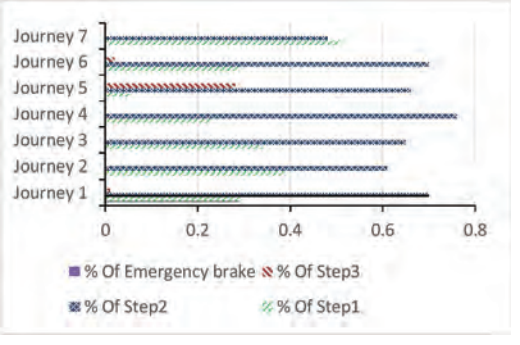


Figure 2. Brake use for a pair of origin-destination pair.

3.2 *Train driver performance indicators*

A bottom-up approach was implemented to develop train driver competency performance indicators (DCPIs) based on driver competency framework using OTDR data.

3.2.1 *OTDR data*

The OTDR data files used in this paper were supplied by Southern Railway. They are for the same route and the same day but different drivers to eliminate the impact of route conditions.

3.2.2 *Initial data handling*

The initial data handling process, presented in Figure 2, includes a number of steps as follows:

- Examine data types and format and correct them if needed. For example, the format of a relative journey time is converted from "+ 01h24 mn26 s 6" to "5066.6" seconds.
- Compress all variables that occurred at the same time in a single data row. Closer inspection of the

data showed a relative journey time record may appear more than once with different groups of variables (i.e. for the same time record, there was more than one input line from different data channels).
- Processing missing data using different logical processes, for example, filling the missing values of train distance with calculated distance based on the available time duration and train speed. This error checking is specific to the Class 455 data used in this study, although it is likely that all OTDR data will need similar error handling and cleaning routines to make it useable.

After this pre-processing, data analysis could commence.

3.2.3 *Detection algorithms*

A number of algorithms have been developed in the R software package to extract relevant indicators for each behavior aspect presented in Figure 1.

3.2.4 *Data analysis*

A number of algorithms have been developed in R to identify relevant scenarios for each behavior aspect presented in Figure 1 and, then, calculate the metric. The metrics are mostly the frequency, the averaging or maximum or minimum values. A number of standard statistical visualizations were used to present the results such as boxplots.

4 RESULTS

4.1 *Proposed DCPIs*

Table 1 a, b and c summarize DCPIs that developed in this work and shows the proposed performance metric and criteria for each indicator assessment.

4.2 *Safety related DCPIs examples*

This section gives a few numerical examples of DCPIs that can be related to safety rules or devices to illustrate the use of OTDR rather than detecting any trend or best practice rule due to the size of the used data sample.

Under the train handling aspect, braking behavior and speed at AWS horn approaching a red aspect are presented here as they are directly related to safety. For braking behavior, Figure 2 shows the percentage of each brake step use calculated by considering the travelled distance using each step. For example, in "Journey 5" the driver applied Step 2 (0.66) in addition to using Step 3 (0.28) as shown in Figure 2 due to the late use of the brake which may create a hazard condition under different circumstance such as low adhesive condition. It should be noted that use of brake Step 3 should be

Table 1. DCPIs based on OTDR data.

(a) Train handling

Aspect	Metric
Braking behavior.	Pattern recognition based on braking curve data
Use of brake-step 3 on approach to stations.	The maximum percentage of distance travelled using brake-step 3 per station during a journey.
Use of brake on approach to stations.	The percentage distance travelled using each brake-step.
Speed	The speed at AWS horn (mph) prior to a red aspect
	The frequency of train speed < = 3 mph when power Notch 4 selected during the journey
Use of power.	The percentage of distance travelled using power notches 1 to 4 (out of 4).

*AWS stand for Automatic Warning System.

(b) Compliance with rules

Aspect	Metric
Use of EBS*	EBS operated event,
Use of TPWS*	TPWS isolated events,
Use of DRA* in front of a red aspect.	The number of DRA operated event comparing with, a number of red aspects the driver experienced.
Use of DRA at the start of a journey	
Use of DRA during the coupling/ uncoupling activity	
Putting the brake controller into Step 3 once the train is at Stand.	Number of Step 3 at stand compared with number of station and red aspect during a journey
Brake test before the first station and the first caution aspect.	The use of brake prior to the first station or an AWS horn.

*EBS, TPWS and DRA stand for Emergency Bypass Switch, Train Protection and Warning System, and Driver's Reminder Appliance, respectively.

(c) Vigilance

Aspect	Metric
Instant cancellation of AWS horn	The percentage of instant AWS cancellation
Wrong side door release	Number of wrong side door release
TPWS Demand application	Number of TPWS Demand application

minimized, as a good practice unless the driver had to use it due to low track adhesion.

For the speed, Table 2 presents the speed at AWS horn when the driver approaching a red aspect. A higher than normal speed when approaching a red aspect may cause a SPAD (signal passed at dangerous without authorization) or lead to a full brake application to stop the train at the correct location. For one train operator this is 20 meters in advance of the red aspect (LM, 2009). For the OTDR sample used in this paper all driver complied with this rule as showed in Table 2.

Under compliance with rules, braking test is checked as the brake test enables the driver to evaluate the performance of train braking system prior to the need to use it. Using OTDR data enables checking this rule prior to the driver first stop (due to a station or a red aspect). For example, Figure 3 shows that the driver carried out the brake test prior to the first AWS horn, in contrast, the driver presented by first AWS horn.

For vigilance aspect, wrong-side door release is presented. Releasing the doors on the wrong side of the train may have serious consequences as it could cause a potential harm to railway passengers. Only one of the journeys showed any instances that appeared to have wrong-side door release. This journey is shown in Figure 5 and is unusual in that the train appears to be stationary for most of the time. It is possible that this file shows a train under maintenance. Figure 5 shows door releases; on each occasion the right-hand side door (shown by the blue circles) is released shortly before the left-hand side door (shown by the red line). Because of the very short period between the two door releases, they appear to occur at the same time in the figure. Whilst this file does not appear to show an instance of an actual hazard—since the train did not appear to be moving—it nevertheless demonstrates that it is possible to use OTDR data to detect instances of wrong-side door release.

Table 2. Maximum speed at AWS horn prior to red aspect.

Journey Number	Number of red aspects	Maximum train Speed approaching a red aspect
Journey 1	0	NA*
Journey 2	2	11
Journey 3	1	14
Journey 4	1	11
Journey 5	1	14
Journey 6	2	13
Journey 7	1	14

*The driver did not have any red aspect signal during his/her journey

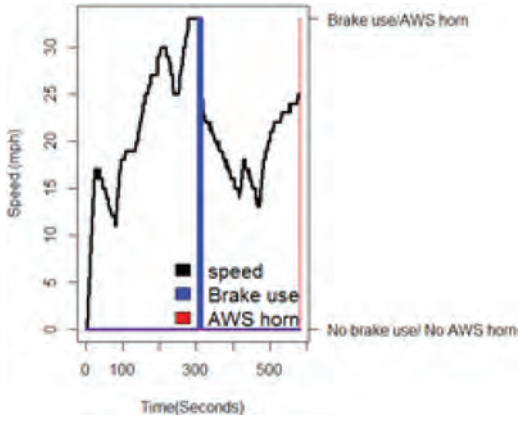


Figure 3. Brake test before stopping prior to the first AWS horn.

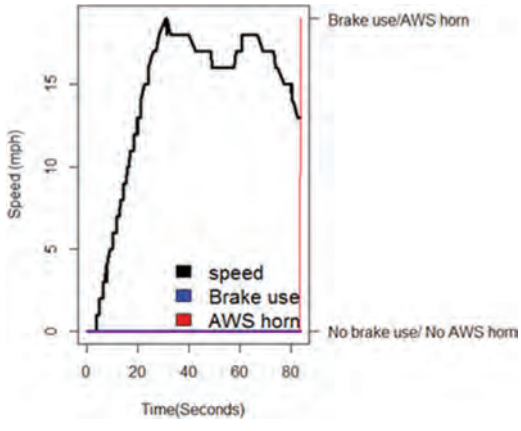


Figure 4. No brake test before stopping prior to the.

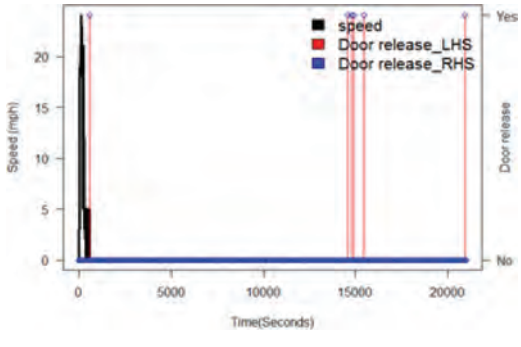


Figure 5. Wrong side door release.

5 DISCUSSION

Considering the growing interest in harvesting data sources such as OTDR, the development of DCPIs

that support that direction is essential. DCPIs proposed in this paper developed not only based the available data from OTDR but also supported by the official documents such as Rule Book and number of professional driving policy documents.

The technological approach described offers sensible solutions for the extraction of DCPIs from OTDR; it offers rapid analysis of the driver performance in contrast to in-cab-riding assessment by a driver manager that normally takes place every six months and only provides the opportunity for a driver to be assessed under a limited range of conditions. Furthermore, the in-cab-riding assessment may cause drivers to behave differently under observation, whilst DCPIs can be calculated without disrupting the driver.

DCPIs do not pass judgement about what is 'good' or 'bad' but illustrate how data is extracted in a sensible way from a huge dataset. For qualitative judgement, the allocation of indicators may need further discussion in the implementation stage to consider related parties point of view.

A few numerical examples of DCPIs are presented to illustrate the practicality of using OTDR to calculated DCPIs. However, the used data sample was very small to detect any trend or best practice rule.

6 CONCLUSION

In this paper, the train driver competency framework was introduced to outline the main areas of a driver function using documentary analysis (e.g. TOCs professional driving policy and the Rule Book), in addition to interviews and in cab-rides. A number of DCPIs have also proposed to assess driver performance under real-life conditions using OTDR data.

OTDR offers great sources to develop a comprehensive list of behavior aspects related to driver performance that can be determined from OTDR data.

REFERENCES

Balfe, N. 2017. A framework for human factors analysis of railway on-train data. In D. de Waard, A. Toffetti, R. Wiczorek, A. Sonderegger, S. Röttger, P., Bouchner, T. Franke, S. Fairclough, Noordzij, M. & Brookhuis, K. (Eds.) (2017). *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2016 Annual Conference. ISSN 2333–4959 (online).* Available from http://hfes-europe.org

El Rashidy, R. & Van Gulijk, C. 2016. Driver competence performance indicators using OTMR. In *Proceedings of CIT2016 Congreso de Ingeniería del Transporte (XII Congress of Transport Engineering, pp. 354–361.*

Green, S.R.& Barkby, S. & Puttock, A. & Craggs, R. 2001. Automatically assessing driver performance using black box OTDR data. *Railway Condition Monitoring and Non-Destructive Testing (RCM 2011), 5th IET Conference, pp.1–5, 29–30 Nov. 2001.*

HAS, The Health and Safety Authority, 2013. *Behaviour based safety guide: doing what we do better, smarter, safer.* HAS, the Health and Safety Authority, Dublin.

London Midland, LM. 2009. Professional driving policy.

McCorquodale, B. & Chissick, C. & McGuffog, A., Rowley, I. Bunting, A. & Page, H. 2002. Driver's Reminder Appliance (DRA) effectiveness study: Final report, Qinetiq/KI/CHS/CAP/CR020937/2.0/2.0, QinetiQ, Farn-borough.

Mcleod, R. & Walker, G.H. & Moray, N. & Love, G. 2003. Driver reliability with extended AWS. Project Summary Report. B/C271/FD.5. Nickleby HFE Ltd.

RSSB. 2004. The Rail Safety and Standards Board, Driver Error Data Collection Project: Final Report, RSSB.

RSSB. 2015. Guidance on Defective On-Train Equipment. Retrieved March 8, 2016, from: https://www.rssb.co.uk/rgs/standards/GOGN3637%20Iss%202.pdf.

Rule Book. 2015. Train Driver Manual. Retrieved March 8, 2016,from: https://www.rssb.co.uk/rgs/rulebooks/GERM8000-traindriver%20Iss%202.pdf.

South West Trains, LM. 2012. *Professional driving policy and braking instructions*. 7, South West Trains.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

A preliminary approach to subsea risk management using sensor network information

M. Bucelli
Alma Mater Studiorum—University of Bologna, Bologna, Italy
Norwegian University of Science and Technology NTNU, Trondheim, Norway

I.B. Utne & N. Paltrinieri
Norwegian University of Science and Technology NTNU, Trondheim, Norway

P. Salvo Rossi
Kongsberg Digital AS, Trondheim, Norway

V. Cozzani
Alma Mater Studiorum—University of Bologna, Bologna, Italy

ABSTRACT: During the last decade, increasing attention has been focused on environmental protection. For instance, the ecological effects of hydrocarbon releases in the sea are of paramount concern. One way to assess their environmental impact is to consider the amount of pollutant discharged. Effective early detection would help in revealing spills in advance and take the necessary mitigating measures to contain the released volume. Standards and guidelines are established for developing effective sensor networks in the subsea templates for monitoring purposes and data collection. Sensors provide a heterogeneous amount of information about the template they are monitoring. According to recent studies on risk assessment, the *level of knowledge* about a specific system is an intrinsic feature that should be considered during the assessment and evaluation phases for better managing potential increments of the risk level. The information provided by sensor networks may be used in this perspective. Sensors may be functionally placed in fault tree analyses and update the information about frequency deviation. The work in this paper is focused on risk management using such information from subsea sensor networks. A real reference case from the oil and gas industry located in an environmentally sensitive area on the Norwegian Continental Shelf is provided for testing the suggested approach. The case study refers to subsea monitoring of oil leakages from the wellhead templates. Insights from the case study highlight how sensor data analysis may improve risk management and support operational decision making.

1 INTRODUCTION

Dynamicity to risk assessment and management is a main challenge that today´s researchers have to face. A quantitative assessment of the level of risk for a production installation is required by law, but it is usually performed during the design phase. Effective support during operations is missing (Villa et al., 2016). The chemical and petrochemical industry requires tools and methods to update the risk picture on a real-time basis and then improving risk management (Paltrinieri and Khan, 2016). Different approaches have been suggested to dynamically update the risk level. Some of these are based on Bayesian networks (Khakzad et al., 2016, 2014) while others are proactive approaches based on indicators (Paltrinieri et al., 2016).

In this perspective, the Dynamic Risk Management Framework (DRMF) has been developed (Paltrinieri et al., 2014). Figure 1 shows the DRMF. DRMF focuses on the continuous systematization of information on new risk evidence. As shown in Figure 1, its shape opens the process to new information and early warnings by means of continuous monitoring. Side information is an input to each step of risk management through communication and consultation.

The available information provided by different sources, such as monitoring and control devices, but also training reports and audits, should be included and exploited when assessing the risk level during operations. As suggested by Aven and Krohn (2014), a new dimension to the definition of risk from Kaplan and Garrick (1981) should be

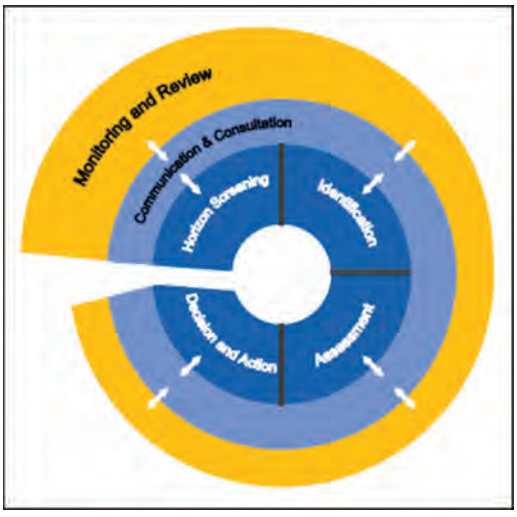


Figure 1. Dynamic risk management framework—clockwise (adapted from (Paltrinieri et al., 2014)).

added. As shown in Eq. (1), risk (R) is a function of the identified scenario (s), its probability (p), its consequences (c) and of what Aven and Krohn (2014) define as level of knowledge (k).

$$R = f(s, p, c, k) \tag{1}$$

The level of knowledge for a specific system is an intrinsic feature that should be considered during the assessment and evaluation phases for better managing potential increments of the risk level.

The information provided by sensor networks may be used in this perspective. Sensors may be functionally placed in fault tree analyses and update the information about frequency deviation. The current analysis in this paper refers to subsea detection networks.

Subsea leak detection is a considerable challenge for the oil and gas offshore industry, although the main concern for subsea templates is blow-out. As shown by Macondo (Deepwater Horizon Study Group, 2011) accident, the effect of a well blow-out due to the large amounts of spilt crude are catastrophic from human, environmental, economic and reputational point of views.

However, as oil and gas offshore production is moving north, towards sub-Arctic and Arctic areas, monitoring and control of crude oil spill are becoming critical issues. These areas are environmentally sensitive (Larsen et al., 2004) and specific requirements (DNV-GL, 2012) must be met during production. For instance, the Barents Sea area is recognized by the World Wildlife Fund (WWF) as critically sensitive from an environmental point of view (Larsen et al., 2004) due to:

- Naturalness;
- Representativeness;
- High biological diversity;
- High productivity;
- Ecological significance for species;
- Source area for essential ecological processes or life-support systems;
- Uniqueness; and
- Sensitivity.

The current development of large oil and gas templates in the Barents Sea may lead to severe pollution and increased risks of large oil spill (Bioforsk Soil and Environment, 2006), constituting a major threat to the biodiversity of this particularly sensitive area.

Detectors are required to show high sensitivity to small amounts of leaking hydrocarbons and to detect a spill in a reasonable time interval. This is the basis for early detection systems. The threshold value of a leakage rate to be detected by the sensors is a critical parameter that influence the choice and the cost of the device. Furthermore, the detectors have to be available and reliable when in place to effectively provide information to the topside control room. Fault logs' information may be gathered to evaluate to which extent the measurement by the sensor is trustable.

Moreover, it would be preferable to locate the leakage source through the detection system. Collecting information about where the template is spilling oil is useful for both intervention and consequent maintenance activities.

The contribution in this paper addresses the main challenges related to subsea oil detection coupled with risk management for a real case of an oil and gas Floating, Production, Storage and Offloading (FPSO) unit located in the Barents Sea. Available sensor network information is used to support risk management.

The paper is organized as follows: Section 2 provides some fundamentals of signal processing useful for a comprehensive understanding of how the subsea leak detector network works. The case study is extensively described in Section 3. The legislative requirements and both the subsea template and sensor network characteristics are included in this Section. The results of the study and their discussion are provided in Section 4 and 5. The paper ends with conclusions in Section 6.

2 FUNDAMENTALS OF SIGNAL PROCESSING FOR OIL DETECTION

The detection system purpose is to reveal hydrocarbon spills in the sea from the subsea equipment.

For the sake of simplicity, this work addresses the oil leakage event in a binary way: the presence of release is associated with the state H_1, while the absence with the state H_0. Sensors detect the presence (H_1) or absence (H_0) of oil leakage from the wellhead. A sensor's local detection is performed by comparing the registered signal with a fixed threshold.

Typically, distributed multiple sensors are in place to detect the oil leakage. Their number is defined as K and everyone is equipped with an acoustic transducer. Every i-th sensor makes a local decision, y_i, and this signal is transmitted to a fusion center (FC), which takes a (theoretically more reliable) global decision, d, about the presence or absence of the binary event. The global decision is derived by appropriately combining the received information on local decisions from different sensors. This type or architecture is defined as centralized and it is represented in Figure 2 (Salvo Rossi et al., 2016; Salvo Rossi and Ciuonzo, 2015).

Referring to Figure 2, the present study considers that the local decision from the i-th sensor, y_i, does not suffer of disturbance and signal attenuation while it is transferred to the FC. The signal transmitted to the FC from the i-th sensor is named r_i. For the assumptions made, the value of r_i corresponds with y_i.

Locally, at sensor level, four different decision situations may result considering a binary leak event. Such decision situations are summarized in Table 1. The present analysis assumes that every sensor senses autonomously the environment in a defined space cell to detect the presence or absence of a target (which in this specific case is the presence of oil leaking from the template).

The probability of detection (P_D), false alarm (P_F) and missed detection (P_M) are defined according to the equations 2–4:

$$P_D = p(y = H_1 \mid H_1) \tag{2}$$

$$P_F = p(y = H_1 \mid H_0) \tag{3}$$

$$P_M = p\left(y = H_0 | H_1\right) = 1 - P_D \tag{4}$$

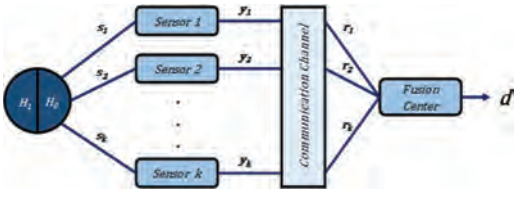


Figure 2. Distributed detection system (K sensors) with fusion center (adapted from Salvo Rossi and Ciuonzo (2015)).

Table 1. Detection and detection errors.

		DECISION	
		$d = H_0$	d = H_1
EVENT	H_0	Correct decision	Error type 2: False alarm
	H_1	Error type 1: Missed detection	Correct decision (detection)

The sensor local performance may be described by means of different parameters. The present work refers to P_D and P_F according to common practice in communication engineering studies (Salvo Rossi et al., 2016; Salvo Rossi and Ciuonzo, 2015). The present work assumes that sensors are independent from each other. Given this hypothesis, P_D and P_F are as well stationary and conditionally independent. The sensors within the network are assumed to have identical local performance (homogeneous network).

The detection system performance is evaluated in terms of the global probability of detection, Q_D, the global probability of false alarm, Q_F, and the global probability of missed detection, Q_M. They are defined according to the following equations 5–7:

$$Q_D = p(d = H_1 \mid H_1) \tag{5}$$

$$Q_F = p(d = H_1 \mid H_0) \tag{6}$$

$$Q_M = p\left(d = H_0 | H_1\right) = 1 - Q_D \tag{7}$$

The FC takes the final decision based on the received decisions and using a Fusion Rule (FR) (Javadi and Peiravi, 2013). This work applies the Counting Fusion Rule (CFR). The sum of sensor decisions is compared to a specific threshold at the FC to make the final decision (Javadi and Peiravi, 2013). The CFR is a simple and intuitive strategy to count the number of reported detections (Niu and Varshney, 2008), but it is far from the optimal performance (Javadi and Peiravi, 2013). However, it is suitable for the purpose of the current analysis as it does not require previous system knowledge and it provides a good basis for trade-off analysis.

3 CASE STUDY

As previously mentioned, this study focuses on the main challenges of subsea oil detection and risk management for a real case of an oil and gas Floating, Production, Storage and Offloading (FPSO) unit located in the Barents Sea.

For this reason, the study aims to evaluate if the facility detection system is able to:

– Improve subsea safety;
– Reduce environmental impact by controlling the released hydrocarbon quantities;
– Reduce the need for remotely operated vehicle (ROV) inspections.

In particular, the focus of this work is on early detection of oil releases in the subsea template on the seabed.

3.1 *Regulations and stakeholders*

Companies operating on the Norwegian Continental Shelf (NCS) are required to carry out environmental monitoring to obtain information about the actual and potential environmental impact of their activities (Norwegian Environment Agency, 2015). Different regulations set the requirements for the monitoring of petroleum activities. The regulations relating to conducting petroleum activities (The Activities Regulations) (Petroleum Safety Authority Norway, 2016a) dedicate Sections 52–57 to special requirements for environmental monitoring. These requirements include the monitoring of the water column and of the benthic habitats, as well as the establishment of an effective remote sensing system to detect and map acute pollution. The Management Regulations (Petroleum Safety Authority Norway, 2016b) require in Section 34 the operators to report the results obtained from monitoring of the external marine environment. These requirements have to be satisfied during oil and gas operations.

In 2014, a Joint Industry Project (JIP) led by DNV-GL was aimed at developing the best practices for designing and implementing detection systems (Leirgulen, 2014). Twenty key partners joined the project, including different operators, integrators and suppliers, as well as authorities, the Norwegian Ministry of Climate and Environment, and the Petroleum Safety Authority Norway (PSA) (Leirgulen, 2014). The JIP identified relevant functional requirements and general specification for a subsea detection system. The outcomes are included in the Recommended Practice F302 (DNV GL, 2016). The key functional requirements identified for the subsea detection system by DNV GL (2016) may be summarized as follows:

- Sensitivity to small releases;
- Responsiveness of the detection system;
- Availability and reliability of the leak detector;
- Ability to locate the leakage source.

Therefore, the detection system must satisfy the requirements set by the standard for oil detection in the subsea template RP-F302 (DNV GL, 2016). The standard sets qualitative requisites to be fulfilled. First, the Best Available Techniques (BAT) approach for leak detection has to be selected. RP-F302 requires a two-step BAT process where the firstly single techniques are assessed and then different configurations are compared to identify the most efficient in cost and risk reduction.

Anyway, the analysis of the different standards does not provide straightforward guidelines for the positioning of subsea leak detectors. Different configurations have to be assessed and redundancy margins to be guaranteed. The main purpose of the subsea network is to strain the detection of oil releases to unit.

Different actors are involved in the response when a subsea leak is detected. The topside operators have to gather relevant information and start preliminary mitigation actions. Moreover, the offshore personnel have to consult experts from the onshore department, and notify the coast guard and to the airborne in case their intervention is needed. From the topside, it is possible to monitor and control the amount of oil released from the subsea equipment. The production system needs a detailed and reliable picture of the situation in the subsea template in case there would be a need for shut-down. The economic impact of unplanned shutdowns can be severe for oil and gas companies (Oil and Gas IQ, 2014). Assessing the risk in a detailed way may allow minimizing time (and costs) of unnecessary stops of production. Moreover, the effectiveness of the subsea detection system is also critical for limiting the number of unplanned ROV inspections. ROVs are operated by a crew on board dedicated vessels and are usually used for maintenance activities on the subsea templates. ROV inspections are extremely expensive and especially dedicated expert personnel is required. A reliable sensor network able to identify releases due to mechanical failures would be helpful in eliminating the costs of unnecessary ROV inspections. With a detection system that works effectively and identifies (and eventually locates) the leakage sources, the number of required interventions from the topside would decrease, leading to a subsequent drop in operation costs.

In addition, different environmental organizations have raised their concern about oil and gas exploration and drilling, particularly in the sensitive Arctic and sub-Arctic areas, which are critical for biodiversity and ecological significance (Greenpeace, 2017). These organizations may affect public opinion towards the environmental protection policy of a company. The impact on reputation of oil and gas operators may be severe. For this reason, implementation of advanced and effective strategies and technologies for environmental protection should be a main priority for the operator company.

For all these reasons, the stakeholders of subsea oil and gas activities within Arctic and sub-Arctic regions may be the following:

- Offshore operator;
- Production system;
- Onshore department;
- Coast guard;
- Airborne;
- ROV operator;
- Sensor supplier;
- Petroleum Safety Authority Norway;
- Environmental protection agency; and
- Non-Governmental Organizations (NGOs).

3.2 *Subsea template overview*

The subsea template on the seabed is a critical area of the oil production installation where a high number of valves and joint points are located. These critical connections may be potential sources of oil leakage due to pressure increments during production disturbances and/or mechanical failures.

Sensors are placed in the template structure to early detect oil releases. Although different types of sensors may be available, this analysis refers only to acoustic oil leak detectors.

Figure 3 shows the physical elements needed for the detection of hydrocarbon leakage at the subsea wellhead and X-Tree (adapted from Røsby (2011)).

3.3 *Sensors characteristics and configurations*

According to RP F302 (DNV GL, 2016), there are no unique guidelines to locate the sensors in the distributed detection network. The only relevant requirement concerns the use of BAT approach for early detection of oil releases.

Two types of sensors are considered named Type A and Type B, respectively. They are set to work with the same P_F (equal to 10^{-2}) as common practice in telecommunication engineering studies. However, the sensors have different Receiver Operating Characteristic (ROC) curves and this results in different P_D. The more performing sensor (Type A) has a P_D of 0.90 and the other (Type B) of 0.50 (Salvo Rossi et al., 2016).

Table 2 summarizes the characteristics of the sensors.

Detection and its reliability are key parameters during oil and gas operations. Reliably assessing that a mechanical rupture has happened and that the template is leaking is critical in efficient ROV intervention management.

The sensors are placed in two different configurations. The area of interest is organized in structured square cell grids, as shown in Figure 4. The first configuration considers one single sensor for each grid cell defined in the sensed environment (namely, single configuration). In the second configuration, redundant N sensors monitor the presence (or absence) of the target of interest (namely, redundant configuration). Figure 4 is shown as representative.

The case study compares the detection performance of the distributed sensor network in two cases. The first scenarios refer to the single configuration using high-performance acoustic sensors in terms of detection probability (Type A). The second considers the redundant configuration applying theoretically cheaper and less performance sensors (Type B). The detection performance of the two sensors are described in Table 2. The single configuration uses one sensor of Type A for each grid node described in Figure 4. The sensor covers the

Table 2. Description of detection performance for sensors Type A and B.

Sensor	P_D	P_F
Type A	0.90	0.01
Type B	0.50	0.01

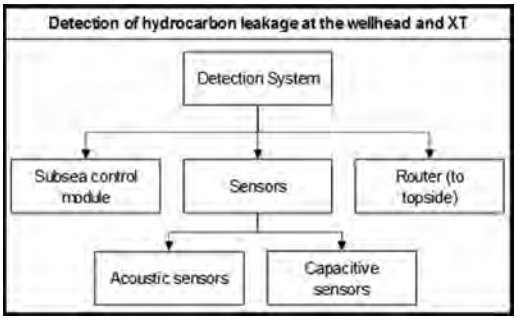


Figure 3. Detection system for the wellhead and X-Tree (adapted from Røsby (2011)).

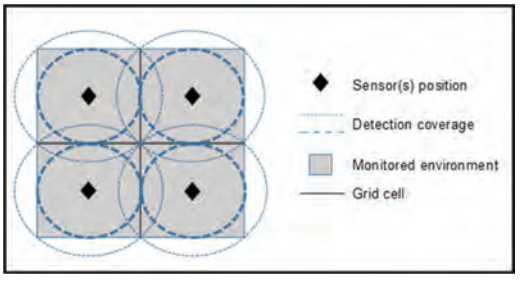


Figure 4. Sensor grid in the monitored environment.

entire grid cell and it sends its local decision about the presence or absence of oil release to the FC. The redundant configuration applies a number of N sensors Type B for each grid cells. The number N of sensors should be defined to approximately match the detection probability obtained with a single Type A sensor. The CFR is applied as fusion rule at the FC. The threshold is set conservatively to 1. This means that the FC conservatively takes a positive decision on the presence of oil leakage when at least one detector monitoring the grid cell sends a signal revealing the presence of the target.

4 RESULTS

The current analysis considers a release trend as the one shown in Figure 5. It is worth to noticing that the release behaviour has been adopted for demonstrative purposes. The sensors detect noises from the subsea template and they record them above a defined threshold. Some oscillations are recorded due to any pressure variation in the reservoir. In that case, the pressure is controlled and reset to its optimal value without any intervention from the topside (see the first 50 time steps in Figure 5). This trend may also be due to some slightly overpressure scenario developing in the first year of production, when the pressure in the reservoir is higher (Kansas Geological Survey, 2000). The oscillations may result in fatigue on mechanical components and induce a mechanical failure of some valve in the X-mas tree and wellhead. The template is then continuously leaking and it needs dedicated inspections and intervention.

The detection and false alarm probabilities are calculated using the fusion rule described in Section 2. Table 3 shows that a number of Type B sensors equal to 4 in the redundant configuration has been found to approximately match the detection probability obtained with a single Type A sensor.

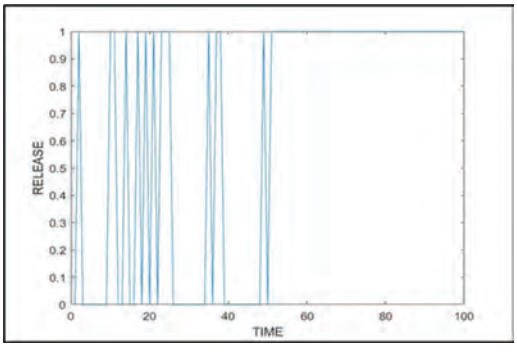


Figure 5. Assumed target trend for the present analysis.

Table 3. Detection performance of the subsea template system in two different configurations.

Configuration	SINGLE	REDUNDANT
Sensor type	Type A	Type B
Number for each grid cell	1	4
P_D	0.90	0.94
P_F	0.01	0.04

The P_D in the redundant configuration is slightly higher than the in the single configuration, while the P_F is four times increased.

5 DISCUSSION

Table 3 highlights a relevant increment of the detection system performance using redundant "cheap" sensors (Type B). The detection probability P_D for a single Type B sensor is 0.50 (see Table 2 in Section 3.3), but it is almost doubled in the redundant configuration (0.94). Moreover, this type of configuration slightly exceeds the single expensive Type A sensor P_D. Redundant configurations of Type A sensors may be also considered. However, the increment in detection performance would not be as relevant as in the case of Type B sensors as their performance is already high.

However, the redundant configuration as described in this work allows the increment of the false alarm probability P_F as shown in Table 3. This may have a negative effect on the organizational levels. False alarms may result in unnecessary unplanned ROV inspections and shut-downs with strong increments of operational costs. The fusion rule for the FC adopts a conservative approach with respect to leak detection for which the global decision is the presence of leak in case *at least* one detector emits a positive signal. That justifies a global P_F for the redundant configuration of four times the single value.

A more sophisticated decision rule should be implemented (Javadi and Peiravi, 2013). The sensor placement should be investigated and optimized to guarantee early detection and to track the oil spill movement in case intervention is needed.

The sensors detect noises from the subsea template and they record them above a defined threshold. Some oscillations are recorded due to any pressure variation in the reservoir. In such scenarios, the pressure is controlled and reset to its optimal value without any intervention from the topside. This trend may also be due to some slightly overpressure scenario developing in the first year of production, when the pressure in the reservoir is higher (Kansas Geological Survey, 2000). These oscillations may result in fatigue on

mechanical component and induce a mechanical failure of some valve in the X-mas tree and wellhead. The template may then continuously leak, needing dedicated inspections and intervention.

According to the results of the performed simulations, the number of missed detections is lower in redundant configurations. It is possible to identify and distinguish if the release is due to well fluctuations or mechanical failures by coupling the signal from the FC and pressure data. This allows recording of early warnings and use them for risk assessment and management. The analysis of near-accident data is a fundamental step in the framework to forecast likely accident scenarios. The information from sensor networks may provide a basis to the reactive update the risk picture of the installation with respect to subsea leakage risk. For instance, the data from sensors may be used as evidence in Bayesian inference network for updating release probabilities (Paltrinieri and Khan, 2016).

Reliable information from the subsea may also improve communication between different stakeholders and decision making processes.

6 CONCLUSION

The threshold leakage rate for the sensors defines its sensitivity and therefore its cost. High sensitivity (detection of lower leakage rate) results in highly sophisticated sensors with substantial cost. A solution would be the application of low cost redundant sensors located in a specific network in order to perform early detection. The decision about the presence of oil leakage into sea from the subsea template determines the need of intervention from the topside. Different (internal and external) stakeholders are involved in oil and gas facilities. A reliable subsea detection system may help avoid unnecessary intervention and improve the overall company risk management. Moreover, every intervention to the subsea template from the topside requires substantial costs that may be reduced with a reliable basis of information.

The analysis suggests the investigation of different sensor placement configurations in order to enhance early detection and oil leakage tracking. Further studies should be considered applying different and more specific decision fusion rules.

The information from decision making may be used in updating the risk picture of the installation and in improving the decision making process.

REFERENCES

Aven, T., Krohn, B.S., 2014. A new perspective on how to understand, assess and manage risk and the unforeseen. *Reliab. Eng. Syst. Saf.* 121, 1–10.

Bioforsk Soil and Environment, 2006. Barentswatch 2006 - The Barents Sea environment and petroleum activity.

Deepwater Horizon Study Group, 2011. Final Report on the Investigation of the Macondo Well Blowout Deepwater Horizon Study Group.

DNV-GL, 2012. Barents 2020 Assessment of international standards for safe exploration, production and transportation of oil and gas in the Barents Sea - Report no 2012–0690.

DNV GL, 2016. Recommended Practice DNVGL-RP-F302 Offshore Leak Detection.

Greenpeace, 2017. Save the Arctic [WWW Document]. URL http://www.greenpeace.org/international/en/campaigns/climate-change/arctic-impacts/ (accessed 12.5.17).

Javadi, S.H., Peiravi, A., 2013. Weighted Decision Fusion vs. Counting Rule over Wireless Sensor Networks : A Realistic Comparison, in: 21st Iranian Conference on Electrical Engineering ICEE, 14–16 May 2013. Mashhad, Iran, pp. 3–8.

Kansas Geological Survey, 2000. Application of Horizontal Wells in Mature Basins - A case study from Kansas [WWW Document]. URL http://www.kgs.ku.edu/Class2/Tulsa/index.htm (accessed 12.8.17).

Kaplan, S., Garrick, B., 1981. On the quantitative definition of risk. Risk Anal. 1, 11–27.

Khakzad, N., Khan, F., Paltrinieri, N., 2014. On the application of near accident data to risk analysis of major accidents. *Reliab. Eng. Syst. Saf.* 126, 116–25.

Khakzad, N., Paltrinieri, N., Khan, F., 2016. Chapter 5 - Reactive approache of Probability Update Base on Bayesian Methods, in: Dynamic Risk Analysis in the Chemical and Petroleum Industry. pp. 51–61.

Larsen, T., Nagoda, D., Andersen, J.R., 2004. The Barents Sea ecoregion: A biodiversity assessment. WWF's Barents Sea Ecoregion Programme.

Leirgulen, S.I., 2014. Joint Industry Project to enhance the offshore leak detection approach [WWW Document]. DNV GL. URL https://www.dnvgl.com/news/joint-industry-project-to-enhance-the-offshore-leak-detection-approach-8096 (accessed 10.26.17).

Niu, R., Varshney, P.K., 2008. Performance analysis of distributed detection in a random sensor field. IEEE Trans. Signal Process. 56, 339–349.

Norwegian Environment Agency, 2015. Environmental monitoring of petroleum activities on the Norwegian continental shelf - Guidelines M–408.

Oil and Gas IQ, 2014. Shutdowns and Turnaround in the Oil and Gas Industry [WWW Document]. Glob. Oil Gas Intell. URL https://www.oilandgasiq.com/integrity-hse-maintenance/articles/shutdowns-and-turnarounds-in-the-oil-and-gas-indus (accessed 11.13.17).

Paltrinieri, N., Khan, F., 2016. Dynamic Risk Analysis in the Chemical and Petroleum Industry: Evolution and Interaction with Parallel Disciplines in the Perspective of Industrial Application, 1st ed. Elsevier Science.

Paltrinieri, N., Khan, F., Amyotte, P., Cozzani, V., 2014. Dynamic approach to risk management: application to the Hoeganaes metal dust accidents. *Process Saf. Environ. Prot.* 6, 669–79.

Paltrinieri, N., Landucci, G., Nelson, W.R., Hauge, S., 2016. Chapter 6 - Proactive Approaches of Dynamic

Risk Assessment Based on Indicators, in: Dynamic Risk Analysis in the Chemical and Petroleum Industry. Butterworth-Heinemann, pp. 63–73. doi:http://dx.doi.org/10.1016/B978-0-12-803765-2.00006-8

Petroleum Safety Authority Norway, 2016a. Regulations relating to conducting petroleum activities (The Activities Regulations) [WWW Document]. URL http://www.ptil.no/activities (accessed 11.11.17).

Petroleum Safety Authority Norway, 2016b. Regulations relating to management and the duty to provide information in the petroleum activities and at certain onshore facilities (The Management Regulations) [WWW Document]. URL http://www.ptil.no/management (accessed 11.6.17).

Røsby, E., 2011. Goliat development project Subsea leak detection design, in: Aker Solution.

Salvo Rossi, P., Ciuonzo, D., 2015. Energy Detection for MIMO Decision Fusion in Underwater Sensor Networks. IEEE Sens. J. 15, 1630–1640.

Salvo Rossi, P., Ciuonzo, D., Kansanen, K., Ekman, T., 2016. Performance Analysis of Energy Detection for MIMO Decision Fusion in Wireless Sensor Networks Over Arbitrary Fading Channels. IEEE Trans. Wirel. Commun. 15, 7794–7806.

Villa, V., Paltrinieri, N., Khan, F., Cozzani, V., 2016. Towards dynamic risk analysis: a review of the risk assessment approach and its limitations in the chemical process industry. *Saf. Sci.* 89, 77–93.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Reliability-based cyber plant

Harald Rødseth & Per Schjølberg
Norwegian University of Science and Technology (NTNU), Norway

Ragnhild Eleftheriadis & Odd Myklebust
Sintef Raufoss Manufacturing, Norway

ABSTRACT: With the onset of Industry 4.0 several technological possibilities are offered in industry such as big data analytics, digital twin and augmented reality. The result is a more digitalised industry where faster and better decisions are possible. In long term this should provide a more reliable production with increased plant capacity and reduced downtime. To succeed with these possibilities a Cyber Physical Systems (CPS) must be established for the company. Currently, an own framework for CPS is under development and is expected to be tailored for Norwegian manufacturing. When building on the principle in Industry 4.0, big data capability with machine learning will be a fundamental model. Nevertheless, Industry 4.0 should also include other models for big data capability such as reliability modelling. The aim in this article is to present the current status of CPS framework and how it could be implemented in manufacturing industries. In particular, the article discusses and demonstrates the balance between machine learning and reliability engineering in big data analytics.

1 INTRODUCTION

The European competitive advantage is under pressure, where customer needs, such as improved delivery accuracy of products, have changed over time (Smart Industry, 2017). It might challenge the future industry in Europe how to implement and digitalize equipment and tools for a safe and reliable environment.

Several initiatives like platforms for Industry 4.0 have been established (Kagermann et al., 2013). Also national strategic initiatives have been established, like "Smart Industry" in Netherland (Smart Industry, 2017) and "Industry, greener, smarter and more innovative" in Norway (Ministry of Trade Industry and Fisheries, 2017) where the focus is adapting systems for data handling and digitalization. Several important elements can be related to Industry 4.0 such as predictive maintenance (McKinsey&Company, 2015). The benefit of predictive maintenance is improved reliability with application of the opportunities from big data and statistics where application of continuous real-time monitoring of assets, with alerts given based on pre-established rules or criticality levels (Pwc, 2017). It remains to investigate more in detail how reliability engineering methods also can be combined with big data analytics in order to improve the reliability of the production plant.

Another important element of Industry 4.0 is cyber-physical systems (CPS) (Kagermann et al., 2013). As an overall understanding, CPS are integrations of computation with physical processes (Lee, 2008). Since manufacturing is one essential application of CPS (Lee, 2008), the notion cyber-physical production systems (CPPS) is often used in manufacturing and production (Monostori, 2014, Lee et al., 2017, Hehenberger et al., 2016, Monostori et al., 2016).

Although the economic impact of applying the CPS in manufacturing is significant, computing and network technologies today may impede the progress towards this application (Lee, 2008). For example, the "best effort" in networking technologies make predictable and reliable real-time performance difficult. Nevertheless, certain efforts have been conducted where structures and architectures of CPS have been constructed, ranging from typical sketches with sensors and actuators (Lee, 2010), towards more generic architectures both as level based CPS (Lee et al., 2015) and CPS architecture with three dimensions (IEC, 2017).

Several challenges have been addressed for CPS, such as physical critical infrastructure that calls for preventive maintenance (Rajkumar et al., 2010). It has also been pointed out as a challenge to have a CPS architectures that are both "globally virtual and locally physical" (Rajkumar et al., 2010). Another challenge is need for standards (Chaâri et al., 2016). Although a pre-standard of CPS has been published (IEC, 2017) the industry has already started to test alternative architectures (Lee et al., 2017) in advent for a standard.

In Norway it is of interest to establish a CPS framework for Norwegian Industry. To create such a framework an ongoing competence project where framework, tools and implementation in demonstrators are in progress (Eleftheriadis and Myklebust, 2017).

The aim of this article is to present the current status of a Norwegian CPS framework and how it could be implemented in manufacturing and process industries.

To achieve this aim, following sub objectives are outlined:

1. Present existing elements for CPS architecture
2. Present existing CPS architecture for Norwegian manufacturing and process industry
3. Evaluate how it can be further developed based on existing CPS theory
4. Propose reliability-based analysis methods and technology for the CPS architecture
5. Discuss how the CPS architecture will be implemented in Norwegian industry.

The remainder of this article is structured as follows: In Section 2 existing elements for CPS architectures are presented. Based on these elements the Norwegian CPS framework is constructed in Section 3. In Section 4 three relevant CPS analysis methods and technologies are proposed and elaborated; life cycle profit (1), Safety perspective (2), and machine learning related to reliability engineering (3). Section 5 elaborates how the CPS framework can be implemented, while concluding remarks are made in Section 6.

2 EXISTING ELEMENTS FOR CPS ARCHITECTURES

To ensure successful application of the breakthrough technologies offered in Industry 4.0 in an organisation, a concrete architecture for CPS must be established. Today, there exist several architectures for CPS. In particular three architectures seem to be of relevance in Industry 4.0.

As a first example of CPS architecture, Lee et al. (2015) has proposed a 5-level CPS architecture denoted as the 5C architecture. This architecture provides a step-by step guideline in rolling out CPS in manufacturing with following levels:

1. *Smart Connection level.* Implementing the necessary instrumentation of machines, "plug & play" sensors, and wireless communication.
2. *Data-to-Information Conversion level.* The data collected from the sensors will be input-data for several models that provide information such as assessment of degradation.
3. *Cyber level.* At this level the digital twin of the plant is established and more advanced analytics is possible with assessment of fleet of machines.
4. *Cognition level.* To support the decision-maker to conduct faster and better decisions, proper presentation of the acquired knowledge is necessary. This level visualises e.g. future factory performance and key performance indicators.
5. *Configuration level.* This level provides feedback from the virtual world back to the physical world based on decisions conducted in level 4. This level also self-optimizes several properties of the plant.

Some successful case studies of the 5C architecture have recently been conducted both for ball screw health monitoring (Lee et al., 2017) and a wire rod machine (Rødseth et al., 2016b).

A second proposed architecture for CPS classifies the digitalization of Industry 4.0 into two types of value chains: Horizontal and vertical value chain (Geissbauer et al., 2014). The horizontal value chain comprises suppliers, the company and its customers, whereas the vertical value chain comprises activities in the company such as sales, manufacturing, service and product development.

A third proposed CPS architecture is "reference architecture model industry 4.0" (RAMI 4.0). Currently, a PAS (publicly available specification) has been developed for RAMI 4.0 (IEC, 2017). This specification does not fulfils the requirements for a standard, but is at least made available to the public. The core in RAMI 4.0 is to ensure cooperation and collaboration between technical assets which has a value for an organisation. RAMI 4.0 comprise a CPS architecture visualised with three dimensions:

1. *Layers*. In total six layers represent the information relevant for the technical asset: Business, functional, information, communication, integration and asset.
2. *Life cycle and value stream.* This dimension represent the life cycle of the technical asset.
3. *Hierarchy.* The hierarchy classifies the enterprise system into following categories: Connected world, enterprise, work centres, station, control device, field device and product.

RAMI 4.0 is developed from a more "Generalized Enterprise Reference Architecture Methodology (GERAM) which later was converted to three standards in late nineties. The GERAM was a extension of Computer Integrated Manufacturing (CIM) models which is an early enterprise or business model (Myklebust, 2002). The integration of GERAM and RAMI 4.0 from (Industrial Internet Consortium, 2016) shows the building block of a enterprise model that has interrelationships between organisational, process and product

structures. Members of the organisation are connected to process roles defining their work tasks. Competence are connected to process roles, goals are connected to the processes and products, and resources are connected to processes.

The GERAM later RAMI has a well-structured design and fit well with the generic demand of product, process and organisation. The link to the manufacturing system theory is therefore the last approach to include the product configuration and design process of disciplines like mechanics, cybernetic and material science on the physical side and planning activities, economical aspects and optimization processes on the logical side. Theoretically based on geometrical foundation and the methods within the theory that are related to concepts of connections. The analysis of the manufacturing systems is the prime area for the usage of this theory and is important to bring a science base into manufacturing. However how to succeed in developing, managing and operating such an enterprise model is still maybe the main challenge.

3 CONSTRUCTING A NEW CPS FRAMEWORK

Figure 1 presents the proposed CPS framework tailored for Norwegian manufacturing and process industry. With the motivation of establishing a value chain between vendor and the user (Geissbauer et al., 2014) a horizontal value chain has been outlined. In addition, inspired partly by the 5C architecture (Lee et al., 2015), a vertical value chain is also proposed to ensure that data from sensors will lead to smarter decisions. In total a CPS framework with two dimensions are developed.

The horizontal value chain consist of vendor (A), the production (B), and the customer where the end-product is consumed (C). At the vendor the asset is created and the support is provided from the vendor. The vendor can e.g. be a machine builder and supports with providing the maintenance programme. As pointed out by (Smart Industry, 2017) the maintenance could be totally outsourced where all maintenance activities are performed by the vendor and the machine is leased by the user. This will also require a more strategic alliance with the vendor (Batran et al., 2017). The production is where the asset, such as the machine is operated. At this location, an own maintenance management is located to ensure that the required technical condition of the asset is achieved with support from both internal and external maintenance resources. It is a crucial decision for the maintenance management to establish the most appropriate maintenance strategy relevant for the vendor. An important issue for the maintenance management to decide is the correct degree of maintenance outsourcing. The end-product is located at the customer where it is consumed. The customer value will be influenced by production where lack of maintenance can reduce the production assurance and result in late product delivery. Also a defect in production can be undetected and finally discovered by the customer. With application of real-time system, changes in customer requests will be ensured.

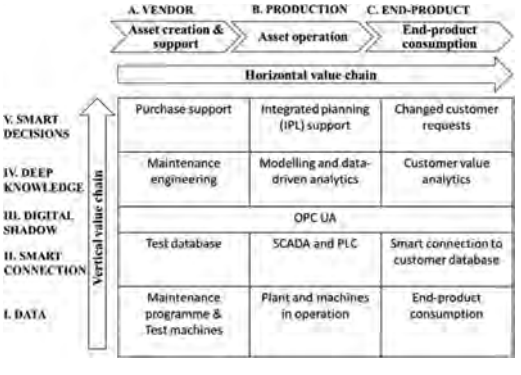


Figure 1. CPS framework.

The vertical value chain consist of six separate levels of data from assets (I) smart connection to assets (II), a digital shadow of the data (III), deep knowledge application (IV), smart decision application (V). At level I the asset is located that provides value for the organsiation. From the asset all relevant raw data is collected. The asset is not only the asset crated by the vendor such as the machine, but also other technical objects such as data servers, ERP-systems, algorithms and software programs. At this level the raw data is extracted from the physical assets, e.g. data capturing from a temperature sensor in a machine. The next level is smart connection (II) where data is extracted with SCADA and PLC systems and organized in databases such as ERP. To ensure that all databases can exchange data, an own level of OPC UA is necessary. In level IV, it will therefore be possible to apply deep knowledge analytics where databases at production and vendor can provide data-driven analytics in e.g. predictive maintenance.

In level V the deep knowledge analytics will support the decision maker with visualization and dashboards. This level can be considered to be a "digital advisor" for the decision maker. As an example in production, application of key performance indicators in integrated planning can support the planner to improve his future activities.

4 CPS ANALYSIS METHODS AND TECHNOLOGIES

4.1 *Life cycle profit*

Life cycle profit (LCP) is in this article defined as "*accumulated profit of a component or system over it's lifetime*". LCP presents the potential financial losses over the lifetime of a system due to the different time losses measured in overall equipment effectiveness (OEE) (Nakajima, 1989). The profit generated from the system after these losses is then LCP. Table 1 presents a proposed correlation between the time losses in OEE and LCP.

Figure 2 illustrates the LCP model, modified from Rolstadås et al. (1999). The area above the line x-x represents the time losses in accordance with Nakajima (1989). In addition, this area also distinguish between planned and unplanned maintenance. The reason for this distinction is that some of the planned maintenance require a shutdown of the machine and if necessary the production plant. If there are no time losses for the machine, it would be no area above the line x-x and maximum turnover would be achieved.

The area bellow the line x-x represents the costs that occurs during operation of the machine. In this model it is assumed that capital expenditure (CAPEX) is constantly scarred over the operation time. Both the maintenance costs and production costs will be decries in the start and increase at the end of the lifetime. The curve of line B-B will for the *bathtub curve* due to its characteristic shape and is due to the failure rate of the system over its lifetime (Sintef and Oreda, 2009, Rausand and Høyland, 2004). The bathtub can be divided into three specific phases:

- *Burn-in period.* This is an initial phase with high failure rate due to undiscovered defects. This is also known as "infant mortality".
- *Useful life period.* This phase is considered to be the useful period of the system where the failure rate is constant due to the maintenance activities.
- *Wear-out period.* In this phase, the regular maintenance activities can no longer keep the failure rate constant and it will decrease until the disposal of the system.

The LCP should be developed by the vendor with support from production in the CPS framework. With support from historical operations and loads it will be possible to achieve more accurate life cycle profit calculations.

Table 1. Time losses and LCP.

Time loss category	LCP element
Availability	Production Maintenance Resources
Performance	Degraded machine, energy loss.
Quality	Value of product before it is scrapped

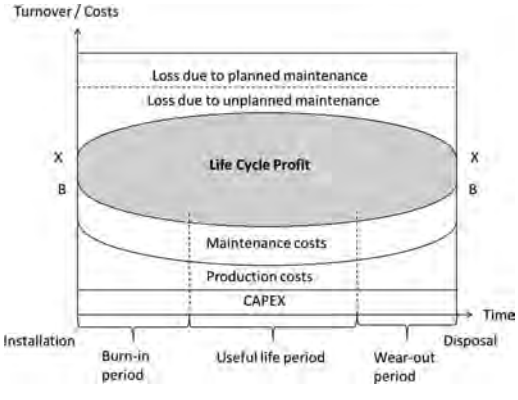


Figure 2. Life cycle profit model, modified from (Rolstadås et al., 1999).

4.2 *Safety perspective*

Regarding the safety perspective, following statements will be important:

- All corrective maintenance is deviation from required function.
- All maintenance activities will have a risk potential.
- Good maintenance is a pillar for effective and safe manufacturing and production.

The safety perspective will be of relevance of following situations:

- Accidents during maintenance
- Wrong type of maintenance
- Lack of maintenance

Table 2 presents a proposal of how these perspectives are relevant for the CPS framework.

4.3 *Machine learning and reliability engineering*

CPS plant position analytics in level IV with deep knowledge. It has been pointed by the European commission that intelligent maintenance systems based on condition prediction mechanisms with computation of remaining useful life (RUL) will increase reliability availability and safety (EFFRA, 2013). Furthermore, more sophisticated techniques for cause-effect and trend analyses are also

Table 2. Safety perspective in the CPS framework.

Safety perspective	Example of position of CPS framework	Example of Application in CPS
Accidents during maintenance	B. Production V. Smart decisions	Application of augmented reality.
Wrong type of maintenance	A. Vendor IV. Deep knowledge	Real-time notification to vendor in maintenance engineering.
Lack of maintenance	B. Production IV. Deep knowledge	Estimation of RUL in real-time with machine learning.

required. The deep analytics has been developed an integrated approach form machine learning and the need for zero defect manufacturing (ZDM). With intelligent sensor system ZDM can be operated for short term, medium term and long term decisions in the EU-project IFaCOM (intelligent fault correction and self-optimizing manufacturing systems) (Rødseth et al., 2016a). It has also been argued that maintenance could be one part of the IFaCOM concept. When advancing towards novel predictive maintenance technologies with reliability-based maintenance approaches, it is pointed out that this should include quality-maintenance methods as well as failure modes, effects, and criticality analysis (FMECA) (European Commision, 2016). Thus it is in this article of interest to investigate how FMECA can be balanced with big data analytics such as machine learning.

The maintenance model called deep digital maintenance (DDM) comprise an artificial intelligence module that tested remaining useful life (RUL) prediction based on dataset of degradation simulation run-to-failure data of jet engines (Rødseth et al., 2017). The output of the prediction model is the probability that RUL is more than 10 cycles in a specific point in time, denoted as $P_t(RUL > 10)$. One cycle is a magnitude for time, e.g. one week.

The prediction model should also include some error estimate to indicate the accuracy. Predictive maintenance should improve the maintenance planning capability in the organization where the plant capacity is increased as well as improved utilization of maintenance resources. The latter can be controlled by capacity overview (Liebstückel, 2014). Due to the operational conditions of degree of prediction in predictive maintenance and the available capacity of the craft technicians, the maintenance window needed by the maintenance planner will vary. Liebstückel (2014) has exemplified the

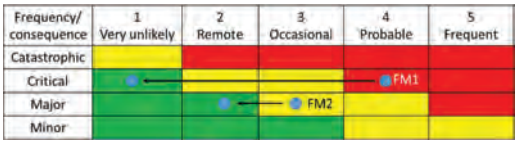


Figure 3. Risk matrix from FMECA, adapted from (Rausand and Høyland, 2004).

maintenance window to be 10 weeks when the planner shall conduct a capacity evaluation.

FMECA evaluates the risk of each failure mode and risk reducing measures. The risk of each failure mode may be positioned in a risk matrix shown in Figure 3 (Rausand and Høyland, 2004). The decision criteria for the risk matrix is as follows:

- *Red area.* The risk is unacceptable and risk reducing measures are required.
- *Yellow area.* Acceptable level of risk. The risk should be as low as reasonable as possible. Further investigations should be considered.
- *Green Area.* Acceptable level of risk. Only consider to keep the risk as low as reasonable as possible.

When the relevant failure modes has been evaluated in FMECA, it is further possible to evaluate to what degree implementation of machine learning in predictive maintenance can reduce the frequency of each failure mode. In the risk matrix, there are two failure modes denoted FM1 and FM2. The failure mode FM1 has non-acceptable risk whereas failure mode FM2 has acceptable but should still be investigate further for risk reducing measures. For both FM1 and FM2 predictive maintenance with machine learning is considered. To reduce the risk for FM1 to the green area, high accuracy in machine learning will be required. For FM2, it is not required the same accuracy in machine learning in predictive maintenance to reduce the risk to the green area.

Following criteria must be considered when evaluating the reduction of frequency due to implementation of predictive maintenance as risk reducing measure:

- The needed maintenance window and the accuracy of the trained data set.
- The similarity of operational conditions from the trained data and the predicted data.

5 ROLLING OUT THE CPS FRAMEWORK IN ORGANISATIONS

In a Norwegian perspective implementation of a CPS framework has to be followed up by guide-

lines and tools for fulfilling the expected impact. The Norwegian industry is probably one of the most organized labour markets in Europe and consist generally of small and medium size businesses. Where the labour policy for decades has been based on a tripartite cooperation between the government, trade unions and enterprise federations. The result is a flat structure where the involvement of skilled and self-dependent workers has been essential for competing in a global market.

The expectation of digitalizing Norwegian industry is therefore improved performance and a higher productivity. However, thru different maturity mappings, literature and surveys we can see the complexity in CPS and Industry 4.0 is broad. There is an image of a leadership which request for change, but do not find the right tools on one side. On the other hand, impatient workers with high digital competence and a mix match of equipment not prepared for digitalization. (Eleftheriadis and Myklebust, 2017)

A development of regulated safety and quality cultures is one of the benefit from such an organised labour where structure for reliable quality systems, preventive maintenance and management methods are implemented and where the improvement is a part of the organised culture.

6 CONCLUDING REMARKS

The aim of this article is to present the current status of CPS framework and how it can be implemented in Norwegian industry. With sound concepts of CPS theory a framework was proposed to be implemented for Norwegian industry which is both vertical and horizontal integrated. Also the analysis methods LCP and FMECA and different technological application for the safety perspective was recommended for the CPS framework.

The benefit of the CPS framework is that it can integrate all relevant data at sensor level up to decisions at plant level and at the same time connect the horizontal value chain including the machine builder, industrial user of the machine and the customer that consumes the end-product. As an impact for the industry it is expected that the value creation of this framework will be measured in terms of improved asset utilization with improved availability as well as improved product quality with reduced scrappage.

CPS plant will require parallel work with both vertical and horizontal integration of the CPS framework. Further work for the vertical integration will require specification of data capturing including establishment of a detailed specification of sensors that are to be applied in the project. For the horizontal integration, identification of interfaces in the horizontal value chain should be identified and mapping the value across companies.

Further work of the safety perspective would be to build a list of recommended application in CPS framework based on Table 2 and evaluate the reduction of risk. For the LCP, the horizontal value chain should be mapped when the vendor develops the LCP with support from production. The further development of FMECA would require more cooperation between the reliability engineering and machine learning. In detail, this would require simulation where the accuracy of the trained algorithms in machine learning calculates the failure rates as an input for the risk matrix.

For the implementation of the CPS framework, further work would require a detailed road map for Norwegian industry based on findings in demonstration of the CPS framework as well as involvement with several Norwegian companies within manufacturing and process industry.

ACKNOWLEDGEMENT

The authors wish to thank for valuable input from the research project CPS-plant.

REFERENCES

Batran, A., Erben, A., Schulz, R. & Sperl, F. (2017) *Procurement 4.0: A Survivial Guide in a Digital, Disruptive World,* Frankfurt, Campus Verlag.

Chaâri, R., Ellouze, F., Koubâa, A., Qureshi, B., Pereira, N., Youssef, H. & Tovar, E. (2016) Cyber-physical systems clouds: A survey. *Computer Networks,* 108, 260–278.

EFFRA (2013) Factories of the future: Multi-annual roadmap for the contractual PPP under Horizon 2020.

Eleftheriadis, R. & Myklebust, O. (2017) Industry 4.0 and Cyber Physical systems in a Norwegian industrial context. *IWAMA2017.*

European Commision (2016) TOPIC: Novel design and predictive maintenance technologies for increased operating life of production systems

Geissbauer, R., Schrauf, S., Koch, V. & Kuge, S. (2014) Industry 4.0 – Opportunities and Challenges of the Industrial Internet.

Hehenberger, P., Vogel-Heuser, B., Bradley, D., Eynard, B., Tomiyama, T. & Achiche, S. (2016) Design, modelling, simulation and integration of cyber physical systems: Methods and applications. *Computers in Industry,* 82, 273–289.

IEC (2017) IEC PAS 63088: Smart manufacturing – Reference architecture model industry 4.0 (RAMI4.0). *Publicly Available Specification Pre-Standard.* Switzerland, The International Electrotechnical Commission.

Industrial Internet Consortium (2016) Cooperation Among Two Key Leaders in the Industrial Internet.

Kagermann, H., Wahlster, W. & Helbig, J. (2013) Recommendations for implementing the strategic initiative Industrie 4.0.

Lee, E.A. (2008) Cyber physical systems: Design challenges. Proceedings – 11th IEEE Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing, ISORC 2008.

Lee, E.A. (2010) CPS foundations. Proceedings – Design Automation Conference.
Lee, J., Bagheri, B. & Kao, H.A. (2015) A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters,* 3, 18–23.
Lee, J., Jin, C. & Bagheri, B. (2017) Cyber physical systems for predictive production systems. *Production Engineering,* 11, 155–165.
Liebstückel, K. (2014) *Plant Maintenance with SAP – Practical Guide,* Bonn & Boston, SAP Press.
Mckinsey& Company (2015) Industry 4.0 – How to navigate digitization of the manufacturing sector.
Ministry Of Trade Industry And Fisheries (2017) The Industry – greener, smarter and more innovative (in Norwegian: St.meld. nr 27 (2016–2017) – Industrien – grønnere, smartere og mer nyskapende). *Report to the Storting.*
Monostori, L. (2014) Cyber-physical Production Systems: Roots, Expectations and R&D Challenges. *Procedia CIRP,* 17, 9–13.
Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., Sauer, O., Schuh, G., Sihn, W. & Ueda, K. (2016) Cyber-physical systems in manufacturing. *CIRP Annals,* 65, 621–641.
Myklebust, O. (2002) Enterprise modelling supported by manufacturing systems theory, Trondheim, NTH.
Nakajima, S. (1989) TPM development program: implementing total productive maintenance, Cambridge, Mass., Productivity Press.
PWC (2017) Predictive Maintenance 4.0 – Predict the unpredictable.
Rajkumar, R., Lee, I., Sha, L. & Stankovic, J. (2010) Cyber-physical systems: The next computing revolution. *Proceedings – Design Automation Conference.*
Rausand, M. & Høyland, A. (2004) System reliability theory: models, statistical methods, and applications, Hoboken, N.J., Wiley-Interscience.
Rolstadås, A., Andersen, B. & Schjølberg, P. (1999) *Produksjons- og driftsteknikk,* Trondheim, Tapir.
Rødseth, H., Myklebust, O., Eleftheriadis, R. & Schjølberg, P. (2016a) Improving maintenance by profit indicators. *Advanced Manufacturing and Automation V.* WIT Press.
Rødseth, H., Schjølberg, P. & Larsen, L.T. (2016b) Industrie 4.0 – A new trend in predictive maintenance and maintenance management. *EuroMaintenance 2016 – Conference Proceedings.* Artion Conferences & Events.
Rødseth, H., Schjølberg, P. & Marhaug, A. (2017) Deep digital maintenance. *Advances in Manufacturing*.
Sintef & Oreda (2009) OREDA: offshore reliability data handbook: Vol. 1: Topside equipment, Trondheim, OREDA Participants.
Smart Industry (2017) Smart indudtry – Durch Industry Fit For The Future.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Integrated analysis system for elevator optimization maintenance using ontology processing and text mining

M. Nagasaka & M. Sato
Corporate Research and Development Center, Toshiba, Japan

E. Kinoshita
Toshiba Elevator and Building Systems Corporation, Japan

ABSTRACT: In recent years, analysis methods using sensor data and record data acquired through the provision of maintenance services for social infrastructure equipment have attracted considerable attention. We focus on optimal replacement of elevator components. Features of elevators include that they move continuously for a long time without any operator and their proportion is high among social infrastructure equipment. We define five steps in the analysis of maintenance services for social infrastructure equipment. We have developed an analysis system consisting of life-limited component analysis, replacement planning simulations, and service performance analysis. The analysis system uses a combination of functionality of machine learning, such as ontology processing, text mining, and facility-type clustering in order to handle various types of facility data.

1 INTRODUCTION

In recent years, analysis methods using sensor and record data acquired through maintenance services for social infrastructure equipment have attracted considerable attention (Mobley, K. 2008, Narayan, V. 2004). Generally, social infrastructure equipment such as elevators and escalators require inspections and repairs by experts with special knowledge and skills. Experts visiting a site to maintain equipment usually prepare maintenance reports. In this paper we consider maintenance service data accumulated from elevators for a period of over twenty years.

Elevators are in continual operation without operators and are a very common form of social infrastructure equipment. Since there are various types of elevators installed in various building environments, it is important to consider these variations when analyzing the elevator data. Experts have compiled large datasets while providing elevator maintenance services. Elevators have thousands of components and database formats change over time, and there are also variations in how individual experts record maintenance data.

The remainder of this paper is organized as follows. In Section 2, we describe an integrated analysis system for maintenance optimization. In Section 3–5, we describe the elements of the system. Finally, we present conclusions in Section 6.

Figure 1. Integration of analyses for maintenance optimization.

2 FAILURE ANALYSIS

We define five steps in the analysis of maintenance services for social infrastructure equipment: (1) function diagnosis, (2) identification of factors that account for irregularities, (3) irregularity prediction, (4) maintenance planning based on predictions, and (5) verification of facility performance. The Plan–Do–Check–Act (PDCA) cycle is a widely known method for facilitating management tasks.

This study aims to improve maintenance services by developing data mining methods and simulation methods that consider the five analysis steps and the varied forms of elevator maintenance data. We pay particular attention to replacement planning for elevator components. We have developed an analysis system consisting of life-limited component analysis (Yano, T. et al. 2013) with statistical

survival analyses (Nelson, W. 2011), a naive Bayes model (Hsu, C.N. et al. 2000), and mixed survival analyses; replacement planning simulations with facility-type clustering; and facility and service performance analysis with a naive Bayes model and ontology processing using maintenance record data and troubleshooting data. To accommodate various facility data, the analysis system combines machine learning with functions such as ontology processing, text mining, and facility-type clustering.

3 CONSTRUCTION OF SURVIVAL MODEL

This section describes failure analysis in our system for diagnosing component failure, identifying equipment attributes that affect failures, and predicting failures. Survival analysis is generally known as a statistical modeling method for irregularities, commonly used in medicine, reliability engineering, and other fields. From the survival model of a component, we can calculate the probability of survival at a future time or after some number of uses following installation. Constructing a survival model requires "right-censored data," in which conditions are specified when the components are replaced. We used survival analysis to construct survival models of elevator components from maintenance records in order to apply those models to prediction of component failure. The maintenance records were composed of equipment master data, troubleshooting data, and regular operation data. However, to specify the conditions under which components were replaced, it was necessary to use handwritten reports of troubleshooting data, which used a variety of expressions to some extent.

To collect troubleshooting data associated with a target component, we use a naive Bayes model often used for text mining. Figure 2 shows a sample of troubleshooting data, which includes the date of occurrence, replacement component ID, trouble classification, treatment classification, and a detailed description. To allow the model to search for motor failure in troubleshooting data, we added "trouble classification" and "treatment classification" categories to the learning data as explanatory variables. Additionally, keywords such as "motor" and "electric mt" (mt: motor) in "detail description" are automatically extracted from maintenance reports in the troubleshooting data and added to the training data. A naive Bayes model is constructed from the data using positive and negative examples. The models can be used to search data sorted from high to low relevance and to select from among them based on a probability threshold parameter. Figure 2 shows how processing occurs. The naive Bayes failure search model thus uses threshold parameters to collect various component failure data.

We next describe construction of component survival models. The upper part of Figure 3 shows right-censored data for survival analysis, where the target component is replaced twice in equipment 1 and once in equipment 2. In the figure, "N" means normal and "F" means component failure at replacement. "N" is collected from regular operation data and "F" is collected using failure search model from troubleshooting data. The short line on the left indicates the start of maintenance service, and the line on the right indicates the analysis date on which the data were collected for survival analysis. There are only three cases (solid lines) in the component replacement database. Right-cen-

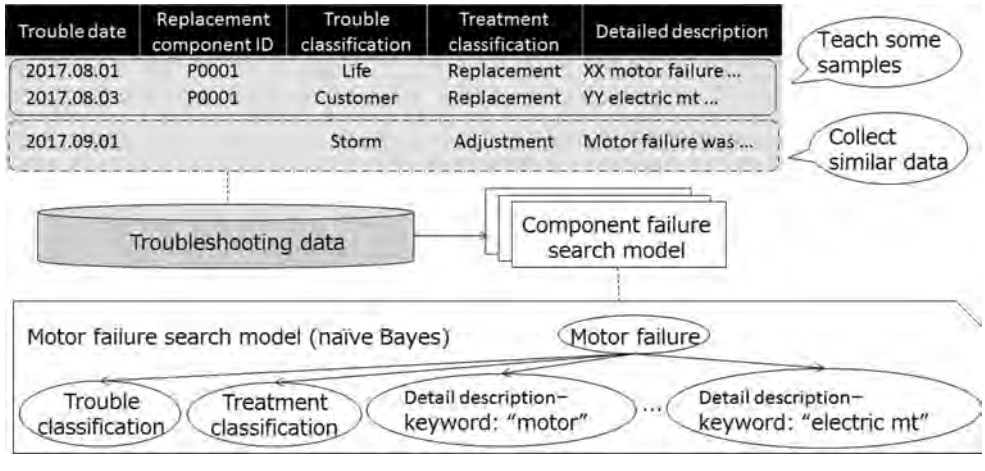


Figure 2. Component failure search model using text mining.

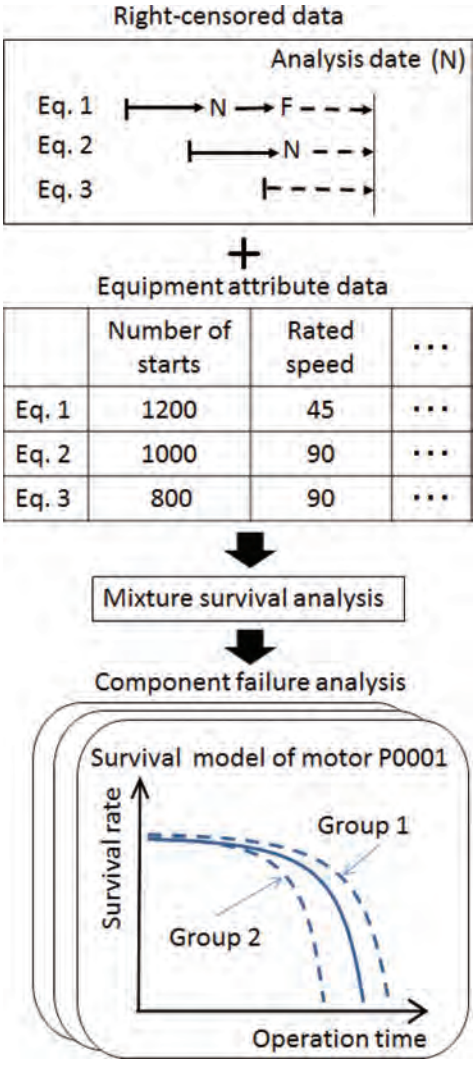


Figure 3. Censored data with attribute data, and survival models.

sored data must include six cases in addition to the above three cases, as indicated by the broken line. These three cases are assumed to be under normal conditions when we analyze the data. The solid curve in Figure 3 shows a survival model obtained from these right-censored data.

We adopt mixture survival analysis for components with many failure cases. Figure 3 shows differing equipment conditions such as number of starts and rated elevator speed. In this case, no components deteriorate over time over a fixed period. We thus split equipment into groups and create a component survival model for each group. A group consists of components with higher tendency to fail.

As described above, we specify component failure according to similar maintenance reports extracted using a naïve Bayes model. We generate right-censored data and conduct mixture survival analyses, and construct a survival model to predict irregularities and to identify factors that account for irregular functioning. We constructed about 100 fault-search models and about 2000survival models using the data mining methods.

4 MAINTENANCE PLANNING SIMULATION

This section describes the maintenance simulation used in our analysis system for maintenance planning. For maintenance planning using component survival models, we adopted simulation for estimating equipment maintenance costs and failure rate. This simulator should be able to accommodate various equipment types and components. Obtained simulation results can be applied to component replacement planning or budget management at individual branch offices.

Figure 4(a) shows an actual system in which users can visually grasp survival models. The left side of Figure 4(a) lists elevator type. When a user selects an elevator type, a list of components that make up the selected type is shown on the right. When the user selects a component from the list, a maintenance rule editor for that component is shown in the center. They can configure the maintenance rules for obtaining the desired simulation results and can set acceptable component replacement years through trial and error. The survival models of individual components are shown on the right. Component IDs and component names are shown in the maintenance rule editor. Users can specify the elevator types to which a rule applies. They can also configure replacement operations, component acquisition costs as replacement cost, and maintenance type, and can perform queries using elevator attributes or replacement years. By executing simulations after configuring the above items, lifecycle maintenance costs and component failure rates are calculated using the survival models, component ontology, and the elevator master data containing the attributes. Relations between equipment types and mounted components are organized as components in ontology data. Users can configure maintenance rules for obtaining the desired simulation results and can set acceptable component replacement years through trial and error. Other values (1) indicate maintenance costs for preventive maintenance:

$$\text{Replacement cost} = \text{Acquisition cost} + \text{Replacement operation cost} \quad (1)$$

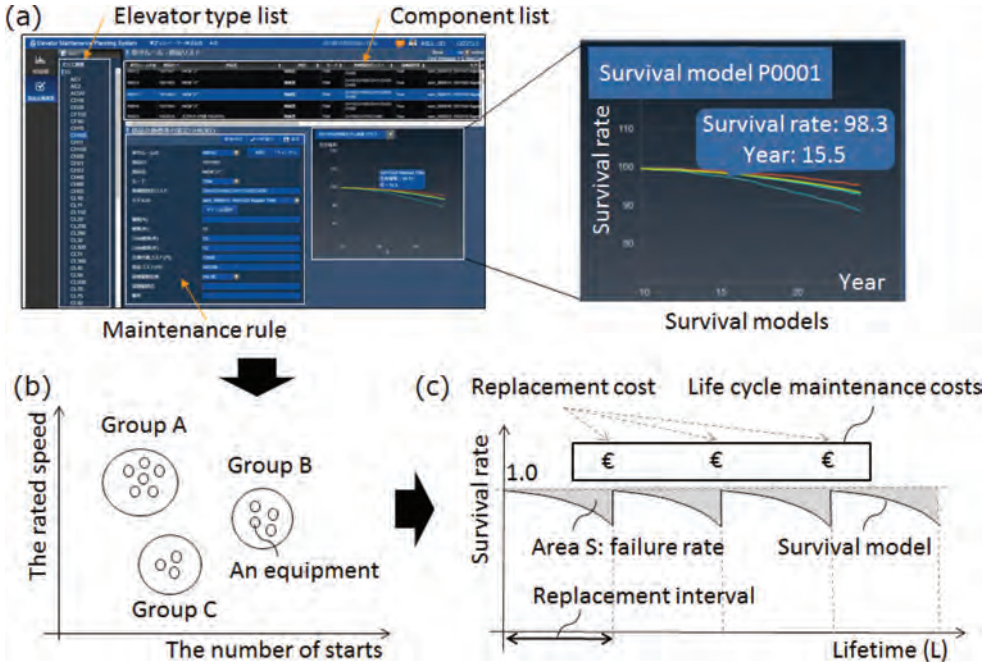


Figure 4. Life-cycle model and maintenance plan simulation. (a) Maintenance planning simulator. (b) Equipment grouping. (c) Life-cycle simulation.

To calculate simulation results rapidly for various elevator types, we previously used elevator master data to find groups with similar attributes and adopted these patterns for the principal elevator types. Figure 4(b) shows the simulation process for maintenance planning using equipment groups. When providing a simulation, it is necessary to know in advance the number of elevators belonging to a group and the patterns of principal types. It is also necessary to calculate the number of components attached to each elevator and to multiply this by the number of elevators to calculate the maintenance costs. After generating patterns for principle types, we proceed to life-cycle simulation. By using the above methods, we can perform simulation faster than by calculating the data for the elevator.

Figure 4(c) shows the calculation of life-cycle maintenance costs and component survival rates using the survival model. In the graph, the horizontal axis represents elevator life-cycle time and the vertical axis represents component survival rate. As the graph shows, the component survival rate gradually decreases. Area S represents the component failure rate, and dividing this by lifetime L gives the annual component failure rate. Costs are incurred at each replacement interval set by the users according to the maintenance rules. Life-cycle maintenance costs represent total costs.

Users can configure complex maintenance rules in the mixture survival model using attributes such as number of starts, and rated speed. Many elevator types require treatment of operation data from more than one hundred thousand elevators. There are too many elevator types to perform exact simulations for all types because of the amount of time needed. Therefore, we prepared about 10–20 typical groups using ontology processing to represent elevator types, principle patterns, and the number of standard components by elevator type. Users can configure the maintenance rules for obtaining the desired simulation results and can set acceptable component replacement years through trial and error. Utilizing these functions, organization staff can analyze information associated with component replacement based on actual maintenance data to determine repair and maintenance strategies. This is expected to aid with drafting plans for better maintenance services. By applying operation data such as those shown in Figure 4, users can analyze and update current maintenance records every day.

5 MAINTENANCE KEY PERFORMANCE INDICATOR ANALYSIS

As the previous section shows, we have focused on elevator component replacement planning

that uses cost and failure rate simulations based on a survival model. This section describes Key Performance Indicator (KPI) analysis applied to maintenance in our analysis system for checking facility performance of maintenance services to verify the consequences of a component replacement plan through simulation. Performance indicators need to include the number of component issues that arise when maintenance rule are changed. Down time, maintenance costs, and workloads are frequently used service indicators for maintenance. We developed analysis functions to calculate trouble incidence rates and component troubleshooting time.

Figure 5 shows the architecture of the maintenance KPI analysis system. These KPI analysis functions use maintenance records such as operating data, troubleshooting data, and equipment master data. Maintenance staff record regular inspections and maintenance reports along with troubleshooting reports describing situations of urgent site visits for unplanned maintenance. These data include handwritten reports as described in Section 3. Because there are various elevators types with tens of thousands of components each, the data analysis requires summarization methods. For this purpose, we prepared various ontology processes and failure search models using naive Bayes modeling.

Figure 6 shows a sample of the ontology processing function. The central table lists the replaced components. We created an equipment component ontology to determine equipment and components with higher tendencies for failure. The ontologies express thousands of component groups in four stages, allowing users to summarize large classifications such as "machine room" or "equipment type" or sub-stages lower in the hierarchy. This allows analysis of various component types among similar groups.

We also introduced a failure search model using a naive Bayes model to treat various handwritten texts. We showed a naive Bayes model for searching through troubleshooting reports of motor failure in Figure 2, and also adopted the model for searching for similar data in the maintenance KPI analysis system because the analyzed data includes handwritten reports. For example, natural disasters such as earthquakes have an effect on the operation of elevators, but are not a diagnosis of the operation. Therefore, we also constructed a naive Bayes model that can treat "earthquake" as a positive example. In fact, the ability to search for incidents of natural disasters allows users to analyze rates of incidence data and remove data that are related to natural disasters. Switching components to types with a longer life also decreased the rate of component failure.

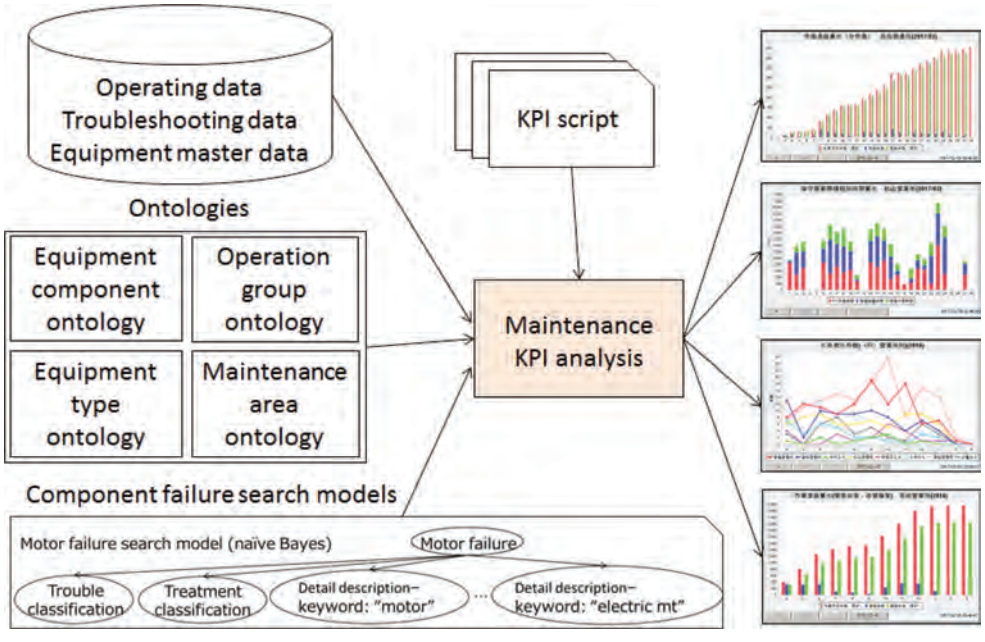


Figure 5. Maintenance KPI Analysis System with ontologies and component failure search models.

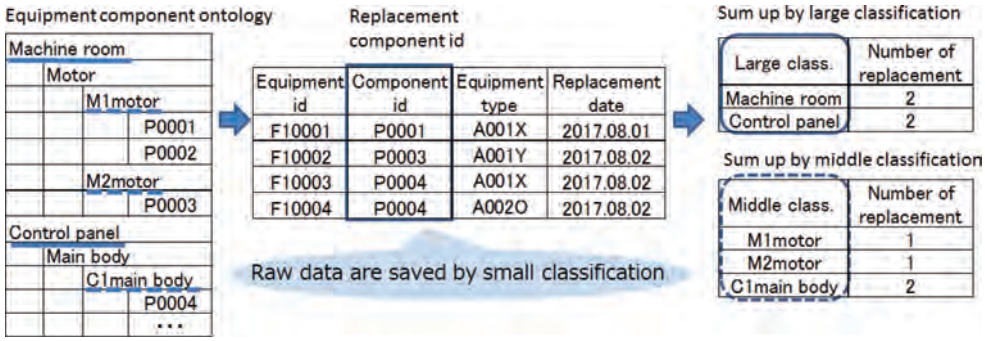


Figure 6. Adopting ontology data.

As stated above, we developed a maintenance KPI analysis system that prompts for updates to the policies and survival models. System functions include analysis of each group using ontology data and similar search models using naive Bayes models. At present, the system uses not only survival analysis components, but also visualizations of various maintenance service circumstances for optimization of field maintenance tasks. It calculates about 1000 KPIs for a few hours every night.

6 CONCLUSION

This paper described algorithms applied to elevator analysis systems for prediction-based maintenance planning. We described construction of failure analysis using a component survival model, and maintenance simulation that utilizes the failure analysis. In addition to the simulation, we also added a maintenance KPI analysis that optimizes maintenance by incorporating long-term maintenance records. The system is designed to include features such as application to various equipment types and long-term operation for social infrastructure equipment. In future research, we intent to investigate replacement planning based on maintenance planning simulations and elucidate factors in component degradation using maintenance KPI analysis with the aim of constructing and adopting an advanced survival model.

REFERENCES

Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. *Journal of machine learning research* 3: 1289–1305.

Hsu, C.N. & Huang, H.J. & Wong, T.T. 2000. Why discretization works for naive Bayesian classifiers. *Proceedings of the 17th ICML*: 399–406.

Mobley, K. 2008. Maintenance engineering handbook, 7th edition. McGraw-Hill professional.

Narayan, V. 2004. Effective maintenance management: Risk and reliability strategies for optimizing performance. Industrial Press Inc.

Nelson, W. 2011. *Applied life data analysis.* Wiley Interscience.

Yano, T. & Sato, M. & Kinoshita, E. 2013. Components replacement decision by survival analysis based on categories of elevator operational parameters. Safety, reliability and risk analysis: Beyond the horizon. CRC Press/Balkema: 697–702.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Safety enterprise architecture approach for a railway safety management system

S. Khan
Riyadh Metro, Hill International ME, Riyadh, Saudi Arabia

C. van Gulijk
Railway Research Institute, University of Huddersfield UK

ABSTRACT: This paper finds an extended set of railway Safety Management System (SMS) goals to manage safety compliance, technological complexities, operational uncertainties and business objectives in a holistic perspective using Enterprise Architecture (EA) approach. It also presents a comparative analysis of EA frameworks for implementation of a railway SMS. We call this system Safety Enterprise Architecture System (SEAS). In the technique, a set of selected capabilities based on the proposed railway SMS goals and system requirements are evaluated from legal, business and information systems perspectives. The SEAS approach establishes that the prevailing new technologies, evolving social realities and changing market dynamics has made the case of railway safety management more complex. Alongside regulatory compliance, it should also mandate explicit goals of standardization, innovation, business transformation, and IT alignment at all levels. It was found that at present not a single EA framework exists that suits the target SEAS approach entirely but the open group standard TOGAF comes closest.

1 INTRODUCTION

The Enterprise Architecture is a strategy to develop an IT backbone to support the safety infrastructure that comprises of railway's people, processes and systems in order to align safety objectives, business requirements and their complex interfaces through a structured IT system. For adequate application in railway safety, the architecture should cover all aspects of safety management by providing a 'whole system' approach to design, plan, delivery and control as part of a company business (Lodzinski et al. 2007 and RTS & TSLG. 2012). It should enable safety compliance, facilitate strategic decision making, support innovation and realize an effective business transformation (NAS 2014, Aier et al. 2016 and Cooney & Paxton 2016).

The innovation and business transformation for safety management systems are part of a continuous improvement process for Railway Undertakings (RUs) and Infrastructure Managers (IMs) alike, but it remains fundamentally based on their existing business functions and operations to remain in conformity with the legal requirements (Lodzinski et al 2007).

This paper describes an EA strategy to facilitate the IT business transformation in the organizational pursuit to institutionalized safety as inbuilt quality. The section 2 of the paper gives the EA overview, discusses some popular EA frameworks and explores railway safety management background. Section 3 list down the target SMS goals. Section 4 discusses criteria selection to examine the proposed strategy. Section 5 is the comparison table of some carefully chosen EA frameworks against the selected capabilities. Section 6 discusses the result and section 7 ends with concluding remarks.

2 BACKGROUND

2.1 *Enterprise architecture background*

The comprehensive introduction of EA terminology is down to J. A. Zachman from his paper title "A Framework for Information Systems Architecture.", published in the IBM systems journal where he presented his Zachman grid to provide EA ontology (Zachman 1987). An EA framework presents a skeletal structure to define suggested architectural artifacts, their relationships and characteristics. Every EA framework typically embraces a reference enterprise architecture, a planning and implementation methodology, guidelines tools and common vocabulary (Ahlemann et al. 2012). For various perspectives, several alternative EA frameworks have been proposed and evaluated by academics and practitioners, each with different scope and activities, to support all aspects of EA lifecycle (Dang & Pekkola 2017, Nikpay et al. 2015, Susanne

& Zellner 2015, Rouhani et al. 2015, Odongo et al. 2010, and Urbaczewski et al. 2006). The case of EA application to railway safety demands appropriate framework selection to ensure safety compliance in the very changing environment of technological complexities, operational uncertainties and evolving social realities.

Previous studies recognized that railways operational safety depends on several data-dependent systems including signaling, infrastructure management, rolling stock, organizational safety, culture and human factors. Many accident investigators agree that lack of linkage in between these systems contributed to the substandard control of multi-causal accidents (Kyriakidis et al. 2012, Silla & Kallberg 2011, Evans 2011 and Baysari et al. 2008). The studies also highlighted the weakness of railway SMSs in implementation especially in the area of information management, organizational structure, role & responsibility and competence management. It also observed significant deficiency of processes for design and improvement related to risk assessment (Shang et al. 2017). Ultimately, EA paves the way for better integration of overall data on the railway through alignment of its business with IT infrastructure.

2.2 *Enterprise architecture frameworks*

This paper assesses four major EA frameworks Zachman, TOGAF, DODAF and FEAF. They are selected on the basis of their popularity and customization. They are described in brief below.

2.2.1 *Zachman framework*

Zachman Framework (Zachman 1987, Sessions 2007) provides a concise way to structure and model enterprise architecture. It is a two-dimensional grid with a set of six perspectives or views and five basic interrogatives. A perspective constitutes a row that depicts role for a stakeholder of the project team. The various stakeholders are the planner, owner, designer, builder, subcontractor and user. The columns of the grid represent a characterization of information for each perspective through what-data, how-function, where-network, who- people, when- timing and why- motive. The rows and columns thus intersect to create cells where each cell is resulting in an architecture activity depending on the system's aspect for a particular stakeholder.

Zachman framework is the most comprehensive of the EA frameworks which offer a series of views and visualization support as a planning tool to help better selections between the alternative options. However, it does not focus on the EA development and governance mechanism. It also could not provide appropriate software tools configuration and alignment to the innovation and new social realities. The framework does not explicitly offer support for nonfunctional requirements and system development lifecycle. The Zachman framework, although self-described as a framework, is more accurately defined as a taxonomy for organizing architectural artifacts.

2.2.2 *TOGAF*

The Open Group Architecture Framework (TOGAF) is very comprehensive with regards to actual process involved (Magoulas 2012, Rouhani et al. 2013). TOGAF's view of an enterprise architecture consists of business, application, data and technical architectures. The most important parts of TOGAF is the Architecture Development Method (ADM) for process development, the enterprise continuum for various architecture views and the knowledge base repository for resources, implementation guidelines, templates and background information.

It provides appropriate strategy and governance supports for the designing, planning and implementing of an architecture based on the enterprise requirements. It utilizes appropriate models for both IT and enterprise activities. TOGAF-ADM is a step-by-step process performed in creating EA and consists of a preliminary phase, followed by eight transformation phases that guide the users through various levels of architecture maturity in a managed manner through the transition. TOGAF is flexible and allows phases to be performed incompletely, skipped, combined, reordered, or reshaped as per the stakeholder's requirement to fit any organization's needs. TOGAF views the world of enterprise architecture as a continuum of architectures, ranging from highly generic to highly specific. It document design rationale to trace design and architecture decisions.

TOGAF is lacking instructions to clearly describe the output specifications of each development cycle and missing the organizational role and responsibilities. Although strong on business and architecture perspectives, it is short in detail from planning and maintenance aspects.

2.2.3 *DODAF*

The Department of Defense Architecture Framework (DODAF) is developed by US defense to provide a holistic support platform for its agencies goals transformation. Its overall orientation is different in solving the EA issues. DODAF (Odongo et al. 2010) uses model templates to collect and disseminate information data on a specific issue resulting in a view or perspective. In DODAF 2.0 there are eight prescribed perspectives. Its architecture development process consists of six steps of context definition, scope, requirements, perspectives, development and application.

DODAF in acceptable manner supports the concept, modeling and process phases of EA life cycle. Its overall characteristics in the target system implementation reside in the same area with TOGAF with little lower in grade.

The issue with the framework is that there is no complete governance guidance, social, financial and technical analysis is available to support the system objectives. The lack of enterprise integration and software configuration support contradict its application in a dynamic environment. Also, it offers limited support to nonfunctional requirements of consistency, design traceability and verifiability.

2.2.4 *FEAF*

Federal Enterprise Architecture Framework (FEAF) of US government is to facilitate the shared development of common processes and information among US federal entities (Odongo et al. 2010, Sessions 2007). FEAF is based on Zachman framework, but refers only to the first three columns that represent what-data, how-function and where-network respectively and focuses on the top three rows presenting various perspectives to provide standard terms and definitions through Business Reference Model (BRM), Components Reference Model (CRM), Technical Reference Model (TRM), Data Reference Model (DRM), and Performance Reference Model (PRM) with each have unique goals. It uses architecture analysis, architectural definition, investment and funding strategy and program management plan as a four-step process for creating an EA.

FEAF is the most complete of all the methodologies discussed here. It has both a comprehensive taxonomy, like Zachman, and an architectural process, like TOGAF. It supports vision, strategy and knowledge base repository for EA planning. It allows flexibility in the use of tools and has standard practices for interoperability.

FEAF is primarily a framework for architecture planning rather EA development and maintenance. A part from system security, FEAF does not explicitly support other non-functional requirements and offer limited support on plan validation and traceability.

2.3 *Safety management background*

Different countries developed different railway Safety Management Systems (SMS) based on the operational circumstances and legal requirements. In Europe, the systems are guided by the Common Safety Methods (CSM) to be used as a solid basis to support the design and implementation of their SMS. The approach thus resulted in a number of SMSs but a reasonable degree of agreement can be achieved on what a standard SMS must cover. It shall contain all the processes and procedures describing activities related directly or indirectly with railway safety both at an organizational and operational level.

Although recent technological advances and development of railway SMS guidelines and standards assist operators to perform their duties efficiently, railway accidents still occur due to a complex interaction of its components systems and interface management shared between many actors functioning in fragmented collections of standalone or silo-based systems and processes. These silo systems include both older in house developed systems and a number of best available in the market commercial off-the-shelf solutions which are selected to meet their local business needs (Aier et al 2016, Clayton 2010, Cooney & Paxton 2016). Therefore, to couple the overall business with the current intelligent communication systems, sophisticated and extensible customer systems with cross-silo integration components are purchased at high cost without any future roadmap but only to provide opportunistic solutions to keep running the system. This approach literally lacks a holistic enterprise perspective of central IT functionality. The future highly automated railway operations anticipate innovation with information systems for business agility and safety requirements, leading to standardization and integration, at all levels. It is the enterprise architecture, SEAS, which can provide the desirable platform to deliver the greatest business value through reliability, interoperability and safety using a whole system unification of enterprise entities, railway systems and people.

3 SAFETY ENTERPRISE ARCHITECTURE

To appraise the proposed SEAS approach for a railway SMS in the ever-changing economic, regulatory and technical environments involves many considerations. A typical safety management system is driven by key guidelines originating from the CSMs for supervision, monitoring, risk evaluation, conformity assessment, and national legislations for effective safety management. Key elements of railway SMS are listed below (EU Railway Safety Directive 2004):

- Staff commitment
- Transition or migration
- Information management
- Emergency and crisis management
- Compliance assurance
- Accident/Incident reporting
- Continual improvement
- Gradual and step by step implementation
- Roles and responsibilities

- Competence management
- Change management
- Tools for monitoring
- Risk assessment and evaluation
- Strategies to achieve targets
- Documentation
- Assets management and maintenance
- Internal audits
- Standard Glossary

Implementation of these elements is essential for regulatory compliance and a railway SMS should establish specific targets and formulate relevant assessment criteria for achieving these targets to ensure the realization of safety complaint operations. A typical railway SEAS, in addition to these legal requirements, should achieve a number of other goals listed below as part of its strategic plan for a financially beneficial railway (NAS 2014, Tang et al. 2004, Aier et al. 2016, RTS & TSLG 2012, Lange & Mendling 2011, Lim et al. 2009, Magoulas et al. 2012, and EU S2R 2015). It will result in a blueprint for Safety Enterprise Architecture. Ideally, this blueprint should support the SMS independent of railway system geography, industry size, operational domain and architecture style.

- Regulatory Compliance
- Transparency
- Complexity Management
- Business-IT Alignment
- Configuration Management and Control
- Harmonization & Standardization
- Business Transformation
- Innovation
- Risk Control Measure for all risks associated with IM/RU, maintenance and assets management, subcontractors and other parties, reputational risks.
- Architecture Maturity
- Agility
- Soft aspects
- Standard Glossary

4 METHOD FOR CRITERIA SELECTION

This research is divided into foundation ground work, system requirements and SEAS capabilities identification and evaluation. For the purpose, a systematic literature review of relevant research articles was carried out. A set of research questions related to SEAS capabilities and framework selection are formulated and discussed with railway experts.

RQ1: What system aspects should be considered for SEAS?

In order to answer the first question, the SEAS can be divided into the following five aspects:

- Operational Safety: Railway SMS external goals are regulatory driven. The IM/RUs should fulfill community requirement to ensure safe railway operations through continuous improvement, adopting a system-based approach and assigning of responsibilities (EU railway safety directive 2004, Lodzinski et al. 2007, ORR 2015).
- Business Support: These are related to railway SMS internal goals related to cost reduction, new business initiatives, improve decision making, and risk management (Aier et al. 2011 and Lange & Mendling 2016).
- Information Technology System: This aspect is related to organization information resources, information systems, architecture tools, implementation models, software configuration and operating platforms (Urbaczewski & Steven 2006).
- Qualitative or Non-functional: It is the system ability that the characteristic of its offered service or a product satisfies the user's requirements. Therefore in railway SMS case we have to define the user's demands to understand the quality. Under the quality or non-functional aspects the following items are collected for EA framework analysis: interoperability, flexibility, reusability, scalability, standardization, Alignment, reduce risk, reduce complexity, integration, better and faster service, communication, innovation (Lim et al 2009).
- Social and Cultural: Safety has relative levels in relation to different situations and environments. Also the increasingly unreliable and demanding customers require through consideration in to which product or services to offer at any given time to ensure safety in operation as well mitigate reputation risks (Magoulas et al. 2012).

RQ2: What capabilities the proposed railway SEAS should have?

From the SEAS goals listed in section 3 in conjunction with required system aspects discussed in question 1 above, forty-two (42) capabilities are identified. The system should attain these capabilities through the best practices implementation by achieving set targets through gradual transformation and improvement of the system (Zamermann et al. 2015, Aier et al. 2016, Tung et al. 2004, Nikpay et al. 2017).

RQ3: Which of the EA frameworks is more relevant for railway SEAS implementation?

The question deals with the already established EA frameworks in various categories. These include Zachman from commercial frameworks, Open Group Architecture Framework (TOGAF) from open group frameworks, FEAF from government frameworks and Department of Defense Framework (DODAF) from defense frameworks (Jacco 2014, Nikpay et al, 2017). TOGAF is capa-

ble to be used as a general framework in any enterprise with modifications and therefore replaces GERAM and RM-ODP in that category. FEAF is the complete methodology with ZF-like classification and TOGAF like structural design process and excels over the TEAF in the federally developed frameworks category (Tang et al, 2004). DoDAF version 2.0 is an evolution of the C4ISR and NATO framework due to a limited application was dropped and replaced by DODAF.

Table 1. Comparison of EA Frameworks for SEAS.

CAPABILITY	ZF	TOGAF	DODAF	FEAF
Strategic Vision and Goals	H	H	M	H
Architecture definition	M	H	H	H
Architecture Development Process	M	H	H	M
Process Completeness	L	H	M	M
Architecture Verifiability	L	H	L	L
Transition Strategy and Plan	L	H	H	H
Tool Support	H	M	M	M
Information Reference Resources	H	H	L	L
Standardization	L	H	H	M
Step by Step guidelines	L	M	M	M
Continual	L	H	L	L
Architecture Evolution Support	L	H	H	H
Governance guidelines	L	M	L	M
Integrating enterprise systems	H	H	M	H
Business Alignment	H	M	M	L
Socio Cultural Alignment	M	M	-	-
Functional Alignment	L	H	-	-
Structural Alignment	L	M	-	-
Infological Alignment	L	L	-	-
Contextual Alignment	L	M	M	M
Dynamic (Innovative)	L	L	L	L
Business Model	H	H	H	H
System Model	H	H	H	H
Information Model	H	H	H	H
Computational Model	H	H	H	H
Software Configuration Model	L	H	L	L
Software Processing Model	H	H	H	H
Implementation Model	M	H	H	M
Platforms	H	H	H	H
Nonfunctional Requirements	M	H	M	M
EA Security Issues	L	M	M	L
Views/perspectives	H	L	M	M
Abstractions	H	L	M	M
System Development Lifecycle	L	M	H	H
Change Management	H	H	H	M
Maintenance Process	L	M	L	L
Costing and Vendor Supporting	M	H	H	L
Meta model	M	L	M	M
Maturity Model	L	M	M	M
Enterprise Knowledge Base	L	H	H	H
Taxonomy Completeness	H	M	M	M
Vendor Neutrality	M	H	L	M

RQ4: How different frameworks are compared?

Three relative conformance levels (High, Medium and Low) based on the support each framework support to a capability are developed. If a framework explicitly supports a capability element in the table 1, it is marked as High (H), if partially support then marked Medium (M) and Low (L) in case of little or no support at all.

5 EVALUATION OF EA SYSTEMS

A SEAS capabilities list was developed as summarized in table 1 below based on the already done research in this area and produced here in citations. It provides a high level comparison and analysis of the selected EA frameworks.

From Table 1 outcomes of various frameworks are summarized based on the support they offer to the capabilities. In the current analysis, only considering the clearly supportive option (H) for suitable EA framework selection, ZF accrued 15 Hs, TOGAF has 30 Hs, DODAF has 16 Hs and FEAF has accumulated 13 Hs. In consequence, TOGAF, although does not meet overall criteria, emerged the most suitable preference.

6 DISCUSSION AND LEARNING

The current studies provide discussion platform for reconsideration of railway SMS goals and implementation approaches.

6.1 *Revamping railway safety management*

The current studies introduce business aspects and soft aspects of the railway safety management pursuit of attaining whole system approach. The typical safety management systems currently in sway in rail and metro industry rely on hard aspects to focus on safety compliance in supervision, monitoring, maintenance and improvement within their existing operational boundaries which is an inappropriate and outdated approach in the very changing environment. In today's world, the railway like all other safety-centric industries especially to mention, airline and space industries, is changing hands from public to privately operated corporations. The underlying drivers behind these handing overs are fundamentally based on business objectives. The market competition and survival threat compel every entity to deliver as per customer satisfaction with excellence in operations where safety matters at the center. We therefore suggest revamping of the European railway agency SMS wheel to cater for business objectives and soft

aspects in order to align with new social realities and market dynamics.

6.2 *EA framework selection*

EA implementation can be independent of any framework however it provides a roadmap and tools to capture requisite information and model them to avoid any panic. The results of the above comparison in perspective of railway SEAS objectives indicate that each enterprise methodology has its strengths and weaknesses and none of them is categorically complete. The Meta model, maintenance, dynamic and soft aspects are the issues which are not supported by all selected EA frameworks.

7 CONCLUSION AND FUTURE WORK

This paper presented EA application to an implementation of a railway SMS in order to develop SEAS. The new approach which is based on holistic perspective is to partner the safety regulatory requirements and business objectives of RU/IMs in a business transformation support in the face of system complexity, operational uncertainty and new social realities. The purpose of the paper is to evaluate different well- known EA frameworks methodologies to narrow down on the selection in order to support a future railway SEAS that is capable to shape norms and values of railway organizations. It also attempt to establish EA as the definitive solution and formulate parameters of a standard SEAS to ensure its optimal realization. Any of the existing EA frameworks can provide guidance for EA implementation to a certain degree and TOGAF scores higher in the case due to a number of distinctive capabilities. It uses architecture development in incremental step by step guidelines and supports the enterprise architecture modeling with detailed descriptions in order to effectively address critical business needs. The methodology also establishes a common set of risk and security concepts and demonstrates mapping to the implementation languages and tools leading to build a shared understanding and response strategies to ensure traceability and verifiability. Nevertheless it is not complete enough to meet the railway RU/IMs specific requirements in the area of business and socio-cultural alignment, risk management implementation and innovations support.

The SEAS goals set in this studies are not final and the related capabilities should evolve with architecture maturity. In the current assessment scheme all the capabilities are given equal weight irrespective of their risk exposure and manageability. In future studies some type of quantitative analysis will be considered based on relative scoring technique for EA framework selection.

REFERENCES

Ahlemann, F. Messerschmidt, M.E. & Legner, C. 2012. Strategic Enterprise Architecture Management: Challenges, Best Practices, and Future Developments. Springer Heidelberg Dordrecht London New York.

Aier, S. Buckl, S. Gleichauf, B. Florian, M.F. Schweda, M.C. & Winter, R. 2011. Towards a More Integrated EA Planning: Linking Transformation Planning with Evolutionary Change. 4th International Workshop on Enterprise Modelling and Information Systems Architectures (EMISA 2011). 190. 23–36.

Aier, S. Weiss, S. Winter, R. & Rytz, B. 2016. Untangling EA's Long Path of Becoming a Partner for Business Transformation: The Case of Swiss Federal Railways. 20th IEEE- Enterprise Distributed Object Computing Workshop (EDOCW). 91–97.

Baysari, T.M. McIntosh, S.A. Wilson, R.J. 2008. Understanding the human factors contribution to railway accidents and incidents in Australia. Accident Analysis and Prevention. 40. 1750–1757.

Clayton, J.R. 2010. Re-integrating Railway Silos. IET Railway Network Young Professionals Best Paper Seminar.

Cooney, C.R. & Paxton, C.M. 2016. Strategic Enterprise Architecture Design and Implementation Plan for the Montana Department of Transportation- Final Report. US.

Dang, D.D. & Pekkola, S. 2017. Systematic Literature Review on Enterprise Architecture in the Public Sector. The Electronic Journal of e-Government. 15(2). 132–154.

European Union (EU)-Shift2Rail Undertaking 2015. Shift2Rail Strategic Master Plan (2015). 01.

Evans, W.A. 2011. Fatal train accidents on Europe's railways: 1980–2009. Accident Analysis and Prevention. 43. 391–401.

Jacco, R. 2014. The Relationship between Enterprise Architecture, Business Complexity and Business Performance. Master Thesis. University of Twenty.

Kyriakidis, M. Hirsch, R. & Majumdar, A. 2012. Metro railway safety: An analysis of accident precursors. Safety Science. 50. 1535–1548.

Lange, M. & Mendling, J. 2011. An Experts' Perspective on Enterprise Architecture Goals, Framework Adoption and Benefit Assessment. Proceedings of the 6th Trends in Enterprise Architecture Research Workshop.

Lim, N. Lee, T. & Park, S. 2009. A Comparative Analysis of Enterprise Architecture Frameworks based on EA Quality Attributes. ACIS international Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing. 10. 283–288.

Lodzinski, P.J. Mcdaid, L. Patacchini, A. & Salander, C. 2007. Position Paper on the Assessment criteria for Safety Management Systems of Part a Safety Certifications and Safety Authorizations. ERA Safety Unit.

Magoulas, T. Hadzic, A. Saarikko, T. & Pessi, K. 2012. Alignment in Enterprise Architecture: A Comparative Analysis of Four Architectural Approaches The Electronic Journal Information Systems Evaluation. 15(1) 88–101. Available online at www.ejise.com

National Academy of Sciences (NAS) - Washington DC US. 2014. Interim Report of a Review of the Next Generation Air Transportation System Enterprise Architecture, Software, Safety, and Human Factors.

Nikpay, F. Ahmad, B.R. Rouhani, D.B. Mahrin, M.N. & Shamshirband, S. 2017. An effective Enterprise Architecture Implementation Methodology. International Journal of Enterprise Information Systems. 11(2). 50–64.

Odongo, A.O. Kang, S. & Ko, I. 2010. Scheme for Systematically Selecting an Enterprise Architecture Framework. 9th IEEE/ACIS International Conference on Computer and Information Science. 665–670.

Office of the Railway Regulation (ORR) UK. 2015. Common Safety Method for risk evaluation and assessment: Guidance on the application of Commission Regulation (EU) 402/2013.

Rouhani, B.D. Mahrin, M.N. Nikpay, F. & Nikfard, P. 2013. A Comparison Enterprise Architecture Implementation Methodologies. International Conference on Informatics and Creative Multimedia. 1–6.

Rouhani, B.D. Mahrin, M.N. Nikpay, F. Najafabadi, M.K. & Nikfard, P. 2015. A Framework for Evaluation of Enterprise Architecture Implementation Methodologies. World Academy of Science, Engineering and Technology, 9(1). 1–6.

RTS & TSLG. 2012. The Future Railway: The Industry's Rail Technical Strategy 2012.

Sessions, R. 2007. A Comparison of the Top Four Enterprise-Architecture Methodologies. Available on: https://msdn.microsoft.com/en-us/library/bb466232(d = printer).aspx

Shang, W.U. Xiaocheng, G. & Yangun, L. 2017. A New Model for Safety Management System for Railway Operation. International Conference on Rail Transportation, China. (Unpublished). www.eprints.hud.ac.uk/id/eprint/32873

Silla, A. Kallberg, V. 2012. The development of railway safety in Finland. Accident Analysis and Prevention. 45. 737–744.

Susanne, L. & Zellner, G. 2006. Evaluation of current architecture frameworks. SAC- Dijon, France. 1546–1553.

Tang, A. Han, J. & Chen, P. 2004. A Comparative Analysis of Architecture Frameworks. Proceedings of the Asia-Pacific Software Engineering Conference, 11, 640–647.

The European Parliament. 2004. Railway Safety Directive 2004/49/EC. Official journal of the European Union.

Urbaczewski, L. & Mrdalj, S. 2006. A Comparison of Enterprise Architecture Frameworks. Issues in Information Systems. 7(2). 18–23.

Zachman, A.J. 1987. A framework for information systems architecture. IBM Systems Journal. 26(3). 276–292.

Zimmermann, A. Schmidt, R. Jugel, D. & Möhring. M. 2015. Evolving Enterprise Architectures for Digital Transformations. Digital Enterprise Computing. 183–194.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

A computer leaning approach to obtain safety information from multi-lingual accident reports

P. Hughes, M. Figueres-Esteban, R.A.H. El Rashidy & C. van Gulijk
Institute of Railway Research, University of Huddersfield, Huddersfield, UK

R. Slovak
Federal Office of Transport, Bern, Switzerland

ABSTRACT: Accident reports provide a valuable source of data for any safety management system. In multi-lingual jurisdictions, accident reports can be provided in more than one language. For example the Swiss transport authority collects accident reports that are written in either German, French, or Italian. The unstructured nature of free-text makes it difficult to extract information from large numbers of accident reports. Machine-reading of text is an emerging area of research, however there are few instances of information being extracted from text in more than one language.

This paper introduces an ontology-based interactive learning method between a human and computer software to identify safety-related information by analysing text written in three different languages. The results of the method were analysed by fluent speakers of each language, who rated the overall accuracy of the method to be 98.5%.

The method stores and processes the data in a NoSQL graph database, which provides a powerful tool to readily integrate the analysis with other data sources, for example train movement data, passenger census data, or even comparative data from other railways.

1 INTRODUCTION

A large amount of information that is useful to safety is contained in natural language text reports, for example accident reports, hazard reports, or safety audit reports. Whist these data sources can contain valuable information, it is not easy to extract the information. Human-reading of large amounts of textual data is slow and error-prone and, if the task is divided amongst a large number of readers, then differences in interpretation can occur. Machine reading of text is an emerging area of research which has the potential to provide useful information from large text sources, although there are still a number of problems to be overcome. No prior work has been found that has attempted to extract information from safety-related documents written in more than one language.

The Swiss Federal Office of Transport (FOT) collects textual data on safety incidents that occur on the nation's transport system. Switzerland is multi-lingual, and the text in the incident reports is provided in either German, French, or Italian. Each report contains information that could be useful to managing safety of the system, however no existing process is known whereby the information can be collated in a way that supports safety management.

This paper describes an ontology-based approach to obtain information from 5065 incident reports provided by the FOT. Incidents were classified into a number of categories including incidents that occurred: whilst passengers were boarding trains; whilst they were alighting trains; or as a result of passengers falling down stairs, caught by closing doors, or struck by falling baggage.

2 BACKGROUND

2.1 *Safety management of big data*

Modern approaches to safety management require organisations to collect information on accidents. It is very common for these data to include text that describes where and when the accident occurred; the context within which the accident occurred (for example weather conditions; activities that were taking place at the time); what injuries and damage resulted; what proximate and underlying causes led to the accident; what risk controls failed to allow the accident to occur; and recommendations to minimise and mitigate recurrence. The purpose of collecting such data is to support decision-making processes that consider information from different sources (for example safety-critical

work procedures, budget data, or legislative requirements) and take action to optimise safety management.

Professional safety management systems often collect information not only on accidents that have been observed, but also on incidents where safety risk controls have broken down but no injury or damage occurred, so called *near-miss* or *close call* events (Gnoni, Andriulo et al. 2013, Andriulo and Gnoni 2014, Macrae 2014). These accident and incident reports themselves can amount to a large quantity of data (Hughes et al., 2016a). Extracting information from these large data sources can be a challenge by itself; the problem is further complicated when the data is provided as free-text rather than structured machine-coded data. Combining data from such a large data source with data from other, potentially very large, data sources can be problematic. This data data management challenge is commonly referred to as *big data*. There is an emerging body of work describing the challenges of big data and techniques that can be used to extract useful information from the data. To obtain information that supports safety management, Van Gulijk et al. (2016) introduce the concept of Big Data Risk Analysis (BDRA), and describe the four *enablers* of BDRA:

- data and data-management,
- visualisation interface,
- analytics and software, and
- ontology and knowledge representation.

Figure 1 presents a schematic overview of the interaction of these enabling components. Each enabler is described below.

Data and data management is the initiating reason for BDRA. Modern organisations and their safety management systems collect large amounts of data with the intention of using these data to improve safety and safety management. Data may be collected from manual processes, such as workers completing forms as part of a safety-critical work process; from automatic systems, such as supervisory control and data acquisition (SCADA) systems; or from external sources, such as information on the internet regarding weather or traffic conditions. Collecting data on hazards, incidents and accidents is at the heart of modern safety management systems. The FOT has a database of thousands of reports describing all detected accidents on the transport network, including minor accidents, such as where a passenger fell over and sustained only very minor injuries. The data in the accident database is a valuable source of information that can be used to understand the causes of accidents and any underlying trends, and to determine actions that may minimise the likelihood of recurrence.

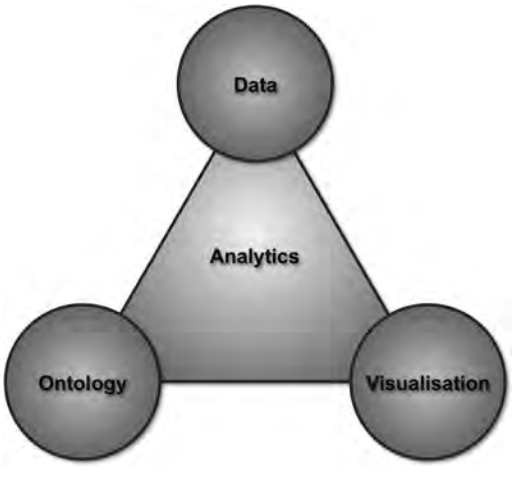


Figure 1. The four enablers of BDRA.

A *visualisation interface* is an essential component for understanding large data sources. Humans have a capacity for visualising concepts and their relationships, which is valuable for understanding big data sources which can contain thousands, or even hundreds of thousands of inter-related concepts. For example the safety management of an operating railway requires an understanding of concepts such as all types of rail vehicles including locomotives, carriages, wagons, and maintenance machines; all railway locations including track locations, stations, sidings, maintenance facilities; all organisations who operate within the railway including law enforcement organisations; users of the railway; routes and timetables *etc*. It is not feasible to expect staff who operate the safety management system to be able to visualise all these concepts and the multitudinous interactions between them without external support. Visual analytic tools facilitate the understanding of these concepts and include applications to help understand different categories of safety risks and the geospatial distribution of incidents and accidents. Such tools may even include visual causation models, such as bow-tie diagrams, that describe the chains of events that can lead to an accident and the risk controls that are in place to reduce the risk.

Analytics and software are the backbone of BDRA: modern data management systems are based on software and BDRA requires large software analytical capability such as that available on modern cluster computers or provided by cloud computing services. Several software services are required for BDRA in order to:

- store data in short-term stores and long-term archives;

- organise data into meaningful data units and transport it to the locations required for processing;
- pre-process data ready for analysis and format it as required;
- analyse the data to produce results that assist safety-related decision-making;
- collate results of the analytics and aggregate them as necessary;
- present the results—in a visual form—to analysts to allow understanding of underlying hazards, risks, controls, accidents and consequences;
- iterate through analysis depending on the findings of earlier analyses;
- store results of analyses to allow subsequent analysis to expand on the results of earlier work;
- oversee and coordinate all of the above processes and distribute computer resources in an optimal way.

The software tools to support these tasks may include traditional tools such as simple graphing tools and SQL databases, as well as technologies that have emerged in the past decade such as interactive visualisation tools, graph databases and massively parallel, distributed analytical tools.

Ontology is a structured method to classify entities within a domain and the interactions between them. An entity is any item that can have properties and interact with other entities. A simple example of an entity in a railway safety context are *trains* which have properties of rolling stock class, furthermore individual instances of trains have train numbers as properties. Trains interact with stations, track and passengers. Other examples of entities include tickets (which have properties such as valid dates and routes) and railway staff (who have properties such the job function and interact with people and organisations). Entities within an ontology may also be abstract, such as dates and times. Regardless of whether an entity is abstract or has a physical form, a defining feature of all entities is that they can be referred to by themselves; this differentiates them from *relationships* that require entities in order to be meaningful. An example relationship is that a passenger can *board* a train. Both the passenger and the train are entities that can exist by themselves. However the boarding relationship requires these entities to be present in order to be meaningful: boarding cannot occur by itself. The ontology provides a structure that allows data to be joined to the key entities that are relevant to safety management. For example as an instance of data, an accident report may be joined in an ontology to the station where it occurred; to the members of staff who responded; and to the date and time where it occurred. If the ontology contains an incident causation model, such as a bow-tie diagram, the accident report can also be linked to the particular risk control breakdowns that led to the accident and the outcomes that occurred. Individual stations in the ontology may be linked to the general concept of a station; individuals may be linked to the organisations for whom they work, and so on. In this way it is possible to structure queries of the data to request accident reports that occurred at particular stations, or at stations in general; or to identify staff responders from particular organisations. The ontology may also be linked to other data in the safety management system such as audit reports, or maintenance logs. In this way it is possible to identify accidents that occurred at stations where audits have not recently been performed; or accidents at locations where particular maintenance activities have occurred. Where the ontology contains dates or chains of causation, these entities can also be queried to extract precise and meaningful information to support safety management.

2.2 *Ontologies for data management and understanding*

Smith & Welty (2001) describe the traditional understanding of the word *ontology*, as being an ancient Greek concept that addresses the fundamental nature of reality. Aristotle established ten categories of reality *viz.*: action; habit; location; passion; position; quality; quantity; relation; substance or essence; and time (Ritter & Kohonen 1989). This ontology was established to address underlying questions in ancient Greek philosophy such as *what is real?* and *what can be said to exist?* A basic method for establishing an ontology can be considered in two stages: firstly observation and conceptualisation of the real-world domain; and secondly explicit formalisation of the of the identified concepts (Evermann & Fang 2010). Such a formalisation of the world into well-defined concepts that can be reasoned about in a meaningful way provides a framework that is well-suited to computational analysis of data (Searle 2006, Smith 1998). An open question in ontology research is the degree to which an ontology needs to be complete in order to support understanding and decision-making. The concept of a *naïve ontology* is discussed by Dahlgren (1995) which is an ontology that considers only objects and their classifications. For example a naïve ontology would consider a *passenger train* as being a type of *train*, which in turn is a type of *vehicle*. However such an ontology may not consider abstract concepts such as the concept of *lateness* in relation to a train running to a timetable. This *naïve* approach underpins the resource description framework established by the World Wide Web

Consortium as well as the approach taken by Noy & McGuinness (2001). Dahlgren (1995) asserts that the approach is sufficient to perform almost 80% of common-sense reasoning. Brewster and O'Hara (2007) argue that ontologies are particularly useful in well-defined domains such as individual organisations. Noy & McGuinness (2001) assert that the ontological approach to data management provides a valuable method to reuse domain knowledge. For example database queries can be stored within the ontology so that information found by one analyst can be found again later by other analysts. In this way the ontology adds to the amount of data that needs to be stored, but reduces the need for repetition of work.

2.3 *Ontologies for natural language processing*

Popping (2000) classifies natural language processing (NLP) techniques in three categories, which are in order of sophistication: *thematic*, *semantic*, and *network*. Thematic analysis considers the relative occurrence of words within the source text and can be used as a broad categorisation method to identify texts (such as accident records) that contain similar words and therefore may relate to the same broad themes. Semantic analysis expands the thematic approach by consider the function of words within a sentence (their *part of speech*, such as whether a word is a noun, a verb, or an adjective) to identify subject-verb-object triples. These triples provide the underpinning for complex formal ontology systems that develop from a mereology of objects or actions (Bateman et al. 2010; Bierwisch & Schreuder 1992; Miller & Fellbaum 1991; Ritter & Kohonen 1989).

Network analysis of text establishes the source text as a graph consisting of nodes (which usually represent single words) and edges (which describe relationships between the nodes such as co-occurrence of words within a single sentence). As such the network approach establishes a form of ontology of text within a document, although such an ontology is based on abstract concepts (using words as labels for objects) and therefore fundamentally differs from naïve ontologies which consider the nodes in the ontology as representations of the objects themselves. Miller and Fellbaum (1991) note that a disadvantage with network analysis can be the need for additional software tools such as graphing and network analysis tools, which can complicate the analysis if there is a need to transfer data between separate software products. The introduction of graphical text analysis tools introduces a new domain of research, for example as discussed by Figueres-Esteban et al. (2015).

For general text analysis, Miller & Fellbaum (1991) propose an ontology of 26 basic concepts, these concepts can be used to form a basis for domain-specific ontologies; they are:

- act, action, activity;
- animal, fauna;
- artefact;
- attribute, property;
- body, corpus;
- cognition, ideation;
- communication;
- event, happening;
- feeling, emotion;
- food;
- group, collection;
- location, place;
- motive;
- natural object;
- natural phenomenon;
- person, human being;
- plant, flora;
- possession, property;
- process;
- quantity, amount;
- relation;
- shape;
- society;
- state, condition;
- substance;
- time.

Finally, Van Gulijk et al. (2016) conclude their discussion of the use of ontologies in computer science by providing the following three counsels. Firstly there is no such thing as a perfect ontology; rather there can be a number of alternative ontologies that serve the same purpose. Secondly, effective ontologies are developed iteratively, perhaps as users interact with an ontology and seek more detail from it. Thirdly, ontologies that relate to physical objects are the easiest to create; ontologies that relate to abstract concepts can provide conceptual difficulties for both the ontology builder and the analyst using the ontology.

3 METHOD

Extraction of information from the text was performed in a process that consisted of four mains steps. The first step being the preparation of the data to allow for efficient completion of the later steps. The second step was the identification of key terms in the text and the construction of an ontology to make explicit the relationships between the terms. The third step involved the execution of queries to identify records that correspond with each of the categories of incident; this is an automated step performed by software. The final step was a review of the returned results and conse-

quent refinement of the ontology and queries until an acceptable result was achieved. Each step is described in detail below.

3.1 *Data preparation*

The accident reports provided by the FOT were imported into a graph database. Graph databases structure data as nodes and edges, rather than using the structure required by Structured Query Language (SQL), which has been a prevalent structure for databases some decades. As such, graph databases belong to a class of databases known as *NoSQL*.

Data relating to an individual incident was imported as a single node in the database: 5065 record nodes were created in the database. An automatic process was used to analyse the text in the source records and create a new node for each sentence in the text. During this process a simplifying assumption was made that a full-stop followed by a space (.) would always mark the end of a sentence. The sentence nodes were linked to the node that contained the data from the record.

In alphabetic languages, such as the European languages used in Switzerland, the basic unit of text analysis is a word. Fundamentally the process to establish meaning from text is performed by analysing the occurrence, frequency, and colocation of words or groups of words. In this work each sentence was broken into individual words: punctuation marks were separated from words by inserting spaces. Each word was converted to lower-case text and added as a word node in the graph. During this process the frequency of occurrence of each word was stored in the word node. Colocations of pairs of words were shown by the creation of an edge marked *next*; the edge also recorded data on the frequency of the colocation of the pair. This process of creating word nodes and *next* relationships is the same as the process applied by Lyon (2015).

3.2 *Step 2: Ontology learning*

The source data were analysed to identify terms (words and bigrams) for inclusion in an ontology of items. Candidate terms were identified by calculating the TFIDF score for each word in the source text. Subsequently, each bigram was considered to be a single token and a TFIDF score was calculated for the bigram. All terms from all the source records were compiled in a table of descending TFIDF score and presented to an analyst for consideration in the ontology.

The analyst reviewed the list of candidate terms, starting with those with the largest TFIDF score, and selected the terms that appeared to be relevant to each category of incident. Since the TFIDF ranking is intended to list terms in order of relevance, the analyst worked through the list until reaching a point where it was determined the that the terms were generally irrelevant and no further terms would be considered. The analyst in this trial spoke only English and had no fluency in any of the source languages (German, French, nor Italian) and used simple on-line translation tools to understand the candidate terms. Terms were selected based solely on the analyst's understanding of railway operations and safety. After identifying terms, the analyst created an ontology that joined matching or similar concepts. The ontology allowed equivalent terms in different languages to be joined to the same node. For example, in records written in German, the words *ambulanz* and *krankenwagen* were both used to refer to an ambulance and were linked to the *ambulance* ontology node. Similarly the French term *ambulance*, and the Italian term *ambulanza* were linked to the same node. A simplification was made to link plural terms to singular ontology concepts as a simple form of lemmatisation of the text.

The ontology learning process was limited to the creation of only a naïve ontology: only a single type of relationship was defined indicating that each ontological element can be *a type of* another element; for example *a woman is a type of person*. The ontology did not contain relationships such as *a door is a part of a train*; nor *a train can arrive at a station*.

3.3 *Step 3: Execution of queries based on the ontology*

Queries were performed on the data in the graph database to identify records related to each category of incident. The queries were started at the ontology nodes that defined each category of incident and traced via edges in the graph to identify records that contained terms relating to the incident category.

3.4 *Step 4: Iteration and reporting*

As noted above, the process of ontology creation is iterative. After execution of each query, the analyst reviewed the results to determine whether the records correctly corresponded to the category. Since the analyst had no fluency in the languages used to write the records, the TFIDF ranking process was re-applied to only the records returned as a result of each query. The analyst reviewed the terms occurring in the subset of records, using simple translation tools again, to determine whether terms were occurring that did not appear to relate to each query.

4 RESULTS

This section presents the results of each stage of the analysis.

4.1 *Data preparation*

The 5065 records were loaded into the database and a total of 16,419 unique word nodes were created. Relationships were created to show collocations of words. Figure 2 shows an example of the pair of words *dame* and *âgée* with the NEXT edge joining them. The figure shows that the word *dame* occurs 620 times in the source text, the word *âgée* occurs 202 times and, as a colocation, *dame âgée* occurs 150 times.

4.2 *Step 2: Ontology learning*

The process of TFIDF ranking returned 82,726 candidate terms, being 16,419 single words terms (one for each word node in the database) and 66,307 bigrams. From this list the analyst identified 389 terms that appeared to be related to the specified categories of incident. It is notable that, in some cases, the TFIDF calculation produced similar scores for terms with similar meaning in different languages. For example of the 82,726 identified terms, the term *ältere dame* (German for *elderly lady*) was ranked 314th in the list; the term *dame âgée* (French for *elderly lady*) was ranked 316th on the list.

The analyst constructed an ontology based on the identified terms. For ease of analysis the ontology was limited to only objects and actions and structured in two layers. Table 1 shows example the entities included in the ontology.

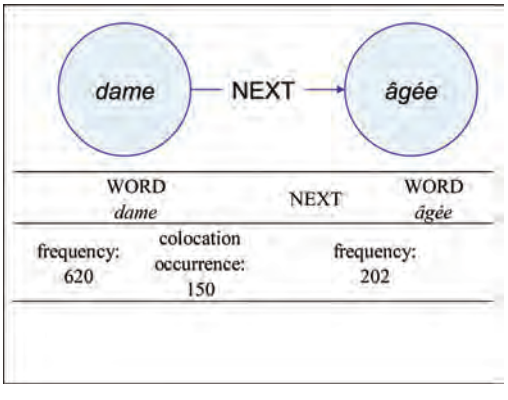


Figure 2. Example of the colocation *dame âgée* and the *next* edge linking the words.

Table 1. Example ontology entries.

Ontology upper layer	Ontology lower layer
person	doctor, self, customer, person, driver, passenger, months old, years old, baby female, old, elderly, young, male, man, lady
places	line, stations, pavement, hospital, ground, platform, stairs
actions	hit, medical, injure, get out, enter, fall, rush
body parts	foot, head, arm, leg
vehicle	carriage, vehicle, ambulance, tram, tain, bus
direction	backwards, direction, in between, in front
items	bag, alcohol, drugs, stairs, footboard, customer information system, ticket, door

4.3 *Step 3: Execution of queries based on the ontology*

Queries were designed and executed for each category of incident based on the terms identified during Step 2. The starting point for each query was the ontological entries that define the key entities being sought in the query. For example to identify records related to injuries caused to passengers by closing doors, the query identified the ontology nodes relating to *passengers*, *closing doors*, and *injury*. The query then identified the terms relating to those concepts, followed by the records that contained those terms. Figure 3 shows an example of records that relate to the ontological concepts of an elderly person and stairs. The query can be thought of executing from the top of the diagram down: firstly the ontology elements for *stairs* and for a *person* – and in particular an old person—are identified. These ontology elements are traced in the query to instances of words that define them, for example the words *Treppe* (being a German word for stairs) and *scale* (Italian for stairs). In turn, these words are traced to instances of accident reports where the words occur. It can be seen that the plural term for men in Italian (*uomini*) has been linked to the singular term in the ontology.

4.4 *Step 4: Iteration and reporting*

The results of the queries were reviewed by the analyst for correctness and the process in Steps 2 and 3 was iterated to refine the terms, ontology and queries to create results that appeared to better align with each category of incident. After iteration of the process, the results of the queries were presented to fluent speakers of each language for review. The reviewers were independent of

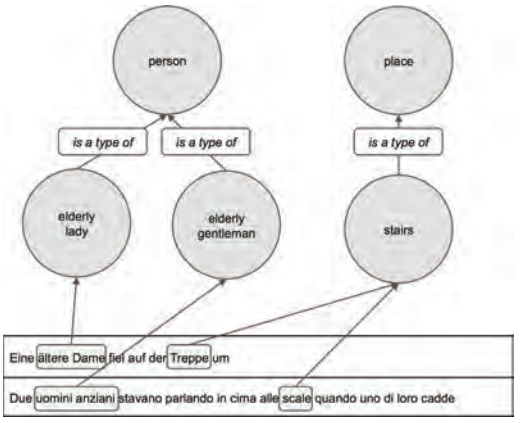


Figure 3. Schematic diagram of an example query.

the process and assessed each record entirely on whether it appeared to correctly describe an event belonging to each category. As such the reviewers' assessed only the rate of true positive matches compared with the overall number of records found to match for each category. The overall results from all reviews indicated that the number of true positives was 98.5% of all positive results returned by the queries.

5 DISCUSSION

The overall finding is that this preliminary study has demonstrated the potential for the technique to be a powerful tool for identifying specific instances of safety-related events from multi-lingual accident reports. The result of 98.5% true positives in all results returned is very strong, especially considering that the analyst did not have fluency in any of the source languages and used only simple translation tools. At this stage, however, it is not clear how many false negatives results occurred as a result of the queries (i.e. records that described one of the categories of incident but were not identified by the queries). Further work would be required to determine the overall accuracy of the process.

The trials of the technique to date have been limited to only a few categories of incident that were specified before the process of ontology creation. Since the process is based on the occurrence of terms in the text, it appears possible that the process could be started by examining the text to determine what categories of incident are being describe, i.e. a *bottom-up* exploration of the text to identify categories rather than starting the analysis with specific categories of incident (a *top-down* analysis). Such a bottom-up approach may be valuable to identify unexpected trends in the data that could not be presupposed; for example unexpected categories of incident caused by emerging technology.

The ontology developed during the process is based on the terms that occur in the text and, as such, it appears that the technique should be applicable to other sources of text that can support safety management such as audit reports, inspection reports; or even to general sources of textual data.

Another limitation of the technique is that it is based on the occurrence of simple concepts being described in the text, but does not consider compounded concepts such as negation. For example when stairs are mentioned in the text, it is presumed that stairs are relevant to the incident, however a record may refer to stairs even though they are not relevant to an incident, (e.g. *an old man, whom I had previously seen on the stairs, fell whilst entering the train*). To address this issue the ontology would need to be updated to include complex ontological concepts. Further work is being carried out to align this study with our previous work (Hughes et al., 2016b) to address these issues in the technique.

REFERENCES

Andriulo, S. and M. G. Gnoni (2014). Measuring the effectiveness of a near-miss management system: An application in an automotive firm supplier. *Reliability Engineering & System Safety* 132(0): 154–162.

Bateman, J. a. et al., 2010. A linguistic ontology of space for natural language processing. *Artificial Intelligence*, 17: 1027–1071.

Bierwisch, M. & Schreuder, R., 1992. From concepts to lexical items. *Cognition*, 42: 23–60.

Brewster, C. & O'Hara, K., 2007. Knowledge representation with ontologies: Present challenges-Future possibilities. *Int. J. Human Computer Studies* 65: 563–568.

Dahlgren, K., 1995. A linguistic ontology. *Int J Human-Computer Studies*, 43: 809–818.

Evermann, J. & Fang, J., 2010. Evaluating ontologies: Towards a cognitive measure of quality. *Information Systems* 35: 391–403.

Figueres-Esteban, M., Hughes, P. & Van Gulijk, C., 2015. Visualising Close Call in railways: a step towards Big Data Risk Analysis. In Fifth International Rail Human Factors Conference: 725–734. London: RSSB.

Gnoni, M. G., Andriulo, S., Maggio, G., & Nardone, P. (2013). Lean occupational safety: An application for a Near-miss Management System design. *Safety science*, 53, 96–104.

Hughes, P., Figueres-Esteban, M., Van Gulijk, C. (2016a) Learning from text-based close call data. *Safety and Reliability: SaRS Journal.* ISSN 0961–7353

Hughes, P., Figueres-Esteban, M., Van Gulijk, C. (2016b) *From negative statements to positive safety.* In: Risk, Reliability and Safety: Innovating Theory and Practice: Proceedings of ESREL 2016. CRC Press. ISBN 9781138029972

Lyon, W. (2015). *Natural Language Processing With Neo4j - Mining Paradigmatic Word Associations*. Retrieved from http://www.lyonwj.com/2015/06/16/nlp-with-neo4 j/. Retrieved 05 May 2017.
Macrae, C. (2014). Close Calls: *Managing Risk and Resilience in Airline Flight Safety*, Palgrave Macmillan.
Miller, G. a & Fellbaum, C., 1991. Semantic networks of English. *Cognition*, 41: 197–229.
Noy, N. & McGuinness, D., 2001. Ontology development 101: A guide to creating your first ontology. *Development*, 32: 1–25.
Popping, R., 2000. *Computer-Assisted Text Analysis*. London, SAGE.
Ritter, H. & Kohonen, T., 1989. Self-organizing semantic maps. *Biological Cybernetics* 61: 241–254.
Ritter, H. & Kohonen, T., 1989. Self-organizing semantic maps. *Biological Cybernetics* 61: 241–254.
Searle, J.R., 2006. Social Ontology: Some Basic Principles. Papers. *Anthropological Theory* 6: 12 –27.
Smith, B. & Welty, C., 2001. Ontology: Towards a New Synthesis. In Proceedings of the international conference on Formal Ontology in Information Systems: 3–9.
Smith, B., 1998. Basic concepts of formal ontology. In *Formal Ontology in Information Systems*: 19–28. Amsterdam: IOS press.
Van Gulijk, C., Hughes, P., & Figueres-Esteban, M. (2016). The potential of ontology for safety and risk analysis. In *Proceedings of ESREL 2016*. CRC Press. Chicago.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)

Building cyber resilience through a discursive approach to "big cyber" threat landscapes

T.O. Grøtan
SINTEF Technology and Society, Trondheim, Norway

ABSTRACT: Cyber safety, security and resilience of Critical Infrastructures (CI) and critical societal functions is a contemporary challenge. To understand the bigger picture, we may build composite threat landscapes in which vulnerabilities and threats combine and travel across distinct domains between which expertise, competence, experience and knowledge horizon related to safety, security and risk may differ substantially. Additional sensitization towards emerging cyber threats is however needed. Inspired by the post-normal "science of what-if", the "BigCyber" model advance threat landscapes further into sensitivity to hidden, dynamic and emergent vulnerabilities. The approach is exemplified in terms of smart metering of household electricity consumption. The need for discursive support for different stakeholders relating to threat landscapes is identified, and a discursive framework for stepwise nurturing of polycentric governance is outlined. The framework can also be used to elaborate and support the idea of resilience landscapes of autonomous entities, facilitating a polycentric approach to cyber resilience.

1 INTRODUCTION

The potential consequences of failure and disturbance of Critical Infrastructures (CI) and societal functions (e.g., energy, water, transport, and logistics) depending on Information and Communication Technology (ICT) are frightening and potentially devastating. The overall risk picture is increasingly blurred, mixed and constantly evolving. It is difficult to maintain a sharp divide between the stable inside of a critical infrastructure, and a more innovative outside. Presumed motivations of potential adversaries and perpetrators span a wide range, encompassing cyber conflict and hybrid warfare, fake information, political influence, cyber-physical damage, cybercrime, sheer vandalism or teenager tricks. This adds to the existing prospects of the accidents, failures and unfortunate incidents in (already) complex systems. Perrow's (1984) notion of the "normal accident" is persistently hard to escape.

The Internet of Things (IoT) is already on the scene, offering new access (= attack) points, new magnitudes of automation and cyber-physical impact, but also boosting the ability to "informate" (Zuboff 1984); to generate electronic texts about the use of the infrastructure. Moreover, the "Internet of Everything" (IoE) has been coined as the "Big Other" surveillance capitalism (Zuboff, 2015), fueling a logic of accumulation. This is signified by the increasing rate of ICT systems rigged for collecting as much (surplus) data as possible. Vendors collecting extensive information from installations without the customer's consent, could be coined as the "industrial Big Other"

In the 1990's, the prospect of "trusted" computer systems prevailed. Today, few if any ICT systems are delivered with assurances that support this. Practically no ICT system, including CI, may preclude the possibility of intrusion, disturbance and hacking. Big-scale consumer innovations, e.g. autonomous cars and home appliances, are seemingly always lagging in computer security. Some voices even claim that "computer security is broken from top to bottom" (Economist, 2017).

Potential countermeasures are often invasive, e.g. on privacy, often unduly playing on strings of fear and anxiety. Public initiatives, e.g. from the EU (Galbusera and Giannopoulos, 2016) aiming for public, semantic web descriptions of critical infrastructures may also be exploited to enable sophisticated attacks.

We cannot expect of holistic, cross-nation, cross-sector approaches to these challenges. The obstacle is not just the tremendous information coordination challenge, but also the incommensurate and diverse motives and objectives across boundaries of private vs public, classified vs unclassified, national vs international. Information cannot be shared, nor trusted, in one "heap". Motives and objectives are incommensurate, increasingly located in an atmosphere of post-fact attitudes, fake news, and information warfare targeting societal trust, in which even security agencies may find it difficult to navigate.

This fundamental challenge demands an attempt of imagining the inconceivable. Societies, organizations and stakeholders habitually directs their hope and faith for dealing with such challenges to risk management and governance, but these are increasingly acknowledging their limitations. Illustratively, a new Specialty Group (SG) on resilience analysis was approved by the Society of Risk Analysis (SRA) Council on December 10, 2017.

In the following, a diverse portfolio of strategies and approaches that can be utilized at several levels, from the national regulator to the infrastructure owner and stakeholder is proposed. The key issues are about building *threat landscapes* to increase sensitivity to hidden, dynamic and emergent ("*h*/*d*/*e*") vulnerabilities and couplings, and to employ the concept of resilience in a polycentric manner catering for diversity, rather as a system-wide property on uniform terms.

An example is offered: smart household electricity metering as part of smart grids. Energy companies strive to use technology to innovate their customer relations, technical maintenance and grid stability, fearing cyber threats, but also fearing a sudden, technology-driven meltdown of their business models.

2 THREAT LANDSCAPES—AND BEYOND

2.1 *Threat landscapes*

Risk management is traditionally not possible without making demarcations about a system regarding boundaries, threats, vulnerabilities, key events, and other inventories (e.g., acting subjects). In today's complex cyber-inflicted systems, such presumptions become increasingly difficult. *H*/*d*/*e* couplings between parts that we traditionally would prefer to keep apart for analytical clarity, or events and conditions that would be considered as unlikely or even irrelevant in conjunction, challenge organizations' and societies' experiences, capabilities and skills regarding imagination as well as actual resilience towards disturbance and surprise.

A societal perspective will have to address the b*igger picture* by recognizing and combining multiple, distinct domains of expertise, competence, experience and knowledge horizons related to, e.g., safety, security, resilience, threat and risk. In this paper, any such distinct domain is "squared out" as a picture, with a frame representing demarcation of inside vs outside, however with the premise that there *may* always be some relevant knowledge missing.

A key challenge is that due to the diversity and h/d/e ICT-induced couplings of physical as well as logical nature, "pictures" may suddenly turn out to be be flawed, and the new threats may travel across such experience-based boundaries in unprecedented ways. The understanding of such composite threat landscapes require methods beyond the practices used to address single domains. Although it is likely that the (more or less professional) risk management approaches per se do not vary dramatically across such "squared" frames, it is likely that the pragmatic knowledge horizon of each domain, e.g., the sensitivity to different phenomena and the ways information and knowledge is recognized, collected, combined and appreciated, will differ substantially more.

Due to the presumed heterogeneity of the total landscape, it is held unlikely that a joint holistic picture of threats and vulnerabilities can be comprehended from a single knowledge horizon. Hence, it is presumed that the "visible landscape" that can be created and shared between domains is constituted by several "squares", each of which representing a specific horizon of knowledge-gathering ("knowledging") strategies and actual experience. To be able to construct such a landscape, three issues are crucial:

1. Explication of the boundary conditions for each horizon "squared out" (Figure 1) in terms of the frame description, the demarcations of the validity of the inside, and the indicators of its saturation (that is, when it cannot accommodate more issues, without losing its pragmatic meaning)
2. the characteristics of overlap zones and the corresponding h/d/e vulnerabilities and couplings that may enable threats to propagate
3. The joint acknowledgement that single frames, as well as their intersections, are not only uncertain, but also influenced by a background landscape encompassing h/d/e phenomena.

The labyrinth background in Figure 1 signifies the persistent hermeneutical challenge of a "moving horizon" (Gadamer 1992) facing each "squared" horizon, as it encounters new phenomena and contesting horizons through the overlap zones.

The resulting threat landscape metaphoric hence implies a loss of traditional presumptions of clear-cut responsibility and authority traditionally associated with single pictures/frames, but also an increased sensitivity to other horizons of understanding.

For taking advantage of this landscape metaphoric, each knowledging agent or community must acknowledge the need to understand the foundations of its own horizon, and be able as well as willing to take a closer look beyond its prevalent presumptions.

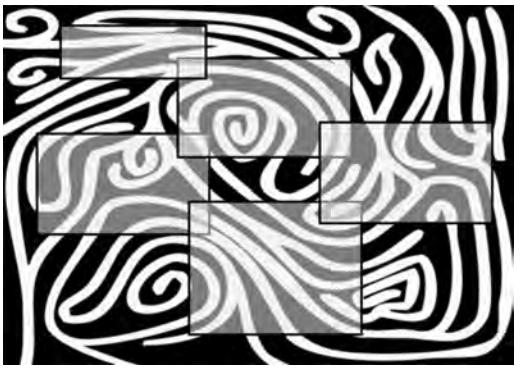


Figure 1. Overlapping threat pictures constituting a threat landscape on a labyrinth background (Grøtan and Antonsen, 2016).

In this way, the threat landscape metaphor may be used to build a "bigger picture" including h/d/e couplings between distinct domains, among which expertise, competence, experience and knowledge horizon related to safety, security, threat and risk are not necessarily commensurate. To establish the grounds for extended understanding of the surrounding landscape, each agent may take advantage of the "take it to the limits" approach (Grøtan and Antonsen 2016) in which a sequence of issues is raised to encircle the boundaries and the saturation points of each frame, and open up for inputs from other domains.

2.2 *"What-if": Sensitization and weak signals*

The threat landscape approach above is primarily useful for utilizing past experience and existing knowledge from professional domains, looking for new combinations, and for revealing or spotting h/d/e vulnerabilities before they have negative impact.

However, many recent events illustrate that cyber (h/d/e) vulnerabilities emerge as a big surprise or a "black swan". A recent example is social media allegedly being used for political communication, tipping elections (Guardian, 2017), indeed exceeding what may be anticipated by means of traditional scientific approaches. The public sphere is very likely to be affected by the consequences, and involved in the key phenomena, e.g., as actually being the product, not the customer, for "Big Others". The threat landscape approach per se may thus not be enough. The residual challenge is thus to be able to raise and pursue the question "what-if" based on hints, weak signals or sheer imagination, and make a serious judgment in time on issues that normally might be discarded as belonging to a risk distribution tail. Also, the "layman" horizon should be included in the process.

To address this residue, inspiration is here gathered from the "post-normal science" (PNS) (also denoted the "science of what-if") introduced by Funtowicz and Ravetz (1993) and Marchi & Ravetz (1999), setting out to resolve a science in crisis (sic!).

Kønig et al. (2017) identify the conditions characterizing a post-normal situation: Irreducible complexity, deep uncertainties, multiple legitimate perspectives, value dissent, high stakes, and urgency of decision-making. The PNS goal is not to attain certain knowledge, but quality, a more robust 'science for policy'. Pointing to how politicization of science renders classical Mertonian scientific norms invalid, they identify an ethos for PNS which they denominate TRUST (Transparency, Robustness, Uncertainty management, Sustainability, and Transdisciplinarity), considering this a nexus for reflexivity practices. They propose that the public trust in science advice can be restored through the PNS ethos.

PNS is also portrayed as "both *descriptive* (describing urgent decision problems – post-normal issues – characterized by incomplete, uncertain or contested knowledge and high decision stakes and how these characteristics change the relationship between science and governance) and *normative* (proposing a style of scientific inquiry and practice that is reflexive, inclusive and transparent in regards to scientific uncertainty and moving into a direction of democratization of expertise)" (Strand 2017).

Here, the PNS challenge is responded to in a more meagre and restricted way; 1) by urging for sensitivity to weak signals, and 2) the proposition of a cyber-vulnerability sensitization model that hopefully make sense to professionals and laymen alike. Hopefully, this is a contribution to the PNS urge to invite "extended peers" into the conversation. Ubiquitous cyber vulnerabilities, and the hope of cyber resilience, should not only be based on a discourse among professionals; ultimately it involves us all.

The turn towards PNS for inspiration, and the return to the less ambitious sensitization model presented below, is thus a natural next step from the idea of the threat landscape as a vehicle for joint comprehension and discourse based on validated of, or even sense-*making* from, weak signals. The notion of "weak signal" is thus not confined to uncertainty within a familiar domain, but include the possibility that something radically outside the normal frames of reference may "travel" trough the landscape, with significant impact at unexpected places.

2.3 *The BigCyber sensitization model*

This sensitization model is intended as a generic tool to support a balanced approach to under-

standing the temptations as well as the possible drawbacks related to utilization of the ever-evolving "cyber space".

2.3.1 *Underlying and formative issues*

2.3.1.1 Potential conflict in cyber space

The actors who own and operate critical infrastructures are usually not directly involved in (military) cyber conflict scenarios, they have traditionally not been seen as military actors. Still, they may be targets for offensive cyber weapons in a potential conflict situation. By making attacks on critical infrastructure from afar technically possible, digital technology also make these types of attacks feasible. It is therefore important that these actors think about the possibility of being targets, and prepare accordingly. An attack of this type could be intended to simply disrupt services, sabotage or even cause physical damage. E.g., Since being coined by CIA Director Leon Panetta in 2016, there has been a persistent concern in the US regarding a potential "Cyber Pearl Harbour" attack.

2.3.1.2 The Internet of Things (IoT)

IoT implies a network of objects able to collect data through embedded sensors and exchanging this information via the internet, but are notoriously hard to secure, and even hard to update when needed.

Both intended, malicious cyber threats and unintended system failures and vulnerabilities of IoT dispersed throughout a CI may lead to severe disruptions in cyber physical systems. In 2016, we also experienced a hint of the future, as the recognized scale of DDoS attacks increased dramatically due to the broad availability of tools for compromising and leveraging the collective, offensive firepower of IoT devices—poorly secured Internet-based security cameras, digital video recorders and Internet routers (Guardian, 2016). The intentions and motives behind may be related to crime and hackers, ranging from teenagers' ploys via organized crime to state actors, but also to cyber conflict and hybrid warfare.

IoT thus sparks the ability to "informate", to generate electronic texts around the use of the infrastructure and technology, boosting the "Big-Other" logic of accumulation.

2.3.1.3 From IoT to Internet of Everything (IoE)

As more and more personal information is being made more or less public, and the possibility for combination increases, a new form of information economy emerges. Zuboff (2015) describes the emergent logic of accumulation in the networked sphere as an "Internet of Everything" (IoE) in which personal information becomes a commodity of high value for a wide range of (unknown) users. This radical new form of surveillance capitalism aims to predict and modify human behavior as a means to produce revenue and market control. Zuboff (2015) launches the need for an 'information civilization' addressing the challenges from "Big Other": "a ubiquitous networked institutional regime that records, modifies, and commodifies everyday experience from toasters to bodies, communication to thought, all with a view to establishing new pathways to monetization and profit" (Zuboff, 2015).

Such an "information civilization" requires a new comprehension of cyber safety and security, including the multifaceted concept of resilience.

2.3.1.4 The lack of assurances

Given the high ambitions related to evaluation criteria for "trusted" computer systems a couple of decades back, there is a striking contemporary silence and numbness related to the lack of assurances about vulnerability of critical computer systems, at least in the non-classified domain. The infamous Stuxnet incident has demonstrated that a widely used industrial control system platform can be used to launch very intricate attacks that are very hard to spot. This is not only about "zero days", it is also about an inherent technological brittleness, and the possibility that industrial plants such as windfarms (Staggs et al. 2017) or smart metering systems (Hansen et al. 2017) demonstrably can be "hacked", with potentially severe consequences. This is also about a flawed marketplace that does not care to ask for such assurances at all, or just to a very minor degree.

2.3.1.5 Privacy

The Norwegian Data Inspectorate have just recently aired their concern regarding the implications of this, and The Norwegian Consumer Council is worried about privacy and consumer rights in a situation where such consumer data has become a "goldmine" for infrastructure operators. The Norwegian telecom operator Telenor is making data from the cellular network to a commodity under the label "mobility analytics". In the US, a new bill is criticized for being a lift of existing legislation that "not only gives cable companies and wireless providers free rein to do what they like with your browsing history, shopping habits, your location and other information gleaned from your online activity, but it would also prevent the Federal Communications Commission from ever again establishing similar consumer privacy protections". It can be doubted whether the individual customer will be able to value his/her privacy sufficiently in relation to the "benefits", or the sheer volume of "user agreements" that are offered.

2.3.1.6 Enter psychology

1. Big Five in Big Data

Psychoinformatics (Montag et al., 2016) is a discipline on the rise. The "Big Five" model has been

a prevalent model for psychological profiling, with alleged predictive power on human behaviour and influence. Recently, the Big Five model has been a driving force in "Big Data" attempts of collecting enough data to reveal patterns from which predictions about human behaviour become quite precise). Some findings seem to suggest strong correlations between Big Five parameters and social media (e.g., facebook) data. E.g., an average of 68 "facebook clicks" seemed to be enough (in 2012) to predict colour of skin, sexual preference, political preference, intelligence, religious belief, use of alcohol/tobacco/drugs, or of having divorced parents, with reasonably high confidence (Grassegger & Krogerus, 2018). With more data, the model predictions beat the assessments of a person from colleagues, friends, parents and spouse. Ultimately, the smart phone is an "enormous psychological questionnaire" feeding us (or someone) with more and more detail. With more information, the prospect is raised that somebody could know "more than the informant think they know about themselves". Inherent in this is the assumed ability to predict an informant's response to a condition/situation.

But it also works the other way around: the user data can also be used as a filter to find and track down users/individuals with specific personality details; providing a method to "profile" people without themselves knowing. It is claimed that this has been used recently in political marketing/communication, by "micro-targeting" through assessments of personalities through Big Five and digital footprints. From which, political messages are organized and based on psychometry rather than demography, by, e.g., designated "messages" as personality-adapted advertisements or "news" (not necessarily "fake"). "Dark posts" are paid fb ads exclusively in the news feeds to users with specific personalities. It can also be about microscopic variations in the same message to accomplish psychological effectiveness, headings, colours, captions, stills or videos, targeting villages, neighbourhoods, or individuals differently. Hence, digital footprints become "real humans" with worries, needs, interests and addresses.

2. Cyber psychology in change

Another issue with possibly unprecedented implications is the potential implications of how digital omnipresence leaks into and potentially changes our psychology as users and operators, e.g. in terms of increased conformity (Størseth 2013).

2.3.2 *The BigCyber sensitization model*

The Big Cyber model summarizes the key issues at large, as illustrated in Figure 2. The model comprises five different "Janus-faces", each of which offering a huge benefit (inside of the dotted pentagon), as well as conducing a severe downside (outside of the dotted pentagon) that can be viewed as an h/d/e threat or vulnerability.

Figure 2. The BigCyber sensitization model.

BigBrother(s) may offer comfort and security in times of crisis and terror, but are giving themselves rather free passes to track down and inflict harm on any instance or person that may be regarded as a present or potentuial adversary. There are no international agreements on ethical conduct of cyber offense.

Personalization and customization of services offers ease of use. But the backside is that we are enrolled into the **BigOther** surveillance capitalism (Zuboff 2015) without being properly asked or informed; "users" are transformed to products and monetized behavioral commodities in a digital economy.

BigData coupled with artificial intelligence and machine learning promise an endless range of new insight and capabilities, but these are not reserved for the "good" purpose. What if the key ideas of "insurance" are jeopardized? Intelligent offense towards CI and ICT systems is as likely as intelligent defense.

The **BigFive** personality model can probably make us even more comfortably numb while effortlessly harvesting the benefits of cyberspace. Will we be able at all to resist the narrowed "alternatives" presented? Will we develop a "cyber psychology" that enables us to recognize and deal with commercially and politically motivated communication?

This is also about an aggregated, unevenly distributed **digital economy and power**. BigOther will have supreme power to utilize BigData as well as BigFive. Evry cyber innovation is aiming for sale, and BigOther is loaded with cash and ready to buy any advantage and "edge" available.

Societies, organizations as well as individuals are always hungry for the **BigInn**(ovation). The lack of basic assurances are hardly noticed, except for the invitation to become an "update junkie", and that the computer industry is not subject to any-

thing near the liability issues that, say, automakers or pharmaceutical industries must consider. Also, outsourcing with fragmented managerial accountabilities is a too easy escape when ambitions of digitalization exceed available competence to deal with the vulnerabilities.

The BigCyber model support understanding of exposure to unfamiliar intentions and motives, and of new attack surfaces and vectors, e.g., cyber-physical impact, small and large, massive profiling, crime and intrusion and an endless stream of "zero days".

3 EXAMPLE: SMART METERING

A smart meter is a physically separate device designed with encrypted communication between the energy supplier and the customer for regular metering at, e.g., an hourly basis. In one way, the Smart Meter is just another Industrial Control System (ICS) or Supervisory Control and Data Acquisition (SCADA) system that is a precondition for even preconceiving the idea of a smart grid, depending on control functions and measurements at an unprecedented scale.

As illustrated in Figure 3, the smart meter also has a "private" physical connector (in Norway denoted a "Home Area Network" (HAN) port) that enable third parties, e.g., providing "smart home" solutions, to read metering data as part of their (innovative) services, connected to the internet. However, do we understand the potential threat landscape of this?

Smart metering is an entry point to the huge challenges of protecting the energy grid as a critical infrastructure. Concerns can be raised, independent on whether the connection is physical or not, on both unauthorised access and to whether the end user oversee the implications of granting additional connections.

The attractiveness of the electricity system for cyber attacks was demonstrated in Ukraine (2015 and 2016). Disruptions may range from massive shutdown (leading to imbalance and potential physical damage) or just poor quality (voltage/frequency). The potential for targeting vast numbers of smart meters simultaneously is demonstrated (Hansen et al. 2017). We have yet to experience the full damage potential, but in the UK, MPs were warned of sabotage threat from smart meter hackers. As experts said rogue programmers could target £11bn system, a massive shutdown will put enormous strain on both the supplier and consumer side (Financial Times, 2016).

Figure 3. The Smart Meter as part of the Smart Grid.

3.1 *The microcosmic threat landscape of ICS*

We start by illustrating the potential of the threat landscape approach at a very small scale. Resting on a similar vocabulary employed by The European Union Agency for Network and Information Security (ENISA), the smart meter seen as an ICS/SCADA system, can be depicted as a "microcosmic" threat landscape in its own respect (Figure 4).

the approach is illustrated in terms of a workshop assessment of an industrial SCADA system in a networked context. The actual "squared" threat pictures (left side of Figure 4) are selected and derived from a similar approach by ENISA. The actual threat landscape composition was conducted as part of the (1st Annual) Workshop on Cyber Safety, Security and Resilience of Critical Energy Infrastructures, Oslo, Norway June 2016. Here, each threat picture was elaborated before combined into the landscape. Both the contents of each "frame" or "horizon", and their overlaps, turned out to be surprisingly complex, and did add weight to the suspicion that the ideal design does not cover every vulnerability in this (microcosmic) threat landscape".

3.2 *The BigCyber-sensitized threat landscape*

Can we conceive a bigger picture, a BigCyber-sensitized smart meter threat landscape?

In addition to necessary functionality for building of smart grids, the smart metering solution is also an excellent example of a connection to "Big

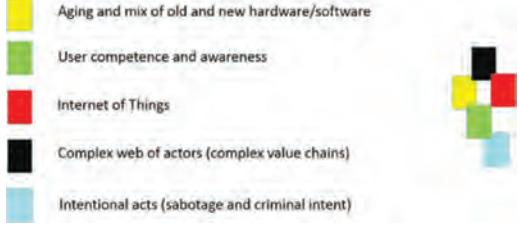


Figure 4. "Microcosmic" ICS/SCADA Threat Landscape.

Other". The joint access to the HAN port it is also a source for building information about "energy behavior" with a huge commercial potential, especially when it is linked to other sources of individual and commercial behavior that can be used to profile targeted individuals or groups. The privacy issues are imminent, but a hostile "BigBrother" may in the ultimate case also weaponize this to trigger collective irregular consumer behavior, and target key personnel, with the intention of disturbing the energy system per se. Another possibility may be conceived through the infamous Stuxnet attack; either by (1) disturbing the crucial grid measurements in order to destabilize trust in grid operation, or (2) initiate (cyber-)physical damage by imposing electrical imbalances.

Hence, we may see the contours of new attack surfaces and vectors of both tangible and intangible kinds, that can be combined and cleverly orchestrated. Vulnerable equipment can be attacked, users and populations can be manipulated and influenced, and key personnel in protection of critical infrastructure services could also be specifically targeted as part of an orchestrated attack, e.g. with a criminal intent. For the defenders, a main vulnerability is the lack of acknowledgement of the coupling.

In Figure 5, a BigCyber-inspired Threat Landscape for smart metering is indicated. The prospects of "clinical" attack vectors, triggering of user behavior as part of attack, optimization of damage and targeting of key personnel on the inside, are simply not refutable one by one. Maybe not even in combination.

3.3 *Weak signals in sight?*

The "metering paranoia" threat landscape (Figure 5) is hypothetical. Are there weak signals that support the likelihood that it may manifest into reality?

- Banks in Asia are already using customers' smartphone data points, like how (often) they drain their battery, to determine whether they're eligible for a loan (CNN, 2016). Can electricity metering derived from smart houses contain behavioral data that could be matched with a personality assessment?
- Will customers care to use the new European privacy legislation to demand insight into smart metering data? Would the trivia of energy consumption draw the necessary attention?
- Energy companies are now increasingly concerned about disruptive competition. Who will take lead in offering homes and companies the dual role of producer and consumer, utilizing solar, wind, (virtual) batteries, e.g. in electrical cars, optimize the energy consumption in a market in which, e.g., consumption based pricing is replaced by capacity-based pricing? Will access to personal energy consumption be part of the "price"? Who will have the data edge in a new market environment? Will we see similar dynamics as when the "Flash Boys" (Lewis 2015) changed the stock markets by means of getting split-second advantages over other actors?
- The grid, however "smart", will still need some supervised electrical stability. Who will be responsible for managing this, with potentially severe consequences in terms of physical damage of electro-mechanical equipment. If for example Google offer an "integrated" energy system to "prosumers", would they care about the grid? If the risk towards the grid level of service is relocated, who will be in charge?

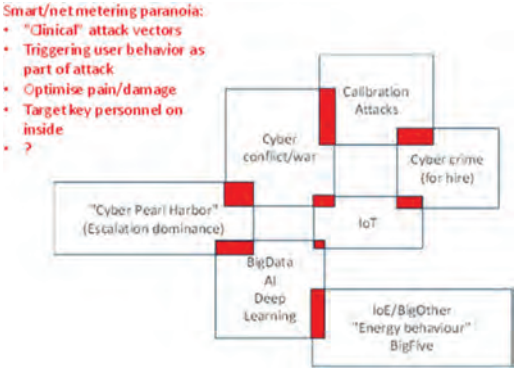


Figure 5. "Smart metering paranoia".

4 DISCURSIVE SUPPORT FOR POLYCENTRIC GOVERNANCE (PCG)

The challenges described above goes beyond the limits of safety and security as traditional disciplines. Petersen (2012) argue that we need an analytical approach "sensitive to conceptual change and diversity" that "enable us to identify innovations in political language" and "provide us with the ability to grasp new developments in the corporate, governmental or organizational conception of risk". There is thus a need for a step change in the way societies and organizations deal with cyber risk, from fragmented to polycentric risk governance (PCB).

The threat landscape metaphoric and the BigCyber sensitization model provide discursive support for PCG. E.g., as in the smart metering example, a regulator can be aware of privacy challenges, but must reach a risk-informed assessment,

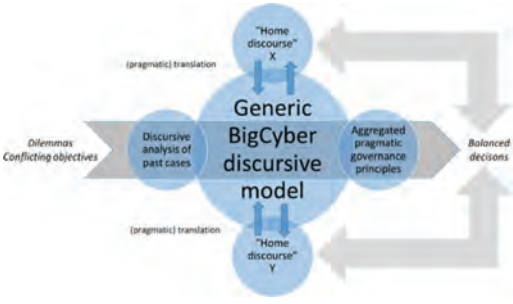


Figure 6. Discursive framework for polycentric governance.

based on current knowledge, without jeopardizing the objective of transforming the grid. We may expect that the regulator take part in a broader, responsible discourse, but we cannot expect them to be voluntarily taken hostage for issues beyond their primary mandate. Hence, we need a discursive framework that sensitizes not only academics and analysts, but also the actual decision makers/processes to the very same issues.

Petersen (2012) claim that "a conceptual discourse does not exist by itself; rather, it will always be defined in interaction with other discourses". Hence, a discursive framework will have to be designed with the following in mind: the users of the framework will come from unique and different "home discourses", and we should enable a sustained resonance between the "home" and the joint discourse, as participants move along their unique trajectories.

As indicated in Figure 6, several stakeholders, each of which bound to a home discourse, e.g., of privacy and consumer rights or of facilitation of smart grid development, can join forces, overlap horizons, share threat landscapes (TL) and challenge themselves and others by using the Generic BigCyber discursive model, which is the simplistic (and recursive) formula of

$$TL \rightarrow BigCyber \rightarrow TL',$$

as exemplified in chapter 3.

Using this discursive framework will contribute to an improved coherence between decisions made by different stakeholders. The "lay" perspective may be voiced through civic participation, NGOs, or proxied through agencies of consumer rights and privacy.

5 FROM PCG TO POLYCENTRIC RESILIENCE

During the past decade, the safety field as well as the societal security and disaster fields, have devoted attention to the concept of "resilience. However, the notion of cyber resilience demands more than a technological fix. Human and organizational issues are more inert than the technological, and also for cyber resilience we must respect the double-hermeneutic scientific principle of understanding *understanding subjects*, rather than explaining them as objects.

It is important that the concept is properly contextualized. Though it sounds normatively good, it carries no guarantee for success. It is an attractive idea that invites fallible practices, and hence it must be brought under managerial supervision, accountability and mandate. If not, we may invite expectations that will victimize those that are not able to thrive from being exposed to risk, or that do not possess the resources or skills in the first place.

We must take the notion of resilience seriously without depriving it from its content and origins through mere re-labelling of traditional risk management practices. Resilience is ultimately a matter of emergent, "bottom-up" and situated solutions to unique and idiosyncratic demands and situations rather that instrumental responses to stereotypical replications of former situations.

By implication of the above, cyber resilience must be translated to a scheme of composite protection comprising a diverse set of (resilient) entities that can be orchestrated to a certain degree. Grøtan and Bergström (2016) propose a theoretical foundation for exploring the concept of resilience landscapes; autonomous but interconnected resilient entities that forms a composite scheme of resilience. Such entities can utilize the same discursive structure as for PCG (Figure 6), and the evolving threat landscape can be a basis for dynamic interfaces and interactive patterns.

6 CONCLUSION

The threat landscape metaphoric and the BigCyber sensitization model is a promising approach that make sense in the smart metering case, and carries a potential for further application for the emerging cyber threat landscapes. The notions of polycentric governance and polycentric resilience landscapes are logical companions to the former, and both can benefit from the discursive support structure presented.

ACKNOWLEDGEMENTS

This paper is funded by the SAMRISK-II project New Strains of Society, under Contract no 238093/H20 with the Norwegian Research Council.

REFERENCES

Financial Times, web pages. 2016. MPs warned of sabotage threat from smart meter hackers. September 24, 2016.

Funtowicz, S.O. & Ravetz, J.R. 1993. *Science for the post-Normal age. Futures*, 25 (7) (1993), pp. 739–755.

Gadamer, H-G. 1992. Truth and Method. 2nd ed. Trans. Joel Weinsheimer and Donald G. Marshall. N.Y.: Crossroad.

Galbusera, L. & Giannopoulos, G. 2017. Exploiting web ontologies for automatic critical infrastructure data retrieval. Eleventh Annual IFIP Working Group 11.10 International Conference on Critical Infrastructure Protection. Arlington, USA, March 2017.

Grassegger, H. & Krogerus, M. 2018. The Data That Turned the World Upside Down. https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win, downloaded 2018–02–15

Grøtan, T.O. & Antonsen, S. 2016. Take it to the limits! Exploring the hidden, dynamic and emergent vulnerabilities of society. ESREL 2016: Taylor & Francis Group, London.

Grøtan, T.O. & Bergström, J. 2016. Calibrated resilience landscapes of composite protection: Theoretical grounding of an empirical approach. ESREL 2016: Taylor & Francis Group, London.

Guardian, The. 2016. DDoS attack that disrupted internet was largest of its kind in history, experts say. October 26th.

Guardian, The. 2017. Did Cambridge Analytica influence the Brexit vote and the US election? March 4th.

Hansen, Aa., Staggs, J. & Shenoi, S. 2017. Security Analysis of an Advanced Metering Infrastructure. . Eleventh Annual IFIP Working Group 11.10 International Conference on Critical Infrastructure Protection. Arlington, USA, March 2017.

Kønig, N., Børsen, T. and C.v Emmeche. 2017. The ethos of Post-normal science. Futures. Volume 91, August 2017, Pages 1–4.

Lewis, M. 2015. Flash Boys. A Wall Street Revolt. Norton Marchi, B., Ravtez, J: 1999. Risk management and governance: A post-normal science approach. *Futures* 31(7):743–757.

Montag, C. et al. 2016. Toward Psychoinformatics: Computer Science Meets Psychology. Computational and Mathematical Methods in Medicine. Volume 2016 (2016), http://dx.doi.org/10.1155/2016/2983685

Perrow, C. 1984. Normal Accidents: living with high-risk technologies. New York: Basic Books.

Petersen, K.L. 2012. Risk analysis–A field within security studies? Eur. J. Int. Relations, vol. 18, no. 4, pp. 693–717.

Staggs, J., Ferlemann, D. & Shenoi, S. 2017. Wind Farm Security: Attack Surface, Targets, Scenarios and Mitigation. Eleventh Annual IFIP Working Group 11.10 International Conference on Critical Infrastructure Protection. Arlington, USA, March 13–15 2017.

Strand, R. 2017. Post-normal science. In C.L. Spash (Ed.), Routledge handbook of ecological economics: Nature and society, Routledge, London, pp. 288–298.

Størseth, F. 2013. Digital culture conformity: contours of a 'new psychology' and its impact on safety. PSAM 2013, Tokyo, Japan, 14–18 April 2013.

Zuboff, S. 1984. In The Age Of The Smart Machine: The Future Of Work And Power.

Zuboff, S. 2015. Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30, 75–89.

CNN. 2016. http://money.cnn.com/2016/08/24/technology/lenddo-smartphone-battery-loan/index.html

The Economist, April 2017. Computer Security is Broken from Top to Bottom.

Foundation of risk and reliability assessment and management

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Safety principles for autonomous driving

H. Schäbe
TÜV Rheinland InterTraffic GmbH, Köln, Germany

ABSTRACT: In safety technology, the application of safety principles (e.g. fail-safe or safe-life) is used to design and implement a safe system that eventually fulfils the requirements of the functional safety standards. Safety principles have already been described and applied to guided transport systems, including systems with immaterial guidance principles. The different responsibility of human driver and technical driving system in different automation levels for autonomous driving vehicles require the application of safety principles. We consider, which safety principles have to be applied using general safety principles and analysing the relevant SAE level based on the experience from projects. For the five levels of automated driving as defined by the SAE, safety principles are derived. For the levels 0–2, the driver is fully responsible for driving, whereas starting from level 3, the automated driving equipment monitors the vehicle. To give the driver the possibility to intervene, means that this must be implemented according to the relevant safety integrity level and that the driver must have enough time to take over control. The latter strongly depends on the level of automation and the speed and the environment in which the vehicle moves. Depending on the level of automation, the technical systems are implemented as fail-silent or as safe-life. There are also exclusions, e.g. when the technical systems can be implemented as fail safe. This is possible, when the vehicle always can be brought to a safe stop, e.g. when driving with low speed and on a controlled territory.

1 INTRODUCTION

Autonomous driving has become a very important subject of research and first pilot projects. In safety technology, the application of safety principles as e.g. fail-safe or safe-life is a very important tool to design and implement a safe system that eventually fulfils the requirements of the standards for functional safety. Safety principles have already been described and applied to guided transport systems, including system with immaterial guidance principles.

In earlier papers, safety principles have been described and later applied to guided driving.

In the present paper, we systematically consider which safety principles have to be applied for which SAE level of autonomously driving systems und we show how an autonomous system could be built. This is partially done with the help of general safety principles, partially by analysing the relevant SAE level based on the experience from several projects.

According to UN resolution /UN/ or SAE / SAE/, autonomous driving on the road knows five different levels:

- 0 No automation
- 1 Driver assistance
- 2 Partial automation
- 3 Conditional automation
- 4 High automation
- 5 Full automation

For the levels 0–2, the driver is fully responsible for driving, starting from level 3 the automated driving equipment monitors the vehicle.

This different responsibility of human driver and technical driving system requires the application of safety principles. In the present paper, we systematically consider which safety principles have to be applied for which level und we show how such a system could be built.

This is partially done with the help of general safety principles, partially by analysing the relevant level.

We start with a very simple and abstract model of the system and show that there exist different possibilities to implement autonomous driving. An important result is that an arbiter needs to be installed that gives the human driver the possibility to override the decisions of the autonomous system to fulfil legal requirements.

For the five levels of automated driving as defined by the SAE (2016), safety principles are derived. For the levels 0–2, the driver is fully responsible for driving, whereas starting from level 3, the automated driving equipment monitors the vehicle. To give the driver the possibility to inter-

vene, means that this must be implemented according to the relevant safety integrity level and that the driver must have enough time to take over control. The latter strongly depends on the level of automation and the speed and the environment in which the vehicle moves.

Depending on the level of automation, the technical system are implemented as fail-silent or as safe-life. There are also exclusions, when the technical systems can be implanted as fail safe, when the vehicle always can be brought to a safe stop, e.g. when driving with low speed and on a controlled territory.

We consider the two main functions of guidance and braking / acceleration and their role for autonomous driving. Moreover, detection and reaction with regard to fixed and moving obstacles is discussed.

Two basic requirements for autonomous systems are that they need to be developed according to the relevant standards of functional safety fulfilling an ASIL (or SIL) level and that the capability of the autonomous driving system must at least on the same level as that of a human driver.

We note that Wachenfeld (2016) has proposed a stochastic approach to show that an autonomous system fulfils a certain level of performance or safety. This, however, can only be seen additional evidence, the main evidence for a safe system is an appropriate safety architecture implemented according to the rules of functional safety, see ISO 26262.

2 GENERAL SAFETY PRINCIPLES

In this chapter we will briefly remember the main safety principles, see Gülker & Schäbe (2006) and and Gayen & Schäbe (2008).

Fail safe: If the system has a safe stopping state, i.e. a safe state in which it is not operational and this state is stable which can be reached fast enough, then the fail safe principle can be applied. It means that a system is brought into this sate if a failure occurs which cannot be tolerated. This principle can be implemented as inherent fail-safety, ractive fail-safety or composite fail-safety.

Safe life (fail operational): If the system does not have a safe stopping state which can reached fast enough, then the safety function has to be ensured. This is mainly done by using redundancies.

Fail silent: The fail silent principle is applied to a function the loss of which is tolerable since it is either an assistance function or the function is implemented in several instances. Then, failure of the function must be such that there is no repercussion on the safe functioning of the system. That means, that a fail-silent system must detect its failures and possible dangerous states and switch itself off without influencing other systems in a dangerous way.

3 ABSTRACT MODEL OF THE SYSTEM

Lotz (2017) proposed an architecture consisting of three levels: a navigational level, a manoeuvring level and a controller level. We will try to discuss a model that is as simple as possible.

For systems that drive automatically, partially automatically or autonomously, we will use the following very simple structure for the system. In fact, this system must be equipped not only with a human driver, but also with a technical driving system, that carries out the driving.

The vehicle consists of driving sub-systems as steering, braking, acceleration systems etc. in a very abstract manner. These sub-systems could be even very simple systems as pure mechanical steering system, pneumatic brake systems etc. The driving is carried out by the human driver using these driving sub-systems directly.

The manoeuvring and navigational level according to Lotz (2017) have here been combined in one system (human driver / technical driving system).

If a vehicle shall be operated by a technical system which does the driving in place of the human driver or supports the human driver, then this system must have access to the driving sub-systems. This is possible only using a driving controller and actuator. That means that these types of systems must be present in the vehicle to allow for driving by a technical system.

Then, this allows also the human driver to access the driving sub-systems via the driving controllers.

Hence, there are different possibilities to operate these subsystems.

a. The driver can directly access the driving subsystems, e.g. the steering wheel is mechanically connected to the steered axle.
b. The driver accesses the driving sub-system via a controller and an actuator which operate the sub-system electronically. A typical example for such a system is an electric parking brake.
c. The technical driving system accesses the driving subsystem via a controller and actuator

We will not consider the aspects of how the driver should best access the driving sub-systems yet.

We need to distinguish two situations:

a. driving on an open road and
b. driving on private territory

Without going into details we must be aware of the fact that for driving on an open road, the Convention requires a driver to be always present

which is implemented in the national law of almost all countries. For driving on private territory, the traffic law is not applicable—the car would be a moving machine. Nevertheless, also her, safety requirements have to be obeyed.

4 SAFETY INTEGRITY LEVELS

Whenever a function might lead to harm, i.e. injury of fatalities to persons, material damage, damage to the environment, functional safety has to be applied. That means that the risk arising from a possible functional failure must be reduced to an acceptable level.

For this sake, safety integrity levels are defined. According to ISO 26262 this can be QM, ASIL A to ASIL D with ASIL D being the most severe. This system has to be applied for road vehicles with a gross weight of up to 3.5 t. For heavier vehicles (still) IEC 61508 would be applicable, which knows the safety integrity levels 1 to 4. Also, for moving machines, IEC 61508 is applicable.

In practice this means, that for all driving functions and all driving sub-systems, the necessary safety level (ASIL or SIL) has to be determined using a risk analysis.

5 LEVEL ANALYSIS

In this section we will analyse the levels (SAE (2016)) of automatisation and draw conclusions for the safety architecture of a vehicle.

In levels 0-1 execution of steering and acceleration and deceleration is in the responsibility of the human driver, the driver is responsible for monitoring and the technical system is able to support some driving modes (level 1).

That means, the human driver is doing the driving and the technical driving system can only add some supporting functionality as warn the driver or react in cases, when he is not able to react (emergency brake assistant). This means that the technical driving system must be fail silent, i.e. upon failure of this system the driving behaviour of the vehicle must not be influenced or only influenced in such a manner that safe driving is still possible. The driver should be warned, if such an assistance system fails to work.

In automation level 2, the system takes responsibility in some driving modes. The human driver monitors the technical driving system and he is the fall back solution. That means, that all technical systems are pure assistance systems and that

R1. The driver must have the technical possibility to interfere, i.e. to override the technical systems. That means, that each controller for acceleration, braking and steering that receives signals from the human driver and from the technical driving system must have a voter which always gives priority to the driver. In fact this means that an electronic control system needs to be present for these function that has an ASIL that coincides with the function, mainly this would be ASIL D. This control system then must have a priority input for the driver and another non-priority input for the technical driving system. The relevant driving controller must detect, when the driver wants to override the technical driving system and has to carry out the required reaction.

R2. The driver must have enough time to detect wrong or faulty behaviour of the technical driving system and react and be able to bring the vehicle back to a safe driving state. That means, that the controllers have to limit the influence of the technical driving system, e.g. limit the level of acceleration, deceleration and the steering angles or angular speed and angular acceleration and jerks so that the driver always has the time to react. Moreover, the driver must be trained for this function or the controllers must be designed in such a manner that they give enough time for reaction for any driver.

Level 3 differs from level 2 in just one point. The technical driving system is responsible for monitoring of the driving equipment. That means that the system must diagnose itself and the environment in order to decide whether it can go on with driving or whether the human driver must act as a fall back solution. The following questions are important

R3. A clear handshake must be defined between human driver and technical driving system. Either the technical driving system must go on with functioning until the human driver has accepted to take over control or

R4. A certain time of e.g. one second is foreseen for the human driver to take over control at any time, if the technical driving system asks him to do so.

In the first case, the technical driving must be safe life, in the second case the latency time for the human driver to take over must be ensured by technical systems—either by the safe life property or just by the driving situations and speed. Timing considerations can be found in Vogelpohl et al. (2016).

Levels 4 and 5 are even more advanced. The difference between levels 4 and 5 is relatively small, since the distribution of responsibilities is the same, only in level 4 some driving modes are excluded which allows the technical driving system to have

limited capabilities. However, when this system is active, it must be able to take full responsibility.

As a consequence, the technical driving system must always ensure safe driving and would need to be safe life.

The relevant requirements are derived the so called GAMÈ principle, which can be found e.g. in EN 50126 "All new guided transport systems must offer a level of risk globally at least as good as the one offered by any equivalent existing system". Here we apply the phrase on guided transport system just to an autonomously driving vehicle. We compare the classical vehicle with a human driver with an autonomously driving vehicle. Then, there are two aspects to be considered:

a. Performance and
b. The technical system (vehicle and technical driving system) are sufficiently free from dangerous failures.

Both aspects are considered separately. For performance, the technical driving system must be at last as good as a human driver in the relevant driving situations, see Mazzega et al (2016). If this cannot be ensured for all driving situations, the set of relevant driving situations must be limited and the human driver must handle the most complex ones.

The second, the safety aspect can be handled as for any technical system by defining an appropriate safety level (ASIL or SIL). This leads to

R5. The performance of the technical driving system as reaction time, detection and handling of traffic situations etc. with an un-failed system must be at least as good as that of a human driver.
R6. The technical driving system must be developed according to a reasonable SIL / ASIL.

For level 4, a clear handshake must be defined how to pass over responsibility between technical driving system and human driver. Especially, the driving modes must be defined, where the technical driving system must not be used for reasons of e.g. insufficient performance. Handshake must be carried out either during standstill or the technical driving system must early enough inform the driver that it wants to pass control to the driver and the driver must take responsibility. If the driver does not take over, the technical driving system must still have the possibility to stop the vehicles as long as it is in a driving mode, where automatic driving is allowed and possible.

If the driver passes responsibility to the technical driving system he must have responsibility until the technical driving system informs him that it has taken over responsibility.

When driving on an open road, the driver must be in full responsibility of the driving behaviour of the vehicle, see the Convention. Then, even if the technical driving system is able to perform up to SAE level 5 with the necessary safety integrity, the driver must have the possibility to intervene. So, the requirements under a) and b) mentioned for SAE level 2 hold if driving on an open road.

Autonomous driving, i.e. driving without intervention of a human driver is in fact only realized in SAE level 4 (partially) or 5 (completely). This holds even if the laws require a driver to be present.

6 IMPLEMENTATION

In this section, we will discuss possible implementation schemes and principles.

It is clear that for technical driving systems in levels up to 2 the systems must and can be fail silent and R1 and R2 must be fulfilled to ensure that the driver has the possibility to take over control.

6.1 *Arbitration between human driver and technical driving system*

Comparing this with Figure 1 it becomes clear that arbitration between the commands of the human driver and the technical driving system must take place.

There are different levels on which arbitration can take place:

a. driving subsystems
   In this case the force applied by the driving controller and actuator must be so small that the driver can always overrule without a problem. However, he would be either required to switch off the driving controller an actuator manually, or those system need to have an in-built function to detect the interference of the driver and switch themselves off.
b. driving controller and actuator
   Here, the driving controller has two inputs with different priority. The high priority input is used by the driver, the low priority input by the tech-

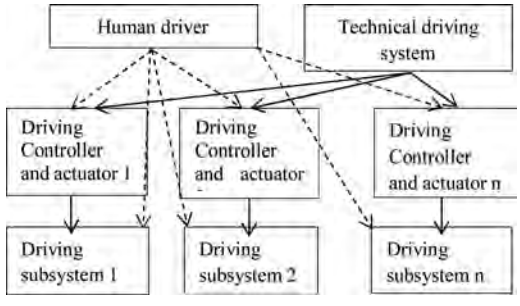


Figure 1. Scheme of a vehicle with automatic driving capabilities.

nical driving system. The arbitration is done by the driving controller which detects overruling by the human driver and switches off the input from the technical driving system. Many controllers in modern cars (brake controller, steering controller etc.) have an additional input for assistance systems which just fulfils this requirement. This approach assumes that the human driver himself controls the vehicle via x-by-wire via the relevant controllers.

c. technical driving system

Arbitration is between the human driver and the technical driving system. If the human driver overrules the technical driving system the latter does not generate its own control signals but simply transfers the signals of the human driver to the driving controllers.

The choice on one of the approaches is a choice of the manufacturer of the vehicle. However, this choice influences the suppliers of the driving controllers and actuators. They need to implement different architectures in their controllers.

In case a) they need to detect intervention of the man driver and deactivate the actuator.

In case b) they need to have two inputs with different priority and need to carry out arbitration

In case c) only one input is necessary and no arbitration is necessary.

We see that x-by-wire is a necessary precondition for solutions b) and c).

We will guide ourselves by the requirements for SAE levels 5 and 2 combined, i.e. for a fully autonomous driving vehicle with a possibility for the human driver to take over responsibility at any time.

6.2 *Technical implementation with safety principles*

For the technical implementation we will discuss the controllers and the technical driving system itself.

Regarding the driving controllers, they need to be developed and implemented according to an adequate SIL. This is for the brake (ABS / ESP) mainly ASIL D, for the steering ASIL B…ASIL D, depending on the function of this controller. With such a choice most of the vehicles can perform with velocities up to 250 km/h.

First of all, we need to determine whether there exists a safe stopping state that can be reached sufficiently quick. Assume the velocity of the vehicle is limited to a value v, the braking deceleration is a and the reaction time t then the vehicle will stop within a distance of

$$s = v \cdot t + v/(2a).$$

Assuming that the steering has no limitation, stopping the vehicle will be a safe action if there is no obstacle within a distance of s from the outer boundaries of the. This area can be made even smaller taking into account that

- actual direction of steering and the (physical) limitations of changes of the steering angle and
- physical limitations for changing the driving direction.

In such case, the technical driving system and the driving controllers could be a complete fail safe system, stopping the vehicle in case that a failure is detected.

Depending on the free space around, the vehicle speed is determined. Obviously, the less free space available, the slower the vehicle must drive.

If the vehicle is intended to move faster, the technical driving system and the driving controllers must be safe life, at least as long as the vehicle is in motion.

The implementation using safety principles differs whether we are talking on a road vehicle or a moving machine. In the first case the environment cannot be assumed to be under control, in the second case this can be ensured since the technical driving system acts on private territory. In this latter case it is much easier to ensure enough free space.

From this consideration it becomes also clear, that not all functions must be always implemented with the highest SIL / ASIL. This depends very much on the speed and the environment. If speed is limited by physical or other means, the also a lower SIL or ASIL can be used. In any case this needs to be shown by the risk analysis that has to be accrued out based on ISO 26262 or IEC 61508.

The following functions are the main functions to be considered:

- Guidance
  How to implement such a function including the steering is described in Bouwman, Schäbe & Vis (2006). Mostly the steering of the axles needs to be safe life and a safe computer has to be used in the technical driving system to determine the steering angles. Another important function is determination of the location, where differential GPS, maps together with ultrasonic sensors, radar or lasers or cameras or different types of marking placed physical on the lane of the vehicle can be used. The safe computer will determine the real location and compare this with the assumed location as a result of its steering activities and correct or stop he vehicle.
- Braking and acceleration
  Assuming that the vehicle moves along the desired trajectory, the vehicle needs to start,

move and stop. So the vehicle needs to react to these commands. It is important to limit the speed e.g. in curves or at narrow places and to be able to perform an emergency stop, if parts of the system fail. In order to perform this function, the system needs to know the location.
Solely with these two functions the vehicle would move without taking into account the environment. Any change in the environment could lead to a collision or the vehicle leaving its track.

- Reaction to unforeseen events (obstacle)
The vehicle must be able to detect obstacles. By an obstacle we denote any object that is in the (planned) or near the (planned) trajectory of the vehicle. We need to distinguish fixed obstacles and moving obstacles. In the beginning we consider as only strategy of the vehicle to stop in front of the obstacle. Moving around the obstacle will be considered later together with moving obstacles
  a. Stationery obstacle: To detect the technical driving system needs to have a blueprint of the environment and needs to compare the real environment with that blueprint and detect differences. This would require certain algorithms for detection and classification of objects. Note that "detection" and "blueprint" does not mean that the technical driving system uses optical means. It can be optical means, but also others or in combination.
  In a first step the obstacle as such needs to be detected. This is possible only at a certain distance and takes a certain time. This performance of the system might limit the speed, since the vehicle must always come to a standstill in front of the object.
  In a second step the technical driving system can classify the obstacles as small. Note that this classification can be present implicitly if the technical driving system will not detect obstacles of small size. Such a classification is always present due to limitations of the system.
  If the obstacle is small enough and not tall, the vehicle might decide to go on with driving.
  b. Moving obstacles: Moving obstacles must be traced and its motion must be predicted using the actual position and speed. It must also be taken into account whether the object can accelerate or decelerate or change its motion direction. The latter factors strongly depend on the nature of the object. E.g. a motorbike can reach other acceleration values as a pedestrian. In order to provide a good prediction, the technical driving system must cluster moving objects according to their capability of motion. Consequently, for each object of the different clusters future positions must be predicted and the technical driving system must define the motion of the vehicle in such a manner that collisions are avoided. This might lead to the decision to stop or to keep the present fixed position.
  Depending on the performance of the clustering and prediction algorithms, the technical driving system would behave more or less conservatively. With better algorithms the technical driving system would stop less frequently. We remind that the performance of these algorithms together with the stopping process in case of doubts about the future trajectory of the obstacle must be as least as good as that of a human driver. This includes of course strategies to drive around an obstacle.
  c. Stationery obstacles that could start moving are in fact a combination of cases a) and b) discussed above. This means, that the technical driving system must not only trace moving obstacles but must also be able to classify stationery obstacles and provide a judgement on whether they might move and with which velocity and in which direction. A most safe strategy would surely be to stop at a safe distance of any unknown object.
  If a proper reaction of the vehicle cannot be ensured for all driving situations, the set of relevant driving situations must be limited and the human driver must handle the more complex ones. This would lead to an SAE level 4 situation. An example would be a strategy, where the technical driving system takes over control on a motorway and the human driver in the city.

7 PROBLEMS

In connection with autonomous driving some problem appear. We will, discuss only some of them and try to describe possible technical solutions.

a. Assume an autonomous vehicle cannot prevent an accident and needs to make a choice, e.g. between material damage, environmental damage and injury or—even worse—injuring or even killing either an older or younger person, another driver, the own passengers etc., see e.g. EK (2017)
This type of discussions automatically comes up when the responsibility for driving is carried over from the human driver to a technical driving system. The ethic problem that is behind this discussion cannot be solved in this paragraph. It is obvious that a technical solution to this problem would require to distinguish between persons and objects or animals, to discriminate between different persons etc.

This would require rather complex algorithms, if it is feasible at all.

The simplest solution to the problem is to apply the principle of driving on sight. That means the rule for the autonomous vehicle would be to drive only with such a velocity that it can stop before each obstacle that appears on the road. This requirement covers:

- Detection of any obstacle above a certain size,
- Prediction of movement of objects (which is the most complicated part),
- Reducing speed if necessary to come to a standstill before such an obstacle.

Based on such a "safety first" approach, later on objects of certain (small) size can be neglected to ensure performance and avoid the vehicle stopping in front of a leaf or a plastic bag.

b. Additional information

A vehicle might optionally use additional information provided by the infrastructure, which might lead to better performance regarding safety.

Let us consider the following example. The vehicle uses information from cameras mounted on the street and has the possibility to "look around the corner". Then, it could e.g. detect a suddenly appearing child running out of the house, what a human driver could not.

c. Safety targets

Since the target of autonomous driving behaviour would always be the performance of a human driver, the technical driving system would have to fulfil this important requirement. However, assume that autonomous systems will set a new target in the future—then the question will arise: Does the driver have the right to switch the automatic system off and decrease the achieved level of safety? It would be somehow equivalent to a train driver switching off automatic train protection, e.g. to use some speed margins. This simple example shows that the way to autonomous driving would be a one-way street, with no return to manual driving at the end.

8 CONCLUSIONS

In this paper we have presented some ideas on possible safety architecture for autonomous driving, deduced from known safety principles and from general requirements. We have analysed the SAE levels and the implication for the safety architecture per level.

Possible implementation principles have been described and specific problems of autonomous driving have been discussed.

REFERENCES

Bouwman, R., H. Schäbe, H. Vis, (2009), Application of safety principles for a guidance system in public transport, *ESREL 2009, Proceedings Reliability, Risk and Safety*, vol. 3, p. 2275–2278.

EK 2017, *ETHIK-KOMMISSION AUTOMATISIERTES UND VERNETZTES FAHRE*N (Ethics Commission for automated and networked driving, in German), Bericht, Juni 2017, WWW.BMVDI.DE

EN 50126 Railway applications – The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS) (EN 50126), 1999

Gayen, J.-T., H. Schäbe, (Miss-) Konzeptionen von Sicherheitsprinzipien, *Signal und Draht*, 100 Nr. 7+8 (2008) pp. 11–18.

Gayen, J.-T., H. Schäbe, (Mis-) conceptions of safety principles, *ESREL 2008, Proceedings Safety, Reliability and Risk analysis*, vol. 2, pp. 1283–1291

Gräfling, S., H. Schäbe, The agri-motive safety performance integrity level – or how do you call it?, *ESREL 2012 / PSAM 11*, paper 26 Fr2_1, 10 p..

Gülker, J., H. Schäbe, 2006, Physical Principles of Safety, *Safety and Reliability for Managing Risk, Proc. of ESREL 2006*, pp. 1045–1050.

IEC 61508 Functional safety of electrical / electronic / programmable electronic safety-related systems, 2010, parts 1–7,

Lotz, G.O. 2017, *Eine Referenzarchitektur für die assistierte und automatisierte Fahrzeugführung mit Fahrereinbindung*, Dissertation Technical University Darmstadt, 2017 (A reference architecture for assisted and automatic driving with driver intervention),

Mazzega, J., Köster, F., Lemmer, K.,. Form, T., *Absicherung hochautomatisierter Fahrfunktionen*, Automobiltechnische Zeitschrift, 118 (2016), no. 10, 48–52 (Safe Implementation of Highly automated Driving Functions)

SAE (2016) Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, *SAE J3016*, September 2016.

Vogelpohl, T., Vollrath, M., Kühn, M. Hummel, T. Gehlert, T., (2016), *Übergabe von hochautomatisiertem Fahren zu manueller Steuerung*, Forschungsbericht Nr. 39, Unfallforschung der Versicherer GDV, August 2016, ISBN 978–3-939163–67–1

Short English Version:

Convention (1973), *Convention on Road Traffic*, 8.11.1968, European Additional Treaty from 1.5.1071 and Protocol 1.3.1973.

ISO 26262 Road vehicles — Functional safety, 2011, parts 1–10.

Kühn, M. Takeover times in highly automated driving Compact accident research No. 57, Unfallforschung der Versicherer GDV, 07/2016

UN (2017) Economic Commission for Europe, Inland Transport Committee, World Forum for Harmonization of Vehicle Regulations, Consolidated Resolution on the Construction of Vehicles, (R.E.3), Revision 6, 11.7.2017

Wachenfeld, H. K. (2016), *How Stochastic can Help to Introduce Automated Driving*, Dissertation, Technical University Darmstadt, 19.10.2016

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Tool for risk reduction at specific component aircraft engine welding

D. Procházková & J. Prochazka
Faculty of Transportation Sciences, Czech Technical University in Prague, Czech Republic

ABSTRACT: The safety is on the first rank in the aero industry. From this reason at the aero components fabrication, the great attention is given on precision and perfection of execution of all production operations. It goes on removing the causes of aircraft accidents caused by errors at production, assembly and maintenance of aero engines. The detail research aimed to improvement of safety was carried out for the GE Aviation Czech company in Praha. It was directed to the aero engine production at which it is necessary to observe the technological procedures and to eliminate the human factor that influences the production process. In the paper there are given results connected with the production of aero engine protective cowling, namely specifically at welding the basic components. It contains special checklist and results of safety audit.

1 INTRODUCTION

In the aero industry, the safety is on the first place. From this reason, at aircraft component production, the great attention is given on precision and perfection of execution of all production operations. It goes on elimination of causes of aircraft traffic accidents caused by faults at production, montage and maintenance of aircraft engine (GE Aviation 2015). The detailed research aimed to safety improvement was performed for the GE Aviation Czech in Praha. It was aimed to domain of aircraft engine production in which it is necessary to respect the technological procedures and to eliminate the human factor troublesome impacts, which influence the production process.

In present paper, the partial results of research are given. These results are connected with the production of protective cowling that is located inside the aircraft engine, and it is attached to the outer aircraft engine shell that fenced the capacity turbine. Its task is to protect the aircraft engine against fragments of blades, discs or of other parts. On the basis of evaluation of information in technical operation notebooks (GE Aviation Czech 2017a) we concentrated the great attention to welding of fundamental components of protective cowling.

For identification and assessment of risks there are used the risk engineering methods, namely the process maps, check lists and safety audits (Procházková 2011a). With regard to the production technological procedure basis (GE Aviation Czech 2016) and the production technical documentation (GE Aviation Czech 2017a), the detailed check list was processed for safety audit that was handed over to experts of the Company Technical Board to the review. After its permission the safety audit was realised with help of checklist under account. The outcomes for individual items of check list at the safety audit were determined as median from results obtained from 3 specialists (company auditor, supervisory department leader, state inspector). By evaluation of audit results the critical spots of welding process were found. Consecutively, the measures for safety improvement of both, the production and the product, were proposed in cooperation with company experts.

2 SAFETY CULTURE, LOSS PREVENTION AND PROCESS SAFETY

The culture denotes the specific material and spiritual values that the humans create by their activities and by which they enhance the life of both, the humans and the whole human society. The society culture is an integral system of substances, values and societal norms, which the members of a given society follow, and which through sharing they transmit to next generation. It is the collection of values, symbols, company heroes, rituals and own histories that act upon exterior, and they have big influence on human behaviour at working positions (Procházková 2011b).

On the basis of just given culture definition, the safety culture means that the human at all his / her roles (control worker, employee, employer, citizen or disaster victim) respects the safety culture, i.e. he / she behaves in a way so he / she may not cause to happen the possible risks realisation, and if risk realisation happens, he / she may contribute to the effective response, the protective interests´

renovation and to start of further development. According to some authors it goes on set of attitudes, surmises, norms and values that exist in a given entity. It is the reflection of way, by which the company is ruled, i.e., they include the general principles of separation of authority and responsibility, principles of management and a certain ration among the accent of working outcomes, authority, care on humans, observance of safety principles and ensuring the given entity functionality (Procházková 2011b).

The effective safety culture is the fundamental element of safety management. It reflects the safety concept and it goes out from values, attitudes and manners of top management workers and from their communication with all involved persons. It is obvious obligation to participate in solving the problems of safety and it promotes so all involved persons perform safely and so they observe the appropriate legal rules, standards and norms. The safety culture rules need to be incorporated into all activities in each entity and in each territory. Their ground is not the concentration to punishment of malefactors / originators of faults, but the lessons learned from the mistakes and the introduction of such corrective measures so mistakes could not repeat, or rather their occurrence frequency might be distinctly reduced.

The safety culture principles with (Česká technologická platforma bezpečnosti průmyslu 2015) are:

1. Outright, open attitude to weak spots, action directed to finding the solution.
2. Diversion from the culture of determination of responsibility for fault and punishment of such person.
3. Employees, employer and top management behave responsibly, separately and with orientation to team. It means that the safety culture is a part of their life.
4. Safety standards are accepted and integrated to everyday company life.
5. Safety and health protection form important value for both, the company workers and the whole company.

The safety culture level is the quantity that cannot be directly and exactly measured, but for all that it has fundamental influence on workers´ behaviours, the management style and the technology level. The definition of weak and strong features in individual parts of safety is important for safety culture level. The comparison of time series of investigations permits to evaluate the effectiveness of corrective measures.

The safety culture for company manufacturing the aircraft engines means that the company undertakes to carry out the manufacturing with the highest safety standards. For reaching such aim, it is crucial and important to have in force the effective and without disincentives the pursued announcement of all accidents, incidents, near misses, random events and cases, experiences, doubts and further information and data that might adversely shaped the element of component of produced aircraft engine. After all, each individual employee is not only kindly encouraged, but also obligated to announce the arbitrary information concerning the safety.

The announcement is not the subject of some imputation and following retributive measure. Its main purpose is the management and govern of risk and preceding the accidents and incidents, i.e. not to impute the blame. They are not made interventions against employee, who reports any data concerning the safety by help of report system, until such announcement without any doubts does not expose that there was commit the criminal act, offensive negligence or intentional and conscious violation of rules or technological procedures. The method of collection, recording and spreading the precaution information secures the protection in the whole width and range according to the law, including the protection of identity of person, who announces the information concerning the safety (Česká technologická platforma bezpečnosti průmyslu 2015). The individual pillars of safety culture are shown in Figure 1.

In link-up with the safety culture there are often in present professional literature connected with technologies used the terms loss prevention

Figure 1. Fundamental pillars of safety culture (DEPOSITPHOTOS 2014). It is necessary to introduce: open communication; accent on prevention; incident analysis; support of safe behaviour; and to provide resources for safety formation.

and process safety. Their definitions we give also for that reason that they are the tools that in connection with technologies serve for protection of persons and property. Loss Prevention is the systematic approach to prevention of accidents, or at least to reduction of their impacts. It includes the means for elimination of sources of risks or for reduction of probability of their realization, and for mitigation of impacts connected with this realization (preventive and consequential measures). Further it includes the identification of suitable supervisory measures, identification and application of suitable remedial measures, by help of which it is ensured the safe entity with appropriate level of security and sustainable development that does not pose unacceptable danger for its vicinity (Procházková 2011c).

The process safety or better the safety of processes is a branch of safety directed to safety in industry, in which there are series of manufacturing and additive processes that are necessary for setting up of final product of a given industry. Together with production it goes on averting the accidents that have special and characteristic features for a given specific industry. It deals e.g. with the prevention of immediate leakage of chemical substances or energies in harmful amount, and in case if such leakages occur with the reduction of their sizes, impacts and consequences. It does not include the questions of classic safety and protection of workers at work, i.e. it deals with purely technical problems, by which it differs from the system safety that is directed to all public assets.

3 AIRCRAFT ENGINE SAFETY

The aircraft is a floating transport vehicle that is heavier than air with the solid wing. It goes on the safe system, which has basic parts: kite (wing, fuselage, tails, flying equipment, landing gear); paraphernalia (life-saving systems, defroster system, air-conditioning, outfit of cockpit, instruments) and propulsive segment (ULT 2014); the schematic technical picture of aircraft is in Figure 2.

So the aircraft might fly, it needs certain flying speed and trend angle, otherwise it might happen the loss of required lift for flight. The required flying speed ensures the tractive component, i.e. the aircraft engine. The aircraft engine consists of generator (engine heart), which is composed from compressor, combustor and generator turbine. The compressor from surrounding atmosphere sucks the air, it compresses it and shift it up to combustor. In combustor it arrives to fuel injection that burns, and developed hot gases drive the generator turbine. Over joint shaft this turbine drives the compressor and by this it obtains the compressed

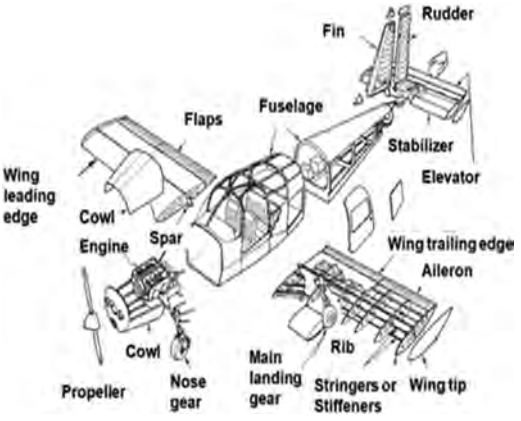


Figure 2. Schematic picture of aircraft (ULT 2014).

air that is necessary for fuel combustion (GE Aviation 2015). For drive of loose turbine that is joined to propeller over the gearbox with constant rate, the aim of which is the reduction of high turns of speed of this loose turbine on the level that is usable for propeller, it is used the remaining energy in hot gases behind the generator turbine (GE Aviation 2015).

The important role at engine protection it fills the protective cowling (GE Aviation 2015). It goes on rigid, i.e. non-rotatory part, the function of which is to catch the prospective fragments of blades, discs or of other parts. It is located inside the aircraft engine, and it is attached to the outer aircraft engine shell that fenced the capacity turbine. Because, it ties to the combustor, from which the hot gases are regulated to the output turbine blades, it is designed by the way, so in addition to absorption of spasmodic performance it would withstand the high temperature and pressure (GE Aviation 2015).

From this reason, the protective cowling needs to have the capability to catch the high kinetic energy. Even if the turbine blade has relatively small mass, so the computation (Marešová 2017) shows that that centrifugal force comes to near 16 000 N, which it is in conversion around 1.6 t.

From the viewpoint of construction of protective cowling, it always goes on compromise. On one side, it is the requirement on sufficiently robust construction, so the safety may be sufficiently ensured. On the other side, it is the requirement, so its mass may be the smallest.

The analysis of faults of rotor (Marešová 2017) shows that on engine failure, they are responsible in 60% fragments of engine parts (56% fragments of blades, 4% fragments of discs, box, gasket etc.), and that the protective cowling does not catch only 15% fragments of blades.

4 DATA FOR PROTECTIVE COWLING

From the technical documentation (GE Aviation 2015), it follows that all parts of hot department of aircraft jet engine, i.e. also the protective cowling need to withstand the impacts of high temperatures and high pressures. From this reason, the only materials´ type, which can be used, is the nickel and cobalt super alloys. To this group, it also belongs the super alloy Nimonic 80 A, which is used for the protective cowling manufacture.

The fabrication of complete protective cowling is time-consuming process. The layout of protective cowling is composed of four parts (Figure 3) that are mutually put together by welding.

The protective cowling composes from 4 parts:

1. Buffer—it is the most important part of protective cowling that comes as the first into contact with blades. Its function is catching the prospective parts of high pressure turbine (blades).
2. Coat—it acts as location of joint (the joint is not part of protective cowling makeup). The joint serves for set up of complete layout of protective cowling with the engine envelope.
3. Whorl—it serves for set up of buffer for tacking to engine envelope.
4. Strut—it serves for set up of coat into proper location compared with buffer.

The manufacturing process of these components includes the great spectrum of manufacturing technologies, as the process map in (Marešová 2017) shows. Among these works, it is possible to find the processing by help of standard lathes or by help of locksmith works, the CNC working as it is use of CNC lathes and CNC milling machines and flection. Further, there are used special working processes as the annealing for remove of internal tension and precipitation case-hardening or the welding by method Tungsten Inert Gas (TIG). The important operations in material processing are inter operational checks for audit of dimensions and correct processing. One of such specific audits, it is special luminescence technology for identification of cracks and surface defects in material. On the basis of analyses of manufacturing procedures in (Marešová 2017) performed on the ground of operation notebooks (GE Aviation Czech 2017a), it was shown that the critical operation is also the TIG welding. From this reason we give below the results of our investigation.

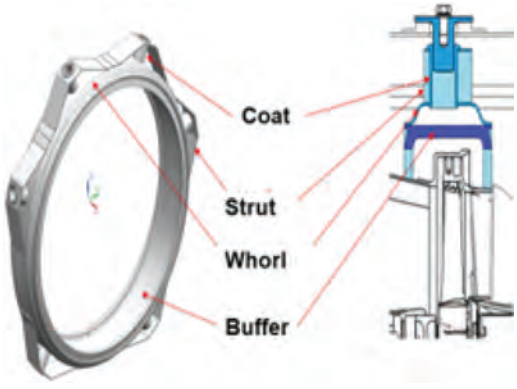


Figure 3. Picture of layout of protective cowling and its components (left) and picture of protective cowling position in aircraft jet engine (right) (GE Aviation Czech 2016).

The TIG welding belongs among the arc welding methods. The electric arc is between the infusible wolfram electrode and molten weld metal in inert protective gas atmosphere. In considered case, there are used argon, helium or their mixtures (Hrivňák 2009).

The TIG welding method was set up at the end of 30 s of twentieth century for welding the magnesium alloys. This welding method partly replaced the fastening with rivets that in that time was the most used method for soldering the components from aluminium and magnesium in aero industry.

The given welding method is the most important at welding the components from stainless steel, aluminium, magnesium, copper and reactive materials (e.g. titanium and tantalum). The thickness of weld materials ranges from several tens of millimetres up to several millimetres (Olson 1993).

The TIG method advantages over against other ways of welding are according (Olson 1993) the following:

1. Manufacture of high quality welds, i.e. low deformation of welded parts, welds with minimum amount of impurities, gasses, respectively pores and cracks by occurrence of which it is reduced the material capacity (resistance to high temperature and high pressure).
2. At welding, the skewness does not origin, and therefore, no need for its removing.
3. Welding is possible to perform with or without additive material.
4. Welding is possible almost for all material kinds and also various materials.
5. Precise audit of weld parameters is possible.

The TIG welding method is used in cases in which there are requirements on high quality of welds. By this method it is possible to weld almost all metal materials. The welder is capable during the welding to perform the precise check of heat taken in weld, because the weld surrounding is not rounded during the welding by vapours and gasses from the process (Olson 1993).

Except of advantages the TIG method has in comparison with other methods also disadvantages, according (Olson 1993) they are:

1. Lower weld output in comparison with other arc welding methods.
2. Higher demands on welder craftsmanship.
3. Higher economic demandingness of production in comparison with welding method by coated electrode.

According to regulation AWS D17.1 (AWS 2013), the welds according to mutual location of welded parts are divided in:

1. Blunt welds—joint of two materials (sheets or pipes), mutually put together by frontal surfaces.
2. Fillet welds—joint of two materials that form an angle and the welding joint is located on edges of these welded parts.

In the next part there are only given results for blunt welds by the TIG method.

5 USED RISK ENGINEERING METHODS

For detection and judgement of risks there were used the following methods: process map, i.e. the scheme picture of production process that shows places in which are possible conflicts in production because of insufficient workplace capacity or bad production planning (ASME 2010); Check list (Procházková 2011a); and safety audit (GE Aviation 2015, Procházková 2011a).

Specific check list was compiled according to manufacturing technique of TIG method (GE Aviation Czech 2016), in which the critical spots were determined on the basis of operating events company book (GE Aviation Czech 2017a). It is given in Table 1.

The judgement of safety audit results was done according the ČSN OHSAS 18001scale, Table 2 (Procházková 2013).

Table 1. Check list used for risks evaluation in welding process; SN—serial number, Y—YES, N—NO, R—remark.

SN	Question	Y	N	R
Preparation before welding				
1	Are connected surfaces of welded parts metallically unalloyed, i.e. are they ridded of tips and rough layers of oxide?			
2	Are connected surfaces of welded parts ridded of smear?			
3	Are connected surfaces of welded parts ridded of other dirt affecting the final weld quality?			
4	Is set up of welded parts, i.e. the size of weld space, in concord with requirements of regulation?			
5	Is set up of welded parts, i.e. the size of overlapped surfaces, in concord with requirements of regulation?			
6	Are the edges of connected materials in place of future weld ridded of splinters?			
7	Are the edges of connected materials in place of future weld without gross shrink?			
8	Are welded parts put down on unpolluted surface of work bench?			
9	Is it manipulated with welded parts in clean cotton gloves that do not release fibres?			
Machines and fittings				
10	Are used for welding only certifiable machines and apparatuses?			
11	Are used for welding only calibrated machines and apparatuses?			
12	Are used for welding only certifiable machines and apparatuses determined in technological procedure (SMC 2004)?			
13	Is the welding source capable to be continuously regulated in the whole range of values of welding parameters given in technological procedure (FLICKER 2012)?			
14	Is the source of protective gas capable to be continuously regulated in the whole range of values of welding parameters given in technological procedure?			
15	Is the source of electric current for welding without symptom of defect?			
16	Is attachment for welding without symptom of defect?			
17	Are cables for welding without symptom of defect?			
18	Is the underlay piece on working table, on which the welded components are located, flat, without sharp roughness and dents?			
19	Is the underlay piece on working table, on which the welded components are located, from the electrically conductive material?			
20	Are the staples of welding apparatus located in close contiguity of welded components?			
21	Is ensured the sufficient conductivity between staple and the underlay piece on working table?			
22	Does the protective gas quality fulfil the minimal demands on pureness according to norms in force and rule (SMC 2004)?			
23	Is used the protective gas according to technological procedure (SMC 2004) with obligatory certificate?			

(*Continued*)

Table 1. (*Continued*).

SN	Question	Y	N	R
24	Is welding burner range for a given electric current load?			
25	Are cable ranges for a given electric current load?			
26	Is the burner spout without noticeable defects?			
27	Is the spout diameter in the range determined in technological procedure (SMC 2004), which enabling the supply of sufficient amount of protective gas for the effective protection of welded metal against influences from surrounding atmosphere?			
28	Does the wolfram electrode type correspond with demands given in technological procedure (SMC 2004)?			
29	Does the wolfram electrode diameter correspond with demands given in technological procedure (SMC 2004)?			
30	Is the wolfram electrode also ranged for maximal electric current load?			
Welding personnel				
31	Does the welding perform the qualified welding personnel?			
32	Is the welding personnel properly trained?			
33	Does welding personnel keep the welding certificate?			
34	Does welding personnel keep the valid medical check?			
Welding process				
35	Are in the case of requirements on tacking, the stitches uniformly arranged?			
36	Are the stitches without cracks and craters?			
37	Is it used as the additive material only certificate weld wire?			
38	Is the additive material degreased?			
39	Is the additive material free from dust and dirt?			
40	Is the additive material properly branded?			
41	Is the additive material stored in original packing?			
42	Does only use the additive material that is properly branded?			
43	Is it suppressed to use the unbranded additive material?			
44	Is used for welding the weld wire with demanded chemical texture?			
45	Is used for welding the weld wire given in technological procedure (SMC 2004)?			
46	Is used for welding the weld wire with diameter in demanded range?			
47	Is the welding current regulated according to the technological procedure or the weld certificate (SMC 2004)?			
48	Is the welding speed set according to technological procedure or the weld certificate (SMC 2004)?			
49	Is the surface of welding formation cleaned after welding by stainless brush?			
50	Is the part influenced by heat cleaned after welding by stainless brush?			
Check after welding				
51	Is the visual check of weld performed in workplace that is determined for such check?			
52	Are accessible the records on minimal demanded value of lighting intensity (300 lx)?			
53	Is the visual check by purely eye performed according to rule for visual weld check?			
54	Is the visual check performed by magnifying glass with demanded enlargement according to rule for visual weld check (SMC 2004)?			
55	Is the personnel pursuing the visual weld check qualified?			
56	Is the personnel pursuing the visual weld check trained?			
57	Are records from weld check get-at-able for next check?			
58	Are in the case of detection of weld defects, these weld defects eliminated?			
59	Are the weld defects eliminated only in range permit by a given rule?			

Table 2. Risk rate according to scale (Procházková 2013).

Risk rate	Number of answers "NO" v%
Extremely high	More than 95%
Very high	70–95%
High	45–70%
Medium	25–45%
Low	5–25%
Negligible	Lower than 5%

6 SAFETY AUDIT RESULTS AND PROPOSALS FOR SAFETY IMPROVEMENTS

The check of welding was performed on workplace "Svarovna" under supervision of technologist and auditor. In case of answer "NO", it was given argument, why this is. The judgement was performed by auditor. Except of judgement by check list, there were drawn up the photo documenta-

tion and taken down the record, that was signed by all participants, i.e. also by welder (GE Aviation Czech 2017c).

From the record (GE Aviation Czech 2017c) it follows that that at answers to 59 questions, six answers were "NO", i.e. 10%. From the risk rate viewpoint, it goes on low rate of criticality.

For removing four problems that affect the weld quality, there were proposed the measures:

1. Problem 1 – the connected spaces of welded parts are not before itself welding ridded of tips, rough layers of oxide, smear and further dirt negatively affecting the quality of final welds (GE Aviation Czech 2016, Hrivňák 2009, Olson 1993). Before the welding, the welded parts are only washed in cleaner bath. On the surface of these parts the residues of cleaner emulsion are evident.
   Proposal for improvement:
   For reach of high quality of weld joint, it is necessary from the surface of parts in place of future weld to remove all grimes. The surface oxides and apposite tips may be removed mechanically by help of stainless brush or SiC abrasive material (abrasive canvas). The part degreasing is suitable to perform chemically, i.e. by acetone or technical ethyl alcohol. This is important, because the smear, especially the sulphur reacts with nickel in welding bath and causes crack in a given weld.
2. The welded parts need to be set up from the reason of reaching the acceptable weld slit (GE Aviation Czech 2016, 2017a).
   Proposal for improvement:
   For reach of high quality of weld joint, it is necessary so weld slit at rim welds was minimal, ideally null-space. The reason is reality that at stiffening the weld metal it arises to its contraction, and by this to origin of tensile stresses. The higher the weld slit is the higher tensile stresses originate. When the size of originated tensile stresses exceeds the material solidity limit, it goes to origin of cracks. From this reason, the quality of performed welds is influenced not only by itself welding process, but also the preparation, i.e. make-up of parts. To reach demanded make-up, the technological procedure (SMC 2004) needs to be kept, i.e. it is necessary to keep the set tolerance, which goes to increase of exactness of dimensions of produced parts.
3. The additive material for welding is used in form of rod with diameter Ø 2.0 mm. In spite of the fact that weld with additive wire Ø 2,0 is certificated as convenient, so for welding the materials from which the protective cowling is produced it would be more suitable to use the additive wire with lower diameter, ideally Ø 1,6 mm. The reason is the lower amount of heat necessary for smelting the welding wire, and so lower amount of heat carried into weld. By this way the part of weld vicinity (TOO) affected by heat is smaller (GE Aviation Czech 2016, SMC 2004). In our case it goes on welding the case-harden material, in which the cracks just originated in the TOO (GE Aviation Czech 2017a), i.e. the smaller TOO, the lower is the probability of defect origin.
   Proposal for improvement:
   On the Czech Republic territory, the supplier of additive material with demanded thickness 1,6 mm is not. The foreign suppliers offer this material in packets with great amount, which leads to wastage. From this reason this proposal has not been realised yet.
4. For execution of welding, it is not positional fittings. By producing such positional fittings with adjustable speed of rotation, it would reach the facilitation of work, increase of comfort, and primary the improvement of work, and by this also the weld (Marešová 2017). The positional fittings can be used for welding the other welded part of aircraft engine.
   Proposal for improvement:
   The purchase of new positional fittings and the judgement of its influence on improvement of weld quality and work ergonomics are in run.

7 CONCLUSION

On the basis of project on co-operation of our faculty with the GE Aviation Czech, there are step by step created the tools for decrease of operational risks with aim to ensure the high safety of aircraft engines. One of such tool connected with welding the protective cowling is described above. The safety audit performed with specific check list revealed four defects. From the safety reasons three defects were fast rectified; the last one is in handling due to the present inaccessibility of economically acceptable product.

ACKNOWLEDGEMENT

Authors thanks to Authors thank for grant to the EU and the Czech Ministry for Education, the RIRIZIBE project, CZ.02.2.69/0.0/0.0/16_018/0002649.

REFERENCES

ASME, 2010. *The Impact Load on Containment Rings During a Multiple Blade Shed In Aircraft Gas Turbine Engines*. http://asme.org

AWS, 2013. *D17.1/D17.1M:2010-AMD1. Specification for Psion welding of aerospace and applications.* Miami: FL: American Welding Society.

Česká technologická platforma bezpečnosti průmyslu, 2015. *Co je to kultura bezpečnosti*. http://www.cztpis.cz/safety-culture-award/kultura-bezpecnosti/

DEPOSITPHOTOS 2014. *Kultura bezpečnosti – Stock obrázek*. http://cz.depositphotos.com/65417987/stock-photo-cult ure-of-safety.html

FLICKR 2012. *Report 4310481794*. https://www.flickr.com /photos/ browndogwelding/431 0481794

GE Aviation 2015. *Výroba turbovrtulového motoru v 6 krocích.* Praha: GE Aviation. https://www.geaviation.cz/cla nky/detail/35_95-vyroba-turbovrtuloveho-motoru-v-6-krocich

GE Aviation Czech, 2016. *Technologický postup výroby ochranného krytu*. Praha: GE Aviation Archives.

GE Aviation Czech, 2017a. Technické provozní deníky. Praha: GE Aviation Archives.

GE Aviation Czech, 2017b. *Technická dokumentace*. Praha: GE Aviation Archives

GE Aviation Czech, 2017c. Zápis z auditu, 24. 4. 2017. Praha: GE Aviation Archives.

Hrivňák, I., 2009. *Zváranie a zvariteľnosť materiálov.* ISBN 978–802–2731–676. Bratislava: STU, 126p.

Marešová, Š. 2017. Nástroj ke snížení rizik při výrobě specifických dílů pro motor letadla. Diplomová práce. Praha: ČVUT, fakulta dopravní, 66p.

Olson, D. L., 1993. *ASM Handbook, Volume 6: Welding, Brazing, and Soldering*. ISBN 978–0-87170–382–8. Washington: ASM, 168p.

Procházková, D. 2011a. *Metody, nástroje a techniky pro rizikové inženýrství*. ISBN 978–80–01–04842–9. Praha: ČVUT, 369p.

Procházková, D. 2011b. *Ochrana osob a majetku*. ISBN 978–80–01–04843–6. Praha: ČVUT, 301p.

Procházková, D. 2011c. *Analýza a řízení rizik*. ISBN: 978–80–01–04841–2. Praha. ČVUT, 405p.

Procházková, D. 2013. *Základy řízení bezpečnosti kritické infrastruktury*. ISBN 978–80–01–05245–7. Praha: ČVUT 2013, 223p.

SMC, 2004. *Publication Number SMC–099. Nimonic alloy 80A.* http://www.haraldpihl.se/globa lasets/pdf/057nimonic- alloy-80a.pdf

ULT, 2014. Letadla. http://www.aerospace.fsik.cvut.cz

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Reliability of supplies in a manufacturing enterprise

J. Żurek & M. Zieja
Air Force Institute of Technology, Warsaw, Poland

J. Ziółkowski
Military University of Technology, Warsaw, Poland

ABSTRACT: The article brings closer the methodology of material requirement planning in a manufacturing enterprise. A manufacturing structure consisting of two independent products, A and B, showing modularity of design was assumed. A mathematical description reflecting the dependencies of individual product components used to execute a delivery schedule was used for the calculations. The article describes how to plan the production of products with a module characteristics and determine its supply moment so as to maintain a continuity and rhythmicality of a manufacturing process. It is of utmost importance due to the prompt execution of external orders (independent demand) and thus it has an impact on the dependability of the execution of orders. The suggested methodology of proceedings eliminates errors, reduces the stock level, enables to fully control particular production stages, improves the utilization of the available resources and synchronizes both order and delivery processes of materials with production needs.

1 MATERIAL REQUIREMENT PLANNING IN TERMS OF RELIABILITY OF SUPPLIES

The functional specificity of industrial companies involves processing raw materials into finished products, as a result of a creative process. This process is associated with changing the form, shape or properties of a product defined as a final product.

The aim of the article is to present a methodology for describing products of a modular structure, and then to utilize it to develop a delivery schedule, ensuring reliability of manufacturing orders fulfilment. It can be achieved through timely execution of assembly tasks, precise determination of stock levels, defining the size of manufacturing batches, and synchronizing and control of the manufacturing-assembly process.

Material Requirement Planning (MRP) is defined by the literature in an ambiguous manner. In a book by Cecil Bozarth and Robert B. Handfield, MRP is defined as: "a planning process enabling to translate the superior manufacturing plan onto the planned orders for the parts and components necessary to manufacture products, of which the completion was included in the superior plan" (Bozarth & Handfield 2007). Whereas Donald Waters, in his book, state that: "(…) material requirement planning utilizes the main manufacturing plan in order to schedule the material supply. Expanding the main manufacturing plan allows planning material supplies at the very moment they are needed" (Waters 2007).

B. Śliwczyński has a slightly different perception of the MRP concept, writing: "(…) planning material requirement covers every element of the final product, in each of the manufacturing stages and defines the material requirements, resulting from the product range, volume and lead time of a manufacturing batch. The lead times of deliveries of the materials and elements necessary to manufacture a finished product are calculated as per the MRP method, according to the main manufacturing schedule. As a result of material requirement planning, a delivery schedule is developed, which is the basis for material supply planning" (Śliwczyński 2008). On the other hand, S. Krawczyk positions MRP within the process of handling a manufacturing enterprise, therefore, relates it to, i.a., the level of customer service. Therefore, the author draws attention to the aspect concerning delivery reliability, which should be understood in a slightly narrower sense as timeliness (Krawczyk 2011). The concept of supply reliability, related to a company, may and should be considered on many levels. For example, supply reliability may be shaped by: lack of damage, relevance, timeliness, conformity, completeness, etc. T. Nowakowski indicates that supply reliability is impacted by: its timely execution, completeness of an order and releasing/receiving undamaged goods (Nowakowski 2004 &2011). J. Żurek points to a

direct relationship between reliability and readiness, and relates these terms to operational processes of facilities and technical systems (Niziński & Żurek 2011, Żurek & Jażwiński 2007). With regard to the reliability of systems associated with stock control, the author implies that one of the basic delivery reliability methods is to maintain an excess stock, understood as a safety reserve. Furthermore, the author differentiates between stock levels, depending on the cost criterion (prices). The author states that stock of cheap elements may be maintained at a high sufficiency level, while the stock of expensive elements at a low or zero sufficiency level. Very expensive elements are often purchased only in the event of a demand (Waters 2007). S. Werbińska-Wojciechowska points also to time redundancy in the aspect of evaluating availability of logistics system resources in companies, drawing attention to the issues of temporary stock availability (Werbińska-Wojciechowska 2007, 2013).

An MRP system improves stock management and facilitates the creation of requirement plans for raw materials and materials necessary for manufacturing, for which the demand depends on market needs. They make up the manufacturing schedule, defined by calendar time, which takes into account both the assembly labor consumption, as well as the lead times of component deliveries. Therefore, its correct execution impacts the delivery reliability understood (to a narrow extent) as timeliness of executing manufacturing tasks. Basic information of the Material Requirement Planning system (MRP) form information streams, with their set containing (Bozarth & Handfield 2007, Krawczyk 2011, Skowronek & Sarjusz-Wolski 2012, Śliwczyński 2008):

- main manufacturing schedule;
- product structure;
- stock main set.

MRP system's information streams are presented in Fig. 1.

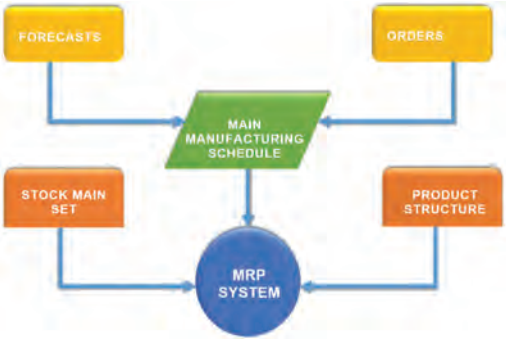


Figure 1. Information streams supplying an MRP system.

As indicated in Fig. 1, information streams making up an MRP system are formed by: the main manufacturing schedule, which is developed in the long term on the basis of sales forecasts, and in the shorter time horizon, it is confirmed by orders coming from the customers. The second stream is a set of stock, expanded by assembly-delivery lead times, while the third one is the modular structure of a product. The components described above shall be used for practical development of a delivery schedule.

2 SCHEDULING DELIVERIES IN PRACTICE

An essential component of an MRP system is the manufacturing schedule, which includes orders from customers, concerning the required number of products and the order lead time. In this context, a correctly drawn up delivery schedule ensures reliability, understood as delivery timeliness. In order to develop such a schedule, we also need the structure of products A and B (Fig. 2 and Fig. 3), and its description method, which shall be utilized for the calculations.

In structural terms, the presented final product consists of products A and B, which include repeating structural fragments, called modules, appearing in at least two locations within a discussed structured.

Two types of functional modules may be distinguished in the structure of product A:

- module of the 1st order C-1 (consists of the mentioned module of the 2nd order D-2 and four elements marked as F);

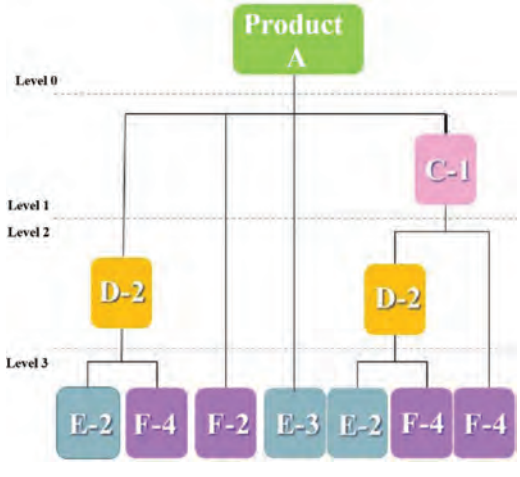


Figure 2. Product A structure.

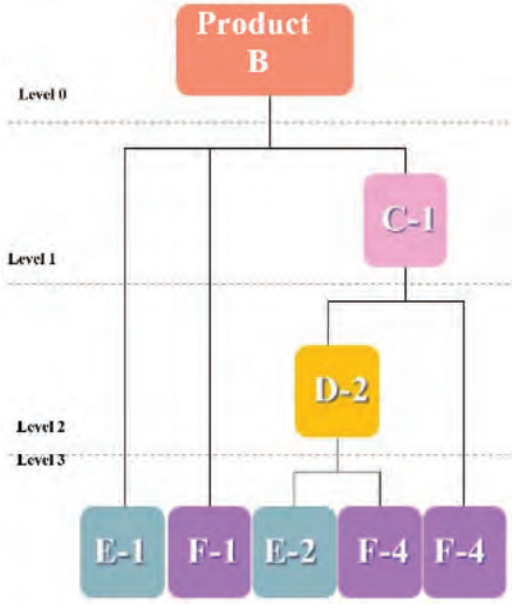


Figure 3. Product B structure.

- module of the 2nd order D-2 (consists of two E type elements and four F type elements).

Furthermore, product A consists of elements located within the structure on three functional levels. The first level contains a C-1 type module, the second one represents the D-2 module, while the third one contains F-2 and E-3 type elements, independent in structural terms, subordinate directly to product A.

The structure of product B should be discussed by analogy (Fig. 3). We should remember that products A and B located in level 0 are structurally independent in relation to each other, but combined form a single final product. The indicated modularity of both products causes a specific description of their structure, used for calculations in the delivery schedule (MRP). Below you can find a presentation of the description methodology, according to which such a schedule is to be executed, starting with level 0, and ending with level 3.

The first structural level of both products making up a finished product, contains one C-1 type module (it is a so-called module of the first order). The second structural level contains a D-2 type module (a so-called module of the second order). The methodology of describing such a structure is as follows:

- products A and B are located at the same level (level 0) and there are no structural links between them, therefore, they are not subject to a description;
- level 1 in both products contains a C type module, the second level contains a D type module, and the third level contains E and F type elements, which are indivisible in terms of structure;
- the fact that module C is present in products A and B is recorded as follows: C(A, B);
- the second product structure level contains a D-2 type module, whereas in relation to the structure of products A and B it has a slightly different location, that is, in the case of product A it is directly subordinate to A (level 0), while in relation to product B, it constitutes a part of a C type module (level 2), and this fact shall be recorded as D(2A, 2C);
- the last level (level 3) contains elements, which should be described according to the principles set out above, therefore: E(3A, B, D) and F(2A, B, 4C, 2D).

Table 1. Main manufacturing schedule.

Gross demand					
Product/ module/ element	Week				
	6	7	8	9	10
A	1254	123	124	125	870
B	456	543	560	678	340
D	274	228	132	143	270
E	376	223	345	678	450

Table 2. Company's stock level.

Available stock							
Product/ module/ element	Week						
	1	2	3	4	5	6	7
A	556						70
B	234						60
C	523						50
D	345						60
E	123						70
F	321						72

Table 3. Delivery lead times.

Delivery time	
A	2 weeks
B	2 weeks
C	1 week
D	1 week
E	1 week
F	1 week

Table 4. Material requirement plan for products A and B.

MRP	Dates/weeks	1	2	3	4	5	6	7	8	9	10
A	gross demand	0	0	0	0	0	1254	123	124	125	870
	available stock	556	556	556	556	556	556	70	0	0	0
	net demand	0	0	0	0	0	698	53	124	125	870
	order	0	0	0	698	53	124	125	870	0	0
B	gross demand	0	0	0	0	0	456	543	560	678	340
	available stock	234	234	234	234	234	234	60	0	0	0
	net demand	0	0	0	0	0	222	483	560	678	340
	order	0	0	0	222	483	560	678	340	0	0
C (A,B)	gross demand	0	0	0	920	536	684	803	1210	0	0
	available stock	523	523	523	523	0	0	50	0	0	0
	net demand	0	0	0	397	536	684	753	1210	0	0
	order	0	0	397	536	684	753	1210	0	0	0
D (2A,2C)	gross demand	0	0	794	2468	1474	2028	2898	1872	143	270
	available stock	345	345	345	0	0	0	60	0	0	0
	net demand	0	0	449	2468	1474	2028	2838	1872	143	270
	order	0	449	2468	1474	2028	2838	1872	143	270	0
E (3A,1B,D)	gross demand	0	449	2468	3790	2670	4146	3148	3438	948	450
	available stock	123	123	0	0	0	0	70	0	0	0
	net demand	0	326	2468	3790	2670	4146	3078	3438	948	450
	order	326	2468	3790	2670	4146	3078	3438	948	450	0
F (2A,1B,4C,2D)	gross demand	0	898	6524	6710	7381	9496	9512	2366	540	0
	available stock	321	321	0	0	0	0	72	0	0	0
	net demand	0	577	6524	6710	7381	9496	9440	2366	540	0
	order	577	6524	6710	7381	9496	9440	2366	540	0	0

The above methodology of description shall be used for calculations in a Manufacturing Requirement Plan (MRP).

Table 1 presents the planned manufacturing of products A and B, a D type module and an element marked as E.

Table 2 presents the stock level of a company regarding individual products, modules and elements on week one and seven. They are used to calculate the net demand, which is defined as the difference between the gross demand and available stock.

Table 3 presents the delivery times for products, modules and elements, which enable the determination of an exact moment, so as to ensure manufacturing continuity and an effective utilization of available stock.

Table 4 below, presents a Material Requirement Plan (MRP), for the development of which the data summarized in Tables 1–3 was used. The above description methodology, which is characteristic for the structure of products A and B showing modularity, was used to calculate the gross demand. This methodology applies to the description of modules marked as C-1 and D-2, and the elements defined as E and F, therefore, it relates to the components assigned to levels 1–3 of a product structure (cf. Figs. 1 and 2).

The presented material requirement plan was developed for a time horizon covering a range of 1–10 weeks. The gross demand for products A and B was submitted in weeks 6–10. After taking into account the delivery lead times and assembly times, the generated internal orders actually covered a time period of 1–4 weeks. It is associated with the so-called reverse planning, and a bill of material, related to the structure of the product in each case.

3 CONCLUSIONS

The objective of this paper was to develop a descriptive methodology for products with a modular structure and its application in manufacturing scheduling. The planned objective was achieved, and the article presents the manner, in which to plan manufacturing products of a modular characteristic and how to determine the moment to order them, with the purpose to maintain continuity and rhythm of the manufacturing process. It is extremely important from the point of view of timely processing of external orders (independ-

ent demand), therefore, it impacts the reliability of delivery execution. The presented methodology eliminates mistakes, decreases the level of maintained stock, enables full control over individual manufacturing stages, improves the utilization of available resources and synchronizes the material ordering and delivery processes with manufacturing needs. The methodology was developed in 1957 by APICS (American Production and Inventory Control Society), and its particular development fell on the 1970s and 1980s. It is dedicated to a variable demand, characterized by irregularity, and an enterprise manufacturing products on a mass scale. For this reason, in the first place it should be applied to serial or direct-line manufacturing, and second of all, to unit manufacturing or a combination of the aforementioned types.

REFERENCES

Bozarth, C. & Handfield, R.B. 2007. *Introduction to operational and supply chain management*. Gliwice: Helion.

DeLurgio, S.A. 1998. *Forecasting principles and applications*. University of Missouri-Kansas City: Irwin/McGraw-Hill.

Jacyna-Gołda, I. & Izdebski, M. & Podviezko, A. 2017. Assessment of efficiency of assignment of vehicles to tasks in supply chains: A case study of municipal company. *Transport* 32(3): 243–251.

Jacyna-Gołda, I. & Lewczuk, K. 2017. The method of estimating dependability of supply chain elements on the base of technical and organizational redundancy of process. *Eksploatacja i Niezawodnosc-Maintenance and Reliability* 19(3): 382–392.

Krawczyk, S. 2011. *Logistics. Theory and practice*. Warsaw: Difin.

Niziński, S. & Żurek, J. 2011. *General Logistics.* Warsaw: WKiŁ.

Nowakowski, T. 2004. Reliability model of combined transportation system. *Probabilistic Safety Assessment and Management.* London: Springer.

Nowakowski, T. 2011. *Reliability of logistics systems.* Wrocław: OWPW.

Pham, H. 2006. *Handbook of Engineering Statistics*. London: Springer-Verlag.

Skowronek, C. & Sarjusz-Wolski Z. 2012. *Logistics in a company,* Warsaw: PWE.

Śliwczyński, B. 2008. *Logistics planning, Poznań*: ILiM.

Tomaszek, H. & Zieja, M. & Ważny, M. 2016. A method for reliability assessment of structural components of aircraft and sea-going ships with taking into account a given failure generation model. *Polish Maritime Research* 23(2): 83–90.

Waters, D. 2007. *Operational planning. Goods and services.* Warsaw: PWN.

Werbińska-Wojciechowska S. 2013. Time resource problem in Logistics systems dependability modeling, *Eksploatacja i Niezawodnosc–Maintenance and Reliability 15(4)*.

Werbińska-Wojciechowska, S. 2007. The availability model of logistic support system with time redundancy. *Eksploatacja i Niezawodnosc-Maintenance and Reliability*.

Zieja, M & Ważny, M. & Stępień, S. 2016. Distribution determination of time of exceeding permissible condition as used to determine lifetimes of selected aeronautical devices/systems. *Eksploatacja i Niezawodnosc-Maintenance and Reliability* 18(1): 57–64.

Zio, E. 2009. *Computational Methods for Reliability and Risk Analysis*. Singapore: World Scientific Publishing.

Żurek, J. & Jaźwiński, J. 2007. *Selected issued of stock control,* Warszawa-Radom: Wydawnictwo Instytutu Technologii Eksploatacji (PIB).

Żurek, J. & Smalko, Z. & Zieja, M. 2010. Methods applied to identify causes of air events. *Reliability, Risk and Safety: Theory and Applications*. CRC Press-Taylor and Francis Group: 1817–1822.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Swimming in a slurry of schemes: Making sense of aquaculture standards and certification schemes

M. Nilsen
Studio Apertura, NTNU Social Research, Norway

V.S. Amundsen & M.S. Olsen
Department of Sociology and Political Science, Faculty of Social and Educational Sciences, NTNU—Norwegian University of Science and Technology, Trondheim, Norway
Studio Apertura, NTNU Social Research, Norway

ABSTRACT: Growth in the number of certification schemes in the aquaculture industry has been attributed to several factors. The schemes contribute to improved traceability of products, provide healthier stocks, and provide more information to customers' decision-making efforts. There is a wide range of certification schemes and standards available, addressing food safety, environmental impact, animal welfare, and worker conditions, to name a few. The abundance of certification schemes has resulted in concerns about consumers becoming confused with the number of labels and that certification schemes themselves may become a barrier to trade. This paper examines 5 major certification schemes in the aquaculture sector and categorizes them according to their purpose, proprietorship, and process. We investigate what has caused this wave of attention to be given to such a diverse range of issues, exploring how the diversity of these certifications is rooted in their inception and the areas they address.

1 INTRODUCTION

By 2050, the world population is predicted to have increased from 7.6 billion people to 9.8 billion (UN 2017). This implies that the need for fish as a source of nutrition will increase, and with that, there will be increased challenges for wild catch and production.

Capture fisheries have become stagnant since the 1980s, while aquaculture of fish and shellfish has more than doubled its growth in the last quarter of the twentieth century (FAO 2016). Salmon is one of the species that has seen spectacular growth, especially in Norway, Chile, Canada, and the UK (FAO 2003).

The growth of aquaculture production plays an important part in international trade and has helped the economy in many developing countries (Prein and Scholz 2014). However, this growth does not come without negative consequences to people or the environment. The "blue revolution" calls for problems to be addressed, such as water pollution, ecosystem degradation, and poor labor conditions. The rapid growth of the salmon farming industry has in many countries raised public concern and critique from stakeholders and politicians regarding social, economic, and environmental impacts. The concerns are both country-specific and/or global, from the effects of aquaculture on biodiversity and wild fish stocks to socio-economic impacts (e.g. competition for ocean space, land, and property value) (Bush et al. 2013). Asche et al. (1999) categorized salmon farming´s sources of environmental problems into three categories: (1) organic material emission; (2) spread of diseases that may affect wild species; and (3) genetic contamination of wild stocks by escapees.

The critiques of salmon aquaculture, combined with a general increased focus on environmental and social issues, have led to a rise in public awareness and a demand for a more sustainable industry (Prein and Scholz 2014). Despite a unified call for 'sustainability', there lacks a shared consensus as to what that actually entails and how it can be accomplished (Davidson 2010). With little agreement beyond the common notion of the three dimensions of sustainability: *environmental* (ecosystem and biodiversity), *economic* (long-term business viability), and *social* (social responsibility and community well-being) (World Bank 2014), the road to 'a sustainable industry' has become a vague and ambiguous one.

While the main production of salmon aquaculture is found in Norway, Chile, the UK, and Canada, farmed salmon is sold to more than 100 countries worldwide. Stakeholders are therefore not only from the producing countries but from quite a large, global marketplace. With demands for sustainability coming from, and the actual production happening in, very different corners of the

world, there has been an increased need for global consistency in the regulation of the industry (Busch 2011, Stanton 2012).

An effort to achieve this is through the use of global standards, certification schemes, and labeling created by NGOs and retailers (e.g. IKEA, Tesco). These are a form of private governance or 'soft law', which entails that their sanctions do not carry the force of law and are therefore not mandatory (Busch 2011). Certification schemes provide different standards for which the producers can voluntarily choose to comply, and in doing so obtain a certification from the chosen scheme. In Europe, the most prevalent standards in aquaculture are the GLOBALG.A.P. Aquaculture Standard and the Aquaculture Stewardship Council (ASC) standards. In North America, on the other hand, the standards set by Global Aquaculture Alliance, the Best Aquaculture Practice, are widely used (Prein and Scholz 2014).

In recent years, the number of certification schemes for food production and processing has increased significantly, along with a variety of actors involved in the development of these standards. Attempting to cover the many rising challenges in aquaculture, these standards and labels relate to issues such as sustainability, food safety, organic production, etc. As a consequence, the types of schemes, their objectives, and their scope vary considerably (Nadvi and Wältring 2002).

This paper aims to illustrate the multitude of standards existing in the market today. As seen from the literature, there is a wide range of certification schemes and standards available and the arguments for the development of these vary between the need for consumer legitimacy, market demands, quality improvement, etc. This paper explores what has caused this wave of attention given to such a diverse range of issues, which has led to this sea of certifications. By doing a comparison of the *proprietorship*, *process,* and *purpose* (hereafter referred to as the 3 P´s of certification) for 5 major certification schemes in use for salmon aquaculture, we seek to understand how differences in their standards and their focus areas can be related to their origin. What arguments are being used for each certification scheme/standard, and why do they differ in focus and demand for improvement?

2 BACKGROUND

Certification and labeling are one type of signal or attribute giving the consumer the opportunity to evaluate a product before purchase/consumption (Chen et al. 2015). FAO differ between ecolabels, and food safety and quality standards (Washington and Ababouch 2011). Ecolabels, also referred to as 'best practice' labels, focus on responsible aquaculture practices, procurement policies of retailers/brand owners, and support to consumers in their purchasing decisions. The food safety standards are schemes that provide assurance in the quality and safety of products and the processes involved.

Numerous reasons for the emergence of such certification schemes have been identified, seen both from consumers, market actors (e.g. retailers), and producers. One argument focuses on a lack of sufficient regulation, arguing that these certification schemes have emerged where the public regulation is perceived as inefficient or ineffective in their response to food safety, quality, and environmental sustainability (Washington and Ababouch 2011).

For the retailers and companies selling seafood, labels are also viewed as a mechanism to reduce risk related to negative publicity concerning production practices (Boyd and Nevin 2011). Achieving trust from consumers and supporting producer legitimacy are an important part of certification schemes (Bush et al. 2013). Summarized by Morris (1997), the possibility to improve the image and/or sales of a company, in addition to encouraging firms to account for the environmental impact of their production, are important arguments to support certification schemes.

Certification usually provides product traceability, standardization among global suppliers, and transparency of production processes (Washington and Ababouch 2011). Standardization can be seen as a form of risk management that extends a company´s liability to a third-party Certification Body (CB), thus, allowing the company to claim due diligence in the event of a predicament (Busch 2011). In addition to allocation of risk, certification may also deter "real and/or perceived risks along the food chain" (Stanton 2012: 247).

Nevertheless, there are uncertainties about the certification schemes' consequences for sustainability. There is little scientific proof that shows a reduction of negative environmental impacts by certified farms compared to noncertified farms (Boyd and Nevin 2011). Though it might be likely to reduce impact on a farm level, this may not contribute to an overall improvement in sustainability (Tlusty & Thorsen 2017). Questions have also been raised as to whether the increased demand for documentation and record-keeping of the aquaculture companies through these schemes actually are making the production more sustainable (Bush et al. 2013).

Another concern regarding certification schemes is that they may act as a barrier to trade for smaller companies or companies from developing countries who cannot afford the costs and documentation requirements of standards originating in the industrialized countries (Busch 2011).

Although private standards are not legally required, international markets demand that companies comply with supposedly voluntary stand-

Table 1. Various schemes and their characteristics.

Scheme and relevant standard	Origin	Year*	Objectives	S	QS	E	AW	3rd party	Stakeholders	Coverage
GLOBAL-G.A.P. Aquaculture Standard	European retailers (EUREGAP-1997)	2004	Safe, sustainable agriculture worldwide. We set voluntary standards for the certification of agricultural products around the globe–and more and more producers, suppliers and buyers are harmonizing their certification standards to match.	*	*	*	*	IAF	Board: 5 retailers, 5 producers, NTWG: 41 countries Aqua.TC: 7 retailers, 7 producers (1 Asian) 2 Certification Body (observers) Focus groups: may be non-member, Board-approved Public: 2 public consultation periods	Producers must source compound feed and hatchery level from reliable suppliers. Farm level. (Also offers standards to entire chain of custody, feed manufacturers).
ASC - Salmon Standard	Salmon Aquaculture Dialogues (2004, WWF and IDH)	2012	To transform aquaculture towards environmental sustainability and social responsibility using efficient market mechanisms that create value across the chain.	*	(*)	*	*	ASI	ASC Board: 2 Industry rep (recruit 2 more), 4 non-industry TAG: 3 industry, 4 non-industry, 3 other TWG: 1 industry, 4 non-industry, 1 other SC: 10 industry, 5 non-industry Public: public consultation/complaints	Salmon standard from feed to farm level. (Also offers standards to entire chain of custody).
IFS - Food Standard	Retailer federation and industry companies International Food Standard (2003)	2003	To establish a common standard with a uniform evaluation system, work with accredited certification bodies and qualified auditors for IFS Food, ensure comparability and transparency in the entire supply chain, and reduce costs and time for both manufacturers and retailers.		*			IAF or EA recognized AB	TC: 15 retailers, 4 CBs, 6 Manufacturers, and 1 food services NWG: national retailers (IT, CHL, FR, GER, USA, ES) EWG: retailers and experts RC: retailers, industry, and CB	Only covers processing or handling of products during primary packaging.
BAP - Aquaculture Standard, Salmon Farms	Global Aquaculture Alliance (1997, Farmers)	2004	Achievable, science-based and continuously improved global performance standards for the aquaculture supply chain that assure healthful foods produced through environmentally and socially responsible means.	*	*	*	*	IAF	GAA Board: 20 members SOC: 4 conservation/social NGOs, 4 academia/regulators, 4 industry TC: 4 conservation/social NGOs, 4 academia/regulators, 4 industry Public: 60 days public comment	Salmon standard from feed (BAP-certified feed mills or declares compliance to BAP feed mill standards 3.1. & 3.3.) to farm level.
RSPCA – Welfare standard for farmed Atlantic salmon	RSPCA Animal Welfare and Rescue (1824)	2002	For all farm animals to have a good life and be treated with compassion and respect. To give people a higher welfare choice by ensuring animals are farmed to RSPCA welfare standards.				*	RSPCA Assured assessor, UKAS, ISO17065	STAG: 24 experts for the salmon standards (RSPCA specialist, farms, consultants, veterinarians, RSPCA field staff, and tech/field operations of Freedom Food Ltd). By selection. WCG: retailers, food companies, livestock farmers, farming associated industries, veterinarians, agricultural economists, environmentalists, and relevant individuals/orgs. By selection.	Salmon standard covers all aspects of the fish´s life including health, diet, environment, handling, and slaughter. Feeds produced according to UK & EU legislation.

S = Social, QS = Food Quality and/or Safety, E = Environment, AW = Animal Welfare
*Year refers to the year the specific standard was launched.
(The Food Ethics Council and Pickett 2014, ASC 2017, Freedom Food Ltd 2017, BAP 2017, IFS 2017, RSPCA Assured 2017, GLOBALG.A.P. 2017)

ards (Stanton 2012). Private standards that have become industry norm no longer provide a real choice for suppliers to comply with in order to participate or remain in a specific market. Hence, private schemes become "de facto mandates" as demarcation between mandatory requirements and voluntary standards becomes obscure (Casey 2009, Stanton 2012).

From the perspective of the consumer, the large amount of certification schemes, standards, and labels available may contribute to confuse and complicate the purchase decision, as well as negatively influence their attitude towards the food producers and owners of the label in use. It has also been shown that many consumers do not know the content of each label so that decisions are often made on other characteristics and heuristics (Grunert 2005). Research shows consumers might prefer sustainable seafood; however, they do not pay much attention to this when buying seafood (Alfnes 2017).

3 METHODS

This paper is based on an analysis of documents from a range of certification schemes, the content of their different standards, and literature on certification. The chosen method is aimed to provide a comparison of a selected number of certification schemes and their origin, motivation for establishment, and content of their standard(s). The selected standards are established at different times, some of them are aquaculture and salmon specific, while others are not, and they differ in their focus on sustainability and/or animal welfare. Common for all is their relevance to salmon aquaculture production. The selection of schemes and standards is also based on their prevalence in the major nations of salmon aquaculture production. To illustrate the muddled sea of certifications in which production companies find themselves, the choice of standards in this study is also meant to reflect the diversity of focus areas, motivation, and actors involved. After gathering data and categorizing them according to characteristics (see Table 1), the background for the inception of these schemes was also analyzed (see Figure 1). The following information, unless otherwise specified, comes from the websites of these schemes.

4 STANDARDS AND CERTIFICATION SCHEMES

4.1 *ASC*

Established in 2009, the Aquaculture Stewardship Council (ASC) originated from the Aquaculture Dialogue, a multi-stakeholder roundtable founded

Figure 1. Development of schemes.

by the World Wide Fund for Nature (WWF) in 2004 (WWF Norge 2016). WWF and The Sustainable Trade Initiative (IDH, includes businesses, trade unions, NGOs, and Dutch Ministries for stimulating sustainable trade) from the Netherlands worked together in establishing the Aquaculture Stewardship Council in 2010 (IDH 2017).

ASC is the only aquaculture certification scheme that is recognized as a full member of the ISEAL Alliance Code of Good Practice for Setting Social and Environmental Standards. Also, the organization develops standards that are in line with FAO guidelines. ASC partners with the Global Aquaculture Alliance (GAA) and GLOBALG.A.P., and is supported by various suppliers, producers, retailers, and food brands. Any stakeholder or individual can raise issues regarding a certification of a facility as the certification documents are available online.

There are currently 8 aquaculture standards that cover 12 different species: abalone, bivalves (clams, mussels, oyster, scallop), freshwater trout, pangasius, salmon, shrimp, tilapia, seriola, and cobia. The ASC Salmon Standard was developed in 2012 by over 500 participants (WWF Norge 2016). The scope of the ASC standard for salmon includes: compliance with national and local laws and regulations, habitat, biodiversity and ecosystem, health and genetic integrity of wild populations, responsible use of resources, managing disease and parasites responsibly, socially responsible development and operations, and community involvement. The review of the standards is conducted regularly to ensure that the standards are compatible with new scientific developments and practices. The ASC supervisory board is composed of representatives from academia, NGOs, and the industry while its Technical Advisory Group (TAG) consists of a group of invited technical experts. The Technical Working Groups (TWG) and Steering Committees also meet and guide ASC standard development.

4.2 *GLOBALG.A.P*

EurepGAP was initiated by European retailers in 1997 with the goal of establishing a generic stand-

ard for Good Agricultural Practice (GAP) (Kalfagianni and Pattberg 2013). Prior to its establishment, European supermarket chains started various "Integrated Crop Managements" (ICMs) as an effort to gain consumers that preferred 'sustainable products' (Casey 2009, Kalfagianni and Fuchs 2012). The suppliers struggled with achieving the many ICMs of different supermarkets. As a way of harmonizing these agricultural processes, EurepGAP was born and was renamed to GLOBALGAP in 2007 as the standard became widespread in the international scene (Kalfagianni and Fuchs 2012).

The GLOBALG.A.P. Aquaculture module was included in GLOBALG.A.P. in 2004 and covers the entire production chain of a variety of farmed fishes, crustaceans, and mollusks from suppliers (brood-stock, feeds, seedlings) to the various activities, such as faring, harvesting, processing, and post-harvest handling operations (Prein and Scholz 2014). GLOBALG.A.P. is a business-to-business standard, and is classified by FAO as both a standard and a code (Washington and Ababouch 2011). The scope of the certification for the aquaculture module includes site management, reproduction, chemical compounds, occupational health and safety, fish welfare, management and husbandry, sampling and testing, feed management, pest control, environmental and biodiversity management, water usage and disposal, harvesting and post-harvest operations, holding and crowding facilities, slaughter activities, depuration, post-harvest mass balance and traceability, and social criteria.

In addition to certification, GLOBALG.A.P. also has a consumer label called GGN (GLOBALG.A.P. Number) for certified aquaculture products that are in accordance with GLOBALG.A.P. (GGN 2017). Feed that includes captured fish should come from fisheries that adhere to the FAO Code of Conduct for Responsible Fisheries.

GLOBALG.A.P. members elect the Board (5 producers and 5 retailers), which guides the Secretariat, the Technical Committees (one, out of eleven representatives, is from Asia in the Aquaculture group), and Focus Groups (voluntary members and non-members). The Secretariat gives directions to the Benchmarking Committee, Certification Body Committee, Integrity Surveillance Committees, and the National Technical Working Groups (41 countries). The Technical Committees give direction to the respective Focus Groups. National Technical Working Groups are responsible for translating the national interpretation guidelines and local adaptation of the standard. There are two public consultations or rounds for submitting comments by interested parties within a period of 40 to 60 days.

4.3 *RSPCA*

The Royal Society for the Prevention of Cruelty to Animals (RSPCA) is an animal welfare charity organization in England and Wales. The RSPCA Assured label which replaced the Freedom Food label in 2015, is an ethical food label established by the RSPCA. A report from The Food and Ethics Council and Picket (2014) identified three drivers for farm assurance schemes. Firstly, the 1980s and 1990s in the UK were overcast by a number of highly publicized food scares such as with BSE in cattle and reports uncovering salmonella-infected egg production. In addition to the aim of restoring consumer confidence, the Food Safety Act in 1990 introduced the requirement of retailers' due diligence which assigned food safety responsibility to retailers. A third reason for farm assurance schemes to proliferate during this time was the desire to promote responsible farming and animal welfare (The Food Ethics Council and Pickett 2014).

Priding itself as being the only farm animal welfare scheme in the UK, the RSPCA welfare standards examine all aspects that are vital to an animal´s welfare, such as farm management, husbandry practices, healthcare, living conditions, nutrition, transport, and humane slaughter. The RSPCA welfare standards include beef cattle and calves, chickens, ducks, hatcheries, laying hens, dairy cattle and calves, pigs, pullets, salmon, sheep, trout, and turkey. Meetings with the Standards Technical Advisory Group (STAG) are conducted by RSPCA once a year for each species to ensure effective accumulation of the latest scientific, veterinary, and industry information. STAG members include retailers, food companies, farming associated industries (e.g. manufacturing), veterinarians, environmentalists, or organizations and individuals advising the RSPCA Farm Animals Department on standard development. STAG membership is by invitation only. Membership for the Wider Consultation Group (WCG) is by invitation only by the Farm Animals Department of RSPCA. RSPCA Assured currently covers more than 140 million salmon. Major retailers in the UK offer more than 2,000 RSPCA Assured products.

4.4 *IFS food standard*

The International Featured Standards (IFS), originally called the International Food Standard, was established in 2003. IFS is an association of retailers and industrial companies that aims to set harmonized standards for their producers, logistics companies, brokers, and agents. Since their expansion, they now have 8 standards for food products and services published in five primary languages (English, German, Spanish, French, and Italian). The IFS Food Standard deals with food safety and quality of the product and the processes of food packing and processing companies. The standard is recognized by the Global Food Safety Initiative (GFSI). The scope of the standard includes sen-

ior management responsibility, quality and food safety management system, resource management, planning and production processes, measurements analysis and improvements, and food defense and external inspections.

Retailers that require suppliers to have IFS certification include Aldi, Lidl, and Metro (Bureau Veritas 2017). The IFS certification is also sought after by retailers from their suppliers in the French and German markets (Washington and Ababouch 2011).

The IFS Technical Committee (TC) is composed of representatives from retailers (17, many from Germany, Italy, France, and Spain), industry (6 manufacturers, 1 food service), and certification bodies (4 from Europe). The TC is responsible for content and requirements of the standards. National Working groups (NWG) from Italy, France, Germany, Chile, USA, and Spain are responsible for supporting and providing the TC technical information to the International Working Group. Examination Working Groups (EWGs) are composed of retailers and experts. A Review Committee is represented by retailers, industry, and CBs. They discuss experiences and discuss changes of requirements of the audit report and training.

4.5 *BAP*

The Global Aquaculture Alliance (GAA), a nonprofit organization attending to issues related to advocacy, education, and leadership in responsible aquaculture, is the owner of the BAP certification scheme. GAA was established in 1997 by shrimp farmers as a response to criticisms from Greenpeace in the 1990s and a global moratorium demanded by NGOs and community organizations in Choluteca, Honduras (Lee and Connelly 2006). According to Aguayo and Barriga (2016), BAP standards were led by the industry corporate actors and there was no participation by stakeholders not belonging to the industry (Aguayo and Barriga 2016).

BAP is an aquaculture standard that promotes codes of conduct through best management practices (Lee and Connelly 2006). The standards are continuously improved through efforts from the Technical Committee, Standards Oversight Committee (SOC) comprised of experts in environ-mental conservation, the academia and the industry, and comments from the public, which are available on their website. The BAP consumer eco-label includes a star rating system that shows the level of integration in the food chain, with one star meaning the product is produced by a BAP-certified processing plant while a 5-stars label means that the product has been produced only by BAP-certified facilities (processing plant, farms, hatchery, and feed mill). The standard covers community property rights and regulatory compliance, community relations, worker safety and employee relations, sediment and water quality, fishmeal and fish oil conservation, control of escapees, predator and wildlife interactions, storage and disposal of farm supplies, animal health and welfare, biosecurity and disease management, control of potential food safety hazards, and traceability.

BAP standards are continuously updated. The GAA is responsible for coordinating the development of the standards. The technical details are developed by the Technical Committee (TC) under the guidance of the Standards Coordinator from GAA and subject to the review and approval from the Standards Oversight Committee (SOC). The 12-member SOC should consist of equal numbers of representatives from academia, conservation groups, and industry groups. After the SOC has reviewed the document (and modified, if needed), the changes are published for a 60-day comment period where the public can participate. The SOC carefully considers all the public comments for possible inclusion in the final draft. The draft is then submitted for approval by the SOC and the GAA Board of Directors before the standard is implemented.

5 DISCUSSION

Figure 1 shows a diagram illustrating how standards are established for different purposes and through diverse processes by distinct proprietors. As discombobulating as the figure seems, the reality is far more confounding. This can be explained by the many different stakeholders involved, with their various motives, interests, and desires to tackle the array of challenges that salmon aquaculture is facing. Despite running the risk of confusing the consumers, and at worst, resulting in label indifference, the schemes continue to evolve with a goal of making themselves distinct from the others while aiming to expand their terrain.

To give a more orderly and comprehensive understanding of the differences and similarities that characterize these schemes and their standards, we here provide a summary divided into the 3 Ps of certification: *purpose*, *proprietorship,* and *process*. Purpose refers to the needs and interests that have motivated the development of the different standards. Proprietorship deals with the owner(s) of the scheme. Process involves how the standards were developed and which actors were involved.

5.1 *Purpose*

Each standard was established with a purpose in mind. Some were intended to cover very specific issues, such as the IFS Food Standard and the RSPCA, while others were meant to be more general and all-encompassing. In the latter category, the GLOBALG.A.P. Aquaculture standard and

the BAP Aquaculture standard are similar in that they both cover aspects of food safety and quality, social, environment, and animal welfare. However, GLOBALG.A.P. was initiated to unify several schemes required by suppliers to provide consumers with sustainable products, while the BAP certification was developed as a response to criticisms from environmental groups and NGOs. The ASC Salmon Standard, also in the latter category, differs as it is a species-specific scheme with less focus on food safety, and was developed as a response to increased focus on the environment and social responsibility of the aquaculture industry. As with many of the more general standards, the IFS Food Standard was also aimed at providing a unified standard for suppliers; however, its focus is on general food safety and quality. The RSPCA Assured was established to improve animal welfare and, therefore, focuses more or less only on concerns regarding this issue.

5.2 *Proprietorship*

GLOBALG.A.P. and IFS schemes were both established by retailers while the ASC and RSPCA standards were both initiated by non-governmental organizations. Of the five schemes, only the BAP Aqua-culture Standard was started by producers. Certification is performed by third-party certification bodies, except for RSPCA, which differentiates itself by certifying farms using their own RSPCA Assured assessors. A majority of these private schemes are mostly owned by retailers and NGOs, which means that they are able to exert power over the producers by demanding that these requirements be met if they are to be recognized as suppliers. Moreover, the schemes come from developed countries and Northern markets, tipping the scales in favor of large companies (Belton et al. 2011).

5.3 *Process*

The development of standards for the different schemes is similar in the sense that they are including different stakeholders and expert groups. Some schemes try to balance the number of representatives from the different stakeholder groups, such as BAP and GLOBALG.A.P. Not all the schemes, however, include public consultation. The IFS scheme, for instance, does not mention any public consultation nor does it say anything about NGO participation. Other schemes only include participants by invitation, such as the RSPCA, selecting the experts for consultation and standard development. Furthermore, the documents stating how many of each stakeholder group should be included in a Technical Group does not apply in practice (e.g. GLOBALG.A.P. and ASC).

According to Fuchs et al. (2011), the retailer-dominated private standards, such as IFS, are dominated by the standard owner. The food industry and certification bodies play only a consultative role, while civil society is not provided with a voice. They categorize GLOBALG.A.P. as a standard that provides an equal partnership between the retailers and producers through elections, and certification bodies only act as associate members, while civil society and the NGOs may participate in the annual meetings. Despite the seemingly equal opportunities for stakeholders to take part in representing their group, in reality, not all of the stakeholders afford to take part in the development process as this requires a lot of time and resources.

6 CONCLUDING REMARKS

As has been shown here, there are countless challenges that follow the proliferation of certifications, standards, and labels in the aquaculture industry. Increasing pressure from both public and private regulatory agencies is causing a continuous build-up of demands for production companies. Since standards purposely differ from one another in some ways and overlap in other aspects, there is often a need to comply with more than one standard. This entails that the new standards which emergence do not replace others, but add yet more layers.

Having just one all-encompassing standard could possibly curtail certification-related work for producers and strengthen consumers' trust in labeling; but would this be attainable? Based on our findings in this study, it is unlikely to happen. This can be attributed to numerous explanations. For one, the different certification schemes are in competition with each other, as certifications, standards, and labels have become big business. Furthermore, the standards are created at different times and continue to be adapted and revised, making a potential unification difficult to achieve. Most importantly, the endeavor to improve the aquaculture industry, currently under the banner of sustainability, is pulling in many different directions. The numerous challenges that the industry is facing are subject to trade-offs and political priorities, as many of them run counter to each other. In order for the standard to cover everything, it would necessarily go against itself.

ACKNOWLEDGMENT

This work has been conducted through the research project SUSTAIN-FISH (project number 254841) financed by The Research Council of Norway.

REFERENCES

Aguayo, B.E.C. & Barriga, J. 2016. Behind certification and regulatory processes: Contributions to a political

history of the Chilean salmon farming. *Global Environmental Change* 39: 81–90.
Alfnes, F. 2017. Selling only sustainable seafood: Attitudes toward public regulation and retailer policies. *Marine Policy*, 78, 74–79.
ASC 2017. ASC Salmon Standard - v.1.1 - April 2017.
Asche, F., Guttormsen, A.G. & Tveterås, R. 1999. Environ-mental problems, productivity and innovations in Nor-wegian salmon aquaculture. *Aquaculture Economics & Management* 3(1): 19–29.
BAP 2017. Aquaculture Facility Certification Salmon Farms.
Belton, B., Haque, M.M., Little, D.C. & Sinh, L.X. 2011. Certifying catfish in Vietnam and Bangladesh: Who will make the grade and will it matter? *Food Policy* 36(2): 289–299.
Boyd, C.E. & Nevin, A.A. 2011. An Early Assessment of the Effectiveness of Aquaculture Certification and Standards. In Steering Committee of the State-of-Knowledge Assessment of Standards and Certification. (2012). *Toward sustainability: The roles and limitations of certification*. Washington, DC: RESOLVE, Inc.
Bureau Veritas 2017. Retrieved 1 September 2017, from http://www.bureauveritas.com/services+sheet/brc-ifs-certification_1105.
Busch, L. 2011. *Standards: Recipes for Reality.* Mit Press.
Bush, S.R., Belton, B., Hall, D., Vandergeest, P., Murray, F.J., Ponte, S., Oosterveer, P., Islam, M.S., Mol, A.P.J., Hata-naka, M., Kruijssen, F., Ha, T.T.T., Little, D.C. & Kusumawati, R. 2013. Certify sustainable aquaculture? *Science* 341(6150): 1067–1068.
Casey, D.K. 2009. Three puzzles of private governance: GlobalGAP and the regulation of food safety and quality. UCD Working Papers in *Law, Criminology & Socio-Legal Sudies Research,* paper no. 22/2009.
Chen, X., Alfnes, F. & Rickertsen, K. 2015. Labeling Farmed Seafood. Working Papers no.10/2015. Norwegian Uni-versity of Life Sciences.
Davidson, K.M. 2010. Reporting Systems for Sustainability: What are they measuring? *Social Indicators Research*, 100 (2): 351–365.
FAO 2003. The ecosystem approach to fisheries. FAO Technical Guidelines for Responsible Fisheries.
FAO 2016. *The state of the world fisheries and aquaculture.* Retrieved 30 August 2017, from http://www.fao.org/3/a-i5692e.pdf.
Freedom Food Ltd 2017. *What is RSPCA Assured?.* Retrieved November 28, 2017, from https://www.berspcaassured.org. uk/.
Fuchs, D., Kalfagianni, A. & Havinga, T. 2011. Actors in private food governance: the legitimacy of retail standards and multistakeholder initiatives with civil society participation. *Agriculture and human values* 28(3): 353–367.
GGN 2017. *What does GGN mean?*. Retrieved 30 August 2017, from http://aquaculture.ggn.org/en/what-does-ggn-mean.html.
GLOBALG.A.P. 2017. *GLOBALG.A.P. - Putting Food Safety and Sustainability on the Map*. Retrieved November 28, 2017, from http://www.globalgap.org/uk_en/who-we-are/ about-us.
Grunert, K.G. 2005. Food quality and safety: consumer perception and demand. *European Review of Agricultural Economics* 32, 3, 369–391.
IDH 2017. *Driving sustainability from nitch to norm*. Retrieved November 23, 2017, from https://www.idh-sustain abletrade.com/about-idh/.
IFS 2017. *IFS Food.* Retrieved November 28, 2017, from https://www.ifs-certification.com/index.php/en/standards/251-ifs-food-en.
Kalfagianni, A. & Fuchs, D. 2012. The GlobalGAP in Reed, D., Utting, P. & Mukherjee-Reed, A. (Eds.) Business Regulation and Non-State Actors: Whose Standards? Whose Development?: 148–160. Routledge.
Kalfagianni, A. & Pattberg, P. 2013. Fishing in muddy waters: Exploring the conditions for effective governance of fisher-ies and aquaculture. *Marine Policy* 38: 124–132.
Lee, D. & Connelly, J. 2006. Global aquaculture alliance on best aquaculture practices: an industry prepares for sustainable growth. *Sustainable Dev. L. & Pol'y* 7, 60–62.
Morris, J. 1997. *Green goods? Consumers, product labels, and the environment*. London, U.K.: Institute of Economic Affairs, Environment Unit.
Nadvi, K. & Wältring, F. 2002. *Making sense of global standards*. INEF-report 58.
Prein, M. & Scholz, U. 2014. The Role of VSS in enhancing the contribution of fisheries and aquaculture to sustainable development. In Schmitz-Hoffmann, C., Schmidt, M., Hansmann, B. and Palekhov, D. (eds.) *Voluntary Standard Systems,* 315–343, Berlin: Springer.
RSPCA Assured 2017. "Farm animal welfare." Retrieved November 28, 2017, from https://www.rspcaassured .org.uk/farm-animal-welfare/.
Stanton, G.H. 2012. Food safety-related private standards: the WTO perspective in Marx, A., Maertens, M., Swinnen, J. and Wouters, J. (eds.) *Private Standards and Global Governance: Economic, Legal and Political Perspe*ctives, 235–254. Cheltenham, UK: Edward Elgar.
The Food Ethics Council & H. Pickett 2014. "Farm animal welfare. Past Present and Future." Report.
Tlusty, M.F., & Thorsen, Ø. 2017. Claiming seafood is 'sustainable'risks limiting improvements. *Fish and Fisher-ies* 18(2): 340–346.
UN 2017. *World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100.* Retrieved 30 August 2017, from https://www.un.org/development/desa/en/news/popula tion/world-population-prospects–2017.html.
Washington, S. & Ababouch, L. 2011. Private standards and certification in fisheries and aquaculture, FAO.
World Bank 2014. *Sustainable Aquaculture*. Retrieved 30 August 2017, from http://www.worldbank.org/en/topic/environment/brief/sustainable-aquaculture.
WWF Norge 2016. AS*C - Miljøsertfisering til havbruk.* Retrieved November 23, 2017, from https://www.wwf.no/dette_jobber_med/hav_og_kyst/havbruk/miljostandard/asc/.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

An ontological and semantic foundation for safety science

P.J. Blokland
Safety and Security Science Group (S3G), Delft University of Technology, The Netherlands

G.L.L. Reniers
Safety and Security Science Group (S3G), Delft University of Technology, The Netherlands
Center for Corporate Sustainability (CEDON), KULeuven, Campus Brussels, Belgium
Department of Engineering Management, Faculty of Applied Economic Sciences (ENM), University of Antwerp, Belgium

ABSTRACT: This article proposes an ontological and semantic foundation for safety science, based on an etymological and etiological study of the concepts of risk and safety. The awareness regarding the concepts of safety and risk have both evolved in similar ways because of increasingly more demanding situations and events that impact society in an economic way, also linked to the value of human lives. From a purely negative view on risk and safety, this awareness has grown into a more systemic and even holistic perspective on these concepts. The proposed foundation is aligned with the semantics and concepts used in the ISO 31000 risk management standard. Based on this foundation, the article also advocates a theoretical model and a metaphor on how to look at safety and performance in any organization.

1 INTRODUCTION

When one talks about safety, risk or performance, everyone understands what is being talked about. There's no one who doesn't grasp what the words mean in one's own perception and how they can be understood. However, when opening a discussion on what these concepts really are, and how one should study or deal with them, it is most likely to end up in ontological and semantic debates due to different views, perceptions and understanding.

Science is served with clear concepts and well-defined parameters. Because, having these concepts and parameters allows for exact measurement of observations and in its turn, this opens the opportunity of accurate analysis, which then can be used to develop sound theories and practices. When studying in the field of safety science (a relatively young field of science), it is hard to find unambiguous definitions and parameters that clearly link safety, performance and risk.

When reviewing the safety science literature, the question "what is safety" is answered in many ways and it is very hard to find a clear definition of its opposite, which we could also name 'unsafety'. Likewise, there is also a lack of standardization when it comes to defining its opposite. Terms like accident, incident, mishap, disaster, catastrophe, etc. have different meanings depending on the persons or fields of knowledge using these commonly used words.

Because there is no commonly accepted way to define safety and its opposite, it becomes very difficult to measure and compare the level of safety of situations and organizations in an unambiguous or objective manner, certainly amongst different sectors or societies. Also, it is more difficult to think of proactive solutions that generate safety instead of developing reactive methods that prevent unsafety.

Although the field of safety science is relatively novel as a separate and independent domain of study, many theories, models and metaphors have already been proposed, attempting to describe what safety is and how it can be achieved. Often these theories are drawn from the investigation of – and lessons learned from– catastrophes and disasters. As such, these theories are often justified by explaining how these mishaps came about. Therefore, in general, efforts to improve safety of systems have mostly been driven by hindsight, both in research and in practice (Woods and Hollnagel, 2006).

2 EVOLVING PERCEPTIONS REGARDING SAFETY (SCIENCE) AND RISK (MANAGEMENT)

Safety and risk are two concepts that are tightly coupled and have known similar evolutions in their development and in how people understood these

concepts. Also, the evolution on how people have dealt with risk and safety is very much comparable. Safety and risk are often perceived in a similar way and are regularly used as antonyms. Risky often means unsafe and safe often means without or protected from risk (as indicated in many dictionary definitions of safety). And, when looking at the past, one can see that ideas about safety and risk have evolved in a very analogous way and for comparable reasons.

2.1 *A historical perspective on risk management, the etymology of risk and its etiology*

2.1.1 *Ancient times*

For thousands of years, people considered all that happened being the will and acts of the gods (Bernstein, 1996). So, the general idea was that whatever one tried, things finally always happened to the will of the gods and there was nothing to do about it but to accept it.

However, this doesn't mean that concepts as risk and safety were strange to people. In their article "Risk analysis and risk management: an historical perspective", Covello and Mumpower (1985) describe how in the Tigris-Euphrates valley, about 3200 B.C. the Asipu already offered a kind of consultancy services related to risk and safety.

2.1.2 *The Renaissance and modern time period*

The serious study of risk started during the Renaissance and it took until the work of Pascal in the 17th Century to see a sudden progress in the understanding of risk and decision making based on numbers.

In this time period science was on the rise and it was a period of expanding trade of new and scarce products, transported overseas. This created a new reality. Trade oversea to distant regions and countries was a high-risk endeavor. This economic factor made people become more aware of the concept of managing risk. Soon, the insurance industry emerged as an effort to manage risk in commerce. Wealth was no longer the privilege of the happy few, but could be earned by investing in trade and making the right decisions (Bernstein, 1996; Covello & Mumpower, 1985).

2.1.3 *20th century*

Although the etymological roots of the term risk, can be traced back as far as the late Middle Ages, the more modern concepts of risk appeared only gradually, with the transition from a traditional to a modern society. With larger and ever more complex technology systems emerging after the second World War (e.g. nuclear installations), the focus on probability and risk supported a scientific, mathematically-based approach toward risk and risk assessment (Zachmann, 2014).

Later in the twentieth century, with standards of living quickly rising after World War II, other objectives became also important and the concept of managing risk expanded from a mathematically-based approach to include also more qualitative methods. Hence, the origins of operational risk management, which can be traced back to the discipline of safety engineering (Raz & Hillson, 2005).

Continuing losses, injuries and casualties, triggered the US Armed Forces and NASA to develop risk management proposing a more comprehensive approach, called Operational Risk Management (ORM), adapting the world of risk management to the human factor involved in day to day operations. However, by the end of the century further development of the concept of operational risk management expanded the view on risk from a loss and probability perspective to a systemic view, shifting attention from probability to achieving goals.

In the same period of time, due to scandals such as the Barings Bank (1995), the dot.com bubble (1997–2001) and ENRON (2001), people became ever more concerned with the management of risk and the good ethical practices in managing organizations. During this last decade of the twentieth century, there has been a major surge of interest in improving the ability to deal with an uncertain future, and at that time, still with a focus on the negative impact at the organizational level. Operational risk scarcely existed as a category of practitioner thinking at the beginning of the 1990's, however, by the end of that decade, regulators, financial institutions and practitioners could talk of little else (Power, 2005). At that time, the first risk-related standards were published (Raz & Hillson, 2005).

2.1.4 *21st century*

The changes that emerged during the last quarter of the 20th century persisted and ongoing changing ideas concerning risk management generated an increasing understanding of the concept of risk, as modern risk management evolved substantially due to factors such as the rise of knowledge-intensive work, an expanding view on stakeholders, a growing importance of project management, the expanded use of technology, increased competitive pressure, increased complexity, globalization and continuing change (Raz & Hillson, 2005).

This growing concern and increasing awareness regarding risk management at the turn of this century led to the development of a whole range of additional risk management standards. These standards were issued by governments (Canada in 1997, United Kingdom 2000, Japan 2001 and Australia/New Zealand 2004), International insti-

tutions (IEEE-USA 2001, CEI/IEC-CH 2001) or professional organizations (IRM/ALARM/ AIRMIC-UK 2002, APM-UK 2004, PMI-USA 2004). Each of these standards, coming from different perspectives, reflect an increasing understanding of risk and risk management, proposing different definitions of risk and comparable processes to manage risks. At that moment in time, a shift occurs from a purely negative view on risk, still expressed in the definitions of some of those (older) standards (CAN/CSA-Q850-97:1997 and IEEE 1540:2001) to more neutral or even very broad definitions of risk in the other, more modern, standards. Another remarkable aspect of the "newer" definitions is the fact that risk is more explicitly linked to objectives and that the effects of uncertainties on objectives (consequences) can be positive, negative or both (Raz & Hillson, 2005).

2.1.5 *Enterprise Risk Management (ERM)*

Also in the first decade of this century, and due to a number of scandals—similar to ENRON, there is an ever-increasing attention for corporate governance and the role of operational risk management in that regard. This resulted in the first internationally used comprehensive corporate standard on risk management, the COSO Enterprise Risk Management Integrated Framework (2004). (Mestchian et Al, 2005). Enterprise Risk Management (ERM), similar to ORM in the military and aviation sectors, is the more holistic approach that is needed to cope with the complex realities and awareness of risks for the corporate world in the 21st century.

However, the COSO ERM framework, developed as an auditing tool to check compliance, failed during the 2008 financial crisis, because organizations implementing ERM would still follow the reductionist approach these organizations were used to. So, the International Standardisation Organisation (ISO), set out to establish a working group to achieve consistency and reliability in risk management by creating a standard (ISO 31000) that would be applicable to all forms of risk and to all kinds of organizations, creating a foundation for risk management (Purdy, 2010).

Little real progress could be made with the ISO standard until all agreed on a definition of risk that arose from a clear and common understanding of what risk is and how it occurs. The working group arrived at: "*risk is the effect of uncertainty on objectives*". When risk is defined like this, it reveals more clearly that managing risk is, quite simply, a process of optimization that makes the achievement of objectives more likely, objectives being understood in the broadest sense of the word. Successfully detecting and understanding risk, including how it is caused and influenced, allows, if necessary, to change it so that it is more likely to achieve objectives and reach them faster, more efficiently, and with improved results. (Purdy, 2010)

The specific way in which risk is regarded by the ISO 31000 standard also broadens the understanding and attention of risk management towards performance, instead of solely focusing on compliance or the prevention of loss.

2.2 *Evolution of awareness in safety science*

Compared to the other sciences, surprisingly little attention has been given to the history of safety. (Guarnieri, 1992). As such, the following sections only try to give an indication of how and why the perspective of safety changed over time.

2.2.1 *The industrial revolution*

Safety science, similarly to risk management, originated because of a need to cope with uncertain profit, the failure of maintaining possession of valuable assets and the accidental injury or loss of workforce. In the same way expanding views impacted the etymology of the concepts of risk and risk management, ever-increasing awareness and knowledge regarding the concepts of safety and safety management has also impacted the etymology of safety and safety science.

The industrial revolution and the appearance of new technologies provoked reoccurring and severe accidents, damaging valuable assets, causing severe casualties and injuries to workers. In the beginning these accidents are just seen as set-backs, caused by workers behavior and part of the business. However, during the second industrial revolution, starting at the end of the 19th Century, the ongoing mechanization and new technological developments are used to develop new industries. Furthermore, production engineering substantially increased productivity with the advent of mass production. As a result, life was getting better, incomes were rising and mortality was declining. (Mokyr, 1998). These rapid economical, technological and social changes also triggered the dawn of safety as a science.

Accidents have always been a problem. Yet they did not appear as a major economic and health issue until the early 1800s when the declining death rate from infectious diseases shifted attention to other causes of mortality (Guarnieri, 1992). Accidents in a production line are costly, not only due to the casualties and lost workforce, but also the loss in production and production capacity are a burden to the profitability of the new factories. Furthermore, these accidents are responsible for a high mortality in the industrial world, leading to a bad reputation. Due to the rising prosperity, this is no longer acceptable. Accidents are no longer acts of God, but man made and can be prevented. (Swuste et al, 2010).

From the start, the awareness about risk and safety, risk management, safety management and safety science were triggered by the possibility of adverse effects, impacting on the profitability of endeavors and related to new emerging sectors. Both trying to accommodate for losses that impact that profitability. Risk, as such, became the domain of insurers and the start of a whole financial industry to compensate for financial losses. Likewise, safety science started with focusing on accidents, injuries and casualties and their prevention. In both cases, people drifted away from what they really needed, which is safeguarding and achieving objectives, getting what they want and being safe.

One of the first theories on safety is about accident proneness (Farmer, 1925). However, the accident proneness theory only looks at one possible cause of accidents and therefore cannot explain accidents in a general way.

Heinrich observed production facilities to discover trends and patterns in occupational accidents, resulting in Heinrich's pyramid or triangle (Heinrich, 1931). Heinrich also proposed his Domino theory on accident causation when studying the cost of accidents and the impact of safety on efficiency, opening up the perspective to the role of management in accident prevention (Heinrich, 1941). Heinrich's domino theory became a basis for many other studies on accident causation and the role of management in accident prevention, dominating the world of safety practitioners well beyond the Second World War (Hosseinian & Torghabeh, 2012).

2.2.2 *After World War II*

Heinrich's research and work inspired other researchers, also to incorporate the role of management in their models. For instance, Petersen (1971) developed a model based on "unsafe acts" and "unsafe conditions" and Weaver (1971) and Bird (1974) updated the domino model with more emphasis on the role of management. (Hosseinian & Torghabeh, 2012; Swuste et al, 2014).

At the beginning of the second half of the twentieth century, Gibson (1961) and Haddon (1970) focused on the causation of injuries, discovering a formula for injury prevention. This shift in focus, also caused safety science to look at engineering to reduce injuries, leading to safety belts, bumpers and many other devices capable of absorbing or deflecting energy (Guarnieri, 1992). In this period of time, also the introduction of the "hazard" – "barrier" – "target" model and tools, such as Failure Mode and Effect Analysis (FMEA), Hazard and Operability Analysis (HAZOP), the Energy Analysis approach and so on, are to be noted. (Swuste et al, 2014)

Similar to the evolutions in risk management, safety science further evolved as a result of a series of accidents which had a huge impact on society. Flixborough (1 June 1974), Seveso (10 July 1976) and Three Miles Island (28 March 1979) are part of the history of safety science, provoking a broader perspective on safety and increased safety regulations. This enlarged awareness about safety, also reflects in the increasing political attention for safety related issues and the rise in associated regulations, clearly demonstrated by the advent of a number of safety related scientific journals in the last quarter of the twentieth century (Hale, 2014). Investigating these accidents, expands the awareness of safety practitioners from the role of management, to interactions in the entire socio-technical system.

2.2.3 *More major accidents and disasters*

The socio-technical concept arose in 1949 (Trist, 1981). However, at that time in the fifties, the societal climate was negative towards socio-technical innovation. This would only change thirty years later (Walton, 1979). Again, alike the development of risk management and operational risk, safety science took up this wider organizational perspective on safety issues as from in the early eighties. At the same time, advances in technology also make safety engineering an indispensable part of safety science, with the development of safety equipment, for instance safety belts, air bags etcetera.

Another result of analyzing, amongst others, the Three Miles Island accident, is Charles Perrow's book, Normal Accidents (1984), in which the `normal accident theory' (NAT) is proposed. It has been particularly influential among researchers concerned to understand the organizational origins of disasters and the strategies which might be used to make organizations safer (Hopkins 1999).

Safety science further developed in the past thirty year as a result of another series of significant disasters, such as Bohpal (2–3 December 1984), Challenger (January 28, 1986), Tsjernobyl (26 April 1986), and The Herald of Free Enterprise (6 March 1987). Each of these accidents show the complexity of socio-technical systems. As a result, scholars try to model systems in order to predict their behavior. Building on the work of Rasmussen (1983), Reason (1990) proposes the Generic Error Modeling System (GEMS), later to become known as the Swiss Cheese model (of defenses) (Reason, 1997, 2016).

By the end of that disastrous decade, people also look at human factors and behavior, by introducing the notion of safety culture. It is loosely used to describe the corporate atmosphere or culture in which safety is understood to be, and is accepted as, top priority (Cullen, 1990). A more specific approach is the concept of Just Culture (Dekker, 2008, 2017). Furthermore, the concepts of 'High Reliability Organizations (HRO)' (LaPorte & Consolini, 1991; La Porte, 1996; Weick & Sutcliffe, 2001) and 'Resilience Engineering' (Woods & Hollnagel, 2006; Hollnagel, 2013) were introduced, looking at the whole organization.

Recent years have seen a whole range of models that try to model the taxonomy and structure of accidents. Some examples are the Systems-Theoretic Accident Modell and Process (STAMP) (Leveson, 2011) and the Functional Resonance Analysis Method (FRAM) (Hollnagel, 2012). Most remarkable is that FRAM is focused on safety instead of unsafety, going beyond the failure concept and the concepts of barriers and controls, aiming at the day to day performance. (Hollnagel, 2012). The idea is to achieve safety proactively. An idea further developed with the advent of the concepts of Safety-I and Safety-II (Hollnagel, 2014a). In his article, 'Is safety a subject for science', Hollnagel (2014b) indicates the difficulty to change the mindset in the safety science community from what is going wrong to what is going right.

In the new millennium, alike risk management expanded into a systemic/holistic view with the advent of Enterprise risk management, today these approaches come together in concepts such as Resilience Engineering, HRO and Safety-I & Safety-II. Ever more these modern concepts in safety are focusing on what people want and how to achieve it, instead of trying to protect against failure. Likewise, increasingly, scientists are looking for significant leading indicators in order to be more proactive in preventing accidents by achieving what is the aim. Concepts therefore also evolved from a purely negative view on risk and safety towards a more encompassing view, also considering the positive sides of risk and safety. Now the focus is gradually more on safety instead of solely concentrating on unsafety.

3 INCREASING AWARENESS SEEN FROM A SYSTEMIC PERSPECTIVE – THE SYSTEMS THINKING ICEBERG

The above review of risk and risk management, safety models, metaphors and theories is far from complete. A more complete overview is to be found in reading the related references. We only wanted to show the changing etymology of the concepts of risk and safety, and its etiology, over a period of time.

A model that can help in getting a comprehensive view on the evolution of risk and safety is the systems thinking iceberg model (Bryan et al, 2006). The systemic iceberg model, is a way to look at reality from a systems perspective. The visible part of the iceberg, (above the waterline) represents *the events* that result from the system(s) involved. When events are observed over time, patterns and trends can be discovered. It is at this easy to perceive level of awareness of systems that safety science has originated. Still today, safety practitioners are driven by the facts that are directly visible, gathered in statistical data, trying to find and understand trends or delineate recognizable patterns related to the observed events. As such, they try to discover causal relationships and produce better predictions to prevent these negative events from happening.

The common approach to safety is to look at the events such as loss of life, injury, harm, damage, or any other event generating negative effects, trying to understand, predict and prevent accidents. As an example, the accident proneness theory can be seen as a result of that process.

Increasing awareness on accidents allows to become aware of the system of causes and effects that produces the unwanted events and that repeat themselves. It is what every accident investigation tries to achieve, i.e. the understanding how elements act together, in order to find ways to prevent them from happening again. When the system is understood, it is possible to proactively alter the system and prevent the same unwanted events happening. Through history, scholars have been searching for ways to explain why unwanted events happen and how disasters can be predicted, trying to discover the system(s) that is (are) behind their occurrence. One of the first to develop a theory on accident causation is Heinrich and his domino theory, naming the elements of the system that are involved in the creation of accidents and using the metaphor of domino blocks to represent the subsystems and their interaction. Also, NAT and the Brownian movements model (Rasmussen, 1995, 1997) can be seen as such.

To be able to predict or to obtain more understanding and control on the systems involved in accidents, scientists also try to determine the structure of the systems involved, getting a clearer view on the dynamics that are at the genesis of accidents. Recent years have seen a whole range of models that try to model and structure the systems that generate unwanted events. STAMP and FRAM can be seen as examples and there is also the Swiss Cheese model (Reason, 1997, 2016) and the models that build on the same human factor approach.

Finally, scientists and practitioners aim at developing understanding on how mental models generate the system(s) and how they can be controlled and managed. An example of a mental model generating safety is for instance the concept of Just Culture (Dekker, 2008, 2017).

4 A MODERN PERSPECTIVE ON RISK AND SAFETY

Safety is often defined as a dynamic non-event and mostly explained by the events that violated that state of dynamic non-events (Weick, 2011). The problem with this approach is that it only covers the domain of unsafety and leaves any interpretation of safety open. When safety thinking is linked with dynamic non-events it solely focusses on pre-

venting bad things from happening. But is this the right approach in pursuing safety?

Is turning away from unsafety the same as aiming for safety? When one considers a situation of 100% safety, is this a situation where nothing is happening? This seems an impossible assumption. There will always be something happening, events and consequences (positive effects on objectives) one wants and events and consequences (negative effects on objectives) one doesn't want, both are important from a modern safety perspective. So, what is distinctive for safety to emerge, exist and persist? A modern perspective on safety, in the same way as a modern perspective on risk and risk management, looks at the whole picture. It starts with what people want to achieve, what needs to be safe and first make sure this will be achieved. It is making certain that the return on investment is attained when pursuing an opportunity. Hollnagel (2014) talks of Safety I and Safety II, where safety I is the traditional approach of avoiding losses while safety II is making sure the objectives are accomplished. In our view, this is how safety and safety science should evolve. It is about both the absence of unsafety and the presence of achieved and safeguarded objectives.

Nancy Leveson says that Safety is an emergent property of systems, not a component property (Leveson, 2011). It means Safety is something that needs to be achieved by the system, repeatedly. Obviously, a component can also be considered as a system on its own. Every system is made up of sub-systems which have other objectives than the overarching system. The safety of these sub-systems is important to the safety of the overarching system and each of them is subjected to a set of risk sources that can affect those more specific objectives.

4.1 *Total respect management*

The modern perspective on safety, and the mental model to achieve safety in a proactive way we propose, is called Total Respect Management (TR³M). Respect, in the way it is used for this model, is an expression originally derived from the Latin word respectus. Respectus comes from the verb respicere, meaning 'to look again,' 'to look back at,' 'to regard' or 'to consider someone or something.' In other words, the original meaning of the word 'respect' holds the connotation of giving someone or something dedicated attention to have a better view on the matter or give it consideration, particularly to come to a better understanding (Blokland & Reniers, 2017). As much as possible the systems, sub-systems (including the human factor) and their objectives need to be known and understood.

The reason for this 'respect' for systems and their sub-systems is the conviction that there is no common structure to all 'accidents'. This is also a way how we intend to look at the Swiss cheese metaphor. In this metaphor, the whole cheese is a reflection and representation of a socio-technical system and its performance. The cheese itself can be understood as excellent performance, where objectives have been achieved and are safeguarded (Safety-II). On the other hand, the holes in the cheese are the sub-systems of which the objectives are not achieved or safeguarded. As such, these are the different factors contributing to accidents (Safety-I). The model's hypothesis is that one can never know for 100% sure which sub-systems will fail at a given time or why and how they will become connected at a given time to produce a major accident and therefore it is important to give attention to all failed objectives and aim at reducing the number and magnitude of these failed objectives by increasing the level of performance.

Each hole in the cheese is considered being a "failed" objective and therefore unsafe. The Swiss cheese is dynamic. The holes constantly change positions and dimensions in an unpredictable way and could be seen as shifting around, coming and going, shrinking and expanding in response to operator actions and local demands (Reason, 1997).

Although the idea of layers of protection is useful to a certain extent, it only covers for Safety-I. Therefore, the TR³M model also focusses on performance as an element of safety. The aim of performance is to achieve objectives and maintain objectives safeguarded. As such performance stands for the whole cheese, or the whole concerned socio-technical system and the aim of TR³M is excellent performance (excellence) in that regard.

The way TR³M approaches the Swiss cheese metaphor is by stating that each of these latent conditions (failed objectives) can be seen as accidents on their own. It is just the level of importance and number of objectives involved that differentiates 'accidents' worth investigating from 'less important' holes. However, for TR³M each one of the holes is meaningful and needs one's respect. Hence the name 'Total Respect Management'. The holes/accidents result from (sub)systems, created by non-aligned or defective mental models existing in, or surrounding, the system (Blokland & Reniers, 2017).

Risk and the level of risk, in this sense, is nothing more than the possible effects on one's objectives due to the decisions taken and performance reached at a given time, representing a possible future reality. This reality will also determine the level of safety. The better this future reality can be imagined, the more it can be shaped to the desires and needs of the beholders by taking the right decisions at the appropriate time, generating safety proactively.

4.2 *A semantic connection between risk and safety and performance*

Regarding risk, ISO 31000 provides a whole set of definitions, also to be found in the ISO Guide 73. However, for Safety, in its broadest sense, there is no such internationally agreed standard and neither a modern and commonly used definition. A random selection of some definitions (Wikipedia, Merriam Webster, Dictionary.com) gives a traditional view on safety. Safety is a state, a condition, a control, a device, a quality or anything else that keeps us safe. Sure, we all know what safety means, it is being protected from harm, from bad things happening, but what does it really mean to be safe? So, how can safety and security be defined, taking into account the most recent ideas on safety and risk, inspired by the ISO 31000standard and its definition of risk? When are systems or people completely safe? Isn't it when one has attained all of ones objectives, when everything performs well and nothing affects ones objectives in a negative way?

Defining risk being the effect of uncertainty on objectives, means that three aspects define risk, namely; 'Objectives'; 'Effects on objectives' and 'Uncertainty related to the effects and the objectives'.

The proposed semantic foundation can be seen as follows: risk is an uncertain effect on objectives, while the actual performance is the result of that uncertain effect. Performance is Safety I + Safety II. It is why both concepts have evolved over time in similar ways and at a comparable timing.

Risk management, in a way, started closely related with gambling activities. Because professional poker players know that they don't win by chance or as a result of acts from the gods, but through carefully gathering information and analyzing/considering options based on that knowledge. It allows them to increase the probability that they make the right decision to support their aim of winning the game by taking more risk when it is appropriate to do so and limit the risks they take when it is the wiser decision, each time counting on the fact that the risks run are low for the decisions they take. However, they will only be safe when the game is over, all effects of uncertainty have their outcome and the profit has been paid. As such, safety and risk are the same, where risk, and how it is managed, determines the future of one's safety. Therefore, the same semantical foundation can and should be used.

4.3 *What is the usefulness of defining SAFETY in this way?*

This perspective on risk and safety, allows to develop a commonly used terminology for safety science and risk management. It will also allow to build systems that can measure safety instantly and in a much broader range than it is actual the case. Or better said, to measure 'unsafety'. Because measuring safety would require knowing all objectives present in a system and its sub-systems. This is impossible, because most of the time people are not aware of all of their objectives and also organizations are unable to know all the objectives of all their stakeholders. On the other hand, it is much easier to discover unsafety. Because when an objective has failed, it is likely to be seen or trigger a reaction. Human nature is designed to recognize unsafety, because it is necessary for survival. To measure unsafety, it is sufficient to measure all the holes to get an idea of the level of safety in the cheese.

4.4 *Safety first or first in safety*

The adagio "safety first", certainly from a traditional perspective, is fiction. When safety is the prevention of bad things happening, this credo is a real show stopper, as the safest thing to do for avoiding losses is to do nothing and prevent any activity. However, when you look at this motto from a fresh and modern perspective, it becomes a helpful mental model in achieving safety proactively. Safety first then means to achieve and protect objectives as a priority, aiming on excellent performance (safety II) and calling to action instead of inaction. When this is the governing paradigm, it will also become possible to be the first in safety, aiming at excellence. Because also safety performance will be an objective to be achieved and safeguarded.

5 CONCLUSION

In this article, we have expounded the evolutions of the concepts of risk and safety, how the awareness grew due to repetitive adverse effects on objectives and looking for ways to understand and cope with what had happened. We also indicated how the meaning of the concepts changed as a result of this increased awareness. To finally propose a new paradigm regarding safety and performance, linking risk and safety rather as synonyms, instead of treating them as antonyms.

REFERENCES

Bernstein, P.L. (1996). *Against the Gods: the remarkable story of Risk*. John Wiley & Sons. Inc. New York.

Blokland, P., & Reniers, G. (2017). Safety and performance: Total Respect Management (TR3M): a novel approach to achieve safety and performance pro-actively in any organisation.

Bryan, B., Goodman, M., & Schaveling, J. (2006). Systeemdenken. Academic Service.

Covello, V.T., & Mumpower, J. (1985). Risk analysis and risk management: an historical perspective. Risk analysis, 5(2), 103–120.
Cullen, H.L. (1990). The public inquiry into the Piper Alpha disaster (Report to the Parliament by the Secretary of State for Energy by Command of Her Majesty Vols. 1 and 2).
Dekker, S. (2017). Just culture: Restoring trust and accountability in your organization. CRC Press, Taylor & Francis Group.
Dekker, S.W. (2008). Just culture: who gets to draw the line?. Cognition, Technology & Work, 11(3), 177–185.
Farmer, E. (1925). The method of grouping by differential tests in relation to accident proneness. Industrial Fatigue Research Board, Annual Report, 43–45.
Guarnieri, M. (1992). Landmarks in the history of safety. Journal of Safety Research, 23(3), 151–158.
Hale, A. (2014). Foundations of safety science: A postscript. Safety Science, (67), 64–69.
Heinrich, H.W. (1931). Industrial Accident Prevention. A Scientific Approach. Industrial Accident Prevention. A Scientific Approach. First ed. McGraw-Hill Book Company, London.
Heinrich, H.W. (1941). Industrial Accident Prevention. A Scientific Approach. Industrial Accident Prevention. A Scientific Approach. Second ed. McGraw-Hill Book Company, London.
Hollnagel, E. (2012). FRAM, the functional resonance analysis method: modelling complex socio-technical systems. Ashgate Publishing, Ltd.
Hollnagel, E. (2014a). Safety-I and safety–II: the past and future of safety management. Ashgate Publishing, Ltd.
Hollnagel, E. (2014b). Is safety a subject for science?. Safety Science, 67, 21–24.
Hollnagel, E. (Ed.). (2013). Resilience engineering in practice: A guidebook. Ashgate Publishing, Ltd.
Hopkins, A. (1999). The limits of normal accident theory. Safety Science, 32(2), 93–102.
Hosseinian, S.S., & Torghabeh, Z.J. (2012). Major theories of construction accident causation models: a literature review. International Journal of Advances in Engineering & Technology, 4(2), 53–66.
La Porte, T.R. (1996). High reliability organizations: Unlikely, demanding and at risk. Journal of contingencies and crisis management, 4(2), 60–71.
La Porte, T.R., & Consolini, P.M. (1991). Working in practice but not in theory: theoretical challenges of" high-reliability organizations". Journal of Public Administration Research and Theory: J-PART, 1(1), 19–48.
Leveson, N. (2011). Engineering a safer world: Systems thinking applied to safety. MIT press.
Leveson, N. (2015). A systems approach to risk management through leading safety indicators. Reliability Engineering & System Safety, 136, 17–34.
Mestchian, P., Makarov, M., & Mirzai, B. (2005). Operational risk–COSO re-examined. Journal of Risk Intelligence, 6(3), 19–22.
Mokyr, J. (1998). The second industrial revolution, 1870–1914. Storia dell'economia Mondiale, 219–45.
Perrow, C. (1984). Normal accidents: Living with high risk systems. Princeton University Press.
Power, M. (2005). The invention of operational risk. Review of International Political Economy, 12(4), 577–599.
Purdy, G. (2010). ISO 31000: 2009—setting a new standard for risk management. Risk analysis, 30(6), 881–886.
Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE transactions on systems, man, and cybernetics, (3), 257–266.
Rasmussen, J. (1995). Risk Management and the Concept of Human Error. Joho Chishiki Gakkaishi, 5(1), 39–70.
Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. Safety science, 27(2), 183–213.
Raz, T., & Hillson, D. (2005). A comparative review of risk management standards. Risk Management, 7(4), 53–66.
Reason, J. (1990). Human error. Cambridge university press.
Reason, J. (1997). Organizational accidents: the management of human and organizational factors in hazardous technologies. England: Cambridge University Press, Cambridge.
Reason, J. (2016). Managing the risks of organizational accidents. Routledge.
Schaveling, J., Bryan, B., & Goodman, M. (2012). Systeemdenken: van goed bedoeld naar goed gedaan.
Swuste, P., van Gulijk, C., & Zwaard, W. (2010). Safety metaphors and theories, a review of the occupational safety literature of the US, UK and The Netherlands, till the first part of the 20th century. Safety Science, 48(8), 1000–1018.
Swuste, P., Van Gulijk, C., Zwaard, W., & Oostendorp, Y. (2014). Occupational safety theories, models and metaphors in the three decades since World War II, in the United States, Britain and the Netherlands: A literature review. Safety science, 62, 16–27.
Trist, E. (1981). The evolution of socio-technical systems. Occasional paper, 2, 1981.
Walton, R.E. (1979). Work innovations in the United States. Harvard Business Review, 57(4), 88–98.
Weick, K.E. (2011). Organizing for transient reliability: the production of dynamic non-events. Journal of contingencies and crisis management, 19(1), 21–27.
Weick, K., & Sutcliffe, K. (2001). Managing the unexpected: Assuring high performance in an age of uncertainty. San Francisco: Wiley, 1(3), 5.
Woods, D.D., & Hollnagel, E. (2006). Prologue: resilience engineering concepts. Resilience engineering. Concepts and precepts, 1–16.
Zachmann, K. (2014). Risk in historical perspective: Concepts, contexts, and conjunctions. In Risk-A Multidisciplinary Introduction (pp. 3–35). Springer International Publishing.

Economic analysis in risk management

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Time-dependent reliability in flood protection decision making in The Netherlands

W.J. Klerk & W. Kanning
Department of Hydraulic Engineering, Delft University of Technology, Delft, The Netherlands
Deltares, Delft, The Netherlands

M. Kok
Department of Hydraulic Engineering, Delft University of Technology, Delft, The Netherlands

ABSTRACT: Since 2017 Dutch flood protection standards are defined as target flood probabilities that all primary flood defences have to comply with by 2050. Explicitly accounting for uncertainties in probability distributions of load and resistance is an integral part of estimating the actual flood probability. Based on such estimates, many flood defences will be reinforced in the coming years, for design lifetimes that are generally 25–100 years. Therefore it is important that we correctly take into account time-dependence of both load and resistance during the lifetime. Loads are typically uncorrelated from year to year, whereas strength parameters exhibit significant correlation over time. This correlation over time of strength parameters can significantly reduce the failure rate and increase the lifetime reliability of a flood protection structure. In this paper we show the implications of time-dependent reliability for a set of illustrative cases. We consider the effect of different degrees of temporal dependence on reliability, lifetime and relative cost savings. The cases show that for common configurations, the inclusion of time-dependent effects, especially the correlation in time of strength variables, can increase the lifetime of a flood protection structure by up to 50%.

1 INTRODUCTION

Since January 2017 the Dutch primary flood defences have to satisfy new risk-based safety standards. Based on economic risk analysis, analysis of societal risk (risk of large numbers of casualties) and individual risk (risk of dying due to a flood), allowable (i.e. target) probabilities of failure for all major flood defences have been derived (Kok et al. 2017). The failure criterion is herein defined as the loss of flood retention capacity resulting in flood of a (defined) neighborhood with an average depth of > 0.2 meters. Safety standards are generalized into main categories with annual allowable failure probabilities 1/300, 1/1000, 1/3000, 1/10000 etcetera.

These safety standards are based upon a Bayesian interpretation of probability, meaning that the failure probability should be interpreted as a state of belief. A change in the magnitude of uncertainties due to e.g. new knowledge or measurements will then cause a change in the estimated failure probability. Hence, when the safety standard is not met, reducing dominant uncertainties can be a very relevant measure.

The new failure probability requirements for flood defences are formulated as annual probabilities, implicating that for each separate year the failure probability has to satisfy the defined standard. In design this is often interpreted as that the failure probability at the end of the design life has to equal the maximum allowable probability of flooding. This is different from for instance the failure probabilities in the Eurocode, where the design criterion is expressed as both an annual target reliability and a reliability for a lifetime, e.g. 50 years (CEN 2002). Depending on the exact definition of the failure probability, the difference between annual and lifetime reliability, which will be discussed in the next section, can have significant implications.

The actual reliability of a flood protection structure can be assessed by doing probabilistic computations using probability distributions of both load and strength variables. Historically, the tools for design and assessment that were used in the Netherlands are based on a semi-probabilistic approach. Slomp et al. (2016) gives a thorough overview of the current safety assessment tools. The current assessment and design tools allow for both probabilistic and semi-probabilistic assessment and are based on an explicit coupling between (old) semi-probabilistic tools and the new probabilistic safety standards.

The quantitative assessment of failure probabilities also enables accounting for time-dependent reliability effects. Next to time-dependent (uncertain) deterioration and uncertain changes in climate, also correlations between years can be taken into account explicitly. Loads are typically independent from year to year (the maximum water level in year i is typically not conditional on the maximum water level in year $i-1$). However, the strength variables are typically correlated from year to year as these uncertainties are merely caused by spatial variability in combination with limited knowledge. Incorporating this correlation could have significant impact on the assessment of reliability during the lifetime.

Currently there is little attention for the actual source of the uncertainty, which poses a problem when using concepts of time-dependent reliability. For instance for hydraulic load models, model uncertainties are used for water level, wave height and wave period, but the source of these uncertainties is not immediately clear. Also in the distributions for strength parameters, there can be significant uncertainty, especially for geotechnical failure mechanisms. For instance, the failure probability for piping is dominated by uncertainty in permeability and grain size (Jongejan and Maaskant 2015). Strength uncertainties can typically consist of natural variability, measurement uncertainty, transformation uncertainty or model uncertainty (Phoon and Retief 2016). The source of the uncertainty is important for two main reasons:

- It determines the optimal method for uncertainty reduction: some methods might have the same source of uncertainty. These will not (efficiently) increase the quality of available data and hence not reduce uncertainty nor improve the reliability estimate;
- It determines the amount of time dependence of subsequent years as some uncertainties might be (fully) correlated in time and others may not.

In this paper we explore different definitions of reliability that can be used for flood defences. Using illustrative cases with different degrees of uncertainties we illustrate the influence of these definitions and how that translates to the lifetime of flood defences and their life cycle costs.

2 METHODOLOGY

2.1 *Time-dependent reliability*

Flood defences are generally constructed for design periods of 25–100 years which, in the context of annual failure probability, implicates that in any given year in such a period the reliability should be higher than the requirement. In reliability engineering in general, concepts such as the survival time, failure rate and hazard function are used to characterize the temporal reliability (Kottegoda and Rosso 2008). Especially the hazard function is of interest, as this provides the failure rate of the system, this is conceptually shown in Figure 1. Here three phases are distinguished for the hazard rate of a system:

- The inception phase: here the hazard rate decreases as due to first experiences and quality control errors are corrected. One could say that at $t_0 = 0$ the constructed system is accepted.
- The phase where neither initial errors, nor deterioration play a role. In Kottegoda and Rosso (2008) this is denoted the *useful life*.
- The deterioration phase where deterioration of the system causes the hazard rate to increase significantly.

The inception phase for a flood defence has two major aspects: first of all there is the experience from initial performance, mainly during construction, that improves the reliability as instantaneous repairs are carried out. In this paper we consider flood defences that have just been delivered, so this phase is not considered. Secondly there is the dependence of failures on preceding years, meaning that if a dike doesn't fail and there is any kind of correlation between the years it yields some information on its performance. This is an effect that will be relevant during the entire life-cycle.

When also considering the other two phases, the distinction between the three phases doesn't fit that well for flood defences. Most of the deterioration processes are gradual and play a role during the entire life-cycle (see e.g. Buijs et al. (2009),

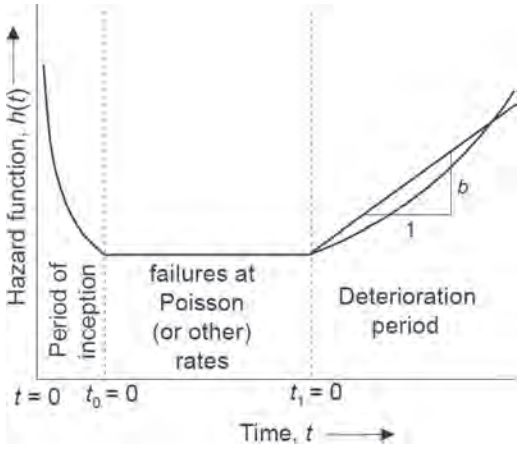


Figure 1. Hazard rate with distinction of three phases (Kottegoda and Rosso 2008).

Speijker et al. (2000)). In practice this means that there is no second phase ("failures at Poisson (or other) rates"), but that after delivery the reliability simultaneously decreases due to deterioration and increases due to information from non-failures. Whether and how these processes are taken into account depends on the definition of the failure probability that is used. For the failure probability in year t $P_f(t)$ the three main ones are:

1. $P_f(t)$ which means that the failure probability in year t is independent from the failure probability in other years.
2. $P(f_t \cap \overline{f}_{1..t-1})$ which denotes the probability of failure in year t and no failures occurred in the previous year.
3. $P(f_t \mid \overline{f}_{1..t-1})$ which denotes the probability of failure in year t given that no failures occurred in the previous year.

where f_t denotes failure at t and $\overline{f}_{1...t-1}$ denotes no failure in the period $1 \ldots t-1$. It has to be noted that most failure rates considered in literature assume a constant failure rate, which is in fact comparable to the first definition. The second and third are best compared to a description of a Decreasing Failure Rate (DFR) as described by Finkelstein (2008). However in order to better connect to current flood defence reliability practice a slightly different description is chosen here.

The choice of definition is dependent on the application and on the specifics of the situation. For cases where either the correlation between $P_f(t)$ and $P_f(t-1)$ is small and/or $P_f(t)$ is small equation 1 holds, and there is little difference in the three definitions.

$$P_f(t) \approx P\left(f_t \cap \overline{f}_{1..t-1}\right) \approx P(f_t \mid \overline{f}_{1..t-1}) \tag{1}$$

For all other cases the first definition is conservative.

Also it has to be noted that dike reinforcements generally do not entail a complete renewal, but rather an improvement of an existing flood defence. This means that part of the flood defence has already passed the inception period (as well as the other periods) and has to some degree proven itself. For this paper we consider a completely new flood defence and do not take that consideration into account although it can be very important if part of the dominant uncertainty is in a part of the dike body that has existed and survived for multiple decades or centuries.

2.2 *Temporal dependence in life-cycle reliability*

We consider a simple reliability problem where the limit state function at time t is given by:

$$Z(t) = R(t) - S(t) \tag{2}$$

with R the resistance and S the strength. In such a case, if we assume the limit state function can be approximated as a linearized hyperplane, the limit state function can be written as:

$$Z(t) = \beta(t) - \alpha_R(t)u_R(t) - \alpha_S(t)u_S(t) \tag{3}$$

where $\beta = \Phi(1 - P_f)$, where $\Phi(\cdot)$ is the inverse standard normal distribution. α_R and α_S are the influence coefficients of the random variables, indicating the respective contribution of their uncertainty towards the failure probability. u_R and u_S are random variables.

If we want to calculate the temporal reliability according to definitions 2 and 3 in the previous section, we need to take into account the correlation between subsequent years. The correlation of a component of the system in equation 2 is defined by:

$$\rho(Z_t, Z_{t-1}) = \alpha_{R,t}\alpha_{R,t-1}\rho_R + \alpha_{S,t}\alpha_{S,t-1}\rho_S \tag{4}$$

where ρ_R and ρ_S is the autocorrelation for the random variables of strength and load. For the strength, provided that there is no deterioration it could be argued that $\rho_R = 1$, the load is independent each year so $\rho_S = 0$.

For combining correlated components one could use numerical integration or probabilistic techniques such as Monte Carlo, but a very fast and efficient method is the Equivalent Planes Method, which is extensively described by Roscoe et al. (2015). In Roscoe et al. (2015) it is shown that this method is accurate for most cases, although some accuracy is lost for very large systems and for very strong correlations. However as the load is uncorrelated and values for α_R^2 are typically at most 0.7, such high values for the correlation will not be encountered when studying temporal reliability of flood defences.

The Dutch flood defense act allows for all three definitions of the previous section to be applied. The Equivalent Planes method therefore provides a fast and reliable method for evaluating the second and third definition of the annual failure probability. For the assessment of existing structures the third definition is most sensible, as in such cases it would be desirable to take into account that the structure didn't fail in the previous years, as has been done in for instance Schweckendiek (2014) and Schweckendiek et al. (2017). The third definition fits best with that. The second definition can be used for design purposes, as it is sensible to not account for the probability of failure in year t when the built structure has already failed in year $t-1$.

It has to be noted that for small probabilities of failure the second and third definition are almost the same as it follows from the definition of conditional probability that the difference between the two definitions for year t equals $1/\left(\prod_{i=1}^{n} 1-P_{f,t}\right)$. This indicates that for small P_f the difference will be negligible.

In order to combine different years, it is important that correlations in time between the limit state function from year $i-1$ to year i are correctly estimated. In many cases the reliability problem will not be so easy as the previously described problem, but will consist of many random variables that are (partially) correlated in time. In order to properly determine the temporal correlation of parameters and uncertainties it is important to classify uncertainties based on their original source, as only then a reliable classification can be made. This is further explained in the following section.

2.3 *Uncertainty in flood defence reliability*

There are various sources of uncertainty in flood defence reliability assessments. These are categorized by Gelder (2000) as inherent (aleatory) in time and space and knowledge (epistemic) uncertainty due to model and statistical uncertainty.Other categories can be used as well, e.g. Walker et al. (2003) distinguishes between different levels of uncertainty, and how these influence a decision problem. In general a distinction is often made between reducible and irreducible uncertainty as these influence the optimal action to deal with unacceptable failure probabilities (see e.g. Slijkhuis et al. (1997) and Schweckendiek (2014)). Inherent uncertainties are typically considered irreducible, where as knowledge uncertainty is considered reducible. This framework works well for typical loads on flood defences (a better model reduces model uncertainty but inherent natural variability in annual maxima of river discharges remains irreducible), but is less trivial for strength uncertainties. The strength uncertainty of flood defences mainly arises due to heterogeneity of the subsoil and dike body combined with limited knowledge of this subsoil, in combination with imperfect models describing the strength of the flood defence. In here, most uncertainty could theoretically be reduced but the question is more whether it is economically feasible to do so than whether it is technically possible (Schweckendiek 2014). It is therefore more applicable to use the classification of Phoon and Retief (2016) where geotechnical strength uncertainties are split into natural variability, measurement uncertainty, transformation uncertainty and model uncertainty. All of these uncertainties can be reduced to some extent, but each requires a different measure. For instance, if the source of uncertainty in Pre-Overburden Pressure (POP) is mainly natural variability, more measurements could be applied. However, if the source is a old and inaccurate measurement method, a more accurate method should be applied as there will also be a lot of measurement uncertainty. Hence it is important to systematically distinguish the main uncertainties based on their original source.

When doing a time-dependent reliability analysis this becomes even more important, as some uncertainties (mainly epistemic strength uncertainties and model uncertainties) will be correlated in time, whereas aleatory uncertainties are not. In order to correctly apply the notion of non-failure in preceding years, these uncertainties should be clearly distinguished. In this paper we will focus on the influence of temporal correlation on time-dependent reliability: it has to be noted that also spatial correlation can be used as information. For instance if the same model is used for different dike sections, and model uncertainty is the dominant parameter, failures and non-failures at location A might provide information on the reliability at location B. However this is out of the scope of this paper.

3 RESULTS

3.1 *Case description*

In order to investigate the effects of different formulations of temporal reliability and the influence of different values of uncertainty and correlation for different values of the reliability index we use fragility curves to describe the strength of a flood defence. This is a broadly used method of aggregating failure probabilities from more complex failure models (see e.g. Bachmann et al. (2013) and Schweckendiek et al. (2017)). The fragility curve expresses the critical height h_c which is an integration of the joint probability of the strength given a certain water level, resulting in the following limit state function

$$Z = h_c - h \tag{5}$$

where h is the water level and h_c the critical height. This approach is sound as long as the water level is (strongly correlated to) the dominant load for the mechanism. In this case we consider flood defence reliability described by the aforementioned limit state function where it holds that h_c and h are normally distributed. The Equivalent Planes method requires information about the influence coefficients (α_{ij}) of all i random variables per year j, reliability indices (β_j) for each year j and, as autocorrelations are constant in time, a correlation matrix with dimensions $i*i$. Using this method we can then combine the non-failure and failure events for subsequent years.

3.2 *Example 1: Life-cycle reliability of a dike without deterioration*

First we investigate the life-cycle reliability of a dike without deterioration, so constant h_c. In this case the value for $P_f(1) = ... = P_f(t-1) = P_f(t)$. For the temporal autocorrelation it holds that the strength is fully correlated $(\rho_{h_c} = 1)$, which is typical for many strength parameters of flood defences. The loads are uncorrelated from year to year $(\rho_h = 0)$, as the maximum water level in year i is typically independent of the maximum in year $i-1$. In the examples we will only consider the second definition for annual reliability index which is $\beta(f_t \cap \bar{f}_{1..t-1})$, as the difference with $\beta(f_t \mid \bar{f}_{1..t-1})$ is very small. For instance, if we assume that $\beta = 3$ and $\alpha_{h_c}^2 = \alpha_h^2 = 0.5$, the relative difference in β after 100 years is only 0.5%.

As h_c is assumed to be fully correlated in time, and h is fully uncorrelated, the influence coefficients will have a significant influence on the difference between $\beta(t)$ and $\beta(f_t \cap \bar{f}_{1..t-1})$.

Figure 2 shows the relative change in reliability index β for different values of α_{h_c} for $\beta = 3$. As expected it is observed that for higher values of α_{h_c} the difference is larger. Typical values for the influence coefficient of the strength for failure due to overflow are very small (order of 0.1 or 0.2), but for geotechnical failures these are often in the order $\alpha_{h_c} = 0.75$, meaning that for a lifetime of 50 years the various definitions of the reliability yield a difference in resulting reliability index of 10%. In terms of failure probability this is approximately a factor 3, which is equal to the difference in safety standard for two subsequent categories as defined in the law (e.g. 1/300 to 1/1000). Another important fact is that the reducing effect diminishes over time, which can be explained from the change in α over time, see Figure 3. The fact that the influence coefficient of the correlated variable reduces in time makes intuitive sense as more of the same information will result in increasingly less new insight.

A last important investigation of this simple case is the level of correlation. As was argued in the preceding sections it is important to distinguish different parameters with different uncertainties and different temporal correlations. However,

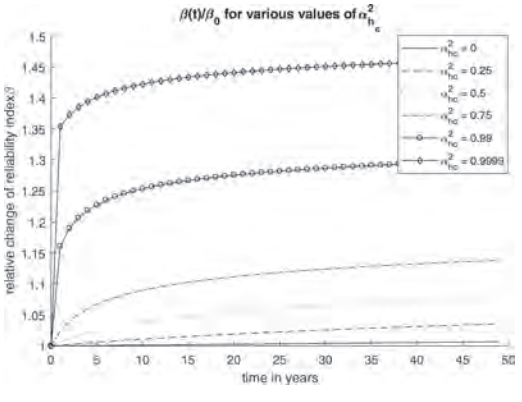


Figure 2. Relative change in $\beta(t)$ for various values of the time correlated α_{h_c} and $\beta(t=0) = 3$.

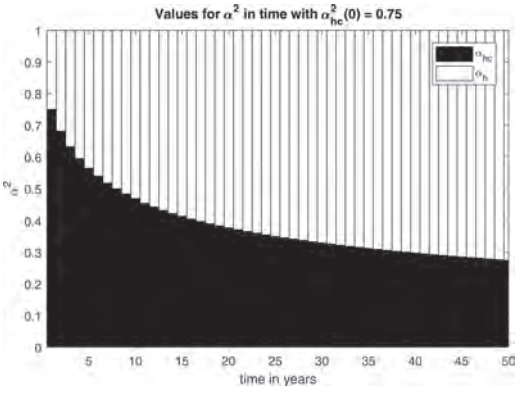


Figure 3. Change in α^2-values over time for a case with $\beta(t=0) = 3$.

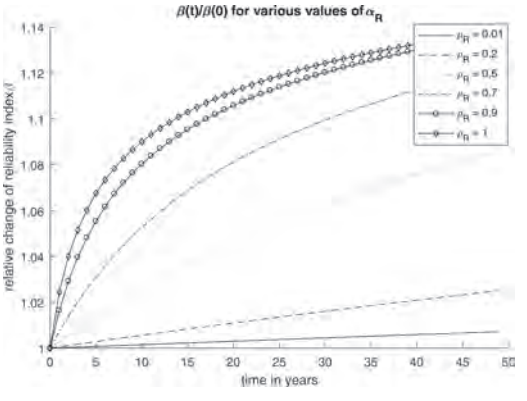


Figure 4. Relative change in $\beta(t)$ for various values of the correlation coefficient ρ_{h_c} and $\beta(t=0) = 3$.

Table 1. Parameters for Example 2.

Variables	Distribution	Parameters	
h_c	$N(\mu, \sigma)$	4.5	1.05
H	$N(\mu, \sigma)$	0	0.6
Δh_c	$\Gamma(\eta, \delta)$	$\frac{1}{var_{\Delta h_c}^2}$	$\frac{1}{\mu_{\Delta h_c} * var_{\Delta h_c}^2}$

Figure 4 shows us that the influence of having a ρ_{h_c} that is slightly smaller than 1 is not that influential on the $\beta(t)$: even for a ρ_{h_c} of 0.7 values close to the ones for full correlation are found. So even if there is a small error due to e.g. combining two (uncertainty) parameters with different time correlations, the influence on the result will often be relatively small.

3.3 *Example 2: Life-cycle reliability of a dike with deterioration*

In practice it will not occur that a dike will remain the same for 50 years. The most common deterioration mechanism for dikes is settlement, which can be described by parametric models (see e.g. Buijs et al. (2009)) or stochastic process models such as the Gamma Process (Pandey and van Noortwijk 2004)). Here we use such a Gamma process. We introduce a new random variable Δh_c which denotes the change in critical height compared to the first year. For the sake of the example we make an important simplification here as we assume that the critical height is fully dominated by the initial crest height and its settlement. For many (geotechnical) failure mechanisms this is not the case, and other types of deterioration will be more dominant. We assume that the average annual settlement is 2 cm with a coefficient of variation of 30%. By splitting the variables we can maintain that $\rho_{h_c} = 1$. We assume $\rho_{\Delta h_c} = 0$ as it is a random process, this leads to the following distributions for the random variables:

The choice of the distributions is such that for the initial situation $\alpha_{h_c} \approx 0.75$, comparable to the first example. The initial $\beta \approx 3.7$. It has to be noted that while we attribute the temporal change to settlement in this case (i.e. decrease in strength), it could also be attributed to an increase in load, for instance due to climate change. The behaviour of such a parameter would be similar: increasing in time with increasing uncertainty.

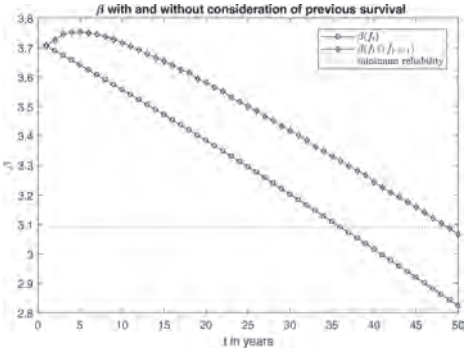


Figure 5. Values for different definitions of β with deterioration

Figure 5 shows the results for the time-dependent reliability. Here it can be seen that for this case the influence of the definition is rather large: when we compare to a minimum required reliability $\beta = 3.09 \left(P_f = 1/1000\ yr^{-1} \right)$, the expected extended life when taking into account survival is approximately 15 years (or: an extension of the lifetime by almost 50%).

This amount of lifetime extension however is dependent on the rate of deterioration, and especially the uncertainty in deterioration. Figure 6 shows the α^2-values for two rates of deterioration, on the left is the same as used in Figure 5, the right is a distribution with higher variation and slightly lower mean, such that $\beta(f_{50})$ is equal for both cases. However for the deterioration with high variation $\beta\left(f_{50} \cap \overline{f}_{1..49}\right)$ is significantly smaller than for the case with smaller variation. This can be explained by a smaller $\alpha^2_{h_c}$ in the design point, meaning that the influence of that uncertainty on the reliability is smaller, resulting in less valuable non-failure information.

3.4 *Economic implications of time dependent reliability*

Generally the goal of flood defence management is to maintain flood defences at a desired level of reliability, against acceptable costs. In many cases Life Cycle Costing (LCC) is used to evaluate costs in time, for which the principles were first reported by Samuelson (1937). In an LCC analysis the Net Present Value, which denotes the value in current day prices, is calculated using the following formula:

$$NPV = \sum_{i=1}^{t} \frac{C_i}{(1+r)^i} \tag{6}$$

where, C_i is the total cost in year i, r is the discount rate and t is the evaluation period. One of the major implications of this economic theory is that postponing an investment yields significant benefits. For instance: if we postpone an investment by 10 years, assuming a discount rate of 3%, the current stand-

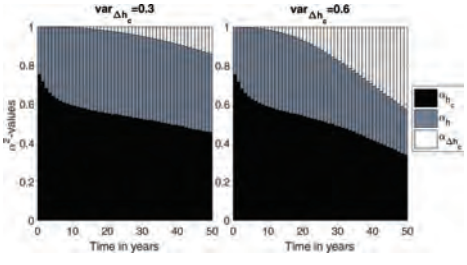


Figure 6. Change of α^2 in time for different rates of deterioration.

ard in the Netherlands (Werkgroep Discontovoet 2015), the cost after ten years is only 75% of the cost, expressed in present day prices. A disadvantage of using LCC is that it is slightly harder to compare investments with different lifetimes. For such comparisons the NPV can be expressed as Equivalent Annual Cost (EAC), which is calculated using the following formulas (Schoemaker et al. 2016):

$$EAC = \frac{NPV}{A_{t,r}} \tag{7}$$

$$A_{t,r} = \frac{1-(1+r)^{-t}}{r} \tag{8}$$

where $A_{t,r}$ is the Annuity factor for year t and discount rate r, which denotes the sum of the discount factors compared to $t = 0$.

In Figure 5 it was shown that the definition of time-dependent reliability can have a significant influence on the lifetime of a structure. To further investigate this we consider 4 situations, and generate distributions for the initial strength h_c corresponding to a wide range of α-values. The 4 cases contain 3 cases (1, 2 and 3) with different reliability requirements and 1 case (case 4) with an adapted uncertainty for the settlement (similar to the comparison of different deterioration rates in the previous section). The different cases are summarized in Table 2. It is expected that the case with low target reliability has the largest life extension, as here a non-failure is more relevant than for the case with a very high reliability. Also, based on the α-plots in Figure 6, where we saw a more rapid decrease in α_{h_c} for a higher variation of the settlement, we would expect the increase in lifetime for the case with small variation in settlement to be larger.

Figure 7 shows the extension of the lifetime in years for the 4 considered cases. Here it can indeed be seen that for a lower reliability larger extensions are gained, and that for lower uncertainty in deterioration the effect of non-failures in preceding years is also larger. It has to be noted that the lines are a bit wobbly, which is due to the fact that the reliability is determined per year, resulting in discrete steps and therefore small wobbles. However the lifetime extension doesn't directly translate into financial benefits. Therefore in Figure 8 the relative savings following from the postponement of a new reinforcement are shown for a discount rate of 3%. Here it follows that in the most extreme case (high $\alpha^2_{h_c}$ for Case 2) the relative savings can amount up to a factor 3. These relative savings are independent of other investments during the life cycle, and also independent of the actual reference year as the relative savings are linear in time (due to the exponential character of the discount rate). For instance for assessments of existing structures this would be a relevant value, as the reference year wouldn't be 0 but somewhere between 0 and the end-of-life. However in design decision making the change in Equivalent Annual Costs is more relevant, as this denotes the economic yearly cost for a design option. If we assume that at t = 0 a reinforcement is made for a lifetime of 50 years, for a discount rate of 3% a reduction in Equivalent Annual Cost of approximately 12.5% is realized for a lifetime extension to 70 years, which is in accordance with Figure 8. Obviously this will not hold for all flood defences in the Netherlands, but mainly in the

Table 2. Cases for analysis of lifetime extension.

Case	$\beta(P_f)$	$\boldsymbol{var}_{\Delta h_c}$
1	$2.326(10^{-2})$	0.3
2	$3.090(10^{-3})$	0.3
3	$3.719(10^{-4})$	0.3
4	$3.090(10^{-3})$	0.05

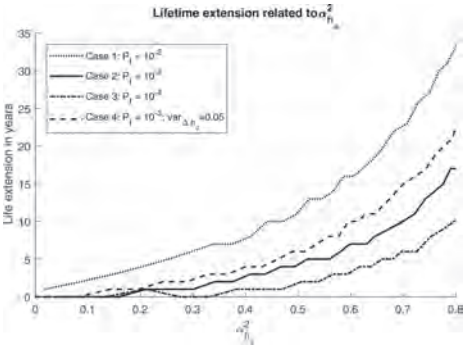


Figure 7. Life extension in years for different α_{h_c}-values for the 4 considered cases.

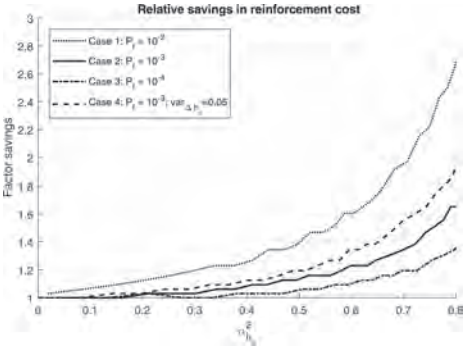


Figure 8. Factor of relative savings for the reinforcement cost for the 4 considered cases

riverine area failure probabilities are dominated by geotechnical failure mechanisms, meaning that such lifetime extensions will not be uncommon.

4 CONCLUSIONS AND RECOMMENDATIONS

In this paper we have explored various aspects of temporal dependence in reliability of flood defences. As flood defence reliability is often determined by highly uncertain but temporally correlated strength parameters, non-failures in previous years constitute information that can improve the estimate of the strength, especially when the estimated reliability is low. In this paper we've explored some parametric cases where it is shown that the savings due to accounting for non-failures rather than the commonly used conservative approach of disregarding temporal dependence in reliability can be significant (up to 20 years in lifetime extension). This is of relevance for many aspects of flood defence management such as design guidelines as well as assessment of existing flood defences where often low reliabilities are found from models. It has to be further investigated for different decision problems in flood defence management what the consequences of accounting for this temporal dependence are, especially by looking in more detail into the sources of uncertainty of actual dike reliability analyses. However, based on the considered cases it is expected that it can significantly improve reliability estimates in assessment and design, especially when there is large uncertainty on strength parameters that are correlated in time. It has to be noted that in such cases it might be necessary to improve the knowledge on strength parameters through obtaining additional information. Due to the high strength uncertainties that are often encountered, the consideration of the value of improved information is one of the major decision problems to be studied for flood defence management.

ACKNOWLEDGEMENTS

This work is part of the Perspectief research programme All-Risk with project number P15–21, which is (partly) financed by NWO Domain Applied and Engineering Sciences.

REFERENCES

Bachmann, D., N.P. Huber, G. Johann, & H. Schüttrumpf (2013). Fragility curves in operational dike reliability assessment. *Georisk 7*(1), 49–60.

Buijs, F., J. Hall, P. Sayers, & P. Van Gelder (2009). Timedependent reliability analysis of flood defences. *Reliability Engineering & System Safety 94*(12), 1942–1953.

CEN (2002). Eurocode—Basis of structural design. *En 3*(December 2008), 89.

Finkelstein, M. (2008). *Failure Rate Modelling for Reliability and Risk* (Springer S ed.), Volume 2008. Springer London.

Jongejan, R.B. & B. Maaskant (2015). Quantifying flood risks in the Netherlands. *Risk analysis: an official publication of the Society for Risk Analysis 35*(2), 252–64.

Kok, M., R. Jongejan, M. Nieuwjaar, & I. Tanczos (2017). Fundamentals of Flood Protection.

Kottegoda, N. & R. Rosso (2008). *Applied statistics for civil and environmental engineers* (2nd ed.). Blackwell Publishing Ltd.

Pandey, M.D. & J.M. van Noortwijk (2004). Gamma process model for time-dependent structural reliability analysis. In E. Watanabe, D.M. Frangopol, and T. Utsunomiya (Eds.), *Bridge maintenance, safety, management and cost,* pp. 1–8.

Leiden: A.A. Balkema. Phoon, K. & J. Retief (2016). *Reliability of Geotechnical Structures in ISO2394.* Number September.

Roscoe, K., F. Diermanse, & T. Vrouwenvelder (2015). System reliability with correlated components: Accuracy of the Equivalent Planes method. *Structural Safety 57,* 53–64.

Samuelson, P.A. (1937). A Note on Measurement of Utility. *The Review of Economic Studies 4*(2), 155–161.

Schoemaker, M.A., J.G. Verlaan, R. Vos, & M. Kok (2016). The use of equivalent annual cost for cost-benefit analyses in flood risk reduction strategies. In *FLOODRisk 2016,* Volume 20005.

Schweckendiek, T. (2014). On reducing piping uncertainties: A Bayesian decision approach.

Schweckendiek, T., M.G. van der Krogt, A. Teixeira, W. Kanning, R. Brinkman, & K. Rippi (2017). Reliability Updating with Survival Information for Dike Slope Stability Using Fragility Curves. In *Geo-Risk 2017,* Reston, VA, pp. 494–503. American Society of Civil Engineers.

Slijkhuis, K., P.H.A.J.M. van Gelder, & J. Vrijling (1997). Optimal dike height under statistical-, construction- and damage uncertainty. In Y.W.N. Shiraishi, M. Shinozuka (Ed.), *Structural Safety and Reliability,* Volume 7, Kyoto, Japan, pp. 1137–1140.

Slomp, R., H. Knoeff, A. Bizzarri, M. Bottema, & W. de Vries (2016). Probabilistic Flood Defence Assessment Tools. *E3S Web of Conferences 7,* 03015.

Speijker, L., J.M. van Noortwijk, M. Kok, & R.M. Cooke (2000). Optimal maintenance decisions for dikes. *Probability in the Engineering and Informational Sciences 14*(1), 101–121.

Van Gelder, P. (2000). Statistical methods for the risk-based design of civil structures.

Walker, W., P. Harremoës, J. Rotmans, J. van der Sluijs, M. van Asselt, P. Janssen, & M. Krayer von Krauss (2003). Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integrated Assessment 4*(1), 5–17.

Werkgroep Discontovoet (2015). Rapport werkgroep discontovoet 2015. pp. 95.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Impact assessment of road infrastructure: A holistic approach

Y.Z. Ayele
Faculty of Engineering, Østfold University College, Fredrikstad, Norway

ABSTRACT: As construction companies attempt to maximize the value of each road project and optimize their investment portfolio, it has become vital to assess the impact of road transport projects for their cost effectiveness in order to support well-informed decision-making. Proper impact assessment will result in road design that is more advanced, efficient operations, and improved environmental protection. Identifying cost-effective road design, which have a minimal environmental footprint, however, is one of the biggest challenges for road construction industries. The purpose of this paper is thus to propose a holistic methodology for assessing the impacts of the road projects, by taking into account the available natural resources, the environment, and by involving all relevant stakeholders. The paper also outlines steps to assess both monetised and non-monetised impacts. By employing, the proposed steps, the pros and cons of new road networks can be identified in a structured way and hereby highlight factors of success and failure.

1 INTRODUCTION

Across the world, inadequate or poorly designed and performing road infrastructures leads to major financial and social challenges, which governments and businesses need to address (Dobbs et al., 2013, Ayele et al., 2013). Further, in broad sense, road network designs without a proper impact assessment, resulted in a road that is not adequate for meeting the current and future road transport needs of a given society (Systematics, 2012). Hence, in investment decisions, the relevant decision-maker often carry out a cost-effectiveness analysis (CEA) and cost-benefit analysis (CBA) of road network projects. When identifying and categorizing the costs and benefits of road design alternatives, a reasonable effort has to be made to identify those costs that will have the most significant implications on the strategic decision.

To identify cost-effective road network design, which have a minimal environmental footprint, it has been argued that two questions are fundamental (Macias and Gadziński, 2013, Hanley, 2001, Ayele et al., 2016b). Firstly, which road design alternative is estimated to be cost-effective and environmentally sustainable, based on the prevailing evidence? Secondly, should further research be carried out in order to minimise the level of uncertainty related to the decision? To answer these questions and, determine the cost-effectiveness of the road network design, number of studies have been carried out, see e.g. Hanley (2001), Macias and Gadziński (2013), Nagurney et al. (2010), and Welde et al. (2013). For instance, Welde et al. (2013) compared the implementation of CEA and CBA, which are the two common economic evaluation methods, as tools for allocation of national public funds in the transport sector. Moreover, Macias and Gadziński (2013) discussed issues of road transport networks influence on the natural environment. Furthermore, for better understanding of the environmental impacts, Nagurney et al. (2010) proposed environmental impact assessment (EIA) indices for evaluating the environmental effects of link capacity degradation in transportation (road) networks.

However, in most of the available CEA and CBA literature, either cost or environmental impact is the only factor considered; and there is a lack of consideration of the available natural resources, the environment, and all relevant stakeholders when assessing the impacts (monetised and non-monetised) of the road projects. Further, the adequacy of the models used to estimate the various costs and benefits of the road network investment projects is increasingly being called into question (Systematics, 2012).

This is considered as a significant drawback since inadequate or poorly designed infrastructure presents major economic and social challenges. Further, understanding how a specific decision or choice of assessment method affects the intended goals is a key in identifying a cost-effective road network design. Proper CEA and CBA will result in more advanced road network design, more efficient operations, and improved environmental

protection. Moreover, it can be used to identify weaknesses or strengths of existing or new road networks in a structured way and hereby highlight factors of success and failure. It is also a core element in examining the overall quality of the road design alternatives.

Hence, the purpose of this paper is to propose a holistic methodology for assessing the impacts of the road projects, by taking into account the available natural resources, the environment, and by involving all relevant stakeholders. The paper also outlines steps to assess both monetised and non-monetised impacts. The rest of the paper is organised as follows: Sect. 2 presents a problem description. Section 3 discusses the concept of CBA and CEA process. Section 4 introduces the proposed holistic methodology for assessing the impacts of the road networks. Section 5 provides the concluding remarks.

2 PROBLEM DESCRIPTION

The problem considered here is an impact assessment problem of the road networks. Suppose we have a finite number of road network design alternatives, each with a different construction, maintenance and HSE (health, safety, and environmental) costs. The idea is to use the most suitable design alternatives with a minimum environmental footprint. However, considering the available natural resources, the environment, and the involving relevant stakeholders when assessing the CEA and CBA of road design alternatives, a decision-maker will face a time-variant decision making process. In other words, the decision maker is faced with an optimisation problem, since these factors can be considered as covariates. A covariate, in the context of this paper, is a factor that can have an influence on the direct and indirect costs as well as the decisions of the road projects.

To optimise the effectiveness and benefit and, to determine what road design alternative will be used, the proposed holistic methodology constantly assesses the impacts of the road projects, by considering the existing natural resources, the environment, and by involving all relevant stakeholders. The proposed methodology attempt to capture the monetised and non-monetised impacts of new road systems, in a structured way and hereby highlight factors of success and failure.

3 A GENERIC CEA AND CBA PROCESS

Assessment tools are structured procedures or methods which, when used, result in objective and replicable information; either on the technological and design appropriateness of the decision choice or the local allowing or restricting conditions that can influence the attainment or non-attainment of the decision choice (it can be technical or managerial) (Zurbrügg et al., 2014). A range of assessment methods are often developed for assessing a specific aspect of the road networks, such as technical, environmental and health, economic and financial, social and institutional, organizational aspects; and, others attempt to provide a more holistic picture by integrating different assessment methods into the same tool; see e.g. (Welde et al., 2013, Hanley, 2001, Nagurney et al., 2010, Macias and Gadziński, 2013). When assessing environmentally sustainable and cost-effective road project, in general, three main issues govern the final decision. Figure 1 illustrates these three main issues, which needs to be considered while assessing environmentally sustainable and cost-effective road project.

3.1 *CBA and CEA*

3.1.1 *Cost-Benefit Analysis (CBA)*

CBA is an analytical method for assessing the total costs and benefits from the planned project (Finnveden et al., 2007, Skovgaard et al., 2007). It provides a starting point from which to begin evaluation of a project; can provide quantitative data to back up qualitative arguments; it does allow interested parties to clearly define the issues involved; and it allows comparisons to be made between investments or projects (Gerald Shively and Marta Galopin, 2014). However, it requires that the analyst assign monetary values to all benefits and costs, however, there are numerous benefits and costs which are intangible and therefore

Figure 1. An overview of the relationship between the main aspects of road network.

difficult to value; CBA results can be very sensitive to the choice of the discount rate; and some future benefits and costs cannot conceived, much less measured (Gerald Shively and Marta Galopin, 2014).

3.1.2 *Cost-Effectiveness Analysis (CEA)*

A generic CEA process involves: i) determining which cost variables affect the cost-effectiveness of the chosen road design solution. This includes determining and analysing: (a) internal cost factors, which arise because of company decisions and goals; the company largely manages these cost factors and, if necessary, they can be changed and (b) external cost factors, which are not controlled by the company but will influence the overall cost & the decisions. ii) Determining inherent risk factors for the chosen design highway practices and the company tolerance for them. iii) Determining the functional interdependence between the cost and risk variables and the degree to which each of these variables can be controlled. Figure 2 illustrates the key steps in the CEA.

CEA focuses on the costs on achieving the goals of the system and the most efficient way of achieving it (Finnveden et al., 2007, Ayele et al., 2014). The benefits of achieving the goals do not have to be quantified (Finnveden et al., 2007). CEA is most useful when analysts face constraints which prevent them from conducting CBA (Tan-Torres Edejer et al., 2003, Ayele et al., 2014). For instance, the most common constraint is the inability of analysts to monetize benefits. Though the basic cost-effectiveness calculation appears to be simple, choices about units of measurement, scope of costs, and prices to be included not only will alter the numerical results but also will affect the interpretation of the cost-effectiveness ratio (The International Bank for Reconstruction and Development and The World Bank Group, 2006).

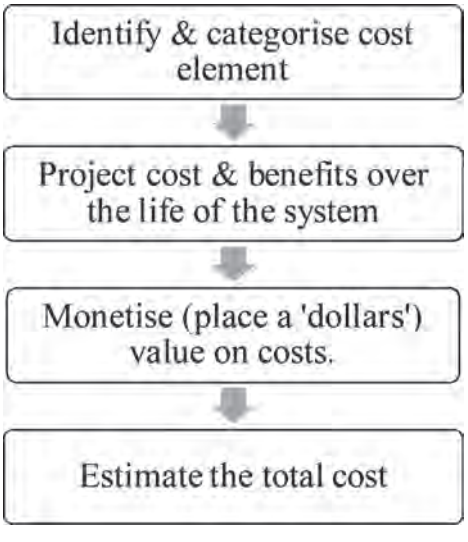


Figure 2. Key steps in the CEA.

3.1.3 *Cost-Effectiveness Ratio (CER)*

The CER is a ratio in which the denominator is the unit of effectiveness and the numerator is the present value of the cost of a particular road infrastructure. Units of effectiveness are a measure of any quantifiable outcome central to the road design objectives. In road construction, the total volume of construction material required would be the most important outcome and, would be an obvious unit of effectiveness. Mathematically, CER for a specific road infrastructure, based on Cellini and Kee (2010), can be described as:

$$CER_{RInf.} = \frac{PVC_{RInf.}}{U_{RInf.}} \tag{1}$$

where:

- $U_{RInf.}$ is the unit of effectiveness of road infrastructure (*RInf.*), and
- *PVCRInf.* is the present value of cost of road infrastructure (*RInf.*), and is given by:

$$PVC_{RInf} = TC_{RInf\cdot 1} + \frac{TC_{RInf\cdot 2}}{(1+r)^1} + \frac{TC_{RInf\cdot 3}}{(1+r)^2} + \ldots + \frac{TC_{RInf\cdot t}}{(1+r)^{t-1}} = \sum_{t-1}^{T} \frac{TC_{RInf\cdot t}}{(1+r)^{t-1}} \tag{2}$$

where:

- $TC_{RInf\cdot 1}$ is the annualised total cost of road infrastructure (RInf.),
- t indicates the year from 1 to T (the last year of the analysis),
- r is discount rate, which is meant to reflect society's impatience or preference for consumption today over consumption in the future.

CER results are very sensitive to the choice of the discount rate, and thus an appropriate choice of the discount rate is critical, and there is ongoing and considerable debate as to the appropriate rate, see e.g. Stern (2006), Lopez (2008) and Cellini and Kee (2010).

3.2 *Environmental Impact Assessment (EIA)*

EIA is a process for assessing environmental consequences (positive and negative) because of the planned project. Conducting an EIA during the project design phase will provide information

about the environmental conditions of the area; early assessment of negative impacts can ensure that appropriate mitigation measures and opportunities are identified and implemented; and it can help to reduce costs in the long term (Jonathan R. and Emma J., 2010). However, some of the reasons given for the non-use of EIA tools are that they are too cumbersome, time consuming, and generalized; another reason is the lack of evidence confirming the actual value and success of EIA's; and in contrast to other assessment methods EIA is generally site-specific (Jonathan R. and Emma J., 2010, Finnveden et al., 2007).

3.3 *Engaging relevant stakeholders*

The impact analysis can only be undertaken when stakeholders agree on a minimum range of prerequisites (points). In addition, during the effectiveness assessment the inclusion of stakeholders, intergenerational equity as well as the satisfaction of the social needs can be assured. That means the involvement of all relevant stakeholders in the decision-making process can be corroborated (Morrissey and Browne, 2004).

4 PROPOSED HOLISTIC METHODOLOGY

The main steps for assessing the impacts of the road project is depicted in Figure 3. The methodology has various key elements; and, it takes into account the available natural resources, the environment, relevant stakeholders. The first stage, in the proposed methodology, is to specify the goals and criteria based on the available standards, regulations, and recommended guidelines. One of the main criteria is that the result from the assessment should support the decision-maker to clearly identify the consequences of the chosen road network design alternative. Moreover, when defining goals for effective and environmentally sustainable road network design, the inclusion of all relevant stakeholders must be ensured.

After defining the objectives, the goals, and the criteria, in the next stage, the road network design options should be specified. In this stage, the aspects of road network design as well as the features of transport network expansion capacity should be outlined. The road network design, in general, deeds with the configuration of network to achieve specified objectives (Tom and Mohan, 2003). Further, there are two variations to the design problem, the continuous network design and the discrete network design. Some of the design problem include, determining the width of the road, setting of road user charges, calculation of signal timings, etc. (Tom and Mohan, 2003).

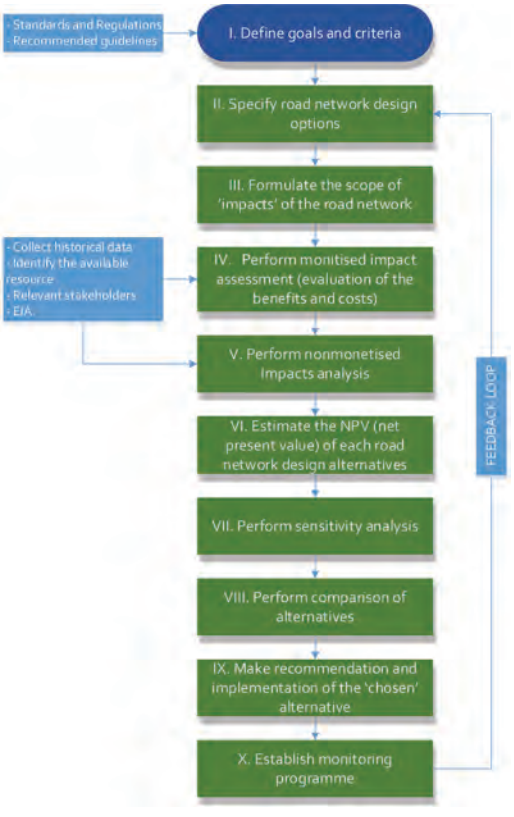


Figure 3. The proposed holistic methodology for assessing the impact of the road infrastructure.

Formulating the scope of the impacts of the road network is the next stage. In this stage, it is essential to identify the scope of the impacts, such as the influence of the road networks in the landscape ecology. In broad sense, with construction of road networks, the range of human activities also expands, which increases the human impact on the natural ecology. Hence, in this stage, the scope should be clearly defined the facet of the impact analysis.

In the next stage, the monetized impact assessment or evaluation of the benefit and cost of the alternatives should be carried out. Figure 4 depicts the key parameters that needs to be quantified. In this stage, the analyst assigns monetary values to all benefits and costs. Since the CBA results can be very sensitive to the choice of the discount rate, the analyst should be vigilant. It is also important to note that some future benefits and costs cannot conceived, much less measured (Gerald Shively and Marta Galopin, 2014).

After assessing and quantifying the impact, thereafter, non-monetised impact assessment

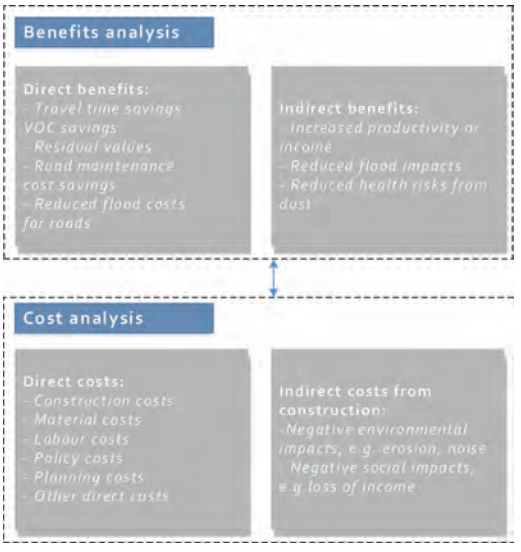


Figure 4. Monitised impact assessment: Benefit and cost analysis, modified from Systematics (2012).

should be performed. In this stage, the main emphases are on assessing the costs on achieving the goals of the network design and the most efficient way of achieving it. Here, the benefits of achieving the goals do not have to be quantified in monetary terms. This stage is most useful when analysts face constraints which prevent them from conducting monetized impact analysis (Tan-Torres Edejer et al., 2003). For instance, the most common constraint is the inability of analysts to monetise benefits.

The next stage, in the proposed methodology, is estimating the net present value (NPV) of each road network design alternatives. Here, by employing the NPV method, the benefit and cost of a given road design alternative over an extended period of time should be calculated and then discounted at a selected discount rate to give the present value (Sinha and Labi, 2011). Benefits are treated as positive and cost as negative and the summation gives the NPV (Sinha and Labi, 2011). In general, any road design with positive NPV is treated as acceptable. In comparing more than one design alternatives, an alternative with higher NPV should be accepted.

After estimating the NPV for each alternative, in the next stage, the sensitivity analysis should be performed. The aim of the sensitivity analysis is to recognize the vital cost variables and their potential impact in terms of changes in the annualised total cost and present value cost. In general, there are two common sensitivity analysis – partial and extreme cases sensitivity analyses. The partial approach varies one assumption (or one parameter or number) at a time, holding all else constant (Cellini and Kee, 2010). On the other hand, extreme case sensitivity analysis varies all of the uncertain parameters simultaneously, picking the values for each parameter that yields either the best- or worst-case scenario (Cellini and Kee, 2010, Ayele et al., 2016a).

Thereafter, the comparison between the alternative design options should be performed. The comparison should consider the strengths and weaknesses of each alternative, check the extent and the validity of the 'results' obtained from the assessment, and investigate whether a given design alternative match the intended plan and outcomes, etc. Further, at this stage, specific effectiveness indicators should be clearly identified; the advantage and disadvantage of the proposed road design alternatives should be assessed; and, the compilation of HSE standards should be ensured, by verifying the consideration and integration of the social aspects.

Thereafter, the recommendation should be made and, afterwards the chosen road design alternative should be implemented. The recommendation should comply with road design standards and guidelines. Afterwards, the approved recommendation should be monitored by establishing the monitoring programme. That means the monitoring programme has to include the follow-up of the implementation of the chosen design alternative, by emphasising what does work, what does not work, and what continues to work. Further, there should be a feedback loop where the recommendations should help to review the impact assessment and, the choice of the alternatives.

5 CONCLUDING REMARKS

To address the challenges related with inadequate or poorly designed and performing road infrastructures, proper impact assessment of road network design alternatives is vital. This consequently helps to ensure cost-effective and environmentally sustainable road infrastructure design. Furthermore, proper impact assessment can have a significant economic, environmental, and social benefit, by reducing the HSE (health, safety, and environmental) impacts, identifying cost-effective and environmentally sustainable road.

This paper proposed a holistic methodology for assessing the impact of the road networks. The proposed methodology is based on the consideration of possible road design alternatives. The first part of the proposed methodology presents the problem description and discuss the concept

of cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA) process. The second part of the paper describes estimation of the net present value (NPV) by categorising the road networks impact into two: monetised and non-monetised. By employing, the proposed methodology the pros and cons of new road systems can be identified in a structured way and hereby highlight factors of success and failure.

REFRENCES

Ayele, Y. Z., Barabadi, A. & Droguett, E. L. 2016a. Risk-Based Cost-Effectiveness Analysis of Waste Handling Practices in the Arctic Drilling Operation. *Journal of Offshore Mechanics and Arctic Engineering*, 138, 031301.

Ayele, Y. Z., Barabadi, A. & Barabady, J. Effectiveness assessment for waste management decision-support in the Arctic drilling. IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2014, 2014, 559–564.

Ayele, Y. Z., Barabadi, A. & Droguett, E. L. 2016a. Risk-Based Cost-Effectiveness Analysis of Waste Handling Practices in the Arctic Drilling Operation. *Journal of Offshore Mechanics and Arctic Engineering*, 138, 031301.

Ayele, Y. Z., Barabadi, A. & Barabady, J. 2016b. Dynamic Spare Parts Transportation Model for Arctic Production Facility. *International Journal of System Assurance Engineering and Management*, 7, 84–98.

Ayele, Y. Z., Barabadi, A. & Markeset, T. 2013. Spare part Transportation Management in High North. Proceedings of the International Conference on Port and Ocean Engineering under Arctic Conditions, POAC2013.

Cellini, S. R. & Kee, J. E. 2010. Cost-effectiveness and cost-benefit analysis. *Handbook of practical program evaluation*, 493.

Dobbs, R., Pohl, H. & Lin, D. 2013. Infrastructure Productivity: How to Save $1 Trillion a Year. McKinsey Global Institute. Print.

Finnveden, G., Björklund, A., Moberg, Å. & Ekvall, T. 2007. Environmental and economic assessment methods for waste management decision-support: possibilities and limitations. *Waste management & research*, 25, 263–269.

Gerald Shively & Marta Galopin. 2014. *An Overview of Benefit-Cost Analysis* [Online]. Department of Agricultural Economics, Purdue University. Available: http://www.agecon.purdue.edu/staff/shively/COURSES/AGEC406/reviews/bca.htm [Accessed 20.04 2014].

Hanley, N. 2001. Cost-Benefit Analysis and Environmental Policymaking. *Environment and Planning C: Government and Policy*, 19, 103–118.

Jonathan R. & Emma J. 2010. *Environmental Impact Assessment Tools and Techniques* [Online]. Available: http://green-recovery.org/wordpress/wp-content/uploads/2010/11/Module-3-Content-Paper.pdf [Accessed 20.04 2014].

Lopez, H. 2008. The social discount rate: Estimates for nine Latin American countries. *World Bank Policy Research Working Paper Series, Vol*.

Macias, A. & Gadziński, J. 2013. Assessment of Road Transport Environmental Impact as Illustrated by a Metropolitan Area. *Polish Journal of Environmental Studies*, 22.

Morrissey, A. & Browne, J. 2004. Waste management models and their application to sustainable waste management. *Waste Management*, 24, 297–308.

Nagurney, A., Qiang, Q. & Nagurney, L. S. 2010. Environmental impact assessment of transportation networks with degradable links in an era of climate change. *International Journal of Sustainable Transportation*, 4, 154–171.

Sinha, K. C. & Labi, S. 2011. *Transportation decision making: Principles of project evaluation and programming*, John Wiley & Sons.

Skovgaard, M., Ibenholt, K. & Ekvall, T. 2007. Nordic guideline for cost-benefit analysis of waste management. *Nordic Council of Ministers*.

Stern, N. H. 2006. *Stern Review: The economics of climate change*, HM treasury London.

Systematics, C. 2012. Norwegian Road Network Strategic Assessment: Re-examining the Estimation of Costs and Benefits of Investments in Road Transport in Norway. Juli.

Tan-Torres Edejer, T., Baltussen, R., Adam, T., Hutubessy, R., Acharya, A., Evans, D. B. & Murray, C. J. L. 2003. Making choices in health: WHO guide to cost-effectiveness analysis.

The International Bank for Reconstruction and Development & the World Bank Group 2006. Priorities in Health: Chapter 3 Cost-Effectiveness Analysis.

Tom, V. & Mohan, S. 2003. Transit route network design using frequency coded genetic algorithm. *Journal of transportation engineering*, 129, 186–195.

Welde, M., Eliasson, J., Odeck, J. & Börjesson, M. 2013The Use of Cost-benefit Analyses in Norway and Sweden: A Comparison. European Transport Conference 2013.

Zurbrugg, C., Caniato, M. & Vaccari, M. 2014. How Assessment Methods Can Support Solid Waste Management in Developing Countries—A Critical Review. *Sustainability*, 6, 545–570.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Harmonizing normative organizational structures and serification & validation concepts for safety critical generic projects

E.H. Dogruguven
Aselsan Inc., Ankara, Turkey

I. Ustoglu
Department of Control and Automation Engineering, Yildiz Technical University, Istanbul, Turkey

ABSTRACT: Development of a Generic Application or Product (GAP) is always challenging, because it should be designed as much as generic and simply configurable for the final specific applications. If such a development is also safety critical, then the complexity increases dramatically as the resulted system will have impacts on human life, property and environment. The safety management plays a major role to keep this tough development process under control by avoiding systematic faults. Setting up correct organizational structure as well as applying Verification & Validation (V&V) concepts, which are two fundamental elements of safety management, in an accurate way are therefore crucial. This paper discusses railway safety management in terms of organizational structure and V&V with regards to the current normative status with its drawbacks. Proposals are provided for an updated organization and more harmonized V&V concepts including relations with safety management and quality assurance by sharing practical experiences.

1 INTRODUCTION

As Murthy et al. (2008) explain regulatory requirements, customer requirements, and technical requirements shall be fulfilled when producing a safety instrumented system. Independent safety assessment against the international and/or European standards shall be realized for the acceptance and approval of the system. If necessary, national standards, regulations and directives are also to be followed. These are shown in a hierarchical manner in Figure 1.

For railway applications, both On-board and trackside, the European Norms EN 50126, EN 50129, EN 50128, and EN 50159, dealing with RAMS, electronic system, SW and transmission, respectively, are referenced in European countries, and at the same time they are also well accepted and applied in many other non-European countries. These norms are derived from the core norm IEC 61508 like many other European and international norms as depicted below in Figure 2.

Despite the fact that they are very detailed and strict as they are related to complex safety critical electronic systems and subsystems, Hokstad and Corneliussen (2004) draw attention to a couple of ambiguities about safety unavailability from quantitative evaluation perspective diminishing the usefulness of the standards. This paper focuses

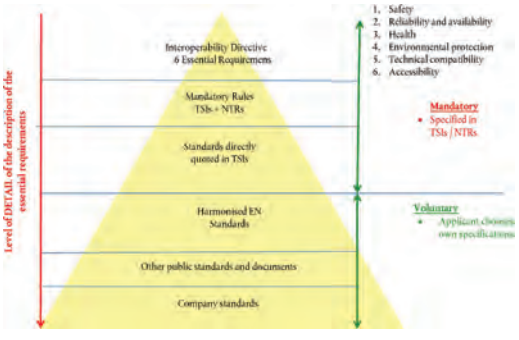


Figure 1. Hierarchy of laws [EU Recommendations commission (2014)].

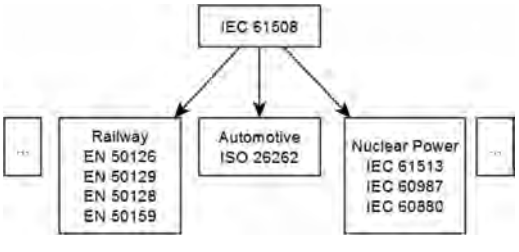


Figure 2. Derivation of standards from IEC 61508.

on two main issues from the qualitative perspective regarding SIL (Safety Integrity Level) 4 development. It is organized as follows. Section I gives a brief information about the standards. In Section II, the organization and responsibilities in the norms are distilled and experiences are highlighted by proposing a new role and responsibilities taking into account the critical duties in a safety critical project. In Section III, Verification & Validation (V&V) concepts confused usually not only on the account of their different understanding in other domains like defense or avionic, but also their different descriptions inside the CENELEC norms themselves are discussed and accordingly the confusing concepts are clarified. Moreover, relations between verification, validation, safety management, quality assurance and assessment are explained. In section II and III, project experiences gained during SIL 4 GAP for mainline On-board signalling part of European Rail Traffic Management System—European Train Control System (ERTMS ETCS) and metro line On-board signalling part of Communication Based Train Control System conjointly with wayside interlocking system assessed independently by a European Notified Body are utilized. Finally, the paper conveys some concluding remarks.

2 ORGANIZATION AND RESPONSIBILITIES

Firstly, the organizational independence for SIL 4 is illustrated as in Figure 3 in EN 50129, the Railway Standard for "Safety related electronic systems for signalling", referred to the System and HW development. In the illustration, same person and same organization arrangements are depicted as solid and dashed lines, respectively. In the figure, PM stands for Project Manager, DI for Designer/ Implementer, VER for Verifier, VAL for Validator. Having provided the illustration and mentioned the independence of roles in Table E.3 in EN 50129, there are no more additional explanations about the independence of roles in this norm.

Secondly, when the norm EN 50128, "Software for railway control and protection systems" is examined, it gives much more detail about the organizational independence (see Figure 4). It defines not only new roles such as Requirements Manager (REQ), Implementer (IMP), Integrator (INT) and Tester (TST), but also new type of vertical lines for the responsibility dependency such that RQM, DES (Designer), IMP shall report to the PM; INT, TST and VER can report to the PM whereas VAL shall not report to PM.

Differently from the norm for electronic systems, this norm elaborates the roles by thirteen

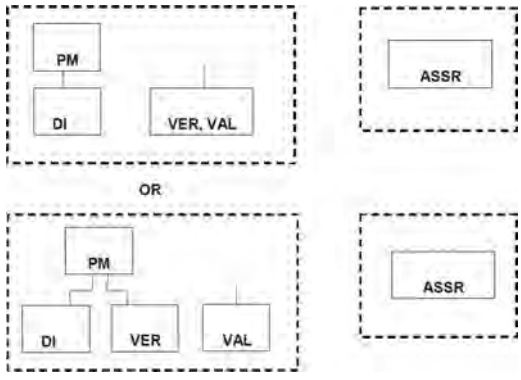


Figure 3. Organizational independence [EN 50129 (2003)].

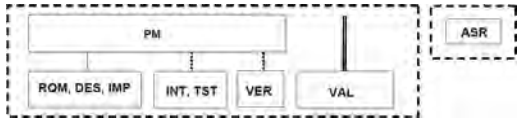


Figure 4. Preferred organizational structure [EN 50128 (2011)].

clauses. In addition, it offers two options for the roles validator and verifier. It allows that the VAL and the VER can be same person guarantying independence from the PM where in this case there shall be a further person who reviews the documents of VER and who does not report to PM. A similar approach is brought to the role of VER in the way that VER and INT/TST can be the same person provided VAL performs verification tasks, too, such that there are double checks.

If considered, the similar tasks are also realized in system and HW development phases. For instance, there will also be HW tests, and a test report and a further verification report, however though the integrity level is the same, the organizational independencies differ from each other which means that the VER can perform test activities and write both test and verification report. We therefore propose at this step that the organizational structures in these norms shall be harmonized and the one stated in EN 50128 can be used for this harmonization purpose, as it declares more roles and therefore more checks which are inevitable in a safety critical system. Another problem for the independency requirements arises from the need of several verifiers in a project, as the verification activities do require specialization in the technique. In this case, a lead verifier can be assigned to the interface and will be responsible for fulfilling all the verification activities. In case there is a limited number of available specialized personnel, this

time there arise personnel availability problems. To overcome this issue, it can be proposed that VER of a work-package can be allocated as DI of another work-package.

Besides, intensive safety analyses (Table 1) shall be performed in safety critical systems to derive safety requirements, both at system and subsystem level. These analyses require broad technical knowledge and long term of experiences, in short special qualification. Therefore it is far more above than the specialization of a designer. Additionally, as Oedewald and Gotcheva (2015) take attention, many safety activities are carried out by subcontractor networks which means that the subcontractors and their deliverables shall be checked for correctness, completeness and consistency by an expert of the main system owner. Hence, we propose a new role, namely the "safety responsible (SR)" to be added to this structure with the tasks provided below:

- Developing the safety plan,
- Performing safety analyses,
- Creating and maintaining the hazard log,
- Ensuring that the entries in the hazard log are successfully closed or further allocated in an appropriate manner,
- Being responsible for ensuring that safety requirements are met successfully,
- Deciding whether a deviation or change is safety relevant or not and ensuring that it is successfully closed,
- Obtaining safety evidence of the third party items including their Safety Related Application Conditions,
- Deriving the Safety Related Application Conditions for the developed item(s) in the project
- Composing the safety case

The addition of the new role "SR" arises from another necessity, as well that in terms of the independence and appropriateness of the tasks of VAL, a considerable number of tasks are usually fulfilled by the VAL which does not coincide with the primary validation tasks. For instance, there is no statement in the norms about the responsible person for the safety plan. It has been observed in the industry that the safety plan is usually written by the VAL. However, the safety plan itself needs both verification and validation. The same is also valid for the safety analyses as well as hazard log and safety case. To make this point clearer, we analyze the V&V concept in the further part of this paper.

Table 1. Failure and hazard analysis [EN 50129 (2003)].

Techniques/Measures	SIL 4
Preliminary hazard analysis	HR
Fault tree analysis	HR
Markov diagrams	HR
FMECA	HR
HAZOP	HR
Cause-consequence diagrams	HR
Event tree	R
Reliability block diagram	R
Zonal analysis	R
Interface hazard analysis	HR
Common cause failure analysis	HR
Historical event analysis	R

Beside the safety responsible as acting person for safety technical pertinent issues, another work package in a safety critical project is the safety management activities such as planning and coordination of the V&V activities, the internal and external audits and assessments, reviews, management of subcontractors concerning RAMS activities. Both the SR and VAL can perform safety management activities as these activities do not contradict with the validation activities. Below, the proposed organizational structure for all three development processes, system, HW, SW, are depicted. The SR is preferable independent from the PM considering his/her activities explained above such as deciding for the safety relevance of a change request to avoid any stress factor that can be caused by the PM due to time and/or costs pressure. Another issue is that EN 50128 does not allow REQ, DES or IMP to be INT, however, considering the role of INT in the norm, this is not contrary to the independency, and thus we propose to update the organization as in Figure 5.

Furthermore, an important role during safety critical system development is allocated to the customer/operator. However, in GAPs, there is neither customer nor operator, at least during the design phase, which increases the difficulty as the customer has in fact significant tasks in safety relevant projects as explained in Table 2. The customer/operator are incorporated once during the specific application projects on the field. Especially when identifying hazards and assessing risks, the customer is needed, as the standard IEC 61882 clarifies that both the contractor and the client should constitute the HAZOP team. To meet these needs, a person inside the company having experience in the domain and technical knowledge about

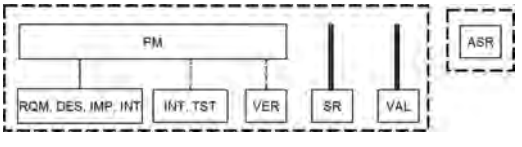


Figure 5. Proposed organizational structure.

Table 2. Responsibilities within the RAMS Process [EN 50126 (1999)], x full, (x) partial responsibility.

	Customer/ Operator	Approval Authority	(Main) Contractor	Sub-Contractor	Suppliers
Concept Phase	X				
System Definition & Application	X				
Risk Analysis	X		X		
System Requirements	X	(X)			
Apportionment of System Requirements	(X)		X		
Design and Implementation			X	(X)	
Manufacture			X	X	X
Installation			X	(X)	
System Validation	X	X	X	(X)	
System acceptance	X	X			
Operation and Maintenance	X		(X)	(X)	
Performance Monitoring	X		(X)	(X)	
Modification and Retrofit	X		X	X	
De-commissioning and Disposal	X		(X)		

the specifications can be assigned as a customer to the project. However, it should be noted that, the customer and the PM have very different responsibilities which shall be performed in an independent manner and it should not be confused that customer and PM can be the same person as there is no limitation in the organizational structure given in the aforementioned norms.

3 V&V CONCEPTS

Although verification and validation definitions should be consistent in the CENELEC Norms, these are defined somehow differently in these norms which causes ambiguity when creating plans and performing activities.

- Verification in EN 50126: confirmation by examination and provision of objective evidence that the specified requirements have been fulfilled.
- Verification in EN 50129: the activity of determination, by analysis and test, at each phase of the life-cycle, that the requirements of the phase under consideration meet the output of the previous phase and that the output of the phase under consideration fulfils its requirements.
- Verification in EN 50128: the process of examination followed by a judgment based on evidence that output items (process, documentation, software or application) of a specific development phase fulfil the requirements of that phase with respect to completeness, correctness and consistency.

As mentioned in the previous section, as TST is not defined in EN 50129, the verification activity here involves test whereas in EN 50128, the verification is based on the review. EN 50129 comments verification as a process, rather than a product. For each phase in EN 50126, there are verification steps of the phase considering the requirements of the phase. EN 50128 makes these more comprehensible and inclusionary, therefore for avoiding misapplication during the project and the harmonization of the verification definition, it is proposed to use the definition in EN 50128 in the whole project.

Coming to the validation, the situation requires the same level of attention. According to the definitions given below, again here, the test is underlined in EN 50129 while EN 50128 uses tests for checking the results. For the harmonization, we propose again that the definition in EN 50128should be used.

- Validation in EN 50126: Confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use have been fulfilled.
- Validation in EN 50129: The activity applied in order to demonstrate, by test and analysis, that the product meets in all respects its specified requirements.
- Validation in EN 50128: Process of analysis followed by a judgment based on evidence to determine whether an item (e.g. process, documentation, software or application) fits the user needs, in particular with respect to safety and quality and with emphasis on the suitability of its operation in accordance with its purpose in its intended environment.

Moreover, validation is depicted (see Figure 6) in all the CENELEC norms in a way that it is realized against system requirements. However, the

validation is to be performed not only at the end against system requirements, but as (Lundteigen et al., 2009) states, verification and validation are important activities in all phases of the project development process. For instance, at the planning phase, the tools shall be validated or if they are already pre-validated, this should be checked against the intended use or the risk analyses are to be validated. Another task can be witnessing the independence during the design and test. Therefore we propose the Figure 7 to represent the V&V activities.

Several management reports are to be created in a safety critical system. The main reports included in the safety case are verification, validation, safety management, quality assurance and assessment reports. An important issue to be tackled is the creation order of these reports, since there must be relations between them. For the order, the following procedure, also approved by an independent assessor, can be applied. After the deliverables are produced, the products of the phase and the process itself shall be verified against EN 50126. Then, the validator can create his/her report considering the deliverables and verification report. Sometimes, the validator can prefer to compile some phases as shown in Figure 8. Following this, the safety management report can be developed using check-list so that every item is closed, or if necessary, transferred to the next phase in an appropriate manner. After this, quality assurance team can verify phase related quality requirements recording the results in their reports. Having applied the illustrated order in Figure 8 during the project, the safety management report to be incorporated in the safety case can only refer to the individual phase relevant safety management reports by summarizing their results.

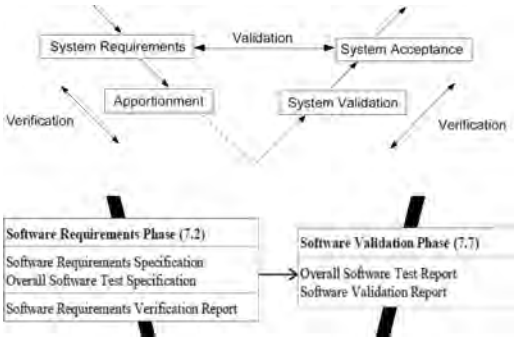


Figure 6. V&V [EN 50126 (2009)] (top), SW Validation against SW Requirements [EN 50128 (2011)] (bottom).

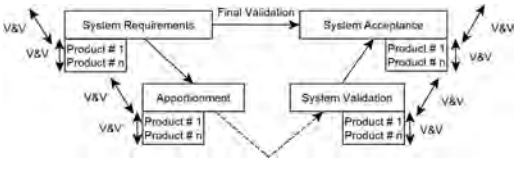


Figure 7. Proposed V&V representation.

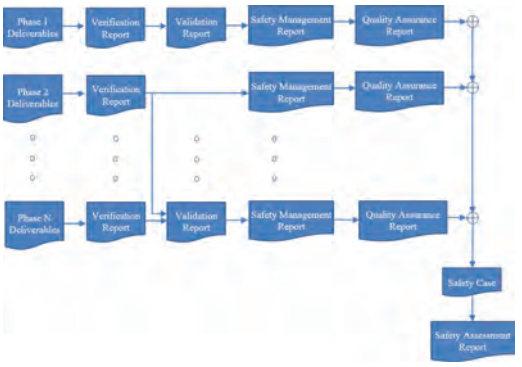


Figure 8. Relations between Verification, Validation, Safety Management, Quality Assurance and Assessment Reports.

4 CONCLUSION

Safety critical systems being developed especially for SIL 4 GAPs require much effort and organizational independences with intensive verification and validation activities. The safety standards are very detailed and evolving continuously. Although the independence and V&V concepts are very crucial and have direct effects on both technical and fiscal success of the project, it has been revealed that these points are neither clearly comprehensible in these reference standards nor discussed in the past studies, but faced once a project is developed. It is therefore very important to place the independency requirements for the roles and set V&V activities at the outset of the project correctly, completely and consistently. Defining a new role as safety responsible is very beneficial as this role fulfills the activities with special qualification and essential independency. The verification covers both process and product. For different workpackage, different verifiers can be assigned in an effective way. Validation is not just performed at the end against system requirements, but mostly at each phase during all system, HW, SW life cycles, hence VAL should play an active role during the whole life cycle with agreements or disagreements affecting the project progress widely. An appropriate relation between verification, validation, qual-

ity and safety management can be constructed in a consecutive way as proposed in this study such that the safety management report is produced step by step resulting in a ready safety management report at the end of the project that could be simply integrated into the safety case to be assessed by the independent assessor.

REFERENCES

EN 50126: 1999. Railway Applications – The specification and demonstration of Reliability, Availability and Safety (RAMS).

EN 50129: 2003. Railway Applications – Communications, signalling and processing systems – Safety related electronic systems for signalling.

EN 50128: 2011. Railway Applications – Communications, signalling and processing systems – Software for railway control and protection systems.

EN 50159: 2010. Railway applications – Communication, signalling and processing systems – Safety-related communication in transmission systems.

EU Recommendations commission 2014. Recommendation of 5 December 2014 on matters related to the placing in service and use of structural subsystems and vehicles under Directives 2008/57/EC and 2004/49/EC of the European Parliament and of the Council.

Hokstad, P., Corneliussen, K. 2004. Loss of safety assessment and the IEC 61508standard, Reliability Engineering and System Safety, v. 83, pp. 111–120.

IEC 61508: 2010. Functional safety of electrical/electronic/programmable electronic safety related systems.

IEC 61882: 2001. Hazard and operability studies (HAZOP studies), Application guide.

Lundteigen, M.A. et al., 2009. Integrating RAMS engineering and management with the safety life cycle of IEC 61508, Reliability Engineering and System Safety, v. 94, pp. 1894–1903.

Murthy, D., Rausand, M., Østeras, T. 2008. Product reliability: specification and performance. Springer, London.

Oedewald, P., Gotcheva, N. 2015. Safety culture and subcontractor network governance in a complex safety critical project, Reliability Engineering and System Safety, v.141, pp. 106–114.

Big data risk analysis

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
 ISBN 978-0-8153-8682-7

Manifestation of ontologies in graph databases for big data risk analysis

M. Figueres-Esteban, P. Hughes, R.A.H. El Rashidy & C. van Gulijk
Institute of Railway Research, University of Huddersfield, Huddersfield, UK

ABSTRACT: Big Data Risk Analysis (BDRA) intends to combine the huge volume of information that railway systems produce from a variety of data sources for safety and risk management. One of the most challenging issues is how safety scientists can use big data techniques. This is especially important in the light of data coming from different systems that hold information about critical events, hazards or controls. Yet, the integration of complex safety-related data is not just another IT problem. Fundamentally, it requires the expertise of safety experts in order to make sense of the data. A solution lies in the use of graph databases and ontologies to access this data for safety purposes. This type of database allows for the handling of huge amounts of data whilst it is still accessible to safety experts that are not gifted programmers. This approach opens up big data for safety management and enables a plethora of possibilities for future safety research.

1 INTRODUCTION

A BDRA safety system is defined as an enterprise safety management system that performs the following:

- Extracts information from mixed data sources.
- Processes it quickly to infer and present relevant safety management information.
- Combines applications to collectively provide sensible interpretation.
- Uses online interfaces to connect the right people at the right time.

In order to:

- Provide decision support for safety and risk management.

This definition guides the development of BDRA systems that are of use to companies that work on the GB railways. BDRA aims to use big data analytics techniques for safety (Van Gulijk et al. 2018; Van Gulijk et al. 2017).

One of the key challenges of BDRA is to store and process that massive amount of data and manage the heterogeneous knowledge from different information systems to obtain safety insight. A solution lies in the use of graph databases that are controlled by ontologies to represent a common framework of understanding and integrate data. (Figueres-Esteban et al. 2016; Van Gulijk et al. 2016). The method explained in this paper opens up big data for safety scientists. The method is straightforward but powerful and does not rely on gifted programmers. In theory, the database is infinitely scalable so it is hard to predict the limitations of the approach.

2 DATABASES AND BIG DATA

In the last decades, relational databases (aka SQL databases) have dominate the market of databases until coming up a standard way. They are structured in tables for access and have been very efficient when it comes to rapid and efficient access to data. Nevertheless, in big data environments where huge amounts of information have to be stored and integrated from new unknown sources, relational databases become unwieldy (Sadalage & Fowler 2013).

A solution to bypass this problem is to omit the relational table by simply storing data in a system that, for lack of a better example, finds its analogy in an infinitely scalable library card catalogue (Van Gulijk et al. 2018). Databases that work in that way are called NoSQL databases.

2.1 *Graph databases*

In a relatively novel development these NoSQL databases have been enriched with a sensible visual interface based on graphs. They are simply called graph databases. A graph database is database management system that store data in the form of a property graph (Robinson et al. 2013). Safety scientists will recognise a property graph as a collection of nodes and links; as we often see them in our work.

The organization of the data in graphs is extremely useful in terms of understanding (Figueres-Esteban, Van Gulijk, et al. 2015; Figueres-Esteban, Hughes, et al. 2015). Graph databases allow to represent different types of data models into a common space in order to integrate diverse type of data (EL Rashidy et al. 2017). This issue is a key aspect in order to implement ontologies that represent the knowledge of technical domains such as railways, risk and safety.

3 KNOWLEDGE REPRESENTATION IN RAILWAYS

Railways are a complex systems that represent a rich tapestry of different types of organisational knowledge, created for different purposes and people with different expertise, skills and competences in many different contexts. Bringing together all the data that railways produce means to make sense of heterogeneous knowledge from different information systems.

The most common technique used by computer scientists to represent a common framework of understanding and manage the knowledge is an ontology. A formal, and broadly accepted, definition of an ontology is provided by (Gruber 1995): "*An ontology is an explicit specification of a conceptualization*."

There are different types of ontologies such domain and application ontologies depending on their specificity of the knowledge (Guarino 1997).

In the railway domain, the FP6 European Integrail project (http://www.integrail.eu/) and the RailML community (http://www.railml.org) proved the utility of ontologies in the communication and integration of data through railway information systems (Van Gulijk & Figueres-Esteban 2016).

4 ONTOLOGIES AND GRAPH DATABASES FOR BDRA

Different ontology languages and frameworks have been developed to support the implementation of an ontology (Corcho et al. 2003). The challenge is that a single ontology should be understood by people and machines.

One of the most used frameworks is showed in the left side of Figure 1. Different data structures represented in formats such as XML, JSON and CSV can be integrated through ontologies implemented in RDF/OWL languages. These languages support the application of Artificial Intelligence (AI) in order to reason with the represented knowledge. The approach that this work is taking

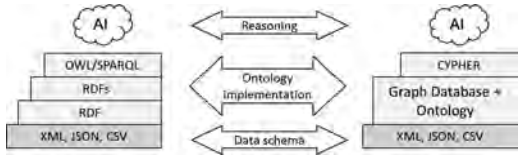


Figure 1. Transformation of the stack of ontology languages for BDRA.

bypasses complicated ontology languages and replaces it with a relatively straightforward visual interface in a graph database. This is where we omit the need for gifted programmers.

This paper demonstrates how to use the framework showed in the right side of Figure 1.

5 METHODOLOGY

The paper describes the implementation of a railway domain ontology by safety experts in order to connect three different data sources to an event related to safety management.

The current BDRA project focuses on understanding SPAD risks (passing red signals) but for the benefit of explaining the method we focus on part of that risk: the "signal obscured" hazard. This means that safety records related to obscured signals and instances of a signal database have to be found and linked to enrich the analysis of these type of events. The methodology has three basic steps:

a. Selection of data sources and storing data in a graph database.
b. Building the signal domain ontology.
c. Implementing the signal ontology for the integration of data.

5.1 *Data sources*

This trial uses four *.csv files extracted from three information systems: three files of text records from the SMIS and IFCS systems of Railway Safety and Standard Board (RSSB) containing around 100,000 incidents and a table of signals from the Ellipse Asset Management tool of Network Rail (NR) containing 40,000 descriptions of signals.

SMIS is a database for recording safety-related events that occur on the rail network in Britain (RSSB 2017). Railway stakeholders such as NR or train/freight operators enter about 75,000 events per year such as derailments and SPADs. In this exercise, we are just using records related to obscured signals. IFCS is a database that focuses on human performance and underlying causes of

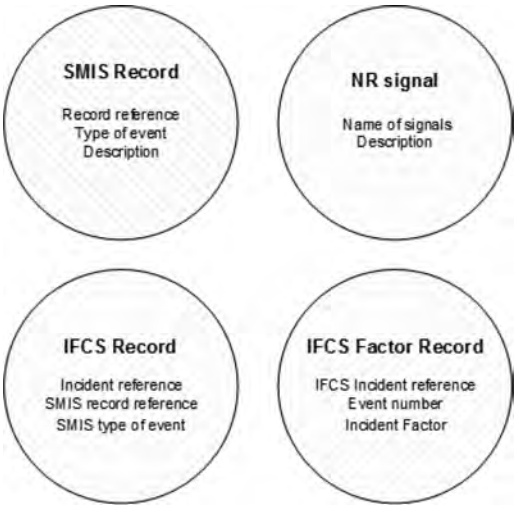


Figure 2. Description of the data sources used to support data integration.

rail incidents. These underlying caused are classified using 10 Incident Factors that are breakdown by different levels of sub-categories (Gibson et al. 2015). The table of signals is a sample of descriptions of signals that is part of the Ellipse Asset Management tool of NR. Figure 2 shows the properties that were used to integrate the data.

5.2 *Signal domain ontology*

The purpose of the signal domain ontology is to align data structures of the information systems with an accepted reference framework by railways. For this exercise, the reference framework for the signal domain has been the railway signal standard in UK (RSSB 2015).

The sources showed below have been used to build the ontology:

- The Signals, handsignals, indicators and signs. Handbook RS/521 Issue 3 (December 2015).
- The data model of the SMIS+ program.
- The schema of the table of signals.

The standard RS/521 provides a classification and description of all railway signals in UK. The data model of the SMIS+ includes a taxonomy of railway signals that is aligned with other reporting systems. The data structure of the table of signals does not provide a signal taxonomy but it can be extracted from the field "Item Name". An example of value of this field is "EZ220 – SIG HEAD – COLOUR LIGHT – LED". Note that programmers don't have the expertise for this exercise, even if they are gifted. The interpretation and consideration of safety-aspects lies within the remit of safety experts.

5.3 *Implementation of the ontology and data integration*

The data from the information systems were stored in a Neo4 j graph database. Each row of the data files represents a node in the graph database and each node has as many properties as columns the data file has. In this first step, the database has no structure and it just stores data under a label (data nodes).

In the same database, the signal domain ontology was implemented in a graph data model (ontology nodes). Using the properties of the ontology nodes and analysing the property of the data nodes that stores the text of the record, the links between each node were created.

The signal obscure event (event node) was connected to data nodes of signals and these ones were connected to the data nodes of SMIS/IFCS records.

6 RESULTS

Table 1 shows an excerpt of the extracted ontology from the table of signals. This ontology was mapped with the explicit ontologies of the standard RS/521 and the SMIS data model. Table 2 shows an excerpt of the mapping table. Figure 3 shows a piece of the final signal ontology.

Figure 4 shows an excerpt of the graph database that contains part of the implementation of the signal ontology and instances of signal and SMIS/IFCS records. The ontology is connected to the

Table 1. Excerpt of the signal taxonomy from the table of signals.

Item name			
First token	Second token	Third token	Fourth token
EZ220	SIG HEAD	COLOUR	LED
EZ101		LIGHT	1 ASPECT
EZ102			2 ASPECT
EZ103			3 ASPECT
EZ104			4 ASPECT

Table 2. Excerpt of the mapping between the signal taxonomies of the RS/521 standard, SMIS and the table of signals.

RS/521	SMIS	Table of signals
SPAD indicator	SPAD indicator	ES100
Limit of shunt signal	Limit of shunt	EZ160
Point indicator	Points indicator	BR100, EZ170

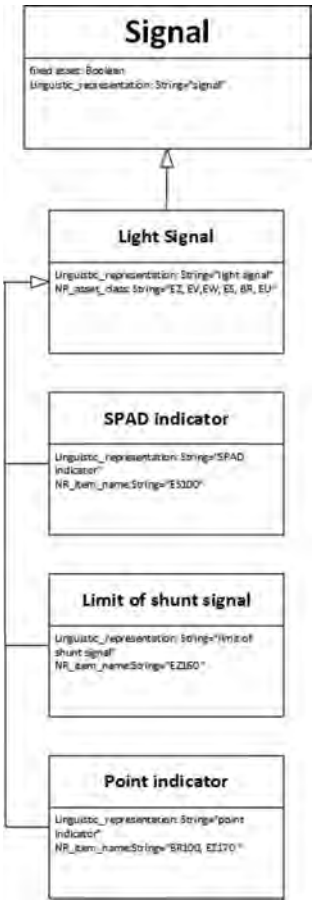


Figure 3. Excerpt of the UML diagram that represents the signal ontology.

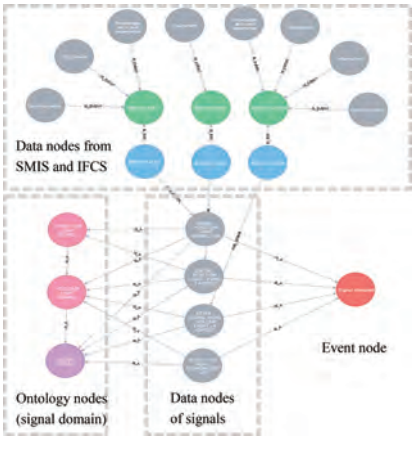


Figure 4. Excerpt of the graph database that integrates different types of instances of data with the event "Signal obscured" by means of the signal domain ontology.

signal nodes that are connected to the SMIS/IFCS records and the signal obscure event.

7 DISCUSSION

This paper shows that NoSQL graph databases guided by ontologies enable safety scientists to work with big data techniques without the intervention of IT experts. Some programming is required but most of it is not much more complicated than excel macros or Matlab. The expertise of safety experts, however, is fundamentally required to build safety and railway domain ontologies to support the integration of data for further safety analysis.

This work demonstrates that graph databases can store complex data structures as single nodes, which helps safety scientists navigate through their data. Figure 4 displays different nodes that represent data from signals and SMIS/IFCS records regardless of the internal structure of the data source.

Domain ontologies can be straightforwardly implemented in the database as a data model to integrate data. These ontologies support the analysis of data nodes that use different semantics about a railway domain. This semantic alignment allows to interconnect data nodes each other or connect them to specific events related to safety management. However, ontologies are far from being populated automatically and require safety expertise and human effort to build them (Figueres-Esteban & Van Gulijk 2016). Table 1 and Table 2 shows the results of this effort in order to align three different data sources with a railway standard in a single ontology that represent the signal domain (Figure 3).

The alignment of different types of data with a domain ontology and events related to safety has important benefits. Firstly, the ontology provides framework in order to query signals. In this case, the standard RS/521 was selected as reference framework. Secondly, the integration of data allows to connect all the information available in the data sources. For example, linking data nodes of signals to the signal obscure event and SMIS/IFCS records allows to filter records by specific types of signals in order to improve the safety understanding related to obscured signals.

8 CONCLUSIONS

This paper demonstrates how safety scientists can enter the realm of big data. It demonstrates that the challenge of storing large amounts of data from diverse railway data sources to extract safety learning requires safety experts that can work with graph databases.

Graph databases allow to store data regardless of the structure of the data. But more fundamentally, it allows the co-location of domain ontologies to integrate different data sources and extract safety learning.

In theory, the database is infinitely scalable so it is hard to predict the limitations of the approach.

REFERENCES

Corcho, O., Fernandez-Lopez, M. & Gomez-Perez, A., 2003. Methodologies, tools and languages for building ontologies. Where is their meeting point? *Data & Knowledge Engineering*, 46, pp. 41–64.

EL Rashidy, R. et al., 2017. A Big Data modeling approach with graph databases for SPAD risk. *Safety Science*.

Figueres-Esteban, M. & Van Gulijk, C., 2016. *Ontology learning for BDRA. Report 110–124 II*, Huddersfield.

Figueres-Esteban, M., Van Gulijk, C. & Hughes, P., 2015. *Visualisation and Risk Communication in Railway Big Data Risk Analysis (BDRA): Literature Review. Report 110–113*, Huddersfield.

Figueres-Esteban, M., Hughes, P. & Van Gulijk, C., 2016. Ontology network analysis for safety learning in the railway domain. In L. Walls, M. Revie, & T. Bedford, eds. *Risk, Reliability and Safety: Innovating Theory and Practice*. London: Taylor & Francis Group, pp. 2937–2942.

Figueres-Esteban, M., Hughes, P. & Van Gulijk, C., 2015. The role of data visualization in railway Big Data Risk Analysis. In *Safety and Reliability of Complex Engineered Systems – Proceedings of the 25th European Safety and Reliability Conference, ESREL 2015.*

Gibson, W.H. et al., 2015. The incident factor classification system and signals passed at danger. In *Fifth International Rail Human Factors Conference*. RSSB, pp. 22–31.

Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43, pp. 907–928. Available at: http://dx.doi.org/10.1006/ijhc.1995.1081.

Guarino, N., 1997. Understanding, building and using ontologies. *International Journal of Human Computer Studies*, 46, pp. 293–310.

Robinson, I., Webber, J. & Eifrem, E., 2013. *Graph Databases*, California, United States of America: O'Reilly Media, Inc.

RSSB, 2017. Safety Management Intelligence System (SMIS). Available at: https://www.rssb.co.uk/risk-analysis-and-safety-reporting/reporting-systems/smis.

RSSB, 2015. Signals, handsignals, indicators and signs. Handbook. RS/521 Issue 3, p. 60.

Sadalage, P.J. & Fowler, M., 2013. *NoSQL Distilled: A brief guide to the emerging word of plyglot persistence*, New Jersey: Pearson Education, Inc.

Van Gulijk, C. et al., 2018. Introduction to IT transformation of safety and risk management systems. In *Handbook of RAMS in railway systems: Theory and Practice*. CRC.

Van Gulijk, C. et al., 2017. The case for IT transformation and Big Data for Safety Risk Management on the GB railways. *Journal of Risk and Reliability*.

Van Gulijk, C. & Figueres-Esteban, M., 2016. *Background of Ontology for BDRA. Report 110–124 I*, Huddersfield.

Van Gulijk, C., Hughes, P. & Figueres-Esteban, M., 2016. The potential of ontologies for safety and risk analysis. In L. Walls, M. Revie, & T. Bedford, eds. *Risk, Reliability and Safety: Innovating Theory and Practice*. London: Taylor & Francis Group, pp. 1315–1322.

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Author index